

Abstract— There are indications in the literature that digital signal processing allows non-invasive diagnosis of laryngeal pathologies, and there is no definition of the most indicated method and the characteristics or parameters best suited to detect the presence of deviations. In this work, a study is performed for the detection of vocal deviations using voices from the Disordered Voice Database. We used 166 signals distributed between healthy voices and pathological voices affected by edema, nodule and paralysis in the vocal folds. Auto-Regressive models (AR and ARMA) were generated for the representation of these signals and, using the parameters of the obtained models, the K-Nearest Neighbors algorithm was used to classify the analyzed signals. The obtained results were compared with a classification only by the Euclidean distance between the signs and indicated a good result in the proposed study, with a rating rate in the classification superior to 71%.

Keywords— Detection of vocal deviations, AR, ARMA, K-nearest neighbor.

Resumo— Existe na literatura indicações de que o processamento digital de sinais permite diagnosticar, de maneira não invasiva, patologias laringeas, não existindo definição do método mais indicado e das características, ou parâmetros, mais adequados para detectar a presença de desvios. Nesse trabalho é realizado um estudo para detecção de desvios vocais utilizando-se vozes do banco de dados *Disordered Voice Database*. Foram utilizados 166 sinais distribuídos entre vozes saudáveis e vozes patológicas afetadas por edema, por nódulo e por paralisia nas pregas vocais. A partir dos sinais de voz, foram gerados modelos Auto Regressivos (AR e ARMA) para representação desses sinais e, utilizando os parâmetros dos modelos obtidos, foi utilizado o algoritmo *K-Nearest Neighbors* para a classificação dos sinais analisados. Os resultados obtidos foram comparados com uma classificação apenas pela distância euclidiana entre os sinais e apontaram um bom resultado no estudo proposto, com uma taxa de acerto na classificação superior a 71%.

Palavras-chave— Detecção de desvios vocais, AR, ARMA, K-nearest neighbor.

1 Introdução

A fala humana é o meio de comunicação e expressão mais importante para o ser humano, sendo um objeto de estudo atrativo em diferentes campos da ciência. As doenças no trato vocal provocam mudanças nos padrões de vibração da voz, além de afetarem significativamente a emissão e o entendimento da informação a ser transmitida.

Os problemas na produção da voz podem surgir devido a dois fatores: (i) uma desordem funcional causada pelo abuso, ou uso incorreto, do sistema vocal, podendo ser corrigidos por meio de terapia vocal; (ii) por patologias laringeas, como nódulos nas pregas vocais, pólipos, úlceras, carcinomas e paralisia do nervo laríngeo, que podem ser corrigidos por meio de terapia vocal, cirurgia e, em alguns casos, radioterapia (Martínez e Rufiner, 2000).

A detecção de patologias na laringe tem ocorrido, basicamente, por meio de diagnósticos médicos, apoiados em exames clínicos invasivos e que causam desconforto ao paciente (Souza, 2010). Comumente, é realizado um exame inicial de audição da voz pelos otorrinolaringologistas, juntamente com a anamnese e, posteriormente, exames físicos e visuais da laringe, tais como: a videolaringoscopia direta e a videostroboscopia (Parraga, 2002). Esses exames são realizados por meio de solicitação médica e, na maioria das vezes, só ocorrem quando as alterações na fala já são marcantes ou estão causando dor ao indivíduo. Nesse ponto, a doença já está em um grau avançado, dificultando o tratamento (Souza, 2010).

Diversos métodos têm sido propostos para avaliar a presença de desvios vocais e quantificar o grau dos distúrbios. Entretanto, a definição do melhor método e das características, ou parâmetros adequa-

dos para detectar a existência, bem como o grau dos desvios, ou a presença de patologias laringeas, ainda é objeto de estudo (Pinho, 2017).

O conceito principal para a análise de sinais de voz evidencia a dependência da estrutura do sistema vocálico, ou seja, qualquer distúrbio que ocorra na laringe é refletido no sinal. Diversas técnicas de processamento digital de sinais têm sido utilizadas ao longo dos anos como uma eficiente ferramenta não invasiva para diagnosticar as alterações na produção dos sons provocados por patologias da laringe, classificação de doenças da voz e sua pré-deteção, auxiliando, dessa forma, no desenvolvimento do processo terapêutico (Costa et al., 2007). A maioria das pesquisas realizadas até o momento baseiam-se em três metodologias: acústica, paramétrica e não paramétrica ou em métodos estatísticos.

