EBioMedicine 39 (2019) 272–279

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Research paper

Developed and validated a prognostic nomogram for recurrence-free survival after complete surgical resection of local primary gastrointestinal stromal tumors based on deep learning

Tao Chen [a,*,1], Shangqing Liu [b,1], Yong Li [d,1], Xingyu Feng [d], Wei Xiong [c], Xixi Zhao [c], Yali Yang [c], Cangui Zhang [a], Yanfeng Hu [a], Hao Chen [a], Tian Lin [a], Mingli Zhao [a], Hao Liu [a], Jiang Yu [a], Yikai Xu [c,*], Yu Zhang [b,*], Guoxin Li [a,*]

[a] Department of General Surgery, Nanfang Hospital, Guangdong Provincial Engineering Technology Research Center of Minimally Invasive Surgery, Southern Medical University, Guangdong Province, Guangzhou 510515, China
[b] School of Biomedical Engineering, Southern Medical University, Guangdong Province, Guangzhou, 510515, China
[c] Medical Image Center, Nanfang Hospital, Guangdong Province, Southern Medical University, Guangzhou 510515, China
[d] Department of General Surgery, Guangdong Academy of Medical Science, Guangdong General Hospital, Guangdong Province, Guangzhou 510080, China

ARTICLE INFO

Article history:
Received 20 November 2018
Received in revised form 8 December 2018
Accepted 14 December 2018
Available online 23 December 2018

Keywords:
Gastrointestinal Stromal Tumors
Deep Learning
Residual Neural Network
Recurrence-free Survival
Imatinib

ABSTRACT

This study aimed to develop and validate a prognostic nomogram for recurrence-free survival (RFS) after surgery in the absence of adjuvant therapy to guide the selection for adjuvant imatinib therapy based on Residual Neural Network (ResNet).

The ResNet model was developed based on contrast-enhanced computed tomography (CE-CT) in a training cohort consisted of 80 patients pathologically diagnosed gastrointestinal sromal tumors (GISTs) and validated in internal and external validation cohort respectively. Independent clinicopathologic factors were integrated with the ResNet model to construct the individualized nomogram. The performance of the nomogram was evaluated in regard to discrimination, calibration, and clinical usefulness.

The ResNet model was significantly associated with RFS. Integrable predictors in the individualized ResNet nomogram included the tumor site, size, and mitotic count. Compared with modified NIH, AFIP, and clinicopathologic nomogram, both ResNet nomogram and ResNet model showed a better discrimination capability with AUCs of 0·947(95%CI, 0·910–0·984) for 3-year-RFS, 0·918(0·852–0·984) for 5-year-RFS, and AUCs of 0·912 (0·851–0·973) for 3-year-RFS, 0·887(0·816–0·960) for 5-year-RFS, respectively. Calibration curve shows the good calibration of the nomogram in terms of the agreement between the estimated and the observed 3- and 5- year outcomes. Decision curve analysis showed that the ResNet nomogram had a higher overall net benefit.

In conclusion, we presented a deep learning-based prognostic nomogram to predict RFS after resection of localized primary GISTs with excellent performance and could be a potential tool to select patients for adjuvant imatinib therapy.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Gastrointestinal stromal tumors (GISTs) are mesenchymal neoplasms that mostly originating from the gastrointestinal tract with varying malignant potential which ranges from the benign lesion to fatal sarcoma [1]. Adjuvant treatment with the tyrosine kinase inhibitor imatinib is recommended for the patients with high risk of recurrence [1]. However, underestimation of recurrence risk might have a negative impact on recurrence-free survival (RFS) due to the inadequacies of treatment [2]. Besides, for patients with an underestimated risk of recurrence, a longitudinal follow-up may not be scheduled. Conversely, patients with low-risk likely to be cured by surgery, rather than receiving further benefits, may suffer toxic effects and unnecessary costs from adjuvant treatment [3]. Thus, accurate assessment of the recurrence risk is vital for the management of GISTs that underwent curative resection. Although the risk stratification standards have been revised and improved, their predictive accuracy is roughly similar [1]. New proposed systems have not been widely applied due to the lack of powerful evidence, sufficient applicability, and particularly substantial performance.

* Corresponding authors at: Department of General Surgery, Nanfang Hospital, Guangdong Provincial Engineering Technology Research Center of Minimally Invasive Surgery, Southern Medical University, No.1838, North Guangzhou Avenue, Guangdong Province, Guangzhou 510515, China.
E-mail addresses: drchentao@163.com (T. Chen), yikaivip@163.com (Y. Xu), yuzhang@smu.edu.cn (Y. Zhang), gzliguoxin@163.com (G. Li).
[1] Tao Chen, Shangqing Liu, and Yong Li contributed equally to this work.

https://doi.org/10.1016/j.ebiom.2018.12.028
2352-3964/© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Research in context

Evidence before this study

We searched articles with the following terms: "(Deep learning OR Radiomics OR ResNet) AND (GISTs OR Gastrointestinal stromal tumors) AND (prognosis OR survival) AND (prediction OR predictive OR predict)" on PudMed and Web of Science. The articles were not limited to English language publications and didn't have date restriction. This search did not identify any study to predict the recurrence risk of GISTs patients by deep learning model.

Added value of this study

To our knowledge, this is the first study to predict the recurrence risk of GISTs patients by deep learning technique. Artificial intelligence (AI) has become a hot topic. Radiomics is a typical and effective case of medical application but relies on multi-step pipelines. Deep learning, as one of the power algorithms of AI, can simplify the procedure by traditional radiomics approach and strongly supports the translation from AI into clinical application. Here, we developed and validated a prognostic nomogram based on a deep learning approach to predict the recurrence-free survival (RFS) of GISTs with satisfactory performance, which may be a potential tool to predict the RFS for GISTs after complete resection, avoiding excessive therapy or missing the optimal timing.

Implications of all the available evidence

Our deep learning-based model, combined with existing evidence, proved that radiomics with deep learning approach did have a better prediction on RFS of GISTs patients. It might contribute to personalized medicine, which may be a potential tool in the search for clinical decision support that is individualized and effective. In the future work, it may be better to validate in additional cohorts and verify in randomized controlled trials.

