

Classificação Automática de Patologias da Laringe Usando Wavelets e Sistemas Inteligentes

Rafael Alberto dos Santos * Paulo Rogério Scalassara * Wagner Endo *

* Departamento Acadêmico da Elétrica, Universidade Tecnológica Federal do Paraná, PR, (e-mail: rafaelasantos.2015@alunos.utfpr.edu.br, [prscalassara, wendo]@utfpr.edu.br).

Abstract: Acoustic analysis is a digital signal processing technique, non-invasive, capable of assisting in the assessment of voice disorders. One of the methods used in this acoustic analysis is to obtain characteristics that discriminate the signals. Therefore, a proposal is presented that uses the sustained vowel “a” as an input signal to extract relevant information from the signals. Information extraction is obtained by the wavelet transform, followed by the calculation of energy and entropy of the resulting coefficients. Finally, these characteristics are classified by the classifier support vector machine. The classification of the healthy and pathological classes obtained an average accuracy of 94.32% for the energy quantity, 93.83% for the entropy quantity and with the combined quantities 94.32%. For the classification of pathologies only, it obtained an average accuracy of 96.30% for the energy quantity, 95.93% for the entropy quantity and 96.67% for the combined quantities.

Resumo: A análise acústica é uma técnica de processamento digital de sinais, não invasiva, capaz de auxiliar na avaliação de distúrbios da voz. Um dos métodos utilizados nesta análise acústica consiste em obter características que discriminem os sinais de forma automática. Portanto apresenta-se uma proposta que utiliza como sinal de entrada a vogal “a” sustentada para extrair informações relevantes dos sinais. A extração de informação é obtida pela transformada *wavelet*, seguida do cálculo de energia e entropia dos coeficientes resultantes. Por fim estas características são classificadas pelo classificador *support vector machine*. A classificação das classes saudáveis e patológicas obteve um acerto médio de 94.32% para a grandeza energia, 93.83% para a grandeza de entropia e com as grandezas juntas 94.32%. Para a classificação apenas das patologias obteve um acerto médio de 96.30% para a grandeza energia, 95.93% para a grandeza entropia e 96.67% para as grandezas juntas.

Keywords: feature; wavelet; pathology; support vector machine; voice..

Palavras-chaves: característica; wavelet; patologia; support vector machine; voz.

1. INTRODUÇÃO

O mau uso e a hiperfunção da voz podem causar patologias da laringe tais como: edemas, nódulos e pólipos (Ramig and Verdolini, 1998). Nódulos e edemas de Reinke são patologias da laringe, mais especificamente, localizadas nas pregas vocais. Estes tipos de patologias são, descritas geralmente como disfonias organofuncionais que apresentam lesões secundárias, com base funcional (Cielo et al., 2011; Dajer, 2010).

A avaliação da voz pode ser feita usando métodos, tais como: avaliação percepto-auditiva, análise visual da laringe e análise acústica (Nemr et al., 2005). A avaliação percepto-auditiva é considerada subjetiva, pois depende de um profissional treinado escutar e interpretar a voz do paciente (Oates, 2009). A análise visual da laringe pode ser feita pela técnica de videolaringoscopia que utiliza um endoscópio rígido, acoplado a um sistema de vídeo para o registrar imagens, portanto uma técnica invasiva. A análise

acústica é uma técnica de processamento digital de sinais não invasiva, pois os sinais são gravados e processados digitalmente em computadores. Esse método pode auxiliar na avaliação dos distúrbios da voz (Dajer, 2010; Scalassara et al., 2009a), como no caso da avaliação pós operatório de cirurgia da tireoide, onde a análise acústica pode ser utilizada como avaliação, diminuindo a necessidade da laringoscopia (Ortega et al., 2009).

Os distúrbios da voz são, geralmente, caracterizados por alterações da qualidade, frequência natural e intensidade da voz (Ramig and Verdolini, 1998). Por causa das alterações nos sinais de voz, alguns métodos de análise acústica buscam medir de forma direta as variações da frequência natural (*jitter*) e da amplitude (*shimmer*) em sinais saudáveis e patológicos, com o objetivo de classificá-los (Scalassara et al., 2007; Teixeira et al., 2013; Verde et al., 2018). Outra abordagem para classificação dos sinais é a utilização de medidas de informação como no trabalho de Scalassara et al. (2009b) que apresenta um método para diferenciar sinais patológicos de saudáveis através da medida da entropia relativa entre os sinais.

* Os autores agradecem a Fundação Araucária pelo suporte ao projeto.

O trabalho de Arjmandi and Pooyan (2012), utiliza *wavelet packet* para decompor os sinais, gerando os coeficientes de aproximação e detalhe. Então, é realizado o cálculo da energia e entropia dos coeficientes, a fim de discriminar os sinais a partir destas características. A premissa de medir a energia e entropia dos coeficientes da transformada *wavelet*, é que estas características fornecem informações das irregularidades das pregas vocais (Arjmandi and Pooyan, 2012). Para a etapa de classificação, uma das metodologias mais populares na análise automática das condições vocais, é o *support vector machine* (SVM) (Gómez-García et al., 2019).

