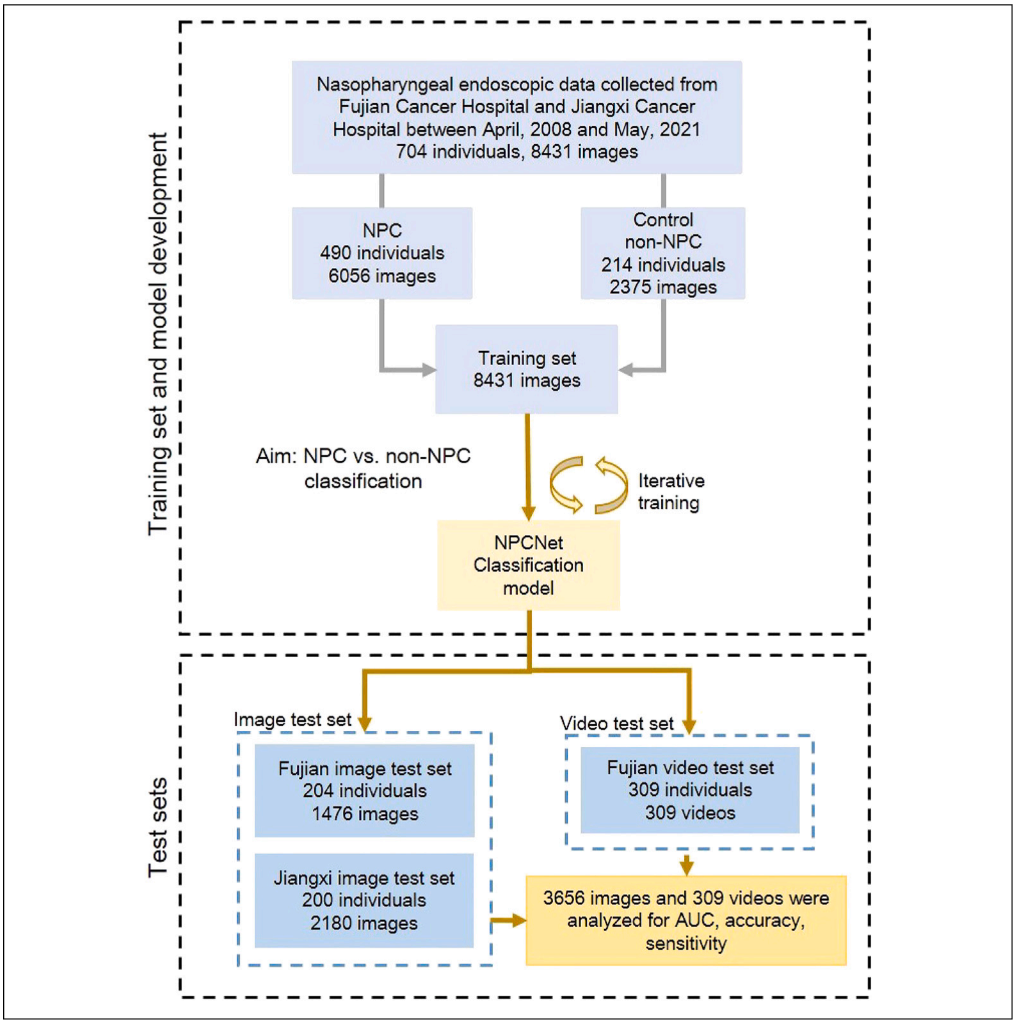
iScience

CellPress
OPEN ACCESS

Article

Advancing precise diagnosis of nasopharyngeal carcinoma through endoscopy-based radiomics analysis



Yun Xu, Jiesong Wang, Chenxin Li, ..., Shaojun Lin, Jingao Li, Dan Wu

wudan_0729@163.com

Highlights

It shows high accuracy in NPC diagnosis, with AUC values of 0.981 and 0.937 for images

It aids early NPC detection, identifying lesions often missed by conventional methods

Performance varies by gender and age, with higher AUC values seen in female patients

Heatmaps highlight malignant NPC regions, aligning with expert endoscopist assessments

Xu et al., iScience *27*, 110590
September 20, 2024
https://doi.org/10.1016/j.isci.2024.110590

iScience

CellPress
OPEN ACCESS

Article

Advancing precise diagnosis of nasopharyngeal carcinoma through endoscopy-based radiomics analysis

Yun Xu,[1,2,11] Jiesong Wang,[3,11] Chenxin Li,[4] Yong Su,[5,6] Hewei Peng,[7] Lanyan Guo,[8] Shaojun Lin,[1,2] Jingao Li,[5,6] and Dan Wu[9,10,12,*]

SUMMARY

Nasopharyngeal carcinoma (NPC) has high metastatic potential and is hard to detect early. This study aims to develop a deep learning model for NPC diagnosis using optical imagery. From April 2008 to May 2021, we analyzed 12,087 nasopharyngeal endoscopic images and 309 videos from 1,108 patients. The pre-trained model was fine-tuned with stochastic gradient descent on the final layers. Data augmentation was applied during training. Videos were converted to images for malignancy scoring. Performance metrics like AUC, accuracy, and sensitivity were calculated based on the malignancy score. The deep learning model demonstrated high performance in identifying NPC, with AUC values of 0.981 (95% confidence of interval [CI] 0.965–0.996) for the Fujian Cancer Hospital dataset and 0.937 (0.905–0.970) for the Jiangxi Cancer Hospital dataset. The proposed model effectively diagnoses NPC with high accuracy, sensitivity, and specificity across multiple datasets. It shows promise for early NPC detection, especially in identifying latent lesions.

INTRODUCTION

Nasopharyngeal carcinoma (NPC)[1] is an epithelial malignancy that exhibits a distinct geographical distribution. In 2018, it was estimated to have affected 130,000 patients globally, with over 70% of the cases occurring in Southeast Asia, particularly in southern China.[2–4] Unlike other head and neck carcinomas, over 70% of NPC patients are diagnosed with locoregionally advanced disease at presentation. This is largely attributed to the hidden anatomical site and nonspecific symptoms.[5,6] Despite the use of intensity-modulated radiotherapy and induction chemotherapy, approximately 5%–15% experience recurrence, and 15%–30% are at risk of distant metastasis.[7] Therefore, early diagnosis is the key to improving the prognosis of NPC.

Endoscopy with histopathological examination of suspected mucosal lesions is the gold standard for NPC diagnosis. Magnetic resonance imaging is advised for T and N classification.[8] However, NPC biopsy and diagnosis are affected by many factors, such as submucosal lesions with blurred images and multifocal distribution of NPC. Moreover, nasopharyngeal lesions concealed in mucosa or with normal lymphoid hyperplasia also hinder diagnosis. Therefore, it is essential to identify a reliable method of accurately detecting primary tumors to assist in NPC diagnosis.

Since 2012, significant advancements have been made in image recognition benchmarks due to deep learning[9–13] This progress is attributed to the amalgamation of computational power, algorithms, data, and application scenarios.[14–20] Deep learning has found extensive applications in the medical field, particularly in clinical image diagnosis, where it has achieved remarkable success in traditional pathology tasks.[21–24] Numerous AI models based on endoscopic images have been published, but the majority of these models are primarily focused on digestive diseases.[25–29] While some studies have utilized other types of images (e.g., biopsies, MRI, and CT images) to train a diagnostic model for NPC[30–37] there are only a few studies that have used endoscopic images for the same purpose. The application of

---

[1] Department of Radiation Oncology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, China
[2] Fujian Key Laboratory of Translational Cancer Medicine, Fuzhou, Fujian, China
[3] Department of Lymphoma & Head and Neck Oncology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, China
[4] Department of Electrical Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China
[5] Department of Radiation Oncology, Jiangxi Cancer Hospital, Jiangxi, China
[6] National Health Commission (NHC) Key Laboratory of Personalized Diagnosis and Treatment of Nasopharyngeal Carcinoma (Jiangxi Cancer Hospital of Nanchang University), Nanchang, China
[7] Department of Epidemiology and Health Statistics, Fujian Provincial Key Laboratory of Environment Factors and Cancer, School of Public Health, Fujian Medical University, Fuzhou, China
[8] School of Medical Imaging, Fujian Medical University, Fuzhou, China
[9] Tianjin Key Laboratory of Human Development and Reproductive Regulation, Tianjin Central Hospital of Gynecology Obstetrics and Nankai University Affiliated Hospital of Obstetrics and Gynecology, Tianjin, China
[10] Tianjin Cancer Institute, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China
[11] These authors contributed equally
[12] Lead contact
*Correspondence: wudan_0729@163.com

https://doi.org/10.1016/j.isci.2024.110590

iScience 27, 110590, September 20, 2024

deep learning, particularly with endoscopic images, holds immense potential for aiding in the diagnosis of NPC due to its wide availability. However, the methodology for this application is still under development and requires further validation.[38–40]

In this study, our objective is to develop a deep learning model for the optical diagnosis of NPC. We have used pathological examination results as the gold standard for NPC diagnosis. As shown in Figures 1, 2 and 3, the developed deep network, which is named with NPCNet, was independently evaluated on two testing sets comprised of static images and one testing set comprised of video data.[41,42]

RESULTS

Baseline characteristics of the training and test datasets

Between April 2008 and May 2021, 12,087 nasopharyngeal endoscopic images were available for this study, including 6,203 images from 573 individuals at Fujian Cancer Hospital (FCH) and 5,884 images from 535 individuals at Fujian Cancer Hospital (JCH). FCH also provided 309 videos from 309 individuals for this study. No videos were available from JCH. The training set consisted of 8,431 images from 704 individuals, including 6,056 images from 490 NPC patients and 2,375 images from 214 non-NPC patients. Male sex accounted for 74.9% (367/490) in the NPC patient group versus 60.7% (130/214) in the control group. The average age of the NPC patient group was 50 versus 52 of the control group. For NPC patients, 1.6% had stage I disease, 16.5% had stage II disease, 36.7% had stage III disease, 41.5% had stage IV disease, and 3.7% had no TNM stage information.

