ACTA OTORHINOLARYNGOLOGICA ITALICA 2023;43:317-323; doi: 10.14639/0392-100X-N2309

Laryngology

Artificial intelligence for the recognition of benign lesions of vocal folds from audio recordings

Il ruolo del machine learning nel riconoscimento delle lesioni cordali benigne dal segnale vocale

Maria Raffaella Marchese[1], Federico Sensoli[2], Silvia Campagnini[2,3], Matteo Cianchetti[2], Andrea Nacci[4], Francesco Ursino[5], Lucia D'Alatri[1,6], Jacopo Galli[1,6], Maria Chiara Carrozza[2], Gaetano Paludetti[1,6], Andrea Mannini[2,3]

[1] Unità Operativa Complessa di Otorinolaringoiatria, Dipartimento di Neuroscienze, Organi di Senso e Torace, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy; [2] Institute of Biorobotics, Scuola Superiore Sant'Anna, Pontedera, Italy; [3] IRCCS Fondazione Don Carlo Gnocchi, Firenze, Italy; [4] U.O. Otorinolaringoiatria Audiologia e Foniatria, Azienda Ospedaliero Universitaria Pisana, Pisa, Italy; [5] Istituto Nazionale di Ricerche in Foniatria "G. Bartalena", Pisa, Italy; [6] Sezione di Otorinolaringoiatria, Dipartimento Universitario Testa-Collo e Organi di Senso, Università Cattolica del Sacro Cuore, Rome, Italy

SUMMARY

Objective. The diagnosis of benign lesions of the vocal fold (BLVF) is still challenging. The analysis of the acoustic signals through the implementation of machine learning models can be a viable solution aimed at offering support for clinical diagnosis.

Materials and methods. In this study, a support vector machine was trained and cross-validated (10-fold cross-validation) using 138 features extracted from the acoustic signals of 418 patients with polyps, nodules, oedema, and cysts. The model's performance was presented as accuracy and average F1-score. The results were also analysed in male (M) and female (F) subgroups.

Results. The validation accuracy was 55%, 80%, and 54% on the overall cohort, and in M and F, respectively. Better performances were observed in the detection of cysts and nodules (58% and 62%, respectively) *vs* polyps and oedema (47% and 53%, respectively). The results on each lesion and the different patterns of the model on M and F are in line with clinical observations, obtaining better results on F and more accurate detection of polyps in M.

Conclusions. This study showed moderately accurate detection of four types of BLVF using acoustic signals. The analysis of the diagnostic results on gender subgroups highlights different behaviours of the diagnostic model.

KEY WORDS: artificial intelligence, machine learning, benign lesions of vocal folds, dysphonia

RIASSUNTO

Obiettivo. *La diagnosi delle lesioni cordali benigne è ancora una sfida. L'analisi dei segnali vocali attraverso l'applicazione di modelli di Machine Learning potrebbe rappresentare una valida soluzione nell'offrire un supporto alla diagnosi clinica.*

Materiali e metodi. *In questo studio una Support Vector Machine è stata addestrata e sottoposta a validazione incrociata 10 volte usando 138 caratteristiche estratte dai segnali acustici di 418 pazienti affetti da polipi, noduli, edema di Reinke e cisti. Le prestazioni del modello sono state espresse in termini di accuratezza e punteggio F1 medio. I risultati sono stati inoltre analizzati nei sottogruppi maschi (M) e femmine (F).*

Risultati. *L'accuratezza era del 55%, 80% e 54% rispettivamente nel campione totale, nei maschi e nelle femmine. Le performances migliori sono state ottenute nel riconoscimento delle cisti e dei noduli (58% e 62% rispettivamente), rispetto ai polipi e agli edemi (47% e 53% rispettivamente). I risultati per ciascuna lesione e i differenti pattern del modello sono in linea con le caratteristiche cliniche nei sottogruppi maschi e femmine per i migliori risultati ottenuti nel gruppo femminile e una sensibile discriminazione dei polipi nei maschi.*

Conclusioni. *Questa ricerca ha dimostrato una capacità di riconoscimento dei quattro tipi di lesioni cordali benigne in base ai segnali acustici moderatamente accurata. L'analisi dei risultati diagnostici nei sottogruppi divisi per genere evidenzia i diversi comportamenti del modello diagnostico.*

