

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL
MESTRADO EM ENGENHARIA CIVIL

GLEDSON FÁBIO COTRIM ROCHA

DISSERTAÇÃO DE MESTRADO

**ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DE TRILHOS FERROVIÁRIOS
POR MEIO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA**

VITÓRIA
2019

GLEDSO N FÁBIO COTRIM ROCHA

DISSERTAÇÃO DE MESTRADO

**ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DE TRILHOS FERROVIÁRIOS
POR MEIO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Civil do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Engenharia Civil, na área de concentração Transportes.

Orientador: Prof. Dr. Rodrigo de Alvarenga Rosa.

Coorientador: Prof. Dr. Ing. Renato Antonio Krohling.

VITÓRIA
2019

Ficha catalográfica disponibilizada pelo Sistema Integrado de
Bibliotecas - SIBI/UFES e elaborada pelo autor

- R672e Rocha, Gledson Fábio Cotrim, 1977-
 Estimação da vida útil remanescente de trilhos ferroviários
 por meio de técnicas de aprendizado de máquina / Gledson
 Fábio Cotrim Rocha. - 2019.
 115 f. : il.
- Orientador: Rodrigo de Alvarenga Rosa.
 Coorientador: Renato Antonio Krohling.
 Dissertação (Mestrado em Engenharia Civil) - Universidade
 Federal do Espírito Santo, Centro Tecnológico.
1. Trilho Ferroviário. 2. Rede Neural Artificial. 3. k
 Vizinhos mais próximos. 4. Superestrutura Ferroviária. I. Rosa,
 Rodrigo de Alvarenga. II. Krohling, Renato Antonio. III.
 Universidade Federal do Espírito Santo. Centro Tecnológico. IV.
 Título.

CDU: 624

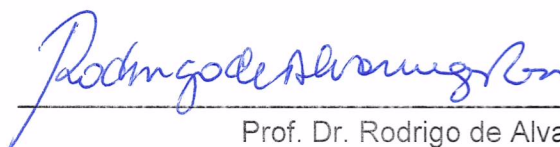
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DE TRILHOS FERROVIÁRIOS POR MEIO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA

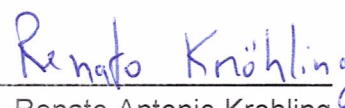
Gledson Fábio Cotrim Rocha

Dissertação apresentada ao Curso de Mestrado em Engenharia Civil do Programa de Pós-Graduação em Engenharia Civil da Universidade Federal do Espírito, como requisito parcial para obtenção do título de Mestre em Engenharia Civil, área de Transportes.

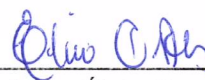
Aprovada no dia **27 de agosto de 2019** por:



Prof. Dr. Rodrigo de Alvarenga Rosa
Doutor em Engenharia Elétrica
Orientador – UFES



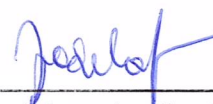
Prof. Dr. Renato Antonio Krohling
Doutor em Engenharia Elétrica
Coordenador – UFES



Prof. Dr. Elcio Cassimiro Alves
Doutor em Engenharia e Tecnologia Espaciais
Examinador Interno – UFES



Prof. Dr. Daniel Cruz Cavaliere
Doutor em Engenharia Elétrica
Examinador Externo - IFES



Prof. Dr. Jodelson Aguiar Sabino
Doutor em Engenharia de Produção
Examinador Externo – Vale S/A

Vitória – ES, agosto de 2019

AGRADECIMENTOS

A Deus por sua infinita sabedoria, força e poder que me impulsionou nesta empreitada de minha vida.

Agradeço a minha esposa Kyla Rocha, pelo apoio, incentivo, companheirismo, compreensão e principalmente por acreditar nos momentos mais difíceis, sendo que sem seu apoio não seria possível a realização deste trabalho. Agradeço ao meu filho João por entender e compreender a ausência de seu pai durante estes momentos.

Agradeço aos meus pais João Rocha e Maria Rocha que sempre me apoiaram e me incentivaram para a realização deste Mestrado e durante a realização deste trabalho.

Agradeço ao professor Rodrigo Rosa por aceitar o desafio na condução deste estudo, por transmitir seus conhecimentos, sugestões, pelo apoio e pelos períodos de orientações, durante a elaboração deste trabalho.

Agradeço ao professor Renato Krohling por todas as sugestões apresentadas na qualificação e elaboração deste trabalho e a banca pela disponibilidade, pelas sugestões apresentadas na qualificação e pelo tempo dedicado a leitura deste trabalho.

Agradeço a Vale S.A. pelo apoio e liberação para cursar o Mestrado, em especial aos meus gestores Edmundo Ferraz e Rogério Ribeiro pelo incentivo.

Aos meus colegas de trabalho da Vale S.A., pelo incentivo na busca do conhecimento, apoio e compreensão.

