Classificação de Problemas da Voz via Processamento Digital de Sinais e Aprendizado de Máquina

Samuel Santos, Luan Oliveira, Adam Santos

***Resumo*—À medida que surgem novas pesquisas e avanços de técnicas na área de aprendizado de máquina, nota-se o significativo impacto dos algoritmos de classificação no processo de detecção de patologias de maneira automatizada. A construção de modelos algorítmicos otimizados fornece conhecimento eficaz em diagnósticos devido sua precisão e capacidade preditiva entre diferentes distúrbios. Neste trabalho, o foco é referente à classificação de pacientes saudáveis e diagnosticados com disfonia e laringite. A metodologia empregada está baseada na utilização da análise da densidade espectral de potência dos sinais de voz de cada paciente em conjunto com a otimização de algoritmos de aprendizado de máquina para automatizar o processo de detecção. Dentre os modelos empregados, aquele que obteve melhor desempenho de classificação foi o *Support Vector Machine*.**

***Palavras-chave*—Densidade Espectral de Potência, Disfonia, Laringite, Aprendizado de Máquina.**

I. Introdução

OS distúrbios vocais afetam as cordas vocais na região da laringe, resultando em vibrações irregulares que consequentemente prejudicam a voz. A abordagem tradicional de diagnóstico é cara e leva em consideração a subjetividade do profissional [1]. Pesquisas buscam automatizar a detecção de patologias, aproveitando o processamento digital de sinais e o aprendizado de máquina para auxiliar no diagnóstico vocal.

O estudo em Pham *et al.* [1] propõe uma metodologia de classificação de distúrbios vocais, incluindo Neoplasia, Fonotrauma e Paralisia vocal. Para isso, utiliza-se a extração dos *Mel Frequency Cepstral Coefficients* de séries temporais, que são então utilizados para treinar e testar diversos algoritmos de aprendizado de máquina, incluindo *Support Vector Machine* (SVM), *Random Forest*, *K-NN*, *Gradient Boosting* e um *Ensemble* de modelos. Além disso, realiza-se um processo de otimização de hiperparâmetros dos classificadores.

No estudo de Pinho *et al.* [2], uma abordagem automatizada para classificar 17 ritmos cardíacos é apresentada. O processo inclui o mapeamento de sinais de ECG para o domínio da frequência via método de Welch, o balanceamento de classes com ADASYN e a normalização Z-score como pré-processamento. Modelos de aprendizado, como SVM, *Multilayer Perceptron*, *K-NN* e *Random Forest*, são utilizados e avaliados com métricas como precisão, sensibilidade, especificidade e o índice *kappa* de Fleiss, com resultados de 98,86%, 99,93%, 98,85% e 89,68%, respectivamente.

A pesquisa de Verde *et al.* [3] apresenta uma metodologia de classificação de distúrbios vocais, com foco em pacientes com disfonia. As características utilizadas para a classificação incluem frequência fundamental, *Jitter*, *Shimmer*, *Harmonic to Noise Ratio* e MFCC. Os algoritmos de classificação selecionados, disponíveis no software WEKA, incluem SVM, árvores de decisão, classificação bayesiana, etc. Essa abordagem aprimorou a detecção automatizada de distúrbios vocais.

O presente trabalho tem como objetivo apresentar um *pipeline* de detecção de distúrbios vocais com técnicas de processamento digital de sinais e algoritmos de classificação otimizados, distinguindo estados saudáveis e patológicos.

II. Materiais e Métodos

Esta seção apresenta o *pipeline* proposto, objetivando detalhar o processo de desenvolvimento deste trabalho. Logo, cada etapa do *pipeline* será explicada nas posteriores subseções, desde o pré-processamento dos dados até a otimização dos classificadores empregados.

A. Base de dados e Pré-processamento

O conjunto de dados desta pesquisa foi fornecido pelo ambulatório de Foniatria e Videolaringoscopia do Hospital Universitário de Nápoles Frederico II, sendo disponibilizado no repositório da plataforma Kaggle [4]. A base de dados é composta por 208 registros de vozes amostrados em 8000 Hz no domínio do tempo, contendo 150 registros de vozes patológicas com disfonia (113) ou laringite (37), e 58 registros de vozes saudáveis.

Foi estimada a densidade espectral de potência (*Power Spectral Density*–PSD) para cada sinal temporal via método de Welch [5] e janela de Hann com tamanho de 4096 amostras, conforme apresentada na Fig. 1. O quantitativo amostral do janelamento contribuiu para uma melhor discriminação dos sinais em termos da alta resolução das componentes de frequência. O tratamento buscou melhorar o desempenho dos modelos de aprendizado de máquina e reduzir a dimensionalidade dos dados de 38720 *features* para 2049.

O balanceamento de classes foi efetuado pela técnica de pré-processamento *Syntetic Over-sampling Technique* (SMOTE) [6], gerando novos registros para as classes minoritárias. Após o pré-processamento, as classes laringite e saudável passaram a ter o mesmo quantitativo de registros da classe disfonia.

