

REDE NEURAL ARTIFICIAL APLICADA NA DETECÇÃO DE DIREÇÃO DO SINAL SONORO

WILLIAM MACHADO, WILIAN SOARES LACERDA, CARLOS ANTÔNIO RUFINO JÚNIOR*

**Universidade Federal de Lavras
Lavras, Minas Gerais, Brasil*

Email: williames.m@outlook.com, lacerda@dcc.ufla.br, carlos.rufinoj@gmail.com

Abstract— A very common problem in the daily life of the hearing impaired is the difficulty of detecting the direction of sound signals such as horns, motors of motor vehicles, among others. Often the hearing impaired can not tell whether the audible alert has been issued on the right or left side. In order to offer an alternative to this problem, the present work proposes an algorithm that implements an artificial neural network (ANN) to identify the direction of the sound signal, using for this the angular lag between the sound alerts. The results obtained are promising, since the One Step Secant (OSS) network topology showed 98,1 % hits.

Keywords— Artificial Neural Network, Sound Detector, Accessibility, Sound Direction.

Resumo— Um problema muito comum no dia-a-dia dos deficientes auditivos é a dificuldade de detectar a direção de sinais sonoros como buzinas, motores de veículos automotores, entre outros. Muitas vezes, os deficientes auditivos não conseguem distinguir se o alerta sonoro foi emitido do lado direito ou do lado esquerdo. No sentido de oferecer uma alternativa para este problema, o presente trabalho propõe um algoritmo que implementa uma Rede Neural Artificial (RNA) para identificar a direção do sinal sonoro, utilizando para isto o defasamento angular entre os alertas sonoros. Os resultados obtidos mostram-se promissores, tendo em vista que a topologia de rede *One Step Secant (OSS)* apresentou 98,1% de acertos.

Palavras-chave— Rede Neural Artificial, Detector de Sinal Sonoro, Acessibilidade, Direção Sonora.

1 Introdução

Não é recente a luta dos deficientes por melhores condições de vida. Em muitos casos, a falta de conscientização da sociedade a respeito de suas necessidades especiais, pode levar pessoas a tomarem ações equivocadas, que podem levar os surdos ao isolamento e exclusão da sociedade.

Lentamente e dolorosamente, os deficientes tem feito progresso ao serem considerados membros de uma minoria com necessidades especiais, cuja inclusão exige dimensões sociais, políticas e de cidadania (Antunes et al., 2011).

O emprego de tecnologias assistivas pode melhorar a qualidade de vida de milhares de pessoas que possuem diversos tipos de deficiências, dentre elas a deficiência auditiva.

A deficiência auditiva consiste na perda parcial ou completa da audição. Sendo assim, a expansão das cidades e a urbanização, assim como, o aumento de automóveis e equipamentos mecânicos tornou o trânsito desorganizado e sujeito a diversos tipos de imprevistos que podem criar problemas para as pessoas, especialmente para os deficientes.

Devido a esta desorganização, um dos desafios enfrentados pelos deficientes auditivos é a percepção do sinal sonoro, podendo ser o som do motor de um veículo que se aproxima, uma buzina, um apito de trem, entre outros sons de trânsito.

No sentido de melhorar a qualidade de vida das pessoas, o presente trabalho apresenta um algoritmo capaz de detectar a direção do sinal sonoro a partir de um sinal captado com diferentes

defasamentos, utilizando Redes Neurais Artificiais (RNA), no sentido de alertar o usuário a respeito de possíveis perigos que podem ocorrer no trânsito.

Uma das linhas de pesquisa na área de processamento de dados sonoros é a inteligência computacional que utiliza técnicas de reconhecimento de padrões como Redes Neurais Artificiais, Sistemas *Fuzzy* e uma das principais técnicas utilizadas recentemente é o *Deep Learning*.

Diversos trabalhos já utilizaram inteligência computacional para realizar o processamento de som. O trabalho proposto por Yu et al. (2016), utiliza uma antena para simular o ouvido humano e detectar micro-ondas sonoras. Uma outra técnica amplamente utilizada é a rede neural. O trabalho proposto por Vesperini et al. (2016) treinou uma rede neural artificial multi-camadas alimentada por Transformação de Fase de Correlação Cruzada, em inglês, *Generalized Cross-Correlation Phase Transform* (GCC-PHAT) com o objetivo de identificar auto-falantes ligados em uma sala.

