

An Artificial Intelligence Pipeline for Diagnosing Hepatocellular Carcinoma Patients with Bile Duct Tumor Thrombus

Jinming Liu

University of Science and Technology of China

Jiayi Wu

Fujian Provincial Hospital

Anran Liu

University of Science and Technology of China

Yannan Bai

Fujian Provincial Hospital

Hong Zhang (✉ zhangh@ustc.edu.cn)

University of Science and Technology of China

Maolin Yan

Fujian Provincial Hospital

Research Article

Keywords: Hepatocellular carcinoma, Bile duct tumor thrombus, Object detection, Artificial intelligence

Posted Date: December 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1126021/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

An artificial intelligence pipeline for diagnosing hepatocellular carcinoma patients with bile duct tumor thrombus

Jinming Liu^{a,1}, Jiayi Wu^{b,c,1}, Anran Liu^a, Yannan Bai^{b,c}, Hong Zhang^{a,*}, Maolin Yan^{b,c,*}

^a*Department of Statistics and Finance, School of Management, University of Science and Technology of China, 96 Jinzhai Road, Hefei, 230026, Anhui, P. R. China*

^b*Shengli Clinical Medical College of Fujian Medical University, Fuzhou, Fujian, P. R. China*

^c*Department of Hepatobiliary Pancreatic Surgery, Fujian Provincial Hospital, 134 Dongjie Road, Fuzhou, 350001, Fujian, P. R. China*

Abstract

Background and purpose: Preoperative diagnosis of bile duct tumor thrombus (BDTT) is clinically important as the surgical prognosis of hepatocellular carcinoma (HCC) patients with BDTT is significantly different from that of patients without BDTT. The preoperative diagnosis of BDTT is usually based on identification of dilated bile ducts (DBDs) on medical images (eg., CT and MRI images). However, it is easy for doctors to ignore DBDs when reporting the imaging scan result, leading to a high misdiagnosis rate in practice. The aim of the present study was to develop an artificial intelligence (AI) pipeline for diagnosing HCC patients with BDTT using medical images.

Methods: The proposed AI pipeline included two stages. First, the object detection neural network Faster R-CNN was adopted to identify DBDs; then, an HCC patient was diagnosed to have BDTT if the proportion of images with at least one identified DBD exceeds some threshold value. Four-fold cross validation was used to evaluate the performance of the proposed AI pipeline.

Results: The proposed AI pipeline was applied on a real dataset consisting of CT images collected from 34 HCC patients (16 with BDTT and 18 without BDTT). The average true positive rate for identifying DBDs per patient was 0.92, while the patient-level true positive rate for diagnosing BDTT was 0.94. The AUC value of patient-level diagnosis of BDTT was 0.92 (95% CI: 0.83, 1.00), compared with 0.71

*Corresponding authors

Email addresses: zhangh@ustc.edu.cn (Hong Zhang), yanmaolin74@163.com (Maolin Yan)

¹These authors contributed equally to this work.

(95% CI: 0.52, 0.89) by random forest.

Conclusions: This study first proposes an AI pipeline to identify DBDs and diagnose BDTT, and the high accuracies demonstrate that it is successful in the diagnosis of BDTT.

Keywords: Hepatocellular carcinoma, Bile duct tumor thrombus, Object detection, Artificial intelligence

1. Introduction

In 1947, Mallory et al. reported the first hepatocellular carcinoma (HCC) patient with bile duct tumor thrombus (BDTT) who had a typical symptom of obstructive jaundice [1]. The clinical incidence of BDTT was shown to be about 0.5%-12.9% among HCC patients [2, 4] (0.5%-9.0% as reported in [3]; 1.2%-12.9% as reported in [1]). HCC patients with BDTT were shown to have worse prognosis than HCC patients without BDTT after liver resection or liver transplantation [1, 2, 3, 4, 5, 6, 7, 8]. Specifically, HCC patients with BDTT were shown to have a higher recurrence rate [4, 5, 6] (eg., one year recurrence rates was estimated to be 70.3% and 34.8% for HCC patients with and without BDTT, respectively, as reported in [6]) and a lower overall survival rate (Table 1) [1, 2, 3, 4, 6, 7, 8]. Early diagnosis is important in improving the prognosis of such HCC patients as it can help doctors make better preoperative decisions to remove the residual tumors in the bile duct as much as possible [9]. Currently, the diagnosis and evaluation of HCC with BDTT before surgery mainly depend on clinical symptom judgment and image examination [2, 5]. However, most HCC patients with BDTT have no specific clinical symptoms at early stages [10]. Furthermore, obstructive jaundice caused by BDTT's invasion into common hepatic duct or common bile duct can be easily misdiagnosed to be a symptom of biliary stones or cholangiocarcinoma [1, 10, 11, 12].

With the development of medical imaging technology, CT scan and MRI scan have been widely accepted as safe and valuable methods for diagnosing BDTT. Although the imaging features of BDTT and their correlations with the corresponding histopathologic manifestations have been reported in the literature [10, 12], it is still hard for doctors to identify BDTT on medical images by naked eyes. Many clinical studies showed that the general clinical feature of HCC with BDTT is the appearance of intrahepatic dilated bile ducts (DBDs) [10, 13]. Based on this fact, doctors can diagnose BDTT through identifying DBDs near intrahepatic tumors. DBDs caused by BDTT usually present dark linear structures on CT images (Figure 1(a)). DBDs only occupy a small part of the CT image, and these inconspicuous structures tend

Table 1 Overall survival rates of HCC patients with and without BDTT

| Reference | HCC with BDTT | | | HCC without BDTT | | |
|---------------------|---------------|--------|--------|------------------|--------|--------|
| | 1 year | 3 year | 5 year | 1 year | 3 year | 5 year |
| Lu et al. [2] | 77.0% | 42.0% | 23.0% | 80.0% | 60.0% | 48.0% |
| Wang et al. [4] | 81.8% | 50.0% | 37.5% | 90.9% | 66.9% | 55.9% |
| Shao et al. [6] | 70.3% | 25.9% | 7.4% | 90.6% | 54.0% | 37.7% |
| Rammohan et al. [7] | 82.0% | 48.0% | 10.0% | 90.0% | 55.0% | 38.0% |
| Shiomi et al. [8] | - | 46.6% | 28.0% | - | 63.2% | 47.5% |

to be easily ignored if doctors are lack of sufficient awareness of their appearance, resulting in a high misdiagnosis rate in practice. To our best knowledge, no automatic method has been developed to identify DBDs, despite its importance in accurate preoperative diagnosis of BDTT.

