Detecção de patologias laríngeas com base na análise dinâmica de sinais de voz utilizando redes neurais profundas

Lucas C. Dias, Luana R. Barros, Suzete E. N. Correia e Silvana L. do N. C. Costa

***Resumo*— Este trabalho trata da aplicação de classificadores baseados em redes neurais profundas (RNPs) na discriminação entre sinais de vozes saudáveis e afetados pelas patologias laríngeas edema, carcinoma, leocoplasia, pólipos e paralisia das pregas vocais. Cada sinal de voz foi particionado em segmentos, sendo estes inseridos em uma RNP de 05 camadas ocultas com 200 neurônios cada e um neurônio na camada de saída. Para a classificação empregou-se uma metodologia que analisa o comportamento dinâmico dos segmentos usados para o teste. O método proposto forneceu uma acurácia de 85,55±4,39%.**

***Palavras-Chave*— Aprendizagem profunda, processamento digital de sinais de voz, patologias laríngeas, redes neurais artificiais.**

***Abstract*— This paper deals with the application of classifiers based on deep neural networks (DNNs) in the discrimination between healthy voice signals and affected by laryngeal pathologies edema, carcinoma, leukoplakia, polyps and vocal fold paralysis. Each voice signal was partitioned into segments, which are inserted in an DNN of 05 hidden layers each of them with 200 neurons and one neuron in the output layer. For the classification a methodology was used that analyzes the dynamic behavior of the segments used for the test. The proposed method provided an accuracy of 85,55±4,39%.**

***Keywords*— Artificial neural networks, deep learning, digital processing of voice signals, pathologies laryngeal.**

I. INTRODUÇÃO

O sinal de voz é uma onda sonora, emitida pelo processo natural de fonação humana, que após ser digitalizada passa a ser representada por um conjunto de amostras em um intervalo finito de tempo, tornando-se uma série temporal discreta que traz consigo informações sobre o sistema de produção vocal [1].

Quando afetadas por patologias laríngeas, as pregas vocais passam a ter seu processo natural de vibração modificado, resultando em alterações no sinal de voz emitido pelo indivíduo [2]. Diante do sinal digitalizado, a análise acústica comumente realiza a extração de atributos (características únicas) do sinal de voz inerentes ao sistema de produção vocal, com o objetivo de obter valores numéricos capazes de quantificar as alterações vocais [1, 3].

Lucas C. Dias, e-mail: lucas.cd@live.com; Luana R. Barros, e-mail: luana.barros@assert.ifpb.edu.br; Suzete E. N. Correia, e-mail: suzete.correia@gmail.com; Silvana L. do N. C. Costa, email: silvanacunhacosta@gmail.com. Programa de Pós-Graduação em Engenharia Elétrica, Instituto Federal da Paraíba, João Pessoa-PB. Este trabalho foi parcialmente financiado pela Pró-Reitoria de Pesquisa, Inovação e Pós-Graduação do IFPB/JP e Capes.

Tais atributos podem ser representados por medidas acústicas tais como: frequência fundamental (F_0), *jitter*, *shimmer*, relação harmônica-ruído (*Harmonic-Noise Ratio* - HNR), excitação do ruído glotal (*Glottal Noise Excitation* - GNE) [4]; medidas com base na análise no domínio da frequência, como os coeficientes cepstrais [5]; medidas com base no modelo linear de produção da fala, como os coeficientes de predição linear [1]; medidas com base na análise dinâmica não linear (passo de reconstrução, primeiro mínimo da informação mútua, dimensão de correlação) e de quantificação de recorrência (determinismo, comprimento médio das linhas diagonais, comprimento médio das linhas verticais e transitividade), entre outras [6, 7].

Por possibilitar uma avaliação clínica não invasiva e com possibilidade de ser realizada a distância, a análise acústica apresenta-se como uma alternativa para a redução da quantidade de exames invasivos aplicados para a inspeção visual da laringe, como por exemplo a laringoscopia a videolaringoscopia. Apesar de apresentar vantagens relevantes para a saúde pública, a análise acústica tradicional possui limitações que podem comprometer a sua expansão no meio clínico [4, 8].

