Pattern Recognition Letters
journal homepage: www.elsevier.com

Deep Learning for Fully Automatic MRI-Based Nasopharyngeal Carcinoma Diagnosis

Yang Wang[a], Hui Ma[b], Yifan Xu[a], Xianbo Deng[b], Chuansheng Zheng[b,**], Dongrui Wu[a,**]

[a]*Ministry of Education Key Laboratory of Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, 430074, China*
[b]*Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China*

Article history:
Received XX XXX 2020
Received in final form XX XXX 2020
Accepted XX XXX 2020
Available online XX XXX 2020

Communicated by X. XXX

Keywords: Nasopharyngeal carcinoma, deep learning, MRI

ABSTRACT

Nasopharyngeal carcinoma (NPC) is common in China and Southeast Asian. This paper proposes the first fully automatic magnetic resonance imaging (MRI) based NPC diagnosis and visualization system, which can effectively accommodate images from different MRI machines and with different resolutions. It first performs adaptive segmentation and cropping of MRI slices to extract the effective brain regions, simultaneously solving the problem of different MRI machines and different image resolutions. Then, it uses a modified residual deep learning network to process the MRI slices and extract features. Finally, the high-level features of different slices are fused to output the NPC classification probability. The main advantages of our proposed approach are: 1) it has extremely high diagnosis precision – the area under the ROC curve (AUC) reached 0.994; and, 2) it can quickly visualize and locate the slices and regions where malignant tumors may exist, significantly saving radiologists' time in reviewing and annotating the MRIs.

1. Introduction

Nasopharyngeal carcinoma (NPC) is a malignant tumor that occurs in the nasopharyngeal cavity or upper throat. It's the most common head and neck cancer in southeast mainland China,Hong Kong, Taiwan, Malaysia, and Singapore [1]. The incidence rate of NPC is about 40 per 100,000 in mainland China, 25 in Hong Kong, and 27 among Malaysian Chinese. In America and Europe, the incidence rate is about 1 per 100,000.

Chemotherapy and radiotherapy are common treatments for NPC. Medical imaging can provide valuable information such as the location, volume, and severity of the tumor, which plays a critical role in clinical diagnosis and treatment of NPC. Common medical imaging modalities include endoscopy, computed tomography (CT), positron emission tomography (PET) and magnetic resonance imaging (MRI).

The current clinical practice diagnoses NPC from medical images by visual examination, and the results depend completely on the experience and subjective judgment of radiologists. Because NPC has complex and irregular structural expressions, the diagnosis results are usually uncertain and inconsistent.

There have been some studies on machine learning for NPC diagnosis. Mohammed *et al.* [2] proposed an endoscopy-based decision support system, which uses local binary pattern, gray-level co-occurrence matrix, histogram of oriented gradients, fractal dimension, etc., as features, and trains a multi-layer neural network for NPC diagnosis. Chuang *et al.* [3] employed a convolutional neural network (CNN) for nasopharyngeal biopsies. Wu *et al.* [4] proposed a staged NPC diagnosis approach from bimodal PET-CT images. It first extracts candidate lesion areas based on PET-CT images and anatomical prior knowledge, and then uses a support vector machine for classification. Zhao *et al.* [5] proposed a full CNN with an auxiliary path to segment NPC tumor regions from bimodal PET-CT images.

Compared with other imaging technologies, MRI has the advantages of non-invasion, high efficiency, and high spatial resolution. MRI can reveal anatomical structures in the nasopharynx, including tissue details deep in the pharyngeal recesses and nasopharynx. It can distinguish soft tissues, retropharyngeal lymph node metastasis, skull base invasion, peripheral nerve infiltration, and so on, and can guide the biopsy of suspicious

**Corresponding authors: Tel.: +86-027-87543130; fax: +86-027-87543130;
e-mail: hqzcsxh@sina.com, drwu@hust.edu.cn (Dongrui Wu)

areas under endoscopy in clinics [6]. As a result, MRI is widely used in NPC diagnosis and beyond [7, 8, 9, 10, 11]. Based on 3D MRI data, Korolev *et al.* [7] used 3D VGGNet [12] and a residual network [13] for the diagnosis of Alzheimer's disease and mild cognitive impairment. Pinaya *et al.* [8] used a deep belief network to extract features from MRI images for schizophrenia diagnosis. Other applications include attention deficit hyperactivity disorder [9], stroke [10], multiple sclerosis [11], etc.

MRI has also been applied to the adjuvant therapy of NPC. Most studies took advantage of MRI's high spatial resolution for the segmentation of NPC tumor regions [14, 15, 16], which is critical in radiotherapy. This paper considers MRI-based NPC diagnosis. To our knowledge, there has been only one study in this direction. Wu *et al.* [17] used an unsharp mask to enhance the edge contrast of MRI images, asked a radiologist to specify the nasopharyngeal area of interest, applied histogram equalization to remove noise, and next employed Otsu's method [18] to extract the nasopharyngeal tumor area. Finally, according to the characteristics of nasopharyngeal tissue hypertrophy and symmetric distribution, texture and geometric features were extracted from the tumor area, and a neural-fuzzy AdaBoost classifier was trained to recognize benign or malignant tumors.

