

Standardisation of an AI-based vocal fold assessment tool on a recurrent respiratory papillomatosis model

Mikolaj Buchwald^{1*}, Piotr Nogal^{2*}, Jan Nowak¹, Szymon Kupinski¹, Wojciech Andrzejewski¹, Juliusz Pukacki¹, Joanna Jackowska², Hanna Klimza³, Cezary Mazurek¹, Alberto Paderno^{4,5}, Cesare Piazza^{6,7}, Małgorzata Wierzbicka⁸

¹ Network Services Department, Poznan Supercomputing and Networking Center, Polish Academy of Sciences, Poznan, Poland; ² Department of Otolaryngology, Head and Neck Surgery, Poznan University of Medical Sciences, Poznan, Poland; ³ Regional Specialist Hospital Wroclaw, Research & Development Centre, Wroclaw, Poland; ⁴ IRCCS Humanitas Research Hospital, Rozzano (Milan), Italy; ⁵ Department of Biomedical Sciences, Humanitas University, Rozzano (Milan), Italy; ⁶ Unit of Otorhinolaryngology – Head and Neck Surgery, ASST – Spedali Civili of Brescia, Brescia, Italy; ⁷ Department of Surgical and Medical Specialties, Radiological Sciences, and Public Health, University of Brescia, School of Medicine, Brescia, Italy; ⁸ Wroclaw University of Science and Technology, Wroclaw, Poland

*MB and PN contributed equally to the work.

Received: January 1, 2024

Accepted: September 2, 2024

Correspondence

Mikolaj Buchwald

E-mail: mikolaj.buchwald@gmail.com

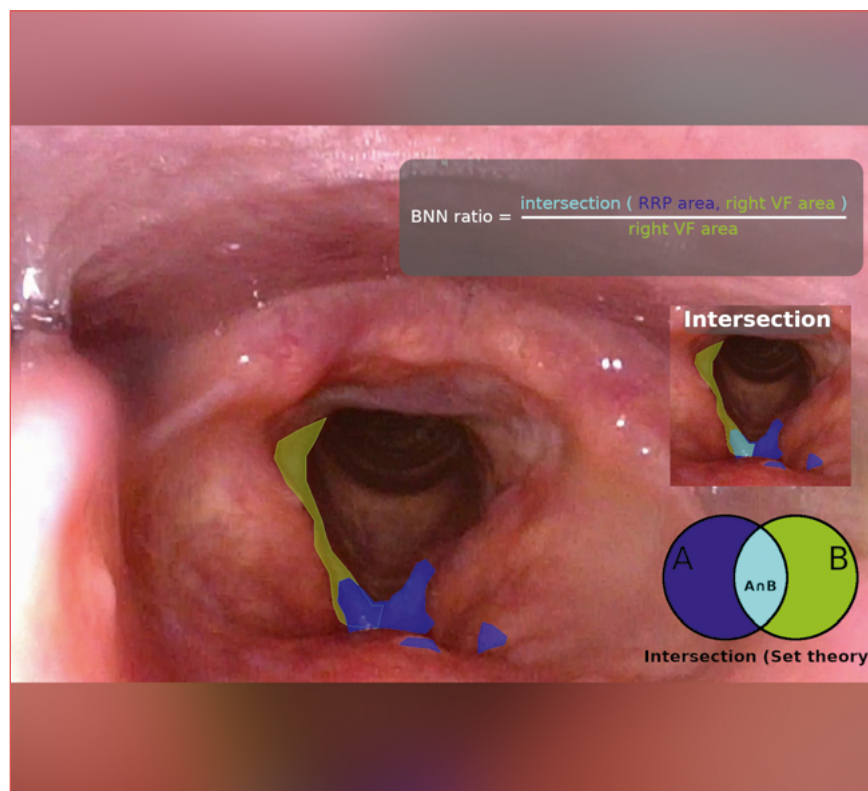
How to cite this article: Buchwald M, Nogal P, Nowak J, et al. Standardisation of an AI-based vocal fold assessment tool on a recurrent respiratory papillomatosis model. Acta Otorhinolaryngol Ital 2025;45:244-251. <https://doi.org/10.14639/0392-100X-N2896>

© Società Italiana di Otorinolaringoiatria e Chirurgia Cervico-Facciale



OPEN ACCESS

This is an open access article distributed in accordance with the CC-BY-NC-ND (Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International) license. The article can be used by giving appropriate credit and mentioning the license, but only for non-commercial purposes and only in the original version. For further information: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>



Cover figure. Endoluminal annotation of the laryngeal pathology in a frame extracted from videos obtained during white light laryngoscopy, showing the intersection of the vocal fold area of interest (green) and the recurrent respiratory papillomatosis area of interest (blue).

