

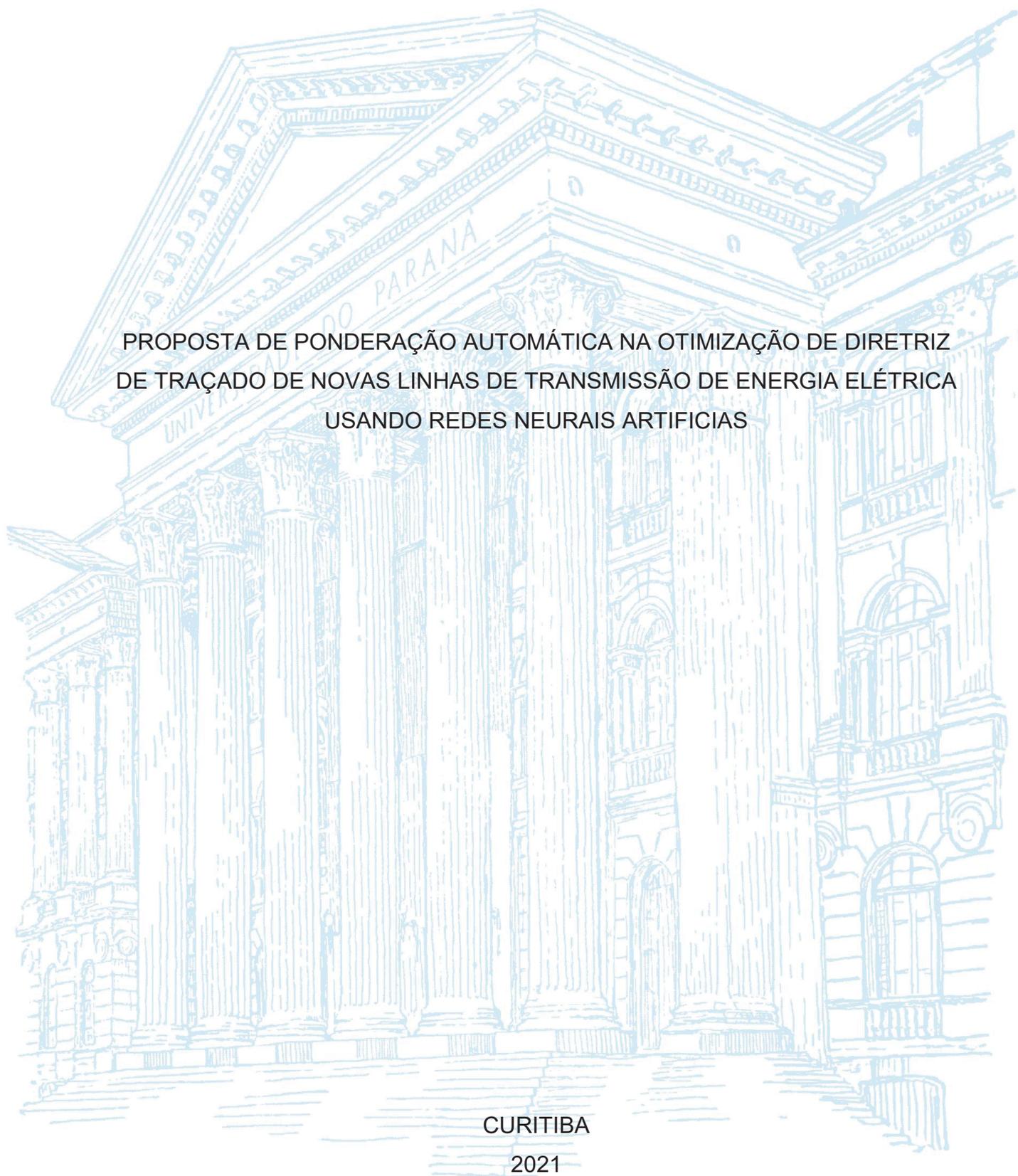
UNIVERSIDADE FEDERAL DO PARANÁ

LUIZ HENRIQUE BOSSOLA

PROPOSTA DE PONDERAÇÃO AUTOMÁTICA NA OTIMIZAÇÃO DE DIRETRIZ
DE TRAÇADO DE NOVAS LINHAS DE TRANSMISSÃO DE ENERGIA ELÉTRICA
USANDO REDES NEURAS ARTIFICIAS

CURITIBA

2021



LUIZ HENRIQUE BOSSOLA

PROPOSTA DE PONDERAÇÃO AUTOMÁTICA NA OTIMIZAÇÃO DE DIRETRIZ
DE TRAÇADO DE NOVAS LINHAS DE TRANSMISSÃO DE ENERGIA ELÉTRICA
USANDO REDES NEURAS ARTIFICIAS

Dissertação apresentada como requisito parcial à
obtenção do grau de Mestre em Ciências
Geodésicas, no curso de Pós-Graduação em
Ciências Geodésicas, Setor de Ciências da Terra,
da Universidade Federal do Paraná.

Orientador: Prof. Dr. Daniel Rodrigues dos Santos

CURITIBA

2021

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

B745p Bossola, Luiz Henrique
Proposta de ponderação automática na otimização de diretriz de traçado de novas linhas de transmissão de energia elétrica usando redes neurais artificiais [recurso eletrônico] / Luiz Henrique Bossola – Curitiba, 2021.

Dissertação - Universidade Federal do Paraná, Setor de Ciências da Terra, Programa de Pós-graduação em Ciências Geodésicas.
Orientador: Prof. Dr. Daniel Rodrigues dos Santos

1. Linhas elétricas. 2. Sistema de informação geográfica. 3. Redes neurais artificiais (Neuralnet). I. Universidade Federal do Paraná. II. Santos, Daniel Rodrigues dos. III. Título.

CDD: 526.31

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS DA TERRA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS
GEODÉSICAS - 40001016002P6

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação CIÊNCIAS GEODÉSICAS da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **LUIZ HENRIQUE BOSSOLA** intitulada: **PROPOSTA DE PONDERAÇÃO AUTOMÁTICA NA OTIMIZAÇÃO DE DIRETRIZ DE TRAÇADO DE NOVAS LINHAS DE TRANSMISSÃO DE ENERGIA ELÉTRICA USANDO REDES NEURAIIS ARTIFICIAS**, sob orientação do Prof. Dr. DANIEL RODRIGUES DOS SANTOS, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 27 de Setembro de 2021.

Assinatura Eletrônica

04/10/2021 15:25:39.0

DANIEL RODRIGUES DOS SANTOS

Presidente da Banca Examinadora

Assinatura Eletrônica

01/10/2021 09:55:56.0

RODRIGO BEZERRA DE ARAUJO GALLIS

Avaliador Externo (UNIVERSIDADE FEDERAL DE UBERLÂNDIA)

Assinatura Eletrônica

04/10/2021 15:58:56.0

ALZIR FELIPPE BUFFARA ANTUNES

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

AGRADECIMENTOS

Primeiramente agradeço minha família, que sempre me apoiou e não me deixou desistir apesar de todas as adversidades vividas ao longo dessa jornada. Aos meus pais Eloá e Antônio Bossola e minha irmã Míria, por todo a paciência, amor e incentivo. Essas pessoas são responsáveis por minha formação como pessoa e por me fornecer a estrutura necessária para eu realizar mais esta etapa na minha vida acadêmica, que mesmo sem condições financeiras às vezes, fizeram de tudo para eu concluir cada etapa da minha vida com sucesso.

Aos meus amigos Alex, Kalima, Kauê, Daniel (Fluniel) e todas as pessoas que me ajudaram com dúvidas e me auxiliaram com ajudas técnicas e de programação quando a pesquisa parecia estagnada ou impossível.

A todos meus amigos próximos, de infância, do ensino médio técnico lá de Pato Branco, da universidade ou da vida, pelos momentos de descontração e pelas conversas nas mesas dos bares de Curitiba.

Aos meus amigos da empresa Smart Sky Tech Hub, pelo incentivo para a conclusão desta etapa, em especial para ao André, Chase, Renata, Marcelo, Josiane e Marcus que sempre me motivaram a terminar.

Aos meus professores em todos os níveis de ensino, responsáveis por minha formação intelectual ao longo da minha vida, mas em especial para os professores do Departamento de Geomática e da Pós-Graduação em Ciências Geodésicas da UFPR, que estão me orientando tecnicamente em todas as disciplinas desde o início da minha graduação, em 2014.

Ao professor Daniel, meu orientador, pelas críticas, conversas, direcionamentos, e pela paciência por receber sempre “em cima da hora” os documentos e materiais.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo fomento da bolsa recebida para a elaboração dos estudos.

A todos que direta ou indiretamente estiveram presentes e me ajudaram de alguma forma na produção da dissertação e no decorrer do mestrado.

RESUMO

De acordo com a ONS (Operador Nacional de Sistema Elétrico), há uma previsão de construção de aproximadamente 39,8 mil quilômetros de novas Linhas de Transmissão (LTs) com tensões entre 230 kV e 500 kV no país. Levando em consideração a largura da faixa de abrangência de uma nova obra que é de 5 km (segundo legislação vigente), têm-se uma área potencial indireta a ser afetada de 199 mil km² e direta de mais de 2 mil km². Um dos fatores preponderantes na construção de LTs é a fase de planejamento do traçado de diretrizes (definição traçado inicial e experimental), que envolve o conhecimento prévio de fatores restritivos presentes na área de instalação do empreendimento. Em abordagens convencionais, o planejamento do traçado de diretriz de LTs é realizado de forma manual, com base em um conjunto de dados extraídos de cartas topográficas do exército, imagens de satélite, entre outras bases de dados disponíveis, dependendo também do profissional que realiza essa análise e a ponderação de critérios são subjetivos. Dessa forma, o objetivo principal deste estudo é utilizar inteligência artificial (IA) para o cálculo dos pesos de cada variável de forma homogênea visando minimizar impactos sociais, ambientais, culturais e econômicos na fase de diretriz de traçados na construção de novas linhas de transmissão. A RNA utilizada é a Neuralnet, baseada em linguagem R, utilizando a entropia cruzada como métrica de erro. Foram utilizados como dados de entrada da rede as variáveis: uso do solo, declividade, áreas de proteção integral, aeródromos, áreas indígenas, altimetria, linhas de transmissão existentes e empreendimentos lineares, composto pelas camadas de ferrovias, rodovias e dutovias. Foi realizado o treinamento da RNA utilizando 2 mil e 20 mil amostras, escolhidas aleatoriamente, sendo 60% delas utilizadas para a checagem da qualidade do modelo treinado. Verificou-se que a utilização de 5 neurônios ocultos é suficiente para a geração automática dos pesos usando a RNA. A taxa de sucesso máxima alcançada pelo modelo treinado foi de 69% utilizando a quantidade maior de amostras. A taxa de sucesso encontrada é resultado da baixa qualidade das amostras empregadas, uma vez que há inúmeras camadas utilizadas para entrada do modelo que não foram consideradas na construção das linhas de transmissão usadas, neste trabalho, como amostras positivas.

Palavras-chave: Estudo de traçado de diretriz. Inteligência artificial. Linhas de transmissão. Neuralnet. Operador Nacional do Sistema Elétrico.

ABSTRACT

As reported by NESO (National Electric System Operator), there are plans to build approximately 39.8 thousand kilometers of new Transmission Lines (TLs) with voltages between 230 kV and 500 kV in the national territory. According to current legislation, the width of the range of a new Transmission Line construction is 5 km, thus, there is a potential indirect area of 199 thousand km² to be changed and a direct area of more than 2 thousand km². One of the most important factors in the construction of TLs is the planning phase of the guideline layout (initial and experimental route definition), which involves prior knowledge of restrictive factors present in the area of the project. In conventional approaches, the LT's guideline path planning is done manually, based on a set of data extracted from Army topographic maps, satellite images, among other available databases. This planning depending on the professional who performs this analysis and, thus, the criteria weighting are subjective. Hence, the main objective of this study is to use Artificial Intelligence (AI) to calculate the weights of each variable in a homogeneous way, in order to minimize social, environmental, cultural, and economic impacts in the guideline path phase of the construction of new Transmission Lines. For this purpose, the Artificial Neural Network (ANN) used is Neuralnet, which is based on the R language, using cross-entropy as an error metric. The following variables were used as input data for the network: land use, slope, integral protection areas, airfields, indigenous areas, altimetry, existing transmission lines, and linear enterprises (layers of railroads, highways, and pipelines). The ANN was trained using 2 thousand and 20 thousand randomly chosen samples, 60% of which were used to check the quality of the trained model. It was found that the use of 5 hidden neurons is sufficient for the automatic generation of the weights using ANN. The maximum success rate achieved by the trained model was 69% using the larger amount of samples. The success rate found is a result of the low quality of the used samples, since there are numerous layers used for model input that were not considered in the construction of the transmission lines used as positive samples in this work.

Keywords: Guideline Planning. Artificial Intelligence. Transmission Lines. Neuralnet.
National Electric System Operator.

LISTA DE FIGURAS

FIGURA 1 – EXTENSÃO DA REDE BÁSICA DE TRANSMISSÃO E EXTENSÕES PREVISTAS	17
FIGURA 2 – SUBSISTEMAS DO SIN	20
FIGURA 3 – REDE DE TRANSMISSÃO DO BRASIL	21
FIGURA 4 – SUPERFÍCIE(S) DE CUSTO(S)	28
FIGURA 5 – ÁLGEBRA DE MAPAS	29
FIGURA 6 – EXEMPLO DE UM GRAFO COM FUNÇÃO DE CUSTO SOBRE AS ARESTAS (a) CONSTRUÇÃO DO GRAFO (b) DEFINIÇÃO DO TRAÇADO OTIMIZADO	30
FIGURA 7 – LEAST COST PATH APLICADO SOBRE UMA BASE RASTER	31
FIGURA 8 – FASES E DADOS NECESSÁRIOS	33
FIGURA 9 – ESQUEMATIZAÇÃO DO MODELO DE CAMPOS (2014)	39
FIGURA 10 – NEURÔNIO ARTIFICIAL	43
FIGURA 11 – ARQUITETURA DE UMA RNA MULTICAMADAS	44
FIGURA 12 – PROCESSO DE APRENDIZAGEM	49
FIGURA 13 – PRINCÍPIO DO ALGORITMO DE BACKPROPAGATION EM UMA FUNÇÃO ERRO UNIVARIADA	51
FIGURA 14 – MATRIZ DE CONTIGÊNCIA PARA MODELOS DE CLASIFICAÇÃO	53
FIGURA 15 – CURVA ROC CALCULADA COM DIVERSOS CONJUNTOS DE TESTE	55
FIGURA 16 – ÁREA DE ESTUDO	59
FIGURA 17 – ARQUITETURA DO MÉTODO PROPOSTO	61
FIGURA 18 – FLUXO DO ALGORITMO DE APLICAÇÃO DA REDE NEURAL	65
FIGURA 19 – EXEMPLO DE MAPA DE PREDIÇÃO GERADO PELA RNA NEURALNET	69
FIGURA 20 – PREPARAÇÃO DOS DADOS	72
FIGURA 21 – PREPARAÇÃO DAS CAMADAS	73
FIGURA 22 – PREPARAÇÃO DAS AMOSTRAS	75
FIGURA 23 – DADOS VETORIAIS, IMAGENS RASTER E AS AMOSTRAS POSITIVAS E NEGATIVAS EMPREGADAS NO TREINAMENTO DA RNA. (A) RESULTADO DA CLASSIFICAÇÃO DA CAMADA USO DO	

SOLO. (B) CAMADA DE DISTÂNCIA À EMPREENDEMENTOS LINEARES. (C) AMOSTRA NEGATIVA. (D) AMOSTRA POSITIVA..	76
FIGURA 24 – RNA COM 5 NEURÔNIOS OCULTOS E 2 MIL AMOSTRAS	80
FIGURA 25 – RNA COM 10 NEURÔNIOS OCULTOS E 20 MIL AMOSTRAS.....	85
FIGURA 26 – CAMADAS DISCRETAS E CATEGÓRICAS (A).....	85
FIGURA 27 – CAMADAS DISCRETAS E CATEGÓRICAS (B).....	86
FIGURA 28 – CAMADAS DISCRETAS E CATEGÓRICAS (A).....	87
FIGURA 29 – CAMADAS DISCRETAS E CATEGÓRICAS (B).....	87
FIGURA 30 – GRÁFICO <i>PAIRWISE</i>	89
FIGURA 31 – GRÁFICO DE SENSIBILIDADE.....	90
FIGURA 32 – TAXA DE SUCESSO COM 20 MIL AMOSTRAS.....	92

LISTA DE TABELAS

TABELA 1 – CAPACIDADE INSTALADA DO SIN E PREVISTA PARA 2024	16
TABELA 2 – ÁREA PARA ANÁLISE DE INTERFERÊNCIAS DE EMPREENDIMENTOS LINEARES	17
TABELA 3 – FAIXA DE SEGURANÇA DE LINHAS DE TRANSMISSÃO DE ENERGIA	18
TABELA 4 – IMPACTOS CAUSADOS PELA CONSTRUÇÃO DE UMA LT	25
TABELA 5 – ESCALA DE PONDERAÇÃO DO MÉTODO AHP	26
TABELA 6 – ÍNDICE DE CONSISTÊNCIA ALEATÓRIA.....	32
TABELA 7 – ESCALA DE PONDERAÇÃO DO MÉTODO EPRI.....	34
TABELA 8 – COMPARAÇÃO ENTRE METODOLOGIAS EMPREGADAS EM ESTUDOS DE TRAÇADOS DE LINHAS DE TRANSMISSÃO DE ENERGIA.....	40
TABELA 9 – DADOS DE REFERÊNCIA E SUAS FONTES	60
TABELA 10 – DADOS DE ENTRADA, INFORMAÇÕES ADICIONAIS.....	62
TABELA 11 – PESOS CALCULADOS PARA 5 NEURÔNIOS OCULTOS	77
TABELA 12 – PESOS CALCULADOS PARA 10 NEURÔNIOS OCULTOS	81
TABELA 13 – MATRIZ DE CONFUSÃO PARA 2 MIL AMOSTRAS	91
TABELA 14 – MATRIZ DE CONFUSÃO PARA 20 MIL AMOSTRAS	92

LISTA DE ABREVIATURAS E SIGLAS

AHP: *Analytic Hierarchy Process* (Método de Análise Hierárquica);
ANA: Agência Nacional de Águas;
ANAC: Agência Nacional de Aviação Civil;
ANEEL: Agência Nacional de Energia Elétrica;
AUC: *Area Under Curve*;
BDGEx: Banco de Dados Geográficos do Exército Brasileiro;
BP: *Backpropagation*;
COPEL: Companhia Paranaense de Energia;
DNIT: Departamento Nacional de Infraestrutura de Transportes;
EMATER: Instituto Paranaense de Assistência Técnica e Extensão Rural;
EMBRAPA: Empresa Brasileira de Pesquisa Agropecuária;
FUNAI: Fundação Nacional do Índio;
IAP: Instituto Ambiental do Paraná;
IBAMA: Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis;
IBGE: Instituto Brasileiro de Geografia e Estatística;
INCRA: Instituto Nacional de Colonização e Reforma Agrária;
INDE: Infraestrutura nacional de Dados Espaciais;
INPE: Instituto Nacional de Pesquisas Espaciais;
IPHAN: Instituto do Patrimônio Histórico e Artístico Nacional;
ITCG: Instituto de Terras, Cartografia e Geologia do Paraná;
LCP: *Least Cost Path*;
LT: Linha de Transmissão;
MAM: Métodos de Análise Multicritério;
ONS: Operador Nacional do Sistema Elétrico;
RBP: *Resilient Backpropagation*;
RNA: Redes Neurais Artificiais;
ROC: *Receiver Operating Characteristic*;
SIG: Sistema de Informações Geográficas;
SIGEF: Sistema de Gestão Fundiária;
SIN: Sistema Interligado Nacional;

SUMÁRIO

1 INTRODUÇÃO	14
1.1 CONSIDERAÇÕES INICIAIS	14
1.2 CONTEXTUALIZAÇÃO DO PROBLEMA E MOTIVAÇÃO	16
1.3 OBJETIVOS	18
2 REVISÃO DE LITERATURA	20
2.1 SISTEMA DE TRANSMISSÃO DE ENERGIA ELÉTRICA NO BRASIL	20
2.1.1 Conceitos básicos e principais componentes de uma LT	22
2.1.2 Desafios dos estudos de traçados de uma LT e aspectos de planejamento	22
2.1.3 Impactos na construção de uma nova LT	24
2.2 ESTADO DA ARTE EM ANÁLISE DE TRAÇADOS DE LINHAS DE TRANSMISSÃO DE ENERGIA	25
2.2.1 Metodologias para a análise de traçados de LT	25
2.2.1.1 Metodologia utilizando o método AHP	26
2.2.1.2 Metodologia Electric Power Research Institute – EPRI	32
2.2.1.3 Metodologia utilizando modelagem ambiental	36
2.2.2 Comparação entre os métodos existentes	40
2.3 REDES NEURAIS ARTIFICIAIS	40
2.3.1 Introdução às Redes Neurais Artificiais	41
2.3.2 Arquiteturas de rede e aprendizagem	43
2.3.2.1 Tipo de tarefas de aprendizagem	46
2.4 NEURALNET	47
2.4.1 Introdução	47
2.4.2 Arquitetura multicamadas	48
2.4.3 Aprendizagem supervisionada	49
2.4.4 Backpropagation e Resilient Backpropagation	50
2.5 MÉTODO DE AVALIAÇÃO DO MODELO FUNCIONAL TREINADO	52
2.5.1 Probabilidade conjunta e condicional	52
2.5.2 Avaliação do modelo	54
2.5.2.1 A curva ROC	55
2.5.2.2 Cálculo da AUC	56
3 MATERIAIS E MÉTODO	58
3.1 MATERIAS	58

3.1.1 Área de estudo	58
3.1.2 Dados de referência utilizados no estudo.....	60
3.2 MÉTODO PROPOSTO	61
3.2.1 Homogeneização dos formatos de dados de entrada e cálculo das estatísticas descritivas	62
3.2.2 Amostragem dos dados de treinamento e de teste	64
3.2.3 Aplicação da RNA	64
3.2.3.1 Normalização dos dados numéricos de entrada e adaptação dos dados categóricos	65
3.2.3.2 Treinamento da RNA.....	66
3.2.3.3 Validação do algoritmo resultante da RNA.....	68
3.2.3.4 Geração do mapa preditivo das áreas propícias à construção de novas linhas de transmissão.....	68
3.2.4 Otimização do modelo predito.....	70
3.2.5 Geração da diretriz de traçado otimizado de uma LT.....	71
3.2.6 Comparação de rotas alternativas.....	71
3.3 PREPARAÇÃO DOS DADOS DE ENTRADA.....	71
3.3.1 Preparação das camadas covariáveis.....	72
3.3.2 Preparação das amostras de treinamento e teste.....	74
4 EXPERIMENTOS E ANÁLISE DOS RESULTADOS	76
4.1 PESOS CALCULADOS.....	77
4.1.1 Utilizando 5 neurônios ocultos.....	77
4.1.2 RNA utilizando 10 neurônios ocultos.....	80
4.2 ANÁLISE DAS COVARIÁVEIS DE ENTRADA.....	85
4.2.1 Treinamento utilizando 2 mil amostras	85
4.2.2 Treinamento utilizando 20 mil amostras	86
4.3 TAXA DE SUCESSO.....	91
4.3.1 Treinamento utilizando 2 mil amostras	91
4.3.2 Treinamento utilizando 20 mil amostras	92
5 CONCLUSÕES E RECOMENDAÇÕES	94
5.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS	96
REFERÊNCIAS.....	98

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

O desenvolvimento social e econômico de uma nação tem uma forte relação com a expansão do setor elétrico, o que torna a energia elétrica e suas diversas aplicações indispensáveis à sociedade moderna. O progresso de uma região é acompanhado pelo aumento na demanda por eletricidade, tanto pelos fornecedores, que necessitam de mais energia para poderem produzir, quanto por parte dos consumidores, que adquirem mais produtos como eletrodomésticos e eletroeletrônicos (MENEZES, 2015).

Por isso, para atender à crescente demanda e assegurar a disponibilidade dos recursos energéticos e garantir as demandas futuras, é necessário construir novas linhas de transmissão (LT) e novas subestações no território nacional. O planejamento do traçado de diretrizes de LT (definição traçado inicial e experimental) envolve o conhecimento prévio de fatores restritivos presentes na área de instalação do empreendimento (FREIMAN, 2019).

A definição da melhor rota de deslocamento sobre uma área é um dos mais antigos problemas de otimização tratados na literatura e também, pode-se dizer, da Engenharia. Deslocar-se, levar avanços tecnológicos e econômicos de uma área já antropizada para outra área qualquer, não é uma tarefa simples. Exemplos como obras lineares de engenharia tais como oleodutos, gasodutos, estradas, ferrovias, canais e linhas de transmissão de energia possuem alto impacto ambiental, social e econômico. Impactos que podem ser benéficos, os quais devem ser maximizados e outros não-benéficos, que devem ser evitados ou minimizados.

Para o estabelecimento das rotas ótimas passam pelo estabelecimento de pesos para cada fator envolvido na análise. Os pesos são na prática uma ponderação dos fatores que são utilizados para descrever um fenômeno ou uma relação matemática entre dois conjuntos de dados. Essas ponderações são muito valiosas e tem grande aplicação na Engenharia e em problemas reais, isto é, saber a relação entre um fenômeno e seus condicionantes pode ser uma questão determinante para a viabilidade ou redução significativa de custos de uma obra, por exemplo.

A atribuição de pesos à fatores condicionantes em obras de engenharia são feitos na maioria das vezes por especialistas com amplo conhecimento e experiência

sobre o fenômeno estudado. Contudo, esta análise se torna subjetiva individual de cada profissional. Automatizar a relação entre variáveis resposta e suas camadas de maneira eficaz, torna-se assim uma tarefa essencial para o desenvolvimento sustentável dado o seu potencial de minimizar impactos ambientais e seguir levando em conta tanto fatores econômicos quanto sociais.

Paralelo a isso, com o desenvolvimento da tecnologia e o surgimento dos computadores, surgiram os primeiros conceitos de Inteligência Artificial (IA) com os estudos do inglês Alan Turing na década de 1950. Desde então, houve enormes avanços na área da informática e as IAs se tornaram um campo do conhecimento presentes em situações do cotidiano, nas decisões que são tomadas e na maneira de interagir com o mundo.

A IA pode auxiliar na automação de tarefas diárias, solucionar problemas complexos e analisar grandes volumes de dados que, geralmente, demanda uma grande quantidade de tempo. Com isso, inúmeros pesquisadores vêm se dedicando ao desenvolvimento de técnicas de aprendizagem de máquina através do treinamento de redes neurais artificiais, que são inspiradas no funcionamento do cérebro humano. A finalidade é desenvolver sistemas especializados para a realização de determinadas tarefas, usadas não só em tarefas simples do dia-a-dia, mas também em quase todas as áreas das ciências aplicadas, sendo uma tecnologia com potencialidade ainda a ser exploradas, principalmente, em empreendimentos lineares.

Tendo todas essas questões em mente relacionadas à otimização de projetos para novas linhas de transmissão, neste trabalho é proposto o emprego dos conceitos da IA através das redes neurais artificiais no problema de traçado de diretrizes de linhas de transmissão. Esta abordagem pode fornecer além da automação do processo de atribuição de pesos subjetivos para as camadas restritivas, um sistema especialista que indica através de um mapa preditivo, as áreas com maior aptidão de se construir novas linhas de transmissão. Isto pode ser feito por meio de aprendizagem com as LTs já existentes em uma área teste. Vale ressaltar que este estudo é parte da etapa de otimização do traçado automático de diretrizes de LTs do Projeto PD-06491-0353/2014, PROGRAMA P&D – Copel Geração e Transmissão S.A.

1.2 CONTEXTUALIZAÇÃO DO PROBLEMA E MOTIVAÇÃO

Segundo os relatórios do Operador Nacional do Sistema Elétrico - ONS, só no Paraná há a previsão de 19 novos planos de obras para expansão das linhas de transmissão de energia do Sistema Interligado Nacional - SIN. Para exemplificar a demanda crescente por energia, a TABELA 1 apresenta dos dados da capacidade da infraestrutura nacional instalada no país em 2020 e uma previsão para os próximos 4 anos.

