CellPress

Article

Deep learning radiomics-based preoperative prediction of recurrence in chronic rhinosinusitis

CRS patients who needed surgical therapy after failure of appropriate medical treatment

Deep Learning Signature
sinus segmentation, recurrence prediction

Clinical Model
recurrence prediction

Allergic diseases   Eosinophil

Combined Nomogram

Shaojuan He, Wei Chen, Xuehai Wang, ..., Xuezhong Li, Anning Li, Xin Feng

anningli00@163.com (A.L.)
drfengxin@sdu.edu.cn (X.F.)

Highlights

3D U-Net can automatically segment nasal cavity and nasal sinuses structures

A nomogram combining a multi-task DLR signature and clinical factors

The fusion model can accurately predict the recurrence of CRS preoperatively

We offered a promising practice of deep learning radiomics in non-cancer diseases
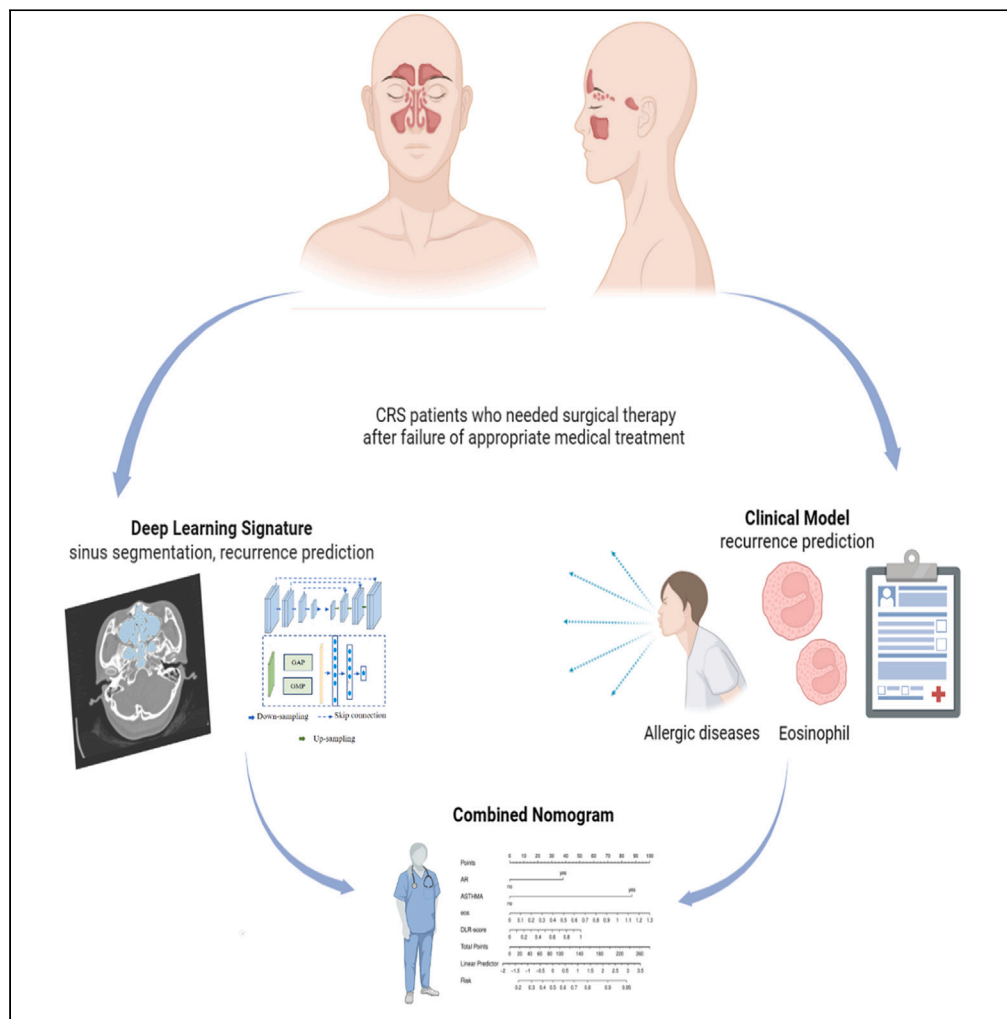
iScience

Article

Deep learning radiomics-based preoperative prediction of recurrence in chronic rhinosinusitis

Shaojuan He,[1,2,5] Wei Chen,[3,5] Xuehai Wang,[4] Xinyu Xie,[1] Fangying Liu,[1] Xinyi Ma,[1] Xuezhong Li,[1] Anning Li,[2,*] and Xin Feng[1,6,*]

SUMMARY

Chronic rhinosinusitis (CRS) is characterized by poor prognosis and propensity for recurrence even after surgery. Identification of those CRS patients with high risk of relapse preoperatively will contribute to personalized treatment recommendations. In this paper, we proposed a multi-task deep learning network for sinus segmentation and CRS recurrence prediction simultaneously to develop and validate a deep learning radiomics-based nomogram for preoperatively predicting recurrence in CRS patients who needed surgical treatment. 265 paranasal sinuses computed tomography (CT) images of CRS from two independent medical centers were analyzed to build and test models. The sinus segmentation model achieved good segmentation results. Furthermore, the nomogram combining a deep learning signature and clinical factors also showed excellent recurrence prediction ability for CRS. Our study not only facilitates a technique for sinus segmentation but also provides a noninvasive method for preoperatively predicting recurrence in patients with CRS.

INTRODUCTION

Chronic rhinosinusitis (CRS) is a common inflammatory disease of the nose and paranasal sinuses with a prevalence of 5–12%, which is characterized by significant disease burden, specific computed tomography (CT) changes, and endoscopic signs. This disease can be divided into CRS without nasal polyps (CRSsNP) and CRS with polyps (CRSwNP).[1] For patients with CRS, continued appropriate medical therapy combined with surgical treatment if necessary is the first-line treatment.[2] However, the prognosis of these patients is still poor with early recurrence, especially for patients with CRSwNP. Moreover, some long-term follow-up studies showed revision polypectomy rates were as high as 20%–50%.[3–5] Hence, for those patients who had failed appropriate medical therapy and needed surgical treatment, accurate detection of relapse risk factors and identification of CRS patients with high risk of recurrence preoperatively would be important in assisting in optimizing surgical case selection and instructing patients to use medicines and endoscopic debridement periodically after surgery as prescribed to reduce the recurrence rate of CRS.

Radiomics, an emerging tool, can extract high-throughput features from noninvasive medical imaging such as CT, positron emission tomography (PET), and magnetic resonance (MR) and construct machine learning models for diagnostic and prognostic prediction.[6] One of the biggest challenges in the radiomics analysis of CRS patients is the complex anatomy of the sinus region, such as many cavities, which is very unfavorable for extracting traditional radiomics features and brings great challenges to the modeling construction. Additionally, sinus segmentation needs to be performed manually by experienced radiologists, which is an extremely tedious and time-consuming process. Furthermore, the usual radiomics studies directly extract features from segmentation and will not consider the implicit information during the segmentation process. Fortunately, deep learning-based radiomics can directly take raw images or rectangular regions containing tumors as input and automatically learns discriminative features.[7] Recently, deep learning-based methods have made outstanding achievements in various medical imaging segmentation,[8,9] and multi-task learning can solve this problem by performing segmentation and predictive analysis simultaneously in an end-to-end model.[10,11]

To date, deep learning has gained remarkable success in multiple fields, such as lung cancer,[12] liver cancer,[13] brain tumor,[14] etc. However, fewer studies have applied deep learning to non-cancer diseases. Furthermore, although many previous studies have explored the clinical risk factors for the recurrence of

Figure 1. The flowchart of this study

CRS,[15,16] these studies only focused on unilateral factors and could not provide a comprehensive and accurate prognostic assessment for CRS. To our knowledge, few studies have tried to comprehensively assess the prognosis prediction for CRS recurrence based on combination of clinical information, laboratory tests, and radiology. Thus, we aimed to develop and validate a comprehensive nomogram combining a multi-task deep learning-based radiomics signature and clinical factors for preoperatively predicting recurrence in CRS patients who needed surgical treatment in this study.