Summary

Objective. The assessment of extension of papilloma growth in recurrent respiratory papillomatosis (RRP) on vocal folds can be performed quantitatively utilising artificial intelligence (AI).

Methods. This study evaluated the efficacy of an AI-based annotation system, Glottis Coverage - Artificial Intelligence and Deep learning (GC-AID) in 4 patients to assess affected mucosa in white light (WL) and narrow band imaging modalities as a case-study for future applications.

Results. In healthy larynges, the mean difference between areas of the right and left vocal folds was minimal (2.6%). For patient # 4, following treatment, RRP coverage in WL decreased from 69.5% to 42.6%. A similar improvement was observed for patient # 1, while no significant benefits were noted for patients # 2 and # 3.

Conclusions. The extent of RRP was precisely measured with GC-AID before and after treatment. Obtaining objective, quantitative results was possible with frame extraction and annotation using the system described herein.

Key words: larynx, papillomatosis, NBI, artificial intelligence, deep learning

Introduction

Recurrent respiratory papillomatosis (RRP) is characterised by the proliferation of benign squamous cell papillomas within the aerodigestive tract^{1,2}. The disease is caused by human papilloma virus (HPV) types 6 and 11, and is associated with a high risk of recurrence after surgical resection^{1,2}. An effective treatment to definitively cure the disease does not currently exist. Surgery mechanically removes the lesions, but does not prevent disease recurrence because the surrounding mucosa is infected. Patients require frequent procedures causing irreversible changes in vocal fold (VF) structure and, as a consequence, in glottic function and phonation. Attempts to introduce new adjuvant treatments have been made for many years, as shown in multiple studies³⁻⁶. One of the key challenges when it comes to quantifying the effectiveness of RRP treatment, based on laser photocoagulation and/or medication, is the automation of the procedure and objective estimation of the papilloma extent. For instance, systematic research on intralesional drug administration and the exact measurement of the influence of these drugs remains scarce in the literature. Even the recent Benedict and Derkay's state-of-the-art article on RRP treatment did not address the issue of automation and accuracy of estimation of papilloma growth on VFs⁷, possibly due to its focus on selecting appropriate treatment measures, with quantitative estimate as the gold standard. Common quantitative parameters, including the number of surgeries at follow-up, Derkay or severity score, and remission status, are not rapid enough, suffer to a certain degree of subjectivity, and require extensive training to properly estimate RRP status. However, patient outcomes after treatment should be measured as accurately and objectively as possible *in situ* during clinical evaluation. Despite the many tools and scales available to measure the severity of RRP, it is still difficult to precisely determine changes in the area covered by the diseased tissue before and after treatment.

Another challenge lies in relating the RRP coverage area to the total surface area of the anatomical site, given the significant interpersonal differences in larynx dimensions, volume, and internal surface area. Although there are some technologies available to take measurements of laryngeal anatomy and pathologies⁸, a definitive model for *in vivo* morphometric studies of the VFs has not been established (despite recent intriguing research in machine learning applied to VFs⁹). Measurements have been conducted using contemporary imaging¹⁰, fresh specimens after total laryngectomy¹¹, and on cadavers¹². While ranges of VF length are well-known¹³⁻¹⁵, this parameter does not capture the surface area, which is essential to describe pathologies that spread superficially along the epithelium.

The absence of standardised, objective methods to compare the extent of RRP before and during follow-up visits hinders precise determination of the treatment effects within 1% accuracy. We hypothesise that a machine learning (ML) system for segmenting videoendoscopic frames can objectively quantify the RRP-affected VF area in relation to the total VF area.

The aim of the present study was to investigate the efficacy of an artificial intelligence (AI) based annotation system to objectively assess the affected mucosa area in laryngeal RRP using white light (WL) and narrow band imaging (NBI) to create the Glottis Coverage – Artificial Intelligence and Deep learning (GC-AID) tool.

Materials and methods

Data from 4 patients with RRP were analysed. Initial presentation at the study institution was between July 2022 and January 2023.

Inclusion criteria were as follows: age > 18 years, bilateral glottic extension of RRP, and written consent for study participation. All patients were treated by transoral laser microsurgery with carbon dioxide laser (CO₂ TOLMS), and the papillomas were surgically removed from one vocal fold. Additionally, bevacizumab was employed as an adjuvant therapy based on clinical findings and patient preference. It was injected bilaterally into both the true and false VFs as required, with doses ranging from 25 to 100 mg. Therefore, one VF was treated both surgically and with bevacizumab, with the contralateral VF acting as control with bevacizumab injection alone. A therapeutic goal was to compare the effect of bevacizumab alone to surgery and bevacizumab. Videoendoscopic follow-up by flexible endoscope with video segmentation, frame annotation, and lesion measurement was conducted for both VFs, including the operated and the contralateral one with persistent RRP. The control group comprised the videolaryngoscopy of 4 healthy patients, performed as a part of routine ENT examination in patients prepared for otosurgical procedures. VFs (right and left) from 80 frames of the 4 healthy larynges were annotated and analysed.

