

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

CARLOS HENRIQUE RODRIGUES ALVES

**AVALIAÇÃO DE MODELOS DE PREVISÃO DO RUÍDO SONORO URBANO
COM REDES NEURAS ARTIFICIAIS**

PONTA GROSSA

2021

CARLOS HENRIQUE RODRIGUES ALVES

**AVALIAÇÃO DE MODELOS DE PREVISÃO DO RUÍDO SONORO URBANO
COM REDES NEURAIAS ARTIFICIAIS**

**Evaluation of urban sound noise forecast models with artificial neural
networks**

Dissertação apresentada como requisito parcial
à obtenção do título de Mestre em Engenharia
de Produção ao Programa de Pós-Graduação
em Engenharia de Produção da Universidade
Tecnológica Federal do Paraná (UTFPR).
Orientador: Prof. Dr. Flávio Trojan

PONTA GROSSA

2021



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite remixe, adaptação e criação a partir do trabalho, para fins não comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



CARLOS HENRIQUE RODRIGUES ALVES

**AVALIAÇÃO DE MODELOS DE PREVISÃO DO RUÍDO SONORO URBANO COM REDES NEURAIAS
ARTIFICIAIS**

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Engenharia De Produção da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Gestão Industrial.

Data de aprovação: 30 de Agosto de 2021

Prof Flavio Trojan, Doutorado - Universidade Tecnológica Federal do Paraná

Prof.a Anna Diva Plasencia Lotufo, Doutorado - Universidade Estadual Paulista - Unesp

Prof Hugo Valadares Siqueira, Doutorado - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 28/11/2021.

Dedico este trabalho à minha família e amigos que sempre me deram total suporte ao desenvolvimento pessoal e profissional. Sem vocês essa conquista teria sido imensuravelmente mais difícil.

AGRADECIMENTOS

Certamente há muitas pessoas que fizeram parte dessa importante fase de minha vida. Todavia, desejo deixar registrado meu agradecimento à família do Prof. Dr. Claudinor Bitencourt Nascimento e sua esposa Rosalba, amigos de longa data com os quais sempre pude contar e ser incentivado a continuar crescendo.

Aos meus colegas de turma e de vida Tatiane Zander, Hugo Leite e Janaína Medeiros pelos conselhos e conversas confortantes.

Gostaria de deixar registrado também, mais uma vez, o meu reconhecimento à minha família, pois acredito que sem o apoio deles seria difícil vencer esse desafio, e em especial à minhas afilhadas Melissa e Maria Fernanda.

Agradeço ao Prof. Dr. Lucas Nonato de Oliveira do Instituto Federal de Goiás (IFG-GO) pela disponibilização da base de dados.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

O saber a gente aprende com os mestres
e os livros. A sabedoria se aprende é com a vida
e com os humildes.

Cora Coralina

RESUMO

O ruído de tráfego urbano é uma das maiores fontes de poluição sonora sendo composto pelos diferentes tipos de veículos que circulam nas vias. Poucos estudos têm sido realizados em busca de uma metodologia capaz de antever o ruído sonoro urbano a partir do volume de tráfego e de diversas outras variáveis presentes no ambiente. Este trabalho objetiva avaliar por meio de comparações os menores erros de previsão do ruído sonoro do tráfego urbano usando os modelos de Redes Neurais Artificiais MLP, ELM e RBF frente ao modelo linear de Regressão Linear Múltipla. Os dados utilizados nesse trabalho são fruto da revisão bibliográfica acerca do tema e das bases de dados coletados nas cidades de Goiânia-GO e Maringá-PR. Os resultados encontrados pelas redes são comparados em termos de seus erros médios para avaliação dos melhores mapeadores do nível equivalente sonoro Leq que é a saída dos modelos e representa o ruído sonoro produzido pelo tráfego. Os modelos neurais nesse estudo se mostraram melhores mapeadores do Leq se comparados ao modelo linear.

Palavras-chave: redes neurais artificiais; poluição sonora; tráfego urbano.

ABSTRACT

The urban traffic noise is one of the biggest sources of noise pollution being composed by the different types of vehicles that circulate on the roads. Few studies have been carried out in search of a methodology capable of predicting urban noise based on the volume of traffic and several other variables present in the environment. This work aims to evaluate through comparisons the smallest noise prediction errors in urban traffic using the MLP, ELM and RBF Artificial Neural Networks models against the linear Multiple Linear Regression model. The data used in this work are the result of a literature review on the subject and databases collected in the cities of Goiânia-GO and Maringá-PR. The results found by the networks are compared in terms of their average errors to evaluate the best mappers of the equivalent sound level L_{eq} , which is the output of the models and represents the noise produced by traffic. The neural models in this study proved to be better L_{eq} mappers when compared to the linear model.

Keywords: artificial neural networks; noise pollution; urban traffic.

LISTA DE ILUSTRAÇÕES

Figura 1 – O neurônio biológico.....	27
Figura 2 – Modelo do neurônio artificial.....	29
Figura 3 – Gráfico da função degrau $g(u)$	31
Figura 4 – Gráfico função degrau bipolar $g(u)$	31
Figura 5 – Gráfico da função rampa simétrica.....	32
Figura 6 – Gráfico da função logística.....	33
Figura 7 – Gráfico da função tangente hiperbólica.....	34
Figura 8 – Gráfico da função gaussiana.....	35
Figura 9 – Gráfico da função linear.....	35
Figura 10 – Rede acíclica com 1 camada neuronal.....	37
Figura 11 – Esquema representativo de uma MLP.....	40
Figura 12 – Rede de função de base radial (RBF).....	44

LISTA DE GRÁFICOS

Gráfico 1 – Distribuição da frequência da variável dependente da base 1.....	54
Gráfico 2 – Distribuição da frequência da variável dependente da base 2.....	54
Gráfico 3 – Valores previstos e real para todos os modelos da base 1.....	58
Gráfico 4 – Valores previstos e real para todos os modelos da base 2.....	59
Gráfico 5 – Box-Plot de dispersão de MSE da base 1.....	60
Gráfico 6 – Box-Plot de dispersão de MSE da base 2.....	60

LISTA DE TABELAS

Tabela 1 – Artigos relacionado ao tema da pesquisa.....	18
Tabela 2 – Parâmetros estatísticos descritivos das bases.....	53
Tabela 3 – Parâmetros estatísticos descritivos das entradas da base 1.....	55
Tabela 4 – Parâmetros estatísticos descritivos das entradas da base 2.....	55
Tabela 5 – Matrizes de correlações da base 1.....	56
Tabela 6 – Matrizes de correlações da base 2.....	56
Tabela 7 – Parâmetros estatísticos das previsões da base 1.....	58
Tabela 8 – Parâmetros estatísticos das previsões da base 2.....	59

LISTA DE SIGLAS E SIGLAS

ASJ - Acoustic Society of Japan
CORTN – Calculation of Road Traffic Noise
ELM – Extreme Learning Machine
FHWA - Federal Highway Administration
LEQ – Equivalent sound level
MA – Moving Average
MAE – Mean Absolute Error
MAPE – Mean Absolute Percentage Error
MLP – Multilayer Perceptron
MME – Média Móvel Exponencial
MMP - Média Móvel Ponderada
MMS – Média Móvel Simplificada
MSE – Mean Square Error
mV - Milivolt
PNAD - Pesquisa Nacional por Amostra de Domicílios
RBF – Radial Basis Function
RLM – Regressão Linear Múltipla
RMSE - Root Mean Square Error
RNA - Redes Neurais Artificiais
WHO – World Health Organization

SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	Objetivos do trabalho.....	17
1.2	Trabalhos relacionados.....	17
1.3	Justificativa.....	22
2	ESTATÍSTICA MULTIVARIADA.....	24
2.1	Regressão Linear Múltipla.....	24
3	REDES NEURAIS ARTIFICIAIS.....	26
3.1	O neurônio artificial.....	28
3.1.1	Tipos de funções de ativação.....	30
3.2	Arquitetura de redes neurais.....	36
3.3	Fase de treinamento da rede.....	38
3.3.1	Treinamento supervisionado.....	38
3.3.2	Validação Cruzada.....	39
3.4	As redes de múltiplas camadas.....	40
3.4.1	O treinamento de uma MLP usando o algoritmo <i>backpropagation</i>	41
3.5	As redes RBF.....	43
3.5.1	Treinamento das RBF.....	43
3.6	As Máquinas de Aprendizado Extremo.....	46
3.6.1	Treinamento de uma ELM.....	46
4	METODOLOGIA.....	49
4.1	Enquadramento metodológico do trabalho.....	49
4.2	Coleta de dados.....	49
4.3	Pré-processamento dos dados.....	51
4.4	Modelos de previsão.....	51
5	RESULTADOS.....	53
5.1	Variáveis.....	53
5.1.1	Análise estatística descritiva da variável dependente.....	53
5.1.2	Análise estatística das variáveis independente.....	53
5.2	Previsão da variável dependente.....	58
6	CONCLUSÕES.....	63
	REFERÊNCIAS.....	65

1 INTRODUÇÃO

A mobilidade urbana tem sido um problema para as populações globais, especialmente em razão do crescimento desordenado que ocorreu nas últimas décadas e devido às mudanças na estrutura econômica e produtiva dos países em desenvolvimento, dentre eles o Brasil, segundo pesquisa do Instituto de Pesquisa Econômica Aplicada – IPEA (2016). Utilizando nosso país como exemplo do crescimento urbano, segundo a Pesquisa Nacional por Amostra de Domicílios - PNAD (2015), 84,72% da população brasileira está domiciliada em regiões urbanas. A porcentagem é ainda maior na região sudeste, a mais urbanizada do país, onde os índices chegam a 93,14%.

Assim, o deslocamento populacional diário somado a popularização do uso dos meios particulares de transporte, também ligado ao inchaço urbano das últimas décadas brasileiras contribuem para os problemas ligados à poluição urbana IPEA (2016). O aumento de veículos circulando nas vias acarreta problemas de diversas espécies, mas especialmente questões relacionadas a poluição têm aumentado nos últimos anos e continuarão aumentando, de acordo com o relatório da Organização Mundial da Saúde (WHO, do inglês *World Health Organization*, 2021) sobre audição e primeiro do gênero na organização.

Também em razão deste crescimento acelerado, problemas como a poluição sonora são pouco considerados no planejamento urbano das cidades. O fato de a poluição sonora não ser cumulativa no ambiente faz com que ela seja, muitas vezes, pouco relevante ou não considerada nociva à saúde humana.

De acordo com Ramírez e Domínguez (2013), existem preocupações e ações de controle da comunidade urbana em relação aos efeitos adversos da intensa movimentação de veículos e pessoas, que se confirmam pelo aumento expressivo de elaboração de leis e regulamentos para controlar os níveis de ruído. Segundo os autores, a poluição sonora gerada pelo trânsito é um dos mais relevantes problemas oriundos da ausência de planejamento urbano e um grande redutor da qualidade de vida da população, por causar a degradação do conforto acústico nas áreas urbanas.

Dados da WHO (2021) destacam que a poluição sonora urbana, resultado do ruído do tráfego e de indústrias, já é o segundo maior poluente por número de pessoas afetadas e perdendo posição apenas para poluentes presentes no ar e na água, além de ser o segundo maior causador de perdas auditivas.

Segundo Barbosa (2015), o ruído pode ser considerado um agressor potencial e pouco reconhecível devido à dificuldade no estabelecimento de correlações concisas entre causas e efeitos. Entretanto, os efeitos adversos mais comuns trazidos pela exposição ao ruído à saúde humana são negativos e resultam em danos crônicos aos sistemas nervoso e endócrino. Em virtude da cronicidade e dos altos níveis, ao longo do tempo, originam problemas de saúde mais graves como doenças cardiovasculares.

A principal justificativa adotada pelos planejadores urbanos para não considerar o estudo da poluição sonora, segundo Ahmed e Pradhan (2019), é que consultorias para esse tema apresentam alto custo e complexidade e exigem profissionais experientes em sistemas avançados de modelagem de som.

O estudo e elaboração de modelos de previsão sonora pode auxiliar os planejadores urbanos a considerar a presença desse poluente e difundir esforços da comunidade acadêmica em prol da elaboração de estudos avançados sobre poluição sonora urbana. Steele (2001) cita mais grupos de profissionais interessados em modelos preditivos sonoros, tais como: engenheiros de tráfego, engenheiros acústicos, testemunhas especializadas que participam de júris nas áreas civis e criminais e especialistas e consultores acústicos que produzem laudos de impacto ambiental.

Nas últimas décadas, alguns modelos de previsão da poluição sonora foram propostos se utilizando de modelos teóricos lineares e, recentemente, não lineares. Segundo Steele (2001) os modelos iniciais, que foram desenvolvidos nas décadas de 1950 e 1960, foram sendo aperfeiçoados com a adição de variáveis independentes e pontos de medição ao longo das vias. Ainda segundo o autor, os estudos também foram modificando suas formas de respostas, que inicialmente estudavam a emissão sonora de apenas um veículo, e, posteriormente, passaram a considerar o tráfego como um sistema sonoro ativo através da medição em um período de tempo do nível equivalente sonoro (L_{eq}).

Um dos relatos mais antigos que se tem na literatura acerca da temática do ruído urbano está no *Handbook of Acoustic Noise Control* (1952), desenvolvido nos Estados Unidos, que trouxe uma abordagem linear em uma equação com variáveis que representavam o volume de tráfego e a distância da fonte sonora até o aparelho de medição. Na década seguinte, Nickson (1965) e Lamure (1965) desenvolveram,

paralelamente, estudos com uma equação linear que incluía uma constante representando o nível individual de emissão sonora da fonte. Posteriormente, Johson *et al.* (1968) formulou uma equação que incluía, pela primeira vez, a velocidade média dos veículos emissores de som. A equação era indicada para composições de tráfego com até 20% de veículos pesados, mas uma variável que indicasse a composição não foi incluída na equação.

Galloway *et al.* (1969) foram os primeiros a separar o fluxo de veículos em categorias. A equação apresentada contém, além das variáveis independentes já utilizadas em modelos antecessores como: *distância da fonte sonora, quantidade de carros por hora, velocidade média dos carros*, a variável T , que representa a porcentagem de veículos pesados (caminhões e ônibus) em relação a quantidade total de veículos.

Em 1976 o Departamento de Transportes Rodoviários dos Estados Unidos publicou o Modelo de Previsão de Ruído do Tráfego FHWA (*Federal Highway Administration*), que considera os veículos como fontes pontuais de emissão sonora se deslocando a velocidades constantes. A partir desse estudo, os modelos começaram a figurar como parte de softwares de previsão do ruído sonoro urbano e então se tornaram metodologias híbridas aliando matemática e computação.

Koyasu (1978) publicou o modelo de previsão do pseudo- L_{50} , proposto pela ASJ (*Acoustic Society of Japan*) em 1975, que separa os veículos em classes e considera fontes sonoras pontuais e distanciadas por espaços físicos iguais em um fluxo livre de carros. O modelo padrão alemão, RLS-90 (*Richtlinien für den Lärmschutz an Straßen - Guidelines for Noise Protection on Streets*), de 1981, inovou ao acrescentar dados de tráfego substitutos para cálculos quando os dados reais não são conhecidos. O Escritório Federal Suíço de Proteção Ambiental (1987), publicou o software STL-86 v1.0 que engloba modelos de previsão de tráfego e os usa como entrada para os modelos de previsão de ruído sonoro urbano. O Departamento de Transporte do Reino Unido (1985), formulou seu método com o Cálculo do Ruído do Tráfego Rodoviário (*Calculation of Road Traffic Noise - CoRTN*) que considera a emissão sonora como uma linha contínua e a velocidade do tráfego constante e é ainda utilizado, com pequenas atualizações, para cálculos da Lei de Compensação de Terras. A empresa francesa 01dBA, em 2002, lançou o software MITHRA. O modelo assume a emissão sonora em formato de linha ao longo da pista e contém vários tipos

de pistas e módulos para cálculos em ferrovias proporcionando uma ampla gama de possibilidades de cálculos em diferentes cenários. A propagação é determinada pelo Princípio da Reciprocidade que define de forma igualitária a propagação das ondas sonoras em qualquer direção e sentido.

Os modelos lineares foram, aos poucos, sendo substituídos por modelos mais complexos e híbridos com tecnologias de aprendizado de máquina. Cammarata, Cavalieri e Fichera (1995) publicaram um estudo com a utilização de redes neurais artificiais na previsão do nível sonoro gerado pelo tráfego urbano. O objetivo era desenvolver modelos de previsão que tivessem uma acurácia maior do que os até então utilizados – modelos de regressão lineares. Estudos similares começaram a surgir e se repetem até a atualidade. Entretanto, as variáveis independentes (que atuam no resultado final dos modelos) tanto dos modelos lineares quanto dos não-lineares sempre foram selecionadas baseando-se nas limitações do estudo, na opinião de especialistas no assunto de medição sonora do tráfego urbano ou baseando-se em estudos anteriores.