O desenvolvimento de instrumentos para monitorar e auxiliar em diagnósticos médicos inclui ferramentas interdisciplinares das áreas de matemática, física e engenharia. A abordagem de sistemas biomédicos como sistemas dinâmicos tem se constituído uma linha de pesquisa de grande relevância, onde alguns autores defendem a ideia de que muitas doenças humanas podem ser determinadas por meio de técnicas de modelagem e de análise de sistemas não lineares (Glass; Mackey, 1988).

2 Identificação de Sistemas

De acordo com Mathias (1998), o campo de abrangência do problema de identificação de sistemas engloba diferentes ramos da ciência e, consequentemente, a terminologia utilizada na literatura não é única. No contexto geral, as aplicações no campo da identificação podem ser divididas como um processo ao longo do tempo, em que informações do passado

são aplicadas na análise de comportamento e interpretação de operação; informações do presente são aplicadas na análise de comportamento e monitoramento; e informações do futuro são aplicadas à análise de comportamento e predição.

Coelho e Coelho (2015) discorre que em controle de processos, denota-se um sistema como um objeto, ou uma coleção de objetos, que realiza determinada função, cujas propriedades pretende-se estudar. Uma das alternativas existentes para representar o comportamento dinâmico de um sistema é o uso de modelos matemáticos.

Assim, o modelo matemático é uma aproximação das características biológicas, químicas e físicas de um sistema dinâmico. Sendo assim, não existe um modelo exato para descrever o comportamento dinâmico e sim um conjunto de modelos que podem atender as especificações necessárias para a modelagem do sistema.

De acordo com Ljung (1987) existem diversas alternativas no processo de construção de estruturas para os modelos matemáticos com o objetivo de representar o comportamento dinâmico de um sistema ou planta da melhor forma possível. Uma destas maneiras, é a representação da dinâmica do processo a partir de equações a diferenças.

2.1 Modelos Auto Regressivos

Para Jacobs e Costa (2011) a análise por meio de séries temporais é um método que mapeia uma determinada variável ao longo do tempo, removendo as variações com causas assinaláveis e utilizando a extrapolação para previsão do comportamento futuro. Tais modelos tentam prever o futuro com base em dados passados. É uma importante área da previsão de demanda onde dados passados de determinada variável são coletados e analisados para o desenvolvimento de um modelo que descreva a relação não descoberta desta variável.

Conforme descrito em Ljung (1998), um modelo discreto linear invariante no tempo, que relaciona a saída, $y(k)$, com a entrada, $u(k)$, e um ruído gaussiano, $v(k)$, conforme equação abaixo:

$$A(z^{-1})y(k) = \frac{B(z^{-1})}{F(z^{-1})}u(k) + \frac{C(z^{-1})}{D(z^{-1})}v(k) \quad (1)$$

em que $A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$, $D(z^{-1})$ e $F(z^{-1})$ são polinômios com ordens na , nb , nc , nd e nf , respectivamente.

A partir da Equação 1, é possível classificar os modelos formados por estes polinômios, conforme mostra a Tabela 1.

Tabela 1. Modelos discretos conforme polinômios.

Polinômios	Modelo
A	AR
A, B	ARX
A, B, C	ARMAX
A, C	ARMA
A, B, D	ARARX
A, B, C, D	ARARMAX
B, F	<i>Output Error</i>
B, F, C, D	<i>Box-Jenkins</i>

Da Equação 1, observa-se que em situações em que a série temporal da componente de excitação não é observada, ela se reduz aos modelos AR e ARMA.

O modelo estrutural ARMA é denominado paramétrico baseado na hipótese que seus parâmetros guardam informações relativas à dinâmica do sistema, uma vez que esses são relacionados com a covariância equivalente do sistema e ao processo de excitação.

Uma vez escolhido o modelo matemático para representar o fenômeno a ser analisado, torna-se interessante a estimação dos parâmetros de tal modelo.

2.2 Estimador dos Mínimos Quadrados

Uma variedade de métodos para a estimação de parâmetros de modelos pode ser encontrada na literatura, mas, dá-se um destaque ao estimador dos mínimos quadrados, visto que ele é o pilar de desdobramento de outros métodos.