A diversidade de metodologias aplicadas na extração de atributos presentes no universo da análise acústica e o fato de um mesmo atributo possuir variações diferentes dentro de uma classe, devido as constituições morfológicas, idade e sexo do indivíduo, torna-se complexa a tarefa de escolha das medidas acústicas que melhor representem os sinais de voz sob análise. Portanto, a seleção ótima de atributos que favoreça o desempenho da aprendizagem e generalização de tais algoritmos não se caracteriza como uma etapa trivial, além de também compartilhar das limitações que uma análise acústica via atributos possui, diante de casos fonoterápicos mais complexos [9, 10].

A aprendizagem de máquina profunda, por meio de implementações de redes neurais profundas (RNPs) para sinais unidimensionais e redes neurais convolucionais (RNC), para sinais bidimensionais, têm sido utilizadas no desenvolvimento de sistemas inteligentes, aptos para realizar o reconhecimento de padrões característicos presentes em sinais complexos [11-12].

O seu potencial despertou o interesse de pesquisadores que passaram a estudar o uso de RNPs na análise acústica de sinais voz. Em sua pesquisa, Wu et al. [13] investigaram a discriminação entre vozes saudáveis e afetadas por patologias laríngeas, utilizando uma RNC composta por 3 camadas convolucionais e 8 profundas. Utilizou a base de dados alemã

Saarbruecken Voice Database (SVD), da qual foram selecionados 482 sinais (vogal sustentada $/a/$) de vozes saudáveis e 482 vozes afetadas pelas patologias edema, pólipos, leucoplasia, carcinoma, e paralisia das pregas vocais. Para cada sinal de voz foi aplicada a transformada rápida de Fourier e extraídos seus respectivos espectrogramas. A imagem de cada espectrograma em foi inserida na RNC durante as etapas de treino, teste e validação, sendo utilizados em cada etapa, respectivamente, 70%, 15% e 15% do conjunto total de dados. Como resultados foram obtidos 71% de acurácia para o conjunto de teste e 68% na etapa de validação.

O trabalho de Harár et al. [14] apresentaram a discriminação entre sinais de vozes saudáveis e patológicos utilizando RNCs. Foram empregadas 687 sinais de vozes saudáveis e 1353 distribuídos entre os 71 tipos de desordens vocais presentes na base de dados SVD, dentre as quais estão incluídas as patologias investigadas por Wu et al. [13]. A metodologia consistiu em particionar cada sinal de voz em segmentos de 3200 amostras e montar uma matriz de dimensões 3200 $\times$ números de segmentos. Tal matriz é inserida em uma RNC com 4 camadas convolucionais e 3 profundas. Foi utilizado 70% do conjunto de dados para treino, 15% para validação e 15% para teste. Obteve-se como resultados 68% de acurácia no teste e 71,36% na validação.

A análise acústica de sinais de voz por meio da utilização de RNPS surgiu para potencializar este tipo de avaliação. No entanto, diversos fatores precisam ser aperfeiçoados para proporcionar sistemas que atendam a necessidade clínica fonoterápica, como a discriminação entre vozes saudáveis e afetadas por patologias laríngeas.

Este trabalho apresenta uma metodologia para realizar a detecção de patologias laríngeas, com base no comportamento dinâmico da série temporal discreta de sinais de voz, utilizando uma RNP classificadora. O método foi avaliado com base nas métricas acurácia, sensibilidade e especificidade obtidas mediante matriz de confusão, após validação cruzada do tipo $k - fold$.

II. Redes Neurais Profundas - RNPs

As RNPs possuem múltiplos processadores paralelamente distribuídos (camadas de entrada, ocultas e de saída). Tais camadas são constituídas de unidades de processamento simples (neurônios) com o objetivo de criar um modelo matemático complexo, capaz de realizar previsões de respostas (saídas) com base na apresentação de informações (entradas) que podem ter ou não uma relação comum [15]. A relação entre o conjunto de valores entrada x e a saída da primeira camada h_1 é descrita pela Equação (1):

$$h_1 = f(w_1 x) + b_1 \tag{1}$$

em que w_1 e b_1 são, respectivamente, os parâmetros ponderadores (pesos sinápticos) de cada neurônio da camada e o vetor de polarização $bias$. $f(\cdot)$ representa a função de ativação dos neurônios, sendo utilizada nesta pesquisa a função ReLu. A relação matemática entre a saída das camadas antecessoras e conseguintes é expressa pela Equação (2).