The above process is complicated, and requires radiologist involvement and rich prior knowledge. This paper uses deep learning to integrate tedious feature engineering into the learning process. An end-to-end MRI-based NPC diagnosis and visualization system is proposed. It requires only a few preprocessing steps and can accommodate MRIs from different machines and with different resolutions. In addition to near-perfect NPC diagnosis accuracy (0.994 AUC), the visualization module can also quickly locate MRI slices and areas where malignant tumors may exist, significantly saving radiologists' time in reviewing and annotating the MRIs.

2. Methods

2.1. Dataset

The MRI dataset used in our experiment were collected from the Union Hospital, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, China. It consisted of 526 subjects, 326 of whom had NPC. The images were collected from different MRI machines with resolutions ranging from 208×256 to 640×640. They were labeled by one of the authors (X. Deng, an experienced radiologist).

For NPC diagnosis, an MRI machine usually scans multiple slices (each slice is an image) along the axial plane, and all these slices constitute a 3D structural scan of the entire brain. Fig. 1 shows the axial MRI images of a typical NPC patient. Several of them show clear tissue hypertrophy and compressed nasopharyngeal structure. Since the diagnosis of NPC does not require high axial density, the spatial interval between adjacent slices can be large. The number of slices per subject was between 14 and 35, mostly between 15 and 21. We only used the axial T1 structure images in our study.

Even for an NPC patient, the tumors may not present in all MRI images; only those containing tumors were labeled positive, and all others negative. For non-NPC subjects, all images were labeled negative. There were a total of 9,708 MRI images, of which 1,470 were positive and 8,238 were negative. The positive to negative ratio was about 1:5.6.

2.2. Preprocessing

As shown in Fig. 1, there are numbers, letters and black borders in the MRI images, which are not useful in NPC diagnosis. We performed preprocessing to automatically remove them. First, an opening operation (i.e., erosion and dilation) was used to remove numbers, letters, and also some tiny tissues. Then, Otsu's method [18] was employed to adaptively select a threshold to segment the foreground and background. Finally, connected component analysis was performed on the binarized image, and the largest connected component was cropped out for further classification.

Considering the low density of MRI axial scans in this problem, it is not appropriate to directly use 3D spatial convolution for classification. So, we developed a two-stage NPC diagnosis procedure. The first stage determines whether there is an NPC area in a particular MRI image (image-level diagnosis), and the second determines whether there is NPC based on all images of the subject (subject-level diagnosis).

2.3. Image-Level Diagnosis

The image-level diagnosis pipeline is shown in Fig. 2. We used a modified ResNet18, summarized in Table 1, as the backbone network. The residual network (ResNet) [13] was proposed by He *et al.* in 2016 and has been widely used in computer vision. ResNet uses identity mapping to provide a direct path for gradient back-propagation, which improves both the training speed and the generalization performance.

Every convolutional layer, except conv_fc, was followed by a batch-normalization layer [19] and a ReLU activation function. Our specific modifications were:

1. The input size was changed from the original 224×224 to 225×225 to achieve better edge alignment and spatial resolution, as NPC diagnosis mainly depends on the structural information.
2. Dropout2d, which randomly discards some channels of the feature map to reduce the feature redundancy, was used to prevent overfitting.
3. A 1×1 convolution layer was employed to give predictions, which were then integrated through global average pooling and *sigmoid* to output the final probability.

These modifications made it easy to know which part of the input image contributes more to the final prediction, which is necessary in visualization.

2.4. Visualization

Visualization of the NPC regions can help radiologists review and annotate the results. Our visualization approach was inspired by Zhou *et al.*'s work [20], with some modifications to make the derivation more rigorous and easier to understand.