The validation set consisted of 1,476 static images of 204 patients from FCH and 2,180 static images of 200 patients from JCH as well as 309 videos of 309 patients from FCH. The latter consisted of 98 videos from individuals with histologically confirmed NPC, 106 videos from healthy nasopharyngeal patients (control group 1) and 105 videos from recuperated patients after radiotherapy (control group 2) who had achieved clinical complete response without residual tumors. The baseline characteristics of the training set and three test sets are shown in Table 1.

Performance of the proposed deep models on test sets

The processes displayed in Figure 4 include the initial phase of developing the deep learning model, where data is inputted and the model is trained to recognize patterns related to postoperative lung cancer prognosis. This is followed by the validation phase where the model is tested on new data to evaluate its accuracy in predicting the prognosis. This step-by-step visualization offers a clear understanding of how deep learning models are developed and validated in the medical field.

The model achieved high performance in identifying NPC patients in the validation set with area under the curve (AUC) values of 0.981 (95% confidence of interval [CI] 0.965–0.996) and 0.937 (0.905–0.970) for the FCH and JCH image datasets, respectively, and 0.953 (0.927–0.980) for the FCH video dataset (Figure 5).

For the FCH image dataset, the accuracy was 0.956 (0.918–0.980), the sensitivity was 0.948 (0.884–0.983), and the specificity was 0.963 (0.907–0.990). For the FCH video dataset, the accuracy was 0.906 (0.868–0.936), the sensitivity was 0.847 (0.757–0.909), and the specificity was 0.934 (0.891–0.963). For the JCH image dataset, the accuracy was 0.880 (0.827–0.922), the sensitivity was 0.931 (0.864–0.972), and the specificity was 0.827 (0.737–0.896). The performance of the model on the three test sets is shown in Table 2.

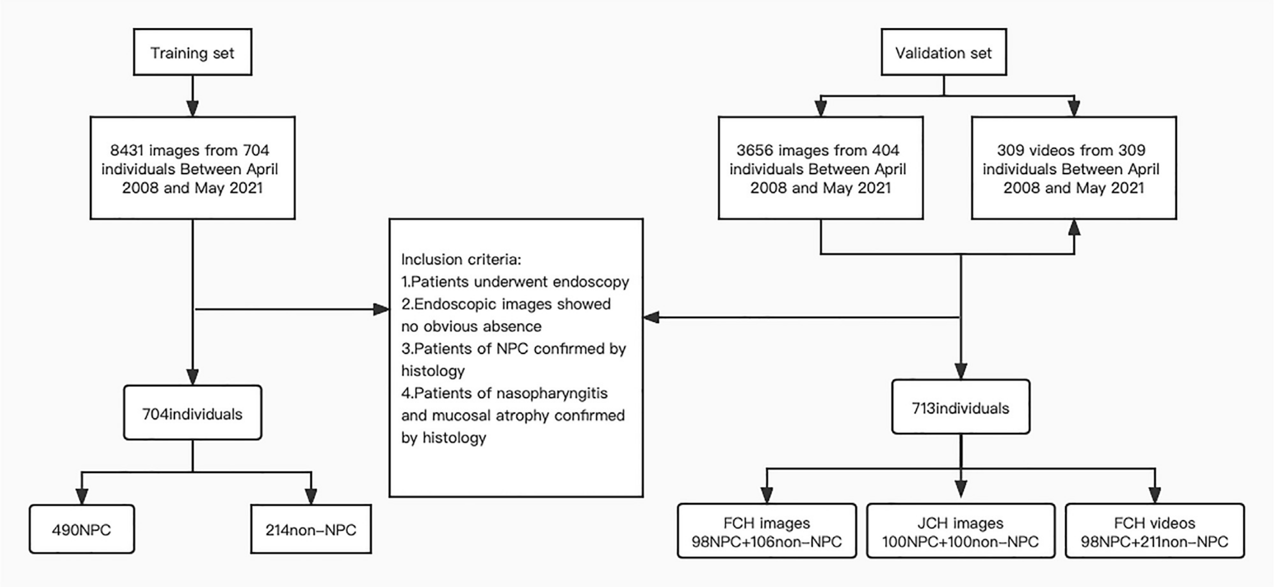


Figure 1. Flowchart illustrating the patient inclusion process for a retrospective diagnostic study conducted using endoscopic images and videos from Fujian Cancer Hospital and Jiangxi Cancer Hospital

The study involved patients aged 12–85 years who underwent pathological examination, with a separate healthy control group.

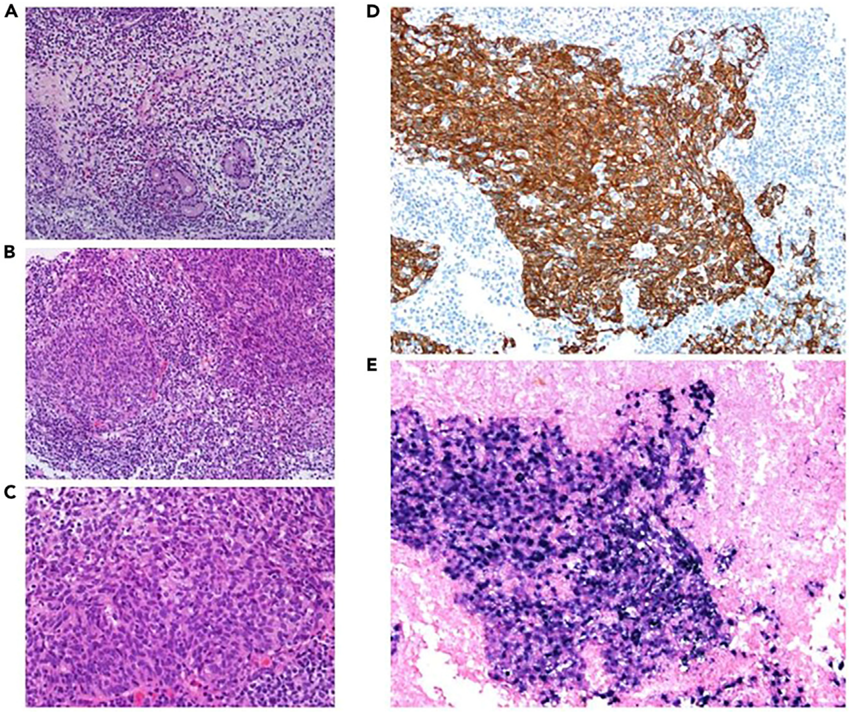


Figure 2. Hematoxylin and Eosin (HE) pathological images of patients with nasopharyngeal carcinoma and nasopharyngeal mucosal inflammation of NPC after radiotherapy

(A) Microscopic morphology of mucosal inflammation after radiotherapy for NPC (20x).

(B and C) Microscopic morphology of NPC (20x and 40x).

(D) CK immunohistochemistry was strongly positive in nasopharyngeal carcinoma.

(E) EBER *in situ* hybridization was positive in nasopharyngeal carcinoma.

Notably, when using 106 videos from healthy nasopharynx as the control (control group 1), the model achieved an AUC of 0.967 (95% CI 0.942–0.991) for detecting NPC in 98 patients, and this AUC value is greater than obtained using postradiotherapy patients as the control (control group 2) [AUC 0.940 (0.908–0.972)] (Figure 6).

Performance of NPCNet in different gender and age groups

NPCNet achieved a higher AUC in the female group than in the male group on all three test sets. In the FCH image dataset, NPCNet achieved an AUC of 0.944 (95% CI 0.903–0.984) for the male group and 0.995 (0.987–1.000) for the female group. In the FCH video dataset, the AUC was up to 0.933 (0.896–0.970) for the male group and 0.988 (0.972–1.000) for the female group. In the JCH image dataset, the AUC was 0.931 (0.889–0.973) for the male group and 0.957 (0.912–1.000) for the female group (Figure 7; Table 3).