RESULTS

Characteristics of the patients

In this study, we designed a multi-task deep learning network for sinus segmentation and CRS recurrence prediction simultaneously and developed a comprehensive nomogram by combining a deep learning signature and clinical factors. The overall study design is shown in Figure 1. Overall, 265 patients were enrolled in this study, and 92 patients were excluded. Further details can be found in Figure 2. The training cohort consisted of 200 patients (mean age, 44years ±15 [SD]; 137 men) from medical center A. The external independent testing cohort comprised 65 patients (mean age, 51years ±11 [SD]; 47 men) from medical center B. There were no statistically significant differences in clinical variables (gender, age, smoking, drinking, hypertension, and diabetes) between the recurrence and no-recurrence groups neither in the training cohort nor in the testing cohort with the exception of asthma (training cohort p < 0.001), allergic rhinitis (training cohort p < 0.001; testing cohort p = 0.002), and circulating eosinophil (training cohort p = 0.001; testing cohort p = 0.009), justifying their use as the training and testing cohorts. Table 1 shows the clinical features of patients in the training cohort and testing cohort.

Development and validation of deep learning signature

As shown in Figure S1, the traditional radiomics model did not obtain satisfactory results with an area under the curve (AUC) of receiver operating characteristic (ROC) of just 0.496 in the testing cohort (the details of the traditional radiomics technology can be found at https://github.com/chenypic/deepSinuses). Hence, a multi-task deep learning network was proposed to perform sinus segmentation and predict CRS recurrence simultaneously (see Figure S2). For the training cohort, we performed stratified 5-fold cross
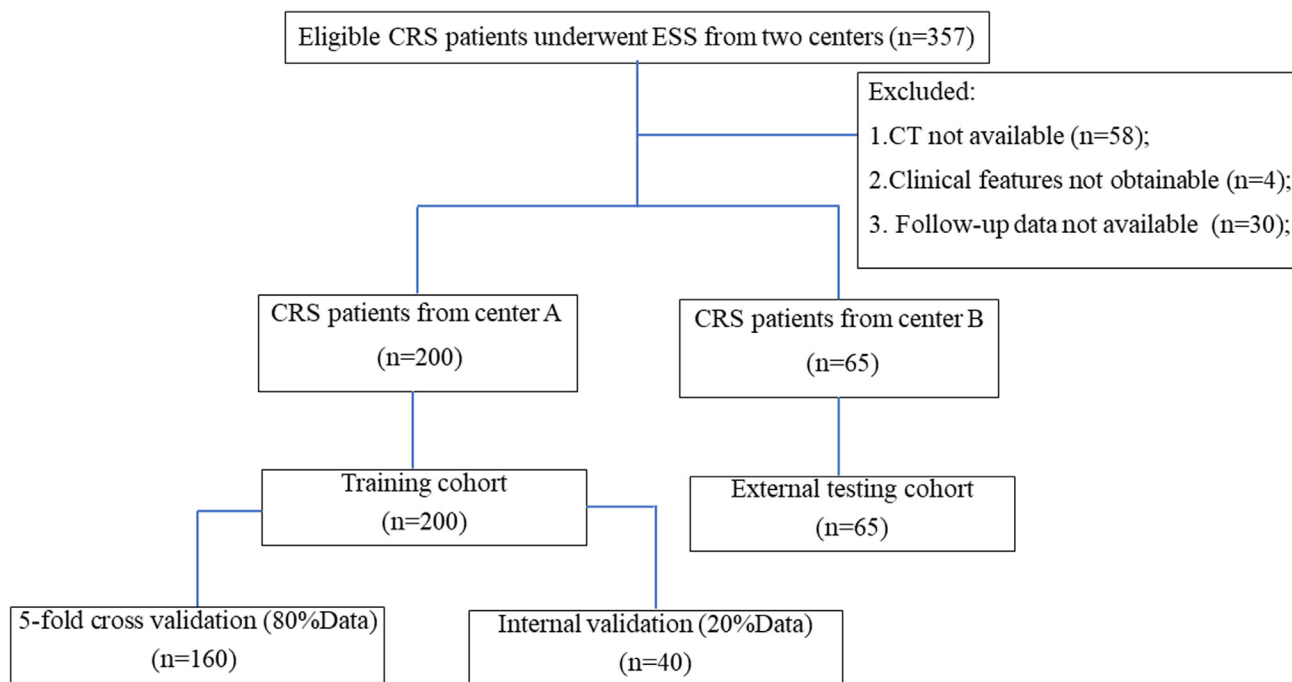
Figure 2. Flow diagram of patient inclusion and exclusion

validation, where each fold had the same proportion of the positive samples for recurrence. The results are shown in Figure 3A. The mean AUC is 0.755, proving that our model has good performance in the internal validation cohort. We then retrained the model using all the training cohorts and achieved an AUC of 0.742 (95% confidence interval [CI], 0.619–0.865) in the testing cohort, as shown in Figure 3B, suggesting that the deep learning signature can be an indicator for the recurrence of CRS. To further verify the consistency of our model, we calculated the optimal cutoff values. For the cross validation, the overall optimal cutoff value was 0.431, and the optimal cutoff value of the testing cohort was 0.435. For sinus segmentation, with the manual segmentation as gold standard, we achieved a mean dice similarity coefficient (DSC) of 0.843 in the 5-fold cross-validation cohort and 0.833 in the testing cohort. Five qualitative segmentation results

Table 1. Patients and preoperative clinical features in the training and testing cohorts

Clinical features	Training (n = 200)			Testing (n = 65)		
	Recurrence (n = 85)	Non-recurrence (n = 115)	p value	Recurrence (n = 14)	Non-recurrence (n = 51)	p value
Age, mean ± SD	44.18 ± 13.23	43.71 ± 15.64	0.933	50.79 ± 8.62	50.49 ± 11.78	0.987
Female, No. (%)	27 (31.76)	36 (31.30)	0.945	5 (35.71)	13 (25.29)	0.449
Asthma, No. (%)	32 (37.65)	6 (5.22)	<0.001*	4 (28.57)	8 (15.69)	0.271
Allergic rhinitis, No. (%)	29 (34.12)	14 (12.17)	<0.001*	11 (78.57)	17 (33.33)	0.002*
Hypertension, No. (%)	10 (11.76)	23 (20.00)	0.121	5 (35.71)	8 (15.69)	0.097
Diabetes Mellitus, No. (%)	4 (4.71)	2 (1.74)	0.224	2 (14.29)	1 (1.96)	0.052
Circulating eosinophil count, mean (SD), cells/μL	390.9 ± 277.73	257.02 ± 192.12	0.001*	621.43 ± 683.44	244.71 ± 189.90	0.009*
Smoking, No. (%)	21 (24.71)	28 (24.35)	0.954	3 (21.43)	11 (21.57)	0.991
Drinking, No. (%)	33 (38.82)	38 (33.04)	0.398	6 (42.86)	12 (23.53)	0.152

Abbreviation: SD: standard deviation. *p < 0.05.
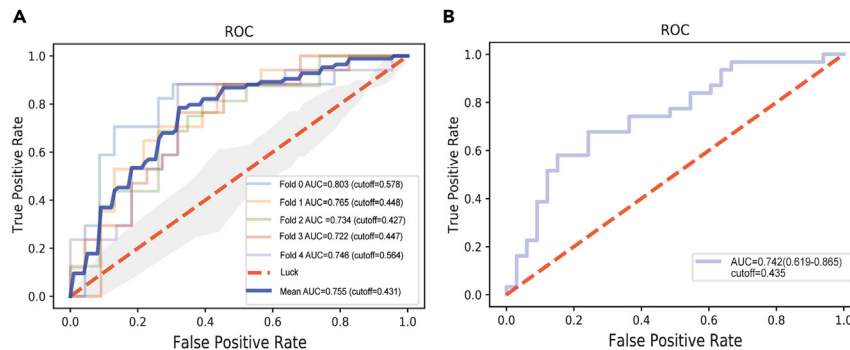
Figure 3. The ROC curves of the deep learning signature
(A) The ROCs of the 5-fold cross validation.
(B) The ROC of the testing cohort.

are shown in Figure 4. The blue contour corresponds to ground truth, while the red contour corresponds to the segmentation results. It was shown that our proposed method achieved good segmentation results.