O estimador dos mínimos quadrados realiza o ajuste de curva para um determinado conjunto de dados obtidos experimentalmente, estimando assim, os coeficientes do modelo matemático para o sistema em estudo.

A saída do sistema em estudo pode ser representada pela Equação 2,

$$\mathbf{y} = \mathbf{\Psi}\mathbf{\theta} + \mathbf{\xi} \quad (2)$$

em que $\mathbf{y}$ é o vetor de dados de saída do sistema, $\mathbf{\Psi}$ é a matriz de regressores, $\mathbf{\theta}$ é o vetor dos parâmetros estimados do sistema e $\mathbf{\xi}$ é o vetor de erro de estimação.

A partir da Equação 2, o objetivo é, portanto, definir os valores da matriz $\mathbf{\theta}$. Para isto, é utilizado o estimador dos mínimos quadrados, definido na Equação 3.

$$\mathbf{\hat{\theta}} = (\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T \mathbf{y} \quad (3)$$

Na Equação 3, a expressão $(\Psi^T \Psi)^{-1} \Psi^T$, é chamada de pseudoinversa de Ψ , e é utilizada, pois, em muitos casos, Ψ é uma matriz não inversível.

Com a utilização da Equação 3, é possível estimar um modelo para os sinais de voz em estudo. Porém, ainda é necessário a utilização de um algoritmo de classificação de tais modelos, baseado nos parâmetros estimados.

3 Classificador KNN

O algoritmo *K-Nearest Neighbor* (KNN) é um método de aprendizagem supervisionado. A metodologia de classificação consiste na identificação de grupos de indivíduos com características similares e seu posterior agrupamento (*clustering*) (Rosa, 2003). A ideia geral consiste em encontrar o número (K) de exemplos rotulados mais próximos do exemplo não classificado e, com base no rótulo desses exemplos mais próximos, é tomada a decisão relativa à classe do exemplo não rotulado (Ferrero, 2009). O valor de K (o número de vizinhos mais próximos que serão considerados pelo algoritmo) é definido pelo usuário, sendo recomendada a escolha de um número ímpar para evitar um empate na classificação.

Se duas classes A e B possuem vários pontos em seus domínios, dado um ponto desconhecido x , este ponto será classificado em função da quantidade de pontos cujas distâncias forem as menores possíveis em relação às classes A e B (Rosa, 2003).

O cálculo da distância da amostra desconhecida em relação às amostras conhecidas é realizado pela raiz quadrada do somatório do quadrado da diferença de cada uma das m variáveis (colunas) da amostra desconhecida (vd) em relação às m variáveis (colunas) de cada uma das i ($i = 1 \dots n$) amostras conhecidas (vc):

$$dist_i = \sqrt{\sum_{j=1}^m (vd_j - vc_{ij})^2} \quad (4)$$

4 Materiais e Métodos

Nesta seção é feita a descrição da base de dados e da metodologia empregada no trabalho.

4.1 Base de Dados

O banco de dados *Disordered Voice Database* utilizado para esse trabalho foi desenvolvido pelo *Massachusetts Eye and Ear Infirmary* (MEEI).

A base contém pronúncias da vogal sustentada /a/, sendo 53 arquivos de falantes com voz saudável e 657 arquivos de falantes com vozes afetadas por patologias laringeas. Para esse trabalho foram utilizados

166 sinais contidos nesta base de dados, dos quais 53 sinais são de vozes saudáveis e 113 sinais são de vozes patológicas, sendo 43 sinais de vozes afetadas por edema, 18 sinais afetados por nódulo e 52 sinais afetados por paralisia nas pregas vocais.

Todos os sinais usados têm a duração de 1 a 3 segundos e foram amostrados na frequência de 25 kHz, preservando o tamanho do segmento em amostras. Esses sinais são previamente processados, não contendo intervalos de silêncio antes ou depois da locução.

4.2 Análise Inicial do Sinal de Voz

A aplicação de análises estatísticas básicas para a investigação da presença de um padrão de união baseado nas características extraídas de cada amostra segue as etapas do procedimento básico de avaliação de um sinal, de tal forma que pode ser descrito nos seguintes passos:

- Separar os sinais em grupos equivalentes;
- Calcular a média de cada sinal de voz;
- Subtrair a média calculada de cada amostra do seu respectivo sinal original.
- Normalizar os dados;
- Delimitar cinco sinais de cada categoria para composição do grupo de controle;
- Criar 10 janelas aleatórias em cada sinal de controle;
- Encontrar a menor distância euclidiana;
- Calcular os valores de média e de desvio padrão dos resultados encontrados;
- Calcular a matriz de confusão.