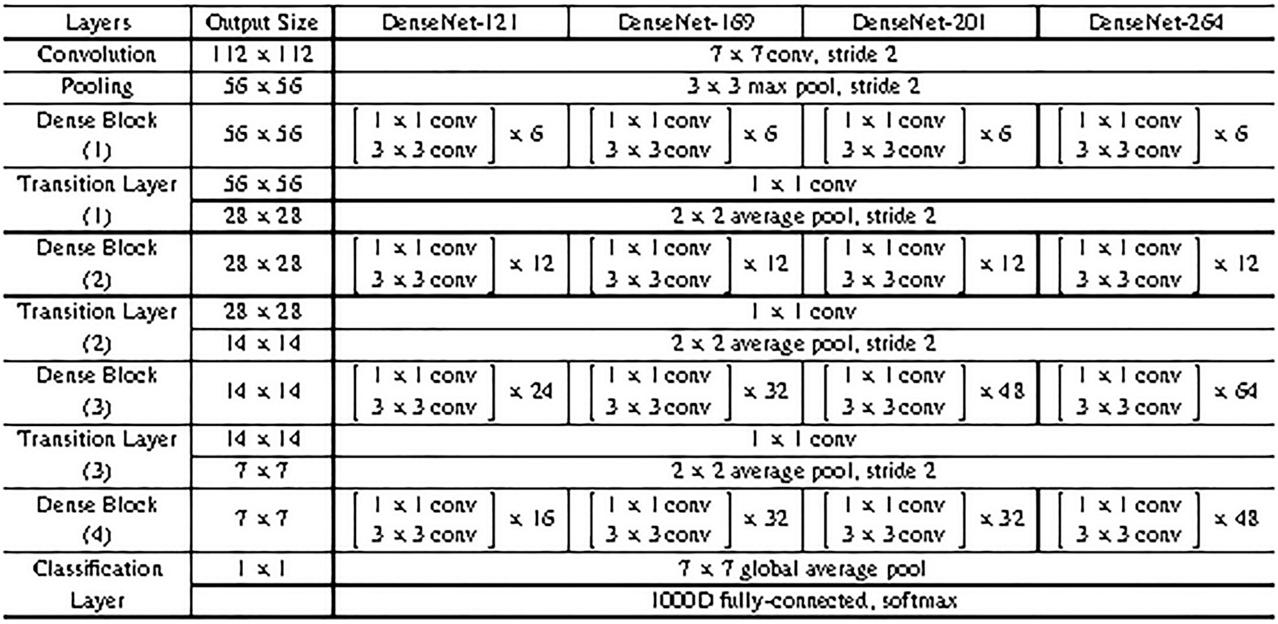
Patients were stratified based on age into >50 and $\leq$50-year-old groups. In the FCH image dataset, NPCNet achieved an AUC of 0.986 (0.969–1.000) for the >50-year-old group and 0.975 (0.949–1.000) for the $\leq$50 years old group. In the FCH video dataset, the AUC was 0.946 (0.912–0.980) for the >50-year-old group and 0.957 (0.915–0.998) for the $\leq$50-year-old group. In the JCH image dataset, the AUC was 0.944 (0.903–0.984) for the >50-year-old group and 0.934 (0.885–0.983) for the $\leq$50-year-old group (Figure 8; Table 4).

Visual explanation of the decision made by NPCNet

We calculated the score of NPC images after partly erasing pixels and subtracted this score from the original score to derive a visual explanation by localizing the image area that most influenced the decision made by NPCNet. Representative NPC images with accompanying heatmaps highlighting the features that most influenced NPCNet prediction are shown in Figure 9. The endoscopists from FCH agreed that the heatmaps captured the regions of malignant NPC lesions.

DISCUSSION

Our study underscores the efficacy of our model in the automated diagnosis of NPC using a comprehensive dataset of images and videos from various hospitals in China. The model demonstrated high accuracy, sensitivity, and specificity in real-world settings, outperforming existing models that predominantly rely on biopsies, MRI, and CT images.[30–37,43,44] By integrating optical diagnosis during endoscopy, our approach facilitates earlier treatment planning, enhances NPC detection rates, and reduces missed diagnoses due to hidden lesions or



Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	[1 × 1 conv 3 × 3 conv] × 6	[1 × 1 conv 3 × 3 conv] × 6	[1 × 1 conv 3 × 3 conv] × 6	[1 × 1 conv 3 × 3 conv] × 6
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	[1 × 1 conv 3 × 3 conv] × 12	[1 × 1 conv 3 × 3 conv] × 12	[1 × 1 conv 3 × 3 conv] × 12	[1 × 1 conv 3 × 3 conv] × 12
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	[1 × 1 conv 3 × 3 conv] × 24	[1 × 1 conv 3 × 3 conv] × 32	[1 × 1 conv 3 × 3 conv] × 48	[1 × 1 conv 3 × 3 conv] × 64
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	[1 × 1 conv 3 × 3 conv] × 16	[1 × 1 conv 3 × 3 conv] × 32	[1 × 1 conv 3 × 3 conv] × 32	[1 × 1 conv 3 × 3 conv] × 48
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

Figure 3. DenseNet architectures for ImageNet

The growth rate for all the networks is k = 32. Note that each "conv" layer shown in the table corresponds the sequence BN-ReLU-Conv.

clinician errors. This study significantly contributes to the literature on NPC diagnostic models and underscores the potential of optical diagnosis in improving patient outcomes.

Early diagnosis of NPC is challenging due to several factors: (1) The similarity in symptoms between benign and malignant tumors can lead to a lack of specificity and potential misdiagnosis. (2) Nasopharyngeal lesions are difficult to detect, resulting in missed diagnoses. (3) Nasopharyngeal biopsy requires expertise and is typically performed by senior doctors. False negative results may necessitate repeated biopsies, thereby delaying treatment. A study by Baran H et al. revealed that only 10.67% of patients had malignancy in nasopharyngeal biopsy, with

Table 1. Baseline characteristics

	Training set		Fujian test set (image)		Fujian test set (video)		Jiangxi test set (image)	
Variables	NPC	non-NPC	NPC	non-NPC	NPC	non-NPC	NPC	non-NPC
Patients	490	214	98	106	98	211	100	100
Images	6056	2375	793	683	98	211	1174	1006
Male sex	367	130	79	27	79	103	70	72
Images	4532	1498	641	185	79	103	759	723
Female sex	123	84	19	79	19	108	30	28
Images	1524	877	152	498	19	108	415	283
Age (years)	50 (12–80)	52 (15–80)	50 (18–75)	48 (14–74)	50 (18–75)	49 (12–74)	52 (12–84)	51 (15–78)
Age >50years male)	177 (35.6%)	71 (14.3%)	42 (39.6%)	12 (11.3%)	42 (23.1%)	52 (28.6%)	44 (31.0%)	33 (23.2%)
Age (≤50years male)	190 (38.2%)	59 (11.9%)	37 (34.9%)	15 (14.2%)	37 (20.3%)	51 (28.0%)	26 (18.3%)	39 (27.5%)
Age >50years female)	58 (28.0%)	49 (23.7%)	5 (5.1%)	34 (34.7%)	5 (3.9%)	46 (36.2%)	14 (24.1%)	17 (29.3%)
Age (≤50years female)	65 (31.4%)	35 (16.9%)	14 (14.3%)	45 (45.9%)	14 (11.0%)	62 (48.9%)	16 (27.6%)	11 (19.0%)
TMN staging								
I	8 (1.6%)	.	3 (3.1%)	.	3 (3.1%)	.	0	.
II	81 (16.5%)	.	11 (11.2%)	.	11 (11.2%)	.	12 (12%)	.
III	180 (36.7%)	.	30 (30.6%)	.	30 (30.6%)	.	27 (27%)	.
IV	203 (41.5%)	.	52 (53.1%)	.	52 (53.1%)	.	61 (61%)	.
Missing	18 (3.7%)	.	2 (2.0%)	.	2 (2.0%)	.	0	.

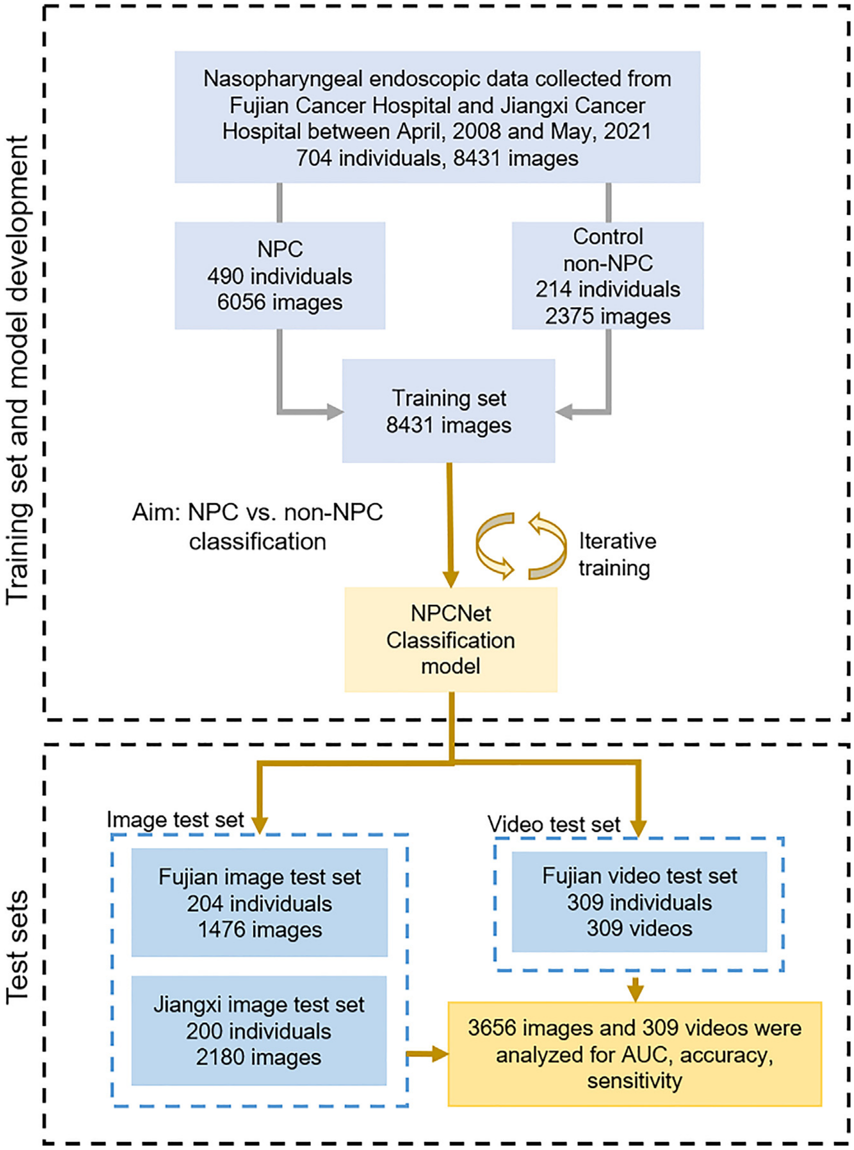


Figure 4. The processes of development and validation of the deep-learning model for NPC diagnosis

The model is trained on data, then tested for accuracy. This depiction offers an insight into the systematic construction and validation of deep learning models in medicine.

various pathological results.[45] Our internally validated NPCNet model has the potential to significantly facilitate early NPC diagnosis if externally validated.

