Grade classification of nasal obstruction from endoscopy videos using machine learning

Nonpawith Phoommanee
Department of Medical Physics and Biomedical Engineering
University College London
London, UK
nonpawith.phoommanee@ucl.ac.uk

Peter J. Andrews
Department of Rhinology and Facial Plastic Surgery
Royal National Throat, Nose and Ear Hospital
London, UK
peterandrews@nhs.net

Terence S. Leung
Department of Medical Physics and Biomedical Engineering
University College London
London, UK
t.leung@ucl.ac.uk

***Abstract*— Nasal obstruction (NO), referring to blockage in the nasal cavity, is prevalent, affecting approximately one-third of the adult population. Consequently, diagnosis typically requires a combination of medical imaging techniques and tests, as NO is often subjective. This study aims to automate the grade classification of NO from anterior nasal cavity images using nasal endoscopy as a standalone diagnostic tool for common NO conditions: allergic rhinitis, chronic rhinosinusitis, and deviated nasal septum. To evaluate this, we examined a proposed method based on a support vector machine (SVM) using two explainable features: the number of the middle turbinate (MT) pixels and the MT contact ratio, derived from the segmentation map in our previous publication, involving 73 participants. This evaluation was compared against deep learning methods - ResNet-50 and Vision Transformers (ViT)-tiny using direct images as input. Our SVM-based method achieved an interrater agreement with the manual grade classification of NO, provided by an ear, nose and throat (ENT) consultant, of 0.46 (moderate agreement) and 0.14 (none or slight agreement) on the validation set and testing set, respectively. While the proposed method introduces the first quantitative approach for differentially diagnosing common NO conditions, further investigation into additional features and strategies to obtain video-level grade classification from frame-level classification data is warranted to achieve a suitable interrater agreement for clinical translation. This has the potential to facilitate the transition of ENT examinations from secondary care to primary care settings, consequently reducing unnecessary ENT referrals.**

***Clinical relevance*— This study showcases the first successful grade classification of NO using anatomical segmentation maps from the anterior nasal cavity. The findings hold significant clinical potential, aiding in the early detection of NO.**

I. Introduction

Nasal obstruction (NO) refers to insufficient airflow through the nose due to a blockage in the nasal cavity. It is prevalent in otolaryngology practices, affecting up to one-third of the global adult population [1]. The primary mucosal causes of NO are allergic rhinitis (AR) and chronic rhinosinusitis (CRS), both with nasal polyps (CRSwNP) and without nasal polyps (CRSsNP). Structural causes include a deviated nasal septum (DNS), often necessitating nasal surgery, such as septoplasty, for NO treatment. The negative health impact of AR and CRS is substantial, leading to exacerbation of obstructive sleep apnea and reduced work efficiency.

Nasal endoscopy (NE) involves a flexible tube with a small camera and bright lighting, providing a detailed assessment of the nasal cavity. This level of examination is not achievable through standard anterior rhinoscopy due to its invasiveness, making NE crucial in the differential diagnosis of NO [2]. Diagnosis using NE is typically complemented by clinical history, allergy testing, CT scans, and quality-of-life evaluation. Additionally, objective examinations such as nasal inspiratory peak flow (NIPF) and acoustic rhinometry are also employed.

To align with NHS England's initiative to streamline diagnostics in primary care and reduce ear, nose and throat (ENT) referrals [3], this paper aims to develop an automated differential diagnosis of major NO conditions (AR, CRS and DNS) from the anterior nasal cavity using nasal endoscopy. The anterior nasal cavity, situated within the nose and separated into left and right parts by the septum along the sagittal plane, plays a crucial role in normal breathing. Conditions like NO can arise if the internal nasal valve (INV) area narrows, contributing to two-thirds of the total nasal airway resistance [4].

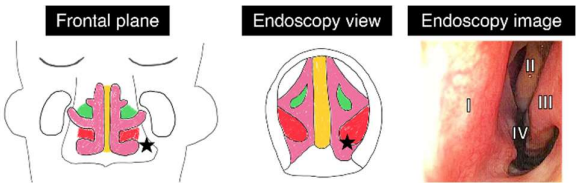

Fig. 1. Different views of the anterior nasal cavity, with the endoscopic image location marked by a black star. Key structures highlighted include the septum (I - yellow), middle turbinate (II - green), inferior turbinate (III - red), and internal nasal valve region (IV - pink).

