

LETTER TO EDITOR

Development of a deep learning-based nomogram for predicting lymph node metastasis in cervical cancer: A multicenter study

Dear Editor,

Cervical cancer is one of the most frequently diagnosed cancers in women and has a high mortality rate worldwide.¹ Lymph node metastasis (LNM) is an important prognostic factor in patients with cervical cancer.²⁻⁴ The assessment of LNM before treatment is essential to guide and tailor the treatment.^{5,6} The morphological examination of lymph nodes via medical images is commonly used for diagnosing LNM. However, it depends mainly on radiologists' experience and has relatively low accuracy. Thus, we collected a multi-center dataset and developed a deep learning-based nomogram (DLN) to improve the accuracy of LNM diagnosis in cervical cancer.

In total, 1123 cervical cancer patients with computed tomography (CT) examination were enrolled from 13 centers in our study (Table S1 and Supplementary A1). As shown in Supplementary A2 and Figure S1, we divided these patients into four cohorts: training cohort, validation cohort, external testing cohort 1, and external testing cohort 2. Detailed information on the four cohorts is presented in Table S2. The clinical characteristics included age, gravidity, histological type, FIGO stage, etc. Moreover, two experienced gynecologists, who were blinded to the pathological report, were invited to diagnose the status of LNM together using only CT images. Additionally, a follow-up cohort including 148 patients from one center was used for survival analysis.

The workflow of this study is described in Figure 1, including region of interest (ROI) segmentation, data preprocessing (Supplementary A3), model construction, and model evaluation (Supplementary A4).

We invited experienced gynecologists to segment ROIs in normalized CT images. Before model construction, data augmentations, including flipping, rotating, and random cropping, were used to generate new training samples to avoid overfitting. Oversampling methods were used to balance the ratio of LNM-positive patients and LNM-negative

patients in the training cohort. Three state-of-the-art deep learning methods, including ResNet18,⁷ ResNet50,⁷ and SE-Net,⁸ were used to construct three candidate models (Supplementary A5). As shown in Table S3, ResNet18 showed the best performance in the validation cohort, and thus it was selected to build the final deep learning signature (Sig_DL). As shown in Supplementary A6, a total of 1407 handcrafted radiomic features were extracted, and three key radiomic features were selected via a series of feature selection methods and integrated them into a radiomic signature (Sig_radiomic).⁹⁻¹⁰ As shown in Table 1 and Figure S2, the AUCs of Sig_DL performed better than Sig_radiomic in all the cohorts.

Additionally, univariate analysis was used to screen for significant clinical features. We noticed that the FIGO stage was significantly associated with LNM ($P < 0.01$). After multivariable logistic regression, we selected the FIGO stage and age as key clinical features and used them to construct a clinical signature (Sig_clin). The area under the receiver operating characteristic curve (AUCs) of Sig_clin reached 0.678 and 0.597 in training and validation cohorts, respectively.

Finally, we integrated Sig_DL, diagnoses of gynecologists, and all significant clinical features into a DLN via multivariate linear regression analysis (Table S4 and Figure 2A). Compared with other models, DLN had the best predictive ability (Figure S3), with AUCs of 0.867, 0.807, 0.781, and 0.804 in the training cohort, validation cohort, external testing cohort1 and external testing cohort2 (Figure 2B-E). As shown in Table 1, the accuracy also indicated the good performance of DLN in these four cohorts.

Meanwhile, the decision curves showed that the patients could benefit more from DLN than both Sig_DL and Sig_clin (Figure 2F). As shown in Figure 2G, the calibration curves demonstrated that the DLN had good consistency with the gold standard of LNM.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Clinical and Translational Medicine* published by John Wiley & Sons Australia, Ltd on behalf of Shanghai Institute of Clinical Bioinformatics.