O presente trabalho utiliza a decomposição por *wavelet packet* (WPD) para decompor os sinais de voz e, a partir dos coeficientes gerados, calcular a energia e entropia. Os valores de energia e entropia formam o vetor de característica, sendo este, o parâmetro de entrada de um classificador do tipo SVM.

2. TEORIA

2.1 Patologias da Laringe

O presente trabalho utiliza duas patologias para a análise, sendo elas: nódulo e edema de Reinke.

O nódulo é definido como uma lesão pequena que ocorre nos dois lados das pregas vocais e geralmente é imóvel durante a fonação (Wallis et al., 2004). Nódulos são associados a um comportamento vocal inadequado e abusivo (Ramig and Verdolini, 1998; Behlau, 2005). Do ponto de vista acústico, a frequência fundamental é mais grave ou normal, com maiores índices de perturbação (*jitter* e *shimmer*) e com maior ruído na fala (Behlau, 2005).

O edema de Reinke é definido como uma lesão difusa na camada superficial da prega vocal, tem coloração rosada, sendo caracterizada pelo acúmulo de fluido irregular, em alguma região da porção membranosa ou em toda ela (Behlau, 2005). O edema de Reinke está associado com o uso intensivo da voz, abusos vocais variados e tabagismo (Behlau, 2005; Omori, 2011). As características acústica causadas pela presença do edema são, frequência fundamental da voz mais baixa que a esperada para sexo e idade e grandes alterações no *shimmer* (Behlau, 2005).

2.2 Decomposição por Transformada Wavelet

A transformada de *wavelet* discreta (DWT) pode ser entendida, do ponto de vista da análise de multirresolução, como a decomposição do sinal em duas bases de funções, geralmente ortogonais. As bases de funções são conhecidas como função escala, ϕ , e *wavelet*, ψ . Estas funções podem ser definidas usando versões deslocadas e dilatadas da função escala e também de um par de filtros passa-baixas, h_n , e passa-altas, g_n , conforme (1) e (2), respectivamente.

$$\phi[n] = \sum_{k \in \mathbb{Z}} h_n \phi[2n - k] \quad (1)$$

$$\psi[n] = \sum_{k \in \mathbb{Z}} g_n \phi[2n - k] \quad (2)$$

Os coeficientes gerados na decomposição são chamados de coeficientes de aproximação e coeficientes de detalhe.

A DWT é um processo iterativo, onde a decomposição ocorre sempre nos coeficientes de aproximação, gerando novos coeficientes para análise.

A implementação das transformadas *wavelets* discretas podem ser vistas como um caso especial de bancos de filtros criticamente decimados (Diniz et al., 2014). Para realizar a decomposição do primeiro nível da DWT, efetua-se a convolução do sinal de entrada, $x[n]$, com os filtros, h_n e g_n . O resultado é a divisão do sinal, $x[n]$, em duas partes: coeficiente de aproximação que contém informações das baixas frequências e coeficientes de detalhe que contém informações das altas frequências. Após a convolução os coeficientes são decimados por 2, visto que, apenas metade da banda permanece nos coeficientes.

A WPD é a generalização entre a aproximação de multirresolução e *wavelets* (Mallat, 2008). Portanto, ao contrário da DWT, a decomposição acontece nos coeficientes de aproximação e de detalhe, o resultado é uma análise mais minuciosa do sinal. A Figura 1 ilustra os componentes do processo WPD de terceiro nível.

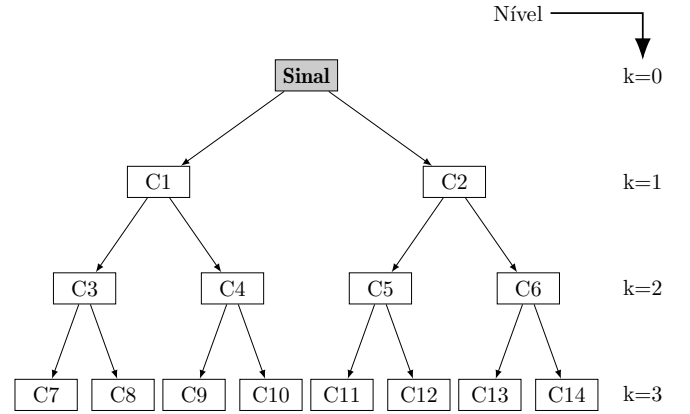


Figura 1. Diagrama do terceiro nível da *wavelet packet decomposition*.