On the other hand, Artificial intelligence (AI) has become a hot topic with reports of breakthroughs not only in industry, finance, but also the medical care support. Radiomics, as a typical and effective case of medical application of AI [4–16], can utilize diggable data via high-throughput extraction of quantitative features based on medical images, but it relies on multi-step pipelines using traditional machine learning techniques. Deep learning, as one of the power algorithms of AI, can simplify the procedure by learning predictive features directly and strongly supports the translation from AI into clinical application [5,17–23].

In this study, we aim to develop and validate a deep learning-based prognostic nomogram to predict the recurrence risk after curative resection of a localized primary GIST in the absence of adjuvant imatinib therapy.

2. Materials and methods

2.1. Patients enrollment

Three independent cohorts consisted of 147 patients with GISTs pathologically diagnosed were enrolled in the study. Eighty cases as the training cohort and 35 cases as internal validation cohort were obtained in our center from January 2005 to December 2015. Moreover, we included one external validation cohort that comprised 32 cases from Guangdong General Hospital with the same criteria between January 2008 to December 2015. Ethical approvals were obtained for this retrospective analysis in two participating centers, and the patient informed consent requirements were waved.

Inclusion criteria: (1) localized primary GIST patients underwent surgical resection with curative intent; (2) GISTs confirmed by postoperative pathology and immunohistochemistry examinations; (3) contrast-enhanced computed tomography (CE-CT) was performed within 15 days before the surgery; (4) complete clinical and pathological data were available. Exclusion criteria: (1) patients received imatinib therapy or other tyrosine kinase inhibitor before and after surgery; (2) presence of metastases at diagnosis; (3) patients with tumor ruptured before or during the operation.

Flow diagram for extracting eligible patients was presented in the Fig. S1. Patients were postoperatively followed up with abdominal CTs every 6–12 months for the first 3 years and then annually. The follow-up duration was measured from the time of operation to the last follow-up date, and the survival status at the last follow-up was recorded. We defined RFS as the time to recurrence at any site.

2.2. Image acquisition and ROI annotation

After a non-contrast CT scan (Scanner: SIEMENS 64-MDCT or GE Healthcare, Hino) with a thickness of 2·0 mm, a dynamic contrast-enhanced scan was performed, with 90 to 100 ml iodine contrast medium (Ultravist 370, Bayer Schering Pharma, Germany) injected intravenously at a rate of 3·0 to 3·5 ml/s. Arterial phase image of contrast-enhanced abdominal CT with manual region of interest (ROI) was selected for analysis. ROI was delineated with the whole data in a blind fashion by two radiologists with 12 (reader 1, W.X.) and 7 years (reader 2, X.X.Z.) of experience in the interpretation of abdominal CT. The annotation results were assessed with satisfactory inter- and intra- observer reliability in our previous study [13]. All outcomes were based on the annotations of the first reader.

2.3. Image pre-processing

Image intensities were rescaled within [0, …,255] by a soft-tissue window of [−110,190]HU to increase the contrast of soft tissues as well as show more details about abdominal organs. Small patches (28×28 voxels) with tumor were extracted. To extract patches, a bounding rectangle derived from the tumor segmentation was drawn around the tumor. This ensured that the entire tumor area was captured as well as a portion of the tumor margin. The patches that tumor area less than 50th percentile patch area were discarded.

Patch samples from the same patient were kept together when randomizing into training cohort, internal validation cohort, and external validation cohort. We augmented the training data by introducing random rotations, translations, shearing, zooming, and flipping (horizontal and vertical), generating "new" training data. The augmentation technique allows us to further increase the size of our training cohort. For every epoch, we augmented the training data before inputting it into the neural network. Augmentation was only performed on the training cohort, not on the internal validation cohort, or the external validation cohort. Data augmentation was performed in real time in order to minimize memory usage.

2.4. Residual neural network

Residual neural network [24] (ResNet) is applied to train the image data and build our neural network model. There are 10 identity blocks and 2 convolution blocks. Each identity block has 2 convolutional layers. Batch normalization (BN) and rectified linear units (ReLu) are adopted after every convolutional layer. Batch standardization forces network activation to follow a unit Gaussian distribution after each update to

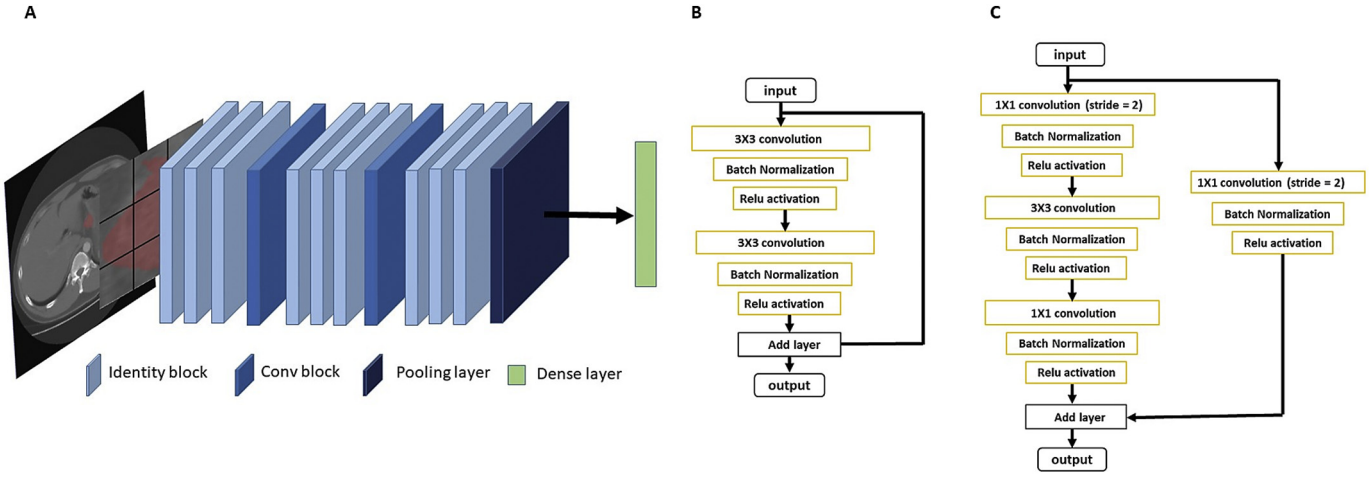

Fig. 1. Residual neural network. (A) Network architecture. (B) Identity block: each identity block has 2 convolutional layers.(C) Convolutional block: each convolution block has 3 convolutional layers and a projection shortcut (convolution with a stride of 2). The weights are initialized with a normal distribution in convolutional layers. *ReLu* rectified linear units.

prevent internal covariate migration and overfitting. The shortcut is directly used because the input and output are of the same dimensions. As the Fig. 1 shows.