Muitas dessas variáveis são comuns à maioria dos estudos: número de carros, motos e veículos pesados e suas proporções no volume total do tráfego, distância da fonte sonora, velocidade do tráfego total ou individual dos veículos, presença de barreiras, coeficientes de absorção do ar, velocidade e direção do vento, coeficientes relacionados ao tipo de asfalto, temperatura ambiente e até mesmo angulação da medição em relação à fonte sonora.

Segundo Ahmed e Pradhan (2019) estudos de ruído do tráfego podem, durante a coleta de dados, colocar os pesquisadores em riscos por se darem muitas vezes em rodovias. O processo também pode ser custoso financeiramente e delongado. Considerando a enorme quantidade de variáveis que têm sido utilizadas nos diversos estudos feitos, mostra-se necessária a construção de uma metodologia que auxilie de maneira formal a seleção das variáveis para estudos futuros tanto na economia de tempo, dinheiro e trabalho quanto na segurança dos pesquisadores.

Com vistas a aprimorar os processos de construção dos modelos de previsão de poluição sonora decorrente do tráfego urbano, neste trabalho foram avaliados os resultados de previsões de L_{eq} de modelos lineares e não lineares. Previsões de L_{eq} foram geradas através de Redes Neurais Artificiais (RNA) e, por fim, a comparação dos resultados gerados pelos modelos neurais foram comparadas com os resultados

dos modelos lineares como forma de avaliar se existe superioridade daqueles modelos em relação a técnicas lineares.

1.1 Objetivos do trabalho

Sabido que a acurácia de modelos de previsão são, na grande maioria dos casos, o fator mais importante na sua escolha, esta pesquisa se ocupa em responder a seguinte pergunta: “Quais arquiteturas dentre as redes neurais e do modelo linear avaliados produz a melhor previsão de L_{eq} oriundo do tráfego urbano?”

No presente trabalho é proposta uma avaliação de modelos de previsão de poluição sonora urbana com a utilização de RNA que são, por fim, comparadas com o modelo previsão linear Regressão Linear Múltipla (RLM).

Os objetivos específicos desse trabalho são:

- ✓ Elencar nos modelos publicados na literatura as variáveis independentes utilizadas em modelos de previsão do ruído sonoro do tráfego urbano e a justificativa que os autores deram para a escolha dessas variáveis;
- ✓ Produzir uma comparação em termos de acurácia dos modelos de RNA frente ao RLM com auxílio de duas bases distintas.

1.2 Trabalhos relacionados

Diversas variáveis são utilizadas pelos autores nos diferentes modelos publicados. Os modelos têm sido feitos com pequenos diferenciais para se adaptarem aos locais de medição para que caracterizem fielmente o ambiente. As variáveis utilizadas nos modelos mais populares propostos podem ser encontradas na tabela 1.

Fortuna, Occhipinti e Xibilia (1994) em um dos primeiros artigos sobre o assunto avaliam a capacidade de previsão de uma rede neuro-fuzzy. Todavia o caminho estudado foi o inverso. O objetivo do estudo era desenvolver um modelo que previsse a quantidade de veículos que passaram pela via a partir das entradas: L_{eq} , largura da via e da altura média dos prédios que cercavam a mesma. Como resultado, os autores ressaltam a grande velocidade com que a rede é capaz de aprender e a simplicidade na aplicação das lógicas *if-then* e *fuzzy*.

Cammarata, Cavalieri e Fichera (1995) propuseram o uso de redes neurais artificiais na previsão do nível sonoro gerado pelo tráfego urbano. Os autores justificam que os modelos lineares até então utilizados não tinham uma boa precisão

dos níveis de pressão sonora gerados pelo tráfego. Desenvolveram, então, um modelo que usa retro propagação e uma arquitetura com dois níveis em cascata. No primeiro nível, uma rede supervisionada (um vetor de quantização de aprendizado - LVQ) filtra os dados e possíveis erros de medição. Já no segundo nível, a rede de retro propagação efetua a previsão do nível de pressão sonora (variável mais representativa do volume sonoro gerado). As variáveis de entrada são: o número de veículos observados no período decomposto em carros, motos e caminhões, a altura média dos prédios ao redor da área de medição e a largura da rua. Os autores pontuaram que a detecção de dados fora dos padrões, utilizada pela rede, feita em um local pode não funcionar em outro devido ao fato de os resultados da rede estarem ligados aos padrões dos dados de treinamento (CAMMARATA, CAVALIERI, FICHERA; 1995)

Já nos anos 2000 Parbat e Nagarnaik (2008) desenvolveram uma rede neural artificial com as variáveis de entrada: tráfego total, composição de tráfego em termos percentuais de carros leves, pesado, de duas rodas e bicicletas e outros, a largura da faixa de rodagem e a distância da fonte sonora. As saídas também são compostas, diferentemente da maioria dos estudos, de várias variáveis: L_{10} , L_{eq} . O objetivo do estudo era testar a viabilidade de um modelo de previsão de ruído do tráfego em uma cidade de médio-porte no interior da Índia. O modelo obteve valor da raiz quadrada do erro médio (RMSE, do inglês *Root Mean Square Error*) que permitiu aos autores julgá-lo como adequado ao problema.

Tabela 1 – Artigos relacionado ao tema da pesquisa

(Continua)

Ano	Autores	Metodologia	Entradas	Saídas	Diferenciais/Resultados
1994	Fortuna, Occhipinti e Xibilia	Rede Neuro-fuzzy	L_{eq} , largura da via e da altura média dos prédios da via.	Total de veículos	Aplicaram a lógica fuzzy.
1995	Cammarata, Cavalieri e Fichera	LVQ filtra erros nos dados e uma RNA prevê a saída. Rede MLP	Total de veículos, altura média dos prédios da via, largura da via. Tráfego classificado, largura da via e a distância da fonte sonora.	L_{eq}	<i>Backpropagation</i> e arquitetura do modelo em 2 níveis
2008	Parbat e Nagarnaik			L_{10} , L_{eq} ,	O modelo propunha ser adaptado a uma cidade indiana de médio porte.

(Continuação)

Ano	Autores	Metodologia	Entradas	Saídas	Diferenciais/Resultados
2010	Givargis e Karimi	Compararam uma MLP com o modelo CoRTN.	Total de veículos por hora, velocidade média do tráfego, percentual de veículos pesados, gradiente de correção e ângulo de visão no local da medição	$L_{eq,1h}$	Os autores não encontraram resultados significativamente melhores quando compararam o modelo ao CoRTN.
2012	Cirianni e Leonardi	Burgess Model, Gercia e Bernal Model e o Revised CoRTN comparados com a MLP	Quantidade de carros leves, médios e pesados, largura da via, velocidade média do tráfego.	L_{eq}	O modelo neural apresentou menor erro médio quando comparado aos modelos lineares.
2014	Nedic <i>et al.</i>	Utilizou uma MLP para comparar os resultados estatisticamente com modelos lineares	Percentual de veículos leves, médios e pesados, velocidade média do tráfego	L_{eq}	A MLP se mostrou melhor previsora do que os modelos tradicionais
2014	Kumar <i>et al.</i>	Utilizaram o algoritmo de <i>Levenberg – Marquardt</i> para o treinamento da rede MLP	Total de carros por hora divididos em sete categorias de acordo com o trânsito indiano, percentual de caminhões ou veículos considerados pesados e velocidade média do tráfego.	L_{eq} e L_{10}	Foram incluídos no estudo os triciclos indianos e sons de buzina como forma de caracterizar o trânsito indiano.
2016	Singh <i>et al.</i>	Compararam o Generalized Linear model, Decision Trees, Randon Forests e Neural Networks	Volume de tráfego, porcentagem de veículos pesados e velocidade média do tráfego.	L_{eq}	Na comparação dos erros médios o modelo Randon Forests demonstrou melhor resultado.

(Conclusão)						
Ano	Autores	Metodologia	Entradas	Saídas	Diferenciais/Resultados	
2016	Tomic <i>et al.</i>	Produziram 2 modelos com <i>soft computing</i> e compararam com dados coletados e com redes MLP.	2 Quantidade de carros leves, médios, pesados, ônibus e motocicletas, velocidade do vento, velocidade máxima da via, ângulo da visão da via, distância do centro da via até a fachada dos prédios laterais, a soma dos ângulos em relação às fachadas e a relação com os ângulos das fachadas opostas.	L_{eq}	Embora tenham usado algoritmos evolucionários na seleção dos parâmetros dos modelos com <i>soft computing</i> . Os modelos neurais demonstraram melhores resultados em termos de acurácia.	
2017	Khalil, Hamad e Shanableh	BSTN e ORNAMENT Foram comparados à MLP	Volume de tráfego, velocidade média do tráfego, temperatura média na via	L_{eq}	A inserção da temperatura da via aumentou a acurácia do modelo dando à MLP resultados melhores do que os modelos lineares comparados.	

Fonte: Autoria própria (2021)

Em estudo feito na cidade de Teerã, no Irã, Givargis e Karimi (2010) elaboraram uma rede neural artificial através de 50 amostras de dados coletados em cinco vias da cidade. As entradas são: quantidade total de veículos por hora, velocidade média do tráfego por hora, percentual de veículos pesados, gradiente de correção e ângulo de visão no local da medição (para se observar possíveis reverberações e distorções na captação do som). A saída da rede foi o $L_{eq,1h}$ (nível equivalente sonoro pelo período de 1 hora). Do montante de dados 60% foi utilizado para o treinamento da rede. O restante foi dividido entre validação e teste. Embora os autores tenham encontrado coerência no modelo utilizado, não observaram significativa diferença quando os resultados foram comparados a abordagem CORTN. Os autores ressaltaram a necessidade de amostras maiores para se avaliar as vantagens das redes neurais frente ao modelo comparado.

Cirianni e Leonardi (2012) propuseram uma abordagem metodológica para a análise quantitativa da poluição sonora do tráfego. Três modelos lineares (Burgess Model, Gercia e Bernal Model e o CoRTN Revisado) foram testados e comparados com o modelo de rede neural. As variáveis de entrada são a quantidade de carros estratificada, a largura da pista e a velocidade média do tráfego. A saída dos modelos

foi o L_{eq} sonoro. O Modelo Geral de Regressão de Rede Neural se mostrou mais adequado ao problema por apresentar menor erro médio nos dados de saída, que constituem a capacidade de previsão da rede. Os autores sugerem a inclusão de mais variáveis para estudos futuros, como: a rugosidade da vida e sua capacidade de reflexão sonora.

Nedic *et al.* (2014) propuseram um modelo no qual as entradas eram compostas por cinco variáveis (quantidade de carros pequenos, caminhões leves, caminhões pesados, de ônibus e pela velocidade média horária do tráfego). A saída de dados foi o nível equivalente sonoro (L_{eq}). Em comparação com os métodos estatísticos tradicionais a rede neural artificial demonstrou ser mais capaz de prever o nível sonoro L_{eq} . Embora os autores tenham ressaltado que as características locais do trânsito são importantes fatores no desenvolvimento do modelo, o desenvolvido por eles supera, segundo os autores, o modelo pioneiro de Cammarata, Cavalieri e Fichera (1995) por esse não ser sensível ao layout do local de medição.

Kumar *et al.* (2014) propuseram um modelo de rede neural artificial do tipo MLP utilizando o algoritmo de *Levenberg – Marquardt* para treinamento para identificação do nível sonoro no tráfego indiano. As entradas da rede são as variáveis convencionais: quantidade de carros observados por hora e divididos em sete categorias de acordo com o trânsito indiano, percentual de caminhões ou veículos considerados pesados e velocidades medidas no período de tempo estudado. Porém, nesse modelo, os autores obtiveram duas saídas de dados: uma delas com o nível sonoro equivalente (L_{eq}) e outra com o nível sonoro que se observou por mais de 10% do tempo de medição (L_{10}). Os resultados mostraram um desempenho melhor do que a observada em modelos de regressão linear puramente estatísticos. Entretanto, a quantidade de dados utilizados no estudo é considerada baixa pelos autores que sugeriram amostras maiores para estudos futuros.

Uma abordagem inovadora foi trazida Singh *et al.* (2016). Os autores utilizaram o *Generalized Linear model*, *Decision Trees*, *Randon Forests* e a *Neural Networks* para prever o L_{eq} em diferentes localidades de uma cidade na Índia. As entradas dos quatro modelos são as mesmas e as comumente utilizadas: volume de tráfego, porcentagem de veículos pesados e velocidade média do tráfego. A performance dos quatro modelos foi comparada pelo MSE, pelo coeficiente de determinação (R^2) e pela acurácia do resultado. Nesse estudo, o modelo *Randon*

Forests obteve melhor desempenho na previsão do ruído de tráfego ao ser comparado com os demais.

Tomic *et al.* (2016) que usaram dois modelos matemáticos de previsão de ruído sonoro do tráfego, produzidos com o auxílio de técnicas de *soft computing*, e os compararam com dados coletados nas ruas e com modelos lineares comumente usados como: *German RLS-90*, *Nordic Prediction Method* e o *Italian Model*. Entretanto, os modelos não apresentaram resultados coerentes. Os autores, então, aumentaram o tamanho da amostra incluindo novas medições e desenvolveram dois novos modelos de predição, esses com a utilização das redes neurais que se mostraram, para o tipo de problema, capazes de prever melhor o ruído do tráfego urbano. Porém, uma ressalva da necessidade de capacidade computacional elevada e de alta quantidade de dados de entrada que os modelos neurais exigem faz com que eles não sejam indicados para problemas de trânsito menos complexos ou que exijam menor acurácia.

Outra inovação surgiu com Khalil, Hamad e Shanableh (2017) ao desenvolverem um modelo para locais de climas quentes. Os autores incluíram como variável de entrada na rede o volume de tráfego, a velocidade média do tráfego, a temperatura média na via e como saída o nível equivalente sonoro L_{eq} . Na comparação de resultados dois modelos foram utilizados: *The Basic Statistical Traffic Noise Model* (BSTN) e *Ontario Ministry of Transport Road Traffic Noise Model* (ORNAMENT). A inserção da variável temperatura trouxe maior acurácia ao modelo. Entretanto, segundo os autores, há muito ainda o que ser melhorado nesse tipo de estudo. A variável volume de tráfego foi a que menos influenciou no resultado do modelo. Como sugestão para artigos futuros, os autores aconselham a inserção de variáveis como aceleração e desaceleração e tipo de pavimento.

1.3 Justificativa

As últimas décadas têm sido marcadas pelo rápido aumento do número de veículos que transitam todos os dias pelas vias urbanas. A grande diversidade de tipos de veículos e o envelhecimento da frota contribuem na geração desordenada de ruído que se propaga livremente por áreas comerciais, escolares e residenciais em todos os períodos do dia.

Muitos estudos têm sido publicados, recentemente, sobre os danos causados

à saúde humana pela exposição excessiva ao ruído do tráfego. Chang *et al.* (2011), Bodin *et al.* (2009), Barregard, L., Bonde, E., Ohrström, E. (2009) estudaram a relação com a hipertensão arterial; Jakovljevic *et al.* (2006) e Pirrera S, De Valck E., Cluydts R. (2010) os distúrbios do sono e Babisch *et al.* (2005), Selander *et al.* (2009) e Babisch *et al.* (2013) as alterações cardiovasculares.

Muitos países, então, têm buscado, ainda que minimamente, planos de redução de danos causados pela elevada exposição humana ao ruído urbano. A literatura científica tem apresentado alguns modelos de previsão do volume sonoro gerado pelo tráfego urbano ao longo dos anos e o desenvolvimento computacional permitiu, além do desenvolvimento de softwares que auxiliam no mapeamento da propagação do ruído, a incorporação de ferramentas de aprendizado de máquina que permitem prever, a partir do comportamento observado, gerações de ruído urbano e determinadas situações futuras hipotéticas.

A maioria dos trabalhos citados nesta seção têm como limitações declaradas a necessidade de estudos mais aprofundados na correlação das variáveis e com amostras maiores. É fundamental ressaltar que os dados do tráfego são repletos de incertezas pois estão associados, muitas vezes, às condições climáticas e não somente à presença ou ausência de fontes sonoras, mas também às suas diferentes intensidades. Para Garg (2014) uma abordagem metodológica mais simplificada e em consonância com uma simples e adequada coleta e que leve em consideração as imprecisões inerentes dos modelos será mais oportuna para os planejadores urbanos, já que a medição de variáveis é sempre trabalhosa.

