INDEX COPERNICUS INTERNATIONAL

ORIGINAL ARTICLE

Endoluminal larynx anatomy model – towards facilitating deep learning and defining standards for medical images evaluation with artificial intelligence algorithms

Piotr Nogal[1], Mikołaj Buchwald[2], Michalina Staśkiewicz[1], Szymon Kupiński[2], Juliusz Pukacki[2], Cezary Mazurek[2], Joanna Jackowska[1], Małgorzata Wierzbicka[3]

[1]Department of Otolaryngology, Head and Neck Surgery, Poznan University of Medical Sciences, Poznan, Poland; Head: prof. Małgorzata Wierzbicka MD PhD
[2]Network Services Department, Poznan Supercomputing and Networking Center, Polish Academy of Sciences, Poznan, Poland; Head: prof. Marek Figlerowicz MD PhD
[3]Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; Head: prof. Michał Witt MD PhD

ABSTRACT:

Aim: The concept of "anatomy in a new perspective" presented in this article has never been proposed before and is a breakthrough in numerical approaches to larynx modeling for endoluminal imaging methods. This new anatomical model can be represented numerically for the segmentation purposes, thus allowing it to be implemented as part of the development of the deep learning models.

Aim: Presented approach is characterized by strict delineation and gradual changes (28 segments), thus allowing the model to be adapted to the needs of videolaryngoscopy segmentation for Machine Learning workflows.

Material and methods: Analysis of the literature concering the subject of larynx segmentation. Preparing a model of the larynx for the need of Artificial Intelligence needs.

Results: The approach described in this article assumes defining a process for data acquisition, integration, and segmentation (labeling), for the needs of a new branch of knowledge: digital medicine and digital diagnosis support expert systems. The first and crucial step of such a process is creating a digital model of the larynx, which has to be then validated utilizing multiple clinical, as well as technical metrics. The model will form the basis for further artificial intelligence (AI) requirements, and it may also contribute to the development of translational medicine.

Conclusions: The model will form the basis for further artificial intelligence (AI) requirements, and it may also contribute to the development of translational medicine.

KEYWORDS: diseases of the salivary glands, parotid gland, salivary gland neoplasm, salivary gland, sublingual gland, submandibular gland

ABBREVIATIONS

AI – artificial intelligence
BAGLS – Benchmark of Automatic Glottis Segmentation
CNN – convolutional neural networks
CT – computed tomography
DL – deep learning
ELAM.AI – Endoluminal Larynx Anatomy Model for Artificial Intelligence
HRCT – high resolution computer tomography
LSTM – long short-term memory
MIMICS – Materialise Interactive Medical Image Control System
ML – machine learning
MRI – magnetic resonance imaging
NBI – narrow band imaging
RRP – recurrent respiratory papillomatosis
VT – Vocal Tract
WL – white light
WLE – white-light endoscopy
XAI – explainable artificial intelligence

INTRODUCTION

The field of contemporary laryngeal endoscopy has to face some cutting-edge problems in the rapidly changing world, both from the scientific, as well as medical practice perspective. Machine learning (ML), or more specifically – deep learning (DL), and artificial intelligence (AI) methodologies are being introduced to modern medicine in order to support the diagnostic and treatment processes in many areas of potential applications by utilizing a range of informatics tools [1–7].

Normal anatomy seems to be a very stable branch of science across nearly all medical fields. Nevertheless, new diagnostic tools enforce a modern approach to canonical anatomy [8, 9]. For example, the endoluminal image of the upper respiratory tract is described in general anatomy but has never been subdivided to meet the requirements of the more thorough videolaryngoscopy diagnosis standardization. Probably it was one of the reasons why the canonical anatomy of the larynx inlet has never been adapted for deep learning purposes [3]. The mucosal elements that make up the gross anatomy have not been previously fragmented in line with the view from the endoluminal camera either.

The segmentation of the larynx box anatomy for deep learning has been initiated first for the purposes of improving imaging methods and planning radiation fields in oncology [10–12]. The segmentation of the upper airway has primarily been performed on three-dimensional radiological models, taking into account all tissues that build up the organ: cartilage skeleton, muscles, vessels, nerves and structures adjacent to the laryngeal box from the outside. However, the mucosa covering the inside of the larynx and airways, formed into folds and creases, is a flat structure, poorly visible in radiological images. Mucosal lesions' endoluminal assessment with videolaryngoscopy, canonically in the white light (WL), has recently been supplemented by narrow band imaging (NBI) modalities [13]. Importantly, the learning curve of videoendoscopy is long, and errors in laryngeal assessment are fraught with negative consequences for the patients [14, 15]. Nevertheless, the endoluminal larynx evaluation is currently the basis of the clinical examination, as all imaging methods are a valuable supplement to diagnostics.

One of the most promising ideas for supporting the NBI learning process, as well as increasing diagnostic accuracy in the WL endoluminal larynx videoendoscopy, is to integrate AI methods into medical procedures [3, 16]. Importantly, in order to be successfully implemented for the purpose of being utilized in endoscopy scenarios, AI methodologies require a preliminary step of creating the model with an expert-annotated ground-truth input dataset. In fact, one of the crucial challenges, from the engineering perspective, in the effort of developing such models, is providing medical domain experts with proper tools facilitating data segmentation and labeling [17]. Although the unsupervised approach, i.e., the procedure which does not require expert input, is also viable, this paper will cover exclusively the supervised deep learning process, because it is believed that the decisions of the artificial agent should be possible to be traced back to the data from which the model knowledge has originated from (see explainable artificial intelligence, XAI; [18]). In other words, although it is possible to create a diagnostic support system, based only on labels – i.e., names of the particular pathologies visible in the images – without the exact segmentation (delineation) of the pathologies, the latter approach almost always provides more reliable results, and requires fewer cases to be segmented by the specialists (see, e.g., [19]).

The challenges mentioned above have been addressed by an interdisciplinary team in the framework of the ELAM.AI project (Endoluminal Larynx Anatomy Model for Artificial Intelligence). The project's research hypotheses assume that the best way to increase the quality and usability of AI in larynx videolaryngoscopy diagnostics is to develop a standardized way of evaluating medical images. The standardization process, apart from improving the medical procedure itself, will also be a significant step towards preparing the data as an input to the ML models. And the other way around, once the validated AI models will be in place, they can also serve to provide a better way to support the whole videoendocscopy diagnostic process. The creation of an endoluminal larynx digital model is the first and crucial step in standardizing the mucosa description, as it will help to constitute one, consistent workflow for training, testing and the deployment of an expert support system in the form of an AI algorithm. Thus, the aim of the project is to create the digital model of the endoluminal anatomy of the larynx for deep learning, based on the segmenting of the laryngeal inlet by means of images' registration and assignment of digital markings/signature.