The advent of the era of big data and the improvement of computer hardware have promoted the development of artificial intelligence (AI), which enables mining disease information in medical images such as CT image and MRI image. Many AI algorithms have been developed to effectively utilize the disease information contained in the medical images. Among these AI methods, neural network methods have shown their competitiveness in many fields. As one of the most representative deep learning methods, convolutional neural networks (CNNs) have been widely applied in image recognition, providing state-of-the-art results in the fields of image classification, semantic segmentation, object detection, and so on [14]. Based on CNNs, more advanced neural networks have been proposed such as fully convolution network (FCN) [15], generative adversarial network (GAN) [16], and variational auto-encoder (VAE) [17], and so on. Many researchers applied these neural networks in medical fields and achieved promising results in identifying bone fracture [18] and tumor region [19], screening high-risk subjects of a specific disease [20], locating biomarkers [21], segmenting tissues and organs [22, 23, 24, 25], and so on.

Many neural network methods have been developed specifically for object detection. Compared with traditional classification methods, object detection methods can not only identify objects but also locate them. Some well-developed object detection neural networks include Faster R-CNN [26], YOLO [27], SSD [28], and so on. These object detection methods have been shown to have a broad application prospect in clinical diagnosis. For example, Thian et al. [29] applied Faster R-CNN to locate the fracture area on wrist radiographs, and achieved sensitivities as high

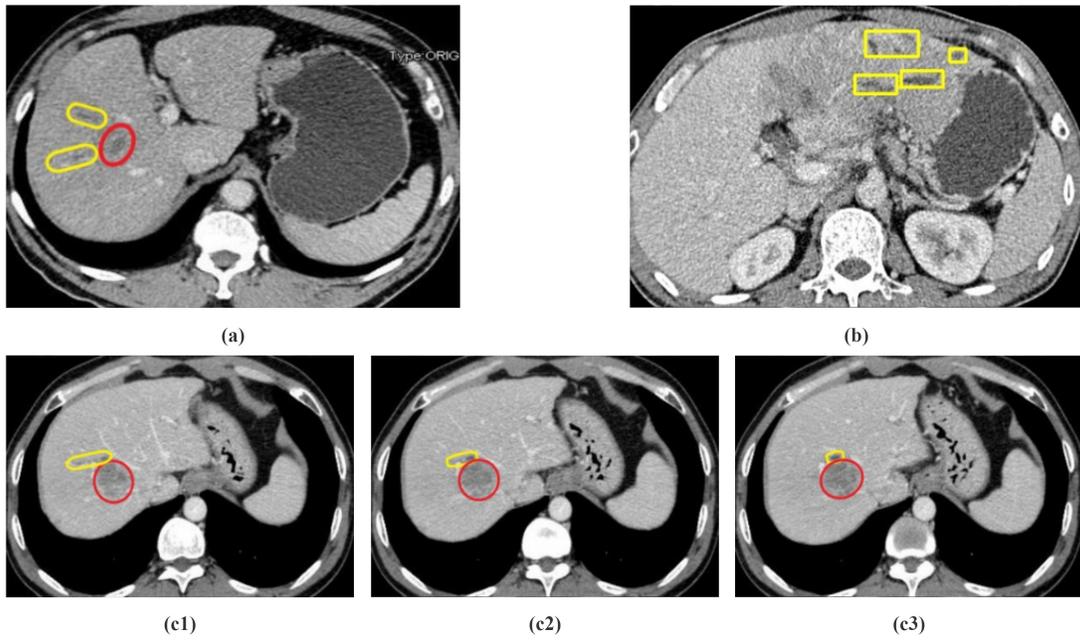


Fig. 1 (a) CT image of an HCC patient with BDTT (DBDs caused by BDTT were marked in capsules, tumor area was marked in an ellipse); (b) CT image with four labeled bounding boxes for DBDs; (c1)-(c3) Three consecutive CT images of an HCC patient with BDTT. In all images, tumors were marked in ellipses and DBDs caused by BDTT were marked in capsules

as 91.2% and 96.3% based on forward images and lateral images, respectively. Their object detection method was shown to outperform the traditional CNN methods such as VGG network used by Olczak et al. [30] and Inception V3 network used by Kim and MacKinnon [31]. Boot and Irshad [32] combined Faster R-CNN with Unet [22] to locate and segment breast tumors, and achieved a micro average F1 score of 0.805 and a macro average F1 score of 0.843.

The current study aimed to develop an AI pipeline for preoperative diagnosis of BDTT through identifying DBDs on CT images. To our best knowledge, this is the first computer-aided method developed for diagnosing BDTT. Based on a total number of 34 HCC patients, the developed pipeline was shown to have an area under ROC curve as high as 0.92, making it a powerful tool in diagnosing BDTT. The proposed AI pipeline can also locate the DBDs caused by BDTT.

2. Materials

CT images of 16 HCC patients with BDTT were retrospectively collected from four hospitals that are well known in diagnosis and treatment of hepatobiliary diseases in China (Fujian Provincial Hospital, $n = 3$; West China Hospital, Sichuan University, $n = 8$; Mengchao Hepatobiliary Hospital, Fujian Medical University, $n = 1$; The First Affiliated Hospital of Fujian Medical University, $n = 4$). As controls, 18 HCC patients without BDTT were randomly collected from Fujian Provincial Hospital. The present study was approved by the institutional review board of all relevant institutions, and informed consents were obtained from the patients or their guardians for their data to be used for research purposes. Throughout this paper, the 16 HCC patients with BDTT are denoted as case group and the 18 HCC patients without BDTT are denoted as control group.

All patients underwent preoperative serological examinations, preoperative imaging examination, collection of demographic information (eg., sex and age), test for hepatitis B virus infection and assessment of liver (reserve) function, surgical resection, postoperative histopathological, and immunohistochemical examinations. The diagnosis of HCC with BDTT was based on postoperative histopathological examination.

