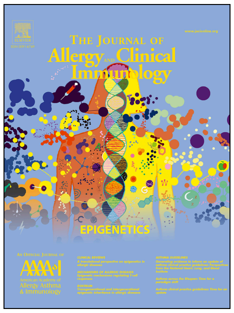


Expert-level Diagnosis of Nasal Polyps Using Deep Learning on Whole-slide Imaging

Qingwu Wu, MD, Jianning Chen, MD, Huiyi Deng, MD, Yong Ren, MSc, Yueqi Sun, MD, PhD, Weihao Wang, MD, Lianxiong Yuan, MD, Haiyu Hong, MD, PhD, Rui Zheng, MD, Weifeng Kong, MD, Xuekun Huang, MD, PhD, Guifang Huang, BEng, Lunji Wang, BEng, Yana Zhang, MD, PhD, Lanqing Han, MEng, Qintai Yang, MD, PhD

PII: S0091-6749(19)31631-8

DOI: https://doi.org/10.1016/j.jaci.2019.12.002

Reference: YMAI 14296

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 21 August 2019

Revised Date: 20 November 2019

Accepted Date: 2 December 2019

Please cite this article as: Wu Q, Chen J, Deng H, Ren Y, Sun Y, Wang W, Yuan L, Hong H, Zheng R, Kong W, Huang X, Huang G, Wang L, Zhang Y, Han L, Yang Q, Expert-level Diagnosis of Nasal Polyps Using Deep Learning on Whole-slide Imaging, *Journal of Allergy and Clinical Immunology* (2020), doi: https://doi.org/10.1016/j.jaci.2019.12.002.

Expert-level Diagnosis of Nasal Polyps Using Deep Learning on Whole-slide Imaging

Qingwu Wu, MD[1, *], Jianning Chen, MD[2, *], Huiyi Deng, MD[1, *], Yong Ren, MSc[3, *], Yueqi Sun, MD, PhD[4], Weihao Wang, MD[1], Lianxiong Yuan, MD[5], Haiyu Hong, MD, PhD[6], Rui Zheng, MD[1], Weifeng Kong, MD[1], Xuekun Huang, MD, PhD[1], Guifang Huang, BEng[3], Lunji Wang, BEng[3], Yana Zhang, MD, PhD[7], Lanqing Han, MEng[3, #], Qintai Yang, MD, PhD[1, #]

[1] Department of Otorhinolaryngology-Head and Neck Surgery, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510630, China

[2] Department of Pathology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510630, China

[3] Artificial Intelligence Innovation Center, Research Institute of Tsinghua, Pearl River Delta, Guangzhou 510735, China

[4] Otorhinolaryngology Hospital, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou 510080, China

[5] Department of Science and Research, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510630, China

[6] Department of Otolaryngology-Head and Neck Surgery, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai 519020, China

[7] Feinberg School of Medicine, Northwestern University, Chicago 60611, United

States

* The authors contributed equally to this work.

# Corresponding authors

1. Lanqing Han, MEng

Artificial Intelligence Innovation Center
Research Institute of Tsinghua, Pearl River Delta
No. 11 Kaiyuan Road, Guangzhou 510735, China
Phone: 020-22213618
E-mail: hanlance@tsinghua-gd.org

2. Qintai Yang, MD, PhD

Department of Otolaryngology-Head and Neck Surgery
The Third Affiliated Hospital of Sun Yat-sen University
No. 600 Tianhe Road, Guangzhou 510630, China
Phone: 020-85252239
E-mail: yang.qt@163.com

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 81670912 and 81870704), the Industry-Academic Cooperation Foundation of

Guangzhou (No. 201704030046), Sun Yat-sen University Clinical Research 5010 Program (NO.2019006) and The Third Affiliated Hospital of Sun Yat-Sen University, Clinical Research Program (No. QHJH201901).

Conflicts of interest

The authors declare that they have no conflicts of interest. This work described was original research that has not been published previously, and not under consideration for publication elsewhere.

Capsule summary

AICEP is the first use of deep learning in combination with WSI in nasal polyp diagnosis and treatment. It can improve the diagnosis and management of nasal polyps more quickly and accurately.

Key words

CRSwNP; deep learning; pathological classification; eosinophils; WSI

Acknowledgements

The authors would like to thank Chunkui Shao (Professor, Department of Pathology, The Third Affiliated Hospital of Sun Yat-sen University) and his colleagues for the help.

Ethical approval

This study was approved by the Research Ethics Committee of the Institute of Basic Research in Clinical Medicine, Third Affiliated Hospital of Sun Yat-sen University

([2019]02-157-01). The research was registered at Chinese Clinical Trails Registry (http://www.chictr.org.cn/index.aspx) with the number ChiCTR1900021601.

To the Editor:

Chronic rhinosinusitis (CRS) is defined as a chronic inflammation of the nose and paranasal sinuses. It is estimated that CRS affects more than 100 million patients worldwide and it involves high management costs and poor quality of life (QOL) in affected subjects[1]. The presence of eosinophils in nasal polyps is linked to higher postoperative visual analogue pain scores (VAS), impaired QOL, and high recurrence rate[2]. A better understanding of the ratio of eosinophils (RE) to infiltrating inflammatory cells in tissue is needed to improve diagnostic and treatment strategies for affected patients[3].

Thus far, there are no uniform standards or rules regarding diagnosis of eosinophilic chronic rhinosinusitis with nasal polyps (eCRSwNP), and a variety of problems exist in practice. Some researchers recommend that the amounts of eosinophils per high power field (HPF) should be more than 15 or 100[4, 2]. Most researchers support the assessment of RE in several random HPFs, with eCRSwNP diagnosed when RE is >10%[5, 6]. The traditional method ($\mathrm{RE}_{\text{slide-tm}}$) dictates that the pathologist assesses the ratio of eosinophils to infiltrating inflammatory cells (which include eosinophils, neutrophils, lymphocytes, plasma cells, etc.) in 10 random HPFs for the tissue[6].

However, RE obviously differs between various HPFs. Preliminary studies have shown sampling errors among the estimates based on 10 random HPFs and in the overall eosinophil counts in the total sample. Therefore, we considered the RE of whole-slide imaging (WSI) as the gold standard ($\mathrm{RE}_{\text{slide-actual}}$) for assessing eCRSwNP for its lack of sampling error. However, it is difficult in practice because it is both time-consuming and subjective.

Artificial intelligence (AI), especially deep learning algorithms, has made great progress and is similar to or even better than humans in terms of visual perception and speech recognition. Therefore, we aimed to establish an artificial intelligence evaluation platform (AICEP, $RE_{slide\text{-}predict}$) to diagnose eCRSwNP rapidly and accurately via deep learning and WSI.