MATERIALS AND METHODS

The method of improving the process of standardization presented in this article assumes binding the approaches from two fields: medical science (laryngology), and computer science (machine learning approach, ML). In order to fulfill this task, below we present definitions of the terms commonly used in medical-informatics interdisciplinary projects:

- Segmentation is the process of localizing and delineating anatomical structures and pathological lesions in a medical image;
- Labeling is assigning particular categories (labels) to the images. Sometimes the terms "segmentation" and "labeling" are used interchangeably, or, e.g., it is said that the image has been segmented – and then labels have been assigned to particular segments. However, for the sake of clarity, labeling is defined throughout this paper as a more general type of procedure than segmentation (unless explicitly stated otherwise). I.e., if there is a frame from videoendoscopy video, on which a particular type of pathology is visible, one can label that frame, by saying, e.g.: "In this frame, we can see a polyp", as opposed to segmenting the pathology – by delineating the pixels in which the polyp is visible on that frame;
- Atlas-based methods. Atlas is a collection of medical images. The atlas contains models of normal anatomy. Such a model is pre-marked with segmentation and serves as the basis of the algorithm. When a new image is to be segmented, it is fitted to the atlas. Deep learning is done by identifying important anatomical structures in the atlas and superimposing pixel intensity or other image transformation methods onto the anatomy model pattern. Once aligned, the segmentation algorithm generates a new image segmentation;
- Digital model – as compared to the descriptive model, the digital model is characterized with much more detailed segmentation of the areas of the given organ. In this kind of model, the canonical (healthy) anatomy is divided into separate areas which enforce more accurate in vivo diagnosis, and more accurate models to be created by the means of the computer science tools (i.e., a more accurate and/or precise classification to be performed by the artificial intelligence algorithms);
- Database – a set of medical data, including: medical images (e.g., videoendoscopy images, radiological images), as well as (anonymized) patient ID, clinical data, test results, medical history, etc.;
- Supervised segmentation algorithms/Unsupervised Image segmentation – Supervised image segmentation assumes that prior to the model learning phase some representative images have to be annotated (labeled or segmented) by the domain expert (e.g., medical doctor). On the other hand, with the unsupervised approach, it is possible to segment the image automatically with computer vision algorithms, e.g., by detecting some repetitive, specific textural pattern. Nevertheless, the unsupervised segmentation cannot provide a meaningful, semantic name for the segmented

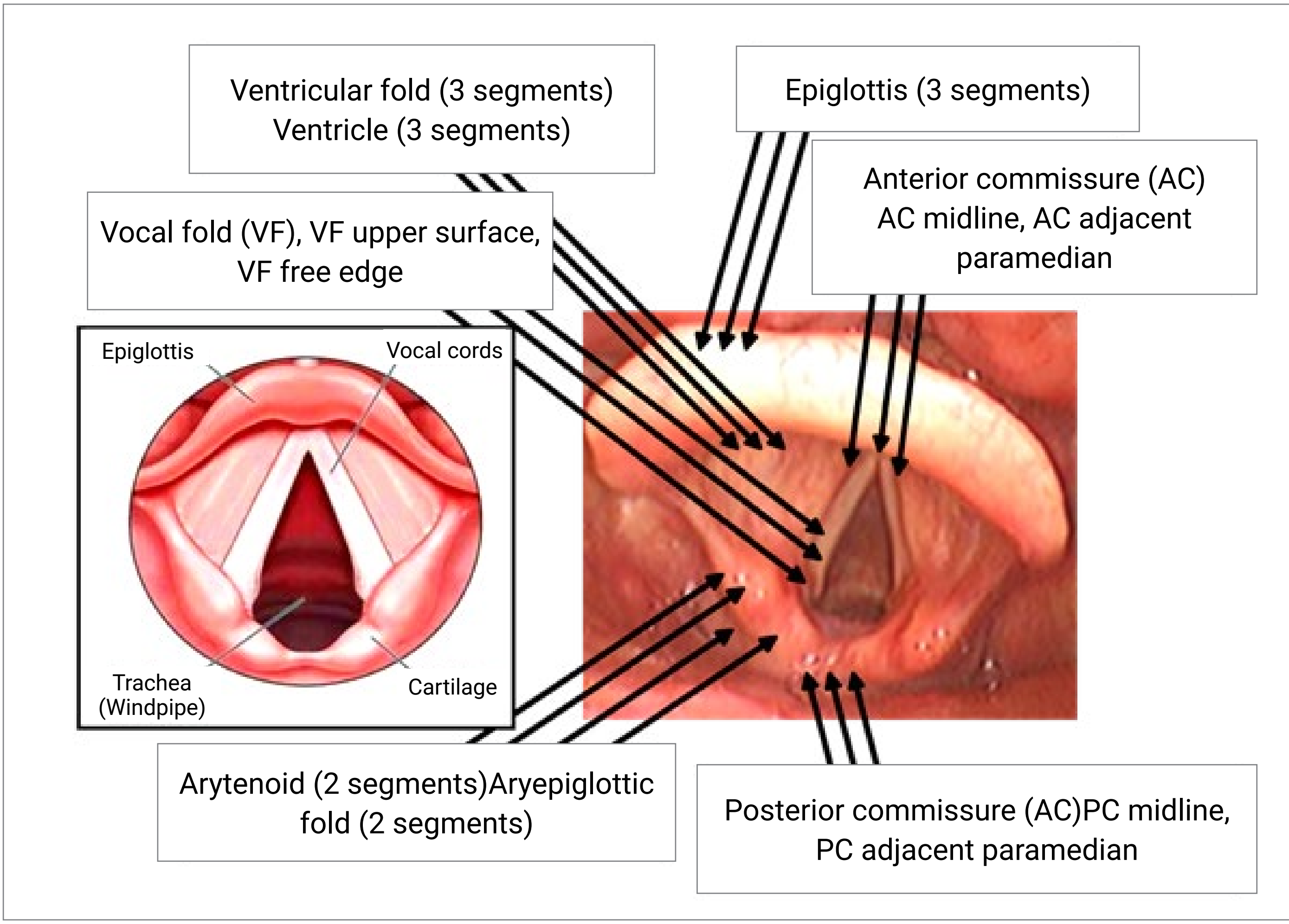

Fig. 1. Endoscopy of the larynx – endoluminal view and gross anatomy picture of the healthy organ. The descriptions of the larynx endoluminal view are presented in normal font for structures with anatomical names, and in italics for structures distinguished additionally, during videoendoscopy in the segmentation process, for the purposes of ML for the larynx images. The picture on the right has relatively low quality, but it is a real-life case, in which the endoscopic images or videos that can be recorded (even when using the top-grade equipment available on the market) often are exactly like that – not in the spatial quality we are used to in modern displays. In fact, the quality of the images is one of the key challenges when creating an endoluminal digital model for practical applications.

