

Received June 29, 2021, accepted July 10, 2021, date of publication July 26, 2021, date of current version September 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3099687

Deep Learning Framework for Preoperative Recognition of Inverted Papilloma and Nasal Polyp

TAO REN¹, XINYAO LI², YICONG TIAN³, AND WEI LI²

¹Department of Software College, Northeastern University, Shenyang 110819, China

²Department of Otolaryngology, The First Hospital of China Medical University, Shenyang 110001, China

³Department of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110819, China

Corresponding author: Wei Li (wli@cmu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61473073 and Grant 61433014; in part by the Fundamental Research Funds for the Central Universities under Grant N161702001, Grant N171706003, Grant 182608003, and Grant 181706001; and in part by the Fundamental Research Funds for the China Medical Universities under Grant HMB201901104.

ABSTRACT Surgery is the most commonly used method of curing inverted papilloma (IP) or nasal polyp (NP). Although accurate preoperative recognition by computed tomography (CT) is a critical aspect of surgical planning, the minor CT imaging differences in such lesions may be a challenge. Therefore, we have devised a deep learning framework for automatic recognition of IP and NP in CT. The proposed framework involves two major steps: (a) use of a convolutional neural network (CNN) to preclassify lesions and (b) automatic IP/NP recognition. The preclassify CNN enables classification of CT slices according to anatomic structure. Separate networks are then implemented to differentiate IP and NP accordingly. Once the framework was trained using a CT dataset (5681 slices) from 136 patients, it outperformed other methods during evaluation, achieving 89.30% accuracy (area under the curve [AUC]=0.95) in classification. The proposed framework has clear potential as a clinical tool, enabling effective and highly accurate preoperative recognition of IP and NP.

INDEX TERMS Deep learning, inverted papilloma, nasal polyp, pre-classify, recognition.

I. INTRODUCTION


Inverted papilloma (IP) is a common but benign sinonasal neoplasm that has recently drawn much attention in the realm of otolaryngology, given its potential for local invasion/recurrence or malignant transformation [1]. IP typically originates from the lateral wall of nasal cavity, middle turbinate, or ethmoid recess [2], often signaled by maxillary and ethmoid sinus dilatation. Frontal and sphenoidal sinuses are rarely affected [3], [4]. As the most common benign nasal mass, nasal polyp (NP) shares clinical symptoms of IP [5], making it difficult for otolaryngologists to distinguish the two. Both IP and NP are cured through surgery, albeit by quite different means. precise preoperative differentiation is thus critical. Although imaging studies, chiefly computed tomography (CT) and magnetic resonance imaging (MRI), are indispensable for preoperative assessments in this setting [6], IP lacks distinctive features on CT (as does NP),

appearing only as a soft tissue density. Partial or complete obstruction of adjacent orifices and concavities caused by IP may produce secondary inflammatory changes (edema and mucosal thickening), hampering the differentiation of IP from other nasal masses (especially NP) [7]. Its tell-tale graphic features (ie, cerebriform pattern) are detectable by MRI [8], but related costs and long waiting times are prohibitive, which is why CT remains the first option. A breakthrough of IP and NP recognition based deep learning by using CT would be highly beneficial.

A number of studies have explored the use of medical imaging features for computer-aided diagnosis (CAD). Such efforts primarily have involved either a hand-crafted feature approach or a deep learning-based network.

A. HAND-CRAFTED FEATURE METHODS

Hand-crafted methods rely upon feature extraction for classification, which in most instances, is aimed at various nodules. Haralick's texture features [9] addresses lung nodules

The associate editor coordinating the review of this manuscript and approving it for publication was Abdullah Iliyasa .

in this way, whereas texture, shape, and context are utilized by Jacobs [10] for this purpose. Statistical distributions [11] have also been used to classify thyroid nodules, applying the sparse fast Fourier transform (SFFT) algorithm as a filter in fusing of features. Similarly, fused histogram of oriented gradient (HOG) and local binary pattern (LBP) descriptors have fueled support vector machine (SVM) classifiers for detecting diabetic macular edema by optical coherence tomography [12]. In the field of breast cancer [13], an AdaBoost classifier applied to Haar-like features has helped to identify a preliminary set of tumor regions. Unfortunately, hand-crafted feature methods are insufficiently robust and require highly selective features for proper classification.

B. DEEP LEARNING BASED METHODS

In recent years, deep learning methods based on convolutional neural networks (CNNs), such as AlexNet [14], GoogleNet [15], VGG [16], residual net (ResNet) [17], U-Net [18], Faster R-CNN [19], Single-Shot Detector (SSD) [20], and YOLO v3 [21], have achieved outstanding performance in fields of image classification, segmentation, and recognition. CNNs in the realm of CAD have been used to classify lung nodules [22], liver tumors [23], and Alzheimer's disease [24]; recognize melanoma [25] and breast masses [26]; segment HEP-2 cells [27], rectal tumors [28], and covid-19 [29]. To date, however, sinonasal studies of this sort are still quite rare. Chang [30] has identified nasal tumors and fibrosis by a neural network, and an automated method of segmentation for MRI radiotherapy of nasopharyngeal carcinoma has been proposed [31]. As for recognition of nasal polyp, Wu [32] has applied a method-based artificial method to classify phenotyping of nasal polyp by whole-slide imaging (WSI). Kim [33], [34] employed a deep learning method to identify maxillary sinusitis or normal cases by Waters' view radiographs. However, the studies of recognition of nasal polyp and inverted papilloma are still blank. Compared with hand-crafted feature methods, the clear advantage of CNNs is the effective extraction of highly discriminating features.

Both IP and NP vary in shape and distribution, showing a morphologic spectrum even the same patient across CT slices. The NP of Figure 1(a) has irregular shapes and boundaries due to anatomic variations, and the areas of morbidity are numerous. Figure 1(b) shows an IP with irregular shapes and boundaries, and the morbidity areas are also varied. Irrelevant areas (brain and eyeballs) cover parts of the CT slices, so it is difficult to recognize NP and IP using one simple classification network of deep learning basis.