A total of 195 nasal polyp specimens were obtained from three affiliated hospitals of Sun Yat-sen University (The Third Hospital=179, The First Hospital=9, The Fifth Hospital=7). After WSI, we automatically extracted 26589 patches in the lamina propria of mucosa and marked the RE in each patch ($RE_{patch\text{-}actual}$, see the Methods section in this article’s Online Repository at www.jacionline.org). The patches were classified as the training dataset, the internal validation dataset and independent external test dataset (Fig. E1).

In this study, our AICEP compared three common architectures (Resnet50, Xception, and Inception V3) for application of a transfer learning algorithm to assess their performance in the classification and regression of patches extracted from WSIs (Fig. 1). Within 100 epochs (iterations through the entire training dataset), the retrained weights were saved due to the absence of further improvement in the mean absolute error (MAE) (Fig. E2, A) and mean square error loss (Fig. E2, B).

First, we completed the qualitative classification of both internal validation and external test datasets using Resnet50, Xception, and InceptionV3 models. WSI results were classified as eosinophilic when RE_{slide} exceeded 10% (see the Methods section in this article's Online Repository at www.jacionline.org). The respective sensitivities for the internal and external datasets were 97.0% and 93.5% for Resnet50, 90.1% and 84.2% for Xception, and 93.9% and 90.3% for InceptionV3 model, respectively. The corresponding specificities were 86.0% and 84.6%, 88.2% and 88.4%, and 88.2% and 86.4%, individually. Our study showed that internal authentication was far superior to external authentication (Fig. E3). The AUCs from internal validation and external test datasets of Inception V3 were 0.974 and 0.957, respectively, which indicated that this was the best model (Fig. 2, A and B).

Second, the convolutional neural network was visualized to identify the region of eosinophils, which confirmed that the model was able to learn from the characteristics of eosinophils only (Fig. 2, C and D).

In addition, for the quantitative analysis of AICEP, we found that the MAEs of $RE_{patch-actual}$ and $RE_{patch-predict}$ in both internal validation dataset and external test dataset were 4.3% and 5.8%. Meanwhile, both the consistency of intraclass correlation coefficient and the agreement of $RE_{patch-predict}$ and $RE_{patch-actual}$ in the internal validation dataset and external test dataset were greater than 0.95, indicating high consistency from AICEP analysis (Table E1).

When compared with $RE_{slide-predict}$ from AICEP, pathologist simulation and $RE_{slide-actual}$ from the internal validation dataset of 12 patients, AICEP can diagnose all the 12 patients correctly, while the traditional method only made 10 correct diagnoses, unfortunately, with two misdiagnosed patients (NO. 4 and 5; Fig. 2, E). Similarly, when we compared $RE_{slide\text{-}predict}$ from AICEP with pathologist simulation and $RE_{slide\text{-}actual}$ from the external test dataset of 16 patients, AICEP correctly diagnosed all 16 patients, while the traditional method may misdiagnosed 4 patients (NO. 6, 7, 8, and 10; Fig. 2, F).

Finally, we compared the diagnostic time between AICEP and pathologist judgement. The result showed that AICEP (5.4 ± 0.87 min) took less time than $RE_{slide\text{-}tm}$ (12.7 ± 2.78 min) and $RE_{slide\text{-}actual}$ (148.6 ± 34.36 min, $P < 0.0001$, Table E2).

In our study, we advocated WSI assessment instead of $RE_{slide\text{-}tm}$. While WSI is undoubtedly more accurate, it costs an immense amount of time. What's worse, in China, the medical resources in the Midwest are significantly worse than those in the eastern coastal areas, and pathologists are inadequate, especially in some primary hospitals. To some extent, AICEP can well solve this problem, as it can diagnose nasal polyp pathological types by WSI and AI more efficiently.

AI-facilitated diagnosis can alleviate doctors' workload and contribute to high-quality medical care provision to patients in need[7, 8]. It is well known that the diagnosis of disease depends on the intuition and experience of pathologists. Moreover, large workload can lead to pathologists' working inefficiency and increasing the chance of making mistakes. Our results showed that $RE_{slide\text{-}tm}$ may result in wrong diagnosis, especially when the proportion of tissue eosinophils was approximately 10%.

However, this problem can be resolved by our AICEP, which can diagnose all patients accurately.

Although AI has already shown great potentials for assisting doctors in diagnosis and decision making, there are still some limitations. For instance, the real-world diagnostic accuracy of AI was lower than that reported in their previous study conducted with screening datasets[9]. Our study showed the similar result that AICEP performed better in the internal validation dataset than in the external test dataset. In our study, the internal validation dataset and training dataset came from a similar process regarding slicing, staining, and WSI scanning, whereas these aspects may differ in the external test dataset. Thus, it is important to optimize AICEP with data from multiple centers.

Overall, AICEP is the first use of deep learning in combination with WSI in nasal polyp diagnosis. It can evaluate the pathological characterizations of nasal polyps in a faster and more accurate way. We believe that AICEP will be used widely in particular in primary hospitals, even all around the world through the cloud platform.

Qingwu Wu, MD[1, *], Jianning Chen, MD[2, *], Huiyi Deng, MD[1, *], Yong Ren, MSc[3, *], Yueqi Sun, MD, PhD[4], Weihao Wang, MD[1], Lianxiong Yuan, MD[5], Haiyu Hong, MD, PhD[6], Rui Zheng, MD[1], Weifeng Kong, MD[1], Xuekun Huang, MD, PhD[1], Guifang Huang, BEng[3], Lunji Wang, BEng[3], Yana Zhang, MD, PhD[7], Lanqing Han, MEng[3, #], Qintai Yang, MD, PhD[1, #]

[1] Department of Otorhinolaryngology-Head and Neck Surgery, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510630, China

[2] Department of Pathology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510630, China

[3] Artificial Intelligence Innovation Center, Research Institute of Tsinghua, Pearl River Delta, Guangzhou 510735, China

[4] Otorhinolaryngology Hospital, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou 510080, China

[5] Department of Science and Research, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510630, China

[6] Department of Otolaryngology-Head and Neck Surgery, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai 519020, China

[7] Feinberg School of Medicine, Northwestern University, Chicago 60611, United States

* The authors contributed equally to this work.

# Corresponding authors: Lanqing Han and Qintai Yang.

E-mail: hanlance@tsinghua-gd.org and yang.qt@163.com.

References

1. Fokkens WJ, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, et al. European Position Paper on Rhinosinusitis and Nasal Polyps 2012. Rhinology Supplement 2012; 23: 3 p preceding table of contents, 1-298.

2. Ikeda K, Shiozawa A, Ono N, Kusunoki T, Hirotsu M, Homma H, et al. Subclassification of chronic rhinosinusitis with nasal polyp based on eosinophil and neutrophil. The Laryngoscope 2013; 123: E1-9.