PAROLE CHIAVE: *intelligenza artificiale, machine learning, lesioni cordali benigne, disfonia*

Received: September 20, 2022
Accepted: March 22, 2023
Published online: July 28, 2023

Correspondence
Maria Raffaella Marchese
Unità Operativa Complessa di Otorinolaringoiatria, Dipartimento di Scienze dell'Invecchiamento, Neurologiche, Ortopediche e della Testa-Collo, Fondazione Policlinico Universitario A. Gemelli IRCCS, l.go "A. Gemelli" 8, 00168 Rome, Italy
Tel. +39 06 30154439. Fax +39 06 3051194
E-mail: mariaraffaella.marchese@policlinicogemelli.it

How to cite this article: Marchese MR, Sensoli F, Campagnini S, et al. Artificial intelligence for the recognition of benign lesions of vocal folds from audio recordings. Acta Otorhinolaryngol Ital 2023;43:317-323. https://doi.org/10.14639/0392-100X-N2309

Introduction

The development of artificial intelligence (AI), the evolution of voice technology, progress in audio signal analysis, and natural language processing/understanding methods have opened the way to numerous potential applications of voice, such as the identification of vocal biomarkers for diagnosis, classification, patient remote monitoring, or to enhance clinical practice [1]. Especially at the beginning of the applications of AI, most studies used automatic systems based on machine learning (ML) on voice recordings to classify healthy subjects according to age and gender [2,3]. More recently, research focused on the role of the audio signal of the voice as a signature of the pathogenic process. Dysphonia indicates that some changes have occurred in voice production [4]. The overall prevalence of dysphonia is approximately 1% even if the actual rates may be higher depending on the population studied and the definition of the specific voice disorder [5]. The impact of a voice disorder has been increasingly recognised as a public health concern because it influences the quality of physical, social, and occupational aspects of life by interfering with communication [6]. Voice health may be assessed by several acoustic parameters. The relationship between voice pathology and acoustic voice features has been clinically established and confirmed both quantitatively and subjectively by speech experts [7]. The automatic systems are designed to determine whether the sample belongs to a healthy or non-healthy subject. The exactness of acoustic parameters is linked to the features used to estimate them for speech noise identification. Current voice searches are mostly restricted to basic questions even if with broad perspectives. Dankovicova et al. [8] demonstrated that ML analysis can recognise pathological speech with high classification accuracy. In the literature, the studies on vocal biomarkers have mainly been performed in the fields of neurodegenerative disorders (Parkison's and Alzheimer's diseases) [6]. On the contrary, the literature on vocal biomarkers of specific vocal fold diseases is anecdotal and related to functional vocal fold disorders or rare movement disorders of the larynx [9] (Tab. I).

The most common causes of dysphonia are benign lesions of the vocal fold (BLVF). The prevalence has been reported to be 2.5%-7.7% in national epidemiologic studies in the United States and South Korea [10,11]. Currently, videolaryngostroboscopy is the gold standard for the diagnosis of BLVF [12]. However, laryngoscopy is an invasive and expensive procedure. Moreover, it is not generally available in primary care units increasing the risk of delayed diagnosis and treatment. BLVF can interfere with vocal folds closure, introducing asymmetry to the vocal folds, and producing severely rough voices and aperiodic acoustic waveforms [13]. The traditional acoustic analysis is suitable only for nearly periodic signals and allows the clinician to have a quantifiable preliminary status for treatment follow-up, but it does not provide information for differential diagnosis or inter-individual comparison. In recent years, research has been oriented towards two principal aims, namely nonlinear dynamic acoustic analysis and the individuation of quantitative instrumental tools for voice assessment [14]. Novel ML algorithms have recently improved the classification accuracy of selected features in target variables compared to more conventional procedures thanks to the ability to combine and analyse large data-sets of voice features [15,16]. Even if the majority of studies focus on the diagnosis of a disorder where they differentiate between healthy and non-healthy subjects, we believe that a more important task is frequently differential diagnosis, where one needs to choose between two or more different diseases. Even though this is a challenging task, it is of crucial importance to move decisional support to this level. To our knowledge, the differential discrimination of BLVF using automatic audio data processing methods has been poorly studied to date. Moreover, there is no dataset of off-the-shelf audio recordings from dysphonic patients affected by BLVF available online. The main aim of this work is the study, development, and validation of ML algorithms to recognise the different BLVF from digital voice recordings. As a side result, the analysis of features' importance for the diagnostic models and their pathophysiological relevance was obtained.