In this publication, a novel framework, unlike any other related works and methods, is proposed for automatic preoperative recognition of IP and NP. This end-to-end deep learning framework has two-component networks, one for pre-classification of anatomic structure and the other for recognition of IP and NP. The pre-classification network serves to categorize sinonasal CT slices by anatomic details, and there are two recognition networks trained for separate

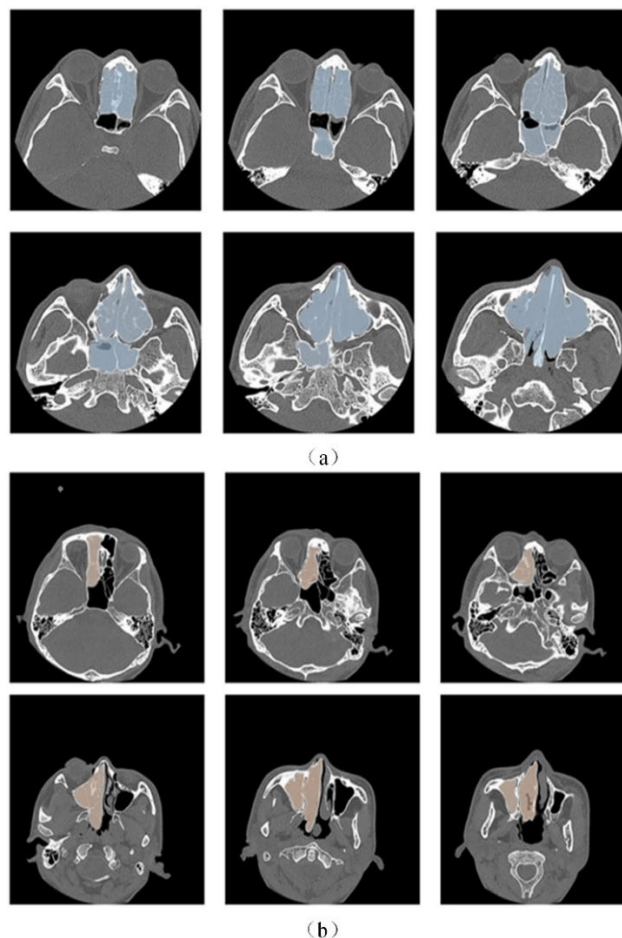


FIGURE 1. Illustration of (a) NP and (b) IP (areas of NP or IP morbidity shown for edification only, not for training or testing).

sinonasal regions. No manual preprocessing is involved, and the pre-classification network ensures greater recognition accuracy.

II. MATERIALS AND METHODS

The proposed end-to-end framework (Figure 2) has two major steps:

- (1) Classify upper (USN) and lower (LSN) sinonasal CT slices by CNN, and
- (2) Extract regions of interest (ROIs) for NP and IP recognition in USN and LSN separately

A. DATA COLLECTION

The available dataset includes 5681 slices from 136 patients (IP, 49; NP, 87), each scanned at the First Hospital of China Medical University (Shenyang, China) using an Aquilion One system (Canon Medical Systems, Otawara, Tochigi, Japan) configured as follows: voltage, 100 kV; current, 400 mA; slice thickness, 0.5 mm; and scanning range, 287 mm. The axial pixel size of each slice was 512×512 . Pathologic reports supplied the ground truth for disease type, and three experienced staff otolaryngologists provided ROI ground truth. The scientific committee in the First Hospital of China Medical

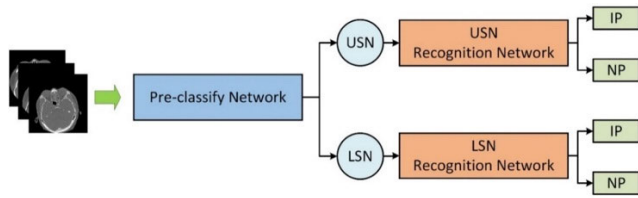


FIGURE 2. Framework of the proposed method.

University waived privacy restrictions, releasing patient data for scientific research.

The framework dataset was allocated as follows: 4376 slices from 114 patients (IP, 39; NP, 73) for training and 1305 slices from 22 patients (IP, 8; NP, 14) for testing. In our experiments, hyperparameters of training data were adjusted using a 10-fold cross-validation strategy. At each fold, we annotated the dataset by patient ID so that all cross-section slices belonging to the same patient could be used for either training or validation.

B. PRECLASSIFICATION ON ANATOMIC BASIS

The sinonasal region is situated roughly above nasal floor plane and beneath base of skull. The nasal cavity sits in mid-sinonasal region and is rimmed by four paranasal sinuses (maxillary, ethmoid, frontal, and sphenoid). This regional anatomy is complex, the variably shaped cavities or air cells presenting quite different imaging features. It is difficult to train a model in IP and NP recognition without considering these elements. We have subsequently devised a preclassification network to categorize CT slices based on anatomic details. The LSN region, extending from plane of nasal floor (Figure 3, line C) to bottom of orbit (Figure 3, line B) [35], is largely occupied by maxillary sinus (spacious cavities) [36]. The USN region, extending from base of orbit to base of skull (Figure 3, line A) [37] harbors ethmoid, frontal, and sphenoid sinuses (small, irregular air cells) [5], [6].

1) PREPROCESSING

Pixel values may differ somewhat among scanners or vary due to illumination, impacting model-driven feature extraction. Ordinarily, they should be normalized. Data augmentation is needed to increase sampling numbers to diversify the dataset, thus strengthening the model’s generalizability. Given the directional aspects of sinonasal slices, mirror flip, random shear (parameter range, 0.2), zoom (parameter range, 0.2), and rotation (0-30°) were utilized.