$$h_{i+1} = f(w_i h_i + b_{i+1}), [i = 1, 2, ..., L-1], \tag{2}$$

em que L representa o número total de camadas que constitui a RNP. As respostas referentes a predição $\hat{y}$ da camada de saída (h_L) são obtidos por meio da Equação 3. $g(\cdot)$ representa a função de ativação utilizado nos neurônios da camada de saída, sendo utilizada nesta pesquisa a função sigmóide.

$$\hat{y} = g(h_L) \tag{3}$$

Durante o processo de treinamento de uma RNP, os pesos sinápticos, θ, são calculados com base na minimização de uma função custo, O, por meio da aplicação de algoritmos de otimização. Por se tratar de um processo de classificação a função custo utilizada, nesta pesquisa, foi a *binary cross-entropy*, conforme definida na Equação (4),

$$O_t = -\frac{1}{N}\sum_{n=1}^{N} y_n log(\hat{y}_n) + (1 - y_n) log(1 - \hat{y}_n) \tag{4}$$

em que y_n é a resposta real para o vetor de entrada x_n; $\hat{y}_n$ da predição da RNP com base no vetor de entrada x_n; N o total de vetores de entrada, que serão inseridos durante a época de treinamento t. Neste trabalho, o algoritmo de otimização Adamax foi empregado para efetuar o cálculo dos valores de θ.

III. Materiais e métodos

Nesta Seção, são descritos os materiais utilizados para o desenvolvimento desta pesquisa, bem como a metodologia empregada nos experimentos computacionais.

A. Base de dados

A base de dados aplicada neste trabalho trata-se da *Saarbruecken Voice Database* (SVD), desenvolvida na Alemanha e distribuída por meio de um repositório digital aberto, na qual possui sinais de vozes e de eletroglotografia (EGG) capturados em laboratório [16].

Da base, foram selecionados 640 sinais de voz, sendo 320 sinais de vozes saudáveis (180 femininas e 140 masculinas), 63 vozes afetadas por edema de Reinke (56 femininas e 7 masculinas), 21 vozes afetadas por carcinoma (1 feminina e 20 masculinas), 41 vozes afetadas por leocoplasia (17 femininas e 27 masculinas), 45 vozes afetadas por pólipos vocais (19 femininas e 26 masculinas) e 150 vozes afetadas por paralisia das pregas vocais (90 femininas e 60 masculinas). Todos os sinais patológicos selecionados são das patologias laríngeas que apresentam uma quantidade considerável de gravações, sendo essas da vogal sustentada $/a/$. Os sinais possuem, originalmente, uma taxa de amostragem de de 50000 amostras por segundo, mas foram re-amostrados, nesta pesquisa, para reduzir a dimensão dos segmentos que serão analisados e a complexidade computacional do classificador. Todas as gravações possuem uma resolução de 16 bits/amostra com tempo de duração de 1 a 3 segundos gravados em formato .wav.

B. Metodologia

O processo discriminativo entre vozes saudáveis e patológicas tem como base a detecção de padrões no comportamento dinâmico do sinal de voz utilizando uma RNP e, por meio desta análise, classificar se tal dinâmica é representativa de um sinal de voz saudável ou patológica. Os aspectos metodológicos referentes a esta aplicação podem ser divididos em 3 etapas: pré-processamento, treinamento e validação.

1) Pré-processamento: Inicialmente, todos os sinais de voz selecionados da base de dados passaram por um filtro linear de suavização do tipo média-móvel [17]. O filtro foi aplicado para reduzir possíveis ruídos presentes nos sinais da base. Este filtro opera calculando a média de vários pontos do sinal de entrada para produzir cada ponto no sinal de saída, conforme definido na Equação (5).

$$s[n] = \frac{1}{M} \sum_{j=0}^{M-1} x[n-j], \tag{5}$$

sendo $s[n]$ o sinal de voz filtrado e $x[n-j]$ as amostras do sinal original, levadas em consideração para calcular $s[n]$. A quantidade dessas amostras é definida de acordo com a ordem do filtro M. Neste trabalho, foi estabelecido o valor de M igual a 5.