Aos professores do Programa de Pós-Graduação em Engenharia Civil da UFES, por partilharem seus conhecimentos e experiências.

Aos meus amigos e demais familiares pelo apoio e incentivo constante durante a realização do Mestrado e deste trabalho.

De maneira geral, agradeço a todos que de alguma maneira contribuíram e torceram para essa realização.

Bem-aventurado o homem que acha
sabedoria, e o homem que adquire
conhecimento.

Provérbios 3:13

RESUMO

Sendo o trilho o elemento de maior custo na via permanente, buscar ferramentas para estimar a vida útil remanescente dos trilhos é importante para utilizá-los ao máximo. Assim, tem-se como hipótese que algoritmos de aprendizado de máquina treinados a partir de bases de dados históricos, possam auxiliar ao planejador na realização dessa estimativa. Foram testados dois algoritmos de aprendizado de máquina: Rede Neural Artificial (RNA) e K-Vizinhos mais próximos (k-NN). As bases de dados históricos utilizados foram da Estrada de Ferro Vitória a Minas (EFVM). O conjunto de dados possui 1.275.034 registros referente ao período de 6 anos e 8 meses. As RNA tiveram 5 neurônios na camada de entrada: 1) o grau da curva; 2) trilho interno ou externo a curva; 3) largura do boleto; 4) altura do boleto; e 5) peso médio transportado. A camada intermediária, independentemente da categoria para a estimativa da vida útil remanescente dos trilhos, teve variações de 1, 2 e 3 camadas e variações na quantidade de neurônios de 30, 50, 100, 200 e 400. A camada de saída depende da categoria de período da vida útil remanescente dos trilhos: 1) mês, com 80 neurônios, 2) trimestre, com 27 neurônios, 3) semestre, com 14 neurônios e 4) ano, com 7 neurônios. Para o algoritmo k-NN, foram testadas configurações variando $k=5, 7$ e 9 . Para ambos algoritmos, aplicou-se a validação cruzada *k-fold*, com $f=10$, e o desempenho foi avaliado por meio do valor da acurácia e do *F1-score*. A linguagem de programação foi *Python* e a biblioteca *Scikit-Learn*. As configurações das RNA e as configurações do k-NN foram comparadas e o k-NN apresentou resultados superiores as RNA. Entretanto, ambos algoritmos atingiram o objetivo proposto nesta dissertação que era a estimativa da vida útil remanescente dos trilhos de maneira a auxiliar o planejador da manutenção da via férrea quanto à substituição dos trilhos, onde obtiveram resultados na acurácia e *F1-score* superiores a 80% em ambos os algoritmos para as categorias de período semestre e ano, sendo estas categorias de períodos mais utilizadas pelos operadores ferroviários. O algoritmo k-NN obteve sempre melhores resultados quando comparados com a RNA.

Palavras-chave: Trilho Ferroviário. Rede Neural Artificial. k-Vizinhos mais próximos. Superestrutura Ferroviária.

ABSTRACT

As the rail is the costliest element on the permanent track, finding tools to estimate the remaining life is important to make the most of them. Thus, it has hypothesized that machine learning algorithms trained from historical databases can assist the planner in performing this estimation. Two machine learning algorithms were tested: Artificial Neural Network (RNA) and Nearest K-Neighbors (k-NN). The historical databases used were from Vitória to Minas Railway (EFVM). The dataset has 1,275,034 records for the period of 6 years and 8 months. The RNA had 5 neurons in the input layer: 1) the degree of the curve; 2) internal or external rail curve; 3) the width of the rail; 4) the height of rail; and 5) average weight carried. The intermediate layer, regardless of the category for estimating the remaining rail life, had variations of 1, 2 and 3 layers and variations in the number of neurons of 30, 50, 100, 200 and 400. The output layer depends on the period category the remaining useful life of the rails: 1) month, with 80 neurons, 2) quarter, with 27 neurons, 3) semester, with 14 neurons and 4) year, with 7 neurons. For the k-NN algorithm, configurations ranging from $k = 5, 7$ and 9 were tested. For both algorithms, k-fold cross validation was applied, with $f = 10$, and performance was evaluated using the accuracy value and F1-score. The programming language was Python and the Scikit-Learn library. RNA configurations and k-NN configurations were compared and k-NN showed superior results to RNA. However, both algorithms reached the objective proposed in this dissertation, which was the estimation of the remaining rail life in order to help the railroad maintenance planner to replace the rails, where they obtained results in accuracy and F1-score higher than 80% for both algorithms for the semester and year period categories, these period categories being most used by rail operators. The k-NN algorithm always obtained better results when compared to RNA.

Keywords: Rail Track. Artificial neural network. k-Nearest neighbors. Railway superstructure.

