Artificial Intelligence In Medicine 120 (2021) 102162

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Replication-based regularization approaches to diagnose Reinke's edema by using voice recordings

Lizbeth Naranjo [a], Carlos J. Pérez [b,*], Yolanda Campos-Roca [c], Mario Madruga [b]

[a] Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, 04510 Ciudad de México, Mexico
[b] Departamento de Matemáticas, Facultad de Veterinaria, Universidad de Extremadura, 10003 Cáceres, Spain
[c] Departamento de Tecnologías de los Computadores y de las Comunicaciones, Escuela Politécnica, Universidad de Extremadura, 10003 Cáceres, Spain

ARTICLE INFO

Keywords:
Acoustic features
Classification
Reinke's edema
Regularization
Replicated measurements
Variable selection

ABSTRACT

Reinke's edema is one of the most prevalent laryngeal pathologies. Its detection can be addressed by using computer-aided diagnosis systems based on features extracted from speech recordings. When extracting acoustic features from different voice recordings of a particular subject at a concrete moment, imperfections in technology and the very biological variability result in values that are close, but they are not identical. This suggests that the within-subject variability must be properly addressed in the statistical methodology. Regularization-based regression approaches can be used to reduce the classification errors by favoring the best predictors and penalizing the worst ones. Three replication-based regularization approaches for variable selection and classification have been specifically designed and implemented to take into account the underlying within-subject variability. In order to illustrate the applicability of these approaches, an experiment has been specifically conducted to discriminate Reinke's edema patients (30 subjects) from healthy people (30 subjects) in a hospital environment. The features have been extracted from four phonations of the sustained vowel /a/ recorded for each subject, leading to a database that has fed the proposed machine learning approaches. The proposed replication-based approaches have been proved to be reliable in terms of selected features and predictive ability, leading to a stable accuracy rate of 0.89 under a cross-validation framework. Also, a comparison with traditional independence-based regularization methods reports a great variability of the latter in terms of selected features and accuracy metrics. Therefore, the proposed approaches contribute to fill a gap in the scientific literature on statistical approaches considering within-subject variability and can be used to build a robust expert system.

1. Introduction

Voice is the main communication tool that human beings have. Misuse or overuse of the vocal folds can damage the vocal function. Voice disorders may affect anyone, but they are especially relevant for voice professionals such as teachers, singers, actors, anchors, coaches, lawyers… Voice professionals are prone to suffer from organic voice disorders and, because of that, they need to avoid potential risks and, eventually, ask for medical care [41].

Reinke's edema is one of the most prevalent laryngeal pathologies [33]. It is the result of the gelatinous fluid accumulation in the Reinke's space, mainly due to vocal abuse and/or heavy tobacco use. It mainly affects women, causing progressive hoarse voice with a lower pitch, less vocal power and a tendency to fatigue in more intense cases [5]. Direct inspection of the larynx through laryngoscopy and videostroboscopy (specialized invasive equipment) and/or subjective listening tests to evaluate voice quality are two common diagnostic tools used by otolaryngologists [45].

In the last years, acoustic features extracted from voice recordings have been considered as a potential biomarker (non-invasive, fast, objective, and low cost) to assist in the diagnosis and tracking of voice-related diseases. Computer-Aided Diagnosis (CAD) systems have been built with this purpose, consisting of an acoustic feature extraction step followed by the use of machine learning algorithms. A perspective on automatic speech signal analysis for clinical diagnosis and assessment of speech disorders is provided by Baghai-Ravary and Beet [1] and Gómez-García et al. [13]. These systems have been developed for several diseases affecting the voice such as, e.g., vocal fold nodules, vocal fold polyps, Reinke's edema, or even neurodegenerative disorders such as Parkinson's disease.

* Corresponding author at: Avda. de las Ciencias, s/n, 10003, Cáceres, Spain.
E-mail address: carper@unex.es (C.J. Pérez).

https://doi.org/10.1016/j.artmed.2021.102162
Received 15 February 2021; Received in revised form 21 August 2021; Accepted 31 August 2021
Available online 8 September 2021

Diagnosis of Reinke's edema can be addressed by using CAD systems based on features extracted from speech recordings. Some authors have considered a mix of different pathologies, including Reinke's edema, to build a unique pathological class to discriminate diseased subjects from healthy ones [6,23,31,58]. Verde et al. [56] focused on Reinke's edema and their results were based on a personalized fundamental frequency estimation and no other acoustic feature was considered. Features based on nonlinear dynamics analysis have not been thoroughly used for Reinke's edema diagnosis in the scientific literature. Tavares et al. [51] combined entropy measures and cepstral analysis to discriminate healthy subjects from people suffering from Reinke's edema. Based on energy, zero-crossing rate and signal entropy, Silva Fonseca et al. [48] presented a speech disorder classification method that handles coexisting pathologies (Reinke's edema and laryngitis) that share the main phonic symptom. Phonation of sustained vowels was used in the previous works because they constitute easy to produce tasks, involving vocal fold vibration [39].

MEEI database, commercialized by Kay Elemetrics, is one of the most used voice database for automatic diagnosis research and covers several voice pathologies, including Reinke's edema [34]. However, it suffers from some disadvantages: the recorded phonations have been performed with high quality equipment in an acoustically controlled environment, and normal and pathological voices were recorded in different locations. Besides, the voice recordings have been selected by experts, which allowed for obtaining the best examples of each disease [44]. This has provided high accuracies when applying machine learning methods, but the results are not transferable to realistic situations where the phonations are recorded in medical centers or occupational health and safety services.

Voice databases used for organic disease diagnosis are generally based on one single utterance per subject, i.e., acoustic features extracted from only one voice recording per subject. However, there exists variability between two or more voice recordings from the same subject at a particular time, so using only one utterance per subject may provide different results depending on the voice recording that has been selected. The imperfections in technology and the very biological variability result in values that are similar (but not identical) for recordings from a particular subject, rather than for recordings from different individuals. For Parkinson's disease diagnosis, many authors considered several replicated voice recordings for each subject, so a collection of related features based on consecutive voice recordings for each subject are used (see, e.g., Little et al. [29]). Although the existing variability among the extracted features from the several voice recordings of each subject has been recognized and the experimental design is based on the within-subject dependence of the recordings of each individual, traditional machine learning techniques based on independence have been usually applied to all the utterances as if they were independent [8,29,54]. This means that the considered experimental unit is the utterance, and not the subject, so a voting system is used to decide if a subject is classified as healthy or having the disease by taking into account the larger number of utterances classified as healthy or diseased for each subject. This leads to an artificial increase of the sample size, a diffuse criterion to make decisions since one subject can have utterances classified as healthy and diseased, and the application of independence-based methods to dependent data.

The replicated measurements must be treated with specifically designed methods that address the existing within-subject variability. Pérez et al. [43] developed a logistic regression-based classification approach that takes into account the underlying within-subject dependence based on 6/7 utterances per subject. Later, Naranjo et al. [37] addressed this problem with a probit regression based on 3 utterances per subject, whereas Naranjo et al. [38] proposed a variable selection and classification approach for the same data. All these three approaches have been developed in the context of Parkinson's disease diagnosis with features extracted from voice recordings.

In this paper, replication-based Bayesian regularization approaches for Reinke's edema diagnosis using acoustic features extracted from speech recordings have been developed and implemented. Variable selection and classification approaches have been widely addressed by Bayesian regularization regression with independent instances (see, e.g., van Erp et al. [55]), which aim to shrink small effects to zero while maintaining true large effects. However, there is a lack of regularization methods able to address within-subject variability. To the best of the authors' knowledge, up to now, it has never been demonstrated that having into account the within-subject variability provides more stable results than the approaches based on independent instances at the same time that relevant features are selected and accuracy metrics keep at good values. This study contributes to fill a gap in the scientific literature on statistical approaches considering replicated data and they can be used to build robust CAD systems. The main contributions of this article are:

- Designing and implementing three Bayesian regularization approaches based on replicated measurements.
- Using Markov Chain Monte Carlo (MCMC) methods to solve the increasingly complex models.
- Conducting an experiment to discriminate subjects suffering from Reinke's edema (30 subjects) from healthy people (30 subjects) in a hospital environment.
- Extracting a variety of relevant features based on perturbation, cepstral analysis, noise, nonlinear dynamics, and entropies.
- Proposing and integrating a 95% Bayesian credible interval-based technique to determine the most relevant acoustic features.
- Reporting a robust performance in terms of feature selection and predictive capability, leading to an accuracy of 0.89 by using cross-validation and 0.93 without it.
- Reporting the outperformance of the replication-based approach based on Ridge regression with respect to the traditional regularization methods based on independent instances, which provide a great variability in terms of selected features and accuracy metrics.