Após essa etapa, duas grandezas que podem ser obtidas para quantificar os coeficientes são a energia e a entropia, sendo calculadas conforme (3) e (4), respectivamente.

$$Energia_j = \frac{1}{K} \sum_{k=1}^K |C_{k,j}|^2 \quad (3)$$

$$Entropia_j = -\frac{1}{K} \sum_{k=1}^K |C_{k,j}|^2 \ln |C_{k,j}|^2 \quad (4)$$

Sendo K o número de amostras do coeficiente no nível e j variando do primeiro coeficiente no nível até o último.

2.3 Classificação Utilizando SVM

O SVM é um algoritmo de aprendizagem supervisionada que utiliza hiperplanos de decisão, a fim de separar classes. A Equação (5) define um hiperplano de decisão para classificar duas classes linearmente separáveis.

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (5)$$

Sendo $\phi(\mathbf{x})$ a transformação fixa do vetor de característica $\mathbf{x}$, $\mathbf{w}$ é conhecido como vetor de peso e b é o parâmetro de tendência. O conjunto de dados de treinamento são M vetores de entrada, $\mathbf{x}_1, \dots, \mathbf{x}_M$, e valores alvo, $t_1, \dots, t_M$, com $t_m \in \{-1, 1\}$ (Bishop, 2006). Novos dados são classificados de acordo com o sinal de $y(\mathbf{x})$.

Supondo que exista ao menos um parâmetro $\mathbf{w}$ e b , tal que (5) seja, $y(\mathbf{x}_m) > 0$ para pontos com $t_m = +1$ e $y(\mathbf{x}_m) < 0$ para pontos $t_m = -1$, então, para todos os dados de treino $t_m y(\mathbf{x}_m) > 0$.

Para resolver (5) é necessário algum critério de otimização. A abordagem do SVM é através do conceito de margem, definido como a menor distância entre o limite de decisão e qualquer uma das amostras (Bishop, 2006). A Figura 2 ilustra o caso de duas dimensões, sendo o hiperplano uma linha e os vetores de suporte na cor vermelha.

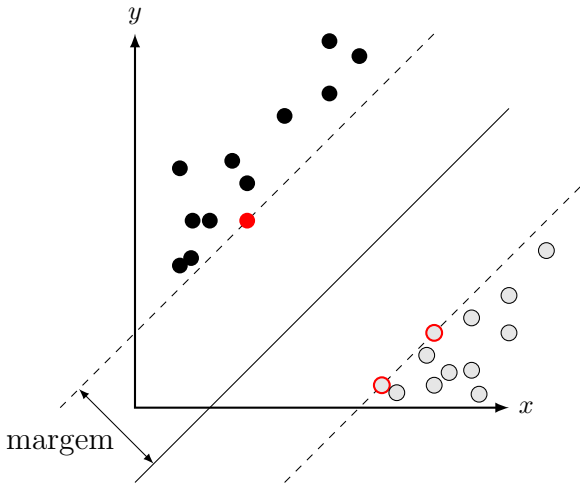


Figura 2. Hiperplano de decisão separando linearmente duas classes.

A distância perpendicular do ponto $\mathbf{x}$ para o hiperplano definido, $y(\mathbf{x}) = 0$, é dado por $|y(\mathbf{x})|/\|\mathbf{w}\|$. Se todos os pontos são corretamente classificados, ou seja, $t_m y(\mathbf{x}_m) > 0$ para todo m , a distância de um ponto $\mathbf{x}_m$ para a superfície de decisão pode ser escrita conforme (6).

$$t_m (\mathbf{w}^T \phi(\mathbf{x}_m) + b) \geq 1, \quad m = 1, \dots, M \quad (6)$$

Pode-se resolver o problema de otimização minimizando $\|\mathbf{w}\|^2$, sujeito as restrição de (6).

Se os dados de treino não são linearmente separáveis, pode não ser viável encontrar $\mathbf{w}$ e b que satisfaça a restrição de (6) (Burges, 1998).

Para dados não linearmente separáveis, pode-se relaxar a restrição (6), pela introdução de variáveis de folga (Bishop, 2006; Burges, 1998), $\xi_m \geq 0$ onde $m = 1, \dots, M$, conforme (7).

$$t_m (\mathbf{w}^T \phi(\mathbf{x}_m) + b) \geq 1 - \xi_m, \quad m = 1, \dots, M \quad (7)$$

Para pontos corretamente classificados $\xi_m = 0$. Pontos dentro da margem, mas do lado correto da classe $0 < \xi_m \leq 1$. Para pontos incorretamente classificados $\xi_m > 1$. Para

maximizar a margem, penalizando os pontos classificados incorretamente, minimiza-se (8) (Bishop, 2006).

$$C \sum_{m=1}^M \xi_m + \frac{1}{2} \|\mathbf{w}\|^2 \quad (8)$$

Onde o parâmetro $C > 0$ controla a troca entre a variável de folga e a margem. Utiliza-se a técnica de multiplicadores de Lagrange para solucionar (8), levando a (9).