4.3 Algoritmos de Identificação

Após a análise inicial dos sinais de voz, procedeu-se para a etapa de identificação. Os dados utilizados foram obtidos de condições experimentais controladas. O procedimento de identificação adotado consistiu em estimar modelos paramétricos AR e ARMA, de 1ª a 5ª ordem, para representar os sinais de voz normalizados.

Para verificar se era possível perceber visualmente um padrão de separação nos sinais de voz analisados através dos resultados da identificação, foi plotada a dispersão dos modelos AR de cento e quarenta e seis sinais estimados, excluindo os vinte sinais do grupo de controle, de 2ª e 3ª ordem em relação aos parâmetros a , considerando as características: normal, edema, nódulo e paralisia.

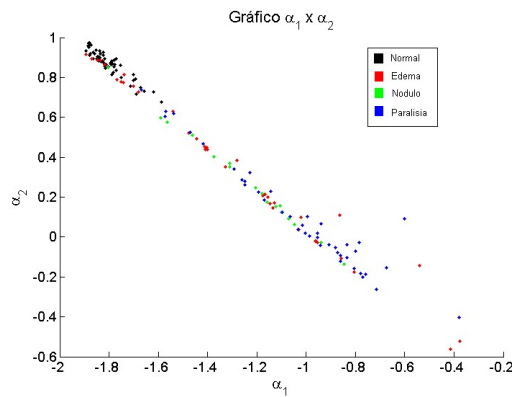


Figura 1. Gráfico de dispersão dos modelos de sinais no plano α_1 x α_2 .

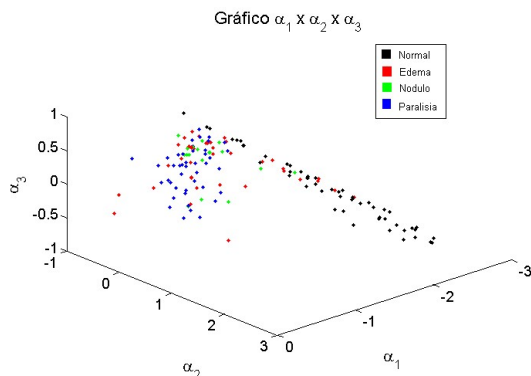


Figura 2. Gráfico de dispersão dos modelos de sinais no plano α_1 x α_2 x α_3 .

Através da representação gráfica dos resultados dos parâmetros estimados do modelo AR, observados nas Figuras 1 e 2, tomou-se a decisão da utilização de um classificador de dados, sendo escolhido o classificador KNN para os fins desse trabalho.

O método dos K vizinhos mais próximos tem sido aplicado na solução de problemas de classificação desde o início das pesquisas nessa área e, apesar de simples, tem se mostrado um dos métodos mais eficazes já propostos.

5 Resultados

Os sinais de voz foram normalizados, a fim de minimizar os problemas oriundos do uso de unidades e dispersões distintas, e divididos em grupos de controle contendo cinco amostras aleatórias de cada característica (normal, edema, nódulo e paralisia). Foi utilizado um índice quantitativo representado pelo cálculo da menor distância euclidiana na convolução das janelas aleatórias para comparar as amostras.

A matriz de confusão ilustrada na Tabela 1 foi calculada para distribuição dos registros nos termos de suas classes reais (R) e suas classes previstas (P), o que indica a qualidade do modelo atual frente a análise inicial dos sinais de voz.

Tabela 2. Matriz de confusão dos sinais analisados pela distância euclidiana.

R \ P	Normal	Edema	Nódulo	Paralisia
Normal	48 (100%)	0 (0%)	0 (0%)	0 (0%)
Edema	29 (76,3%)	3 (7,9%)	1 (2,6%)	5 (13,2%)
Nódulo	11 (84,6%)	1 (7,7%)	0 (0%)	1 (7,7%)
Paralisia	41 (87,2%)	3 (6,4%)	0 (0%)	3 (6,4%)

O acerto percentual foi calculado baseado no resultado da diagonal principal da matriz de confusão apresentada na Tabela 2, em que foram analisados os resultados de agrupamento do total de sinais analisados daquela característica em relação ao resultado real encontrado para a característica analisada. A precisão encontrada foi de 36,98% em relação à análise utilizando apenas a distância euclidiana entre os sinais de voz, ou seja, dos cento e quarenta e seis sinais analisados, excluindo os sinais do grupo de controle, apenas cinquenta e quatro foram associados ao grupo real ao qual pertencem.