The rest of this paper is structured as follows. Section 2 shows the necessary information to collect the dataset, i.e., participants, equipment, speech recordings, and feature extraction procedures. In Section 3, the general Bayesian approach is presented, including the binary response model, the way the replications are addressed in the model, the prior distributions for the different approaches, the Bayesian analysis, and the variable selection method. Section 4 shows the experimental settings and results. In Section 5, a discussion is presented, and the conclusions can be found in Section 6.

2. Data collection

This section provides details on the different aspects related to the generation of the acoustic feature database, i.e., the participants, protocol, recording equipment, vocal task, and feature extraction process.

2.1. Participants

A total of 60 people participated in the study. Half of them were diagnosed as suffering from Reinke's edema and the other half were healthy control subjects. The general eligibility criteria for participation were to be volunteers, native Spanish speakers, aged from 18 to 65, and to properly perform the phonation task in the research protocol.

The group of people suffering from Reinke's edema comprised 27 women and 3 men, with mean (standard deviation) age of 47.9 (11.8) years. They were recruited among the volunteers who attended the voice disorder program at the San Pedro de Alcántara Hospital. Note that there is a gender imbalance due to the fact that women are more affected by organic vocal-fold pathologies than men (see, e.g., Hunter et al. [21]). The gender rate in this study is approximately the same as in people attending the voice disorder program at the moment of the recruitment.

On the other hand, the healthy control group was selected among people with good vocal health status, who had never suffered from any voice pathology or used their voices in a professional way. It comprised 26 women and 4 men, with mean (standard deviation) age of 40.8 (11.2) years.

All the subjects were informed and provided their consent by signing an informed consent letter.

2.2. Protocol and equipment

The participants were asked to fill out a questionnaire for assessment of part of the general and specific eligibility criteria. They provided information such as sex, age, smoking habits, use of medication, and previous surgical interventions. They also underwent a medical examination consisting of a laryngological evaluation by videostroboscopy performed by an otorhinolaryngologist. For the subjects suffering from Reinke's edema, it was confirmed that Reinke's edema was the only existing voice pathology.

A portable computer with an external sound card (TASCAM US322) and a headband microphone (AKG 520) featuring a cardioid pattern was used to record the phonations. The digital recording was performed using Audacity software (release 2.0.5). The sampling frequency was 44.1 kHz and the resolution 16 bits/sample.

This research protocol was approved by the bioethics committees of the San Pedro de Alcántara Hospital and the University of Extremadura.

2.3. Speech recordings

The voice recordings were performed in an ordinary diagnostic room at San Pedro de Alcántara Hospital. The room was not sound-proof, but a certain isolation from the aisles and waiting halls was obtained by regular walls and closed doors. No specific measures for acoustic isolation were implemented.

The participants were asked to perform a sustained voicing of the /a/ vowel, at a comfortable pitch and loudness, as constantly as possible. This phonation was kept up as long as they could after a deep breath. A segment of one second was considered for feature extraction. This procedure was repeated four consecutive times per individual to address the within-subject variability after feature extraction.

2.4. Feature extraction

Different types of acoustic features were considered. The idea was to measure different aspects of speech degradation caused by the voice disorder.

Two conventional perturbation measures (jitter and shimmer) were extracted based on the high values observed in patients with Reinke's edema in previous studies [47]. Fundamental frequency and amplitude perturbations also produce an impact on the cepstral peak prominence (CPP). This measure, originally proposed by Hillenbrand et al. [19], is considered more robust than time-domain techniques, since it does not require pitch tracking and can be reliably extracted even from highly aperiodic signals. For this reason, CPP has been included in the list of features.

Voice roughness is a characteristic symptom of Reinke's edema because the swelling alters the elasticity of the vocal folds [7]. Two noise measures have been included in the feature set to assess roughness: glottal-to-noise excitation (GNE) ratio and the harmonic-to-noise ratio (HNR). These noise measures have been considered suitable for the detection of voice pathologies [12].

According to previous scientific studies, vocal fold pathologies lead to changes in vocal tract configuration during phonation. Lee et al. [27] pointed out that the reason is related to physiological or psychological compensations. Mel-frequency cepstral coefficients (MFCCs) have been widely used to characterize the vocal tract configuration in different application areas of speech classification, also for the detection of vocal-fold disorders [10]. A total of 13 MFCCs were calculated and included in the feature set.

Furthermore, it has been emphasized that nonlinear behaviors play a relevant role in the voice production process, especially in the case of disordered voices [12,32,53]. Therefore, the classical source-filter theory is not sufficient to describe all important aspects of speech that can be useful to detect pathologies. Orozco-Arroyave et al. [40] state different reasons which lead to a nonlinear speech behavior: nonlinear pressure-flow in the glottis, nonlinear stress-strain curves of vocal fold tissues, and nonlinearities in vocal fold collisions. These authors also consider the compensatory movements mentioned in the previous paragraph as nonlinear effects. Based on this nonlinear assumption, some authors have proposed acoustic features taken from the field of time-series analysis to predict diseases affecting voice [25,30,40]. The following ones have been used: Hurst exponent (HURST), correlation dimension (D2), permutation (PERMUTATION) and shannon entropy (SHANNON), pitch period entropy (PPE), and recurrence period density entropy (RPDE). Finally, the zero-crossing rate (ZCR) was also included. This adds up to a total of 25 acoustic features extracted from each voice sample. The extraction methods were coded in Python.

Gender is also important in this topic. Yamauchi et al. [59] used glottal area waveform analysis based on high-speed digital imaging to emphasize the relevant role of gender when deciding whether a vocal fold pattern is normal or pathological. Previous studies [26,52] had already identified gender differences in vocal fold configuration during phonation: in glottal flow, glottal area or contact area waveforms. These anatomical and physiological differences have motivated the inclusion of the gender label as an additional feature, giving a total number of 26 features.

The feature extraction procedure provides a dataset with 240 rows (60 subjects $\times$ 4 utterances) and 27 columns (number of features plus health status).

3. Methodology

In the following subsections the methodology is described. Firstly, a hierarchical model to deal with binary responses and replicated covariables is formulated. This provides a general framework for replication-based classifiers. Then, three Bayesian regularization methods are considered through their respective prior distributions. Next, the posterior distribution is estimated and the posterior predictive probabilities are calculated. Finally, a variable selection method based on Bayesian credible intervals is proposed to determine the most relevant features.

3.1. Binary response model

In order to define the hierarchical model, the first level corresponds to the binary response variable. Let $Y_1, \ldots, Y_n$ be the n independent binary random variables:

$$Y_i \sim \text{Bernoulli}(\theta_i)$$

The probabilities $\theta_i = \text{P}(Y_i = 1)$ are related to two sets of covariates, $\boldsymbol{w}_i$ and $\boldsymbol{z}_i$ by:

$$\Psi^{-1}(\theta_i) = w_i^{'}\beta + z_i^{'}\gamma,$$

where $w_i = (w_{i1}, \ldots, w_{iK})'$ and $z_i = (z_{i1}, \ldots, z_{iH})'$ are covariate vectors of dimension K and H, respectively. The parameters β and γ are vectors of unknown parameters, of dimensions K and H, respectively. $\Psi^{-1}(\cdot)$ is the inverse of the cumulative distribution function (cdf) of the normal distribution.