LISTA DE FIGURAS

Figura 1 – Pseudocódigo da RNA em Backpropagation	23
Figura 2 – Divisão dos dados para treinamento, <i>validação e teste</i>	25
Figura 3 – Divisão dos dados por k-fold cross validation	26
Figura 4 – Obtenção dos valores da precision e recall.....	28
Figura 5 – Pseudocódigo do k-NN	31
Figura 6 - Partes do trilho ferroviário tipo Vignole	35
Figura 7 – Seção da superestrutura ferroviária	37
Figura 8 – Distribuição das categorias de estimação da vida útil remanescente para um horizonte de 6 anos e 8 meses.....	39
Figura 9 – Mapa geral do traçado da EFVM.....	41
Figura 10 – Vagão de minério GDE	41
Figura 11 – Perfil do trilho TR-68	42
Figura 12 – Carro Controle EM-80 - EFVM	43
Figura 13 – Dimensões do Carro Controle EM-80 - EFVM	43
Figura 14 – Localização do sistema de leitura do Carro Controle EM-80 - EFVM	44
Figura 15 – Sistema de leitura laser do Carro Controle EM-80 - EFVM.....	45
Figura 16 – Leitura da geometria do trilho na tela do computador de bordo do Carro Controle.....	46
Figura 17 – Leitura da geometria do trilho pelo Carro Controle EM-80	47
Figura 18 – Trecho da coletada de dados na EFVM	48
Figura 19 – Distribuição dos registros para o período de 2011 a 2017 trilho esquerdo – linha 1 – EFVM	50
Figura 20 – Distribuição dos registros referentes aos trechos em tangentes e curvas para o trilho esquerdo – linha 1 - EFVM	51
Figura 21 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2011 - linha 1 - EFVM	52
Figura 22 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2012 - linha 1 - EFVM.....	53
Figura 23 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2013 - linha 1 - EFVM.....	54

Figura 24 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2014 - linha 1 - EFVM	55
Figura 25 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2015 - linha 1 - EFVM	56
Figura 26 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2016 - linha 1 - EFVM	57
Figura 27 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2017 - linha 1 - EFVM	58
Figura 28 – Distribuição das categorias de estimação da vida útil remanescente para um horizonte de 6 anos e 8 meses	71
Figura 29 – Distribuição dos registros no conjunto de dados – categoria mês	73
Figura 30 – Distribuição dos registros no conjunto de dados – categoria trimestre ..	74
Figura 31 – Distribuição da categoria trimestre para os seus respectivos meses	75
Figura 32 – Distribuição dos registros no conjunto de dados – categoria semestre..	76
Figura 33 – Distribuição da categoria semestre para os seus respectivos meses	76
Figura 34 – Distribuição dos registros no conjunto de dados – categoria ano	77
Figura 35 – Distribuição da categoria ano para os seus respectivos meses	78
Figura 36 – RNA investigadas para estimação da vida útil remanescente de trilhos ferroviários	86
Figura 37 – Matriz de confusão da Configuração 39 - Semestre	96
Figura 38 – Matriz de confusão da Configuração 39 - Semestre	96
Figura 39 – Matriz de confusão da Configuração 54 - Ano	97
Figura 40 – Matriz de confusão da Configuração 54 - Ano	97
Figura 41 – Matriz de confusão da Configuração 61 - Semestre	103
Figura 42 – Matriz de confusão da Configuração 61 - Semestre	103
Figura 43 – Matriz de confusão da Configuração 64 – Ano	104
Figura 44 – Matriz de confusão da Configuração 64 – Ano	104
Figura 45 – Comparação das configurações com melhores desempenho da RNA e K-NN	106

LISTA DE TABELAS

Tabela 1 – Matriz de confusão	29
Tabela 2 – Raios e grau da curva - km 16+696 ao km 45+742 - EFVM.....	58
Tabela 3 – Peso da carga transportada sobre a linha 1 - EFVM.....	61
Tabela 4 – Quantidade de vagões que trafegaram sobre a linha 1 - EFVM.....	61
Tabela 5 – Peso total dos vagões que trafegaram na linha 1 - EFVM	62
Tabela 6 – Quantidade de locomotivas que trafegaram sobre a linha 1 – EFVM.....	62
Tabela 7 – Peso total das locomotivas que trafegaram sobre a linha 1 - EFVM	63
Tabela 8 – Peso do carregamento total trafegado sobre a linha 1 - EFVM.....	63
Tabela 9 – Conjunto de Dados.....	66
Tabela 10 – Conjunto de dados final da linha 1 - EFVM	67
Tabela 11 – Levantamento realizado para o trilho esquerdo da linha 1 – EFVM (35+737).....	69
Tabela 12 – Levantamento do peso médio trafegado da linha 1 – EFVM (35+737) .	70
Tabela 13 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – mês.....	82
Tabela 14 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – trimestre.....	82
Tabela 15 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – semestre.....	83
Tabela 16 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – ano	83
Tabela 17 – Configurações do k-NN para cada categoria de período.....	88
Tabela 18 – Resultados de treinamento/validação da categoria de período mês – grupo G1	90
Tabela 19 – Resultados de treinamento/validação da categoria de período trimestre – grupo G2	90
Tabela 20 – Resultados de treinamento/validação da categoria de período semestre – grupo G3	91
Tabela 21 – Resultados de treinamento/validação da categoria de período ano – grupo G4	91