3. Snidvongs K, Lam M, Sacks R, Earls P, Kalish L, Phillips PS, et al. Structured histopathology profiling of chronic rhinosinusitis in routine practice. International forum of allergy & rhinology 2012; 2: 376-385.

4. Wen W, Liu W, Zhang L, Bai J, Fan Y, Xia W, et al. Increased neutrophilia in nasal polyps reduces the response to oral corticosteroid therapy. The Journal of allergy and clinical immunology 2012; 129: 1522-1528.e1525.

5. Mahdavinia M, Suh LA, Carter RG, Stevens WW, Norton JE, Kato A, et al. Increased noneosinophilic nasal polyps in chronic rhinosinusitis in US second-generation Asians suggest genetic regulation of eosinophilia. The Journal of allergy and clinical immunology 2015; 135: 576-579.

6. Cao PP, Li HB, Wang BF, Wang SB, You XJ, Cui YH, et al. Distinct immunopathologic characteristics of various types of chronic rhinosinusitis in adult Chinese. The Journal of allergy and clinical immunology 2009; 124: 478-484, 484.e471-472.

7. Luo H, Xu G, Li C, He L, Luo L, Wang Z, et al. Real-time artificial intelligence

for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. The Lancet Oncology 2019.

8. Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, et al. Diagnostic Efficacy and Therapeutic Decision-making Capacity of an Artificial Intelligence Platform for Childhood Cataracts in Eye Clinics: A Multicentre Randomized Controlled Trial. EClinicalMedicine 2019; 9: 52-59.

9. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nature Biomedical Engineering 2017; 1.

Figure Legend

Figure 1. Schematic of processes. A, nasal endoscopic examination. B, samples of nasal polyps obtained by functional endoscopic sinus surgery (FESS). C, made HE slides. D, digitized the slides into the whole slide images (WSI) by scanner. E, delineated the lamina propria to obtain region of interest (ROI). F, patches extracted from ROI of WSI and corresponding RE tags. G, trained transfer learning models that can be deployed to diagnose eCRSwNP. H, $RE_{slide\text{-}predict}$ of patients according to the model. I, chose the appropriate treatment strategy.

Figure 2. Performance of AICEP. A and B, the receiver operating characteristic curves (ROC) and the area under ROC (AUC) for detection of patches with RE⩾10% from patches with RE<10%. A, comparison of AUC/ROC for Resnet50, Xception and Inception V3 models using internal test dataset. The Inception V3 model had an AUC (0.974) significantly greater than the other two models. B, comparison of AUC/ROC for Resnet50, Xception and Inception V3 models in independent external test dataset. The Inception V3 model also provided the best AUC (0.957) compared to the other ones. C and D, visualization and explainability of CNN models using Grad-CAM to classify patches with RE_{patch}⩾10% from patches with RE_{patch}<10%. C, RE_{patch}=86.66%, eosinophils were marked by red arrows. D, corresponding Grad-CAM images, the highlighted areas were discriminative features of eosinophils. E and F, diagnostic efficiency comparison of AICEP and current method. Black dot represented current method result, and 50 times bootstrap were performed to evaluate its random error, blue line was the actual value of WSI and yellow line was the

AICEP predicted value of WSI, red dashed line was the diagnostic boundary of 10%. E, internal validation dataset: all patients were accurately diagnosed by AICEP while current method may make wrong diagnosis in patient NO. 4 and 5. F, external test dataset: all patients were accurately diagnosed by AICEP while current method may make wrong diagnosis in patient NO. 6, 7, 8 and 10.

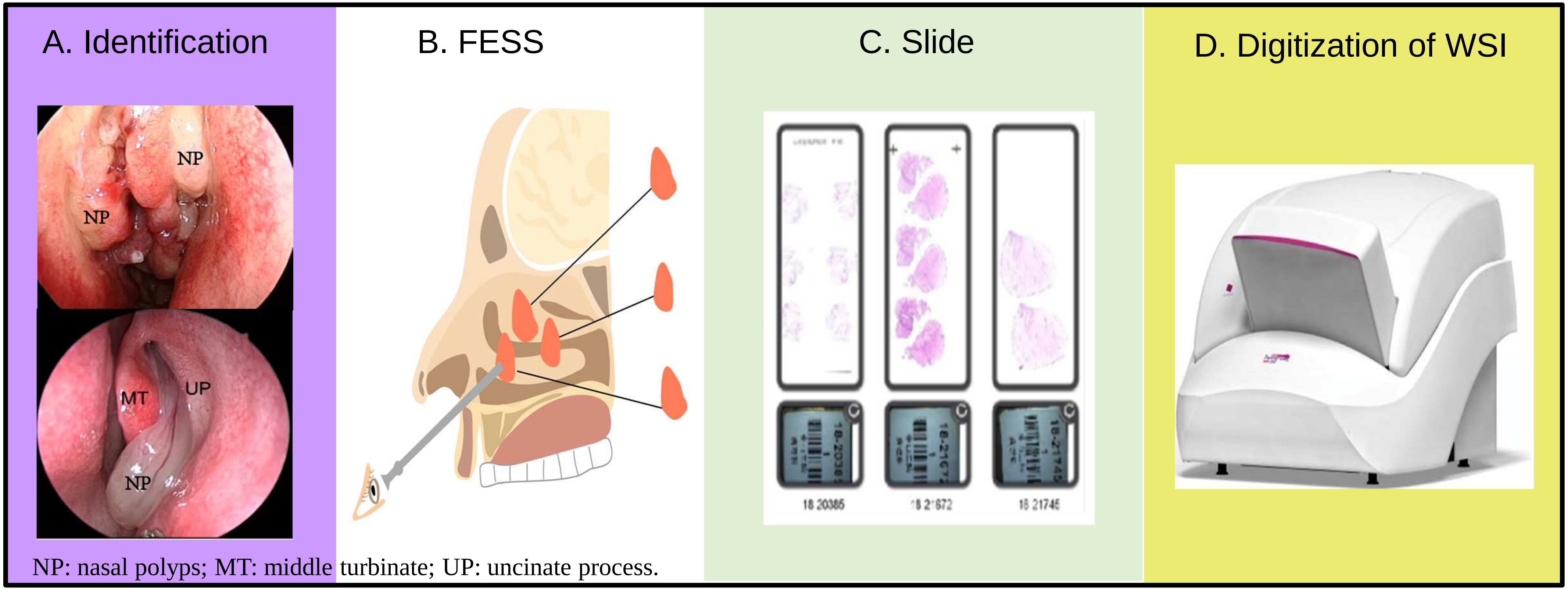
Table E1. Consistency assessment for AICEP in internal validation dataset and external test dataset according to the $RE_{patch-actual}$ and $RE_{slide-actual.}$

Level	Internal Validation Dataset		External Test Dataset	
	ICC Consistency	ICC Agreement	ICC Consistency	ICC Agreement
RE_{patch}	0.981 （0.979,0.983）	0.981 （0.979,0.982）	0.977 （0.975,0.979）	0.976 （0.970,0.980）
RE_{slide}	0.999 （0.997,1.000）	0.999 （0.998,1.000）	0.995 （0.985,0.998）	0.993 （0.973,0.998）

RE, the ratio of eosinophils; ICC, intraclass correlation coefficient.

