European Radiology
https://doi.org/10.1007/s00330-020-07451-y

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Convolutional neural network for discriminating nasopharyngeal carcinoma and benign hyperplasia on MRI

Lun M. Wong[1] · Ann D. King[1] · Qi Yong H. Ai[1] · W. K. Jacky Lam[2] · Darren M. C. Poon[3] · Brigette B. Y. Ma[3] · K. C. Allen Chan[2] · Frankie K. F. Mo[3]

Received: 4 June 2020 / Revised: 25 September 2020 / Accepted: 30 October 2020

Abstract

Objectives A convolutional neural network (CNN) was adapted to automatically detect early-stage nasopharyngeal carcinoma (NPC) and discriminate it from benign hyperplasia on a non-contrast-enhanced MRI sequence for potential use in NPC screening programs.

Methods We retrospectively analyzed 412 patients who underwent T2-weighted MRI, 203 of whom had biopsy-proven primary NPC confined to the nasopharynx (stage T1) and 209 had benign hyperplasia without NPC. Thirteen patients were sampled randomly to monitor the training process. We applied the Residual Attention Network architecture, adapted for three-dimensional MR images, and incorporated a slice-attention mechanism, to produce a CNN score of 0–1 for NPC probability. Threefold cross-validation was performed in 399 patients. CNN scores between the NPC and benign hyperplasia groups were compared using Student's *t* test. Receiver operating characteristic with the area under the curve (AUC) was performed to identify the optimal CNN score threshold.

Results In each fold, significant differences were observed in the CNN scores between the NPC and benign hyperplasia groups ($p < .01$). The AUCs ranged from 0.95 to 0.97 with no significant differences between the folds ($p = .35$ to $.92$). The combined AUC from all three folds ($n = 399$) was 0.96, with an optimal CNN score threshold of > 0.71, producing a sensitivity, specificity, and accuracy of 92.4%, 90.6%, and 91.5%, respectively, for NPC detection.

Conclusion Our CNN method applied to T2-weighted MRI could discriminate between malignant and benign tissues in the nasopharynx, suggesting that it as a promising approach for the automated detection of early-stage NPC.

Key Points

- *The convolutional neural network (CNN)–based algorithm could automatically discriminate between malignant and benign diseases using T2-weighted fat-suppressed MR images.*
- *The CNN-based algorithm had an accuracy of 91.5% with an area under the receiver operator characteristic curve of 0.96 for discriminating early-stage T1 nasopharyngeal carcinoma from benign hyperplasia.*
- *The CNN-based algorithm had a sensitivity of 92.4% and specificity of 90.6% for detecting early-stage nasopharyngeal carcinoma.*

Keywords Nasopharyngeal carcinoma · Hyperplasia · Deep learning · Computational neural network · Early detection of cancer

✉ Ann D. King
king2015@cuhk.edu.hk

[1] Department of Imaging and Interventional Radiology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR

[2] Department of Chemical Pathology, State Key Laboratory of Translational Oncology, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR

[3] Department of Clinical Oncology, State Key Laboratory of Translational Oncology, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR

Published online: 25 November 2020

Abbreviations

3D	Three-dimensional
AUC	Area under the curve
CI	Confidence interval
CNN	Convolutional neural network
DNN	Deep neural network
GPU	Graphic processing unit
MRI	Magnetic resonance imaging
NPC	Nasopharyngeal carcinoma
NPV	Negative predictive value
PPV	Positive predictive value
RAN	Residual Attention Network
ROC	Receiver operator characteristic

Introduction

Recent advancements in computational power, especially graphic processing units and the invention of gradient backward propagation [1], have allowed artificial neural networks to achieve exceedingly deep architectures while remaining relatively easy to train. These networks, often termed deep neural networks (DNNs), frequently show superior performance compared to other machine learning techniques [2–4]. Among the various DNNs, the convolutional neural network (CNN), which exploits extensive stacks of learnable convolutional filters, has been shown to be one of the most efficacious for medical imaging [5].

Broad adaptations of CNNs are being used to address problems in a wide range of areas in radiology, including denoising [6], super-resolution [7], and pseudo-modality conversion [7, 8]. Specifically, in magnetic resonance imaging (MRI), CNNs have already contributed to the reduction of gadolinium dose [9], correction of artifacts [10], and acceleration of scan time [11]. In cancer imaging using MRI, CNNs have shown promise as an effective tool for automating tumor segmentation [12], detection [13], and classification [14]. The direct use of CNNs to discriminate between benign and malignant tissues on MRI is an emerging area of research for cancers of the brain [15], breast [16], liver [17], prostate [18], and kidney [19].