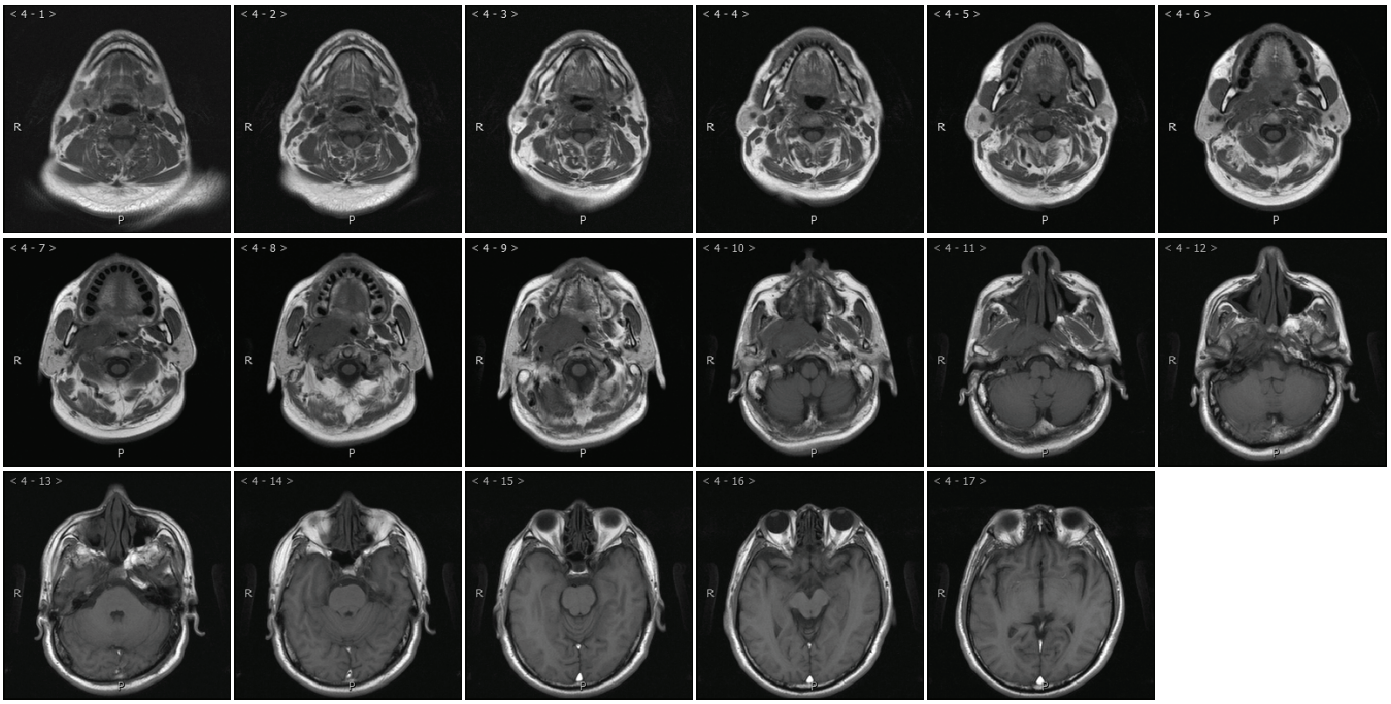

Fig. 1. MRI images from a typical NPC patient. An MRI machine scans multiple slices along the axial plane, which constitute a 3D structural scan of the entire brain.

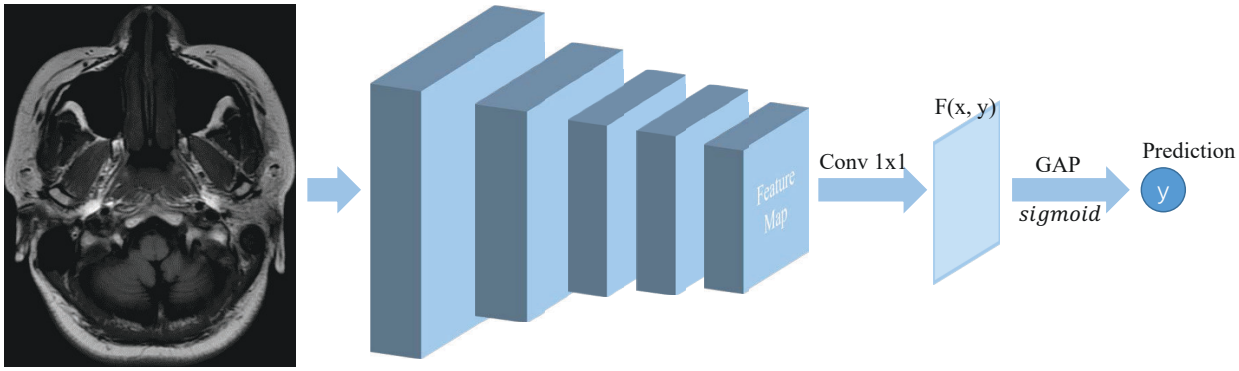

Fig. 2. Image-level diagnosis pipeline.

Let $\boldsymbol{f}$ be the feature map output by res5 in ResNet18, and $\boldsymbol{f_c}(x, y)$ the feature value at location (x, y) of Channel c. Then, the output of conv_fc is $\boldsymbol{F}(x, y) = \sum_c \boldsymbol{w}_c \boldsymbol{f}_c(x, y) + b$, where $\boldsymbol{w}_c$ is the weight of conv_fc at Channel c, and b the bias. The global average pooling output is $\boldsymbol{S} = \frac{1}{N} \sum_{x,y} \boldsymbol{F}(x, y)$, where N is the number of cells in feature map $\boldsymbol{f}$. The final prediction probability is $P = sigmoid(S)$.

