

Classificação de Sons Pulmonares com Técnica de *Few-Shot Learning*

Rafael Ávila dos Santos
Departamento de Automática
Universidade Federal de Lavras
Lavras, MG, Brasil
rafael.santos5@estudante.ufla.br

Danielle Abreu Fortunato
Departamento de Automática
Universidade Federal de Lavras
Lavras, MG, Brasil
danielle.fortunato@estudante.ufla.br

Danton Diego Ferreira
Departamento de Engenharia
Universidade Federal de Lavras (UFLA)
Lavras, MG, Brazil
danton@ufla.br

Bruno H.G. Barbosa
Departamento de Engenharia
Universidade Federal de Lavras (UFLA)
Lavras, MG, Brazil
brunohb@ufla.br

Abstract—A pandemia em decorrência da COVID-19 evidenciou para a humanidade a gravidade que uma doença pulmonar pode ter e a consequência se a mesma não for diagnosticada e contida o mais rápido possível. Dessa maneira, o reconhecimento de doenças pulmonares, o tratamento e a contenção são alvos de estudo há anos. Porém, as limitações de identificação manual utilizando os estetoscópios nos procedimentos de auscultas pulmonares acarreta na necessidade de outros exames, e consequentemente mais tempo para discernir a doença, e em casos graves onde isso não é imediato pode provocar a morte do paciente. Por esse motivo, este trabalho visa a elaboração de dois modelos de classificação de sons pulmonares utilizando o banco de dados ICBHI 2017 e as técnicas *Random Forest* e *Rede Siamesa*, comparando seus resultados e seus métodos. Utilizando o critério de validação cruzada k -fold, tais modelos obtiveram acurácias médias de 84.84% e 85.74% nos melhores métodos, respectivamente, mostrando ser viável a identificação de doenças a partir da análise de sinais de áudio de auscultas pulmonar.

Index Terms—lung sounds, siamese network, empirical mode decomposition, feature extraction, few shot learning

I. INTRODUÇÃO

Algumas doenças pulmonares são de alta letalidade, sendo a terceira principal causa responsável por mortes em todo o mundo [16] devido à sua forma de propagação que ocorre através do ar, como foi constatado pela descoberta da COVID-19, declarado pela Organização Mundial da Saúde (OMS) como emergência global em 2020 [12], a qual contaminou mais de 760 milhões de pessoas e quase 7 milhões de mortes em todo o mundo [11]. Além disso, outras doenças como doença pulmonar obstrutiva crônica (DPOC), asma, câncer de pulmão e tuberculose e infecções agudas do trato respiratório inferior contabilizam mais de 3 milhões de mortes por ano, segundo a OMS [13]. Dessa forma, é compreensível que tais doenças são uma preocupação humanitária e deve-se existir meios de evitá-los ou tratá-los de forma mais eficiente possível desde seu diagnóstico.

À vista disso, a exigência da identificação das doenças pulmonares o mais breve possível e de forma correta refletirá na recuperação e saúde dos próprios indivíduos. Com

isso, as técnicas de Machine Learning e outros algoritmos computacionais contornam as limitações existentes nas auscultas pulmonares que são feitas por médicos utilizando os estetoscópios, proporcionando melhorias e facilitações para os diagnósticos [14], [15].

Entretanto, fatores como a escassez de banco de dados e amostras com informações sobre as doenças pulmonares, além das próprias variações das doenças dificultam a criação de modelos de classificação de forma abrangente. Uma alternativa à isso é trabalhar com algoritmos que não dependam de um grande espaço amostral, como é o caso de técnicas *Few-Shot Learning*. Com essa técnica é possível desenvolver redes de classificação eficientes, mesmo se tratando de um banco de dados pequeno ou que possua desequilíbrio de amostras por classe.

Utilizando o banco de dados proveniente do desafio ICBHI 2017, este trabalho visa o desenvolvimento e comparativo de classificação dos sons pulmonares por duas abordagens diferentes, utilizando *Random Forest* e também *Redes Siamesas*, analisando suas respectivas acurácias e também a contribuição que o tratamento do áudio via EMD (*Empirical Mode Decomposition*) e DWT (*Discrete Wavelet Transform*) assim como a técnica de *oversampling ADASYN* refletem nas acurácias.