Tabela 22 – Resultados das configurações da RNA na fase de treinamento/validação e teste	93
Tabela 23 – Resultados das configurações da RNA na fase de teste.....	95
Tabela 24 – Resultados das configurações do k-NN na fase de treinamento e validação	99
Tabela 25 – Resultados das configurações dos k-NN na fase de treinamento/validação e teste	100
Tabela 26 – Resultados das configurações do k-NN na fase de validação.....	102
Tabela 27 – Comparação das configurações da RNA com as configurações do k-NN para cada uma das categorias de períodos	106

LISTA DE SIGLAS

ABNT – Associação Brasileira de Normas Técnicas

AREMA – American Railway Engineering and Maintenance-of-Way Association

CCD – Charge-Coupled Device

EFVM – Estrada de Ferro Vitória a Minas

GPS – Sistema de Posicionamento Global

k-NN – K-Vizinhos mais próximos

MPL – Perceptron Multicamadas

NBR – Norma Brasileira Registrada

RAM – Memória de Acesso Aleatório

RNA – Rede Neural Artificial

SUMÁRIO

1	INTRODUÇÃO	13
1.1	OBJETIVOS	16
1.1.1	Objetivo Geral	16
1.1.2	Objetivos Específicos	17
1.2	JUSTIFICATIVA.....	17
1.3	ESTRUTURA DA PROPOSTA DE DISSERTAÇÃO	18
2	REFERENCIAL TEÓRICO	20
2.1	REDE NEURAL ARTIFICIAL	20
2.2	K-VIZINHOS MAIS PRÓXIMOS (K-NN)	30
2.3	REDES NEURAS ARTIFICIAIS APLICADAS EM FERROVIAS.....	31
3	DESCRIÇÃO E CONCEITUAÇÃO DO PROBLEMA	34
3.1	O PROBLEMA ANALISADO	38
3.2	DESCRIÇÃO DA FERROVIA ESTUDADA.....	40
3.3	VEÍCULO DE INSPEÇÃO FERROVIÁRIA - CARRO CONTROLE	42
4	LEVANTAMENTO DE DADOS	48
4.1	TRATAMENTO DOS DADOS	65
5	REDE NEURAL ARTIFICIAL DE CLASSIFICAÇÃO PARA A ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DO TRILHO FERROVIÁRIO.....	79
6	K-VIZINHOS MAIS PRÓXIMOS (K-NN) DE CLASSIFICAÇÃO PARA A ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DO TRILHO FERROVIÁRIO.....	87
7	RESULTADOS EXPERIMENTAIS	89
7.1	RESULTADOS E ANÁLISES DA REDE NEURAL ARTIFICIAL (RNA)	89
7.2	TESTE DA REDE NEURAL ARTIFICIAL (RNA).....	92
7.3	ANÁLISES E RESULTADOS PARA O K-VIZINHO MAIS PRÓXIMO (K-NN)	98

7.4	TESTE DO K-VIZINHOS MAIS PRÓXIMOS (K-NN)	99
7.5	COMPARAÇÃO ENTRE A REDE NEURAL ARTIFICIAL (RNA) E O K- VIZINHOS MAIS PRÓXIMOS (K-NN)	105
8	CONCLUSÕES	108
8.1	TRABALHOS FUTUROS.....	110
	REFERÊNCIAS.....	111

1 INTRODUÇÃO

Os trilhos de aço estão presentes nos sistemas ferroviários ao redor do mundo, desde seus primórdios. O transporte ferroviário tem uma característica que o diferencia dos demais modos de transporte, as rodas dos veículos ferroviários não entram em contato com o solo, tendo os trilhos a função de realizar o suporte, e ser o guia das rodas dos veículos ferroviários (HAY,1982; PROFILLIDIS, 2006; FAIZ, 2010; ELKHOURY et al., 2018). Pode-se afirmar que sem os trilhos não há o transporte ferroviário.

Os trilhos, portanto, exercem função essencial neste sistema de transporte e devem apresentar alto índice de confiabilidade e segurança (ZUMPANO e MEO, 2006). Devido às constantes passagens dos veículos ferroviários sobre os trilhos, gerando o contato entre as rodas das composições e os trilhos, estes apresentam desgastes em sua geometria, sendo obrigatória as suas substituições, para que a ferrovia continue em condições de operação com segurança (PROFILLIDIS, 2006; HAY,1982).