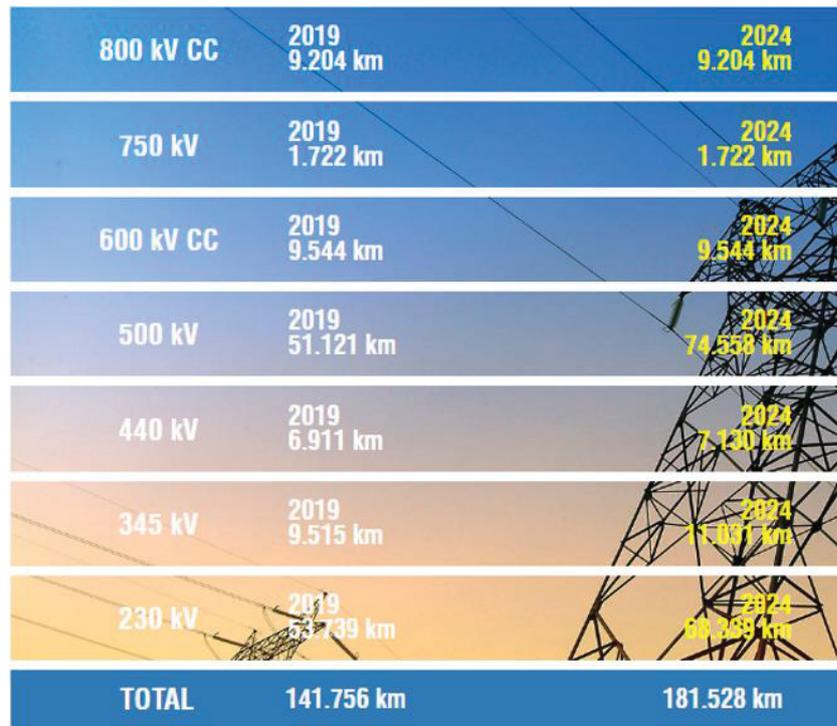
TABELA 1 – CAPACIDADE INSTALADA DO SIN E PREVISTA PARA 2024

Fonte de energia	% Nacional em 2020	Capacidade instalada (MW)	% Nacional prevista	Capacidade instalada prevista (MW)
Hidrelétrica	65,9	108.400	62,0	109.218
Eólica	9,2	15.092	11,1	19.627
Term. Gás + GNL	8,6	14.208	10,3	18.176
Biomassa	8,3	13.654	8,3	14.549
Term. Óleo + Diesel	2,7	4.404	2,7	4.692
Solar	1,8	3.033	2,4	4.280
Term. Carvão	1,8	3.017	1,7	3.017
Nuclear	1,2	1.990	1,1	1.990
Outra	0,4	590	0,4	745
Total	100	164.388	100	176.294

FONTE: ONS (2020).

A FIGURA 1 complementa a informação apresentada na TABELA 1, dando o dimensionamento físico da rede de transmissão básica instalada no país no ano de 2019 (dados mais recentes disponibilizados) e também a previsão para o ano de 2024.

FIGURA 1 – EXTENSÃO DA REDE BÁSICA DE TRANSMISSÃO E EXTENSÕES PREVISTAS



FONTES: ONS (2020).

Conforme a Portaria nº 419/2011, a abrangência dos estudos relacionados aos tipos de empreendimentos lineares e suas respectivas restrições pode ser observada na TABELA 2.

TABELA 2 – ÁREA PARA ANÁLISE DE INTERFERÊNCIAS DE EMPREENDIMENTOS LINEARES

Tipo de empreendimento	Largura da faixa de abrangência	
	Amazônia Legal	Demais Regiões
Empreendimentos lineares (exceto rodovias)		
Ferrovias	10 km	5 km
Dutos	5 km	3 km
LTs	8km	5 km

FONTES: Adaptado da Portaria Interministerial nº 419/2011 (2011).

As distâncias de segurança (horizontal e vertical) para a operação de uma LT são definidas em função da tensão transmitida, de acordo com a NBR 5422, de 1985. Com relação à faixa de segurança horizontal de operação de uma LT, as distâncias mais usuais adotadas por empresas que gerenciam o setor de transmissão de energia podem ser vistas na TABELA 3.

TABELA 3 – FAIXA DE SEGURANÇA DE LINHAS DE TRANSMISSÃO DE ENERGIA

Tensão (kV)	Largura da faixa de segurança
70	12 m – 20 m
140	16 m – 20m
230	30 m – 40 m
345	40m - 50 m
400	50 m – 60 m
500	60 m

FONTE: Companhia de Transmissão de Energia Elétrica Paulista - CTEEP (2015).

Analisando os dados da ONS, a previsão de aumento do Sistema Interligado Nacional é de cerca de 7,2% da capacidade e 28% em extensão. Isso significa uma previsão de construção de aproximadamente 39,8 mil quilômetros de novas LTs com tensões entre 230 kV e 500 kV no país. Levando em consideração a largura da faixa de abrangência de uma nova obra que é de 5 km (segundo legislação vigente), têm-se uma área potencial indireta a ser afetada de 199 mil km² e direta de mais de 2 mil km². Por isso, realizar estudos que possibilitem a minimização de impactos ambientais, culturais e socioeconômicos nesta área potencial é de fundamental importância para o desenvolvimento da sociedade, uma vez que além de atender a demanda crescente, fará isso com responsabilidade social, ambiental e financeira.

Além do potencial uso do método proposto neste trabalho, há também outra importante contribuição nesta pesquisa: a atribuição automática de pesos para cada variável considerada no processo. Tradicionalmente, a avaliação do peso dado às camadas geralmente é feita de maneira empírica ou mesmo arbitrária.

A aplicação de aprendizado de máquina em um problema de engenharia pode se refletir na otimização de recursos e de tempo na construção de novas obras de LTs, dado que tem capacidade de avaliar múltiplas variáveis de forma rápida e eficiente (o que demandaria muito tempo, caso feito por seres humanos) acelerando assim o processo de expansão. Somado a isso, a abordagem utilizando redes neurais artificiais na resolução de problemas relacionados a traçados de novas LTs é pioneira na literatura, podendo assim gerar novas pesquisas no futuro.

1.3 OBJETIVOS

O objetivo geral deste trabalho é desenvolver um método automático para ponderação das camadas e de restrições em análise de diretrizes de traçados para construção de novas linhas de transmissão de energia elétrica, sem a subjetividade e

a interferência humana no processo, a fim de se obter um esboço de um sistema especializado. Para alcançar o objetivo desta pesquisa são definidos como específicos as seguintes metas, a saber:

- A. Adaptar uma rede neural artificial para o aprendizado das contribuições das camadas consideradas em um estudo de traçado, em relação à LTs já existentes;
- B. Calcular, automaticamente, os pesos de cada covariável;
- C. Gerar uma diretriz de traçado desta nova obra a partir de uma superfície probabilística;
- D. Avaliar os resultados obtidos e discutir a potencialidade da modelagem empregada para a definição de diretrizes de traçado de LTs.

2 REVISÃO DE LITERATURA

2.1 SISTEMA DE TRANSMISSÃO DE ENERGIA ELÉTRICA NO BRASIL

Segundo o Operador Nacional do Sistema Elétrico – ONS, o Brasil conta com uma grande extensão de redes de LT, denominada como Sistema Interligado Nacional (SIN). O SIN é subdividido em regiões ou subsistemas: Sul, Sudeste/Centro-Oeste, Nordeste e Grande parte da região Norte. Esta divisão pode ser observada na FIGURA 2. Além dos 4 subsistemas há sistemas isolados, ou seja, que não possuem conexão com essas regiões, localizados na sua totalidade na região Norte do país. A causa pode ser explicada com as características naturais locais, tal como florestas densas e cidades e povoados afastados ou a obstáculos como rios de grande porte, como por exemplo.

FIGURA 2 – SUBSISTEMAS DO SIN

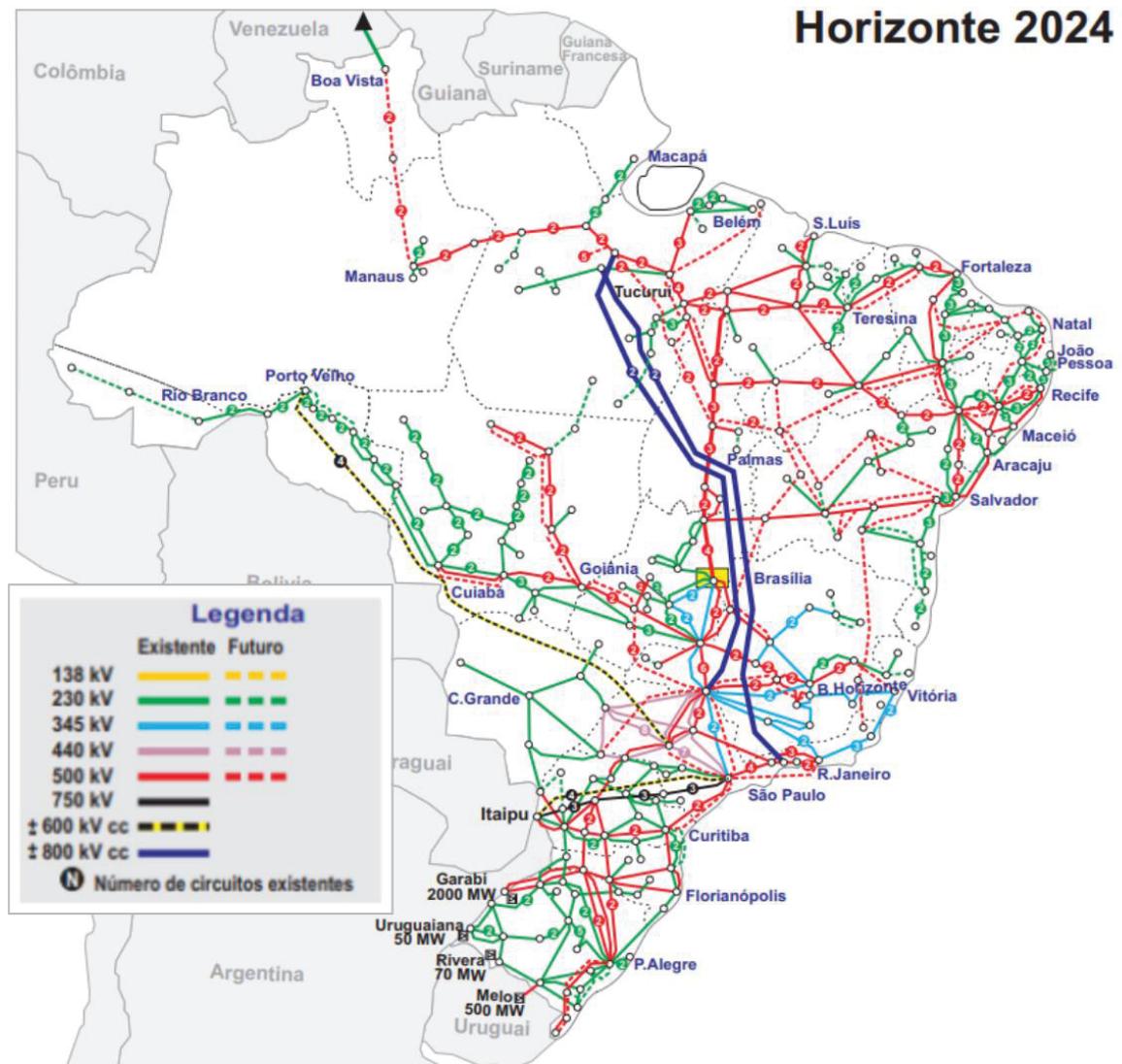


FONTE: Freiman (2019).

A interconexão dos sistemas elétricos, por meio da malha de transmissão, proporciona a transferência energética entre subsistemas e como isso permite a obtenção de ganhos energéticos e explora a diversidade e potencial dos regimes

hidrológicos das bacias em território nacional. A extensão total da malha de transmissão soma aproximadamente 142 mil quilômetros, com capacidade entre 800 KW e 230 KW ONS (2020). A FIGURA 3 apresenta a rede básica de transmissão do Brasil e suas respectivas capacidades energéticas.

FIGURA 3 – REDE DE TRANSMISSÃO DO BRASIL



Segundo dados da ONS, já apresentados na TABELA 1, atualmente cerca de 66% da energia produzida no Brasil e que alimenta o SIN é oriunda de usinas hidrelétricas, distribuídas em 16 bacias hidrográficas nas diferentes regiões geográficas do Brasil. Hidrelétricas, seguido de energia eólica, térmicas a gás e

biomassa são as mais relevantes atualmente com relação ao suprimento da demanda do SIN, totalizando juntas cerca de 92% da energia total produzida.

2.1.1 Conceitos básicos e principais componentes de uma LT

A tensão de uma LT é determinada de acordo com a potência a ser transmitida para uma determinada área a ser suprida. A energia é transmitida em níveis elevados de tensão, que é justificado pela capacidade de transmissão de uma linha, que aumenta proporcionalmente ao quadrado da tensão (KIESSLING et al., 2003 *apud* LIMA, 2015). No Brasil são consideradas tensões de transmissão valores acima de 230kV, de acordo com a Resolução Normativa ANEEL n° 67 de 2004 (ANEEL, 2004 *apud* LIMA, 2015).

Segundo Lima (2015), uma LT é composta basicamente de torres, cabos condutores, cadeias de isoladores e ancoragem, sendo os últimos com a função de fixação e estabilização dos condutores nas torres. Diversos são os tipos de torres, fundação, e demais componentes de uma linha de transmissão, que são dimensionados de acordo com as características elétricas como tensão, potência, e também características locais como relevo, estabilidade do solo, presença de nível de água superficial, dentre outras características.

2.1.2 Desafios dos estudos de traçados de uma LT e aspectos de planejamento

As ciências geodésicas e as geotecnologias constituem um papel de grande importância na construção de novas linhas de transmissão de energia. Um projeto pré-estabelecido de forma pensada e bem executado permite um melhor aproveitamento da terra utilizada para a passagem da linha entre as unidades geradoras e as unidades consumidoras, minimizando assim as perdas ambientais, sociais e econômicas públicas, uma vez que grande parte das empresas que operam na geração e distribuição são governamentais.

Os profissionais da área atuam no estudo do traçado, no cadastro fundiário das propriedades afetadas pela construção de uma nova linha (levantamento, registro e averbação) e na topografia para a construção (levantamentos aerofotogramétricos, perfilhamentos utilizando a tecnologia LiDAR do corredor de implantação. Outras atividades como laudos de avaliações de propriedades, topografias de subestações e

gerenciamento de obras em campo também são atividades que podem ser desenvolvidas por engenheiros desta área.

Para o primeiro dos passos, o estudo do traçado, não é uma atividade simples pois pode exigir muitos estudos preliminares e atividades em campo e é de fundamental importância para estudos específicos, identificando as áreas com restrições legais que impedem a construção ou com maior propensão a conflitos. Este estudo deve ser minucioso e proporciona conhecimento do meio físico, biótico e antrópico, buscando o menor custo social, ambiental e econômico para o empreendimento.

O estudo de traçado utiliza imagens aéreas obtidas através de sensores aéreos ou orbitais e levantamentos de órgãos oficiais tais como: INCRA, FUNAI, DNIT, ANAC, IBAMA, IPHAN, ANEEL, prefeituras e governos estaduais, entre outros que possuem dados da infraestrutura e de fenômenos que estão dentro do corredor potencial para construção, a fim de compatibilizar e apresentar alternativas que minimizem os custos somados.

Segundo Laureano (2019), dentre os desafios da definição do estudo de traçados de LT, pode-se citar:

- A. Evitar deflexões e diminuir a extensão: a extensão reduz a quantidade de cabos, torres e materiais de serviço e construção associados. Deflexões fortes (ângulos muito agudos) necessitam, por consequência, de estruturas reforçadas tanto nas torres quanto nas suas fundações;
- B. Evitar travessias de percursos: evitar ou minimizar o cruzamento dos cabos com grandes obstáculos naturais ou artificiais como rios, grandes interseções de ruas ou rodovias, ferrovias e áreas alagadas, uma vez que estas áreas geram trâmites burocráticos que demandam tempo, além de gastos maiores com a estrutura. Por outro lado, há autores que mencionam que a construção de novas LTs próximas à rodovia e/ou ferrovia facilita o transporte de materiais, equipamentos trabalhadores para os canteiros de obras. Além do mais, ao construir paralelamente a outras linhas de transmissão existentes ou projetos lineares pode-se aproveitar a faixa de domínio e acessos construídos;
- C. Solo e relevo ideais: optar por solos apropriados podem resultar em uma economia significativa no final do projeto, assim como evitar torres

com alturas muito elevadas ou vão de comprimento reduzidos evitam gastos desnecessários com material na construção;

- D. Minimizar impactos na flora: durante a construção pode haver impacto sobre a vegetação através do desmatamento, limpeza de áreas para montagem de torres e construção de praças de lançamentos de cabos. Por isso deve-se adotar medidas como corte seletivo (remoção apenas de árvores de porte maior) ou construir torres de sustentação mais altas;
- E. Minimizar impactos na fauna: conseqüentemente ao impacto na vegetação, a fauna também é afetada, uma vez que pode fragmentar o habitat, causar redução do número de animais e mudança na estrutura das comunidades. Além do mais, aumenta o risco de colisão de aves com cabos das torres das torres de transmissão, principalmente nas espécies migratórias em áreas de travessias de rios e alagadiças;
- F. Evitar Áreas de Proteção Permanentes (APPs): os traçados devem garantir o distanciamento de APPs e de regiões com fragmentos de vegetação nativa;
- G. Levantamento fundiário: cadastrar as propriedades abrangidas na faixa de estudo é essencial para avaliar e minimizar a burocracia em áreas especiais como terras indígenas, aeroportos, assentamentos e unidades de conservação por exemplo, estudando a situação cartorial e jurídica dos locais;

2.1.3 Impactos na construção de uma nova LT

De acordo com Freiman (2019), para a definição das diretrizes de uma nova LT, os meios físico, biótico, socioeconômico e cultural devem ser considerados, assim como levantamentos aspectos meteorológicos e geotécnicos, visando escolher a alternativa que minimize os impactos e custos associados. A TABELA 4 resume os impactos citados pelo autor em seu estudo.

TABELA 4 – IMPACTOS CAUSADOS PELA CONSTRUÇÃO DE UMA LT

Meio	Impactos
Físico	Erosão do solo Interferência com recursos hídricos Interferência com áreas legalmente protegidas
Biótico	Retirada de cobertura vegetal Interferência na fauna e na flora Interferência na rota de migração de pássaros
Socioeconômico e cultural	Degradação da paisagem Desapropriações de terras Limitação do uso do solo (servidões) Deslocamento nas populações afetadas Interferência nas atividades agropecuárias Influência em populações indígenas/quilombolas Interferências em locais de interesse histórico e cultural

FONTE: Freiman (2019).

A seguir serão abordadas as metodologias utilizadas para analisar traçados de LT de energia presentes na literatura.

2.2 ESTADO DA ARTE EM ANÁLISE DE TRAÇADOS DE LINHAS DE TRANSMISSÃO DE ENERGIA

2.2.1 Metodologias para a análise de traçados de LT

As metodologias para a análise multicriterial de traçados, consolidadas classicamente e empregadas até os dias atuais, utilizam-se de ferramentas de Sistemas de Informações Geográficas (SIG) como auxílio na tomada de decisões aliados com Métodos de Análise Multicritério (MAM). Uma das principais diferenças entre as abordagens está relacionada com o tipo de MAM utilizado: *Analytic Hierarchy Process - AHP* definido por Saaty (1982), método *Elimination et Choix Traudisant la Réalité – ELECTRE* definido por Roy (1968) ou método *Preference Ranking Organization Method Enrichment Evaluations - PROMETHEE* definidos por Brans e Vincke (1985) (FREIMAN, 2019).

A outra diferença relaciona-se nos tipos de variáveis ou camadas utilizadas no estudo como a proximidade de áreas urbanas (GRASSI et al, 2014), áreas de proteção ambiental, degradação da paisagem (BAGLI et al; 2011; GRASSI et al, 2014), limitações de engenharia (MONTEIRO et al, 2005) e custos associados à construção do empreendimento (SHU et al, 2012; LIMA et al, 2016) e velocidade do vento (FREIMAN, 2019).

O ELECTRE baseia-se na concordância e na medição de vantagens relativas entre variáveis (POHEKAR, RAMACHANDRAN; 2003 *apud* FREIMAN, 2019). Já o PROMETHEE baseia-se na relação binária entre critérios atribuindo a cada um deles um peso relativo à sua importância quanto à otimização (BRANS e VINCKE, 1985 *apud* FREIMAN, 2019). O método AHP é o mais amplamente utilizado para a finalidade de estudos de traçados de empreendimentos lineares e será abordado detalhadamente na seção 2.2.1.1.

2.2.1.1 Metodologia utilizando o método AHP

A metodologia AHP estrutura o problema de otimização em uma hierarquia de importância a partir da realização de comparações pareadas e possibilita a análise de consistência das ponderações atribuídas a cada critério pelo analista (FREIMAN, 2019). Apesar de haver uma hierarquização, na prática há uma homogeneização das dimensões dos arquivos *raster* das variáveis envolvidas no processamento, porém a escala dos dados inseridos não é compatibilizada. Isso significa que a análise é feita em múltiplas escalas, relacionando dados nas superfícies de custo e adequabilidades em um único ciclo, assumindo a menor escala como a precisão do traçado final.

Freiman (2019) também explica que para o cálculo dos pesos para os diferentes critérios é realizado a partir da criação de uma matriz A de comparação pareada de dimensão $m \times m$, em que m é o número de critérios a serem avaliados. Cada componente de A (a_{ij}) representa a importância do critério j em relação ao critério i . A análise da adequabilidade dos critérios é realizada a partir de uma escala numérica entre 1 a 9 proposta por Saaty (1980), conforme apresentado na TABELA 5. Estes valores são atribuídos na matriz A de acordo com a concepção do analista em relação ao problema de otimização, subjetivamente.

TABELA 5 – ESCALA DE PONDERAÇÃO DO MÉTODO AHP

Intensidade de importância	Definição	Explicação
1	Mesma importância	As duas variáveis contribuem igualmente para o objetivo.
3	Importância pequena de uma sobre a outra	A experiência e o julgamento favorecem levemente uma atividade em relação a outra
5	Importância grande ou essencial	A experiência e o julgamento favorecem fortemente uma atividade em relação à outra

Intensidade de importância	Definição	Explicação
7	Importância muito grande ou demonstrada	Uma atividade é muito fortemente favorecida em relação à outra.
9	Importância absoluta	A evidência favorece uma atividade em relação à outra com o mais alto grau de certeza.
2,4,6,8	Valores intermediários	Quando se procura uma condição de compromisso entre duas definições

FONTE: Saaty (1980).

Saaty (1980) recomenda que para atribuir os pesos relativos às variáveis observadas, é necessário normalizar a matriz A , dividindo o valor numérico de cada elemento de A pela soma dos valores da coluna a qual pertence, a fim de se ter valores entre 0 e 1, conforme a EQUAÇÃO (1).

$$\bar{a}_{ij} = \frac{a_{ij}}{\sum_{ij=1}^m a_{ij}} \quad (1)$$

Os pesos normalizados de cada critério são calculados através do vetor de prioridade W , como mostrado na EQUAÇÃO (2).

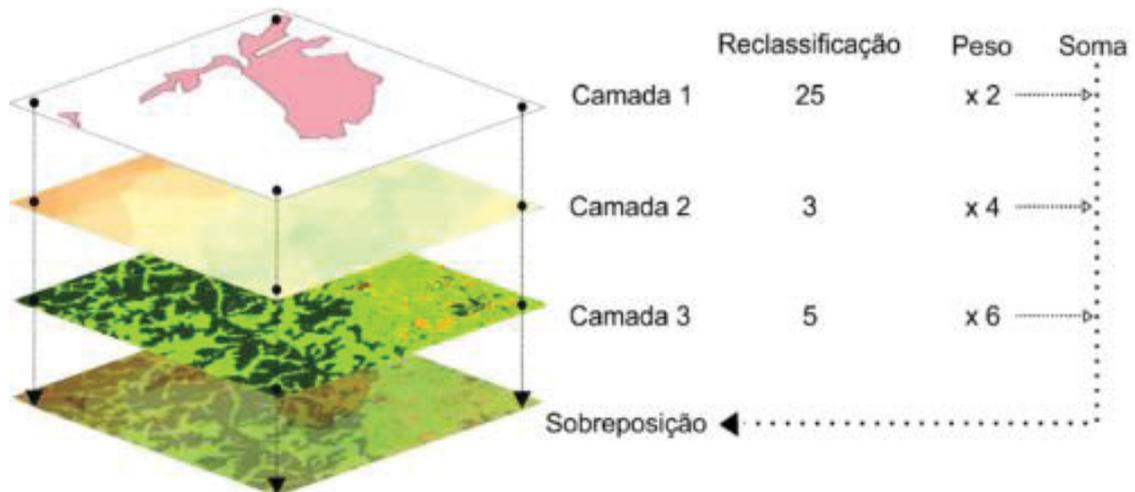
$$W = \frac{\sum_{ij=1}^m \bar{a}_{ij}}{m} \quad (2)$$

Concluídas o cálculo dos pesos, faz-se a reclassificação.

2.2.1.1.1 Superfícies de custo e de adequabilidade

Os MAMs são utilizados para a reclassificar os atributos internos de uma base de dados cartográficos a partir de ferramentas de SIG. Nos processos de modelagem espacial multicritério, a organização é composta por pixels georreferenciados, como unidades mínimas de área passível de análise. Isso se deve ao fato de que o formato matricial simplifica o processo de ponderação e por isso, as camadas vetoriais são convertidas para matriciais e após isso, são reclassificadas de acordo com o peso de cada variável e combinadas a fim de gerar as chamadas “superfícies de custo”, conforme o esquema apresentado na FIGURA 4.

FIGURA 4 – SUPERFÍCIE(S) DE CUSTO(S)



FONTE: Freiman (2019).

A associação de todas as camadas da análise de traçados em uma única superfície (S), ou superfície de adequabilidade, e a geração das superfícies de custo são realizadas a partir do processo conhecido como álgebra de mapas. Segundo Tomlin (1990), o termo “álgebra de mapas” é utilizado para indicar o conjunto de procedimentos de análise espacial em Geoprocessamento que produz novos dados, a partir de funções manipulação aplicadas a um ou mais mapas; usando um conjunto de operações matemáticas sobre mapas, analogamente aos ambientes de álgebra e estatística tradicional. Os mapas são tratados como variáveis individuais, e as funções definidas sobre estas variáveis são aplicadas de forma homogênea a todos os pontos do mapa. A FIGURA 5 mostra um exemplo de um resultado de uma operação de adição utilizado entre uma camada de uso do solo e outra de APP.

O modelo matemático mais utilizado para a geração da superfície de adequabilidade é o da combinação linear ponderada, como expressa na EQUAÇÃO (3).