II. MATERIAIS E MÉTODOS

A. Banco de Dados

Afim de realizar os trabalhos aqui descritos, utilizou-se o banco de dados proveniente do desafio ICBHI 2017 [16]. O banco de dados disposto pelo desafio é composto por 920 áudios de 9 classes diferentes, sendo elas: Asthma, COPD (*Chronic Obstructive Pulmonary Disease*), URTI (*Upper Respiratory Tract Infection*), LRTI (*Lower Respiratory Tract Infection*), Pneumonia, Bronchiectasis, Bronchiolitis e Saudável. Os áudios, por sua vez, possuem taxas de amostragem que variam entre 44,1 kHz, 10 kHz e 4 kHz e foram capturados por quatro estetoscópios distintos, sendo eles: Microfone AKG C417L, Estetoscópio 3M Littmann Classic II SE, Estetoscópio

Eletrônico 3M Litmann 3200 e Estetoscópio eletrônico WelchAllyn Meditron Master Elite.

Do banco de dados original, selecionou-se 785 áudios com taxa de amostragem de 44,1 kHz e que variam entre as classes Saudável (30 amostras), COPD (697 amostras), URTI (21 amostras) e Pneumonia (37 amostras).

Pode-se distinguir dois grupos para os dados demográficos das amostras de sons pulmonares coletados: adultos e crianças. A idade das crianças variam entre 3 meses e 16 anos, sendo 21 do sexo feminino e 20 do sexo masculino. O peso varia entre 8,24 kg e 80 kg. Já a altura varia entre 68 cm e 183 cm.

A idade dos adultos variam entre 45 e 93 anos, sendo 17 do sexo feminino e 51 do sexo masculino. Neste caso, o IMC (Índice de Massa Corporal) variou entre 16,5 kg/m² e 53,5 kg/m².

B. Pré-Processamento

O pré-processamento dos dados foi realizado utilizando as técnicas de EMD (*Empirical Mode Decomposition*) [4], [8] e DWT (*Discrete Wavelet Transform*) [6], [9], [10], conforme sugerido em [7], [19] que aplicaram essas técnicas no problema de classificação de sons pulmonares.

A EMD é uma técnica para tratar sinais não-lineares e não estacionários como uma soma de componentes que recebem o nome de IMF (*Intrinsic Mode Functions*). Dessa forma, cada IMF é vista como uma decomposição que representa um intervalo de frequências de determinado som no tempo, sem perder o comprimento original do mesmo, decompondo um áudio que possui um intervalo grande de frequências em intervalos menores, como pode ser visto na figura 1. Isso torna possível reconstruir o áudio apenas com os intervalos desejados, possibilitando também outros feitos, por exemplo, a exclusão de componentes de alta ou baixa frequência de um áudio.

A DWT, por sua vez, é uma técnica bastante empregada para a remoção de ruídos indesejados, podendo ser representada pela equação:

$$W_L[a, b] = \frac{1}{\sqrt{a} \sum_R L[n] \Phi\left[\frac{n-b}{a}\right]} \quad (1)$$

em que a , b , Φ e W_L representam o fator de escala, o fator de translação, a *Wavelet* mãe e a função de transformação de *Wavelet* da série temporal de entrada $L[n]$, respectivamente. Neste trabalho, o *Coiflet5* foi usado como *Wavelet* mãe como também pode ser visto sua utilização em [7].

Os sons pulmonares traqueais, os quais são de grande importância para fins de classificação de doenças pulmonares, são identificados em frequências na faixa de 60-600 Hz [5], [8]. Sendo assim, o pré processamento dos áudios foi realizado no intuito de remover os ruídos em alta frequência e componentes de frequência indesejada.

Após aplicar a EMD, reconstruiu-se o sinal utilizando as IMF's 2, 3 e 4, conforme implementado de forma semelhante em [7], [8]. A figura 1 apresenta o exemplo do procedimento onde é possível obter a extração de 5 IMF's e no final o sinal reconstruído com as IMF's 2, 3 e 4 as quais correspondem

às componentes desejadas para o caso. A partir disso, utilizando a DWT, o sinal primeiramente é decomposto em dois coeficientes: de Aproximação e de Detalhes. O coeficiente de aproximação guarda informações de baixa frequência, enquanto que o coeficiente de detalhe se refere às componentes de alta frequência. Como o interesse é apenas nas baixa frequência, a limiarização remove todas as componentes do coeficiente de detalhe. Na sequência, o sinal é reconstruído mantendo apenas as baixas frequências desejadas.