2) NETWORK ARCHITECTURE

CNNs have been used in a series of medical imaging classification tasks [22]–[25], so we opted for CNN as the pre-classification network (depicted in Figure 4). The usual CNN model has three blocks, including two-dimensional (2D) convolutions and 2D pooling finishes, with fully connected layers, to classify features extracted from convolutional

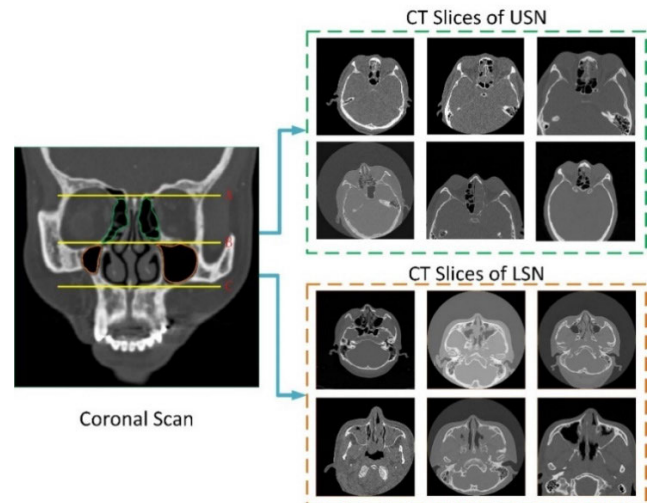


FIGURE 3. Illustration of USN and LSN regions (green circles marks USN air cells; orange circles marks LSN cavities).

layers. The pre-classification task was simpler and expended less time than customary procedures, so we set the layer number at 8. To enhance convergence speeds of the training model and prevent overfitting, we converted the non-linear CNN activation function from ReLU to leakyReLU. We also added batch normalization (BN) layers between conventional and max-pooling layers. Finally, we engaged softmax to determine network predictions, assigning slices by class.

C. IP AND NP RECOGNITION BASED ON YOLOv3 MODEL

In NP and IP recognition, otolaryngologists focus on the sinonasal area in CT, ignoring concomitant brain and eye contributions. Features of these irrelevant regions hinder feature extraction and training of networks. Sinonasal sizes also visibly differ in CT slices. To detect multiscale ROIs and classify disease types of ROIs effectively, we chose YOLO v3 for end-to-end recognition network modeling. Because USN and LSN slices differ substantially in lesion distribution and shape, the two networks were thus trained to recognize IP and NP in USN and LSN separately.

1) MULTI-SCALE RECOGNITION NETWORK

To extract multiscale ROIs in down-sampling, three feature map sizes were set in YOLO v3, each feature map having three corresponding prior boxes (see Table 1).

There are three components of YOLO v3, the first being input, and feature extraction. Image input of the network was 512 × 512, but prior to inputting slices, pixel values were normalized. Feature extraction was enabled by Darknet-53, without fully connected layers. Unlike the down-sampling process of traditional CNN, Darknet-53 holds the convolution core at a stride of 2 × 2 to replace max-pooling. Feature map sizes were then reduced to 1/32 of originals through five down-samplings. In a series of residual blocks, Darknet-53 achieves higher training

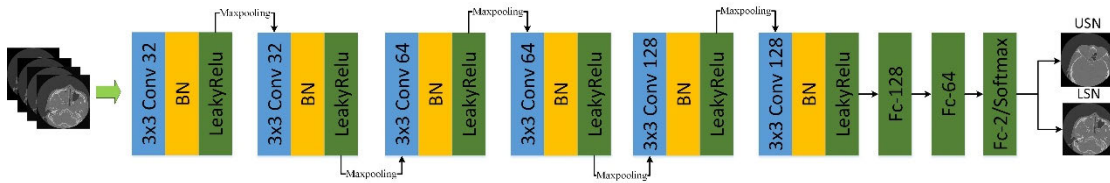


FIGURE 4. The architecture of the preclassification network, including six 2D convolutional layers, two fully connected layers, and a softmax layer.

TABLE 1. Relations between feature maps and prior differently sized boxes.

FEATURE MAP	13×13	26×26	52×52
RECEPTIVE FIELD	BIG	MEDIUM	SMALL
PRIOR BOX	(116×90) (156×198)	(373×326) (30×61) (62×45) (59×119)	(10×13) (16×30) (33×23)

efficiency and deeper network structure than other CNNs. The structure of the recognition network is shown in Figure 5.

The second component is the fusion of multiscale features. To detect and classify multiscale regions, YOLO v3 has three scales of feature maps. Down-sampled and up-sampled tensors are fused by concatenation method. In this way, the network integrates multiscale contextual imaging features.

The third component is output. According to the three different scales of feature maps, outputs of YOLO v3 are designed as tensors with three corresponding sizes. Nine prior anchors are divided equally by three different scales of output tensors according to the size of each prior anchor. The three output tensors' sizes are $y1$: $13 \times 13 \times 21$, $y2$: $26 \times 26 \times 21$, and $y3$: $52 \times 52 \times 21$.

Among the output tensors, size was determined by the sizing of three feature maps. Each predicted box had five parameters ($x, y, w, h, conf$), where x is lower left-hand abscissa, y is lower left-hand longitudinal coordinate, w is width, h is height, and $conf$ is confidence degree for classification. The depth of the tensor was determined by the following formula:

$$Ten_{deep} = 3 \times (Par_{num} + Class_{num}) \quad (1)$$

where Par_{num} is the number of each predicted box, $Class_{num}$ is the number of our classification tasks.

Finally, we used Logistic regression to score 9 prediction boxes and ultimately selected the boxes with the highest score as ROI.

2) MULTI-TASK LOSS

Each training image was ground-truth annotated, including a class label y and a ground-truth bounding-box regression target l , expressed as a 4-dimensional vector (x -position, y -position, width, height). Multitask loss variables l_B and l_C of each labeled bounding box were utilized to train bounding-box regression and classification. In bounding-box

regression, we calculated the offset between the predicted box and ground truth, using the L2 norm to make predicted boxes approach their ground truth l , with L^B as the loss of all bounding-box regression as below.

$$L^B(l, l^*) = \sum_{r=1}^R (l_r - l_r^*)^2 \quad (2)$$

with l_r and l_r^* representing ROI ground-truth labels and predicted outputs separately.

In ROI classification, distinction between IP and NP was formulated as a binary problem, applying cross-entropy to make predicted category y^* approach its ground truth y , using L^C to describe the loss as below.

$$L^C(y, y^*) = \sum_{r=1}^R - (y_r (\log(y_r^*)) + (1 - y_r) (1 - \log(y_r^*))) \quad (3)$$

with y_r and y_r^* representing ground-truth labels and predicted outputs of classification tasks respectively.

III. RESULTS

A. IMPLEMENTATION DETAILS

Open-source platforms were chosen as the deep learning frameworks of preclassification (Keras) and recognition (Tensorflow) networks. As for the training processes, the pre-classification and recognition networks are separate. The training processes of all experiments were conducted on a workstation powered by an NVIDIA (Santa Clara, CA, USA) GeForce GTX 1060 GPU (6 GB of memory), using the compute unified device architecture (CUDA) toolkit 8.0. Parameters of the Adam optimizer were set at $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 1e - 8$ to minimize the loss function. In the training of the USN recognition network, we used parameter files trained by LSN to fine-tune the model, with all established layers trainable.