In summary,

$$
\begin{aligned}
P &= sigmoid\left(\frac{1}{N}\sum_{x,y}\boldsymbol{F}(x,y)\right) \\
&= sigmoid\left(\frac{1}{N}\sum_{x,y}\left(\sum_{c}\boldsymbol{w}_c\boldsymbol{f}_c(x,y)+b\right)\right)
\end{aligned}
\tag{1}
$$

For a given image, $\boldsymbol{F}(x, y)$ represents the contribution of the location (x, y) to the final prediction. The activation of each cell can be represented by the visual mode of its corresponding receptive field. By up-sampling $\boldsymbol{F}$ to the original image resolution, we can figure out which region contributes more to the final prediction.

2.5. Subject-Level Diagnosis

Since each subject has multiple MRI images, it is intuitive to integrate the image-level diagnoses to obtain more reliable subject-level diagnosis.

We used a CNN, shared among different images, to extract their features. After that, different feature fusion strategies can be used to compute the subject-level diagnosis probability. Fig. 3 shows three feature fusion strategies considered in our study. The previous image-level model was used as a feature extractor. More specifically, the output of res5 was used as the feature, represented by a blue rectangle with text **C** in Fig. 3.

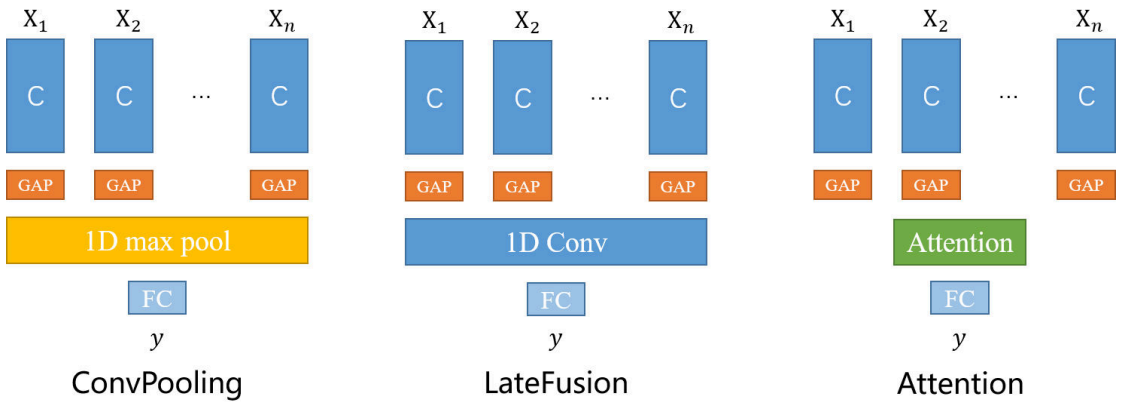

Fig. 3. Three different feature fusion strategies in subject-level diagnosis. Each feature extractor is represented by a blue rectangle with text C. 'GAP', '1D max pool', '1D Conv' and 'Attention' denote global average pooling, one-dimensional max-pooling, one-dimensional convolution and the attention mechanism, respectively. 'FC' denotes fully connected layer and *sigmoid* function.

Table 1. The modified ResNet18.

Layer	Parameter	Output Size
conv1	7×7, 64, stride 2	$113 \times 113 \times 64$
res2	3×3 max pool, stride 2	$57 \times 57 \times 64$
	$\begin{bmatrix} 3 \times 3, & 64 \\ 3 \times 3, & 64 \end{bmatrix} \times 2$	
res3	$\begin{bmatrix} 3 \times 3, & 128 \\ 3 \times 3, & 128 \end{bmatrix} \times 2$	$29 \times 29 \times 128$
res4	$\begin{bmatrix} 3 \times 3, & 128 \\ 3 \times 3, & 128 \end{bmatrix} \times 2$	$15 \times 15 \times 128$
res5	$\begin{bmatrix} 3 \times 3, & 512 \\ 3 \times 3, & 512 \end{bmatrix} \times 2$	$8 \times 8 \times 512$
dropout2d	drop rate 0.5	$8 \times 8 \times 512$
conv_fc	1×1, 1	$8 \times 8 \times 1$
Global average pooling + *sigmoid*		$1 \times 1 \times 1$

ConvPooling: The ConvPooling scheme utilized 1D max-pooling to integrate the features from different MRI images. The pooling layer had stride 1 and kernel size n (n is the number of MRI images from a subject). In ConvPooling, for each feature (channel), only the largest value in all images was selected as the feature value. Because simply taking the maximum may be sensitive to input noise, ConvPooling may lead to high false positive rate.

LateFusion: LateFusion used 1D convolution to capture spatial information between adjacent image features. It employed local connection and weight sharing, suitable for handling spatial hierarchies. The kernel size was set to 5, and the channel number 128. The output of the convolutional layer was vectorized and sent to the fully connected layer for final classification.