Outras técnicas baseiam-se no efeito Doppler para identificar a aproximação de veículos híbridos e elétricos que são silenciosos. Neste caso, os veículos transmitem um sinal acústico modulado digitalmente que contém a sua identidade (ID), de modo que o dispositivo móvel do pedestre registra esse sinal, alertando sobre a aproximação do veículo individual em tempo real (Umezawa et al., 2017).

No cenário atual, a acessibilidade de deficientes auditivos é um grande desafio. De acordo

com a Organização Mundial de Saúde (OMS), a quantidade de deficientes auditivos no mundo é de aproximadamente 360 milhões. Estima-se que, um terço desta população, ou um total de 165 milhões é formado por pessoas acima de 65 anos e 32 milhões por crianças e adolescentes com idade igual ou inferior a 15 anos de idade (Pondé, 2009; Porsinal, 2013; Exame, 2016).

No Brasil, o censo de 2010 realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), 9,7 milhões de pessoas têm deficiência auditiva. Deste total, 2.147.366 milhões apresentam deficiência auditiva severa, situação em que ocorre perda entre 70 e 90 decibéis (dB). Estima-se que cerca de um milhão são jovens até 19 anos. (Brasil, 2016).

Com esse grande número de pessoas enfrentando diversos desafios, surge possibilidade de desenvolver um sistema para detectar a direção do sinal sonoro a partir de um sinal captado com diferentes defasamentos, utilizando Redes Neurais Artificiais (RNA), e técnicas eficientes e eficazes para reduzir as dificuldades enfrentadas por deficientes auditivos no cotidiano.

2 Revisão Teórica

A presente seção apresenta uma revisão dos conceitos para a compreensão de Redes Neurais, que são importantes para o algoritmo proposto.

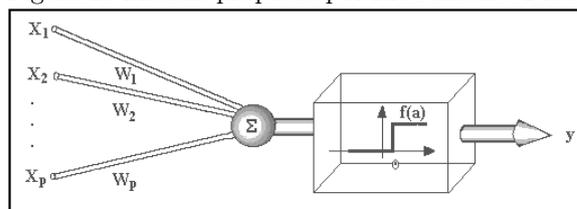
2.1 Neurônio Artificial

Um neurônio biológico pode ser interpretado como um circuito de entradas binárias combinadas por uma soma ponderada produzindo uma entrada efetiva, proposto por McCulloch and Pitts (1943) (Figura 1), o processamento pode ser resumido da seguinte maneira:

- Sinais são apresentados à entrada;
- Cada sinal é multiplicado por um número, ou peso, que indica sua influência na saída da unidade;
- É feito a soma ponderada dos sinais que produz um nível de atividade;
- Se este nível de atividade exceder um certo limite (*threshold*) a unidade produz uma determinada resposta de saída.

No modelo geral de neurônio, como mostrado na Figura 1, as entradas W_i , são combinadas usando uma função F , para produzir um estado de ativação do neurônio, as entradas chegam através dos dendritos e tem um peso atribuído pela sinapse (Carvalho, 2009).

Figura 1: Modelo proposto por McCulloch e Pitts



Fonte: (Carvalho, 2009)

2.2 Redes Neurais Artificiais

Uma rede neural artificial é composta por várias unidades de processamento, cujo funcionamento é bastante simples. Essas unidades, geralmente são conectadas por canais de comunicação que estão associados a determinado peso. As unidades fazem operações apenas sobre seus dados locais, que são entradas recebidas pelas suas conexões. O comportamento inteligente de uma Rede Neural Artificial vem das interações entre as unidades de processamento da rede.

Assim como o sistema nervoso é composto por bilhões de células nervosas, a rede neural artificial também seria formada por unidades que nada mais são que pequenos módulos que simulam o funcionamento de um neurônio. Estes módulos devem funcionar de acordo com os elementos em que foram inspirados, recebendo e retransmitindo informações.