To further verify the performance of our proposed multi-task deep learning network, we also presented single-task methods for segmentation and recurrence prediction of CRS in the testing cohort. Our experimental results showed that the single-task segmentation model achieved a DSC of 0.837, which is slightly higher than 0.833 of the multi-task model, but without a statistically significant difference (p = 0.477). For recurrence prediction, the single-task classification model never achieved results with an AUC greater than 0.6 in multiple experiments, which may be due to the severe overfitting problem. In conclusion, the result indicated that the proposed multi-task deep learning network significantly improved the predicted accuracy of CRS without the loss of segmentation accuracy.

Development and validation of the combination nomogram
In univariate analysis, clinical parameters in Table 1 were explored, and asthma, allergic rhinitis (AR), and circulating eosinophil showed significant differences between the recurrence group and non-recurrence group in the training cohort (p < 0.05), which were then used to conduct a clinical model with an AUC of 0.767 (95% CI: 0.699–0.835) in the training cohort and 0.766 (95% CI: 0.642–0.881) in the testing cohort.

Subsequently, we enrolled the deep learning radiomic score (DLR-score), asthma, AR, and circulating eosinophil as factors in a multivariable logistic regression analysis to build the personalized prediction model. According to multivariate logistic analysis, a clinical-radiomic model was built by integrating the three independent clinical factors and DLR-score.

Performance of the combination nomogram
The fusion model showed the best evaluation performance in both the two cohorts, with an AUC of 0.790 (95% CI: 0.727–0.854) in the training cohort and 0.842 (95% CI: 0.740–0.943) in the testing cohort (Figures 5A and 5B), and was visualized into a nomogram for clinicians (Figure 5C).

Furthermore, the calibration curves of the nomogram (Figures 6A and 6B) indicated that the predicted values generated by the nomogram were in good agreement with the actual results. Finally, we used a decision curve analysis (DCA) curve to assess whether this nomogram would help with clinical treatment strategies. The DCA curves (Figures 6C and 6D) demonstrated that the nomogram had a good net benefit in predicting the recurrence of CRS. Patients with CRS can benefit more from using the nomogram when the threshold probability lies across a wide range of 0.1–1. The clinical benefit of the nomogram was significantly better than that of the clinical model.

As shown in Table 2, the clinical model achieved slightly better predictive performance than the DLR model in the testing cohort (Clinical model AUC = 0.766, 95% CI [0.642, 0.881]; DLR model AUC = 0.742, 95% CI [0.619, 0.865]) with less false-negative rate (FNR). When combining DLR-score with clinical factors, the signature achieves an improved performance than either of them alone with an AUC of 0.842 (95% CI
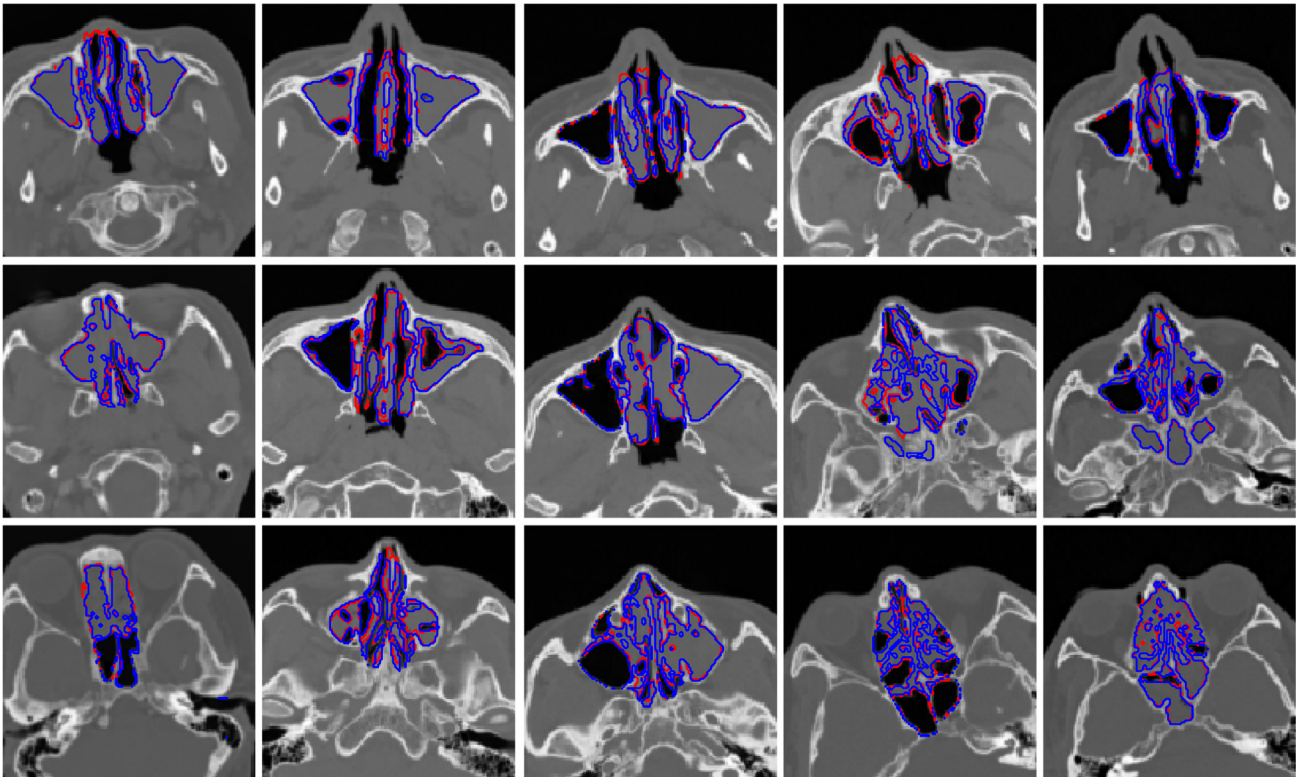
Figure 4. Five qualitative segmentation results in the cross-validation cohort
The blue contour corresponds to ground truth, while red contour corresponds to the automatic segmentation results. Note: The different-layer CT images listed in the same column belong to the same patient.

[0.740, 0.943]). Additionally, with the cutoff value of the training cohort, the nomogram also obtained the highest accuracy of 0.813 and the best specificity of 0.848. Furthermore, the significant difference from De-long test suggested significant improvement in predictive performances between the combination nomogram and the clinical model (p = 0.029), while it indicated no significant difference between the nomogram and DLR model (p = 0.130). The receiver operating characteristic (ROC) curves of the combination nomogram, clinical model, and DLR model for the overall group in the testing cohort are shown in Figure 7.

