BMC Cancer

RESEARCH Open Access

Application of deep learning models in gastric cancer pathology image analysis: a systematic scoping review

Sijun Xia[1], Yuanze Xia[2], Ting Liu[2], Yiming Luo[2] and Patrick Cheong-Iao Pang[2*]

Abstract

Background Accurate diagnosis and prognosis stratification of gastric cancer (GC) are crucial for effective treatment. However, traditional histopathological image analysis relies on the subjective judgment of pathologists, which is time-consuming and prone to errors. The emergence of deep learning (DL) models provides new ways to automate and improve the analysis of GC pathology images. This systematic review aims to evaluate the current application, challenges, and future directions of DL in GC pathology image analysis.

Methods The study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Extension for Scoping Reviews (PRISMA-ScR) guidelines and searched four databases: PubMed, Scopus, Web of Science, and IEEE Xplore (as of June 19, 2025).

Results The initial search identified 520 articles, and 22 studies that met the inclusion criteria were finally included. The results show that DL models have performed excellently in GC detection, histological classification, and prognosis prediction. Some models even reached an accuracy of over 95% in GC detection. Convolutional neural networks (CNN) are the most commonly used DL models. However, current studies still have limitations, such as limited dataset size, lack of external validation, and insufficient data diversity. The applicability to different types and stages of GC is also unclear.

Conclusions Future research must build larger, more diverse, and more representative datasets. These should cover a wider range of GC types and stages, and undergo rigorous clinical validation. This will help fully realize the potential of DL in GC pathology image analysis and ultimately improve clinical practice.

Keywords Deep learning, Gastric cancer, Pathology image, Systematic review

*Correspondence:
Patrick Cheong-Iao Pang
mail@patrickpang.net
[1]Department of Pathology, Sheyang County People's Hospital, Yancheng 224300, Jiangsu, China
[2]Faculty of Applied Sciences, Macao Polytechnic University, Macao 999078, China

Background

DL is increasingly applied in medical diagnosis and treatment, covering areas such as disease screening, diagnosis, prognosis prediction, and treatment response evaluation [53, 60, 67, 71]. In particular, DL models have reached a high level in certain specific tasks in the field of medical image analysis [7, 36]. For example, DL has been successfully applied in tasks such as lung nodule detection [14], diabetic retinopathy screening [63], and skin cancer diagnosis [15]. The widespread use of DL is mainly due to its powerful feature learning ability. It can automatically extract complex and multi-level feature representations from large amounts of data without relying on manual intervention or feature engineering [31]. In contrast, traditional machine learning (ML) models, such as support vector machines and random forests, rely on manually designed features. Their performance is limited by the methods used for feature selection and extraction [2, 52].

GC has a high incidence and mortality rate worldwide. Early diagnosis and precise treatment are very important for improving patient prognosis [48, 57, 58]. Traditional diagnosis of GC mainly relies on the experience and subjective judgment of pathologists, which has limitations such as low efficiency and poor reproducibility [13, 16]. DL, as a powerful image analysis technology, has shown great potential in the field of medical imaging [25, 55]. It provides a new way to improve the accuracy and efficiency of GC diagnosis.

In cancer image diagnosis, DL models, especially CNN, have become the mainstream method [12, 74]. CNN can effectively capture local features and global context information in images through convolution and pooling operations [59]. Researchers have developed various CNN-based models for the diagnosis of different types of cancer, such as breast cancer [41], lung cancer [10], and prostate cancer [50]. These models can identify and classify cancer cells, assess tumor grade and stage, predict prognosis, and identify potential biomarkers [43]. In GC diagnosis, DL models can learn and identify various key histopathological features, such as nuclear pleomorphism, nuclear-to-cytoplasm ratio, degree of cell arrangement disorder, and stromal response [17]. This helps pathologists make more accurate diagnoses and prognosis assessments.

In recent years, there have been several review articles exploring the applications of DL in different medical fields. For example, Nawab et al. [46] summarized artificial intelligence (AI) applications across gastrointestinal diseases by organizing data from relevant studies. Renjith et al. [51] focused on explainable AI techniques in gastrointestinal cancer image analysis, evaluating their strengths and limitations. Zhuang et al. [72] reviewed the applications of DL in gastrointestinal endoscopy, imaging, and pathology. They discussed model types, challenges, and limitations. Li et al. [34] conducted a meta-analysis on AI's diagnostic accuracy for gastric intestinal metaplasia using statistical approaches. Park et al. [49] systematically reviewed AI-based prediction of microsatellite instability (MSI) and defective DNA mismatch repair deficiency (dMMR), assessing model performance and research methods. Turtoi et al. [62] provided an overview of AI in gastritis diagnosis, emphasizing its clinical value. Kuntz et al. [29] reviewed CNN models analyzing histological images of gastrointestinal cancers and precursor lesions, including molecular features and prognosis. Li et al. [35] summarized ML, DL, and transfer learning applications in medical image analysis. They focused on introducing the advancements in classification, segmentation, and detection. Alam et al. [1] studied the prediction of MSI for different cancers using AI models. They analyzed the trends, performance, and research challenges. However, there are some limitations in the existing review articles. First, most reviews cover a broad range of topics and lack a specialized analysis of GC. They do not deeply explore the effectiveness of DL in GC diagnosis [1, 35, 49, 72]. Second, the literature included in these reviews is often not up-to-date enough to capture new developments and trends [29, 51]. Finally, some reviews lack in-depth discussions on the challenges faced by DL models in clinical applications [34, 46, 49, 62].

The aim of this study is to conduct a systematic review of relevant literature from 2020 to 2025 and provide a comprehensive summary of the current research on the application of DL in GC pathological image analysis. Although the applications of DL in histopathology are rapidly expanding to encompass areas beyond initial diagnosis, such as prognostication, treatment response prediction, and biomarker discovery [45, 64, 66], this systematic scoping review deliberately focuses on the application of DL for fundamental diagnostic tasks in GC pathology. This targeted scope was chosen to provide a comprehensive foundational overview of how DL is being utilized to assist in the primary detection and classification of GC from pathological images. This remains a critical and evolving area in GC pathology and serves as a cornerstone for subsequent, more complex applications (e.g., prognostic prediction). The study believes that before delving into more advanced applications, it is crucial to first systematically map and understand the current landscape, methodologies, and challenges of DL in these core diagnostic processes. Therefore, this review aims to systematically delineate the existing evidence, methodologies, and challenges specifically related to these diagnostic applications, thereby providing a solid foundation for future research. The study will perform a Scoping Review and attempt to answer the following research questions:

(1) What are the publication trends of DL research in GC diagnosis?

(2) What research designs have been applied to DL in GC diagnosis?

(3) What DL models are used in GC image diagnosis, and what are their performance results?

(4) What are the characteristics of the images used for GC analysis with DL?

(5) What challenges and impacts are encountered when using DL models to analyze GC images?

By answering the above questions, this study aims to provide a clearer picture of the application of DL in GC pathological image analysis and offer a reference for future research.

Methods

The study adopted the Scoping Review method. This aim is to provide a comprehensive overview of current research on using DL models to analyze GC pathology images. It also seeks to identify the similarities and differences between studies in diagnostic performance, application scenarios, and technical methods. In addition, this study strictly followed the guidelines specified in the PRISMA-ScR guidelines [42] to ensure the rigor and reproducibility of the study.

Search strategy

To ensure a comprehensive literature search, this systematic review used the four databases: PubMed, Web of Science, IEEE Xplore, and Scopus. The search was conducted up to June 19, 2025. The core search terms used were "Gastric Cancer" AND "Deep Learning" AND "Pathology" AND"Transformer"AND"Foundation Model," adapted to the specific syntax and functionality of each database to maximize retrieval comprehensiveness and accuracy. The specific search formula is shown in Table 1.

Table 1 Selected databases and search formats

Database	Search Formula
Web of Science	Deep learning (All Fields) and Gastric cancer (All Fields) and Pathology (All Fields) and Transformer (All Fields) and Foundation Model (All Fields)
Scopus	deep AND learning AND gastric AND cancer AND pathology AND transformer AND foundation AND model
PubMed	((((Deep Learning) AND (Gastric Cancer)) AND (Pathology)) AND (Transformer)) AND (Foundation Model)
IEEE Xplore	("All Metadata": Deep learning) AND ("All Metadata": Gastric cancer) AND ("All Metadata": Pathology) AND ("All Metadata": Transformer) AND ("All Metadata": Foundation Model)

Inclusion and exclusion criteria

This systematic review included studies published up to June 19, 2025, that used DL models to analyze pathological images for the prediction or diagnosis of GC and reported relevant outcome measures, such as diagnostic accuracy, sensitivity, and specificity. Inclusion criteria include: (1) analyzing pathological images using DL models to predict or diagnose GC; (2) Relevant outcome measures reported, such as diagnostic accuracy. Exclusion criteria include: (1) studies not related to GC or studies whose purpose was not pathology image-based diagnosis or prognosis of GC; (2) studies not using DL models or pathological images; (3) reviews, commentaries, and conference abstracts; (4) studies lacking essential data, such as whether pathological images or DL models were used. The literature screening process was managed using Covidence systematic review platform. Two reviewers independently screened titles and abstracts, and subsequently, full-text articles of potentially eligible studies. During both the title/abstract screening and full-text review stages, reviewers were blinded to each other's decisions. Any disagreements between the two reviewers were resolved through discussion or, if necessary, by consulting a third reviewer. The screening process, detailing the number of studies at each stage, will be presented using a PRISMA flow diagram.

