

Diferenciação entre Edema de Reinke e Nódulos Vocais através de Parâmetros Não-Lineares da Voz

Gabriel A. R. Silva, Marco A. R. Alves, Bruno C. Bispo, María E. Dajer e Pedro M. Rodrigues

Resumo— Este trabalho tem como objetivo diferenciar distúrbios vocais relacionados com nódulo vocal e edema de Reinke através de parâmetros não-lineares. Os parâmetros são obtidos da banda completa e, por meio da transformada wavelet, de sub-bandas de vogal sustentada. A wavelet que maximiza a capacidade discriminante individual dos parâmetros é buscada. Duas seleções dos parâmetros em sub-bandas são realizadas. Os parâmetros em banda completa e em sub-bandas, e os conjuntos selecionados são aplicados a vários classificadores com validação cruzada *leave-one-out*. Acurácias de classificação de até 86,2% são obtidas sem seleção de parâmetros enquanto acurácias até 100% são alcançadas com parâmetros selecionados.

Palavras-Chave— Nódulo vocal, edema de Reinke, parâmetros não-lineares, wavelet, classificação.

Abstract— This work aims to detect vocal disorders related to vocal nodule and Reinke's edema through non-linear features. The features are computed from the fullband and, by means of the wavelet transform, from sub-bands of sustained vowel. The wavelet that maximizes the individual discriminating capacity of the features is sought. Two sub-band parameter selections are performed. The fullband and sub-band parameters, and the selected sets are applied to several classifiers with leave-one-out cross-validation. Classification accuracies of up to 86.2% are obtained without parameter selection while accuracies of up to 100% are achieved with selected parameters.

Keywords— Vocal nodule, Reinke's edema, nonlinear features, wavelet, classification.

I. INTRODUÇÃO

A voz é uma das principais ferramentas de comunicação humana e sua qualidade impacta diretamente a vida profissional e social, podendo suscitar o desenvolvimento de problemas de ordem psicológica e emocional [1]. Acredita-se que cerca de 10% da população em geral apresenta distúrbios vocais, sendo que este valor aumenta para os 50% quando contabilizados apenas os profissionais com funções comunicacionais [2].

Duas patologias frequentes na clínica de voz são os nódulos vocais e o edema de Reinke. Estas patologias são lesões benignas causadas por abuso, mal uso vocal ou hábitos vocais inadequados [3]. Os nódulos vocais são massas calosas bilaterais localizadas no terço médio das pregas vocais, resultado do choque constante das pregas vocais causado pela contração excessiva dos músculos intrínsecos da laringe [4]. O edema de Reinke é um inchaço das pregas vocais, precisamente na camada superficial da lâmina própria, preenchido por um

Gabriel A. R. Silva, Marco A. R. Alves e Pedro M. Rodrigues, Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina - Laboratório Associado, Escola Superior de Biotecnologia, Porto, Portugal, e-mail: pmrodrigues@ucp.pt; Bruno C. Bispo, Departamento de Engenharia Elétrica e Eletrônica, Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil, e-mail: bruno.bispo@ufsc.br; María E. Dajer, Departamento de Engenharia Elétrica, Universidade Tecnológica Federal do Paraná, Cornélio Procopio-PR, Brasil, e-mail: medajer@utfpr.edu.br.

líquido de viscosidade variável [5]. Esta lesão apresenta uma grande variabilidade clínica podendo ser uni ou bilateral, simétrica ou assimétrica e móvel durante a emissão vocal [6].

Ambas as patologias são lesões de massa que afetam a mesma estrutura anatômica das cordas vocais (espaço de Reinke) e podem apresentar aparência macroscópica semelhante [5], podendo apresentar características acústicas e perceptivas similares. As características compartilhadas por essas lesões são tão amplas que, apesar de diferenças histológicas, sua unificação em um único grupo de estudo foi proposta em [5]. Também há autores que relataram que os nódulos podem ser um estágio subsequente de edemas [7], [8]. Esses fatos evidenciam que, dependendo do estágio evolutivo, diferenciar essas patologias por meio do exame visual laríngeo ou da análise perceptual auditiva pode ser uma tarefa desafiadora. Assim, o desenvolvimento de métodos computacionais para auxiliar na diferenciação dessas patologias se faz necessário.