3.2. Introducing replications

Assume that the covariates $\boldsymbol{z}_i$ are exactly known (e.g. sex), but the covariates $\boldsymbol{w}_i$ are not (acoustic features), instead they have been

measured with J replicates. Let $x_{ij} = (x_{i1j}, \ldots, x_{iKj})'$ be the jth replication of the unknown covariate vector $w_i = (w_{i1}, \ldots, w_{iK})'$, $j = 1, \ldots, J$, and assume that they have a linear relationship specified as an additive measurement error model (see, e.g. Buonaccorsi [3]), i.e.:

$$x_{ikj} = w_{ik} + \varepsilon_{ikj},$$
$$\varepsilon_{ikj} \sim \text{Normal}\left(0, \delta_k^2\right),$$

where the errors ε_{ik} are independent of w_{ik}, and x_{ikj} can be considered as surrogates of w_{ik}.

The rationale under this formulation is that the observed replicated features can be considered as measurement with errors of the underlying real acoustic feature, which is unknown for each individual. This latent variable-based structure is the key idea to address the within-subject variability.

3.3. Integrating regularization

Regularization methods simultaneously perform estimation and variable selection. They favor the best predictors and penalize the worst ones through parameter regularization. A wide variety of regularization methods have been developed (see e.g., Hastie et al. [15] and Hastie et al. [16]). The most usual regularization methods are Least Absolute Shrinkage and Selection Operator (LASSO), Ridge, and Elastic Net. They have been widely used for independent instances, but now they are considered for data with dependent nature in a framework that addresses the within-subject variability of replicated measurements, and therefore for a different type of statistical design.

In typical Bayesian regression, the prior distribution for the regression parameters is normal. When regularization methods are considered, different prior distributions are used. LASSO is one of the most commonly used penalized regression methods (see Park and Casella [42]). The prior distribution for the regression parameters β_k is based on the proposal of Genkin et al. [9], i.e., a Laplace distribution is considered, i.e.:

$$\beta_k \sim \text{Laplace}\left(0, \lambda_1^{-1}\right),$$

with mean 0 and variance $2/{\lambda_1}^2$, for k=1,…, K.

The Laplace pdf is proportional to:

$$\text{p}(\beta_k) \propto exp\{-\lambda_1|\beta_k|\},$$

and it can be represented as a scale mixture of normal distributions with independent exponentially distributed variances, i.e.:

$$\text{p}(\beta_k) = \int_0^\infty \text{p}(\beta_k|\tau_k)\text{p}(\tau_k)\text{d}\tau_k,$$

where

$$\beta_k \,|\, \tau_k^2 \sim \text{Normal}\left(0, \tau_k^2\right),$$
$$\tau_k^2 \sim \text{Exp}\left(\lambda_1^2/2\right),$$

being the exponential distribution parameterized so the mean is $2/{\lambda_1}^2$.

Ridge regression is another regularization model (Hoerl and Kennard [20]). In this case, the prior distribution for the regression parameters β_k is:

$$\beta_k \sim \text{Normal}\left(0, \lambda_2^{-1}\right),$$

i.e., its pdf is proportional to:

$$\text{p}(\beta_k) \propto exp\left\{\frac{\lambda_2}{2}\beta_k^2\right\}$$

The prior distribution restricts the regression parameters (with high probability) to a sphere of radius determined by λ_2.

Finally, the Elastic Net method combines LASSO and Ridge regularization methods [60]. The prior distribution for the regression parameters β_k is:

$$\text{p}(\beta_k) \propto exp\left\{-\lambda_1|\beta_k| - \frac{\lambda_2}{2}\beta_k^2\right\}$$

By using latent variables, it is possible to obtain a scale mixture of normal distributions representation:

$$\beta_k \,|\, \sigma_{\beta_k}^2 \sim \text{Normal}\left(0, \sigma_{\beta_k}^2\right),$$
$$\sigma_{\beta_k}^2 = \left(\tau_k^{-2} + \lambda_2\right)^{-1},$$
$$\tau_k^2 \sim \text{Exp}\left(\lambda_1^2/2\right)$$

3.4. Exploring the posterior distribution

Firstly, the prior distributions are presented. The prior distributions for the regression parameters related to the acoustic features β_k, $k = 1, \ldots, K$, have been defined in Section 3.3. Besides, normal distributions are assumed for the regression parameters related to the exactly known covariates, i.e. $\gamma_h \sim \text{Normal}(c_h, C_h)$, for $h = 1, \ldots, H$, where $c = (c_1, \ldots, c_H)$ and $C = (C_1, \ldots, C_H)$ are fixed values. Inverse Gamma distributions are considered for variances ${\delta_k}^2$, i.e., ${\delta_k}^2 \sim \text{InvGamma}(s_k, r_k)$, where S_k and r_k are the shape and rate parameters, respectively.

Normal distributions are considered for the latent variables, i.e., $w_{ik} \sim \text{Normal}(\mu_k, {\tau_k}^2)$. For the hyperparameters of the latent variables, the prior distributions are defined as $\mu_k \sim \text{Normal}(m_k, {v_k}^2)$ and ${\tau_k}^2 \sim \text{InvGamma}(u_k, t_k)$. The hyperparameters of the regularization methods, ${\lambda_1}^2$ and λ_2, can be fixed values, but they may have hyperprior distributions, e.g., ${\lambda_1}^2 \sim \text{Gamma}(a_1, d_1)$ and $\lambda_2 \sim \text{Gamma}(a_2, d_2)$.

The binary hierarchical model with replications defined in Sections 3.1 and 3.2 results in the likelihood function, considering the observed and the latent variables, given by:

$$\begin{aligned}
&\mathscr{L}\left(\beta, \gamma, \delta^2, \mu, \tau^2 \,|\, y, x, z, w\right) \\
&= \text{p}(\mathbf{y}|z, w, \beta, \gamma)\text{p}\left(x|w, \delta^2\right)\text{p}\left(w|\mu, \tau^2\right) \\
&= \prod_{i=1}^{n}\left\{\text{p}(y_i|z_i, w_i, \beta, \gamma)\left[\prod_{k=1}^{K}\left\{\prod_{j=1}^{J}\text{p}\left(x_{ikj}|w_{ik}, \delta_k^2\right)\right\}\text{p}\left(w_{ik}|\mu_k, \tau_k^2\right)\right]\right\}
\end{aligned} \tag{1}$$

The joint posterior distribution is obtained by using the likelihood function (1) and the prior distributions previously defined, and it is given by:

$$\begin{aligned}
&\text{p}\left(\beta, \gamma, \delta^2, \mu, \tau^2 \,|\, y, x, z, w\right) \\
&\propto \mathscr{L}\left(\beta, \gamma, \delta^2, \mu, \tau^2 \,|\, y, x, z, w\right)\text{p}(\beta)\text{p}(\gamma)\text{p}\left(\delta^2\right)\text{p}(\mu)\text{p}\left(\tau^2\right)\text{p}(\lambda)
\end{aligned} \tag{2}$$

A Markov Chain Monte Carlo (MCMC) algorithm has been implemented in JAGS[1] through the R platform[2] to estimate the posterior distribution. The source code and instructions that allow to run the approach for a simulation-based dataset can be found in the GitHub repository through the link https://github.com/lizbethna/ClassificaReplicaRegulariza.git.

Other Monte Carlo approaches could be applied. For instance, particle filtering could be considered [11]. It deals with targets that are influenced by the proximity and/or behavior of other targets. Also, Hamiltonian Monte Carlo methods can be used. They utilize techniques from differential geometry to generate transitions spanning the full marginal variance [2] or the No-U-Turn sampler, which is an adaptive form of Hamiltonian Monte Carlo sampling [4].

3.5. Determining the most relevant features

After the chain has converged, a random sample for each parameter from the posterior distribution is obtained. Based on the estimated

[1] http://mcmc-jags.sourceforge.net/http://mcmc-jags.sourceforge.net/
[2] https://cran.r-project.org/https://cran.r-project.org/

posterior densities for the regression parameters, a variable selection method based on Bayesian credible intervals is proposed here. For the estimated posterior density of each parameter, this method considers a 95% Bayesian credible interval, being the lower interval limit the 2.5% percentile and the upper one the 97.5% percentile (see, e.g., Hespanhol et al. [18]). The features related to the regression parameters that do not contain 0 in the Bayesian credible interval are selected as relevant features, given the important contribution for predicting the response. Then, the approach is applied to these features to provide accuracy rate, sensitivity, specificity, and AUC-ROC (Area Under the Curve Receiver Operating Characteristic).