Additionally, a total of 80 frames from 4 RRP patients, in the non-surgically treated vocal fold, only injected with bevacizumab, were analysed. Follow-up examinations for the 4 patients were at 55, 29, 85, and 90 days, respectively (mean 64 days; standard deviation 28.4).

The main goal was to standardise the AI/ML-based method for the precise assessment of VF area and its involvement by RRP. Four consecutive steps were undertaken to achieve the study objectives:

1. Assess and describe the tool to measure the size of the surface of a small anatomical region. The 4 healthy larynges constituted the standard for gross anatomy of the VFs;
2. Measure the precise VF area in 4 RRP patients for both right and left VFs;
3. Calculate the exact RRP-covered VF area in 4 RRP patients for both right and left VFs;
4. Estimate the percentage of RRP coverage during initial examination and in follow-up using 2 videolaryngoscopic modalities: (A) WL and (B) NBI. This involves determining the precise RRP-affected VF area and comparing coverage areas measured in WL and NBI.

The entire frame selection and annotation process was conducted using our custom-built platform, which enabled cloud-based access to all frames stored in the central database, automated selection of informative frames from endoscopic videos, and seamless in-platform annotation capabilities. Using our system for segmenting videoendoscopic frames, we delineated the papilloma area involved before the first operation, and during follow-up visits in both WL and NBI (Cover figure and Fig. 1). We then calculated the average coverage from multiple frames per patient, both during the initial visit and at follow-up (Fig. 2).

In detail, the delineation procedure utilised in the current study consisted of: (1) a clinical expert selecting informative frames from a set chosen by the AI – at least 10 to be annotated (resident annotating first, then a specialist correcting, if needed); (2) the papilloma annotation process, i.e., delineating where the papilloma is visible in the given frame; (3) annotating the left and the right VFs separately (Cover figure, Figs. 1-2). Subsequently, the data were processed by data scientists and AI specialists. This way, the information from the images annotated by the medical experts could be used to calculate an objective, quantitative measure of the coverage of a VF by papillomas.

The procedure to calculate the coverage of the VF was as follows: (1) the areas delineating the VFs were imported to a Jupyter Notebook instance running on a Python 3.10 kernel; (2) the area delineating the RRP was also imported to a Jupyter Notebook; (3) separately, for the left and right VFs, the intersection (a set-theoretic procedure performed on the 2 areas of interest) of either the left, or right VF, and the RRP area was established.

The process of calculating the area change ratio of a VF papilloma before the first intervention and at the follow-up visit is shown in Figure 3. Indices calculated independently for the left and right VFs allow to determine how the ratio of the papilloma area delineated at the frames from the

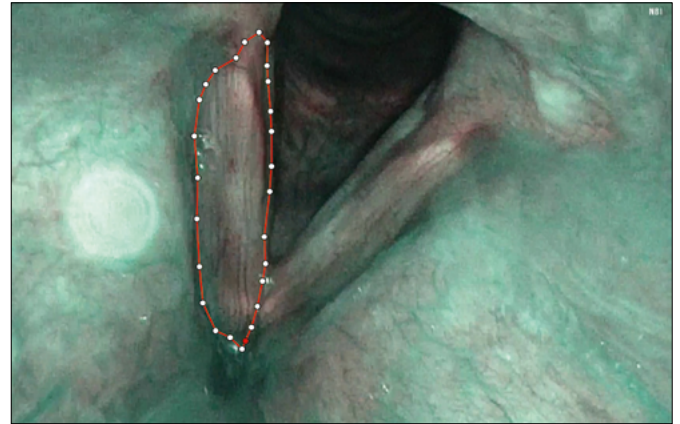


Figure 1. Endoluminal annotation of a vocal fold in narrow band imaging modality of a frame extracted from videos obtained during laryngoscopy. The outline of annotation for the right vocal fold is shown.