$$L(\mathbf{a}) = \sum_{m=1}^M a_m - \frac{1}{2} \sum_{m=1}^M \sum_{j=1}^M a_m a_j t_m t_j k(\mathbf{x}_m, \mathbf{x}_j) \quad (9)$$

Sujeito as restrições de (10).

$$\sum_{m=1}^M a_m t_m = 0; \quad 0 \leq a_m \leq C, \quad m = 1, \dots, M \quad (10)$$

Sendo definido $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}')$ como função núcleo. Pode-se classificar novos dados modificando (5), conforme (11).

$$y(\mathbf{x}) = \sum_{m=1}^M a_m t_m k(\mathbf{x}, \mathbf{x}_m) + b \quad (11)$$

3. MATERIAIS E MÉTODOS

3.1 Banco de Dados

Os sinais de voz utilizados neste trabalho foram obtidos do banco de dados Saarbruecken Voice Database (Barry and Pützer, 2007). Os sinais obtidos são gravações da vogal “a” sustentada, com frequência de amostragem de 50000 Hz no formato wav. Foram obtidos 45 sinais, sendo 32 de mulheres e 13 de homens.

Os sinais são divididos em três grupos, contendo 15 sinais de cada, com pessoas em seguintes situações: saudável, nódulo e edema de Reinke. A Tabela 1 mostra os grupos e a variação de idade.

Tabela 1. Sinais da vogal “a” sustentada.

Grupos	Sinais	Idades
Saudável	15	45-65
Edema	15	43-65
Nódulo	15	19-64

A Figura 3 mostra um sinal de cada classe, sendo a Figura 3(a) saudável, 3(b) edema e 3(c) nódulo. Estes sinais foram limitados em 30 milissegundo para melhor visualização.

A Figura 3(a) mostra a quase periodicidade da vogal “a” sustentada. Na Figura 3(b) é possível ver a quase periodicidade, entretanto, nota-se a deformação no sinal em relação a Figura 3(a). O sinal da Figura 3(c) não se parece com os demais sinais, apesar de, ainda demonstrar a quase periodicidade.

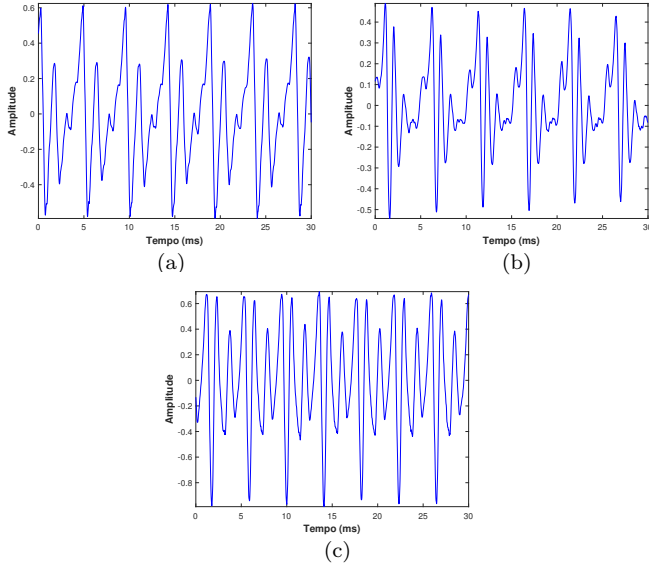


Figura 3. Sinais da vogal /a/ sustentada. (a) saudável. (b) edema. (c) nódulo.

3.2 Metodologia

No presente trabalho são realizados dois experimentos. No primeiro experimento é realizado a classificação da classe saudável em relação as duas patologias. O segundo experimento é a classificação somente das patologias.

O fluxograma da Figura 4 apresenta os passos para o desenvolvimento do método proposto. Todas as etapas serão explicadas a seguir.

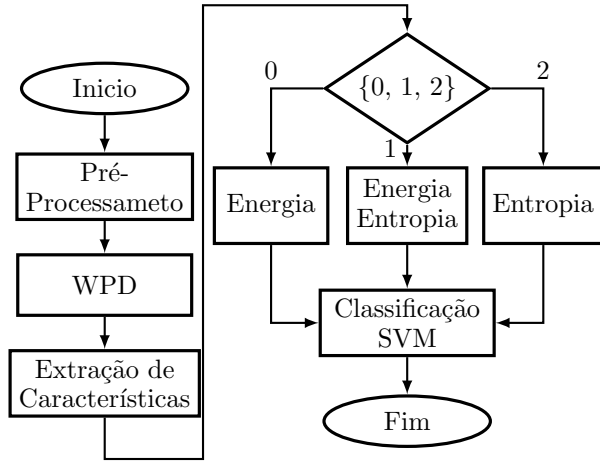


Figura 4. Método proposto para identificar patologias da laringe.

• Pré-Processamento

Devido a variação da quantidade de amostras nos sinais de voz, optou-se pela padronização da quantidade em $N = 25000$ amostras, que equivale a um sinal de meio segundo.