A determinação da vida útil remanescente de um trilho, ou seja, determinar o momento que os trilhos deverão ser substituídos, torna-se essencial para o modo ferroviário, e de suma importância para os operadores ferroviários. Uma vez que, determinar essa vida útil remanescente do trilho pode ser entendido não somente como uma problemática de ordem técnica e de planejamento para os operadores ferroviários, mas também de ordem econômica tendo em vista que os trilhos são os elementos de maior custo financeiro na manutenção na superestrutura ferroviária (ELKHOURY et al., 2018; FALAMARZI et al.,2018; GULER, JOVANOVIĆ e EVREN, 2011; PROFILLIDIS, 2006; ANDRADE e TEIXEIRA, 2012; FALAMARZI et al., 2018; AUDLEY e ANDREWS, 2013, HAY,1982).

Técnicas que possam auxiliar na estimação da vida útil remanescente dos trilhos são vitais, e de suma importância na gestão da manutenção da via permanente ferroviária (FALAMARZI et al., 2018; ANDRADE e TEIXEIRA, 2012). Assim, como hipótese de pesquisa, deseja-se saber se a utilização de algoritmos de aprendizado de máquina pode realizar a estimação da vida útil remanescente dos trilhos ferroviários. Uma vez que, apresentando os dados levantados desta ferrovia específica e sendo os

algoritmos de aprendizado de máquina treinado com estes dados, ou seja, com as características geométricas do trilho, da geometria da via férrea e carregamentos, os algoritmos de aprendizado de máquina poderá estimar a vida útil remanescente dos trilhos ferroviários, de maneira a auxiliar ao planejador na tomada de decisão na substituição destes trilhos na via férrea. Além disso, não foi identificado no campo de pesquisa, trabalhos que buscaram a realização da estimação da vida útil remanescentes de trilhos ferroviários utilizando algoritmos de aprendizado de máquina.

Essa dissertação propõe uma Rede Neural Artificial (RNA) e um algoritmo K-Vizinhos mais próximos (k-NN) que seja capaz de realizar a estimação da vida útil remanescente dos trilhos por meio de categorias de períodos e com base em bancos de dados históricos. Foram definidas quatro categorias: 1) mês; 2) trimestre; 3) semestre; e 4) ano. Essas categorias de vida útil remanescentes do trilho são obtidas dividindo-se a vida útil remanescente, representada por um número contínuo, por, respectivamente, 12, 4, 2 e 1. Assim, têm-se as categorias representadas por uma variável categórica. Essas variáveis categóricas foram denominadas “categorias de períodos de vida útil remanescente”.

Para esta dissertação foram desenvolvidas e testadas configurações das RNA de classificação, utilizando variáveis categóricas. As RNA, independentemente da categoria para a estimação da vida útil remanescente dos trilhos ferroviários, terão 5 variáveis na camada de entrada, ou seja, 5 neurônios na camada de entrada que são: 1) grau da curva; 2) trilho interno ou externo a curva; 3) largura do boleto; 4) altura do boleto; e 5) peso médio transportado.

A camada intermediária, independentemente da categoria para a estimação da vida útil remanescente dos trilhos ferroviários, foi avaliada variando-se a quantidade de camadas intermediárias e a quantidade de neurônios nestas camadas intermediárias. Testou-se configurações de RNA com 1, 2 e 3 camadas intermediárias. Estabeleceu-se que as 3 camadas sempre teriam a mesma quantidade de neurônios. Assim, testou-se camadas intermediárias com: 30, 50, 100, 200 e 400 neurônios para as configurações das RNA para as categorias trimestre, semestre e ano, e 100, 200 e 400 neurônios para as configurações das RNA para a categoria mês. Perfazendo um total de 54 configurações de RNA para serem avaliadas nessa dissertação.

A camada de saída representa a categoria de período da vida útil remanescente dos trilhos ferroviários. Sendo que o número de neurônios da camada de saída é dependente da categoria de período da vida útil remanescente dos trilhos. Assim para a categoria de período mês tem 80 neurônios, para a categoria de período trimestre tem 27 neurônios, para a categoria de período semestre tem 14 neurônios e para a categoria de período ano tem 7 neurônios.

A RNA deverá identificar com base nos 5 neurônios de entrada, qual a categoria de períodos de vida útil que o trilho se classifica. Da mesma maneira deverá ser procedido com o k-NN, que por intermédio das mesmas variáveis de entrada utilizadas para a RNA e com os valores de $k=5$, $k=7$ e $k=9$ deverá categorizar os períodos de vida útil que o trilho se classifica. Assim a utilização da RNA e do k-NN podem auxiliar o planejador da manutenção ferroviária a obter a estimativa da vida útil remanescente dos trilhos.

Utilizaram-se para o treinamento, teste e validação das configurações da RNA os dados levantados da Estrada de Ferro Vitória a Minas (EFVM), para um trecho de 29 km de ferrovia e durante um período de 6 anos e 8 meses. Os dados referentes a geometria dos trilhos e a geometria da via férrea foram levantados por meio de um veículo de inspeção ferroviária denominado Carro Controle. As toneladas médias trafegadas pela ferrovia foram obtidas por meio de relatórios de volumes trafegados da EFVM.