Nasal endoscopy is recognized as the gold standard for diagnosing deviated nasal septum (DNS), with various classification systems based on observed septum anatomical shapes [5]. For AR and CRS classification, physical appearances such as nasal polyps, mucopurulent secretions, inflamed mucosa, bluish discoloration, and mucosal edema are used [6], [7]. Nasal decongestants may help differentiate AR from CRSsNP patients, as AR patients show improved unilateral NIPF (uNIPF) after applying decongestants, while CRS patients show the opposite outcome [8].

There are already attempts to classify polyps in the anterior nasal cavity, a characteristic of CRSwNP, using deep learning [9], but no studies so far focusing on the classification of ARS, CRSsNP, and DNS. Hence, we review existing subjective grading systems based on key anatomical structures of the anterior nasal cavity [10], [11], [12]. The proposed grading method by Patel et al. has demonstrated high reliability and moderate correlation with unilateral NIPF measurements [10]. Given the association between the INV area, uNIPF, and NO grading, NO grading could aid in distinguishing between AR and CRSsNP, and identifying DNS patients by asymmetry in grading on each side of the nose.

Our validated rule-based expert system [13] will use polyp appearance and NO grading as inputs, allowing for the use of video imaging, potentially digital anterior rhinoscopy, in primary care settings. By implementing this objective method, a standardized approach to assessing NO grading, diagnosing

conditions, and determining appropriate treatment interventions can be established.

We aim to utilize explainable features derived from the semantic segmentation of key anatomical structures in the anterior nasal cavity (see Fig. 1) for the grade classification of NO. We will employ a basic machine learning method, namely Support Vector Machine (SVM), due to its efficiency with small datasets, lower computational requirements, and suitability for primary care settings. This paper demonstrates the grade classification of NO, contributing to the study of NO by collecting and annotating a dataset from participants with or without NO. We aim to develop machine learning models that accurately grade the severity of NO. This approach not only contributes to the advancement of NO research but also provides valuable insights for the practical implementation of ENT examination in primary care settings.

II. Proposed Framework

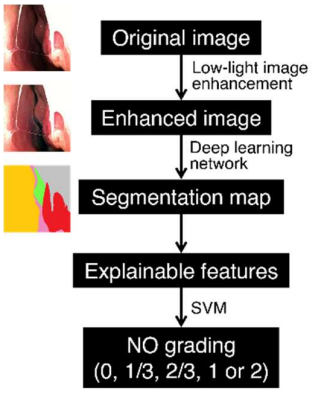

Fig. 2. Framework for grade classification of NO: A custom deep learning network, coupled with low-light image enhancement, was employed to generate the segmentation map [14]. The explainable features, specifically MT pixel and MT contact ratio, were extracted as inputs for SVM analysis to determine NO grading.

Fig. 2 illustrates the framework for grade classification of NO. In our previous publication [14], we applied a low-light image enhancement algorithm to the original images from the anterior nasal cavity recording. Subsequently, we employed a fine-tuned Mask2Former to generate the corresponding semantic segmentation map. Each pixel in the map corresponds to one of the following classes: septum, inferior turbinate (IT), middle turbinate (MT), polyp, airway, and others (structures not present in the key structures mentioned in Fig. 1 or located in regions with excessive brightness).

In this study, we employed the NO grading system proposed by Patel et al. [10], in which they categorized the INV into three grades based on MT visibility: 0, 1, and 2 (see TABLE I. for description). We have modified this grading, allowing doctors to assign a certain level if they are uncertain whether the patient belongs to grade 0 (maybe 0: 1/3) or grade 1 (maybe 1: 2/3). This adjustment aims to better measure the INV's response level after applying a decongestant, as validated in our publication on the rule-based expert system [13].

TABLE I. Description of Validated NO Grading System

Grade	Description
0	The head of the MT can be seen clearly
1	The MT is partially blocked from the view
2	The MT cannot be seen at all

Based on the NO grading description, we can quantify the MT in two ways: the size of the MT and the extent to which it is covered by other anatomical structures (septum, IT, polyp, and others classes). Therefore, we propose two explainable features as measures of NO grading by utilizing the number of pixels and the MT contact (MTC) ratio obtained from semantic segmentation as the main inputs for SVM.