A pesquisa bibliográfica realizada mostrou que poucos trabalhos tentam diferenciar nódulos vocais e edema de Reinke. A maioria diferencia um grupo saudável e outro patológico, onde nódulos e edema estão incluídos, como [9], [10]. Outros distinguem diferentes grupos patológicos mas, neste caso, nódulos e edema estão no mesmo grupo, como [11], [12]. Em [13], nódulo e edema são considerados grupos diferentes, mas outros grupos de estudo são incluídos em classificações multi-classe. Uma revisão detalhada sobre detecção automática de distúrbios de voz pode ser encontrada em [14].

Esse trabalho tem como objetivo avaliar a aplicação de parâmetros não-lineares de vogal sustentada para diferenciar sinais de voz de pacientes diagnosticados com nódulo vocal e edema de Reinke. Com o intuito de explorar uma possível maior capacidade discriminante de bandas frequenciais, os parâmetros são obtidos da banda completa e, por meio da transformada Wavelet, de sub-bandas dos sinais de voz.

O artigo está organizado da seguinte maneira: a Seção II descreve a base de sinais de voz; a Seção III explica a decomposição multibanda realizada; a Seção IV apresenta os parâmetros não-lineares utilizados; a Seção V apresenta e discute os resultados obtidos; a Seção VI conclui o trabalho.

II. BASE DE DADOS DE VOZ

A base de dados de voz foi fornecida pelo Grupo de Pesquisa em Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (GP/EM/CNPq). Ela consiste em gravações da vogal sustentada /a/ proferidas por homens e mulheres adultos (1 gravação por indivíduo).

As gravações foram realizadas no Ambulatório de Voz do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP) com autorização do

comitê de ética em pesquisa de seres humanos da Universidade Federal de São Carlos, protocolo 256/2010. Elas foram digitalizadas com uma frequência de amostragem de 22050 Hz. A vogal /a/ foi escolhida devido à sua ampla utilização na análise acústica da voz [15]. Como recomendado em [16], a duração de cada fonema foi sempre superior a 2 s.

Neste trabalho, dois grupos de estudo são considerados: 16 pacientes diagnosticados com edema de Reinke e 15 pacientes diagnosticados com nódulos vocais. Os sinais de voz tiveram amplitude normalizada segundo

$$x(n) = \frac{x(n)}{\sum_{n=0}^{N-1} x^2(n)}, \quad (1)$$

onde N é o tamanho do sinal, e componente contínua removida. Em seguida, os sinais foram segmentados em trechos não-sobrepostos com duração de 20 ms usando a janela de Hamming, resultando em 10 segmentos por sinal/indivíduo.

III. DECOMPOSIÇÃO MULTIBANDA

A transformada wavelet discreta (DWT, do inglês *Discrete Wavelet Transform*) de um sinal de tempo discreto é a sua decomposição em um conjunto de funções base obtidas a partir de um número finito de seqüências protótipas e suas versões deslocadas no tempo [17]. Ela é implementada por meio de um banco de filtros em bandas de oitavas e criticamente decimado [17], [18]. Considerando apenas as frequências positivas, a m -ésima sub-banda está limitada a [18]

$$W_m = \begin{cases} [0, \pi/2^S], & m = 0, \\ [\pi/2^{S-m+1}, \pi/2^{S-m}], & m = 1, 2, \dots, S, \end{cases} \quad (2)$$

onde S é o nível de decomposição, $S + 1$ é o número de sub-bandas e π é a frequência angular normalizada.

A DWT utiliza uma função escala de análise $\tilde{\phi}_1(n)$ e uma função wavelet de análise $\tilde{\psi}_1(n)$ que são respostas ao impulso dos filtros de análise passa-baixa e passa-alta de meia banda, respectivamente. Definindo as seguintes fórmulas de recursão

$$\tilde{\phi}_{i+1}(n) = \tilde{\phi}_i(n/2) * \tilde{\phi}_1(n) \quad (3)$$

$$\tilde{\psi}_{i+1}(n) = \tilde{\phi}_i(n) * \tilde{\psi}_1(n/2^i), \quad (4)$$

o filtro de análise equivalente da m -ésima sub-banda é dado por

$$h_m(n) = \begin{cases} \tilde{\phi}_S(n), & m = 0, \\ \tilde{\psi}_{S+1-m}(n), & m = 1, 2, \dots, S. \end{cases} \quad (5)$$