3. Results

Table 2 Demographic, clinical, and pathological variables of HCC patients

| | Case ($n = 16$) | Control ($n = 18$) | P value |
|---|-------------------|----------------------|---------|
| Sex (male/female) | 13/3 | 15/3 | 0.61 |
| Age (years) | 52.75 \pm 12.84 | 59.50 \pm 12.78 | 0.10 |
| HBsAg(-/+) | 0/16 | 4/14 | 0.07 |
| Liver cirrhosis (-/+) | 0/16 | 3/15 | 0.14 |
| Total bilirubin (μ mol/L) | 17.90 \pm 6.38 | 13.62 \pm 6.30 | 0.01* |
| Albumin (g/L) | 39.82 \pm 5.10 | 43.28 \pm 4.59 | 0.02* |
| ALT (U/L) | 60.81 \pm 41.55 | 45.39 \pm 61.19 | 0.02* |
| AST (U/L) | 85.50 \pm 69.64 | 47.67 \pm 68.65 | 0.00** |
| Child–Pugh class (A/B) | 15/1 | 18/0 | 0.47 |
| AFP (<400/ \geq 400 ng/mL) | 5/11 | 10/8 | 0.14 |
| Tumor number (single/ multiple) | 8/8 | 10/8 | 0.51 |
| Maximum tumor size (cm) | 8.23 \pm 5.54 | 6.02 \pm 4.19 | 0.12 |
| Microvascular invasion (-/+) | 7/9 | 5/13 | 0.91 |
| Edmondson–Steiner grade (I-II/III-IV) | 4/12 | 15/3 | 0.00** |
| The AJCC 8th Staging System (I-II/III-IV) | 5/11 | 11/7 | 0.08 |

HBsAg, hepatitis B surface antigen; ALT, alanine aminotransferase; AST, aspartate transaminase; AFP, α -fetoprotein; AJCC, American Joint Committee on Cancer.

*P < 0.05; **P < 0.01.

Table 2 summarizes the demographic, clinical, and pathological variables of the patients included in this study. The demographic variables included age and sex. It seems that sex was balanced between the two groups (P = 0.61). On the other hand, the control group was suggestively more aged than the case group (mean age: 52.75 vs. 59.50, P = 0.10). This suggests that potential confounding effects due to both sex and age were well controlled (the age effect is controlled if the age of the control group equals or is larger than the case group). As far as the clinical variables are concerned, the difference between the case group and the control group was significant for the total bilirubin level (P = 0.01), albumin level (P = 0.02), ALT level (P = 0.02), and AST level (P < 0.01). Specifically, compared with the control group, the patients in the case group had significantly higher serum levels of total bilirubin, ALT, and AST, while their serum albumin levels were significantly lower. The differences in the above clinical variables demonstrate a worse liver function of the HCC patients with BDTT. The pathological variables included microvascular invasion, Edmondson–Steiner grade, and the AJCC 8th Staging System. The difference between the

case group and the control group was significant for the Edmondson–Steiner grade ($P < 0.01$), which is a sign of poor differentiation of tumor cells. The patients in the case group had a suggestively more advanced AJCC cancer stages (stage III/IV) compared with the control group (68.8% versus 38.9%, $P = 0.08$).

After filtering and pre-processing, 1160 and 1451 CT images were retained for the case group and the control group, respectively. After labeling the CT images for the case group, there were totally 756 bounding boxes labeled on the 553 CT images, of which 366 (66.2%) containing one bounding box, 172 (31.1%) containing two bounding boxes, 14 (2.5%) containing three bounding boxes, and only 1 (0.2%) containing four bounding boxes (refer to Figure 1 (b)). The largest annotated bounding box only accounted for less than 6.0% of the whole image, suggesting that DBDs were inconspicuous and could be easily ignored by naked eyes.

Table 3 shows the DBD-level true positive rates (D-TPRs) for the 16 patients in the case group. The average D-TPR value was 0.92 and the range of D-TPR was 0.80~1.00, indicating a high success rate of detecting DBDs. Figure 2 shows several examples of successfully detected DBDs. In these examples, for most DBDs, Faster R-CNN output multiple bounding boxes that overlapped the DBDs with a high IoU value (refer to Appendix A of the Supplementary Materials for the definition of IoU).

Table 3 D-TPR of 16 patients in the case group (average D-TPR: 0.92)

| ID | tDBD | dDBD | D-TPR | ID | tDBD | dDBD | D-TPR |
|----|------|------|-------|----|------|------|-------|
| 1 | 5 | 5 | 1.00 | 9 | 18 | 15 | 0.83 |
| 2 | 10 | 8 | 0.80 | 10 | 11 | 11 | 1.00 |
| 3 | 11 | 11 | 1.00 | 11 | 4 | 4 | 1.00 |
| 4 | 11 | 9 | 0.82 | 12 | 18 | 17 | 0.94 |
| 5 | 10 | 9 | 0.90 | 13 | 9 | 8 | 0.89 |
| 6 | 5 | 5 | 1.00 | 14 | 8 | 7 | 0.88 |
| 7 | 7 | 7 | 1.00 | 15 | 17 | 14 | 0.82 |
| 8 | 7 | 7 | 1.00 | 16 | 16 | 13 | 0.81 |

tDBD: the number of true DBDs, dDBD: the number of successfully detected DBDs, $D-TRP = dDBD/tDBD$.

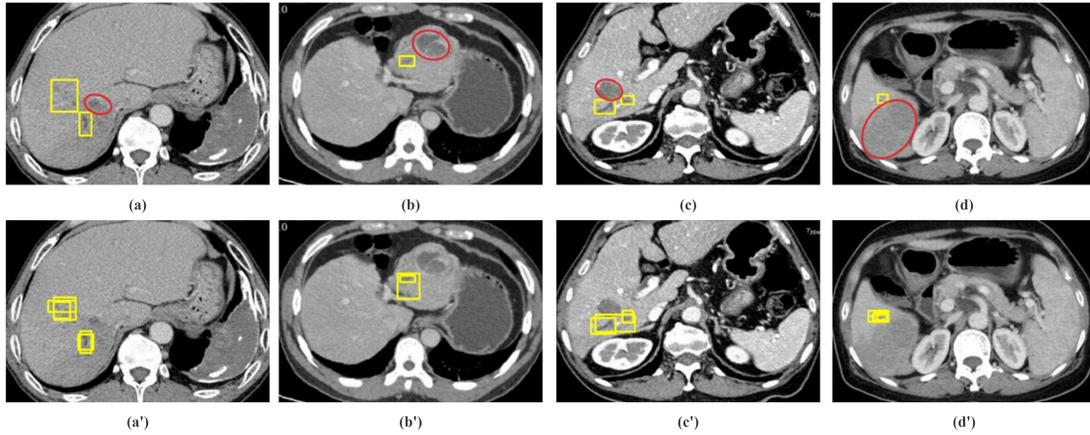


Fig. 2 Several examples of successfully detected DBDs: (a)-(d) were the ground truth annotations of CT images of four different HCC patients with BDTT (tumors were marked in ellipses and DBDs caused by BDTT were marked in rectangles); (a')-(d') were the corresponding bounding boxes output by Faster R-CNN