The basic learning rate was generally set to 0.0001 and reduced by 90% if the validation error did not decline after

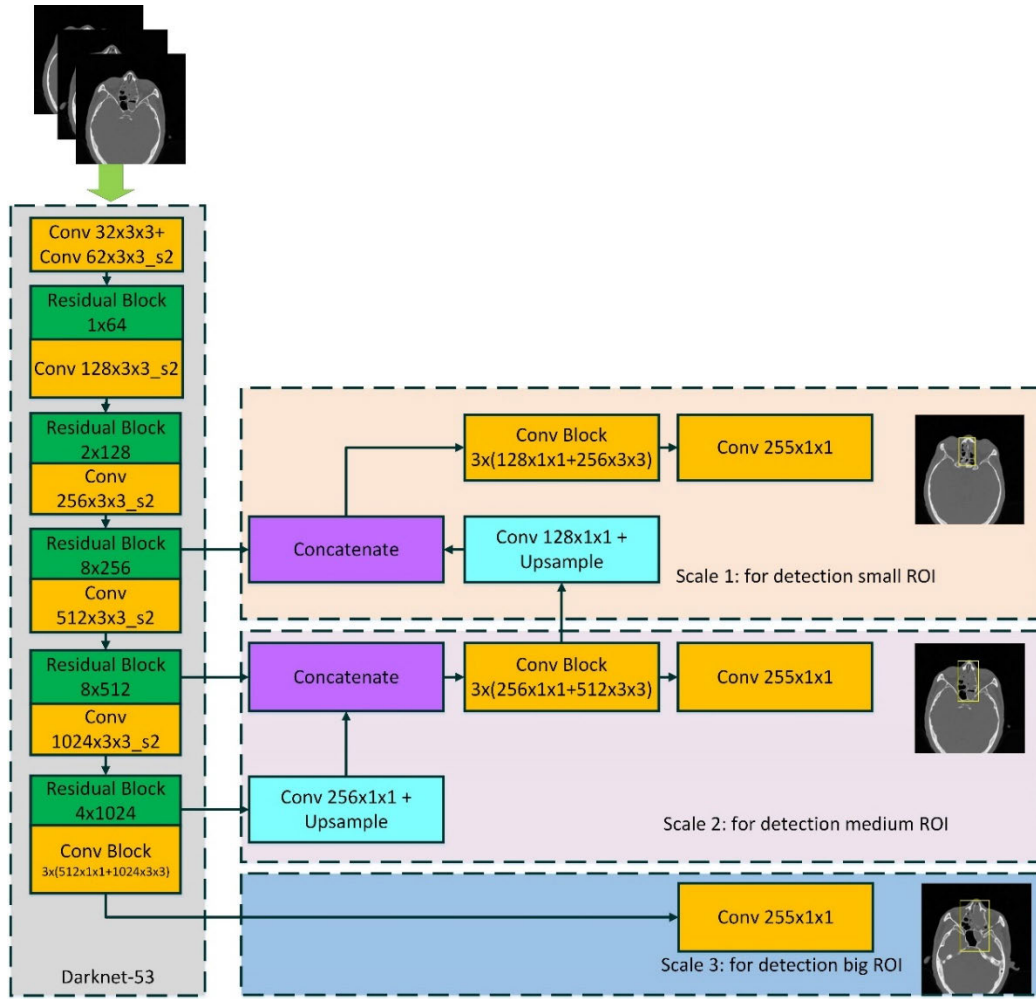


FIGURE 5. The architecture of YOLOv3, including Darknet-53 and multiscale outputs.

10 epochs. To prevent overfitting, a patience of 20 epochs and a minimum delta of 0.001 served as criteria for early stopping. Batch sizes of pre-classification and recognition networks were set to 32 and 4, respectively. The epoch for each network was 100. The hyperparameters for each trained network are shown in Table 2.

B. EVALUATION METRIC

We evaluated the accuracy, recall/sensitivity, and specificity of classification results, defining IP as positive and NP as negative. Accordingly, TP, TN, FP, and FN signified true positive, true negative, false positive, and false negative outputs, respectively. The above metrics were calculated as shown below.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$recall/sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$specificity = \frac{TN}{TN + FP} \quad (6)$$

C. ANALYSIS OF RESULTS

First, the performance of the preclassification network is analyzed. Moreover, a series of three experiments were performed to assess the effects of feature extraction, ROI extraction, and pre-classify on recognition results. The outcomes are discussed below.

1) PERFORMANCE OF PRECLASSIFICATION

To analyze the performance of different activation functions, the ablation experiments were designed. Experimental results are shown in Table 2.

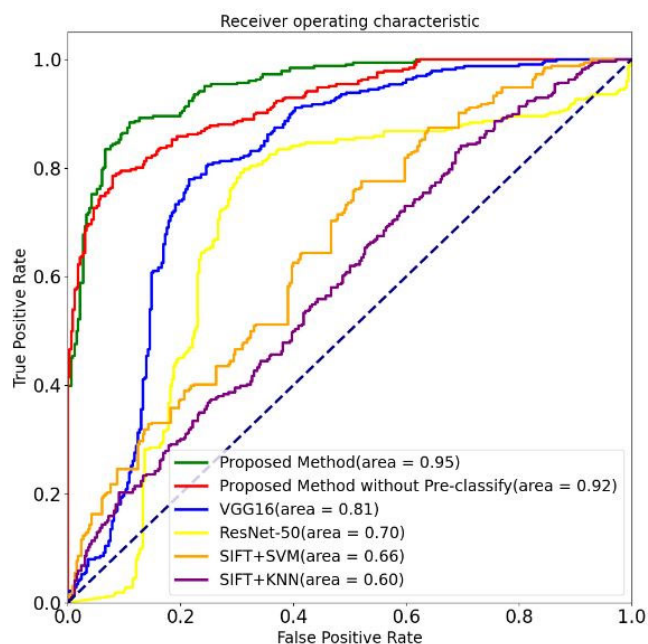
As shown in Table 2, the use of leakyrelu is advantageous to increase the performance of preclassification. Leakyrelu can alleviate the problem of neuron death to some extent, which can raise the network performance.