O algoritmo de classificação do KNN foi empregado no reconhecimento de padrões dos sinais de voz. O KNN classificou um dado elemento de acordo com as respectivas classes dos K ($K \geq 1$) vizinhos mais próximos pertencentes à base de treinamento dada.

Em modelos AR de 1ª a 5ª ordem foi realizada a análise da matriz de confusão com base nos valores de $K = 1; 3$ e 5 , visto que valores de K pares poderiam resultar em empates indesejados no processo de classificação.

Na Tabela 3 é apresentada a matriz de confusão de um modelo AR de 2ª ordem com $K=1$, onde o acerto percentual é de 71,69%, resultado bem superior quando comparado à análise inicial e corrobora para comprovação da qualidade da proposta apresentada no trabalho.

Tabela 3. Matriz de confusão do KNN com $K=1$ para modelos AR 2ª ordem.

R \ P	Normal	Edema	Nódulo	Paralisia
Normal	51 96,2%	2 3,8%	0 0%	0 0%
Edema	8 18,6%	18 41,9%	5 11,6%	12 27,9%
Nódulo	1 5,6%	2 11,1%	12 66,7%	3 16,6%
Paralisia	1 1,9%	6 11,5%	7 13,5%	38 73,1%

A precisão encontrada para o modelo AR de 3ª, 4ª e 5ª ordem, considerando $K = 1$ foi de 59,04%, 56,63% e 56,02%, respectivamente.

Para modelos ARMA de 1ª a 5ª ordem com $K=1$, a precisão encontrada variou de 43,37%, em um modelo ARMA de 3ª ordem, a 53,01% em um modelo ARMA de 1ª ordem.

Para modelos AR de 1ª a 5ª ordem com $K=3$, a precisão encontrada variou de 32,53%, em um modelo AR de 1ª ordem a 65,66% em um modelo AR de 2ª ordem.

Para modelos ARMA de 1ª a 5ª ordem com $K=3$, a precisão encontrada variou de 32,16%, em um modelo AR de 3ª ordem, a 59,04% em um modelo AR de 2ª ordem.

Para modelos AR de 1ª a 5ª ordem com $K=5$, a precisão encontrada variou de 23,49%, em um modelo AR de 1ª ordem a 63,86% em um modelo AR de 2ª ordem.

Para modelos ARMA de 1ª a 5ª ordem com $K=5$, a precisão encontrada variou de 37,35%, em um modelo ARMA de 2ª ordem, a 60,24% em um modelo ARMA de 1ª ordem.

Foram analisados também os resultados da classificação de sinais normais versus patológicos, a fim de verificar a viabilidade do estudo proposto em detectar apenas a presença de distúrbios vocais.

A Tabela 4 apresenta os resultados da análise dos sinais utilizando apenas a distância euclidiana para verificar o acerto na classificação entre vozes normais e patológicas. Os resultados mostram um acerto de 100% na classificação das vozes normais, mas apenas um acerto de 28,3% em relação a classificação das vozes patológicas.

Tabela 4. Classificação de sinais Normais versus Patológicos utilizando a distância euclidiana.

R \ P	Normal	Patológico
Normal	53 100%	0 0%
Patológico	81 71,7%	32 28,3%

Os melhores resultados da classificação de sinais normais versus patológicos utilizando os modelos AR e ARMA de 1ª a 5ª ordem associados ao classificador KNN com $K=1$, 3 e 5 são apresentados detalhadamente nas Tabelas 5 a 10.

Os resultados encontrados mostram que a metodologia proposta apresenta altas taxas de acerto na classificação entre vozes normais e patológicas, e, resultados superiores na discriminação de patologias, quando comparados à análise inicial.

Tabela 5. Classificação de sinais Normais versus Patológicos para o modelo AR de ordem 2 com $K=1$.

R \ P	Normal	Patológico
Normal	51 96,2%	2 3,8%
Patológico	10 8,8%	103 91,2%

Tabela 6. Classificação de sinais Normais versus Patológicos para o modelo ARMA de ordem 1 com $K=1$.