Data extraction

A standardized form was used to extract data from studies meeting the inclusion criteria, including article title, first author, publication year, study country or region, study design, study subjects, GC pathology, use of pathology images, DL models, sample size, diagnostic accuracy, and study results and conclusions. In addition to the previously listed data items, the study specifically extracted information regarding the use of Multiple Instance Learning (MIL) frameworks or other pooling or aggregation strategies (e.g., max pooling, average pooling, fully supervised approaches) for studies employing whole-slide image (WSI) predictors. After data extraction, a second review will be conducted. If any data are missing or unclear, the corresponding authors will be contacted. Studies excluded due to unobtainable data will be noted in the results section.

Collating, summarizing, and reporting the results

The study follows the PRISMA-ScR guidelines and employs a narrative synthesis approach to organize, summarize, and report the findings of the included studies. The focus will be on study characteristics, DL models and algorithms, application scenarios, outcome measures, study quality, and future research directions. We will present this information using tables, figures, and

narrative text, and discuss the meaning of the results and their implications for clinical practice and future research.

Result

The search period for this study was until June 19, 2025, and a total of 520 documents were retrieved. 97 duplicate papers were automatically eliminated through EndNote, 384 irrelevant papers were eliminated by reading titles and abstracts, 8 irrelevant papers were eliminated by reading the full text, and 9 papers were eliminated by selecting articles from the last five years. Finally, a total of 22 papers were included in the study, as shown in Table 2. The flowchart of the study selection is shown in Fig. 1.

Study characteristics

Publication year

Figure 2 shows the distribution of publication years for the included studies, displaying a fluctuating trend with a decline followed by an increase. The results indicate that three articles were published in 2020, the number decreased to two in 2021, rose to five in 2022, dropped again to three in 2023, and significantly increased to nine in 2024. This fluctuation trend may be influenced by multiple factors. The decrease in publications in 2021 may be related to the global outbreak of COVID-19, which disrupted research activities. The increase in 2022 reflects the growing attention to DL in the field of medical image analysis, as more researchers began applying DL models to GC pathology image analysis. The brief decline in 2023 may just be a random fluctuation, while the significant increase in 2024 could indicate that research in this field is entering a period of rapid development. This could be due to the continuous progress in DL models, such as the emergence of more efficient model architectures and training algorithms, as well as the release of larger and higher-quality GC pathology image datasets. Additionally, the growth in clinical demand and increased research funding may have also contributed to the development of this field.

Publication country/region

Figure 3 shows the distribution of the regions where the first author of the articles included in this study is based, covering Asia, North America, the Middle East, and Europe. The results indicate that China is the most active in this field, with nine articles published, followed by South Korea and Malaysia, with three articles each. China's high output in the field of DL applied to GC pathology image analysis may be related to its relatively high GC incidence rate [58] and significant investment in AI research [53]. South Korea and Malaysia also show high research activity, which may reflect their focus on the field of medical AI and their growing research capabilities. The relatively low publication output in this specific field from the United States, a leading country in DL research, may be related to its lower GC incidence rate and different research priorities. This geographic imbalance highlights the importance of promoting collaboration and resource sharing globally. It can drive the application and development of DL in GC pathology image analysis, ultimately benefiting GC patients in all regions.

Study designs

As shown in Table 2, all included studies focused on model development and validation, encompassing the development, training, validation, and performance evaluation of DL models. While all studies aimed to develop and validate DL models for GC pathology image analysis, their specific methodologies varied. These variations included different model development strategies: 15 studies employed single models with single-task learning, four used ensemble learning, two used hybrid models, and two employed transfer learning. One study incorporated both ensemble and transfer learning, as shown in Fig. 4. The research objectives also differed, with 11 studies focusing on GC detection, ten on classification and segmentation, and one on metastasis prediction, as shown in Fig. 5.

Study subjects

Regarding data sources, the included studies used various datasets, including the public dataset GasHisSDB, the TCGA database, the BOT gastric biopsy database, and self-collected clinical data, as detailed in Table 3. GasHisSDB is a database focused on gastric pathology images. It is commonly used in GC research because of its high image quality. TCGA contains a wide range of cancer genomic data and large sample sizes. It is suitable for cancer research, but is usually chosen based on specific cancer types due to its complexity. The BOT gastric biopsy database specializes in gastric pathology data. It is mainly used in gastric disease research, but has a narrower scope and is used less frequently. The self-collected clinical data comes from the hospital where this study was conducted. It has strong regional and personalized characteristics. Although the sample size is small, it provides actual clinical conditions. This variation in data sources may introduce biases and affect the generalizability of the developed models.

Pathological image features used

All included studies used WSI, which provides rich histological information. H&E-stained WSI were used in 21 studies, making it the most commonly used type. This is because H&E staining is widely used in pathology. It is cost-effective and provides important structural

Table 2 Overview of study characteristics

Author/ Year/ Country or Region	Research design	Study Subjects	Gastric cancer pathology	Use of pathology images	Diagnostic accuracy	Sample size	Deep learning model	Results or conclusion	Use MIL
Govind et al. [44] 2024 USA	Development and evaluation of ensemble deep learning models for gastric cancer detection	Gastric cancer pathology images from the GasHisSDB dataset	Yes	WSI images and patch images after H&E staining	80 × 80 px: ResNet34: 0.93, VGGNet: 0.94, ensemble model: 0.99; 120 × 120 px: ensemble model: 0.99, VGGNet: 0.97, ResNet50: 0.97; 160 × 160 px, ensemble model: 0.99, VGGNet: 0.98, ResNet50: 0.98, EfficientNet: 0.92.	245,196 cropped gastric cancer pathology images derived from 600 original images	Ensemble model based on ResNet34, VGGNet, and other CNN architectures	Ensemble model shows high diagnostic accuracy, outperforming individual models and supporting early gastric cancer detection.	No, max pooling is used
Muhammad et al. [73] 2024 Pakistan	Design and validation of a new framework for gastric histology classification and segmentation (GHCS)	Gastric cancer pathology images from the GasHisSDB dataset	Yes	WSI images and patch images after H&E staining	Classification: GasHisSDB dataset: Validation set: 0.9887, Test set: 0.9847 HCRF dataset: Validation set: 0.9728, Test set: 0.9731 Segmentation: IFCM: 0.8468	GasHisSDB: 245,196 images. HCRF: 700 Images.	Naïve BaYes with Gaussian Mixture Model (GMM) and Expectation Maximization (EM). Improved Fuzzy C-Means (IFCM) for segmentation.	Proposed framework shows improved accuracy in classification and segmentation of GC compared to some existing techniques. Grad-CAM visualizations enhance interpretability.	No, the research uses the technical route of strong supervised feature engineering + statistical learning model + clustering segmentation
Jonghyun et al. [32] 2024 South Korea	To develop and validate a deep learning-based algorithm to identify lymphovascular invasion (LVI) lesions associated with prognosis in gastric cancer	Gastric adenocarcinoma	Yes	WSI images and patch images after H&E staining	Classification: ResNet50:0.9319,EfficientNetB3:0.9281,ConViT:0.9348 Detection: YOLOv3:0.927,YOLOX:0.9353 Ensemble: 0.9514	100 WSIs (internal) + 48 WSIs (external)	Classification: ConViT, ResNet50, EfficientNet B3. Detection: YOLOX, YOLOv3. Ensemble: Weighted average of classification and detection model outputs.	The ensemble deep learning model effectively predicted LVI in gastric cancer WSIs, achieving high accuracy and potentially reducing interobserver variability.	No, the research uses the technical route of strong supervision patch classification + target detection + model integration
Huang et al. [22] 2024 Malaysia	A dilated CNN with late fusion strategy (DCNNLFS) was developed and validated for gastric histopathology image classification.	Gastric cancer pathology images	Yes	WSI images and patch images after H&E staining	DCNNLFS: 0.935	700 gastric histopathology images (560 cancer, 140 normal)	Dilated Convolutional Neural Network with Late Fusion (DCNNLFS)	The DCNNLFS model outperforms traditional CNN, AlexNet, VGG, and GoogLeNet models in classifying gastric histopathology images. Dilated convolutions and late fusion strategy improve feature extraction and classification accuracy. Future work will explore multi-modal data and neuroheuristic analysis.	No, the research adopts the technical route of dilated convolution + maximum pooling + late feature fusion

Table 2 (continued)

Author/ Year/ Country or Region	Research design	Study Subjects	Gastric cancer pathology	Use of pathology images	Diagnostic accuracy	Sample size	Deep learning model	Results or conclusion	Use MIL
Danial et al. [26] 2024 Iran	Development and evaluation of a hybrid deep learning and gradient boosting model for gastric cancer detection.	Gastric cancer pathology images from the GasHisSDB dataset	Yes	WSI images and patch images after H&E staining	EfficientNetV2B0-CatBoost combination: 80px: 0.897, 120px: 0.931, 160px: 0.939	245,196 images (GasHisSDB dataset); split 80/20 for training/testing	EfficientNetV2B0 for feature extraction; CatBoost for classification.	Proposes a hybrid approach achieving high accuracy in classifying gastric histopathology images. Grad-CAM visualizations confirm focused learning on relevant features. t-SNE shows clear clustering of normal and abnormal cases.	No, the research adopts the technical route of pre-trained model + average pooling + gradient boosting classification
YONG et al. [70] 2024 Malaysia	Development and evaluation of deep learning models based on intra-domain transfer learning and ensemble learning.	Histopathology image patches and WSIs from multiple datasets. GasHisSDB gastric dataset Chaoyang colorectal dataset CPTAC-CCRCC kidney dataset.	Yes	WSI images and patch images after H&E staining	GasHisSDB: 0.9978 Chaoyang: 0.8569 CPTAC-CCRCC: 0.9917	GasHisSDB: 245,196 patches from 600 WSIs. Chaoyang: 6,160 patches. CPTAC-CCRCC: 783 WSIs.	Ensemble learning with pre-trained CNN base models (MobileNet, MobileNetV2, EfficientNetB0/B1, DenseNet121/169, InceptionV3, Xception, ResNet50/101). Intra-domain transfer learning used to enhance feature extraction.	Models achieved state-of-the-art accuracy on multiple histopathology datasets. Intra-domain transfer learning and ensemble learning enhance feature extraction and improve performance on low-resolution images.	No, the research adopts the technical route of pre-trained model + mixed pooling + in-domain transfer learning + model integration
Andrea et al. [38] 2024 Italy	Comparative analysis of machine learning and deep learning models for gastric cancer image classification	Gastric cancer pathology images from the GasHisSDB dataset	Yes	WSI images and patch images after H&E staining	Manual feature support vector machine (SVM): CH_1: 0.7592, CH_2: 0.725 Random forest using local binary pattern features: 0.7957. Deep features: SVM using DenseNet201: 0.9193 Feature fusion experiment: fusion of LBP with DenseNet201 and EfficientNetB0 features, SVM: 0.9503 Cross-amplification experiment: on 120*120 test set, RF combined with some features: 0.8904, on 80*80 test set: 0.7889	245,196 sub-sized gastric histopathology images (97,076 abnormal, 148,120 normal) from 600 WSIs. Source: GasHisSDB	Various machine learning classifiers (DT, kNN, SVM, RF) were used with both handcrafted (invariant moments, texture, color features) and deep features (from pre-trained CNNs like AlexNet, DenseNet, etc.). Feature fusion strategies were also explored	Feature fusion, particularly combining handcrafted and deep features, improves classification accuracy. SVM performed best among the classifiers. The study highlights the potential of using pre-trained CNN features without fine-tuning for gastric cancer image classification. Cross-magnification experiments show promising results for generalizability across different resolutions	No, the research adopts the technical route of pre-trained model pooling + manual features + traditional classifier