Table E2. Comparison of time-consuming between AICEP and pathologists.

Method	Mean time ± SD (min)	95% CI
$RE_{slide\text{-}predict}$	5.4±0.87	[5.28, 5.52]
$RE_{slide\text{-}tm}$	12.7±2.78	[12.31, 13.09]
$RE_{slide\text{-}actual}$	148.6±34.36	[143.78, 153.42]



A. Identification

B. FESS

C. Slide

D. Digitization of WSI

NP: nasal polyps; MT: middle turbinate; UP: uncinate process.

Figure 1,A-D

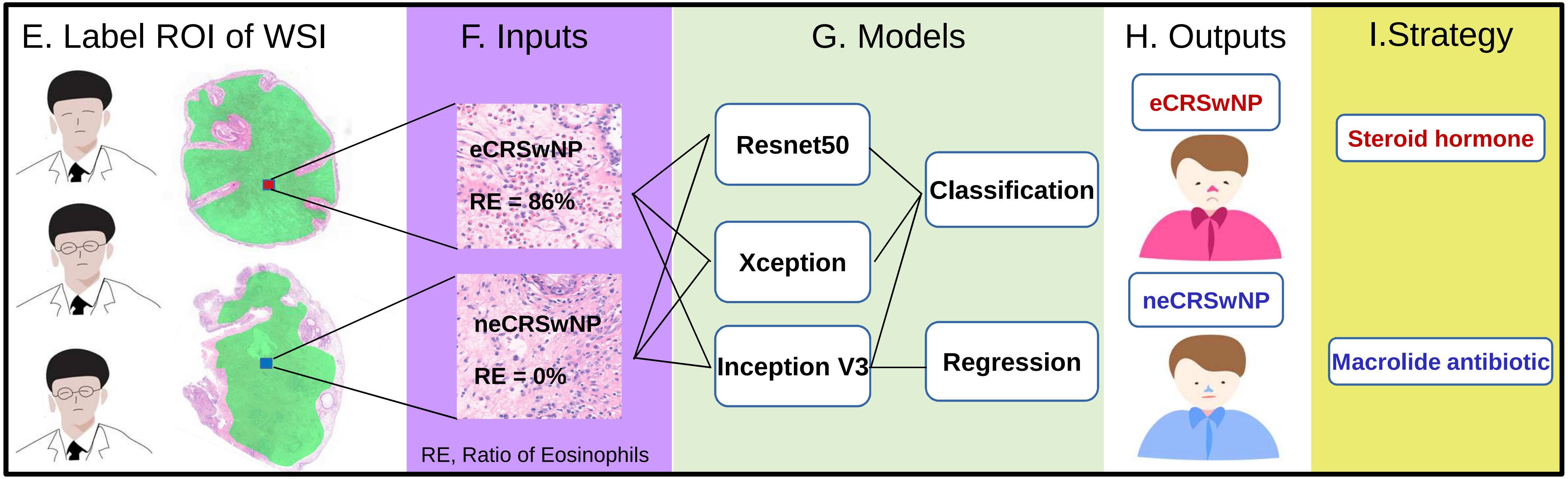

E. Label ROI of WSI
F. Inputs
eCRSwNP
RE = 86%
neCRSwNP
RE = 0%
RE, Ratio of Eosinophils
G. Models
Resnet50
Xception
Inception V3
Classification
Regression
H. Outputs
eCRSwNP
neCRSwNP
I.Strategy
Steroid hormone
Macrolide antibiotic

Figure 1,E-I

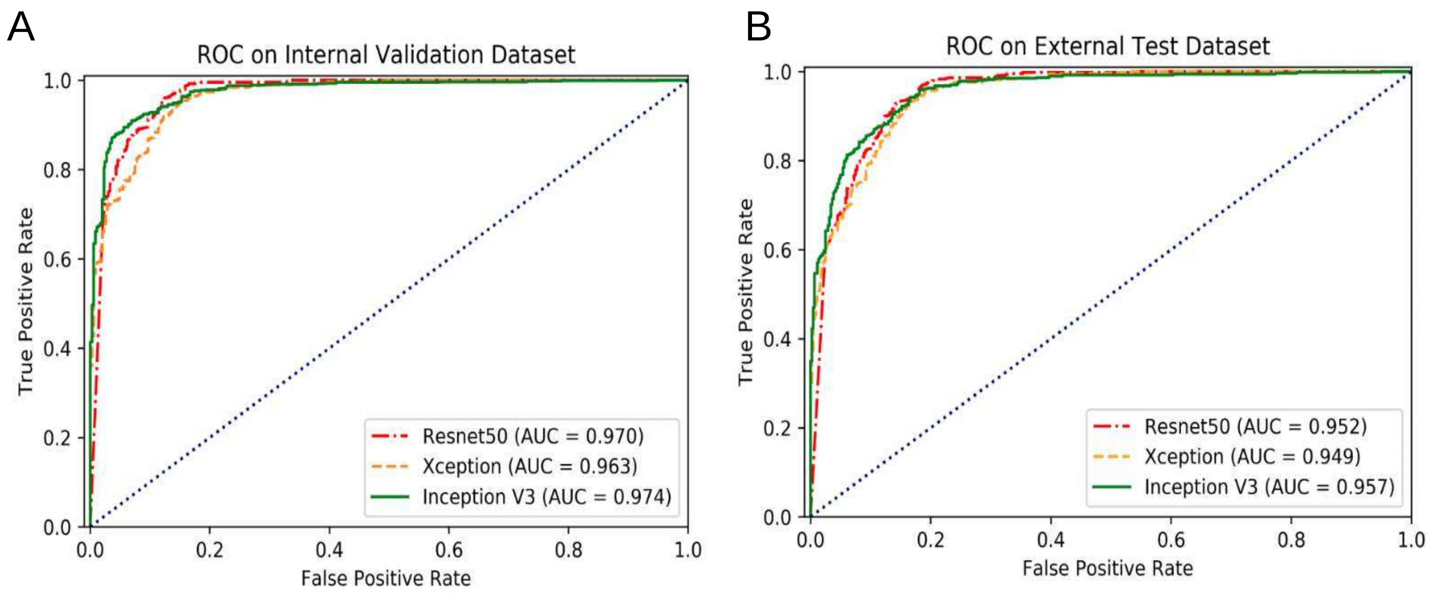

A
ROC on Internal Validation Dataset
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Resnet50 (AUC = 0.970)
Xception (AUC = 0.963)
Inception V3 (AUC = 0.974)
B
ROC on External Test Dataset
True Positive Rate
False Positive Rate
Resnet50 (AUC = 0.952)
Xception (AUC = 0.949)
Inception V3 (AUC = 0.957)