Givargis e Karimi (2010) pontuam que a inserção do uso das redes neurais nos modelos de previsão do ruído urbano permite a inclusão de uma quantidade muito maior de variáveis do que os modelos tradicionais (autorregressivos) pois as redes neurais têm a capacidade de modelar complexas funções multivariáveis.

A literatura do assunto é carente de modelos que tenham sido testados sobre mais de uma base de dados e, diante desse impasse, a proposição de uma avaliação que permita comparar modelos de previsão de ruído sonoro urbano com redes neurais artificiais e seja efetuado sobre duas bases distintas preenche uma lacuna da literatura e auxilia, em grande medida, a agilidade nos estudos da poluição sonora urbana.

2 ESTATÍSTICA MULTIVARIADA

As análises de regressão têm por objetivo investigar o comportamento de uma variável dependente em relação à outras variáveis envolvidas e consideradas independentes. Modelos matemáticos são utilizados para correlacionar as variáveis.

Para Montgomery (1991) a análise multivariada tem a sua importância centrada na resolução de problemas em que nos deparamos com múltiplas variáveis, as quais podem ter seus efeitos interdependentes, parcialmente independentes ou fortemente independentes a depender dos coeficientes calculados para correlação dos parâmetros.

Especialmente em problemas multidimensionais, com grandes quantidades de dados e variáveis, os métodos de análise multivariada têm sua eficácia comprovada porque permitem o confronto entre duas ou mais variáveis frente aos resultados dependentes e a aplicação dos métodos pode levar à extração de tendências nos dados permitindo melhores análises (CORRAR *et al.*, 2007)

Johnson e Wichern (1992) afirmam que a análise multivariada pode explorar um grande campo de finalidades como redução de dados, simplificação estrutural de problemas, classificação e agrupamento de dados, elaboração de hipóteses e investigação de interdependência de variáveis.

2.1 Regressão Linear Múltipla

Conforme Dantas (2003) o uso da técnica estatística de Regressão Linear Múltipla (RLM) é bastante usado na comunidade científica e na Engenharia de Avaliações, que trata dos múltiplos fatores que influenciam no preço de um bem, por exemplo.

Hair *et al.* (2005) afirma que a RLM é o método estatístico adequado quando o principal objetivo do avaliador é prever alterações na variável dependente de acordo com alterações das variáveis independentes, ou seja, o impacto das variáveis regressoras na variável resposta.

O modelo genérico da Regressão Múltipla é dado pela equação (1).

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad (1)$$

em que Y_i é o termo que representa a variável dependente ou explicada, com índice i variando dentro de um conjunto $i = \{1, 2, \dots, n\}$ com n amostras, β_0 é o valor de

intercepto, ou seja, um termo independente de variável, β_k o coeficiente que representa a inclinação da reta de Y_i em relação a variável X_{ki} (mantendo constantes as demais variáveis do conjunto) e ε_i o erro aleatório em Y_i sempre presente, mesmo que pequeno, porque corresponde a diferença entre os valor real e o predito.

Para estimar os coeficientes de regressão na RLM pode ser utilizado o método dos quadrados mínimos (MONTGOMERY, 1991). A estimação pode ser dada pela equação (2).

$$\hat{\beta} = (X^t X)^{-1} (X^t Y) \quad (2)$$

Em que $\hat{\beta}$ representa a matriz que contém os regressores, X^t a matriz transposta das variáveis, X a matriz que contém o valor das variáveis e Y a matriz com os valores resposta.

A matriz $X^t X$ é geralmente não singular, de modo que os métodos descritos sobre determinantes e métodos para inverter essas matrizes podem ser usados para encontrar o inverso do produto da matriz transposta de X pela própria matriz X (MONTGOMERY, 1991).

3 REDE NEURAIS ARTIFICIAIS

As Redes Neurais artificiais tiveram seus estudos iniciados a partir da ideia de se fazer uma analogia dos cálculos matemáticos com os neurônios biológicos e, assim, criar uma espécie de neurônio artificial. Essa primeira analogia foi publicada por McCulloch e Pitts (1943) na obra *A Logical Calculus of the Ideas Immanent in Nervous Activity* (1943) na qual foi proposta a primeira relação lógico-matemática para o comportamento de um neurônio humano. Entretanto, a modelagem da relação foi feita com módulos de processamento binários.

Rosenblatt (1958) propôs uma abordagem, até aquele momento inédita para problemas de reconhecimento de padrões, chamada *perceptron*, e a partir da ideia inicial de que o cérebro humano trabalha como um associador de padrões ao invés de um circuito lógico.

Para Haykin (2001), desde o começo, a motivação na comparação às células nervosas animais estava no reconhecimento do cérebro humano como um computador complexo, não-linear, paralelo e incomparavelmente rápido, que é uma prova viva da extraordinária capacidade do processamento paralelo tolerante às falhas. A comparação aos mecanismos nervosos se deu porque os neurônios são células presentes por todo o corpo humano, especialmente no sistema nervoso, onde estão em maior quantidade, e têm papel elementar por conduzirem impulsos (sob a forma de estímulos elétricos resultantes de reações físico-químicas), respeitando, sempre, determinadas condições de operação (SILVA; SPATTI, FLAUZINO, 2010). Ainda segundo os autores, o neurônio biológico é dividido teoricamente em três partes, a saber:

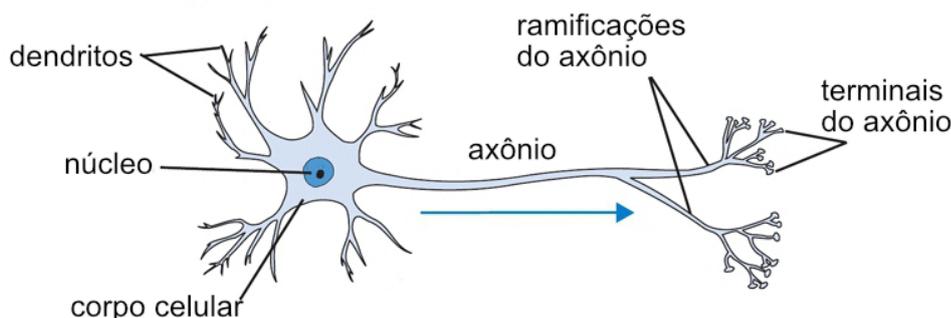
- **Dendritos:** formados por longos filamentos delgados, constituindo uma árvore dendrital, têm a função de captação contínua dos estímulos oriundos de diversos outros neurônios, quando do tipo conector, ou das adjacências onde estão inseridos, quando do tipo sensitivo.
- **Corpo celular ou Soma:** é o segmento onde ocorre o processamento dos estímulos recebidos pelos dendritos. É no Soma que ocorre a decisão, a partir de um potencial elétrico de ativação pré-definido, por excitação ou inibição, da resposta a ser gerada como forma de impulso elétrico e que será propagada através do segmento seguinte chamado de axônio.
- **Axônio:** é o segmento da célula nervosa responsável pela propagação do

impulso elétrico para os neurônios seguintes da rede, ou, no caso dos neurônios efetadores, diretamente para o tecido muscular. A extremidade do axônio, se localizam nas sinapses, ao contrário de seu comprimento monofilamentar essa extremidade é ramificada e desconexa das células seguintes.

Lent (2010) define o neurônio como a unidade morfofuncional fundamental do sistema nervoso. O autor afirma que a célula nervosa produz e veicula pequenos sinais elétricos que podem ser considerados verdadeiros bits de informações e são capazes de codificar tudo o que percebemos a nossa volta e o que acontece no interior do nosso organismo, os comandos que damos, sejam direta ou indiretamente, aos neurônios efetadores (presentes nos músculos e nas glândulas) e tudo o que se realiza a partir de nossa atividade mental.

Relvas (2010) ensina que os pontos de contatos sinápticos, chamados de fendas sinápticas, são regiões altamente especializadas na transmissão de informação de uma célula à outra, informações, essas, oriundas dos dendritos da célula nervosa vizinha. Para o autor, os dendritos são uma espécie de ramificação que tanto recebe quanto libera sinais elétricos e um neurônio pode ter milhares de dendritos.

Figura 1 – O neurônio biológico



Fonte: Github (2021)¹.

Shepherd e Koch (1990), na sua obra sobre a organização sináptica do cérebro, definem as sinapses como unidades mediadoras entre dois neurônios e que convertem os sinais elétricos pré-sinápticos em estímulos químicos nos neurotransmissores e, novamente, os converte em estímulos elétricos no processo

¹ Disponível em: < <http://cs231n.github.io/neural-networks-1> > Acesso em: 12 mar. 2021.

pós-sináptico. Isso acontece porque a fenda sináptica não contempla conexão física entre um axônio e o dendrito do neurônio seguinte.

Essas células, de acordo com Haykin (2001), são entre cinco e seis grandezas menos velozes que as portas lógicas em circuitos de silício. Porém, a alta taxa de aproveitamento do cérebro é obtida pelo número de neurônios presentes na massa cerebral, que, de acordo com Lent (2010) é de cerca de 85 bilhões de unidades e, também, pelo paralelismo extremo de suas conexões.

Lent (2010) defendem, ainda, que o neurônio trabalha em um sistema de cooperação celular, e não em um sistema isolado. A cada nova experiência a que o ser humano é submetido, o seu cérebro se modifica. Para os autores, a principal causa de aumento das sinapses nervosas entre os neurônios é a interação humana com o meio e essas experiências provocam mudanças constantes no cérebro.

Pia (1985) usa o termo plasticidade sináptica para nomear a capacidade do cérebro adaptar-se por meio dos processos de criação de novas conexões ou por meio de modificação nas conexões preexistentes e essa é a principal característica deste tipo celular porque permite a ele o aprendizado.

Relvas (2009) afirma que um neurônio pode receber ou enviar sinais para um número entre 1.000 a 100.000 conexões sinápticas de outros neurônios adjacentes, dependendo do seu tipo e localização no sistema nervoso humano. A quantidade e a qualidade dessas sinapses podem variar, entre outros fatores, pela experiência vivenciada e pela intensidade da aprendizagem, demonstrando a enorme capacidade plástica do Sistema Nervoso.

Para Lucas (2012) o processo de aprendizado pode ser exemplificado fisicamente pelas diferentes respostas que um indivíduo pode apresentar diante de uma mesma situação. O processo é quase sempre carecido de ensinamento. A figura de um professor é necessária em um primeiro momento para demonstrar como uma tarefa deve ser executada. O processo, então, é seguido por tentativas e erros até que a execução aconteça sem problemas.

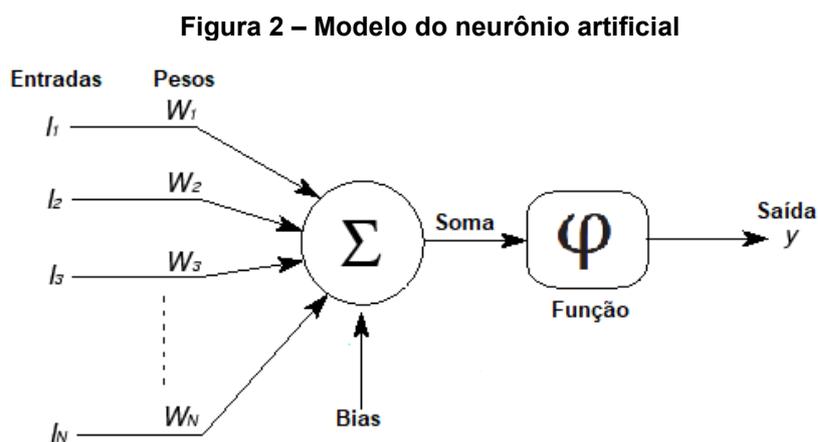
3.1 O neurônio artificial

Os neurônios artificiais são elementos computacionais ou unidades processadoras de arquitetura semelhantes a uma simplificação dos neurônios biológicos. De acordo com Silva, Spatti e Flauzino (2010) a inspiração veio a partir da

análise da geração, recebimento e envio de pulsos elétricos pela membrana celular dos neurônios oriundos do trabalho de Hodgkin e Huxley (1952)

Esses elementos computacionais implementados por códigos de programação realizam funções simples, como receber os sinais em suas entradas, processá-los de acordo com a função operacional usada naquele elemento e apresentarem uma resposta para o processamento.

Braga, Carvalho e Ludermir (2007) tratam o primeiro neurônio artificial concebido por McCulloch & Pitts (1943) como uma simplificação do conhecimento científico da época acerca dos neurônios biológicos. Esse modelo inicial serviu de base para o desenvolvimento das arquiteturas de redes que hoje são utilizadas. A figura (2) mostra um esquema representativo de um neurônio artificial.



Fonte: Adaptado de Haykin (2001)

As entradas do neurônio estão representadas na figura (2) pelo conjunto de elementos $I = \{I_1, I_2, \dots, I_n\}$. As entradas são os sinais que o neurônio recebe do meio externo. Em uma analogia aos neurônios biológicos da retina humana, por exemplo, as entradas seriam os estímulos visuais luminosos que esta receberia. Silva, Spatti e Flauzino (2010) defendem que haja uma normalização dos valores do conjunto de entrada visando incrementar a eficiência computacional dos algoritmos de aprendizagem que conterão esses neurônios.

As substâncias neurotransmissoras, que preenchem as fendas sinápticas nos neurônios biológicos e são responsáveis pelo avanço do sinal elétrico ajustando o caminho correto do impulso até seu destino, podem ser representadas pelo conjunto $W = \{w_1, w_2, \dots, w_n\}$ de pesos das conexões, do neurônio artificial representado na

figura (2) (SILVA; SPATTI; FLAUZINO, 2010).

As entradas $I_i \in I(I_n)$ serão ponderadas por seus pesos sinápticos associados presentes do conjunto W . Como é possível ver na figura (2), o valor da entrada I_i é associado ao peso w_1 .

Após serem ponderados os sinais passam pelo Combinador Linear, representado na figura (2) pela letra grega sigma maiúscula (Σ) e também chamado de Junção Somadora. Na versão biológica do neurônio, esse passo do processamento do sinal ocorreria no Soma ou Corpo celular. A função do combinador linear é justamente agregar todos os sinais ponderados pelos pesos sinápticos e oferecer um valor de saída.

O neurônio da figura (2) possui um valor de bias associado. Esse valor também é chamado de limiar de ativação e é somado ao valor obtido como resposta no Combinador Linear. Bellotti (2019) afirma que o valor mais comum associado ao termo é 1 e que ele pode também ser interpretado como mais um peso do neurônio.

O valor oriundo da soma do combinador linear e do bias será argumento da função de ativação $\varphi(\cdot)$ presente na figura (2) que, por fim, retornará a saída y . A equação (3) pode representar o neurônio artificial da figura (2) matematicamente.

$$y = \varphi \left(\left[\sum_{i=1}^n I_i w_i \right] + bias \right) \quad (3)$$

3.1.1 Tipos de Funções de Ativação

As funções de ativação dos neurônios artificiais podem ser de diferentes tipos e apropriadas de acordo com o domínio de definição das mesmas. Silva, Spatti e Flauzino (2010) as dividem em dois grupos principais: funções parcialmente diferenciáveis e funções totalmente diferenciáveis. Essa será também a divisão utilizadas nesse trabalho.

a) Funções de ativação parcialmente diferenciáveis:

Funções parcialmente diferenciáveis são aquelas em que a derivada de primeira ordem inexiste em alguns de seus pontos. Silva, Spatti e Flauzino (2010) elegem as três principais como:

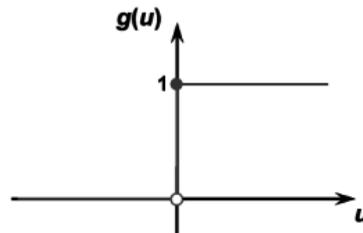
- i. Função degrau (*heavyside/hardlimiter*): sua imagem será dada por valores unitários e positivos, se o valor do potencial de ativação (u) for maior ou igual a zero e nulo no caso contrário; assim temos a notação (1) que

representa a função.

$$g(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ 0, & \text{se } u < 0 \end{cases} \quad (1)$$

A Função degrau também pode ser representada graficamente pela figura (3).

Figura 3: Gráfico da Função degrau $g(u)$.



Fonte: Silva, Spatti e Flauzino (2010)

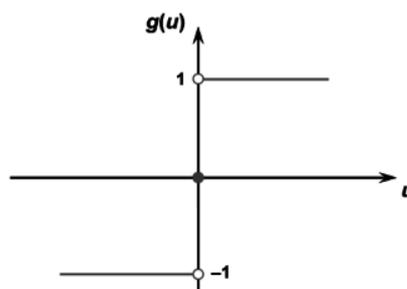
Haykin (2001) ressalta que, na literatura de engenharia, a Função degrau é um tipo de função de limiar.