Na figura 2 pode ser visto o comparativo do sinal após aplicado o método EMD e reconstruído com as IMF's selecionadas, com o sinal onde houve a aplicação de DWT. Com a DWT o sinal se apresenta mais definido, com amplitudes menores e sem interferências de alta frequência.

Após o pré-processamento, a extração de características foi fundamentada no plano temporal, na qual se extraiu 17 características temporais as quais foram utilizadas para o treinamento e validação dos modelos e serão expostas na próxima seção. A figura 3 representa a metodologia realizada para o pré-processamento dos sons pulmonares e a extração das características desejadas, assim como as variações utilizando as redes de classificação Random Forest e Siamesa para fins de análise dos comportamentos nestes casos distintos.

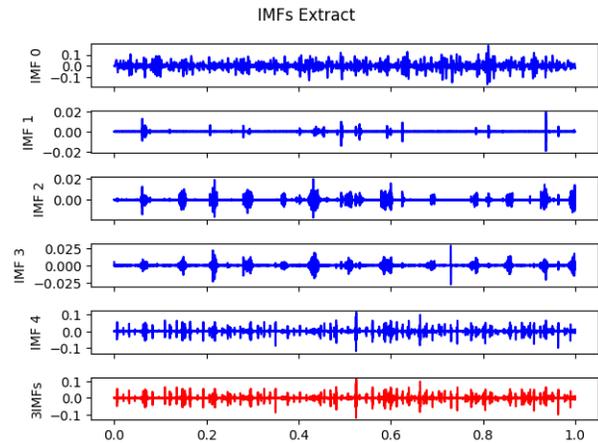


Fig. 1. Extração de IMFs e Reconstrução do Sinal

C. Características Temporais

As características temporais extraídas dos áudios após o tratamento com EMD e DWT foram escolhidas seguindo o critério semelhante visto em [7] as quais são: Média (2), Variância (3), Máximo Absoluto (6), Desvio padrão (4), Distorção (7), Curtose (8), Pico a pico, Raiz Quadrada Média (5), *Crest Factor* (9), *Margin Factor* (10), *Shape Factor* (11), *Impulse Factor* (12), Desvio Abstrato Mediano, Média Abs. Derivação, Raiz da Soma dos Quadrados (13), Energia (14) e Logaritmo da Energia (15):

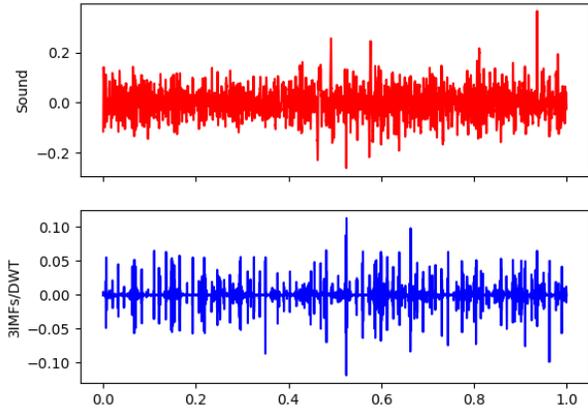


Fig. 2. Comparativo de Sinal com e sem Tratamento (EMD/DWT)