Cabe ressaltar que os dados utilizados na realização desta dissertação foram obtidos da empresa Vale S.A. por intermédio do convênio com a Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES), número 75528452/2016. Neste contrato, a empresa Vale S.A. autoriza a utilização de alguns dados operacionais para fins acadêmicos. A fim de viabilizar a publicação, os valores foram transformados por uma função não divulgada com o objetivo de manutenção do sigilo dos dados apresentados.

O conjunto de dados (*dataset*) para a realização desta pesquisa foi composto de 1.275.034 registros tratados para o período de 6 anos e 8 meses. Os 1.275.034 registros tratados foram separados em 2 partes, dos quais 1.100.000 registros foram utilizados para o treinamento e validação das configurações da RNA e do k-NN, e 175.034 registros foram utilizados para o teste das configurações da RNA e do k-NN.

A RNA e o k-NN foram implementados em Python, versão 3.7, no ambiente *Spyder* versão 3.2.6, onde utilizou-se a biblioteca *Scikit-Learn*, versão 0.20.3. Utilizou-se no modelo o neurônio artificial *Perceptron* e o algoritmo de *Backpropagation*, com a estrutura tipo *Perceptron Multicamadas* (MPL). Para o k-NN utilizou-se o classificador *K-Neighbors Classifier*, os valores utilizados para o reconhecimento do vizinho mais próximos, ou seja, os valores de k , utilizados nas configurações testadas do k-NN foram $k = 5$, $k = 7$ e $k = 9$. Aplicou-se a validação cruzada denominada de *k-fold* para divisão do conjunto de registros, sendo que nesta pesquisa foi utilizado $f=10$.

Para a validação das configurações da RNA e do k-NN, escolheu-se a configuração da RNA e do k-NN que apresentaram o melhor desempenho dentre as configurações investigadas, utilizando-se o valor da acurácia e do *F1-score*. Realizou-se uma análise comparativa dos resultados obtidos das configurações das RNA que apresentaram o melhor desempenho, com os resultados obtidos das configurações do k-NN que apresentaram o melhor desempenho, com o objeto de se verificar o desempenho das configurações das RNA investigadas.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Investigar algoritmos de aprendizado de máquina para estimação da vida útil remanescente dos trilhos ferroviários medida como categorias de períodos de estimação (mês, trimestre, semestre e ano), tomando como base a geometria do trilho, a geometria da via férrea e o carregamento total trafegado e a data de substituição dos trilhos.

1.1.2 Objetivos Específicos

Os objetivos específicos dessa dissertação de mestrado são:

- Investigar e analisar diferentes configurações da RNA e do k-NN e determinar para cada algoritmo, a configuração que obtém o melhor desempenho para estimação das categorias de períodos de vida útil remanescente de trilhos;
- Utilizar dados reais da Estrada de Ferro Vitória a Minas (EFVM) para avaliar as configurações propostas;
 - Realizar o levantamento e coleta dos dados das medições realizadas nos trilhos ferroviários por meio do Carro Controle para o período de 6 anos e 8 meses;
 - Realizar o levantamento das toneladas trafegadas para o período de 6 anos e 8 meses para a ferrovia em análise;
 - Realizar o tratamento de todos os dados levantados para o período de 6 anos e 8 meses;
- Realizar o treinamento, validação e teste das configurações da RNA para a estimação da vida útil remanescente dos trilhos na EFVM;
- Realizar o treinamento, validação e teste das configurações do k-NN para a estimação da vida útil remanescente dos trilhos na EFVM;
- Comparar os resultados das configurações da RNA com a do k-NN para analisar qual dos dois algoritmos de aprendizado de máquina teve melhor desempenho.

1.2 JUSTIFICATIVA

Na literatura científica, encontram-se vários estudos realizados e em desenvolvimentos referentes à utilização dos algoritmos de aprendizado de máquina para as mais diversas áreas do conhecimento, i.e. exatas, humanas e biológicas (JAIN et al., 2013; ADELI, 2001). Entretanto, para a análise das condições da estimação de vida útil remanescente em trilhos ferroviários, não foi identificada nenhuma pesquisa.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL
MESTRADO EM ENGENHARIA CIVIL

GLEDSON FÁBIO COTRIM ROCHA

DISSERTAÇÃO DE MESTRADO

**ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DE TRILHOS FERROVIÁRIOS
POR MEIO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA**

VITÓRIA
2019

GLEDSO N FÁBIO COTRIM ROCHA

DISSERTAÇÃO DE MESTRADO

**ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DE TRILHOS FERROVIÁRIOS
POR MEIO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Civil do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Engenharia Civil, na área de concentração Transportes.