E o sinal da m -ésima sub-banda do sinal $x(n)$ é dado por

$$x_m(n) = \begin{cases} \sum_{k=-\infty}^{\infty} x(k)h_m(2^S n - k), & m = 0, \\ \sum_{k=-\infty}^{\infty} x(k)h_m(2^{S-m+1} n - k), & m = 1, 2, \dots, S. \end{cases} \quad (6)$$

Neste trabalho, a DWT foi aplicada até o terceiro nível, $S = 3$, aos segmentos dos sinais de voz. Os sinais das sub-bandas foram reamostrados para a frequência de amostragem original através do método de interpolação wavelet [19].

IV. ANÁLISE NÃO-LINEAR

Os parâmetros não-lineares utilizados neste trabalho são brevemente descritos abaixo, assumindo que $x(n)$ possui tamanho N e média \bar{x} . Maiores detalhes sobre os parâmetros podem ser encontrados na literatura.

A. Medidas do Atrator Reconstruído

Em cada instante de tempo, o estado de um sistema dinâmico definido por m variáveis pode ser representado por um ponto no espaço m -dimensional. Nesse espaço, denominado espaço de estado ou de fase, a seqüência de estados ao longo do tempo define curvas denominadas trajetórias. Estas trajetórias, quando observadas por longos períodos, podem convergir para uma estrutura geométrica específica e independente das condições iniciais do sistema, denominada atrator [20].

Entre as técnicas para representar o espaço de estado de um sistema dinâmico a partir de uma única série temporal (sinal), processo denominado reconstrução, utilizou-se a técnica de incorporação de atraso. Nesse caso, uma reconstrução do espaço de estado é dada por [20]

$$\mathbf{x}_i = [x(i), x(i + \tau), \dots, x(i + (m - 1)\tau)], \quad (7)$$

onde τ é o atraso de incorporação. Os valores de τ e m foram obtidos de $x(n)$ por meio dos métodos descritos em [20], [21].

A seqüência de vetores $\mathbf{x}_i, i = 1, 2, \dots, M$ onde $M = N - (m - 1)\tau$, formam o atrator reconstruído.

1) *Dimensão da Correlação*: Medida que caracteriza a distribuição dos pontos do atrator, refletindo a complexidade do sistema dinâmico, estimada como [20], [21]

$$D_2 = \lim_{r \rightarrow 0} \frac{\log(C(r, M))}{\log(r)}, \quad (8)$$

onde

$$C(r, M) = \frac{2}{M(M-1)} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M \Theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|) \quad (9)$$

é a probabilidade do par de pontos $\{\mathbf{x}_i, \mathbf{x}_j\}$ no atrator estar separado por uma distância menor ou igual a r e Θ é a função Heaviside [20], [21].

2) *Expoentes de Lyapunov*: Medidas que fornecem informação sobre a evolução das trajetórias ao longo do tempo [21], refletindo a estabilidade do sistema dinâmico [20]. Coeficientes positivos e negativos indicam divergência e convergência exponencial de trajetórias próximas no atrator, respectivamente [22]. Um sistema caótico tem pelo menos um expoente positivo [21], [22]. O maior expoente de Lyapunov (LLE, do inglês *Largest Lyapunov Exponent*) pode ser estimado ao, para cada estado \mathbf{x}_i , encontrar o estado \mathbf{x}_j que satisfaça $\min_j \|\mathbf{x}_i - \mathbf{x}_j\|$ tal que $|i - j| > T_m$, onde T_m é o período médio de $x(n)$. As estimativas são dadas por [23]

$$\lambda(i) = \frac{1}{M+2} \sum_{k=1}^M \frac{1}{kT_s} \ln \frac{\|\mathbf{x}_{i+k} - \mathbf{x}_{j+k}\|}{\|\mathbf{x}_i - \mathbf{x}_j\|}, \quad (10)$$

onde T_s é o período de amostragem. E o LLE é definido pela inclinação da melhor aproximação linear de $\lambda(i)$ [23].