- ii. Função degrau bipolar ou função sinal (*symmetric hard limiter*): neste tipo de função, o resultado da aplicação será positivo se o potencial de ativação (u) for maior que zero; o resultado será nulo caso o potencial seja também nulo; e assumirá valores unitários negativos quando o potencial de ativação (u) for também menor que zero. Matematicamente temos a notação (2) que representa a Função degrau bipolar.

$$g(u) = \begin{cases} 1, & \text{se } u > 0 \\ 0, & \text{se } u = 0 \\ -1, & \text{se } u < 0 \end{cases} \quad (2)$$

A figura (4) que ilustra a representação gráfica desta função.

Figura 4 – Gráfico da função degrau bipolar $g(u)$.



Fonte: Silva, Spatti e Flauzino (2010)

Silva, Spatti e Flauzino (2010) alertam que a função degrau bipolar, quando usada para classificação de padrões (uma das aplicações das RNA), pode ser aproximada para a notação (3). Para Haykin (2001) essa notação também pode ser usada caso essas sejam as respostas que melhor se encaixem nos padrões desejados de extensão de saída do neurônio, sendo a Função degrau bipolar também um tipo de função de limiar.

$$g(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ -1, & \text{se } u < 0 \end{cases} \quad (3)$$

Em casos em que seja necessário manter o valor do potencial de ativação (u) inalterado e usá-lo como saída do neurônio (se não houve ativação), pode-se usar a função na forma da notação (3).

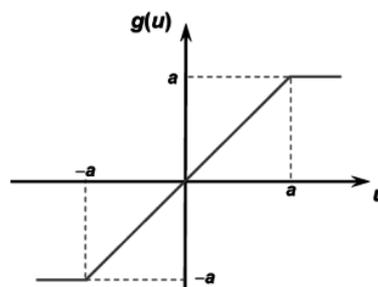
$$g(u) = \begin{cases} 1, & \text{se } u > 0 \\ \text{saída inalterada}, & \text{se } u = 0 \\ -1, & \text{se } u < 0 \end{cases} \quad (3)$$

- iii. Função rampa simétrica: nesse tipo de função, quando usado como função de ativação, os valores de saída do neurônio serão os próprios valores dos potenciais de ativação dentro do intervalo $[-a, a]$. A notação matemática (3) representa a função rampa simétrica.

$$g(u) = \begin{cases} a, & \text{se } u > a \\ u, & \text{se } -a \leq u \leq a \\ -a, & \text{se } u < -a \end{cases} \quad (5)$$

A ilustração gráfica da Função rampa simétrica é dada pela figura (5)

Figura 5 – Gráfico da função rampa simétrica.



Fonte: Silva, Spatti e Flauzino (2010)

- b) Funções de ativação totalmente diferenciáveis

Funções totalmente diferenciáveis são aquelas que têm todas as

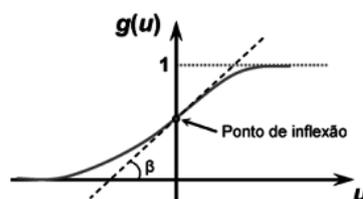
derivadas de primeira ordem conhecidas em todos os pontos de seu domínio de definição. Silva, Spatti e Flauzino (2010) elencam as quatro principais funções desse grupo e que podem ser usadas nas RNA.

- i. Função logística: neste tipo de função, o resultado apresentado na saída do neurônio que a usa como função de ativação assume valores reais dentro de um intervalo contínuo que vai de zero a um. A equação matemática (4) representa a função logística.

$$g(u) = \frac{1}{1 + e^{-\beta \cdot u}} \quad (4)$$

Na equação genérica (17) e representa o número de Euler, u é o potencial da ativação da função de ativação do neurônio e β é a representação de uma constante real ligada ao nível de inclinação do gráfico da Função logística, quando comparada ao seu ponto de inflexão, como pode ser visto na figura (6) que ilustra o gráfico da função.

Figura 6 – Gráfico da função logística



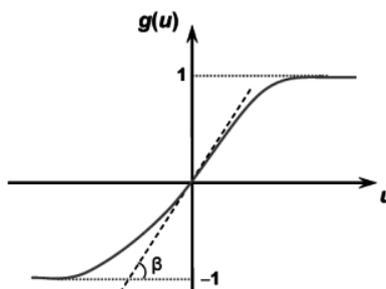
Fonte: Silva, Spatti e Flauzino (2010)

De acordo com Haykin (2001) a Função logística é uma das mais utilizadas em RNA; o autor chama a constante β de “parâmetro de inclinação da função sigmoide” (grupo de funções com gráfico em formato de “s” do qual a Função logística faz parte) e que ao ser variado produz um gráfico sigmoide com diferentes inclinações. As sigmoides são as funções de ativação mais utilizadas em RNA. O autor pontua que essas funções apresentam comportamento estritamente crescente e exibem um bom balanceamento linear e não-linear. (HAYKIN, 2001).

- Função tangente hiperbólica: nessa função o resultado de saída sempre assumirá valores reais dentro do intervalo contínuo $[-1, 1]$. Assim como na Função logística, o parâmetro β se associa ao nível de inclinação da função

tangente hiperbólica em relação ao seu ponto de inflexão, como se vê na figura (7).

Figura 7 – Gráfico da função tangente hiperbólica



Fonte: Silva, Spatti e Flauzino (2010)

A Função tangente hiperbólica tem equação matemática dada por (5).

$$g(u) = \frac{1 - e^{-\beta \cdot u}}{1 + e^{-\beta \cdot u}} \quad (5)$$

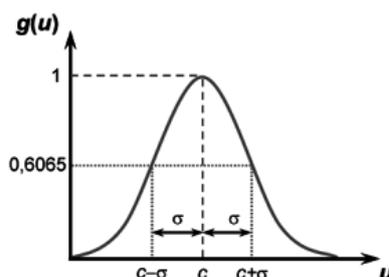
Silva, Spatti e Flauzino (2010) discorrem que tanto na Função logística, quanto na Função tangente hiperbólica (ambas sigmoideais) o incremento do parâmetro β leva a um aumento da inclinação da curva da função no gráfico. Haykin (2001) ressalta que, no limite, quando β se aproxima do infinito, a função sigmoide transforma-se em uma função de limiar; para o autor, a diferença das duas funções consiste na diferenciabilidade de cada uma, pois a função de limiar assume valores de 0 ou 1, enquanto que a função sigmoide pode assumir valores reais dentro de um intervalo contínuo de valores $[-1, 1]$.

- Função gaussiana: nesse tipo de função, quando usada como função de ativação, a saída do neurônio será composta por valores iguais aos valores do potencial de ativação (u) que se encontrarem a uma mesma distância do centro do conjunto (média dos valores), produzindo um gráfico com curva simétrica em relação ao centro. Matematicamente, a função gaussiana é dada por (6).

$$g(u) = e^{-\frac{(u-c)^2}{2\sigma^2}} \quad (6)$$

Na equação (19) c é um parâmetro cujo valor define o centro da função gaussiana (a média dos valores da imagem da função) e σ denota o desvio padrão da média da imagem da função. A figura (8) ilustra o gráfico de uma função gaussiana.

Figura 8 – Gráfico da função gaussiana



Fonte: Silva, Spatti e Flauzino (2010)

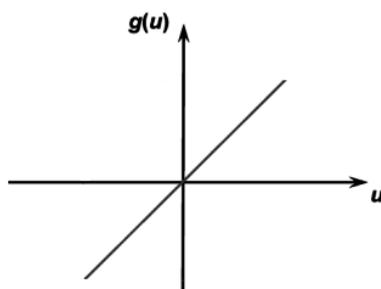
O desvio padrão da média é uma medida de dispersão dos valores e, portanto, de acordo com Silva, Spatti e Flauzino (2010) está diretamente associado com o ponto de inflexão da função gaussiana.

- Função linear (identidade): a aplicabilidade desse tipo de função de ativação em neurônios artificiais está nas RNA como aproximadoras universais de funções com o objetivo de mapear as relações entre entradas e saídas da rede (SILVA; SPATTI; FLAUZINO, 2010). Essa função tem seus resultados de saída iguais aos valores de potencial de ativação (u), que são argumentos da função. A notação matemática (7) a representa.

$$g(u) = u \quad (7)$$

A figura (9) representa graficamente a Função Linear.

Figura 9 – Gráfico da função linear



Fonte: Silva, Spatti e Flauzino (2010)

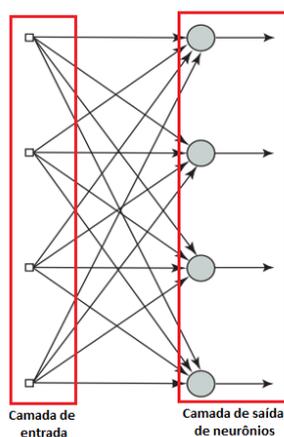
3.2 Arquitetura de redes neurais

As RNA não são simples junções de neurônios artificiais, elas requerem um desenho correto para que o sinal propagado tenha um resultado coerente no ponto de saída e, assim, respondam às necessidades pelas quais a rede foi proposta. Esse desenho é conhecido como arquitetura da rede. Essa organização de uma rede neural artificial é que define a forma como seus neurônios estão posicionados, uns em relação aos outros dentro da rede, sendo a essência desse arranjo o direcionamento das conexões sinápticas dos neurônios. (SILVA; SPATTI; FLAUZINO, 2010)

Outro termo utilizado para a forma como as redes neurais estão organizadas, mas que se refere especificamente às composições estruturais das redes, e não à sua arquitetura, é topologia. Silva, Spatti e Flauzino (2010) exemplificam a topologia de uma rede mostrando que em uma mesma rede pode haver mais de uma topologia, sendo uma composta de 10 neurônios e outra de 20 neurônios e, ainda assim, todos pertencerem à mesma arquitetura e o termo “topologia” pode ainda estar relacionado a presença ou ausência do bias ou dos tipos de função de ativação assumidas pelos neurônios da rede.

Apenas duas camadas são obrigatórias para se ter uma RNA: a de entrada e a de saída. Portanto, apenas uma camada neuronal é suficiente. Haykin (2001) afirma que nessa arquitetura mais simples de rede, a camada de entrada, que é formada por nós (pontos de entrada) de fonte, projeta-se sobre a camada de saída, formada por nós computacionais (neurônios), mas nunca de forma vice-versa. A camada neuronal será então, ao mesmo tempo, de entrada e de saída. Esse tipo de arquitetura pode ser visto na figura (10).

Figura 10 – Rede acíclica com 1 camada neuronal



Fonte: Adaptado de Haykin (2001)

É importante ressaltar que uma rede neural pode ter tantas camadas intermediárias quantas forem necessárias, ou mesmo não conte nenhuma como no exemplo da figura (10), e essas não precisam ser iguais entre si em número de neurônios ou funções de ativação. Anochi (2015) menciona que a topologia das redes, ou seja, a quantidade de camadas e neurônios é determinada principalmente pelo tipo de problema a ser tratado. Todavia, para Haykin (2008) a divisão básica de uma RNA para fins de estudo pode ser composta de 3 camadas nomeadas da seguinte forma:

- **Camada de entrada:** é a responsável pelo recebimento do tipo de informação que os neurônios irão processar, geralmente amostras de dados, portanto, nessa camada, sinais advindos do meio externo. A camada de entrada geralmente não possui neurônios.
- **Camadas intermediárias ou ocultas:** como mencionado, elas podem existir em qualquer quantidade e até mesmo não existir. Essas camadas possuem neurônios, e se existirem não apresentarão a saída da rede, apenas sinais para camada seguinte. Haykin (2008) afirma que elas são responsáveis não apenas pela maior parte do processamento interno das RNA, quanto pela inserção da não linearidade aos dados que estão sendo processados.
- **Camada de saída:** essa será sempre a última camada de uma RNA e terá tantos neurônios quantas forem as saídas da rede, justamente por ser a camada que processa os sinais oriundos das camadas anteriores, transforma-os e os apresenta como resultados finais da rede neural.

3.3 Fase de treinamento da rede

O treinamento de uma rede neural pode ser definido como a execução de um algoritmo, que são passos ordenados em função de um objetivo. No caso das redes neurais esse objetivo é o seu principal atrante, o aprendizado. Silva, Spatti e Flauzino (2010) afirmam que em função do conjunto de passos desse algoritmo de treinamento, ele pode ser chamado de algoritmo de aprendizagem e que o objetivo dessa aprendizagem é alcançar a generalização de soluções produzidas pela rede.

Na fase de treinamento o conjunto de amostras dos dados a ser utilizado é chamado de conjunto de treinamento e têm função essencialmente ligada aos ajustes dos pesos sinápticos. O as amostras restantes do conjunto total ainda

serão divididas em conjunto de validação e teste.

O objetivo dessa separação é evitar que a rede, ao ser testada, processo posterior aos ajustes de pesos, seja alimentado com dados que ela já teve acesso, ou seja, os conjuntos de teste e treinamento não podem ser o mesmo (BELOTTI, 2019).

3.3.1 Treinamento supervisionado

Haykin (2008) define o treinamento supervisionado com uma analogia a um professor que corrige a resposta da rede na direção para uma saída com o menor erro possível, para atingir esse resultado é necessário conhecer a resposta desejada, no caso da rede a saída esperada.

Esse processo ocorre por meio de cálculos da diferença entre a saída da rede em relação a saída esperada e eles são feitos a cada entrada inserida na rede. Esse ajuste que proporciona o treinamento é feito baseado na defasagem entre essas saídas e é considerado completo quando atinge um nível aceitável de acordo com a aplicação da rede (Haykin 2008).

Todavia, pode-se pensar que aumentando-se a quantidade de camadas ou neurônios de uma rede sempre se atingirá um resultado ideal porque ter-se-á uma capacidade muito alta de processamento dos dados. Essa ideia é equivocada porque a rede poderá ser exaustivamente treinada. Haykin (2008) afirma que ao se inserir dados inéditos em uma rede exaustivamente treinada, a resposta da rede tenderá a ser muito ruim e terá ocorrido um sobre-treinamento da mesma, que perdeu seu poder de generalização.

Esse excessivo ajuste da rede também é conhecido como overfitting e para se evitar esse comportamento indesejado existem técnicas como a validação cruzada.

3.3.2 Validação Cruzada

A busca pela máxima capacidade de generalização de uma RNA pode ser a busca por uma condição que garanta que a rede apresentará a melhor resposta para entradas desconhecidas e, nesse contexto, é necessário controlar adequadamente o grau de flexibilidade do ajuste dos pesos (SIQUEIRA, 2013). Esse controle é justamente para evitar problemas como o overfitting.

Haykin (2008) define a validação cruzada como uma ferramenta capaz maximizar a capacidade de generalização das redes, por meio da indicação do melhor

ponto de parada de ajuste dos pesos, ou número de iterações ideal.

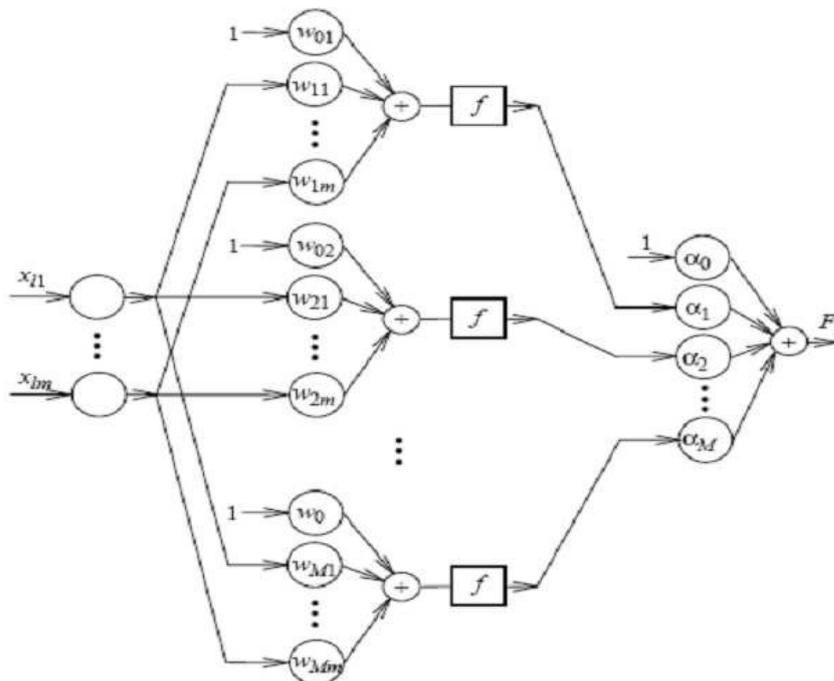
O método consiste em, ao final de época de treinamento, submeter a rede a um conjunto de dados desconhecidos por ela. Esse conjunto é oriundo de uma repartição do conjunto de teste em 2 subconjuntos: um subconjunto de validação com dados que serão apresentados à rede durante o treinamento e outro subconjunto de teste que será utilizado para medir o desempenho da rede após o processo de treinamento (BELOTTI, 2019).