2.5. Implementation details

Our implementation was based on the Keras package with the TensorFlow library as the backend. During training, the probability of each patch sample belonging to the recrudescence or none class was computed with a sigmoid classifier. The weight of the network was optimized via Rmsprop algorithm with a mini-batch size of 32. The objective function used was binary cross-entropy. The initial learning rate was set to 0·001. The learning rate was reduced to 0·2 of its value after 50 consecutive epochs without an improvement of the validation loss.

At the end of training phase, the model was reverted back to the model with the lowest validation loss up until that point in training. The final model was the one with lowest validation loss at any point during training. Kernel weight was initialized randomly using the Glorot uniform initializer. Biases was initialized with zero. We ran our code on a graphics processing unit to exploit its computational speed. Our algorithm was trained on a NVIDIA TITAN X graphics processing unit.

2.6. Performance assessment of ResNet model

We assessed the prognostic accuracy of the risk prediction in both training cohort and validation cohort (internal and external validation cohorts) using time-dependent receiver operator characteristics (ROC) analysis at different follow-up times. The GISTs patients were classified in to high and low risk score groups. The thresholds of classification were identified by using X-title [25]. We evaluated the potential association of the ResNet model with RFS in the training cohort and validated it in validation cohort by using Kaplan-Meier survival analysis.

2.7. Development and evaluation an individualized nomogram

To testify the incremental value of the ResNet model to the independent clinicopathological factors for individualized assessment of RFS, we developed a ResNet nomogram and a clinicopathologic nomogram in the whole cohort based on the multivariate Cox analysis [26]. The calibration was applied to compare the predicted survival with the actual survival [27]. The Net Reclassification Improvement (NRI) was calculated to quantify the improvement of usefulness with the addition of ResNet model [28]. To determine the clinical usefulness of our ResNet model, a decision curve analysis (DCA) which quantify the net benefits at different threshold probabilities was conducted [29].

Table 1
Clinical pathological characteristics and followed- up results of patients in different cohorts.*

Variables	Training cohort($n = 80$)			Internal validation cohort($n = 35$)			External validation cohort($n = 32$)		
	Low-score(%)	High- score(%)	*P*-value	Low- score(%)	High- score(%)	*P*-value	Low- score(%)	High- score(%)	*P*-value
Gender			0.655			0.803			1.000
Male	32(78.0)	9(22.0)		12(66.7)	6(33.3)		8(88.9)	1(11.1)	
Female	32(82.1)	7(17.9)		12(70.6)	5(29.4)		19(82.6)	4(17.4)	
Age(mean ± SD,years)	53.83 ± 12.59	59.94 ± 12.37	0.086	51.58 ± 12.12	63.82 ± 10.99	0.007*	62.41 ± 11.38	59.00 ± 9.03	0.533
Tumor site			0.144			0.709			0.673
Gastric	50(84.7)	9(15.3)		19(65.5)	10(34.5)		22(84.6)	4(15.4)	
Non-gastric	14(66.7)	7(33.3)		5(83.3)	1(16.7)		5(83.3)	1(16.7)	
Tumor size(cm)	4.78 ± 2.71	11.34 ± 5.11	< 0.0001*	4.79 ± 3.13	9.54 ± 4.77	0.009*	6.03 ± 3.05	10.50 ± 5.27	0.012*
Mitotic count			0.073			0.007*			0.642
≤5/50HPFs	49(86.0)	8(14.0)		21(84.0)	4(16.0)		17(89.5)	2(10.5)	
>5/50HPFs	15(65.2)	8(34.8)		3(30.0)	7(70.0)		10(76.9)	3(23.1)	
Recurrence			< 0.0001*			< 0.0001*			0.011*
Absent	62(95.4)	3(4.6)		22(88.0)	3(12.0)		23(95.8)	1(4.2)	
Present	2(13.3)	13(86.7)		2(20.0)	8(80.0)		4(50.0)	4(50.0)	

Independent samples t-test was applied in continuous variables. Chi-Squared test was applied in categorical variables.
SD standard deviation, *HPF* high-power field.
* *P* value <0.05.