Table 2 (continued)

Author/ Year/ Country or Region	Research design	Study Subjects	Gastric cancer pathology	Use of pathology images	Diagnostic accuracy	Sample size	Deep learning model	Results or conclusion	Use MIL
Armand et al. [3] 2024 South Korea	Development and evaluation of a vision transformer model for GC detection on weakly annotated WSI	Gastric cancer pathology images from the GasHisSDB dataset	Yes	WSI images and patch images after H&E staining	ViT-16: 0.859, ResNet50: 0.7059, DenseNet121: 0.7451, EfficientNetB1: 0.7871, ViT: 0.77	560 cancerous and 140 normal histopathological images, each 2048 × 2048 pixels (GasHisSDB dataset). 40 images used for testing	Vision Transformer (ViT) base model with a patch size of 16 × 16	The ViT model achieved competitive performance in classifying gastric cancer WSIs using only weak, region-level annotations. The study suggests that ViTs can be effective for gastric cancer detection and could support pathologists in decision-making	No, the research adopts the technical route of weak supervision ViT + global feature classification
Sung Hak Lee et al. [33] 2023 South Korea	Development and validation of deep learning for fully automated classification of microsatellite instability (MSI) status in gastric cancer tissue sections	75 MSI-H and 256 MSS patients from TCGA; 383 patients from SSMH	Yes	WSI images and patch images after H&E staining	The Cancer Genome Atlas(TCGA) frozen by TCGA frozen: 0.863, TCGA FFPE by TCGA FFPE:0.862 TCGA FFPE by both FFPE: 0.847 Seoul St. Mary's Hospital(SSMH) FFPE by TCGA FFPE: 0.802 SSMH FFPE by both FFPE: 0.895	331 (TCGA) + 383 (SSMH)	Inception v3	Deep learning can accurately predict MSI status in GC from H&E slides. Model trained on TCGA data generalizes well to an independent Asian cohort (SSMH). MSI classifiers appear incompatible across cancer types (GC vs. CRC).	No, the research adopts the technical route of CNN + traditional pooling + three-level strong supervision classification
Lan et al. [30] 2023 China	Development and validation of a deep learning-based auxiliary diagnostic system	Gastric cancer pathology images	Yes	WSI images and patch images after H&E staining	WSI: 0.9034, Patch: 0.9552。	2020 specimens total (various splits for training, validation, and testing).	DeepLabv3 + for segmentation; ResNet-50 for classification	Proposed a system that achieves high accuracy and significantly aids pathologists in diagnosing gastric cancer, especially on biopsies, while reducing annotation workload. Active learning improved data efficiency.	No, the research adopts the technical route of strong supervision segmentation + classification + ASPP multi-scale pooling
Yong et al. [69] 2023 Malaysia	Development and evaluation of transfer learning-based CNNs for gastric cancer detection	Gastric cancer histopathology image patches	Yes	WSI images and patch images after H&E staining	DenseNet121: 0.9868	245,196 patches (97,076 abnormal, 148,120 normal) across 3 sub-databases (80 × 80, 120 × 120, 160 × 160 pixels), derived from 600 WSIs. Source: GasHisSDB	Various pre-trained CNN architectures (MobileNet, EfficientNet, DenseNet, Inception V3, Xception) fine-tuned for binary classification. DenseNet121 achieved the best performance	Transfer learning is effective for gastric cancer detection from histopathology images. DenseNet121 architecture performed best, achieving high accuracy on the GasHisSDB dataset. The study suggests potential for assisting pathologists in gastric cancer diagnosis	No, the research adopts the technical route of pre-trained CNN + traditional pooling + strong supervised fine-tuning

Table 2 (continued)

Author/ Year/ Country or Region	Research design	Study Subjects	Gastric cancer pathology	Use of pathology images	Diagnostic accuracy	Sample size	Deep learning model	Results or conclusion	Use MIL
Hu et al. [20] 2022 China	Comparison of the performance of various algorithms in the classification of gastric pathology subscale images	Gastric cancer histopathology image patches	Yes	WSI images and patch images after H&E staining	Machine learning, random forest classification of color histogram features: 0.8599 Deep learning: ResNet50: 0.9609, VGG16: 0.959, InceptionV3: 0.9457, ViT: 0.8621	GasHisSDB database: Total 245,196 images 97,076 abnormal images 148,120 normal images	VGG16, InceptionV3, ResNet50, ViT	The study demonstrates performance differences among various classifiers on the GasHisSDB dataset, suggesting their complementarity for ensemble learning. Classical machine learning models showed good performance for abnormal category classification, while some excelled in normal category classification. Deep learning models also exhibited complementarity.	No, the research adopts the technical route of CNN traditional pooling + ViT attention mechanism + strong supervision classification
Fu et al. [11] 2022 China	Development and validation of a deep learning-based multi-classification model, StoHisNet, for classification of gastric pathology images	Pathological images of gastric cancer patients, breast cancer patients, and endometrium	Yes	WSIs stained with H&E	Gastric: 0.9469 BreakHis: 0.9164 Endometrium: 0.8174	Gastric: 4511 images, Breast Cancer: 7160 images, Endometrial Cancer: 15,806 images	StoHisNet (proposed CNN-Transformer hybrid), EfficientNet-B4, ResNet-50, Xception, GoogLeNet, ViT, DeiT, TNT, PiT, AFF + ResNet.	StoHisNet achieves high performance and generalizability in multi-class classification of gastric histopathology images, outperforming other CNN and Transformer models. Combining CNN and Transformer architectures leverages both local and global information.	No, the research adopts the technical route of CNN traditional pooling + Transformer attention + strong supervision multi-classification
Han et al. [19] 2022 China	Development and validation of a deep learning algorithm for HER2 scoring in gastric cancer.	183 HER2 stained WSIs from Fujian Cancer Hospital.	Yes	WSIs stained with HER2	RepVGG-GCT + MLP:0.94	183 WSIs; split for TLCN and WHSPN training, validation, and testing.	Two-stage framework: Tile-level classification network (TLCN) using RepVGG with GCT or SE attention; WSI-level HER2 Score Prediction Network (WHSPN) using SVM or MLP. Re-parameterization used for efficiency.	Proposed a novel deep learning algorithm for HER2 scoring of gastric cancer, demonstrating high accuracy. The use of re-parameterization improved inference speed.	No, the research adopts the technical route of strong supervision patch classification + re-parameterization optimization + statistical aggregation

Table 2 (continued)

Author/ Year/ Country or Region	Research design	Study Subjects	Gastric cancer pathology	Use of pathology images	Diagnostic accuracy	Sample size	Deep learning model	Results or conclusion	Use MIL
Xue et al. [68] 2022 China	Development and validation of an ensemble deep learning model for predicting distant metastasis	Gastric cancer pathology images	Yes	WSIs stained with H&E	Ensemble model: 0.84	492 gastric cancer patients (256 with distant metastasis, 256 without). 110,202 tiles extracted after sliding window slicing	Ensemble of ResNet-18, MobileNetV3-small, and EfficientNetV2-S. Majority Voting ensemble strategy used.	The ensemble deep learning model achieved higher accuracy in predicting distant metastasis compared to single models. Pre-training a tumor region detection model and using transfer learning for individual classifiers contributed to improved performance. This approach can potentially assist clinicians in selecting treatment strategies.	No, the research adopts the technical route of strong supervised learning framework + CNN + integrated learning
Tung et al. [61] 2022 Taiwan	Development and validation of a deep learning model for GC detection on slide images	Images of GC slices from patients diagnosed with GC	Yes	WSI images and patch images after H&E staining	Validation set:0.91	13,600 images from 50 patients (2200 images from 40 patients for training, 550 images from 10 patients for testing)	YOLOv4 object detection model. A 3D model is subsequently generated to refine GC localization	The deep learning model successfully identified GC in pathological slide images with high accuracy. The use of a 3D model helped to further pinpoint the location of GC. The study suggests AI can assist pathologists in GC diagnosis, particularly for screening and quality control	No, the research adopts the technical route of image patching + YOLOv4 target detection + spatial pyramid pooling (SPP)
Kanavati et al. [24] 2021 Japan	A deep learning model was trained to classify gastric diffuse adenocarcinoma in WSI.	Patients with diffuse gastric adenocarcinoma	Yes	WSIs stained with H&E	Hospital 1:0.924 Hospital 2:0.946 Hospital 3:0.925 Hospital 4:0.888 Hospital 5:0.903	Training: $n = 353$ (diffuse-type ADC) Testing: n = variable across 5 test sets	Two-stage (InceptionV3, RNN); One-stage (EfficientNetB1)	Deep learning shows promise for classifying diffuse-type ADC in WSIs, potentially aiding pathologists.	No, the research adopts the technical route of strong supervised learning framework + CNN + global average pooling