Orientador: Prof. Dr. Rodrigo de Alvarenga Rosa.

Coorientador: Prof. Dr. Ing. Renato Antonio Krohling.

VITÓRIA
2019

Ficha catalográfica disponibilizada pelo Sistema Integrado de
Bibliotecas - SIBI/UFES e elaborada pelo autor

- R672e Rocha, Gledson Fábio Cotrim, 1977-
 Estimação da vida útil remanescente de trilhos ferroviários
 por meio de técnicas de aprendizado de máquina / Gledson
 Fábio Cotrim Rocha. - 2019.
 115 f. : il.
- Orientador: Rodrigo de Alvarenga Rosa.
 Coorientador: Renato Antonio Krohling.
 Dissertação (Mestrado em Engenharia Civil) - Universidade
 Federal do Espírito Santo, Centro Tecnológico.
1. Trilho Ferroviário. 2. Rede Neural Artificial. 3. k
 Vizinhos mais próximos. 4. Superestrutura Ferroviária. I. Rosa,
 Rodrigo de Alvarenga. II. Krohling, Renato Antonio. III.
 Universidade Federal do Espírito Santo. Centro Tecnológico. IV.
 Título.

CDU: 624

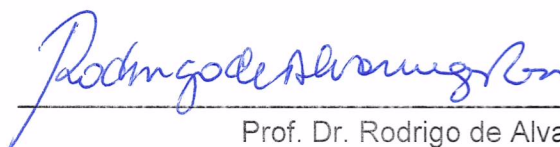
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DE TRILHOS FERROVIÁRIOS POR MEIO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA

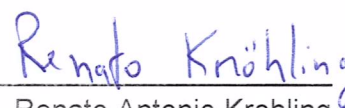
Gledson Fábio Cotrim Rocha

Dissertação apresentada ao Curso de Mestrado em Engenharia Civil do Programa de Pós-Graduação em Engenharia Civil da Universidade Federal do Espírito, como requisito parcial para obtenção do título de Mestre em Engenharia Civil, área de Transportes.

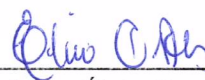
Aprovada no dia **27 de agosto de 2019** por:



Prof. Dr. Rodrigo de Alvarenga Rosa
Doutor em Engenharia Elétrica
Orientador – UFES



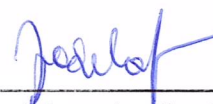
Prof. Dr. Renato Antonio Krohling
Doutor em Engenharia Elétrica
Coordenador – UFES



Prof. Dr. Elcio Cassimiro Alves
Doutor em Engenharia e Tecnologia Espaciais
Examinador Interno – UFES



Prof. Dr. Daniel Cruz Cavaliere
Doutor em Engenharia Elétrica
Examinador Externo - IFES



Prof. Dr. Jodelson Aguiar Sabino
Doutor em Engenharia de Produção
Examinador Externo – Vale S/A

Vitória – ES, agosto de 2019

AGRADECIMENTOS

A Deus por sua infinita sabedoria, força e poder que me impulsionou nesta empreitada de minha vida.

Agradeço a minha esposa Kyla Rocha, pelo apoio, incentivo, companheirismo, compreensão e principalmente por acreditar nos momentos mais difíceis, sendo que sem seu apoio não seria possível a realização deste trabalho. Agradeço ao meu filho João por entender e compreender a ausência de seu pai durante estes momentos.

Agradeço aos meus pais João Rocha e Maria Rocha que sempre me apoiaram e me incentivaram para a realização deste Mestrado e durante a realização deste trabalho.

Agradeço ao professor Rodrigo Rosa por aceitar o desafio na condução deste estudo, por transmitir seus conhecimentos, sugestões, pelo apoio e pelos períodos de orientações, durante a elaboração deste trabalho.

Agradeço ao professor Renato Krohling por todas as sugestões apresentadas na qualificação e elaboração deste trabalho e a banca pela disponibilidade, pelas sugestões apresentadas na qualificação e pelo tempo dedicado a leitura deste trabalho.

Agradeço a Vale S.A. pelo apoio e liberação para cursar o Mestrado, em especial aos meus gestores Edmundo Ferraz e Rogério Ribeiro pelo incentivo.

Aos meus colegas de trabalho da Vale S.A., pelo incentivo na busca do conhecimento, apoio e compreensão.

Aos professores do Programa de Pós-Graduação em Engenharia Civil da UFES, por partilharem seus conhecimentos e experiências.

Aos meus amigos e demais familiares pelo apoio e incentivo constante durante a realização do Mestrado e deste trabalho.

De maneira geral, agradeço a todos que de alguma maneira contribuíram e torceram para essa realização.

Bem-aventurado o homem que acha
sabedoria, e o homem que adquire
conhecimento.