The details about the concrete practical implementation considering cross-validation frameworks for variable selection and accuracy metrics are provided in the experimental setting subsection of the results section.

4. Results

4.1. Experimental settings

The replication-based Bayesian regularization approaches in Section 3 are applied to the dataset described in Section 2. The response variable Y takes values Y=0 for healthy subjects and Y=1 for people suffering from Reinke's edema, whereas the 25 acoustic variables have been individually normalized to have mean 0 and standard deviation 1, and the variable sex Z takes values $Z = 0$ for men and $Z = 1$ for women.

The MCMC sampling is applied using the following hyperparameters for the prior distributions. For the regression parameters of the covariates exactly known $\gamma_h \sim$ Normal(0,0.01), for $h = 1, \ldots, H$. For the latent variables in the replications, $w_{ik} \sim$ Normal($\mu_k, {\tau_k}^2$), where $\mu_k \sim$ Normal $(0,1)$, ${\tau_k}^2 \sim$ InverseGamma(1,1), and ${\delta_k}^2 \sim$ InverseGamma(0.01,0.01), for $k = 1, \ldots, K$. For the parameters in the regularization methods, ${\lambda_1}^2 \sim$ Gamma(1,1) and $\lambda_2 \sim$ Gamma(1,1).

A total of 30,000 iterations with a burn-in of 10,000 and a thinning period of 10 generated values are used, providing a sample of length 2000. With these specifications, the chains generated by using the MCMC sampling algorithm seem to have converged. Bayesian Output Analysis (BOA) package was used to perform the convergence analysis [49]. The previous specifications are enough to provide evidence of convergence for all parameters in the three regularization approaches.

Posterior predictive probabilities are obtained for the accuracy metrics. The used metrics are accuracy rate ((TP + TN)/n), sensitivity (TP/(TP + FN)), specificity (TN/(TN + FP)), where TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative. AUC-ROC is also considered.

A stratified cross-validation framework is considered. Specifically, the dataset is randomly split into a training subset composed of 75% of the control subjects (3 men and 20 women healthy) and 75% of the people with Reinke's edema (2 men and 20 women with Reinke's edema) for each iteration. The remaining individuals constitute the testing subset, 25% of healthy people (1 man and 6 women) and 25% with Reinke's edema (1 man and 7 women). This framework is applied for variable selection and, later, for evaluating accuracy metrics by using the selected variables in an independent way, i.e., in each one of the iterations, the partitions are independent. In the first case, the model parameters are determined using the training subset, and the 95% Bayesian credible intervals (built as specified in the Section 3.5) for the model parameters are computed using the testing subset. This procedure is independently repeated 100 times. Then, the variables associated to parameters having more than one non-null 95% credible intervals out of the 100 iterations are selected. This leads to one only set of selected features for the whole cross-validation process. In the second case, once the variables have been selected, the model parameters are determined using the training subset, and the accuracy metrics are computed using the testing subset. This procedure is repeated 100 times and the accuracy metrics are then averaged. Each regularization approach has been trained independently. Note that the second stage has been introduced to test if the concrete set of acoustic features performs well in an independent cross-validation framework. In practical applications, the first stage is applied to select the features, then the classification of the new subjects is done by applying the proposed approach with the selected features without cross-validation.

Three scenarios have been independently considered for each regularization approach, all of them start with the 25 acoustic features plus gender:

1. All the features were used by training and testing with the whole dataset, and later the previously described cross-validation scheme was performed.
2. Common principal components (CPCs) [22] were used to reduce the dimension of the variable space and, then, the approaches were applied to the selected CPCs under the defined cross-validation scheme.
3. The 95% Bayesian credible interval-based approach defined in Section 3.5 was applied to provide the most relevant features based on the previously defined cross-validation framework for variable selection. Then, the approaches are applied to the selected features under the defined cross-validation framework for accuracy metrics.

Finally, an analogous Bayesian credible interval-based approach is applied for the corresponding Bayesian regularization approaches based on independent instances (LASSO, Ridge, and Elastic Net). These methods are designed to be applied to individual instances, i.e., each subject is represented by a feature vector extracted from a single voice recording. Since the database consists of four replications of the sustained /a/ phonation for each subject, four independent cases are considered. The first one uses the first feature vector of each subject, the second case considers the second feature vector of each subject and so on, i.e., the cases are R_1, R_2, R_3 and R_4, where R_j means that only the Jth replication for each individual is used. This leads to four independent experiments with independence-based regularization approaches. The same cross-validation framework for variable selection and accuracy metrics as those defined for the replication-based approaches are used for comparison purposes. Fig. 1 summarizes the experiment capturing within-subject variability and the four experiments based on independent instances, which do not capture the within-subject variability.

Next subsection shows the experimental results obtained for the three scenarios based on replications, and for the four cases of independent instances as well as the comparison among them.

4.2. Experimental results

4.2.1. Replication-based approaches

Firstly, all the acoustic features plus gender were considered by training and testing with the whole dataset, i.e., all the subjects were considered for training and all of them for testing. No differences were found for accuracy rate, sensitivity and specificity, with the three approaches providing the same value of 0.9333 for these three metrics. AUC-ROC results were very close, larger than 0.98. Specifically, 0.9944 for LASSO, 0.9933 for Ridge, and 0.9844 for Elastic Net.

The approaches were applied to all 26 variables with the defined cross-validation scheme for accuracy metrics, and the results are shown in Table 1. The accuracy rates, sensitivities, and specificities are around 0.79, 0.81, and 0.76, respectively, for the three regularization models. The best result was obtained by Elastic Net with an accuracy rate of 0.7927, a sensitivity of 0.8150, and a specificity of 0.7671. The AUC-ROC measures are very close and around 0.88. In general, the differences are very small, so in this scenario very similar results are obtained for the three regularization methods.

The second scenario considers CPCs. Specifically, 75% of the total variability is obtained with eight CPCs. The three regularization methods with the defined cross-validation scheme were applied to these

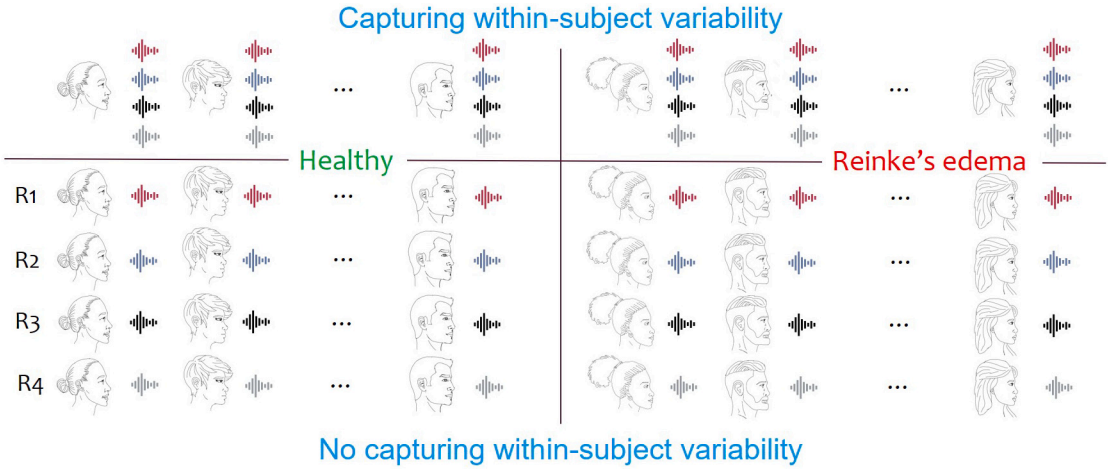

Fig. 1. Graphical scheme of the experiment capturing within-subject variability (top) and the four experiments based on independent instances (bottom).