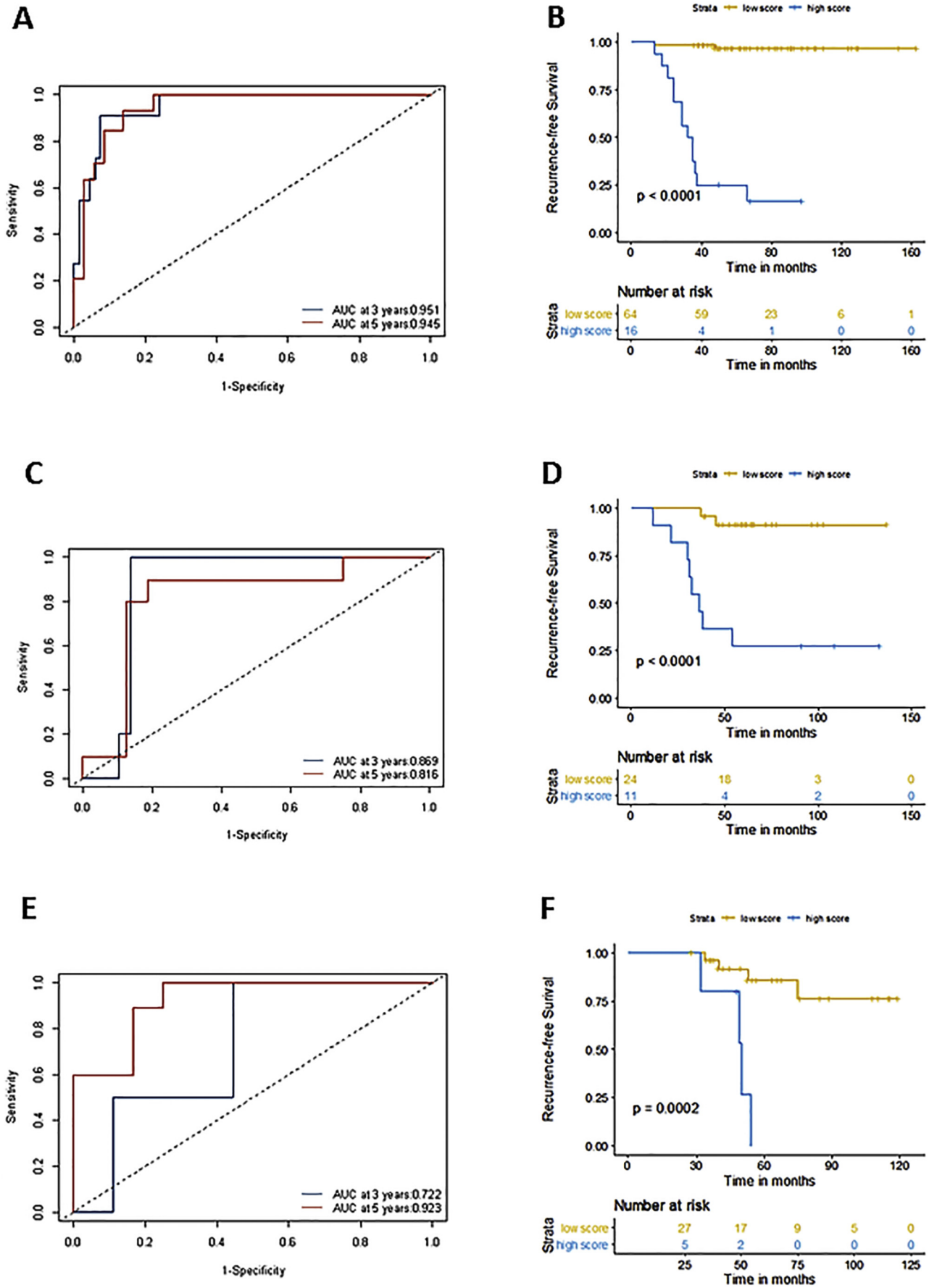

Fig. 2. ResNet model risk prediction measured by time-dependent ROC curves and Kaplan-Meier survival. (A-B) Training cohort. (C-D) Internal validation cohort. (*E*-F) External validation cohort. The prognostic accuracy is evaluated by the AUCs 3 and 5 years in training, internal, and external validation cohorts. *P*-values are calculated by the log-rank test. *ROC* receiver operator characteristic, *AUC* area under the curve.

The discrimination performance of ResNet nomogram was assessed compared with ResNet model, modified National Institutes of Health (NIH) criteria [30], Armed Forces Institute of Pathology (AFIP) criteria [31], and clinicopathologic nomogram on the basis of ROC curves and AUC values. The Akaike information criterion (AIC) was used to evaluate the risk of overfitting.

2.8. Statistical analysis

t-test and chi-square were applied for continuous variables and categorical variables separately. The Kaplan-Meier method and log-rank test were performed to estimate the RFS of GISTs. And the Cox proportional hazards model was used for multivariate analyses. We evaluated the prognostic accuracy of our model using time-dependent receiver operator characteristics (ROC) analysis. Statistical analyses were conducted with R software (version 3·5·1) and SPSS software (version 22·0). We used 10 packages of R software, which were survminer, survival, timeROC, rms, VIM, nomogramEX, Hmisc, Formula, ggplot2, and rmda. A two-sided *P* value <0·05 was considered significant.

3. Results

3.1. Patient characteristics

A total of 147 GISTs patients' contrast-enhanced abdominal CT images were applied in this study. Patient characteristics in the training, internal validation and external validation cohorts were presented in Table 1. The median (interquartile rang [IQR]) survival times for RFS were 57(13–163) months in training cohort, 59(11–137) months in internal validation cohort, and 53(28–119) months in external validation cohort, respectively.

3.2. Model performance and validation

The ResNet model was significantly associated with the RFS (Table S1). In training cohort, AUC at 3 years is 0·951 (95% CI: 0·901–0·999), while AUC at 5 years is 0·945 (95%CI: 0·887–0·999) (Fig. 2A). In internal validation cohort, AUC at 3 years is 0·869 (95% CI: 0·747–0·991), while AUC at 5 years is 0·816 (95%CI: 0·628–0·999) (Fig. 2C). In external validation cohort, AUC at 3 years is 0·722 (95% CI: 0·453–0·991), while AUC at 5 years is 0·923 (95%CI: 0·812–0·999) (Fig. 2E).

The optimum cutoffs generated by the X-tile plot was 0·9819. Accordingly, patients were classified into a low risk score group (score<0·9819) and a high risk score group (score ≥ 0·9819).

In the training cohort, the 3-year RFS and 5-year RFS were 98·44% and 51·56%, respectively, for the low score group; 37·50% and 18·75%, respectively, for the high score group (all P<0·0001) (Fig. 2B). In the internal validation cohort, the 3-year RFS and 5-year RFS were 100·00% and 58·33%, respectively, for the low score group; 54·55% and 27·27%, respectively, for the high score group. (all P<0·0001) (Fig. 2D). In the external validation cohort, the 3-year RFS and 5-year RFS were 88·89% and 44·44%, respectively, for the low score group; In high score group, the 3-year RFS was 80·00%, and there is no 5-year RFS patient. (all P<0·0002) (Fig. 2F).

3.3. Assessment of Incremental Value of model in Individual RFS Performance

Three statistically significant clinicopathologic indicators were obtained: the tumor site, size, and mitotic count (Table S1, 2). The ResNet and clinicopathologic nomograms were presented in Fig. 3A and Fig. S2, respectively. The calibration curves of the nomograms were shown in Fig. 3B and Fig. S3. This curve showed the good calibration of the nomogram in terms of the agreement between the estimated and the observed 3- and 5- year outcomes. The inclusion of the ResNet model in the clinicopathologic nomogram yielded a total NRI of 0·605(95% CI: 0·243-0·966; $p = 0·001$) for RFS, indicating that improved classification accuracy for survival outcomes.