DISCUSSION

Chronic rhinosinusitis, a multifactorial heterogeneous disorder, is characterized by great quality of life (QoL) impairment, propensity for recurrence, and the pain of revision polypectomy.[1,17] According to the European Position Paper on Rhinosinusitis and Nasal Polyps (EPOS2020), rhinosinusitis is one of the top 10 most costly health conditions for United States (US) employer. The highest direct costs and QoL impairment were associated with patients who had recurrent CRS after surgery. Exploring favorable methods to accurately identify those CRS patients who have failed appropriate medical therapy with high risk of relapse preoperatively will contribute to personalized treatment recommendations to reduce the recurrence rate. In this study, we designed a multi-task deep learning network to detect relevant mass lesions of the nasal cavity and nasal sinuses, extracted multi-scale features to generate deep risk scores, and finally constructed a comprehensive nomogram which showed excellent prediction ability in CRS recurrence.

In recent decades, there are advanced understandings of the recurrence of CRS: Lou et al. demonstrated that a large number of tissue eosinophils played an important role in polyp recurrence.[18] Meng et al. suggested that the ratio of Lund-Mackay scores for the ethmoid sinus and maxillary sinus was an indicator for the recurrence of CRSwNP.[19] Beyond these, asthma and AR were also reported to be the risk factors for the relapse of CRS.[16,17] Additionally, with the development of artificial intelligence (AI), Wu et al. established an evaluation platform 2.0 (AICEP 2.0) for CRS to obtain the proportion of inflammatory cells of nasal polyps
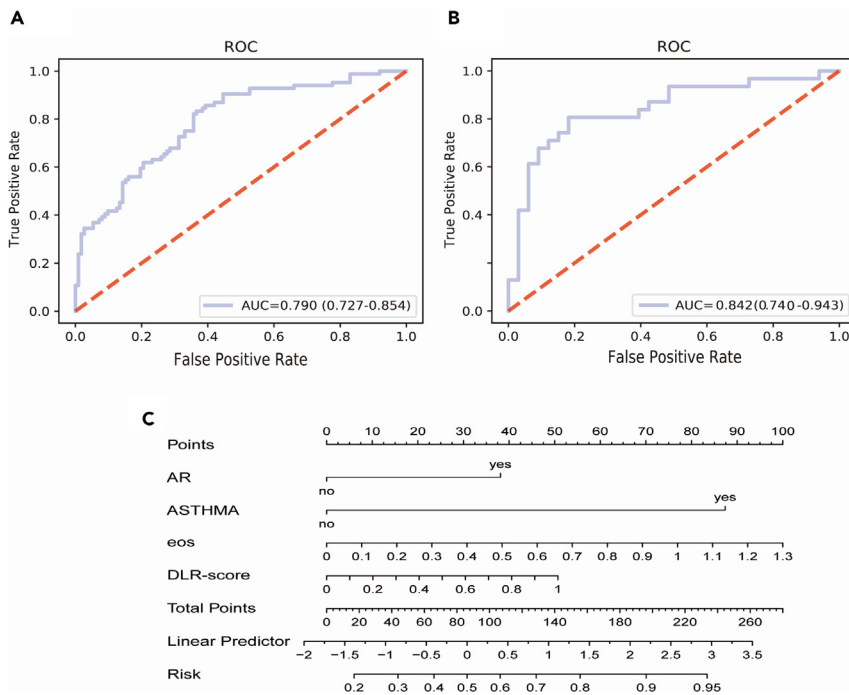
Figure 5. The ROC curves and nomogram of the clinical-radiomic model
(A) The ROC curves of the clinical-radiomic model in the training cohort and (B) the testing cohort.
(C) The developed nomogram to predict recurrence in CRS.

and found eosinophil-dominant phenotypes with the highest recurrence rate of 75% based on whole-slide imaging.[20] However, these described methods are either invasive or unilateral.

In the present study, we pioneerly explored available factors from radiology, laboratories, and physical examination in clinics to enhance the prognosis prediction for CRS patients who needed surgical treatment. For sinus segmentation, we applied 3D U-Net as the backbone network, which achieved a good segmentation result with a mean DSC of 0.843 in the 5-fold cross-validation cohort and 0.833 in the testing cohort, which could contribute to free radiologists from the time-consuming manual segmentation. Recently, an enhanced gradient level set algorithm proposed by Lao et al. showed a good segmentation effect for sinusitis lesions and demonstrated that the radiomics model can effectively predict the prognosis of endoscopic treatment of sinusitis.[21] Additionally, some studies have developed convolutional neural network (CNN) algorithm to perform automatic, volumetric segmentation of the paranasal sinuses or classification of osteomeatal complex inflammation on CT, which further confirmed the potential of deep learning-based algorithm on the disease of CRS.[22–24] However, given the shortcomings of CNNs that they cannot effectively model global and long-range semantic interactions due to the locality of convolution operations, we tried to develop a 3D U-Net algorithm to automatically segment nasal cavity and nasal sinuses structures, which also achieved promising performance. Another CT segmentation study by Pallanch and his colleagues demonstrated that volumetric CT scoring, which calculated the total volume percentage of mucosal thickening and air by using the Analyze software, was an objective outcome measure for the improvement of CRS after medical treatment, but whether the volumetric scoring can be used for predicting recurrence of CRS after surgery was not explored.[25]

In our research, we developed a multi-task deep learning structure that performed segmentation and CRS recurrence prediction simultaneously in an end-to-end model. Impressively, the processed deep radiomic signature also obtained satisfactory risk-evaluation ability for the recurrence of CRS with an AUC of 0.742 in the external testing cohort, which was better than the AUC of just 0.496 obtained by traditional radiomics technology (see Figure S1). The reasonable interpretation of this result is that the features extracted through the traditional radiomics method require prudent engineering and considerable domain expertise to design feature extractors, and it is difficult to cover the comprehensive features of the images.[26] Different from traditional radiomics technology, deep learning is making major advances in solving these
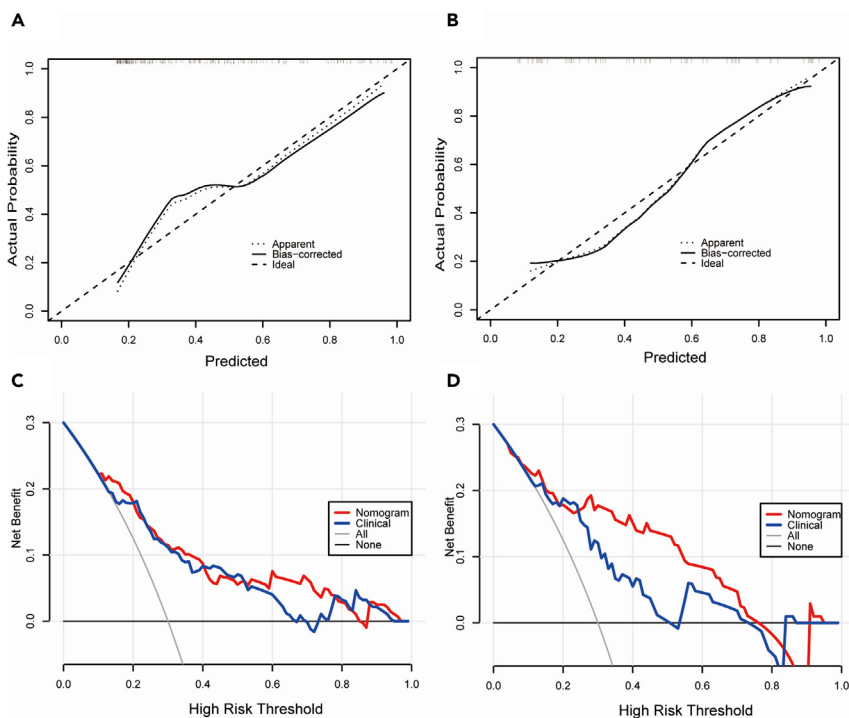
Figure 6. Calibration curves and DCA curves of the combined nomogram
(A) Calibration curves of nomogram in the training cohort and (B) the testing cohort.
(C) DCA curves of nomogram in the training cohort and (D) the testing cohort.

problems by learning from data using a general-purpose learning procedure and does well in discovering intricate signatures in high-dimensional data.[27] However, deep learning methods are often restrained by overfitting problems due to small datasets. Fortunately, multi-task learning can solve this problem to a certain extent. Furthermore, compared with single-task models, the correlation and heterogeneity between the segmentation and prognosis prediction can also be modeled by the multi-task learning, which were confirmed by our results.