Figure 2, A-B

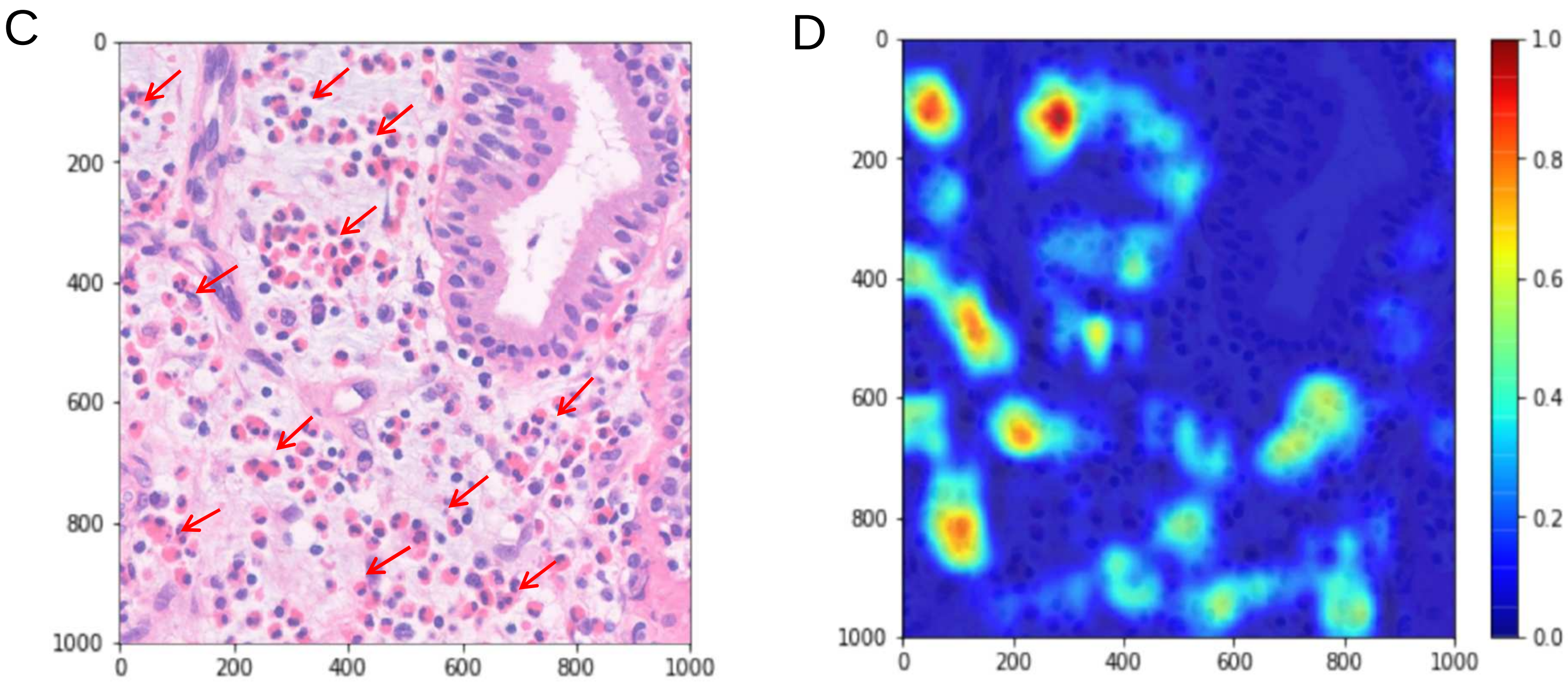

C
0
200
400
600
800
1000
0
200
400
600
800
1000
D
0
200
400
600
800
1000
0
200
400
600
800
1000
1.0
0.8
0.6
0.4
0.2
0.0

Figure 2, C-D

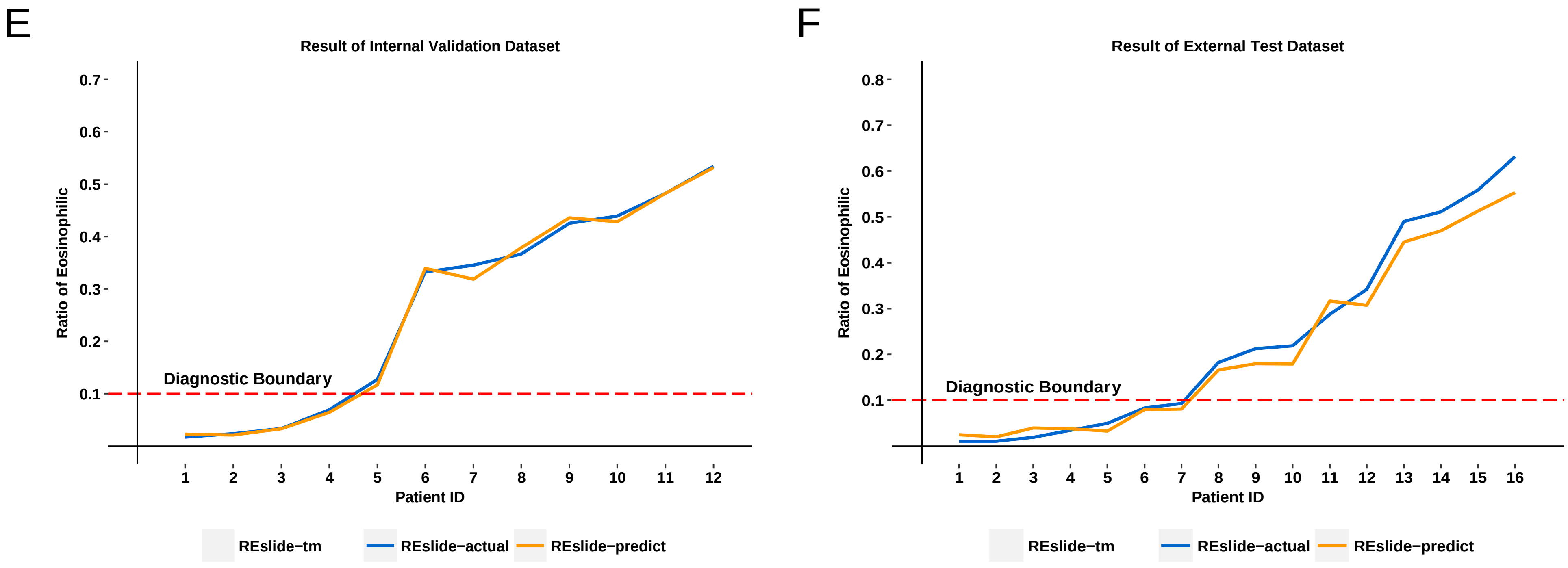

E
Result of Internal Validation Dataset
Ratio of Eosinophilic
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Diagnostic Boundary
1
2
3
4
5
6
7
8
9
10
11
12
Patient ID
REslide−tm
REslide−actual
REslide−predict
F
Result of External Test Dataset
Ratio of Eosinophilic
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Diagnostic Boundary
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Patient ID
REslide−tm
REslide−actual
REslide−predict