2) IMPACT OF FEATURE EXTRACTION

Feature extraction impact was explored through hand-crafted feature experimentation. The SIFT descriptor, which is stable in this regard, was duly implemented. We used pixels in

TABLE 2. Comparison of Pre-classification network.

	EPOCH	BATCH SIZE	LEARNING RATE	PATIENCE	MINIMUM DELTA	β_1	β_2	ε
Preclassification	100	32	0.0001	20	0.001	0.9	0.999	$1e-8$
Expriement2	100	32	0.0001	20	0.001	0.9	0.999	$1e-8$
Expriement3	100	4	0.0001	20	0.001	0.9	0.999	$1e-8$
Expriement4	100	4	0.0001	20	0.001	0.9	0.999	$1e-8$

**FIGURE 6.** ROC curves of experimental data.

16×16 neighborhoods to extract features, dividing blocks into 16 sub-blocks of 4×4 and generating an 8-bin orientation histogram for each sub-block. A 128-bin features vector was thereby extracted from each slice. K-means clustering ($k = 16$) was also applied, generating a 16-dimension codebook for these 128-bin features. To adjust hyperparameters, a 10-fold cross-validation training strategy was exercised on training data. Finally, both k-Nearest Neighbor (KNN) and SVM were trained in these features for binary classification tasks. The neighbors-num of KNN was set to 5, with an SVM radial basis function (RBF) kernel. As shown in Table 2 and Figure 6, the proposed method performed considerably better by comparison. SIFT features were extracted for classification training, collecting only low-level features (shapes and pixel values) of images. However, these low-level hand-crafted features are not sufficiently distinctive to classify IP and NP, both having irregular contours. Deep convolutional networks are more robust than hand-crafted descriptor methods, with greater capacity for representation.

3) IMPACT OF ROI EXTRACTION

To investigate the impact of ROI extraction, we experimented with deep learning classification networks VGG16 and

ResNet-50, both performing well via ImageNet. Batch sizes were set to 32, and the epoch for training was 100. In the training process, we applied 10-fold cross-validation to adjust hyperparameters. Patience and minimum delta of early stopping were set to 20 epochs and 0.001, respectively. Adam was used to optimize the loss function. L2 regularization was invoked to prevent overcomplicated parameters in fully connected layers to discourage overfitting. BN layers were used to normalize convolutional layer outputs, and data augmentation was achieved as above (Section II.B.1). As shown in Table 2 and Figure 6, the proposed method again performed considerably better by comparison.

Otolaryngologists only focus on part of a CT slice (ie, sinonasal region) in recognition of IP and NP. The low signal-to-noise ratios of whole slices are problematic for traditional deep learning classification networks, which often perform poorly in this setting. Recognition networks avoid features in irrelevant areas by ROI extraction. Moreover, noise features fed to training models promote overfitting. This underscores the need for ROI extraction in deep learning frameworks for the recognition of IP and NP in CT.

4) IMPACT OF PRE-CLASSIFY

We experimented with our proposed model, omitting the pre-classification network for comparison. The batch size was set to 4, and the epoch for training was set to 100. As for the setting of batch size, experiments with different batch sizes are employed. The results are shown in Table 5. Other training details were identical to those of comparator methods. Classification outcomes appear in Table 2 and recognition outcomes are shown in Figure 7. ROC curves across experimental systems are shown in Figure 6.

Except for specificity, the proposed method outperformed other techniques, providing better IP recognition. It was noted that high recall is more critical than high specificity because IP has some malignant potential. In terms of recognition results, Figure 7(a) depicts the ground truth of ROI detection. Figures 7(b) and 7(c) illustrate the outcomes of experimentation without pre-classification and the proposed method respectively. Recognition results in rows 1 (LSN) and 3 (USN) indicate that the proposed method recognizes IP more effectively. Rows 2 (LSN) and 4 (USN) show results of NP recognition. Although outputs of both methods resulted in the correct classification (relative to ground truth), the proposed method without pre-classification proved less effective in ROI extraction. Thus, the pre-classification network-enabled more accurate IP and NP recognition.