Materials and methods

We collected the audio recordings of dysphonic patients affected by BLVF who referred to the Phoniatric Unit of the Fondazione Policlinico Universitario A. Gemelli - IRCCS of Rome from June 2015 to December 2019. All voice samples were divided into the following groups based on the endoscopic diagnosis: vocal fold cysts, Reinke's oedema, nodules and polyps. We excluded patients younger than 18 years or older than 65 years, with previous laryngeal or thyroid surgery, speech therapy, pulmonary diseases, gastro-oesophageal reflux, laryngeal movement disorder or recurrent laryngeal nerve paralysis. We also excluded non-native Italian speakers. The audio tracks were obtained by asking to pronounce with usual voice intensity, pitch and quality the word */aiuole/* three times in a row. Voices were acquired using a Shure model SM48 microphone (Evanston IL) positioned at an angle of 45° at a distance of 20 cm from the patient's mouth. The microphone saturation input was fixed at 6/9 of CH1 and the environmental noise was < 30 dB sound pressure level (SPL). The signals were recorded in ".nvi" format with a high-definition audio-recorder Computerized Speech Lab, model 4300B, from Kay Elemetrics (Lincoln Park, NJ, USA) with a sampling rate of 50 kHz

Table I. The list of researches that used ML analysis of voice recordings.

Author and reference	Size of sample	Study aim	Method of analysis	Author's key findings
Li et al. [2]	Training data set: 472 speakers; development data set: 300 speakers	To present an automatic speaker age and gender identification approach which combines different methods at both acoustic and prosodic levels to improve the baseline performance	Baseline subsystems: GMM based on mel-frequency cepstral coefficient features, SVM based on GMM mean supervectors and SVM based on 450-dimensional utterance level features; four subsystems	Minimum 3.1% and maximum 5.6% improvement of accuracy compared to the SVM baseline system
Berardi et al. [3]	From the archive of voice recordings given at Brigham Young University of a single individual spanning about 50 years were used	To investigate the progressive degeneration of a single talker's voice quality by comparing the results of a computer model trained on actual age with a model trained on perceived age with the goal of determining acoustic features related to the perception of aging voice	The acoustic features were used to train two random Forest regression models. One model used actual (chronological) age as the response variable with the other using estimated age	The ML model estimated the age of the talker as well as the human listeners. The acoustic feature related to the perception of aging voice is the fundamental frequency
Dankovicova et al. [7]	1560 speech features extracted and used to train the classification model	To utilise ML methods to recognise dysphonia	The classifiers, used: K-nearest neighbors, random Forest plots, and SVM	91.3% classification accuracy
Zhan et al. [8]	6148 smartphone activity assessments from 129 individuals	Ability of the mPDS to detect intraday symptom fluctuations, the correlation between the mPDS and standard measures, and the ability of the mPDS to respond to dopaminergic medication	ML based approach to generate a mPDS that objectively weighs features derived from each smartphone activity	Effective and objective Parkinson disease severity score derived from smartphone assessments
Suppa et al. [9]	60 patients with adductor-type spasmodic dysphonia pre-BoNT-A therapy and 60 healthy subjects; 35 patients were evaluated after BoNT-A therapy	To evaluate with cepstral analysis and ML adductor spasmodic voices and to compare the results with those of healthy subjects; to investigate the effect of BoNT-A	Cepstral analysis and ML algorithm classification techniques	ML and cepstral analysis differentiate healthy subjects and adductor-type spasmodic dysphonic voices. ML measures correlated with the severity of dysphonia pre and post-BoNT-A therapy
Asci et al. [15]	Voice samples of 138 younger adults and 123 older adults collected at home using smartphones	To examine the age-related changes in voice in ecological setting	ML analysis through a SVM	ML analysis demonstrates the effect of ageing on voice

ML: Machine Learning; mPDS: mobile Parkinson disease score; BoNT-A: botulinum neurotoxin A; SVM: support vector machine; GMM: Gaussian mixture model.

frequency and converted to ".wav" format. Each audio file was anonymously labelled with gender and type of BLVF.