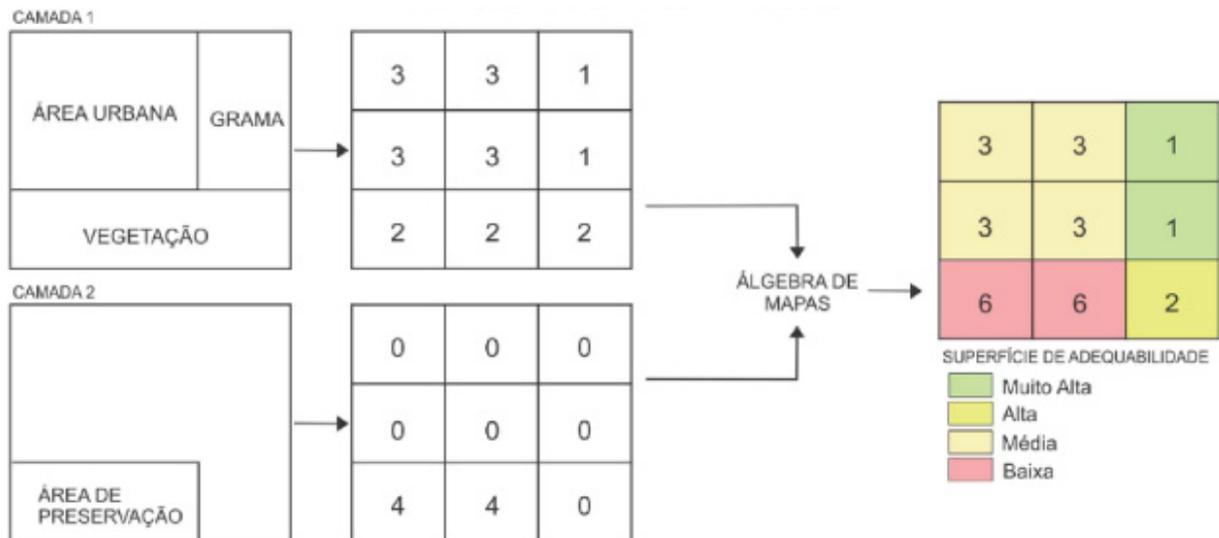
$$S = \sum_1^n W_i x_i, \quad \text{com } 1 \leq x_i \quad (3)$$

Onde:

W_i : é o peso atribuído a variável cartográfica;

x_i : é o valor normalizado da variável.

FIGURA 5 – ÁLGEBRA DE MAPAS



FONTE: Freiman (2019).

2.2.1.1.2 Least Cost Path

O método *Least Cost Path* – *LCP* amplamente utilizado em estudos de traçados é baseado no algoritmo baseado no trabalho de Dijkstra (1959), e segundo Taha (2008) *apud* Freiman (2019) a metodologia se aplica a grafos orientados ou não. O algoritmo admite que todas as arestas possuem pesos não negativos e possui uma premissa que exista pelo menos um caminho possível entre o ponto de partida e o ponto de destino, otimizando o custo em múltiplas iterações.

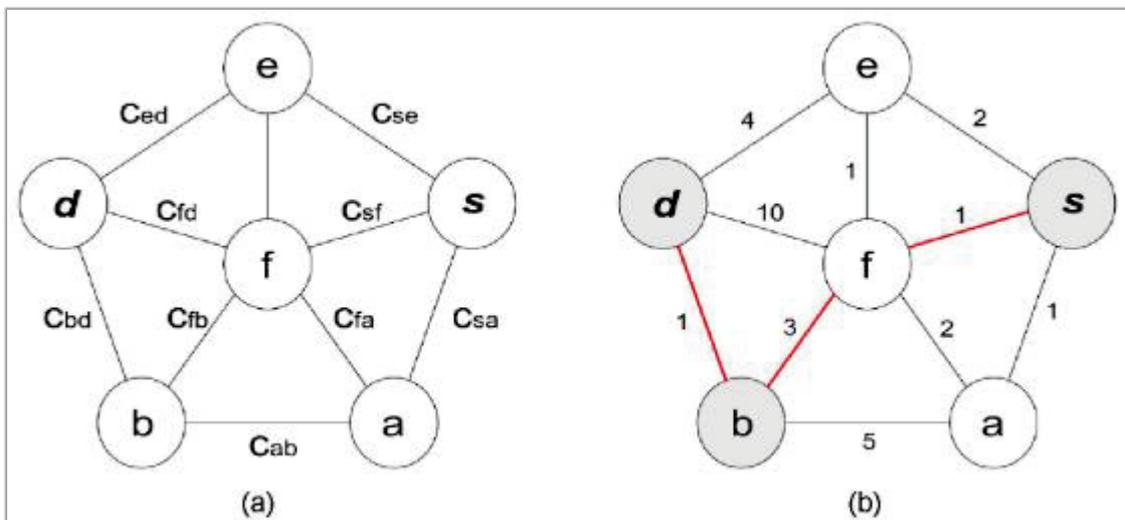
Rees (2004) *apud* Sari e Sen (2017) descreve as etapas do LCP de Dijkstra como:

- Atribuir um valor de custo zero para a célula alvo inicial;
- Identificar todas as células vizinhas à célula alvo e colocá-las na lista de células "ativas". Para cada uma destas células, calcular e atribuir o custo de alcançar a célula de destino, e atribuir um ponteiro que aponte para a célula de destino;
- Encontre a célula na lista que tem o menor custo, nomeie esta célula como *C* e o custo como *k*;
- Determine todas as células vizinhas de *C* como *S*. Para cada célula *C'* em *S*, calcule o custo *l* da mudança para *C*.
 - Se *C* não estiver incluído na lista, adicione-o à lista com um custo $k + l$ e um ponteiro que aponte para *C*;

- b. Se C já é membro da lista, compare o valor de $k + l$ com o custo provisório desta célula. Se $k + l$ for maior ou igual ao custo provisório, não faça nada. Entretanto, se $k + l$ for menor que o custo provisório, altere os atributos da célula C para que seu custo seja agora $k + l$ e seu ponteiro aponte agora para a célula C .
- E. Mudar os atributos da célula C de provisória para definitiva, e removê-la da lista;
- F. Repetir o passo D e E até que a lista esteja vazia.

Freiman (2019) elaborou uma esquematização para explicar o algoritmo em termos de grafos com função de custos sobre as arestas. Na FIGURA 6 busca-se definir o caminho mínimo entre o vértice s (origem) e o vértice d (destino), onde C_{ij} representa o custo de atravessamento contidos em cada aresta que liga nós adjacentes (i e j).

FIGURA 6 – EXEMPLO DE UM GRAFO COM FUNÇÃO DE CUSTO SOBRE AS ARESTAS (a) CONSTRUÇÃO DO GRAFO (b) DEFINIÇÃO DO TRAÇADO OTIMIZADO

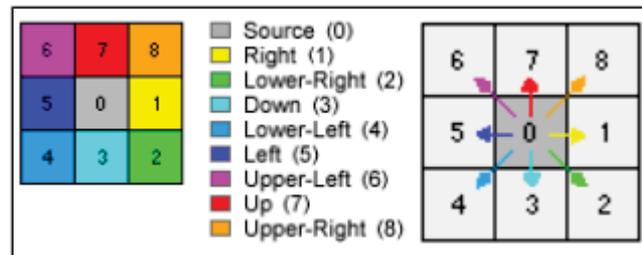


FONTE: Freiman (2019).

Na primeira iteração, o algoritmo designa o rótulo permanente $[0]$ a s . Na segunda iteração os vértices a , e e f são rotulados como temporários e são analisados os valores associados as arestas C_{sa} , C_{se} e C_{sf} . Para os três vértices temporários, f apresenta o menor custo de atravessamento, assim o rótulo de f muda para

permanente (b). Com base no vértice f os vértices e , b e d podem ser alcançados e são classificados como temporários. Esse processo se repete até que o vértice de destino (d) seja atingido, finalizando o processo de otimização.

FIGURA 7 – LEAST COST PATH APLICADO SOBRE UMA BASE RASTER



FONTE: Sari e Sen (2017).

Utilizando este algoritmo em uma base matricial, como por exemplo, em uma camada de custos acumulados, cada vértice do grafo seria um dos 8 pixels adjacentes ao pixel rotulado como permanente. O caminho mínimo final calculado sobre um *raster* conforme apresentado na FIGURA 8, partindo do pixel central o caminho escolhido pelo algoritmo seria do pixel adjacente à direita (amarelo).

2.2.1.1.3 Análise da consistência

Saaty (1980) descreve a metodologia para analisar as inconsistências do processo de ponderação dos critérios avaliados. Para a análise da consistência das avaliações realizadas pelo analista responsável pelo geoprocessamento, utiliza-se o índice de consistência (CI), conforme a EQUAÇÃO (4).

$$CI = \frac{x - m}{m - 1} \quad (4)$$

Onde:

x : é o somatório do produto de cada elemento do vetor de prioridade pelo total da respectiva coluna da matriz comparativa (A);

m : é o número de critérios avaliados.

Para verificar então se o *CI* encontrado por (4), calcula-se a taxa de consistência (*CR*), determinada pela razão entre o valor de *CI* e o índice de consistência aleatória (*RI*), conforme a EQUAÇÃO (5).

$$CR = \frac{CI}{RI} \quad (5)$$

Os valores de *RI* são apresentados na TABELA 6.

TABELA 6 – ÍNDICE DE CONSISTÊNCIA ALEATÓRIA

m	1	2	3	4	5	6	7	8	9	10
<i>RI</i>	0,00	0,00	0,58	0,9	1,12	1,24	1,32	1,41	1,45	1,49

FONTE: Saaty (1980).

Segundo Saaty (1980), consideram-se adequadas as ponderações quando $CR < 0,10$. O valor da *RI* tem como base o número ($m < 10$) de critérios avaliados.

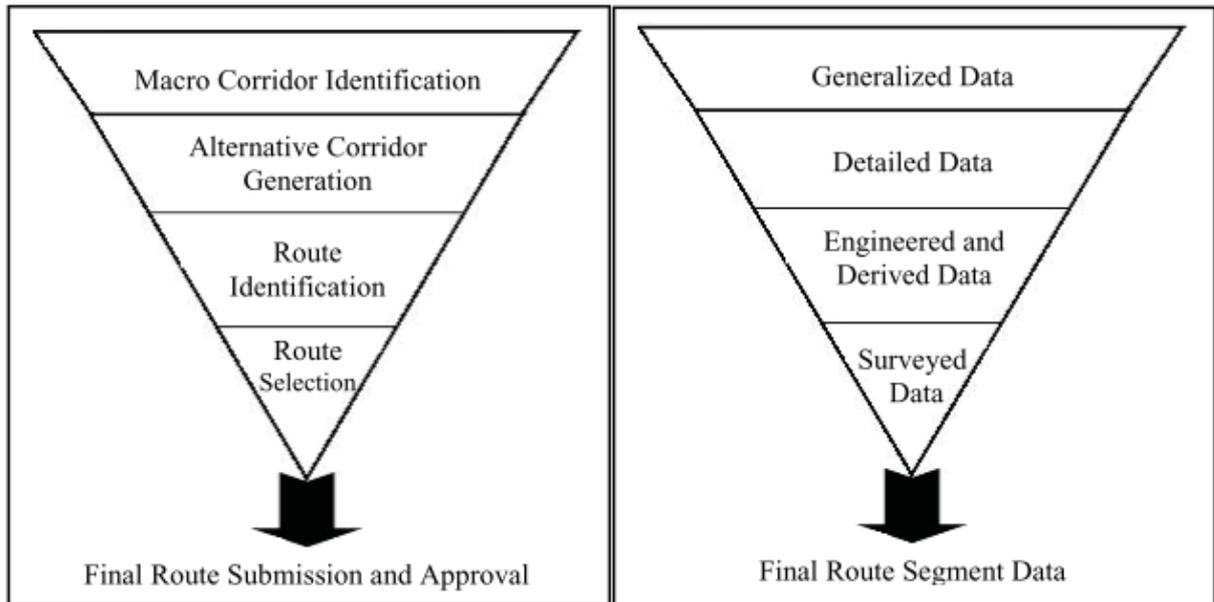
2.2.1.2 Metodologia Electric Power Research Institute – EPRI

Esta metodologia foi desenvolvida analisando dados do estado da Georgia, nos Estados Unidos, em meados da década de 2000. Segundo EPRI (2006), o método consiste em três fases:

1. Geração de um Macro Corredor: uma larga faixa geográfica que define o limite do projeto;
2. Geração de Corredores Alternativos: são áreas lineares contidas no Macro Corredor, consideradas mais adequadas quanto ao meio ambiente natural, ambiente antrópico, requisitos de engenharia e perspectivas combinadas destes aspectos;
3. Análise das Rotas Alternativas: áreas construtíveis contidas nos Corredores Alternativos e seleção das Rotas Preferíveis ou mais adequadas.

O método funciona como um afunilamento de informações e dados avaliando nas primeiras etapas dados em escalas pequenas e no final do processo em escalas maiores pontuais ou de engenharia, conforme mostrado por Schmidt (2009) e sintetizado pelo autor na FIGURA 8.

FIGURA 8 – FASES E DADOS NECESSÁRIOS



FONTE: Schmidt (2009).

A seguir serão explanadas as considerações previstas pela metodologia em cada uma das três fases de projeto.

2.2.1.2.1 Macro Corredores

O Macro Corredor (MC) permite uma rápida identificação da área do projeto, normalmente utilizando dados de cobertura da terra derivados de imagens de satélite (com resolução de 30 metros aproximadamente), modelos digitais de elevação (MDT), infraestrutura viária e LTs já existentes, sendo esses dois últimos classificados quanto ao atravessamento e paralelismo de LTs e caminhos existentes. A identificação dos MCs nas áreas mais adequadas, o método determina três cenários:

- A. Reconstrução ou paralelismo de linhas de transmissão existentes;
- B. Paralelismo de direitos de passagem existentes;
- C. Travessia de terrenos não urbanizados;

O sistema utilizado para a ponderação dos pesos segue uma escala similar ao apresentado por Saaty (1982), onde valores atribuídos variam de 1 a 9, onde os valores de 1 identificam áreas com maior aptidão e 9 identificam uma área de menor aptidão. Os autores descrevem as áreas agrupando a cada três valores, sendo:

- A. Áreas com alta susceptibilidade (1, 2 e 3): são áreas que não contêm recursos sensíveis conhecidos ou restrições físicas e, portanto, devem

- ser consideradas como áreas adequadas para o desenvolvimento dos MC. Exemplos incluem terras não desenvolvidas, pastagens e reconstrução de uma linha de transmissão existente;
- B. Áreas com susceptibilidade moderada (4, 5 e 6): são áreas contêm recursos ou usos do solo que são moderadamente sensíveis a distúrbios ou apresentam uma restrição física moderada à construção e operação de linhas. Os conflitos de recursos ou restrições físicas nessas áreas geralmente podem ser reduzidos ou evitados utilizando medidas de mitigação. Um exemplo é um traçado primário de estradas;
- C. Áreas com baixa susceptibilidade (7, 8 e 9): são áreas que contêm recursos ou usos do solo que apresentam um potencial para impactos adversos significativos que não podem ser prontamente mitigados. A localização de uma linha de transmissão nessas áreas exigiria uma localização cuidadosa ou medidas especiais de projeto. Exemplos incluem zonas úmidas e áreas urbanas densamente povoadas. Estas áreas podem ser utilizadas, mas são evitáveis se possível.

TABELA 7 – ESCALA DE PONDERAÇÃO DO MÉTODO EPRI

Intensidade de importância	Definição	Explicação	Exemplos
1, 2 e 3	Áreas com alta susceptibilidade	São áreas que não contêm recursos sensíveis conhecidos ou restrições físicas e, portanto, devem ser consideradas como áreas adequadas para o desenvolvimento dos MC	Terras não desenvolvidas, pastagens e reconstrução de uma linha de transmissão existente
4, 5 e 6	Áreas com susceptibilidade moderada	São áreas contêm recursos ou usos do solo que são moderadamente sensíveis a distúrbios ou apresentam uma restrição física moderada à construção e operação de linhas. Os conflitos de recursos ou restrições físicas nessas áreas geralmente podem ser reduzidos ou evitados utilizando medidas de mitigação.	Traçado primário de estradas
7, 8 e 9	Áreas com baixa susceptibilidade	São áreas que contêm recursos ou usos do solo que apresentam um potencial para impactos adversos significativos que não podem ser prontamente mitigados. A localização de uma linha de transmissão nessas áreas exigiria uma localização cuidadosa ou medidas especiais de projeto	Zonas úmidas e áreas urbanas densamente povoadas

FONTE: EPRI (2006).

Nesta escala de análise há também áreas que os autores recomendam evitar na hora de estabelecer os MC. Estas áreas são:

- Grandes aeroportos;
- Áreas militares;
- Distritos e estruturas históricas registrados;
- Distritos e sítios arqueológicos registrados;
- Interior de parques nacionais e estaduais;
- Área de refúgio de vida selvagem ou áreas silvestres;
- Corpos d'água não canalizáveis;
- Áreas de pedreiras e mineração;
- Áreas contaminadas com substâncias perigosas;

2.2.1.2.2 Corredores Alternativos

Durante a segunda fase, normalmente é feito um levantamento aerofotogramétrico das áreas potenciais para auxílio na classificação e tomada de decisão, recolhendo dados mais detalhados de zonas úmidas, planícies aluviais, cobertura de terras e outras características e introduzidos na base de dados SIG. Estes dados mais detalhados são usados para identificar quatro Corredores Alternativos (CA) dentro dos limites dos MCs, padronizando a seleção em um dos seguintes rótulos:

- Proteção de lugares e recursos culturais (Perspectiva do Ambiente Construído);
- Proteção de recursos hídricos, plantas e animais (Perspectiva do Ambiente Natural);
- Minimização de custos e atrasos de prazos (Perspectiva dos Requisitos de Engenharia);
- Um composto das alternativas Construído, Natural e de Engenharia (Perspectiva Combinada Simples);

Nesta escala de análise há também áreas que os autores recomendam evitar na hora de estabelecer os CAs, além do que já foi citado na escala anterior. Estas áreas são:

- Áreas edificadas e *buffers* de segurança;

- Áreas de escolas, igrejas, cemitérios e hospitais;
- Rios silvestres e cênicos;
- Parques municipais;
- Áreas de importância ritual;

2.2.1.2.3 Rotas Alternativas

Na etapa final, a equipe identifica um conjunto de Rotas Alternativas dentro dos Corredores Alternativos. Cada rota é então pontuada usando um conjunto padrão de critérios de avaliação e comparada. A rota preferida é selecionada com base nesta comparação.

A metodologia EPRI foi desenvolvida para que todos os agentes atingidos ou participantes da tomada de decisão escolhessem consensualmente a melhor alternativa de cenário. Schmidt (2009) complementa esta metodologia adicionando uma quarta etapa, chamada de “seleção de rota preferida” na qual é aplicado ao modelo uma matriz de ponderação estatística de proporcionalidade aplicando um valor de peso para valores de atributos delineados por segmento de percepção social, custo de construção e cronograma; em seguida, combinando as pontuações para a rota completa afim de determinar a melhor rota ótima.

A diferença entre o método EPRI e os métodos clássicos consolidados é que há uma sistematização de análise com relação à escala das variáveis utilizadas nos estudos de traçados, podendo desfrutar de uma maior discriminação e possibilidades de redução de custos associados dentro da área de interesse.

2.2.1.3 Metodologia utilizando modelagem ambiental

Este método, representado na figura 9, foi apresentado por Campos (2014), onde através de funções probabilísticas bayesianas, busca a parametrização do conhecimento tácito aplicado no planejamento de LT já instaladas em uma determinada região. No estudo, o foco principal é a identificação de padrões espaciais da paisagem, ou seja, se a existência de linhas de energia for considerada um fenômeno que ocorre na paisagem e determinado por variáveis mensuráveis, o autor

propõe que é possível que estes traçados de LT obedeam a padrões espaciais nesta região.

A modelagem matemática é baseada na Teoria da Probabilidade Condicional, cuja aplicação fornece no final a probabilidade de ocorrer a existência de LTs (Z), dado a ocorrência de um conjunto de variáveis explicativas e relacionadas ao fenômeno ($A \cap B \dots \cap n$). Os pesos utilizados nesta abordagem são descritos como “Pesos de Evidência”, sendo estes a influência das variáveis explicativas no fenômeno analisado. A metodologia pode ser descrita em seis equações, onde a (6) representa a probabilidade de ocorrência de Z, assim como (7) o comportamento das variáveis espaciais onde não ocorre Z.

$$P(Z|A \cap B \dots \cap n) = P(Z) \frac{P(A \cap B \dots \cap n|Z)}{P(A \cap B \dots \cap n)} \quad (6)$$

$$P(Z'|A \cap B \dots \cap n) = P(Z') \frac{P(A \cap B \dots \cap n|Z')}{P(A \cap B \dots \cap n)} = 1 - P(Z|A \cap B \dots \cap n) \quad (7)$$

Segundo Mendes e Canales (2011) *apud* Campos (2014), esta função tem características não lineares de regressão e necessita de linearização quando o intervalo de valores possíveis da variável dependente deixar de ser binário. Assim, divisão da probabilidade de $P(Z|A \cap B \dots \cap n)$ por $P(Z'|A \cap B \dots \cap n)$ fornece a chance de ocorrência da variável dependente, cujo conceito é denominado de Odds (O) e tende ao $+\infty$ quando a probabilidade de Z tende a 1, conforme (8).

$$O = \frac{P(Z|A \cap B \dots \cap n)}{1 - P(Z|A \cap B \dots \cap n)} \quad (8)$$

Dessa forma, o intervalo de variação do fenômeno (6) passa a ser $0 \leq (6) \leq +\infty$ e aplicando uma transformação logit da Odds, elimina-se o limite inferior, conforme (9), onde W_n^+ representa o Peso de Evidência.

$$\text{logit}O = \ln e \left[\frac{P(Z|A \cap B \dots \cap n)}{1 - P(Z|A \cap B \dots \cap n)} \right] = W_n^+ \quad (9)$$

Como as variáveis são analisadas em conjunto, os pesos são determinados para cada variável, ou seja, representam a ocorrência de Z dado $(A \cap B \dots \cap n)$. Então converte-se o *logit*O para probabilidade substituindo em (6), conforme (10) e (11).

$$\frac{P(Z|A \cap B \dots \cap n)}{1 - P(Z|A \cap B \dots \cap n)} = e^{\sum W_n^+} \quad (10)$$

$$P(Z|A \cap B \dots \cap n) = \frac{e^{\sum W_n^+}}{1 + e^{\sum W_n^+}} \quad (11)$$

A aplicação do método baseia-se em 4 etapas sequenciais, sendo:

- A. Realizar para cada período de tempo a categorização em classes das variáveis contínuas, a definição dos pesos de evidência e os mapas de probabilidade de ocorrência de LT;
- B. Calibração dos pesos de evidência e recálculo dos mapas de probabilidade;
- C. Validação do modelo;
- D. Otimização do traçado através de variáveis não observáveis na paisagem;

Esta abordagem também se utilizou de testes estatísticos como Coeficiente de Cramer e o *Joint Information Uncertainty* para a verificação da correlação entre todas as variáveis apresentadas em EPRI (2006), destacou as variáveis mais explicativas e optou por utilizar as seguintes variáveis:

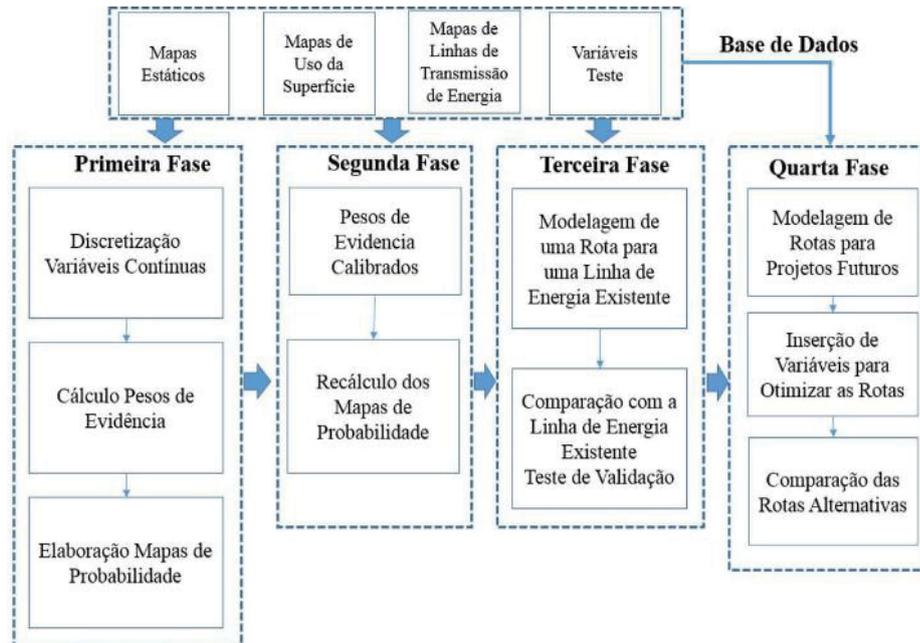
- Distância a rodovias, LT existentes e a principais rios e cursos d'água;
- Uso e cobertura do Solo;
- Declividade.

Utilizando esta metodologia, a superfície de custo torna-se uma superfície de probabilidades e o traçado final busca caminhos (pixels) que possuam maiores probabilidades de ocorrência ao invés de menores valores de custos acumulados.

O tamanho de pixel utilizado no estudo foi de 60 metros, ou seja, o corredor final mínimo determinado pelo algoritmo possui uma largura mínima também de 60 metros. Na fase de desenho do traçado, o autor também fez uso do algoritmo *Least Cost Path*.

Para o cálculo e geração dos mapas o autor fez uso da plataforma conhecida como Dinâmica Geo, desenvolvida e disponibilizada gratuitamente pela Universidade Federal de Minas Gerais - UFMG, e utilizando o módulo *Modeling Environmental Dynamics*, desenvolvido por Soares-Filho *et al* (2008).

FIGURA 9 – ESQUEMATIZAÇÃO DO MODELO DE CAMPOS (2014)



FONTE: Campos (2014).

2.2.2 Comparação entre os métodos existentes

TABELA 8 – COMPARAÇÃO ENTRE METODOLOGIAS EMPREGADAS EM ESTUDOS DE TRAÇADOS DE LINHAS DE TRANSMISSÃO DE ENERGIA

Método de ponderação das camadas	Vantagens	Desvantagens
Modelagem tradicional	Método amplamente difundido e utilizado para a modelagem de problemas multicriteriais. Permite a análise de consistência do processo de ponderação das variáveis de entrada.	A definição dos pesos das variáveis de entrada é feita de forma empírica e está relacionada com a visão e a experiência do analista responsável. Avaliação e utilização de variáveis de entrada de forma aleatória (vários estudos utilizando variáveis distintas, porém usando o mesmo modelo matemático ou suas variações)
Modelagem EPRI	Organização sistêmica e hierárquica do pensamento a partir de pequenas escalas até escalas maiores, de projetos de engenharia. Por utilizar a mesma metodologia funcional que a modelagem tradicional, permite a análise de consistência do processo de ponderação das variáveis de entrada.	Assim como os métodos tradicionais, a definição dos pesos das variáveis é empírica, trazendo subjetividade à análise
Modelagem ambiental (Campos, 2014)	Cálculo de “pesos de evidência” de maneira automática e determinística a partir das equações de probabilidade do modelo.	Utilização de apenas cinco camadas no modelo (consideradas as mais representativas), o que não engloba a completude necessária que o estudo exige. Cálculos e resultados atrelados à plataforma Dinâmica GEO. Não considera variáveis que não estão presentes ou mensuráveis diretamente na paisagem. Trabalha em macro escala

FONTE: O autor (2020).

A TABELA 8 sumariza o estado da arte em relação às metodologias presentes na literatura a respeito do estudo de traçados de linhas de transmissão de energia. Além de tudo, traz-se uma breve descrição a respeito das vantagens e desvantagens de se utilizar cada um dos métodos.

2.3 REDES NEURAS ARTIFICIAIS

Ao contrário dos modelos lineares generalizados (GLM), como é o caso das regressões e suas derivações, para utilizar as redes neurais artificiais para a determinação a relação entre as camadas e variáveis resposta não é necessário pré-

especificar o tipo de relação (Günther e Fritsch, 2010). Isso quer dizer que os dados utilizados são usados para treinar a rede, e esta por sua vez aprende a variação aproximada da relação através da adaptação iterativa de seus parâmetros.

A seguir serão abordados os principais conceitos envolvidos em uma rede neural artificial.

2.3.1 Introdução às Redes Neurais Artificiais

Redes Neurais Artificiais (RNA) é um modelo computacional baseado no comportamento de redes neurais de organismos vivos (HAYKIN, 2001). A diferença entre RNAs e outros algoritmos de inteligência computacional ou sistemas especializados é que as redes neurais não são baseadas em regras. Uma RNA é treinada para reconhecer e generalizar relacionamentos entre um conjunto de insumos e saídas (PECHANEC *et al.*, 2014). A teoria e a base matemática das RNAs são explicadas detalhadamente por muitos autores, como Haykin (2001) e Bishop (1994), todavia uma descrição resumida é apresentada.