Table 4 shows the image-level positive rates (I-PPs) for both case group and control group. The average I-PP values were 0.78 and 0.50 for the case group and the control group, respectively. Most I-PPs for the case group were equal to or above 0.63 (93.8%), while the majority of I-PPs for the control group were below 0.63 (77.8%). The I-PPs of the patients are graphically displayed in Figure 3 (a) via box plots, which intuitively reflect the difference between the case group and the control group. Among those HCC patients with BDTT, only one patient had a relatively low I-PP of 0.55. The poor identification result for this patient could be due to his/her significantly worse image quality compared with other patients.

Table 4 I-PPs of patients in the case group and the control group

| Control group | | | | Case group | | | |
|---------------|------|----|------|---------------|------|----|------|
| ID | I-PP | ID | I-PP | ID | I-PP | ID | I-PP |
| 1 | 0.41 | 10 | 0.54 | 1 | 0.95 | 9 | 0.64 |
| 2 | 0.58 | 11 | 0.43 | 2 | 0.76 | 10 | 0.63 |
| 3 | 0.72 | 12 | 0.60 | 3 | 0.88 | 11 | 0.63 |
| 4 | 0.62 | 13 | 0.33 | 4 | 0.85 | 12 | 0.81 |
| 5 | 0.55 | 14 | 0.64 | 5 | 0.68 | 13 | 0.69 |
| 6 | 0.43 | 15 | 0.20 | 6 | 1.00 | 14 | 0.88 |
| 7 | 0.22 | 16 | 0.68 | 7 | 0.55 | 15 | 0.81 |
| 8 | 0.32 | 17 | 0.54 | 8 | 1.00 | 16 | 0.72 |
| 9 | 0.60 | 18 | 0.66 | | | | |
| Average: 0.50 | | | | Average: 0.78 | | | |

ROC curves for the proposed pipeline and the random forest algorithm are displayed in Figure 3 (b). For the proposed pipeline, the AUC value was 0.92 (95% CI: 0.83, 1.00) and the optimal I-PP threshold value was 0.63, with the corresponding sensitivity and specificity being 0.94 and 0.78, respectively. For random forest, the AUC value was 0.71 (95% CI: 0.52, 0.89) and the optimal I-PP threshold value was 0.59, with the corresponding sensitivity and specificity being 0.63 and 0.83, respectively. The AUC of the proposed pipeline was significantly larger than the random forest algorithm ($P = 0.02$).

4. Discussion

In this study, an AI pipeline was developed for accurate preoperative diagnosis of HCC patients with BDTT through automatically identifying DBDs on CT images by the object detection neural network Faster R-CNN. Only CT images of 16 HCC patients with BDTT were used in this study due to the rarity of BDTT and the difficulty in obtaining complete imaging data of HCC patients with BDTT. The current study could be the first attempt to applying object detection method to identify DBDs.

Analysis results of the multicenter data demonstrates that the trained Faster R-CNN had a promising performance in identifying DBDs. As shown in Table 3, for each patient in the case group, most of their DBDs were successfully detected and nearly half (43.8%) of the patients' DBDs were fully identified. In addition, the bounding boxes output by Faster R-CNN were quite precise (refer to Figure 2). From

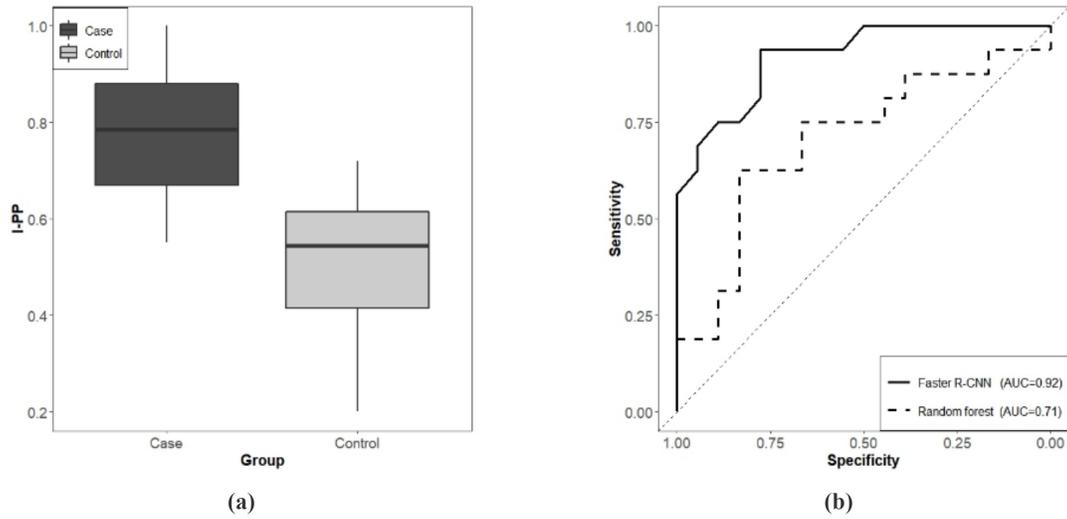


Fig. 3 (a) Box plots of image-level positive proportions (I-PPs) for the case group and the control group. (b) ROC curves of two methods for BDTT diagnosis (solid line: Faster R-CNN using CT images; dashed line: random forest using clinical data). The AUCs were 0.92 and 0.71 for Faster R-CNN and random forest, respectively. The sensitivities corresponding to the optimal threshold values were 0.94 for Faster R-CNN and 0.63 for random forest, respectively, and the specificities corresponding to the optimal threshold values were 0.78 for Faster R-CNN and 0.83 for random forest, respectively