Compared with modified NIH, AFIP, and clinicopathologic nomogram in the whole cohort, both the ResNet nomogram and ResNet model showed a better predictive capability in 3- and 5-year RFS in ROC curves (Fig. 4 A, B). For 3-year RFS, the AUCs of ResNet nomogram, ResNet model, clinicopathologic nomogram, modified NIH, AFIP were 0·947 (95%CI:0·910–0·984), 0·912 (0·851–0·973), 0·852(0·783–0·921), 0·822(0·765–0·879), 0·812(0·726–0·898), respectively. The AUC results of 5-year RFS of these models were 0.918(95%CI:0·852–0·984), 0·887 (0·816–0·960), 0·772(0·679–0·865), 0·754(0·667–0·841), 0·739(0·643–0·835), respectively (Table 2).

According to the decision curve analysis, ResNet nomogram was superior to the current risk predicted criteria and clinicopathologic nomogram over most of the range of rational threshold probabilities, indicating the incremental value of ResNet model in the individualized prognostic prediction. (Fig. 5).

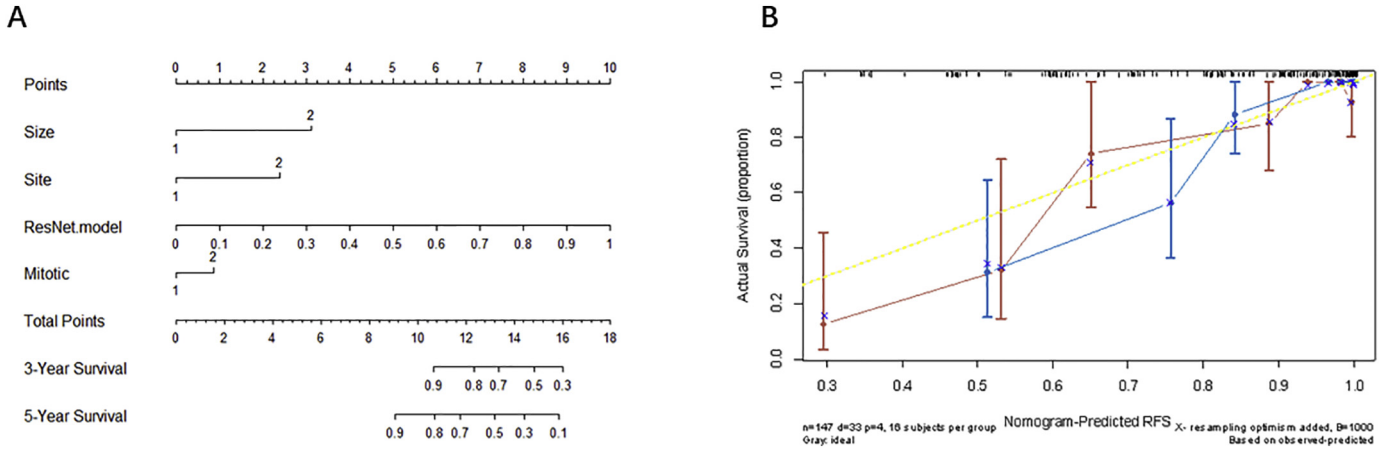

Fig. 3. ResNet nomogram for RFS and calibration curve. (A) ResNet nomogram for RFS. This nomogram was developed integrating with ResNet model and significant clinicopathologic indicators: tumor site, size, and mitotic count. The probability of each predictor can be converted into the points axis at the top of the nomogram. After adding up the points of each predictor in total points axis, we can find the patient's probability of RFS at the bottom of the nomogram. (B) Calibration curves of ResNet nomogram for RFS. Estimated RFS is plotted on the x-axis, and the observed tumor relapse rate is plotted on the y-axis. Yellow dotted line represents a perfect estimated outcome by an ideal model and perfectly association with the actual outcome. Solid line represents estimated outcome of the model, a closer alignment of which with the yellow dotted line represents a better performance. The blue and red solid lines represent the estimations of 3-year RFS and 5-year RFS, respectively. *RFS* recurrence-free survival, *ResNet* Residual Neural Network.