Figure 2, E-F

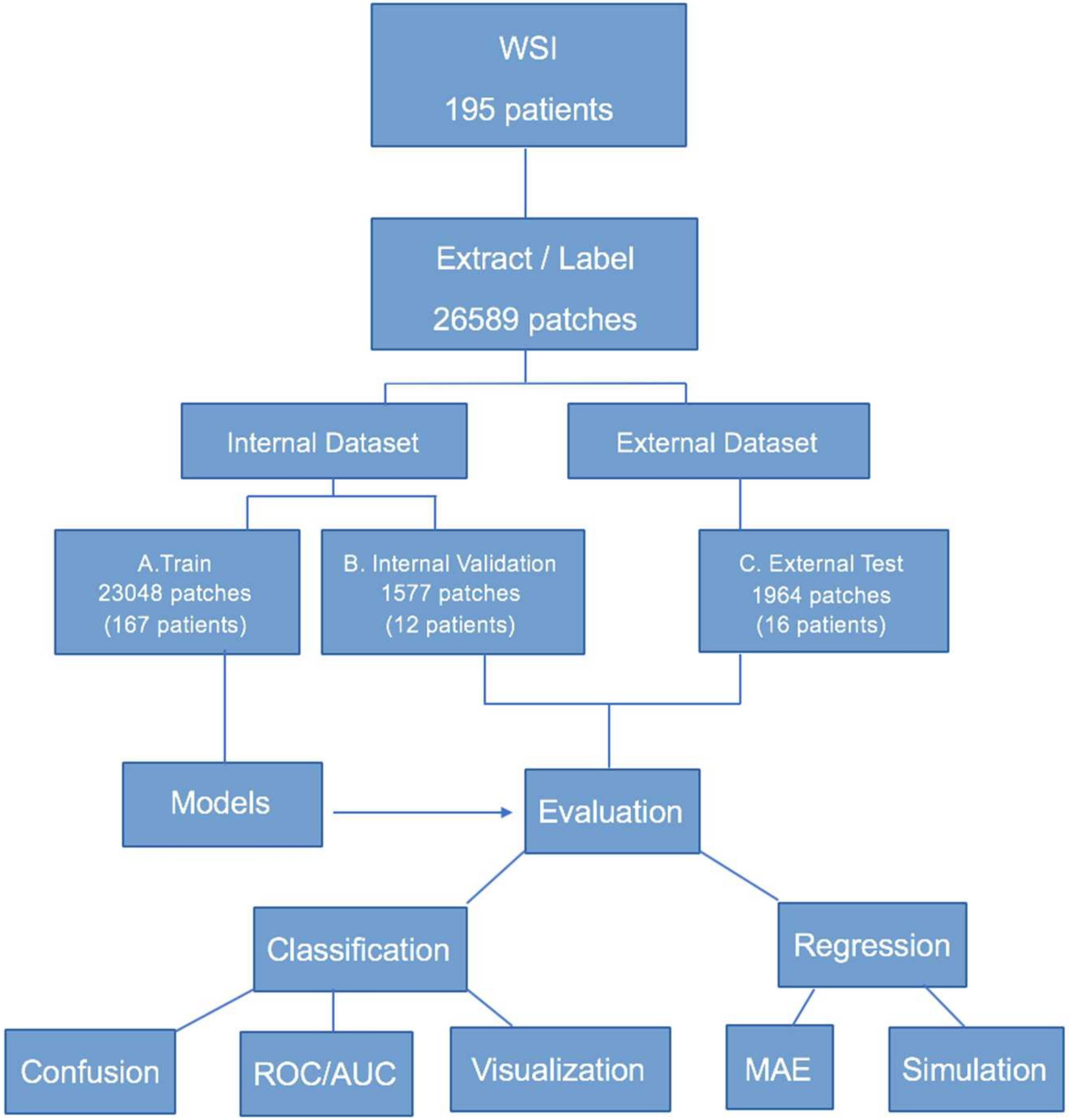

WSI
195 patients
Extract / Label
26589 patches
Internal Dataset
External Dataset
A.Train
23048 patches
(167 patients)
B. Internal Validation
1577 patches
(12 patients)
C. External Test
1964 patches
(16 patients)
Models
Evaluation
Classification
Regression
Confusion
ROC/AUC
Visualization
MAE
Simulation

Figure E1

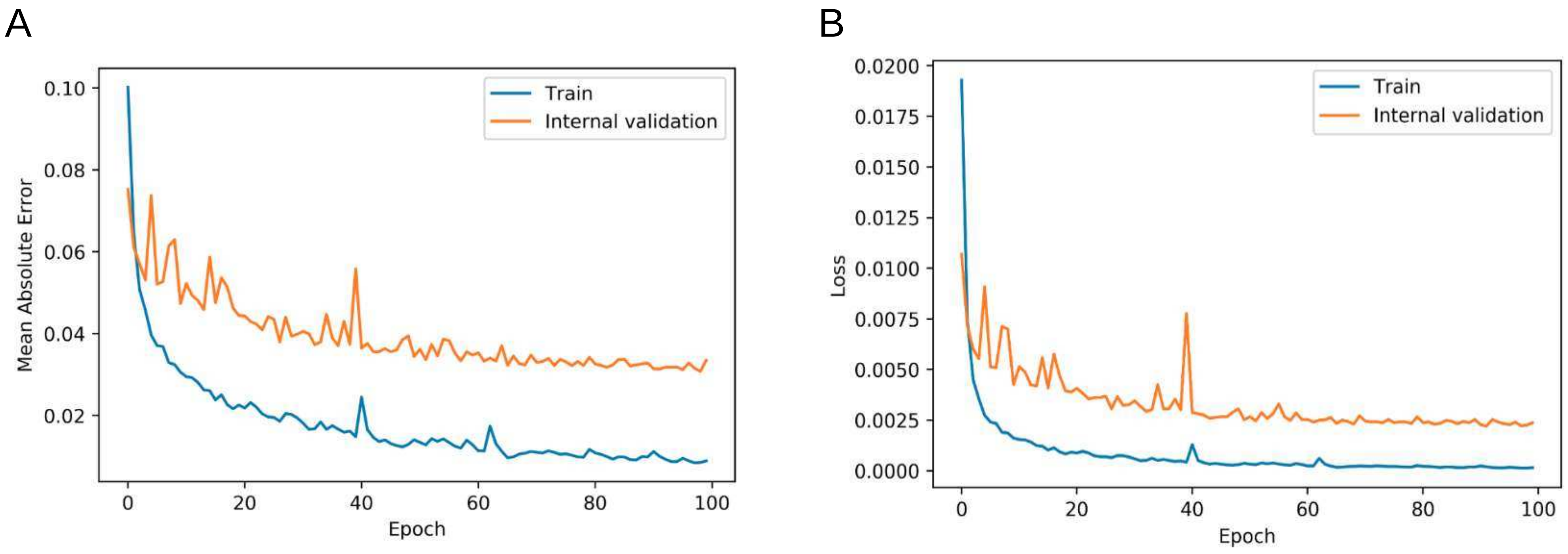

A
0.10
0.08
0.06
0.04
0.02
Mean Absolute Error
0
20
40
60
80
100
Epoch
Train
Internal validation
B
0.0200
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
Loss
0
20
40
60
80
100
Epoch
Train
Internal validation