A maioria dos modelos de redes neurais possui alguma regra de treinamento, onde os pesos de suas conexões são ajustados de acordo com os padrões apresentados. Em outras palavras, elas aprendem através de exemplos. Arquiteturas neurais são tipicamente organizadas em camadas, com unidades que podem estar conectadas às unidades da camada posterior.

Existem diferentes topologias de rede neural artificial, como por exemplo a *Feed-Forward* e *Cascade-Forward Back-Propagation*. As redes neurais do tipo *Feed-Forward*, são redes cujas camadas se conectam as próximas sem retroalimentação. Sendo assim, a informação é propagada pela rede em apenas uma direção e uma camada pode atualizar apenas o peso da próxima camada.

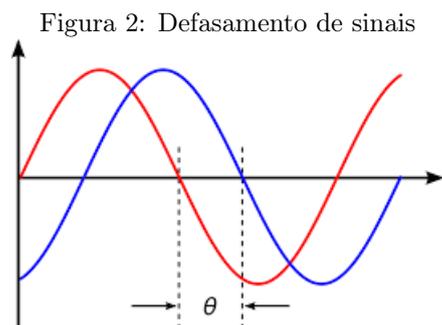
A topologia de rede *Cascade-Forward Back-Propagation* são semelhantes às redes *Feed-Forward*, porém uma camada pode atualizar os pesos de todas camadas posteriores.

3 Método proposto

O trabalho proposto tem por objetivo desenvolver um algoritmo para detectar a direção do sinal sonoro a partir de um sinal captado com diferentes defasamentos, utilizando RNA.

Um exemplo de sinal defasado é apresentado

na Figura 2, onde o Θ representa a distância do defasamento entre dois sinais.



Fonte: Elaborado pelo autor

O algoritmo, realiza a amostragem do sinal sonoro a uma frequência de 22 225 Hz. A partir disso, o algoritmo compara os sinais, uma vez que um sinal apresenta um atraso em relação a outro, identificando o defasamento. Posteriormente, é utilizado uma rede neural, para classificar os sinais em normais ou defasados.

3.1 Base de Dados

Os dados foram obtidos no repositório do site <<http://www.get-sounds.com/>>, que contém diversos sons, como buzina, barulho de motores, derrapagem de pneus, trens, carros, ônibus, avião, até sons de animais, entre outros.

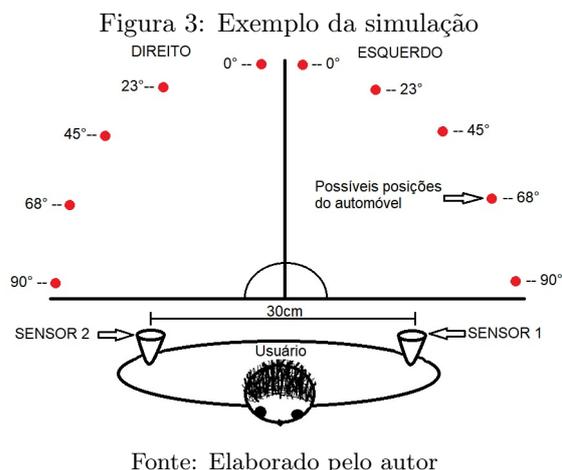
Foram coletados 40 tipos de sons deste repositório, cuja frequência de amostragem do sinal foi de 22 225 Hz. Em seguida mediu a distância entre o ombro de um usuário, e desta forma constatou que a largura média estimada entre o ombro direito e o ombro esquerdo é de aproximadamente 30 cm.

A partir disso, definiu os 5 tipos de atrasos em 0 cm, 8 cm, 15 cm, 23 cm, 30 cm, onde cada atraso foi relacionado em cada grau de defasamento do sinal, que são 0° , 23° , 45° , 68° e 90° , como mostrado na Figura 3, que representa as algumas das possíveis direções sonoras da fonte, cada direção estimada possui um atraso associado.

Desta forma elaborando os vetores de amostras para cada possível direção, compondo de 80 vetores normais e 80 vetores atrasados, totalizando 160 vetores de sons, com 370 linhas cada um que representa a quantidade de amostras, e deste modo criando a matriz de 740×160 para a entrada na rede neural.