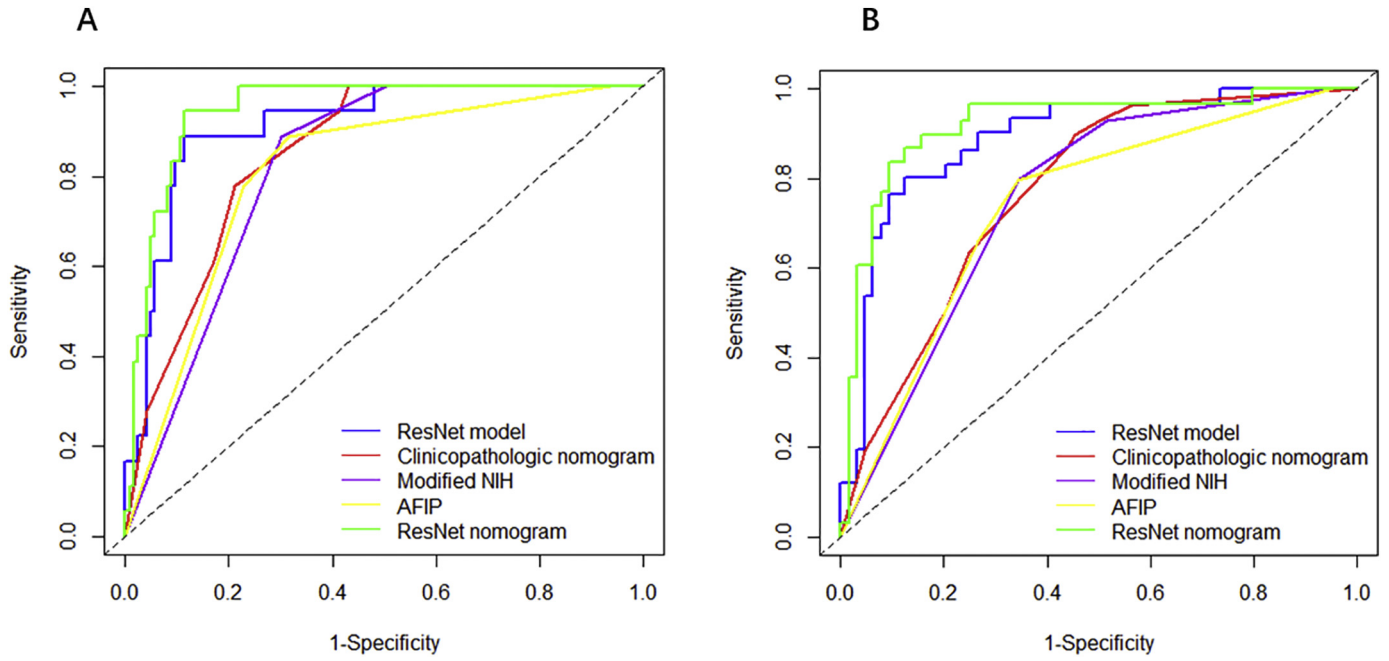

Fig. 4. Receiver operating characteristic (ROC) curves of predictive performances of different methods. (A) ROC curve of 3-year RFS prediction. (B) ROC curve of 5-year RFS prediction. The curves of five colors represent different methods: green, ResNet nomogram; blue, ResNet model; red, clinicopathologic nomogram; purple, modified NIH; yellow, AFIP. *ROC* receiver operator characteristic, *RFS* recurrence-free survival, *ResNet* Residual Neural Network, *NIH* National Institutes of Health, *AFIP* Armed Forces Institute of Pathology.

4. Discussion

We presented a deep learning-based prognostic model of GISTs, which can successfully classify those patients into high and low predicted score groups with significant differences in RFS and was demonstrated to be an independent risk factor of prognosis in the patients with GISTs. Our ResNet nomogram performed better than the modified NIH, AFIP, and clinicopathologic nomogram and showed incremental value of ResNet model for individualized RFS estimation. The nomogram might be useful for the selection of GISTs patients for adjuvant imatinib therapy.

As described previously, accurately evaluating the risk of GIST recurrence after surgery is very important to determine the appropriateness of adjuvant treatment and the intensity of postoperative surveillance. However, current schemes of risk-stratification can't explain all the biological behavior and clinical outcomes of GISTs. Mitotic count is the most important prognostic factor for GISTs in these criteria, but its reliability is controversial. Mitotic count relies on the subjective identification by pathologists, so the number detected may be affected by different visual field of microscopes, tissue fixation, and sampling [3,32]. The quantified analysis in CT features using deep learning technique could eliminate the subjective factors to certain extent, and it might work as a complement of subjective pathology results. Tumor size is another important independent prognostic factor for GISTs, patients with larger tumor size are more likely to have an adverse prognosis. However, some small GISTs may also be aggressive [33,34]. The application of AI in medical data such as the radiomics approach could well capture the intratumoral heterogeneity and might have the potential to perform better preoperatively in some cases with small size [13]. Additionally, the same as mitotic count, tumor size also has potential variability. Because when the specimen is measured in relation to fixation, tumor size could be affected. In our nomogram, the weight of ResNet model is greater than both mitotic count and tumor size.

Risk criteria for GISTs have always been being revised due to the exploration of new significant variables [30,31]. New integrating approaches of prognostic factors could also increase the accuracy of prognostic prediction [35–37] such as the nomogram and non-linear model. However, these criteria mainly depend on different combination of the traditional clinicopathological factors such as the tumor size, site, and, mitotic count, which means they can't improve the performance distinctly [1]. The deep learning approach based on medical images couldn't only provide a novel prognostic multi-feature factor, but also a powerful and efficient algorithmic technique. The comparison results in our study demonstrated the discrimination of deep learning model was not only superior to the modified NIH and AFIP criteria, but also to the nomogram integrating the significant clinicopathologic factors.

Radiomics signature has been demonstrated in various studies, which could assess the biological behavior of a tumor comprehensively and potentially improve the accuracy of diagnosis, prognosis, and prediction [7–16]. Deep learning can simplify the multi-step pipeline of conventional radiomics by training and testing the predictive features directly from the images with greater reproducibility. Convolutional

Table 2
Performance of Models: the values of AUC and AIC.

Model	3 years Disease-free survival		5 years Disease-free survival	
	AUC (95% CI)	AIC	AUC (95% CI)	AIC
ResNet nomogram	0.947(0.910–0.984)	1411.883	0.918(0.852–0.984)	1411.883
ResNet model	0.912 (0.851–0.973)	1416.413	0.887(0.816–0.960)	1416.413
Clinicopathologic nomogram	0.852(0.783–0.921)	1417.826	0.772(0.679–0.865)	1417.826
Modified NIH	0.822(0.765–0.879)	1418.545	0.754(0.667–0.841)	1418.545
AFIP	0.812(0.726–0.898)	1420.848	0.739(0.643–0.835)	1420.848

ResNet Residual neural network, *NIH* National Institutes of Health, *AFIP* Armed Forces Institute of Pathology, *AIC* Akaike information criterion.