Our model achieved lower AUC values in the FCH video validation set (AUC: 0.953, 0.927–0.980) compared to the FCH image validation set (AUC: 0.981, 0.965–0.996). This may be attributed to the presence of irrelevant images outside of the nasopharynx in the video set, as it captures the entire nasopharyngoscopy process, unlike the more focused images taken by doctors. Furthermore, the control group in the FCH video dataset included two subgroups: healthy individuals and post-radiotherapy patients whose NPCs had resolved and whose mucosa had recovered. We included post-radiotherapy patients to account for the challenge of distinguishing recurrences based solely on mucosal appearance, as the nasopharyngeal mucosa can appear dirty and have secretions attached after radiotherapy for NPC.

The performance of the NPCNet model in distinguishing NPC through optical diagnosis varied when comparing two different control groups. When videos from healthy individuals were used as controls, the model achieved a higher AUC compared to when videos from post-radiotherapy patients were used. This difference in performance could be attributed to the varying appearances of the nasopharyngeal mucosa in these two groups. Healthy mucosa appears clean, smooth, and uniformly red, while mucosa in post-radiotherapy patients is often atrophic and may be covered by thick and unclean secretions if nasal irrigation is inadequate. Additionally, post-radiotherapy patients may have undetected microscopic residual lesions. It is crucial to compare these groups as the NPCNet model, an artificial intelligence method,

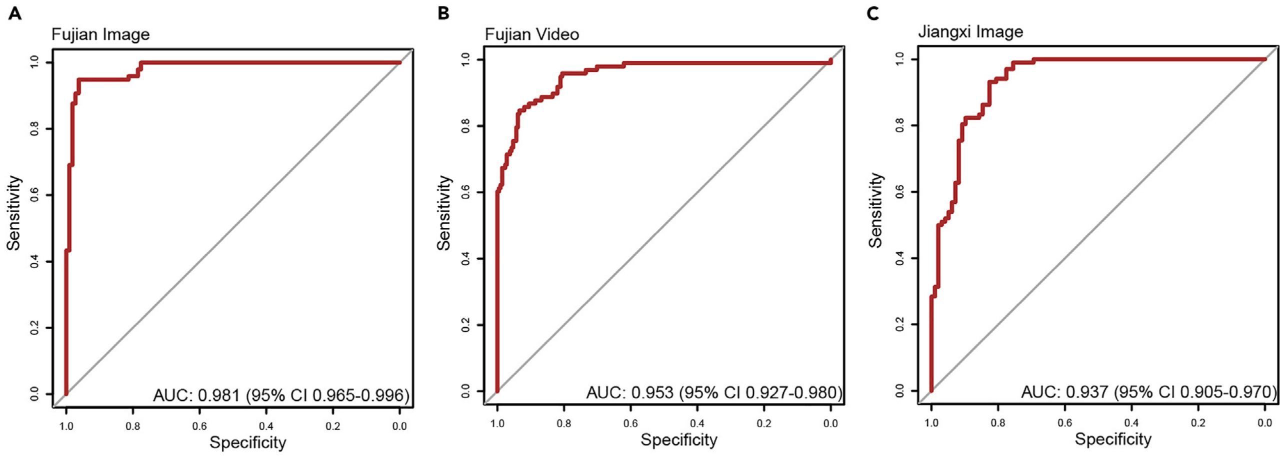


Figure 5. Graphical representation of the model's performance in identifying NPC patients in the validation set

Performance of the NPCNet model in identifying patients with NPC using three test sets.

(A) Performance using the Fujian image dataset.

(B) Performance using the Fujian video dataset.

(C) Performance using the Jiangxi image dataset.

can provide additional information on residual and post-radiotherapy changes that may not be captured by standard MRI examinations. This is particularly valuable for radiotherapy doctors or less-experienced physicians.

In our model, the female group exhibited a higher AUC than the male group (Figure 2). Some researchers have observed that female NPC patients tend to have a better prognosis than male patients, possibly due to intrinsic factors specific to female patients, such as hormones. However, the underlying mechanism remains largely unknown.[46–49]

We calculated the score of NPC images after partially erasing pixels and subtracted this score from the original score to derive a visual explanation. This process localized the image area that most influenced the decision made by NPCNet. The endoscopists concurred that the heatmaps captured the regions of malignant NPC lesions. The interpretability of NPCNet may assist doctors in identifying NPC lesions in the future.[50–52]

Prior to the pathology report, it is challenging to determine whether MRI or nasopharyngeal endoscopic images are more effective for NPC diagnosis. Nasopharyngoscopy, when combined with biopsy, is a simple and cost-effective method for identifying NPC. However, our model demonstrates high accuracy, sensitivity, and specificity in the Fujian video test set, indicating its potential for real-time video diagnosis. This method could significantly reduce doctors' workload, provide faster diagnosis for patients, and improve the detection rate of hidden lesions.

Our study had several limitations. We lacked image data from northern China, which may have resulted in the model learning less about mucosal features in different regions. The training dataset included only 704 individuals and 8,431 images, smaller than those used in other similar cancer diagnosis models. Most patients in our dataset had advanced stages, so there is uncertainty about the model's performance in detecting small tumors. Increasing data size and including real-world data from other areas could improve our model's performance. Prospective, multicenter, large-scale trials are needed for validation. Additionally, our study did not explore the correlation of stage and prognosis, which are crucial factors in the treatment regimen selection based on MRI and EB-DNA copy number, among other factors. It is important to note that we have not included other malignancies that can affect the nasopharynx, such as lymphomas or minor salivary gland tumors. These are significant potential conditions that could play a role in the nasopharynx, and their exclusion from the study is an important limitation that needs to be thoroughly discussed. This omission could potentially impact the overall generalizability of the deep learning model

Table 2. Performance metrics for NPCNet on three test sets

	Fujian image dataset	Fujian video dataset	Jiangxi image dataset
Accuracy (95% CI)	0.956 (0.918–0.980)	0.906 (0.868–0.936)	0.880 (0.827–0.922)
Sensitivity (95% CI)	0.948 (0.884–0.983)	0.847 (0.757–0.909)	0.931 (0.864–0.972)
Specificity (95% CI)	0.963 (0.907–0.990)	0.934 (0.891–0.963)	0.827 (0.737–0.896)
Positive Prediction Value (95% CI)	0.958 (0.897–0.989)	0.856 (0.770–0.919)	0.848 (0.768–0.909)
Negative Prediction Value (95% CI)	0.954 (0.895–0.985)	0.929 (0.886–0.960)	0.920 (0.843–0.967)
Kappa (95% CI)	0.912 (0.853–0.961)	0.783 (0.709–0.850)	0.759 (0.671–0.849)
F1 (95% CI)	0.953 (0.923–0.981)	0.851 (0.798–0.899)	0.888 (0.840–0.928)

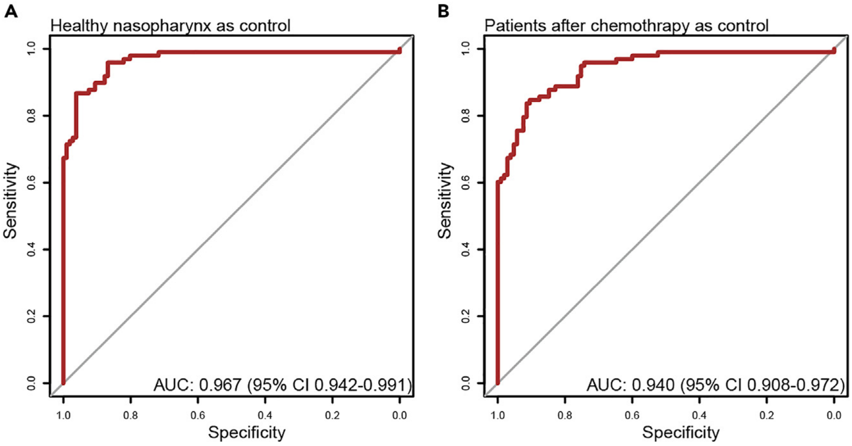


Figure 6. Performance of NPCNet using the Fujian video dataset

(A) Performance using videos from patients with healthy nasopharynx.

(B) Performance using videos from recuperated patients after chemotherapy.

that has been developed to distinguish NPC from these other nasopharyngeal pathologies. It's crucial to consider these factors because they could potentially alter the effectiveness of the model when applied in a broader context, which includes these other types of nasopharyngeal conditions.

The newly developed NPCNet model demonstrated high accuracy, sensitivity, and specificity in identifying patients with NPC. The performance of our model indicates that this method is valuable and has the potential to assist in clinical diagnosis. In the future, this diagnostic model should be tested in independent datasets and prospective clinical trials.