portions of the images – this, ultimately, is a role of an expert who is able to interpret the results;

- Deep learning is a type of machine learning methodology, in which more complex, multiple-layer/multiple-step processing takes place – as opposed to, e.g., a single-layer feedforward artificial neural network, in which only a limited amount of representations can be encoded. Interestingly, the human brain is an example of a biological system that utilizes deep learning principles, as there are multiple layers of processing: from receiving e.g., video (or audio) stimuli, through decoding colours and shapes of the objects, up to naming the objects (or people) that are perceived;
- Neural networks (artificial neural networks) – it is a paradigm in computer science (in the domain machine learning/artificial intelligence) by which (almost) all functions can be expressed as a set of input and output nodes, with a finite number of nodes in between them, and with inter-connections between these nodes (weights). The defined set of the connections, their weights (strengths), and bias of nodes, can represent virtually any algorithm (and thus any operation/function) ranging from simple arithmetical operations, such as addition and subtraction, to the tasks such as image classification, weather forecasting, stock-trading decisions, etc.;
- AI in the endoluminal laryngoscopy – it concerns issues such as: the ability of AI to identify healthy tissue, estimating the accuracy of AI in assessing laryngeal lesions, differentiating between benign lesions from potentially malignant changes, and diagnostic performance of AI using NBI and WLE images.

The methodology of this work consists of four successive tasks: (1) comparison of gross anatomy description and fiberscopic image, (2) presentation of existing models of the larynx, (3) presentation of organ segmentation models, and finally (4) model of the endoluminal anatomy of the normal larynx. Below each of these tasks is described in detail.

Comparison of gross anatomy description and fiberscopic image

Clinical and gross anatomy is based on a three-dimensional model and divides the larynx box into three levels: supraglottic, glottic, and subglottic regions. The mucosa covering the inside of the larynx and airways, formed into folds and creases, is a flat structure.

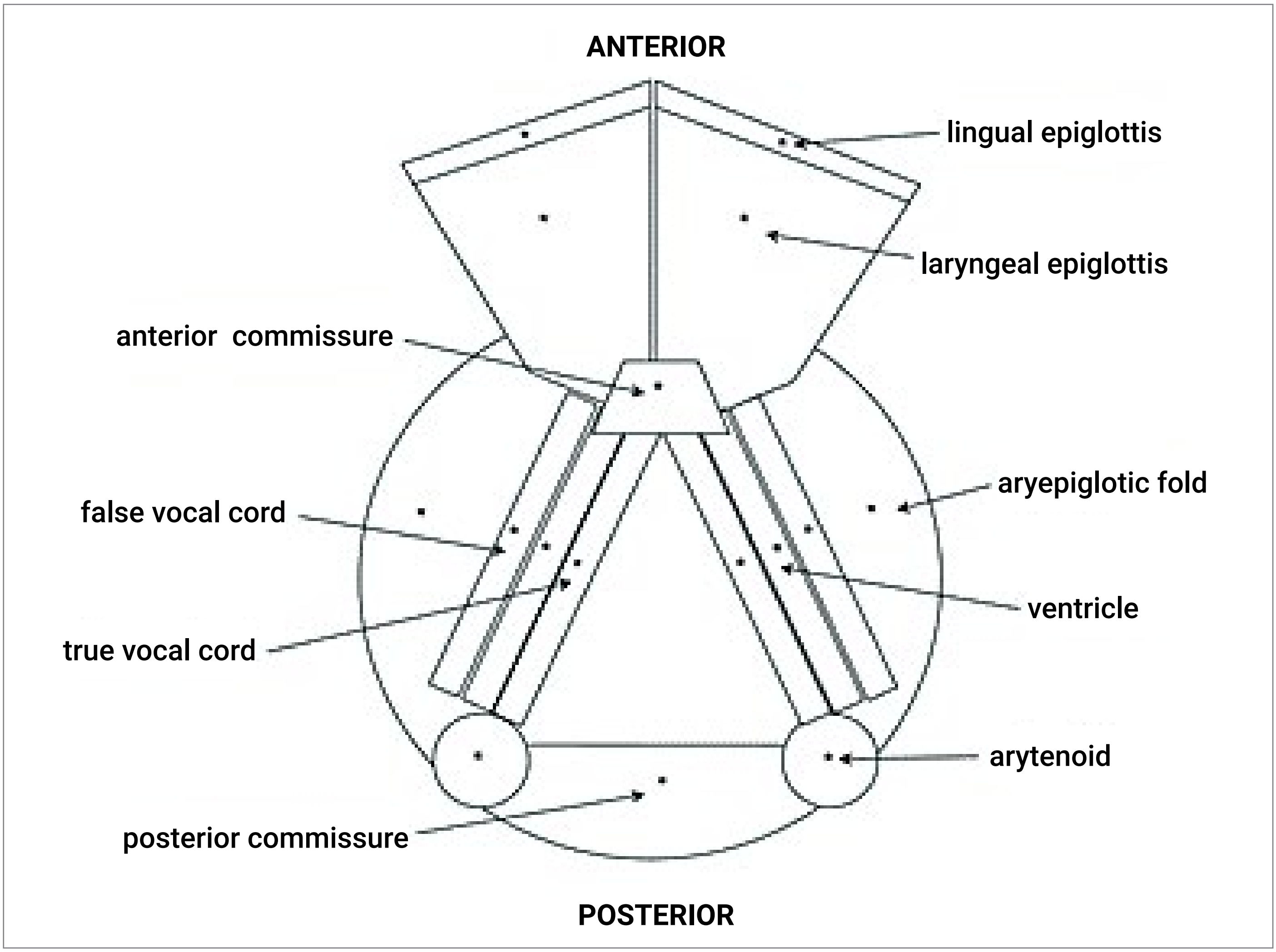

Fig. 2. Derkay's model of the larynx for papillomatosis severity evaluation (acc. to: demarcantonio M., Derkay C. (2013) Laryngeal Papillomatosis. In: Kountakis S.E. (eds) Encyclopaedia of Otolaryngology, Head and Neck Surgery. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23499-6_425).