Attention: MRI scans are performed in a certain direction along a specific axis. Different images show different brain regions, some of which may not contain any nasopharynx area at all. Thus, it is desirable to give different images different weights.

Intuitively, images containing the nasopharynx area should be given larger weights. However, it is difficult to reliably determine whether there is nasopharyngeal area and its size. Attention mechanism, which has been widely used in applications such as machine translation [21], image caption generation [22], video question and answer [23], sleep stage classification [24], etc., can be used to calculate the weights automatically.

Let $\boldsymbol{X_i}$ be a subject's i-th MRI image, and its feature after C and GAP be $\boldsymbol{h_i}$. Then, the final feature $\boldsymbol{c}$ is the weighted sum of different image features:

$$\boldsymbol{c} = \sum_{i=1}^{n} \alpha_i \boldsymbol{h}_i \tag{2}$$

The attention weight α_i for feature $\boldsymbol{h}_i$ is computed by:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{i=1}^{n} \exp(e_i)} \tag{3}$$

where e_i can be given by a trainable fully-connected layer with $\boldsymbol{h}_i$ as input.

2.6. Training and Inference

Data partition: We used stratified sampling to partition the subjects into 60% training, 20% validation and 20% test. The models were trained on the training set. The best hyper-parameters and model were selected on the validation set. Finally, different models were compared on the test set.

Training and optimization: Since the model's final prediction was given by *sigmoid*, we adopted binary cross-entropy as the loss function. Nesterov SGD [25] was used to optimize the models, with momentum 0.9, weight decay 0.001, and number of epochs 50. A weight was given to the positive samples in the loss function due to significant class imbalance of the dataset. For image-level diagnosis, the batch size was 32, the learning

rate was 0.0005 and decayed 10 times every 20 epochs, and the weight for the positive samples was 5. For subject-level diagnosis, the batch size was 16, the learning rate was 0.0001 without decay, and the positive sample weight was 0.5. We used the parameters of the trained image-level model to initialize the feature extractor **C** in the subject-level model, accelerating the training speed and improving the generalization performance.

Early-stopping was used to decide when to stop training, by monitoring the AUC metric on the validation set. Simple hyper-parameter search was performed on the validation set. The proposed models were implemented using PyTorch and trained with a single Nvidia GeForce GTX 1080 GPU.

Data augmentation: Data augmentation can increase the size of the dataset and reduce the risk of overfitting. Before sending images to the network, we resized each image's shorter edge to 256 pixels while keeping its aspect ratio. During training, a 225×225 square image was randomly cropped from the resized image, and then flipped horizontally or vertically with equal probability. During inference (i.e., test), the image was first resized, and then the center 225×225 region was cropped out as the input.

Each subject may have different number of MRI images. During the training of the subject-level model, $n = 13$ adjacent images were randomly selected. During inference, $n = 13$ adjacent images in the middle of the image sequence were used.

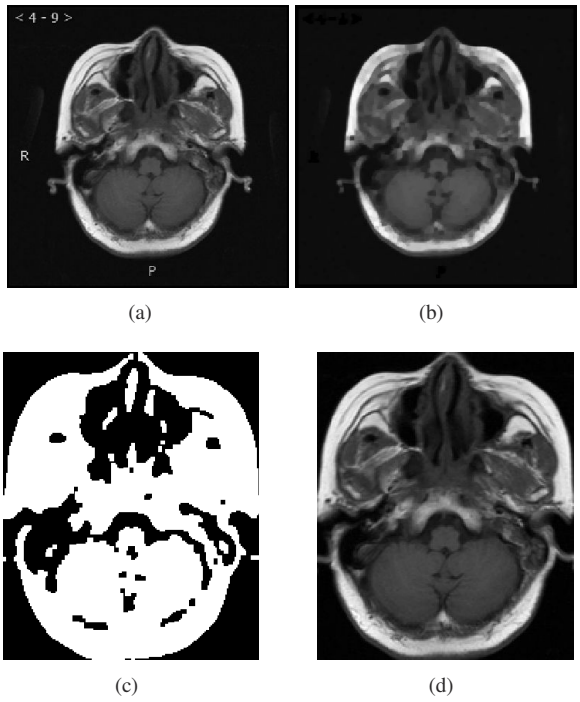

Fig. 4. Illustration of the MRI image preprocessing pipeline. (a) The original image; (b) after opening operation; (c) after Otsu method and cropping; (d) the preprocessed image.

3. Results

3.1. Performance Measures

We used the classification accuracy (ACC), sensitivity (also called recall, or true positive rate) and specificity (also called true negative rate) as performance measures. Because of significant class imbalance in the dataset, AUC and the balanced classification accuracy (BCA) were also used.

More specifically, ACC, sensitivity, specificity and BCA are defined as:

$$ACC = \frac{TP + TN}{N} \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

$$BCA = \frac{Sensitivity + Specificity}{2} \tag{7}$$

where TP, TN, FN, FP and N are the true positive, true negative, false negative, false positive and total number of samples, respectively.