Analysis pipeline

All the following analyses were performed using MatLab R2019b, the MathWorks, Natick MA, USA. The analysis pipeline included signal pre-processing, features extraction, screening of the features, and model implementation (Fig. 1).

Signal pre-processing

The unvoiced segments of signals were removed through a threshold-based algorithm, which accounts for the signal loudness and computed using the MathWorks built-in function acousticLoudness. This function implements two different methods accounted in the ISO 532 standard which allow estimation of loudness as perceived by persons with ontologically normal hearing under specific listening conditions. More specifically, three different features characterising the

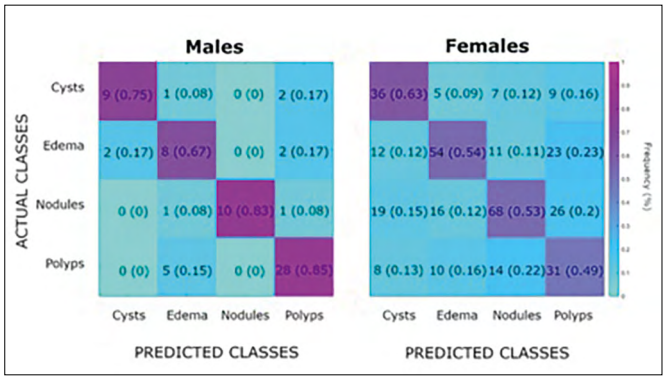

Figure 1. Analysis pipeline.

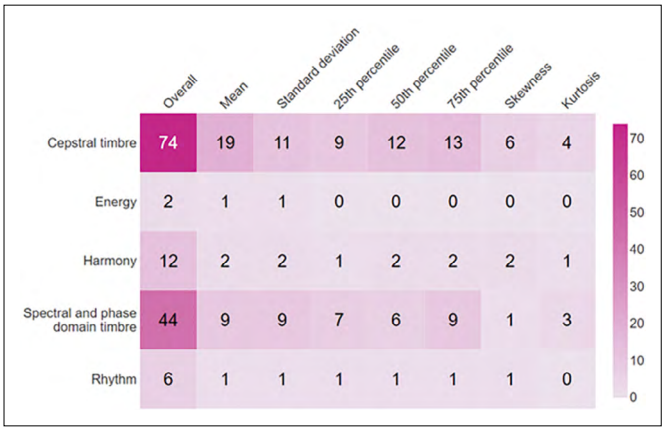

Figure 2. Overview of statistically significant features after the screening process.

human features are implemented, i.e. the dependency on frequency, the concept of critical band and spectral masking (the hearing feature for which frequencies can be swamped by a louder tone with close frequency). Further details on the implementation of these methods can be found in the following references [17,18]. Afterwards, a 1024-point Hamming window was used to segment the signals, with half-window overlap, to reduce the spectral leakage.

Feature extraction

On the segmented signal, 66 different features in the time, frequency, and cepstral domain were extracted. Next, seven statistical measures were computed on the extracted features, namely: mean, standard deviation, skewness, kurtosis, 25th, 50th, and 75th percentiles. In addition, jitter, shimmer, and tilt of the power spectrum were obtained from the whole unsegmented signal (see supplementary material).

Feature screening

Feature screening was applied using biostatistical analyses on the whole dataset to reduce the extended number of features to give as input to the classifier. Two statistical tests were used to screen relevant features for the classification task: the one-way analysis of variance (ANOVA), when all the groups were normally distributed, and the Kruskal-Wallis test, otherwise. The groups' normality was verified with the Kolmogorov-Smirnov test. For all the tests, a p-value < 0.05 was considered significant. An overview of the screened features entering the classification model is presented in Figure 2.