Laryngeal endoscopy is a dynamic examination, and the human eye, armed with a magnifying glass or a camera, examines the interior of the organ from many angles and assesses it in many aspects: colour, vascularity, smooth surface, mobility of individual elements and their arrangement in relation to each other. The picture from videolaryngoscopy does not correspond to an analogous gross anatomy picture but what is more, both visualizations do not reflect the complexity of the interior of the larynx when viewed endoscopically. Detailed locations that do not have an exact equivalent in the descriptions of normal anatomy are: the anterior and posterior commissure in the midline and adjacent paramedian areas, division into sections (anterior, middle, posterior) the levels of the vestibular, vocal folds, and ventricles, distinguishing two surfaces in the structure of the arytenoid, ary-epiglottic folds, and epiglottis segmentation (Fig. 1.).

Presentation of existing models of the larynx

Upper airway has never been divided into subregions for purposes of endoluminal examinations but there is only one schedule available, created for recurrent respiratory papillomatosis (RRP) (Fig. 2.). The schedule presented for RRP has never been detailed, segmented or adapted for ML purposes. It was also not taken into account and used for other pathologies, for instance to precisely qualify the location of precancerous conditions and low-stage cancers.

Presentation of organ segmentation models

Segmentation of images of the upper airways and larynx has a two-way approach: the first was based on a three-dimensional shape assessed in imaging studies and the other, in endoluminal endoscopic evaluation. Automated detection and segmentation methods used for the larynx on computed tomography images, namely, the description of deep learning-based auto-segmentation models for swallowing and chewing structures in CT [20] and in MRI-based 3D segmentation of the VT at a functional state have been presented [21]. The study by Storck et al. [22] concludes that HRCT provides excellent data for three-dimensional visualization of the laryngeal anatomy, and the combined technology of HRCT and MIMICS is useful to study the biomechanics on 3D images and for preoperative planning of laryngeal framework surgery. We do not discuss these issues more extensively as they are not the subject of this study.

Tab. I. Larynx segmentation models based on literature review.

NO.	PUBLICATION	IMAGING MODALITY	PURPOSE	DATA	PARAMETER
1.	Iyer et al., 2020	CT	Segmentation of swallowing and chewing structures	24 CT scans	Dose-volume histogram (DVH) metrics
2.	Chen et al., 2020	CT	Masticatory muscles' auto-segmentation	56 CT images	Dice similarity coefficient (DSC), recall, precision, Hausdorff distance (HD), HD95, and mean surface distance (MSD)
3.	Vrtovec et al., 2020	CT	Review: Auto-segmentation for radiotherapy planning	78 publications on auto-segmentation of oars in the H&N region (from 2008 to 2020)	Dice coefficient as the standard volumetric overlap metrics – looked for in the review
4.	Wong et al., 2019	CT	Assessment of deep learning-based auto-segmentation in clinical practice (radiotherapy planning)	60 radiotherapy planning CT scans	Dice similarity coefficient (DSC), and Hausdorff distance (HD)
5.	Korte et al., 2021	MRI	Automatic segmentation of the parotid glands, submandibular glands, and level II and level III lymph nodes	31 MRI auto-contouring (RT-MAC) scans	Dice similarity coefficient (DSC), Mean surface distance (MSD)
6.	Gómez et al., 2020	Videoendoscopy	A single structure segmentation – glottis	640 videos of healthy and disordered subjects	Intersection over union (IOU)
7.	Moccia et al., 2018	Videoendoscopy	Frame quality classification	720 images from 18 laryngoscopic videos	Classification recall
8.	Paderno et al., 2021	Videoendoscopy	Cancer segmentation	Lesions: 34 NBI OC videos, and 45 NBI OP videos	Dice similarity coefficient (DSC)
9.	Araujo et. al, 2019	Videoendoscopy	NBI frames' classification	33 NBI videos of 33 patients affected by SCC, 10 images selected from each video; a total of 330 images	Class-specific recall
10.	Cho et al., 2021	Videoendoscopy	Analysis of laryngeal pathologies' segmentation	4106 images of larynx pathologies	Stratified K-fold cross-validation
11.	Esmaeili et al., 2019	Contact Endoscopy + NBI	Analysis of vascular patterns	Data from 32 patients involving 1485 images.	Consistency of the vessel direction, corresponding to histogram of gradient direction (HGD) and rotational image averaging (RIA)

The idea of preparing a segmentation model for endoluminal larynx examination has been already highlighted in some aspects of larynx anatomy and function. However, previous authors have been dealing mostly with functional aspects of the larynx: vocal fold movement and larynx elevation in the deglutition act. Kuo et al. [23] prepared a segmentation method where a support vector machine was used to classify laryngeal lesions based on a decision tree. The concept included an assessment of the glottis. Laves, Bicker, and Luder assessed CNN-based methods for semantic segmentation of the larynx and concluded that they are applicable to endoscopic images of laryngeal soft tissue [24]. In 2020 an international collaboration of researchers from the USA and EU prepared a Benchmark of Automatic Glottis Segmentation (BAGLS) – a dataset of 640 recordings of high-speed videoendoscopy of healthy and disordered larynxes [25]. The dataset is publicly available and aims to allow an objective comparison of automatic segmentation methods and support the machine learning community. In 2019 Fehling et al. [26] proposed a fully automated segmentation of the glottis and the vibrating vocal fold tissue in high-speed video using a deep convolutional LSTM network. They evaluated the system and proved that automatic segmentation is as precise or even superior to manual assessment performed by experts. A very similar conclusion derives from another study analyzing the vibratory function of vocal folds, conducted by Lohscheller et al. [27]. Another project, an automated segmentation using Mask R-CNN for laryngeal elevation, prepared by Hyun Haeng, Bo Mi, Chenk-Kun aimed to prepare a quick-to-use neural network model for measuring laryngeal elevation quantitatively using anatomical structures auto-segmented by Mask region-based convolutional NN in videofluoroscopic swallowing study [28, 29] (Tab. I.).

Model of the numerical endoluminal anatomy of the normal larynx (digital endoluminal model).

Conceptualization of the digital endoluminal model of the larynx is presented in Fig. 3. The rules for segmenting and labeling (i.e., assigning names to particular segments of the larynx area) are based on the following four assumptions: (1) the labeling order (i.e., the order of assigning names to particular segments) follows the direction of inspection during routine videolaryngoscopy; (2) the number of labels given to certain regions is higher than to others due to their anatomical and functional importance and the incidence of pathology there; (3) the "density" of labels does not depend on the size of the area which we mark; and (4) additional labels are given to chosen locations (e.g. the free edge of the vocal fold) which will be of key importance for their function and applied assessment techniques (e.g., to assess stroboscopy findings, phonation gap). The final output, numerical endoluminal model of the larynx is shown in Fig. 4.