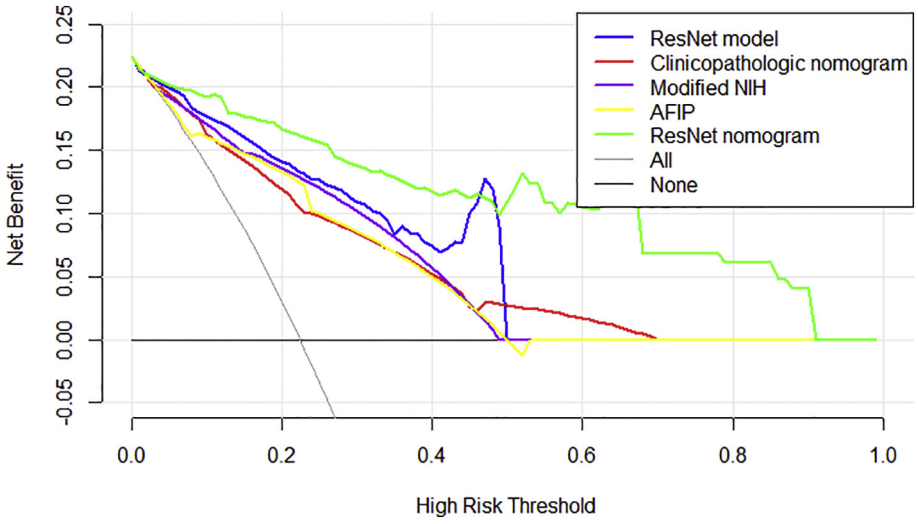

Fig. 5. Decision curve analysis for each method. The y-axis measures the net benefit. The net benefit is calculated by adding up the true positive results and subtracting the false positive results, weighting the latter by a factor relevant to the relative harm of an undetected caner compared with the harm of unnecessary treatment. The ResNet nomogram has the highest net benefit compared to both the other methods and simple strategies such as follow-up of all patients (grey line) or no patients (horizontal black line) across the full range of threshold probabilities at which a patient would choose to undergo imaging follow-up. *ResNet* Residual Neural Network, *NIH* National Institutes of Health, *AFIP* Armed Forces Institute of Pathology.

neural networks (CNNs) is a typical network for learning hierarchical representations of imaging data [38]. Neural networks are inspired by the connectivity pattern between neurons in biological processes, transforming input image through a series of chained convolutional layers and then resulting in output vector of class probability. Compared to conventional machine learning classification, it can obtain higher accuracy but with relatively little pre-processing. Residual neural network won the 2015 Large Scale Visual Recognition Challenge in image classification and substantially superior to previous network of deep learning [24]. ResNet permanently utilizes shortcut connections between shallow and deep networks to adjust training error rate and improve the accuracy of classification. To date, ResNet has been used more and more due to its utility and simplicity [22,39,40].

The need for a large size of training data is one of the challenges of deep learning. The low incidence of GISTs might lead to the insufficient training data. Therefore, in the study, we extracted multiple image samples from one patient using patch pre-processing. In addition, the data augmentation was also used to increase the size of training data and prevent overfitting. Furthermore, the architecture we choose was relatively simple, and the satisfactory results in our study demonstrated that it was complex enough to learn the predictive features. The ResNet with simple architecture and short time consumption might make the proposed method more likely to be applied in GISTs filed. In addition, nomogram provided an individual and quantitative approach for clinic application by integrating the ResNet model and other risk factors. Combined ResNet nomogram acquired better discrimination performance than either the ResNet model or the clinicopathologic nomogram alone with positive NRI.

Automatic image segmentation is one of the applications of deep learning, but still under developing. Therefore, in this study, we used the manual annotation of tumor ROI rather than deep learning algorithm. Moreover, our ROI annotations by hand were obtained a satisfactory inter- and intraobserver assessment in previous study [13], make the subsequent analysis more reliable.

There are several limitations in this study. First, our data were collected retrospectively, and further prospective research is needed. Second, relatively small sample size is also a limiting factor in deep learning, but data pre-processing using patches and data augmentation were performed to increase size of the training cohort. Moreover, the simple architecture of ResNet we applied also benefited the small input data. Third, the process of model development may be time-consuming, but this ResNet nomogram can be programmed into accessible software or websites, which could easily facilitate its clinical application. Nevertheless, to our knowledge, this is the first study to predict the recurrence risk of GISTs patients by deep learning technique, which might supply a valuable reference for deep learning application in gastrointestinal tumor. More cohort validation and more integrable factors such as KIT and PDGFRA mutations should be considered in future research [1,41].

In this study, we developed a ResNet model to predict the recurrence risk of the GISTs with satisfactory performance. Compared with a radiomics approach, our deep learning model don't need pre-engineered features processing. This model may have the potential to become an applicable image biomarker or integrable factor to improve the accuracy of risk prediction for GISTs. Incorporating the ResNet model and clinicopathologic risk factors into an easy-to-use nomogram was more likely to predict the individual RFS for patients after complete resection of localized primary GISTs, consequently avoiding excessive targeted therapy or missing the optimal timing.

Funding

This work was supported by the State's Key Project of Research and Development Plan (2017YFC0108300 and 2017YFC0108303).

Declaration of Interests

The authors have no conflicts of interest to declare.

Authors' contributions

Guarantor of the article.
Guoxin Li, Yu Zhang, Yikai Xu, Tao Chen.

Specific author contributions

Conception and design: Guoxin Li, Yu Zhang, Yikai Xu, Tao Chen, Yong Li.

Collection and assembly of data: Tao Chen, Xingyu Feng, Wei Xiong, Xixi Zhao, Yali Yang, Hao Chen, Tian Lin, Mingli Zhao.

Data analysis and interpretation: Tao Chen, Shangqing Liu, Cangui Zhang, Yanfeng Hu, Hao Liu, Jiang Yu.

Manuscript writing: All authors.

Final approval of manuscript: All authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2018.12.028.