Table 1
Means and standard deviations of accuracy rate, sensitivity, specificity, and AUC-ROC by using the replication-based regularization models with all the features under the defined cross-validation scheme for accuracy metrics (Scenario 1).

	LASSO	Ridge	Elastic Net
Accuracy rate	0.78733 (0.08352)	0.78533 (0.08128)	0.79267 (0.08254)
Sensitivity	0.80625 (0.13574)	0.80250 (0.13551)	0.81500 (0.13702)
Specificity	0.76571 (0.13548)	0.76571 (0.13548)	0.76714 (0.13568)
AUC-ROC	0.88553 (0.06604)	0.88553 (0.06643)	0.88642 (0.06717)

eight CPCs and the results are shown in Table 2. It can be observed how the loss of information provided lower accuracy rates, being now close to 0.76. The same happens for sensitivity, and specificity, which are around 0.73, and 0.80, respectively. In summary, the accuracy metrics have decreased, but they are still very similar for the three regularization approaches.

The third scenario considers the variable selection based on Bayesian credible intervals that has been previously described. Each replication-based regularization approach selects its own feature set under the defined cross-validation framework for variable selection. Table 3 shows the features selected for the three approaches. Note that LASSO selects 7 features, Ridge 7, and Elastic Net 5. Note that CPP, MFCC4, MFCC7, MFCC10, and SHANNON are selected by the three approaches.

Once the feature sets have been defined for each method, the regularization approaches are applied with the defined cross-validation scheme for evaluating accuracy metrics. The results are presented in Table 4. It can be observed how the best performance is provided by Ridge regression for the four accuracy metrics. The accuracy rate is 0.8893, larger than the ones corresponding to LASSO and Elastic Net, which are 0.8240 and 0.8253, respectively.

Table 5 shows the posterior estimations for the model parameters of the three considered replication-based regularization approaches. These are the mean and standard deviation of the parameter estimates obtained from the 100 iterations in the cross-validation framework. Note that standard deviations for intercept parameters, and parameters associated to GNE and sex are higher than the absolute value of the estimate itself. Therefore, the estimations of these three parameters

Table 2
Means and standard deviations of accuracy rate, sensitivity, specificity, and AUC-ROC by using the replication-based regularization models with eight CPCs under the defined cross-validation scheme for accuracy metrics (Scenario 2).

	LASSO	Ridge	Elastic Net
Accuracy rate	0.76466 (0.09776)	0.76600 (0.09639)	0.76733 (0.10005)
Sensitivity	0.72875 (0.12818)	0.73125 (0.12609)	0.73000 (0.13614)
Specificity	0.80571 (0.14574)	0.80571 (0.14574)	0.81000 (0.14649)
AUC-ROC	0.85410 (0.08973)	0.85464 (0.08977)	0.85625 (0.10005)

Table 3
Acoustic features selected by considering the replication-based regularization approaches under the defined cross-validation framework for variable selection (Scenario 3).

Features	LASSO	Ridge	Elastic Net
GNE	■		
Jitter			
Shimmer			
HNR			
CPP	■	■	■
MFCC1			
MFCC2		■	
MFCC3			
MFCC4	■	■	■
MFCC5			
MFCC6			
MFCC7	■	■	■
MFCC8			
MFCC9			
MFCC10	■	■	■
MFCC11			
MFCC12			
MFCC13			
HURST			
PERMUTATION			
PPE			
RPDE	■	■	
SHANNON	■	■	■
D2			
ZCR			
Total	7	7	5

Table 4
Means and standard deviations of accuracy rate, sensitivity, specificity, and AUC-ROC by using the replication-based regularization models considering the selected features under the defined cross-validation framework for accuracy metrics (Scenario 3).

	LASSO	Ridge	Elastic Net
Accuracy rate	0.82400 (0.09064)	0.88933 (0.07104)	0.82533 (0.07971)
Sensitivity	0.84375 (0.13690)	0.90750 (0.09504)	0.82125 (0.13560)
Specificity	0.80142 (0.14198)	0.86857 (0.12463)	0.83000 (0.13111)
AUC-ROC	0.92232 (0.05568)	0.95500 (0.04450)	0.92160 (0.05930)

Table 5
Means and standard deviations of the parameters for the replication-based regularization models considering the selected features under the defined cross-validation scheme (Scenario 3).

Parameters	LASSO	Ridge	Elastic Net
β_0 Intercept	0.26411 (1.19126)	−0.23393 (0.99422)	−0.46030 (0.62526)
β_1 GNE	1.39850 (1.41908)	–	–
β_5 CPP	8.52840 (5.00575)	9.85189 (5.67987)	2.21863 (0.60491)
β_7 MFCC2	–	8.84516 (5.11999)	–
β_9 MFCC4	−2.72866 (2.37346)	−5.54057 (3.08061)	−1.16507 (0.53925)
β_{12} MFCC7	4.21886 (2.65877)	5.20452 (2.54884)	1.56190 (0.47818)
β_{15} MFCC10	−5.81843 (4.36594)	−6.46859 (4.24449)	−1.94313 (0.70239)
β_{22} RPDE	−3.38629 (2.21472)	−8.18662 (4.04477)	–
β_{23} SHANNON	−2.19405 (1.42865)	−3.67750 (2.58735)	−1.09540 (0.32860)
γ Sex	0.10964 (1.16550)	0.23033 (1.11815)	0.59197 (0.60488)
λ_1	0.59822 (0.13337)	–	0.65412 (0.06362)
λ_2	–	0.18626 (0.07188)	0.39791 (0.09816)

come from dispersed values.

4.2.2. Independence-based approaches

Now the traditional independence-based regularization approaches LASSO, Ridge, and Elastic Net are applied to the four cases R_1, R_2, R_3 and R_4, where R_j means that only the jth replication for each individual is used. Each case is treated independently of the others, so each case contain independent instances.

An analogous Bayesian credible interval-based approach is applied and the features are independently selected for each case. The cross-validation framework defined in Section 4.1 has been also applied in this case. Table 6 shows the selected features for the three traditional regularization-based methods in the four cases. Within each method, the selected features are different for each dataset. There are important differences in the chosen features and in the number of them. The four cases select between 8 and 10 features for LASSO, with only 4 common features. For Ridge, between 7 and 10 features are selected, with 5 common features. Finally, for Elastic Net, there are between 5 and 9 features selected with only 3 of them common. This shows a great variability in number and kind of features within each method for the different cases constituted by the individual replications.

Table 6
Acoustic features selected by considering the traditional independence-based regularization approaches in the four cases under the defined cross-validation framework for variable selection.

Features	LASSO				Ridge				Elastic Net			
	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
GNE	■	■	■		■	■				■		
Jitter												
Shimmer				■				■				■
HNR												
CPP	■	■	■	■	■	■	■	■		■	■	■
MFCC1												
MFCC2	■	■	■		■	■	■	■	■	■		
MFCC3												
MFCC4	■	■		■		■		■		■		■
MFCC5												
MFCC6												
MFCC7	■	■	■	■	■	■	■	■	■	■	■	■
MFCC8												
MFCC9												
MFCC10	■	■	■	■	■	■	■	■	■	■	■	■
MFCC11	■			■	■			■				■
MFCC12												
MFCC13	■		■				■					
HURST												
PERMUTATION	■	■	■	■	■	■	■	■	■	■	■	■
PPE				■								
RPDE		■	■	■		■	■	■		■	■	
SHANNON	■	■		■	■	■		■	■	■		
D2												
ZCR												
Total	10	9	8	10	8	9	7	10	5	9	5	7

The variability in the feature selection considering the four cases is translated into the accuracy metrics. The defined cross-validation scheme is independently applied to each case with their selected features and the results are shown in Table 7. In LASSO approach, accuracy rates ranging from 0.8100 to 0.8580 are obtained for the different cases. Ridge approach provides accuracy rates ranging from 0.8160 to 0.8720, whereas accuracy rates for Elastic Net approach range from 0.8326 to 0.8560. Different results are also obtained for sensitivities, specificities, and AUC-ROC through the four cases.

With this experiment, it has been shown how different results for the selected variables and the accuracy metrics are obtained, depending on the concrete voice recording for each subject being considered. For the first time, it has been demonstrated that having into account the within-subject variability provides more stable results at the same time that relevant features are selected and accuracy metrics keep at good values.