Nomograms fusing clinical risk factors have been widely used to predict medical outcomes and prognosis. Currently, our study incorporated multi-scale prognostic features based on deep learning and clinical factors to construct a nomogram. In agreement with previous studies, AR, asthma, and higher levels of circulating eosinophils were independently related to the recurrence of CRS with the exception of asthma in the external cohort because of the small sample sizes. When DLR-score was included in the proposed clinic-radiomic nomogram, the AUC in the testing cohort was significantly increased from 0.766 to 0.842 (p = 0.029), which showed favorable performance and improved risk-evaluation accuracy. In addition, the DCA showed that more patients will benefit from the clinic-radiomic nomogram rather than the clinical model, suggesting that the DLR-score adds incremental value to the clinical usefulness of clinical predictors. Thus, the pleasant preoperative prediction power of the proposed noninvasive method made it a potential preoperative evaluation tool in clinical practice.

Table 2. The results of different models in the testing cohort

	AUC	Accuracy	Sensitivity	Specificity	FPR	FNR
DLR model	0.742 (0.619–0.865)	0.719	0.677	0.758	0.242	0.322
Traditional radiomics model	0.496 (0.360–0.648)	0.446	0.903	0.029	0.971	0.097
Clinical model	0.766 (0.642–0.881)	0.750	0.806	0.697	0.303	0.193
Nomogram	0.842 (0.740–0.943)	0.813	0.774	0.848	0.152	0.226

Abbreviations: FPR = False-Positive Rate; FNR = False-Negative Rate; DLR = deep learning radiomic.
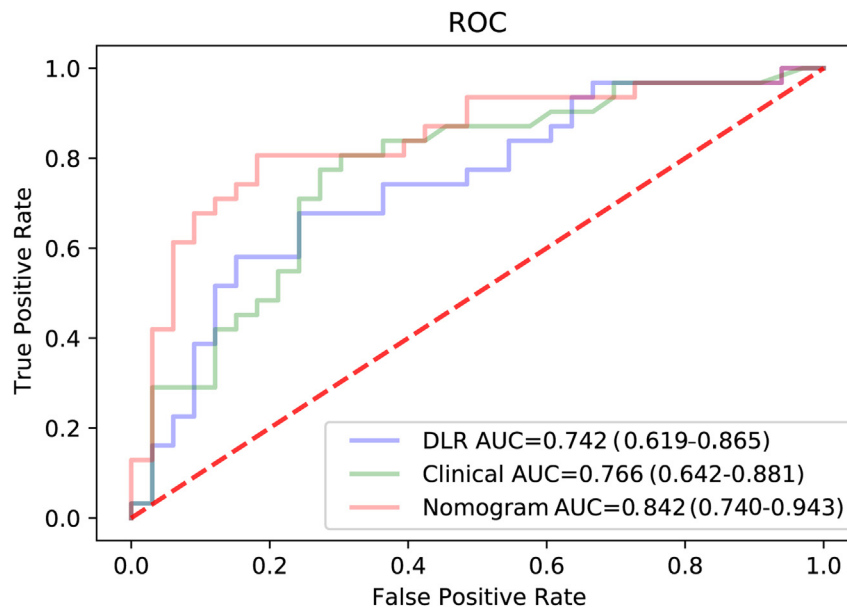
Figure 7. ROC curves of the deep learning radiomic (DLR) model, clinical model, and combination nomogram in the testing cohort

It is well known that asthma, AR, and higher levels of circulating eosinophils are the risk factors of CRS relapse in accordance with our findings.[28–31] For the precise mechanism, previous research demonstrated that type 2 immune response may be involved in the pathogenic process, for which upregulation of type 2 cytokines and immunoglobulin E (IgE)-mediated release of immune mediators have been observed.[32] Additionally, since the infiltration of eosinophils can remodel the nasal mucosa tissue[33] and promote the formation of CRS nasal polyps,[34] eosinophils are recognized as pivotal biomarkers for monitoring disease severity and postsurgical recurrence in patients with CRSwNP.[35–37] The combination of these independent clinical risk indicators and radiomics features leads us to the best estimation of the reaction to treatment. Consistent with this study, To-kunaga et al. also found that asthma, higher levels of circulating eosinophils, and signatures of CT images can be predictors for the recurrence of CRS.[38] While they demonstrated that shadows of sinuses in CT scans (ethmoid ≥ maxillary) were associated with refractoriness of the disease by using Cox proportional hazards models, our study further used deep learning to extract radiomic features, which were difficult to be quantified or recognized by the human eye, including size, shape, and complex patterns at the macroscopic scale.

In conclusion, we proposed a multi-task deep learning network for sinus segmentation and CRS recurrence prediction simultaneously. The predicted DLR-score was then combined with clinical factors to construct a nomogram, which showed excellent predictive ability for the recurrence of CRS. The result demonstrated that our model could noninvasively predict the prognosis of CRS patients who needed surgical treatment preoperatively and provide valuable information for individualized treatment and optimizing manage-ment, which also offers a promising practice for image segmentation and prognostic prediction in non-cancer diseases and broadens the application of DLR to some extent.

Limitations of the study

Despite satisfied sinus segmentation performance and promising recurrence prediction model, our study has several limitations. Despite the potential of deep learning-based methods in CT images, they still have limited capability in handling images heterogeneity of different CT scanners, which may generate under-lying bias during the analysis. Furthermore, the small and imbalanced testing cohort that only 14 of 65 pa-tients relapsed (21.538%) may limit the strength of our obtained conclusions to a certain extent. Addition-ally, for the generality of this nomogram, we recruited CRS patients including both CRSwNP and CRSsNP, but the prognoses of these two subtypes are different; further studies are needed to confirm the accuracy of the nomogram in the specific CRS subtype. Finally, the deep learning signature was abstract; even though it can perform automatic segmentation and discover intricate signatures which achieved promising prediction performance that traditional radiomics model cannot attain, it was difficult to provide direct

reference information for the prediction of prognostic outcomes. Larger amounts of data and multicenter tests are needed for further demonstrating and optimizing our proposed model to establish an AI chronic rhinosinusitis prognosis prediction platform, which will obtain fast and accurate prediction of CRS recurrence by AI when clinicians or patients import digital CT into this platform.

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
- METHOD DETAILS
  - Patients
  - CT images acquisition and ROI segmentation
  - Deep learning radiomic score construction and validation
  - Evaluation of images
  - Radiomic nomogram development
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106527.