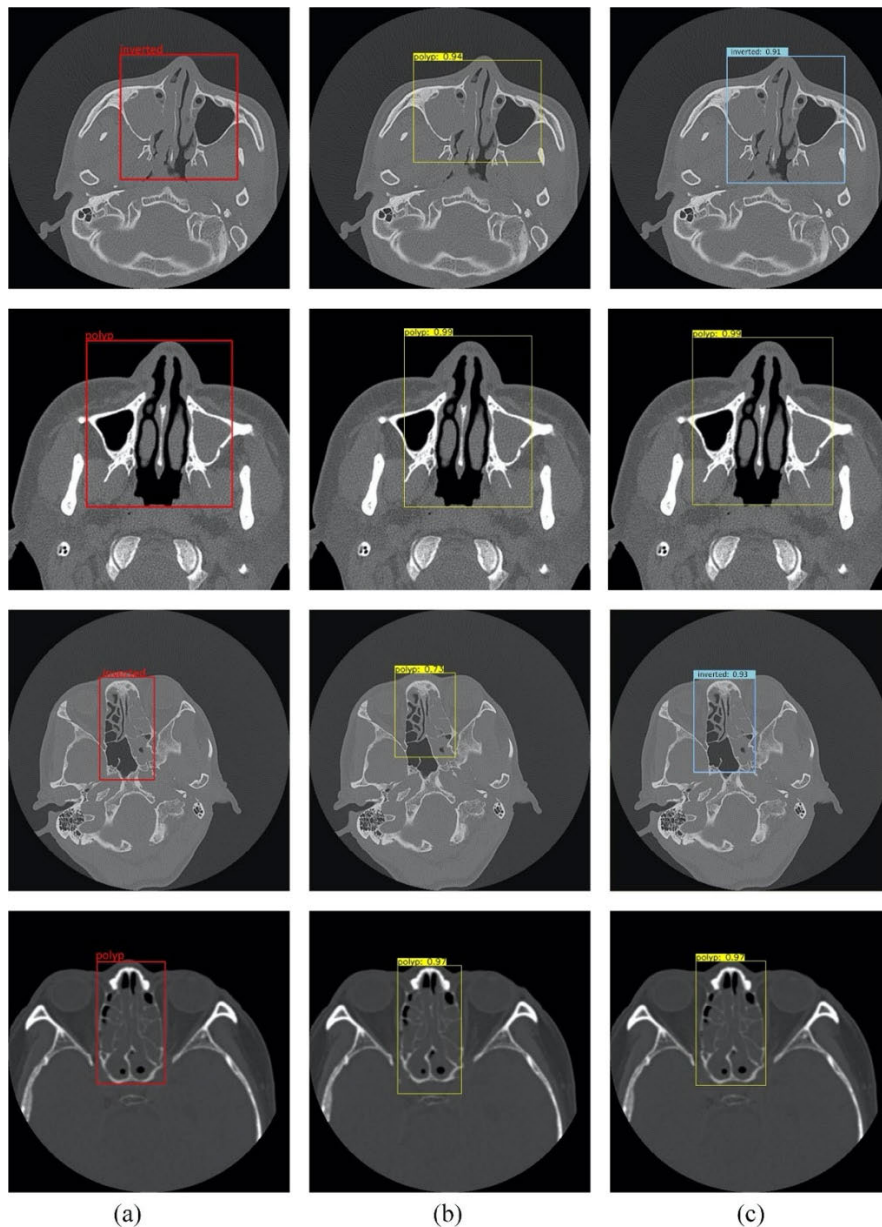


FIGURE 7. Recognition outcomes in CT slices: (a) ground truth; (b) results of experimentation without pre-classification; and (c) results of the proposed method.

TABLE 3. Comparison of Pre-classification network.

NUMBER	ALGORITHMS	ACCURACY(%)	RECALL(%)	SPECIFICITY(%)
1	PRE-CLASSIFICATION NETWORK WITHOUT RELU	97.72%	95.63%	96.89%
2	PRE-CLASSIFICATION NETWORK	99.52%	99.34%	99.33%

Moreover, the detection accuracy is shown in Table 6. We applied average precision (AP) in 2 sizes of intersection over union (IOU) to evaluate the regression of bounding boxes. The report of situations with pre-classification and omitting pre-classification is shown in Table 6. As shown in

Table 6, the pre-classification is beneficial to the regression of bounding boxes.

Organ and skeletal structures within USN and LSN regions are quite different, creating vast morphologic differences in morphologic expressions of the same disease. These are

TABLE 4. Comparison of binary classification methods on testing data.

NUMBER	ALGORITHMS	ACCURACY (%)	RECALL (%)	SPECIFICITY (%)	AUC
1	SIFT+SVM	65.61	64.20	66.13	0.66
	SIFT+KNN	61.58	62.27	60.91	0.60
2	VGG16	78.05	77.61	78.48	0.81
	RESNET-50	73.63	76.99	70.30	0.70
3	PROPOSED METHOD	85.90	80.30	91.50	0.92
	WITHOUT PRE-CLASSIFY				
4	PROPOSED METHOD	89.30	89.01	89.70	0.95

TABLE 5. Comparison Of experiments on different batch sizes.

BATCH SIZE	ACCURACY(%)	RECALL(%)	SPECIFICITY(%)
2	88.89	88.62	89.21
4	89.30	89.01	89.70
8	87.90	87.67	89.65
16	87.55	87.99	86.76

TABLE 6. Comparison of detection accuracy on testing data.

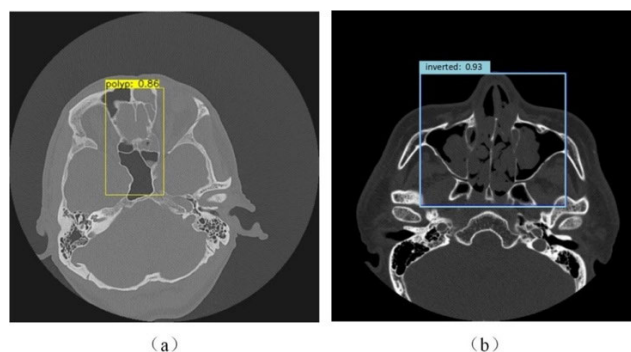
NUMBER	ALGORITHMS	AP@0.5	AP@0.8
1	PROPOSED METHOD	98.72%	94.23%
	WITHOUT PRE-CLASSIFY		
2	PROPOSED METHOD	99.83%	97.33%

problematic in training a robust network to recognize IP and NP. Without pre-classification, results of ROI are not quite accurate. Such errors may then affect further classification performance. ROC curves of Figure 6 demonstrate that the proposed framework is more promising than other methods. Ultimately, pre-classification heightens ROI extraction accuracy thereby ensuring more accurate recognition results.

IV. DISCUSSION

We have presented a novel deep learning framework for effectively distinguishing IP and NP in CT slices. A series of experiments were conducted to evaluate the efficacy of our framework. To our knowledge, this is the first deep learning-based framework for IP and NP recognition. Beyond these results, there are several issues worthy of mention. First, the features extracted by deep convolutional networks are more discriminating than hand-crafted features. Our analysis has nevertheless shown that a classification network of deep convolutional basis is deficient. The results are inadequate, despite a series of methods intended to alleviate overfitting. ROI extraction thus plays an important role in the recognition process. Imaging ROIs are beneficial for IP and NP recognition results, perhaps mitigating overfitting largely due to feature redundancy. Furthermore, anatomically based pre-classification is crucial for accurate ROI extraction. It is an effective means of achieving more accurate ROI extraction.

Although this proposed framework shows promise, there are certain limitations. In Figure 8, our model generated invalid recognition results in both slices. Usually, the bony details of sinuses are relatively symmetric by CT,

**FIGURE 8.** Abnormal recognition results of our model in slices from (a) IP and (b) NP.

but gasification of sphenoid sinuses in Figure 8(a) differs substantially (R>L), owing to disparate skeletal structures. The near-absence of this asymmetry in the training set has prevented accurate ROI detection by the network. The CT slice of Figure 8(b), from a patient with NP, has leaf-like features of IP [37]. This pattern interferes with model results, even if ROI is extracted accurately. Differing skeletal structures of left and right sinuses and abnormal performance in some instances of NP may therefore impact the recognition capacity of this model. Proportionate increases of such slices in the training set would enhance the model's generalizability, reducing these errors. Also, both IP and NP are variably distributed in CT slices. Herein, we regarded all possible distributions of IP and NP as the ground truth of ROIs in our dataset. However, the ROIs of some slices are larger than actual focus areas, impacting recognition results. In the future, we will address these types of problems.