Figure E2

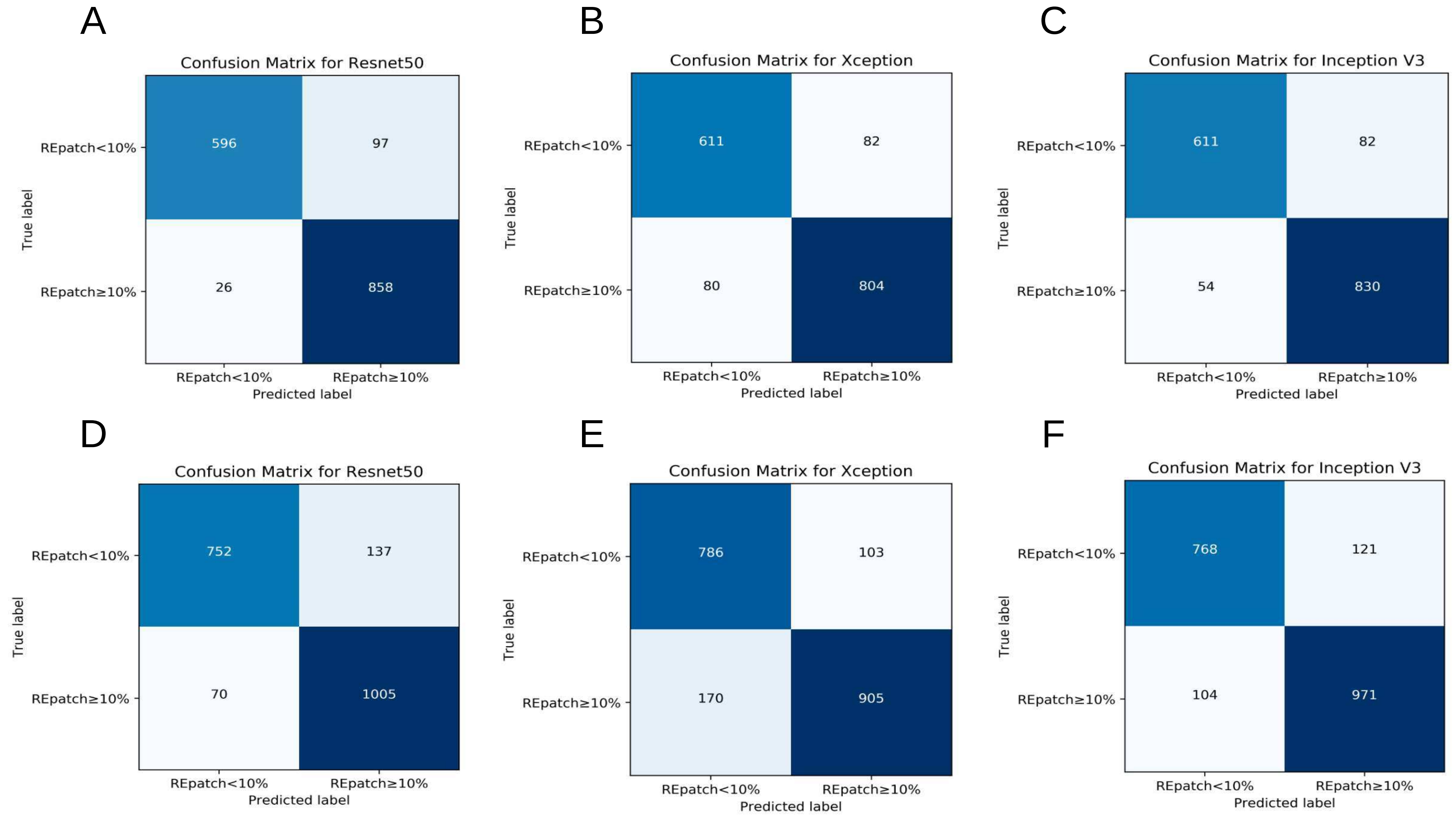

A
Confusion Matrix for Resnet50
REpatch<10%
REpatch≥10%
True label
596
97
26
858
REpatch<10%
REpatch≥10%
Predicted label
B
Confusion Matrix for Xception
REpatch<10%
REpatch≥10%
True label
611
82
80
804
REpatch<10%
REpatch≥10%
Predicted label
C
Confusion Matrix for Inception V3
REpatch<10%
REpatch≥10%
True label
611
82
54
830
REpatch<10%
REpatch≥10%
Predicted label
D
Confusion Matrix for Resnet50
REpatch<10%
REpatch≥10%
True label
752
137
70
1005
REpatch<10%
REpatch≥10%
Predicted label
E
Confusion Matrix for Xception
REpatch<10%
REpatch≥10%
True label
786
103
170
905
REpatch<10%
REpatch≥10%
Predicted label
F
Confusion Matrix for Inception V3
REpatch<10%
REpatch≥10%
True label
768
121
104
971
REpatch<10%
REpatch≥10%
Predicted label

Figure E3

Article's Online Repository at www.jacionline.org

METHODS

Training and internal validation datasets

Biopsies of patients with CRSwNP (n = 1465) were obtained from the Department of Otolaryngology in the Third Affiliated Hospital of Sun Yat-sen University (SYSU) in China from January 2008 to December 2018. Following screening for staining, size, and quality of specimens, 179 patients were used in this analysis. The patients were randomly divided into two groups: 167 patients in the training dataset and 12 patients in the internal validation dataset. After all slides were scanned through an automatic digital slide scanner (Panoramic 250 FLASH, 3DHISTECH Ltd., Budapest, Hungary), we obtained 179 digital whole slide images (WSIs). The lamina propria of mucosa were sketched, excluding large glands, through an automated slide analysis platform (ASAP) (Radboud University Medical Center, The Netherlands) to yield regions of interest (ROI). Patches in ROI were automatically extracted under 400X high-power field using Openslide (version 3.4.1, University of Pittsburgh, Pittsburgh, PA, USA). There were 167 WSIs containing 23048 patches for the training dataset and 12 WSIs containing 1577 patches for the internal validation dataset (Fig. E1).