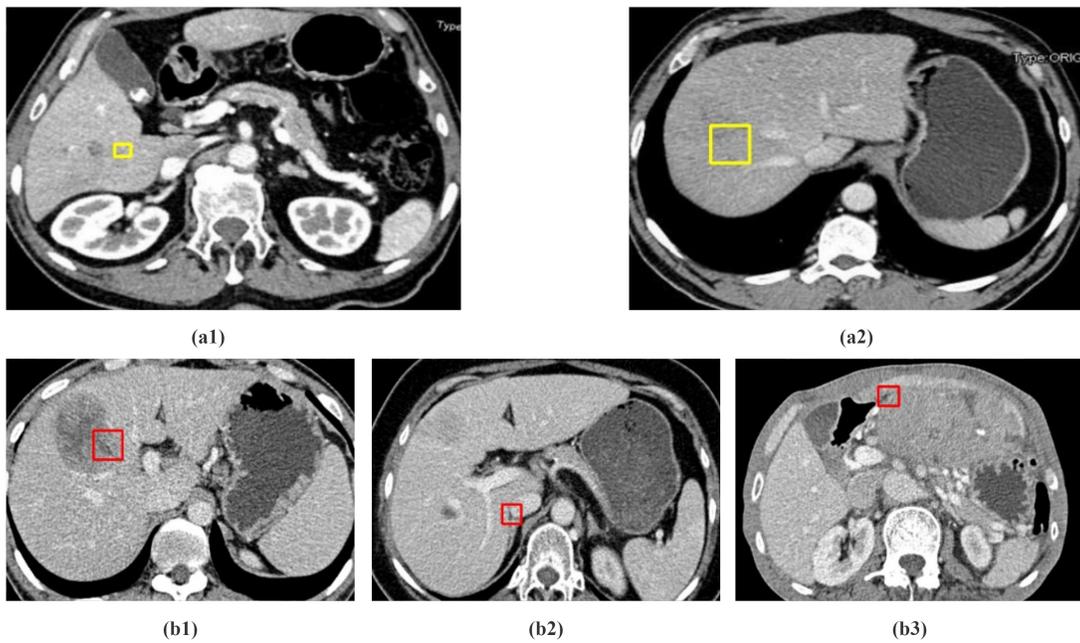


Fig. 4 The first row shows two samples of undetected DBDs: (a1) tiny DBD, (a2) inconspicuous DBD. The second row shows three types of dominant false positive bounding boxes: (b1) DBD-like liver tumor region, (b2) gap between liver and other tissues, (b3) irrelevant structures outside the liver region

the DBD identification result, false negative results (undetected DBDs) arose when their sizes were too small (Figure 4 (a1)) or they were too inconspicuous (Figure 4 (a2)), while the false positive results were mainly due to some confusing structures such as DBD-like tumor region (Figure 4 (b1)), gap at the junction between liver and other tissues (Figure 4 (b2)), and irrelevant structures outside the liver region (Figure 4 (b3)). Nevertheless, most of the false positive findings such as those shown in Figure 4 (b2)-(b3) were easy to distinguish by doctors.

A significant difference of identified DBD proportions (I-PPs) was observed between the case group and the control group (Table 4 and Figure 3 (a)), in accordance with the conclusion drawn in the literature [13]. This motivated a pipeline for diagnosing HCC patients with BDTT (Figure 5). The proposed pipeline was shown to significantly outperform the traditional machine learning method random forest that did not utilize image data (AUCs: 0.92 vs. 0.71), demonstrating the superiority of the proposed pipeline.

The attractive performance of the proposed pipeline could be attributed to two aspects. On one hand, the two-stage anchor-based object detection neural network Faster R-CNN was applied to identify DBDs, which first output object proposals then extracted regional features based on these proposals for detecting DBDs. Such design philosophy of looking and thinking twice is consistent with human vision system and it can result in a higher detection accuracy than one-stage networks [36]. On the other hand, the CT images were first manually filtered and processed then labeled by an experienced doctor. The heterogeneity of CT images from different centers was eliminated as much as possible, so that the CT image difference between the case group and the control group was mainly due to the presence or absence of DBDs. Furthermore, the DBD positions were accurately located, providing a prerequisite for training Faster R-CNN.

Although postoperative histopathological review of surgical specimen remains the golden standard for diagnosing and evaluating BDTT, it is of great significance to determine the existence of BDTT (and their locations if any) before surgical resection. This motivated us to use object detection neural network instead of the conventional classification neural networks. The advantage of using the object detection method is obvious, that is, it is difficult for the classification neural networks to focus on such inconspicuous structure of DBDs, while the object detection method not only identify which images have DBDs but also frame the DBD positions. Since the DBD positions are informative for typing and positioning BDTT, the proposed pipeline also provide useful information for doctors to evaluate the condition of BDTT.

The current AI pipeline developed in this study can potentially be improved from several aspects. First, since the scope of object detection is the whole image, irrele-

vant structures outside the liver region can become a source of false positive findings. Accordingly, the false positive rate can be effectively reduced by segmenting liver region from CT images in advance. However, in order to train the organ segmentation neural network such as Unet [22], pixel-level labels are required but it is much more difficult to obtain than bounding box labels. Second, individual images in portal venous phase were treated as the research objects in this study, which means that the correlation between the CT images was not utilized. Therefore, the DBD identification accuracy could be improved by mining the correlation among the sequential images in the portal venous phase. Third, the proposed pipeline only utilizes CT images of the HCC patients, it can be combined with other clinical evidences to potentially improve the diagnosis of BDTT. Finally, low-quality images may lower the detection accuracy, as demonstrated in the previous section. Consequently, the proposed pipeline can potentially be improved by using higher quality images.

5. Conclusion

In summary, the present study demonstrates the feasibility of AI-assisted preoperative diagnosis of BDTT through the identification of DBDs on CT images. The proposed AI pipeline achieves an average true positive rate for identifying DBDs per patient of 0.92, while achieving a patient-level true positive rate for diagnosing BDTT of 0.94. It is noteworthy that the AUC value of patient-level diagnosis of BDTT achieved by the proposed method was 0.92 (95% CI: 0.83, 1.00), compared with 0.71 (95% CI: 0.52, 0.89) achieved by random forest. Furthermore, the accurate locations of DBDs located by trained Faster R-CNN can help doctors evaluate the condition of BDTT.