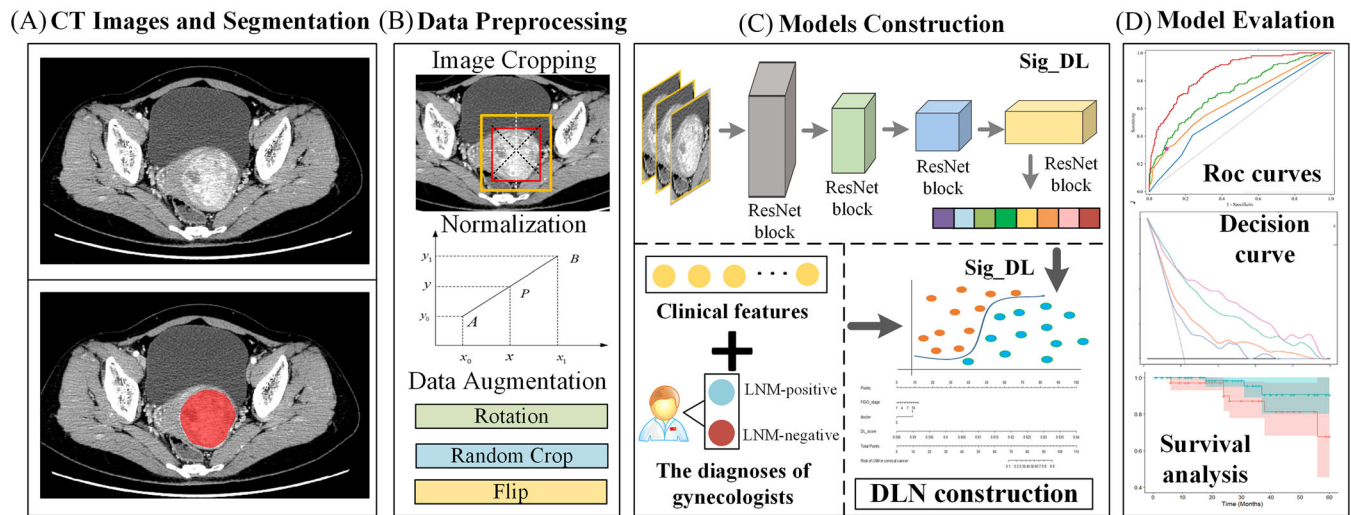


FIGURE 1 Workflow of the development of deep learning-based nomogram (DLN). (A) CT images and segmentation, (B) Data preprocessing. (C) Model construction. (D) Model evaluation

TABLE 1 Performance of models in all cohorts

	Specificity	Sensitivity	Accuracy	AUC (95% CI)	True negative	True positive	False negative	False positive
Sig_clin								
Training cohort	0.578	0.678	0.600	0.678 (0.619–0.726)	289	97	46	218
validation cohort	0.710	0.423	0.665	0.597 (0.481–0.722)	97	11	15	41
External testing1	0.579	0.348	0.544	0.489 (0.367–0.605)	73	8	15	53
External testing2	0.615	0.572	0.609	0.626 (0.485–0.755)	80	12	9	50
Sig_radiomic								
Training cohort	0.527	0.545	0.531	0.575 (0.520–0.626)	269	78	65	247
validation cohort	0.522	0.577	0.530	0.621 (0.505–0.746)	72	15	11	66
External testing 1	0.508	0.652	0.530	0.616 (0.497–0.735)	64	15	8	62
External testing 2	0.562	0.619	0.570	0.595(0.475–0.714)	73	13	8	57
Sig_DL								
Training cohort	0.734	0.818	0.753	0.853 (0.821–0.885)	379	117	26	137
validation cohort	0.710	0.731	0.713	0.787 (0.702–0.878)	98	19	7	40
External testing 1	0.651	0.739	0.664	0.776 (0.677–0.877)	82	17	6	29
External testing 2	0.777	0.714	0.768	0.768 (0.662–0.874)	101	15	6	29
DLN^a								
Training cohort	0.793	0.790	0.792	0.867 (0.839–0.897)	412	113	30	104
validation cohort	0.783	0.654	0.762	0.807 (0.713–0.889)	108	16	10	30
External testing 1	0.714	0.739	0.718	0.781 (0.669–0.876)	91	17	6	35
External testing 2	0.808	0.667	0.788	0.804 (0.705–0.892)	105	14	7	25

^aDLN, deep learning-based nomogram.

It is worth noting that the diagnoses of the gynecologists had high specificity but low sensitivity in our cohorts. Therefore, we modified the cutoff value so that DLN could have the same specificity as the gynecologists' diagnoses. Then, we found that DLN had better accuracy and sensitiv-

ity than the gynecologists (Table S5). The Venn diagrams also showed that DLN had more true positive cases than the gynecologists (Figure S4). Four typical cases are shown in Figure 3, which indicates that DLN could help the clinician reduce the risk of misdiagnosis.