Importantly, the detailed locations included in the digital model are the exact mapping of the areas assessed in the routine videolaryngoscopic examination. According to the symmetry of the

Tab. II. Comparison of functionalities of the anatomical models of the larynx.

STRUCTURE GROSS ANATOMY	DESCRIPTION – ANATOMICAL REGIONS	MODEL OF THE LARYNX ACC. DERKAY	ENDOLUMINAL DIGITAL MODEL (28 SEGMENTS)
EPIGLOTTIS	A thin, leaf-like plate of elastic cartilage that projects obliquely upwards behind the tongue and hyoid body, and in front of the laryngeal inlet.	1. Lingual (left and right); 2. Laryngeal (left and right).	s1– free edge; s2 – middle part; s3 – petiole of epiglottis.
ANTERIOR COMMISSURE	The site where the vocal folds meet anteriorly; the region where fibers of the vocal ligament pass through the thyroid cartilage; the point at which the vocal ligaments attach to the thyroid cartilage known as Broyles' ligament – contains blood vessels and lymphatics	Anterior commissure	3 segments (midline – s23, adjacent paramedial regions – s24, s25)
FALSE VOCAL (VESTIBULAR) FOLD	Vestibular ligament representing the thickened lower border of the quadrangular membrane, fixed in front to the thyroid angle below the epiglottic cartilage and behind to the anterolateral surface of the arytenoid cartilage above its vocal process covered with mucosa,	False vocal folds (left and right)	s8, s9, s10
TRUE VOCAL FOLD	The free thickened upper edge of the conus elasticus forming the vocal ligament which stretches back on either side from the midlevel of the thyroid angle to the vocal processes of the arytenoids covered by mucosa	True vocal fold (left and right)	vocal fold upper surface: s17, s18, s19, vocal fold free edge: s20, s21, s22
VENTRICLE	A slit between the vestibular folds above and the vocal folds below	Ventricle (left and right)	s11, s12, s13
ARYTENOID	A pyramidal in shape cartilage covered with mucosa to which the vocal ligament is attached	Arytenoid (left and right)	s4 – inner surface; s5 – outer surface.
ARYEPIGLOTIC FOLD	Represents the free upper border of the quadrangular membrane; it borders the laryngeal inlet	Aryepiglottic fold (left and right)	s6, s7
POSTERIOR COMMISSURE	The transverse mucosal fold between the two arytenoids	Posterior commissure	3 segments* (midline – s26, adjacent paramedial regions – s27, s28)
SUBGLOTTIS	The lower portion of the larynx, extending 1 cm below the lateral margin of the ventricle where the superior surface of the true vocal fold ends down to the top of the trachea.	None	s14, s15, s16

* anterior commissure and posterior commissure as odd digits for unpaired segments

organ, 22 of the 28 segments were divided into: left (L), and right (R) symmetrical areas (segments no. 1–22). The anterior and posterior commissures are, as an exception from that general left-right symmetry rule, divided into three segments, each (segments no. 23, 24, 25; and 26, 27 28).

The comparison of three different approaches to the larynx anatomy: the gross anatomy descriptions, Derkay larynx diagram (Fig. 2.), and the digital model presented in the current paper, has been summarized in Tab. II. The most important difference between the newly proposed model and Derkay's approach is that the former model is digital, not a descriptive one. Moreover, the digital model provides a more detailed (denser) representation of the larynx – which is crucial from the point of view of the potential applications for the development of the AI-based models. Finally, the predominant feature of the digital model is that canonical anatomy regions are divided into additional, separate larynx areas. Descriptions/labels in digital models are short and precise, yet it clearly relates to the pattern.

The functionalities of the digital model of the larynx based on the segmentation are as follows: (1) Created model allows for an orderly grading system in the assessment of the larynx during videolaryngoscopy, and unambiguous registration of particular segments in the system. Such an organization of the organ allows for devising a more accurate statistical model for the population. Moreover, in the proposed model the larynx is divided into 28 dense segments, as compared to 6 segments in Derkay's/classic anatomical description models. Such a more gradual approach is more promising in terms of developing an accurate ML system based on the input from the standardized mucosa evaluation for diagnostic purposes; (2) The digital model of the larynx allows for an uneven spatial division of the organ with an emphasis on areas of special importance. This is particularly important from the point of view of asymmetries, and in the context of providing crucial-areas-oriented expert knowledge to the ML model; (3) The size of the organ does not influence the proposed segmentation scheme. In other words, the proposed digital larynx model can fit the needs of an entire population, thus allowing the comparison of larynx pathologies between patients, or groups of patients. From the perspective of the process of developing a deep learning model, assigning pathologies to particular segments of the larynx can help automatically narrow the search for pathologies in the whole images – thus elevating the need of, or decreasing the amount of the time needed from the expert for the manual segmentation process; (4) Numerical data is reproducible for a follow-up for an individual patient, and provides uniform criteria on when to perform comparisons between patients. The digital model, thanks to organ segmentation, allows for automatic system learning (in terms of localization), narrowing time resources needed for the whole process, and the margin of error.