Redes Neurais, tanto biológicas quanto artificiais, empregam elementos de processamento simples massivos e interconectados, ou neurônios. De maneira simples, ele é composto pelo corpo, dendritos e um axônio e funciona da seguinte forma: um dendrito longo que conecta os outros neurônios; o axônio usa sinapses para se ligar e enviar sinais a outro neurônio, onde cada neurônio recebe sinais dos demais. Uma vez que o sinal completo (de todos os neurônios) excede um certo limite, o neurônio se torna ativo e começa a enviar um sinal (PECHANEC *et al.*, 2014).

O modelo de um neurônio artificial (equação 13) é mostrado na figura 10. Então, como o neurônio é uma unidade de processamento de informações fundamental para uma RNA, Oliveira (2005) cita três elementos básicos da sua modelagem:

- Sinapse: uma conexão caracterizada por um “peso” ou “força própria”. Uma sinapse m , que está conectada a um neurônio k , recebe um sinal x_m que por sua vez é multiplicado pelo peso sináptico W_{km} ;
- Somador: é um combinador linear, que realiza o somatório das entradas do neurônio, ponderadas pela ação dos respectivos pesos sinápticos;

- Função de Ativação: é o fator restritivo da amplitude do sinal de saída após o condicionamento da ativação do sinal ao ultrapassar um determinado limiar oriundo valor da soma ponderada das entradas. Intervalos típicos após esta normalização são $[0,1]$ e $[-1,1]$.

Para completar, há o *bias* (também conhecido como tendência, em português) que é capaz de aumentar ou diminuir a entrada da função de ativação, além de possuir aplicações externas.

$$v_k = \sum_{j=1}^m W_{kj} x_j \quad (13)$$

Onde:

$W_{k1}, W_{k2}, \dots, W_{km}$: são os pesos sinápticos do neurônio k ;

x_1, x_2, \dots, x_m : são os vetores de entrada;

v_k : é a saída do combinador linear;

b_k : é a tendência;

$F(\cdot)$: é a função de ativação;

y_k : é a saída do neurônio k .

FIGURA 10

Segundo Oliveira (2005), matematicamente o neurônio pode ser descrito pelas EQUAÇÕES (12) e (13).

$$y_k = F(v_k + b_k) \quad (12)$$

$$v_k = \sum_{j=1}^m W_{kj} x_j \quad (13)$$

Onde:

$W_{k1}, W_{k2}, \dots, W_{km}$: são os pesos sinápticos do neurônio k ;

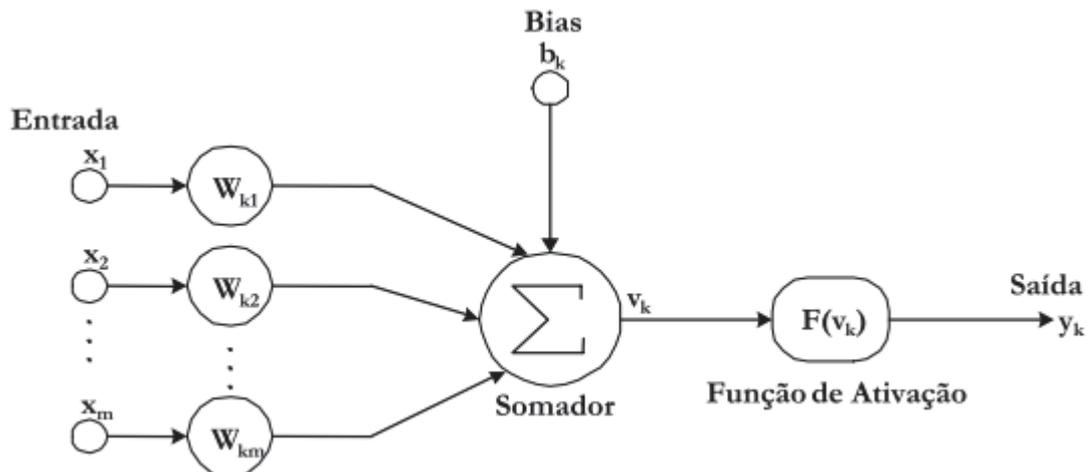
x_1, x_2, \dots, x_m : são os vetores de entrada;

v_k : é a saída do combinador linear;

b_k : é a tendência;

$F()$: é a função de ativação;
 y_k : é a saída do neurônio k .

FIGURA 10 – NEURÔNIO ARTIFICIAL



FONTE: Oliveira (2005).

Oliveira (2005) cita que há pelo menos quatro tipos de funções de ativação, sendo elas: lineares, limiars, rampas (lineares por partes) e sigmoidais. Já Pechanec *et al.* (2014) fala que as funções mais usadas comumente são a função sigmóide ou a função *Radial Basis Function* – *RBF*, sendo que a primeira delas apresenta um comportamento que varia tanto linearmente quanto não-linearmente.

Pechanec *et al.* (2014) fala que um dos primeiros passos para a utilização de uma RNA para resolver um problema é treinar essa rede para “aprender” a relação entre as camadas de entrada e de saída conhecidas (se houver), em um dado conjunto com uma regra de aprendizagem. Isso é feito mapeando e modificando as funções internas a fim de determinar o melhor relacionamento.

Hewitson e Crane (1994) *apud* Pechanec *et al.* (2014) fala que a princípio os pesos são aleatórios, o cálculo é realizado e segue-se uma análise se a rede estimou um resultado correto. Se este não for o caso, os pesos são modificados até que o resultado correto seja alcançado. Esta matriz ponderada de interconexões permite que a rede neural aprenda e se lembre (SKAPURA, 1996).

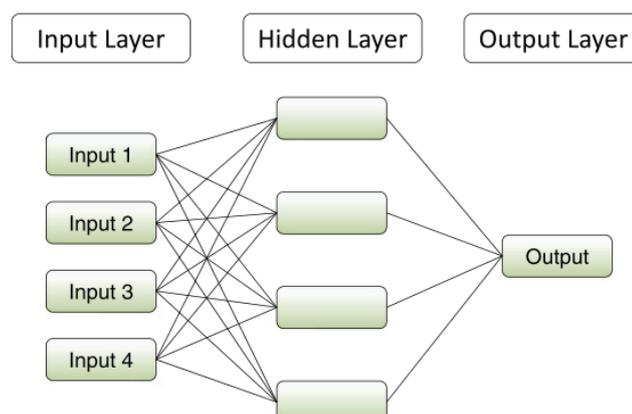
2.3.2 Arquiteturas de rede e aprendizagem

Segundo Oliveira (2005), há três tipos básicos de arquitetura, sendo:

- A. Redes diretas de uma camada: são redes neurais constituídas por apenas um agrupamento de neurônios, ou camada, que recebem informação simultaneamente. Há uma camada de nós de fonte que recebem sinais de excitação do meio externo e estes se conectam com a camada de neurônios, que por sua vez, ligam-se aos nós de saída da rede (HAYKIN, 2001). São tipicamente usadas na telefonia;
- B. Redes diretas de camadas múltiplas (*feed-forward*): possuem camadas ocultas cujos nós computacionais são chamados de neurônios ocultos. Nas redes de múltiplas camadas, como a mostrada na FIGURA 11, cada camada tem uma função específica. A camada de saída final recebe os estímulos da camada intermediária e constrói a resposta global para o padrão de ativação fornecido pelos nós de fonte da camada de entrada. As camadas intermediárias funcionam como extratoras de características, codificando através de seus pesos as características apresentadas nos padrões de entrada, fazendo com que a rede crie uma representação própria do problema, com mais riqueza e complexidade (Rezende, 2003 *apud* Oliveira, 2005);
- C. Redes recorrentes (*Feed-backward*): são redes que possuem pelo menos um laço de alimentação, devido ao seu comportamento dinâmico. São geralmente utilizadas em algoritmos destinados à controle de qualidade de produtos.

Para estudos em GIS, a arquitetura mais usual são as redes diretas multicamadas, conforme mostrado na Figura 11.

FIGURA 11 – ARQUITETURA DE UMA RNA MULTICAMADAS



FONTE: Kia *et al.* (2012).

AlThuwaynee (2020) explica esse tipo de arquitetura da seguinte forma:

- *Input Layers*: são todas as variáveis de entrada representadas como nós de entrada;
- *Hidden Layers*: são nós onde uma ou mais variáveis são combinadas. Tipicamente todos os nós são conectados em todos os *hidden layers*, criando assim novas características derivadas dos *inputs* fornecidos;
- *Output Layers*: usa os nós dos *hiddens* para fazer uma previsão ou classificação. Previsão numérica usa um *output layer*, enquanto classificação usa $c-1$ *layers* (onde c é o número de classes possíveis);
- *Bias Node*: é uma constante aprendida fora dos dados de entrada. Isto permite alterar o modelo aprendido, pois sem isso, qualquer valor de zero ou entradas negativas resultariam em uma camada de saída zerada;
- *Connection Lines*: representam o peso de um nó até um outro nó no próximo *layer*.

Com relação ao tipo de aprendizagem, Oliveira (2005) cita dois tipos de cenários: a aprendizagem “com professor” e a “sem professor”. A aprendizagem com “professor” ou “Supervisionada” onde há a presença de amostras de saídas, ou seja, o “professor” possui conhecimento a respeito de determinado ambiente. A rede então utiliza como orientação o ajuste dos pesos sinápticos a discrepância entre a saída esperada e a saída calculada, na tentativa de minimizar essa diferença. Widrow *et al.* (1994) *apud* Oliveira (2005) fala que um dos parâmetros para aferição do desempenho deste tipo de RNA é a soma dos erros quadráticos das saídas da rede para todas as amostras. O algoritmo mais usado para aprendizagem supervisionada é conhecido como *Backpropagation* ou retropropagação.

Já a aprendizagem sem professor recebe os estímulos do ambiente sem os valores de resposta esperados e pode ser subdividida em (Oliveira, 2005):

- Aprendizagem por Reforço: baseada em “satisfação”. Este tipo permite a interação da rede com o ambiente, desenvolvendo a habilidade para realização das tarefas a partir dos resultados de sua própria experiência, caracterizando assim um problema de aprendizagem baseado em comportamento. Isso faz com que a rede aprenda a tomar decisões a partir da observação de seu próprio comportamento e

melhora a qualidade das suas ações através de um mecanismo de reforço. Essas redes buscam normalmente resolver o problema de adaptação ao usuário (típica aplicação em jogos eletrônicos);

- **Aprendizagem Não-Supervisionada:** o objetivo desta categoria de algoritmos é a descoberta de padrões significativos através de correlações entre os exemplos de treinamento. Isso significa que a rede tem que achar atributos estatísticos relevantes, e há a necessidade redundância de dados de entrada. São aplicadas em problemas como a identificação de usuários de cartão de crédito baseadas em seus perfis de compra, por exemplo.

2.3.2.1 Tipo de tarefas de aprendizagem

Oliveira (2005) elenca tarefas básicas e relaciona o tipo de aprendizagem a ser utilizada, sendo elas:

- A. Associação de Padrões:** Pode-se classificar a associação em autoassociação e heteroassociação. No primeiro caso, apresenta-se um conjunto de padrões para que ocorra o armazenamento desse conjunto. Depois são apresentados à rede os mesmos padrões, porém com características distorcidas, testando o poder associativo da rede, necessitando aprendizado não-supervisionado. Já no segundo caso, como existem pares de entrada e saída esperados utiliza-se aprendizado supervisionado;
- B. Reconhecimento de Padrões:** o processo de reconhecimento passa por uma fase de treinamento onde se apresenta à rede um conjunto de padrões (repetidamente), juntamente com a classe à qual cada padrão pertence. A partir disso, a rede adquire a capacidade de identificar a classe de uma nova entrada a partir da extração das características. Para este caso utiliza-se redes de múltiplas camadas e aprendizado supervisionado;
- C. Aproximação de Funções:** tem por objetivo mapear uma determinada relação funcional que contemple o universo amostral analisado. Neste caso, as saídas esperadas e obtidas geram um sinal de erro que é usado para orientar a calibração dos pesos da rede. O processo é dado

por encerrado quando o sinal de erro é reduzido a valores inferiores a uma determinada tolerância. Nesse caso é uma aplicação de aprendizagem supervisionada;

- D. Controle: também utilizada para controle de processos em um determinado sistema. De forma geral, a saída do processo é alimentada de volta à entrada a fim de calcular um “sinal de erro”, subtraindo de um sinal de referência externo. Esta tarefa requer aprendizagem supervisionada.

Haykin (2001) também cita mais dois tipos de tarefas apresentadas como “filtragem” e “formação de feixes”. Os filtros podem ser utilizados para extrair informações sobre um determinado conjunto de dados ruidosos. Já a formação de feixes é um tipo especial de filtro que é utilizado na acústica para distinguir entre as propriedades espaciais de um sinal alvo e o ruído de fundo.

2.4 NEURALNET

2.4.1 Introdução

Segundo Günther e Fritsch (2010), a Neuralnet foi construída para treinar neurônios multicamadas (*multi-layer perceptrons*) no contexto de uma análise de regressão, ou seja, uma aproximação funcional. O pacote neuralnet foi desenvolvido (utilizando a linguagem *R*) para lidar com um número arbitrário de camadas e variáveis resposta, assim como camadas e neurônios ocultos, porém os custos computacionais podem aumentar de forma exponencial com o aumento da ordem de complexidade. Um outro problema surge com o aumento desta complexidade é que, se o usuário definir um número máximo de iterações e esse valor não for o suficiente, isto pode causar uma parada precipitada no processo de iteração, antes do algoritmo alcançar a convergência.

Através da função *compute* pode ser aplicada para o cálculo de predições para uma nova combinação de camadas, ou seja, com o modelo definido através do processo de treinamento, pode-se prever o fenômeno estudado usando novas entradas, em épocas diferentes por exemplo. Isto quer dizer que, se os dados de treinamento forem de boa qualidade e se houver completude na avaliação de um

determinado fenômeno, esta rede torna-se um sistema especializado na resolução ou estudo daquele evento.

2.4.2 Arquitetura multicamadas

A arquitetura da rede segue os conceitos abordados no item B da seção 2.3.2 e de acordo com Günther e Fritsch (2010), a formulação matemática de um simples neurônio multi-layer, com n camadas de entrada pode ser expressa de acordo com a EQUAÇÃO (14).

$$o(x) = f\left(w_o + \sum_{i=1}^n w_i x_i\right) = f(w_o + w^T x) \quad (14)$$

Onde:

w_o : denota a interrupção ou parada;

$w = w_1, \dots, w_n$: vetor contendo os pesos sinápticos sem parada;

$x = x_1, \dots, x_n$: são as camadas;

f : é a função de ativação;

o : é a variável resposta.

Da mesma forma, pode calcular uma camada de j neurônios ocultos, multiplicando a EQUAÇÃO (14) por uma somatória de pesos acumulados até aquela camada, de acordo com a EQUAÇÃO (15).

$$o(x) = f\left(w_o + \sum_{j=1}^J w_j f(w_{oj} + w_j^T x)\right) \quad (15)$$

Onde neste caso:

w_o : denota a interrupção ou parada do neurônio de saída;

w_{oj} : denota a interrupção ou parada no j -ésimo neurônio oculto;

w_j : denota o peso sináptico correspondente ao j -ésimo neurônio oculto;

$w_j = w_{1j}, \dots, w_{nj}$: é o vetor de pesos sinápticos correspondentes a todos os j -ésimos neurônios ocultos;

No pacote neuralnet há três versões diferentes de *resilient backpropagation* implementadas (versões otimizadas em relação aos retropropagações tradicionais),

além da *backpropagation* tradicional propriamente dita. A função de ativação f é, para o estudo em questão, função não-linear e diferenciável não-diminutiva e, como a variável resposta está no intervalo entre $[0,1]$, utilizou-se a função logística apresentada na EQUAÇÃO (16).

$$f(y) = \frac{1}{1 + e^{-y}} \quad (16)$$

Onde y é dado pela somatória dos pesos acumulados:

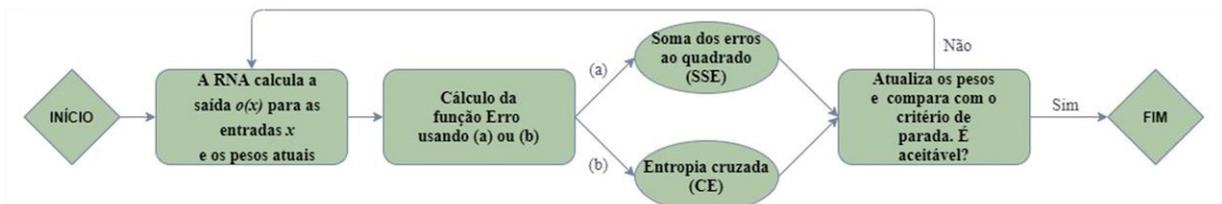
$$y = w_o + \sum_{j=1}^J w_j f(w_{oj} + w_j^T x) \quad (17)$$

A seguir, será explanado sobre a aprendizagem do algoritmo.

2.4.3 Aprendizagem supervisionada

Segundo Günther e Fritsch (2010), a neuralnet utiliza algoritmos de aprendizagem supervisionada. Neste tipo de processo, uma determinada saída calculada é comparada com uma saída prevista e através desta métrica, e os pesos são iterativamente adaptados. Todos os pesos são inicializados com valores aleatórios retirados de uma distribuição normal padrão. O processo de aprendizagem pode ser observado na FIGURA 12.

FIGURA 12 – PROCESSO DE APRENDIZAGEM



FONTE: Adaptado de Günther e Fritsch (2010).

As equações matemáticas que descrevem as relações das métricas de erros E entre uma saída predita $o(x)$ dadas as entradas x e os pesos na interação analisada e a saída y observada são dadas pela Equação 18 e pela Equação 19.

a) Soma dos erros ao quadrado (SSE):

$$E = \frac{1}{2} \sum_{l=1}^L \sum_{h=1}^H (o_{lh} - y_{lh})^2 \quad (18)$$

b) Entropia cruzada (CE):

$$E = - \sum_{l=1}^L \sum_{h=1}^H ((y_{lh} \log o_{lh}) + (1 - y_{lh}) \log(1 - o_{lh})) \quad (19)$$

Onde:

$l = 1, 2, \dots, L$: são os pares de observações entrada-saída;

$h = 1, 2, \dots, H$: são os nós de saída;

O processo é interrompido se um critério pré-estabelecido for atingido, ou seja, se todas as derivadas parciais absolutas da função E em relação aos pesos w são menores que um limiar z , conforme a EQUAÇÃO (20).

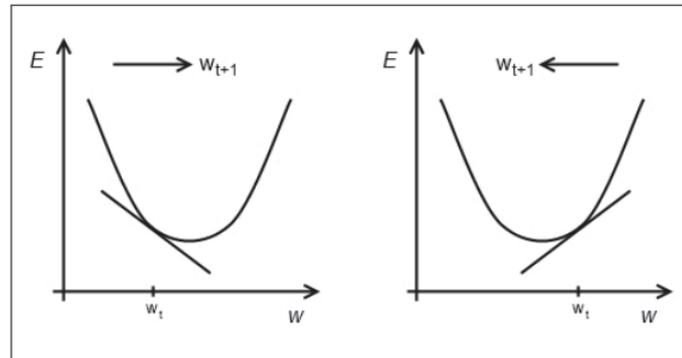
$$\frac{\partial E}{\partial w} < z \quad (20)$$

A seguir, serão abordados conceitos sobre os diferentes tipos de algoritmos de retropropagação presentes no pacote neuralnet.

2.4.4 Backpropagation e Resilient Backpropagation

Segundo Günther e Fritsch (2010), o algoritmo de *resilient backpropagation* – *RBP* (ou retropropagação resiliente) é baseada na retropropagação (*backpropagation* - *BP*) tradicional e tem por objetivo modificar os pesos da rede neural de modo a encontrar um mínimo local da função de erro. Para isso, o gradiente da função E é calculado em relação aos pesos para encontrar o caminho até esse valor, modificando os valores de peso na direção contrária às suas derivadas parciais (se a derivada parcial for negativa, o valor do peso é aumentado e vice-versa). A ideia básica pode ser observada na FIGURA 13.

FIGURA 13 – PRINCÍPIO DO ALGORITMO DE BACKPROPAGATION EM UMA FUNÇÃO ERRO UNIVARIADA



FONTE: Günther e Fritsch (2010).

Segundo Günther e Fritsch (2010), o pacote neuralnet proporciona a capacidade de alternar entre BP e RBP, além usar ou não recuo de peso e a convergência global segundo Anastasiadis *et al.* (2005). De maneira geral, todos os algoritmos tendem a minimizar a função erro adicionando uma taxa de aprendizagem aos pesos na direção oposta ao gradiente. Para fins de otimização o algoritmo RBP usa para cada peso, uma taxa de aprendizado separada (n_k), que pode ser alterada durante o treinamento. Segundo Riedmiller e Braun (1993) *apud* Günther e Fritsch (2010), isso resolveria o problema de definição de uma taxa de aprendizagem geral (apropriada para todo o treinamento e para qualquer trabalho da rede). Adicionalmente, isso garantiria uma influência uniforme da taxa de aprendizagem em toda a rede, pois em vez de se utilizar a magnitude das derivadas parciais, apenas os sinais (*sign*) são usados para a atualização dos pesos.

Na RBP, os pesos são ajustados na rede neural através da EQUAÇÃO (21).

$$w_k^{(t+1)} = w_k^{(t)} - n_k^{(t)} \text{sign} \left(\frac{\partial E^{(t)}}{\partial w_k^{(t)}} \right) \quad (21)$$

Onde:

t : são as iterações;

k : é o índice do peso.

Já para o algoritmo tradicional do *backpropagation*, os pesos são atualizados através da EQUAÇÃO (22).

$$w_k^{(t+1)} = w_k^{(t)} - n \left(\frac{\partial E^{(t)}}{\partial w_k^{(t)}} \right) \quad (22)$$

Uma outra otimização inserida no pacote *RBP* em relação tradicional, é a aceleração da convergência em áreas consideradas “rasas”, ou seja, onde duas derivadas consecutivas possuem o mesmo sinal, adicionando ou diminuindo o valor da taxa de aprendizado (n_k), usando uma técnica chamada de “recuo de peso”. Sem essa melhoria, o algoritmo poderia “saltar” várias vezes sobre o mínimo local procurado (Riedmiller e Braun, 1993 *apud* Günther e Fritsch, 2010).

A convergência global que é utilizada no algoritmo de *RBP* foi desenvolvida por Anastasiadis *et al.* (2005), com uma introdução de taxa de aprendizagem em relação a todas as outras calculadas. É a taxa de aprendizagem a menor derivada parcial absoluta ou a menor taxa de aprendizagem (i), que é alterada de acordo com a EQUAÇÃO (23).

$$n_i^{(t)} = - \frac{\sum_{k; k \neq i} n_k^{(t)} \frac{\partial E^{(t)}}{\partial w_k^{(t)}} + \delta}{\frac{\partial E^{(t)}}{\partial w_i^{(t)}}} \quad (23)$$

Se as condições a seguir forem satisfeitas (EQUAÇÃO (24)):

$$\frac{\partial E^{(t)}}{\partial w_i^{(t)}} \neq 0 \quad e \quad 0 < \delta \ll \infty \quad (24)$$

Para mais informações, consultar Anastasiadis *et al.* (2005).

2.5 MÉTODO DE AVALIAÇÃO DO MODELO FUNCIONAL TREINADO

2.5.1 Probabilidade conjunta e condicional

Para problemas de classificação binária, onde há amostragens positivas e negativas a respeito de um fenômeno, a maneira natural de se apresentar as estatísticas para avaliação de um modelo preditor/classificador é utilizando a tabulação cruzada entre a classe prevista e a classe real de exemplos/amostras. Essa

tabela de contingência, também conhecida como matriz de confusão, mostra a as contagens de frequências absolutas (PRATI *et al.*, 2008).

Quando um exemplo positivo é classificado como positivo, ele é denominado verdadeiro positivo. Da mesma forma, quando um exemplo negativo é classificado como positivo, ele é denominado falso positivo. A nomenclatura similar é utilizada no caso dos exemplos classificados como negativos. Nessa tabela, *TP*, *FP*, *FN* e *TN* correspondem, respectivamente, às quantidades de verdadeiro/falso positivo/negativo. *PP* e *PN* correspondem ao número de exemplos preditos como positivos/negativos e *POS* e *NEG* ao número real de exemplos positivos/negativos na amostra; assim como *N* é o número total de amostras (FIGURA 14 (a)).

Segundo Prati *et al.* (2008), se dividirmos cada entrada na matriz mostrada na FIGURA 14 (a) pelo tamanho da amostra, cada entrada dessa matriz representará uma estimativa da probabilidade conjunta da classe real do exemplo e da predição dada pelo modelo. Dessa forma, pode-se dizer que as matrizes contidas na FIGURA 14 são equivalentes.

FIGURA 14 – MATRIZ DE CONTIGÊNCIA PARA MODELOS DE CLASSIFICAÇÃO

	predito				<i>Y</i>	\bar{Y}	
real	<i>TP</i>	<i>FN</i>	<i>POS</i>	\bar{X}	$p(X, Y)$	$p(X, \bar{Y})$	$p(X)$
	<i>FP</i>	<i>TN</i>	<i>NEG</i>	\bar{X}	$p(\bar{X}, Y)$	$p(\bar{X}, \bar{Y})$	$p(\bar{X})$
	<i>PP</i>	<i>PN</i>	<i>N</i>		$p(Y)$	$p(\bar{Y})$	1
	(a) Frequência absoluta				(b) Probabilidade conjunta		

FONTE: Prati *et al.* (2008).

De acordo com Prati *et al.* (2008), na matriz da FIGURA 14 (b), *X* representa a variável aleatória da classe real positiva, *Y* representa a variável aleatória da classe predita positiva, \bar{X} e \bar{Y} representam a negação de *X* e *Y*.

Em termos de probabilidade condicional (EQUAÇÃO (25)), tem-se que:

$$P(X, Y) = P(X|Y)P(Y) = P(Y|X)P(X) \quad (25)$$

Onde $P(X, Y)$ é a probabilidade condicional de *X* ser verdade dado que *Y* é verdade, sendo $P(X, Y) = P(Y, X)$, mas $P(X|Y) \neq P(Y|X)$. As EQUAÇÕES (29) e (27) mostram como são calculados esses valores a partir da matriz de confusão.

$$P(X|Y) = \frac{P(X, Y)}{P(Y)} = \frac{TP}{PP} \quad (26)$$

$$P(Y|X) = \frac{P(X, Y)}{P(X)} = \frac{TP}{POS} \quad (27)$$

A probabilidade $P(X|Y)$ é importante para o usuário do modelo pois fornece a confiança do modelo, ou seja, a probabilidade de a classe ser positiva, dado que a previsão feita pelo modelo é positiva. Já em termos de avaliação do modelo, $P(Y|X)$ é mais útil pois fornece a crença ou verossimilhança da variável predita, ou seja, a probabilidade de uma predição feita dado a ocorrência de uma observação específica (indica o quanto o modelo é capaz de distinguir os casos entre as possíveis classes, 0 ou 1 no caso).

2.5.2 Avaliação do modelo

A avaliação é baseada na análise da matriz de contingência. A partir dela, pode-se calcular a taxa de erro de classificação (EQUAÇÃO (28)).

$$Erro = P(\bar{Y}, X) + P(Y, \bar{X}) = \frac{FP + FN}{N} \quad (28)$$

Porém segundo Prati *et al.* (2008), essa taxa de erro assume custos iguais para erros em ambas as classes do modelo (*POS* ou *NEG*), e para problemas na área da medicina e de algumas áreas da engenharia por exemplo, esta métrica não é apropriada. Para soluções contínuas, como os resultados de modelos treinados através de redes neurais, os métodos gráficos são mais apropriados, pois permitem a visualização da multidimensionalidade do problema.