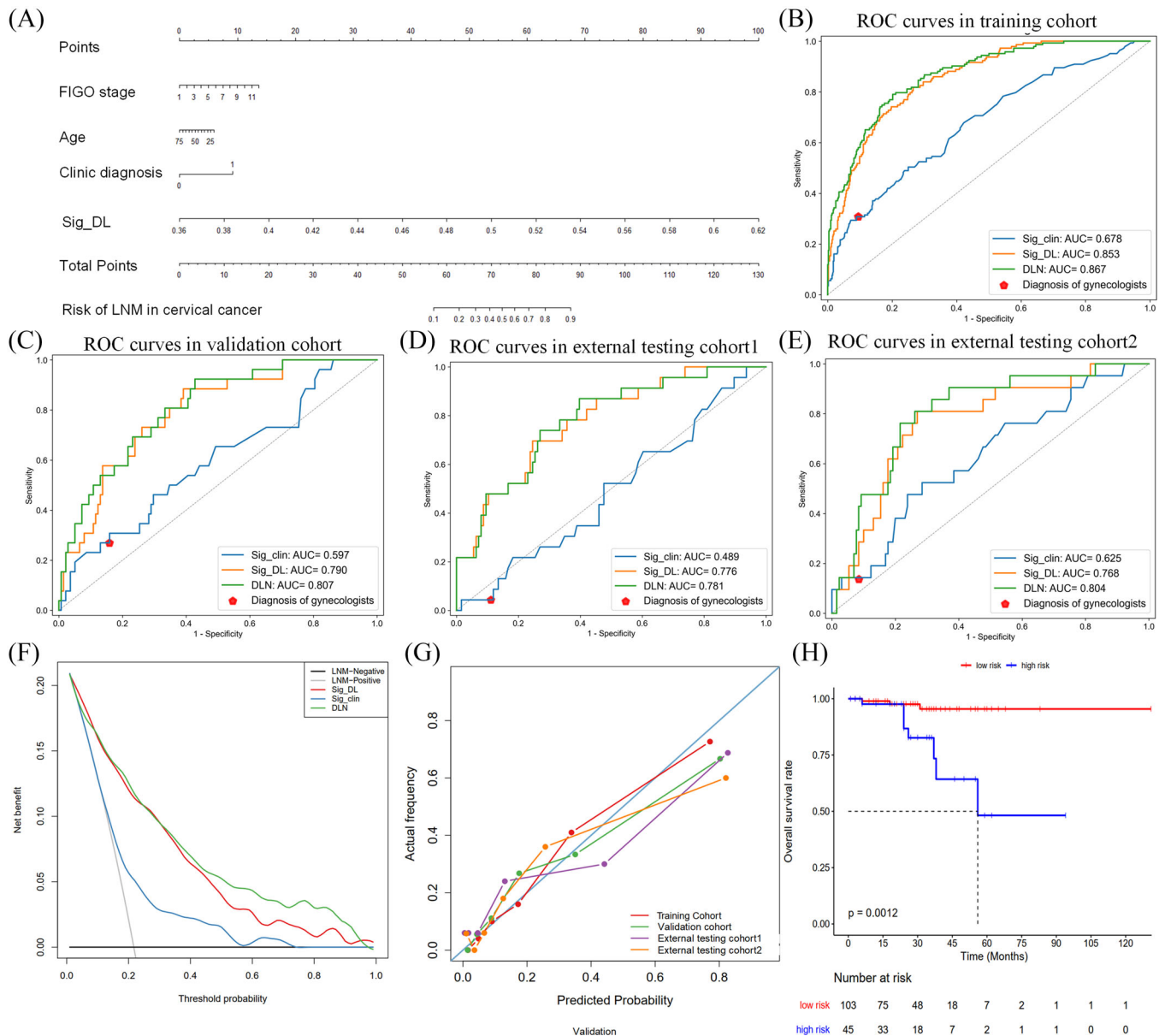


FIGURE 2 Evaluation of the deep learning-based nomogram (DLN) model. (A) The DLN model. The ROC curves in the training cohort (B), validation cohort (C), external testing cohort 1 (D), and external testing cohort 2 (E). (F) The decision curve of all models. (G) The calibration curves of the DLN. (H) Kaplan-Meier curves of DLN in the follow-up cohort. ROC, receiver operating characteristic curve

Subgroup analysis was performed on the data of the enrolled patients, including their clinical characteristics, the CT manufacturers, and the centers. As shown in Figure S5A-F, the subgroup analysis indicates that the DLN was not affected by age, times of pregnancy, human papillomavirus (HPV) testing result, and histological type. Especially, we selected 614 cervical cancer patients for human papillomavirus (HPV) testing. Subgroup analysis revealed that our DLN showed good performance in both HPV-positive subgroup and HPV-negative subgroup (Figure S5G-H). Our model also was

minimally affected by the CT manufacturers and centers (Figure S6A,B).