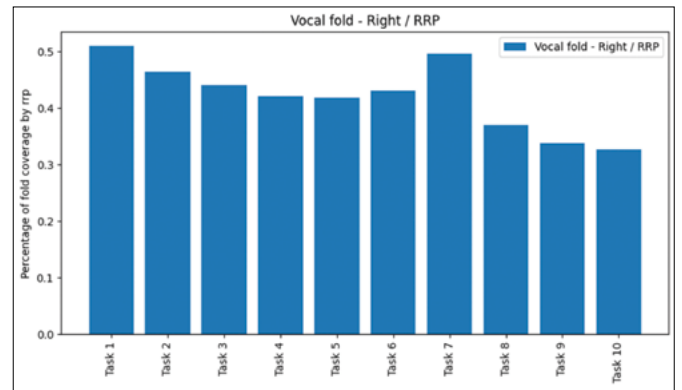


Figure 2. Enlisting the tasks (frames) on the example of one patient, for right vocal fold.

follow-up video compared to the state at the initial examination. In order to determine the ratio of change in the area of the papilloma, the following indicators are calculated: $BNNv_z$ – the ratio of the coverage by the intersection of the papilloma area ($RRParea_z$) and the v vocal fold area ($vVFarea_z$) to the entire v vocal fold area ($vVFarea_z$), where v is the left or right vocal fold and $_z$ is the annotated video frame (task) taken from the patient’s examination. The formula for the $BNNv_z$ indicator is therefore:

$$BNNv_z = \frac{RRParea_z \cap vVFarea_z}{vVFarea_z}$$

$RRPv$ – the average for all $BNNv$ values for a given exami-

nation. The formula for the RRPv indicator, where z is the number of all tasks for the patient, is:

$$RRPv = \frac{\sum_{i=1}^z BNNv_i}{z}$$

DIFv – difference of the RRPv_y value and the RRPv₁ value, where ₁ is the RRPv value for the first study and _y is the RRPv value for the last study. The formula for the DIFv indicator is as follows:

$$DIFv = RRPv_y - RRPv_1$$

AVGv – an average change in papilloma area for all studies for all patients. A positive value indicates an increase in area while a negative value indicates a decrease in the area covered by the papilloma on the vocal fold. The formula for the AVGv indicator, where x is the number of patients, is:

$$AVGv = \frac{\sum_{i=1}^x DIFv_i}{x}$$

STDEVv – a standard deviation for all studies for all patients. The indicator allows to see how much the data for individual patients differ from each other. The smaller the value of the indicator, the more similar the data are to each other, which means that the changes in the size of the papilloma area are similar for all patients. Formula for the STDEVv indicator, where x is the number of all patients, is presented below:

$$STDEVv = \sqrt{\frac{\sum_{i=1}^x (DIFv_i - AVGv)^2}{x}}$$

To determine the degree change in papilloma area, a dataset labelled by a clinical expert was needed. This dataset included VF and papilloma areas for frames obtained from patient examinations before each operation and at each follow-up visit. To ensure high quality results, a video for a single patient had multiple labelled frames (tasks) from different camera angles.

The algorithm calculated coverage ratio values separately for the left and right VFs for each patient. It began with

the first patient, processed data from the oldest video, and calculated the BNN ratio for each marked frame (Stage 1 – Calculate BNN ratio). After calculating BNN values for each frame, it proceeded to calculate the RRP value for each study (Stage 2 – Calculate RRP ratio), followed by the DIF ratio for each patient (Stage 3 – Calculate difference). Finally, after processing data for all patients, the algorithm calculated the final AVG (average) and STDEV (standard deviation) values (Fig. 3).

The outcome measure of the study was the validation of the method to precisely delineate the VF area in healthy and diseased larynges, estimating frame-to-frame variability and standard error measure (i.e., the technical validity of the system).

The primary outcome measure constituted the percentage change of the affected vocal fold area by RRP in 2 videolaryngoscopic modalities: (A) WL, and (B) NBI, focusing on differences in coverage areas at 2 time points.

The following descriptive statistics were utilised to report VF area in healthy larynx: average percentage of the frame (an abstract, relative unit representing the 2-dimensional space of the frame), standard deviation of the area, difference in the averages for the left and right VF, and the relative difference between VFs (i.e., the ratio between the absolute value of the difference, and the higher of the 2 averages – for the left and right VF, in percent points). In the case of patients with RRP, the statistics were: percent coverage of the left VF (at 2 time points: initial evaluation, and follow-up), percent coverage of the right VF (also 2 time points), and the differences before and after intervention (surgery and bevacizumab treatment, either left or right, depending on a patient, in percent points). Moreover, for the VFs covered with RRP, the statistics were reported separately for WL and NBI modalities.

Results

The results are presented in 2 parts: assessment of VF area in healthy larynges and evaluation of VF coverage by RRP before and during follow-up visits.

The first part showed a precise assessment of the VF area in videolaryngoscopic images in 4 healthy larynges. The gross anatomy of the VF, with upper, lower surface, and free edge was annotated. Table I presents the differences in relative areas between the left and right VFs in healthy larynges for both WL and NBI modalities. VF relative size – percent of the frame from the videolaryngoscopic video covered with the delineation for the right and the left VF, for both WL and NBI, is shown.