The proposed framework is aimed at the preoperative recognition of nasal polyp and inverted papilloma. The framework is not suitable for healthy patients yet. Therefore, our framework will be improved to support a system for nasal polyp, inverted papilloma, and healthy patients.

Various aspects of sinonasal diseases based on deep learning are yet to be explored. For example, recognition of other benign or malignant sinonasal lesions and recognition/segmentation of invasiveness in malignant sinonasal tumors is open to investigation.

V. CONCLUSION

In this study, we propose a framework for automatic recognition of NP and IP. We have compared our method with

In this study, a framework for automatic recognition of NP and IP has been tested, comparing our proposed method with others as follows: (1) Hand-crafted feature methods (SVM and KNN classifiers using SIFT features); (2) Deep learning-based classification neural networks (VGG16 and Resnet-50); and (3) Our proposed method without preclassification. The method as proposed delivered 89.30% accuracy, 89.03% recall, and 89.70% specificity, outperforming all comparator methods (with exception of specificity). To our knowledge, this is the first deep learning solution for automatic recognition of IP and NP reliant on sinonasal CT slices. The end-to-end framework provided is easily implemented by clinicians, enabling more definitive preoperative recognition of IP and NP.

ACKNOWLEDGMENT

(Tao Ren and Xinyao Li contributed equally to this work.)

REFERENCES

- [1] J. K. Han, T. L. Smith, T. Loehrl, R. J. Toohill, and M. M. Smith, "An evolution in the management of sinonasal inverted papilloma," *Laryngoscope*, vol. 111, no. 8, pp. 1395–1400, Aug. 2001.
- [2] J. S. Kim and S. H. Kwon, "Different characteristics of a single sinonasal inverted papilloma from sequential PET-CT: A case report," *Medicine*, vol. 96, no. 52, p. e9557, Dec. 2017.
- [3] M. Akkari, J. Lassave, T. Mura, G. Gascou, G. Pierre, C. Cartier, R. Garrel, and L. Crampette, "Atypical presentations of sinonasal inverted papilloma: Surgical management and influence on the recurrence rate," *Amer. J. Rhinol. Allergy*, vol. 30, no. 2, pp. 149–154, Mar./Apr. 2016.
- [4] E. Iida and Y. Anzai, "Imaging of paranasal sinuses and anterior skull base and relevant anatomic variations," *Radiol. Clinics North Amer.*, vol. 55, no. 1, pp. 31–52, Jan. 2017.
- [5] M. Fraczek, M. Guzinski, M. Morawska-Kochman, and T. Krecicki, "Investigation of sinonasal anatomy via low-dose multidetector CT examination in chronic rhinosinusitis patients with higher risk for perioperative complications," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 274, no. 2, pp. 787–793, Feb. 2017.
- [6] C. L. Sham, A. D. King, A. van Hasselt, and M. C. F. Tong, "The roles and limitations of computed tomography in the preoperative assessment of sinonasal inverted papillomas," *Amer. J. Rhinol.*, vol. 22, no. 2, pp. 144–150, Mar./Apr. 2008.
- [7] A. Chawla, J. Shenoy, K. Chokkappan, and R. Chung, "Imaging features of sinonasal inverted papilloma: A pictorial review," *Current Problems Diagnostic Radiol.*, vol. 45, no. 5, pp. 347–353, Sep./Oct. 2016.
- [8] G. Fang, H. Lou, W. Yu, X. Wang, B. Yang, J. Xian, X. Song, E. Fan, Y. Li, C. Wang, and L. Zhang, "Prediction of the originating site of sinonasal inverted papilloma by preoperative magnetic resonance imaging and computed tomography," *Int. Forum Allergy Rhinol.*, vol. 6, no. 12, pp. 1221–1228, Dec. 2016.
- [9] F. Han, H. Wang, G. Zhang, H. Han, B. Song, L. Li, W. Moore, H. Lu, H. Zhao, and Z. Liang, "Texture feature analysis for computer-aided diagnosis on pulmonary nodules," *J. Digit. Imag.*, vol. 28, no. 1, pp. 99–115, 2015.
- [10] C. Jacobs, E. M. van Rikxoort, T. Twellmann, E. T. Scholten, P. A. de Jong, J.-M. Kuhnigk, M. Oudkerk, H. J. de Koning, M. Prokop, C. Schaefer-Prokop, and B. van Ginneken, "Automatic detection of sub-solid pulmonary nodules in thoracic computed tomography images," *Med. Image Anal.*, vol. 18, no. 2, pp. 374–384, 2014.
- [11] C.-Y. Chang, S.-J. Chen, and M.-F. Tsai, "Application of support-vector-machine-based method for feature selection and classification of thyroid nodules in ultrasound images," *Pattern Recognit.*, vol. 43, no. 10, pp. 3494–3506, 2010.
- [12] K. Alsaih, G. Lemaitre, J. M. Vall, M. Rastgoo, D. Sidibe, T. Y. Wong, E. Lamoureux, D. Milea, C. Y. Cheung, and F. Meriaudeau, "Classification of SD-OCT volumes with multi pyramids, LBP and HOG descriptors: Application to DME detections," in *Proc. IEEE 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 1344–1347.
- [13] P. Jiang, J. Peng, G. Zhang, E. Cheng, V. Megalooikonomou, and H. Ling, "Learning-based automatic breast tumor detection and segmentation in ultrasound images," in *Proc. 9th IEEE Int. Symp. Biomed. Imag.*, May 2012, pp. 1587–1590.
- [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [16] C. Szegedy, W. Liu, Y. Jia, and P. Sermanet, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Assist. Intervent.*, 2015, pp. 234–241.
- [19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [20] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
- [21] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: <http://arxiv.org/abs/1804.02767>
- [22] P. Sahu, D. Yu, M. Dasari, F. Hou, and H. Qin, "A lightweight multi-section CNN for lung nodule classification and malignancy estimation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 3, pp. 960–968, May 2019.
- [23] E. Trivizakis, G. C. Manikis, K. Nikiforaki, K. Drevellegas, M. Constantinides, A. Drevellegas, and K. Marias, "Extending 2-D convolutional neural networks to 3-D for advancing deep learning cancer classification with application to MRI liver tumor differentiation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 3, pp. 923–930, May 2019.
- [24] C. Feng, A. Elazab, P. Yang, T. Wang, F. Zhou, H. Hu, X. Xiao, and B. Lei, "Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBI-LSTM," *IEEE Access*, vol. 7, pp. 63605–63618, 2019.
- [25] Z. Yu, X. Jiang, F. Zhou, J. Qin, D. Ni, S. Chen, B. Lei, and T. Wang, "Melanoma recognition in dermoscopy images via aggregated deep convolutional features," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 1006–1016, Apr. 2019.
- [26] S. Y. Shin, S. Lee, I. D. Yun, S. M. Kim, and K. M. Lee, "Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images," *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 762–774, Mar. 2019.
- [27] M. Wang, P. Xie, Z. Ran, J. Jian, R. Zhang, W. Xia, T. Yu, C. Ni, J. Gu, X. Gao, and X. Meng, "Full convolutional network based multiple side-output fusion architecture for the segmentation of rectal tumors in magnetic resonance images: A multi-vendor study," *Med. Phys.*, vol. 46, no. 6, pp. 2659–2668, Apr. 2019.
- [28] Y. Li and L. Shen, "cC-GAN: A robust transfer-learning framework for HEp-2 specimen image segmentation," *IEEE Access*, vol. 6, pp. 14048–14058, Mar. 2018.
- [29] R. C. Joshi, S. Yadav, V. K. Pathak, H. S. Malhotra, H. V. S. Khokhar, A. Parihar, N. Kohli, D. Himanshu, R. K. Garg, M. L. B. Bhatt, R. Kumar, N. P. Singh, V. Sardana, R. Burget, C. Alippi, C. M. Travieso-Gonzalez, and M. K. Dutta, "A deep learning-based COVID-19 automatic diagnostic framework using chest X-ray images," *Biocybern. Biomed. Eng.*, vol. 41, no. 1, pp. 239–254, Jan./Mar. 2021.
- [30] C.-Y. Chang, P.-C. Chung, and P.-H. Lai, "Using a spatiotemporal neural network on dynamic gadolinium-enhanced MR images for diagnosing recurrent nasal papilloma," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 1, pp. 225–238, Feb. 2002.