Outras métricas também usadas em *Machine Learning* que são derivadas da matriz de confusão são a sensibilidade do modelo (ou Taxa de verdadeiros positivos – EQUAÇÃO (29)), a especificidade (EQUAÇÃO (30)) e a taxa de falsos positivos (EQUAÇÃO (31)).

$$\text{Sensibilidade} = \frac{TP}{TP + FN} \quad (29)$$

$$\text{Especificidade} = \frac{TN}{TN + FP} \quad (30)$$

$$\text{TxFalsosPositivos} = \frac{FP}{TN + FP} = 1 - \text{Especificidade} \quad (31)$$

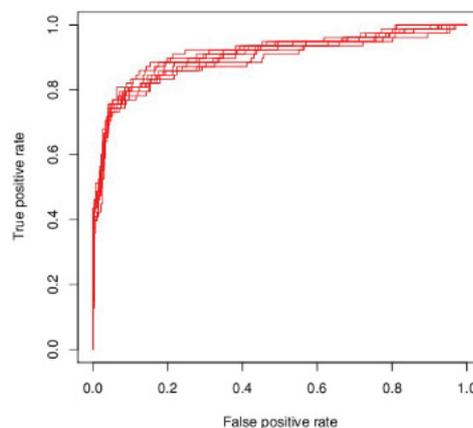
A seguir serão abordados conceitos de análise gráfica.

2.5.2.1 A curva ROC

Prati *et al.* (2008) descrevem a Curva ROC (do inglês *Receiver Operating Characteristic*) como um método gráfico para a avaliação, organização e seleção de sistemas de diagnóstico e/ou predição.

O gráfico ROC é baseado na probabilidade de detecção ou taxa de verdadeiros positivos ($tpr = P(Y|X)$) e na probabilidade de falsos alarmes ou taxa de falsos positivos ($fpr = P(Y|\bar{X})$). Para construir o gráfico, plota-se fpr o eixo x e tpr no eixo y (FIGURA 15). A maneira mais eficiente de gerar essa curva é ordenar todos os casos de teste de acordo com o valor contínuo predito pelo modelo. A partir desse conjunto ordenado, para cada caso desse conjunto e seguindo-se essa ordem, dá-se um passo de tamanho $1/POS$ na direção do eixo y se o exemplo for positivo ou um passo de tamanho $1/NEG$ caso o exemplo seja negativo.

FIGURA 15 – CURVA ROC CALCULADA COM DIVERSOS CONJUNTOS DE TESTE



FONTE: Prati *et al.* (2008).

Quanto mais distante a curva estiver da diagonal principal, melhor será o desempenho do sistema de aprendizado para o fenômeno avaliado. Ainda se tem que, quanto mais próximo a curva estiver do ponto (0.0,1.0), mais acurado será o modelo treinado.

2.5.2.2 Cálculo da AUC

O valor ou coeficiente AUC (*Area Under Curve*) é uma métrica que acompanha a curva ROC na análise de modelos treinados por uma RNA. Esta métrica é invariante em escala, uma vez que trabalha com precisão das classificações ao invés valores absolutos. Além disso, segundo Bradley (1997), este método é invariante ao critério de decisão selecionado e às probabilidades prévias também, cujo valores tem um significado simples.

Segundo Bradley (1997), quando o limiar de decisão na classificação é variado (como é o caso das RNAs) e os pontos na curva forem obtidos, a EQUAÇÃO (32) descreve uma maneira de calcular a área abaixo da curva ROC. A somatória também pode ser entendida como uma integração. Então, se $P(FP) = \alpha$ e $P(TP) = \beta$, tem-se:

$$AUC = \sum_i \left\{ (1 - \beta_i \Delta\alpha) + \frac{1}{2} [\Delta(1 - \beta) \Delta\alpha] \right\} \quad (32)$$

Onde:

$$\Delta(1 - \beta) = (1 - \beta_i) - (1 - \beta_{i-1}) \quad (33)$$

$$\Delta\alpha = \alpha_i - \alpha_{i-1} \quad (34)$$

O significado do valor da área encontrado está associado a probabilidade de que um exemplo positivo escolhido ao acaso é corretamente estimado seja maior que a probabilidade de um exemplo negativo aleatoriamente escolhido. Além disso, essa probabilidade de classificação correta é equivalente às quantidades estimadas pelas estatísticas de Wilcoxon (testa a hipótese de que uma ou mais variáveis x de uma população A sejam iguais a de uma outra população B , como por exemplo, amostras de testes e de treinamento).

Ainda é possível calcular o erro padrão da estimativa AUC. Para mais informações, consultar Bradley (1997).

3 MATERIAIS E MÉTODO

3.1 MATERIAS

Os materiais utilizados nesta pesquisa encontram-se descritos a seguir:

Recursos de hardware:

- A. 01 Computador. S. O. Linux Ubuntu 14.04, processador Intel Core I7, 32GB de memória RAM, HD de 1TB, placa de vídeo Radeon R7 2GB.
Processamento dos dados brutos;

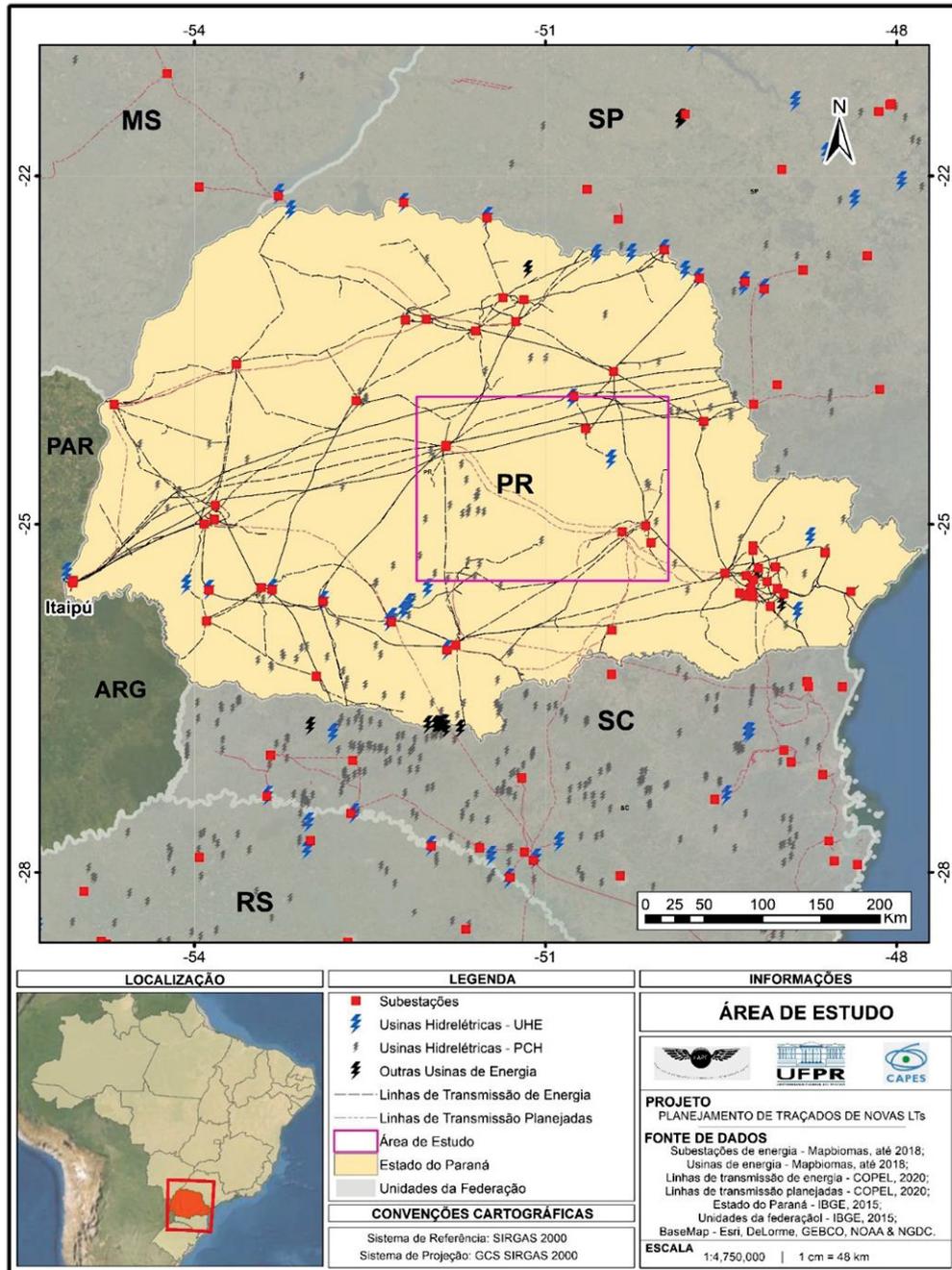
Recursos de software:

- A. Visual Studio Code: editor de código fonte *open source*, com extensões para linguagens como C++, C#, Java, Python, PHP e Go. Versão 1.52;
- B. RStudio: editor de código fonte *open source*, especializado no desenvolvimento de codificação na linguagem R. Versão 1.3.1056;
- C. QGIS: software *open source*, sendo este uma multiplataforma de sistemas SIG que permite a visualização, edição e análise de dados georreferenciados. Versão 2.18.16;
- D. ArcGIS: software proprietário, sendo este uma multiplataforma de sistemas SIG que permite a visualização, edição e análise de dados georreferenciados. Versão 10.5;
- E. Linguagem de programação Python e R;
- F. Pacote *Microsoft Office*: editor de planilha, apresentações e de texto (proprietário).

3.1.1 Área de estudo

A área de estudo utilizada como amostra para a elaboração desta pesquisa é uma parte do estado do Paraná, conforme visto na FIGURA 16. Um dos objetivos do presente trabalho é treinar e uma rede neural capaz de se tornar, em um segundo momento, um sistema especializado no estudo de traçados de linhas de transmissão no estado do Paraná, que então poderia ser empregada pela Companhia Paranaense de Energia – Copel.

FIGURA 16 – ÁREA DE ESTUDO



FONTE: O autor (2020).

Além da área de estudo, a FIGURA 16 traz informações a respeito das usinas de energia e das subestações presentes no território paranaense. A camada vetorial das LTs contida no mapa (Linhas de Transmissão de Energia), foi disponibilizada pela COPEL, assim como as LTs planejadas em toda a extensão de sua concessão (Linhas de Transmissão Planejadas).

A seguir serão abordadas as fontes de dados utilizados, seguido da descrição da metodologia a ser usada.

3.1.2 Dados de referência utilizados no estudo

As camadas empregadas na análise de traçados de LT são oriundas de bases de dados oficiais de órgãos da Cartografia de Referência ou que são oficialmente responsáveis pela coleta e disponibilização de cada tipo de dado dentro da área de estudo utilizada, estando com suas respectivas fontes esquematizadas na Tabela 9.

TABELA 9 – DADOS DE REFERÊNCIA E SUAS FONTES

Descrição	Órgão/Ente Responsável	Link para Busca
Dados Vetoriais De Hidrografia	ANA	https://dadosabertos.ana.gov.br/datasets
Dados De Aeródromos Civis Brasileiros	ANAC	https://www.anac.gov.br/aceso-a-informacao/dados-abertos/areas-de-atuacao/aerodromos/lista-de-aerodromos-publicos-v2
Dados De Áreas Militares	BDGEx	https://bdgex.eb.mil.br/bdgexapp
Dados Vetoriais De LTs Existentes	COPEL	https://www.copel.com/hpcopel/root/nivel2.jsp?endereco=%2Fhpcopel%2Froot%2Fpagcopel2.nsf%2Fdocs%2F2D32A614F65CCAF60325740A005ED0CC
Dados Vetoriais Rodoviários	DNIT	http://servicos.dnit.gov.br/vgeo/
Dados de Uso da Terra (2019)	MAPBIOMAS	https://mapbiomas.org/colecoes-mapbiomas-1
Dados Vetoriais De Áreas Estratégicas Para a Conservação Da Biodiversidade Paranaense	IAP	http://www.iat.pr.gov.br/Pagina/Dados-e-Informacoes-Geoespaciais-Tematicos
Dados Vetoriais Das Divisões Políticas Estaduais e Municipais	IBGE	https://www.ibge.gov.br/geociencias/downloads-geociencias.html
Dados De Imóveis Certificados No SIGEF (Dados de Áreas de Preservação Particulares)	INCRA	https://certificacao.incra.gov.br/csv_shp/export_shp.py
Dados Vetoriais de Terras Indígenas, das Unidades de Conservação de Proteção Integral, de Áreas Embargadas Pelo Ibama	INDE	https://inde.gov.br/AreaDownload
Dados Vetoriais de Elevação e Dados Matriciais de Declividade	INPE	http://www.webmapit.com.br/inpe/topodata/

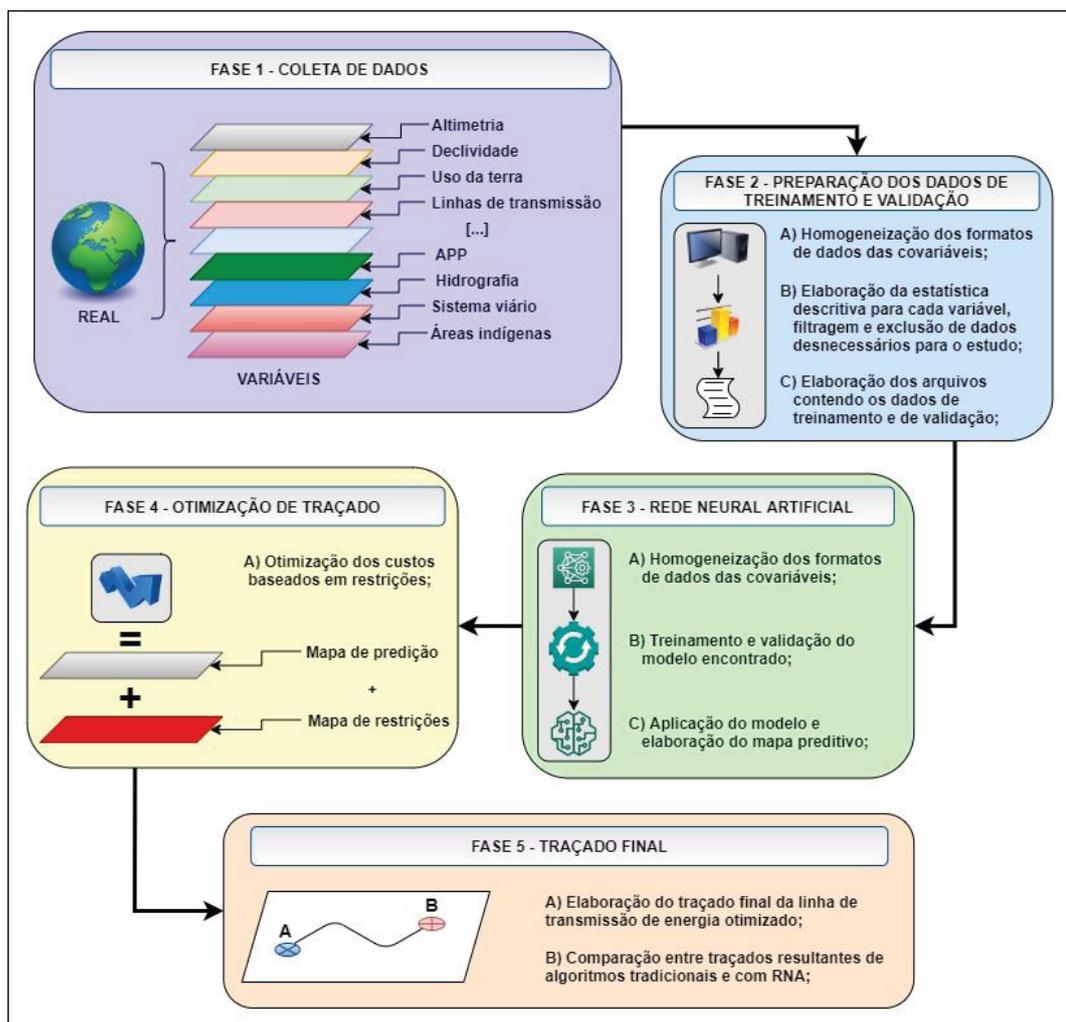
FONTE: Elaborado pelo autor (2020).

3.2 MÉTODO PROPOSTO

O método proposto está dividido em cinco fases distintas (FIGURA 17), sendo elas: coleta de dados, preparação dos dados, aplicação da RNA, otimização do traçado e elaboração do traçado final.

As duas primeiras fases (coleta e preparação dos dados) compreendem as etapas mais morosas do trabalho, uma vez que alguns dados importantes para o trabalho podem demorar para serem adquiridos, como por exemplo, os vetores de LTs existentes. Além disso, a fase de preparação dos dados compreende a manipulação das camadas de entrada, que apesar de se ter algumas rotinas implementadas em *Python*, ainda necessita de muito trabalho manual de preparação, seleção e eliminação de dados referentes às camadas de entrada.

FIGURA 17 – ARQUITETURA DO MÉTODO PROPOSTO



FONTE: O autor (2020).

Os principais tópicos de todas as fases do estudo serão explicados nas seções a seguir.

3.2.1 Homogeneização dos formatos de dados de entrada e cálculo das estatísticas descritivas

Os dados de referência utilizados no estudo possuem fontes diferentes, formatos diferentes e, conseqüentemente, a escala dos dados é diferente. Para avaliar a participação de cada covariável de maneira uniforme na RNA, é necessária a homogeneização dos dados, tanto em formato e, se possível em escala.

Para facilitar a aplicação das relações matemáticas entre variáveis e também as operações algébricas entre as camadas finais do estudo, optou-se pela utilização do formato *raster* para a execução dos estudos. Isto porque já existem RNAs que trabalham com formatos vetoriais ou mistas, porém traz complicações devido as relações topológicas entre layers.

A TABELA 10 complementa as informações apresentadas na Tabela 9, contendo especificações adicionais a respeito dos dados de entrada já apresentados.

TABELA 10 – DADOS DE ENTRADA, INFORMAÇÕES ADICIONAIS

Covariável	Escala	Formato
Declividade	1:60.000	Matricial
Hidrografia	~1:50.000	Vetorial
Áreas de proteção particulares	~1:25.000	Vetorial
Áreas estratégicas para a conservação da biodiversidade paranaense	1:25.000	Vetorial
Elevação	1:60.000	Matricial
Rodovias	1:50.000	Vetorial
Ferrovias	Não informado	Vetorial
Dutovias	Não informado	Vetorial
Uso da terra	~1:50.000	Matricial
Aeródromos	Não informado	Vetorial
Áreas indígenas e quilombolas	Não informado	Vetorial
Unidades de conservação	Não informado	Vetorial

FONTE: O autor (2020).

Com relação à escala, optou-se por utilizar um tamanho de pixel de 30 m para o desenvolvimento dos estudos, uma vez que a faixa de domínio de empreendimentos lineares estão compreendidos nesta faixa de valor.

Um outro tipo de homogeneização necessária é em relação aos dados categóricos, que possuem na maioria das vezes, descrição de características importantes que justificam sua utilização em seus atributos. Como a neuralnet demanda valores numéricos para o treinamento, foi realizada a conversão dos dados relevantes na tabela de atributos, para posterior conversão desses dados categóricos para dados numéricos, utilizando a ferramenta de reclassificação do *software* QGIS, na linguagem *Python*. Vale ressaltar também que todas as variáveis devem estar georreferenciadas em uma mesma projeção e sistema de referência.

Por exemplo, a camada uso do solo será classificada em:

1. Floresta;
2. Formação natural não florestal;
3. Agropecuária;
4. Área não vegetada (antropizadas);
5. Corpo d'água;

Essas classes são originadas da própria fonte dos dados de referência (o projeto MapBiomass), onde nesses macros grupos estão representados todos os 17 tipos de cobertura da terra disponibilizado pelo órgão. Tal agrupamento possui características semelhantes intragrupos e diferentes entre as 5 macros classes, sendo assim o ideal para essa fonte de dados.

Como resultado da reclassificação, haverá uma imagem *raster* contendo pixels cujos valores variam de 1 a 5 que representam as classificações apresentadas na tabela de atributos.

Para a utilização das informações referentes a distâncias à empreendimentos lineares já existentes (como rodovias, ferrovias e dutovias), se faz necessária a confecção da camada de *buffers* a estes empreendimentos existentes, composto pela faixa de domínio de cada tipo de empreendimento.

Finalizadas as conversões, faz-se o cálculo das estatísticas descritivas sobre cada uma das camadas de entrada. Para isso, utiliza-se um *script* desenvolvido em Python e aplicado na plataforma QGIS. O resultado desse algoritmo será um arquivo *.txt* contendo dados como: média, desvio-padrão, moda, mediana, contagem de pixels, entre outros.

Para o estudo, propõem-se a utilização da média (para dados de natureza numérica) e da mediana (para dados de natureza categórica). Isso porque a mediana

ressalta informações de baixo ruído e a média diminui o efeito destes ruídos (ou outliers) presentes nos dados de entrada do modelo.

3.2.2 Amostragem dos dados de treinamento e de teste

Para a amostragem dos dados de teste e de validação, utilizou-se outro *script* em Python (esse já implementado na plataforma) para selecionar aleatoriamente pixels para a entrada no algoritmo da RNA. Para garantir a completude (relacionadas a presença ou não de LTs), fez-se a separação da camada dos vetores das linhas de transmissão já existentes na área de estudo (rasterizadas) em dois mapas: um contendo apenas as amostras positivas e outro com as negativas e, a partir deles, faz-se a coleta de dados.

Vale lembrar que, para o processo de validação, utiliza-se informações de vetores rasterizados que não entram no processo de treinamento e, portanto, é necessária uma pré-separação dos dados antes da geração das amostras, para que não haja tendência no modelo funcional calculado.

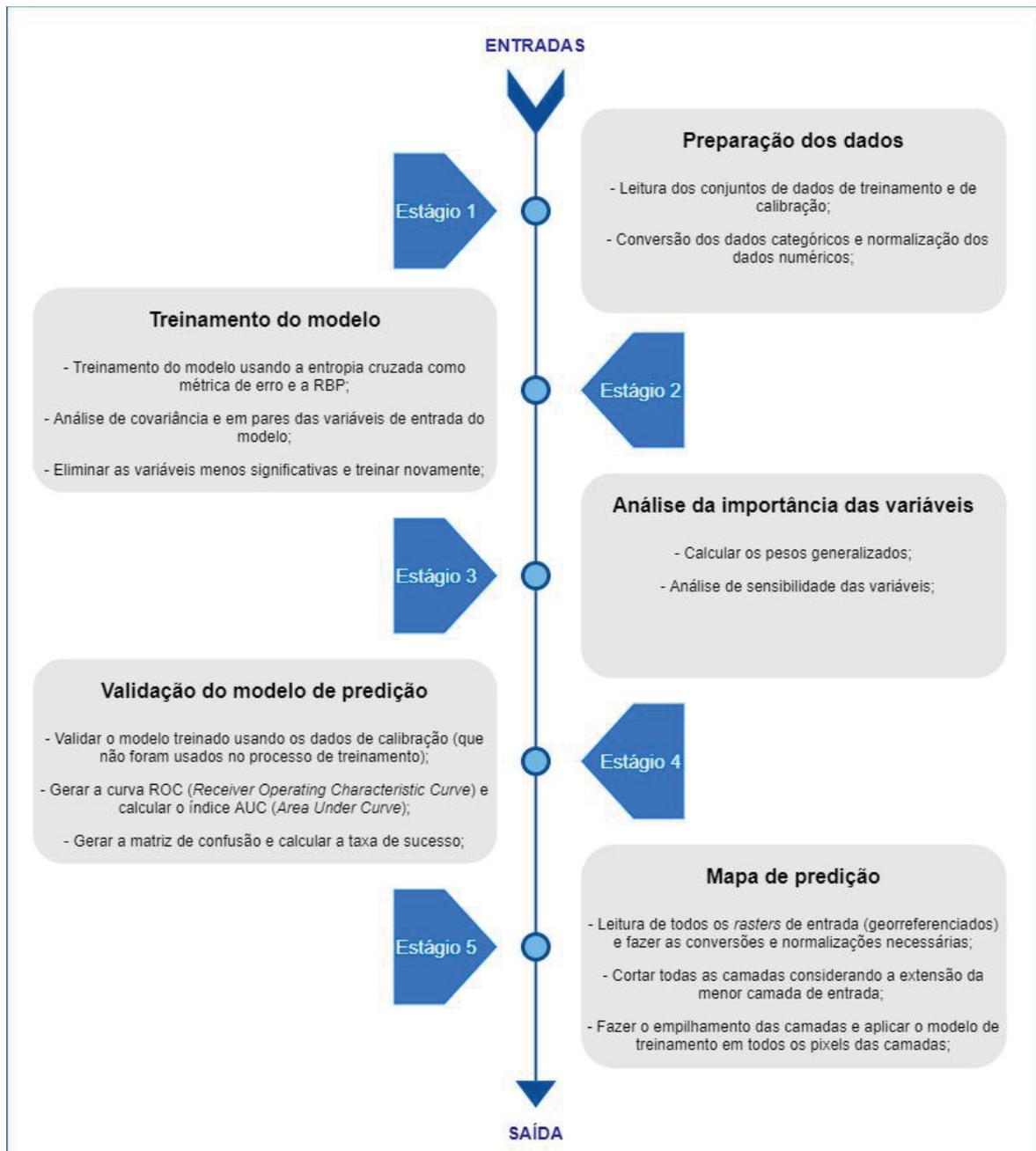
O resultado final da amostragem são dois arquivos .csv (um para o treinamento e outro para a validação) contendo a primeira coluna com informações sobre a amostra (0 se for negativa ou 1 se for positiva) e as demais colunas contêm as informações numéricas das demais variáveis (cada covariável em uma coluna) e o número de linhas é a quantidade de amostras utilizadas. Vale ressaltar que quanto mais amostras, melhor a estimativa preditiva. Porém, o aumento do número de amostras refletirá exponencialmente no aumento do custo computacional da RNA.

3.2.3 Aplicação da RNA

As etapas da utilização do algoritmo Neuralnet contendo suas etapas pode ser observado na FIGURA 18.

As particularidades dos estágios mostrados na FIGURA 18 serão abordados entre as seções 3.2.3.1 a 3.2.3.4 a seguir.

FIGURA 18 – FLUXO DO ALGORITMO DE APLICAÇÃO DA REDE NEURAL



FONTE: O autor (2020).

3.2.3.1 Normalização dos dados numéricos de entrada e adaptação dos dados categóricos

Para que a análise dos diferentes fatores que influenciam na definição do traçado de uma linha de transmissão, todos os dados devem estar em uma mesma escala de comparação. Então, para solucionar esse problema nas variáveis numéricas, utiliza-se a relação matemática conforme a EQUAÇÃO (35).

$$x_{i_{norm}} = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} \quad (35)$$

Com isso, o novo intervalo dos valores será de [0,1], porém positivos e negativos, dependendo o tipo de dados numérico de entrada.

Além desta etapa, as variáveis categóricas também deverão ser transformadas para o intervalo supracitado. Neste caso, uma camada contendo esse tipo de dados dá origem à n outras camadas binárias (sendo 0 a ausência daquele fenômeno e 1 a presença na área de estudo). No caso da variável “Uso da terra”, haverá cinco entradas binárias na RNA ao invés de uma apenas, o que resultará um peso para cada classificação originada por essa camada.

3.2.3.2 Treinamento da RNA

Com os dados já no intervalo correto faz-se a aplicação da RNA. Para a análise, optou-se por utilizar o algoritmo da *Resilient Backpropagation*, e como a métrica de erros sendo a *cross-entropy*, conforme descrito na seção 2.4, uma vez que suas otimizações em relação ao tradicional implicam na redução de custo computacional e de tempo de treinamento. Os pesos de entrada para a primeira iteração são inicializados com valores aleatórios retirados de uma distribuição normal padrão de acordo com a recomendação de Günther e Fritsch (2010).

Após a realização do primeiro treinamento com as variáveis, ou seja, quando o erro na saída calculada for menor que 10^{-3} (ou 0,1%) ou quando o número de iterações atinja o valor de 10^{+8} (obedecendo a EQUAÇÃO (20)), faz-se a elaboração dos fatores de análise de qualidade dos dados de entrada para a RNA. Para isso, é construído um gráfico denominado *pairwise* ou análise em pares de variáveis, o gráfico de análise de sensibilidade de cada covariável e por fim, a geração dos pesos generalizados.