S. Santos é graduando na FEC, Unifesspa, Marabá-PA, Brasil (e-mail: samuel.patrick@unifesspa.edu.br).

L. Oliveira é mestrando no PPGCC, UFG, Goiânia-GO, Brasil (e-mail: luan.silva@discente.ufg.br).

A. Santos é professor na FACSI, Unifesspa, Marabá-PA, Brasil (e-mail: adamdreyton@unifesspa.edu.br).

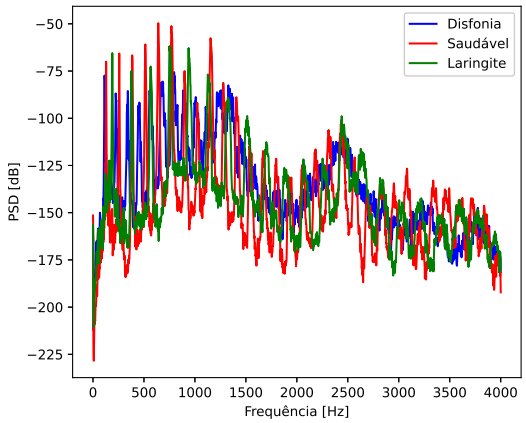

Fig. 1. Exemplos de PSD para as classes saudável, disfonia e laringite.

B. Otimização/Validação e Normalização

A etapa em questão consiste na implementação da técnica de *Cross-validation* [7] para estimar a performance dos modelos para diferentes subconjuntos de dados, sendo considerados cinco *folds* no conjunto de treino. Além disso, foi empregada a técnica *Randomized Search* [8] para seleção do melhor conjunto de hiperparâmetros dos modelos aplicados.

Os modelos selecionados para a classificação dos sinais foram *Random Forest* (RF), *Support Vector Classification* (SVC) e *Multi-layer Perceptron Classifier* (MLP), disponibilizados pela biblioteca *Scikit-learn* [9].

Os hiperparâmetros que possibilitaram um melhor desempenho do algoritmo RF foram o critério de entropia e o número de árvores igual a 250. Enquanto que no SVC, foi a função *kernel* polinomial e o parâmetro de regularização igual a 0.75. Para o MLP, foram selecionadas uma camada oculta com 100 neurônios, a função de ativação ReLU e a quantidade máxima de épocas igual a 2000. Visando a otimização do aprendizado da MLP e a redução de *Underfitting*, foi realizada uma normalização, onde os dados foram transformados pela remoção da média e em seguida pela divisão em relação ao desvio padrão, para cada *feature*.

C. Treino, teste e obtenção de métricas

A proporção do conjunto de dados foi distribuída em 80% para o conjunto de treino e 20% para o conjunto de teste.

Para a avaliação de desempenho de cada modelo, foram utilizadas as métricas de *precision*, *recall* e *f1-score* [7].

III. Resultados

A Tabela I demonstra os resultados de *precision*, *recall* e *f1-score* no conjunto de teste. Em relação à precisão, o algoritmo RF obteve melhor desempenho para a classe Laringite (96%), enquanto que SVC obteve na classificação das classes Disfonia (100%) e Saudável (88%). Em *recall*, o algoritmo SVC se mostrou mais robusto, para todas as classes, apesar da diferença de desempenho na classe Disfonia (75%) em relação as demais (100%). Na métrica de *f1-score*, o SVC se apresentou com alto desempenho nas três classes, com destaque paras as classes que foram balanceadas pelo SMOTE (94% e 96%).

TABLE I
Resultados dos classificadores para o conjunto de teste.

Modelos	**Classes**	***precision***	***recall***	***f1-score***
RF	Disfonia	88%	75%	81%
	Saudável	78%	91%	84%
	Laringite	**96%**	92%	94%
SVC	Disfonia	**100%**	**75%**	**86%**
	Saudável	**88%**	**100%**	**94%**
	Laringite	93%	**100%**	**96%**
MLP	Disfonia	92%	60%	73%
	Saudável	81%	96%	88%
	Laringite	89%	100%	94%

IV. Conclusão

Tendo em vista a metodologia aplicada, o algoritmo SVC apresentou o melhor desempenho, que pode ser justificado pelo fato dos dados serem linearmente separáveis via hiperplano do SVC. O algoritmo RF realiza sucessivas associações entre registros, logo a tentativa de associar valores sequenciais de mesma natureza pode ter implicado na redução de desempenho. Enquanto que para o algoritmo MLP, a otimização de poucos hiperparâmetros pode ter implicado na restrição de desempenho, apesar da normalização de dados aplicada.

Um dos desafios mais pertinentes na metodologia, inicialmente, foi o pré-processamento da base dados, dada a alta dimensionalidade no domínio do tempo. A transformação dos dados para o domínio da frequência apresentou *features* com alto desvio padrão, o que direcionou ao emprego da PSD.

Em trabalhos futuros, as restrições de desempenho citadas servirão para aprimoramento da metodologia aplicada e refinamento dos modelos para obtenção de resultados mais robustos.

Agradecimentos

Este trabalho foi financiado com recursos da Fundação Amazônia de Amparo a Estudos e Pesquisas (Fapespa).

Referências

[1] M. Pham, J. Lin, and Y. Zhang, "Diagnosing voice disorder with machine learning," in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 5263–5266.

[2] N. Pinho, D. Azevedo, and A. Santos, "Classifying cardiac rhythms by means of digital signal processing and machine learning," *Journal of Communication and Information Systems*, vol. 35, no. 1, 2020.

[3] L. Verde, G. De Pietro, and G. Sannino, "Voice disorder identification by using machine learning techniques," *IEEE Access*, vol. 6, pp. 16 246–16 255, 2018.

[4] A. Panigrahi, "Voiced: A database for health and pathological voices," Jul 2023, https://www.kaggle.com/datasets/abhranta/voiced.

[5] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[6] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[7] M. A. Maloof, *Some Basic Concept of Machine Learning and Data Mining*. London: Springer London, 2006, pp. 23–43.

[8] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, no. 10, pp. 281–305, 2012.

[9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.