One of the goals of the current study is the standardisation of the delineation process for gross anatomy (e.g., VF area) and pathological tissues, for the purpose of the application of AI/ML methods in videolaryngoscopy. Table I shows that the average relative differences between left and right VF areas were minimal, ranging from 0.2% to 4.9%, indicating a low margin of error suitable for AI/ML annotations. On average, the difference of the areas between the left and right VFs was 2.6%.

In the second part, the percentage of the VF coverage with RRP before (i.e., during initial examination) and during the follow-up visit was calculated. The data were presented separately for WL and NBI in Tables II and III, respectively. The negative values in the Difference column indicate a decrease in the RRP-affected area.

In the WL modality, the number of frames per patient was:

- Patient # 1: 4 initial examination, 4 follow-up;
- Patient # 2: 7 initial, 2 follow-up;
- Patient # 3: 2 initial, 1 follow-up;
- Patient # 4: 3 initial, 7 follow-up.

For the NBI modality, the frames per patient were:

- Patient # 1: 7 initial examination, 7 follow-up;
- Patient # 2: 0 initial (unavailable), 6 follow-up;
- Patient # 3: 9 initial, 11 follow-up;
- Patient # 4: 5 initial, 5 follow-up.

The WL and NBI modalities were also compared in order to establish whether a different segmentation outcome would be observed, depending on the type of imaging used. Unfortunately, there were no NBI frames available in the recording for patient # 2 for the initial visit, hence comparison for this patient was not possible to be performed. The absolute differences between the 2 imaging modalities were on average 12.7% (Tab. IV; $p = 0.9$).

In summary, in healthy larynges, the average relative differences between the left and right VF areas were minimal, ranging from < 0.2% to < 5%, or approximately 2.6% on

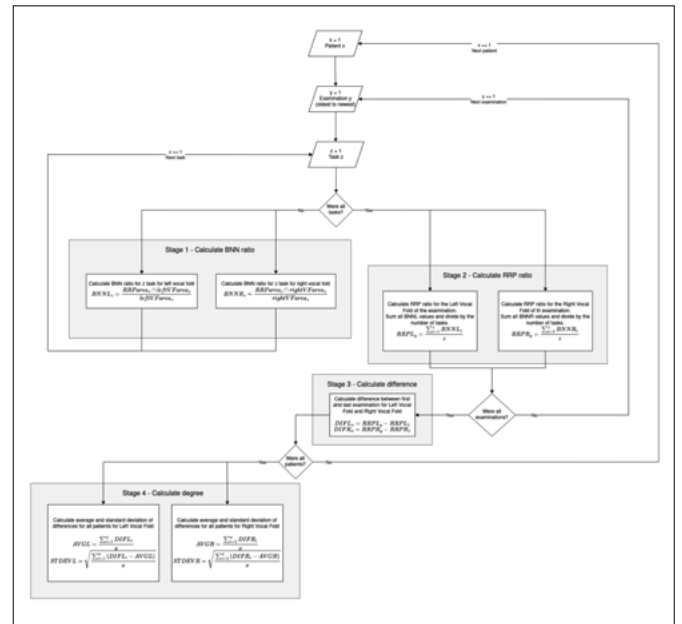


Figure 3. Diagram of the algorithm that calculates the degree of change of the papilloma coverage area located on the vocal folds.

average. This level of precision allows for accurate determination of the difference in RRP coverage, and the annotation tool provides a standardised means of quantitatively assessing treatment effects in videolaryngoscopic images. Using the NBI-ML tool, we were able to precisely quantify changes observed during follow-up visit in RRP coverage of the VFs. For instance, in patient # 4, the percentage of left VF coverage by RRP in WL decreased from 69.5% before treatment to 42.6% after treatment, which represents the most significant difference for the treated VF. However, no changes were observed for patients # 2 and # 3.

Table I. Healthy larynx vocal folds. The differences between the relative areas of the left and right vocal folds.

Patient # (healthy larynx group)	Left VF		Right VF		Difference in the average percentage on the frame	Relative difference between vocal folds
	Average percentage of the frame	Standard deviation	Average percentage of the frame	Standard deviation		
1	2.057	1.866	2.062	1.478	-0.005	0.2%
2	1.593	0.603	1.614	0.543	-0.021	1.3%
3	1.307	0.579	1.361	0.607	-0.054	4.1%
4	1.308	0.698	1.373	0.689	-0.065	4.9%

VF: vocal fold. For example results of one of the patients, i.e., patient # 2 with a healthy larynx, see Figure 2.