5. Discussion

Bayesian independence-based regularization regression methods have been widely used in many contexts (see, e.g., Kadoya et al. [24]). These methods are based on independent instances as input data. When there exists a dependent nature among some instances, methods that are able to properly address this dependency are demanded. Imperfections in technology and the very biological variability result in acoustic features that are not identical for one specific individual in a particular recording time. This leads to the concept of replication that tries to address the within-subject variability underlying the experimental design. The recording of only one phonation per individual introduces lack of confidence in the process, because if other phonations had been performed, different feature vectors representing the subject would have been obtained and, therefore, the results would have been different. Using independence-based approaches has been the common way to address automatic detection of laryngeal pathologies from speech recordings in the scientific literature [23,31,56].

Table 7
Means and standard deviations of accuracy rate, sensitivity, specificity, and AUC-ROC by using the traditional independence-based regularization approaches in the four cases under the defined cross-validation scheme.

	LASSO	Ridge	Elastic Net
R_1			
Accuracy rate	0.85800 (0.08614)	0.87200 (0.08022)	0.85600 (0.08135)
Sensitivity	0.83875 (0.13680)	0.85875 (0.12264)	0.83875 (0.12346)
Specificity	0.88000 (0.11614)	0.88714 (0.11175)	0.87571 (0.12786)
AUC-ROC	0.92696 (0.06097)	0.94035 (0.05128)	0.92785 (0.05484)
R_2			
Accuracy rate	0.83333 (0.08658)	0.83600 (0.08602)	0.84066 (0.08781)
Sensitivity	0.85000 (0.13176)	0.85375 (0.13301)	0.86000 (0.13328)
Specificity	0.81428 (0.14285)	0.81571 (0.14683)	0.81857 (0.14900)
AUC-ROC	0.91964 (0.06104)	0.92071 (0.06050)	0.92375 (0.05753)
R_3			
Accuracy rate	0.81000 (0.08958)	0.81600 (0.08995)	0.83266 (0.08554)
Sensitivity	0.83000 (0.12997)	0.81250 (0.13588)	0.85125 (0.12897)
Specificity	0.78714 (0.13993)	0.82000 (0.14303)	0.81142 (0.14050)
AUC-ROC	0.91910 (0.05740)	0.93482 (0.05191)	0.94500 (0.04826)
R_4			
Accuracy rate	0.84800 (0.07875)	0.83533 (0.07896)	0.83866 (0.06974)
Sensitivity	0.87625 (0.11581)	0.86875 (0.11148)	0.85000 (0.10952)
Specificity	0.81571 (0.12728)	0.79714 (0.13187)	0.82571 (0.12934)
AUC-ROC	0.92839 (0.05369)	0.93196 (0.05297)	0.91303 (0.07204)

Three regularization-based approaches have been implemented and applied to detect Reinke's edema based on features extracted from replicated voice recordings. The existing within-subject variability for each subject has been statistically addressed by considering that the replicated observations from a feature are measurements with errors of the real underlying feature, which is unknown. In this way, the observed replicated features act as surrogates. This idea allows to build hierarchical models based on latent variables that are handled with Bayesian methodology. Due to the way that the models have been designed, MCMC methods can be used to generate from the posterior predictive distribution.

The three replication-based regularization approaches (LASSO, Ridge, and Elastic Net) consist of variable selection and classification. A total of 26 variables have been considered (25 acoustic features plus gender). Each one provides information that may be useful for voice disorder detection. However, there are many variables to feed the classifiers, some of them highly correlated. This may produce a multicollinearity problem and overfitting. To avoid this, two variable selection approaches have been considered. The first one uses CPCs [22]. Note that this is not a conventional principal component analysis, since CPC analysis allows to properly consider the replicated measurements, because the extracted features display a correlation structure that is stable throughout the replications. This kind of analysis has been widely used in other contexts (see, e.g., [28]). However, it has the disadvantage that none of the CPCs is a feature itself, so no interpretation can be obtained in terms of the disease's effects. The second variable selection approach has been specifically proposed for this problem and it is based on Bayesian credible intervals. Relevant features are obtained from those whose regression parameter estimations do not contain 0. This variable selection method within a cross-validation scheme has provided the selection of relevant features related to the malfunctioning of the voice production system under Reinke's edema. Note that the second stage under the defined cross-validation framework is independently applied to the selected features from the first stage to test if this concrete set of selected features works well for metric performance. This step is not necessary for realtime applications, once the selected variables have been tested.

An analysis of selected features from the experiments based on Bayesian credible intervals reveals that the following five features are selected in the three considered replication-based approaches: CPP, MFCC4, MFCC7, MFCC10 and SHANNON. In the case of the method providing the best accuracy metric results, Ridge, two additional features (MFCC2 and RPDE) have been also selected to complete a seven-feature set. However, when considering the independence-based counterparts applied to the four datasets (each one composed by only one of the four replicated feature vectors for each individual) a great variability of selected features is obtained depending on the voice recording considered, ranging from 5 to 10 features per experiment and a total of 15 different features out of the 25 available acoustic features. This contrasts with the previously reported results for feature selection with replication-based regularization approaches.

The selected features provide information about how the voice production system is failing under Reinke's edema disease. CPP, obtained from the cepstrum of a sound, has shown promising results as an acoustic biomarker of dysphonia [17]. High CPP values correspond to a well-defined harmonic structure, whereas periodicity perturbations (either in amplitude or frequency) lead to a lower amplitude of the cepstral peak. Reinke's edema produces an alteration of vocal-fold vibration patterns which has been quantified by means of CPP. The important role played by MFCCs (with three coefficients selected in the three cases, or even four in the case of Ridge method) may be related to the fact that Reinke's edema patients may produce compensatory articulatory changes in response to altered vocal-fold vibration. These compensatory movements modify the resonance properties of the vocal tract. The selection of SHANNON feature lines up with previous results in the literature showing that entropy measures produce higher values in people with vocal-fold disorders in comparison to healthy ones Scalassara et al. [46]. Pathological speech is characterized by an increase in the signal unpredictability that can be quantified by the use of entropy measures. Finally, RPDE also uses the concept of entropy, in this case, to measure the uncertainty in pitch period estimation. Some physiological aspects of this pathology, such as vocal-fold asymmetry, make it difficult for these patients to maintain a stable vocal fold oscillation. These physiological aspects of Reinke's edema have been shown through the use of high-speed digital imaging and videostroboscopy by Watanabe et al. [57].

From an accuracy metric perspective, the application of the independence-based regularization approaches has also provided a great variability within each regularization method, attaining the best accuracy rate with Ridge regression for the dataset with the first voice recordings (R1). This has shown that different results can be obtained depending on the voice recording considered for each individual. In contrast, the replication-based regularization approaches have provided a reduced number of features, and greater agreement regarding selected features among the three methods, at the same time that good accuracy metrics have been obtained. The best approach has been obtained with Ridge regression, providing an accuracy rate of 0.8893, sensitivity 0.9075, specificity 0.8686, and AUC-ROC 0.9550 (see Table 4). All the four metrics outperform those obtained with the independence-based regularization approaches (see Table 7). Even more, the other comparable approach for variable selection and classification that considers replications, that was developed for Parkinson's disease detection [38], provides worse results in this context. Specifically, when applying that methodology to this dataset with the same cross-validation scheme, lower accuracy metrics were obtained, specifically, an accuracy rate of 0.8120, sensitivity of 0.80375, specificity of 0.82142, and AUC-ROC of 0.8750. Finally, it is remarkable that the combination of selecting a reduced number of relevant features, good accuracy metrics and a rigorous statistical basis make the replication-based regularization approaches worthwhile.