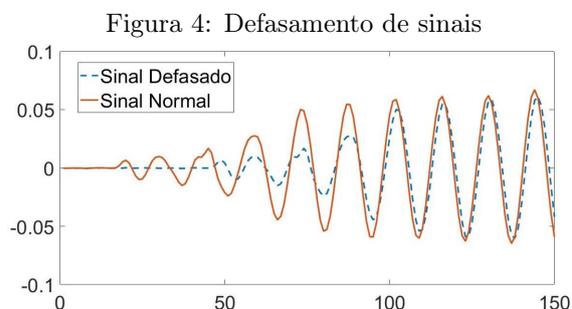
O banco de dados possui o total de 160 amostras, e foi dividido em 70% para dados de treinamento, 15% de dados para teste e 15% de dados de validação.

A Figura 4, apresenta uma forma de onda de um sinal normal e um sinal atrasado, que é apresentado para a rede neural. O sinal em vermelho, indica o som que é captado primeiro pelo sensor,



Fonte: Elaborado pelo autor

indicando assim, que a fonte sonora está mais próximo deste sensor.



Fonte: Elaborado pelo autor

3.2 Estrutura da Rede Neural Artificial

A rede neural testada é projetada com 3 camadas, sendo uma camada de entrada composta por 720 entradas (360 amostras do lado direito e 360 amostras do lado esquerdo), representando 1 vetor de som do lado direito, e 1 vetor de som do lado esquerdo, uma camada escondida composta por 20 neurônios e uma camada de saída composta por 2 neurônios, com o objetivo de classificar o atraso pela direita e pela esquerda. Por fim, definiu-se 1000 épocas de treinamento.

São implementadas duas técnicas de treinamentos supervisionados sendo elas: *Cascade-Forward*, *Feed-Forward*. Adicionalmente, quatro algoritmos de treinamentos são comparados, sendo eles: *One-Step-Secant (NOSS)*, *Gradient Descent with Momentum (GDM)*, *Gradient Descent with Momentum Adaptive LR (GDX)* e *Gradient Descent with Adaptive Learning Rate (GDA)*.

3.3 Avaliação de desempenho

Para avaliação da performance do trabalho, serão utilizados três métricas que são amplamente utilizadas na área de aprendizado de máquinas, sendo

elas: o Erro Médio Quadrático (MSE), o gráfico *Receiver Operating Characteristic* (ROC) e a matriz de confusão.

O erro médio quadrático é obtido através da diferença entre o valor obtido pela estimação e o valor real ponderado pelo número de termos, mostrado pela Equação 1.

$$MSE = \frac{(\hat{y} - y)^2}{n}, \quad (1)$$

onde \hat{y} é o valor estimado, y é o valor real e n é o número de amostras.

Outra métrica utilizada neste trabalho é o gráfico ROC. O gráfico ROC é um método gráfico amplamente utilizado para avaliar o desempenho de classificadores.

O gráfico ROC é geralmente utilizada para medir a sensibilidade e a especificidade do classificador. Ela traça um diagrama onde o eixo da abscissa representa a especificidade e o eixo das ordenadas representa a sensibilidade.

Segundo Prati et al. (2008), a gráfico ROC é baseado na probabilidade de detecção, ou taxa de verdadeiros positivos, e na probabilidade de falsos alarmes, ou taxa de falsos positivos.

A construção de um gráfico ROC dá-se pela disposição da taxa de falsos positivos no eixo X e pela taxa de verdadeiros positivos no eixo Y (Neto, 2016).

Além das técnicas supracitadas, uma outra técnica utilizada para avaliar o desempenho foi a matriz de confusão. Nesta técnica, as linhas da matriz representam as instâncias de uma classe de saída ou real e as colunas representam a classe prevista ou esperada.

A matriz de confusão é uma forma intuitiva de avaliar o desempenho do classificador, principalmente em algoritmos supervisionados.