Table II. Percentage of vocal fold RRP coverage per patient in white light imaging.

Patient #	% LVF covered initial visit	% LVF covered follow-up visit	% RVF covered initial visit	% RVF covered follow-up visit	Bevacizumab-only fold	Difference in the injected fold
1	52.21	56.08	33.08	17.07	right	-16
2	19.37	48.42	40.62	61.60	right	20.97
3	13.43	6.64	6.45	13.69	right	7.24
4	69.53	42.63	47.19	19.83	left	-26.9

LVF: left vocal fold; RVF: right vocal fold.

Table III. Percentage of the vocal fold RRP coverage per patient in narrow band imaging.

Patient #	% LVF covered initial visit	% LVF covered follow-up visit	% RVF covered initial visit	% RVF covered follow-up visit	Bevacizumab-only fold	Difference in the injected fold
1	50.65	58.51	36.95	32.65	right	-4.31
2	NA	41.14	NA	42.25	right	NA
3	34	6.67	19.04	19.67	right	0.63
4	71.66	62.6	42.75	16.33	left	-9.05

LVF: left vocal fold; RVF: right vocal fold.; NA: not available.

Table IV. Comparison of the treatment influence for white light (WL) and narrow band imaging (NBI) modalities.

Patient #	WL	NBI	Difference
1	-16	-4.31	11.69
2	20.97	NA	NA
3	7.24	0.63	6.61
4	-26.9	-9.05	17.85

WL: white light imaging; NBI: narrow band imaging; NA: not available.

Discussion

This study presents 3 interrelated aspects that align with recent advancements in the field of laryngology. The first aspect are the relatively new and poorly documented quantitative methods to assess the extent of RRP on the VFs. We focused our evaluation on the VF approached with this treatment model instead of surgical removal of RRP to avoid interference from mechanical manipulation of the healing process, which may distort surface assessment. The second aspect, resulting from the first point, involves the possibility of precise evaluation of the treatment outcome. The third but crucial aspect is to create a GC-AID tool dedicated to precise, unambiguous and repeatable assessment of the VF surface area. In this context, we utilised the RRP model with progression or regression patterns, which are measurable using the NBI-ML system.

AI/ML tool for precise assessment of the vocal fold area

The deep learning-based approach has been established as a more general and flexible tool to capture glottic edges during any phonatory events and has been shown to be a promising technique in high-speed videolaryngoscopy (HSV) data processing¹⁶⁻¹⁸. Complementary techniques include active contour modelling to detect glottal edges on automatically extracted kymographs at different VF intersections in HSV data¹⁹. A glottic flow model is also available that provides better prediction of the pressure distribution and flow rate in the idealised VF geometry²⁰. A different problem is the differentiation of morphology of pathologies on the examined surfaces¹⁹. Although DL-based computer-aided diagnosis systems to distinguish laryngeal neoplasms from benign conditions have been presented²¹, the efficacy of using AI to identify the other morphologically different laryngeal lesions still remains unknown²². As for utilising AI models with different types of laryngoscopes, one study showed that such systems can provide a reliable auxiliary tool to screen for laryngeal carcinoma and may improve and standardise the diagnostic capacity of laryngologists²³. However, the accuracy of these systems in assessing the nuances of the field occupied by papillomas is still too unpredictable. In the current study, the average relative difference in the areas of the left vs right VF was only 2.6%. Moreover, utilising this method we found differences from 7% to 27% between the initial and follow-up visits. Hence, it can be

concluded that the margin of error of the tool described is relatively low, and thus it is sufficient for the annotations for AI/ML purposes. The difference of 2.6% in the area of left vs right VF can be also explained by the type of the endoscope utilised in the current study (flexible endoscope), which can slightly distort the image. As the primary goal of the current work was to establish the margin of error for segmentation method, and this was shown to be relatively low, further research will determine the segmentation error for different types of endoscopes (e.g., rigid vs flexible endoscopes).

NBI-ML tool for videolaryngoscopic frame segmentation

The implementation of the GC-AID method created by the authors is a web application (NBI-ML system) that allows to upload and process videos from the videolaryngoscope. In this system, the AI model selects the most informative frames from the video with a visible glottis, adequate definition, and no blurring, overexposure, or limited visibility. With this tool, medical experts can conveniently perform segmentations and share results, which ultimately reduces the time required for data synchronisation, annotation standardisation, and VF segmentation in both healthy larynges and those affected by RRP (Cover figure and Fig. 1). Moreover, the entire procedure is highly standardised, meaning that experts performing annotation have exactly the same set of available classes of annotation/segmentation to choose from. In the current article, a very precise description of an algorithm to calculate the VF coverage with papilloma was provided. In a broader context, such an approach can also be applied to other types of laryngeal pathologies covering different parts of the glottis. We believe that creating standardised procedures and tools for laryngeal videoendoscopic frame annotation and segmentation is one of the key challenges when creating large-scale databases of consistent information on laryngeal pathologies for the development of AI models in laryngology.