The MTC ratio is defined as follows: MTC ratio = MT contact length excluding the airway / MT perimeter. The representation of the lengths used for MTC ratio calculations is presented in Fig. 3. The MTC ratio ranges from 0 to 1. In specific cases, the MTC ratio is defined as 2, accounting for grade 2 images where the MT is not visible.

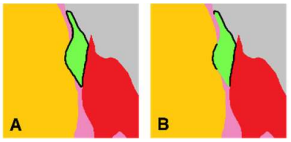

Fig. 3. A black line represents the MT perimeter (A) and MT contact length excluding the airway (B). Classes are color-coded as follows: septum (yellow), IT (red), MT (green), others (gray), and airway (pink).

Fig. 4 shows example images from grades 0, 1, and 2 with corresponding MTC ratios. The MTC ratio is more likely to be 0 for grade 0 images, indicating less contact between the MT and other anatomical structures. Conversely, grade 1 images yield an MTC ratio closer to 1.

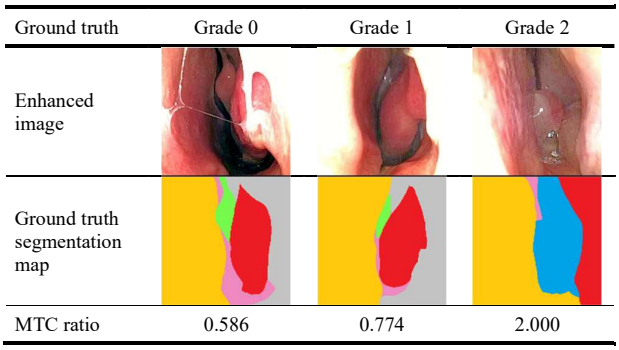

Fig. 4. Testing set images display various NO grades alongside their corresponding ground truth segmentation map and MTC ratio. Additionally, as color-coded in Fig. 3, polyp is represented in blue.

It is expected that we would achieve a grade 0 view bilaterally for participants without NO and NO patients after treatment. However, one nasal cavity may appear blocked (misdiagnosed as grade non-zero) due to the cyclic enlargement of erectile tissue and asymmetrical blood flow known as the nasal cycle. In such cases, considering the NO grading after applying nasal decongestant in the rule-based expert system enhances the accuracy of differential diagnosis of NO. Nasal decongestants play a vital role in restoring normal physiology to the nasal cavity and paranasal sinuses [15].

III. Methods

A. Data collection

We obtained full ethical approval from the London - City & East Research Ethics Committee, with reference number 15/LO/0187, and written consent was obtained from all participants. Seventy-three participants were recruited at Royal National ENT and Eastman Dental Hospitals, including

7 controls and 66 patients. The participants comprised 41 males and 32 females, with a mean age of 42.4 years (95% CI: 38.8-46.0). Controls aged 18 years or above had no history of rhinological conditions or symptoms. Based on their medical history and visual inspection, patients were diagnosed by an otolaryngologist with one or more of the following conditions: AR, CRS, or DNS. We recorded the nasal cavities using a Video Naso-Pharyngo-Laryngoscope (VNL9-CP) (Pentax Medical), and the videos were processed using a VIVIDEO Video Processor (CP-1000) (Pentax Medical). We also applied a nasal decongestant spray (xylometazoline hydrochloride 0.1% w/v) and repeated the recording 10 minutes later.

B. Datasets

An ENT consultant graded each video based on the NO grading described in TABLE I. This assessment served as the ground truth for both training and evaluation purposes. In the selection process, every 5th frame was chosen for each video, excluding frames located outside the nostril, featuring excessive nasal hair, or positioned too deeply. Since nasal asymmetry is minor in most cases, all images from the right nasal side were horizontally flipped. The dataset is utilized using low-light image enhancement with alpha blending.

For the training and evaluation of the model, every 4th participant was designated as part of the testing set, with the remaining participants assigned to the training set. The distribution of the NE-UCLH dataset is outlined in TABLE II.

TABLE II. THE BREAKDOWN OF THE NE-UCLH DATASET IN THE GRADE CLASSIFICATION OF NO.