AUTHOR CONTRIBUTIONS

Conceptualization: Xin Feng and Anning Li.; Resources: Shaojuan He, Wei Chen, Xuehai Wang, Xin Feng, Xinyu Xie, Fangying Liu, Xinyi Ma, and Xuezhong Li.; Methodology: Wei Chen and Shaojuan He.; Formal Analysis: Wei Chen and Shaojuan He.; Investigation: Wei Chen and Shaojuan He.; Data Curation: Shaojuan He, Wei Chen, Xin Feng, and Anning Li.; Validation: Xin Feng and Anning Li.; Project Administration: Xin Feng and Anning Li.; Supervision: Xin Feng and Anning Li.; Funding Acquisition: Xin Feng and Anning Li.; Writing-Original Draft: Shaojuan He and Wei Chen.; Writing- Review and Editing: Xin Feng and Anning Li.; Final approval of the manuscript: all authors.

DECLARATION OF INTERESTS

The authors declare no competing interest.

REFERENCES

1. Fokkens, W.J., Lund, V.J., Hopkins, C., Hellings, P.W., Kern, R., Reitsma, S., Toppila-Salmi, S., Bernal-Sprekelsen, M., Mullol, J., Alobid, I., et al. (2020). European position paper on rhinosinusitis and nasal polyps 2020. Rhinology 58, 1–464. https://doi.org/10.4193/Rhin20.600.

2. Orlandi, R.R., Kingdom, T.T., Hwang, P.H., Smith, T.L., Alt, J.A., Baroody, F.M., Batra, P.S., Bernal-Sprekelsen, M., Bhattacharyya, N., Chandra, R.K., et al. (2016). International consensus statement on allergy and rhinology: rhinosinusitis. Int. Forum Allergy Rhinol. 6 (Suppl 1), S22–S209. https://doi.org/10.1002/alr.21695.

3. Wu, A.W., Ting, J.Y., Platt, M.P., Tierney, H.T., and Metson, R. (2014). Factors affecting time to revision sinus surgery for nasal polyps: a 25-year experience. Laryngoscope 124, 29–33. https://doi.org/10.1002/lary.24213.

4. Hopkins, C., Slack, R., Lund, V., Brown, P., Copley, L., and Browne, J. (2009). Long-term outcomes from the English national comparative audit of surgery for nasal polyposis and chronic rhinosinusitis.

Laryngoscope *119*, 2459–2465. https://doi.org/10.1002/lary.20653.

5. Mendelsohn, D., Jeremic, G., Wright, E.D., and Rotenberg, B.W. (2011). Revision rates after endoscopic sinus surgery: a recurrence analysis. Ann. Otol. Rhinol. Laryngol. *120*, 162–166. https://doi.org/10.1177/000348941112000304.

6. Mayerhoefer, M.E., Materka, A., Langs, G., Häggström, I., Szczypiński, P., Gibbs, P., and Cook, G. (2020). Introduction to radiomics. J. Nucl. Med. *61*, 488–495. https://doi.org/10.2967/jnumed.118.222893.

7. Afshar, P., Mohammadi, A., Plataniotis, K.N., Oikonomou, A., and Benali, H. (2019). From handcrafted to deep-learning-based cancer radiomics: challenges and opportunities. IEEE Signal Process. Mag. *36*, 132–160.

8. Hesamian, M.H., Jia, W., He, X., and Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. J. Digit. Imaging *32*, 582–596. https://doi.org/10.1007/s10278-019-00227-x.

9. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., and Maier-Hein, K.H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods *18*, 203–211. https://doi.org/10.1038/s41592-020-01008-z.

10. Cheng, J., Liu, J., Kuang, H., and Wang, J. (2022). A fully automated multimodal MRI-based multi-task learning for glioma segmentation and IDH genotyping. IEEE Trans. Med. Imaging *41*, 1520–1532. https://doi.org/10.1109/TMI.2022.3142321.

11. Zhou, Y., Chen, H., Li, Y., Liu, Q., Xu, X., Wang, S., Yap, P.-T., and Shen, D. (2021). Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. Med. Image Anal. *70*, 101918. https://doi.org/10.1016/j.media.2020.101918.

12. Vliegenthart, R., Fouras, A., Jacobs, C., and Papanikolaou, N. (2022). Innovations in thoracic imaging: CT, radiomics, AI and x-ray velocimetry. Respirology *27*, 818–833. https://doi.org/10.1111/resp.14344.

13. Liu, B., Zeng, Q., Huang, J., Zhang, J., Zheng, Z., Liao, Y., Deng, K., Zhou, W., and Xu, Y. (2022). IVIM using convolutional neural networks predicts microvascular invasion in HCC. Eur. Radiol. *32*, 7185–7195. https://doi.org/10.1007/s00330-022-08927-9.

14. Chen, W., Liu, B., Peng, S., Sun, J., and Qiao, X. (2018). Computer-aided grading of gliomas combining automatic segmentation and radiomics. Int. J. Biomed. Imaging *2018*, 2512037. https://doi.org/10.1155/2018/2512037.

15. DeConde, A.S., Mace, J.C., Levy, J.M., Rudmik, L., Alt, J.A., and Smith, T.L. (2017). Prevalence of polyp recurrence after endoscopic sinus surgery for chronic rhinosinusitis with nasal polyposis. Laryngoscope *127*, 550–555. https://doi.org/10.1002/lary.26391.

16. Sella, G.C.P., Tamashiro, E., Sella, J.A., Aragon, D.C., Mendonça, T.N.D., Arruda, L.K.D.P., Anselmo Lima, W.T., and Valera, F.C.P. (2020). Asthma is the dominant factor for recurrence in chronic rhinosinusitis. J. Allergy Clin. Immunol. Pract. *8*, 302–309. https://doi.org/10.1016/j.jaip.2019.08.007.

17. Marcus, S., DelGaudio, J.M., Roland, L.T., and Wise, S.K. (2019). Chronic rhinosinusitis: does allergy play a role? Med. Sci. *7*, 30. https://doi.org/10.3390/medsci7020030.

18. Lou, H., Meng, Y., Piao, Y., Wang, C., Zhang, L., and Bachert, C. (2015). Predictive significance of tissue eosinophilia for nasal polyp recurrence in the Chinese population. Am. J. Rhinol. Allergy *29*, 350–356. https://doi.org/10.2500/ajra.2015.29.4231.

19. Meng, Y., Zhang, L., Lou, H., and Wang, C. (2019). Predictive value of computed tomography in the recurrence of chronic rhinosinusitis with nasal polyps. Int. Forum Allergy Rhinol. *9*, 1236–1243. https://doi.org/10.1002/alr.22355.

20. Wu, Q., Chen, J., Ren, Y., Qiu, H., Yuan, L., Deng, H., Zhang, Y., Zheng, R., Hong, H., Sun, Y., et al. (2021). Artificial intelligence for cellular phenotyping diagnosis of nasal polyps by whole-slide imaging. EBioMedicine *66*, 103336. https://doi.org/10.1016/j.ebiom.2021.103336.

21. Li, Y., and Tao, Y. (2022). Radiomics model based on enhanced gradient level set segmentation algorithm to predict the prognosis of endoscopic treatment of sinusitis. Comput. Math. Methods Med. *2022*, 9511631. https://doi.org/10.1155/2022/9511631.

22. Humphries, S.M., Centeno, J.P., Notary, A.M., Gerow, J., Cicchetti, G., Katial, R.K., Beswick, D.M., Ramakrishnan, V.R., Alam, R., and Lynch, D.A. (2020). Volumetric assessment of paranasal sinus opacification on computed tomography can be automated using a convolutional neural network. Int. Forum Allergy Rhinol. *10*, 1218–1225. https://doi.org/10.1002/alr.22588.

23. Massey, C.J., Ramos, L., Beswick, D.M., Ramakrishnan, V.R., and Humphries, S.M. (2022). Clinical validation and extension of an automated, deep learning–based algorithm for quantitative sinus CT analysis. AJNR. Am. J. Neuroradiol. *43*, 1318–1324. https://doi.org/10.3174/ajnr.A7616.