Besides, 148 cervical cancer patients with follow-up from Center 2 were used for exploring the association between DLN score and overall survival (OS) using Kaplan-Meier curves (Supplementary A7). We divided them into low-risk and high-risk groups using the mean value of DLN score as a cutoff. As shown in Figure 2H, we found that the high-risk group exhibited shorter OS (log-rank test: $P = 0.0012$). Furthermore, we stratified patients via the FIGO stage for comparison, however, the FIGO stage showed no sig-

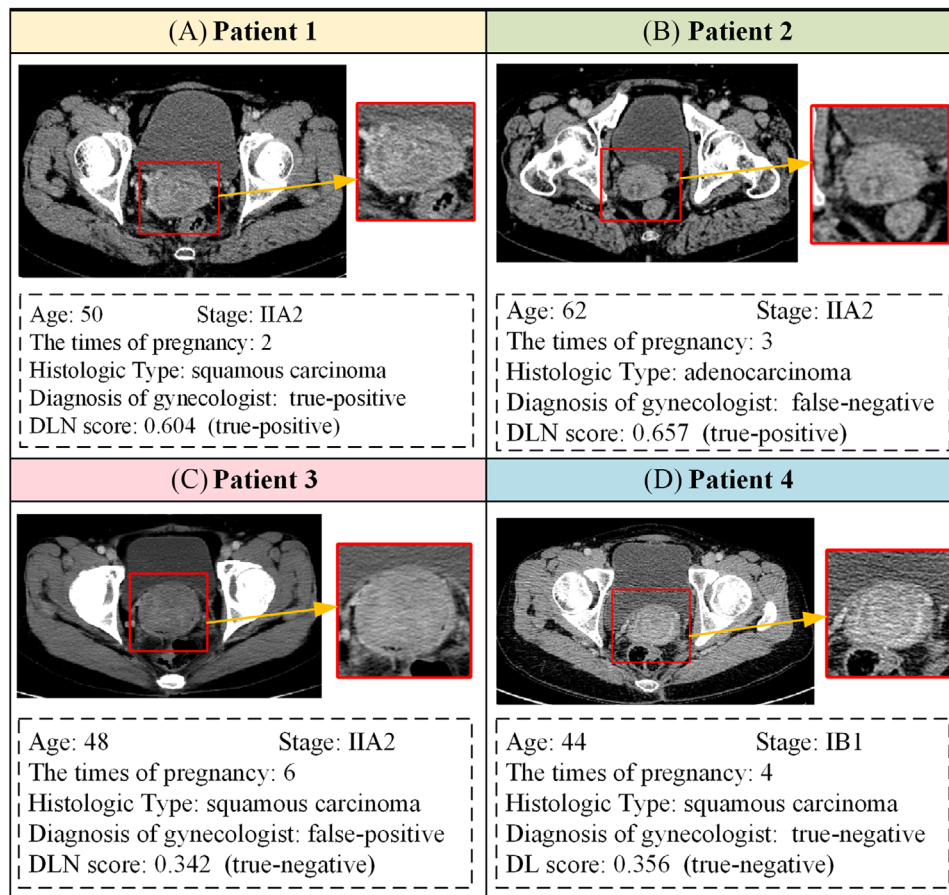


FIGURE 3 Four typical cases to show the performance of deep learning-based nomogram (DLN) and gynecologists. (A) A lymph node metastasis (LNM)-positive patient correctly diagnosed by both DLN and gynecologists; (B) An LNM-positive patient only correctly diagnosed by DLN; (C) An LNM-negative patient only correctly diagnosed by DLN; (D) An LNM-positive patient correctly diagnosed by both DLN and gynecologists

nificant association with OS (Figure S7). Hence, DLN could serve as a significant prognostic factor for cervical cancer.

In conclusion, we developed a deep learning model for the preoperative prediction of LNM in cervical cancer and validated it in a large-scale and multicenter dataset. The performance of DLN surpassed the diagnosis of experienced gynecologists. Therefore, DLN can serve as a non-invasive tool for LNM determination and thus assist treatment decision-making.