B. Medidas de Memória Longa

1) *Expoente Hurst*: Medida da dependência estatística de longo alcance de uma série temporal, utilizada para avaliar a sua suavidade, propriedades de correlação e grau de auto-similaridade [20], [21]. Pode ser definida como [21]

$$H = \frac{\log(R/S)}{\log(N)}, \quad (11)$$

onde R/S é uma medida estatística denominada intervalo reescalado. Na prática, H é estimado pela inclinação da melhor aproximação linear de $\log[R(n)/S(n)]$ em função de $\log(n)$, onde detalhes sobre o cálculo de $R(n)/S(n)$ podem ser encontrados em [24].

2) *Análise de Flutuação Destendenciada*: Método de análise que fornece um parâmetro para quantificar as correlações de longo alcance (auto-similaridade) de séries temporais aparentemente não-estacionárias [25], [26]. A partir de $x(n)$, calcula-se a série de desvio cumulativo $y(k) = \sum_{i=1}^k [x(i) - \bar{x}]$. Em seguida, para cada trecho de comprimento m de $y(k)$, calcula-se uma aproximação linear denotada por $y_m(k)$. A flutuação média do sinal é definida como

$$F(m) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_m(k)]^2}. \quad (12)$$

A inclinação da melhor aproximação linear de $\log[F(m)]$ em função de $\log(m)$ é o expoente de escala α que representa as propriedades de correlação do sinal $x(n)$ [26].

C. Dimensão Fractal

A dimensão fractal (FD, do inglês *Fractal Dimension*) de um objeto geométrico é uma medida de sua complexidade - extensão espacial, complicação ou propensão ao preenchimento do espaço - e está intimamente ligada à sua forma [27].

Numa forma de onda, a FD é uma ferramenta poderosa para a detecção de eventos transitórios [21]. Ela pode ser calculada diretamente a partir dos sinais, não havendo necessidade de reconstruir o atrator. Entre os algoritmos para determinar a FD de uma forma de onda, os brevemente descritos abaixo são conhecidos por serem adequados para análise de EEG [21].

1) *Algoritmo de Higuchi*: Para $m = 1, \dots, k$ e $k = 1, \dots, k_{max}$, onde k_{max} é obtido experimentalmente apesar de $k_{max} = 8$ ter sido inicialmente proposto, calcula-se a seguinte medida de distância [20], [21]

$$L_m(k) = \frac{N-1}{[a]k} \sum_{i=1}^{[a]} |x(m+ik) - x(m+(i-1)k)|, \quad (13)$$

onde $a = (N-m)/k$ e $[a]$ representa o maior número inteiro igual ou menor que a . A distância média é calculada como $L(k) = \sum_{m=1}^k L_m(k)/k$ para $k = 1, \dots, k_{max}$. A estimativa da FD, denotada por FD_H , é então dada pela inclinação da melhor aproximação linear de $\ln[L(k)]$ em função de $\ln(1/k)$.

2) *Algoritmo de Katz*: Segundo [27], a FD de uma forma de onda $x(n)$ pode ser definida como

$$FD_K = \frac{\log(L/a)}{\log(d/a)}, \quad (14)$$

onde L é a soma das distâncias entre os pontos sucessivos de $x(n)$, a é a distância média entre os pontos sucessivos e d é a maior distância entre $x(1)$ e os restantes pontos de $x(n)$.

D. Energia

A energia das sub-bandas está entre os parâmetros mais utilizados na detecção de disfonias vocais. A energia de $x(n)$ é definida como

$$EN = \sum_{n=1}^N |x(n)|^2. \quad (15)$$

E. Entropia

A entropia é uma medida que descreve a quantidade de desordem, incerteza, em um sistema [21]. No caso de uma variável aleatória, a entropia quantifica a informação média associada às suas observações. As entropias de Shannon e Logarítmica podem ser estimadas como [28], [29]

$$ET_S = - \sum_{n=1}^N |x(n)|^2 \log[|x(n)|^2] \quad (16)$$

e

$$ET_L = \sum_{n=1}^N \log[|x(n)|^2], \quad (17)$$

respectivamente.

A entropia aproximada (ApET) é uma modificação na entropia de Kolmogorov-Sinai, que descreve a taxa na qual informações sobre o estado do sistema dinâmico são perdidas ao longo do tempo, para evitar divergências numéricas quando o sinal é contaminado por ruído, sendo definida como [21]

$$ApET(m, r, N) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log[C_i^m(r)] - \frac{1}{N-m} \sum_{i=1}^{N-m} \log[C_i^{m+1}(r)], \quad (18)$$

onde

$$C_i^m(r) = \frac{1}{N-m+1} \sum_{j=1}^{N-m+1} \Theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|) \quad (19)$$

é a probabilidade do ponto \mathbf{x}_i no atrator estar separado dos outros pontos por uma distância menor ou igual a r .