4 Resultados Simulados

Com objetivo de obter melhores resultados, diferentes tipos de topologias e algoritmos de treinamento foram implementados para validação do método proposto. A rede neural é uma técnica heurística, por isso com o objetivo de realizar uma análise estatística dos resultados, executou-se a mesma rotina por um total de 10 vezes.

A Tabela 1, mostra o erro médio quadrático obtido para a topologia de rede *Feed-Forward Back-Propagation*, para diferentes tipos de algoritmos de treinamento.

Analisando a Tabela 1, o algoritmo de treinamento *GDX* obteve melhores resultados para a topologia de rede *Feed-Forward Back-Propagation*. A média do Erro Médio Quadrático (MSE) obtida para esse algoritmo foi de 0.0599 com 219 épocas de treinamento.

Por isso, a matriz de confusão mostrada na Figura 5, mostra os resultados da classificação para

Tabela 1: Erro médio quadrático para topologia de rede *Feed-Forward Back-Propagation*.

Algoritmo de treinamento	MSE		Iterações [Nº de épocas]	
	Média	Desvio Padrão	Média	Desvio Padrão
TrainNOSS	0.0717	0.0490	33	39
TrainGDX	0.0599	0.0206	219	375
TrainGDA	0.1559	0.0321	76	72
TrainGDM	0.1713	0.0180	312	474

o algoritmo GDX, para mostrar o comportamento do melhor classificador para a *Feed-Forward Back-Propagation*, e verificar se as instâncias estão corretas.

Figura 5: Matriz de confusão do Algoritmo GDX

Classe Real	Classe Estimada		
	1	2	
1	66 41.3%	2 1.3%	97.1% 2.9%
2	14 8.8%	78 48.8%	84.8% 15.2%
	82.5% 17.5%	97.5% 2.5%	90.0% 10.0%

Fonte: Elaborado pelo autor

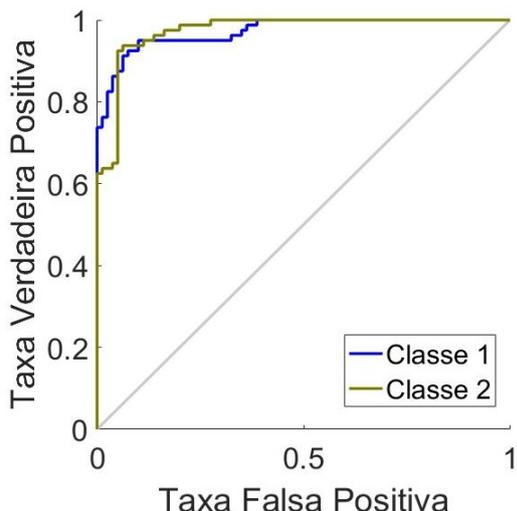
Observando a matriz de confusão do algoritmo GDX, mostrado na Figura 5, ela classificou 66 instâncias como '1'(direito) que realmente são '1'(direito), 78 instancias como '2'(esquerdo) que são '2'(esquerdo), 2 instância como '2'(esquerdo) que na verdade é '1'(direito), e 14 instâncias como '1'(direito) que na verdade são '2'(esquerdo), e assim obtendo 90.0% de acertos, e os erros foram 10.0%.

A Curva ROC do algoritmo GDX é mostrada na Figura 6. É possível perceber que a rede obteve um bom desempenho, pois a ROC esteve próximo do ponto (0, 1), ou (0%, 100%) que representa um bom resultado.

A Tabela 2, mostra os resultados obtidos para a topologia de rede *Cascade-Forward Back-Propagation*, para diferentes tipos de algoritmos de treinamento.

Os melhores resultados obtidos com os dados de teste foram obtidos com a topologia de rede *Cascade-Forward Back-Propagation* treinada com o algoritmo *One-Step-Secant* (NOSS). A média

Figura 6: Curva ROC do Algoritmo GDX



Fonte: Elaborado pelo autor

Tabela 2: Erro médio quadrático para topologia de rede *Cascade-Forward Back-Propagation*.