The high accuracies make our AI pipeline a powerful tool in the diagnosis of BDTT patients and locating DBDs caused by BDTT. The pipeline proposed in this study can be potentially extended in other medical applications such as diagnosis of lymph node enlargement and so on.

6. Methods

6.1. Descriptive statistical analysis

Continuous variables were expressed as mean \pm standard deviation and the categorical variables were expressed as their numbers. The Wilcoxon-Mann-Whitney test and the Fisher exact test were adopted to compare continuous variables and categorical variables, respectively, between the case group and the control group. The DeLong test implemented in the R package pROC [33] was adopted to compare

the AUC values between two ROC curves. All the statistical analysis was performed using R version 4.0.3 [34]. $P < 0.05$ was considered statistically significant.

6.2. The proposed BDTT diagnosis pipeline

The proposed AI pipeline for diagnosing BDTT is displayed in Figure 5, which consists of three steps: 1) image filtering, pre-processing, and label making (Section 6.3), 2) DBD identification (Section 6.4), and 3) patient-level diagnosis of BDTT (Section 6.6).

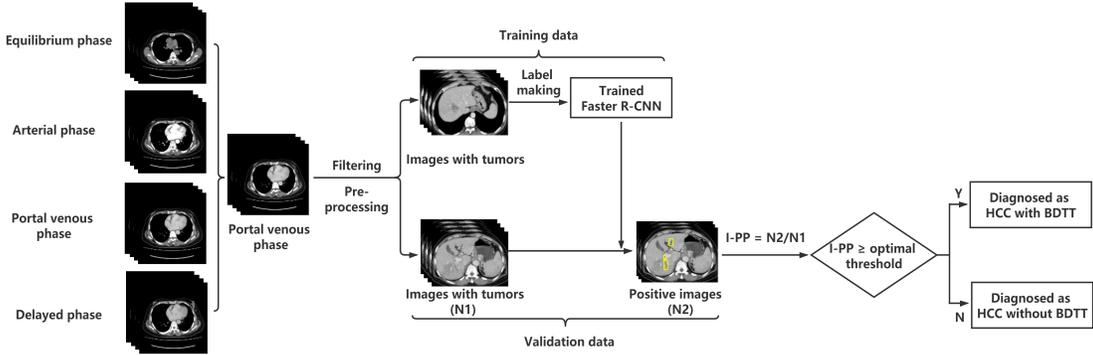


Fig. 5 The proposed AI pipeline for diagnosing BDTT. First, CT images in the portal venous phase with tumors were selected, then the selected images were center-cropped and resized to a unified size. After filtering and pre-processing, the trained Faster R-CNN was applied to identify DBDs on the resulting images. Finally, a patient was diagnosed as HCC with BDTT if the image-level positive proportion (I-PP) was greater than the optimal threshold

6.3. Image filtering, pre-processing, and label making

Some raw CT images have grayscale JPEG (Joint Photographic Experts Group) format while some have DICOM (Digital Imaging and Communications in Medicine) format. To remove potential confounding effect due to the formats, it is necessary to convert the raw CT images so as they have the same format. First, all the DICOM files were converted to the lossless grayscale JPEG images. Then, the resulting grayscale images were converted to pseudo-RGB images by tripling their channels with each channel having the same pixel value, so that they could serve as inputs of the object detection neural network.

With CT scan, each patient’s images can be divided into four phases: equilibrium phase, arterial phase, portal venous phase, and delayed phase. The brightness contrast between DBDs and normal tissues are most obvious in the portal venous phase, as shown in Figure 6. Note that images in the portal venous phase comprise

of a series of consecutive abdominal scan slices, and liver region generally occupies the middle part of those slices. Furthermore, intrahepatic DBDs usually present near tumor regions. Therefore, after format conversion, only those CT images in the portal venous phase with tumors present were retained. Such filtering could balance data between the case group and the control group. The image filtering procedure was reviewed by an experienced doctor.

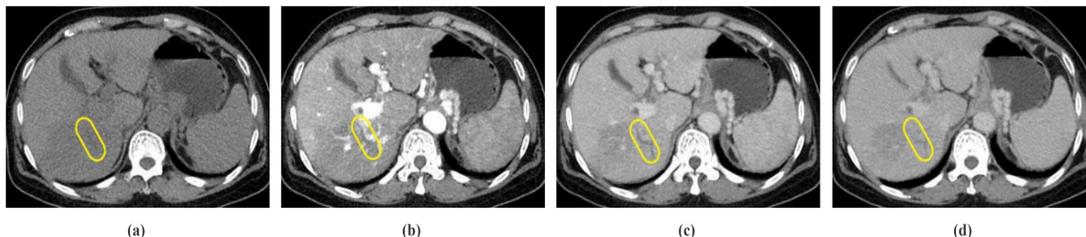


Fig. 6 CT images of an HCC patient with BDTT at the same location from (a) equilibrium phase, (b) arterial phase, (c) portal venous phase, and (d) delayed phase. DBDs caused by BDTT were marked in capsules

Since CT images of the case group were provided by four hospitals, their sizes vary to some extent due to different scanning equipments. Therefore, all images were resized to have a unified size and noninformative parts were discarded.

After filtering and pre-processing, the heterogeneity of CT images from different centers was largely eliminated and the main difference between images of the case group and the control group is whether DBDs are present or not. Several examples of preprocessed CT images with tumors and DBDs are presented in Figure 1, which shows that most DBDs can appear on several consecutive images in the portal venous phase since CT images are usually generated by continuous cross-section scanning of human body (Figure 1 (c1)-(c3)). Finally, DBDs were labeled by an experienced doctor using the software LabelImg [35], which extracted the location information of DBDs, i.e. the coordinates of the upper left and lower right corner of the bounding boxes as XML files. The resulting CT images combined with the label information were used to train the object detection neural network.

6.4. Object detection

Faster R-CNN [26], a popular two-stage neural network designed for object detection, was adopted to identify DBDs. Faster R-CNN consists of four components, namely Data Augmentation Module (DAM), CNN backbone, Region Proposal Network (RPN), and Region CNN (R-CNN). Compared with one-stage object detection

methods such as YOLO [27] and SSD [28], Faster R-CNN tends to be more accurate, albeit at a higher computational cost [36].