V. METODOLOGIA E RESULTADOS

Para cada um dos 31 indivíduos, os 10 parâmetros não-lineares (D_2 , LLE , α , FD_H , FD_K , EN , ET_S , ET_L e $ApET$) foram calculados para todos os segmentos (banda completa) e para os sinais das 4 sub-bandas de todos os segmentos. E aplicados a métodos de aprendizado de máquina para diferenciar edema de Reinke e nódulos vocais.

A. Seleção da Wavelet

Como os valores de cada parâmetro calculado em sub-banda dependem da wavelet $\psi_1(n)$ utilizada na decomposição, uma busca foi realizada para encontrar a wavelet que resulta em parâmetros com maior capacidade discriminante. As famílias de wavelets avaliadas foram Biortogonais, Biortogonais Reversas, Daubechies, Coiflets, Symlets e Fejer-Korovkin.

Os valores dos parâmetros foram separados por cada combinação de wavelet, parâmetro e sub-banda. Em cada combinação, isso resulta em 10 valores (1 parâmetro calculado de 1 sub-banda de 10 segmentos) por indivíduo. Dentro de cada combinação, considerando todos os indivíduos, os valores foram normalizados utilizando a Equação (1).

Para cada combinação, os valores normalizados foram aplicados ao classificador binário X-ROC [30]. Esse classificador utiliza a distância quadrática entre médias de distribuições para encontrar um bom limiar para discriminar duas classes

e a curva ROC (do inglês *Receiver Operating Characteristic*) para analisar o resultado da classificação [30]. Selecionou-se a wavelet que, considerando todas as combinações de parâmetro e sub-banda, resultou na maior acurácia média de classificação. A wavelet selecionada foi a Daubechies 36.

B. Seleção de Parâmetros

Como será discutido mais à frente, utilizando a wavelet selecionada, a discriminação entre os grupos de estudo se demonstrou mais eficiente ao utilizar todos os parâmetros obtidos das sub-bandas dos segmentos do que todos os obtidos dos segmentos completos (banda completa). Assim, com o intuito de melhorar a classificação, uma busca foi realizada para encontrar os parâmetros, calculados nas sub-bandas, com maior capacidade discriminante. Não houve seleção dos parâmetros calculados a partir da banda completa dos segmentos.

Os valores não-normalizados dos parâmetros, calculados das sub-bandas dos segmentos usando a wavelet selecionada, foram separados para cada combinação de parâmetro e sub-banda. Em cada combinação, isso resulta em 10 valores (1 parâmetro de 1 sub-banda de 10 segmentos) por indivíduo. Para cada combinação, considerando todos os indivíduos, os valores foram normalizados utilizando a Equação (1).

Em seguida, um algoritmo genético (GA) com critério de entropia [31] e a um algoritmo de seleção *f-score* (FS) [32] foram aplicados de forma a selecionar conjuntos de 2 a 30 parâmetros, extraídos de todos os indivíduos, a serem utilizados como entradas de diferentes classificadores.

C. Processos de Classificação

Diferentes conjuntos de dados foram utilizados como entrada de vários classificadores binários. Primeiramente, foram utilizados todos (sem seleção) os parâmetros obtidos dos segmentos completos (banda completa). Em seguida, todos (sem seleção) os parâmetros obtidos das sub-bandas dos segmentos. Por fim, os conjuntos de parâmetros, calculados em sub-bandas, selecionados pelos algoritmos GA e FS. Nos dois últimos casos, a wavelet selecionada foi utilizada na decomposição dos segmentos.

Quatro classificadores foram avaliados: máquinas de vetor suporte (SVM, do inglês *Support Vector Machines*), K vizinhos mais próximos (KNN, do inglês *K-nearest neighbors*), análise discriminante e árvores de decisão. Diferentes *kernels* foram utilizados em cada classificador. Para verificar a capacidade de generalização dos classificadores, uma validação cruzada *leave-one-out* foi utilizada. Devido ao seu tamanho limitado, toda a base de dados foi usada na validação cruzada.