We were interested in the potential of using CNN in the head and neck MRI to discriminate between benign and malignant tissues in the nasopharynx to facilitate the MRI detection of endoscopically occult early-stage nasopharyngeal carcinoma (NPC) [20, 21]. We have established the complementary role of MRI to endoscopic examination in the detection of NPC [20, 21] and published the only MRI grading system for MRI detection and discrimination of NPC from benign hyperplasia [20, 22]. A technique that can discriminate early-stage NPC confined within the nasopharynx from the wall thickening and enlargement of the adenoid caused by benign hyperplasia could improve the diagnostic performance of MRI [23, 24] and facilitate the automatic detection of NPC in patients referred for investigation from plasma Epstein–Barr virus DNA screening programs [25]. In this regard, we were especially interested in evaluating the performance of T2-weighted MRI, which is widely available, rapid, and does not require intravenous injection of an MRI contrast agent. We also planned to incorporate three-dimensional (3D) volume analysis into our CNN architecture. Notably, all of the currently established CNN architectures were designed for natural two-dimensional images, but for cancer imaging, a 3D approach is more desirable to cover the whole tumor and allow appreciation of inter-slice relationships.

The aim of this MRI study, therefore, was to investigate the use of a 3D CNN for the discrimination of early-stage NPC from benign hyperplasia using a T2-weighted sequence.

Materials and methods

Patients

This retrospective study was approved by the local institutional board, and the requirement of written consent was waived. Two groups of ethnically Chinese patients who underwent T2-weighted fat-suppressed MRI of the nasopharynx in our institution between 2010 and 2019 were included in the study: (i) those with newly biopsy-proven undifferentiated NPC whose primary tumors were staged as T1 with a thickness ≧ 5 mm on at least one axial slice and (ii) those with benign hyperplasia with a thickness ≧ 5 mm on at least one axial slice without the evidence of NPC on MRI or endoscopic examination. In addition, patients with benign hyperplasia were followed up for a minimum of 12 months (mean of 41 months, median of 35 months, and range 12 to 80 months). A total of 412 patients (mean age = 53.6 ± 9.5 years; 371 men and 41 women), i.e., 203 with stage T1 NPC (mean age = 53.4 ± 11.8 years; 162 men and 41 women) and 209 with benign hyperplasia (mean age = 53.8 ± 6.8 years; all men), were included. Primary tumors ranged in size from 5.0 to 19.6 mm (median of 7.0 mm) and benign hyperplasia ranged in size from 5.0 to 20.0 mm (median of 6.7 mm). The study flow diagram of patients who underwent T2-weighted fat-suppressed MRI is provided in Fig. 1.

MRI acquisition

MRI was performed using a Philips Achieva TX 3.0 T machine containing a body coil for radiofrequency transmission and a 16-channel Philips neurovascular phased-array coil for reception. Patients were scanned with standard sequences centered at the nasopharynx, including an axial MRI T2-weighted fat-suppressed, using SPAIR (spectral attenuated inversion recovery), turbo spin-echo sequence acquired using a repetition time/echo time of 4060/80 ms, turbo spin-echo factor of