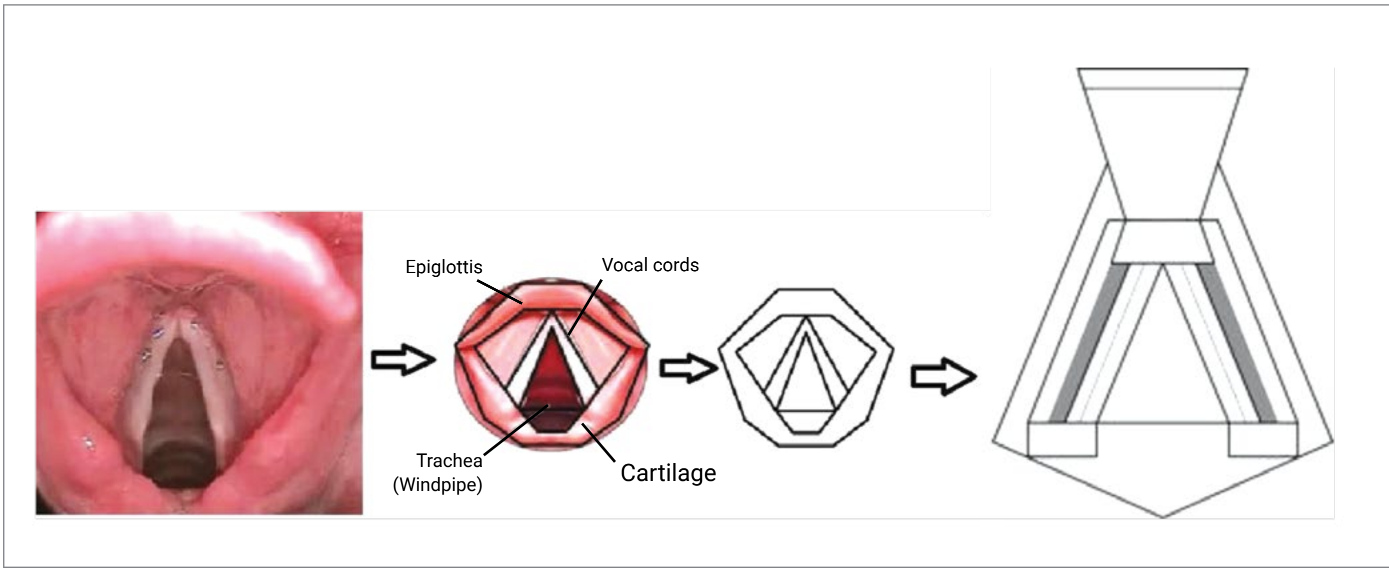

Fig. 3. Conceptualization of the geometrical and digital endoluminal model of the larynx.

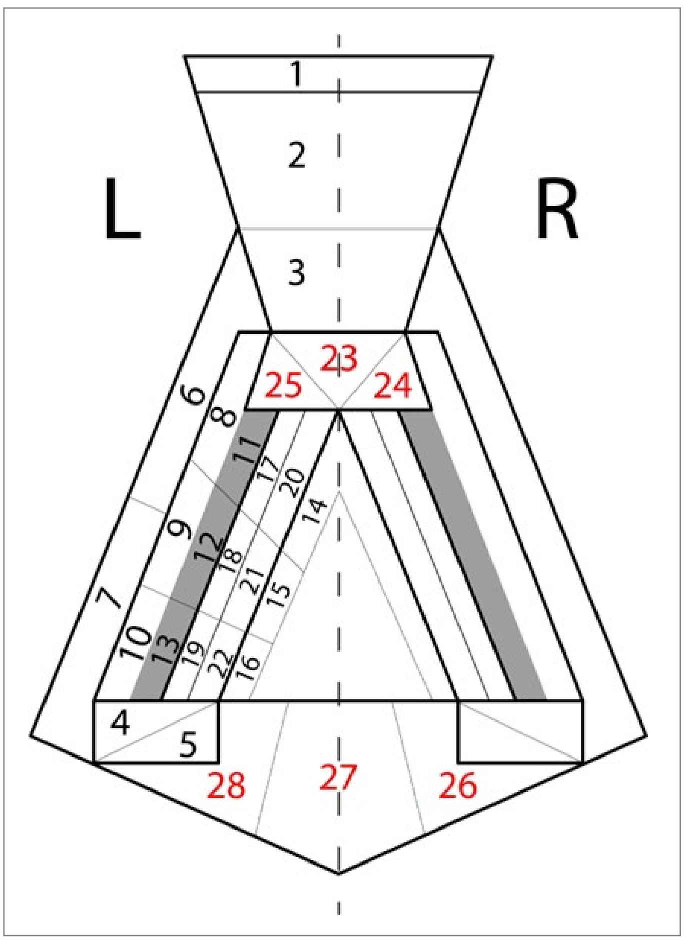

Fig. 4. Digital endoluminal model of the larynx. Segments (labels) have been assigned to: epiglottis: s1 (i.e., segment no. 1) – free edge, s2 – middle part, s3 – petiole of epiglottis; arytenoid: s4 – inner surface, s5 – outer surface; aryepiglottic fold: s6, s7; ventricular fold: division into sections (anterior, middle, posterior) s8, s9, s10; ventricle: division into sections (anterior, middle, posterior) s11, s12, s13; subglottis: division into sections (anterior, middle, posterior) s14, s15, s16; vocal fold upper surface: division into sections (anterior, middle, posterior) s17, s18, s19; vocal fold free edge: division into sections (anterior, middle, posterior) s20, s21, s22; anterior commissure: midline – s23, adjacent paramedial regions – s24, s25; posterior commissure: midline – s26, adjacent paramedial regions – s27, s28.

PLANNED CLINICAL VALIDATION OF THE MODEL

The creation of a digital model of the larynx is a preparation for the next research steps. Further stage of the project will be a clinical trial on a group of healthy people to validate the reliability and the external validity of the proposed model. Then, by disseminating and wide-spreading the validated model, we hope to extend our research attempt to other research groups and institutions, thus providing a multicentric approach to the standardization effort for creating a common data preparation and processing workflow for artificial intelligence-based support systems. The next and final step will be the creation, based on the multicentric dataset, the AI model which will encompass the knowledge on both: the normal larynx, as well as larynx with pathological changes.

DISCUSSION

In this article, the authors proposed the digital model of the gross anatomy of the larynx, which can serve as a uniform common basis for organ segmentation. ML and AI tools enter modern medicine and support the diagnostics and treatment processes. The new tools need a modern approach to canonical anatomy.

Endoluminal picture of the upper airway is described in gross anatomy but it has never been divided into subregions for the needs of videolaryngoscopy, nor has it been adopted for deep learning purposes. A new chapter is needed to complete the gross anatomy of the larynx to include and organize those elements that can be seen and assessed by video laryngoscopy and translated for ML.

There are few larynx segmentation models available, as presented in Tab. I. In head and neck oncology, all existing approaches to segmentation of the larynx are in their preliminary stages, and the majority of them are based on imaging studies. These studies showed a fragmentary approach to the anatomy, function and pathology of the larynx. None of them concerns the organ endoluminal anatomy as a whole. None of them is focused on the assessment of details of the mucosa in terms of the inflammatory, pre-neoplastic and early cancerogenesis phenomenon of all three levels of the larynx. Despite the significant conceptual and technical advancement of the above studies, none of them presents a digital model of the entire interior of the organ.

A digital model of normal anatomy is based on WL imaging; however, it could also assist in describing the NBI of the larynx, and would have a unique application there. Such a numerical system and protocol with the digital larynx atlas can be a useful tool for learning endoscopy methods. Using the digital model will help to minimize the differences between the specialist and the resident in terms of accuracy (possibly at the expense of the time of examination, but the overall accuracy will be higher for the resident when using the digital model, as compared to the routine protocol).