ACKNOWLEDGEMENTS

This work was supported by Strategic Priority Research Program of Chinese Academy of Sciences (XDB 38040200), National Key R&D Program of China (2017YFA0205200), National Natural Science Foundation of China (82022036, 91959130, 81971776, 81771924, 62027901, 81930053, 81227901), the Beijing Natural Science Foundation (Z20J00105), Chinese Academy of Sciences under Grant

No. GJJSTD20170004 and QYZDJ-SSW-JSC005, the Project of High-Level Talents Team Introduction in Zhuhai City (Zhuhai HLHPTP201703), and the Youth Innovation Promotion Association CAS (Y2021049). The authors would like to acknowledge the instrumental and technical support of Multi-modal biomedical imaging experimental platform, Institute of Automation, Chinese Academy of Sciences.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Yujia Liu^{1,2} 
 Hui Duan³ 
 Di Dong^{1,2,4} 
 Jiaming Chen^{3,5} 
 Lianzhen Zhong^{1,2} 
 Liwen Zhang^{1,2} 
 Runnan Cao^{1,2} 

Huijian Fan³ 
 Zhumei Cui⁶
 Ping Liu³ 
 Shan Kang⁷
 Xuemei Zhan⁸
 Shaoguang Wang⁹
 Xun Zhao^{1,2}
 Chunlin Chen³ 
 Jie Tian^{2,4,10,11} 

¹*School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China*

²*CAS Key Laboratory of Molecular Imaging, the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China*

³*Department of Obstetrics and Gynecology, Nanfang Hospital, Southern Medical University, Guangzhou, China*

⁴*Beijing Key Laboratory of Molecular Imaging, Beijing, China*

⁵*Huizhou Municipal central Hospital, Huizhou, China*

⁶*The affiliated hospital of Qingdao University, Qingdao, China*

⁷*Department of Gynecology, Fourth Hospital Hebei Medical University, Shijiazhuang, China*

⁸*Jiangmen central Hospital, Jiangmen, China*

⁹*Department of Gynecology, Yantai Yuhuangding Hospital, Yantai, China*

¹⁰*Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Engineering Medicine, Beihang University, Beijing, China*

¹¹*Zhuhai Precision Medical Center, Zhuhai People's Hospital (Affiliated with Jinan University), Zhuhai, China*












Correspondence

Chunlin Chen, Department of Obstetrics and Gynecology, Nanfang Hospital, Southern Medical University, No. 1838, Guangzhou Avenue, Guangzhou 510515, China.
 Email: ccl1@smu.edu.cn

Jie Tian Ph.D, Beijing Advanced Innovation Centre for Big Data-Based Precision Medicine, School of Engineering Medicine, Beihang University, Beijing, 100191, China.
 Email: jie.tian@ia.ac.cn

Yujia Liu, Hui Duan, Di Dong and Jiaming Chen contributed equally as co-first authors.

ORCID

Yujia Liu  <https://orcid.org/0000-0002-1561-2582>
 Hui Duan  <https://orcid.org/0000-0002-8335-268X>
 Di Dong  <https://orcid.org/0000-0003-0783-3171>
 Jiaming Chen  <https://orcid.org/0000-0002-0073-8174>
 Lianzhen Zhong  <https://orcid.org/0000-0002-8426-5289>
 Liwen Zhang  <https://orcid.org/0000-0001-9816-1344>
 Runnan Cao  <https://orcid.org/0000-0003-1682-3960>
 Huijian Fan  <https://orcid.org/0000-0002-9520-3067>
 Ping Liu  <https://orcid.org/0000-0001-6769-3479>
 Chunlin Chen  <https://orcid.org/0000-0002-7140-5668>
 Jie Tian  <https://orcid.org/0000-0003-0498-0432>

REFERENCES

- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209-249.
- Koh WJ, Abu-Rustum NR, Bean S, et al. Cervical cancer, version 3.2019, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network.* 2019;17(1):64-84.
- Kan Y, Dong D, Zhang Y, et al. Radiomic signature as a predictive factor for lymph node metastasis in early-stage cervical cancer. *J Magn Reson Imaging.* 2019;49(1):304-310.
- Gien LT, Covens A. Lymph node assessment in cervical cancer: prognostic and therapeutic implications. *J Surg Oncol.* 2009;99(4):242-247.
- Liu Y, Fan H, Dong D, et al. Computed tomography-based radiomic model at node level for the prediction of normalized lymph node metastasis in cervical cancer. *Translational Oncology.* 2021;14(8):101113.
- Chen J, He B, Dong D, et al. Noninvasive CT radiomic model for preoperative prediction of lymph node metastasis in early cervical carcinoma. *Br J Radiol.* 2020;93(1108):20190558.
- He K, Zhang X, Ren S, et al. *Deep residual learning for image recognition.* IEEE; 2016.
- Hu J, Shen L, Sun G. *Squeeze-and-excitation networks.* IEEE; 2018.
- Bluemke DA, Moy L, Bredella MA, et al. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers—from the radiology editorial board. *Radiology.* 2020;294(3):487-489.
- Zwanenburg A, Leger S, Vallières M, et al. Image biomarker standardisation initiative. arXiv; 2016.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.