- [31] T. Zhong, X. Huang, F. Tang, S. Liang, X. Deng, and Y. Zhang, "Boosting-based cascaded convolutional neural networks for the segmentation of CT organs-at-risk in nasopharyngeal carcinoma," *Med. Phys.*, vol. 46, no. 12, pp. 5602–5611, Dec. 2019.
- [32] Q. Wu, J. Chen, Y. Ren, H. Qiu, L. Yuan, H. Deng, Y. Zhang, R. Zheng, H. Hong, Y. Sun, X. Wang, X. Huang, C. Shao, H. Lin, L. Han, and Q. Yang, "Artificial intelligence for cellular phenotyping diagnosis of nasal polyps by whole-slide imaging," *EBioMedicine*, vol. 66, Apr. 2021, Art. no. 103336.
- [33] Y. Kim, K. J. Lee, L. Sunwoo, D. Choi, C.-M. Nam, J. Cho, J. Kim, Y. J. Bae, R.-E. Yoo, B. S. Choi, C. Jung, and J. H. Kim, "Deep learning in diagnosis of maxillary sinusitis using conventional radiography," *Invest. Radiol.*, vol. 54, no. 1, pp. 7–15, Jan. 2019.
- [34] C. Wuttiwongsanon, P. Chaowanapanja, R. J. Harvey, R. Sacks, R. J. Schlosser, S. Chusakul, S. Aejumjaturapat, and K. Snidvongs, "The orbital floor is a surgical landmark for the Asian anterior skull base," *Amer. J. Rhinol. Allergy*, vol. 29, no. 6, pp. e216–e219, Nov./Dec. 2018.
- [35] A. Whyte and R. Boeddinghaus, "The maxillary sinus: Physiology, development and imaging anatomy," *Dentomaxillofacial Radiol.*, vol. 48, no. 8, Dec. 2019, Art. no. 20190205.
- [36] R. A. Crosbie, W. A. Clement, and H. Kubba, "Paediatric orbital cellulitis and the relationship to underlying sinonasal anatomy on computed tomography," *J. Laryngol. Otol.*, vol. 131, no. 8, pp. 714–718, Aug. 2017.
- [37] V. Bortoli, R. Martins, and K. Negri, "Study of anthropometric measurements of the anterior ethmoidal artery using three-dimensional scanning on 300 patients," *Int. Arch. Otorhinolaryngol.*, vol. 21, no. 2, pp. 115–121, Apr. 2017.



TAO REN received the B.S. degree in automatic control and the M.S. and Ph.D. degrees in control theory and control engineering from Northeastern University, Shenyang, China, in 2003, 2005, and 2007, respectively.

He is currently a Professor with Northeastern University. He is in charge of 14 projects, such as the National Natural Science Foundation of China. He has published more than 30 high qualified academic articles in several high-ranking journals or conferences. His current research interests include complex networks and intelligent optimization algorithm. He has received the "Chinese People's Liberation Army Science and Technology Progress Award" and several other academic awards.



XINYAO LI received the M.D. degree in otolaryngology from China Medical University, in 2016. She is currently pursuing the Ph.D. degree in genetics. She is currently working with The First Hospital of China Medical University.



YICONG TIAN received the B.S. degree in software engineering from Northeastern University, Shenyang, China, in 2018, where he is currently pursuing the master's degree in biomedical engineering. His research interests include deep learning and image processing.



WEI LI received the M.D. and Ph.D. degrees from China Medical University, in 2004 and 2010, respectively. He studied at the Department of Otolaryngology, Leiden University Medical Center, in 2006. He was a Visiting Scholar with the University of California San Francisco, in 2016. He is currently a Professor with the Department of Otolaryngology, The First Hospital of China Medical University, major in rhinology and expanded endonasal endoscopic approach surgery. He has

published articles in different journals, including the *BMC Cancer*, the *World Neurosurgery*, and the *Journal of Neurological Surgery Part B: Skull Base*. He is a member of the Otolaryngology Branch, Chinese Endoscopist Association, and the Skull Base Branch, Chinese Otolaryngologist Association.

• • •