O pré-processamento também inclui janelar os sinais para análise, sendo utilizado no método a janela de Hanning.

Escolheu-se o comprimento da janela de $L = 5000$ amostras com 50% de sobreposição, resultando em nove janelas para cada sinal. Portanto o total de janelas a ser classificada são de 405, ou seja, nove vezes 45. Todo sinal em análise tem sua energia média normalizada.

• WPD

Como no trabalho Arjmandi and Pooyan (2012), que busca a classificação de desordem da voz, a decomposição é realizada até o sexto nível da WPD. A escolha da função *wavelet* para efetuar a análise, foi baseado no trabalho Guido et al. (2005), que testou diferentes *wavelets* para a classificação de desordens da voz. A família Daubechies de filtros *wavelet* de ordem 20, mostrou bons resultados e foi a escolhida para o presente trabalho.

Para cada janela do sinal é realizada a WPD até o sexto nível, que resulta nos componentes C63 a C126. A informação contida no sinal de voz, geralmente, tem uma banda frequência de aproximadamente 4000 Hz, portanto, não seria útil utilizar todos os coeficientes da transformada. Embora o sinal contenha informações relevantes em apenas alguns componentes, é necessário saber quais os componentes relevantes para o método. Portanto, foi realizada uma análise prévia e notou-se que os componentes C63 a C72 apresentam maiores informações para discriminação dos sinais. Então, para o presente trabalho utiliza-se apenas os componentes C63 a C72.

• Extração de Características

De posse dos coeficientes da WPD, é realizado o cálculo da energia e entropia, conforme (3) e (4), sendo gerado os vetores de características que são os parâmetros de entrada do algoritmo de classificação. Como a decomposição por *wavelet packet* gerou dez coeficientes, o vetor de características contém dez características para cada grandeza. Portanto para a etapa de classificação pode-se escolher o vetor de característica em três distintas configurações: energia, entropia e energia com entropia, sendo que para energia com entropia o vetor de característica contém com o dobro de características.

• Classificação SVM

Nessa etapa, o vetor de característica é separado em conjuntos de treino e teste para os seguintes casos: energia, entropia e energia com entropia. A separação do conjunto de treino e teste foi de 80% do vetor para treino e 20% para teste. Nota-se que, quando são utilizadas ambas as características para a classificação, o vetor tem o dobro de características. O *kernel* utilizado no algoritmo SVM foi o *Radial Basis Function* (RBF), devido a testes prévios.

Para validar o modelo foi utilizado validação cruzada, especificamente, o método *k-fold* com $k = 5$. Antes da separação dos *k-folds*, cada classe tem suas características embaralhadas aleatoriamente.

4. RESULTADOS

A Figura 5 mostra a média e o desvio padrão para a energia e entropia de cada classe. Estas figuras ajudam a identificar os componentes que mais discriminam as classes. Observa-se na Figura 5 que, apesar de baixos

valores, os componentes C70, C71 e C72 apresentam informações para a discriminação dos sinais.

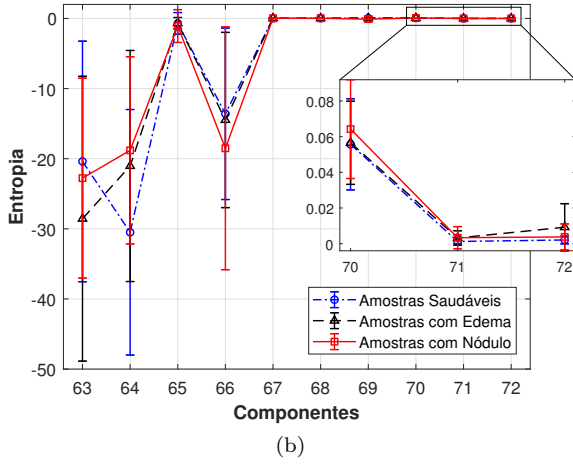
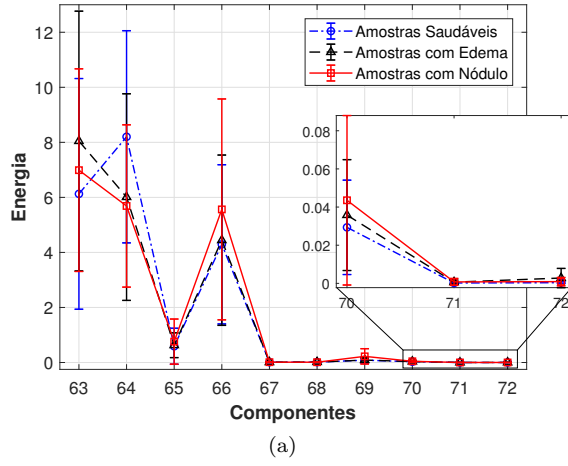


Figura 5. Comparação visual da média e desvio padrão de cada classe. (a) grandeza energia. (b) grandeza entropia.