Além da filtragem, foi realizado um escalonamento entre as amostras que compõe a série temporal discreta de cada sinal de voz. O escalonamento é um requisito comum para muitos estimadores de aprendizado profundo. De acordo com Raschka [18] e Gron [19], algoritmos classificadores e preditores podem ter seu desempenho comprometido ao processar dados com altos níveis de dispersão ou sob a presença de *outliers*.

Para efeitos de escalonamento dos dados fez-se o uso da distribuição normal padrão. Desse modo, cada amostra da série temporal, do seu respectivo sinal de voz, será inserida na Equação (6), a fim de calcular os valores referentes ao escore-z, sendo esta uma medida que indica o quanto uma medida se afasta da média em termos de desvios padrão [20].

$$z[n] = \frac{s[n] - \mu}{\sigma}, \tag{6}$$

sendo μ, Equação (7), a média dos valores das amostras $s[n]$ e σ o desvio padrão das amostras, Equação (8).

$$\mu = \frac{1}{N} \sum_{n=1}^{N} s[n], \tag{7}$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (s[n] - \mu)^2}, \tag{8}$$

em que N é o total de amostras.

2) Treinamento: Foi empregada a validação cruzada do tipo $k - fold$, no qual o conjunto total de sinais de vozes selecionados da base são, aleatoriamente, divididos em k subconjuntos [21]. Desse modo, $k-1$ conjuntos são empregados para o treinamento do classificador e o subconjunto restante é aplicado na etapa de validação. Ao fim, são efetuados k treinos, de tal modo que todos os subconjuntos sejam utilizados nas fases de validação do classificador.

Durante o treinamento, cada sinal de voz pré-processado ($z[n]$) é particionado em 40 segmentos de 16ms (quadros com 400 amostras) sob janelamento retangular e sem taxas de sobreposição. O valor de tempo de duração dos segmentos foram estabelecidos levando-se em consideração o intervalo em que o sinal de voz é considerado estacionário (16ms - 32ms) [4].

Cada segmento do sinal pré-processado foi utilizado como o vetor de dados de entrada da RNP. Por sua vez, a RNP utilizada nesta pesquisa foi estruturada com 05 camadas ocultas, compostas por 200 neurônios cada, ativados pela função Relu e 01 neurônio ativado pela função sigmóide na camada de saída. A quantidade de neurônios em cada camada oculta foi definida com base na metodologia de cálculo utilizada em Vieira et al. [1]. O algoritmo de otimização Adamax foi empregado para efetuar a computação dos pesos sinápticos dos neurônios da RNP por meio da minimização da função custo $binary\ cross - entropy$, apresentada na Equação (4). Para fins de implementação, a parametrização do algoritmo Adamax foi estabelecida de acordo com a metodologia proposta em Ruder [22]. Além disso, os valores iniciais dos pesos sinápticos foram definidos aleatoriamente.

Por se tratar de um processo de aprendizagem supervisionado, a classe verdadeira (saudável ou patológica) do sinal de voz, no qual o segmento analisado pertence, é apresentada durante o treinamento. Desta forma, durante o treinamento, a RNP tem como objetivo principal, por meio de classificação de padrões, conseguir mapear cada segmento a uma determinada classe, sendo esta, por sua vez, a do sinal de voz apresentado a rede.

C. Validação

Assim como no treinamento, na validação o sinal de voz pré-processado foi particionado em 40 segmentos de 16ms. Individualmente, cada um destes segmentos foram colocados na entrada da RNP treinada a fim de serem classificados como pertencente a uma das classes (saudável ou patológica). Após a apresentação dos 40 segmentos do sinal de voz ao classificador, identificou-se a computação da classe predominante dos segmentos, sendo esta a classe à qual o sinal de voz analisado pertence, conforme ilustrado na Figura 1. Este procedimento foi realizado em todos os sinais de voz presentes em cada conjunto de validação.

Ao fim desta etapa, o processo foi avaliado pelas métricas: 1) acurácia: taxa do diagnóstico final correto para todos os sinais de voz, dada pela Equação (9); 2) sensibilidade: taxa do diagnóstico final correto para os sinais afetados por patologias, dada pela Equação (10); e especificidade: taxa do diagnóstico final correto para os sinais saudáveis, dada pela Equação (11). Tais métricas foram calculadas mediante matriz de confusão (Tabela I), sendo esta matriz uma ferramenta importante que determina se o diagnóstico efetuado corresponde ou não ao diagnóstico real, de modo que todos os casos são aferidos e os mesmos são exibidos na matriz [19].