Table 2 (continued)

Author/ Year/ Country or Region	Research design	Study Subjects	Gastric cancer pathology	Use of pathology images	Diagnostic accuracy	Sample size	Deep learning model	Results or conclusion	Use MIL
Huang et al. [21] 2021 China	Development and validation of a deep learning-based model to assist in predicting the diagnosis and overall survival of GC patients using pathological images.	Gastric cancer candidates and gastric cancer patients	Yes	WSIs stained with H&E	GastroMIL: 0.92	2333 images (RHWU & TCGA for model development) + 175 images (NHGRP for external validation)	GastroMIL (diagnostic model): CNN with RNN aggregation. MIL-GC (prognostic model): CNN with MLP. Both utilize Multiple Instance Learning (MIL).	Developed two AI models (GastroMIL and MIL-GC) for accurate diagnosis and prognosis prediction of GC, respectively. GastroMIL showed high accuracy, comparable to expert pathologists. MIL-GC effectively stratified patients into risk groups for overall survival. An online platform makes these models readily accessible.	Yes
Sai Chandra Kosaraju et al. [27] 2020 USA	Development and validation of a deep learning model called Deep-Hipo for histopathology image analysis.	Patients who underwent endoscopic biopsy between February 2016 and July 2017	Yes	WSIs stained with H&E	Deep-Hipo, well-differentiated (WD): 0.951 ± 0.0021, moderately differentiated: 0.949 ± 0.0024, poorly differentiated: 0.918 ± 0.0029 and poorly adherent adenocarcinoma: 0.931 ± 0.0019, overall: 0.937 ± 0.0032.	94 (gastric biopsies)	Multi-scale receptive field CNN (Deep-Hipo), InceptionV3, other benchmark models (N-Net, VGG16, DenseNet, EfficientNet, MRD-Net, GB-INCV3, CAT-Net)	Deep-Hipo outperformed other deep learning models in classifying gastric histopathology images. Multi-scale analysis improves accuracy.	No, the research adopts the technical route of multi-scale feature extraction (through parallel CNN branches) + feature connection (Concatenation) + local pooling operation (Max/Average Pooling)

Table 2 (continued)

Author/ Year/ Country or Region	Research design	Study Subjects	Gastric cancer pathology	Use of pathology images	Diagnostic accuracy	Sample size	Deep learning model	Results or conclusion	Use MIL
Ma et al. [39] 2020 China	Development and validation of a method combining staining normalization, deep convolutional neural network (CNN), and random forest (RF) classifier for the detection and classification of normal gastric mucosa, chronic gastritis, and intestinal type gastric cancer.	Patients with normal gastric tissue, chronic gastritis, and intestinal type gastric cancer	Yes	WSI images and patch images after H&E staining	Patch image: binary classification (benign and cancer) test set without color normalization: 0.981, after color normalization: 0.984, in the three-category (normal mucosa, chronic gastritis and gastric cancer) prediction, the test set: 0.945. WSIs: 0.96	763 WSIs	Inception v3 CNN, Random Forest Classifier	CNN-based system can effectively classify gastric mucosal lesions. Stain normalization improves CNN performance. Features extracted by CNN are useful for predicting prognosis.	No, the research adopts the technology route of Global Average Pooling
Song et al. [56] 2020 China	Development and validation of a deep learning diagnostic system	Patients with gastric cancer or suspected gastric cancer	Yes	WSIs stained with H&E	Daily gastric cancer dataset: PLAGH dataset: 0.873, Multi-center dataset: PUMCH dataset: 0.943, CHCAMS dataset: 0.976	2123 annotated WSIs for training (PLAGH), 3212 WSIs for testing (PLAGH, 3 scanners), 1582 WSIs for multicenter validation	DeepLab v3 with ResNet-50 backbone	Developed and validated a clinically applicable system for gastric cancer detection on WSIs. System achieved high sensitivity and good specificity, aiding pathologists in diagnosis and reducing misdiagnoses. Demonstrated robustness through multicenter testing.	No, the research adopts the technical route of void space pyramid pooling
Wang et al. [65] 2025 China	Model development and experimental validation study with comparative evaluation	Whole slide images from gastric cancer patients (PDGC-YNCH, TCGA-STAD, HE-GHI-DS datasets)	Yes	WSIs stained with H&E	Accuracy up to 96.51% (HE-GHI-DS), 93.46% (TCGA-STAD), 90.31% (PDGC-YNCH); AUC up to 0.9837	PDGC-YNCH: 233 WSIs, 119,200 tiles TCGA-STAD: 30 WSIs, 2786 tiles HE-GHI-DS: 280 original images, 1680 after augmentation	CFI-ViT (a custom vision transformer using a coarse-to-fine two-stage inference strategy)	CFI-ViT achieved superior accuracy and efficiency compared to state-of-the-art CNNs and MIL methods; demonstrates strong generalization and clinical potential	No, we are researching the technical route of using global average pooling + differentiable patch selection module

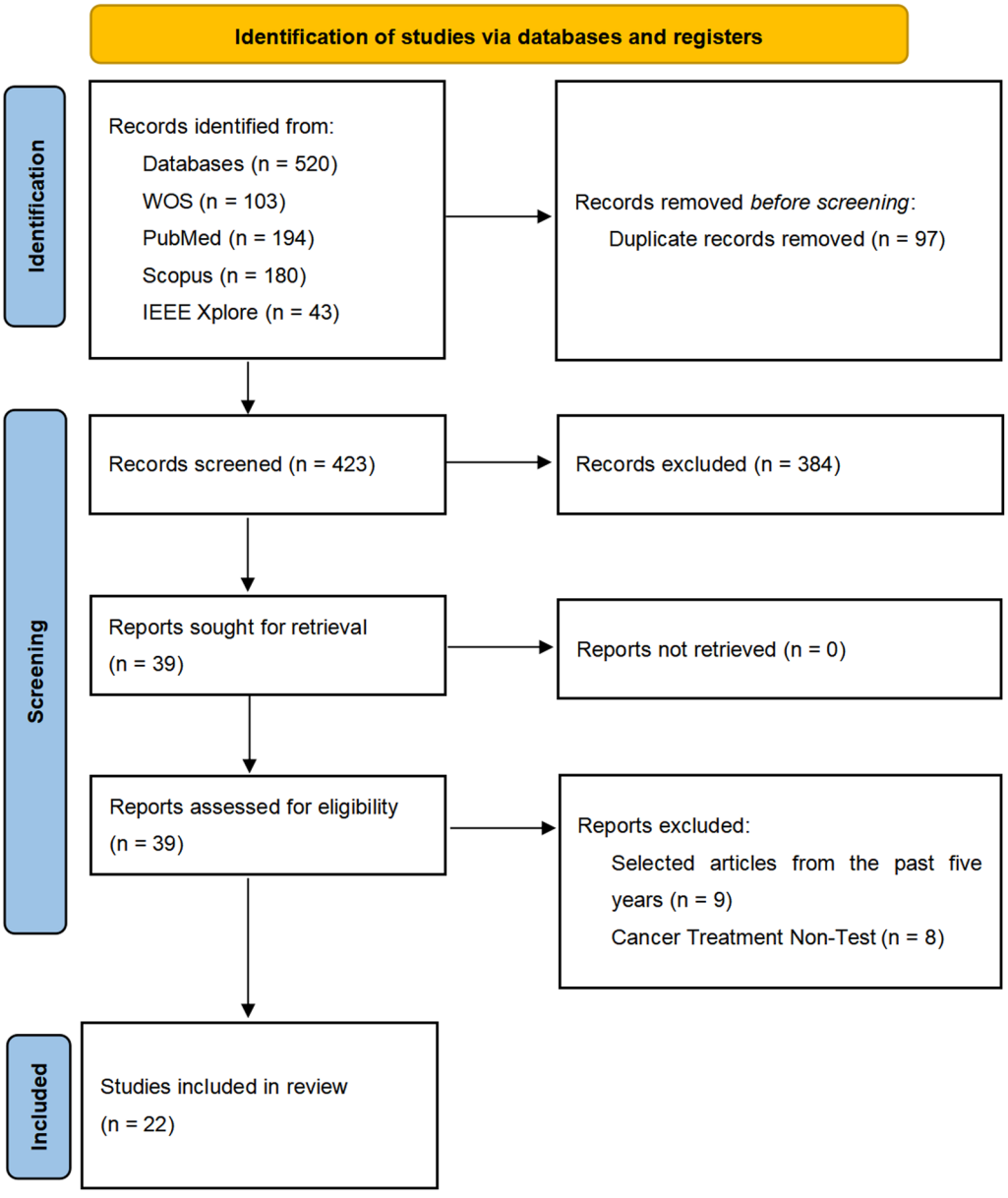

Fig. 1 PRISMA flow chart

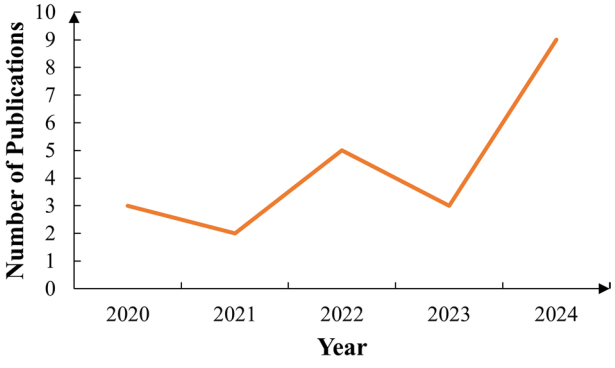

Fig. 2 Annual number of publications from 2020–2024, showing research activity trends relevant to the study

and morphological information for diagnosis. In contrast, only one study used WSI of immunohistochemical (IHC) staining. This method provides molecular and protein-level insights but is less commonly used due to its high cost, complexity, and the need for specific markers. To overcome challenges such as the large data sizes and computational costs associated with WSI, 17 studies extracted smaller image patches from WSI for model training and validation, as shown in Fig. 6. The frequent use of WSI patches reflects their practical advantages, such as reducing computational load and focusing on specific regions of interest, which are essential for efficient and accurate model development. This trend highlights the balance researchers aim to achieve between