Dataset	NO grading					Total
	0	Maybe 0	Maybe 1	1	2	
Frame-level	**1933**	**445**	**1153**	**2503**	**2938**	**8972**
Train + val	1560	245	759	1821	2330	6715
Test	373	200	394	682	608	2257
Video-level	**61**	**14**	**29**	**88**	**100**	**292**
Train + val	49	8	20	65	74	216
Test	12	6	9	23	26	76

C. Implementation details

The networks were implemented in MATLAB R2023b using the Machine Learning Toolbox and Image Processing Toolbox. We employed a multiclass error-correcting output codes (ECOC) model, trained using explainable features (MT pixel and MTC ratio) as inputs for SVM learners in a one-vs-one configuration [16].

To compare performance with deep learning networks, we pre-trained ResNet-50 [17] and the vision transformer (ViT)-tiny [18] models on the PASCAL VOC 2012 dataset [19]. Subsequently, fine-tuning was performed on our NE-UCLH dataset, where all images were resized as specified in TABLE III. The Adam optimization algorithm updated the network weights, initiating an initial learning rate of 0.0001, which decayed by 0.3 every 5 epochs. The cross-entropy loss function was employed as the optimization criterion. The gradient decay factor β_1 was set to 0.9, and the squared gradient decay factor β_2 was set to 0.999. Training took place on a single GPU (NVIDIA GeForce RTX 3050 Ti, Nvidia Corp., Santa Clara, CA, USA) for 30 epochs, utilizing a mini-batch size of 1. L2 normalization with a weight decay rate of 10^{-4} served for regularization. To prevent overfitting, early stopping was implemented if the validation loss did not improve for three consecutive epochs. The weight configuration from the final epoch was then chosen.

For validation and testing, a 5-fold cross-validation was executed on the training set. Data augmentation included rotations between -5° and +5° while avoiding vertical flipping and additional rotations as they were inappropriate for clinical images. All trained networks from cross-validation were used with the testing set for evaluation. Frames-per-second (fps) measurement was conducted on an RTX 3050 GPU with a mini-batch size of 1.

To assess the performance of NO grading, the evaluation employed weighted multiclass metrics, including accuracy, sensitivity, specificity, and F-1 score, as outlined in reference [20]. Additionally, the interrater agreement between the model predictions and the ground truth given by an ENT consultant for the grade classification of NO was compared using Cohen's kappa score. In general, a κ-score of zero or less indicates no agreement, while higher scores fall into the following categories: none to slight ($\kappa \leq 0.20$), fair ($0.21 \leq \kappa \leq 0.40$), moderate ($0.41 \leq \kappa \leq 0.60$), substantial ($0.61 \leq \kappa \leq 0.80$), and almost perfect agreement ($0.81 \leq \kappa \leq 1.00$) [21].

IV. RESULTS AND DISCUSSION

TABLE III. THE MEAN PERFORMANCE METRICS (IN % EXCEPT FOR Κ SCORE) FOR THE GRADE CLASSIFICATION OF NO. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Method	Crop size	Accuracy	Sensitivity	Specificity	F-1 score	κ score
Validation set						
SVM [16]	400×400	79.6	60.8	84.3	60.3	0.46
ResNet-50 [17]	224×224	85.0	69.1	89.7	69.4	0.59
ViT-tiny [18]	384×384	**94.8**	**89.5**	**96.5**	**89.4**	**0.86**
Testing set						
SVM [16]	400×400	69.6	35.3	78.0	31.8	0.14
ResNet-50 [17]	224×224	66.7	27.8	76.9	25.9	0.05
ViT-tiny [18]	384×384	**71.1**	**35.4**	**80.8**	**33.1**	**0.16**

TABLE III. compares the performance of different methods in grade classification of NO. On the validation set, ViT-tiny achieved the highest scores for all evaluation metrics, demonstrating almost perfect interrater agreement ($\kappa = 0.86 \pm 0.12$) with the grade given by an ENT consultant. It also achieved the highest specificity (96.5%), making it the preferred method for a grade classification of NO when minimizing false positives is crucial to avoid unnecessary invasive procedures or treatments. ResNet-50 and our method, based on SVM and explainable features, had the second and the lowest performance, respectively.