Por se tratar de uma validação cruzada do tipo $k - fold$, as métricas foram calculadas para mensurar os resultados obtidos em cada conjunto utilizado na etapa de validação. No final, a

média e o desvio padrão das métricas foram calculadas para representar o desempenho geral do processo.

Todo o procedimento descrito foi implementado em linguagem de programação $Python$. A RNP foi implementada por meio da biblioteca de computação científica $TensorFlow$ e programada utilizando a biblioteca $Keras$.

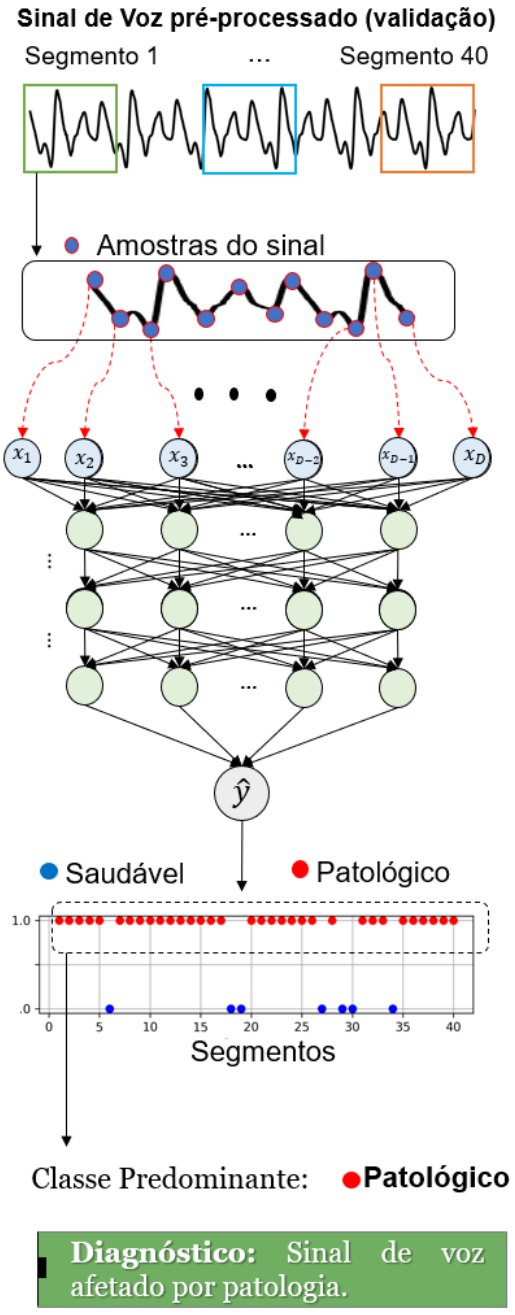

Fig. 1. Diagnóstico com base na análise dinâmica de segmentos de sinais de voz.

$$Acuracia = \frac{VP + VN}{VP + VN + FP + FN} \quad (9)$$

$$Sensibilidade = \frac{VP}{VP + FN} \quad (10)$$

$$Especificidade = \frac{VN}{VN + FP} \quad (11)$$

TABELA I

MATRIZ DE CONFUSÃO EM UM TESTE DE DIAGNÓSTICO DA PRESENÇA OU AUSÊNCIA DA PATOLOGIA.

Resultado	Patologia	
	Presente	Ausente
Positivo	Verdadeiro Positivo (VP)	Falso Positivo (FP)
Negativo	Falso Negativo (FN)	Verdadeiro Negativo (VN)

IV. RESULTADOS E DISCUSSÃO

Nesta Seção são apresentados os resultados obtidos no processo de classificação de vozes saudáveis e patológicas, realizado a fim de explorar o potencial uso do classificador baseado em RNP no correto diagnóstico de patologias laríngeas. Para o classificador foram aplicadas 40 épocas de treinamento, valor este definido empiricamente. A validação cruzada foi implementada utilizando o método $k - fold$ para k = 10.