In certain related contexts such as in Parkinson's disease detection by voice recordings, it has become usual to use features extracted from replicated recordings of each subject as if they were independent (see, e.g., Little et al. [29] and Hariharan et al. [14], and references therein). This means that the experimental unit becomes the phonation and not the subject. Given the fact that each subject has several consecutive feature vectors (each one coming from a phonation), which are dependent, a voting-based system is usually established to decide if a subject is classified as healthy or diseased after applying an independence-based classifier to each phonation. In our case, this increases the sample size from 60 subjects (30 healthy and 30 suffering from Reinke's edema) to 240 feature vectors, which are not all independent. This artificial increase of the sample size may or may not provide better accuracy rates, but it provides incoherent results. Specifically, applying a voting system based on independence-based Ridge regularization regression, it is obtained that, for the 30 healthy subjects, 12 of them (40%) had incoherences in their own voice recording classification (not all the voice recordings were assigned to the healthy group), whereas for the 30 people suffering from Reinke's edema 11 of them (36.67%) had incoherences in a similar way. Regarding the accuracy rate, it was obtained 0.8566, which is lower than the corresponding counterpart based on replications, 0.8893. However, this is not always true, for LASSO, the voting system provides an accuracy rate of 0.8440, which is larger than 0.8240, the one from the corresponding counterpart considering within-subject variability. In order to avoid this conceptual and methodological concern, the methods addressing within-subject variability provide an only response for each subject containing all the information from all voice recordings.

The proposed CAD system relies on a voice recording experiment to detect Reinke's edema based on the phonation of the vowel /a/ in a sustained way, a feature extraction process considering a variety of relevant features and a statistical methodology for variable selection and

classification based on Bayesian regularization for replicated covariates. Any of these components could be modified or replaced to try a better approach in different ways. For example, regarding the phonation protocol, other authors have considered other vowels and their combination for detecting voice disorders (see, e.g., Oliveira et al. [39]). It would be interesting to check if it is possible to further decrease the within-subject variability and improve stability by using recordings of different sustained vowels. Another relevant CAD component is feature extraction, since it provides the main ingredient for the classifiers. We have considered an initial set of features that had shown potential in the scientific literature about vocal-fold pathologies, mixing features based on perturbation, cepstral analysis, noise, nonlinear dynamics, and entropies. The proposed variable selection procedure selected the most relevant ones for Reinke's edema detection. However, classification approaches based on replications could be applied with the same benefits to other feature sets as well. For example, PLP coefficients constitute an interesting option to test. Also filtering as RASTA could be studied for PLP coefficients providing RASTA-PLP features [36] that could be tested on databases recorded under mismatched acoustic conditions for Reinke's edema detection. Robustness on environmental noise and recording channel effects in realistic environments is a research topic of great interest that has not been fully addressed up to now for voice disorder detection. Finally, the third CAD component to discuss is the statistical methodology. The regularization-based approaches considered in this paper can be easily modified to handle other methods different from the most usual ones: LASSO, Ridge and Elastic Net. In this Bayesian context, this is achieved through the use of other shrinkage prior distributions. For example, van Erp et al. [55] provided a theoretical and conceptual comparison of nine different shrinkage prior distributions that included local Student's *t*, group LASSO, hyperLASSO, horseshoe, and discrete normal mixture in addition to LASSO, Ridge and Elastic Net. An approach that would need a different framework to handle within-subject variability would be based on nonlinearity. For example, it would be interesting extending artificial neural networks and support vector machine for replicated covariates. In an independent-based approach, they have been used in the diagnosis of voice diseases by automatic speech recognition [50]. The idea of considering replications in a proper way could also be extended to the construction of kernels, which have been successfully developed for independent instances in the problem of semi-supervised learning using a small number of training samples [35].

There is a scientific and technological challenge to develop robust CAD systems that can be incorporated into medical center protocols in such a way that they provide assistance in the diagnosis and monitoring of voice diseases to the health professionals. The proposed system, including or not modifications of its components, could be integrated into a protocol that could be used in primary care as a triage method. This would enable the family doctor to refer the patient to the appropriate hospital department based on an objective criterion that supports his or her basic knowledge of the symptoms.

6. Conclusion

The proposed CAD system capturing within-subject variability due to the multiple replications of voice recordings for each individual constitutes a robust system to address the detection of voice disorders by using acoustic features. The system relies on a voice recording experiment to detect Reinke's edema, a feature extraction process, and variable selection and classification approaches based on Bayesian regularization considering replications.

The replication-based regularization methods provide a more robust approach to the solution of the current problem than the independence-based methods, at the same time that good accuracy metrics and a relevant set of features are selected, which can be interpreted in relation to the effects of Reinke's edema on the voice production mechanisms. This study constitutes a contribution to fill in the gap provided by the lack of within-subject variability management in the scientific literature. Although the approaches have been applied in the context of an experiment specifically designed for Reinke's edema detection, they can be applied to different contexts where the replications play a key role.

Larger experiments containing different voice recording protocols in mismatched acoustic conditions and the study of other signal processing algorithms for feature extraction are issues of interest to improve the CAD system, as well as trying to explore the possible nonlinearity through the development of new replication-based variable selection and classification approaches based on kernels.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Dr. Moreno for his medical advising, and Sandra Paniagua and Esther de la O. for their work recording part of the speech database. It is also acknowledged the collaboration of the patients and healthy people who voluntarily participated in this study.

This research has been funded by *Agencia Estatal de Investigación*, Spain (Project MTM2017-86875-C3-2-R), *Junta de Extremadura*, Spain (Projects IB16054, GR18108 and GR18055), and the *European Union* (European Regional Development Funds). Lizbeth Naranjo has also been partially funded by *UNAM-DGAPA-PAPIIT* (Project IN118720), Mexico. Mario Madruga has been funded by *Ministerio de Universidades* under the doctoral fellowship FPU18/03274.

References

[1] Baghai-Ravary L, Beet SW. *Automatic speech signal analysis for clinical diagnosis and assessment of speech disorders*. Springer briefs in electrical and computer engineering - speech technology. New York: Springer; 2013.
[2] Betancourt M, Girolami M. Hamiltonian Monte Carlo for hierarchical models. In: Dipak US, Dey K, Loganathan A, editors. Current trends in Bayesian methodology with applications. Chapman & Hall/CRC Press; 2015.
[3] Buonaccorsi JP. Measurement error: models, methods and applications. Boca Raton, FL: Chapman and Hall/CRC; 2010.
[4] Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. J Stat Softw 2017;76(1):1–32.
[5] Çomunoglu N, Batur S, Önenerk AM. Pathology of nonneoplastic lesions of the vocal folds. In: Ahmed M, editor. Voice and swallowing disorders. IntechOpen; 2019. p. 126–75.
[6] Cordeiro H, Fonseca J, Guimarães I, Meneses C. Hierarchical classification and system combination for automatically identifying physiological and neuromuscular laryngeal pathologies. J Voice 2017;31(3) (384.E9–384.E14).
[7] Cordeiro HT, Fonseca JM, Ribeiro CM. LPC spectrum first peak analysis for voice pathology detection. Proc Technol 2013;9:1104–11.
[8] Das R. A comparison of multiple classification methods for diagnosis of Parkinson's disease. Expert Syst Appl 2010;37(2):1568–72.
[9] Genkin A, Lewis DD, Madigan D. Large-scale Bayesian logistic regression for text categorization. Technometrics 2007;49(3):291–304.
[10] Godino-Llorente JI, Gomez-Vilda P, Blanco-Velasco M. Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. IEEE Trans Biomed Eng 2006;53(10):1943–53.
[11] Godsill S. Particle filtering: the first 25 years and beyond. In: 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019. p. 7760–4.
[12] Gómez-Garca J, Moro-Velázquez L, Arias-Londoño JD, Godino-Llorente J. On the design of automatic voice condition analysis systems. Part III: Review of acoustic modelling strategies. Biomed Signal Process Control 2021;66:102049.
[13] Gómez-Garca J, Moro-Velázquez L, Godino-Llorente J. On the design of automatic voice condition analysis systems. Part I: review of concepts and an insight to the state of the art. Biomed Signal Process Control 2019;51:181–99.
[14] Hariharan M, Polat K, Sindhu R. A new hybrid intelligent system for accurate detection of Parkinson's disease. Comput Methods Prog Biomed 2014;113(3):904–13.
[15] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Data mining, inference, and prediction. Springer series in statistics. 2nd ed. Springer; 2009.
[16] Hastie T, Tibshirani R, Wainwright M. Statistical learning with sparsity: the lasso and generalizations. In: Chapman & Hall/CRC Monographs on Statistics and Applied Probability. 1st ed. Chapman and Hall/CRC; 2015.