Fig. 3 Number of publications by country/region over the past five years, showing the geographic distribution of research output relevant to this study

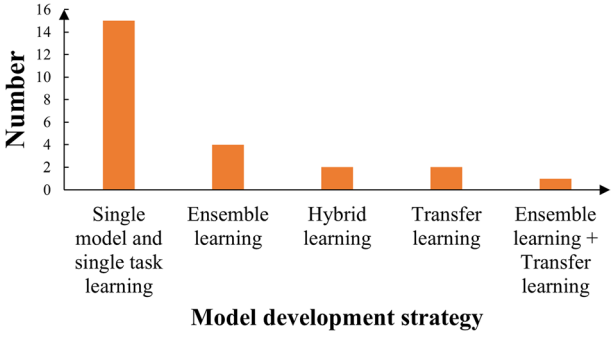

Fig. 4 Distribution of different model development strategies in the included studies, showcasing the prevalence of each approach

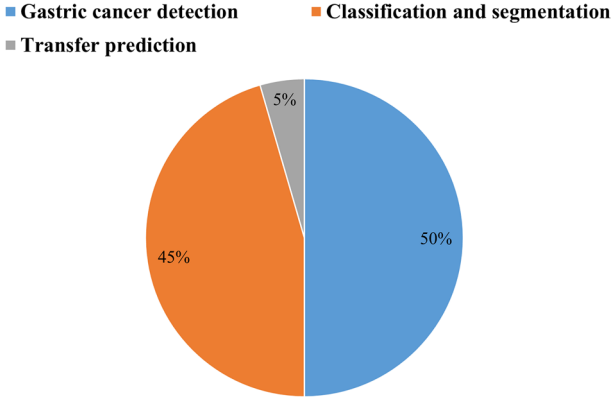

Fig. 5 Distribution of research tasks in GC studies, with proportions of detection, classification-segmentation, and transfer prediction

Table 3 Data sources

Dataset	Number
Self-collected clinical data	11
GasHisSDB	8
TCGA	2
BOT	1

leveraging the detailed information of WSI and addressing computational limitations.

DL models employed

Table 2 summarizes the DL models used in the included studies. CNN architectures constituted the primary model type, with ResNet, VGG, and Inception, along with their variants (e.g., ResNet50, VGG16, Inception v3), being widely employed. Beyond these classic CNN architectures, some studies explored other models, such as the Vision Transformer (ViT) and its variants like PiT, CFI-ViT, and Swin Transformer.

To enhance performance, various DL models were employed. These included ensemble learning, such as ensembles of ResNet34, VGGNet, and other CNN architectures, as well as ensembles of ResNet18, MobileNetV3-small, and EfficientNetV2-S. Transfer learning approaches included using pre-trained MobileNetV2, EfficientNetB0, DenseNet121, InceptionV3, Xception, and ResNet50 models, enhanced by in-domain transfer learning for feature extraction. Fine-tuning of various pre-trained CNN architectures for binary classification was also employed. Hybrid models were also explored, such as a combination of EfficientNetV2B0 and CatBoost, and a CNN-Transformer hybrid model called StoHisNet.

Regarding the aggregation of patch-level predictions to a WSI level, study review found that only one of the 22 included studies explicitly employed a MIL framework. The remaining studies utilizing WSI-based approaches adopted alternative strategies for deriving slide-level predictions. These strategies mainly include maximum pooling, average pooling, and specified methods such as attention-based pooling or voting mechanisms.

Furthermore, some studies have used other technologies, such as DeepLabv3 + combined with ResNet50 and YOLO target detection, as shown in Table 4.

Applications of DL models in GC pathology image analysis

The included studies demonstrated diverse applications of DL models in GC pathology image analysis, reflecting the active and multifaceted research in this area. These applications include:

GC detection

This was the most common application, with numerous studies developing DL models for automated GC detection. For example, Mudavadkar et al. [44] proposed an ensemble model, combining the strengths of multiple DL models and leveraging the aggregated knowledge of numerous models to enhance classification performance. Their results demonstrated an average precision exceeding 0.99 for GC detection. Yong et al. [70] introduced a DL model based on in-domain transfer learning and ensemble learning, achieving a detection accuracy of

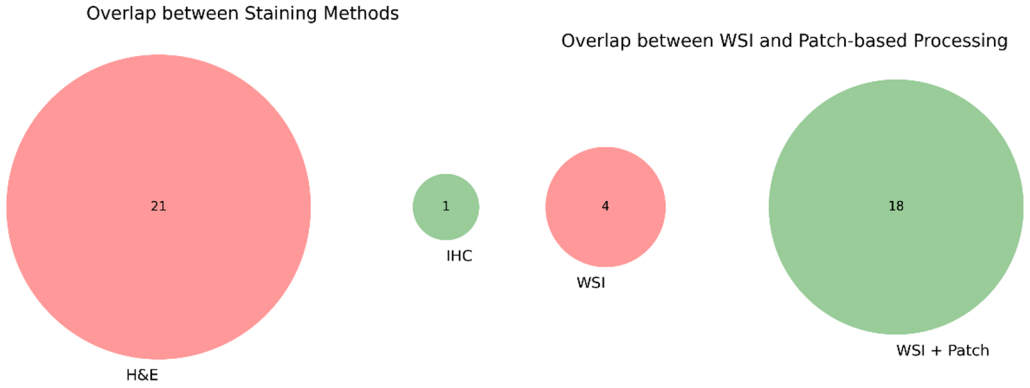

Fig. 6 Venn diagrams illustrating the overlap between staining methods (H&E vs. IHC) and data processing strategies (WSI vs. patch-based)

Table 4 DL models in the included studies

Model Type	Name	Number
CNN Architecture	ResNet、VGGNet、Inception、EfficientNet、MobileNet、DenseNet、Xception	21
Other Architecture	Vision Transformer and the Extend model	6
Ensemble learning model	Integration of ResNet34, VGGNet and other CNN architectures、Integration of ResNet18, MobileNetV3-small and EfficientNetV2-S	2
Transfer Learning Model	MobileNetV2、EfficientNetB0、DenseNet121、InceptionV3、Xception and ResNet50	2
Hybrid Model	EfficientNetV2B0 + CatBoost、StoHisNet: CNN-Transformer	2
Other technologies	DeepLabv3 + and ResNet50, YOLO	3

0.9978. Lan et al. [30] developed a DL-based pathology-assisted diagnostic system using a dataset of 1514 H&E-stained gastric specimens with complete diagnostic information. Their system achieved an accuracy of 0.9034 on WSI and 0.9552 on image patches. Han et al. [19] proposed a DL model for quantitative HER2 assessment in GC to detect the disease, achieving an accuracy of 0.94 in HER2 score prediction. Xue et al. [68] developed an ensemble DL model for predicting distant metastasis in GC. Using pathological images, their model predicted distant metastasis in patients with gastric ulcers with high accuracy, achieving a precision of 0.84. Tung et al. [61] used AI to automatically locate GC within GC tissue slide images, achieving a detection accuracy of 0.91 with their DL model. Huang et al. [21] aimed to develop DL-based models to assist in predicting diagnosis and overall survival (OS) in GC patients using pathological images. Developing their algorithms on collected data, they achieved an accuracy of 0.92 with their DL model. Kosaraju et al. [27] proposed a novel multi-task-based HIstoPatholOgy DL model (called Deep-Hipo) that employs multi-scale patches for accurate histopathology image analysis. Their results demonstrated an overall accuracy of 0.937 ± 0.0032 with their DL model. Ma et al. [39] presented a CNN-based system to detect abnormalities in gastric mucosa. Integrating digitized histopathology of WSI, stain normalization, a deep CNN, and a random forest classifier, they achieved an accuracy of 0.981 on unstained image patches, which improved to 0.984 after staining. Their system achieved an accuracy of 0.945 for three-class prediction (normal mucosa, chronic gastritis, and GC) using patches and 0.96 using WSI. Song et al. [56] reported a clinically applicable system developed at the Chinese PLA General Hospital. This system uses a deep convolutional neural network trained on 2,123 pixel-level annotated H&E-stained whole-slide images. Their results demonstrated accuracies of 0.873 on a routine GC dataset, 0.943 on the multicenter PUMCH dataset, and 0.976 on the CHCAMS dataset.

Image classification

Many studies focused on developing DL models for classifying GC pathology images. This included differentiating benign from malignant tissues, as well as typing and grading GC. For example, Zubair et al. [73] proposed a novel gastric histology classification and segmentation framework by adaptively focusing on relevant features of the image. The results showed that it produced classification accuracies of 0.9887 and 0.9728 on the validation set and 0.9847 and 0.9731 on the test set. Huang et al. [22] designed a dilated CNN with late fusion strategy (DCNNLFS) for gastric histopathology image classification. The results showed that the accuracy of the DCNNLFS model reached 0.935. Khayatian et al. [26] proposed a hybrid DL and gradient boosting approach, extracting features by examining two classifiers (called pre-trained models) from six networks. Their results showed that the EfficientNetV2B0-CatBoost combination performed best, achieving an accuracy of 0.939 at 160px. Loddo et al. [38] used ML and DL models to classify histopathology images into healthy and cancerous categories. The results showed that an accuracy of 0.95 was achieved using a SVM classifier. Armand et al. [3] detected and

classified GC from weakly annotated tissue images using a visual transformer, achieving an accuracy of 0.859 on the ViT-16 model. Yong et al. [69] proposed a CNN based on transfer learning for binary classification of GC pathology images. Experimental results show that the DenseNet121 model achieved an accuracy of 0.9868 on the 160px database. Hu et al. [20] compared the performance of multiple algorithms in order to apply ensemble learning to the actual GC classification problem. The results showed that the performance of different classifiers on GasHisSDB was different. The accuracy of the classic ML model was 0.8599, and the accuracy of the DL model was 0.9609. Fu et al. [11] proposed a classification method based on DL and developed a new multi-scale model called StoHisNet based on Transformer and CNN. The results showed that the effect of this model reached 0.9469 in a public GC dataset. Kanavati et al. [24] trained a DL model to classify diffuse gastric adenocarcinoma in WSI. The results showed that good results were achieved in GC datasets from five different hospitals. Wang et al. [65] proposed a ViT-based model called CFI-ViT for GC subtype classification. Using a two-stage coarse-to-fine strategy, the model achieved accuracies of 0.9031, 0.9346, and 0.9651 on PDGC-YNCH, TCGA-STAD, and HE-GHI-DS datasets, respectively, showing strong performance and generalizability.

Lymphovascular invasion prediction

Lee et al. [33] proposed a DL-based LVI (+) detection method using H&E-stained whole slide images. The results showed that the integrated model accuracy reached 0.9514. This study used a DL model to predict lymphovascular invasion, which is of great significance for the prognostic assessment of GC [40].

Other application scenarios

In addition to the above application scenarios, some studies have also explored other applications. For example: Assessing Microsatellite Instability Using DL Models: Lee et al. [32] used DL to achieve fully automatic classification of microsatellite instability states in GC tissue sections. The results showed that DL can automatically learn the best features to distinguish microsatellite instability states in GC tissue sections.

In summary, the included articles cover multiple aspects of DL in GC pathology image analysis. From basic GC detection to more refined prognostic prediction and biomarker assessment, these applications demonstrate the broad potential of DL in this field.

Discussion

Findings and key conclusions

The main findings of this review focus on the potential and advantages of DL models in GC pathological image analysis. The included studies show that these models can effectively process complex pathological image data and extract key features related to GC. First, DL models have shown outstanding performance in improving diagnostic accuracy. Several studies have shown that these models achieve high accuracy in GC detection, which is crucial for increasing early diagnosis rates and reducing misdiagnosis and missed diagnoses. Second, DL models can learn the pathological features of different GC subtypes, helping pathologists make more accurate classifications. This aids in developing personalized treatment plans and improving prognosis [47]. Some studies also show that DL models can predict prognostic indicators such as lymph node metastasis, distant metastasis, and survival rate. These models have the potential to assess patients' responses to specific treatments and guide clinical decision-making [33]. Third, DL models can automate the assessment of biomarker expression levels, such as HER2, improving evaluation efficiency and consistency while reducing human errors. This is crucial for guiding targeted therapy [19]. Finally, DL models can assist in analyzing WSI, quickly identifying and locating areas of abnormality, which helps improve the efficiency of pathological diagnosis [28]. These findings suggest that DL models have the potential to become powerful tools for pathologists in GC diagnosis and prognosis assessment. By automating and standardizing key steps, they can improve diagnostic efficiency and accuracy, ultimately enhancing clinical management of GC patients. However, the performance of the models largely depends on the quality and quantity of the training data, as well as the model design and training strategies [37]. Future research needs to further optimize DL models and conduct larger-scale clinical validation to fully realize their potential in GC diagnosis and treatment.

Interactions and challenges of algorithms, images and diagnosis

The application of DL in GC pathology image analysis relies on the complex interaction between algorithms, images, and diagnosis. The algorithm extracts feature from the images and analyzes them to assist in diagnosis, but its effectiveness depends on its ability to extract and interpret features [25]. An ideal algorithm should be able to identify and interpret key image features related to GC, such as cell morphology, tissue structure, and staining patterns. However, the diversity and complexity of GC pathology, along with artifacts and noise that may be present in the images, pose challenges for feature extraction and interpretation [4]. Image quality also directly affects the performance of the algorithm. High-resolution, uniformly stained, and clear images without artifacts provide more reliable information, which helps the algorithm accurately identify and analyze pathological

features. In contrast, low-quality images may lead to performance decline and even misdiagnosis [8]. Therefore, it is crucial to acquire high-quality pathological images and perform necessary preprocessing, such as noise removal and color correction [54]. Finally, the algorithm's output, such as cancer tissue probability prediction, needs to be aligned with existing GC diagnostic standards. How to translate the algorithm's output into clinically actionable recommendations and integrate the algorithm into existing pathological diagnosis processes remains a challenge. This requires collaboration between pathologists and clinicians to develop clear guidelines and standards to ensure that DL models can effectively assist in diagnosis.

Challenges of dataset and model selection

DL models have shown potential in GC pathology image analysis, but selecting appropriate datasets and models remains a challenge. The selection and construction of datasets directly affect the training effect and generalization ability of the model. The studies included in this review used a variety of different datasets, including public datasets, such as GasHisSDB, and self-collected clinical datasets. These datasets vary significantly in size, quality, and representativeness. Some studies use datasets with a small sample size, which may lead to overfitting and insufficient generalization in real-world applications [6]. In addition, the quality of the datasets varies, and differences in image quality, resolution, staining methods, and labeling standards can all affect the model's learning performance. The representativeness of the dataset is also a key issue. An ideal dataset should include GC cases with different stages, grades, pathological types, and molecular subtypes to ensure that the model can handle the variations seen in real-world patients. Therefore, future research needs to focus more on building high-quality, large-scale, and representative GC pathological image datasets, as well as establishing standardized data collection and annotation processes. This is crucial for developing more reliable and generalizable DL models. In addition to the dataset, the choice of DL model also has a significant impact on the effectiveness of GC pathological image analysis. The studies included in this review used various DL models, including CNN, recurrent neural network (RNN), and ViT. The architecture, parameters, and training strategies of different models can all affect their performance on specific tasks. For example, CNN excel at image feature extraction [23], while RNN are more suitable for handling sequential data [9]. ViT, as an emerging DL model, has also shown strong capabilities in tasks like image classification and object detection [18]. However, there is currently no clear evidence to suggest which model is the best choice for all GC pathological image analysis scenarios. A notable methodological observation pertinent to WSI analysis is the limited adoption of MIL, with only one of the 22 studies explicitly employing this paradigm. This is surprising given MIL's increasing consideration as a standard for WSI analysis in other oncological domains, especially for weakly-labeled data [5]. The predominance of alternative strategies (e.g., max/average pooling or direct fully supervised WSI models) suggests a potential slower uptake of MIL in GC-specific research, a focus on tasks amenable to simpler aggregation, or underreporting. This reliance on simpler methods or fully supervised WSI approaches (which often require extensive annotation) may not fully leverage WSI information or efficiently handle slide-level prediction from patch features. This disparity highlights a need for greater exploration of MIL in GC pathology and more transparent reporting of WSI aggregation strategies. In addition, model interpretability and clinical validation are important factors to consider. A lack of interpretability may limit the model's application in clinical practice, while clinical validation is crucial to ensure the model's reliability and effectiveness. Therefore, future research needs to systematically compare the performance of different models on various datasets and in different application scenarios. It should also explore more effective model selection strategies to develop DL models that are more accurate, reliable, interpretable, and clinically validated.

Limitation

This review also has some limitations. First, the number of included studies is relatively small, which may not fully reflect all the applications and progress of DL in GC pathological image analysis. Second, due to differences between studies, such as different datasets, models, and evaluation metrics, it is difficult to perform a meta-analysis or draw more accurate conclusions. Finally, this review mainly focuses on English literature, which may have missed some important studies in other languages.

Future efforts in using DL to analyze GC pathological images need to focus on building larger, higher-quality, and more representative datasets, along with establishing standardized data collection and annotation processes. Future development and validation also require more interpretable DL models, and it will be important to explore how to combine the model's predictions with clinical information to assist clinical decision-making. Conducting multi-center, prospective clinical studies to validate the effectiveness and practicality of DL models in real-world clinical settings is also a future research direction. In addition, future research needs to explore the performance differences of DL models across different GC subtypes, stages, and grades, and investigate how to optimize the models for different patient groups.

Conclusion

This scoping review systematically evaluates the current application of DL in GC pathological image analysis. Through the analysis of 22 relevant studies, it was found that DL models demonstrate significant potential in GC detection, classification, and prognosis prediction. Commonly used models include CNN, RNN, and ViT, often combined with techniques such as transfer learning and ensemble learning. However, current research is still limited by issues such as small dataset size, lack of standardized evaluation metrics and external validation, and insufficient model interpretability. The significance of this review lies in three aspects: first, it confirms the value of DL in GC pathological image analysis; second, it identifies the limitations of current research, especially regarding data quality, model interpretability, and clinical validation; third, it points out directions for future research, including building high-quality datasets, developing interpretable models, conducting multi-center clinical validation, and exploring new application scenarios. In conclusion, DL has the potential to transform the diagnosis and treatment of GC, but existing challenges need to be addressed, and further research is required.

Abbreviations

GC	Gastric cancer
DL	Deep learning
PRISMA-ScR	Preferred Reporting Items for Systematic Reviews and Meta-Analysis Extension for Scoping Reviews
CNN	Convolutional neural networks
ML	Machine learning
MSI	Microsatellite instability
AI	Artificial intelligence
dMMR	Defective DNA mismatch repair
WSI	Whole-slide digital scanned images
ViT	Vision Transformer
OS	Overall survival
DCNNLFS	Dilated CNN with late fusion strategy
RNN	Recurrent neural network
MIL	Multiple Instance Learning
IHC	Immunohistochemical

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-14662-3.

Supplementary Material 1.

Acknowledgements

Not applicable.

Author contributions

S.X. and Y.X. proposed and designed this study. T.L. helped with data analyses and resolved disagreements in the review process. S.X. wrote the manuscript with the assistance from Y.L. P.P. supervised and provided resources for this work. All authors discussed the results and contributed to the final manuscript.

Funding

This research is supported by Macao Polytechnic University research grant (project code: RP/FCA-10/2022).

Data availability

The author confirms that all data generated or analyzed during this study are included in this published article. Furthermore, primary and secondary sources and data supporting the findings of this study were all publicly available at the time of submission.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 10 February 2025 / Accepted: 10 July 2025
Published online: 01 August 2025

References

1. Alam MR, Abdul-Ghafar J, Yim K, Thakur N, Lee SH, Jang HJ, Jung CK, Chong Y. Recent applications of artificial intelligence from histopathologic image-based prediction of microsatellite instability in solid cancers: A systematic review. Cancers. 2022;14:2590. https://doi.org/10.3390/cancers14112590.
2. Ao Y, Li H, Zhu L, Ali S, Yang Z. The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling. J Pet Sci Eng. 2019;174:776–89. https://doi.org/10.1016/j.petrol.2018.11.067.
3. Armand T, Bhattacharjee S, Kim H-J, Hussain A, Ali S, Choi H-K, Kim H-C. Vision transformer-based model for gastric cancer detection and classification using weakly annotated histopathological images. 26th Int Conf Adv Commun Technol (ICACT). 2024. https://doi.org/10.23919/icact60172.2024.10471994.
4. Ayyad SM, Shehata M, Shalaby A, Abou El-Ghar M, Ghazal M, El-Melegy M, Abdel-Hamid NB, Labib LM, Ali HA, El-Baz A. Role of AI and histopathological images in detecting prostate cancer: A survey. Sensors. 2021;21:2586. https://doi.org/10.3390/s21082586.
5. Cao R, Gong L, Dong D. Pathological diagnosis and prognosis of gastric cancer through a multi-instance learning method. EBioMedicine. 2021;73:103647. https://doi.org/10.1016/j.ebiom.2021.103671.
6. Cawley GC, Nicola. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res. 2010;11:2079–107. https://doi.org/10.5555/1756006.1859921.
7. Chan H-P, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. In: Advances in Experimental Medicine and Biology. 2020. pp. 3–21. https://doi.org/10.1007/978-3-030-33128-3_1
8. Chow LS, Paramesran R. Review of medical image quality assessment. Biomed Signal Process Control. 2016;27:145–54. https://doi.org/10.1016/j.bspc.2016.02.006.
9. Fang W, Chen Y, Xue Q. Survey on research of RNN-based spatio-temporal sequence prediction algorithms. J Big Data. 2021;3:97–110. https://doi.org/10.32604/jbd.2021.016993.
10. Faruqui N, Yousuf MA, Whaiduzzaman M, Azad AKM, Barros A, Moni MA, LungNet. A hybrid deep-CNN model for lung cancer diagnosis using CT and wearable sensor-based medical IoT data. Comput Biol Med. 2021;139:104961. https://doi.org/10.1016/j.compbiomed.2021.104961.
11. Fu B, Zhang M, He J, Cao Y, Guo Y, Wang R, StoHisNet. A hybrid multi-classification model with CNN and transformer for gastric pathology images. Comput Methods Programs Biomed. 2022;221:106924. https://doi.org/10.1016/j.cmpb.2022.106924.
12. Gao F, Wu T, Li J, Zheng B, Ruan L, Shang D, Patel B. SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. Comput Med Imaging Graph. 2018;70:53–62. https://doi.org/10.1016/j.compmedimag.2018.09.004.
13. Giryes A, Oweira H, Mannhart M, Decker M, Abdel-Rahman O. Exploring the differences between early-onset gastric cancer and traditional-onset gastric cancer. J Gastrointest Oncol. 2018;9:1157–63. https://doi.org/10.21037/jgo.2018.06.11.
14. Golan R, Jacob C, Denzinger J. Lung nodule detection in CT images using deep convolutional neural networks. In: 2016 International Joint Conference

on Neural Networks (IJCNN). 2016. https://doi.org/10.1109/ijcnn.2016.7727205
15. Gouda W, Sama NU, Al-Waakid G, Humayun M, Jhanjhi NZ. Detection of skin cancer based on skin lesion images using deep learning. Healthcare. 2022;10:1183. https://doi.org/10.3390/healthcare10071183.
16. Grundmann E, Grunze H, Witte S. Early gastric cancer. Berlin: Springer Science & Business Media; 2012. ISBN 9783642658914.
17. Gullo I, Oliveira C, Rachel, Jolanda, Davis JL, Bougen-Zhukov N, Guilford P, Carneiro F. Updated perspective and directions on hereditary diffuse gastric cancer. In: Elsevier eBooks. 2021. pp. 217–258. https://doi.org/10.1016/b978-0-323-85563-1.00011-3
18. Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y, et al. A survey on vision transformer. IEEE Trans Pattern Anal Mach Intell. 2022;45:1. https://doi.org/10.1109/tpami.2022.3152247.
19. Han Z, Lan J, Wang T, Hu Z, Huang Y, Deng Y, Zhang H, Wang J, Chen M, Jiang H, et al. A deep learning quantification algorithm for HER2 scoring of gastric cancer. Front Neurosci. 2022;16. https://doi.org/10.3389/fnins.2022.877229.
20. Hu W, Chen H, Liu W, Li X, Sun H, Huang X, Grzegorzek M, Li C. A comparative study of gastric histopathology sub-size image classification: from linear regression to visual transformer. Front Med. 2022;9. https://doi.org/10.3389/fmed.2022.1072109.
21. Huang B, Tian S, Zhan N, Ma J, Huang Z, Zhang C, Zhang H, Ming F, Liao F, Ji M, et al. Accurate diagnosis and prognosis prediction of gastric cancer using deep learning on digital pathological images: A retrospective multicentre study. EBioMedicine. 2021;73:103631. https://doi.org/10.1016/j.ebiom.2021.103631.
22. Huang J, Shier Nee Saw, He T, Yang R, Qin Y, Chen Y, Loo Chu Kiong. DCNNLFS: A dilated convolutional neural network with late fusion strategy for intelligent classification of gastric histopathology images. IEEE J Biomed Health Inf. 2023;1–10. https://doi.org/10.1109/jbhi.2023.3334709.
23. Jogin M, Mohana, Madhulika MS, Divya GD, Meghana RK, Apoorva S. Feature extraction using convolution neural networks (CNN) and deep learning. In: 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). 2018. pp. 2319–2323. https://doi.org/10.1109/RTEICT42901.2018.9012507
24. Kanavati F, Tsuneki M. A deep learning model for gastric diffuse-type adenocarcinoma classification in whole slide images. Sci Rep. 2021;11. https://doi.org/10.1038/s41598-021-99940-3.
25. Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. IEEE Access. 2018;6:9375–89. https://doi.org/10.1109/ACCESS.2017.2788044.
26. Khayatian D, Maleki A, Nasiri H, Morteza D. Histopathology image analysis for gastric cancer detection: A hybrid deep learning and catboost approach. Multimedia Tools Appl. 2024. https://doi.org/10.1007/s11042-024-19816-2.
27. Kosaraju SC, Hao J, Koh HM, Kang M, Deep-Hipo. Multi-scale receptive field deep learning for histopathological image analysis. Methods. 2020;179:3–13. https://doi.org/10.1016/j.ymeth.2020.05.012.
28. Kumar N, Gupta R, Gupta S. Whole slide imaging (WSI) in pathology: current perspectives and future directions. J Digit Imaging. 2020;33. https://doi.org/10.1007/s10278-020-00351-z.
29. Kuntz S, Krieghoff-Henning E, Kather JN, Jutzi T, Höhn J, Kiehl L, Hekler A, Alwers E, von Kalle C, Fröhling S, et al. Gastrointestinal cancer classification and prognostication from histology using deep learning: systematic review. Eur J Cancer. 2021;155:200–15. https://doi.org/10.1016/j.ejca.2021.07.012.
30. Lan J, Chen M, Wang J, Du M, Wu Z, Zhang H, Xue Y, Wang T, Chen L, Xu C, et al. Using less annotation workload to Establish a pathological auxiliary diagnosis system for gastric cancer. Cell Rep Med. 2023;4:101004. https://doi.org/10.1016/j.xcrm.2023.101004.
31. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44. https://doi.org/10.1038/nature14539.
32. Lee SH, Lee Y, Jang H. Deep learning captures selective features for discrimination of microsatellite instability from pathologic tissue slides of gastric cancer. Int J Cancer. 2022;152:298–307. https://doi.org/10.1002/ijc.34251.
33. Lee J, Cha S, Kim J, Kim JJ, Kim N, Gal J, Kim JH, Lee JH, Choi Y-D, Kang SR, et al. Ensemble deep learning model to predict lymphovascular invasion in gastric cancer. Cancers. 2024;16:430. https://doi.org/10.3390/cancers16020430.
34. Li N, Yang J, Li X, Shi Y, Wang K. Accuracy of artificial intelligence-assisted endoscopy in the diagnosis of gastric intestinal metaplasia: A systematic review and meta-analysis. PLoS ONE. 2024;19:e0303421. https://doi.org/10.1371/journal.pone.0303421.
35. Li J, Jiang P, An Q, Wang GG, Kong HF. Medical image identification methods: A review. Comput Biol Med. 2024;169:107777. https://doi.org/10.1016/j.compbiomed.2023.107777.
36. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88. https://doi.org/10.1016/j.media.2017.07.005.
37. Liu T, Pang PCI, Lam CK. Public health education using social learning theory: A systematic scoping review. BMC Public Health. 2024;24:1906. https://doi.org/10.1186/s12889-024-19333-9.
38. Loddo A, Usai M, Ruberto CD. Gastric cancer image classification: A comparative analysis and feature fusion strategies. J Imaging. 2024;10:195. https://doi.org/10.3390/jimaging10080195.
39. Ma B, Guo Y, Hu W, Yuan F, Zhu Z, Yu Y, Zou H. Artificial intelligence-based multiclass classification of benign or malignant mucosal lesions of the stomach. Front Pharmacol. 2020;11:572372. https://doi.org/10.3389/fphar.2020.572372.
40. Marano L, Verre L, Carbone L, Poto GE, Fusario D, Venezia DF, et al. Current trends in volume and surgical outcomes in gastric cancer. J Clin Med. 2023;12(7):2708. https://doi.org/10.3390/jcm12072708.
41. Masud M, Eldin Rashed AE, Hossain MS. Convolutional neural network-based models for diagnosis of breast cancer. Neural Comput Appl. 2020;34. https://doi.org/10.1007/s00521-020-05394-5.
42. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009;62:1006–12. https://doi.org/10.1016/j.jclinepi.2009.06.005.
43. Mostavi M, Chiu Y-C, Huang Y, Chen Y. Convolutional neural network models for cancer type prediction based on gene expression. BMC Med Genomics. 2020;13. https://doi.org/10.1186/s12920-020-0677-2.
44. Mudavadkar GR, Deng M, Awad M, Arora IH, Breggia A, Ahmad B, Christman R, Ryan ST, Amal S. Gastric cancer detection with ensemble learning on digital pathology: use case of gastric cancer on GasHisSDB dataset. Diagnostics. 2024;14:1746. https://doi.org/10.3390/diagnostics14161746.
45. Muti HS, Heij LR, Keller G, Kohlruss M, Langer R, Dislich B, et al. Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: a retrospective multicentre cohort study. Lancet Digit Health. 2021;3:e10654–664. https://doi.org/10.1016/S2589-7500(21)00133-3.
46. Nawab K, Athwani R, Naeem A, Hamayun M, Wazir M. A review of applications of artificial intelligence in gastroenterology. Cureus. 2021;13. https://doi.org/10.7759/cureus.19235.
47. Ning X, Liu R, Wang N, Xiao X, Wu S, Wang Y, Yi C, He Y, Li D, Chen H. Development of a deep learning-based model to diagnose mixed-type gastric cancer accurately. Int J Biochem Cell Biol. 2023;162:106452. https://doi.org/10.1016/j.biocel.2023.106452.
48. Niu C, Zhang J, Okolo PI III, Daglilar E. Plant polyphenols in gastric cancer: nature's healing touch. Semin Oncol. 2025;52:152333. https://doi.org/10.1053/j.seminoncol.2025.01.002.
49. Park JH, Kim EY, Luchini C, Eccher A, Tizaoui K, Smith L, Lim BJ. Artificial intelligence for predicting microsatellite instability based on tumor histomorphology: A systematic review. Int J Mol Sci. 2022;23:2462. https://doi.org/10.3390/ijms23052462.
50. Reda I, Ayinde BO, Elmogy M, Shalaby A, El-Melegy M, El-Ghar MA, El-fetouh AA, Ghazal M, El-Baz A. A new CNN-based system for early diagnosis of prostate cancer. In: 2018 IEEE 19th International Symposium on Biomedical Imaging (ISBI). 2018. pp. 207–210. https://doi.org/10.1109/isbi.2018.8363556
51. Renjith VR, Judith JE. A review on explainable artificial intelligence for gastrointestinal cancer using deep learning. In: 2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS). 2023. pp. 1–6. https://doi.org/10.1109/aicera/icis59538.2023.10420139
52. Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas M. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol Rev. 2015;71:804–18. https://doi.org/10.1016/j.oregeorev.2015.01.001.
53. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Summers RM, Giger ML. Deep learning in medical imaging and radiation therapy. Med Phys. 2018;46:e1–36. https://doi.org/10.1002/mp.13264.
54. Salvi M, Acharya UR, Molinari F, Meiburger KM. The impact of pre- and post-image processing techniques on deep learning frameworks: A