Nota-se, ainda, na Figura 5(a) e 5(b) que na média as classes apresentam boa separação, entretanto, observa-se um grande desvio padrão dos componentes. A Figura 5(a) mostra os componentes relevantes para o estudo, sendo os primeiros componentes deste nível a concentrar maior parte da energia. A partir do componente C66, tem-se uma queda na energia dos componentes. Este fato mostra que o sinal em análise tem maior energia e, conseqüentemente, informação nas baixas frequências.

O componente C63 corresponde a banda de frequência de 0 a 390.63 Hz, sendo essa a banda de frequência, onde se encontra a frequência fundamental da voz de homens e mulheres. Como a frequência fundamental é gerada nas pregas vocais e, as patologias as afetam, observa-se uma boa separação entre os sinais patológicos e saudáveis neste componente. O componente C64 corresponde a banda de frequência de 390.63 a 781.25 Hz. A diferença entre os sinais, pode ser entendida, como a alteração que as patologias causam nos harmônicos, assim, o sinal saudável tem maior riqueza em harmônicos. Essa interpretação explica a maior energia e menor entropia neste componente. As outras diferenças nos componentes são, possivelmente, devido ao ruído introduzido nos sinais patológicos.

A Tabela 2 apresenta os resultados do primeiro experimento, sendo a classificação de saudáveis e patológicos.

Tabela 2. Acertos de classificação dos k -folds para energia, entropia e ambas características.

Iteração	Energia	Entropia	Energia e Entropia
	Acerto%	Acerto%	Acerto%
k=1	95.06	91.36	95.06
k=2	97.53	95.06	91.36
k=3	95.06	92.59	96.30
k=4	92.59	95.06	98.77
k=5	91.36	95.06	90.12

Observa-se que o método proposto apresenta bons resultados em todos os testes, sendo 97.53%, 95.06% e 98.77% as maiores taxas de acerto para energia, entropia e energia com entropia, respectivamente. Em média a taxa de acerto do método é de 94.32% para a grandeza energia, 93.83% para a grandeza entropia e 94.32% para as grandezas juntas. Apesar da característica energia com entropia apresentar na média mesma taxa de acerto que a grandeza energia, esta tem maior variação nos teste efetuados.

O trabalho de Arjmandi and Pooyan (2012), conseguiu maiores taxas de acerto na classificação de sinais saudáveis de patológicos, chegando ao caso ótimo de 100% de acerto. Entretanto esse trabalho utiliza outro banco de dados para a análise dos sinais e também, mais patologias da laringe, tais como: cisto e paralisia das pregas vocais. O trabalho Muhammad et al. (2017) utiliza três banco de dados para a classificação, um deles é o mesmo banco de dados do presente trabalho, entretanto, as patologias em análise são diferentes, sendo utilizado cisto, paralisia e polipo nas pregas vocais. Portanto é necessário cautela na comparação direta com outros trabalhos, visto que, apresentam diferenças nos materiais utilizados.

A matriz de confusão mostra os acertos e erros do algoritmo classificador para cada classe. As linhas da matriz de confusão mostram os resultados preditos, as colunas correspondem a classe alvo. A Figura 6 mostra a matriz de confusão dos piores resultados da Tabela 2.

O total de janelas utilizadas para as classes são 54 para patológica e 27 para saudável. Para entender as porcentagens da matriz de confusão, observa-se a Figura 6(a). Nota-se que a soma da primeira coluna é igual a classe patológica, e diz o acerto e erro para esta classe. A primeira linha são os resultados preditos, então a classe patológica predita apresenta 61 janelas, sendo dessas apenas 54 corretos o que dá uma porcentagem de acerto na predição desta classe de 88.5%.

A matriz de confusão das grandezas apresentaram uma baixa taxa de falso negativo, ou seja, sinais patológicos classificados como saudáveis. Portanto, nos casos onde o método tem menor taxa de acerto, o maior erro acontece em falso positivo, sendo menos prejudicial que o falso negativa nesta análise.

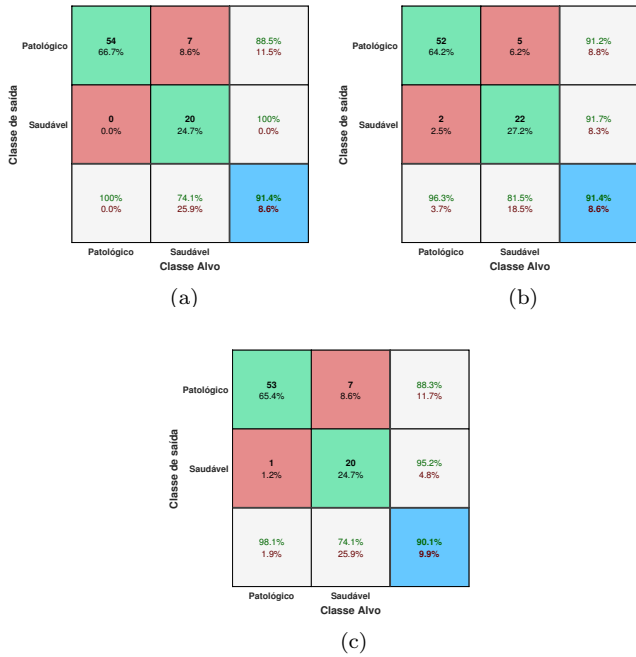


Figura 6. Matriz confusão dos piores testes. (a) grandeza energia. (b) grandeza entropia. (c) grandeza entropia e energia.