Model implementation

A non-linear Support Vector Machine (SVM) with a Gaussian kernel was the algorithm chosen in this work. The classifier was trained on the whole dataset (*Mtot*) and the male (*Mm*) and female (*Mf*) subsamples separately. On the three models, 10-fold cross-validation was used both for parameter optimisation and further feature selection. Specifically, two hyper-parameters of the model were set using a grid-search and in a range 1-1000, the box-constraint C, which is a regularisation parameter, and the kernel parameter γ, which controls the radius of influence of the kernel. Additionally, a sequential feature selection (SFS) algorithm was implemented to find an optimal feature set. Lastly, a class balancing method was employed to overcome the unbalanced frequencies of the pathological classes. The augmentation algorithm of the Synthetic Minority Oversampling Technique [19] was utilised for this purpose.

The classification performance was measured through the accuracy and the average F1-score. Both metrics were provided for the description of the overall classification performances and those obtained on gender subgroups.

Results

Data were collected on 418 patients with a mean age of 48 years (range 19-65 years), 349 of whom (83.5%) were females. The classification domain included four pathological groups: among recruited patients, 69 (16.5% - 12 males and 57 females) were diagnosed with vocal fold cysts, 112 (26.8% - 12 males and 100 females) with Reinke's oedema, 141 (33.7% - 12 males and 129 females) with nodules and 96 (23.0% - 33 males and 63 females) with polyps (Tab. II).

The seven statistical descriptors calculated for each of the 66 extracted features led to a total number of 462 features. Three additional features (jitter, shimmer, and tilt of the power spectrum) were computed on the whole signal, for a total of 466 features. Given the extensive number of features extracted (overcoming the number of patients), a first statistical screening was performed. From the statistical tests, 138 of

Table II. Sample numerosities on the whole dataset and pathology and gender-specific.

		Overall sample	Pathology-specific samples (N)			
			Cysts	Oedemas	Nodules	Polyps
Overall sample		418	69	112	141	96
Gender-specific numerosity	Males	69	12	12	12	33
	Females	349	57	100	129	63

466 features were significantly associated with the outcome and were thus fed into the SVM model. In Figure 1, a summary of the analysis pipeline and the number of patients and features in each step is provided.

On the overall model *Mtot*, the optimised 4 classes classifier showed an accuracy of 55.0% and an average F1-score of 0.54. When splitting the sample by gender, an accuracy and average F1-score of 80% and 0.78 were obtained for *Mm*, and 54% and 0.54 for *Mf*, respectively. These results are summarised in Table III, along with the single pathology accuracies. In Figure 3, the confusion matrices for the developed models are presented.

Discussion

In this work, an SVM was trained and cross-validated with the aim to obtain differential diagnosis of BLVF types given the acoustic signals of 418 patients, obtaining a global accuracy of 55.0%.

Compared to the majority of studies in the literature, our results showed reduced performance, although it is worth noticing that only a few have focused on pathological groups only for differential diagnosis [20,21]. Even the comparison with the latter authors' works should be done in light of the differences among the studies. Namely, non-pathological cases were also included, with the exception of Pham et al. [19] in which deep learning algorithms were applied, different pathological cases were tested and different types of the acoustic signal were processed (sustained vowels or combinations of them).

The taxonomy of the types of BLVF is still in evolution, as well as the tools and technologies available for diagnosis. For these reasons, and due to overlapping characteristics of the groups, differential diagnosis can be challenging [22]. Thus, supporting the clinical experience with a tool based on objective data can be crucial, and acoustic signals are promising non-invasive solutions for such a purpose.

Going into detail about the specific types of BLVF, the model had better performance on cysts and nodules (58% and 62% accuracy, respectively), and less so for recognition of polyps and oedema definition of identity (47% and 53% accuracy, respectively). While more investigation should be dedicated to these results, a possible explanation of these differences may be due to the diverse effects of the above-mentioned lesions on phonation. In fact, while polyps and oedema tend to be very mobile during phonation, cysts and nodules are usually less mobile [23] and this may have an effect on the features obtained from the acoustic signals.

Lower accuracy could also be due to different ages of patients, to fact that the pathological samples were at different stages of disease evolution, or because the morphological features varied and impacted differently on acoustic parameters (e.g. size, the ratio of vocal lesion base to the width of the lesion, the implantation site etc.). In future studies, the system with a larger dataset and new classification model could be used to classify the stage of the disease.