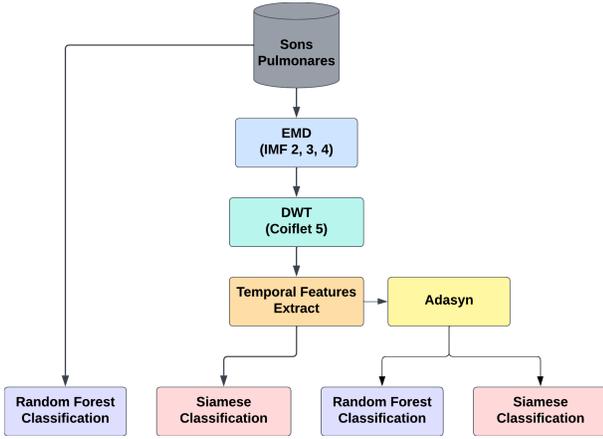


Fig. 3. Metodologia e Variação dos Classificadores

$$X_1 = \frac{\sum_{n=1}^N L[n]}{N},$$

$$X_2 = \left(\frac{\sum_{n=1}^N \sqrt{|L[n]|}}{N} \right)^2,$$

$$X_3 = \sqrt{\frac{\sum_{n=1}^N (L[n] - X_1)^2}{N - 1}},$$

$$X_4 = \sqrt{\frac{\sum_{n=1}^N (L[n])^2}{N}},$$

$$X_5 = \max |L[n]|,$$

$$X_6 = \frac{\sum_{n=1}^N (L[n] - X_1)^3}{(N - 1)(X_3)^3},$$

$$X_7 = \frac{\sum_{n=1}^N (L[n] - X_1)^4}{(N - 1)(X_3)^4},$$

$$X_8 = \frac{X_5}{X_4},$$

$$X_9 = \frac{X_5}{X_2},$$

$$X_{10} = \frac{X_4}{\frac{1}{N} \sum_{n=1}^N |L[n]|},$$

$$X_{11} = \frac{X_5}{\frac{1}{N} \sum_{n=1}^N |L[n]|}, \quad (12)$$

$$X_{12} = \sqrt{\sum_{n=1}^N (L[n])^2}, \quad (13)$$

$$X_{13} = \sum_{n=1}^N (|L[n]|)^2, \quad (14)$$

$$X_{14} = \text{Log} \left(\sum_{n=1}^N (|L[n]|)^2 \right). \quad (15)$$

D. Random Forest

A técnica *Random Forest* [20], [21] como seu nome presuppõe, visa a classificação de uma amostra com base na soma da distribuição de classificações dadas pelas N árvores de decisão que compõem a floresta. A aleatoriedade se dá pela escolha da instância a ser utilizada para as classificações de cada árvore, respectivamente. Ou seja, cada árvore interna escolhe o seu parâmetro a ser utilizado para classificar a amostra de forma arbitrária. Essa aleatoriedade contribui para prevenir a possibilidade de *overfitting* em treinamento da rede. A figura 4 exemplifica o funcionamento da *Random Forest*. Dado o conjunto de dados proveniente do áudio cada uma das 500 árvores de decisão que compõem a *Random Forest* neste trabalho se responsabilizará por escolher uma instância e utilizará esta para fazer sua classificação. Feito as classificações, a *Random Forest* contabiliza e verifica qual é a classificação majoritária, e assim retorna o resultado final.

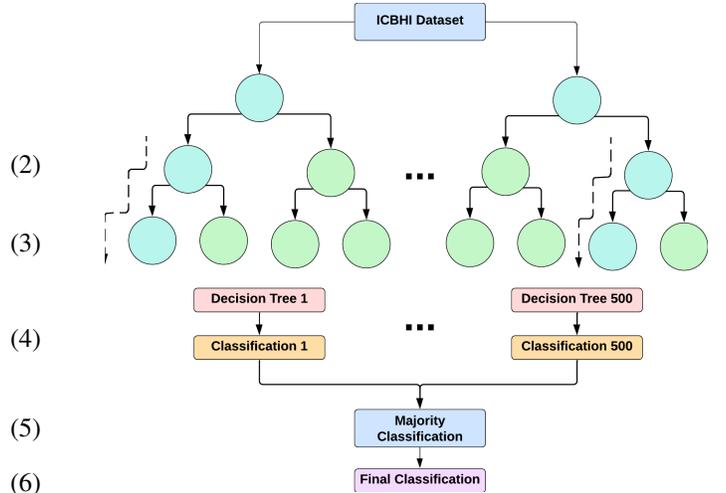


Fig. 4. Diagrama da Random Forest

(8) E. Few-Shot Learning

(9) O processo de aprendizagem com uso de poucas amostras é uma tarefa complexa para as redes neurais tradicionais e, principalmente, para as redes profundas [24].

(10) As redes profundas conseguem bons desempenhos em uma variedade de tarefas mas necessitam de uma grande quantidade de dados [25]. No caso de bases de dados com poucas amostras

(11)

por classes, o processo de treinamento é mais difícil devido ao risco de *overfitting*, reduzindo o desempenho do modelo. Assim, é necessário implementar métodos de aprendizagem de máquina específicos que sejam capazes de realizar o treinamento das redes com número pequeno de amostras, estes métodos de aprendizagem são conhecidos como *Few-Shot Learning* (FSL) [22], [23].