Por fim, os valores escolhidos de pesos sinápticos serão aqueles que promoverem o melhor desempenho da rede, ou seja, o menor erro nas saídas, com a utilização do subconjunto de validação e não com o subconjunto de treinamento porque esses dados já foram utilizados nos ajustes dos pesos e são conhecidos pela rede (HAYKIN, 2008).

3.4 As redes de múltiplas camadas

As redes MLP (*Multilayer Perceptron*) são assim chamadas porque, segundo Silva, Spatti e Flauzino (2010) possuem uma ou mais camadas intermediárias de neurônios, também chamadas de camadas escondidas. As redes MLP são uma extensão do Perceptron de Rosenblat (1958), considerada a forma mais simples de uma rede neural por ser composta apenas pelas camadas de entrada e de saída, ou seja, apenas uma camada neuronal. Um esquema genérico representativo pode ser visualizado na figura (11).

Figura 11 – Esquema representativo de uma MLP



Fonte: Von Zuben e Attux (2007)

Dentre as vantagens das MLP estão a sua versatilidade de aplicabilidade pela possibilidade de se alterar o tamanho da rede pela inclusão ou adição de mais camadas neuronais. Silva, Spatti e Flauzino (2010) ainda ressaltam, como potencialidades das MLP: a aproximação universal de funções, o reconhecimento de padrões, a identificação e controle de processos, a previsão de séries temporais e a otimização de sistemas.

3.4.1 O treinamento de uma MLP usando o algoritmo *backpropagation*.

Rumelhart, Hinton e Williams (1986) descreveram no artigo *Learning representations by back-propagating errors* várias aplicações muito eficientes do algoritmo *backpropagation*, que é também chamado de Algoritmo de Retropropagação do Erro. Nos anos seguintes a publicação do artigo, de acordo com Nielsen (2015), a utilização desse algoritmo no processo de treinamento das redes neurais impulsionou a aplicabilidade e o desempenho das mesmas e o *backpropagation* se tornou um dos algoritmos angulares quando se fala em redes neurais e *deep learning*.

O processo de treinamento de uma MLP que utiliza o algoritmo de *backpropagation* é formado por duas fases distintas, a de propagação, conhecida como fase *forward*, onde os valores de entrada da rede seguem um fluxo unidirecional

até a saída da mesma e vão sendo ajustados por pesos inicialmente definidos de forma randômica. Na segunda fase, com o uso do algoritmo *backpropagation*, ocorre a retropropagação, ou fase *backward*, em que os pesos e limiares passam a ser ajustados por uma combinação dos erros quadráticos de treinamento (dados pela diferença quadrática entre os valores desejados, já que o processo é supervisionado, e as saídas da fase *forward*), da taxa de aprendizagem do algoritmo, das derivadas de primeira ordem das funções de ativação dos neurônios da camada anterior no fluxo *backward* e das saídas que a camada seguinte ao fluxo originou na fase *forward*.

A fase *backward* ainda pode ser dividida em duas etapas, sendo a primeira delas o ajuste dos pesos que geram o vetor alimentador da camada de saída, este contém os argumentos das funções de ativação dos neurônios da última camada. Essa primeira fase é diferenciada porque sendo um processo supervisionado existe um conjunto de saídas desejadas e o erro de treinamento é gerado a partir desses valores. Nas correções das camadas seguintes não existe o valor desejado de saída de cada camada intermediária e por isso existe uma diferença no processo de ajuste dos pesos. O ajuste nos pesos da camada de saída pode ser dado pela equação (8).

$$W_{ji}^{(L)}(t+1) = W_{ji}^{(L)}(t) + \eta \cdot \delta_j^{(L)} \cdot Y_i^{(L-1)} \quad (8)$$

Onde $W_{ji}^{(L)}(t+1)$ representa a matriz de pesos da camada L ajustada, $W_{ji}^{(L)}(t)$ a matriz de pesos da camada L antes do ajuste, η a taxa de aprendizagem do algoritmo *backpropagation*, $\delta_j^{(L)}$ o gradiente local do j-ésimo neurônio da camada L e $Y_i^{(L-1)}$ o vetor de saídas da camada anterior à camada L. O vetor $Y_i^{(L-1)}$ é formado por saídas do i-ésimo neurônio da camada L-1, geradas na fase *forward*, e como o processo de treinamento usando o *backpropagation*, nesta etapa, é uma minimização do erro quadrático entre a camada de saída e o valor desejado, o ajuste da matriz $W_{ji}^{(L)}(t)$ é feito em uma direção oposta ao gradiente e, assim, no sentido contrário ao da fase *forward* da MLP. Em função dessa direção e desses ajustes a serem feitos a partir do erro de treinamento é dito que há retropropagação do erro.

Um dos componentes desse ajuste, o gradiente local do j-ésimo neurônio da camada L (de saída), pode ser encontrado através da seguinte equação (9).

$$\delta_j^{(L)} = (d_j - Y_j^{(L)}) \cdot g'(I_j^{(L)}) \quad (9)$$

Onde d_j representa o valor desejado, fornecido pelo conjunto de treinamento,

$Y_j^{(L)}$ o valor fornecido pelo j-ésimo neurônio da camada L, $g'(\cdot)$ a derivada de primeira ordem da função de ativação do j-ésimo neurônio e $I_j^{(L)}$ o vetor com elementos de entrada do j-ésimo neurônio. Estes serão argumentos para a função de ativação $g(\cdot)$ na fase *forward* e que são oriundos de ponderações (dos pesos sinápticos da camada L) das saídas da camada L – 1.

Outro componente do ajuste de pesos é a taxa de aprendizagem do algoritmo, dada por η e definida à priori. Ela é um fator de elevada influência na qualidade do processo de treinamento de uma MLP. Por estar multiplicando um dos termos que compõe o ajuste de pesos, nos casos em que a taxa for expressa por um valor muito baixo, o treinamento pode se tornar demorado, embora mais preciso, porque a variação no valor de correção será muito pequena. Por outro lado, caso a taxa de aprendizagem tenha um valor muito elevado, as oscilações no processo de aprendizagem afetarão a capacidade do treinamento de convergir para um resultado otimizado. Usualmente, a taxa assume valores entre 0.1 e 1 e é fixa durante todo o processo de treinamento.

O processo de ajuste das camadas intermediárias da MLP, como mencionado anteriormente, é baseado na retropropagação do erro porque não há valores desejados de saída dessas camadas.

Nesse caso, o gradiente de do j-ésimo neurônio da camada L – 1, que é a penúltima camada no sentido *forward*, é dado pelo produto do somatório dos pesos corrigidos para a camada L com a derivada parcial da função de ativação do j-ésimo neurônio da camada L – 1, a qual tem como argumento o vetor de valores oriundos do passo *forward* e pode ser representado pela equação (10).

$$\delta_j^{(L-1)} = \left(\sum_{k=1}^n \delta_k^L \cdot W_{kj}^{(L)} \right) \cdot g' \left(I_j^{(L-1)} \right) \quad (10)$$

Ou ainda pelo procedimento iterativo da equação (11).

$$W_{ji}^{(L-1)}(t+1) = W_{ji}^{(L-1)}(t) + \eta \cdot \delta_j^{(L-1)} \cdot Y_i^{(L-2)} \quad (11)$$

O processo de retropropagação do erro de treinamento oriunda da camada de saída segue por todas as camadas que a MLP tiver até que alcance a camada de entrada. Os ajustes da camada (L – (n-1)) serão feitos a partir dos ajustes feitos na camada (L - n) e assim sucessivamente. Rumelhart, Hinton e Williams (1986) afirmam que esses ajustes nas matrizes de pesos podem ser generalizados para qualquer tipo

de topologia de MLP, independentemente da quantidade de camadas.

3.5 As redes RBF

As Redes de Base Radial (*Radial Basis Function* – RBF) ou Funções de Base Radial são redes compostas tipicamente por uma camada intermediária. Essa camada tem neurônios que processam utilizando função de ativação gaussiana. Para Silva, Spatti e Flauzino (2010) as RBF podem ser utilizadas em quase todos os tipos de problemas em que é possível se empregar as MLP.

A rede RBF possui treinamento supervisionado, ou seja, necessita de um conjunto de dados de respostas para ajustar seus pesos no processo de aprendizados. Também as RBF possuem uma arquitetura *feedforward* com dados entrando pela camada de entrada, passando pela intermediária e seguindo para a camada de saída.

3.5.1 Treinamento das RBF

Embora as redes de base radial (RBF), assim como as MLP, tenham um padrão de conexão *feedforward* e tenham também parte de seu treinamento feito de forma supervisionada, uma das diferenças mais notáveis entre as duas redes é que a primeira tem estrutura típica composta por apenas uma camada neuronal intermediária, a qual tem em seus neurônios funções de ativação de base radial, daí sua denominação, comumente utilizando funções do tipo gaussiana e representada pela equação (8). A figura (12) exemplifica a arquitetura de uma RBF.

Figura 12 – Rede de função de base radial (RBF)

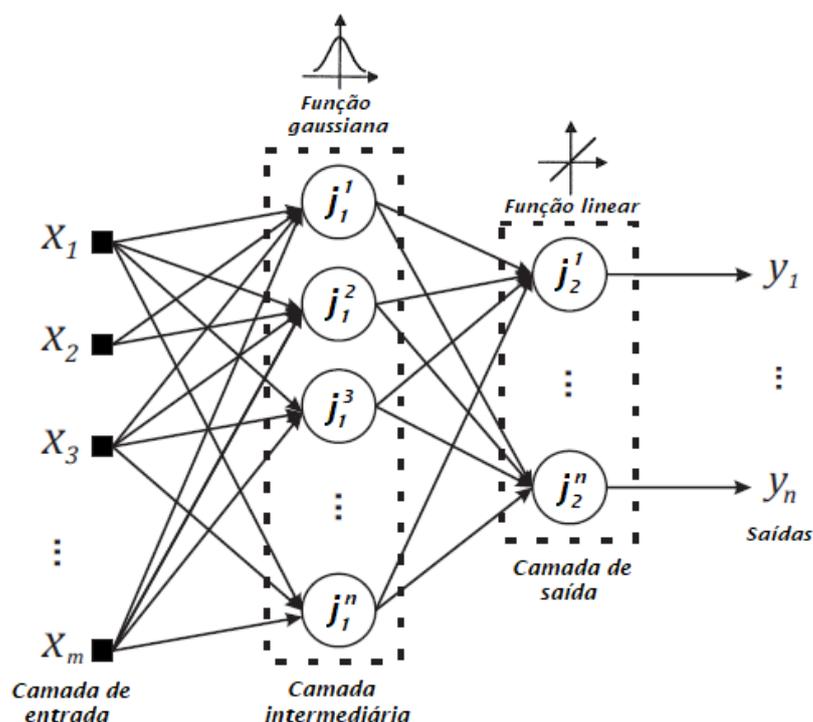


Figura 12 – Adaptado de Silva, Spatti e Flauzino (2010).

Haykin (2001) define uma função de base radial como qualquer função que satisfaça a equação (12). O que, na prática, significa dizer que são funções que têm seus valores funcionais dependentes apenas da distância do seu argumento até à origem.

$$f(x) = f(\|x\|) \quad (12)$$

Na primeira fase do treinamento cada um dos n neurônios da camada intermediária da RBF tem um vetor associado que inicialmente recebe os valores dos conjuntos de entradas. Por exemplo, o conjunto de entradas $\{x_1, x_2, \dots, x_n\}$ em uma RBF com 2 neurônios na camada intermediária, teria os valores de x_1 e x_2 atribuídos aos neurônios j_1 e j_2 na forma de seus correspondentes pesos sinápticos, porque na prática, de acordo com Belotti (2019), cada neurônio da camada intermediária acaba por ser uma função gaussiana, o autor ainda exemplifica que caso a camada intermediária dessa rede possuísse 4 neurônios, por exemplo, os pesos sinápticos dos neurônios j_3 e j_4 também receberiam os valores das entradas x_3 e x_4 , e assim sucessivamente.

Ainda na primeira fase, o passo seguinte é a criação, para cada neurônio j da camada intermediária um correspondente grupo de entradas, representado por Ω^j que

receberá todas as entradas do conjunto de treinamento respeitando a seguinte ordem:

- Um valor de entrada x_i será adicionado ao grupo Ω^j do neurônio j se a base da função gaussiana do neurônio j for a mais próxima do valor de entrada x_i , ou seja, tenha a menor distância euclidiana dentre as outras bases das funções dos outros neurônios da camada intermediária.
- Ao passo que todas as entradas tenham sido alocadas em grupos Ω^j , os centros das gaussianas de cada neurônio (c) são atualizados para que representem todas as entradas x que agora compõem o grupo Ω^j .

Com os centros dos neurônios atualizados, todos os grupos Ω^j serão esvaziados e o passo da adição de entradas x_i a grupos Ω^j será repetido, pois agora temos novos valores de bases para as funções de cada neurônio j . Sequencialmente essas atualizações dos valores de centro das gaussianas será efetuado até que não haja mais alterações na composição dos grupos Ω^j .

É importante ressaltar que após a atualização do centro de uma gaussiana é necessário o recálculo de suas variâncias, essas são calculadas pela equação (13)

$$\sigma_j^2 = \frac{1}{|\Omega^j|} \sum_{x^k \in \Omega^j} \sum_{i=1}^m (x_i^k - w_{ji}^1)^2 \quad (13)$$

em que σ_j^2 representa a variância da base da função gaussiana do j -ésimo neurônio, $|\Omega^j|$ é o valor absoluto do tamanho do conjunto Ω do j -ésimo neurônio, x^k é a amostra k pertencente ao conjunto Ω , m é o número de entradas de cada amostra que comporá o grupo, ou seja, o número de entradas da própria rede, x_i^k representa a entrada i da amostra k e w_{ji}^1 é o peso correspondente ao j -ésimo neurônio da camada intermediária oriundo da entrada i .

O treinamento de uma RBF consiste basicamente, então, na definição da posição dos centros das gaussianas e do grau de abertura das bases das funções gaussianas, que é influenciado pela variância dada por σ^2 , porque $(c - \sigma^2)$ e $(c + \sigma^2)$ delimitam as extremidades da função.

A segunda fase do treinamento da RBF, que corresponde ao treinamento da camada de saída, acontece somente após o treinamento da camada intermediária, outra diferença em relação ao treinamento das MLP. Nessa camada as funções são do tipo linear.

O processo de treinamento da camada de saída é similar ao da MLP, porque

se baseia na regra delta generalizada, é supervisionado porque utiliza as saídas do conjunto de dados de treinamento e também pode ser feito com o uso do algoritmo *backpropagation*.

3.6 As Máquinas de Aprendizado Extremo (ELM)

As Máquinas de Aprendizado extremo (*Extreme Learning Machines - ELM*) foram propostas por Huang, Zhu e Siew (2006) e são redes *feedforward* que não têm suas camadas intermediárias treinadas, os pesos são aleatórios e não ajustados, apenas a camada de saída é treinada.

3.6.1 Treinamento de uma ELM

Além da diferença no número possível de camadas intermediárias entre os modelos MLP e ELM, outra notável diferença entre as redes está no fato de o ajuste dos pesos sinápticos, ou seja, o treinamento, na ELM ser realizado apenas nos pesos da camada de saída. Para tanto, de acordo com Huang, Zhu e Siew (2006) os pesos sinápticos e os limiares da camada intermediária são definidos de forma aleatória e permanecem inalterados.

Como apenas os pesos da camada de saída são atualizados, o processo se inicia com a saída dos valores da camada intermediária se tornando entradas para a camada de saída. De acordo com Huang, Zhu e Siew (2004) essas saídas da camada intermediária podem ser representadas por x_t^h dado pela equação (14).

$$x_t^h = f^h(W^h X u_t + b) \quad (14)$$

Além da diferença no número possível de camadas intermediárias entre os modelos MLP e ELM, outra notável diferença entre as redes está no fato de o ajuste dos pesos sinápticos, ou seja, o treinamento, na ELM ser realizado apenas nos pesos da camada de saída. Para tanto, de acordo com Huang, Zhu e Siew (2006) os pesos sinápticos e os limiares da camada intermediária são definidos de forma aleatória e permanecem inalterados.