Another aspect we considered in our analyses was the difference in results obtained on gender split. Differences between male and female voices and acoustic are related to many factors, such as physiology, anatomy, and even sociology and psychology, in terms of identity definition and behavioural characteristics [24]. In this regard, as a result, differences in frequency content between genders could emphasise the effects in the voice of different pathologies.

In this work, where only features extracted from the spectral analysis entered the model, it is reasonable to obtain very different behaviours of the model on the two subgroups. Specifi-

Table III. Classification results.

Models	Overall sample		Pathology-specific accuracy			
	Accuracy	F1-score	Cysts	Oedema	Nodules	Polyps
Mtot	0.55	0.54	0.58	0.53	0.62	0.47
Mm	0.80	0.78	0.75	0.67	0.83	0.85
Mf	0.54	0.54	0.63	0.54	0.53	0.49

Mtot: Whole dataset; Mm: male dataset; Mf: female dataset.

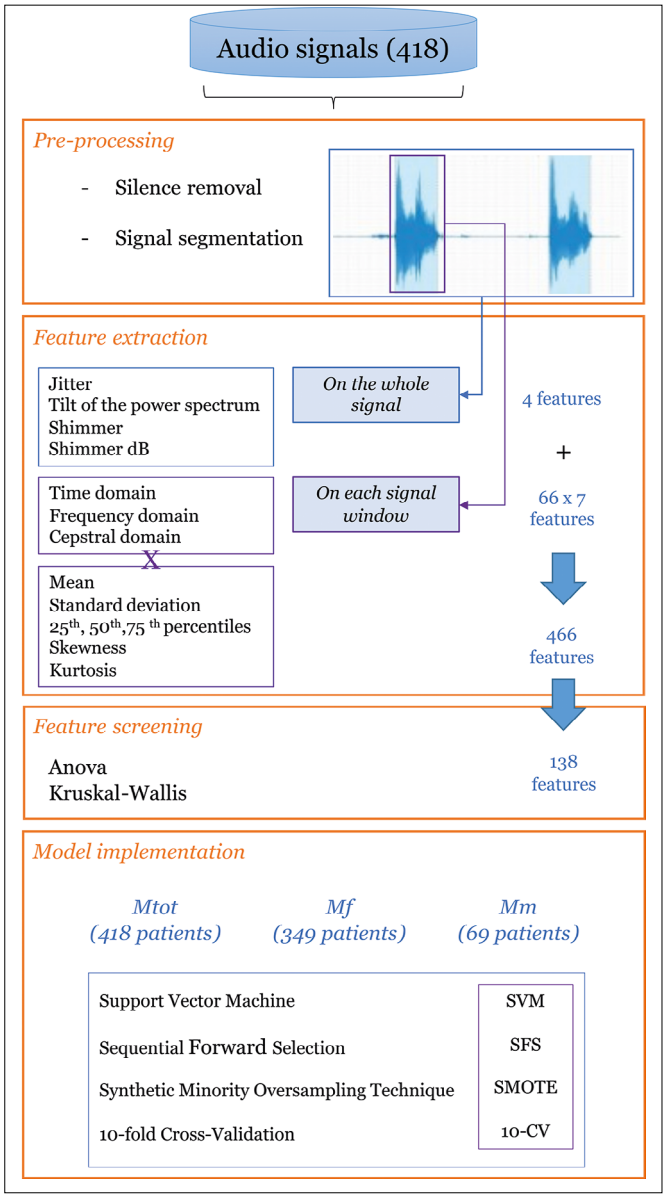

Figure 3. Confusion matrices for the developed model of male and female subgroups.