Existem vários métodos de aprendizado FSL, tais como: *Memory-Augmented Neural Networks* MANN [28], *Matching Networks* MN [26], *Model Agnostic Meta-Learning* MAML [27] e as Redes Siamesas [29].

As Redes Siamesas surgem com o intuito de reconhecer e classificar a partir da similaridade entre dois conjuntos de dados, estes podendo ser duas imagens, dois áudios, etc [17], [18]. Para isso, a arquitetura da Rede Siamesa é feita de forma a possuir duas redes idênticas e que compartilham seus pesos de ponderação interna, conforme apresentado na figura 5. A métrica para definir o grau de similaridade é dada pela Distância Euclidiana dos dois conjuntos de entradas X_1 e X_2 , a qual pode ser equacionada por:

$$D_w = \sqrt{(G_w(X_1) - G_w(X_2))^2}, \quad (16)$$

em que G_w é uma função que recebe os dados de entrada e representa as redes idênticas. A figura 6 apresenta a estrutura interna da rede siamesa desenvolvida neste trabalho. Por sua vez, a função de perda comumente usada para treinamento das siamesas é a Perda Contrastiva (*Contrastive Loss*) [29].

Como o intuito da rede siamesa é classificar com base em similaridades de dois conjuntos de dados, para realizar o seu treinamento é necessário gerar os pares e as rotulações que, por sua vez, foram descritas neste trabalho como 1 se o par é da mesma classe e 0 caso contrário.

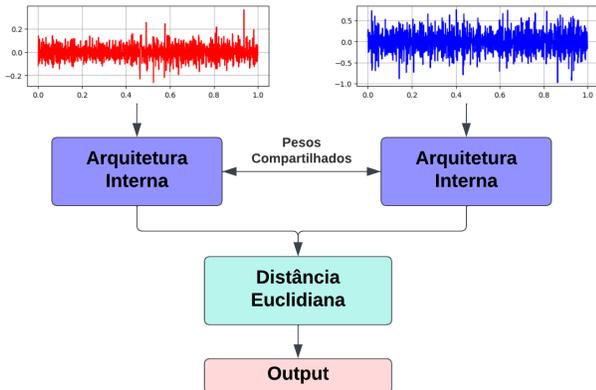


Fig. 5. Diagrama da Rede Siamesa

Para comparação com as Redes Siamesas, foi implementado o classificador do tipo *Random Forest*. Sua construção foi padronizada com 500 árvores, e cultivo de árvores como sendo 16. A construção foi feita utilizando *Google Colab* baseada em linguagem Python e fazendo uso da biblioteca *Sklearn* para sua implementação.

Os treinamentos se deram considerando três situações distintas: dados sem pré-processamento, dados com pré-

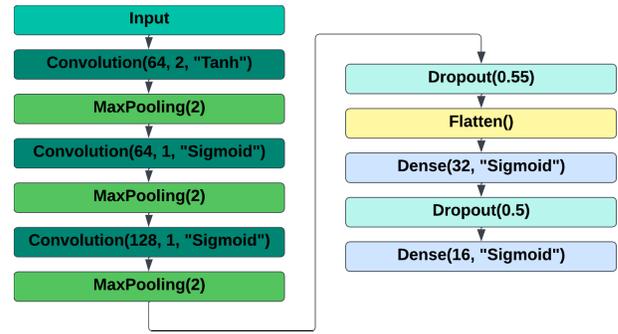


Fig. 6. Arquitetura Interna da Rede Siamesa

processamento e dados com pré-processamento e *data augmentation*. Para realizar o processo de *data augmentation*, utilizou-se da técnica de *over-sampling Adaptive Synthetic (ADASYN)* [3]. Neste caso, a ADASYN leva em consideração os 5 vizinhos mais próximos como vizinhança a ser usada para gerar a amostra sintética.

Para a Rede Siamesa, adotou-se os conjuntos de dados pré-processados com *data-augmentation* e sem *data-augmentation* uma vez que trata-se de uma técnica de *few-shot learning* e o objetivo era verificar seu desempenho em poucas amostras.

Afim de validar a classificação da Rede Siamesa e da *Random Forest* utilizou-se o procedimento de validação cruzada *k-fold*. Com isso, o conjunto de dados inicial é dividido em $K = 5$ grupos distintos. Para cada subconjunto de dados é realizado o treinamento e validação, analisando o comportamento do modelo de classificação.