3.2. Preprocessing Results

A representative image after preprocessing is shown in Fig. 4(d). Comparing Figures 4(a) and 4(d), the numbers, letters and black borders in Fig. 4(a) are successfully removed after preprocessing, and only the center informative area of the brain structure is preserved.

Although MRI images from different machines have different resolutions, annotations, and black border sizes, our preprocessing procedure can automatically accommodate them to ensure their consistency, making subsequent algorithm design much simpler.

3.3. Image-Level Diagnosis Results

Table 2 shows image-level diagnosis results with and without preprocessing. The performances without preprocessing were quite high, suggesting the effectiveness of the proposed image-level classifier. Preprocessing further improved all five performance measures.

Table 2. Image-level diagnosis results with and without preprocessing.

	Without preprocessing	With preprocessing
AUC	0.942	**0.972**
ACC (%)	90.42	**92.76**
BCA (%)	90.30	**92.44**
Sensitivity (%)	89.83	**91.99**
Specificity (%)	90.77	**92.90**

Fig. 5 shows the classification probability and visualization results of image-level diagnosis for MRI images from a patient. The classification probability is close to zero for non-NPC images, and close to one for NPC images, which are desirable. This visualization can quickly guide the radiologist to the abnormal areas, improving their diagnosis and annotation efficiency.

3.4. Subject-Level Diagnosis Results

The image-level diagnosis determines whether there is NPC based on only one image of the subject. Each subject has multiple MRI images from different brain locations. By integrating

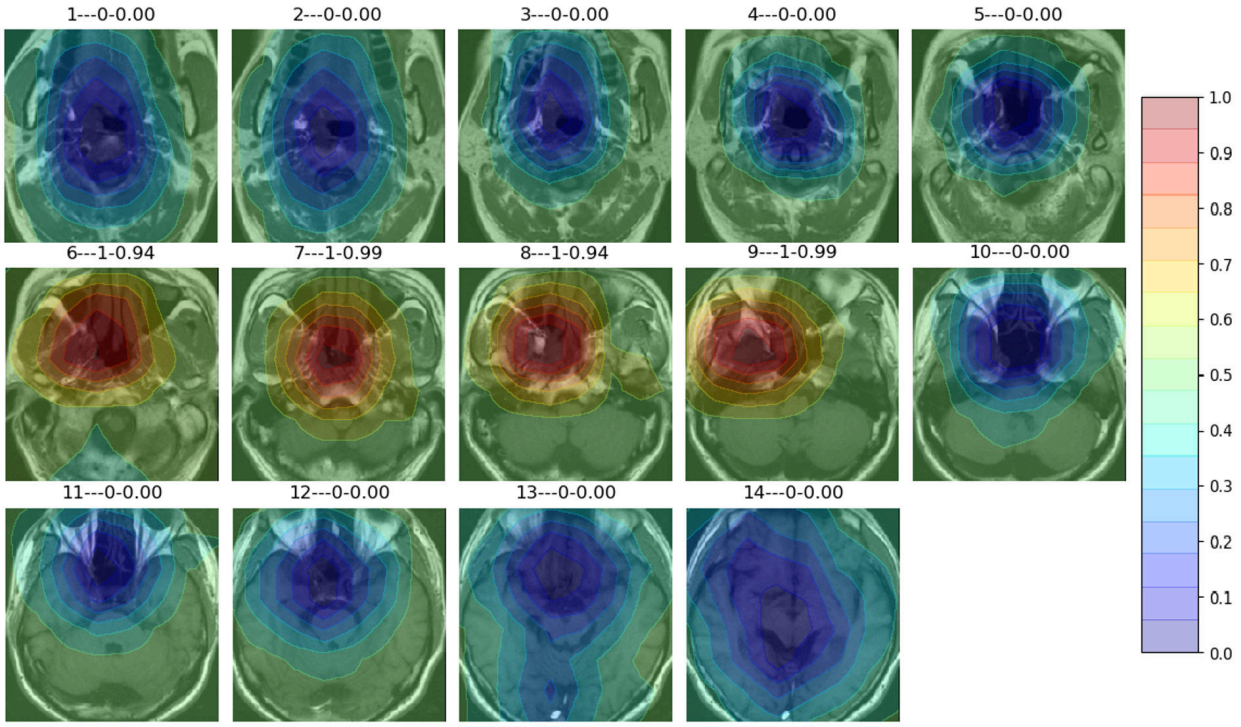

Fig. 5. Visualization of MRI images from a patient. The color bar indicates the contribution to the final classification probability. The title of each panel means 'Image index—true label (1 for positive)-the classification probability'.

the feature information of different images, more stable and accurate prediction results could be achieved.