Na Tabela II são apresentados os resultados mensurados pelas métricas acurácia, sensibilidade e especificidade obtidas da classificação dos sinais de vozes saudáveis e dos patológicos.

Com base na Tabela II, tem-se que o método proposto obtém desempenho em torno de 80% para todas as métricas de avaliação. Apresenta melhor desempenho para métrica acurácia, alcançando a taxa de 85,55±4,39%. Os resultados mensurados para sensibilidade e especificidade são equiparáveis. Esta relação pode ser melhor observada analisando a Tabela III, na qual apresenta a matriz de confusão final obtida após a computação de todos os casos referentes a VP, VN, FP e FN de cada conjunto utilizado na validação.

Em comparação com a pesquisa desenvolvida por Wu et al. [13], que também emprega as 05 patologias laríngeas utilizadas neste trabalho, tem-se que o desempenho do processo discriminativo proposto foi superior na etapa de validação, como pode ser visto na Figura 2. Além disso, esta mesma pesquisa baseia-se no uso de RNC aplicada na análise de sinais bidimensionais originados de espectrogramas.

TABELA II

DESEMPENHOS DO CLASSIFICADOR COM BASE NAS MÉTRICAS ACURÁCIA, SENSIBILIDADE E ESPECIFICIDADE.

Caso/Métricas	**Acurácia**	**Sensibilidade**	**Especificidade**
SDVxPTL	85,55±4,39%	84,39±7,76%	86,87±9,36%

TABELA III

MATRIZ DE CONFUSÃO FINAL APÓS A VALIDAÇÃO CRUZADA

Resultado	Patologia	
	Presente	Ausente
Positivo	277	43
Negativo	45	275

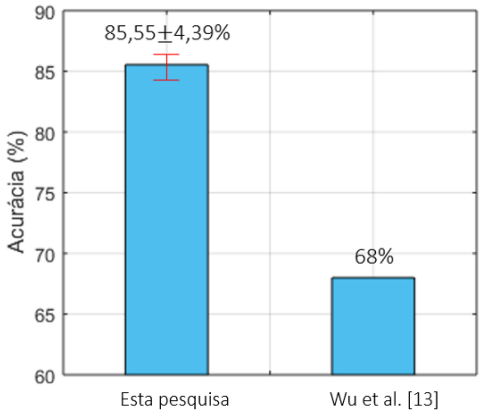

Fig. 2. Comparação entre valores de acurácia na etapa de validação desta pesquisa e Wu et al. [13].

V. Conclusões

Os resultados obtidos, nesta pesquisa, indicam que o método proposto para discriminação entre sinais de vozes saudáveis e afetados por patologias laríngeas, utilizando RNP, tem possibilidade de contribuir com a aplicação da análise acústica por profissionais da área de saúde e podendo de fato auxiliar na redução da quantidade de exames médicos referentes a inspeção visual da laringe. Além disso, o método proposto apresenta diferenças significativas com grande parte dos trabalhos apresentados na literatura, devido a possibilidade de discriminar sinais de voz sem a necessidade de aplicar técnicas como análise dinâmica linear ou não linear, em frequência, ou de extração de medidas acústicas. Dentre os pontos a serem investigados em trabalhos futuros, tem-se a exploração do método na discriminação entre diferentes tipos de patologias laríngeas; discriminação entre vozes saudáveis e afetadas por desvios vocais; e discriminação entre tipos de desvios vocais.

Agradecimentos

Agradecemos ao Instituto Federal de Educação, Ciência e Tecnologia - *Campus* João Pessoa pela disponibilização de laboratórios e equipamentos para o desenvolvimento desta pesquisa; a Pró-Reitoria de Pesquisa, Inovação e Pós-Graduação do IFPB/JP e a Capes pelo apoio financeiro.

Referências

[1] V. J. D Vieira, S. L. do N. C. Costa, W. C. de A. Costa, Suzete E. N. Correia and Joseana M. F. R de Araújo, "Avaliação de desempenho na classificação de patologias laríngeas por análise LPC de sinais de voz e redes neurais MLP,"in *XI Congresso Brasileiro de Inteligência Computacional (CBIC 2013)*, Porto de Galinhas - PE, Setembro. 2013, pp. 1-6.