A Tabela 3 apresenta os resultados do segundo experimento, sendo a classificação das patologias edemas e nódulos.

Tabela 3. Acertos de classificação dos *k-folds* para energia, entropia e ambas características.

Iteração	Energia	Entropia	Energia e Entropia
	Acerto%	Acerto%	Acerto%
k=1	98.15	96.30	98.15
k=2	92.60	92.60	96.30
k=3	100.0	96.30	96.30
k=4	92.60	94.44	94.44
k=5	98.15	100.0	98.15

Nota-se que em todos os testes, e para todas as grandezas o método proposto apresentou bons resultados. Em média a taxa de acerto do método é de 96.30% para a grandeza energia, 95.93% para a grandeza entropia e 96.67% para as grandezas juntas.

Com os resultados da Tabela 3 é possível notar que, apesar das duas patologias afetarem as pregas vocais, a deformação no sinal é diferente, o que permite discriminá-los.

A matriz de confusão mostra os acertos e erros do algoritmo classificador por classe, conforme a Figura 7.

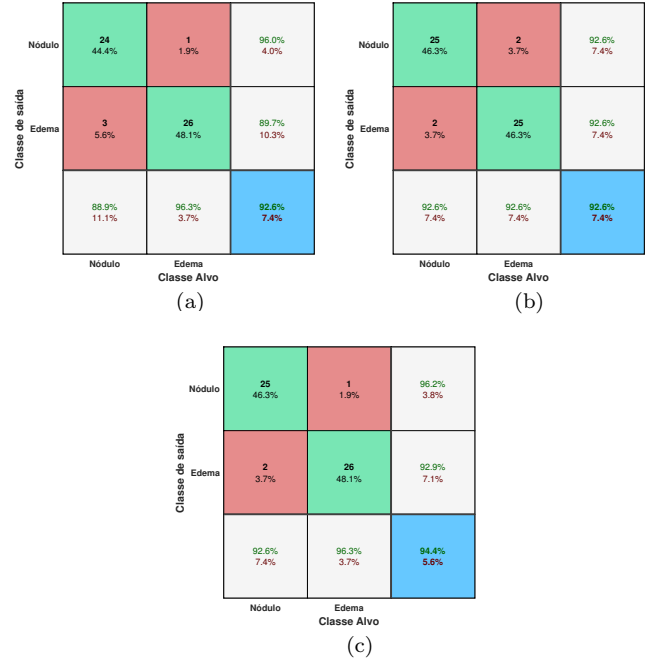


Figura 7. Matriz confusão dos piores testes. (a) grandeza energia. (b) grandeza entropia. (c) grandeza entropia e energia.

Nos piores casos, as matrizes de confusão da Figura 7 são parecidas, com maior erro na classificação de nódulo.

5. CONCLUSÃO

No presente trabalho foi desenvolvido um método para classificar sinais de voz. A abordagem foi extrair informações dos coeficientes da WPD, através do cálculo de energia e entropia, obtendo o vetor de características, sendo este entrada do classificador SVM.

O método obteve sucesso na classificação de sinais patológicos e saudáveis com um acerto médio para o primeiro experimento de 94.32% para a grandeza energia, 93.83% para a grandeza de entropia e 94.32% para as grandezas juntas. Para a classificação apenas das patologias, o método obteve um acerto médio de 96.30% para a grandeza energia, 95.93% para a grandeza entropia e 96.67% para as grandezas juntas.

O estudo mostrou que no primeiro experimento o método apresentou maiores taxas de erro de classificação em falsos positivos, sendo nesta análise menos nocivo que o falsos negativos. Para o segundo experimento o estudo mostrou que as patologias afetam de forma diferente os sinais da vogal “a” sustentada, o que permite classificá-los com boa taxa de acerto.

Este trabalho é uma contribuição para a área de análise de voz, apresentando um método de classificação aplicado a um banco de dados específico. Embora existam estudos com estas grandezas como características para a classificação, é necessário estudos que comprovem a eficácia e validade em outros bancos de dados e outros métodos, que utilizam as mesmas grandezas, mas que as calculam de

forma diferente. Em trabalhos futuros, pretende-se testar o método em outras patologias, buscar a construção de uma família *wavelet* dedicada a sinais de voz e testar outro tipo de classificador.