Os melhores resultados de classificação sem seleção de parâmetros são mostrados na Tabela I. Utilizando todos os parâmetros obtidos dos segmentos completos (banda completa), a maior acurácia alcançada foi de 82,6% com um classificador discriminante linear. Utilizando todos os parâmetros obtidos das sub-bandas dos segmentos, a maior acurácia alcançada foi de 86,2% com um classificador SVM quadrático. Os resultados indicam que os parâmetros não-lineares avaliados nesse trabalho apresentam uma maior capacidade para discriminar edema de Reinke e nódulos vocais quando extraídos das sub-bandas dos segmentos de voz. Esse fato, como mencionado na Seção V-B, motivou a seleção dos parâmetros em sub-bandas.

TABELA I

MAIOR ACURÁCIA DE CLASSIFICAÇÃO SEM SELEÇÃO DE PARÂMETROS.

	Banda completa	Sub-bandas
Acurácia	82,6%	86,2%
Classificador	Discriminante linear	SVM quadrática

TABELA II

MAIOR ACURÁCIA DE CLASSIFICAÇÃO COM SELEÇÃO DE PARÂMETROS.

	Seleção com GA	Seleção com FS
Acurácia	96,6%	100%
Classificador	Regressão logística	KNN cúbico
# Parâmetros	29	11

TABELA III

CONJUNTO DE 11 PARÂMETROS SELECIONADOS PELO FS.

Parâmetro	Sub-banda m
LLE	3
FD_H	0
$ApET$	1
D_2	0
ET_L	1
DFA	1
D_2	1
ET_L	2
EN	1
DFA	2
FD_H	2

Os melhores resultados de classificação com seleção de parâmetros, calculados em sub-bandas, são mostrados na Tabela II. A maior acurácia alcançada com os conjuntos de parâmetros selecionados pelo GA foi de 96,6% com um classificador regressão logística e o conjunto de 29 parâmetros. A maior acurácia alcançada com os conjuntos de parâmetros selecionados pelo FS foi de 100% com um classificador KNN cúbico e o conjunto de 11 parâmetros indicados na Tabela III.

Em comparação com os métodos do estado-da-arte para discriminação entre edema de Reinke e nódulos vocais a partir de sinais de voz, os resultados sem seleção de parâmetros superam os 74,20% obtidos em [33]. E os resultados com seleção de parâmetros também superam os 87% obtidos em [34] e os 88,72% obtidos em [35]. Os trabalhos mencionados da literatura utilizam parâmetros cepstrais multibanda ou coeficientes de predição linear de vogal sustentada.

Os resultados obtidos demonstraram que a combinação de parâmetros não-lineares multibanda de vogal sustentada pode ser adequada para diferenciar distúrbios vocais relacionados com nódulo vocal e edema de Reinke. No entanto, apesar dos resultados promissores, o número de gravações por grupo de estudo disponível na base de dados utilizada é um pouco limitado. Em trabalhos futuros, os resultados devem ser atualizados com uma maior população para garantir a generalização.

VI. CONCLUSÕES

Este trabalho investigou a diferenciação entre distúrbios vocais relacionados com nódulo vocal e edema de Reinke através de parâmetros não-lineares. Os parâmetros são obtidos da banda completa e, por meio da transformada wavelet, de sub-bandas de segmentos de vogal sustentada.

Todos os parâmetros foram utilizadas como entrada para vários classificadores com validação cruzada *leave-one-out*.

Classificações com acurácias de 82,8% e 86,2% foram obtidas com os parâmetros calculados em banda completa e sub-bandas, respectivamente, indicando que os parâmetros não-lineares avaliados apresentam uma maior capacidade discriminante quando extraídos das sub-bandas da voz.

Seleções dos parâmetros calculados em sub-bandas foram realizadas através de um algoritmo genético e um algoritmo *f-score*. Acurácias de classificação de 96,6% e 100% foram obtidas com parâmetros em sub-bandas selecionados, superando os resultados dos métodos do estado-da-arte. Apesar de indicarem que a combinação de parâmetros não-lineares multibanda de sinais de vogal sustentada é adequada para discriminar os grupos de estudo, os resultados devem ser atualizados com uma população maior para garantir a generalização.

AGRADECIMENTOS

Esse trabalho foi apoiado pelos Fundos Nacionais da Fundação para a Ciência e a Tecnologia (FCT), Portugal, através do projeto UIDB/50016/2020.