Algoritmo de treinamento	MSE		Iterações [Nº de épocas]	
	Média	Desvio Padrão	Média	Desvio Padrão
TrainNOSS	0.0450	0.0657	55	75
TrainGDX	0.1120	0.0336	41	48
TrainGDA	0.1384	0.0269	115	252
TrainGDM	0.1537	0.0257	415	503

do Erro Médio Quadrático (MSE) obtida foi de 0.0450 com 55 épocas de treinamento.

A Figura 7, mostra a classificação entre os dados de entrada e o dados estimados pelo algoritmo de treinamento que obteve melhor resultado *One Step Secant* (NOSS) para a rede *Cascade-Forward*, através da matriz de confusão.

Figura 7: Matriz de confusão do Algoritmo NOSS

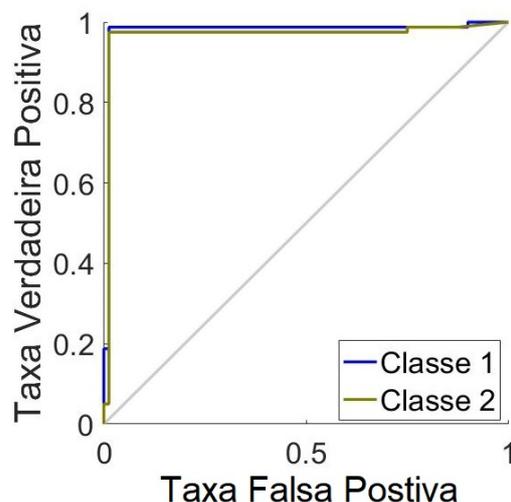
Classe Real	Classe Estimada		
	1	2	
1	79 49.4%	2 1,3%	97.5% 2.5%
2	1 0.6%	78 48.8%	98.7% 1.3%
	98.8% 1.2%	97.5% 2.5%	98.1% 1.9%

Fonte: Elaborado pelo autor

Examinando a matriz de confusão do algoritmo NOSS, mostrado na Figura 7, ela classificou 79 instâncias como '1'(direito) que realmente são '1'(direito), 78 instancias como '2'(esquerdo) que são '2'(esquerdo), 1 instância como '2'(esquerdo) que na verdade é '1'(direito), e 2 instâncias como '1'(direito) que na verdade são '2'(esquerdo), e assim obtendo 98.1% de acertos, e os erros foram 1.9%.

A Figura 8 mostra a curva ROC da rede neural NOSS, que obteve o menor Erro Médio Quadrático (MSE) tanto em relação aos demais algoritmos de treinamento para a topologia de rede *Cascade-Forward Back-Propagation* quanto para a topologia de rede *Feed-Forward Back-Propagation*.

Figura 8: Curva ROC do Algoritmo NOSS



Fonte: Elaborado pelo autor

Analisando a Curva ROC do algoritmo NOSS, mostrado na Figura 8, a rede obteve um desempenho ótimo, pois esteve próximo do ponto (0, 1), ou (0%, 100%) que representa o modelo perfeito, isto é, quase não obteve erros na classificação.

De acordo com os resultados obtidos, a topologia de rede *Cascade-Forward Back-Propagation* treinada com o algoritmo NOSS, obteve o melhor desempenho em relação às outras topologias de rede e algoritmos de treinamento testados.

É possível perceber comparando a Tabela 1 e a Tabela 2, que a topologia de rede *Feed-Forward Back-Propagation* necessitou menor números de épocas para treinar (exceto para o algoritmo de treinamento TrainGDA), porém a topologia de rede *Cascade-Forward Back-Propagation* treinada com o algoritmo NOSS apresentou um erro médio quadrático menor que a topologia de rede *Feed-Forward Back-Propagation* treinada com os algoritmos analisados.

5 Conclusão

Diversas topologias de redes neurais artificiais foram comparadas com o objetivo de obter melhores resultados. A rede que obteve menor erro médio quadrático (MSE) foi o *Cascade-Forward Back-Propagation* com o algoritmo *One Step Secant*, e a matriz de confusão plotada para este algoritmo, mostra uma taxa de acerto de 98,1%. Enquanto para os demais algoritmos como *Gradient Descent with Momentum Adaptive LR (GDX)* apresentaram uma taxa de acerto de 90,0%.