A análise em pares de variáveis é um gráfico que imprime a relação entre todas as variáveis, contendo na diagonal principal as camadas, no triângulo superior as correlações e no triângulo inferior o gráfico da covariância entre as variáveis analisadas. Segundo Wasserman (2004), a formulação matemática para estes índices para duas variáveis A e B , podem ser escritas pelas EQUAÇÃO (36) e pela EQUAÇÃO (37).

$$Cov(A, B) = \frac{\sum(A_i - \bar{A})(B_i - \bar{B})}{n} \quad (36)$$

Um outro índice que pode ser obtido com a ajuda da covariância e que também deve ser considerado na análise do modelo é o índice de correlação ρ , que pode ser obtido pela EQUAÇÃO (37).

$$\rho(A, B) = \frac{Cov(A, B)}{\sigma_A \sigma_B} \quad (37)$$

Onde:

$\sigma_A \sigma_B$: são os respectivos desvios-padrão das variáveis A e B avaliadas.

Já o valor do peso generalizado é calculado de acordo com Intrator e Intrator (2001) *apud* Günther e Fritsch (2010), e é definido de acordo com a contribuição de cada i -enésima covariável à log-odds (EQUAÇÃO (38)).

$$\tilde{w}_i = \frac{\partial \log \left(\frac{o(x)}{1 - o(x)} \right)}{\partial x_i} \quad (38)$$

O peso generalizado \tilde{w}_i expressa o efeito de cada x_i covariável na variável resposta $o(x)$. Este fator é dependente de todas as variáveis e sua distribuição indica se o efeito da covariável é linear no modelo (valores próximos de zero).

O último índice a ser considerado na avaliação das camadas é a sensibilidade. Esta análise mostra o relacionamento entre a variável resposta e seus fatores condicionantes. O estudo consiste na elaboração de um gráfico para cada variável envolvida no estudo, plotando uma linha para cada situação a seguir:

- a) 100% da variável analisada e 0% das demais variáveis;
- b) 80% da variável analisada e 20% das demais variáveis;
- c) 60% da variável analisada e 40% das demais variáveis;
- d) 40% da variável analisada e 60% das demais variáveis;
- e) 20% da variável analisada e 80% das demais variáveis;

A partir do comportamento do gráfico, é possível identificar se a variável analisada permite explicar por si só grande parte do fenômeno avaliado ou se é afetada pelas demais camadas. Se for afetada, significa que esta não pode ser

analisada isoladamente, ou seja, caso não haja uma correlação significativa (acima de 80% por exemplo), esta variável não poderá ser eliminada do cálculo do modelo preditivo.

A somatória da análise de todos os indicadores resulta na eliminação de camadas não-significantes para a análise e com isso, tendem a melhorar a taxa de sucesso do modelo calculado. A taxa de sucesso é dada pela EQUAÇÃO (39).

$$Tx_{sucesso} = \frac{Acertos_{falsos} + Acertos_{verdadeiros}}{total\ de\ amostras\ de\ treinamento} \quad (39)$$

A cada eliminação de fatores, deve-se realizar o treinamento novamente e refazer as análises acerca das variáveis e é necessária a repetição até que não haja mais ganho na taxa de sucesso ou que não haja mais variável não-significativa no cálculo do modelo preditivo.

3.2.3.3 Validação do algoritmo resultante da RNA

O processo de validação é feito utilizando conceitos de curva ROC e de AUC utilizadas em algoritmos de aprendizagem, de acordo com as descrições discutidas na seção 2.5. Estas métricas irão apresentar a qualidade do modelo treinado, assim como sua eficiência.

De maneira geral, o valor que o cálculo de AUC pode assumir está entre os valores de 0 e 1. AUC máximo = 1 significa que o modelo calculado pela RNA é perfeito para predição de novas LTs, dados aquelas variáveis de entrada. Isso implica que os erros associados com falsos positivos e falsos negativos é nulo. Como é improvável que isso aconteça, assume-se que quando mais próximo de um, mais eficiente o modelo de predição calculado será para o objetivo do estudo.

3.2.3.4 Geração do mapa preditivo das áreas propícias à construção de novas linhas de transmissão

Para compor a superfície preditiva a partir do modelo de predição treinado, é necessário aplicar este em todos os dados de entrada, na extensão total da área de pesquisa. Para tanto, faz-se a leitura de todos os arquivos *rasters* que compõem as entradas do aprendizado. A partir da variável que está descrita pela menor imagem,

faz-se o recorte das demais imagens que representam as outras camadas. Caso não haja uma imagem que contenha as menores dimensões, faz-se o recorte a partir das maiores coordenadas em X e de Y (contado a partir do canto inferior esquerdo) até as menores em X e de Y (contado a partir do canto superior direito), garantindo assim, a maior área de sobreposição total das camadas.

Tal procedimento se faz necessário, uma vez que, para aplicar o modelo treinado através da função “*compute*” da neuralnet, é necessário garantir que haja valores para todas as variáveis em todos os pixels no momento da realização do empilhamento das camadas *raster*.

FIGURA 19 – EXEMPLO DE MAPA DE PREDIÇÃO GERADO PELA RNA NEURALNET



FONTE: O autor (2020).

Então, garantido isso, aplica-se o modelo no conjunto de *rasters* de entrada e gera-se o modelo predito para toda a área analisada. Na figura 19 está apresentado um exemplo amostral de uma predição gerada para uma pequena área.

Este modelo predito consiste em uma imagem com valores que estão compreendidos entre 0 e 1, cujos valores podem ser interpretados como a probabilidade de ocorrência de linhas de transmissão de energia naquele pixel. Valores mais próximos de 1 são ditos como mais propícios a se ter a passagem de uma LT, enquanto valores no outro extremo (próximo de 0) representam áreas com fatores limitantes em relação à construção de novas linhas.

3.2.4 Otimização do modelo predito

A otimização do modelo predito é feita para que se minimize os custos ambientais, sociais e econômicos de um futuro empreendimento, evitando ou minimizando a passagem do novo traçado em áreas proibidas ou com restrições legais. Estas áreas que compõe a superfície de restrição são:

- Áreas alagadas;
- Áreas militares;
- Áreas de proteção integral;
- Parques estaduais e municipais;
- Áreas de mineração;
- Sítios arqueológicos;
- Áreas de aeroportos e aeródromos;
- Áreas com restrição jurídica;
- Áreas com declividade acima de 45° ou 100%;

Estas camadas são definidas como base em legislação municipal, estadual ou federal ou, devido à restrição física.

Para cada área, são geradas superfícies binárias (0 se não há a presença e 1 se houver) para toda a área. Em seguida essas superfícies são somadas e, caso haja valores superiores a 1, faz-se a reclassificação para a que o intervalo esteja compreendido entre [0,1].

O modelo matemático para a geração da superfície preditiva ótima a partir das superfícies predita e restritiva, é dado pela EQUAÇÃO (40).

$$p_{\text{ótimo}} = (1 - p_{\text{predito}}) + p_{\text{restritivo}} \quad (40)$$

Tendo a superfície de previsão ótima, segue-se para a elaboração do traçado final.

3.2.5 Geração da diretriz de traçado otimizado de uma LT

A confecção do traçado final é feita sobre a camada ou superfície ótima anteriormente calculada, aplicando-se o algoritmo *Least Cost Path* conforme explicado na seção 2.2.1.1.1. Sendo assim, a partir de dois pontos (subestações de energia), procura-se o caminho que minimize os custos acumulados sobre a superfície, ligando os centroides dos pixels vizinhos que descrevem o caminho calculado.

Os valores dos pixels sobre a superfície otimizada final podem variar entre [0,2] e, invertendo a lógica depois da aplicação da EQUAÇÃO (40), buscam-se os menores valores como sendo as áreas com maior aptidão para receber uma nova LT.

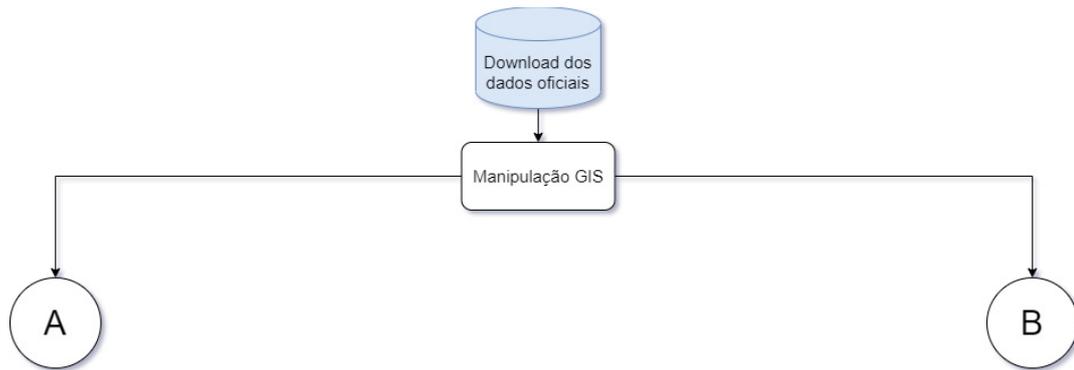
3.2.6 Comparação de rotas alternativas

Ao final da geração da diretriz de traçado otimizado, a qual não participou dos processos de treinamento e de validação, será feita a comparação das linhas geradas pelo método proposto e por metodologias tradicionais que usam pesos empíricos na sua concepção. As possíveis diferenças serão discutidas e avaliadas de acordo com as suas respectivas causas e consequências. E, através destas análises faz-se a discussão dos resultados e as conclusões da pesquisa.

3.3 PREPARAÇÃO DOS DADOS DE ENTRADA

A partir dos dados oficiais do Brasil, fez-se a manipulação dos dados utilizando Sistemas de Informação Geográficas em dois momentos da pesquisa: na preparação das camadas das covariáveis e na geração das amostras de treinamento e teste, conforme mostrado na Figura 20.

FIGURA 20 – PREPARAÇÃO DOS DADOS



FONTE: O autor (2020).

Com a finalidade de facilitar a visualização, a Figura 20 foi dividida para mostrar as etapas de preparação dos dados para entrada na RNA.

3.3.1 Preparação das camadas covariáveis

As camadas de dados espaciais disponibilizados pelos órgãos oficiais brasileiros são de formato vetorial, na maior parte, principalmente devido à natureza do fenômeno representado. Essas informações vetoriais precisam passar por etapas de rasterização para serem manipuladas pela maioria dos algoritmos de RNA, que normalmente trabalham com números.

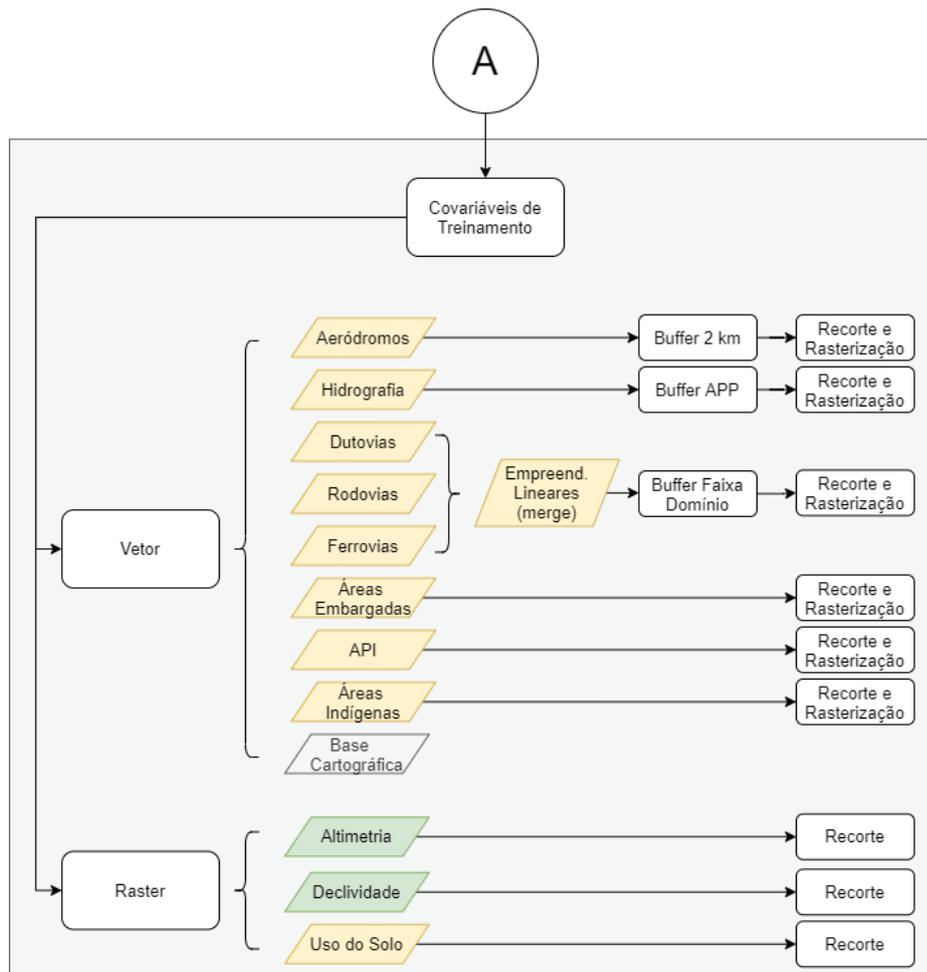
Então, para converter dados espaciais pontuais e dados lineares, fez-se o uso de *buffers* com relação à ocorrência dos fenômenos. As dimensões dos *offsets* foram baseadas nas legislações vigentes, como o Código Florestal Brasileiro, Normas do DNIT entre outros órgãos responsáveis pela infraestrutura brasileira. Em caso de não enquadramento ou dúvidas sobre a classificação da linha, considerou-se o valor mínimo de 30 metros (15 metros para cada lado), a fim de se adequar ao tamanho mínimo de detalhamento adotado na pesquisa. Já para os aeródromos, adotou-se 2 mil metros de *buffer*, pois os dados eram pontuais e representavam o centro geométrico da área da pista que poderiam possuir grandes extensões.

As informações de dutovias, rodovias e ferrovias foram unidas em um único *layer* antes da entrada na RNA, pois o significado de se ter ou não esses fenômenos era o mesmo: todos possuem faixa de domínio pública ou preservada já consolidada.

As camadas com fenômenos representados com a primitiva gráfica de área (ou polígono) foram rasterizadas diretamente da informação original, utilizando um tamanho de pixel de 30 metros.

A Figura 21 apresenta a preparação das camadas de entrada em ambiente GIS.

FIGURA 21 – PREPARAÇÃO DAS CAMADAS



FONTE: O autor (2020).

Na Figura 20, as camadas representadas na cor verde foram tratadas como dados discretos na entrada da rede, ou seja, não é necessária a homogeneização do fenômeno representado. Já para as camadas em amarelo, fez-se necessário a subdivisão em camadas binárias para a entrada na rede dos *layers* de uso do solo e de áreas ambientais embargadas ou com restrição ambiental.

Já os dados espaciais, disponibilizados em formato *raster*, já possuíam as características necessárias e o tamanho de pixel compatível com a análise e, por isso, o processo de preparação foi limitado apenas ao recorte da área.

3.3.2 Preparação das amostras de treinamento e teste

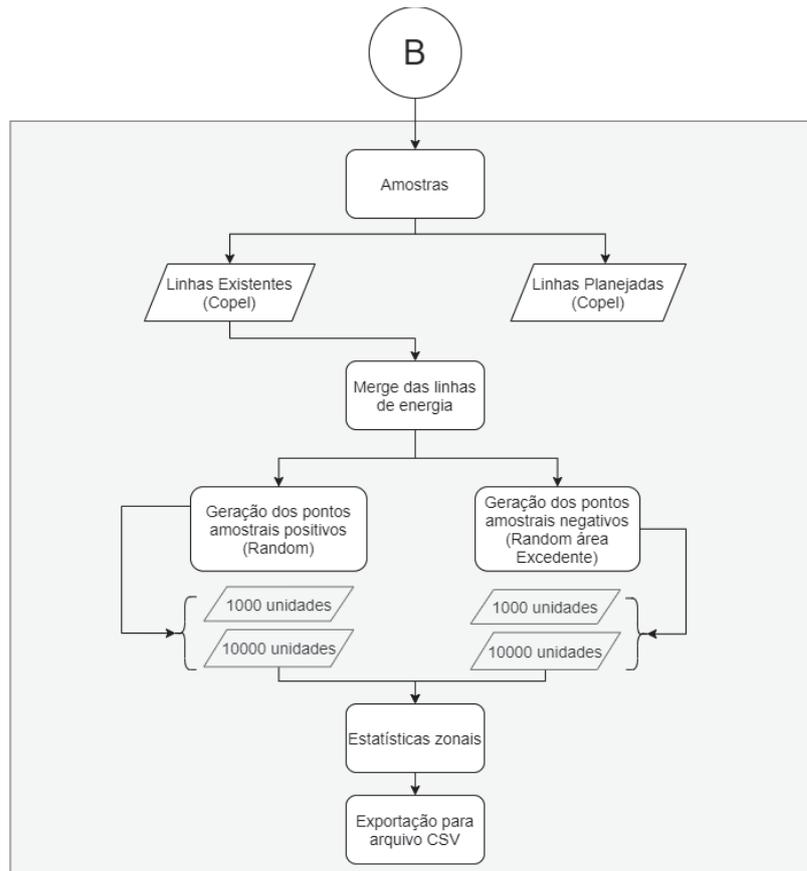
Assim como os dados dos empreendimentos lineares, fez-se a utilização da operação *merge* para os dados recebidos da Companhia Paranaense de Energia, pois para a pesquisa, o que necessitava era a informação da presença ou não do fenômeno no espaço geográfico. A Figura 22 mostra as etapas de geração das amostras.

A camada unida de linhas de energia existentes foi utilizada para a geração das amostras positivas e negativas de entrada na RNA. Para as amostras positivas, fez-se a amostragem randômica utilizando como base de localização, os vértices das linhas existentes, a fim de garantir a região central das faixas de ocupação dos empreendimentos. Já para as amostras negativas, fez-se a amostragem randômica nas regiões fora da faixa de domínio dos vetores existentes, gerado a partir de uma operação entre as camadas de *merge* das faixas de domínio e a camada da área de estudo.

Após a geração das amostras positivas e negativas, fez-se a aplicação de dois algoritmos de geração automática dos arquivos de treinamento e teste da rede, disponíveis nos apêndices desta pesquisa, desenvolvido em linguagem Python. Este algoritmo calcula as estatísticas zonais (média para dados discretos e mediana para dados categóricos) em uma região com raio de 30 metros ao redor do ponto amostral, sumarizando em uma lista, tais estatísticas referentes à todas as imagens das camadas de entrada. Em seguida, as amostras positivas e negativas são unidas em uma única tabela e randomicamente subdivididas em amostras de treinamento e de teste, de acordo com uma porcentagem pré-estabelecida.

Concluído o processo de amostragem, parte-se para o treinamento propriamente dito e para a análise dos resultados.

FIGURA 22 – PREPARAÇÃO DAS AMOSTRAS

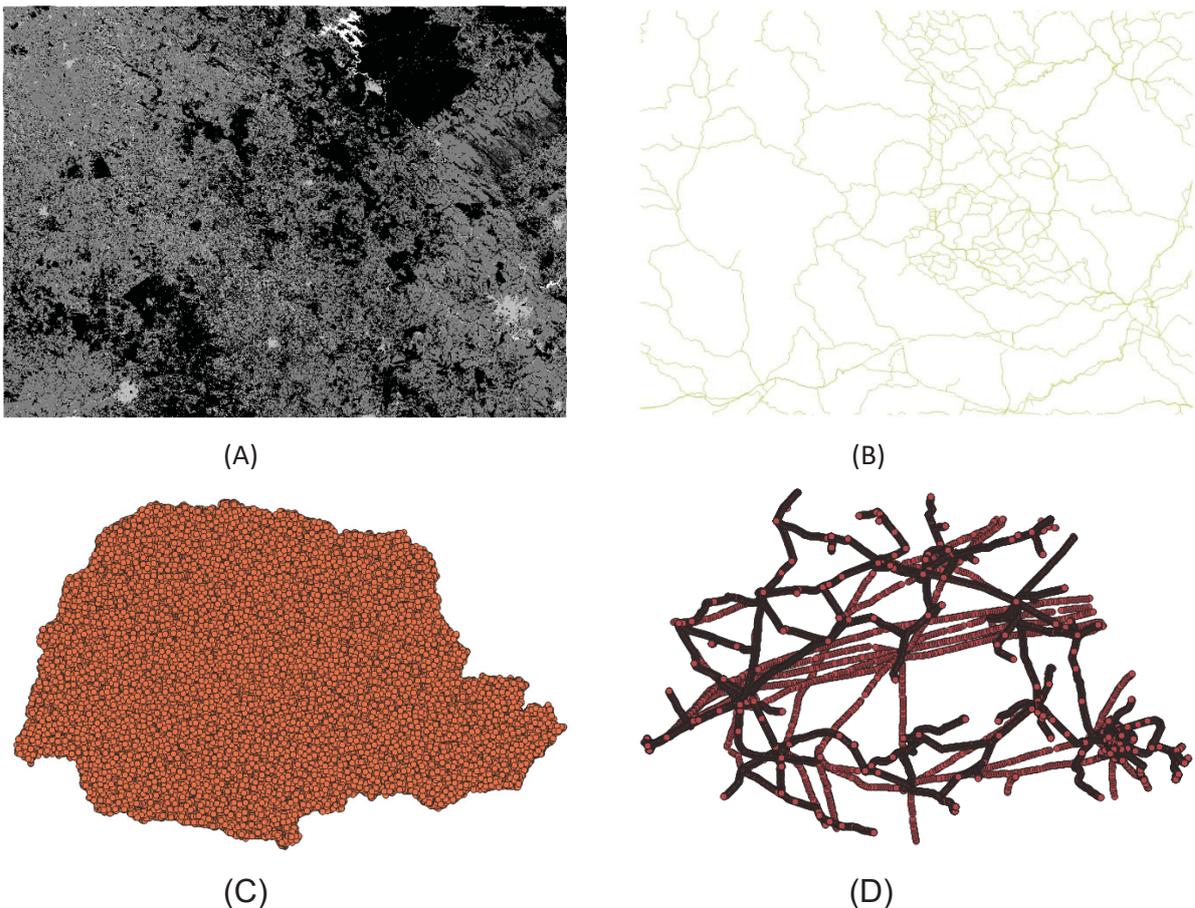


FONTE: O autor (2020).

4 EXPERIMENTOS E ANÁLISE DOS RESULTADOS

Para fins comparativos, fez-se a variação da aplicação da RNA alterando o número de neurônios ocultos empregado na construção da rede e foi definido o número máximo de iterações para a convergência dos pesos no modelo matemático. A Figura 23 mostra exemplos de dados vetoriais e *raster*, bem como as amostras positivas e negativas empregadas no treinamento da RNA.

FIGURA 23 – DADOS VETORIAIS, IMAGENS RASTER E AS AMOSTRAS POSITIVAS E NEGATIVAS EMPREGADAS NO TREINAMENTO DA RNA. (A) RESULTADO DA CLASSIFICAÇÃO DA CAMADA USO DO SOLO. (B) CAMADA DE DISTÂNCIA À EMPREENDIMENTOS LINEARES. (C) AMOSTRA NEGATIVA. (D) AMOSTRA POSITIVA.



Ao observar a Figura 23 é possível notar que as amostras positivas são representadas por um conjunto de pontos extraídos aleatória e automaticamente sobre os buffers do *shapefile* que contém linhas de transmissão já existentes no

estado do Paraná (este arquivo foi disponibilizado pela COPEL), enquanto as amostras negativas são representadas por um conjunto de pontos extraídos aleatória e automaticamente no *shapefile* do estado do Paraná. Isto é feito usando a função de “pontos randômicos em polígonos” do QGIS”.

4.1 PESOS CALCULADOS

4.1.1 Utilizando 5 neurônios ocultos

A utilização de mais ou menos *layers* ocultos implicam em sumarizar (mais ou menos, da mesma forma) relações e interações entre as diferentes camadas de entrada. A utilização de um número insuficiente de *layers* implica em problemas de *underfitting*, ou seja, o modelo está “subjugando” os dados de treinamento quando o desempenho não é satisfatório. Isso significa que a RNA é incapaz de capturar a relação entre o modelo e os exemplos de entrada.

Entretanto, a utilização de quantidades grandes de *layers* intrínsecos à RNA pode hiperparametrizar (*overfitting*) as relações entre as variáveis e o modelo matemático. Isso pode melhorar as métricas de erro calculado pela rede nos dados de treinamento, porém não reflete esse mesmo ganho nos dados de teste, ou seja, o modelo se torna incapaz de generalizar os exemplos diferentes das amostras de entrada.

Para iniciar as análises, fez-se o estudo da quantidade ideal de neurônios ocultos e número de iterações máximas necessárias para o treinamento da RNA. O primeiro treinamento foi utilizando 2 mil amostras e 5 neurônios ocultos. A sumarização dos pesos ou relacionamentos entre as camadas está apresentado na Tabela 11. Na tabela, a sigla NO significa “Neurônio Oculto”.

TABELA 11 – PESOS CALCULADOS PARA 5 NEURÔNIOS OCULTOS

INFORMAÇÃO	VALOR 2 MIL	VALOR 20 MIL
Erro	7.491114 e+02	1.704209 e+03
Limiar	9.978749 e-03	9.665460 e-03
Iterações	1.476850 e+05	1.324940 e+05
Bias para NO 1	2.309333 e+01	-1.508528 e+00
Altimetria para NO 1	-3.555345 e+01	8.268082 e+00
Declividade para NO 1	-5.420385 e+00	1.760198 e+01

INFORMAÇÃO	VALOR 2 MIL	VALOR 20 MIL
APP para NO 1	7.495117 e+02	1.841511 e+01
Empr. Linear para NO 1	7.531792 e+02	1.937183 e+01
Área de Proteção Integral para NO 1	3.701058 e+01	-7.061584 e+00
Área Indígena para NO 1	-5.334376 e+02	-6.065164 e+00
Aeródromo para NO 1	-1.945468 e+02	-7.129296 e+02
Floresta (L0) para NO 1	1.401224 e+01	3.808518 e+01
Formação Natural não Florestal (L1) para NO 1	-7.134154 e+02	-1.706166 e+00
Agropecuária (L2) para NO 1	1.163408 e+01	4.420132 e-01
Área Antropizada (L3) para NO 1	2.341770 e+02	-7.368206 e+02
Corpos D'Água (L4) para NO 1	-7.111312 e+02	-2.477176 e+01
Outros usos (L5) para NO 1	4.202387 e+01	-7.329565 e+02
Bias para NO 2	7.060168 e+00	-7.134203 e+00
Altimetria para NO 2	-5.652711 e+01	4.099654 e+01
Declividade para NO 2	-3.784331 e+01	1.773234 e+01
APP para NO 2	6.419474 e+00	-2.847143 e+01
Empr. Linear para NO 2	3.036483 e+00	-2.718972 e+01
Área de Proteção Integral para NO 2	-7.031399 e+02	-4.738822 e+01
Área Indígena para NO 2	5.128942 e+01	1.622304 e+00
Aeródromo para NO 2	5.359448 e+00	4.481377 e+01
Floresta (L0) para NO 2	-2.623822 e+00	-7.432938 e+02
Formação Natural não Florestal (L1) para NO 2	2.039478 e+01	-2.359791 e+01
Agropecuária (L2) para NO 2	2.914305 e+01	2.301865 e+01
Área Antropizada (L3) para NO 2	2.015931 e+01	7.265701 e+00
Corpos D'Água (L4) para NO 2	2.363349 e+01	-1.903721 e+01
Outros usos (L5) para NO 2	2.684818 e+01	-1.980380 e+01
Bias para NO 3	-5.839060 e+01	-3.007698 e+00
Altimetria para NO 3	2.485541 e+00	7.517545 e+00
Declividade para NO 3	1.204082 e+03	7.581632 e+00
APP para NO 3	-1.002449 e+03	-1.280570 e+01
Empr. Linear para NO 3	2.074837 e+02	3.654666 e+01
Área de Proteção Integral para NO 3	-8.218122 e+02	9.504192 e-01
Área Indígena para NO 3	1.890049 e+02	-2.788304 e+01
Aeródromo para NO 3	-7.918275 e+02	6.572206 e+01
Floresta (L0) para NO 3	9.423833 e+01	-7.182428 e+02
Formação Natural não Florestal (L1) para NO 3	-6.922096 e+02	2.186462 e+01
Agropecuária (L2) para NO 3	-1.594120 e+02	-4.909468 e+00
Área Antropizada (L3) para NO 3	7.840214 e+01	-3.101417 e+00
Corpos D'Água (L4) para NO 3	-2.829972 e+02	1.299095 e+01
Outros usos (L5) para NO 3	-8.186297 e+02	6.553985 e+01
Bias para NO 4	5.218623 e-01	-2.097460 e+00
Altimetria para NO 4	-2.148064 e+03	-2.942418 e+00
Declividade para NO 4	2.352894 e+03	-3.464997 e+00
APP para NO 4	-1.560128 e+03	-6.552866 e-01
Empr. Linear para NO 4	4.185730 e+02	-1.043028 e+00

INFORMAÇÃO	VALOR 2 MIL	VALOR 20 MIL
Área de Proteção Integral para NO 4	-6.904300 e+02	-1.792213 e+02
Área Indígena para NO 4	1.274232 e+02	3.800685 e+01
Aeródromo para NO 4	-3.923996 e+02	7.847405 e+02
Floresta (L0) para NO 4	-4.226895 e+02	-7.099171 e+02
Formação Natural não Florestal (L1) para NO 4	1.837806 e+03	3.598686 e+02
Agropecuária (L2) para NO 4	1.174708 e+03	1.829661 e+02
Área Antropizada (L3) para NO 4	7.197934 e+02	3.051384 e+00
Corpos D'Água (L4) para NO 4	1.671102 e+03	-7.438647 e+02
Outros usos (L5) para NO 4	-5.919708 e+02	2.542812 e+00
Bias para NO 5	1.326481 e+02	6.015155 e+00
Altimetria para NO 5	-9.914054 e+02	-1.974240 e+01
Declividade para NO 5	4.858915 e+03	-8.147145 e+00
APP para NO 5	-1.837474 e+03	8.590489 e+00
Empr. Linear para NO 5	4.757561 e+02	5.528520 e+00
Área de Proteção Integral para NO 5	-1.120290 e+03	4.797879 e+01
Área Indígena para NO 5	3.189514 e+01	-7.105314 e+02
Aeródromo para NO 5	-9.591503 e+02	7.555606 e+02
Floresta (L0) para NO 5	-3.515842 e-01	-7.157617 e+02
Formação Natural não Florestal (L1) para NO 5	1.949991 e+03	7.577983 e+02
Agropecuária (L2) para NO 5	5.081643 e+00	5.424539 e+01
Área Antropizada (L3) para NO 5	5.787583 e+02	-1.246654 e+00
Corpos D'Água (L4) para NO 5	-8.765207 e+02	5.206143 e+00
Outros usos (L5) para NO 5	3.799902 e+02	3.263937 e+00
Bias para Predição	3.103788 e+00	7.732555 e+00
NO 1 para Predição	1.986534 e+01	1.431349 e+01
NO 2 para Predição	-6.719429 e+03	-4.954394 e+00
NO 3 para Predição	-2.225250 e+01	-4.821245 e+00
NO 4 para Predição	-3.049980 e+00	-1.249596 e+01
NO 5 para Predição	-6.712991 e-01	-6.403180 e+00

FONTE: O autor (2020).