Como apenas os pesos da camada de saída são atualizados, o processo se inicia com a saída dos valores da camada intermediária se tornando entradas para a camada de saída. De acordo com Huang, Zhu e Siew (2004) essas saídas da camada intermediária podem ser representadas por x_t^h dado pela equação (15).

$$x_t^h = f^h(W^h X u_t + b) \quad (15)$$

O índice h faz menção à camada escondida (do inglês hidden), a função de ativação $f^h(\cdot)$ é a função dos neurônios da camada intermediária, W^h a matriz de pesos gerados aleatoriamente, $X u_t = [X u_{t1}, X u_{t2}, \dots, X u_{t-(k+1)}]^T$ é o vetor com as entradas que serão enviados à camada intermediária, b é o limiar (bias) de cada neurônio da camada intermediária.

Os valores encontrados para as saídas da camada intermediária, representados por x_t^h serão multiplicados pela matriz de pesos sinápticos da camada de saída W^{out} , e aqui foi mantida a grafia original no inglês, com o termo out fazendo menção à saída, para que se obtenha a saída y_t de acordo com a equação (16).

$$y_t = W^{out} x_t^h \quad (16)$$

Portanto, tendo-se um vetor d , com as saídas desejadas presentes no conjunto de treinamento, treinar a ELM será determinar uma matriz de saídas W^{out} que gere o menor erro possível. Belotti (2019) afirma que uma interessante visão sobre o processo de treinamento de uma ELM, é resumi-lo na resolução do operador pseudo-inversa de Moore-Penrose conforme equação (17).

$$W^{out} = (X_h^T X_h)^{-1} X_h^T d \quad (17)$$

em que X_h representa a matriz com as saídas da camada intermediária e $(X_h^T X_h)^{-1} X_h^T$ a pseudo-inversa de X_h .

Huang, Zhu e Siew (2006) provaram que uma rede ELM é capaz de produzir uma aproximação universal de funções, pois o erro de aproximação produzido por essas redes pode ser sempre reduzido pela adição de um neurônio à cada intermediária, desde que haja um controle dos pesos da camada de saída. Além disso, também o uso da pseudo-inversa de Moore-Penrose trouxe eficiência computacional para as ELM.

Todavia, Huang *et al.* (2012) sugeriram a adição de um coeficiente de regularização C tornando a equação (17) na equação (18) e que pode melhorar a capacidade de aproximação da ELM.

$$W^{out} = \left(\frac{1}{C} X_h^T X_h\right)^{-1} X_h^T d \quad (18)$$

Huang *et al.* (2012) propõem para o cálculo do Coeficiente de Regularização o uso de um conjunto de validação e $C = 2^\lambda$. Sendo o expoente λ um vetor com 52

elementos e $\lambda \in \{-25, -24, \dots, 25, 26\}$. Para seu uso, cada valor do coeficiente C deve ser testado com as entradas da rede pertencentes ao conjunto de validação. O C que apresentar menor erro quando comparado ao conjunto de validação deverá ser o escolhido e, caso a técnica de validação cruzada não seja utilizada no modelo, o Coeficiente de Regularização pode ser testado no conjunto de treinamento da rede.

4 METODOLOGIA

4.1 Enquadramento metodológico do trabalho

Utilizando-se das definições de Gil (2008), em relação aos objetivos desse trabalho ele é enquadrado como explicativo, em razão de utilizar a literatura técnica tanto na busca das variáveis utilizadas nos artigos científicos e modelos de previsão do ruído sonoro urbano publicados, quanto na construção da base teórica que serve de referencial para o trabalho.

Por coletar, reunir e amplificar os conhecimentos acerca do tema previsão ruído sonoro urbano, esse trabalho pode ser classificado como aplicado, pois gera conhecimento, através dos modelos, que pretende ser aplicado à sociedade.

Em relação a abordagem do problema, essa pesquisa caracterizou-se como qualitativa e quantitativa.

A pesquisa pode ser considerada, através da abordagem do problema, como quantitativa.

E quantitativa, porque faz a análise dos parâmetros estatísticos e se utiliza de modelos preditivos puramente matemáticos/estatísticos.

Pelo ponto de vista dos procedimentos técnicos, essa pesquisa pode ser enquadrada como bibliográfica, porque foi fundamentada na análise de estudos científicos sobre modelos de previsão de ruído sonoro urbano. Documental, pois as previsões foram geradas a partir de materiais científicos e dados de bases já publicados. Também operacional, por fornecer além de simulações, ferramentas quantitativas estatísticas de apoio ao processo de decisão pelos melhores modelos.

4.2 Coleta de dados

Os dados desse estudo foram coletados em momentos e locais distintos. As medições dos L_{eq} realizadas em Goiânia e em Maringá foram feitas com um sonômetro com *data logger*, modelo DEC-490 da marca INSTRUTHERM®, e um calibrador acústico, de modelo SC-05 (CEM), ambos com certificado de calibração em acordo com a norma NBR/ISO/IEC 17025:2005. O medidor de nível de pressão sonora utilizado possui precisão de $\pm 1,4$ dB e opera na escala de frequência de 31,5 até 8000 Hz, com ponderações A e C. O modo de resposta *fast* do sonômetro foi utilizado para coletar registros a cada 1 segundo. As respostas apresentadas pelo

aparelho foram em dB(A), na ponderação “A” e calculadas pela equação (19), que têm como variável dependente o resultado observações sonoras somadas segundo a segundo, sendo a variável T o tempo total de medição, n como número de amostras e L_i o valor medido no instante.

$$L_{eq} = 10 \cdot \log_{10} \left(\frac{1}{T} \sum_{i=1}^n 10^{L_i/10} \right) \quad (19)$$

As regiões de estudo deste trabalho, foram a sub-região Setor Marechal Rondon, da cidade de Goiânia, capital do estado de Goiás e seus dados compõem a base 1. A base 2 é composta por dados da Zona 07, região estudantil da cidade de Maringá, no norte do estado do Paraná. As regiões foram escolhidas por serem compostas de avenidas comerciais, áreas residenciais e escolares e serem dotadas de grandes avenidas com intensa movimentação veicular

As amostras de L_{eq} da base 1 foram coletadas em dias de semana comuns (segunda-feira, terça-feira, quarta-feira, quinta-feira, e sexta-feira, excetuando-se os feriados e dias com chuva), em 2016 entre os dias 18 de fevereiro e 29 de abril, das 08:00 às 12:00h e das 13:00 às 17:00h.

As amostras de L_{eq} da base 2 foram coletadas seguindo o mesmo padrão de dias e horários, porém entre 06 e 24 de agosto de 2018.

A escolha dos horários se deu para evitar que possíveis engarrafamentos, horários de pico ou condições atípicas de trânsito pudessem afetar as medições. A norma técnica NBR-10151 da Associação Brasileira de Normas Técnicas (ABNT) foi usada para que as medições seguissem um padrão preestabelecido do volume sonoro. A distância entre pontos de medição variou entre 28 e 150 metros, a depender do tamanho de cada quadra, na coleta de dados da cidade de Goiânia e foi fixa em um ponto na cidade de Maringá.

Os estudos elencados no referencial teóricos deste estudo usaram medições entre 30 segundos a 45 minutos. Ambas as coletas desse trabalho usaram como referencial de tempo de medição estudo de Romeu *et al.* (2011). Neste, os autores recomendaram o tempo de medição de 15 minutos, pois seria o tempo mínimo para se ter um nível de erro aceitável e usual. Foi também realizada a contagem manual do fluxo de veículos separando os veículos em 3 categorias, sendo elas: as motocicletas, os veículos leves (carros e caminhonetes de pequeno porte) e os

veículos pesados (caminhões, camionetes carregadas e ônibus).

A base 1 é composta por 67 amostras, cada uma com 3 entradas categorizadas como descrito acima. A base 2 é composta por 30 amostras com os mesmos tipos de entradas da base 1. As 3 categorias citadas e utilizadas como variáveis de entrada foram contemplam a totalidade dos dados coletados, não houve, portanto, seleção de variáveis, amostras ou dados das amostras: ambas foram utilizadas em sua totalidade.

4.3 Pré-processamento dos dados

As bases de dados usadas nesse estudo foram analisadas segundo a distribuição de frequência normal da variável dependente (L_{eq}). Por serem bases pequenas não foram efetuados testes de normalidade dos dados. Apenas análises estatísticas e gráfica das frequências com comparação com a curva de normalidade, (vide Gráfico 1, Gráfico 2 e Tabela 2 da subseção 5.1.1) a qual demonstrou que as amostram seguiam uma distribuição aproximada da normal.

Posteriormente, foi feita a matriz de correlação de Pearson e também a matriz de correlação de Spearman para verificar se as variáveis de entrada realmente tinham correlação com a variável de saída.

4.4 Modelos de previsão

Silva, Spatti e Flauzino (2010) recomendam que a aplicação de métodos de previsão seja feita com bases divididas em 3 conjuntos: um para treinamento, um para validação e outro para teste dos modelos treinados. É recomendável também, que a amostra de dados utilizadas para treinamento seja composta por uma porcentagem aproximada de 60% a 90% do total das amostras disponíveis, e conseqüentemente, o conjunto para validação e teste deve ser composto pelas amostras restantes (entre 10% e 40% das amostras disponíveis) (SILVA; SPATTI; FLAUZINO, 2016).

Para as amostras do conjunto de teste, não podem ser utilizadas as mesmas amostras de treinamento, pois a função do conjunto de testes é medir a precisão de acerto das redes e elas não podem, então, serem testados com dados já conhecidos porque isso geraria uma medida de erros viesada.

Para ambas as bases foram utilizados 60% dos dados na fase de treinamento das redes neurais e 40% para teste, subdivididos em 20% no subconjunto de

validação e outros 20% no subconjunto de testes.

Para a Regressão Linear Múltipla foram utilizados 80% dos dados da base na fase de ajuste dos coeficientes e os 20% restantes para testes.

Os coeficientes do modelo RLM, foram ajustados pelo método dos quadrados mínimos da equação (2) da subseção 2.1.

A determinação da quantidade de neurônios na camada intermediária das redes foi definida de forma empírica. Definiu-se a arquitetura com apenas 1 camada intermediária devido ao tamanho da base de dados e a fase de treinamento das redes considerou o critério de parada no MSE de 10^{-6} ou 2000 épocas.

A rede MLP foi treinada utilizando-se o gradiente descendente, em 30 épocas e taxa de aprendizagem de 0.01. A função de ativação foi a tangente hiperbólica e o número de neurônios variou entre 3 e 14 na única camada intermediária com incremento de +1 neurônio por teste.

A rede RBF foi treinada utilizando um spread = 1 (abertura da função), em 30 épocas. A função de ativação foi do tipo gaussiana. Utilizou-se o algoritmo *k-means* no treinamento da camada intermediária e o gradiente descendente na camada de saída. O número de neurônios variou entre 3 e 14 na única camada intermediária com incremento de +1 neurônio por teste.

A rede ELM foi treinada utilizando-se a função tangente hiperbólica, em 30 épocas e o número de neurônios variou entre 3 e 14 neurônios na única camada intermediária com incremento de +1 neurônio por teste.

Por fim, escolheu-se a melhor execução como aquela que apresentasse menor MSE frente ao conjunto de testes.

Tanto a configuração do modelo linear RLM quanto as configurações das RNA MLP, RBF e ELM foram desenvolvidas utilizando-se a linguagem Python®.

5 RESULTADOS

Nesse capítulo estão expostas as previsões de valores obtidas pelo modelo linear RLM e pelas RNA para a variável dependente referente ao L_{eq} .

5.1 Variáveis

Nesse subcapítulo é discutida a aplicação prática do modelo de RLM e das RNA elencados para comporem esse trabalho. A aplicação dos modelos foi feita sobre duas bases de dados do tráfego urbano veicular de duas cidades brasileiras. As coletas foram efetuadas pelos pesquisadores e compiladas manualmente.

Os dados da variável dependente dos modelos referem-se ao L_{eq} medido nos pontos de medição com o uso de um sonômetro já descrito na Seção 6.2. Após a compilação dos dados, prosseguiram-se os procedimentos de análise estatística e definição dos parâmetros dos modelos (Seção 5.1.1) e comparação dos desempenhos de previsão (Seção 5.2).

5.1.1 Análise estatística descritiva da variável dependente

Os dados compilados foram analisados estatisticamente e os resultados plotados para identificação de possíveis *outliers* ou fatores que pudessem enviesar as medições.

Na tabela 2, têm-se os parâmetros estatísticos descritivos das bases.

Tabela 2 - Parâmetros estatísticos descritivos das bases na variável dependente

	Média (dBA)	Desvio Padrão (dBA)	Máximo (dBA)	Mínimo (dBA)
Base 1	63,18	6,32	78,60	49,50
Base 2	66,48	1,76	69,80	63,34

Fonte: Autoria própria (2021)

O gráfico (1) demonstra a frequência de distribuição das medições da variável dependente (L_{eq}) da base 1.

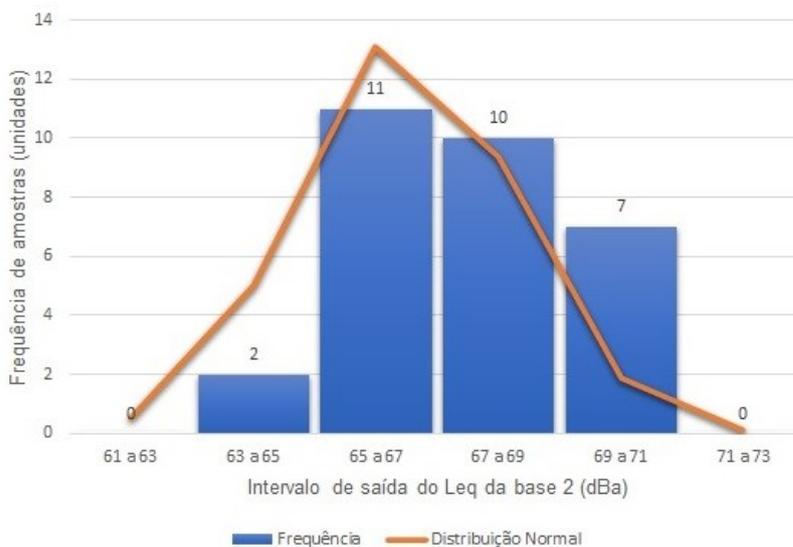
Gráfico 1 – Distribuição da frequência da variável dependente da base 1



Fonte: Autoria própria (2021)

O gráfico 2 demonstra a frequência de distribuição da variável dependente (L_{eq}) da base 2.

Gráfico 2 – Distribuição da frequência da variável dependente da base 2



Fonte: Autoria própria (2021)

Como é possível verificar pelos gráficos, ambas as bases possuem uma distribuição das variáveis dependentes próxima da distribuição normal porque têm curvas que se aproximam do formato sinoidal e centradas nas médias presentes na Tabela 2.

5.1.2 Análise estatística descritiva das variáveis independentes

Os estudos citados na subseção 1.2, constituintes do referencial bibliográfico deste trabalho, nomearam mais de 30 possíveis variáveis independentes usadas como entradas dos modelos de previsão L_{eq} gerada pelo tráfego urbano. Todavia, tanto a base 1 quanto a base 2 deste estudo se limitaram às medições do tráfego urbano estratificado em veículos, veículos pesados e motocicletas como suas variáveis de entrada. Outro diferencial deste estudo é a utilização de duas bases distintas de dados para a avaliação dos modelos previsores.

A Tabela 3 mostra os parâmetros estatísticos descritivos das variáveis de entrada relativos à base 1.

Tabela 3 – Parâmetros estatísticos descritivos das entradas da base 1

	Média	Desvio Padrão	Máximo	Mínimo
Veículos leves	96,21	103,60	337	1
Veículos pesados	3,40	4,43	17	0
Motocicletas	31,87	34,48	105	0

Fonte: Autoria própria (2021)

A Tabela 4 mostra os parâmetros estatísticos das variáveis de entrada relativos à base 2.

Tabela 4 – Parâmetros estatísticos descritivos das entradas da base 2

	Média	Desvio Padrão	Máximo	Mínimo
Veículos leves	548,57	36,03	632	474
Veículos pesados	35,73	5,08	48	28
Motocicletas	206,20	83,40	342	122

Fonte: Autoria própria (2021)

Para se apurar a correlação das variáveis de entrada foram realizados os cálculos para a matriz da correlação. Essa matriz mostra os coeficientes de Pearson, que medem o grau de relação linear entre os pares de variáveis. Esses valores podem estar dentro do intervalo $[-1; 1]$. Valores de coeficiente de Person positivos significam correlação positiva, ou seja, tendem a aumentar ou diminuir juntos. Os coeficientes também medem a força da relação entre as duas variáveis: valores próximos de 1 indicam alta correlação, porém podem significar também que as variáveis medem as mesmas características. Valores afastados de 1, dentro do intervalo $[-1;1]$ podem significar que ou as variáveis medem características diferentes da amostra ou que a correlação não está claramente definida.