cally, the results showed an accuracy of 79.7% in the model trained on males and 54.4% in the models trained on females. The overall F1-score and accuracy and the pathology-specific accuracies suggest how better performances can be obtained in the male group, with respect to females. In fact, particularly high accuracies of 83% and 85% were obtained on detection of nodules and polyps, respectively (Tab. III). The promising results on polyp detection, in opposition to the poor accuracy obtained on the *Mf* (49%), may be explained by the fact that BLVF, and especially vocal polyps, present a higher incidence in the male population [25]. Nevertheless, these results, and more generically the role of gender in differential diagnosis, should be deepened further. Indeed, while reading these last findings, the sample size of the male subgroup should be taken into account as a limitation reducing the generalisability of the results. To optimise the overall accuracy, other improvements could involve the training of other classification algorithms and the application of nested cross-validation for a better generalisability of the results [26]. Additionally, a larger prospective dataset could allow for the exploration of deep learning solutions and a complete selection of the features through automatic methods, avoiding the need for biostatistical screening. In future works, introducing novel features for classifiers or identifying the best performing ones can provide new information. More sophisticated features capturing hidden patterns or nonlinear relationships can significantly boost prediction accuracy. We believe that the clinical usefulness of the classification accuracy achieved by our model could be understood by comparing with further studies how our algorithm performs to those of human experts and non-experts.
Above all, given the promising results in diagnostic problems obtained from clinical and endoscopic high-speed videos [27], we believe that a bimodal analysis system that integrates both audio recording and video-endoscopic imaging data would be helpful to improve accuracy and to explain the correlations of the SVM.
Finally, our preliminary work overall suggests that machine learning, in combination with telemedicine, could provide a strategy to support screening between BLVF and malignant glottic lesions. Undoubtedly, a larger data set is needed before reaching such a target but this could be a concrete development.

Conclusions

This work focused on the development and cross-validation of a diagnostic model for the identification of four different BLVFs: polyps, cysts, nodules, and oedema. Although further efforts could be deployed on the technical implementation, when larger datasets will be available, the results appear promising, with an overall accuracy in the automatic differential diagnosis of 55.3%. Moreover, the behavioural patterns of the models developed on the gender subgroups were particularly interesting. Specifically, a good sensitivity was found in polyp detection for males (85%), and in general better performance in the detection of each BLVF type than in females. Given these results and the analysis of the literature, future research should focus on the combined use of clinical and instrumental data for the development of diagnostic models of BLVF.

Conflict of interest statement

The authors declare no conflict of interest.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions
MRM: conceptualisation, data curation, writing; FS: formal analysis, software; SC: preparation and creation of the published work, specifically data presentation, writing; MC, AN, FU, JG, MCC, GP: supervision; LD: data collection; AM: methodology, data curation.

Ethical consideration
This study was approved by the Institutional Ethics Committee of "Fondazione Policlinico Universitario A. Gemelli IRCCS" of Rome (ID 4519 - prot. n.0041543/21).
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

1 Robin J, Harrison JE, Kaufman LD, et al. Evaluation of speech-based digital biomarkers: review and recommendations. Digit Biomark 2020;4:99-108. https://doi.org/10.1159/000510820

2 Li M, Han KJ, Narayanan S. Automatic speaker age and gender recognition using acoustic and prosodic level information fusion. Comput Speech Lang 2013;27:151-167. https://doi.org/10.1016/j.csl.2012.01.008

3 Berardi ML, Hunter EJ, Ferguson SH. Talker age estimation using machine learning. Proc Meet Acoust 2017;30:040014. https://doi.org/10.1121/2.0000921

4 Lopez-de-Ipina K, Satue-Villar A, Faundez-Zanuy M, et al. Advances in a multimodal approach for dysphagia analysis based on automatic voice analysis. In: Bassis S, Esposito A, Morabito F, et al., editors. Advances in Neural Networks. WIRN 2015. Smart Innovation, Systems and Technologies, vol 54. Springer, Cham. https://doi.org/10.1007/978-3-319-33747-0_20

5 Cohen SM, Dupont WD, Courey MS. Quality-of-life impact of non-neoplastic voice disorders: a meta-analysis. Ann Otol Rhinol Laryngol 2006;115:128-134. https://doi.org/10.1177/000348940611500209

6 Zhan A, Mohan S, Tarolli C, et al. Using smartphones and machine learning to quantify Parkinson disease severity: the mobile Parkinson disease score. JAMA Neurol 2018;75:876-880. https://doi.org/10.1001/jamaneurol.2018.0809

7 Mekyska J, Janousova E, Gomez-Vilda P, et al. Robust and complex approach of pathological speech signal analysis. Neurocomputing 2015;167:94-111. https://doi.org/10.1016/j.neucom.2015.02.085

8 Dankovicová Z, Sovák D, Drotár P, et al. Machine learning approach to dysphonia detection. Appl Sci 2018;8:1927. https://doi.org/10.3390/app8101927