III. RESULTADOS

Para a *Random Forest* e para a Rede Siamesa, a Tabela I apresenta os valores de acurácia encontrados para os respectivos conjuntos de dados:

TABLE I
ACURÁCIAS COM FEATURES TEMPORAIS PELOS MÉTODOS: 1 - RANDOM FOREST; 2 - RANDOM FOREST COM EMD-DWT + ADASYN; 3 - SIAMESAS COM EMD/DWT; 4 - SIAMESAS COM EMD/DWT + ADASYN;

	<i>Fold</i> ₁	<i>Fold</i> ₂	<i>Fold</i> ₃	<i>Fold</i> ₄	<i>Fold</i> ₅	Média
1	67.18%	63.36%	59.54%	52.67%	54.20%	59.39%
2	84.71%	85.35%	84.71%	85.35%	84.08%	84.84%
3	85.18%	89.01%	85.31%	83.12%	89.17%	85.74%
4	86.52%	86.47%	80.38%	83.72%	81.42%	83.70%

Analisando os resultados da *Random Forest*, pode-se observar que o método que apresentou maior acurácia média nos 5 *folds* foi aquele que empregou o tratamento com EMD/DWT, o qual apresenta um valor de 84.84%, representando um ganho médio de mais de 25% se comparado com os resultados sem tratamento que chegaram em 59.39% de acurácia. Tamanho diferença e ganho em acurácia evidencia que o tratamento removendo os ruídos e componentes de alta frequência dos dados foi significativo para uma melhor representação das características temporais, e assim resultando no aumentando da taxa de classificação do modelo.

Como a variação do método com *Random Forest* apresenta *data-augmentation*, ou seja, o conjunto de dados inicial utilizado para treinamento, que era desbalanceado tendo em vista a distribuição das amostras onde das 785, 697 eram de COPD, passando a ter todas as classes com amostras próximas a 697, torna-se pouco plausível a existência de *Overfitting* somado ainda à natureza aleatória dessa rede, explicada na metodologia.

Por sua vez, utilizando a Rede Siamesa e validação com 5 *folds*, para cada *fold* gerou-se 1256 pares para treinamento e 314 pares para teste, sendo treinadas por 250 épocas para os dados que apresentam apenas tratamento EMD/DWT (Caso 3 da Tabela I), e 350 épocas para os dados utilizando EMD/DWT em conjunto com a técnica de *oversampling Adasyn* (Caso 4 da Tabela I). As acurácias médias para os dados apenas com EMD/DWT e com EMD/DWT + Adasyn foram de 85.74% e 83.70% respectivamente. Dessa forma, pode-se observar que a normalização de amostras por classe proporcionada pela técnica de *oversampling* não refletiu em ganhos notórios de acurácia. Isso pode ser justificado tendo em vista que o treinamento e validação da Rede Siamesa se dá por pares de amostras, o que torna o número de amostras muito maior do que a própria classe super-amostrada. Isso também mostra a robustez das redes siamesas frente a problemas com menos amostras, como esperado.

É possível observar que em todos os *folds* de treinamento da Rede Siamesa, antes mesmo de 100 épocas de treinamento, as redes já tinham convergido para próximo de 80% de acurácias em treinamento e validação. Dos 5 *folds*, observou-se que os 3 primeiros apresentam menos variações de acurácias entre uma época e outra, apresentando um gráfico mais contínuo. Já nos dois últimos *folds* pode-se observar uma variação maior de acurácia entre épocas, porém, ainda sim permanecendo próximo a 80%.

IV. CONCLUSÕES

Por meio do banco de dados ICBHI 2017 de áudios pulmonares capturados por estetoscópios, foi feito o tratamento para a remoção de ruídos e componentes de frequência indesejadas utilizando as técnicas EMD/DWT. Com esses dados foram feitos os treinamentos de duas redes de classificação, Random Forest e Siamesas, comparando os seus respectivos resultados utilizando os dados em algumas variações: dados brutos (sem tratamento), com tratamento EMD/DWT, com tratamento EMD/DWT e *oversampling*.

A partir da tabela de acurácias observou-se uma discrepância considerável em relação aos valores da *Random Forest* com os dados tratados e sem tratamento. Dessa forma, é possível inferir que para esse modelo de classificação, o tratamento EMD/DWT melhorou as representações das características temporais dos áudios.