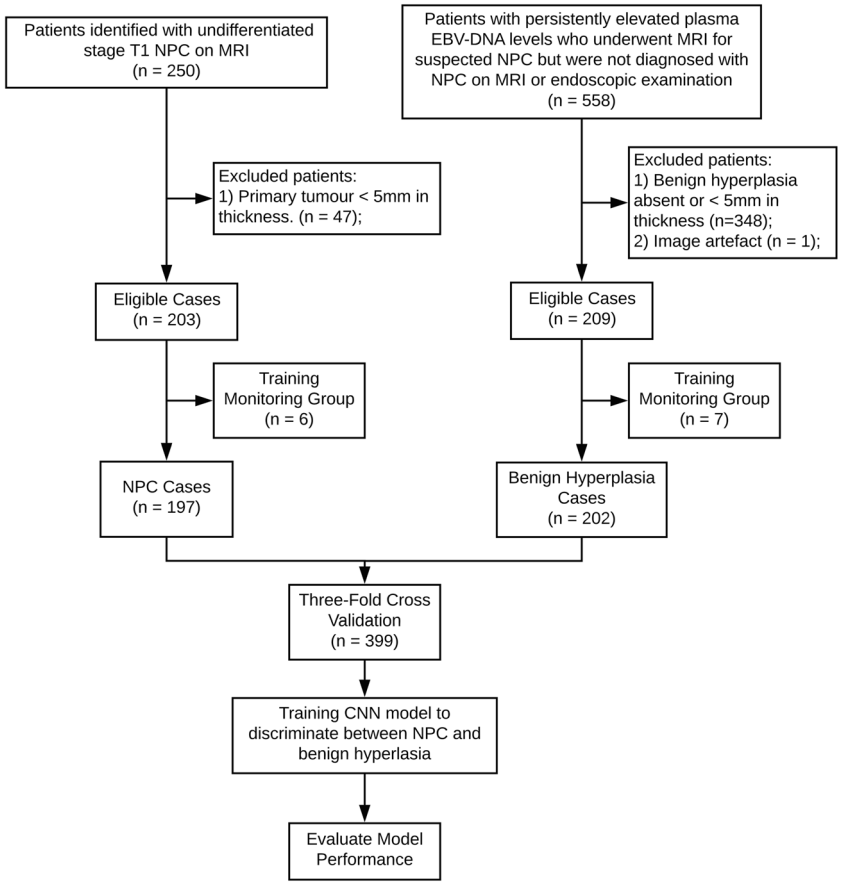

Fig. 1 Study flow of participants. MRI, magnetic resonance imaging; NPC, nasopharyngeal carcinoma; EBV, Epstein–Barr virus; CNN, convolutional neural network

17, field of view of 230 × 230 mm, number of slices of 30–36, sensitivity encoding factor of 1, slice thickness of 4 mm, and final dimension of 512 × 512 for each axial image with a pixel size of 0.45 × 0.45 mm.

Algorithm architecture

Preprocessing

The images were normalized using an automatic pipeline in which the input T2-weighted fat-suppressed axial images were centered at the center of mass (defined as the intensity-weighted mean coordinate of the image) and rotated to align with the plane of symmetry determined using Extraction of Robust Orientation using Symmetry [26]. Subsequently, the intensity profiles were normalized with Nyul normalization [27] and finally cropped to a uniform dimension of 20 × 444 × 444.

Classification with deep neural network

We adopted a CNN architecture called the Residual Attention Network (RAN) [28] as our backbone and modified the network to accept and process 3D inputs. The RAN is a CNN for the classification of 2D images with learnable attention modules. The backbone of the RAN adopts an auto-encoder architecture, where each layer is a residual attention module composed of two parallel branches, a soft-mask branch and a residual trunk branch. While the residual trunk branch learns global features, the soft-mask branch learns to compute a weighted mask for local features. Specifically, we modified part of the convolutional kernels to 3D, added dropout layers to improve the robustness of the learned model, and included a slice-attention mechanism by adding slice-wise attention modules to weight the importance of each axial slice according to its features. The details of network architectures are described in the supplementary materials. For each patient, the learned network produced a score for the probability of NPC positive (from 0 to 1).

Training and validation

The network was implemented and trained with PyTorch v1.2, an open-source Python package, dedicated to artificial neural network computation [29], on a machine equipped with 4 Nvidia 2080 Ti graphic processing units (GPUs). A small set of data unseen by the network was used to monitor it for overfitting during the course of training [30] comprising 13 patients who were randomly sampled from the total group of 412, 6 of whom had NPC and 7 of whom had benign hyperplasia. For threefold cross-validation, the remaining patients

were divided randomly with stratification into three partitions, namely folds 01, 02, and 03, each containing 133 patients. In each round of cross-validation, the network was re-trained from scratch and allowed to iterate for 5500 epochs. The network was initialized with the Kaiming initialization [31] except for linear layers, in which the weights were uniformly set to one with a zero bias, and then was optimized with cross-entropy loss by Adam [32].

Results were compared with those of a radiologist with more than 20 years of experience blinded to the results, who classified the MRI into NPC or benign hyperplasia based on the assessment of only the T2-weighted images according to a previously published grading system for NPC detection designed for use with a non-contrast-enhanced MRI sequence [22]. In brief, MRI scans designated grades 1–3 were negative for NPC, which included scans showing a normal nasopharynx, diffuse symmetrical wall thickening or diffuse asymmetrical wall thickening that is non-expansile, adenoid symmetrically enlarged or composed entirely of cysts, while MRI scans designated grade 4 or 5 were positive for early-stage NPC which included scans showing a diffuse asymmetrical wall thickening that is expansile, focal lesion, or asymmetrically enlarged adenoid [22].