24. Chowdhury, N.I., Smith, T.L., Chandra, R.K., and Turner, J.H. (2019). Automated classification of osteomeatal complex inflammation on computed tomography using convolutional neural networks. Int. Forum Allergy Rhinol. *9*, 46–52. https://doi.org/10.1002/alr.22196.

25. Pallanch, J.F., Yu, L., Delone, D., Robb, R., Holmes, D.R., 3rd, Camp, J., Edwards, P., McCollough, C.H., Ponikau, J., Dearking, A.C., et al. (2013). Three-dimensional volumetric computed tomographic scoring as an objective outcome measure for chronic rhinosinusitis: clinical correlations and comparison to Lund-Mackay scoring. Int.

26. Lao, J., Chen, Y., Li, Z.C., Li, Q., Zhang, J., Liu, J., and Zhai, G. (2017). A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. Sci. Rep. *7*, 10353. https://doi.org/10.1038/s41598-017-10649-8.

27. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444. https://doi.org/10.1038/nature14539.

28. Batra, P.S., Tong, L., and Citardi, M.J. (2013). Analysis of comorbidities and objective parameters in refractory chronic rhinosinusitis. Laryngoscope *123* (*Suppl 7*), S1–S11. https://doi.org/10.1002/lary.24418.

29. Feng, X., Ramsden, M.K., Negri, J., Baker, M.G., Payne, S.C., Borish, L., and Steinke, J.W. (2016). Eosinophil production of prostaglandin D2 in patients with aspirin-exacerbated respiratory disease. J. Allergy Clin. Immunol. *138*, 1089–1097.e3. https://doi.org/10.1016/j.jaci.2016.04.042.

30. Jarvis, D., Newson, R., Lotvall, J., Hastan, D., Tomassen, P., Keil, T., Gjomarkaj, M., Forsberg, B., Gunnbjornsdottir, M., Minov, J., et al. (2012). Asthma in adults and its association with chronic rhinosinusitis: the GA2LEN survey in Europe. Allergy *67*, 91–98. https://doi.org/10.1111/j.1398-9995.2011.02709.x.

31. ten Brinke, A., Grootendorst, D.C., Schmidt, J.T., De Bruïne, F.T., van Buchem, M.A., Sterk, P.J., Rabe, K.F., and Bel, E.H. (2002). Chronic sinusitis in severe asthma is related to sputum eosinophilia. J. Allergy Clin. Immunol. *109*, 621–626. https://doi.org/10.1067/mai.2002.122458.

32. Min, Y.G., Lee, C.H., Rhee, C.S., Hong, S.K., and Kwon, S.H. (1999). Increased expression of IL-4, IL-5, IFN-gamma, IL-6, IL-8, and TGF-beta mRNAs in maxillary mucosa of patients with chronic sinusitis. Am. J. Rhinol. *13*, 339–343. https://doi.org/10.2500/105065899781367546.

33. Feng, X., Payne, S.C., Borish, L., and Steinke, J.W. (2019). Differential expression of extracellular matrix components in nasal polyp endotypes. Am. J. Rhinol. Allergy *33*, 665–670. https://doi.org/10.1177/1945892419860634.

34. Xiang, R., Zhang, Q.P., Zhang, W., Kong, Y.G., Tan, L., Chen, S.M., Deng, Y.Q., Tao, Z.Z., and Xu, Y. (2019). Different effects of allergic rhinitis on nasal mucosa remodeling in chronic rhinosinusitis with and without nasal polyps. Eur. Arch. Oto-Rhino-Laryngol. *276*, 115–130. https://doi.org/10.1007/s00405-018-5195-x.

35. Sharbel, D., Li, M., Unsal, A.A., Tadros, S.Y., Lee, J., Biddinger, P., Holmes, T., and Kountakis, S.E. (2020). Use of mucosal eosinophil count as a guide in the management of chronic rhinosinusitis. Int. Forum Allergy Rhinol. *10*, 474–480. https://doi.org/10.1002/alr.22517.

36. Pan, L., Liao, B., Guo, C.L., Liu, J.X., Wang, H., Long, X.B., and Liu, Z. (2021).

Inflammatory features and predictors for postsurgical outcomes in patients with nasal polyps stratified by local and systemic eosinophilia. Int. Forum Allergy Rhinol. *11*, 846–856. https://doi.org/10.1002/alr.22702.

37. Ye, P., He, S., Tang, S., Xie, X., Duan, C., Zhang, L., Steinke, J.W., Borish, L., Li, X., and Feng, X. (2022). Improvement of subjective olfactory dysfunction in chronic rhinosinusitis with nasal polyps after endoscopic sinus surgery. Front. Surg. *9*, 870682. https://doi.org/10.3389/fsurg.2022.870682.

38. Tokunaga, T., Sakashita, M., Haruna, T., Asaka, D., Takeno, S., Ikeda, H., Nakayama, T., Seki, N., Ito, S., Murata, J., et al. (2015). Novel scoring system and algorithm for classifying chronic rhinosinusitis: the JESREC Study. Allergy *70*, 995–1003. https://doi.org/10.1111/all.12644.

39. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., and Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, S. Ourselin, L. Joskowicz, M.R. Sabuncu, G. Unal, and W. Wells, eds. (Springer International Publishing), pp. 424–432. held in Cham.

40. Fokkens, W.J., Lund, V.J., Mullol, J., Bachert, C., Alobid, I., Baroody, F., Cohen, N., Cervin, A., Douglas, R., Gevaert, P., et al. (2012). European position paper on rhinosinusitis and nasal polyps 2012. Rhinology *23*, 1–298. 3 p preceding table of contents.

41. Bruns, N. (2019). 3D Slicer : universal 3D visualization software. Unfallchirurg *122*, 662–663. https://doi.org/10.1007/s00113-019-0654-4.

42. Flannery, S.W., Kiapour, A.M., Edgar, D.J., Murray, M.M., Beveridge, J.E., and Fleming, B.C. (2022). A transfer learning approach for automatic segmentation of the surgically treated anterior cruciate ligament. J. Orthop. Res. *40*, 277–284. https://doi.org/10.1002/jor.24984.

43. Balachandran, V.P., Gonen, M., Smith, J.J., and DeMatteo, R.P. (2015). Nomograms in oncology: more than meets the eye. Lancet Oncol. *16*, e173–e180. https://doi.org/10.1016/s1470-2045(14)71116-7.

44. Yang, L., Gu, D., Wei, J., Yang, C., Rao, S., Wang, W., Chen, C., Ding, Y., Tian, J., and Zeng, M. (2019). A radiomics nomogram for preoperative prediction of microvascular invasion in hepatocellular carcinoma. Liver Cancer *8*, 373–386. https://doi.org/10.1159/000494099.

45. Kramer, A.A., and Zimmerman, J.E. (2007). Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. Crit. Care Med. *35*, 2052–2056. https://doi.org/10.1097/01.CCM.0000275267.64078.B0.

46. Molodianovitch, K., Faraggi, D., and Reiser, B. (2006). Comparing the areas under two correlated ROC curves: parametric and non-parametric approaches. Biom. J. *48*, 745–757. https://doi.org/10.1002/bimj.200610223.

47. Vickers, A.J., and Elkin, E.B. (2006). Decision curve analysis: a novel method for evaluating prediction models. Med. Decis. Making *26*, 565–574. https://doi.org/10.1177/0272989X06295361.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw paranasal sinuses CT data	This paper	N/A
Code for model and network training and analysis	This paper	https://github.com/chenypic/deepSinuses
Software and algorithms		
Python (version 3.7.6)	Python Software Foundation	https://www.python.org/
R (version 4.1.0)	R software	http://www.R-project.org
3D U-Net	(Çiçek et al., 2016)[39]	https://doi.org/10.1007/978-3-319-46723-8_49

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Xin Feng (Email: drfengxin@sdu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Paranasal sinuses CT data reported in this paper will be shared by the lead contact upon request.