Conclusions

This study demonstrates that NPCNet, tested on two image datasets and one video dataset, can achieve high accuracy, sensitivity, and specificity in automated NPC diagnosis in a real-world setting. The optical model holds promise for early detection of NPC, particularly in detecting latent lesions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Study design and participants
  - Procedures
- QUANTIFICATION AND STATISTICAL ANALYSIS

ACKNOWLEDGMENTS

This study was supported by the National Clinical Key Specialty Construction Program; Fujian Provincial Clinical Research Center for Cancer Radiotherapy and Immunotherapy (no. 2020Y2012); Training Project of Young Talents in Health System of Fujian Province (no. 2020GGB013); Joint Funds for the innovation of science and Technology, Fujian province (no. 2021Y9207), and Tianjin Municipal Health Commission Foundation (no. RC20027).

AUTHOR CONTRIBUTIONS

Y.X.: Conducted experiments, analyzed data, and contributed to manuscript writing. J.W.: Conducted experiments and contributed to data analysis. C.L.: Assisted in experimental design and data collection. Y.S.: Provided critical revisions and contributed to the design of the study. H.P.: Assisted in manuscript editing and data interpretation. L.G.: Contributed to data collection and initial data analysis. S.L.: Assisted in

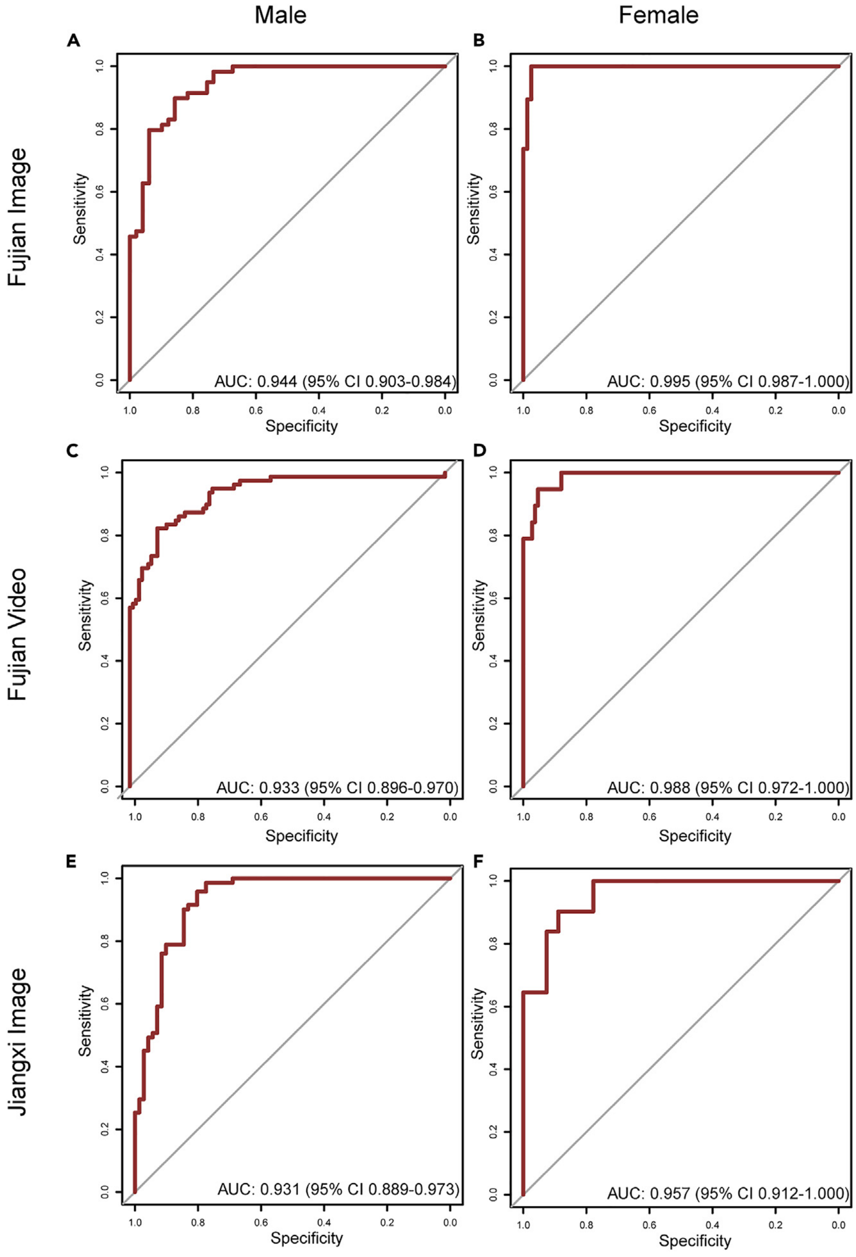


Figure 7. Performance of NPCNet in the male group and female group

(A) Performance on the male group from the Fujian image dataset.
(B) Performance on the female group from the Fujian image dataset.
(C) Performance on the male group from the Fujian video dataset.
(D) Performance on the female group from the Fujian dataset.
(E) Performance on the male group from the Jiangxi image dataset.
(F) Performance on the female group from the Jiangxi image dataset.

experimental procedures and data analysis. J.L.: Provided critical feedback and contributed to the study design. D.W.: Designed the experiments, supervised the project, and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Table 3. Performance metrics of NPCNet of Fujian Video dataset

	Healthy nasopharynx as control	Patients after chemotherapy as control
Accuracy (95% CI)	0.917 (0.870–0.951)	0.877 (0.824–0.919)
Sensitivity (95% CI)	0.867 (0.784–0.927)	0.847 (0.760–0.912)
Specificity (95% CI)	0.962 (0.906–0.990)	0.905 (0.832–0.953)
Positive Prediction Value (95% CI)	0.955 (0.889–0.988)	0.892 (0.811–0.947)
Negative Prediction Value (95% CI)	0.887 (0.814–0.938)	0.864 (0.785–0.922)
Kappa (95% CI)	0.832 (0.753–0.902)	0.753 (0.651–0.842)
F1 (95% CI)	0.869 (0.813–0.919)	0.909 (0.865–0.950)

Received: April 30, 2024
Revised: May 25, 2024
Accepted: July 23, 2024
Published: July 26, 2024