Effectiveness of intralesional application of bevacizumab

The original approach utilising the GC-AID tool, with the segmentation of frames from endoscopic videos and delineation of the affected surfaces, allow for assessment of the papilloma growth before (initial visit) and during follow-up. However, the primary endpoint of our study was not to assess the effect of treatment (laser surgery and bevacizumab) on the change in area occupied by RRP after topical application of the drug. The initial extension and the percentage changes in the coverage of the VF with the disease process were only a convenient model to develop a surface

annotation ML tool. The average 5% reduction of the area of the fold covered with RRP for the 4 patients examined definitely does not reflect the real data on the therapeutic effect. This score rather highlights the large differences in interpersonal sensitivity to RRP treatment. The therapeutic response in 2 patients was clinically relevant, while in 2 others the lesions actually progressed. Undoubtedly, the small number of patients and the single administration of bevacizumab limit the reliability of properly assessing the drug's effect.

Conclusions

In summary, we present a reliable and effective AI-based method to quantify RRP extent over VFs, which proved useful for the case-study of follow-up for patients with RRP. The process of delineating specific tissue types by experienced clinicians is a vital step in developing ML models for diagnostic decision support. This study represents a step forward in creating standardised annotation tools and processes for laryngeal videoendoscopic frames, paving the way for more comprehensive databases to support the advancement of AI models in the field of laryngology.

Acknowledgements

The authors would like to thank Prof. Wioletta Pietruszewska, from the Department of Otolaryngology, Medical University of Lodz, Lodz, Poland, for supporting the idea of a system for annotation and segmentation of the laryngeal pathologies. Although she did not author the current publication, she actively supported our multi-centric attempt to develop a clinical decision support system for endolaryngoscopy, that is based on standardised medical knowledge.

Conflict of interest statement

The authors declare no conflict of interest.

Funding

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Author contributions

PN was responsible of implementing the conception and design; PN, HK, JJ, and AP were responsible for data acquisition, while MW and CP supervised data acquisition; MB, PN, and JN prepared the standard for calculating vocal fold coverage with RRP, under JJ's, MW's, CP's, and CM's supervision; MB and JN supported data acquisition process by providing technological means of uploading, securely storing, and processing the data, with SzK supervising the

process; PN and JJ were responsible for data annotation/segmentation, with MW's guidance and support. JN implemented the coverage calculation algorithms, based in the initial code prototype by WA, and WA supervised the whole implementation process; SzK, JP and CM supervised the overall technological design and calculation implementation; MB and PN drafted the initial version of the manuscript, MB worked on the manuscript, with PN, MW, CP, and AP supervising and validating the writing process. Moreover, AP provided critical insights for the design and the main text of the manuscript. All authors read and approved the final version of the manuscript to be published. All authors hereby agree to be accountable for all aspects of the work in ensuring the questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethical consideration

This study was approved by the Committee on Bioethics, Poznan University of Medical Sciences (resolution no. 690/22).

The research was conducted ethically, with all study procedures being performed in accordance with the requirements of the World Medical Association's Declaration of Helsinki.

Written informed consent was obtained from each participant/patient for study participation and data publication.