REFERÊNCIAS

- [1] S. Krischke, S. Weigelt, U. Hoppe, V. Köllner, M. Klotz, U. Eysholdt, and F. Rosanowski, "Quality of life in dysphonic patients," *Journal of Voice*, vol. 19, no. 1, pp. 132–137, Mar. 2005.
- [2] R. H. G. Martins, H. A. do Amaral, E. L. M. Tavares, M. G. Martins, T. M. Gonçalves, and N. H. Dias, "Voice disorders: Etiology and diagnosis," *Journal of Voice*, vol. 30, no. 6, pp. 761.e1–761.e9, Nov. 2016.
- [3] R. T. Sataloff, M. J. Hawkshaw, and J. B. Sataloff, "Common medical diagnoses and treatments in patients with voice disorders: An introduction and overview," in *Clinical assessment of voice*, 2nd ed., R. T. Sataloff, Ed. San Diego, CA: Plural Publishing, 2017.
- [4] R. H. G. Martins, J. Defaveri, M. A. C. Domingues, R. de Albuquerque e Silva, and A. Fabro, "Vocal fold nodules: morphological and immunohistochemical investigations," *Journal of Voice*, vol. 24, no. 5, pp. 531–539, Sep. 2010.
- [5] A. Hantzakos, M. Remacle, F. G. Dikkers, J.-C. Degols, M. Delos, G. Friedrich, A. Giovanni, and N. Rasmussen, "Exudative lesions of reinke's space: a terminology proposal," *European Archives of Otorhinolaryngology*, vol. 266, pp. 869–878, 2009.
- [6] B. M. J. Neves, J. G. Neto, and P. Pontes, "Diferenciação histopatológica e imunoistoquímica das alterações epiteliais no nódulo vocal em relação aos pólipos e ao edema de laringe," *Revista Brasileira de Otorrinolaringologia*, vol. 70, no. 4, pp. 439–448, Jul. 2004.
- [7] K. W. Altman, "Vocal fold masses," *Otolaryngologic Clinics of North America*, vol. 40, no. 5, pp. 1091–1108, Oct. 2007.
- [8] R. E. Hillman, E. B. Holmberg, J. S. Perkell, M. Walsh, and C. Vaughan, "Phonatory function associated with hyperfunctionally related vocal fold lesions," *Journal of Voice*, vol. 4, no. 1, pp. 52–63, Jan. 1990.
- [9] M. K. Arjmandi and M. Pooyan, "An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine," *Biomedical Signal Processing and Control*, vol. 7, no. 1, pp. 3–19, Jan. 2012.
- [10] P. Saidi and F. Almasganj, "Voice disorder signal classification using m-band wavelets and support vector machine," *Circuits, Systems, and Signal Processing*, vol. 34, no. 8, p. 2727–2738, Aug. 2015.
- [11] H. Cordeiro, J. Fonseca, I. Guimarães, and C. Meneses, "Voice pathologies identification speech signals, features and classifiers evaluation," in *Signal Processing: Algorithms, Architectures, Arrangements, and Applications*, Poznan, Poland, Sep. 2015.
- [12] —, "Hierarchical classification and system combination for automatically identifying physiological and neuromuscular laryngeal pathologies," *Journal of Voice*, vol. 31, no. 3, pp. 384.e9–384.e14, May 2017.
- [13] R. Behroozmand and F. Almasganj, "Comparison of neural networks and support vector machines applied to optimized features extracted from patients' speech signal for classification of vocal fold inflammation," in *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology*, Athens, Greece, Dec. 2005, pp. 844–849.
- [14] S. Hegde, S. Shetty, S. Rai, and T. Dodderi, "A survey on machine learning approaches for automatic detection of voice disorders," *Journal of Voice*, vol. 33, no. 6, pp. 947.E11–947.E33, Nov. 2019.
- [15] A. Castellana, A. Carullo, S. Corbellini, and A. Astolfi, "Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 3, pp. 646–654, Mar. 2018.
- [16] R. F. Coleman, "Sources of variation in phonetograms," *Journal of Voice*, vol. 7, no. 1, pp. 1–14, Mar. 1993.
- [17] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, New Jersey: Prentice Hall, 1995.
- [18] H. S. Malvar, *Signal Processing with Lapped Transforms*. Norwood, Massachusetts: Artech House, 1992.