Digital model coherence validation should be performed in the context of the machine learning approach. Segment similarity analysis (segment coherence analysis), performed in silico, has to confirm that the model is consistent, and representative across multiple subjects. The computer analysis of the digital model, based on the data from healthy participants, should answer the question about the external validity of the model (its coherence within the studied group of participants).

Technical validation of the digital model in the context of deep neural networks has to be performed in order to quantitatively estimate model effectiveness on the sample of participant data. In accordance with the recommendations by Thiyagalingam et al. [33], the empirical data acquired from healthy participants will be used to perform the segmentation, and validate the database/theory-driven model.

Although general insights for endoluminal data segmentation and machine learning processing has been established (see, e.g., [30-32; 34-36]), a consistent, universal approach encapsulating the whole process – including the model validation step – is still missing. The work on the validation of the model, and the statistical analyses will involve creating a tool for data integration, segmentation (labeling) process, as wells as for obtaining model validation measures, such as: correlation, precision, recall, dice coefficient, F1, and accuracy [37–39].

Although the larynx digital model has undoubtful advantages, it also has its limitations and weak points. For instance, the larynx mucosa stays under the influence of many factors, such as: hormonal, metabolic changes, use of medication, or voice overuse. These factors should be also included in the process of obtaining the image data in association with the medical history of each participant (a step towards a personalized medicine approach).

CONCLUSION

To summarize, the concept of "anatomy in a new perspective" presented in this article has never been proposed before and is a breakthrough in numerical approaches to larynx modeling for endoluminal imaging methods. This new anatomical model can be represented numerically for the segmentation purposes, thus allowing it to be implemented as part of the development of the deep learning models. Although the endoluminal view of the upper airway has previously been described in gross anatomy, our approach is characterized by strict delineation and gradual changes (28 segments), thus allowing the model to be adapted to the needs of videolaryngoscopy segmentation for ML workflows. Furthermore, the majority of the existing approaches to larynx segmentation that are based on medical imaging, are in fact dedicated to the diagnosis of a pathology exceeding the mucosa. Premalignant conditions, carcinoma in situ, and T1 flat cancers are not detectable by radiological imaging and they elude the modern approach to segmentation for deep learning purposes. Despite the promising possibilities that the proposed endoluminal digital model of the larynx offers, it still requires a strict technical validation, experimental in its nature, and never performed before. All in all, the larynx digital model that has been proposed in this paper will provide the means for further developments towards AI applications in videoendoscopy diagnostics, and it may be an important contribution to the translational medicine approach.

REFERENCES

1. Hamet P., Tremblay J.: Artificial intelligence in medicine. Metabolism, 2017; 69: S36–S40.
2. Mahmood H., Shaban M., Rajpoot N., Khurram S.A.: Artificial Intelligence-based methods in head and neck cancer diagnosis: An overview. Br J Cancer, 2021; 124(12): 1934–1940.
3. Żurek M., Jasak K., Niemczyk K., Rzepakowska A.: Artificial Intelligence in Laryngeal Endoscopy: Systematic Review and Meta-Analysis. J. Clin. Med., 2022; 11(10): 2752. doi: https://doi.org/10.3390/jcm11102752.
4. Wasilewicz R., Mazurek C., Pukacki J.: Influence of cardiovascular system on 24 hour ocular volume changes, measured with contact lens sensor in healthy and POAG subjects. 8 World Glaucoma Congress, Melbourne, 27–30.03.2019.
5. Buchwald M., Przybylski Ł., Króliczak G.: Decoding brain states for planning functional grasps of tools: a functional magnetic resonance imaging multivoxel pattern analysis study. J Int Neuropsychol Soc, 2018; 24(10): 1013–1025.
6. Buchwald M.: Functional Magnetic Resonance Imaging Signal Modelling and Contrasts: an Example of Manual Praxis Tasks. CMST, 2021; 27(4): 159–167.
7. Behnke M., Buchwald M., Bykowski A., Kupiński S., Kaczmarek L.D.: Psychophysiology of positive and negative emotions, dataset of 1157 cases and 8 biosignals. Sci Data. 2022; 9(1): 10. doi: 10.1038/s41597-021-01117-0.
8. Currie G.: Intelligent imaging: anatomy of machine learning and deep learning. J. Nucl. Med. Technol., 2019; 47(4): 273–281.
9. Cerrolaza J.J., Picazo M.L., Humbert L.: Computational anatomy for multi-organ analysis in medical imaging: A review. Med. Image Anal., 2019; 56: 44–67.
10. Rao D., Prakashini K., Singh R., Vijayananda J.: Automated segmentation of the larynx on computed tomography images: a review. Biomed. Eng. Lett., 2022; 12: 175–183. doi: https://doi.org/10.1007/s13534-022-00221-3.
11. Wong J., Fong A., McVicar N. et al.: Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. Radiother Oncol., 2020; 144: 152–158. doi: 10.1016/j.radonc.2019.10.019.
12. Vrtovec T., Močnik D., Strojan P., Pernuš F., Ibragimov B.: Auto- segmentation of organs at risk for head and neck radiotherapy planning: from atlas- based to deep learning methods. Med Phys, 2020; 47(9): e929–e950.
13. Witkiewicz J., Klimza H., Piersiala K., Jackowska J., Wierzbicka M.: The usefulness of the narrow band imaging (NBI) in decision-making process regarding second look procedure (SL) in laryngeal cancer follow-up after transoral laser microsurgery. PLoS ONE, 2020; 15(8): e0236623.
14. Arens C., Piazza C., Andrea M. et al.: Proposal for a Descriptive Guideline of Vascular Changes in Lesions of the Vocal Folds by the Committee on

Endoscopic Laryngeal Imaging of the European Laryngological Society. Eur Arch Otorhinolaryngol, 2016; 273: 1207–1214. doi: 10.1007/s00405-015-3851-y.