Statistical analysis

Analysis of variance and independent Student's *t* test were used to test for differences in the CNN scores across the three folds and between the NPC and benign hyperplasia groups in each fold, respectively.

For each fold, a receiver operating characteristic (ROC) curve was plotted to obtain the area under the curve (AUC) for discriminating between NPC and benign hyperplasia. The DeLong test was used to compare the AUCs in each fold [33]. The optimum threshold for the CNN score to detect NPC, obtained using the Youden index, was used to calculate the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy. The 95% confidence intervals (CIs) of AUC, sensitivity, and specificity were estimated by bootstrapping. McNemar's test was used to compare the performance of the radiologist and the CNN.

Multivariable linear regression was used to determine whether the presence or absence of a nasopharyngeal biopsy was a confounding factor in the CNN score.

Statistical analysis was performed using the commercial software IBM SPSS Statistics for Windows, version 25.0, and an open-source library pROC [34]. The significance of differences was accepted at $p < .05$.

Results

No significant differences were observed in the CNN scores across the folds for patients with NPC ($p = .71$) and for those with benign hyperplasia ($p = .15$). In each fold, the CNN score for patients with NPC was significantly higher than that for patients with benign hyperplasia ($p < .01$). Details of the comparisons are shown in Table 1. Figure 2 shows T2-weighted MR images of a patient with NPC (Fig. 2a) and one with benign hyperplasia (Fig. 2b).

The AUC of the CNN score in each fold ranged from 0.95 to 0.97 (Table 1) and showed no significant difference between the folds ($p = .41$ to $.80$) (Table 2). The combined AUC from all folds showed no significant difference from the AUC of any of the three folds ($p = .48$ to $.77$) (Table 2).

The receiver operating characteristic (ROC) curves were plotted in Fig. 3. The combined AUC was 0.96 (95% CI: 0.94, 0.98) with an optimal CNN score threshold of > 0.71, producing 182 true-positive, 183 true-negative, 19 false-positive, and 15 false-negative results for NPC. Table 3 summarizes the computed CNN scores of patients with NPC and benign hyperplasia. The sensitivity, specificity, PPV, NPV, and accuracy of the CNN were 92.4% (95% CI: 85.8%, 95.9%), 90.6% (95% CI: 72.8%, 95.1%), 90.5%, 92.4%, and 91.5%, respectively, for NPC detection.

The sensitivity, specificity, PPV, NPV, and accuracy of the radiologist were 91.9%, 82.2%, 83.4, 91.2%, and 87.0% respectively and showed no significant difference for NPC detection when compared to the CNN ($\chi^2 = 3.37$, $p = 0.07$).

Nasopharyngeal biopsy had been performed before the MRI in 197/197 (100%) patients with NPC and 12/202 (5.9%) patients with benign hyperplasia. Furthermore, 182/182 (100%), 2/19 (10.5%), 10/183 (5.5%), and 15/15 (100%) patients with true-positive, false-positive, true-negative, and false-negative results for NPC, respectively, had undergone a nasopharyngeal biopsy. Multivariate linear regression showed that biopsy was not a significant variable ($p = .84$) for the CNN scores when adjusted for the actual presence/absence of NPC ($p < .01$).

Discussion

We developed a promising and effective CNN-based fully automated algorithm for discrimination between malignant and benign tissues in the nasopharynx. Using a CNN scoring system from 0 to 1, a threshold of > 0.71 was found to discriminate between NPC confined within the nasopharynx and benign hyperplasia, enabling the CNN to detect early-stage NPC with a high sensitivity of 92.4% and a specificity of 90.6%. Most cancers were classified with a high level of confidence, as indicated by CNN scores close to 1 for NPC, backed by a high AUC of 0.96. The results show promises for expanding the role of MRI in plasma Epstein–Barr virus DNA NPC screening programs, which are key to improving the prognostic outcome of this disease [25]. Notably, for NPC detection, the overall performance of the CNN was similar to

Table 1 Computed CNN scores for NPC detection (range 0–1). CNN scores of the three folds were combined for the overall analysis