Os resultados mostraram-se satisfatórios, o algoritmo NOSS apresentando uma taxa de erro de 1,9% para o problema de detecção da direção do sinal sonoro, tornando viável a aplicação da rede neural neste problema.

Como trabalhos futuros, objetiva-se captar sons em ambiente urbano, fazendo uma implementação de uma *deep-learning* aplicada em um banco de dados com maior número de amostras com sons atuais. Objetiva-se também aplicar um pré-processamento dos dados com redução de quantidade de atributos e a utilização de um software mais robusto para a amostragem e filtragem de sinais, e um sistema embarcado com uma capacidade maior de armazenamento e processamento para os dados.

Desta forma possibilitando a testes em ambientes reais, tornando possível a melhoria da qualidade de vida dos deficientes auditivos.

Por fim pretende-se embarcar o sistema, após alcançados bons resultados em testes nos ambientes reais, criando um dispositivo, onde o usuário possa utilizar, e assim auxiliando no ambiente urbano, evitando acidentes.

6 Agradecimentos

Os autores agradecem à UFLA pelo suporte financeiro dado ao projeto sem os quais seria improvável a concretização deste trabalho.

Referências

- Antunes, D. R., Guimarães, C., García, L. S., Oliveira, L. E. S. and Fernandes, S. (2011). A framework to support development of sign language human-computer interaction: Building tools for effective information access and inclusion of the deaf, *2011 Fifth International Conference on Research Challenges in Information Science*, pp. 1–12.
- Brasil (2016). Apesar de avanços, surdos ainda enfrentam barreiras de acessibilidade, Portal Brasil. <http://www.brasil.gov.br/cidadania-e-justica/2016/09/apesar-de-avancos-surdos-ainda-enfrentam-barreiras-de-acessibilidade> Acessado em 11/01/2018.
- Carvalho, A. P. L. F. (2009). Redes neurais artificiais, Tutorial Introdutório sobre Redes Neurais Artificiais em especial sobre as Redes Multi Layer Perceptron treinadas com Back-Propagation. <http://conteudo.icmc.usp.br/pessoas/andre/research/neural/> Acessado em 20/01/2018.
- Exame (2016). OMS afirma que 60% dos casos de surdez podem ser prevenidos, Dos 60% dos casos evitáveis, 31% se devem a doenças como sarampo, caxumba, rubéola ou meningite. <https://exame.abril.com.br/mundo/oms-afirma-que-60-dos-casos-de-surdez-podem-ser-prevenidos/> Acessado em 15/01/2018.
- McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity, *The bulletin of mathematical biophysics*, Vol. 5, pp. 115–133.
- Neto, H. S. (2016). *Sistema de detecção de intrusão em redes de computadores com técnicas de inteligência computacional*, Master's thesis, Universidade Federal de Lavras.
- Pondé, L. F. (2009). 42 milhões de pessoas no mundo têm deficiência auditiva de moderada a profunda, Íntegras. <http://integras.blogspot.com.br/2009/02/42-milhoes-de-pessoas-no-mundo-tem.html> Acessado em 10/02/2018.
- Porsinal (2013). Segundo a OMS 360 milhões de pessoas no mundo sofrem de perda auditiva incapacitante, Porsinal. <http://www.porsinal.pt/index.php?ps=destaques&idt=not&iddest=154> Acessado em 01/02/2018.
- Prati, R. C., Batista, G. E. A. P. A. and Monard, M. C. (2008). Curvas roc para avaliação de classificadores, *IEEE Latin America Transactions* 6(2): 215–222.
- Umezawa, Y., Ebihara, T., Mizutani, K. and Wakatsuki, N. (2017). Detection of individual vehicles approaching for pedestrians using digital modulated warning sounds, *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*, pp. 1–5.
- Vesperini, F., Vecchiotti, P., Principi, E., Squartini, S. and Piazza, F. (2016). A neural network based algorithm for speaker localization in a multi-room environment, *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6.

Yu, X., Zhou, R., Zhang, H. and Xin, H. (2016).
A microwave direction of arrival estimation
technique using a single antenna, *IEEE
Transactions on Antennas and Propagation*
64(7): 3189–3195.