A representação gráfica da estrutura da RNA utilizando 5 *layers* ocultos pode ser observada na Figura 24. Cada linha representa uma relação de peso entre as covariáveis de entrada, os *layers* ocultos e a variável resposta (preditiva), além de apresentar os valores do *Bias*, o limiar e o erro calculado pela rede.

Vale ressaltar que a utilização de redes neurais artificiais para a estimativa automática de pesos de variáveis para esse tipo de aplicação preditiva não resulta em um único para uma variável, como encontrado nas metodologias tradicionais. Isso se deve ao fato que para aplicações de aproximação de função com a utilização de neurônios ocultos - NO, calcula-se todas as relações entre as variáveis de entrada e

variação do número de amostras também foi utilizada para entender o comportamento das métricas de erro com a inserção de amostras diferentes dentro do modelo.

TABELA 12 – PESOS CALCULADOS PARA 10 NEURÔNIOS OCULTOS

INFORMAÇÃO	VALOR 2 MIL	VALOR 20 MIL
Erro	7.165131 e+02	1.624318 e+03
Limiar	9.960286 e-03	9.715591 e-03
Iterações	1.203610 e+05	2.390660 e+05
Bias para NO 1	3.952688 e+00	-8.801458 e-01
Altimetria para NO 1	-5.305115 e+00	-8.997449 e+00
Declividade para NO 1	-1.946822 e+01	7.621662 e+00
Hidrografia para NO 1	2.627928 e+02	-7.857537 e-02
Empr. Linear para NO 1	-1.500463 e+02	2.047762 e+00
Áreas de Proteção Integral para NO 1	3.841058 e+01	5.405178 e+00
Áreas Indígenas para NO 1	-5.323376 e+02	9.207708 e+01
Aeródromo para NO 1	3.303876 e+01	-7.103296 e+02
Floresta (L0) para NO 1	4.822586 e+00	4.618518 e+01
Formação Natural não Florestal (L1) para NO 1	1.913577 e+02	4.717853 e+00
Agropecuária (L2) para NO 1	3.949812 e+00	-7.405013 e-01
Área Antropizada (L3) para NO 1	4.270953 e+01	-7.967686 e+02
Corpos D'Água (L4) para NO 1	6.519174 e+00	-5.102526 e+01
Outros usos (L5) para NO 1	3.897561 e+01	-7.102634 e+02
Bias para NO 2	-3.233063 e-01	3.974382 e+02
Altimetria para NO 2	-1.545182 e+00	-1.412006 e+03
Declividade para NO 2	-2.260206 e+01	3.382698 e+02
Hidrografia para NO 2	8.842371 e+01	1.218500 e+01
Empr. Linear para NO 2	-7.132922 e+02	-6.913542 e+02
Áreas de Proteção Integral para NO 2	4.316012 e+01	2.808018 e+02
Áreas Indígenas para NO 2	-4.767106 e+02	1.551567 e+02
Aeródromo para NO 2	3.646017 e+01	-6.901862 e+02
Floresta (L0) para NO 2	-9.309561 e-01	1.185062 e+02
Formação Natural não Florestal (L1) para NO 2	-8.204884 e+01	3.007928 e+01
Agropecuária (L2) para NO 2	5.037759 e+00	2.965955 e+02
Área Antropizada (L3) para NO 2	2.828849 e+00	-1.125213 e+03
Corpos D'Água (L4) para NO 2	8.491481 e+00	3.562434 e+01
Outros usos (L5) para NO 2	4.120332 e+01	-1.132290 e+03
Bias para NO 3	-9.884298 e+01	-7.486419 e-01
Altimetria para NO 3	4.447894 e+02	-2.186863 e+00
Declividade para NO 3	2.638372 e+02	-9.784521 e+00
Hidrografia para NO 3	5.195818 e+01	3.711812 e+00
Empr. Linear para NO 3	3.017518 e+01	1.346238 e+01
Áreas de Proteção Integral para NO 3	2.875621 e+01	-7.132064 e+02
Áreas Indígenas para NO 3	-5.595635 e+02	-1.757794 e+00
Aeródromo para NO 3	-5.233456 e+02	-7.131463 e+02

INFORMAÇÃO	VALOR 2 MIL	VALOR 20 MIL
Floresta (L0) para NO 3	2.209446 e+00	5.522555 e+01
Formação Natural não Florestal (L1) para NO 3	-1.974302 e+02	4.695391 e+00
Agropecuária (L2) para NO 3	2.499973 e+01	3.251583 e+00
Área Antropizada (L3) para NO 3	3.294473 e+02	-7.202748 e+02
Corpos D'Água (L4) para NO 3	-2.604975 e+02	-1.489685 e+01
Outros usos (L5) para NO 3	-1.862663 e+02	-4.237558 e-02
Bias para NO 4	-1.600781 e-01	7.693538 e-01
Altimetria para NO 4	3.324683 e+02	6.961068 e+00
Declividade para NO 4	-1.188575 e+04	6.275476 e+00
Hidrografia para NO 4	4.369460 e+02	-1.479794 e+00
Empr. Linear para NO 4	-5.995270 e+02	-2.155886 e+00
Áreas de Proteção Integral para NO 4	-6.681300 e+02	-5.292398 e+00
Áreas Indígenas para NO 4	4.759023 e+03	-3.744303 e-01
Aeródromo para NO 4	-6.555080 e+02	3.804052 e+01
Floresta (L0) para NO 4	-1.223424 e+02	-7.241171 e+02
Formação Natural não Florestal (L1) para NO 4	3.901846 e+02	6.815914 e+01
Agropecuária (L2) para NO 4	-7.346748 e+02	6.408199 e+01
Área Antropizada (L3) para NO 4	-1.002738 e+02	-2.614314 e+00
Corpos D'Água (L4) para NO 4	4.690192 e+02	-3.824772 e+00
Outros usos (L5) para NO 4	2.237322 e+02	3.837752 e+01
Bias para NO 5	1.624210 e+02	-3.327067 e+00
Altimetria para NO 5	-5.570895 e+02	5.525412 e+02
Declividade para NO 5	4.339024 e+02	4.013715 e+03
Hidrografia para NO 5	6.925529 e+01	-4.634297 e+02
Empr. Linear para NO 5	7.249071 e+01	-4.430978 e+02
Áreas de Proteção Integral para NO 5	-7.072903 e+02	-2.934803 e+02
Áreas Indígenas para NO 5	5.989514 e+01	3.404130 e+01
Aeródromo para NO 5	2.179699 e+01	4.406606 e+02
Floresta (L0) para NO 5	-1.003137 e+03	-8.299617 e+02
Formação Natural não Florestal (L1) para NO 5	-4.058548 e+00	2.261390 e+02
Agropecuária (L2) para NO 5	1.848117 e+02	4.406212 e+01
Área Antropizada (L3) para NO 5	9.479586 e+01	-5.816098 e+02
Corpos D'Água (L4) para NO 5	2.015265 e+02	-7.057434 e+01
Outros usos (L5) para NO 5	-6.965697 e+02	3.942820 e+01
Bias para NO 6	-6.985510 e+01	-4.445948 e-02
Altimetria para NO 6	1.012619 e+03	1.251089 e+01
Declividade para NO 6	-1.117897 e+03	4.428639 e+00
Hidrografia para NO 6	2.337878 e+01	-4.208438 e+00
Empr. Linear para NO 6	-2.894776 e+02	-2.906613 e+00
Áreas de Proteção Integral para NO 6	9.105410 e+01	-7.183524 e+02
Áreas Indígenas para NO 6	-5.578828 e+02	-9.846875 e-01
Aeródromo para NO 6	2.781635 e+01	-7.189384 e+02
Floresta (L0) para NO 6	2.709407 e+02	4.176902 e+01
Formação Natural não Florestal (L1) para NO 6	-2.158185 e+01	-5.421676 e+00

INFORMAÇÃO	VALOR 2 MIL	VALOR 20 MIL
Agropecuária (L2) para NO 6	-6.482401 e+02	7.459202 e+02
Área Antropizada (L3) para NO 6	-2.773884 e+02	-2.544938 e+00
Corpos D'Água (L4) para NO 6	-4.785108 e+02	-5.078152 e+00
Outros usos (L5) para NO 6	1.408277 e+01	-4.740950 e+00
Bias para NO 7	-1.018925 e+02	7.211595 e-01
Altimetria para NO 7	2.179544 e+02	-1.008867 e+01
Declividade para NO 7	-8.684938 e+00	-1.871676 e+01
Hidrografia para NO 7	-7.600236 e+02	4.140368 e+00
Empr. Linear para NO 7	-7.018406 e+02	1.771672 e+00
Áreas de Proteção Integral para NO 7	-7.079975 e+02	-7.075904 e+02
Áreas Indígenas para NO 7	8.402995 e+01	4.144992 e+01
Aeródromo para NO 7	8.200043 e+02	5.675498 e+01
Floresta (L0) para NO 7	8.160917 e+01	-7.124458 e+02
Formação Natural não Florestal (L1) para NO 7	-7.271580 e+01	7.567104 e+02
Agropecuária (L2) para NO 7	-1.019183 e+02	6.637935 e+00
Área Antropizada (L3) para NO 7	-5.727578 e+01	4.336329 e+00
Corpos D'Água (L4) para NO 7	-8.544836 e+01	3.701652 e+00
Outros usos (L5) para NO 7	-7.620348 e+02	8.714048 e+00
Bias para NO 8	-7.790665 e+01	2.095452 e+00
Altimetria para NO 8	7.905809 e+01	-2.225624 e-01
Declividade para NO 8	3.927376 e+02	-6.337592 e+02
Hidrografia para NO 8	1.260303 e+02	1.455877 e+02
Empr. Linear para NO 8	-7.333625 e+01	7.483766 e+01
Áreas de Proteção Integral para NO 8	4.278650 e+01	9.466717 e+02
Áreas Indígenas para NO 8	-5.505630 e+02	5.582516 e+01
Aeródromo para NO 8	-8.313132 e+02	-7.947132 e+02
Floresta (L0) para NO 8	-6.605326 e+01	3.869594 e+02
Formação Natural não Florestal (L1) para NO 8	3.536753 e+01	-2.377833 e+01
Agropecuária (L2) para NO 8	-2.070106 e+00	-8.300910 e+02
Área Antropizada (L3) para NO 8	-1.062383 e+01	-4.647251 e+00
Corpos D'Água (L4) para NO 8	-1.444884 e+02	3.874501 e+01
Outros usos (L5) para NO 8	3.893934 e+01	1.909953 e+02
Bias para NO 9	1.488435 e+00	1.540999 e+00
Altimetria para NO 9	-3.290832 e+00	3.581452 e+00
Declividade para NO 9	-4.088908 e+00	-1.673964 e+02
Hidrografia para NO 9	-2.694041 e+00	-5.428714 e+00
Empr. Linear para NO 9	-1.577104 e+02	6.354140 e+01
Áreas de Proteção Integral para NO 9	-9.037229 e+02	6.539827 e+01
Áreas Indígenas para NO 9	3.793913 e+01	-1.087052 e+02
Aeródromo para NO 9	-4.888737 e+01	1.746630 e+02
Floresta (L0) para NO 9	-1.444396 e+00	-7.116382 e+02
Formação Natural não Florestal (L1) para NO 9	8.660541 e+01	1.695389 e+02
Agropecuária (L2) para NO 9	1.956314 e+02	1.392132 e+02
Área Antropizada (L3) para NO 9	3.261524 e+00	-3.947123 e+00

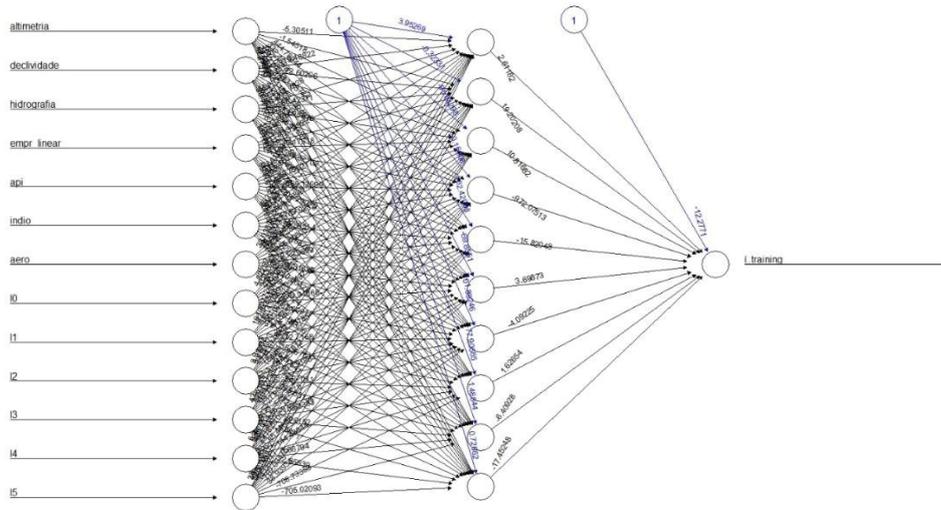
INFORMAÇÃO	VALOR 2 MIL	VALOR 20 MIL
Corpos D'Água (L4) para NO 9	6.679450 e-01	7.406535 e+00
Outros usos (L5) para NO 9	-7.087335 e+02	-3.118249 e-01
Bias para NO 10	7.266165 e-01	-1.506823 e+02
Altimetria para NO 10	-6.275189 e+00	6.747329 e+02
Declividade para NO 10	-2.077666 e+01	-1.939209 e+02
Hidrografia para NO 10	1.615824 e+00	2.005166 e+02
Empr. Linear para NO 10	9.900755 e+01	1.981272 e+02
Áreas de Proteção Integral para NO 10	-7.094743 e+02	-1.039106 e+03
Áreas Indígenas para NO 10	8.504294 e+02	-1.011323 e+03
Aeródromo para NO 10	2.243261 e-01	-9.030810 e+02
Floresta (L0) para NO 10	-1.816017 e+00	6.457469 e+01
Formação Natural não Florestal (L1) para NO 10	-8.011924 e+02	-5.727810 e+00
Agropecuária (L2) para NO 10	4.322561 e+00	-3.890529 e+00
Área Antropizada (L3) para NO 10	4.687100 e+00	-1.290706 e+03
Corpos D'Água (L4) para NO 10	2.656392 e+00	-4.084466 e+02
Outros usos (L5) para NO 10	-7.050209 e+02	-1.020588 e+03
Bias para Predição	-1.227710 e+01	1.488287 e+00
NO 1 para Predição	2.611825 e+00	1.280460 e+01
NO 2 para Predição	1.920208 e+01	3.119254 e+00
NO 3 para Predição	1.081682 e+01	1.351450 e+01
NO 4 para Predição	-9.720751 e+02	-2.029552 e+01
NO 5 para Predição	-1.582048 e+01	-9.596087 e-01
NO 6 para Predição	3.698726 e+00	1.747168 e+01
NO 7 para Predição	-4.092248 e+00	-9.507744 e+00
NO 8 para Predição	1.626539 e+00	1.114704 e+01
NO 9 para Predição	-6.409280 e+00	-1.196022 e+01
NO 10 para Predição	-1.745248 e+01	7.083084 e+00

FONTE: O autor (2020).

Nota-se nas tabelas 11 e 12 que a métrica de erro diminuiu levemente com a duplicação no número de neurônios, porém aumentou com a inserção de maiores quantidades de amostras de entrada. Isso significa que a utilização de mais de 5 camadas intrínsecas não representa ganhos significativos de qualidade do modelo.

A representação gráfica da estrutura da RNA ao utilizar 10 *layers* ocultos pode ser observada da mesma forma na Figura 25. Observe que com o acréscimo do número de *layers* ocultos aumenta significativamente o nível de relações internas da rede, o que aumenta exponencialmente o tempo de treinamento.

FIGURA 25 – RNA COM 10 NEURÔNIOS OCULTOS E 20 MIL AMOSTRAS



FONTE: O autor (2020).

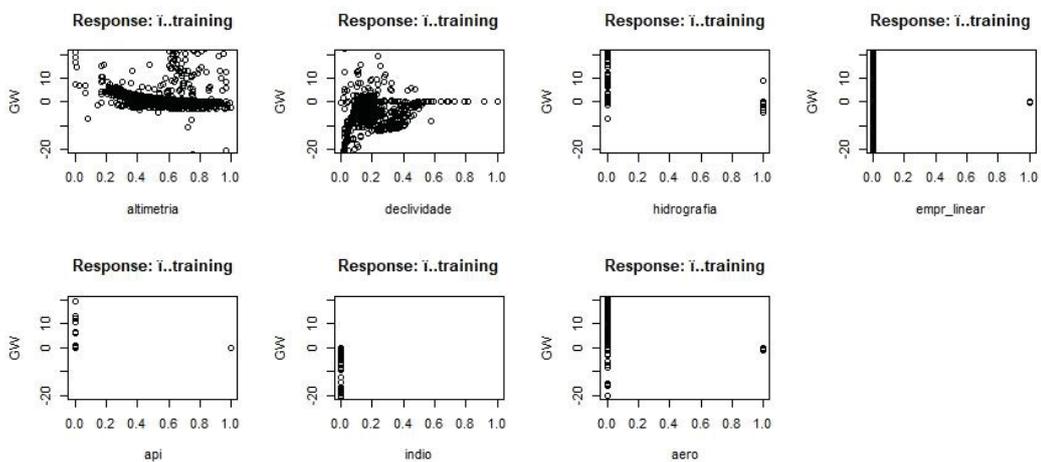
A seção seguinte abordará a análise a respeito das variáveis de entrada da rede.

4.2 ANÁLISE DAS COVARIÁVEIS DE ENTRADA

4.2.1 Treinamento utilizando 2 mil amostras

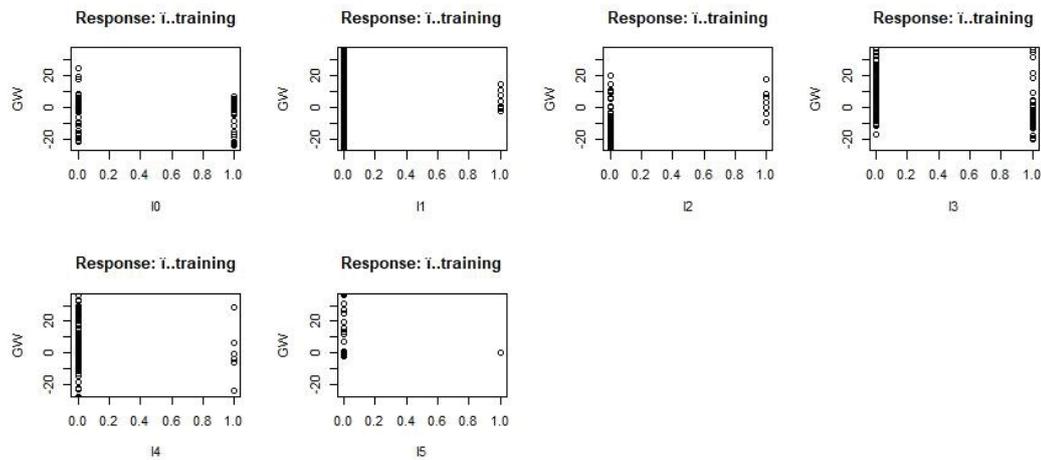
Outra análise passível de ser feita é o quanto as variáveis explicam o modelo matemático utilizado. A Figura 26 e a Figura 27 mostram essa análise para 2 mil amostras.

FIGURA 26 – CAMADAS DISCRETAS E CATEGÓRICAS (A)



FONTE: O autor (2020).

FIGURA 27 – CAMADAS DISCRETAS E CATEGÓRICAS (B)



FONTE: O autor (2020).

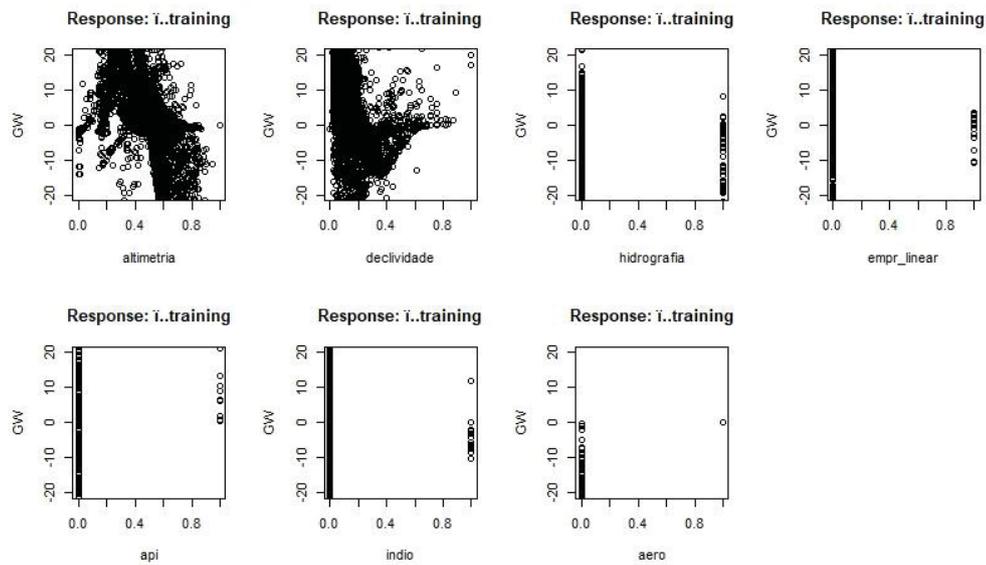
Observando o gráfico para as variáveis numéricas, um coeficiente explicativo adequado em relação à variável resposta (predição) aquelas que possuem uma distribuição menos dispersa dos pesos generalizados (GW) ao longo do valor de 0 e das variáveis binárias, o coeficiente ótimo explicativo é quando a dispersão está próxima de 0, na maior parte das amostras utilizadas (representando um gráfico na direita). Dessa forma, tem-se como variáveis mais explicativas a declividade e as camadas de uso do solo correspondentes às Florestas e Áreas Não Vegetadas.

Nota-se ainda que não há informações das áreas estratégicas ambientais nos gráficos das variáveis categóricas. Isso se deve ao fato de que, pela amostra ser de natureza aleatória, não havia informações desta covariável nos 2 mil exemplos estudados, e seu valor na RNA ficou como “*NaN – Not a Number*”. Isso significa que o valor das operações matemáticas envolvendo esta variável não é um valor definido ou não é representável e por isso deve ser removido da análise.

4.2.2 Treinamento utilizando 20 mil amostras

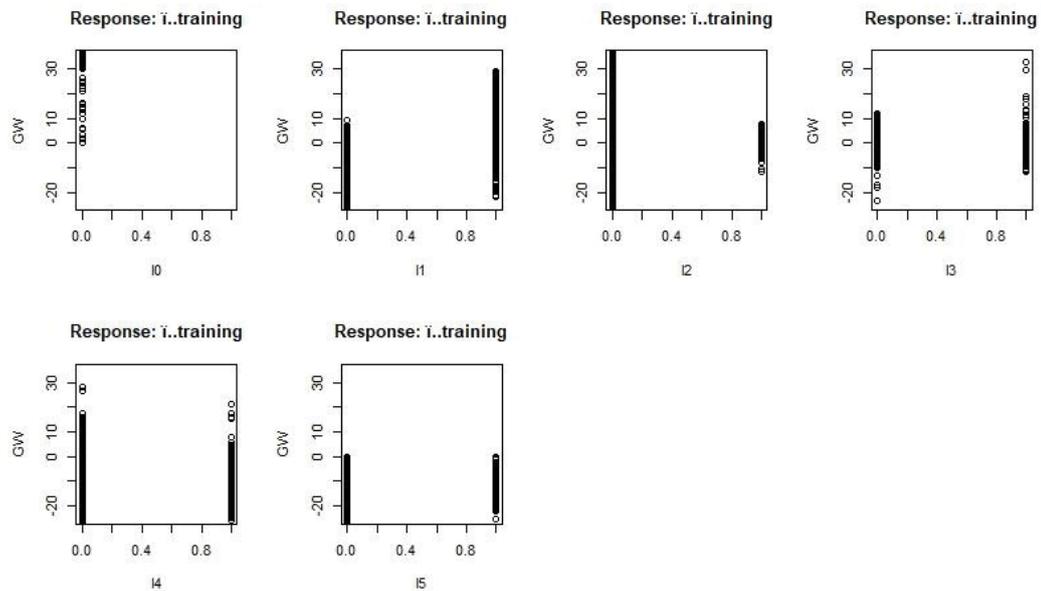
Semelhante ao caso anterior, fez-se a geração gráfica dos coeficientes explicativos da predição final em relação a cada covariável de entrada. As Figuras 28 e Figura 29 mostram essa análise para 20 mil amostras, e da mesma maneira que a anterior, utilizando apenas 5 neurônios ocultos.

FIGURA 28 – CAMADAS DISCRETAS E CATEGÓRICAS (A)



FONTE: O autor (2020).

FIGURA 29 – CAMADAS DISCRETAS E CATEGÓRICAS (B)



FONTE: O autor (2020).

Por coincidência, não houve amostras com informações de áreas com restrições ambientais ou embargadas e da mesma forma, foram excluídas as camadas correspondentes a essa covariável.