References

[1] Joensuu H, Hohenberger P, Corless CL. Gastrointestinal stromal tumour. LANCET [Journal Article; Review] 2013;382(9896):973–83.
[2] Guerin A, Sasane M, Keir CH, Gauthier G, Macalalad AR, Wu EQ, et al. Physician Underestimation of the Risk of Gastrointestinal Stromal Tumor Recurrence After Resection. JAMA ONCOL Journal Article; Observational Study; Research Support, Non-US Gov't 2015;1(6):797–805.
[3] Joensuu H. Physician Estimations of the risk of Gastrointestinal Stromal Tumor Recurrence–not Accurate enough?: more Education May be Needed. JAMA Oncol 2015;1(6):805–6 Comment; Journal Article.
[4] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48(4):441–6.
[5] Aerts HJWL. Data Science in Radiology: a Path Forward. Clin Cancer Res 2018;24(3):532–4.
[6] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than Pictures, they are Data. Radiology 2016;278(2):563–77.
[7] Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. J CLIN ONCOL [Journal Article; Research Support, Non-US Gov't] 2016;34(18):2157–64.
[8] Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome based on Radiomics of Multiparametric MRI. Clin Cancer Res 2016;22(21):5256–64.
[9] Kickingereder P, Gotz M, Muschelli J, Wick A, Neuberger U, Shinohara RT, et al. Large-scale Radiomic Profiling of Recurrent Glioblastoma Identifies an Imaging Predictor for Stratifying Anti-Angiogenic Treatment Response. Clin Cancer Res 2016;22(23):5765–71.
[10] Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. Clin Cancer Res 2017;23(15):4259–69.
[11] Wu S, Zheng J, Li Y, Yu H, Shi S, Xie W, et al. A Radiomics Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer. Clin Cancer Res 2017;23(22):6904–11 Journal Article.
[12] Liu Z, Zhang X, Shi Y, Wang L, Zhu H, Tang Z, et al. Radiomics Analysis for Evaluation of Pathological complete Response to Neoadjuvant Chemoradiotherapy in locally Advanced Rectal Cancer. Clin Cancer Res 2017;23(23):7253–62.
[13] Chen T, Ning Z, Xu L, Feng X, Han S, Roth HR, et al. Radiomics nomogram for predicting the malignant potential of gastrointestinal stromal tumours preoperatively. Eur Radiol 2018 Aug 16. https://doi.org/10.1007/s00330-018-5629-2.
[14] Wu S, Zheng J, Li Y, Wu Z, Shi S, Huang M, et al. Development and Validation of an MRI-Based Radiomics Signature for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer. EBioMedicine 2018;34:76–84.
[15] Jiang Y, Chen C, Xie J, Wang W, Zha X, Lv W, et al. Radiomics signature of computed tomography imaging for prediction of survival and chemotherapeutic benefits in gastric cancer. EBioMedicine 2018;36:171–82.
[16] Wu Y, Xu L, Yang P, Lin N, Huang X, Pan W, et al. Survival Prediction in high-grade Osteosarcoma using Radiomics of Diagnostic Computed Tomography. EBioMedicine 2018;34:27–34.
[17] Beam AL, Kohane IS. Translating Artificial Intelligence into Clinical Care. JAMA 2016;316(22):2368–9 Editorial; Comment.
[18] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA Journal Article; Research support, Non-US Gov't; Validation 2016;316(22):2402–10.
[19] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. NATURE [Journal Article; Research support, NIH, Extramural; Research support, Non-US Gov't; Validation studies] 2017;542(7639):115–8.
[20] Golden JA. Deep Learning Algorithms for Detection of Lymph Node Metastases from Breast Cancer. Helping Artificial Intelligence Be Seen JAMA [Editorial; Comment] 2017;318(22):2184–6.
[21] Chaudharyl K, Poirionl OB, Lu L, Garmire LX. Deep Learning-based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. Clin Cancer Res 2018;24(6):1248–59.
[22] Chang K, Bai HX, Zhou H, Su C, Bi WL, Agbodza E, et al. Residual Convolutional Neural Network for the Determination of IDH Status in Low- and High-Grade Gliomas from MR Imaging. CLIN CANCER RES [Journal Article] 2018;24(5):1073–81.
[23] Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. Gut 2018 May 5. https://doi.org/10.1136/gutjnl-2018-316204 pii: gutjnl-2018-316204.
[24] Kaiming He XZSR. Deep Residual Learning for image Recognition. In: Computer Vision and Pattern Recognition, editor. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016; Las Vegas, NV, United States. Pub Place; 2016. p. 70–8.
[25] Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res 2004;10(21):7252–9.
[26] Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008;26(8):1364–70.
[27] Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. CRIT CARE MED [Journal Article; Research Support, Non-US Gov't] 2007;35(9):2052–6.
[28] Pencina MJ, D'Agostino RBS, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011;30(1):11–21.
[29] Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC MED INFORM DECIS 2008;8(53).
[30] Joensuu H. Risk stratification of patients diagnosed with gastrointestinal stromal tumor. Hum Pathol 2008;39(10):1411–9.
[31] Miettinen M, Lasota J. Gastrointestinal stromal tumors: Pathology and prognosis at different sites. Semin Diagn Pathol 2006;23(2):70–83.
[32] Gal R, Rath-Wolfson L, Rosenblatt Y, Halpern M, Schwartz A, Koren R. An improved technique for mitosis counting. Int J Surg Pathol 2005;13(2):161–5.
[33] Nishida T, Goto O, Raut CP, Yahagi N. Diagnostic and Treatment Strategy for Small Gastrointestinal Stromal Tumors. CANCER-AM CANCER SOC 2016;122(20):3110–8.
[34] Tanaka J, Oshima T, Hori K, Tomita T, Kim Y, Watari J, et al. Small gastrointestinal stromal tumor of the stomach showing rapid growth and early METASTASIS to the liver. DIGEST ENDOSC 2010;22(4):354–6.
[35] Rossi S, Miceli R, Messerini L, Bearzi I, Mazzoleni G, Capella C, et al. Natural history of Imatinib-naive GISTs: a Retrospective Analysis of 929 cases with long-term follow-up and Development of a Survival Nomogram based on Mitotic Index and size as Continuous Variables. Am J Surg Pathol 2011;35(11):1646–56.
[36] Gold JS, Goenen M, Gutierrez A, Martin Broto J, Garcia-Del-Muro X, Smyrk TC, et al. Development and validation of a prognostic nomogram for recurrence-free survival after complete surgical resection of localised primary gastrointestinal stromal tumour: a retrospective analysis. Lancet Oncol 2009;10(11):1045–52.
[37] Joensuu H, Vehtari A, Riihimaki J, Nishida T, Steigen SE, Brabec P, et al. Risk of recurrence of gastrointestinal stromal tumour after surgery: an analysis of pooled population-based cohorts. Lancet Oncol 2012;13(3):265–74.
[38] Yann Lecun LBYB. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278–324.
[39] Qiangqiang Yuan QZJL. Hyperspectral image Denoising Employing a Spatial-Spectral Deep Residual Convolutional Neural Network. IEEE Transactions on Geoscience & Remote Sensing 2018:1–14.
[40] Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, et al. Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network. IEEE T MED IMAGING 2017;36(12):2524–35.
[41] Joensuu HWESH. Effect of KITand PDGFRA mutations on survival in patients with gastrointestinal stromal tumors treated with adjuvant imatinib an exploratory analysis of a randomized clinical trial. JAMA Oncol 2017;3(5):602–9.