R \ P	Normal	Patológico
Normal	41 77,4%	12 22,6%
Patológico	8 7,1%	105 92,9%

Tabela 7. Classificação de sinais Normais versus Patológicos para o modelo AR de ordem 2 com $K=3$.

R \ P	Normal	Patológico
Normal	53 100%	0 0%
Patológico	14 12,4%	99 87,6%

Tabela 8. Classificação de sinais Normais versus Patológicos para o modelo ARMA de ordem 1 com $K=3$.

R \ P	Normal	Patológico
Normal	52 98,1%	1 1,9%
Patológico	21 18,6%	92 81,4%

Tabela 9. Classificação de sinais Normais versus Patológicos para o modelo AR de ordem 2 com $K=5$.

R \ P	Normal	Patológico
Normal	52 98,1%	1 1,9%
Patológico	14 12,4%	99 87,6%

Tabela 10. Classificação de sinais Normais *versus* Patológicos para o modelo ARMA de ordem 1 com $K=5$.

$\begin{matrix} \text{P} \\ \text{R} \end{matrix}$	Normal	Patológico
Normal	50 94,3%	3 5,7%
Patológico	15 13,3%	98 86,7%

6 Conclusão

Os resultados encontrados apontam que o método proposto neste trabalho apresenta um resultado satisfatório, gerando uma taxa de acerto na classificação de vozes patológicas acima de 71% (maior que os 36% a partir do uso da distância euclidiana). Em relação a classificação entre vozes normais e patológicas os resultados ficaram entre 98% e 100% para vozes normais e 87% a 92% para vozes patológicas utilizando modelos AR e ARMA associados ao KNN.

Além disto, o método utilizado é de fácil implementação, podendo ser utilizado em hardwares mais simples. Logo, essa pesquisa tem potencial para gerar um classificador barato e acessível para utilização em larga escala por profissionais de saúde como uma alternativa de pré análise não invasiva para detecção de patologias otorrinolaringológicas que afetem a voz.

Referências Bibliográficas

- Coelho, A. A. R. and Coelho, L. S (2015). Identificação de sistemas dinâmicos lineares. Editora da UFSC, 2015. ISBN 978.85.328.0730-4.
- Costa, S. C. *et al.* (2007). Uso da entropia na discriminação de vozes patológicas. II Congresso de Inovação da Rede Norte Nordeste de Educação Tecnológica, João Pessoa, Paraíba.
- Ferrero, C. A (2009). Algoritmo KNN para previsão de dados temporais: funções de previsão e critérios de seleção de vizinhos próximos aplicados a variáveis ambientais em limnologia. Dissertação (Mestrado) — Instituto de Ciências Matemáticas e de Computação, ICMC-USP, São Carlos, SP.
- Glass, L. and Mackey, M. C (1988). From clocks to chaos: the rhythms of life. : Princeton University Press. ISBN 9780691084961.
- Ljung, L (1987). System Identification: Theory for the User. New Jersey: Pearson Education. ISBN 9780132440530.
- Ljung, L (1998). System Identification: Theory for the User. 2 ed. New Jersey: Pearson Education.
- Martinez, C. E. and Rufiner, H. L (2000). Acoustic analysis of speech for detection of laryngeal pathologies. Proceedings of the 22th Annual EMBS Conference, pp. 2369 – 2372.
- Mathias, M. H (1998). Aplicação da técnica paramétrica ARMA de processamento de sinais na caracterização de estruturas metálicas. Tese (Doutorado) — Universidade Estadual de Campinas - UNICAMP, Campinas, São Paulo.
- Parraga, A (2002). Aplicação da transformada wavelet packet na análise e classificação de sinais de vozes patológicas. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia Elétrica - UFRGS, Porto Alegre, RS.
- Pinho, P. H. U (2017). Avaliação de desordens vocais por meio de características do espaço de fase reconstruído. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia Elétrica - IFPB, João Pessoa, PB.
- Rosa, J. L. A (2003). Classificação de dados através da otimização do método KNN-Fuzzy em ambiente de computação paralela. Tese (Doutorado) — Universidade Federal do Rio de Janeiro - UFRJ, Rio de Janeiro, RJ.
- Souza, L. M (2010). Detecção inteligente de patologias na laringe baseada em máquinas vetores de suporte e na transformada wavelet. Dissertação (Mestrado) — Programa de PósGraduação em Física - USP, São Carlos, SP.