As pointed out in Methods, pre-training is very important in constructing the subject-level model. Table 3 shows the performance of various feature fusion strategies on the validation set without pre-training, after 200 training epochs. Attention and LateFusion converged on the validation set after about 150 epochs. ConvPooling reached AUC 0.99 on the training set at 75 epochs, but converged on the validation set at 200 epochs, indicating serious overfitting.

Table 3 shows that without pre-training, Attention had the best fusion performance, whereas ConvPooling the worst. Since the performance of ConvPooling was much worse than the other two, it was not considered in further experiments.

Table 3. Subject-level diagnosis results without pre-training.

	ConvPooling	LateFusion	Attention
AUC	0.952	0.959	**0.989**
ACC (%)	86.21	89.52	**93.26**
BCA (%)	85.08	88.45	**94.89**
Sensitivity (%)	90.02	**92.57**	90.05
Specificity (%)	80.14	84.33	**99.73**

Table 4 shows the test results of subject-level diagnosis with pre-training. Overall, Attention outperformed LateFusion. For both fusion strategies, all five performance metrics were improved over the image-level diagnosis results. In particular, the specificity (i.e., true negative rate) increased by more than 5%, reducing the misdiagnosis risk of non-patients.

Fig. 6 shows the receiver characteristic curve (ROC) of

Table 4. Subject-level diagnosis results with pre-training.

	Image-level diagnosis	LateFusion	Attention
AUC	0.972	0.991	**0.994**
ACC (%)	92.76	94.85	**95.68**
BCA (%)	92.44	95.43	**96.01**
Sensitivity (%)	91.99	92.44	**93.72**
Specificity (%)	92.90	**98.42**	98.30

the Attention model. Its AUC reached an astonishing 0.994, demonstrating the effectiveness of the proposed approach. For comparison, a previous study that is most similar to ours [17], achieved an accuracy of 92.78% on a much smallest dataset (31 patients), and it required a radiologist to specify an ellipse region of interest of nasopharynx in the MRI image.

4. Conclusions

NPC has a high incidence rate in China and other Southeast Asian regions. This paper proposed a deep learning based NPC diagnosis and visualization system to assist radiologists in analyzing MRI images. It first performs adaptive segmentation and cropping of MRI slices to extract informative brain regions, making images from different MRI machines and of different resolutions more consistent. Then, it uses a modified ResNet18 to process each MRI slice, and visualizes slices and areas where malignant tumors may exist. Finally, all image-level features are integrated by an attention mechanism to give the subject-level NPC probability. Our best model reached an astonishing

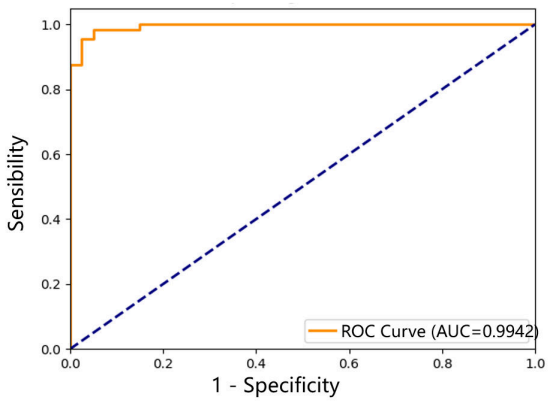

Fig. 6. ROC curve of the attention model for subject-level diagnosis.

AUC of 0.994, and the visualization can significantly save a radiologist's time in reviewing and annotating the MRI images.

To our knowledge, our proposed approach is the first fully automatic pipeline that can directly diagnose and visualize NPC from MRI images of different machines and different resolutions. It was tested on the largest NPC dataset, and achieved the highest classification accuracy.

Acknowledgement

This research was supported by the Hubei Technology Innovation Platform (2019AEA171), the National Natural Science Foundation of China (U1913207 and 61873321), and the International Science and Technology Cooperation Program of China (2017YFE0128300).

Declaration of Competing Interest

The authors declare no conflict of interest.

References

[1] S. Abdul-Kareem, S. Baba, Y. Z. Zubairi, U. Prasad, M. Ibrahim, A. Wahid, Prognostic systems for NPC: A comparison of the multi layer perceptron model and the recurrent model, in: Proc. 9th Int'l Conf. on Neural Information Processing, Orchid Country Club, Singapore, 2002, pp. 271–275.

[2] M. A. Mohammed, M. K. A. Ghani, N. a. Arunkumar, R. I. Hamed, S. A. Mostafa, M. K. Abdullah, M. Burhanuddin, Decision support system for nasopharyngeal carcinoma discrimination from endoscopic images using artificial neural network, The Journal of Supercomputing 76 (2020) 1086–1104.

[3] W.-Y. Chuang, S.-H. Chang, W.-H. Yu, C.-K. Yang, C.-J. Yeh, S.-H. Ueng, Y.-J. Liu, T.-D. Chen, K.-H. Chen, Y.-Y. Hsieh, Y. Hsia, T.-H. Wang, C. Hsueh, C.-F. Kuo, C.-Y. Yeh, Successful identification of nasopharyngeal carcinoma in nasopharyngeal biopsies using deep learning, Cancers 12 (2) (2020) 507.

[4] B. Wu, P.-L. Khong, T. Chan, Automatic detection and classification of nasopharyngeal carcinoma on PET/CT with support vector machine, International Journal of Computer Assisted Radiology and Surgery 7 (4) (2012) 635–646.

[5] L. Zhao, Z. Lu, J. Jiang, Y. Zhou, Y. Wu, Q. Feng, Automatic nasopharyngeal carcinoma segmentation using fully convolutional networks with auxiliary paths on dual-modality PET-CT images, Journal of Digital Imaging 32 (3) (2019) 462–470.

[6] A. D. King, A. C. Vlantis, K. S. Bhatia, B. C. Zee, J. K. Woo, G. M. Tse, A. T. Chan, A. T. Ahuja, Primary nasopharyngeal carcinoma: Diagnostic accuracy of MR imaging versus that of endoscopy and endoscopic biopsy, Radiology 258 (2) (2011) 531–537.

[7] S. Korolev, A. Safiullin, M. Belyaev, Y. Dodonova, Residual and plain convolutional neural networks for 3D brain MRI classification, in: International Symposium on Biomedical Imaging (ISBI), Melbourne, Australia, 2017, pp. 835–838. doi:10.1109/ISBI.2017.7950647.

[8] W. H. Pinaya, A. Gadelha, O. M. Doyle, C. Noto, A. Zugman, Q. Cordeiro, A. P. Jackowski, R. A. Bressan, J. R. Sato, Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia, Scientific Reports 6 (2016) 38897.

[9] X. Han, Y. Zhong, L. He, S. Y. Philip, L. Zhang, The unsupervised hierarchical convolutional sparse auto-encoder for neuroimaging data classification, in: Proc. 8th Int'l Conf. on Brain Informatics and Health (ICBIH), London, UK, 2015, pp. 156–166.

[10] T. Schmah, G. E. Hinton, S. L. Small, S. Strother, R. S. Zemel, Generative versus discriminative training of RBMs for classification of fMRI images, in: Proc. Advances in Neural Information Processing Systems (NIPS), Vancouver, Canada, 2009, pp. 1409–1416.

[11] T. Brosch, L. Y. W. Tang, Y. Yoo, D. K. B. Li, A. Traboulsee, R. Tam, Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation, IEEE Trans. on Medical Imaging 35 (5) (2016) 1229–1239.

[12] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proc. Int'l Conf. on Learning Representations (ICLR), San Diego, CA, 2015, pp. 1–14.

[13] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770–778.

[14] W. Huang, K. L. Chan, J. Zhou, Region-based nasopharyngeal carcinoma lesion segmentation from MRI using clustering-and classification-based methods with learning, Journal of Digital Imaging 26 (3) (2013) 472–482.

[15] N. Nabizadeh, M. Kubat, Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features, Computers & Electrical Engineering 45 (2015) 286–301.

[16] Z. Ma, S. Zhou, X. Wu, H. Zhang, W. Yan, S. Sun, J. Zhou, Nasopharyngeal carcinoma segmentation based on enhanced convolutional neural networks using multi-modal metric learning, Physics in Medicine & Biology 64 (2) (2019) 025005.

[17] Ming-Chi Wu, Wen-Chi Chin, Ting-Chen Tsan, Chiun-Li Chin, The benign and malignant recognition system of nasopharynx in MRI image with neural-fuzzy based adaboost classifier, in: Int'l Conf. on Information Management (ICIM), London, UK, 2016, pp. 47–51.

[18] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. on Systems, Man, and Cybernetics 9 (1) (1979) 62–66.

[19] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proc. 32nd Int'l Conf. on Machine Learning, Lille, France, 2015, pp. 448–456.

[20] B. Zhou, A. Khosla, À. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2921–2929.

[21] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, in: Proc. Int'l Conf. on Learning Representations (ICLR), San Diego, CA, 2014.

[22] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, Y. Bengio, Show, attend and tell: Neural image caption generation with visual attention, in: Proc. 32nd Int'l Conf. on Machine Learning (ICML), Lille, France, 2015, pp. 2048–2057.

[23] K. M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, P. Blunsom, Teaching machines to read and comprehend, in: Proc. Advances in Neural Information Processing Systems (NIPS), Montreal, Canada, 2015, pp. 1693–1701.

[24] Y. Wang, D. Wu, Deep learning for sleep stage classification, in: Chinese Automation Congress (CAC), Xi'an, China, 2018, pp. 3833–3838.

[25] Y. Nesterov, A method of solving a convex programming problem with convergence rate $o(1/k^2)$, in: Soviet Mathematics Doklady, Vol. 27, 1983, pp. 372–376.