9 Suppa A, Asci F, Saggio G, et al. Voice analysis in adductor spasmodic dysphonia: objective diagnosis and response to botulinum toxin. Park Relat Dis 2020;73:23-30. https://doi.org/10.1016/j.parkreldis.2020.03.012

10 Byeon H. Prevalence of perceived dysphonia and its correlation with prevalence of clinically diagnosed laryngeal disorders: the Korea National health and Nutrition Examination Surveys 2010-2112. Ann Otol Rhinol Laryngol 2015;124:770-776. https://doi.org/10.1177/0003489415583684

11 Hah JH, Sim S, An SY, et al. Evaluation of the prevalence of and factors associated with laryngeal diseases among the general population. Laryngoscope 2015;125:2536-2542. https://doi.org/10.1002/lary.25424

12 Bohlender J. Diagnostic and therapeutic pitfalls in benign vocal fold diseases. GMS Curr Top Otorhinolaryngol Head Neck Surg 2013;12:Doc01. https://doi.org/10.3205/cto000093

13 Channon F, Stone RE. Nodules and polyps. In: Brown WS, Vinson BP, Crary MA, editors. Organic voice disorders: assessment and treatment. San Diego, CA: Singular; 2000.

14 Heman-Ackah YD, Sataloff RT, Laureyns G, et al. Quantifying the cepstral peak prominence, a measure of dysphonia. J Voice 2014;28:783-788. https://doi.org/10.1016/j.jvoice.2014.05.005

15 Asci F, Costantini G, Di Leo P, et al. Machine learning analysis of voice samples recorded through smartphones: the combined effect of ageing and gender. Sensors 2020;20:5022. https://doi.org/10.3390/s20185022

16 Hegde S, Shetty S, Rai S, et al. A survey on machine learning approaches for automatic detection of voice disorders. J Voice 2019;33:947. https://doi.org/10.1016/j.jvoice.2018.07.014

17 ISO 532-1:2017(E). "Acoustics – methods for calculating loudness – Part 1: Zwicker method". International Organization for Standardization. ISO, Geneva; 2017.

18 ISO 532-2:2017(E). "Acoustics – methods for calculating loudness – Part 2: Moore-Glasberg method". International Organization for Standardization. ISO, Geneva; 2017.

19 Chawla N, Bowyer K, Hall L, et al. Smote: synthetic minority oversampling technique. J Artif Intell Res 2002;16:321-357. https://doi.org/10.1613/jair.953

20 Hu HC, Chang SY, Wang CH, et al. Deep learning application for vocal fold disease prediction through voice recognition: preliminary development study. J Med Internet Res 2021;23:E25247. https://doi.org/10.2196/25247

21 Pham M, Lin J, Zhang Y. Diagnosing voice disorder with machine learning. IEEE International Conference on Big Data (Big Data) 2018:5263-5266. https://doi.org/10.1109/BigData.2018.8622250

22 Naunheim M, Carroll T. Benign vocal fold lesions: update on nomenclature, cause, diagnosis, and treatment. Curr Opin Otolaryngol Head Neck Surg 2017;25:1. https://doi.org/10.1097/MOO.0000000000000408

23 Dikkers FG, Nikkels PJ. Benign lesions of the vocal folds: histopathology and phonotrauma. Ann Otol Rhinol Laryngol 1995;104:698-703. https://doi.org/10.1177/000348949510400905

24 Pépiot E, Arnold A. Cross-Gender Differences in English/French Bilingual Speakers: A Multiparametric Study. Percept Mot Skills 2021;128:153-177. https://doi.org/10.1177/0031512520973514.

25 Malik P, Yadav S, Sen RD, et al. The clinicopathological study of benign lesions of vocal cords. Indian J Otolaryngol 2017;71:212-220. https://doi.org/10.1007/s12070-017-1240-0

26 Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 2006;8. https://doi.org/10.1186/1471-2105-7-91

27 Unger J, Schuster M, Hecker D, et al. A multiscale product approach for an automatic classification of voice disorders from endoscopic high-speed videos. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:7360-7363. https://doi.org/10.1109/EMBC.2013.6611258