[2] C. A. Cielo, V. V. Ribeiro, G. R. Bastilha and N. de O. Schilling, "Qualidade de vida em voz, avaliação perceptivoauditiva e análise acústica da voz de professoras com queixas vocais,"*Audiol., Commun. Res.*, São Paulo, v. 20, n. 2, p. 130-140, Jun. 2015.

[3] J. Moon and J, Lee, "Development of medical/electrical convergence software for classification between normal and pathological voices,"*Journal of Digital Convergence*, v. 13, p. 187–192, Dez. 2015.

[4] S. L. d. N. C. Costa, "Análise acústica, baseada no modelo linear de produção da fala, para discriminação de vozes patológicas,"*Universidade Federal de Campina Grande*. Tese de Doutorado, 161 p., 2008.

[5] S. Fang, Y. Tsao, M. Hsiao, J. Chen, Y. Lai, F. Lin and C. Wang, "Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach", *Journal of Voice*, v. 33, n. 5, p.634-641, Set. 2019.

[6] T. D. Prado, G. Z. Lima, B. L. Soares, G. do Nascimento, G. Corso, J. Fontenele-Araujo, J. Kurths, and S. R. Lopes, "Optimizing the detection of nonstationary signals by using recurrence analysis," Chaos, v. 28, n. 8, p.085703, 2018.

[7] L. W Lopes, V. J. D. Vieira, S. L. N. C. Costa, S. E. N, Correia and M. Behlau, "Effectiveness of Recurrence Quantification Measures in Discriminating Subjects With and Without Voice Disorders,"*Journal of Voice*, v. 34, p. 208-220, 2018.

[8] J. Grigorjevs, "Problems using the traditional acoustic cues for the phonological interpretation of vowels,"*Baltistica*, v. 48, n. 2, p. 301-312, 2013.

[9] M, Alhussein, G. Muhammad, "Automatic voice pathology monitoring using parallel deep models for smart healthcare,"in *IEEE Access*, v. 7, p. 46474–46479, 2019.

[10] Z. Chuang, X. Yu, J. Chen, and Y Hsu, "DNN-based Approach to Detect and Classify Pathological Voice", in *IEEE International Conference on Big Data (Big Data)*, Washington - EUA, Dezembro, 2018.

[11] A. Vahadane, A. Joshi, K. Madan and T. R. Dastidar, "Detection of diabetic macular edema in optical coherence tomography scans using patch based deep learning,"in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington - EUA, pp. 1427-1430, 2018.

[12] S. Turan and G. Bilgin, "Semantic nuclei segmentation with deep learning on breast pathology images,"in *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, Istanbul, Turkey, pp. 1-4, Dez. 2019.

[13] H. Wu, J. Soraghan, A. Lowit and G. Di Caterina, "A Deep Learning Method for Pathological Voice Detection using Convolutional Deep Belief Network", in *Interspeech*, Hyderabad - IN, Sep. 2018.

[14] P. Harar, J. B. Alonso-Hernandezy, J. Mekyska, Z. Galaz, R. Burget and Z. Smekal, "Voice Pathology Detection Using Deep Learning: a Preliminary Study,"in *2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI)*, Funchal, pp. 1-4, 2017.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.

[16] W. J. Barry and M. Putzer, "Saarbrucken Voice Database", *Institute of Phonetics*, Univ. of Saarland, http://www.stimmdatenbank.coli.uni-saarland.de/.

[17] P. Luo, M. Zhang, Y. Liu, D. Han and Q. Li, "A moving average filter based method of performance improvement for ultraviolet communication system,"in *2012 8th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)*, Poznan, pp. 1-4, 2012.

[18] S. Raschka, *Python Machine Learning*. Packt Publishing, 2015.

[19] A. Gron. *Hands-On Machine Learning with Scikit-Learn and Tensor-Flow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., 2017.

[20] R. Larson and B. Farber. *Estatística Aplicada*. Pearson Education do Brasil, 2004.

[21] P. Duchesne and B. Remillard. *Statistical modeling and analysis for complex data problems*. Springer Science & Business Media, 2005.

[22] S. Ruder. *An overview of gradient descent optimization algorithms*. ArXiv abs/1609.04747, 2016.