- [19] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine*, vol. 8, no. 4, pp. 14–38, October 1991.
- [20] G. R. Bermúdez and P. J. G. Laencina, "Analysis of EEG signals using nonlinear dynamics and chaos - a review," *Applied Mathematics & Information Sciences*, vol. 9, pp. 1–13, 01 2015.
- [21] O. Faust and M. G. Bairy, "Nonlinear analysis of physiological signals: a review," *Journal of Mechanics in Medicine and Biology*, vol. 12, no. 04, pp. 1–21, Sep. 2012.
- [22] C. J. Stam, "Nonlinear dynamical analysis of EEG and MEG: review of an emerging field," *Clinical Neurophysiology*, vol. 116, no. 10, pp. 1388–2457, Oct. 2005.
- [23] M. T. Rosenstein, J. J. Collins, and C. J. D. Lucaa, "A practical method for calculating largest Lyapunov exponents from small data sets," *Physica D: Nonlinear Phenomena*, vol. 65, no. 1, pp. 117–134, May 1993.
- [24] B. Qian and K. Rasheed, "Hurst exponent and financial market predictability," in *Proceedings of IASTED International Conference on Financial Engineering and Applications*, Berkeley, USA, September 2007, pp. 203–209.
- [25] P. Gifani, H. R. Rabiee, M. Hashemi, P. Taslimi, and M. Ghanbari, "Optimal fractal-scaling analysis of human EEG dynamic for depth of anesthesia quantification," *Journal of the Franklin Institute*, vol. 344, no. 3–4, pp. 212–229, 2007.
- [26] J.-M. Lee, D.-J. Kim, I.-Y. Kim, K.-S. Park, and S. I. Kim, "Detrended fluctuation analysis of EEG in sleep apnea using MIT/BIH polysomnography data," *Computers in biology and medicine*, vol. 32, no. 1, pp. 37–47, 2002.
- [27] M. J. Katz, "Fractals and the analysis of waveforms," *Computers in Biology and Medicine*, vol. 18, no. 3, pp. 145–156, January 1988.
- [28] J. Rodríguez-Sotelo, A. Osorio-Forero, A. Jiménez-Rodríguez, D. Cuesta-Frau, E. Cirugeda-Roldán, and D. Peluffo, "Automatic sleep stages classification using EEG entropy features and unsupervised pattern analysis techniques," *Entropy*, vol. 16, no. 12, pp. 6573–6589, Dec. 2014.
- [29] A. B. Das and M. I. H. Bhuiyan, "Discrimination and classification of focal and non-focal EEG signals using entropy-based features in the EMD-DWT domain," *Biomedical Signal Processing and Control*, vol. 29, pp. 11–21, August 2016.
- [30] C. T. Nakas and C. T. Yiannoutsos, "Ordered multiple-class ROC analysis with continuous measurements," *Statistics in Medicine*, vol. 23, no. 22, pp. 3437–3449, Nov. 2004.
- [31] S. K. Smit and A. E. Eiben, "Using entropy for parameter analysis of evolutionary algorithms," in *Experimental Methods for the Analysis of Optimization Algorithms*, T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuss, Eds. Berlin, Heidelberg: Springer, 2010, pp. 287–310.
- [32] Y.-W. Chen and C.-J. Lin, "Combining SVMs with various feature selection strategies," in *Feature Extraction: Foundations and Applications*, I. Guyon, M. Nikravesh, S. Gunn, and L. A. Zadeh, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 315–324.
- [33] M. A. R. Alves, G. A. R. Silva, B. C. Bispo, M. E. Dajer, and P. M. Rodrigues, "Detecção de distúrbios vocais através de parâmetros cepstrais multibanda de vogal sustentada," in *Anais do XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, Florianópolis, Brazil, Nov. 2020.
- [34] H. T. Cordeiro, J. M. Fonseca, and C. M. Ribeiro, "LPC spectrum first peak analysis for voice pathology detection," *Procedia Technology*, vol. 9, pp. 1104–1111, Dec. 2013.
- [35] M. Alves, G. Silva, B. C. Bispo, M. E. Dajer, and P. M. Rodrigues, "Voice disorders detection through multiband cepstral features of sustained vowel," *Journal of Voice (in press)*, 2021.