Tanto o tipo de área de estudo à que pertence a amostra quanto o tamanho da amostra podem influenciar nos valores de correlação. Entretanto, em geral, esperam-se valores maiores que 0,7 para que uma correlação seja considera forte.

Um segundo teste de correlação foi efetuado com as variáveis. O teste de Spearman procura a correlação entre variáveis contínuas ou ordinais. Os valores das variáveis das amostras utilizadas nesse estudo representavam o valor total da entrada durante 15 minutos de medição. Ademais o L_{eq} também é fruto de uma aplicação da equação (19) para o período amostral de 15 minutos, pois é o resultado de medições feitas segundo a segundo pelo sonômetro. Portanto, essa variável pode ser considerada contínua e avaliada pelo coeficiente de Spearman.

A correlação de Spearman é calculada após uma ordenação crescente dos dados porque trabalha com a posição dos valores em uma escala crescente. Por isso, fala-se também em Correlação dos postos de Spearman para o coeficiente de correlação (ρ). Quanto maior for o valor absoluto de ρ dentro do intervalo $[-1;1]$ mais forte é a correlação das variáveis. A título de exemplo, um $\rho=1$ entre duas variáveis (a , b) significa que o maior valor para a variável a está diretamente ligado ao maior valor para a variável b . Quando o coeficiente possui valores negativos a direção da relação se inverte, mas segue proporcionalmente ao seu valor absoluto.

A tabela 5 mostra a matriz de correlação de Pearson e a matriz de correlação de Spearman para as variáveis de entrada da base 1.

Tabela 5 – Matrizes de correlações da base 1

Correlação de Pearson - R				
	Veículos leves	Veículos Pesados	Motocicletas	L_{eq} (dB)
L_{eq} (dB)	0,765	0,72	0,798	1
Correlação de Spearman - ρ				
	Veículos leves	Veículos Pesados	Motocicletas	L_{eq} (dB)
L_{eq} (dB)	0,748	0,813	0,785	1

Fonte: Aatoria própria (2021)

A tabela 6 mostra a matriz de correlação de Pearson e a matriz de correlação de Spearman para as variáveis de entrada da base 2.

Tabela 6 – Matrizes de correlações da base 2

Correlação de Pearson - R				
	Veículos leves	Veículos Pesados	Motocicletas	L_{eq} (dB)
L_{eq} (dB)	0,192	0,223	0,539	1
Correlação de Spearman - ρ				
	Veículos leves	Veículos Pesados	Motocicletas	L_{eq} (dB)
L_{eq} (dB)	0,31	0,274	0,579	1

Fonte: Aatoria própria (2021)

Embora a base 2 tenha apresentado menores valores de correlação, todas as

variáveis de ambas as bases apresentaram correlação positiva e foram, então, utilizadas para avaliação dos modelos preditivos.

Constatada a regularidade dos dados foram feitos os ajustes dos coeficientes da RLM. Os testes foram realizados com aplicação do método dos quadrados mínimos.

Nos modelos de RNA, como já abordado na seção de metodologia, os testes foram empíricos com a variação de neurônios na camada intermediária dentro de um intervalo de 3 a 14 neurônios com incremento de +1 neurônio para a MLP, ELM e RBF, lembrando que os testes partiram de 3 neurônios porque são 3 entradas nas redes.

A comparação dos resultados obtidos nas previsões geradas com os dados separados para teste utilizou as medidas de avaliação de erros:

- Erro Quadrático Médio (MSE, do inglês *Mean Square Error*): medida muito utilizada na comparação de eficiência de estimadores. Todavia, esse modelo estatístico de medição penaliza mais modelos que apresentem uma menor quantidade de erros grandes do que modelos que apresentem maior quantidade de erros pequenos (MORETTIN; TOLOI, 2006). O MSE é utilizado conforme a equação (20).

$$MSE = \frac{1}{n} \sum_{t=1}^n (d_t - y_t)^2 \quad (20)$$

em que n é o número total de amostras utilizadas na medição, d_t é o valor real e desejado e y_t o valor previsto pelo modelo.

- Erro Absoluto Médio (MAE, do inglês *Mean Average Error*): outra métrica utilizada para comparação de erros. De acordo com Morettin e Tolo (2006) ela indica o quão afastados estão os valores previstos pelo modelo em relação dos valores desejados e penaliza de forma igual erros grandes e pequenos. O MAE pode ser utilizado conforme a equação (21).

$$MAE = \frac{1}{n} \sum_{t=1}^n |d_t - y_t| \quad (21)$$

- Erro Percentual Absoluto Médio (MAPE, do inglês *Mean Absolute Percentual Error*): indica o erro médio percentual (desde que multiplicado por 100), ou seja, uma média percentual dos erros cometidos na previsão. O MAPE pode ser calculado pela equação (22).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{d_t - y_t}{d_t} \right| \quad (22)$$

Os valores de erro apresentados neste estudo correspondem ao menor valor das 30 épocas realizadas para cada modelo.

5.2 Previsão da variável dependente

Tendo os parâmetros sido definidos e os dados avaliados estatisticamente procedeu-se à geração das previsões. Na tabela (7) verificam-se os erros encontrados para cada modelo na geração das previsões com os dados da base 1. Nos parâmetros há o número de neurônios de cada rede que produziu os menores erros.

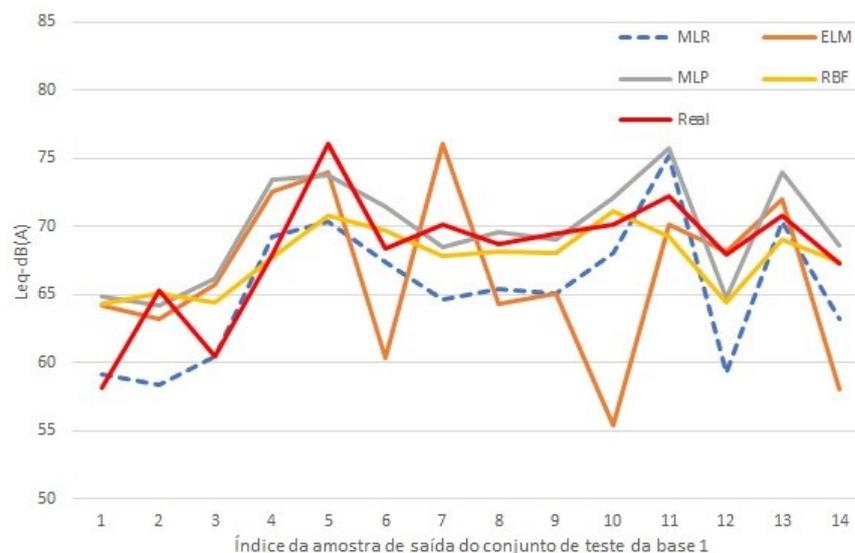
Tabela 7 – Parâmetros estatísticos das previsões da base 1

Ranking	Modelo	Parâmetros	MSE	MAE	MAPE
1	RBF	Neurônios=14 [3,14]	8,196660	2,183331	3,273727
2	MLP	Neurônios=11 [3,14]	11,88230	2,898425	4,388793
3	RLM	-----	17,97152	3,401180	4,934700
4	ELM	Neurônios=10 [3,14]	38,78537	5,042772	7,486070

Fonte: Autoria própria (2021)

O gráfico (3) se refere às previsões para a base 1.

Gráfico 3 - Valores previstos e real para todos os modelos na base 1.



Fonte: Autoria própria (2021)

Na tabela (8) verificam-se os erros encontrados para cada modelo na geração das previsões com a base 2.

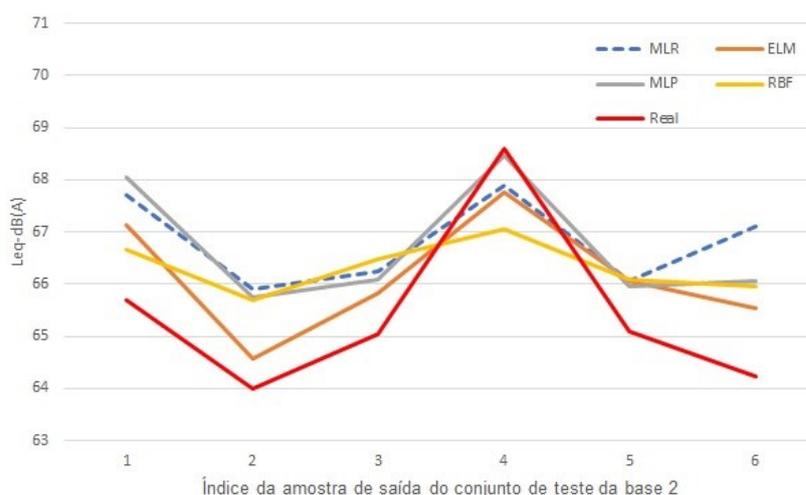
Tabela 8 – Parâmetros estatísticos das previsões da base 2

Ranking	Modelo	Parâmetros	MSE	MAE	MAPE
1	ELM	Neurônios= 1 [3,14]	1,05208	0,98097	1,4998
2	MLP	Neurônios= 7 [3,14]	2,29201	1,32376	2,04075
3	RBF	Neurônios= 9 [3,14]	2,01175	1,38509	2,11925
4	RLM	-----	3,11374	1,60243	2,46634

Fonte: Autoria própria (2021)

O gráfico (4) se refere às previsões para a base 2.

Gráfico 4 – Valores previstos e real para todos os modelos na base 2.

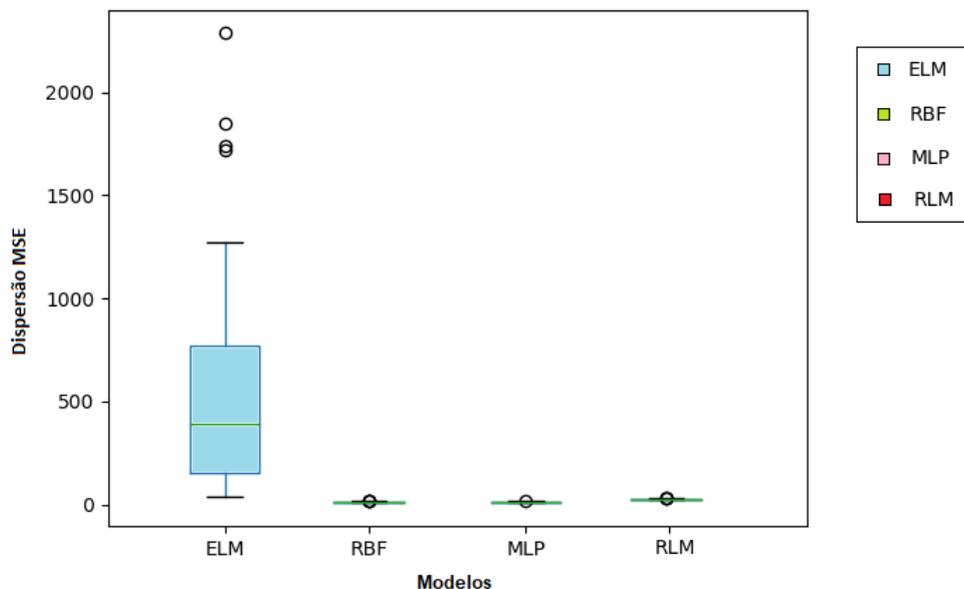


Fonte: Autoria própria (2021)

Quando ranqueados os resultados apresentados pelos modelos demonstraram pequenas divergências entre as bases. Para base 1 de dados, os menores erros de previsão foram apresentados pelo modelo neural RBF, já os maiores erros com o modelo neural ELM. A divergência entre as bases ocorreu justamente nesse quesito: um modelo neural perdeu para o modelo RLM, que ocupou a terceira posição. A base 2, por sua vez, apresentou o modelo ELM com o menor erro de previsão, mas com todos os modelos neurais superando o modelo linear, que ficou em última posição.

O gráfico (5) é um gráfico de Box-Plot da dispersão verificada nas 30 simulações realizadas com cada modelo nos dados da base 1.

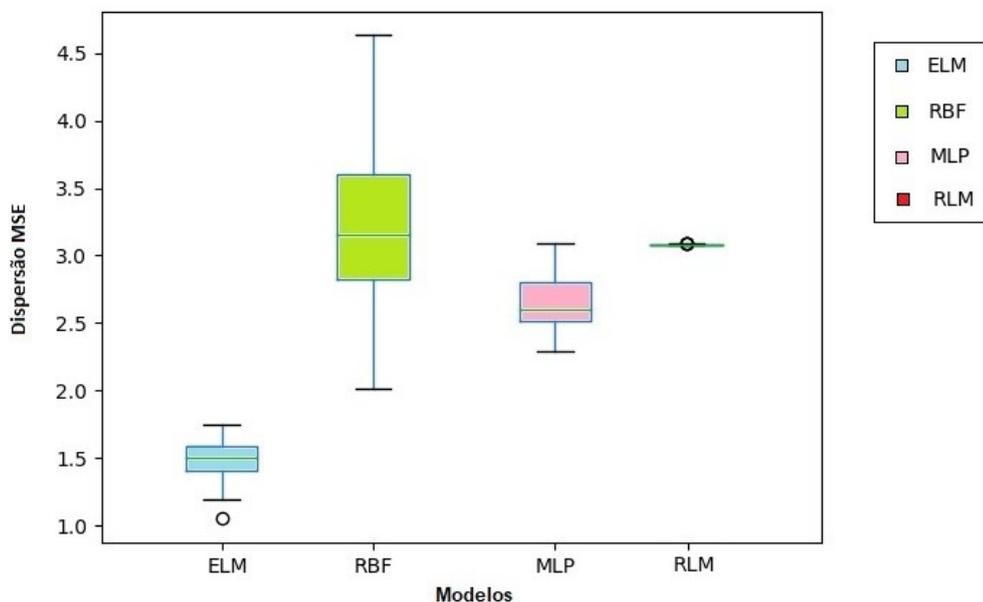
Gráfico 5 – Box-Plot de dispersão de MSE da base 1



Fonte: Autoria própria (2021)

O gráfico (6) é um gráfico de Box-Plot da dispersão verificada nas 30 simulações realizadas com cada modelo nos dados da base 2.

Gráfico 6 – Box-Plot de dispersão de MSE da base 2



Fonte: Autoria própria (2021)

Deina (2020) afirma que nas RNA no processo de ajuste de seus pesos sinápticos, que conhecemos por treinamento (para que a rede adquira capacidade de generalização e mapeamento), as redes têm seus pesos sinápticos inicializados de forma aleatória, o que, como afirma a autora, é a causa de distintos resultados em

cada uma das 30 simulações executadas para cada modelo, gerando a dispersão de resultados mostradas nos gráficos de Box-Plot.

Na base 1 a maior dispersão interquartilica dos erros médios quadráticos das previsões foi observada no modelo ELM. Já na base 2 ela aconteceu com o modelo RBF. Em ambos os casos, dentre os modelos neurais na base 1 a ELM e na base 2 o RBF apresentaram os maiores erros de generalização. Embora os parâmetros estatísticos de erro envolvidos demonstrem os valores médios, e consequentemente estejam ligados à distância interquartilica, é possível concluir que as redes com maiores erros foram as menos capazes de generalizar uma solução ideal porque apresentaram maiores valores de erro a cada iteração.

Na base 1, o modelo ELM apresentou valor de erro muito acima da média dos outros modelos. Todavia, na base 2 ele apresentou o menor erro médio.

O fato de nesse estudo a base 1 apresentar melhor desempenho com um modelo ELM e a base 2 o pior resultado com um modelo ELM, pode ser resultado da diferença de tamanho das bases, já que a base 1 era mais de 120% maior que a base 2. Também, havia discrepância na amplitude da variável dependente associada às bases. Enquanto na base 1 a amplitude do L_{eq} foi de 29,12 dB, na base 2 a amplitude verificada foi de 6,46 dB. Por fim ainda houve uma enorme diferença no fluxo de motocicletas, que foi a variável independente com maior correlação com a dependente. Na base 1 a média das amostras foi de 31,87 motocicletas, na base 2 ela foi de 206,20 motocicletas por amostra. Esses fatores citados contribuem para aumentar ou mesmo mitigar a sensibilidade dos modelos e, portanto, podem gerar melhores treinamentos e, consequentemente, melhores previsões em arquiteturas e topologias diferentes.