Provérbios 3:13

RESUMO

Sendo o trilho o elemento de maior custo na via permanente, buscar ferramentas para estimar a vida útil remanescente dos trilhos é importante para utilizá-los ao máximo. Assim, tem-se como hipótese que algoritmos de aprendizado de máquina treinados a partir de bases de dados históricos, possam auxiliar ao planejador na realização dessa estimativa. Foram testados dois algoritmos de aprendizado de máquina: Rede Neural Artificial (RNA) e K-Vizinhos mais próximos (k-NN). As bases de dados históricos utilizados foram da Estrada de Ferro Vitória a Minas (EFVM). O conjunto de dados possui 1.275.034 registros referente ao período de 6 anos e 8 meses. As RNA tiveram 5 neurônios na camada de entrada: 1) o grau da curva; 2) trilho interno ou externo a curva; 3) largura do boleto; 4) altura do boleto; e 5) peso médio transportado. A camada intermediária, independentemente da categoria para a estimativa da vida útil remanescente dos trilhos, teve variações de 1, 2 e 3 camadas e variações na quantidade de neurônios de 30, 50, 100, 200 e 400. A camada de saída depende da categoria de período da vida útil remanescente dos trilhos: 1) mês, com 80 neurônios, 2) trimestre, com 27 neurônios, 3) semestre, com 14 neurônios e 4) ano, com 7 neurônios. Para o algoritmo k-NN, foram testadas configurações variando $k=5, 7$ e 9 . Para ambos algoritmos, aplicou-se a validação cruzada $k-fold$, com $f=10$, e o desempenho foi avaliado por meio do valor da acurácia e do $F1-score$. A linguagem de programação foi *Python* e a biblioteca *Scikit-Learn*. As configurações das RNA e as configurações do k-NN foram comparadas e o k-NN apresentou resultados superiores as RNA. Entretanto, ambos algoritmos atingiram o objetivo proposto nesta dissertação que era a estimativa da vida útil remanescente dos trilhos de maneira a auxiliar o planejador da manutenção da via férrea quanto à substituição dos trilhos, onde obtiveram resultados na acurácia e $F1-score$ superiores a 80% em ambos os algoritmos para as categorias de período semestre e ano, sendo estas categorias de períodos mais utilizadas pelos operadores ferroviários. O algoritmo k-NN obteve sempre melhores resultados quando comparados com a RNA.

Palavras-chave: Trilho Ferroviário. Rede Neural Artificial. k-Vizinhos mais próximos. Superestrutura Ferroviária.

ABSTRACT

As the rail is the costliest element on the permanent track, finding tools to estimate the remaining life is important to make the most of them. Thus, it has hypothesized that machine learning algorithms trained from historical databases can assist the planner in performing this estimation. Two machine learning algorithms were tested: Artificial Neural Network (RNA) and Nearest K-Neighbors (k-NN). The historical databases used were from Vitória to Minas Railway (EFVM). The dataset has 1,275,034 records for the period of 6 years and 8 months. The RNA had 5 neurons in the input layer: 1) the degree of the curve; 2) internal or external rail curve; 3) the width of the rail; 4) the height of rail; and 5) average weight carried. The intermediate layer, regardless of the category for estimating the remaining rail life, had variations of 1, 2 and 3 layers and variations in the number of neurons of 30, 50, 100, 200 and 400. The output layer depends on the period category the remaining useful life of the rails: 1) month, with 80 neurons, 2) quarter, with 27 neurons, 3) semester, with 14 neurons and 4) year, with 7 neurons. For the k-NN algorithm, configurations ranging from $k = 5, 7$ and 9 were tested. For both algorithms, k-fold cross validation was applied, with $f = 10$, and performance was evaluated using the accuracy value and F1-score. The programming language was Python and the Scikit-Learn library. RNA configurations and k-NN configurations were compared and k-NN showed superior results to RNA. However, both algorithms reached the objective proposed in this dissertation, which was the estimation of the remaining rail life in order to help the railroad maintenance planner to replace the rails, where they obtained results in accuracy and F1-score higher than 80% for both algorithms for the semester and year period categories, these period categories being most used by rail operators. The k-NN algorithm always obtained better results when compared to RNA.

Keywords: Rail Track. Artificial neural network. k-Nearest neighbors. Railway superstructure.

LISTA DE FIGURAS

Figura 1 – Pseudocódigo da RNA em Backpropagation	23
Figura 2 – Divisão dos dados para treinamento, <i>validação e teste</i>	25
Figura 3 – Divisão dos dados por k-fold cross validation	26
Figura 4 – Obtenção dos valores da precision e recall.....	28
Figura 5 – Pseudocódigo do k-NN	31
Figura 6 - Partes do trilho ferroviário tipo Vignole	35
Figura 7 – Seção da superestrutura ferroviária	37
Figura 8 – Distribuição das categorias de estimação da vida útil remanescente para um horizonte de 6 anos e 8 meses.....	39
Figura 9 – Mapa geral do traçado da EFVM.....	41
Figura 10 – