

Deep learning in gastric tissue diseases: a systematic review

Wanderson Gonçalves e Gonçalves ,^{1,2} Marcelo Henrique de Paula dos Santos,³ Fábio Manoel França Lobato,⁴ Ândrea Ribeiro-dos-Santos,^{1,2} Gilderlano Santana de Araújo ¹

To cite: Gonçalves WGe, Santos MHP, Lobato FMF, *et al.* Deep learning in gastric tissue diseases: a systematic review. *BMJ Open Gastro* 2020;**7**:e000371. doi:10.1136/bmjgast-2019-000371

Received 31 December 2019
Revised 14 February 2020
Accepted 24 February 2020

ABSTRACT

Background In recent years, deep learning has gained remarkable attention in medical image analysis due to its capacity to provide results comparable to specialists and, in some cases, surpass them. Despite the emergence of deep learning research on gastric tissues diseases, few intensive reviews are addressing this topic.

Method We performed a systematic review related to applications of deep learning in gastric tissue disease analysis by digital histology, endoscopy and radiology images.

Conclusions This review highlighted the high potential and shortcomings in deep learning research studies applied to gastric cancer, ulcer, gastritis and non-malignant diseases. Our results demonstrate the effectiveness of gastric tissue analysis by deep learning applications. Moreover, we also identified gaps of evaluation metrics, and image collection availability, therefore, impacting experimental reproducibility.

INTRODUCTION

Gastric cancer is the fourth most common tumour worldwide, and the second most malignant tumour with a higher mortality rate.^{1,2} Thus, an early and accurate diagnosis is critical for treatment effectiveness and morbidity and mortality rates reduction.^{3,4}

Glass microscope slide images have been mainly used for gastric cancer diagnosis, supporting the application of pharmacotherapies and neoadjuvant therapies.⁵ Digital pathology analysis is now considered one of the most promising fields in digital health due to advancements in precision medicine, imaging data analysis and robust computational methods.⁶

Among medical image analysis methods, deep learning has gained remarkable attention in the last decade for providing comparable results to humans in classification tasks.^{7,8} Essentially, this machine learning paradigm implements multilayer architectures based on artificial neural networks. It is possible that the first application of a deep learning system implemented a feedforward

Summary box

What is already known about this subject?

- ▶ Computer-assisted systems for health image analysis have improved the medical decision-making process for diagnosing and analysing the progression of various diseases.
- ▶ Diseases affecting gastric tissue are a worldwide health problem.
- ▶ Deep learning applications presented good results in different domains, however its application on gastric tissue analysis is recent, poorly analysed, and standardised.

What are the new findings?

- ▶ We provide a literature categorisation, based on the method and related tasks, identifying the most widely adopted deep learning architecture and data source used.
- ▶ This is the first systematic review dedicated to map gastric tissue deep learning applications covering a broad spectrum, also listing and evaluating open source tools.
- ▶ We identified gaps evaluation metrics, image collection availability and, consequently, implications for experimental reproducibility.

How might it impact on clinical practice in the foreseeable future?

- ▶ Deep learning applications can provide greater and more efficient workflow support and extraction of important information from histological images, consequently, replicable studies need to be conducted clearly, and transparently, also providing the data used.

multilayer perceptron, dating from 1965.⁹ Additionally, earlier works developed a single-layer hidden neural network.¹⁰ However, this machine learning subfield has long been limited to computational costs, amount of available data, and classical techniques limitations.⁸

Deep learning was highly reintegrated in research studies in approximately 2006 for handwriting classification by computational scientists assembled by the Canadian



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Laboratório de Genética Humana e Médica - Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, Pará, Brazil

²Núcleo de Pesquisas em Oncologia, Universidade Federal do Pará, Belém, Pará, Brazil

³Engenharia da Computação, Universidade Federal do Pará, Belém, Pará, Brazil

⁴Instituto de Engenharia e Geociências, Universidade Federal do Oeste do Pará, Santarém, Pará, Brazil

Correspondence to

Dr Gilderlano Santana de Araújo; gilderlano@gmail.com



Institute of Advanced Research,⁸ and it has been applied in innumerable fields such as economics,¹¹ healthcare,¹² environmental analysis, tracking land patterns changes in tropical forests,¹³ information security, network intrusion detection systems,¹⁴ computational biology, molecular expression inference,¹⁵ social networks, and online services recommendation.¹⁶

In this way, some studies highlight the emerging role of deep learning in precision medicine, suggesting the expansion of its utility for future practice in histopathology.^{17–19} The potential of deep learning paradigm has been widely explored as a diagnostic assistance tool, mainly due to the constant improvements in hardware and in statistical inference methods,²⁰ especially in the case of computer vision,^{21–23} which drives the life science community to use medical imaging.

Deep learning analysis showed superior accuracy over specialists in the classification of lung cancer as well as for melanoma by analysing prior computing tomography and skin photography, respectively.^{24 25}

However, deep learning applications are not intended to replace professionals who are directly responsible for the diagnosis, but they provide a more efficient workflow for extracting important information from histology images and on detecting patterns that are implicit to the human eye.²⁶ In addition, these computational tools are important for providing a significant impact on cost and error reduction for healthcare, which maximises reliability and diagnostic quality. Despite the emergence of deep learning research in gastric tissues, there are few intensive reviews addressing publications on this topic. To the best of our knowledge, the available reviews are restricted to machine learning applications in gastrointestinal endoscopy problems and do not provide a general overview of the scenario of deep learning in gastric diseases.^{20 27} Thus, the present study aims to reduce this gap by mapping, inspecting and discussing the literature content regarding the applications of deep learning methods based on gastric tissue medical images. To achieve our goal, we systematically structured and analysed data extracted from public deep learning studies to answer the following research questions:

- What deep learning applications have been reported gastric tissue diseases studies?
- What deep learning architectures and models were used in studies involving gastric tissue image analysis?
- How deep learning architectures and models have been evaluated?
- Which aspects are lacking regarding transparency and reproducibility of deep learning applications?

METHODS

We performed a systematic literature review of research studies regarding the application of deep learning in gastric tissue analysis. An overview of our systematic review process can be seen in [figure 1](#). We conducted this review based on the PICO mnemonic (Population, Intervention,

Context, Outcomes) to determine the review protocol. In this section, we describe the methods by which the works in our study were selected and analysed, pointing out the data sources and research strategies. The criteria adopted for studies inclusion and exclusion are also explained. Finally, we present how data extraction was performed.

Data sources and search strategy

Deep learning can be considered a quite new research topic. Thus, studies published at any time were included for data collection, as described in Section Inclusion and exclusion criteria.

The search terms were selected to find substantial number of papers that applied deep learning in gastric tissue. The search strategy consisted of automatic queries using a string comprised of two terms: ‘deep learning’ and ‘gastric’. The data sources included PubMed, Institute of Electrical and Electronics Engineers (IEEE), Association for Computing Machinery (ACM), and Digital Bibliography Library Project (DBLP). The retrieved papers’ titles and abstracts were analysed, and only those papers that met the inclusion criteria were further considered.

The titles and abstracts were individually analysed by two reviewers to be considered elected for the present study. After independent reviews regarding the inclusion and exclusion criteria, the reviewers agreed and were able to choose which papers would be considered in the study. Disagreements between reviewers were solved by consensus or, if necessary, a third reviewer was consulted.

Inclusion and exclusion criteria

Published studies were selected based on the inclusion criteria, which are defined as follows:

- Papers that present the terms ‘deep learning’ and ‘gastric’;
- Papers reporting the application of deep learning methods for gastric tissue analysis;
- Full papers.

The protocol excluded articles that met at least one of these eight exclusion criteria:

- Publications written in a language other than English;
- One-page publications (abstract works), posters, pre-sentences, procedures and scientific events programme;
- Publications that do not consider deep learning algorithms applied to gastric tissue;
- Theses, dissertations, and monographs;
- Literature reviews and surveys;
- Tutorial slides;
- Incomplete documents, drafts, presentation slides and extended summaries;
- Duplicate publications by the same authors, with similar titles, abstracts, the results, or text. In this case, only one research study remained for analysis.

Data extraction

Data extraction was performed separately through the evaluation of full-text articles by two independent reviewers.

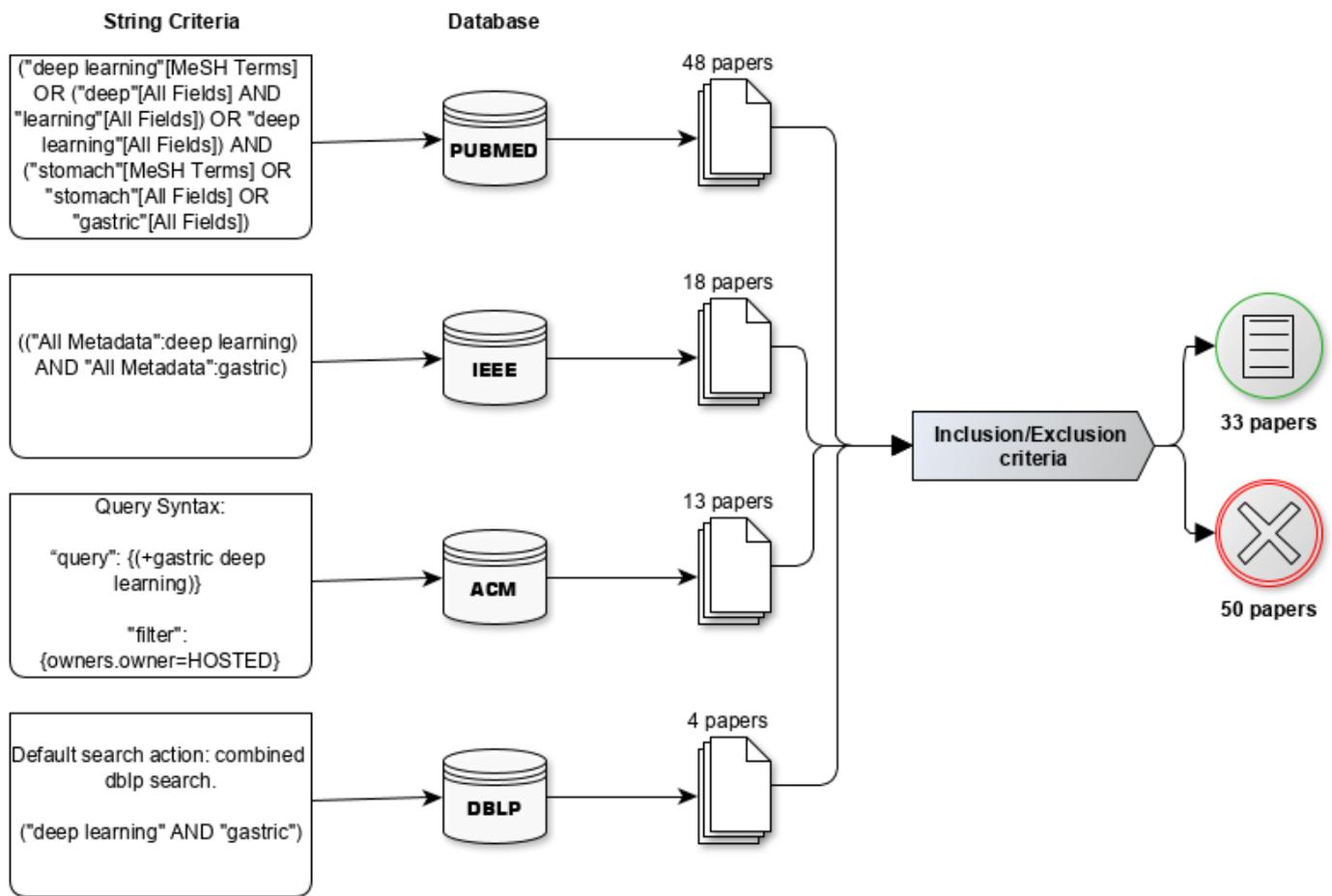


Figure 1 An overview of our systematic review process. In the first and second layer the application of query strings and their respective databases, respectively. In the third layer, the quantitative output of papers and at last the total number of papers after performing inclusion and exclusion criteria. ACM, Association for Computing Machinery; DBLP, Digital Bibliography Library Project; IEEE, Institute of Electrical and Electronics Engineers; MeSH, Medical Subject Headings.

The features described in [table 1](#) were collected for each research study and were organised in a spreadsheet.

RESULTS

The queries returned 83 articles in total, of which 48 were found in PubMed, 18 in IEEE, 13 in ACM and 4 in DBLP. By applying the inclusion and exclusion criteria, 33 papers remained for data extraction. Six studies were found in duplicate. Three studies were duplicated in PubMed and IEEE, and three studies were duplicated in IEEE and DBLP (see [figure 1](#)).

Descriptive information

As mentioned before, deep learning methods have made a major impact in the field of medical informatics. In 2015, applications for medical imaging analysis exceeded the number of research studies compared with other subareas, such as bioinformatics, pervasive sensing, medical informatics, and public health, according to Google Scholar's publication statistics.²⁸

Publication timeline

The first study was published by Malon *et al*²⁹ in 2008 at the beginning of the deep learning techniques revival. The authors demonstrate how convolutional neural networks

can be combined with histological image analysis to achieve higher accuracies regarding breast and gastric cancer classification tasks. The problems were counting mitotic figures in the breast, recognising epithelial layers in the stomach, and detecting signet ring cells. This was possibly the first use of a deep learning system for analysing gastric tissue histological images. In particular, from 2018 to August 2019 (see [figure 2](#)), we observed an increase in publications in deep learning associated with gastric tissue image data. This observed trend is possibly due to the interest of the scientific community in health areas and to the promising results of automatic diagnosis based on digital medical images.^{8 30}

Geographical distribution of papers

Papers are distributed in eight different countries in a non-uniform manner (see [figure 3](#)). Most of the papers catalogued in this review, approximately 70%, are concentrated in East Asian countries, namely, Japan, China, and South Korea, with publication ratios of 30%, 27%, and 12%, respectively. This high interest from Asian research centre in deep learning associated with gastric diseases corroborates with Globocan data,³¹ where gastric cancer incidence rates are markedly high in Eastern Asia. The



Table 1 Information extracted from the studies under scrutiny

Data	Description
Input data type	Initial data into the system
Data set	Information about the data set and its availability
Title	Title of the paper
Year	Publication paper year
Author	Authors of the paper
Publisher journal or conference	Where the paper was published
Network model	Deep learning model used in paper
Parameter	Parameters used in the deep learning model(s)
Evaluation metric	Evaluation metrics adopted in the paper
Software and library	Software and libraries described in the paper.
Computational cost	Time or computational resources for model compilation
Ethics committee	Presence or absence of the ethics committee in the paper
Research centre location	Country where the research was performed
Main result	Main results of the paper
Sample size	Sample size used in paper
Application	The applications of the model in paper

mentioned countries with the most publications have the highest rates of gastric cancer in the world for both genders.^{32 33}

Studies related to deep learning and gastric cancer

Characteristics, such as methods and applications for all studies, are summarised in table 2.

Out of 33 publications selected for analysis in this review, 15 articles related to histology, 11 were related to the application of deep learning to endoscopy images, immunohistochemistry or pathology images, and 5 publications applied deep learning methods to stomach X-ray

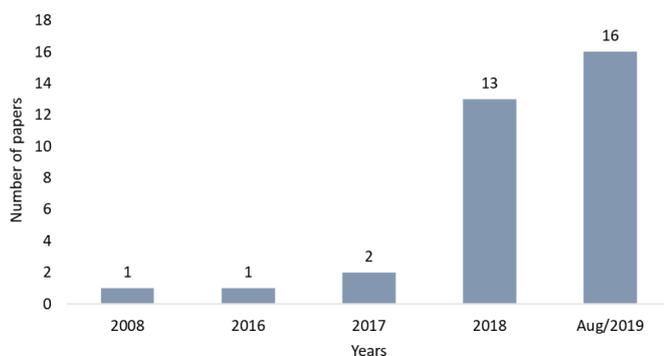


Figure 2 Number of papers over the years about deep learning applied to gastric tissue.

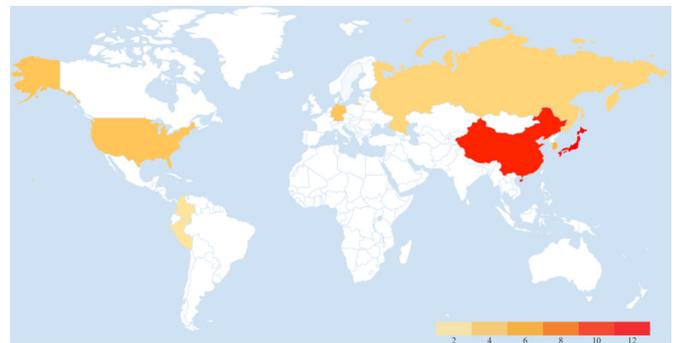


Figure 3 Global map presenting the distribution of deep learning papers associated with gastric tissue. The scale represents the number of papers per country, the redder more papers.

images. One article used CT images using deep learning, and another did not use images in their studies. The proportions of the data sources can be seen in figure 4.

There are few studies with direct application in gastric cancer, either for classification, detection or segmentation. Tumour heterogeneity between individuals and the difficulty of finding large databases of annotated medical images is a challenge for researchers. However, most studies addressed progress analysis of developing gastric cancer, such as classification and detection of gastritis,^{34–37} *Helicobacter pylori*,^{38–40} gastric neoplasms⁴¹ and gastric ulcers.^{42 43}

Two of the papers were restricted reviews regarding only images of endoscopy and addressed the effects of artificial intelligence on gastroenterology, and both were not included in the final database. Also, both reviews addressed challenges in the development of computer-aided diagnostic systems.^{20 27}

Our review is distinguished regarding analysis of four different databases, specific to computing and health-care, involving all research studies related to deep learning and gastric tissue with no time limitation.

Most used models

As mentioned in the previous section, we found that deep learning models were built under training with histological (45%) or endoscopic (33%) images (see figure 4). The convolutional neural network (CNN) is a widely used and state-of-the-art model. Approximately 76% of articles proposed applications of CNN for the diagnosis of gastric cancer or the detection of lesions in gastric tissue. This is a noteworthy consequence of CNN performance over other traditional machine learning methods that require human expertise for feature extraction.⁴⁴ CNNs are neural networks inspired by the process of human vision, in which the stimulus for image recognition is made by means of a region without considering a single point (pixel); that is, the spatial structure of the images and the dependence among their neighbours are considered. The CNN model belongs to the feedforward network class, in which the feed stream only follows one direction.

Table 2 Summary of the different deep learning methods by applications in gastric problems

Based method	Task	Application	Reference
Convolutional neural network	Classification	Benign or malign of stomach biopsy specimens	79
		Benign or malignant images	58
		Gastric cancer or non-cancer	32
			64
		Gastritis or non-gastritis	34
		<i>Helicobacter pylori-positive or helicobacter pylori-negative</i>	38
			39
		Helicobacter pylori-related gastritis, reactive gastropathy and histologically normal gastric mucosa	40
		Neoplasm or non-neoplasm	80
		Normal gastric images or early gastric cancer images	60
Convolutional neural network and residual neural network		Normal or abnormal gastric slow wave	61
		With and without histology-proven atrophic gastritis	35
Convolutional neural network and deep generalised multi-instance learning		Benign ulcer and gastric cancer	42
		Early gastric cancer, advanced gastric cancer, high grade dysplasia, low grade dysplasia or non-neoplasm	41
Convolutional neural network with deep reinforcement learning		Differentiation degree (poorly and well/moderately) and lauren type (intestinal, diffuse and mixed)	81
Residual neural network		Gastric sites	62
Recurrent neural network		Gastric cancer type (intestinal type or diffuse type)	52
		Microsatellite instable or microsatellite stability	53
Convolutional neural network	Classification/ detection	Live or dead probability	51
		Benign or malignant gastric ulcer/gastric ulcer	43
	Detection	HER2+ tumour, HER2 tumour or non-tumour/necrosis detection	5
		Gastric cancer	65
		Gastritis or non-gastritis	36
			37
		Lymphocyte or non-lymphocyte	66
		Normal mucosa, non-cancerous pathology, cancer	63
		Signet ring cell cancer	29
		Signet-ring cell carcinoma component intramucosal or advanced	82
Gastric ulcer	43		
Necrosis detection	5		
Generative adversarial network	Generation	Gastritis image generation	56
			57
Residual neural network	Segmentation	Gastric cancer	54
Fully convolutional network		Gastric tumour	48
		Recognise small cancerous tissues	49

The CNN layers are divided into input, feature extraction, activation and classification, which precedes the output layer. A graphical CNN model structure can be seen in [figure 5](#).

The input layer sets the original images of the problem to be solved on the neural network. The next step is to insert feature layers, which correspond to convolutional and grouping layers (or subsampling), both working

together using kernels or spatial filters and extracting important features.

The activation function makes the non-linear transformation in the input data, making it able to learn and perform more complex tasks. In this way, each layer becomes a specialist in identifying essential image features, such as colours, contours, curves, formats, and others that depend on the image application domain.

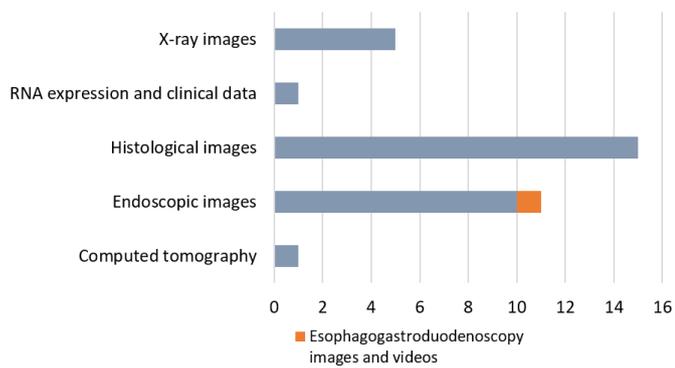


Figure 4 Number of publications by image data type.

The convolutional kernel values are learnt by the network in the training phase using the backpropagation algorithm.⁴⁵ This process generates known activation maps, which are produced by each layer and aid in the resource extraction process. Since the pooling filters do not have defined values, they are used as a logical operation to reduce the processing load of the network. These filters produce another image (matrix of pixels) with smaller dimensions than the previous filter, which can extract the maximum, average or minimum value of a pixel inside the filter that will be applied in the image.

After each layer, the original image dimensions are reduced within the network; these images are used to feed a neural network, usually a multilayer perceptron, to perform the classification of calculated resources in the previous layers.

In this way, a CNN can adjust the image data set by reducing parameters and reusing weights, enabling the successful capture of spatial and temporal image dependencies by applying relevant filters.⁴⁶

Other approaches used were fully connected networks (FCNs)⁴⁷ with two applications: segmentation of gastric tumours and gastric cancer;^{48,49} recurrent neural networks (RNNs)⁵⁰ with one application: predicting survival rate after gastric surgery;⁵¹ deep residual networks²² with three applications: prediction of microsatellite instability in the gastrointestinal cancer, gastric cancer segmentation, and classification of gastric cancer type (intestinal and diffuse type);^{52–54} and generative adversarial networks⁵⁵ with

one application: generation of synthetic and anonymous images of gastritis.^{56,57}

HETEROGENEITY ON DEEP LEARNING MODEL PARAMETERS

The prediction task of tumour (malignant or non-malignant) are the most abundant in literature, followed by detection and segmentation tasks (see table 2). More than 80% of the papers use convolutional neural networks with hyperparameter adjustments, and better model performance were achieved by combining with other algorithms.

Real-world applications, especially in the medical field, often result in a deficit of training data, mainly due to difficulties in obtaining well-annotated data.^{34,35,52,56–59}

This fact impacts model training and consequently, model reconstruction. Most studies applied alternative methods such as transfer learning, secondary training and fine-tuning to overcome the lack of insufficiency of image samples.^{35,38,39,41–43,52,53,58,60–63}

Transfer learning allows a new classification task to be performed on a pre-trained model.^{52,59} This method can make it possible to train a neural network and perform a classification task on a very small data set. Therefore, it is possible to perform secondary training or fine-tune the weights learnt from new data sets.⁶¹

Secondary training consists freezing low-specificity weights from initial subset of layers from the neural network, while subsequent layers remain for data training. Then, new data are used to adjust and model fine-tune with high-specificity weights.⁶¹ Fine-tuning is the process of adjusting the parameters of the entire neural network to adjust the model for different applications.³⁷

Nakashima *et al*³⁹ used GoogleNet fine-tuning to diagnose *Helicobacter pylori* infection (*H. pylori*) and showed a promising diagnostic model with area under the curve (AUC) of 0.96 and 0.95. The authors used a pre-trained CNN model, trained with 1.2 million general images of the The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014 classification challenge, and 2000 training endoscopic images, including image rotation magnification. The images were bright blue and linked colour image laser image data.

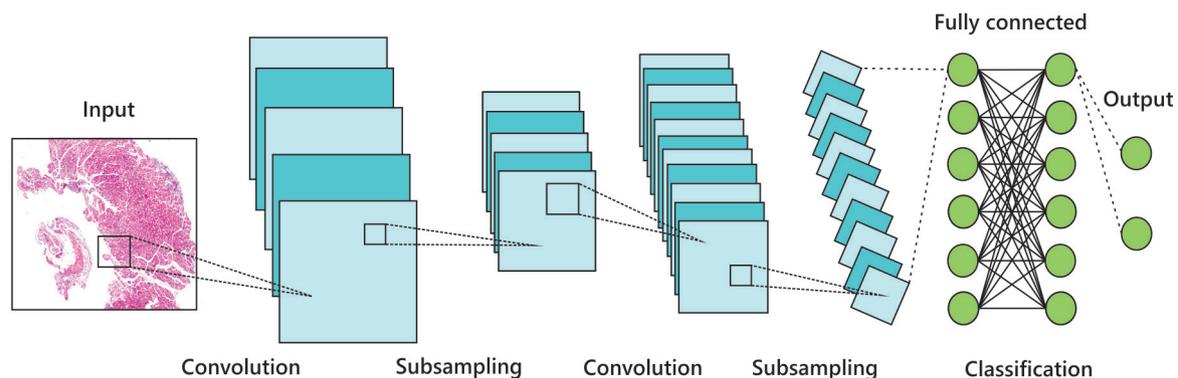


Figure 5 Typical convolutional neural network architecture.

Lee *et al*⁴² proposed detecting malignancies of gastric endoscopic images using deep learning by transfer learning. The authors used a data set consisting of 200 normal cases, 367 cancers and 220 ulcers, and applied the pre-trained ResNet and VGGNet models to ImageNet. A binary classification model, considering normal samples versus cancer cases and normal versus ulcers, resulted in accuracy above 90%. The case of ulcer versus cancer classification resulted in a lower accuracy of 77.1%. According to the authors, this result was possibly due to the smaller difference in image appearance when compared with a normal case.

Lee *et al*⁵¹ developed a deep learning-based model to predict survival rate after surgery. Although most studies used images as input data, Lee *et al*⁵¹ adopts clinical data to feed the deep neural network. The authors proposed a network based on a RNN called a survival recurrent network. This model uses time-sequential data only in the training step, and on being trained, it receives the initial data from the first visit and then sequentially predicts the outcome at each time point until it reaches 5 years. Training data were obtained from patients with gastric cancer by including all available clinical and pathological information and treatment regimens.

We also found a gastritis detection study in double-contrast upper gastrointestinal barium X-ray radiography images with sensitivity, specificity, and harmonic mean equal to 0.962, 0.983, and 0.972, respectively,³⁶ and one automatic segmentation study in histopathological images of gastric cancer. A neural network architecture combined with a new algorithm, known as overlap region prediction, resulted in an intersection over union coefficient of 0.883 and an average accuracy of 91.09% in the data set.⁴⁹

Among the new deep learning network, GastricNet⁶⁴ is a deep learning-based framework for gastric cancer detection. The GastricNet adopts different architectures for superficial and deep layers. This network showed superior results when compared with the model proposed by Liu *et al*.³² GastricNet achieved average classification accuracy of 100% based on image slices.

For automatic segmentation, GT-Net were proposed by Li *et al*⁴⁸ that is constructed under the BOT gastric slice data set, also used in both GastricNet⁶⁴ and Liu *et al*.³² GT-Net adopts different architectures for shallow and deep layers, multiscale module, characteristics pyramid and upsampling convolutional module to improve the extraction of resources. GT-Net performed better than next-generation networks such as FCN-8s and Unet, achieving a new F1-score of 90.88% and state-of-the-art gastric tumour segmentation.⁴⁸

Malignancy detection was conducted by Hirasawa *et al*,⁶⁵ which proposed a CNN based on the single-shot multibox detector architecture, by training the model with 13584 endoscopic images of gastric cancer. CNN were evaluated with 2296 test images in 47s and correctly diagnosed 71 of 77 gastric cancer lesions with an overall sensitivity of 92.2%. Approximately 161 non-cancerous

lesions were detected as gastric cancer, resulting in a positive predictive value of 30.6%. Exactly 70/71 lesions with a diameter of 6mm or more, as well as all invasive cancers were correctly detected. An important detail in this work is that single-shot multibox detector did not have changes in its original algorithm, evidencing a high potential of CNN models in different database types.

For automatic detection both CNN and RNN have been proposed. CNN model for automatic lymphocyte detection were constructed under immunohistochemical images of gastric cancer and achieved a precision of 96.88% regarding test sets.⁶⁶ In other way, Liu *et al*⁵² demonstrated a RNN for automatic detection of intestinal or diffuse gastric cancer by using gastric pathology images, with an F1-score of 96%.

In summary, a high potential for deep learning applications was observed associated with endoscopic, histological and X-ray images of gastric tissue combined with transfer learning, in contrast to traditional classification or detection algorithms.

MODEL EVALUATION

Model evaluation is a mandatory step in machine learning-based applications. Commonly, a model is trained and tested with distinct samples, and evaluation metrics are extracted for performance analysis. Accuracy is the predominant metric in selected studies. Although widely used, accuracy is not enough to model evaluation. Global accuracy may be predisposed to bias in training or test data, thus producing non-generalised models. In this way, other evaluation metrics are combined for a solid decision regarding the proposed model. To fill this gap, other metrics have been combined, such as recall, precision, sensitivity, F1 score, receiver operating characteristic and AUC curve analyses.

Additionally, both mean intersection over union (IoU) and average classification accuracy (ACA) have been applied for evaluation analysis.^{49 64} IoU is a performance measure commonly used in image segmentation. Basically, the IoU is defined as the size of the intersection divided by the union of the two image regions⁶⁷ and represents the similarity between a predicted region and a true region related to an object. Additionally, ACA is the overall correct classification rate of all the testing images.⁶⁴

OPEN SOURCE TOOLS AND THEIR APPLICATIONS

Several ready-made and optimised tools for deep learning project development are currently available in a variety of programming languages. Such solutions are mostly open source and have contributed significantly to the growing popularity of machine learning methods. Based on GitHub activity, among the most popular software and libraries, the most used in 2018 were TensorFlow,⁶⁸ followed by Keras,⁶⁹ PyTorch,⁷⁰ Caffe,⁷¹ and Theano⁷² (see figure 6).

Github Activity

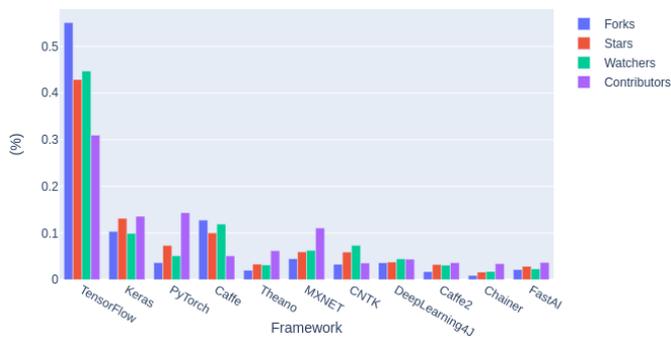


Figure 6 The most popular libraries used in 2018.

In this review, we found that 13 papers did not report their libraries (see figure 7). Keras were used in eight studies, Caffe were applied in six studies, TensorFlow were used in five, and PyTorch had one application. In addition, Theano was not applied directly in any article.

Below is a brief commentary about the five most popular libraries in 2018 and their main areas of application.

- 1. TensorFlow.** Developed by researchers and developers of the Google Brain team; it has pre-written codes for many of the complex deep learning models and has applicability mainly in textual data such as language detection, text summary and image recognition such as image caption, recognition and object detection, and sound recognition, time series analysis and video analysis.
- 2. Keras.** Written in Python, it can work with TensorFlow, Microsoft Cognitive Toolkit, Theano or PlaidML as the back end. Its main focus is its ease of use, it is modular and extensible and has well-known models of high-performance famous networks such as VGG16 and VGG19,⁷³ InceptionV3,⁷⁴ among others; therefore, it is more useful for learning transfer tasks.
- 3. PyTorch.** Elaborated by Facebook, it is more focussed on the area of academic research because it is more transparent and flexible. It is useful mainly in tasks with images (detection, classification, among others), text processing and learning by reinforcement.

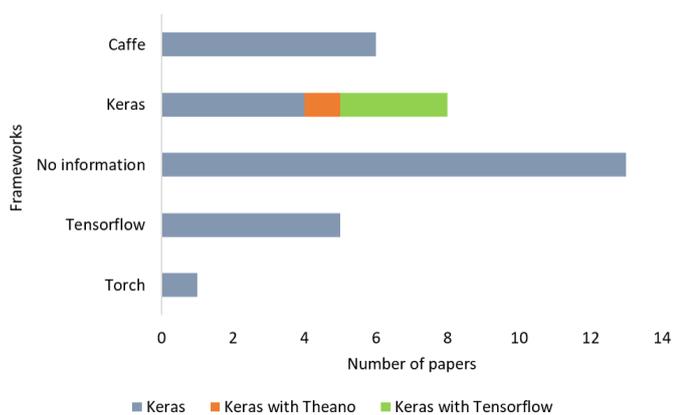


Figure 7 Frameworks used in the selected papers.

- 4. Caffe.** Facing the image processing field and therefore focussed on computer vision, it has a high processing speed compared with other solutions in imaging applications. It is basically applied in simple regression, large-scale classification of images, Siamese networks for image similarity, and speech and robotic applications.
- 5. Theano.** It is one of the oldest and most stable libraries available and low level, with applications in the most varied problems, such as image classification, object detection, chatbots, automatic translation, reinforcement agents or generator models.

A study conducted in 2016 on the performance of the most popular frameworks at the time, with the exception of Keras, can be found in more detail in.⁷⁵ In addition, a more complete list of frameworks can be found at (http://deeplearning.net/software_links/). In the case of programming languages, a variety of them can be used to solve machine learning problems, but some have greater applicability due to the optimisation of algorithms in regard to computational performance issues, such as memory management and the use of processors, as well as the ease of syntax that becomes an advantage.

THE LACK OF MODEL REPRODUCIBILITY

Transparency and reproducibility are two important aspects of all research medical studies,⁷⁶ moreover, both aspects have been neglected. The scientific process comprises not only a research paper as a product, but it must also be composed of technical documents, as well as scripts, and public and private databases might be accessible and clearly documented to provide an easy reproduction of results.

Most research teams build their own image collections, and a low number, less than 18%, of research studies constructed deep-learning models with public data. Most research studies restricted the use of their own data, being that only two databases could be accessed.

References or links were not provided, producing a gap for reproducibility and transparency in medical imaging studies. It is important to emphasise that accessible image collections have fundamental importance for the improvement of transparent models, allowing the reproducibility of the experiments. It is worth noting that the growth of deep learning applications requires a large volume and reliable image collections, and the construction of this mass of data allows to perform comparative model studies.

In addition to the lack of data sharing, there is a deficiency in most of the works regarding computational cost and hardware specifications, as well as libraries that were not reported by the aforementioned papers. For reproducibility, we indicate that studies in deep learning must report all model parameters, mainly regarding training, model test and evaluation.

CONCLUSION

The development of tools capable of leading to an accurate and efficient diagnosis based on artificial intelligence has been pursued in recent years. With the wide availability of graphical processing units, medical researchers are using machine learning approaches and have been achieving exciting results. Thus, we observed a marked increase in the number of publications related to deep learning applied to the gastric tissue in the last year.

The US Food and Drug Administration (FDA) published a discussion document that proposes a regulatory framework for modifications to software based on artificial intelligence/machine learning (AI/ML) as a medical device (SaMD).⁷⁷ The recent advances and developments in software based on artificial intelligence will imply methodological requirements that can provide safety, quality and efficiency of AI-based software as a final product to the patient. The document explores the proposal for a new regulatory approach to the total product life cycle, to promote a rapid cycle of continuous product improvements. Thus, it directs manufacturers to be constantly vigilant in the maintenance, safety, and effectiveness of their SaMD.

Two AI-based medical devices have already been approved by the FDA. One, the IDx-DR analyses images of the retina to detect diabetic retinopathy and the other, called Viz.AI analyses images in search of indicators of stroke.⁷⁸ Both algorithms are categorised as 'blocked algorithms'. These algorithms are not modified each time they are used and improvements are made over periods, using specific training data and an evaluation process that allows the product to function fully.

The performance of these 'blocked' medical devices is essential for decision-making and the choice of accurate and reliable diagnostics. However, it is a difficult task to evaluate the performance of products that learn and evolve on their own whenever they are presented with new real-world data. Applications with promising results are being developed for gastric disease diagnostics and in other domains of health. In this way, the FDA asks for effective, safe and reliable forms of regulation of these adaptable systems so that they can enter the market more quickly and improve medical care for the patient.

Deep learning applications have been reported for analyses of endoscopy, histology, immunohistochemistry or pathology, stomach X-ray, and CT images. Some studies have addressed classification or detection task in gastric cancer. Most of the papers addressed classification or detection of gastritis, *H. pylori*, gastric neoplasms, and gastric ulcers.

For real-world applications, especially in the medical field, there is often a shortage of training image collections sufficiently necessary for model reconstruction, mainly due to difficulties in obtaining well-annotated data. Thus, one of the most commonly used methods in the analysed works was transfer learning, secondary training, and fine-tuning, as well as comparison with the results of self-designed networks.

Additionally, as presented above, insufficient evaluation metrics are being used, and data sets vary widely. Also, there is a shortcoming in most work regarding computational costs, hardware specifications, and applied libraries. Data sets should be used with as many evaluation metrics as possible to allow fair comparisons. Therefore, while the results of studies have the potential for deep learning associated with various types of gastric tissue images, further studies may need to be clearly and transparently performed, with database availability and reproducibility, to develop applicable tools that assist health professionals.

Acknowledgements We would like to thank reviewers Rede de Pesquisa em Genômica Populacional Humana/Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES (Bio. Computacional, No. 3381/2013); Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

Contributors WGG, MHPS and GSA wrote the paper. WGG and GSA designed the research. WGG and MHPS collected the data. WGG, MHPS and GSA analysed the data. FL and ARS reviewed the paper. WGG, MHPS, FL, ARS and GSA agree with manuscript results and conclusions.

Funding Fundação Amazônia Paraense de Amparo à Pesquisa – FAPESPA (No. 008/2017) and PROPESP/UFGA for the financial support and scholarships.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Wanderson Gonçalves e Gonçalves <http://orcid.org/0000-0003-0918-1728>
Gilderlano Santana de Araújo <http://orcid.org/0000-0001-9199-9419>

REFERENCES

- 1 Van Cutsem E, Sagaert X, Topal B, *et al.* Gastric cancer. *The Lancet* 2016;388:2654–64.
- 2 Jemal A, Bray F, Center MM, *et al.* Global cancer statistics. *CA Cancer J Clin* 2011;61:69–90.
- 3 Takahashi T, Saikawa Y, Kitagawa Y. Gastric cancer: current status of diagnosis and treatment. *Cancers* 2013;5:48–63.
- 4 Yada T, Yokoi C, Uemura N. The current state of diagnosis and treatment for early gastric cancer. *Diagn Ther Endosc* 2013;2013:1–9.
- 5 Sharma H, Zerbe N, Klempert I, *et al.* Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. *Comput Med Imaging Graph* 2017;61:2–13.
- 6 Bui MM, Asa SL, Pantanowitz L, *et al.* Manning, digital and computational pathology: bring the future into focus. *J Pathol Inform* 2019;10:10.
- 7 Hecht-Nielsen R. Theory of the backpropagation neural network. In: *Neural networks for perception*. Elsevier, 1992: 65–93.
- 8 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
- 9 Ivakhnenko A, Lapa VG. Cybernetic predicting devices, CCM information Corporation 1965.
- 10 Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;61:85–117.
- 11 Heaton JB, Polson NG, Witte JH. Deep learning for finance: deep portfolios. *Appl Stoch Models Bus Ind* 2017;33:3–12.



- 12 Miotto R, Wang F, Wang S, *et al.* Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform* 2018;19:1236–46.
- 13 Rakshit S, Debnath S, Mondal D. Identifying land patterns from satellite imagery in Amazon rainforest using deep learning. *arXiv* 2018.
- 14 Javaid A, Niyaz Q, Sun W, *et al.* A deep learning approach for network intrusion detection system, in: proceedings of the 9th EAI International Conference on bio-inspired information and communications technologies (formerly BIONETICS) 2016:21–6.
- 15 Chen Y, Li Y, Narayan R, *et al.* Gene expression inference with deep learning. *Bioinformatics* 2016;32:1832–9.
- 16 Wang H, Wang N, Yeung D-Y. Collaborative deep learning for recommender systems, in: proceedings of the 21th ACM SIGKDD International Conference on knowledge discovery and data mining. *ACM* 2015:1235–44.
- 17 Lee J-G, Jun S, Cho Y-W, *et al.* Deep learning in medical imaging: general overview. *Korean J Radiol* 2017;18:570–84.
- 18 Grapov D, Fahrman J, Wanichthanarak K, *et al.* Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine. *OMICS* 2018;22:630–6.
- 19 Nezhad MZ, Zhu D, Li X, *et al.* A deep feature selection approach for precision medicine, in: 2016 IEEE International Conference on bioinformatics and biomedicine (BIBM). *IEEE* 2016:501–6.
- 20 de Lange T, Halvorsen P, Riegler M. Methodology to develop machine learning algorithms to improve performance in gastrointestinal endoscopy. *World J Gastroenterol* 2018;24:5057–62.
- 21 Deng L, Yu D, *et al.* Deep learning: methods and applications, foundations and trends in signal processing 2014;7:197–387.
- 22 He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition, in: proceedings of the IEEE conference on computer vision and pattern recognition 2016:770–8.
- 23 Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks, in: advances in neural information processing systems 2012:1097–105.
- 24 Ardila D, Kiraly AP, Bharadwaj S, *et al.* End-To-End lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019;25:954–61.
- 25 Haenssle HA, Fink C, Schneiderbauer R, *et al.* Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* 2018;29:1836–42.
- 26 Robertson S, Azizpour H, Smith K, *et al.* Digital image analysis in breast pathology—from image processing techniques to artificial intelligence. *Transl Res* 2018;194:19–35.
- 27 Min JK, Kwak MS, Cha JM. Overview of deep learning in gastrointestinal endoscopy. *Gut Liver* 2019;13:388–93.
- 28 Ravi D, Wong C, Deligianni F, *et al.* Deep learning for health informatics. *IEEE J Biomed Health Inform* 2017;21:4–21.
- 29 Malon C, Miller M, Burger HC, *et al.* Identifying histological elements with convolutional neural networks, in: proceedings of the 5th International Conference on soft computing as transdisciplinary science and technology. *ACM*, 2008:450–6.
- 30 Hosny A, Parmar C, Quackenbush J, *et al.* Artificial intelligence in radiology. *Nat Rev Cancer* 2018;18:500–10.
- 31 Bray F, Ferlay J, Soerjomataram I, *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
- 32 Liu B, Zhang M, Guo T, *et al.* Classification of gastric slices based on deep learning and sparse representation, in: 2018 Chinese control and decision conference (CCDC), *IEEE* 2018:1825–9.
- 33 Torre LA, Siegel RL, Ward EM, *et al.* Global cancer incidence and mortality rates and trends—an update. *Cancer Epidemiol Biomarkers Prev* 2016;25:16–27.
- 34 Li Z, Togo R, Ogawa T, *et al.* Semi-supervised learning based on tri-training for gastritis classification using gastric X-ray images, in: 2019 IEEE International Symposium on circuits and systems (ISCAS), *IEEE* 2019:1–5.
- 35 Guimarães P, Keller A, Fehlmann T, *et al.* Deep-learning based detection of gastric precancerous conditions. *Gut* 2020;69:4–6.
- 36 Togo R, Yamamichi N, Mabe K, *et al.* Detection of gastritis by a deep convolutional neural network from double-contrast upper gastrointestinal barium X-ray radiography. *J Gastroenterol* 2019;54:321–9.
- 37 Kanai M, Togo R, Ogawa T, *et al.* Gastritis detection from gastric X-ray images via fine-tuning of patch-based deep convolutional neural network, in: 2019 IEEE International Conference on image processing (ICIP), *IEEE* 2019:1371–5.
- 38 Itoh T, Kawahira H, Nakashima H, *et al.* Deep learning analyzes *Helicobacter pylori* infection by upper gastrointestinal endoscopy images. *Endosc Int Open* 2018;6:E139–44.
- 39 Nakashima H, Kawahira H, Kawachi H, *et al.* Artificial intelligence diagnosis of *Helicobacter pylori* infection using blue laser imaging-bright and linked color imaging: a single-center prospective study. *Ann Gastroenterol* 2018;31:462.
- 40 Martin DR, Hanson JA, Gullapalli RR, *et al.* A deep learning convolutional neural network can recognize common patterns of injury in gastric pathology. *Arch Pathol Lab Med* 2020;144:370–8.
- 41 Cho B-J, Bang CS, Park SW, *et al.* Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. *Endoscopy* 2019;51:1121–9.
- 42 Lee JH, Kim YJ, Kim YW, *et al.* Spotting malignancies from gastric endoscopic images using deep learning. *Surg Endosc* 2019;33:3790–7.
- 43 Sun JY, Lee SW, Kang MC, *et al.* A novel gastric ulcer differentiation system using convolutional neural networks, in: 2018 IEEE 31st International Symposium on computer based medical systems (CBMS), *IEEE* 2018:351–6.
- 44 Wang J, Ma Y, Zhang L, *et al.* Deep learning for smart manufacturing: methods and applications. *J Manuf Syst* 2018;48:144–56.
- 45 Rumelhart DE, Hinton GE, Williams RJ, *et al.* Learning representations by back-propagating errors, cognitive modeling 1988;5:1.
- 46 Alom MZ, Taha TM, Yakopcic C, *et al.* A state-of-the-art survey on deep learning theory and architectures, electronics 2019;8:292.
- 47 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation, in: proceedings of the IEEE conference on computer vision and pattern recognition 2015:3431–40.
- 48 Li Y, Xie X, Liu S, *et al.* Gt-net: a deep learning network for gastric tumor diagnosis, in: 2018 IEEE 30th International Conference on tools with artificial intelligence (ICTAI), *IEEE* 2018:20–4.
- 49 Liang Q, Nan Y, Coppola G, *et al.* Weakly supervised biomedical image segmentation by reiterative learning, *IEEE Journal of biomedical and health informatics* 2018;23:1205–14.
- 50 Graves A, Liwicki M, Fernández S, *et al.* A novel connectionist system for unconstrained handwriting recognition. *IEEE Trans Pattern Anal Mach Intell* 2008;31:855–68.
- 51 Lee J, An JY, Choi MG, *et al.* Deep learning-based survival analysis identified associations between molecular subtype and optimal adjuvant treatment of patients with gastric cancer. *JCO clinical cancer informatics* 2018;2:1–14.
- 52 Liu B, Yao K, Huang M, *et al.* Gastric pathology image recognition based on deep residual networks, in: 2018 IEEE 42nd annual computer software and applications conference (COMPSAC), volume 2, *IEEE* 2018:408–12.
- 53 Kather JN, Pearson AT, Halama N, *et al.* Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* 2019;25:1054–6.
- 54 Sun M, Zhang G, Dang H, *et al.* Accurate gastric cancer segmentation in digital pathology images using deformable convolution and multi-scale embedding networks, *IEEE access* 2019.
- 55 Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets, in: advances in neural information processing systems 2014:2672–80.
- 56 Togo R, Ishihara K, Ogawa T, *et al.* Anonymous gastritis image generation via adversarial learning from gastric X-ray images, in: 2018 25th IEEE International Conference on image processing (ICIP), *IEEE* 2018:2082–6.
- 57 Togo R, Ogawa T, Haseyama M. Synthetic gastritis image generation via loss function-based conditional pggan, *IEEE access* 2019;7:87448–57.
- 58 Qu J, Hiruta N, Terai K, *et al.* Gastric pathology image classification using stepwise fine-tuning for deep neural networks. *J Healthc Eng* 2018;2018:1–13.
- 59 Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2009;22:1345–59.
- 60 Liu X, Wang C, Hu Y, *et al.* Transfer learning with convolutional neural network for early gastric cancer classification on magnifying narrow-band imaging images, in: 2018 25th IEEE International Conference on image processing (ICIP), *IEEE* 2018:1388–92.
- 61 Agrusa AS, Gharibans AA, Allegra A, *et al.* A deep convolutional neural network approach to classify normal and abnormal gastric slow wave initiation from the high resolution electrogastrogram, *IEEE transactions on biomedical engineering* 2019.
- 62 Wu L, Zhang J, Zhou W, *et al.* Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. *Gut* 2019;68:2161–9.
- 63 Khrushchev V, Stepanova O, Lebedev A, *et al.* Deep learning for gastric pathology detection in endoscopic images, in: proceedings of the 2019 3rd International Conference on graphics and signal processing, *ACM* 2019:90–4.

- 64 Li Y, Li X, Xie X, *et al.* Deep learning based gastric cancer identification, in: 2018 IEEE 15th International Symposium on biomedical imaging (ISBI 2018). *IEEE* 2018:182–5.
- 65 Hirasawa T, Aoyama K, Tanimoto T, *et al.* Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* 2018;21:653–60.
- 66 Garcia E, Hermoza R, Castanon CB, *et al.* Automatic lymphocyte detection on gastric cancer IHC images using deep learning, in: 2017 IEEE 30th International Symposium on computer-based medical systems (CBMS), IEEE 2017:200–4.
- 67 Rahman MA, Wang Y. Optimizing intersection-over-union in deep neural networks for image segmentation, in: International Symposium on visual computing, Springer 2016:234–44.
- 68 Abadi M, Barham P, Chen J, *et al.* Tensorflow: A system for large-scale machine learning, in: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI}) 2016:265–83.
- 69 Chollet F. Keras: deep learning library for theano and tensorflow, 2015. Available: <https://keras.io/>
- 70 Paszke A, Gross S, Chintala S, *et al.* Automatic differentiation in pytorch 2017.
- 71 Jia Y, Shelhamer E, Donahue J, *et al.* Caffe: Convolutional architecture for fast feature embedding, in: proceedings of the 22nd ACM International Conference on multimedia, ACM 2014:675–8.
- 72 Bergstra J, Breuleux O, Bastien F, *et al.* Theano: a CPU and GpU math expression compiler, in: proceedings of the python for scientific computing conference (SciPy), volume 4, Austin, tx 2010.
- 73 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014.
- 74 Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the inception architecture for computer vision, in: proceedings of the IEEE conference on computer vision and pattern recognition 2016:2818–26.
- 75 Bahrampour S, Ramakrishnan N, Schott L, *et al.* Comparative study of caffe, neon, theano, and torch for deep learning 2016.
- 76 Buck S. Solving reproducibility. *Science* 2015;348:1403.
- 77 FDA US. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) 2019.
- 78 Ratner M. FDA backs clinician-free AI imaging diagnostic tools. *Nat Biotechnol* 2018;36:673–4.
- 79 Leon F, Gelvez M, Jaimes Z, *et al.* Supervised classification of histopathological images using convolutional neuronal networks for gastric cancer detection, in: 2019 XXII Symposium on image, signal processing and artificial vision (STSIVA). *IEEE* 2019:1–5.
- 80 Radi K, Sergey S, Vyacheslav M. The real time endoscopic image analysis algorithm, in: proceedings of the fifth International Conference on network, communication and computing, ACM 2016:73–7.
- 81 Fang M, Zhang W, Dong D, *et al.* Predicting histopathological findings of gastric cancer via deep generalized multi-instance learning. *International Society for Optics and Photonics* 2019.
- 82 Mori H, Miwa H. A histopathologic feature of the behavior of gastric signet-ring cell carcinoma; an image analysis study with deep learning. *Pathol Int* 2019;69:437–9.