REFERENCES

1. Rotolo, F., Pignon, J.P., Bourhis, J., Marguet, S., Leclercq, J., Tong Ng, W., Ma, J., Chan, A.T.C., Huang, P.Y., Zhu, G., et al. (2017). Surrogate end points for overall survival in loco-regionally advanced nasopharyngeal carcinoma: an individual patient data meta-analysis. J. Natl. Cancer Inst. *109*, djw239.
2. Chen, Y.P., Chan, A.T.C., Le, Q.T., Blanchard, P., Sun, Y., and Ma, J. (2019). Nasopharyngeal carcinoma. Lancet *394*, 64–80.
3. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA A Cancer J. Clin. *71*, 209–249.
4. Zhang, L.F., Li, Y.H., Xie, S.H., Ling, W., Chen, S.H., Liu, Q., Huang, Q.H., Cao, S.M., and Cao, S.M. (2015). Incidence trend of nasopharyngeal carcinoma from 1987 to 2011 in Sihui County, Guangdong Province, South China: an age-period-cohort analysis. Chin. J. Cancer *34*, 350–357.
5. Pan, J.J., Ng, W.T., Zong, J.F., Lee, S.W.M., Choi, H.C.W., Chan, L.L.K., Lin, S.J., Guo, Q.J., Sze, H.C.K., Chen, Y.B., et al. (2016). Prognostic nomogram for refining the prognostication of the proposed 8th edition of the AJCC/UICC staging system for nasopharyngeal cancer in the era of intensity-modulated radiotherapy. Cancer *122*, 3307–3315.
6. Wu, F., Wang, R., Lu, H., Wei, B., Feng, G., Li, G., Liu, M., Yan, H., Zhu, J., Zhang, Y., and Hu, K. (2014). Concurrent chemoradiotherapy in locoregionally advanced nasopharyngeal carcinoma: treatment outcomes of a prospective, multicentric clinical study. Radiother. Oncol. *112*, 106–111.
7. Zhang, Y., Chen, L., Hu, G.Q., Zhang, N., Zhu, X.D., Yang, K.Y., Jin, F., Shi, M., Chen, Y.P., Hu, W.H., et al. (2019). Gemcitabine and Cisplatin Induction Chemotherapy in Nasopharyngeal Carcinoma. N. Engl. J. Med. *381*, 1124–1135.
8. Bossi, P., Chan, A.T., Licitra, L., Trama, A., Orlandi, E., Hui, E.P., Halámková, J., Mattheis, S., Baujat, B., Hardillo, J., et al. (2021). Nasopharyngeal carcinoma: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up(dagger). Ann. Oncol. *32*, 452–465.
9. Li, C., Zhang, Y., Liang, Z., Ma, W., Huang, Y., and Ding, X. (2021). Consistent posterior distributions under vessel-mixing: a regularization for cross-domain retinal artery/vein classification. In 2021 IEEE International Conference on Image Processing (ICIP) (IEEE), pp. 61–65.
10. Li, C., Ma, W., Sun, L., Ding, X., Huang, Y., Wang, G., and Yu, Y. (2022). Hierarchical deep network with uncertainty-aware semi-supervised learning for vessel segmentation. Neural Comput. Appl. *34*, 3151–3164.
11. Shen, D., Wu, G., and Suk, H.I. (2017). Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. *19*, 221–248.
12. Chen, Y., Liu, C., Liu, X., Arcucci, R., and Xiong, Z. (2024). BIMCV-R: A Landmark Dataset for 3D CT Text-Image Retrieval. Preprint at arXiv. https://doi.org/10.48550/arXiv.2403.15992.
13. Chen, Y., Liu, C., Huang, W., Cheng, S., Arcucci, R., and Xiong, Z. (2023). Generative Text-Guided 3d Vision-Language Pretraining for Unified Medical Image Segmentation. Preprint at arXiv. https://doi.org/10.48550/arXiv.2306.04811.
14. Sun, L., Li, C., Ding, X., Huang, Y., Chen, Z., Wang, G., Yu, Y., and Paisley, J. (2022). Few-shot medical image segmentation using a global correlation network with discriminative embedding. Comput. Biol. Med. *140*, 105067.
15. Zhang, Y., Li, C., Lin, X., Sun, L., Zhuang, Y., Huang, Y., Ding, X., Liu, X., and Yu, Y. (2021). Generator versus segmentor: Pseudo-healthy synthesis. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24 (Springer International Publishing), pp. 150–160.
16. Li, C., Zhang, Y., Li, J., Huang, Y., and Ding, X. (2021). Unsupervised anomaly segmentation using image-semantic cycle translation. Preprint at arXiv. https://doi.org/10.48550/arXiv.2103.09094.
17. Chen, Y., Huang, W., Zhou, S., Chen, Q., and Xiong, Z. (2023). Self-supervised Neuron Segmentation with Multi-Agent Reinforcement Learning. Preprint at arXiv. https://doi.org/10.48550/arXiv.2310.04148.
18. Li, C., Liu, X., Li, W., Wang, C., Liu, H., and Yuan, Y. (2024). U-KAN Makes Strong Backbone for Medical Image Segmentation and Generation. Preprint at arXiv. https://doi.org/10.48550/arXiv.2406.02918.
19. Liu, H., Liu, Y., Li, C., Li, W., and Yuan, Y. (2024). LGS: A Light-weight 4D Gaussian Splatting for Efficient Surgical Scene Reconstruction. Preprint at arXiv. https://doi.org/10.48550/arXiv.2406.16073.
20. Li, C., Feng, B.Y., Liu, Y., Liu, H., Wang, C., Yu, W., and Yuan, Y. (2024). EndoSparse: Real-Time Sparse View Synthesis of Endoscopic Scenes Using Gaussian Splatting. Preprint at arXiv. https://doi.org/10.48550/arXiv.2406.16073.
21. Goodman, A. (2020). Deep-Learning-Based Evaluation of Dual Stain Cytology for Cervical Cancer Screening: A New Paradigm. J. Natl. Cancer Inst. *112*, djaa067.
22. Nagpal, K., Foote, D., Tan, F., Liu, Y., Chen, P.H.C., Steiner, D.F., Manoj, N., Olson, N., Smith, J.L., Mohtashamian, A., et al. (2020). Development and Validation of a Deep Learning Algorithm for Gleason Grading of Prostate Cancer From Biopsy Specimens. JAMA Oncol. *6*, 1372–1380.
23. Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J.A.W.M., Hermsen, M., and Manson, Q.F. (2017). Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA *318*, 2199–2210.
24. Yamashita, R., Long, J., Longacre, T., Peng, L., Berry, G., Martin, B., Higgins, J., Rubin, D.L., and Shen, J. (2021). Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. Lancet Oncol. *22*, 132–141.
25. Stidham, R.W., Liu, W., Bishu, S., Rice, M.D., Higgins, P.D.R., Zhu, J., Nallamothu, B.K., and Waljee, A.K. (2019). Performance of a Deep Learning Model vs Human Reviewers in Grading Endoscopic Disease Severity of Patients With Ulcerative Colitis. JAMA Netw. Open *2*, e193963.
26. Nakashima, H., Kawahira, H., Kawachi, H., and Sakaki, N. (2020). Endoscopic three-categorical diagnosis of Helicobacter pylori infection using linked color imaging and deep learning: a single-center prospective

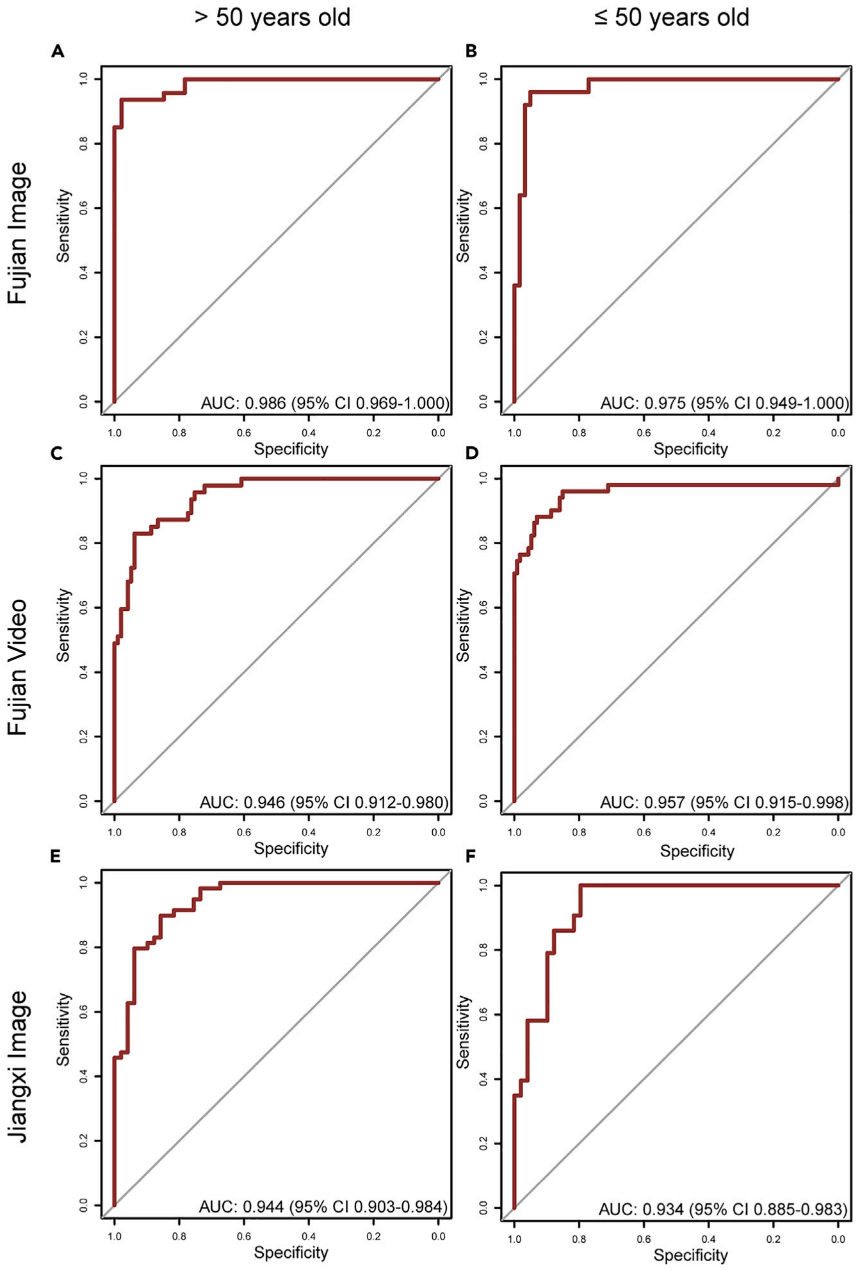


Figure 8. Performance of NPCNet in the >50-year-old group and ≤50-year-old group

(A) Performance on the >50-year-old group from the Fujian image dataset.

(B) Performance on the ≤50-year-old group from the Fujian image dataset.

(C) Performance on the >50-year-old group from the Fujian video dataset.

(D) Performance on the ≤50-year-old group from the Fujian video dataset.

(E) Performance on the >50-year-old group from the Jiangxi image dataset.

(F) Performance on the ≤50-year-old group from the Jiangxi image dataset.

study (with video). Gastric Cancer *23*, 1033–1040.

27. Huang, L., Lu, X., Huang, X., Zou, X., Wu, L., Zhou, Z., Wu, D., Tang, D., Chen, D., Wan, X., et al. (2021). Intelligent difficulty scoring and assistance system for endoscopic extraction of common bile duct stones based on deep learning: multicenter study. Endoscopy *53*, 491–498.

28. Bang, C.S., Lim, H., Jeong, H.M., and Hwang, S.H. (2021). Use of Endoscopic Images in the Prediction of Submucosal Invasion of Gastric Neoplasms: Automated Deep Learning Model Development and Usability Study. J. Med. Internet Res. *23*, e25167.