Portanto, pode-se inferir que na presença de uma variável independente com alto valor de correlação com a variável dependente somada à presença de uma amplitude maior entre os dados o modelo ELM não foi capaz de generalizar uma solução ideal melhor que os outros modelos avaliados, produzindo erro médio maior que todos os modelos, inclusive o modelo linear. A particularidade do modelo ELM, de ter apenas a sua camada de saída treinada com o conjunto de treinamento pode ser uma das causas dos diferentes desempenhos do modelo entre as bases: na base 1 o desvio padrão dos dados da base foi de 6,32 dBA. Já na base, na qual a ELM apresentou-se vencedora, o desvio padrão foi de 1,76 dBA.

O modelo MLP, que é treinado pela retropropagação do erro tanto para a camada de saída quanto para a intermediária, não teve alteração de posição no ranking de erros, permanecendo na segunda posição. Pode-se inferir que as diferenças entre as bases, em termos de desvio padrão e amplitude do conjunto de treinamento, não afetaram a capacidade de generalização do modelo e nos ajustes usando as duas bases deste estudo, quando comparado aos demais.

O modelo RBF, que tem seu treinamento da camada intermediária basicamente na definição da posição dos centros das gaussianas (funções de ativação dos neurônios da camada intermediária), apresentou melhor resultado de generalização na base onde o desvio padrão dos dados do conjunto de treinamento foi maior, ou seja, havia uma distância euclidiana maior entre os dados. Onde o desvio padrão foi menor, o modelo RBF só não teve um desempenho pior do que o modelo linear.

Em todos os casos, entretanto, algum modelo de neural foi melhor do que o modelo linear, o que já era esperado porque é conhecida a maior capacidade de generalização de soluções dos modelos neurais frente aos lineares.

6 CONCLUSÕES

Há, na literatura, alguns estudos buscando resolver esse problema com modelos de previsão lineares, e até mesmo com o uso de RNA. Todavia, as limitações dos modelos estão quase sempre ligadas a carência de dados e ao excesso de variáveis que influem na problemática da poluição sonora.

Algumas das variáveis dos modelos propostos na literatura científica são, também, muito específicas de cada local gerador do modelo. Sendo o som uma onda mecânica, a densidade do ar local, variável essa influenciada também pela temperatura, pressão e umidade locais, altera a propagação da onda sonora e, portanto, torna ainda mais complexa a definição de modelos genéricos. Características locais como o uso frequente da buzina, a idade da frota de veículos e até mesmo as características de motorização dos mesmos, também podem ser fatores que influenciarão na geração do ruído urbano do tráfego veicular.

Devido ao grande número de variáveis relacionadas à problemática, a utilização de modelos lineares torna-se, como comprovado pela literatura científica, laboriosa e imprecisa. As RNA, então, assumem o papel de modelagem de problemas complexos pela altíssima flexibilidade que possuem para trabalhar com muitas variáveis de entradas, arquiteturas e topologias.

Neste estudo, as bases de dados utilizadas para comprovar a capacidade de previsão superior das RNA frente ao modelo linear a partir do L_{eq} oriundo do tráfego urbano foram construídas especificamente para o objetivo de previsão desta variável e em duas diferentes regiões do Brasil.

Diferentemente dos estudos realizados e publicados até hoje, duas bases com variáveis iguais, mas níveis de tráfego distintos e em duas regiões de um país não haviam sido utilizadas na comparação da capacidade de previsão das RNA, frente a um modelo linear, para o L_{eq} do tráfego veicular. Tendo isso em vista, esse trabalho realizou essa comparação utilizando o modelo preditivo RLM e as RNA: MLP, RBF e ELM.

Em ambas as bases os resultados das RNA foram superiores, por apresentarem menores erros quando comparados ao do modelo linear, exceto pelo modelo ELM na base 1 que teve erro maior do que o modelo linear RLM. Esse resultado superior das RNA já era esperado, visto que os modelos neurais já conhecidos cientificamente pelas suas altas capacidade de generalização de

soluções.

Assim como a maioria dos trabalhos utilizados como referencial teórico nesse estudo, aconselha-se para estudos futuros a utilização de bases de dados maiores. As redes neurais artificiais podem ser usadas como métodos muito robustos de previsão, mas necessitam de uma quantidade grande de dados, especialmente para seu processo de treinamento.

Na fase de construção da base de dados é aconselhável o cômputo do maior número de variáveis possível, para que se verifique a correlação das mesmas com o L_{eq} e se aprimore o modelo desenvolvido.

Alguns estudos utilizam métodos de seleção de variáveis para diminuir o custo computacional da geração das previsões e especialmente em bases com elevada quantidade de dados e variáveis esses métodos seriam muito bem aproveitados

REFERÊNCIAS

- AHMED, A. A.; PRADHAN, B. Vehicular traffic noise prediction and propagation modelling using neural networks and geospatial information system. **Environmental Monitoring and Assessment**, v. 191, n. 3, 2019.
- ANOCHI, J. A. **Previsão climática de precipitação por redes neurais autoconfiguradas**. 2015. Tese (Doutorado). Curso de Pós-Graduação em Computação Aplicada, Instituto Nacional de Pesquisas Espaciais – INPE. São José dos Campos, 2015.
- Anon. FHWA. Traffic noise prediction model US. Washington: Department of Transportation, Federal Highway Administration National Technical Information Service, 1978.
- Anon. Handbook of acoustic noise control WADC technical report 52-204. **Wright Air Development Center**, 1952.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 17025:2005**: requisitos gerais para a competência de laboratórios de ensaio e calibração. General requirements for the competence of testing and calibration laboratories. Rio de Janeiro: ABNT, 2005.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **Acústica Avaliação do ruído em áreas habitadas, visando o conforto da comunidade - Procedimento: NBR 10151:2000**. Rio de Janeiro: ABNT, 2000.
- BABISCH, W., BEULE, B., SCHUST, M., KERSTEN, N., ISING, H. Traffic noise and risk of myocardial infarction. **Epidemiology**. v.16, p.33–40.2005
- BABISCH, W., PERSHAGEN, G., SELANDER, J., HOUTHUIJS, D., BREUGELMANS, O., CADUM, E. Noise annoyance— a modifier of the association between noise level and cardiovascular health, **Scientific Total Environment** v.50, n.7, p.452–453. 2013
- BARBOSA, A. L. S.. **Estudo de Barreiras Acústicas para a Atenuação do Ruído Aeronáutico no Aeroporto de Congonhas em São Paulo**. 2015. 346 f. Tese (Doutorado) - Curso de Arquitetura e Urbanismo, Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo, São Paulo, 2015.
- BARREGARD, L., BONDE, E., OHRSTRÖM, E.. Risk of hypertension from exposure to road traffic noise in a population-based sample. **Occup. Environ. Med.** V.66, p.410–415, 2009.
- BELOTTI, J. T. **Previsão de vazões afluentes utilizando redes neurais artificiais e ensembles**. 2019. Dissertação (Mestrado). Programa de Pós-Graduação em Ciência da Computação, Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2019.
- BODIN, T., ALBIN, M., ARDÖ, J., STROH, E., Ostergren, P.-O., Björk, J.. Road traffic noise and hypertension: results from a cross-sectional public health survey in southern Sweden. **Environ. Health: Glob. Access Sci.** v.38, n.8. 2009.

- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. 2. ed. Rio de Janeiro: LTC Editora, 2007.
- CAMMARATA, G.; CAVALIERI, S.; FICHERA, A. A neural network architecture for noise prediction. **Neural Networks**, v. 8, n. 6, p. 963–973, 1995.
- CHANG, T.-Y., LIU, C.-S., BAO, B.-Y., LI, S.-F., CHEN, T.-I., LIN, Y.-J. Characterization of road traffic noise exposure and prevalence of hypertension in central Taiwan. **Sci. Total Environ.** v.409, p.1053–1057, 2011.
- CIRIANNI, F.; LEONARDI, G. Environmental modeling for traffic noise in urban area. **American Journal of Environmental Sciences**, v. 8, n. 4, p. 345–351, 2012.
- DANTAS, R. A. Engenharia de Avaliações: uma introdução à metodologia científica. 1.^a Ed. São Paulo: Pini, 2003
- DEINA, C. **Modelo para previsão de demanda de eletricidade com redes neurais artificiais integradas a métodos multicritério**. 2020. Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Universidade Tecnológica Federal do Paraná. Pato Branco, 2020.
- FORTUNA, L. *et al.* **Neuro-fuzzy model of urban traffic**. Midwest Symposium on Circuits and Systems. **Anais...**Universita di Catania, Catania, Italy: 1994
- GALLOWAY, W. J. *et al.* Urban highway noise: measurement, simulation, and mixed reactions. **Nchrp Report**, [s.l.], p. 78. 1969.
- GARG, N. *et al.* Comparison of ANN and Analytical Models in Traffic Noise Modeling and Predictions. **Acoustics Australia**, v. 43, n. 2, p. 179–189, 9 ago. 2014.
- GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2002.
- GITHUB. **O neurônio biológico**. Disponível em: <http://cs231n.github.io/neural-networks-1>. Acesso em: 12 mar. 2021.
- GIVARGIS, S.; KARIMI, H. A basic neural traffic noise prediction model for Tehran's roads. **Journal of Environmental Management**, v. 91, n. 12, p. 2529–2534, 2010.
- HAIR, J. F., ANDERSON, R. E., TATHAM, R. L., BLACK, W. C. **Análise multivariada de dados**. São Paulo: Bookman, 2005.
- HAYKIN, Simon. **Redes neurais: princípios e práticas**. Trad. Paulo Martins Engel. 2° ed. -Porto Alegre: Bookman. 2001.
- HODGKIN, A. L.; HUXLEY, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. **The Journal of physiology, Wiley-Blackwell**, v. 117, n. 4, p. 500, 1952.
- HUANG, G.-B., Q.-Y. ZHU , AND C.-K. SIEW. "Extreme Learning Machine: A New Learning Scheme of *Feedforward* Neural Networks." **Proc. of Intern. Joint Conf. on Neural Networks**. p. 985-990. 2006.

IBGE. **PNAD - Pesquisa Nacional por Amostra de Domicílios**. 2015. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/educacao/9127-pesquisa-nacional-por-amostra-de-domicilios.html?=&t=o-que-e>. Acesso em: 26 maio 2020.

IPEA – Instituto de Pesquisa Econômica Aplicada (Brasil). **Desafios da mobilidade urbana no Brasil**. v. 1, n. 2198, Brasília, 2016. 30p. Disponível em: http://repositorio.ipea.gov.br/bitstream/11058/6664/1/td_2198 Acesso em: 25 set. 2021

JAKOVLJEVIC´ , B., BELOJEVIC´ , G., PAUNOVIC´ , K., STOJANOV, V., Road traffic noise and sleep disturbances in an urban population: cross-sectional study. **Croat. Med. J.** v.47, p.125–133. 2006

JOHSON, D R; SAUNDERS, E G. The evaluation of noise from freely flowing road traffic. **Journal of Sound And Vibration**. v. 7, n. 2, p. 287-309, 1968.

KHALIL, A.; M.; HAMAD, K.; SHANABLEH, A. Developing Machine Learning Models to Predict Roadway Traffic Noise: An Opportunity to Escape Conventional Techniques. **Transportation Research Record: Journal of the Transportation Research Board**, 2017 v. 2673, n. 4, p. 158–172, 4

KOYASU M. Method of prediction and control of road traffic noise in Japan. **Inter-noise 78**, San Francisco, 1978.

KUMAR, K.; PARIDA, M.; KATIYAR, V. K. Prediction of urban traffic noise using artificial neural network approach. **Environmental Engineering and Management Journal**, v. 13, n. 4, p. 817–826, 2014.

LAMURE C. Niveaux de bruit au voisinage des autoroutes. In: **Proc. Fifth International Congress on Acoustics**, 1965.

LENT, Roberto. **Cem Bilhões de Neurônios?: conceitos fundamentais de neurociência**. 2. ed. Rio de Janeiro: Atheneu, 786 p., 2010.

L'Office fédéral de la protection de l'environnement. **Modèle de calcul de bruit du trafic routier pour ordinateur, lére partie: manuel d'utilisation du logiciel STL-86**. Berne: version 1.0, 1987.

LUCAS, M. **Sensor virtual inteligente para estimação de composições em colunas de destilação**. Dissertação (Mestrado). Escolha de Engenharia de São Carlos da Universidade de São Paulo, São Carlos, 2012.

MCCULLOCH, W. S., PITTS, W.; A Logical Calculus of Ideas Immanent in Nervous Activity”, **Applied Mathematics and Computation** v.5, p.115–133.1943

MONTGOMERY, D.C. **Introduction to statistical quality control**. 2. Ed. New York: John Willey & Sons, 674 p., 1991

MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. 2. ed. São Paulo: Blucher, 2006.

NEDIC, V. Comparison of classical statistical methods and artificial neural network in traffic noise prediction. **Environmental Impact Assessment Review**, v. 49, p. 24–30, 2014.

NICKSON, A. F.. Can community reaction to increased traffic noise be forecast? In: FIFTH INTERNATIONAL CONGRESS ON ACOUSTICS, 1965, [s.l]. **Proceedings...** . [s.l]: [s.e]., 1965.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (Suíça). **The World Health Report 2003: shaping the future**. Genebra: Who Library, 2003.

PARBAT, D. K.; NAGARNAIK, P. B. **Ann modeling of noise levels due to vehicular traffic flow in Indian intermediate city**. 15th International Congress on Sound and Vibration 2008, ICSV 2008. **Anais...**Civil Eng Department, Government Polytechnic, Nagpur, India: 2008

PIA, H.W Plasticity of the central nervous system- a neurosurgeon's experience of cerebral compensation and decompensation. **Acta Neurologica**, n. 77, p. 81-102. 1985

PIRRERA S, DE VALCK E, CLUYDTS R. Nocturnal road traffic noise: a review on its assessment and consequences on sleep and health. **Environ International**. v.36, n.5, p. 492–8. 2010.

RAMÍREZ, A.; DOMÍNGUEZ, E. Modeling urban traffic noise with stochastic and deterministic traffic models. **Applied Acoustics**, v. 74, n. 4, p. 614-621. 2013

RELVAS, Marta Pires. **Fundamentos Biológicos da Educação: Despertando Inteligências e Afetividade no Processo de Aprendizagem**. 2. ed. Rio de Janeiro: Wak, 2009.

RELVAS, Marta Pires. **Fundamentos Biológicos da Educação: potencialidades dos gêneros humanos na sala de aula**. 2. ed. Rio de Janeiro: Wak, 2010.

ROMEU, J., GANESCA, M., PÀMIES, T., JIMÉNEZ, S. "Street categorization for the estimation of day levels using short-term measurements", **Applied Acoustics**., v. 72, no. 8, p. 569-577, 2011

ROSENBLATT, F. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", **Psychological Review** v.65, p.386–408.1958

SELANDER, J., NILSSON, M.E., BLUHM, G., ROSENLUND, M., LINDQVIST, M., NISE, G., PERSHAGEN, G., Long-term exposure to road traffic noise and myocardial infarction. **Epidemiology (Cambridge, Mass.)** v.20. p.272–279. 2009.

SHEPHERD, G. AND KOCH, C. Dendritic electrotonus and synaptic integration. In: *The Synaptic Organization of the Brain*, edited by G. M. Shepherd. **New York: Oxford Univ. Press**, p. 439–473, 1990.

SILVA, I. N.; SPATI, D. H.; FLAUZINO, R. A. 2010. **Redes Neurais Artificiais: para engenharia e ciências aplicadas**. São Paulo: Artliber, 399 p., 2010.

SINGH, D. *et al*. Vehicular traffic noise prediction using soft computing approach. **Journal of Environmental Management**, v. 183, p. 59–66, 2016.

SIQUEIRA, H. V. **Máquinas desorganizadas para previsão de séries de vazões**. 2013. 218 f. Tese (Doutorado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 2013.

STEELE, Campbell. A critical review of some traffic noise prediction models. **Applied Acoustics**, [s.l.], v. 62, n. 3, p.271-287, 2001.

TOMÍĆ, J.; BOGOJEVIĆ, N.; PLJAKIĆ, M.; IUMARAC-PAVLOVIĆ, D. Assessment of traffic noise levels in urban areas using different soft computing techniques. **The Journal Of The Acoustical Society Of America**, [S.L.], v. 140, n. 4, p. 340-345, out. 2016.

Von Zuben, F. J. e Attux, R. R. F. (2007). “Notas de Aula de Redes Neurais Artificiais”, Disponível em: <http://www.dca.fee.unicamp.br/~vonzuben/courses/ia353.html>. Acesso em 18 de setembro de 2020.

WORLD HEALTH ORGANIZATION (Suíça). **World Report on hearing**. 14. ed. Genebra, 2021. 252 p. Disponível em: <https://apps.who.int/iris/handle/10665/339913>. Acesso em: 25 set. 2021.