[17] Heman-Ackah YD, Michael DD, Goding Jr GS. The relationship between cepstral peak prominence and selected parameters of dysphonia. J Voice 2002;16(1):20–7.
[18] Hespanhol L, Vallio CS, Menezes Costa L, Saragiotto BT. Understanding and interpreting confidence and credible intervals around effect estimates. Braz J Phys Ther 2019;23(4):290–301.
[19] Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. J Speech Lang Hear Res 1994;37(4):769–78.
[20] Hoerl A, Kennard R. Ridge regression. In: Encyclopedia of statistical sciences. vol. 8. New York: Wiley; 1988. p. 129–36.
[21] Hunter EJ, Tanner K, Smith ME. Gender differences affecting vocal health of women in vocally demanding careers. Logopedics Phoniatrics Vocol 2011;36(3): 128–36.
[22] Jolliffe IT. Principal component analysis. 2nd ed. New York: Springer; 2002.
[23] Kadiri SR, Alku P. Analysis and detection of pathological voice using glottal source features. IEEE J Select Top Signal Process 2019;14(2):367–79.
[24] Kadoya S, Nishimura O, Kato H, Sano D. Regularized regression analysis for the prediction of virus inactivation efficiency by chloramine disinfection. Environ Sci Water Res Technol 2020;6:3341–50.
[25] Kantz H, Schreiber T. Nonlinear time series analysisvol. 7. Cambridge University Press; 2004.
[26] Kob M, Dejonckere P, Calderon E, Kaynar S. Simulation of differences between male and female vocal fold configuration during phonation. In: NAG/DAGA; 2009. p. 1755–6.
[27] Lee J-W, Kang H-G, Choi J-Y, Son Y-I. An Investigation of Vocal Tract Characteristics for Acoustic Discrimination of Pathological Voices. BioMed Research International; 2013 (page ID 758731).
[28] Li H. Accurate and efficient classification based on common principal components analysis for multivariate time series. Neurocomputing 2016;171:744–53.
[29] Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng 2009;56(4):1015–22.
[30] Little MA, McSharry PE, Roberts SJ, Costello DA, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomed Eng Online 2007;6(1):23.
[31] Lopes L, Vieira V, Behlau M. Performance of different acoustic measures to discriminate individuals with and without voice disorders. J Voice 2020 (In press).
[32] Madruga M, Campos-Roca Y, Pérez CJ. Multicondition training for noise-robust detection of benign vocal fold lesions from recorded speech. IEEE Access 2021;9: 1707–22.
[33] Martins RHG, do Amaral HA, Tavares ELM, Martins MG, Gonçalves TM, Dias NH. Voice disorders: etiology and diagnosis. J Voice 2016;30(6) (761.E1–761.E9).
[34] Massachusetts Eye and Ear Infirmary. Voice disorders database, version 1.03 (cd-rom). Lincoln Park, NJ: Kay Elemetrics Corporation; 1994.
[35] Mhaskar H, Pereverzyev SV, Semenov VY, Semenova EV. Data based construction of kernels for semi-supervised learning with less labels. Front Appl Math Stat 2019; 5:21.
[36] Moro-Velazquez L, Gómez-Garca JA, Godino-Llorente JI, Villalba J, Orozco-Arroyave JR, Dehak N. Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson's disease. Appl Soft Comput 2018;62:649–66.
[37] Naranjo L, Pérez CJ, Campos-Roca Y, Martín J. Addressing voice recording replications for Parkinson's disease detection. Expert Syst Appl 2016;46:286–92.
[38] Naranjo L, Pérez CJ, Martín J, Campos-Roca Y. A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications. Comput Methods Prog Biomed 2017;142:147–56.
[39] Oliveira BF, Magalhães DM, Ferreira DS, Medeiros FN. Combined sustained vowels improve the performance of the Haar wavelet for pathological voice characterization. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE; 2020. p. 381–6.
[40] Orozco-Arroyave JR, Belalcazar-Bolaños EA, Arias-Londoño JD, Vargas-Bonilla JF, Skodda S, Rusz J, et al. Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases. IEEE J Biomed Health Inform 2015;19(6):1820–8.
[41] Paniagua MS, Pérez CJ, Calle-Alonso F, Salazar C. An acoustic-signal-based preventive program for university lecturers' vocal health. J Voice 2020;34(1): 88–99.
[42] Park T, Casella G. The Bayesian LASSO. J Am Stat Assoc 2008;103(482):681–6.
[43] Pérez CJ, Naranjo L, Martín J, Campos-Roca Y. A latent variable-based Bayesian regression to address recording replication in Parkinson's disease. In: EURASIP, editor, *proceedings of the 22nd European signal processing conference, EUSIPCO 2014*. Lisbon, Portugal: IEEE; 2014. p. 1447–51.
[44] Sáenz-Lechón N, Godino-Llorente JI, Osma-Ruiz V, Gómez-Vilda P. Methodological issues in the development of automatic systems for voice pathology detection. Biomed Signal Process Control 2006;1:120–8.
[45] Sataloff RT. Clinical assessment of voice. Plural publishing; 2017.
[46] Scalassara PR, Dajer ME, Maciel CD, Guido RC, Pereira JC. Relative entropy measures applied to healthy and pathological voice characterization. Appl Math Comput 2009;207(1):95–108.
[47] Schyberg YM, Bork KH, Sørensen MK, Rasmussen N. Cold-steel phonosurgery of Reinke edema evaluated by the multidimensional voice program. J Voice 2018;32 (2):244–8.
[48] Silva Fonseca E, Capobianco Guido R, Barbon Junior S, Dezani E, Rosseto Gati R, Mosconi Pereira DC. Acoustic investigation of speech pathologies based on the discriminative paraconsistent machine (DPM). Biomed Signal Process Control 2020;55:1–7.
[49] Smith BJ. BOA: an R package for MCMC output convergence assessment and posterior inference. J Stat Softw 2007;21(11):1–37.
[50] Souissi N, Cherif A. Artificial neural networks and support vector machine for voice disorders identification. Int J Adv Comput Sci Appl 2016;7(5):339–44.
[51] Tavares R, Brunet N, Costa SC, Correia S, Neto BGA, Fechine JM. Combining entropy measurements and cepstral analysis for pathological voice assessment. In: ISSNIP biosignals and biorobotics conference 2011; 2011. p. 1–5.
[52] Titze IR. Physiologic and acoustic differences between male and female voices. J Acoust Soc Am 1989;85(4):1699–707.
[53] Travieso CM, Alonso JB, Orozco-Arroyave JR, Vargas-Bonilla J, Nöth E, Ravelo-Garca AG. Detection of different voice diseases based on the nonlinear characterization of speech signals. Expert Syst Appl 2017;82:184–95.
[54] Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Trans Biomed Eng 2012;59(5):1264–71.
[55] van Erp S, Oberski DL, Mulder J. Shrinkage priors for Bayesian penalized regression. J Math Psychol 2019;89:31–50.
[56] Verde L, De Pietro G, Sannino G. A methodology for voice classification based on the personalized fundamental frequency estimation. Biomed Signal Process Control 2018;42:134–44.
[57] Watanabe T, Kaneko K, Sakaguchi K, Takahashi H. Vocal-fold vibration of patients with Reinke's edema observed using high-speed digital imaging. Auris Nasus Larynx 2016;43(6):654–7.
[58] Wu H, Soraghan J, Lowit A, Di Caterina G. Convolutional neural networks for pathological voice detection. In: 2018 40th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018. p. 1–4.
[59] Yamauchi A, Yokonishi H, Imagawa H, Sakakibara K-I, Nito T, Tayama N, et al. Age-and gender-related difference of vocal fold vibration and glottal configuration in normal speakers: analysis with glottal area waveform. J Voice 2014;28(5): 525–31.
[60] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B 2005;67(2):301–20.