As acurácias providas utilizando a Rede Siamesa pelos dois métodos distintos, embora apresente diferença na metodologia de treinamento, não resultou em diferenças significativas para as acurácias. Com isso, pode-se inferir que a rede siamesa

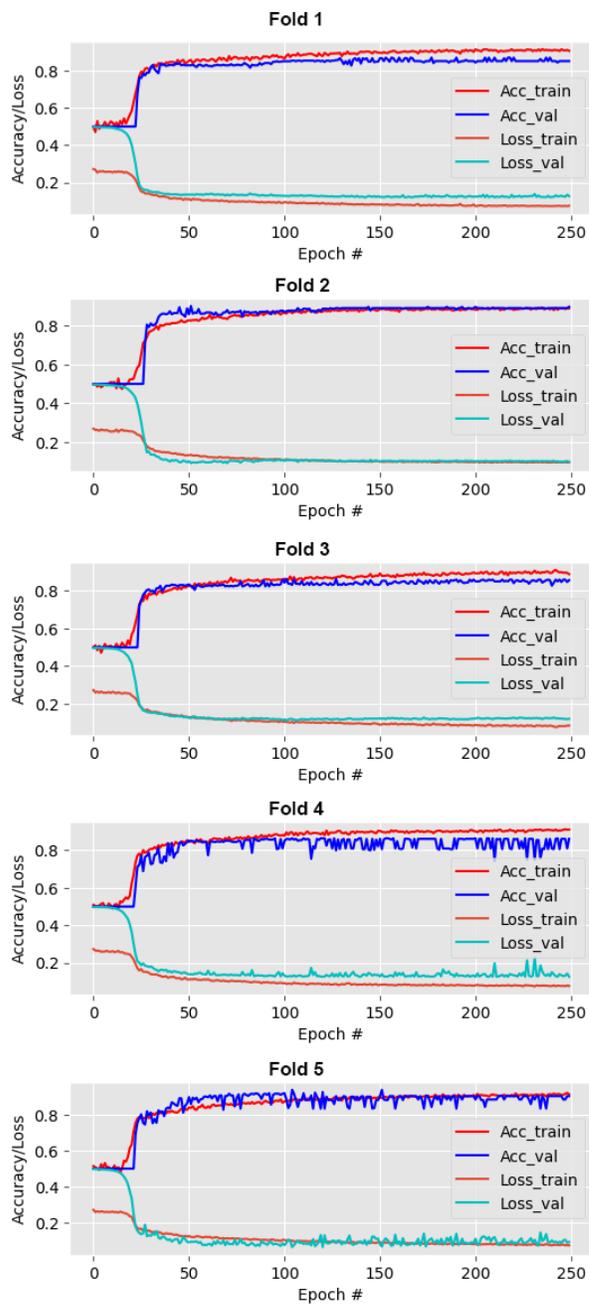


Fig. 7. Treinamento da Rede Siamesa com Para Dados com EMD/DWT

conseguiu classificar adequadamente as informações sem precisar de um banco de dados equilibrado, isto é, onde todas as classes apresentam o mesmo número de amostras, para o seu treinamento.

AGRADECIMENTOS

Agradecimento à FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais), ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e à UFLA (Universidade Federal de Lavras)

pelos recursos financeiros destinados e toda a infraestrutura necessária para execução do projeto.