External test dataset

Sixteen patients (16 WSIs) with nasal polyps were randomly selected from the First Affiliated Hospital of SYSU (n=9) and the Fifth Affiliated Hospital of SYSU (n=7) from January 2017 to December 2018. Independent preparations by each hospital were used for hematoxylin and eosin staining as well as WSI scanning. In total, 1964 patches were obtained using the same method mentioned above.

Labeling

In total, 26,589 patches were independently described and labeled by a committee comprising two competent pathologists with more than 10 years of experience, and an expert pathologist with more than 30 years of experience who was consulted in case of disagreement. The two competent pathologists identified and counted the number of eosinophils (n1), number of lymphocytes (n2), number of neutrophils (n3), and number of plasma cells (n4) in each patch. The number of infiltrating inflammatory cells was regarded as the sum (t), and the ratio of eosinophils ($RE_{patch\text{-}actual}$) was n1/t. When the two pathologists' assessment of $RE_{patch\text{-}actual}$ differed by ≤5%, the average value was used. If the difference was greater than 5%, the patch was rechecked by the expert pathologist, and the value was corrected as necessary. These assessments yielded the average of all patches from WSI, designated as $RE_{slide\text{-}actual}$. CRSwNP patients were classified as eosinophilic when the proportion of tissue eosinophils exceeded 10% of total infiltrating inflammatory cells as previously reported[1]; otherwise, they were regarded as non-eosinophilic CRSwNP.

Deep learning and transfer learning methods

In this study, our artificial intelligence chronic rhinosinusitis evaluation platform (AICEP) compared three commonly used architectures (Resnet50, Xception, and Inception V3) for application of a transfer learning algorithm to assess their performance in the classification and regression of patches extracted from WSIs. Each model loaded the weights pre-trained on the ImageNet dataset, then removed their top layer. Next, to distinguish patches with RE_{patch} values greater or less than the truncated value using a classification algorithm, a full-connection layer with two neurons was added and each neuron contained weights and an activation function, so it can map input value to output value nonlinearly. To predict exact RE_{patch} values with a regression algorithm, we chose the model with the greatest area under the curve (AUC) and added a full-connection layer containing only one neuron. Importantly, no activation function was used at this time to ensure that the model exhibited a broader output value. Within 100 epochs (iterations through the entire training dataset), the retrained weights were saved due to the absence of further improvement in the mean absolute error (MAE) (Fig. E2, A) and the mean square error loss (MSEL) (Fig. E2, B). Finally, the parameters of all layers of quantitative regression architecture were fine-tuned in accordance with the input images and corresponding labels (Fig. 1).

To train and evaluate our models, we adopted the Keras (version 2.2) framework using Tensorflow (version 1.8) backend within Python (version 3.6) programming language, including libraries such as numpy, matplotlib, and Scikit-learn. Computing power was provided by one Tesla V100 GPU with 32GB memory on a Nvidia DGX1 server, which had eight Tesla V100 GPUs, 512 GB DDR4 memory, and 7 TB SSD.

Model and algorithm performance evaluation

Qualitative classification

For the internal validation dataset and external test dataset using Resnet50, Xception, and InceptionV3 for data training, AICEP provided an effective approach for qualitative classification. WSI results were classified as eosinophilic when RE_{slide} exceeded 10%, as previously mentioned. The sensitivity (true positive rate) and specificity (false positive rate) of the confusion matrices of these three models were calculated, as were the areas under the receiver operating characteristic curve (AUC). The model with the highest AUC value was selected for subsequent quantitative analyses. In addition, to verify whether the model was trained correctly based on the characteristics of eosinophils, we used visual gradient-weighted class activation mapping (Grad-CAM).

Quantitative analysis

Evaluation of RE_{patch} in internal validation and external test datasets of AICEP

All patches in both internal validation and external test datasets were input into the AICEP model for simulation, which produced $RE_{patch-predict}$. In addition, the MAE of $RE_{patch-predict}$ and $RE_{patch-actual}$ was calculated. The concordance between $RE_{patch-predict}$ and $RE_{patch-actual}$ was evaluated using the intraclass correlation coefficient.

RE_{slide} comparison between internal validation and external test datasets

For the internal validation and external test datasets, we compared $RE_{slide-predict}$ and $RE_{slide-actual}$ separately. The concordance between $RE_{slide-predict}$ and $RE_{slide-actual}$ was evaluated via intraclass correlation coefficient. In addition, we randomly selected 10 RE_{patch} values of each WSI analysis by a bootstrap method and calculated the average. The bootstrap process was repeated 50 times for each WSI analysis to evaluate and compare the diagnostic effect of the traditional method and of AICEP.

Diagnostic time comparison between AICEP and pathologists

Times for $RE_{slide-predict}$, $RE_{slide-tm}$, and $RE_{slide-actual}$ were calculated.

Statistical analysis

Using a bootstrap simulation of 10 random fields for diagnosis, each WSI was repeated 50 times and compared with $RE_{slide-actual}$. The intraclass correlation coefficient was used to assess agreement between $RE_{predict}$ with RE_{actual}. Receiver operating characteristic curves (ROC) were adopted to evaluate the diagnostic results of AICEP on eCRSwNP. All tests were two-sided, and $P < 0.05$ was considered statistically significant.

References

E1. Cao PP, Li HB, Wang BF, Wang SB, You XJ, Cui YH, et al. Distinct immunopathologic characteristics of various types of chronic rhinosinusitis in adult Chinese. The Journal of allergy and clinical immunology 2009; 124: 478-484, 484.e471-472.

Figure E1. Workflow diagram. It illustrated the overall experimental design, and the flow of whole slide images through extraction and labeling process, the training of transfer learning models using internal dataset, and the evaluating of the models with internal validation dataset and independent external test dataset.

Figure E2. Plot showed the performance in the training and internal validation datasets. Mean absolute error was plotted against the training epoch (A) and mean square error loss was plotted against the training epoch (B) during training the quantitative regression architecture over the course of 100 epochs. The mean absolute error and loss of validation showed great performance with little overfitting due to the diversity of the training dataset.

Figure E3. Confusion matrix of models' classification of patch with RE≥10% from patch with RE<10%. A, B, C, Confusion matrix of internal validation dataset for models of Resnet50, Xception and Inception V3, respectively. D, E, F, Confusion matrix of independent external test dataset for models of Resnet50, Xception and Inception V3, respectively.

Table E1. Consistency assessment for AICEP in internal validation dataset and external test dataset according to the RE$_{\text{patch-actual}}$ and RE$_{\text{slide-actual.}}$

Table E2. Comparison of time-consuming between AICEP and pathologists.