References

- San Giorgi MRM, Van Den Heuvel ER, Tjon Pian Gi REA, et al. Age of onset of recurrent respiratory papillomatosis: a distribution analysis. *Clin Otolaryngol* 2016;41:448-453. <https://doi.org/10.1111/coa.12565>
- Wierzbicka M, Jackowska J, Bartochowska A, et al. Effectiveness of cidofovir intralesional treatment in recurrent respiratory papillomatosis. *Eur Arch Otorhinolaryngol* 2011;268:1305-1311. <https://doi.org/10.1007/s00405-011-1599-6>
- Zeitels SM, Barbu AM, Landau-Zemer T, et al. Local injection of Bevacizumab (Avastin) and angiolytic KTP laser treatment of recurrent respiratory papillomatosis of the vocal folds: a prospective study. *Ann Otol Rhinol Laryngol* 2011;120:627-634. <https://doi.org/10.1177/000348941112001001>
- Rogers, DJ, Ojha, S, Maurer, R, et al. Use of adjuvant intralesional bevacizumab for aggressive respiratory papillomatosis in children. *JAMA Otolaryngol Head Neck Surg* 2013;139:496-501. <https://doi.org/10.1001/jamaoto.2013.1810>
- Best SR, Friedman AD, Landau-Zemer T, et al. Safety and dosing of bevacizumab (avastin) for the treatment of recurrent respiratory papillomatosis. *Ann Otol Rhinol Laryngol* 2012;121:587-593. <https://doi.org/10.1177/000348941212100905>
- Hall SR, Thiriveedi M, Yandrapalli U, et al. Sublesional Bevacizumab injection for recurrent respiratory papillomatosis: evaluation of utility in a typical clinical practice. *Ann Otol Rhinol Laryngol* 2021;130:1164-1170. <https://doi.org/10.1177/0003489421998215>
- Benedict JJ, Derkay CS. Recurrent respiratory papillomatosis: a 2020 perspective. *Laryngoscope Invest Otolaryngol* 2021;6:340-345. <https://doi.org/10.1002/lio2.545>
- Neitsch M, Horn IS, Hofe, M, et al. Integrated multipoint-laser endoscopic airway measurements by transoral approach. *Biomed Res Int* 2016;2016:6838697. <https://doi.org/10.1155/2016/6838697>
- Tran BA, Dao TTP, Dung HDQ, et al. Support of deep learning to classify vocal fold images in flexible laryngoscopy. *Am J Otolaryngol* 2023;44:103800. <https://doi.org/10.1016/j.amjoto.2023>
- Wu L, Zhang Z. A parametric vocal fold model based on magnetic resonance imaging. *J Acoust Soc Am* 2016;140:EL159. <https://doi.org/10.1121/1.4959599>
- Mobashir MK, Mohamed AERS, Quriba AS, et al. Linear measurements of vocal folds and laryngeal dimensions in freshly excised human larynges. *J Voice* 2018;32:525-528. <https://doi.org/10.1016/j.jvoice.2017.08.024>
- Clarós P, Sobolewska AZ, Doménech-Clarós A, et al. CT-based morphometric analysis of professional opera singers' vocal folds. *J Voice* 2019;33:583.e1-583.e8. <https://doi.org/10.1016/j.jvoice.2018.02.010>
- Schuberth S, Hoppe U, Döllinger M, et al. High-precision measurement of the vocal fold length and vibratory amplitudes. *Laryngoscope* 2002;112:1043-1049. <https://doi.org/10.1097/00005537-200206000-00020>
- Su MC, Yeh TH, Tan CT, et al. Measurement of adult vocal fold length. *J Laryngol Otol* 2002;116:447-449. <https://doi.org/10.1258/0022215021911257>
- Nogal P, Buchwald M, Staśkiewicz M, et al. Endoluminal larynx anatomy model – towards facilitating deep learning and defining standards for medical images evaluation with artificial intelligence algorithms. *Otolaryngol Pol* 2022;76:37-45. <https://doi.org/10.5604/01.3001.0015.9501>
- Fehling MK, Grosch F, Schuster ME, et al. Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep convolutional LSTM network. *PLoS One* 2020;15:E0227791. <https://doi.org/10.1371/journal.pone.0227791>
- Kist AM, Zilker J, Gómez P, et al. Rethinking glottal midline detection. *Sci Rep* 2020;10:20723. <https://doi.org/10.1038/s41598-020-77216-6>
- Kist AM, Gómez P, Dubrovskiy D, et al. A deep learning enhanced novel software tool for laryngeal dynamics analysis. *J Speech Lang Hear Res* 2021;64:1889-1903. https://doi.org/10.1044/2021_JSLHR-20-00498
- Yousef AM, Deliyiski DD, Zacharias SRC, et al. A deep learning approach for quantifying vocal fold dynamics during connected speech using laryngeal high-speed videoendoscopy. *J Speech Lang Hear Res* 2022;65:2098-2113. https://doi.org/10.1044/2022_JSLHR-21-00540
- Li Z, Chen Y, Chang S, et al. A one-dimensional flow model enhanced by machine learning for simulation of vocal fold vibration. *J Acoust Soc Am* 2021;149:1712. <https://doi.org/10.1121/10.0003561>
- Ren J, Jing X, Wang J, et al. Automatic recognition of laryngoscopic images using a deep-learning technique. *Laryngoscope* 2020;130:E686-E693. <https://doi.org/10.1002/lary.28539>
- Wróbel MJ, Lewandowski B. Artificial Intelligence, or just statistics done different? *Otolaryngol Pol* 2022;76:1-5. <https://doi.org/10.5604/01.3001.0016.0540>
- Yan P, Li S, Zhou Z, et al. Automated detection of glottic laryngeal carcinoma in laryngoscopic images from a multicentre database using a convolutional neural network. *Clin Otolaryngol* 2023;48:436-441. <https://doi.org/10.1111/coa.14029>