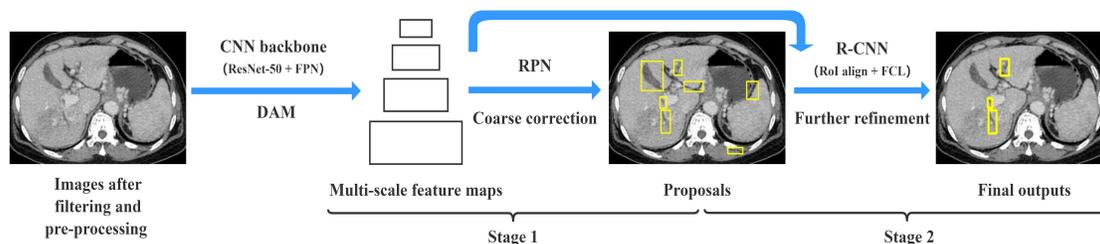


Fig. 7 Flow chart of Faster R-CNN. The CT images were first transformed via DAM, then CNN backbone extracted multi-scale feature maps from the transformed image. Based on the extracted features, at the first stage, RPN output proposal regions which tend to contain DBDs, then at the second stage, R-CNN further classified the proposal regions and refined their coordinates as the final outputs

The flowchart of Faster R-CNN for identifying DBDs is shown in Figure 7. To enrich data, multiple data augmentation techniques (eg., random shift, random rotation, etc.) were implemented in the DAM component during the training process. Meanwhile, transfer learning was adopted in this study, with ResNet-50 [37] being the CNN backbone for extracting feature maps, which was pretrained on the ImageNet dataset [38] with feature pyramid network (FPN) [39]. By combining FPN within the backbone, multi-scale feature maps were introduced and strong semantic features were propagated through a top-down pathway, thus the localization ability could be greatly enhanced. FPN was shown to be very effective in detecting small objects [40, 41]. After feature extraction, RPN and R-CNN were adopted to identify DBDs sequentially. First, RPN was used to identify regions that tend to contain DBDs as proposals and correct their coordinates. Then, R-CNN was used to classify the proposals and further refine the coordinates of these proposals. In R-CNN, RoI Align was used to extract uniform features from the proposals and fully connected layers (FCL) was used for classification and further refinement. Model configurations and training details of Faster R-CNN are included in the Supplementary Materials. We made the code for training Faster R-CNN available at <https://github.com/Daniel-1997/AI-pipeline-for-diagnosing-BDIT>.

6.5. Evaluation metrics for DBD identification

Due to limited data, four-fold cross validation was used to evaluate model’s performance. Specifically, the 16 case patients were evenly and randomly divided into four subsets. Each of the four subset was used for validation and the remaining three

subsets were used for training. (Thus, such training process was repeated for four times).

DBD-level true positive rate (D-TPR) was used as an evaluation metric to assess model’s performance in identifying DBDs. Specifically, each DBD was considered to be successfully detected if at least one bounding box output by Faster R-CNN overlapped with the DBD. Then, for each patient, D-TPR was defined as the ratio of the number of successfully detected DBDs to the number of true DBDs.

In each of the four training processes, image-level positive rate (I-PP) was defined as the proportion of images with at least one detected DBD for each patient in both case validation set and control group. I-PPs for each patient in the control group were averaged over the four training processes.

6.6. Patient-level diagnosis of BDTT

After identifying DBDs, an HCC patient was diagnosed to have BDTT if his/her I-PP was larger than a given threshold. Sensitivity, specificity, receiver operating characteristic (ROC) curve, and area under ROC curve (AUC) were used as diagnosis method evaluation metrics. For each given I-PP threshold, sensitivity was defined as the proportion of positive cases in HCC patients with BDTT, and specificity was defined as the proportion of negative cases in HCC patients without BDTT. The ROC curve was then drawn based on sensitivity vs. specificity pairs across various I-PP thresholds. The AUC value was defined as the area under the ROC curve. All evaluation metrics were calculated using the R package pROC [33].

6.7. Random forest as an alternative method

Random forest [42] was used as an alternative diagnosis method, which has become one of the most commonly used machine learning algorithms in classification and regression due to its high accuracy and fast training speed. In the task of classification, random forest ensembles multiple decision trees, and its output category is determined by the mode of categories output by individual trees. Those preoperative clinical variables significantly different between the case group and the control group were selected as inputs of random forest.

Abbreviations

BDTT: bile duct tumor thrombu; HCC: hepatocellular carcinoma; DBD: dilated bile duct; AI: artificial intelligence; CNN: convolutional neural network; FCN: fully convolution network; GAN: generative adversarial network; VAE: variational auto-encoder D-TPR: DBD-level true positive rate; IoU: intersection over union; I-PP:

image-level positive rate; ROC: receiver operating characteristic; AUC: area under ROC curve; CI: confidence interval; DAM: data augmentation module; RPN: region proposal network; RCNN: region CNN; FPN: feature pyramid network.

Acknowledgement

We would like to thank Fujian Provincial Hospital, the West China Hospital of Sichuan University, Mengchao Hepatobiliary Hospital of Fujian Medical University, and the First Affiliated Hospital of Fujian Medical University for providing valuable patient imaging data for this study.

Statements and Declarations

Funding

The work of JL and HZ was supported by the National Natural Science Foundation of China (No. 12171451, 11771096, 72091212). The work of JW and MY was supported by the Natural Science Foundation of Fujian Province (No. 2020J011105).

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

The present study was approved by the institutional reviewboard of all relevant institutions, and informed consents were obtained from the patients or their guardians for their data to be used for research purposes.

Consent for publication

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Availability of data and materials

The data used in this study are available from the corresponding author on reasonable request.

Authors' contributions

JL and AL performed data filtering, pre-processing and deep learning model development. JL and HZ accomplished statistical analysis, model evaluation and the manuscript writing. MY and YB collected data, provided clinical expertise and revised the manuscript. JW reviewed the data processing and the model evaluation procedure and provided the labels of the CT images. All authors read and approved the final manuscript.