comprehensive review for digital pathology image analysis. Comput Biol Med. 2020;128:104129. https://doi.org/10.1016/j.compbiomed.2020.104129.
55. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221–48. https://doi.org/10.1146/annurev-bioeng-071516-044442.
56. Song Z, Zou S, Zhou W, Huang Y, Shao L, Yuan J, Gou X, Jin W, Wang Z, Chen X, et al. Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning. Nat Commun. 2020;11. https://doi.org/10.1038/s41467-020-18147-8.
57. Takahashi T, Saikawa Y, Kitagawa Y. Gastric cancer: current status of diagnosis and treatment. Cancers. 2013;5:48–63. https://doi.org/10.3390/cancers5010048.
58. Thrumurthy SG, Chaudry MA, Hochhauser D, Mughal M. The diagnosis and management of gastric cancer. BMJ. 2013;347:f6367. https://doi.org/10.1136/bmj.f6367.
59. Tolias G, Sicre R, Jégou H. Particular object retrieval with integral max-pooling of CNN activations. ArXiv preprint. 2016. ArXiv:1511.05879. https://doi.org/10.48550/arxiv.1511.05879
60. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med. 2021;13. https://doi.org/10.1186/s13073-021-00968-x.
61. Tung C-L, Chang H-C, Yang B-Z, Hou K-J, Tsai H-H, Tsai C-Y, Yu P-T. Identifying pathological slices of gastric cancer via deep learning. J Formos Med Assoc. 2022;121:2457–64. https://doi.org/10.1016/j.jfma.2022.05.004.
62. Turtoi DC, Brata VD, Incze V, Ismaiel A, Dumitrascu DI, Militaru V, Munteanu MA, Botan A, Toc DA, Duse TA, et al. Artificial intelligence for the automatic diagnosis of gastritis: A systematic review. J Clin Med. 2024;13:4818. https://doi.org/10.3390/jcm13164818.
63. Tymchenko BI, Marchenko P, Spodarets D. Deep learning approach to diabetic retinopathy detection. arXiv preprint. 2020. arXiv:2003.02261. https://doi.org/10.48550/arxiv.2003.02261
64. Veldhuizen GP, Röcken C, Behrens HM, Cifci D, Muti HS, Yoshikawa T, et al. Deep learning-based subtyping of gastric cancer histology predicts clinical outcome: a multi-institutional retrospective study. Gastric Cancer. 2023;26:708–20. https://doi.org/10.1007/s10120-023-01398-x.
65. Wang X, Tao H, Wang B, Jin H, Li Z. CFI-ViT: A coarse-to-fine inference based vision transformer for gastric cancer subtype detection using pathological images. Biomed Signal Process Control. 2025;100:107160. https://doi.org/10.1016/j.bspc.2024.107160.
66. Wang X, Jiang Y, Yang S, Wang F, Zhang X, Wang W, et al. Foundation model for predicting prognosis and adjuvant therapy benefit from digital pathology in GI cancers. J Clin Oncol. 2025;JCO–24. https://doi.org/10.1200/JCO-24-01501.
67. Xie S, Yu Z, Lv Z. Multi-disease prediction based on deep learning: A survey. Comput Model Eng Sci. 2021;128:489–522. https://doi.org/10.32604/cmes.2021.016728.
68. Xue F, Jin H, Li Z, Tao H. Novel prediction model of gastric cancer distant metastases based on ensemble deep learning. In: 2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI). 2022. pp. 435–440. https://doi.org/10.1109/iwecai55315.2022.00091
69. Yong MP, Hum YC, Lai KW, Goh CH, Yap W-S, Tee YK. Histopathological gastric cancer detection using transfer learning. In: 11th Int Conf Bioinformatics Comput Biol (ICBCB). 2023. pp. 123–129. https://doi.org/10.1109/icbcb57893.2023.10246524
70. Yong MP, Hum YC, Lai KW, Lee YL, Goh CH, Yap WS, Tee YK. Histopathological cancer detection using intra-domain transfer learning and ensemble learning. IEEE Access. 2024;12:1434–57. https://doi.org/10.1109/access.2023.3343465.
71. Zhang X, Zhang Y, Zhang G, Qiu X, Tan W, Yin X, Liao L. Deep learning with radiomics for disease diagnosis and treatment: challenges and potential. Front Oncol. 2022;12:773840. https://doi.org/10.3389/fonc.2022.773840.
72. Zhuang H, Zhang J, Liao F. A systematic review on application of deep learning in digestive system image processing. Vis Comput. 2021;39. https://doi.org/10.1007/s00371-021-02322-z.
73. Zubair M, Owais M, Mahmood T, Iqbal S, Usman SM, Hussain I. Enhanced gastric cancer classification and quantification interpretable framework using digital histopathology images. Sci Rep. 2024;14. https://doi.org/10.1038/s41598-024-73823-9.
74. Zuluaga-Gomez J, Al Masry Z, Benaggoune K, Meraghni S, Zerhouni N. A CNN-based methodology for breast cancer diagnosis using thermal images. Comput Methods Biomech Biomed Eng Imaging Vis. 2020;9:131–45. https://doi.org/10.1080/21681163.2020.1824685.

Publisher's note