REFERENCES

- [1] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface." *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
- [2] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [3] Ele, Haibo, Yang Bai, Eduardo A. Garcia e Shutao Li. "ADASYN: Abordagem de amostragem sintética adaptativa para aprendizado desequilibrado", na IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 1322-1328, 2008.
- [4] Quinn, Andrew J., et al. "EMD: empirical mode decomposition and Hilbert-Huang spectral analyses in Python." *Journal of open source software* 6.59 (2021).
- [5] Gross, Volker, et al. "The relationship between normal lung sounds, age, and gender." *American journal of respiratory and critical care medicine* 162.3 (2000): 905-909.
- [6] Singh, D.; Singh, B.K.; Behera, A.K. Comparative analysis of Lung sound denoising technique. In *Proceedings of the First International Conference on Power, Control, and Computing Technologies (ICPC2T)*, Raipur, India, 3–5 January 2020; pp. 406–410.
- [7] Naqvi, Syed Zohaib Hassan, and Mohammad Ahmad Choudhry. "An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis." *Sensors* 20.22 (2020): 6512.
- [8] Aziz, Sumair, et al. "An automated system towards diagnosis of pneumonia using pulmonary auscultations." *2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*. IEEE, 2019.
- [9] Singh, Neha, Neha Baranwal, and G. C. Nandi. "Implementation and evaluation of DWT and MFCC based ISL gesture recognition." *2014 9th International Conference on Industrial and Information Systems (ICIIS)*. IEEE, 2014.
- [10] Naing, Hay Mar Soe, et al. "Discrete wavelet denoising into MFCC for noise suppressive in automatic speech recognition system." *International Journal of Intelligent Engineering and Systems* 13.2 (2020): 74-82.
- [11] WHO Coronavirus (COVID-19) Dashboard, Accessed: Jun 08, 2023. [Online]. Available: <https://covid19.who.int/>
- [12] Strabelli, Tânia Mara Varejão, and David Everson Uip. "COVID-19 e o Coração." *Arquivos Brasileiros de Cardiologia* 114 (2020): 598-600.
- [13] Nguyen, Truc, and Franz Pernkopf. "Lung sound classification using cotuning and stochastic normalization." *IEEE Transactions on Biomedical Engineering* 69.9 (2022): 2872-2882.
- [14] Nguyen, Truc, and Franz Pernkopf. "Lung sound classification using snapshot ensemble of convolutional neural networks." *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*. IEEE, 2020.
- [15] Gurung, Arati, et al. "Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis." *Respiratory medicine* 105.9 (2011): 1396-1403.
- [16] Rocha, Bruno M., et al. "An open access database for the evaluation of respiratory sound classification algorithms." *Physiological measurement* 40.3 (2019): 035001.
- [17] Droghini, Diego, et al. "Few-shot siamese neural networks employing audio features for human-fall detection." *Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence*. 2018.
- [18] Manocha, Pranay, et al. "Content-based representations of audio using siamese neural networks." *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018.
- [19] Shuvo, Samiul Based, et al. "A lightweight cnn model for detecting respiratory diseases from lung auscultation sounds using emd-cwt-based hybrid scalogram." *IEEE Journal of Biomedical and Health Informatics* 25.7 (2020): 2595-2603.
- [20] Spathis, Dimitris e Panayiotis Vlamos. "Diagnosticando asma e doença pulmonar obstrutiva crônica com aprendizado de máquina." *Revista de informática em saúde* 25.3 (2019): 811-827.
- [21] Bodduluri, Sandeep, et al. "Deep neural network analyses of spirometry for structural phenotyping of chronic obstructive pulmonary disease." *JCI insight* 5.13 (2020).
- [22] Wang, Yaqing, et al. "Generalizing from a few examples: A survey on few-shot learning." *ACM computing surveys (csur)* 53.3 (2020): 1-34.
- [23] Parnami, Archit, and Minwoo Lee. "Learning from few examples: A summary of approaches to few-shot learning." *arXiv preprint arXiv:2203.04291* (2022).
- [24] JADON, S.; GARG, A. **Hands-On One-shot Learning with Python Learn**. [S.l.: s.n.], 2020. v. 1. 1–136 p. ISSN 1098-6596. ISBN 9788578110796.
- [25] ZHENG, Y. et al. Principal characteristic networks for few-shot learning. *Journal of Visual Communication and Image Representation*, Elsevier Inc., v. 59, p. 563–573, 2019. ISSN 10959076.
- [26] VINYALS, O. et al. Matching networks for one shot learning. *Advances in Neural Information Processing Systems*, p. 3637–3645, 2016. ISSN 10495258.
- [27] FINN, C.; ABBEEL, P.; LEVINE, S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In: *34th International Conference on Machine Learning, ICML 2017*. [S.l.: s.n.], 2017. v. 3, p. 1856–1868. ISBN 9781510855144.
- [28] SANTORO, A. et al. Meta-Learning with Memory-Augmented Neural Networks. *33rd International Conference on Machine Learning, ICML 2016*, v. 4, p. 2740–2751, 2016.
- [29] KOCH, G.; ZEMEL, R.; SALAKHUTDINOV, R. Siamese Neural Networks for One-shot Image Recognition Gregory. In: *32nd International Conference on Machine Learning*. [S.l.: s.n.], 2015. v. 37. ISSN 00071447.