29. Hirasawa, T., Ikenoyama, Y., Ishioka, M., Namikawa, K., Horiuchi, Y., Nakashima, H., and Fujisaki, J. (2021). Current status and future perspective of artificial intelligence applications in endoscopic diagnosis and

Table 4. Performance metrics of NPCNet on >50years old group and ≤50 years old group

	Fujian image dataset		Fujian video dataset		Jiangxi image dataset	
	>50years	≤ 50 years	>50years	≤ 50 years	>50years	≤ 50 years
Accuracy (95% CI)	0.957 (0.894–0.988)	0.955 (0.898–0.985)	0.903 (0.842–0.946)	0.915 (0.862–0.953)	0.880 (0.803–0.934)	0.891 (0.809–0.947)
Sensitivity (95% CI)	0.936 (0.825–0.987)	0.960 (0.863–0.995)	0.830 (0.692–0.924)	0.882 (0.761–0.956)	0.898 (0.792–0.962)	1.000 (0.933–1.000)
Specificity (95% CI)	0.978 (0.885–0.999)	0.951 (0.863–0.990)	0.938 (0.870–0.977)	0.930 (0.866–0.969)	0.857 (0.728–0.941)	0.796 (0.657–0.898)
Positive Prediction Value (95% CI)	0.978 (0.882–0.999)	0.941 (0.838–0.988)	0.867 (0.732–0.949)	0.849 (0.724–0.933)	0.883 (0.774–0.952)	0.811 (0.680–0.906)
Negative Prediction Value (95% CI)	0.938 (0.828–0.987)	0.967 (0.885–0.996)	0.919 (0.847–0.964)	0.946 (0.887–0.980)	0.875 (0.748–0.953)	1.000 (0.926–1.000)
Kappa (95% CI)	0.914 (0.826–0.978)	0.909 (0.828–0.981)	0.776 (0.658–0.882)	0.803 (0.694–0.895)	0.757 (0.631–0.869)	0.785 (0.662–0.908)
F1 (95% CI)	0.957 (0.909–0.990)	0.950 (0.897–0.990)	0.848 (0.765–0.920)	0.865 (0.792–0.929)	0.891 (0.824–0.944)	0.896 (0.825–0.951)

management of gastric cancer. Dig. Endosc. *33*, 263–272.

30. Chuang, W.Y., Chang, S.H., Yu, W.H., Yang, C.K., Yeh, C.J., Ueng, S.H., Liu, Y.J., Chen, T.D., Chen, K.H., Hsieh, Y.Y., et al. (2020). Successful Identification of Nasopharyngeal Carcinoma in Nasopharyngeal Biopsies Using Deep Learning. Cancers *12*, 507.
31. Diao, S., Hou, J., Yu, H., Zhao, X., Sun, Y., Lambo, R.L., Xie, Y., Liu, L., Qin, W., and Luo, W. (2020). Computer-Aided Pathologic Diagnosis of Nasopharyngeal Carcinoma Based on Deep Learning. Am. J. Pathol. *190*, 1691–1700.
32. Yang, Q., Guo, Y., Ou, X., Wang, J., and Hu, C. (2020). Automatic T Staging Using Weakly Supervised Deep Learning for Nasopharyngeal Carcinoma on MR Images. J. Magn. Reson. Imag. *52*, 1074–1082.
33. Jing, B., Deng, Y., Zhang, T., Hou, D., Li, B., Qiang, M., Liu, K., Ke, L., Li, T., Sun, Y., et al. (2020). Deep learning for risk prediction in patients with nasopharyngeal carcinoma using multi-parametric MRIs. Comput. Methods Progr. Biomed. *197*, 105684.
34. Qiang, M., Li, C., Sun, Y., Sun, Y., Ke, L., Xie, C., Zhang, T., Zou, Y., Qiu, W., Gao, M., et al. (2021). A Prognostic Predictive System Based on Deep Learning for Locoregionally Advanced Nasopharyngeal Carcinoma. J. Natl. Cancer Inst. *113*, 606–615.
35. Lin, L., Dou, Q., Jin, Y.M., Zhou, G.Q., Tang, Y.Q., Chen, W.L., Su, B.A., Liu, F., Tao, C.J., Jiang, N., et al. (2019). Deep Learning for Automated Contouring of Primary Tumor Volumes by MRI for Nasopharyngeal Carcinoma. Radiology *291*, 677–686.

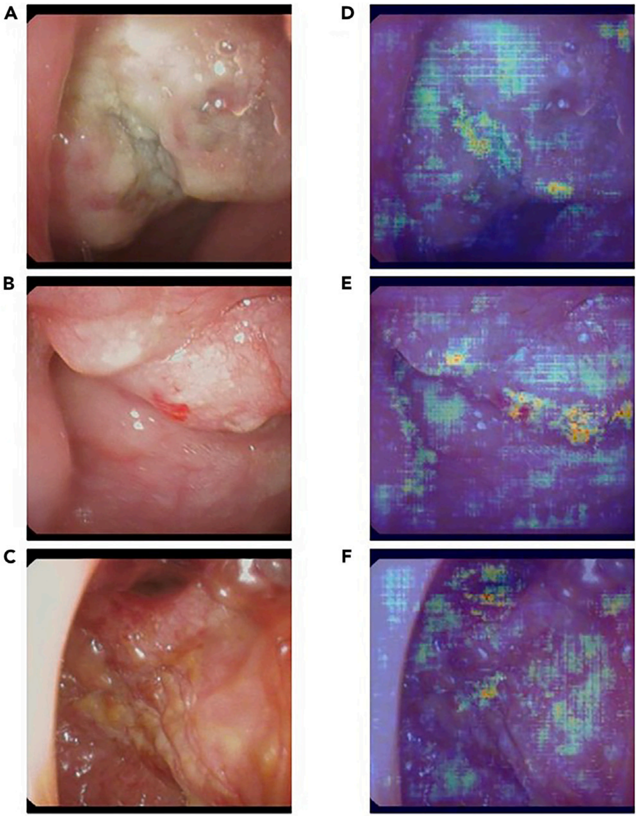


Figure 9. Visual explanation of the decision made by NPCNet

(A–C) Representative NPC images.

(D–F) NPC images with heatmaps to highlight features that most influence NPCNet prediction.

36. Peng, H., Dong, D., Fang, M.J., Li, L., Tang, L.L., Chen, L., Li, W.F., Mao, Y.P., Fan, W., Liu, L.Z., et al. (2019). Prognostic Value of Deep Learning PET/CT-Based Radiomics: Potential Role for Future Individual Induction Chemotherapy in Advanced Nasopharyngeal Carcinoma. Clin. Cancer Res. *25*, 4271–4279.
37. Daoud, B., Morooka, K., Kurazume, R., Leila, F., Mnejja, W., and Daoud, J. (2019). 3D segmentation of nasopharyngeal carcinoma from CT images using cascade deep learning. Comput. Med. Imag. Graph. *77*, 101644.
38. He, Z., Zhang, K., Zhao, N., Wang, Y., Hou, W., Meng, Q., Li, C., Chen, J., and Li, J. (2023). Deep learning for real-time detection of nasopharyngeal carcinoma during nasopharyngeal endoscopy. iScience *26*, 107463.
39. Li, C., Jing, B., Ke, L., Li, B., Xia, W., He, C., Qian, C., Zhao, C., Mai, H., Chen, M., et al. (2018). Development and validation of an endoscopic images-based deep learning model for detection with nasopharyngeal malignancies. Cancer Commun. *38*, 59.
40. Wang, S.X., Li, Y., Zhu, J.Q., Wang, M.L., Zhang, W., Tie, C.W., Wang, G.Q., and Ni, X.G. (2024). The Detection of Nasopharyngeal Carcinomas Using a Neural Network Based on Nasopharyngoscopic Images. Laryngoscope *134*, 127–135.
41. Li, C., Liu, H., Liu, Y., Feng, B.Y., Li, W., Liu, X., Chen, Z., Shao, J., and Yuan, Y. (2024). Endora: Video Generation Models as Endoscopy Simulators. Preprint at arXiv. https://doi.org/10.48550/arXiv.2403.11050.
42. Liu, Y., Li, C., Yang, C., and Yuan, Y. (2024). EndoGaussian: Gaussian Splatting for Deformable Surgical Scene Reconstruction. Preprint at arXiv. https://doi.org/10.48550/arXiv.2401.12561.
43. Ding, Z., Dong, Q., Xu, H., Li, C., Ding, X., and Huang, Y. (2022). Unsupervised Anomaly Segmentation for Brain Lesions Using Dual Semantic-Manifold Reconstruction. In International Conference on Neural Information Processing (Springer International Publishing), pp. 133–144.
44. Xu, H., Zhang, Y., Sun, L., Li, C., Huang, Y., and Ding, X. (2022). AFSC: Adaptive Fourier Space Compression for Anomaly Detection. Preprint at arXiv. https://doi.org/10.48550/arXiv.2204.07963.
45. Baran, H., Aydın, S., and Elibol, E. (2021). A Retrospective Analysis of Nasopharyngeal Biopsy Results in Adult Patients. Cureus *13*, e13455.
46. Li, W.Z., Lv, S.H., Liu, G.Y., Liang, H., Xia, W.X., and Xiang, Y.Q. (2021). Age-dependent changes of gender disparities in nasopharyngeal carcinoma survival. Biol. Sex Differ. *12*, 18.
47. Zhou, D., Tian, F., Tian, X., Sun, L., Huang, X., Zhao, F., Zhou, N., Chen, Z., Zhang, Q., Yang, M., et al. (2020). Diagnostic evaluation of a deep learning model for optical diagnosis of colorectal cancer. Nat. Commun. *11*, 2961.
48. Shorten, C., and Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. J. Big Data *6*, 60.
49. Lu, X., Wang, F.L., Guo, X., Wang, L., Zhang, H.B., Xia, W.X., Li, S.W., Li, N.W., Qian, C.N., and Xiang, Y.Q. (2013). Favorable prognosis of female patients with nasopharyngeal carcinoma. Chin. J. Cancer *32*, 283–288.
50. Li, C., Lin, M., Ding, Z., Lin, N., Zhuang, Y., Huang, Y., Ding, X., and Cao, L. (2022). Knowledge condensation distillation. In European Conference on Computer Vision (Springer Nature Switzerland), pp. 19–35.
51. Li, C., Lin, X., Mao, Y., Lin, W., Qi, Q., Ding, X., Huang, Y., Liang, D., and Yu, Y. (2022). Domain generalization on medical imaging classification using episodic training with task augmentation. Comput. Biol. Med. *141*, 105144.
52. Li, C., Feng, B.Y., Fan, Z., Pan, P., and Wang, Z. (2023). Steganerf: Embedding invisible information within neural radiance fields. In Proceedings of the IEEE/CVF International Conference on Computer Vision (IEEE), pp. 441–453.
53. Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.
54. Li, C., Lin, M., Ding, Z., Lin, N., Zhuang, Y., Huang, Y., Ding, X., and Cao, L. (2022). Knowledge condensation distillation. In European Conference on Computer Vision (Springer Nature Switzerland), pp. 19–35.
55. Pan, P., Fan, Z., Feng, B.Y., Wang, P., Li, C., and Wang, Z. (2023). Learning to Estimate 6dof Pose from Limited Data: A Few-Shot, Generalizable Approach Using Rgb Images. Preprint at arXiv. https://doi.org/10.48550/arXiv.2306.07598.
56. Liang, Z., Rong, Y., Li, C., Zhang, Y., Huang, Y., Xu, T., Ding, X., and Huang, J. (2021). Unsupervised large-scale social network alignment via cross network embedding. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 1008–1017.
57. Li, C., Liu, H., Fan, Z., Li, W., Liu, Y., Pan, P., and Yuan, Y. (2024). GaussianStego: A Generalizable Stenography Pipeline for Generative 3D Gaussians Splatting. Preprint at arXiv. https://doi.org/10.48550/arXiv.2407.01301.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Nasopharyngeal endoscopic images and videos	Fujian Cancer Hospital (FCH) and Jiangxi Cancer Hospital (JCH)	N/A
Pathological examination reports	FCH and JCH pathology departments	N/A
Deposited data		
Raw and analyzed data	This paper	N/A
Software and algorithms		
R	R Foundation for Statistical Computing	https://www.r-project.org/
caret package	CRAN	https://cran.r-project.org/web/packages/caret/index.html
pROC package	CRAN	https://cran.r-project.org/web/packages/pROC/index.html
DenseNet	Huang et al 2017[53]	N/A
Other		
PENTAX Medical endoscope (Model No. VNL-1190STK, VNL-1070STK, VNL-1570STK)	PENTAX Medical	N/A
Endoscopy capture recorder (Model No. EPK-i5000)	PENTAX Medical	N/A