However, testing set results indicate that the performance of all models did not transfer as well compared to the validation set. Specifically, while ViT-tiny still achieved the highest scores for all evaluation metrics, it only showed none or slight agreement ($\kappa = 0.16 \pm 0.06$) with the grade given by an ENT consultant. Additionally, our SVM-based method outperformed ResNet-50 for all evaluation metrics, with an interrater agreement $\kappa = 0.14 \pm 0.01$, demonstrating no significant difference in κ compared to ViT-tiny ($p = 0.43$) and suggesting that our model is stable and less likely to overfit.

The current poor generalization indicates that the consistent ground truth NO grading may not be suitable for frame-based classification due to variations in MT visibility caused by different scope angles (see Fig. 5 for an example). Since the current set of explainable features used in SVM is

static, the significant variations in MT visibility further render the method unsuitable.

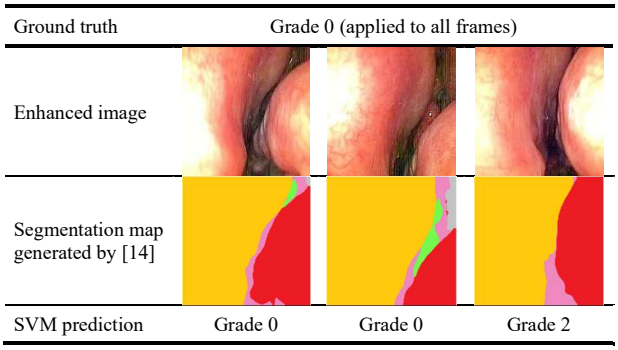

Fig. 5. Example images of the anterior nasal cavity from the testing set, captured from the same participant. The ground truth is consistent across all frames but yields different predictions due to variations in MT visibility. The head of the MT can be seen clearly (Grade 0: middle) or the MT is partially obscured (Grade 2/3: left, Grade 1: right) caused by different scope angles.

Based on these insights, along with incorporating more clinically validated explainable features, two potential strategies could be explored: 1) automatically selecting a keyframe from a video, or 2) revising the ground truth to allow for varied NO grades throughout the video. Subsequently, observe the performance of various algorithms for obtaining video-level classification based on frame-level data, such as using minimum grading, averaging and rounding to the nearest grading, or employing voting methods. Specifically for method 2), determine the optimal number of frames required to achieve a suitable interrater agreement with manual NO grading given by an ENT consultant.

TABLE IV. COMPARISON OF THE NUMBER OF PARAMETERS AND INFERENCE TIME OF DIFFERENT MACHINE LEARNING NETWORKS

Method	#params	fps
SVM [16]	**< 0.1M**	**7114**
ResNet-50 [17]	25.6M	124
ViT-tiny [18]	5.7M	27

Our SVM-based method provides some advantages in interpretability, despite lower overall performance. It utilizes two clinically validated, explainable features extracted from segmentation maps, providing a highly interpretable and transparent model. Additionally, as shown in TABLE IV. , our method offers a lightweight model and considerably shorter inference time compared to deep learning networks. Ensembling this model after the fine-tuned Mask2Former network used for anatomical segmentation, which reported 47M parameters and an inference speed of 6.28 fps in our previous publication [14], enables real-time segmentation and corresponding grade classification of NO.

V. CONCLUSION

This paper introduces an objective grade classification of NO based on explainable features - the size of the MT and MTC ratio derived from a semantic segmentation map from anterior nasal cavity images. The classification aids in differentiating common NO causes (AR, CRS, and DNS). Ultimately, we aim to use NO grading and polyp appearance pre- and post-nasal decongestant as inputs for a rule-based expert system in differential diagnosis of NO. Thus, the system could be translational to primary care settings, utilizing digital anterior rhinoscopy for ENT examination.

REFERENCES

[1] A. Valero et al., "Position paper on nasal obstruction: Evaluation and treatment," Journal of Investigational Allergology and Clinical Immunology, vol. 28, no. 2. ESMON Publicidad S.A., pp. 67–90, 2018. doi: 10.18176/jiaci.0232.

[2] Y. Maru and Y. Gupta, "Nasal Endoscopy Versus Other Diagnostic Tools in Sinonasal Diseases," Indian Journal of Otolaryngology and Head and Neck Surgery, vol. 68, no. 2, pp. 202–206, Jun. 2016, doi: 10.1007/s12070-014-0762-y.