Para essas amostras, têm-se que as variáveis mais relevantes para a variável permanecem as mesmas: declividade e os usos do solo correspondentes às Florestas

e às Áreas Não Vegetadas, porém com participação maior das áreas de APP e do uso do solo correspondentes à Áreas Alagadas.

Estas análises são suportadas por duas outras análises, que também são saídas da Neuralnet: a análise do gráfico *pairwise* (Figura 30) e o gráfico de sensibilidade das variáveis (Figura 31).

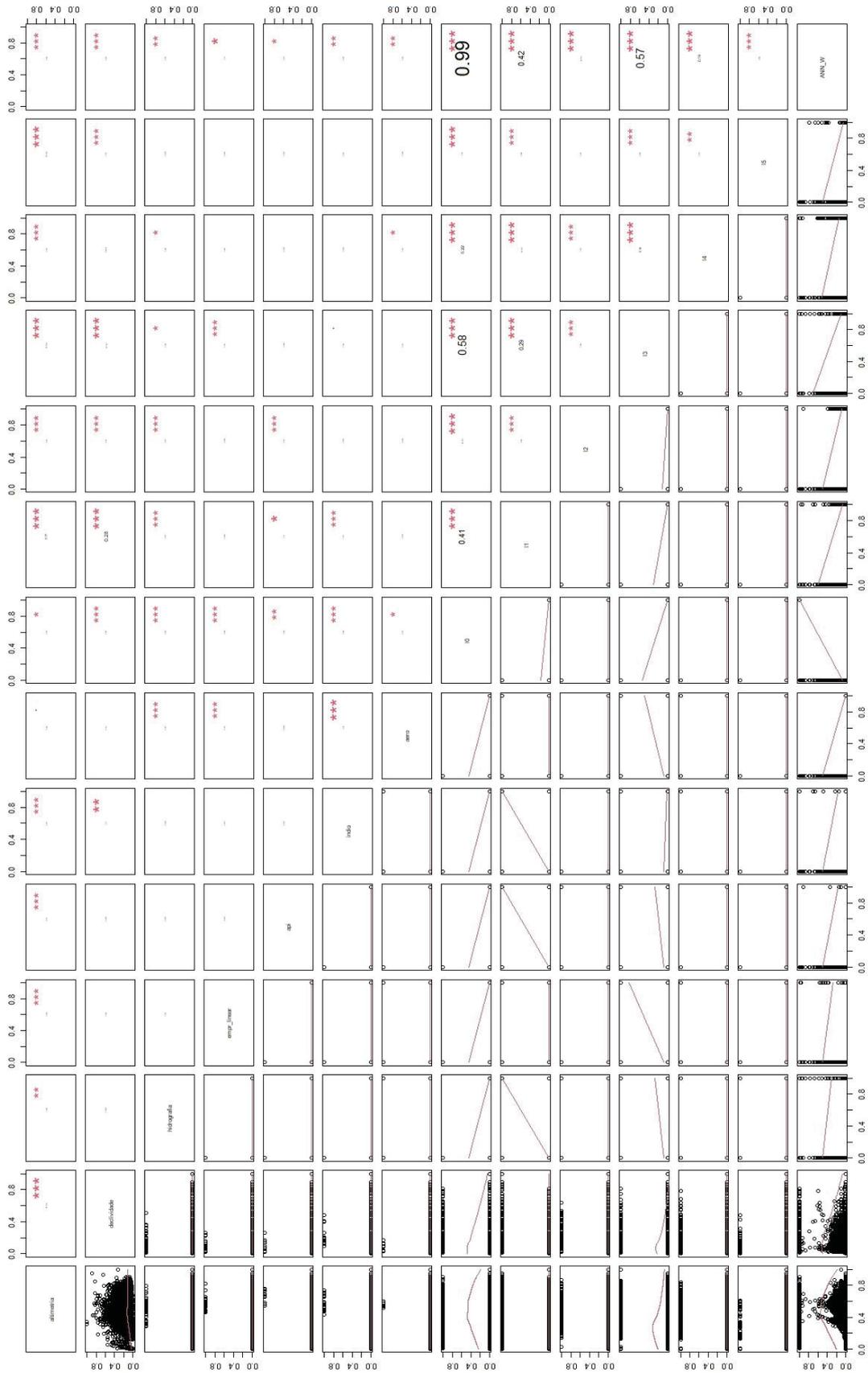
O gráfico *pairwise* possui a sumarização de todo o universo estatístico da rede, como a as correlações entre todas as variáveis de entrada e os pesos generalizados do modelo matemático. Ao analisá-lo, nota-se que para a área de estudo e as amostras utilizadas, as variáveis L0, L1 e L3 são as mais relevantes para a predição. As correlações entre as covariáveis não se mostraram significativas (todas abaixo de 80%), ou seja, nenhuma é totalmente explicada estatisticamente por outra.

No gráfico de sensibilidade, tem-se as variáveis: aeródromos, altimetria, áreas de proteção integral, declividade, empreendimentos lineares, APP, áreas indígenas, L0, L1, L2, L3, L4 e L5 representando o uso do solo, nesta ordem. No eixo das abcissas o índice explicativo de 0 a 100%, no eixo das ordenadas o comportamento das respostas e as cores das linhas mostram a porcentagem da variável aplicada na amostra de treinamento, sendo:

- Vermelha = 0% das demais variáveis aplicadas;
- Amarela = 20% das demais variáveis aplicadas;
- Verde = 40% das demais variáveis aplicadas;
- Ciano = 60% das demais variáveis aplicadas;
- Azul = 80% das demais variáveis aplicadas;
- Magenta = das demais variáveis aplicadas;

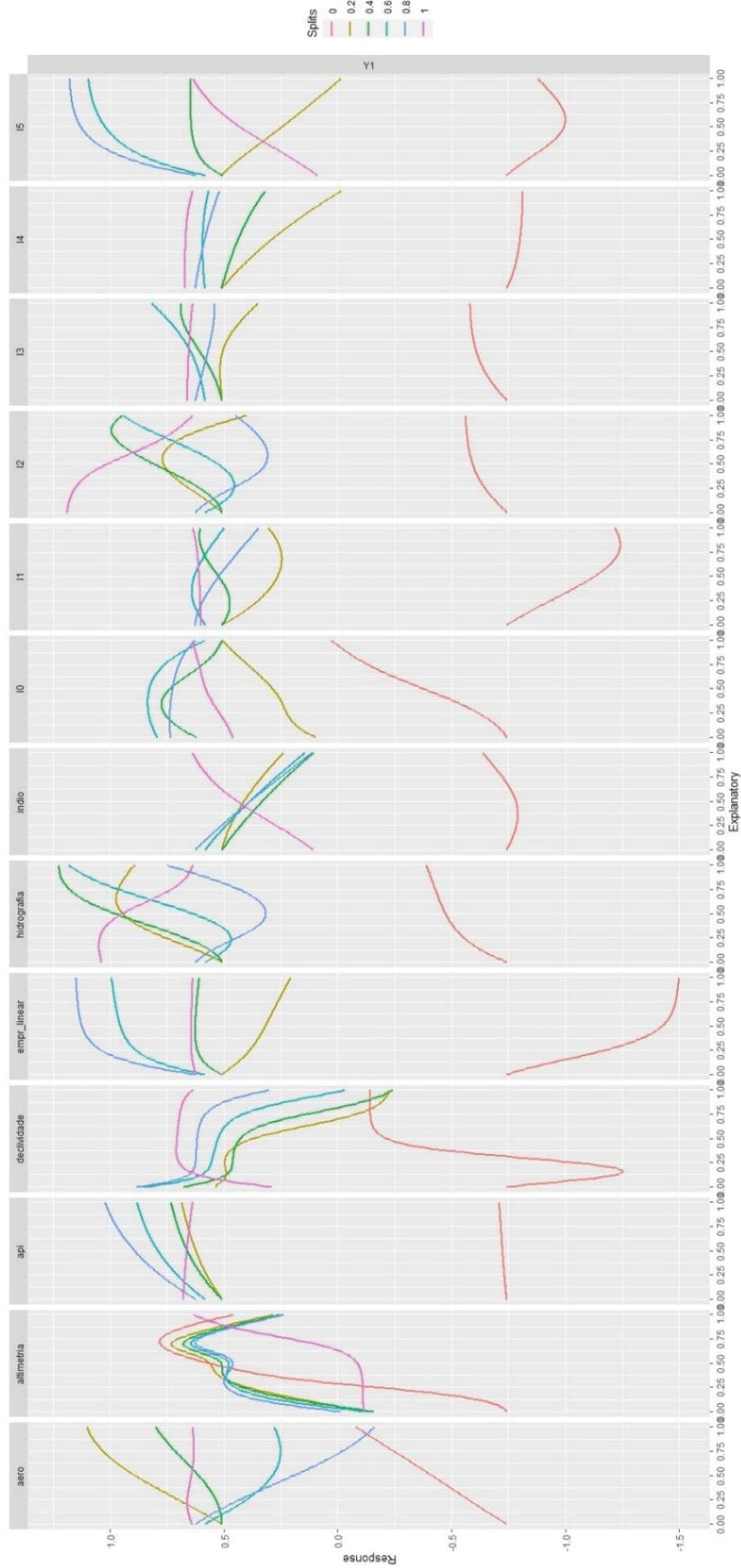
Este gráfico ajuda a entender o comportamento individual das variáveis e se há a possibilidade de analisar tais variáveis de modo isolado das outras covariáveis. Analisando a Figura 31, observa-se que nenhuma variável sozinha pode explicar o fenômeno estudado ou predizer a ocorrência ou não de linhas de transmissão individualmente.

FIGURA 30 – GRÁFICO PAIRWISE



FONTE: O autor (2020).

FIGURA 31 – GRÁFICO DE SENSIBILIDADE



FONTE: O autor (2020).

Isoladamente, a variável correspondentes aos aeródromos, a altimetria, a declividade e o uso do solo correspondentes a florestas se sobressaem em relação às demais, porém não conseguem explicar o fenômeno por completo. Quando são combinadas com as demais covariáveis, elas aumentam seu poder explicativo em relação à predição. Apenas a variável altimetria não sofreu alterações significativas com o acréscimo das demais, mostrando que ela é mais forte que as demais, ou seja, é menos influenciada pela outras. Porém, isso não significa que sua contribuição ou peso é mais importante que as restantes variáveis na predição.

4.3 TAXA DE SUCESSO

4.3.1 Treinamento utilizando 2 mil amostras

O gráfico completo que representa a curva ROC só foi gerado para o número de 20 mil amostras, uma vez que o comportamento dele em relação ao aumento da amostra de entrada se manteve. A taxa de sucesso para essa quantidade de amostras foi de 61%, o que é um índice de assertividade considerado baixo.

Uma informação complementar à curva ROC e ao valor AUC é a matriz de Contingência ou Matriz de Confusão a respeito das amostras utilizadas. Como foi trabalhado com a proporção de 60% de amostras de teste, ou seja, 1.200 (mil e duzentos) exemplos foram sumarizados na Tabela 13.

TABELA 13 – MATRIZ DE CONFUSÃO PARA 2 MIL AMOSTRAS

	True	False
Amostras Positivas (1)	322	270
Amostras Negativas (0)	186	419

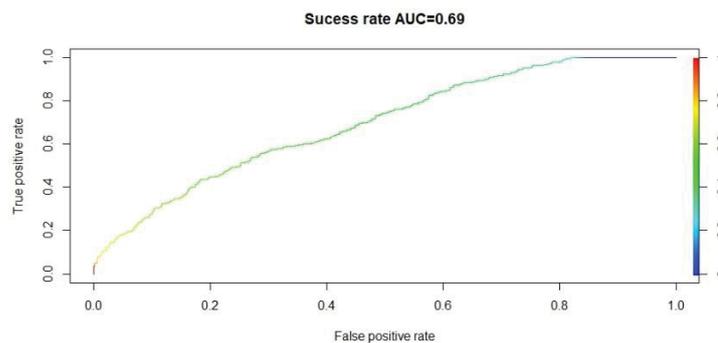
FONTE: O autor (2020).

Os resultados mostraram que para esses dados de entrada, houve muitas amostras positivas que foram consideradas “falsas” pelo modelo treinado. Isso significa que há pouca distinção entre as amostras positivas e negativas em relação às informações presentes nos *layers* das camadas.

4.3.2 Treinamento utilizando 20 mil amostras

Igualmente, fez-se as análises para a quantidade de 20 mil amostras. O comportamento do gráfico que apresenta a curva ROC, assim como a o valor AUC associado a ele podem ser observados na Figura 32.

FIGURA 32 – TAXA DE SUCESSO COM 20 MIL AMOSTRAS



FONTE: O autor (2020).

Houve uma pequena melhora de 8% na taxa de assertividade com o aumento da quantidade de amostras, porém fica evidente na curva ROC que o modelo treinado não é suficiente para explicar a predição usando os dados de entrada. Isso porque, quando o modelo é altamente explicativo e treinado com amostras de boa qualidade, há um efeito de convergência de verdadeiros positivos próximos a 1 e de falsos positivos próximo a 0, o que não se observa com esses dados.

A matriz de Contingência ou Matriz de Confusão a respeito das amostras utilizadas foi gerada mantendo a proporção de 60% de amostras de teste, ou seja, 12.000 (doze mil) exemplos. Estas informações estão sumarizadas na Tabela 14.

TABELA 14 – MATRIZ DE CONFUSÃO PARA 20 MIL AMOSTRAS

	True	False
Amostras Positivas (1)	3.866	2.176
Amostras Negativas (0)	1.573	4.385

FONTE: O autor (2020).

Nota-se que a proporção entre as amostras positivas classificadas como “falsas” pelo modelo treinado (erro) aumentou com o acréscimo no número de amostras de entrada, apesar do aumento da taxa de sucesso.

5 CONCLUSÕES E RECOMENDAÇÕES

Neste trabalho foi proposto o emprego dos conceitos da IA através das redes neurais artificiais no problema de traçado de diretrizes de linhas de transmissão. Esta abordagem propicia a automação do processo de atribuição de pesos subjetivos para as camadas restritivas.

Apesar de não ser possível o estabelecimento de um valor unitário para os pesos das variáveis de entrada em relação à variável preditiva, resposta procurada pela rede neural, a metodologia apresenta vantagens em relação a metodologia tradicional, justamente por proporcionar a modelagem das relações internas entre as próprias covariáveis de treinamento, de forma mais robusta que apenas em relação à cálculos de correlação.

Os resultados obtidos mostraram que a utilização de 5 neurônios ocultos é o suficiente para esse tipo RNA treinada, não alterando significativamente os índices das métricas de erros de estimativa e nem refletido na taxa de sucesso da predição. Esses números de *layers* utilizados nesta pesquisa foram obtidos ou usados de maneira empírica ou por tentativa e erro.

Com relação à contribuição das camadas, o comportamento explicativo individual reforçou os resultados encontrados no trabalho de Campos (2014), onde o autor utilizou apenas a declividade, o uso do solo e a distância às rodovias, LT existentes e aos principais rios e cursos d'água existentes. Isso mostra que as LTs existentes atualmente, foram construídas tendo esses fatores com maior peso decisivo. Ou seja, fatores sociais, culturais, históricos e até ambientais foram preteridos em relação ao fator econômico.

De maneira geral, o algoritmo utilizado no estudo e utilizado na pesquisa se mostrou muito eficaz para a estimativa preditiva de traçados de novas linhas de transmissão de energia ou de qualquer outro fenômeno linear ou pontual. Porém, a linguagem R base do algoritmo não possui muitas bibliotecas e funções estatísticas (fundamentais para análises de RNAs) que gerenciem a alocação de memória RAM de um computador de maneira efetiva. Isso dificulta a utilização de imagens *rasters* de grandes resoluções e, conseqüentemente, limita a análise à grandes extensões de maneira generalizada ou à pequenas extensões para utilização de informações com maior detalhamento.

Devido à impossibilidade de gerenciamento e alocação de memória de máquina não foi gerada a superfície de probabilidade final, base para a geração do traçado final na aplicação do algoritmo *Least Coast Path* e, conseqüentemente, a diretriz de traçado desta nova obra a partir dessa superfície probabilística. Essa superfície é uma imagem *geotif* que contém as mesmas dimensões que os *rasters* correspondentes de todas as camadas de entrada da rede, resultado da aplicação do modelo matemático treinado em uma nova área ou novo traçado na área de estudo. A partir desta imagem, seria possível a aplicação da operação do “Caminho de Menor Custo” em qualquer software GIS nas últimas versões.

Analisando os resultados estatísticos do estudo feito, tem-se que uma taxa de aprendizado que pode ser considerada boa em muitas outras aplicações atuais, possuindo assertividade superior a 90% em relação às análises realizadas. O fato de a taxa de aprendizado ser baixa neste caso se atribui a baixa qualidade da amostra utilizada. Isso porque no momento da concepção dos estudos de traçados de LTs já existentes, muitas camadas aqui testadas não foram levadas em consideração há décadas, ou seja, os dados não são ótimos para as circunstâncias atuais. Então, para melhorar o sucesso de uma RNA de predição de traçados, faz-se necessário utilizar amostras não-tendenciosas para o treinamento do modelo matemático da rede neural.

Além disso, não foram utilizados no estudo alguns fatores de muita relevância para estudos de construção de redes de transmissão, que descreveriam de maneira mais completa o fator econômico, tais como: custo de construção por uma unidade de medida, custos de construções médios por combinação de fatores como tipo de solo, declividade, uso do solo, velocidade do vento; não foram utilizados no estudo. Essa informação, que é de propriedade das empresas privadas e de órgãos e concessionárias que gerenciam a geração e distribuição da energia elétrica no Brasil, não foi passível de obtenção.

Ademais, a utilização de uma amostragem em área para esse tipo de situação sem uma métrica de pertencimento ou de proximidade em relação a um percurso ou linha de transmissão existente, não se mostrou eficaz para melhorar a taxa de sucesso. Esse fenômeno pode ser observado mais facilmente na análise da matriz de contingência dos experimentos, que mostra a indefinição da predição mais evidente nas amostras positivas que a RNA classificou como *false* (ou negativa predita pelo modelo). Ou seja, as condições que se tem para explicar uma região no espaço geográfico que possui a ocorrência de uma linha de transmissão de energia são muito

semelhantes às regiões em que não há a presença do fenômeno. Se a RNA propagasse um histórico linear ou métrico de outras amostras próximas, ou utilizasse abordagens lineares, os resultados talvez poderiam ser otimizados.

Nenhuma covariável foi removida do estudo que continha informação de entrada na rede, pois o intuito é que o algoritmo seja um sistema especialista no estudo de traçado de novas linhas de transmissão de energia. Então, fazendo a remoção de algumas variáveis menos explicativas poderia gerar um pequeno ganho na assertividade, porém poderia deixar o algoritmo menos adaptativo para amostras em locais com características diferentes de que a área de teste escolhida.

Apesar da baixa taxa de aprendizagem apresentada pelos testes realizados ao longo da pesquisa, a metodologia proposta tem uma série de vantagens a serem aproveitadas, sobretudo na economia de recursos, a saber:

- De Tempo, pois diante de condições favoráveis, demandará poucas horas pra treinar, podendo ser aplicada diretamente;
- Humanos, ao reduzir a dependência de profissionais altamente treinados e que ainda podem apresentar um viés, advindo da natureza humana;
- Financeiros, decorrentes da economia dos demais recursos. Vale lembrar que muitas vezes, tais recursos são públicos.

O melhoramento de um sistema com esse abordado proporcionará a criação de um sistema especialista para análises desta natureza, que proporcionará a economia financeira e de tempo, além de contribuir para a criação de obras mais sustentáveis e responsáveis no sentido social e de preservação da diversidade cultural no país.

5.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Para fins de melhoramento dos resultados, sugere-se:

1. Utilizar um conjunto de dados simulados e ótimos como dados para a entrada na rede neural, a fim de minimizar a utilização de dados tendenciosos na fase de treinamento da rede. Isso é de fundamental importância para o melhoramento das métricas de aprendizagem e índice

de assertividade, reduzindo assim os erros de estimativas dentro da matriz de confusão;

2. Testar outros algoritmos de RNA para explicação das camadas;
3. Fazer uso de métodos de amostragem matemáticos mais robustos, a fim de minimizar o efeito randômico das amostras;
4. Fazer uso de dados mais acurados ou com maior resolução como parâmetro de entrada. Isso porque a utilização de dados muito generalizados e em escala pequena como os disponibilizados por órgãos responsáveis trazem consigo erros e interpolações que podem afetar diretamente nas estimativas da rede neural;
5. Fazer a manipulação do algoritmo a fim de otimizar e aumentar o gerenciamento de memória RAM durante a fase de treinamento, e geração da superfície preditiva, a fim de permitir a utilização de uma maior quantidade de amostras, além de permitir a utilização de imagens *rasters* maiores e mais detalhadas;
6. Traduzir o algoritmo para outra linguagem de programação que permita um melhor gerenciamento do hardware de treinamento ou realizar a tradução para treinamento em ambientes em nuvem, onde os servidores possuem configurações mais robustas e melhores que o computador utilizado para o desenvolvimento da presente pesquisa;
7. Desenvolver uma métrica para o melhoramento da explicação de variáveis fundamentais para a análise, como distância à empreendimentos lineares existentes, áreas de preservação e principalmente financeira;
8. Testar a abordagem de amostragem em linhas.

REFERÊNCIAS

- ABNT. **Projeto de linhas aéreas de transmissão de energia elétrica - NBR 5.422**. Rio de Janeiro: ABNT, 1985.
- ALTHUWAYNEE, O. **How to use ANN for prediction mapping using GIS data**. Disponível em: <<https://www.udemy.com/course/how-to-use-ann-for-prediction-mapping-using-gis-data/learn/lecture/14031034#overview>>. Acesso em: julho de 2020.
- ANASTASIADIS A., MAGOULAS G., VRAHATIS M. **New globally convergent training scheme based on the resilient propagation algorithm**. *Neurocomputing*, 64:253–270, 2005.
- BAGLI, S., et al. **Routeing of power lines through least-cost path analysis and multicriteria evaluation to minimise environmental impacts**. *Environmental Impact Assessment Review*, [s.l.], v. 31, n. 3, p. 234-239. Elsevier BV. Disponível em: <<http://dx.doi.org/10.1016/j.eiar.2010.10.003>>. Acesso em: out. 2018.
- BISHOP, C. M. **Neural networks and their application**. *Rev. Sci Instrum* 65(6):1803–1830. 1994.
- BONAMENTE, M. **Statistics and analysis of scientific data**. Editora Springer, 2ª edição. 71p. 2017.
- BRADLEY, A. P. **The use of the area under the ROC curve in the evaluation of machine learning algorithms**. *Pattern Recognition*, Vol. 30, No. 7, pp. 1145-1159, 1997.
- CAMPOS, P. B. R. **Modelagem de corredores ótimos para linhas aéreas de transporte de energia elétrica**. Dissertação. Programa de Pós-graduação em Análise e Modelagem de Sistemas Ambientais. Universidade Federal de Minas Gerais. 50p. Belo Horizonte – MG, 2014.
- DIJKSTRA, E. W. **A note on two problems in connexion with graphs**. *Numerische Mathematik* 1(1), 269-271, 1959.
- EPRI. **Overhead electric transmission line siting methodology**. EPRI, Palo Alto, CA, and Georgia Transmission Corporation, Tucker, GA, 2006.
- FREIMAN, F. P. **Modelagem espacial multicriterial para a definição automática de diretrizes de linha de transmissão de energia elétrica**. Dissertação. Programa de Pós-graduação em Ciências Geodésicas. Universidade Federal do Paraná. 75p. Curitiba, 2019.
- GRASSI, S. et al. **A GIS-based process for calculating visibility impact from buildings during transmission line routing**. *Connecting a digital Europe through location and place*, [s.l.], p.383-402, 2014. Springer International Publishing. Disponível em: <http://dx.doi.org/10.1007/978-3-319-03611-3_22>. Acesso em: nov. 2018.

GÜNTHER, F; FRITSCH S. **Neuralnet: training of neural networks**. The R Journal. ISSN 2073-4859. Vol. 2/1, June 2010.

HAYKIN, S. **Redes neurais: princípios e prática**. Editora Bookman, 2ª edição. 908p. 2001.

LAUREANO, A. **Os Desafios do Estudo de Traçado para a Construção de Linhas de Transmissão**. Disponível em: <<https://www.visaogeo.com.br/os-desafios-do-estudo-de-tracado-para-a-construcao-de-linhas-de-transmissao>>. Acesso em: junho de 2020.

LABEGALINI, P. R.; LABEGALINI, J. A.; FUCHS, R. D., ALMEIDA, M. T. **Projetos Mecânicos das Linhas Aéreas de Transmissão**. 2ª. Edgard Blücher, 1992.

LIMA, R. M. **Otimização de Traçado de Linhas de Transmissão por Meio de Avaliação Multicritério e Sistema de Informações Geográficas (SIG)**. Dissertação. Programa de Pós-graduação em Engenharia de Energia. Universidade Federal de Itajubá. 73p. Itajubá - MG, 2015.

MENEZES, V. P. **Linhas de Transmissão de Energia Elétrica: Aspectos Técnicos, Orçamentários e Construtivos**. Departamento de Engenharia Elétrica. Universidade Federal do Rio de Janeiro - Escola Politécnica. 87p. Rio de Janeiro, 2015.

MINISTÉRIO DO MEIO AMBIENTE (BR). **Portaria nº 419, de 26 de outubro de 2011. Regulamenta a atuação da Fundação Nacional do Índio - FUNAI, da Fundação Cultural Palmares - FCP**. Instituto do Patrimônio Histórico e Artístico Nacional - IPHAN e do Ministério da Saúde.

MONTEIRO, C. et al. **GIS Spatial Analysis Applied to Electric Line Routing Optimization**. IEEE Transactions on Power Delivery, [s.l.], v. 20, n. 2, p.934-942, abr. 2005. Institute of Electrical and Electronics Engineers (IEEE). Disponível em: <<http://dx.doi.org/10.1109/tpwrd.2004.839724>>. Acesso em: junho de 2020.

OLIVEIRA, A. R. **Redes Neurais Artificiais Aplicadas na Detecção, Classificação e Localização de Defeitos em Linhas de Transmissão**. Dissertação. Universidade Federal de Juiz de Fora. 132p. Juiz de Fora - MG, 2005.

ONS. **O Sistema Interligado Nacional**. Disponível em: <<http://www.ons.org.br/>>. Acesso em: junho de 2020.

SAATY, T.L. **The Analytic Hierarchy Process**. McGraw-Hill, New York, NY. 1980.

SCHMIDT, A, J. **Implementing a GIS Methodology for Siting High Voltage Electric Transmission Lines**. Volume 11, Papers in Resource Analysis. 17 pp. Saint Mary's University of Minnesota University Central Services Press. Winona, MN. 2009.

SHU, J. et al. **A New Method for Spatial Power Network Planning in Complicated Environments**. IEEE Transactions on Power Systems, [s.l.], v. 27, n. 1, p.381-389,

fev. 2012. Institute of Electrical and Electronics Engineers (IEEE). Disponível em: <<http://dx.doi.org/10.1109/tpwrs.2011.2161351>>. Acesso em: out. 2018.

SKAPURA, D. M. **Building neural networks**. New York: ACM Press, 1996.

SOARES-FILHO, B.S; RODRIGUES, H.O; FALIERI, A; COSTA, W. L., **Dinâmica EGO**, 2008. Tutorial disponível em: <<http://www.csr.ufmg.br/dinamica>>.

PECHANEC, V.; BURIAN, J.; DOBEŠOVÁ, Z. **Integration Neural Networking and GIS in Modeling Landscape Changes**. International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM, v. 1, n. 2, p. 651-658, 2014.

PRATI, R.C.; BATISTA, G. H.; MONARD, M. C. **Curvas ROC para avaliação de classificadores**. 2008. Disponível em: <https://sites.icmc.usp.br/gbatista/files/ieee_la2008.pdf>. Acesso em: Jul. 2020.

TOMLIN, D. **Geographic information systems and Cartographic Modeling**. Prentice Hall, New York, 1990.

WASSERMAN, L. **All of Statistics: a concise course in statistical inference**. Editora Springer, 1ª edição. 446p. 2004.