15. Chang C., Lin W.N., Hsin L.J. et al.: Reliability of office-based narrow-band imaging-guided flexible laryngoscopic tissue samplings. Laryngoscope., 2016; 126(12): 2764–2769. doi: 10.1002/lary.26016.
16. Paderno A., Piazza C., Del Bon F. et al.: Deep Learning for Automatic Segmentation of Oral and Oropharyngeal Cancer Using Narrow Band Imaging: Preliminary Experience in a Clinical Perspective. Front Oncol., 2021; 11: 626602. doi: 10.3389/fonc.2021.626602.
17. Bhadoria S., Aggarwal P., Dethe C.G., Vig R.: Comparison of segmentation tools for multiple modalities in medical imaging. J. Adv. Inf. Technol., 2012; 3(4): 197–205.
18. Gunning D., Stefik M., Choi J. et al.: XAI-Explainable artificial intelligence. Sci Robot., 2019; 4(37): eaay7120. doi: 10.1126/scirobotics.aay7120.
19. Sekuboyina A., Husseini M.E., Bayat A. et al.: VerSe: A Vertebrae labelling and segmentation benchmark for multi-detector CT images. Med Image Anal., 2021; 73: 102166. doi: 10.1016/j.media.2021.102166.
20. Iyer A., Thor M., Haq R., Deasy J.O., Apte A.P.: Deep learning-based auto-segmentation of swallowing and chewing structures. bioRxiv, 2020; 772178.
21. Poznyakovskiy A.A., Mainka A., Platzek I., Mürbe D.: A Fast Semiautomatic Algorithm for Centerline-Based Vocal Tract Segmentation. Biomed Res Int., 2015; 2015: 906356. doi: 10.1155/2015/906356.
22. Storck C., Gugatschka M., Friedrich G. et al.: Developing a 3D model of the laryngeal cartilages using HRCT data and MIMICS's segmentation software. Logoped Phoniatr Vocol., 2010; 35(1): 19–23. doi: 10.3109/14015430903552378.
23. Kuo C.F.J., Lai W.S., Barman J., Liu S.C.: Quantitative laryngoscopy with computer-aided diagnostic system for laryngeal lesions. Sci Rep., 2021; 11(1): 10147. doi: 10.1038/s41598-021-89680-9.
24. Laves M.H., Bicker J., Kahrs L.A., Ortmaier T.: A dataset of laryngeal endoscopic images with comparative study on convolution neural network-based semantic segmentation. Int J Comput Assist Radiol Surg., 2019; 14(3): 483–492. doi: 10.1007/s11548-018-01910-0.
25. Gómez P., Kist A.M., Schlegel P. et al.: BAGLS, a multihospital Benchmark for Automatic Glottis Segmentation. Sci Data., 2020; 7(1): 186. doi: 10.1038/s41597-020-0526-3.
26. Fehling M.K., Grosch F., Schuster M.E., Schick B., Lohscheller J.: Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep Convolutional LSTM Network. PLoS ONE, 2020; 15(2): e0227791.
27. Lohscheller J., Toy H., Rosanowski F., Eysholdt U., Döllinger M.: Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. Med. Image Anal., 2007; 11(4): 400–413. doi: https://doi.org/10.1016/j.media.2007.04.005.
28. Lee H.H., Kwon B.M., Yang C.K., Yeh C.Y., Lee J.: Measurement of laryngeal elevation by automated segmentation using Mask R-CNN. Medicine (Baltimore)., 2021; 100(51): e28112. doi: 10.1097/MD.0000000000028112.
29. Korte J.C., Hardcastle N., Ng S.P. et al.: Cascaded deep learning-based auto-segmentation for head and neck cancer patients: Organs at risk on T2-weighted magnetic resonance imaging. Med Phys., 2021; 48(12): 7757–7772. doi: 10.1002/mp.15290.
30. Araújo T., Santos C.P., De Momi E., Moccia S.: Learned and handcrafted features for early-stage laryngeal SCC diagnosis. Med Biol Eng Comput., 2019; 57(12): 2683–2692. doi: 10.1007/s11517-019-02051-5.
31. Cho W.K., Lee Y.J., Joo H.A. et al.: Diagnostic Accuracies of Laryngeal Diseases Using a Convolutional Neural Network-Based Image Classification System. Laryngoscope., 2021; 131(11): 2558–2566. doi: 10.1002/lary.29595.
32. Esmaeili N., Illanes A., Boese A. et al.: Novel automated vessel pattern characterization of larynx contact endoscopic video images. Int J Comput Assist Radiol Surg., 2019; 14(10): 1751–1761. doi: 10.1007/s11548-019-02034-9.
33. Thiyagalingam J., Shankar M., Fox G., Hey T.: Scientific machine learning benchmarks. Nat. Rev. Phys, 2022; 4: 413–420.
34. Paderno A., Piazza C., Del Bon F. et al.: Deep Learning for Automatic Segmentation of Oral and Oropharyngeal Cancer Using Narrow Band Imaging: Preliminary Experience in a Clinical Perspective. Front Oncol, 2021; 11: 1–12. doi: https://doi.org/10.3389/fonc.2021.626602.
35. Moccia S., Vanone G.O., Momi E. et al.: Learning-based classification of informative laryngoscopic frames. Comput Methods Programs Biomed, 2018; 158: 21–30. doi: https://doi.org/10.1016/j.cmpb.2018.01.030.
36. Patrini I., Ruperti M., Moccia, S. et al.: Transfer learning for informative-frame selection in laryngoscopic videos through learned features. Med Biol Eng Comput, 2020; 58(6): 1225–1238. doi: https://doi.org/10.1007/s11517-020-02127-7.
37. Vabalas A., Gowen E., Poliakoff E., Casson A.J.: Machine learning algorithm validation with a limited sample size. PLoS ONE, 2019; 14(11): e0224365. doi: https://doi.org/10.1371/journal.pone.0224365.
38. El Naqa I., Ruan D., Valdes G. et al.: Machine learning and modeling: Data, validation, communication challenges. Med Phys, 2018; 45(10): e834–e840.
39. Breck E., Polyzotis N., Roy S., Whang S., Zinkevich M.: Data Validation for Machine Learning. MLSys., 2019; https://mlsys.org/Conferences/2019/doc/2019/167.pdf.

Table of content: https://otolaryngologypl.com/resources/html/articlesList?issueId=0 Tables: 2 Figures: 4 References: 39

Competing interests: The authors declare that they have no competing interests.

Corresponding author: Piotr Nogal MD; Department of Otolaryngology, Head and Neck Surgery, Poznan University of Medical Sciences, Poznan; Przybyszewskiego street 49, 60-355 Poznan, Poland; E-mail: pionogch@gmail.com

Cite this article as: Nogal P., Buchwald M., Staskiewicz M., Kupinski S., Pukacki J., Mazurek C., Jackowska J., Wierzbicka M.: Endoluminal larynx anatomy model – towards facilitating deep learning and defining standards for medical images evaluation with artificial intelligence algorithms; Otolaryngol Pol 2022; 76 (5): 1-9; DOI: 10.5604/01.3001.0015.9501