References

- [1] Navadgi S, Chang C C, Bartlett A, et al. Systematic review and meta-analysis of outcomes after liver resection in patients with hepatocellular carcinoma (HCC) with and without bile duct thrombus[J]. *HPB*, 2016, 18(4): 312-316.
- [2] Lu W, Tang H, Yang Z, et al. A proposed modification for the Barcelona clinic liver cancer staging system: adding bile duct tumor thrombus status in patients with hepatocellular carcinoma[J]. *The American Journal of Surgery*, 2020, 220(4): 965-971.
- [3] Meng K W, Dong M, Zhang W G, et al. Clinical characteristics and surgical prognosis of hepatocellular carcinoma with bile duct invasion[J]. *Gastroenterology Research and Practice*, 2014, 2014.
- [4] Wang D D, Wu L Q, Wang Z S. Prognosis of hepatocellular carcinoma with bile duct tumor thrombus after R0 resection: a matched study[J]. *Hepatobiliary & Pancreatic Diseases International*, 2016, 15(6): 626-632.
- [5] Wang C, Yang Y, Sun D, et al. Prognosis of hepatocellular carcinoma patients with bile duct tumor thrombus after hepatic resection or liver transplantation in Asian populations: A meta-analysis[J]. *PLoS One*, 2017, 12(5): e0176827.
- [6] Shao W, Sui C, Liu Z, et al. Surgical outcome of hepatocellular carcinoma patients with biliary tumor thrombi[J]. *World Journal of Surgical Oncology*, 2011, 9(1): 1-5.
- [7] Rammohan A, Sathyanesan J, Rajendran K, et al. Bile duct thrombi in hepatocellular carcinoma: is aggressive surgery worthwhile?[J]. *HPB*, 2015, 17(6): 508-513.
- [8] Shiomi M, Kamiya J, Nagino M, et al. Hepatocellular carcinoma with biliary tumor thrombi: aggressive operative approach after appropriate preoperative management[J]. *Surgery*, 2001, 129(6): 692-698.
- [9] Zhou X, Wang J, Tang M, et al. Hepatocellular carcinoma with hilar bile duct tumor thrombus versus hilar Cholangiocarcinoma on enhanced computed tomography: a diagnostic challenge[J]. *BMC Cancer*, 2020, 20(1): 1-11.
- [10] Liu Q Y, Zhang W D, Chen J Y, et al. Hepatocellular carcinoma with bile duct tumor thrombus: dynamic computed tomography findings and histopathologic correlation[J]. *Journal of Computer Assisted Tomography*, 2011, 35(2): 187-194.

- [11] Zeng H, Xu L, Wen J, et al. Hepatocellular carcinoma with bile duct tumor thrombus: a clinicopathological analysis of factors predictive of recurrence and outcome after surgery[J]. *Medicine*, 2015, 94(1).
- [12] Liu Q, Chen J, Li H, et al. Hepatocellular carcinoma with bile duct tumor thrombi: correlation of magnetic resonance imaging features to histopathologic manifestations[J]. *European Journal of Radiology*, 2010, 76(1): 103-109.
- [13] Wu J Y, Huang L M, Bai Y N, et al. Imaging features of hepatocellular carcinoma with bile duct tumor thrombus: a multicenter study[J]. *Frontiers in Oncology*, 2021: 4618.
- [14] Yue-Hei Ng J, Hausknecht M, Vijayanarasimhan S, et al. Beyond short snippets: Deep networks for video classification[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 4694-4702.
- [15] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 3431-3440.
- [16] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. *Advances in Neural Information Processing Systems*, 2014, 27.
- [17] Kingma D P, Welling M. Auto-encoding variational bayes[J]. *arXiv preprint arXiv:1312.6114*, 2013.
- [18] Pranata Y D, Wang K C, Wang J C, et al. Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images[J]. *Computer Methods and Programs in Biomedicine*, 2019, 171: 27-37.
- [19] Li W. Automatic segmentation of liver tumor in CT images with deep convolutional neural networks[J]. *Journal of Computer and Communications*, 2015, 3(11): 146.
- [20] Basu S, Wagstyl K, Zandifar A, et al. Early prediction of alzheimer’s disease progression using variational autoencoders[C]//*International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2019: 205-213.
- [21] Zhang R, Tan S, Wang R, et al. Biomarker localization by combining CNN classifier and generative adversarial network[C]//*International Conference on Medical*

Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 209-217.

- [22] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015: 234-241.
- [23] Vorontsov E, Tang A, Pal C, et al. Liver lesion segmentation informed by joint liver segmentation[C]//2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018: 1332-1335.
- [24] Christ P F, Ettliger F, Grün F, et al. Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks[J]. arXiv preprint arXiv:1702.05970, 2017.
- [25] Alirri O I. Deep learning and level set approach for liver and tumor segmentation from CT scans[J]. Journal of Applied Clinical Medical Physics, 2020, 21(10): 200-209.
- [26] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.
- [27] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 779-788.
- [28] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [29] Thian Y L, Li Y, Jagmohan P, et al. Convolutional neural networks for automated fracture detection and localization on wrist radiographs[J]. Radiology: Artificial Intelligence, 2019, 1(1): e180001.
- [30] Olczak J, Fahlberg N, Maki A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs: deep learning algorithms—are they on par with humans for diagnosing fractures?[J]. Acta Orthopaedica, 2017, 88(6): 581-586.
- [31] Kim D H, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks[J]. Clinical Radiology, 2018, 73(5): 439-445.

- [32] Boot T, Irshad H. Diagnostic Assessment of Deep Learning Algorithms for Detection and Segmentation of Lesion in Mammographic Images[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2020: 56-65.
- [33] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, Müller M (2011). “pROC: an open-source package for R and S+ to analyze and compare ROC curves.” *BMC Bioinformatics*, 12, 77.
- [34] R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org> [accessed 09 November 2021].
- [35] Tzutalin. LabelImg. Git code. <https://github.com/tzutalin/labelImg>; 2015 [accessed 09 November 2021].
- [36] Zhang S, Chi C, Yao Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 9759-9768.
- [37] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [38] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, 2009: 248-255.
- [39] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2117-2125.
- [40] Pang J, Chen K, Shi J, et al. Libra r-cnn: Towards balanced learning for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 821-830.
- [41] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8759-8768.
- [42] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterial.pdf](#)