AGRADECIMENTOS

Os autores gostariam de agradecer a Universidade Tecnológica Federal do Paraná (UTFPR), a Fundação Araucária pelo suporte ao projeto e a Saarbruecken Voice Database por disponibilizar os sinais.

REFERÊNCIAS

- Arjmandi, M.K. and Pooyan, M. (2012). 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*, 7(1), 3–19.
- Barry, W. and Pützer, M. (2007). Saarbrücken voice database. *Institute of Phonetics, Universität des Saarlandes*, <http://www.stimmdatenbank.coli.uni-saarland.de>.
- Behlau, M. (2005). *Voz: o livro do especialista*. Revinter, Rio de Janeiro.
- Bishop, C.M. (2006). *Pattern recognition and machine learning*. springer, Berlin, Heidelberg.
- Burges, C.J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2), 121–167.
- Cielo, C.A., Finger, L.S., de Caldas Rosa, J., and Brancalioni, A.R. (2011). Lesões organofuncionais do tipo nódulos, pólipos e edema de reinke. *Revista CEFAC*, 13(4), 735–748.
- Dajer, M.E. (2010). *Análise de sinais de voz por padrões visuais de dinâmica vocal*. Ph.D. thesis, Universidade de São Paulo.
- Diniz, P.S., da Silva, E.A., and Netto, S.L. (2014). *Processamento Digital de Sinais: Projeto e Análise de Sistemas*. Bookman Editora.
- Gómez-García, J., Moro-Velázquez, L., and Godino-Llorente, J.I. (2019). On the design of automatic voice condition analysis systems. part i: Review of concepts and an insight to the state of the art. *Biomedical Signal Processing and Control*, 51, 181–199.
- Guido, R.C., Pereira, J.C., Fonseca, E., Sanchez, F.L., and Vieira, L.S. (2005). Trying different wavelets on the search for voice disorders sorting. In *Proceedings of the Thirty-Seventh Southeastern Symposium on System Theory, 2005. SSST'05.*, 495–499. IEEE, Tuskegee, USA.
- Mallat, S. (2008). *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, Inc., Orlando, FL, USA, 3 edition.
- Muhammad, G., Alsulaiman, M., Ali, Z., Mesallam, T.A., Farahat, M., Malki, K.H., Al-nasheri, A., and Bencherif, M.A. (2017). Voice pathology detection using interlaced derivative pattern on glottal source excitation. *Biomedical signal processing and control*, 31, 156–164.
- Nemr, K., Amar, A., Abrahão, M., Leite, G.C.d.A., Köhle, J., Santos, A.d.O., and Correa, L.A.C. (2005). Comparative analysis of perceptual evaluation, acoustic analysis and indirect laryngoscopy for vocal assessment of a population with vocal complaint. *Revista Brasileira de Otorrinolaringologia*, 71(1), 13–17.
- Oates, J. (2009). Auditory-perceptual evaluation of disordered voice quality. *Folia Phoniatrica et Logopaedica*, 61(1), 49.
- Omori, K. (2011). Diagnosis of voice disorders. *Japan Medical Association Journal*, 54(4), 248–253.
- Ortega, J., Cassinello, N., Dorcaratto, D., and Leopaldi, E. (2009). Computerized acoustic voice analysis and subjective scaled evaluation of the voice can avoid the need for laryngoscopy after thyroid surgery. *Surgery*, 145(3), 265–271.
- Ramig, L.O. and Verdolini, K. (1998). Treatment efficacy: voice disorders. *Journal of Speech, Language, and Hearing Research*, 41(1), S101–S116.
- Scalassara, P.R., Maciel, C.D., and Pereira, J.C. (2009a). Predictability analysis of voice signals. *IEEE Engineering in Medicine and Biology Magazine*, 28(5), 30–34.
- Scalassara, P.R., Dajer, M.E., Maciel, C.D., Guido, R.C., and Pereira, J.C. (2009b). Relative entropy measures applied to healthy and pathological voice characterization. *Applied Mathematics and Computation*, 207(1), 95–108.
- Scalassara, P.R., Maciel, C.D., Guido, R.C., Pereira, J.C., Fonseca, E.S., Montagnoli, A.N., Júnior, S.B., Vieira, L.S., and Sanchez, F.L. (2007). Autoregressive decomposition and pole tracking applied to vocal fold nodule signals. *Pattern recognition letters*, 28(11), 1360–1367.
- Teixeira, J.P., Oliveira, C., and Lopes, C. (2013). Vocal acoustic analysis-jitter, shimmer and hnr parameters.
- Verde, L., De Pietro, G., and Sannino, G. (2018). A methodology for voice classification based on the personalized fundamental frequency estimation. *Biomedical Signal Processing and Control*, 42, 134–144.
- Wallis, L., Jackson-Menaldi, C., Holland, W., and Giraldo, A. (2004). Vocal fold nodule vs. vocal fold polyp: answer from surgical pathologist and voice pathologist point of view. *Journal of Voice*, 18(1), 125–129.