Folds	Patients with NPC (N = 197)			Patients without NPC (N = 202)			p values
01	0.90	±	0.26	0.11	±	0.28	$p < .01$
02	0.92	±	0.26	0.17	±	0.33	$p < .01$
03	0.94	±	0.22	0.08	±	0.23	$p < .01$
Overall	0.92	±	0.25	0.12	±	0.29	$p < .01$

A p value < .05 indicates statistical significance. Data are presented as means ± standard deviations. *CNN*, convolutional neural network; *NPC*, nasopharyngeal carcinoma

that of an experienced radiologist in NPC screenings [20, 22] with improved specificity which is important in screening programs where false-positive results should be kept to a minimum to reduce unnecessary biopsies. Furthermore, this is one of only a few MRI studies on cancer that have used a CNN as an automated machine learning method to discriminate between malignant and benign tissues in the nasopharynx [35], whereas the other related studies have examined cancers outside the head-and-neck region [15–19].

This method also has the advantage of not requiring manual delineation of nasopharyngeal lesions, which is both laborious and requires a high level of expertise [36]. In this regard, our automated method might overcome the practical issues of manual contouring that currently is a major hurdle to the implementation of traditional machine learning models into clinical practice. Recently, an automated method has been successfully implemented for the assessment of prostate cancer [37]. Our method is also rapid; specifically, using a GPU with at least one Nvidia 2080 Ti, the computational time was less than a second for each case in our study, making this method way faster than one that requires manual delineation. Moreover, the method could be applied using a T2-weighted sequence, which is desirable in a screening setting as it avoids T1-weighted contrast-enhanced MRI sequences, thereby reducing costs and, wherever possible, avoids the injection of MRI contrast agents in patients who turn out to have no cancer. Non-contrast-enhanced MRI sequences, such as diffusion-weighted imaging [36, 38], are also candidates for screening, but this study focused on T2-weighted imaging because it is a standard, widely available sequence with few artifacts in the head-and-neck region.

The most challenging component of the proposed method was to redesign the DNN architecture so that the network could accept and process 3D image volumes rather than two-dimensional images. Currently, no established classification DNN architecture is available for 3D image volume analysis because 3D analysis, with the exponential increase in its complexity due to higher number of dimensions, adds burden to both hardware requirements and training. Nevertheless, a few pioneers have described the detect-then-classify strategy by expanding a renowned two-dimensional DNN architecture, Faster-RCNN, to 3D [13, 39, 40]. This was achieved by first identifying a rectangular cuboid region of interest, followed by classifying the cropped region of interest. This strategy is particularly advantageous when the region of interest is much smaller than the whole image volume and has a more-or-less constant size with a clearly identifiable boundary. Unfortunately, this is not the case when imaging the nasopharynx. However, we were able to exploit the thick MRI slices acquired through the nasopharynx by using an anisotropic convolutional kernel in the CNN. This substantially reduces the hardware requirements for 3D analysis, allowing us to expand and redesign the RAN to overcome the limitation in analyzing our 3D data. In addition, our method introduced a

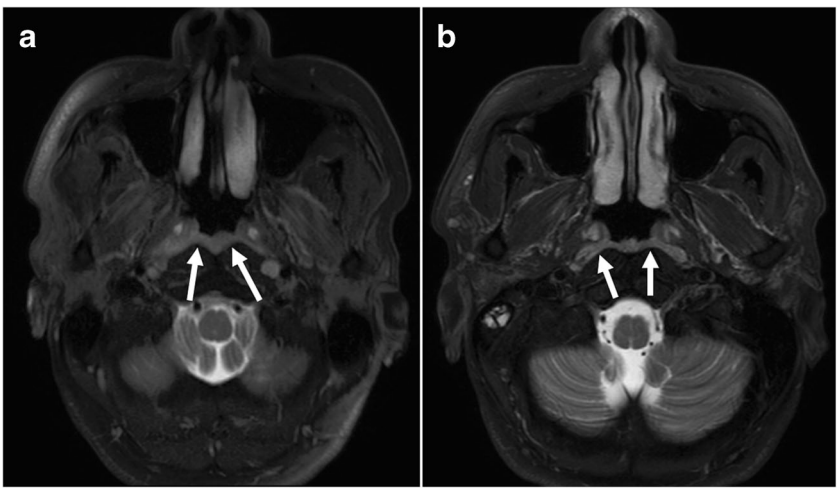

Fig. 2 MR T2-weighted fat-suppressed axial image of the nasopharynx in a patient with stage T1 NPC (white arrows) (CNN score ~ 1.00) (**a**) and benign hyperplasia (white arrows) (CNN score < 0.01) (**b**). NPC, nasopharyngeal carcinoma; CNN, convolutional neural network

Table 2 The DeLong test *p* values for differences in the AUC of the ROC plots of each fold. No significant differences in AUCs across folds and between individual folds and the combined analysis were observed

		AUC	Folds		Combined (AUC = 0.96)
			01	02	
Folds	01	0.97	-	-	.48
	02	0.95	.41	-	.72
	03	0.97	.80	.60	.77

A *p* value < .05 indicates statistical significanc. *ROC*, response operator characteristic; *AUC*, area under the ROC curve

slice-wise attention module, inspired by the Squeeze-and-Excitation network [41], on top of the original attention mechanism, that weighted the importance of each input axial slices. In doing so, the network could take into consideration that not all slices in a volume are equally important, which is similar to the approach followed in clinical practice wherein the radiologist tends to select the most relevant slices for diagnosis. We believe that this slice-wise attention module could be a key player in the future development of 3D classification DNNs.

The performance of our study is lower than that published by Ke et al in which the sensitivity, specificity, PPV, NPV, and accuracy were 99.7%, 91.7%, 97.5%, 98.9%, and 97.8% in detecting NPC and discriminating NPC from benign hyperplasia [35]. However, that study evaluated CE-T1W images in a large group of NPC patients with primary tumors ranging from early-stage (T1 and T2) to late-stage (T3 and T4) disease. Their subgroup analysis of stage T1 cancers, involving 64 patients with stage T1 NPC and 96 patients with benign

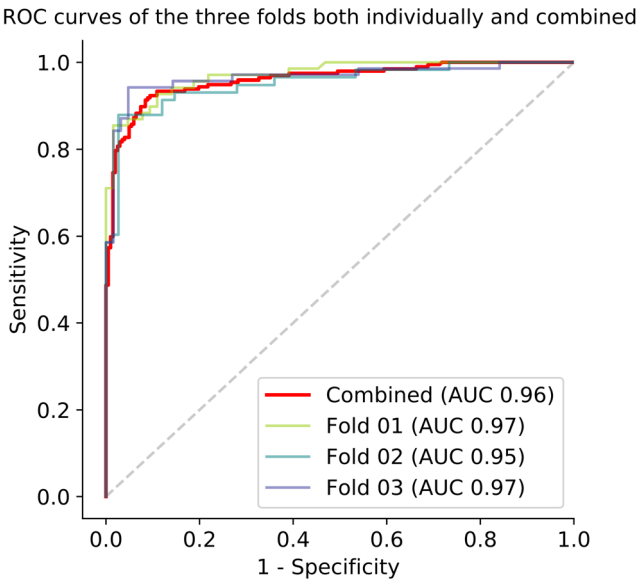

Fig. 3 The ROC plot of the three-fold cross-validation experiment. As there is no significant difference between the curves of the three folds, a combined ROC curve was considered in further analysis. ROC, receiver operator characteristic; AUC, area under the ROC curve

disease (89 patients with benign hyperplasia and 7 patients with chronic inflammation), found similar performance but there was a reduction in the PPV from 97.5 to 88.7%. It is known that the MRI grading systems for NPC detection perform better on contrast-enhanced images, so it is not surprising that our T2-weighted imaging-based CNN did not perform as well as one based on contrast-enhanced images, although the PPV was similar. However, our study has the advantage that it was designed with NPC screening programs in mind with the aim of avoiding the use of an intravenous contrast agent. Moreover, our CNN did not require manual delineation of the nasopharyngeal lesions for training and included only the most difficult group of NPCs to detect and discriminate from benign hyperplasia, i.e., stage T1 cancers which are the early cases where MRI adds value to the endoscopic examination.

Our study has limitations. First, the technique is only applicable to MRI scans of the head-and-neck region in which the field of view is centered on the nasopharynx. Deviating from this setting will result in inaccurate cropping and, consequently, impaired network performance. Second, we could not validate our findings on an external database because screening is a new area for MRI; therefore, it is difficult to obtain MRI scans centered on the nasopharynx in patients with benign hyperplasia. Third, it is inevitable that patients with NPC undergoing radiomic or deep learning analysis will have a biopsy before being referred for the MRI staging scan. However, our analysis showed no association between the CNN score and the presence or absence of a nasopharyngeal biopsy. Fourth, it is not possible to entirely exclude the possibility of very early NPC in patients with benign hyperplasia even after biopsy, but based on MRI and endoscopic examination plus follow-up with a mean of 41 months, we believe the risk of an undetected NPC in this group is very small.

In conclusion, we have tested a fully automated CNN method to discriminate between malignant and benign tissues

Table 3 CNN results of the differentiation between NPC and benign hyperplasia on MRI. Data are the number of NPC or benign hyperplasia patients who fall within the CNN score intervals

CNN scores intervals	NPC	Benign
0–0.1	12	166
> 0.1–0.2	1	3
> 0.2–0.3	0	5
> 0.3–0.4	0	4
> 0.4–0.5	0	1
> 0.5–0.6	2	1
> 0.6–0.7	0	3
> 0.7–0.8	2	1
> 0.8–0.9	6	3
> 0.9–1.0	174	15
Total	197	202

CNN, convolutional neural network; *NPC*, nasopharyngeal carcinomas

in the head and neck region using a T2-weighted MRI sequence. Our results showed that in the discrimination between early-stage NPC and benign hyperplasia in the nasopharynx, the CNN achieved an overall sensitivity of 92.4% and specificity of 90.6%, with an overall AUC of 0.96 for NPC detection. The ability of the network to automatically classify lesions based on a sequence that does not require the injection of an intravenous MRI contrast agent shows the potential of using CNN in NPC screening programs.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-020-07451-y.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Ann D. King.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the co-authors is a biostatistic expert and had provided support in the statistical analysis of this study.

Informed consent Written informed consent was waived by the local Institutional Review Board.

Ethical approval Local Institutional Review Board approval was obtained.

Study subjects or cohorts overlap A total of 412 patients were included in this study, of which 151 with nasopharyngeal carcinoma and 102 with benign hyperplasia have been included in two previous studies published on American Journal of Neuroradiology entitled "MR imaging criteria for the detection of nasopharyngeal carcinoma: discrimination of early-stage primary tumors from benign hyperplasia" (10.3174/ajnr.A5493) and "Early detection of cancer: evaluation of mr imaging grading systems in patients with suspected nasopharyngeal carcinoma" (10.3174/ajnr.A6444).

Methodology
- Retrospective
- Case-control study
- Performed at one institution

References

1. Werbos PJ (1990) Backpropagation through time: what it does and how to do it. Proc IEEE 78:1550–1560
2. Soniya PS, Singh L (2015) A review on advances in deep learning. IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions 2015:1–6
3. Rachmadi M, Valdés-Hernández M, Agan M, Komura T (2017) Deep learning vs. conventional machine learning: pilot study of WMH segmentation in brain MRI with absence or mild vascular pathology. J Imaging 3:66
4. Rabhi S, Jakubowicz J, Metzger M-H (2019) Deep learning versus conventional machine learning for detection of healthcare-associated infections in French clinical narratives. Methods Inf Med 58:31–41
5. Bakator M, Radosav D (2018) Deep learning and medical diagnosis: a review of literature. Multimodal Technol Interact 2:47
6. Kaur P, Singh G, Kaur P (2018) A review of denoising medical images using machine learning approaches. Curr Med Imaging Rev 14:675–685
7. Kaji S, Kida S (2019) Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. Radiol Phys Technol 12:235–248
8. Florkow MC, Zijlstra F, Willemsen K et al (2020) Deep learning–based MR-to-CT synthesis: the influence of varying gradient echo–based MR images as input channels. Magn Reson Med 83:1429–1441
9. Gong E, Pauly JM, Wintermark M, Zaharchuk G (2018) Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J Magn Reson Imaging 48:330–340
10. Tamada D, Kromrey M-L, Onishi H, Motosugi U (2018) Method for motion artifact reduction using a convolutional neural network for dynamic contrast enhanced MRI of the liver. Available via https://arxiv.org/abs/1807.06956. Accessed 18 Jul 2018
11. Lee D, Yoo J, Tak S, Ye JC (2018) Deep residual learning for accelerated MRI using magnitude and phase networks. IEEE Trans Biomed Eng 65:1985–1995
12. Li Q, Xu Y, Chen Z et al (2018) Tumor segmentation in contrast-enhanced magnetic resonance imaging for nasopharyngeal carcinoma: deep learning with convolutional neural network. Biomed Res Int 2018:1–7
13. Nasrullah N, Sang J, Alam MS, Mateen M, Cai B, Hu H (2019) Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors (Basel) 19:3722
14. Mazurowski MA, Buda M, Saha A, Bashir MR (2019) Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI. J Magn Reson Imaging 49:939–954
15. Mohsen H, El-Dahshan E-SA, El-Horbaty E-SM, Salem A-BM (2018) Classification using deep learning neural networks for brain tumors. Futur Comput Informatics J 3:68–71
16. Antropova N, Huynh BQ, Giger ML (2017) A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Med Phys 44:5162–5171
17. Hamm CA, Wang CJ, Savic LJ et al (2019) Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. Eur Radiol 29:3338–3347
18. Yang X, Liu C, Wang Z et al (2017) Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI. Med Image Anal 42:212–227
19. Xi IL, Zhao Y, Wang R et al (2020) Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging. Clin Cancer Res 26:1944–1952
20. King AD, Vlantis AC, Bhatia KSS et al (2011) Primary nasopharyngeal carcinoma: diagnostic accuracy of MR imaging versus that of endoscopy and endoscopic biopsy. Radiology 258:531–537
21. King AD, Woo JKS, Ai Q-Y et al (2019) Complementary roles of MRI and endoscopic examination in the early detection of nasopharyngeal carcinoma. Ann Oncol 30:977–982
22. King AD, Woo JKS, Ai Q-Y et al (2020) Early detection of cancer: evaluation of MR imaging grading systems in patients with suspected nasopharyngeal carcinoma. AJNR Am J Neuroradiol 41:515–521
23. King AD, Wong LYS, Law BKH et al (2018) MR imaging criteria for the detection of nasopharyngeal carcinoma: discrimination of

early-stage primary tumors from benign hyperplasia. AJNR Am J Neuroradiol 39:515–523
24. Wang M-L, Wei X-E, Yu M-M, Li W-B (2017) Value of contrast-enhanced MRI in the differentiation between nasopharyngeal lymphoid hyperplasia and T1 stage nasopharyngeal carcinoma. Radiol Med 122:743–751
25. Chan KCA, Woo JKS, King A et al (2017) Analysis of plasma Epstein-Barr virus DNA to screen for nasopharyngeal cancer. N Engl J Med 377:513–522
26. Smith S, Jenkinson M (1999) Accurate robust symmetry estimation. Lect Notes Comput Sci 1679:308–317
27. Sun X, Shi L, Luo Y et al (2015) Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions. Biomed Eng Online 14:73
28. Wang F, Jiang M, Qian C et al (2017) Residual attention network for image classification. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2017:6450–6458
29. Paszke A, Gross S, Massa F et al (2019) PyTorch: an imperative style, high-performance deep learning library. Adv Neural Inf Process Syst 32:8024–8035
30. Park SH, Han K (2018) Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology 286:800–809
31. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. Available via http://arxiv.org/abs/1502.01852. Accessd 6 Feb 2015
32. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. Available via http://arxiv.org/abs/1412.6980. Accessed 30 Jan 2017
33. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837–845
34. Robin X, Turck N, Hainard A et al (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12:77
35. Ke L, Deng Y, Xia W et al (2020) Development of a self-constrained 3D DenseNet model in automatic detection and segmentation of nasopharyngeal carcinoma using magnetic resonance images. Oral Oncol 110:104862
36. Ai Q-Y, King AD, Chan JSM et al (2019) Distinguishing early-stage nasopharyngeal carcinoma from benign hyperplasia using intravoxel incoherent motion diffusion-weighted MRI. Eur Radiol 29:5627–5634
37. Reda I, Ghazal M, Shalaby A et al (2019) Detecting prostate cancer using a CNN-based system without segmentation. IEEE 16th International Symposium on Biomedical Imaging 2019:855–858
38. Hirshoren N, Damti S, Weinberger J et al (2019) Diffusion weighted magnetic resonance imaging of pre and post treatment nasopharyngeal carcinoma. Surg Oncol 30:122–125
39. Roblot V, Giret Y, Bou Antoun M et al (2019) Artificial intelligence to diagnose meniscus tears on MRI. Diagn Interv Imaging 100:243–249
40. Fan M, Li Y, Zheng S et al (2019) Computer-aided detection of mass in digital breast tomosynthesis using a faster region-based convolutional neural network. Methods 166:103–111
41. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2018:7132–7141