[3] "Transforming elective care services ear, nose and throat (ENT) Learning from the Elective Care Development Collaborative."

[4] D. W. Hsu and J. D. Suh, "Anatomy and Physiology of Nasal Obstruction," Otolaryngologic Clinics of North America, vol. 51, no. 5. W.B. Saunders, pp. 853–865, Oct. 01, 2018. doi: 10.1016/j.otc.2018.05.001.

[5] J. Teixeira, V. Certal, E. T. Chang, and M. Camacho, "Nasal Septal Deviations: A Systematic Review of Classification Systems," Plast Surg Int, vol. 2016, pp. 1–8, Jan. 2016, doi: 10.1155/2016/7089123.

[6] W. J. Fokkens et al., "European Position Paper on Rhinosinusitis and Nasal Polyps 2020."

[7] G. K. Ziade et al., "Reliability Assessment of the Endoscopic Examination in Patients with Allergic Rhinitis," Allergy and Rhinology, vol. 7, no. 3, 2016, doi: 10.2500/ar.2016.7.0176.

[8] C. H. Li, A. Kaura, C. Tan, K. L. Whitcroft, T. S. Leung, and P. Andrews, "Diagnosing nasal obstruction and its common causes using the nasal acoustic device: A pilot study," Laryngoscope Investig Otolaryngol, vol. 5, no. 5, pp. 796–806, Oct. 2020, doi: 10.1002/lio2.445.

[9] B. Ay, C. Turker, E. Emre, K. Ay, and G. Aydin, "Automated classification of nasal polyps in endoscopy video-frames using handcrafted and CNN features," Comput Biol Med, vol. 147, Aug. 2022, doi: 10.1016/j.compbiomed.2022.105725.

[10] B. Patel, J. S. Virk, P. S. Randhawa, and P. J. Andrews, "The internal nasal valve: a validated grading system and operative guide," European Archives of Oto-Rhino-Laryngology, vol. 275, no. 11, pp. 2739–2744, Nov. 2018, doi: 10.1007/s00405-018-5142-x.

[11] G. J. Tsao, N. Fijalkowski, and S. P. Most, "Validation of a Grading System for Lateral Nasal Wall Insufficiency," Allergy & Rhinology, vol. 4, no. 2, p. ar.2013.4.0054, Jan. 2013, doi: 10.2500/ar.2013.4.0054.

[12] M. Camacho et al., "Inferior turbinate classification system, grades 1 to 4: Development and validation study," Laryngoscope, vol. 125, no. 2, pp. 296–302, Feb. 2015, doi: 10.1002/lary.24923.

[13] N. Phoommanee, V. Acharya, A. Navaratnam, T. S. Leung, and P. J. Andrews, "Rule-based expert system for the differential diagnosis of nasal obstruction using nasal endoscopy," Unpublished.

[14] N. Phoommanee, P. J. Andrews, and T. S. Leung, "Segmentation of endoscopy images of anterior nasal cavity using deep learning," SPIE-Intl Soc Optical Eng., in press.

[15] E. Wong, N. Deboever, J. Chong, N. Sritharan, and N. Singh, "Isolated Topical Decongestion of the Nasal Septum and Swell Body Is Effective in Improving Nasal Airflow," Am J Rhinol Allergy, vol. 34, no. 3, pp. 417–421, May 2020, doi: 10.1177/1945892420902913.

[16] A. Rocha and S. K. Goldenstein, "Multiclass from binary: Expanding One-versus-all, one-versus-one and ECOC-based approaches," IEEE Trans Neural Netw Learn Syst, vol. 25, no. 2, pp. 289–302, Feb. 2014, doi: 10.1109/TNNLS.2013.2274735.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.03385

[18] A. Dosovitskiy et al., "An image is worth 16x16 words: transformers for image recognition at scale," [Online]. Available: https://github.com/

[19] M. Everingham and J. Winn, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Development Kit," 2012.

[20] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: an Overview," Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.05756

[21] Anthony J. Viera and Joanne M. Garrett, "Understanding Interobserver Agreement: The Kappa Statistic," Fam Med, vol. 37, no. 5, pp. 360–363, 2005.