RESOURCE AVAILABILITY

Lead contact

Dan Wu wudan_0729@163.com.

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The nasopharyngeal endoscopic images and videos used in this study are not publicly available due to patient privacy concerns. However, these data may be available from the corresponding author upon reasonable request and with appropriate data use agreements.
- The code used for training and evaluating the NPCNet model will be available from the corresponding author upon reasonable request.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This retrospective diagnostic study analyzed nasopharyngeal endoscopic images and videos from 1108 participants (predominantly Han Chinese, aged 12–85 years) at Fujian Cancer Hospital (FCH) and Jiangxi Cancer Hospital (JCH). The dataset, randomly divided into training and validation sets at the patient level, included 704 individuals in the training set (NPC patients: mean age 50 years, 74.9% male; controls: mean age 52 years, 60.7% male). The validation set comprised 1476 static images from 204 FCH patients, 2180 static images from 200 JCH patients, and 309 videos from FCH (98 NPC patients, 106 healthy controls, 105 post-radiotherapy patients). Nasopharyngoscopies were performed by a senior radiologist with over 15 years of experience in radiation oncology, while an instrumentation engineer extracted the images and videos. Pathological examination reports were obtained from both hospitals' pathology departments. All data, including a distinct healthy control group in the video test set, were anonymized before analysis. The study subject inclusion process is illustrated in Figure 1.

METHOD DETAILS

Study design and participants

We carried out a retrospective diagnostic study utilizing endoscopic images and videos from Fujian Cancer Hospital (FCH) and Jiangxi Cancer Hospital (JCH) in China. A senior radiologist, with over 15 years of experience in radiation oncology, conducted the nasopharyngoscopies, while an instrumentation engineer was responsible for the extraction of the images and videos. The amassed image dataset was randomly segregated into training and validation sets at the patient level, ensuring there was no overlap between the two sets. The study encompassed patients aged between 12 and 85 years who had undergone pathological examination, with a distinct healthy control group included in the video test set. Pathological examination reports were procured from the pathology departments at FCH and JCH. All data and reports were anonymized prior to their utilization in the research. The process of study subject inclusion is illustrated in Figure 1.

This study was approved by the institutional review board (IRB) of both hospitals and undertaken according to the Declaration of Helsinki. Informed consent from patients with NPC and controls was not required by the IRB given the retrospective nature of this study.

Procedures

All nasopharyngeal endoscopic images were procured from the Imaging Databases at Fujian Cancer Hospital (FCH) and Jiangxi Cancer Hospital (JCH). The static images were provided in JPEG format, while the video streams were in AVI format. The endoscopic equipment utilized for the nasopharyngeal examination was manufactured by PENTAX Medical. The images were captured using an endoscope (Model No. VNL-1190STK, VNL-1070STK, VNL-1570STK) and an endoscopy capture recorder (Model No. EPK-i5000). The image quality control process involved the removal of out-of-focus and blurred images based on the review by the endoscopist.

All participants in this study underwent a pathological examination, which included a control group of healthy nasopharynx and patients who had undergone radiotherapy. The control group comprised 106 healthy patients and 105 patients who had completed radiotherapy with a minimum duration of six months. These patients had been clinically evaluated and demonstrated complete tumor elimination. Post-radiotherapy, changes in the nasopharynx, such as inflammation and mucosal atrophy, were commonly observed. Figure 2 illustrates Hematoxylin and Eosin (HE) pathological images of NPC patients and inflammation in the nasopharyngeal mucosa post-radiotherapy.

In previous studies, a deep convolutional neural network model for colorectal cancer was trained based on DenseNet,[53] a neural network framework. The model was pretrained with 464,105 colonoscopic images. In this NPC image research, we fine-tuned the last two layers of this model, setting the last two layers of the model as fully connected layers for training.[54] The number of classes was set to 2 (i.e., benignity and malignancy determined from pathological reports). The parameters of the last two layers of the pretrained model were updated by stochastic gradient descent with a momentum of 0.9, a learning rate of 0.0001, and a batch size of 256 for 30 epochs, while the parameters of the other layers were frozen.[38,55–57] The data augmentations included random resized crops, random horizontal flipping, color jittering, and mixup. The architecture of DenseNet we use is illustrated in Figure 3.

Upon completion of the model training, we computed the malignancy score for each individual in the validation sets. Videos were converted to images with one image per frame. After importing the net and weights, all images of the test sets were read to calculate the malignancy score. Each patient had multiple images, resulting in multiple scores. After comparing the average value and weighted average value, we selected the highest score as the patient malignancy score, which was used to calculate the Area Under the Curve (AUC), accuracy, and sensitivity. The malignancy score was computed as follows:

$$p = \max \left[p1, p2, p3, p4, \ldots, pn\right]$$

We also calculated the score of NPC images after partly erasing pixels and subtracted this score from the original score to derive a visual explanation by localizing the image area that most influenced the decision.

QUANTIFICATION AND STATISTICAL ANALYSIS

The performance of our deep learning model (NPCNet) was evaluated using Receiver Operating Characteristic (ROC) curve analysis, a standard method for assessing binary classification models. The ROC curve was constructed by plotting the true positive rate (sensitivity) against the false-positive rate (1-specificity) across a range of probability thresholds. This approach allows for a comprehensive evaluation of the model's discriminative ability at various operating points. The Area Under the ROC Curve (AUC) was calculated to quantify the overall performance of the model, with values ranging from 0.5 (no discriminative power) to 1.0 (perfect discrimination).

To provide a measure of statistical uncertainty, we calculated 95% confidence intervals (CIs) for sensitivity and specificity using the Clopper-Pearson method. This method, also known as the exact binomial method, is particularly suitable for proportion data and provides conservative interval estimates. Sensitivity, specificity, and accuracy were calculated at various thresholds to assess the model's performance under different operating conditions, allowing for a nuanced understanding of its strengths and limitations.

A confusion matrix was constructed to provide a detailed breakdown of the model's predictions, reporting true positives (correctly identified NPC cases), false positives (non-NPC cases incorrectly classified as NPC), false negatives (NPC cases missed by the model), and true negatives (correctly identified non-NPC cases). This matrix offers insights into the types of errors made by the model and helps in assessing its clinical utility.

All statistical analyses were performed using R software (version 4.1.2, R Foundation for Statistical Computing, Vienna, Austria), a widely-used platform for statistical computing and graphics. We utilized the 'caret' package for calculating performance metrics such as AUC, accuracy, sensitivity, and specificity. The 'pROC' package was employed for ROC curve analysis and visualization, providing a robust framework for model evaluation.

To assess the model's performance across different patient subgroups, we conducted stratified analyses based on demographic factors (such as age and gender) and clinical characteristics (including disease stage and histological subtypes). These subgroup analyses help identify any variations in model performance that might be clinically relevant and guide future refinements.

For comparisons between subgroups or different models, we employed appropriate statistical tests, including McNemar's test for paired nominal data and DeLong's test for comparing AUCs. A p-value threshold of 0.05 was used to determine statistical significance, balancing the risks of Type I and Type II errors. All analyses were performed in accordance with established guidelines for reporting diagnostic accuracy studies, ensuring transparency, reproducibility, and adherence to best practices in biomedical research.