- All original code has been deposited at github and is publicly available as of the date of publication. DOI is listed in the key resources table.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Patients

This retrospective study was performed at two medical institutions Qilu Hospital of Shandong University (center A) and Weihai Municipal Hospital (center B), and approved by the Medical Ethics Committee of the two centers. We systematically reviewed medical records and paranasal sinuses computed tomography (CT) images of consecutive patients (n =265) who satisfied the diagnostic criteria for CRS established by the European Position Paper on Rhinosinusitis and Nasal Polyps (EPOS2012)[40] and required endoscopic sinus surgery (ESS) because of failed medical treatment. ESS was performed in compliance with uniform standards by four surgeons with more than 10-years' experience at two centers between January 2016 and December 2020 and the extent of surgery was defined relative to the extent of the disease by each sinus based on paranasal sinuses CT images.

The inclusion criteria: (a) patients diagnosed of CRS, (b) patients with medication alone fail to control the disease, (c) patients who needed surgical treatment.

The exclusion criteria: (a) patients who did not receive a CT scan of the paranasal sinuses within a week before surgery, (b) patients with odontogenic sinusitis, cystic fibrosis, fungal sinusitis, immunodeficiency diseases, or tumors, (c) patients without follow-up data, (d) patients without complete clinical data.

Baseline clinical features including gender, age, asthma, allergic rhinitis (AR), blood absolute eosinophil counts, smoking, drinking, hypertension, and diabetes were obtained from medical records of each center. The pre-operation CT images of the paranasal sinuses were used to extract the radiomic features. Patients

from center A were used as the training cohort (n = 200) and patients from medical center B were used as the testing cohort (n = 65). All postoperative patients were instructed to use daily nasal saline irrigation, antibiotics, topical corticosteroids (budesonide spray 400 μg once daily for 12 weeks), and endoscopic debridement periodically after surgery. Patients were consistently followed-up after ESS based on endoscopic examinations. We defined recurrence as endoscopic signs (nasal polyps, mucopurulent secretions, and / or inflamed mucosa), bothersome symptoms (nasal blockage, rhinorrhea / postnasal drip, smell impairment, facial pain / headache, and / or sleep disturbance / fatigue) that persisted for at least 1 month and the need for a systemic medical treatment (long-term antibiotics or systemic steroids) in that month, or both of endoscopic signs and symptoms.[40] Finally, according to a minimum of 1-year follow-up, patients were divided into a recurrence group (the symptoms and endoscopic signs relapsed) and a non-recurrence group (the symptoms were controlled and without endoscopic signs).

CT images acquisition and ROI segmentation

In center A, all preoperative axial CT scans were performed using a 64 detector row CT scanner (Discovery CT750 HD, GE), a 256 detector row CT scanner (Brilliance iCT, Philips) or a dual-source spiral CT scanner (SOMATOM Force, Siemens). The scanning parameters were routinely set as follows: tube voltage, 80–140 kVp; tube current, automatic tube current modulation (maximum, 450 mAs); pitch factor, 1.0; slice thickness, 1 mm; and slice interval, 1 mm. The patients from center B underwent preoperative axial CT scan using a dual-source spiral CT scanner (SOMATOM Force, Siemens), a 256 detector row CT scanner (Brilliance, Philips), a 256 detector row CT scanner (Revolution CT, GE) or a 64 detector row CT scanner (Revolution Frontier, GE). The scanning parameters were the same as center A. All the CT images were reconstructed with a standard kernel at a slice thickness of 1 mm and an increment of 1 mm.

The images were imported into the 3D-Slicer software (version 4.11.2; https://www.slicer.org) in DICOM format from the PACS system and the horizontal axial plain was selected to perform the nose and paranasal sinuses region manual segmentation in each CT slice by two radiologists (S. H. and N. L., with 5 and 9 years CT experience) who were blinded to the clinical outcome.[41] The edge of ROI was 1-2 mm away from the edge of the lesion to avoid air, bone, and other regions.

Deep learning radiomic score construction and validation

Due to the complexity of the organizational structure of the paranasal sinus region, we proposed a multitask deep learning structure that can simultaneously perform the segmentation and recurrence prediction of the sinus region in an end-to-end model to fully exploit the shared global information between these two tasks. Taking the manual segmentation results of radiologists as the gold standard, we first utilized a 3D U-Net as the backbone network for multi-task learning because of its excellent performance in 3D medical image segmentation. The 3D U-Net architecture consisted of three parts: (i) encoding path, (ii) decoding path, and (iii) skip connections between them. In the bottleneck layer of 3D U-Net, the hidden layer features were extracted through a multilayer perceptron, and the classification probability was generated through softmax function. Finally, we got the deep learning radiomic score (DLR-score). During the training process, a multi-task learning loss function was designed to balance the two tasks by combining segmentation and classification loss functions with uncertain weights. Further details can be found at https://github.com/chenyipc/deepSinuses. In order to further investigate the performance of our proposed multi-task model, we designed two single-task networks for sinus segmentation and recurrence prediction in the testing cohort respectively. Specifically, we dropped the classification module for sinus segmentation, so the model was similar to 3D U-net. For recurrence prediction, we dropped the decoder module.

Evaluation of images

As previously noted, the Dice similarity coefficient (DSC) was used to measure the degree of similarity between the automatic and manual (ground truth) segmentations to evaluate the segmentation performance.

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

where A and B denote the automatic segmentation and the ground truth. A score (x) of $0 \leq x < 0.7$ would be considered "poor" performance, $0.7 \leq x < 0.8$ "fair" performance, $0.8 \leq x < 0.9$ "good" performance, and $x \geq 0.9$ "excellent" performance.[42]

Radiomic nomogram development

Univariate analysis was first used to identify statistically clinical characteristics ($P < 0.05$). Then multivariable analysis was applied to combine the clinical features and the DLR-score with multivariable logistic regression model. Finally, a combination nomogram was generated based on the proposed multivariate model. We assessed and compared the performance of the models in terms of discrimination ability, calibration ability, and clinical usefulness. The area under the curves (AUCs) of receiver operating characteristic (ROC) were used to quantify the discriminative efficacy for recurrence prediction. Multiple comparisons of the curves were also performed. Calibration curves were plotted to analyze the calibration performance and displayed to show the agreement between the predicted and observed risks of recurrence of the nomogram.[43] The decision curve analysis (DCA) was conducted to estimate the clinical utility of the models by quantifying the net benefits at different threshold probabilities.[44]

QUANTIFICATION AND STATISTICAL ANALYSIS

All clinical characteristics were statistically analyzed using the Mann–Whitney U test or chi-square test under R software (version 4.1.0; http://www.R-project.org) or Python (version 3.7.6; https://www.python.org/). A P-value of <0.05 was considered statistically significant. The univariate and multivariate logistic regression analyses were conducted using the "glm" function. The "rms" package was used for nomogram construction and calibration plotting.[45] The ROC curves plotting and area under curve calculation were performed using the "pROC" package. ROC curves were compared using the Delong test.[46] The DCA was conducted with the "rmda" package.[47] The DSC was used to evaluate the performance of segmentation.

ADDITIONAL RESOURCES

This study was conducted with the approval of the Medical Ethics Committee of Qilu Hospital of Shandong University (reference number: KYLL-202206-006) and Weihai Municipal Hospital (reference number: 2022037).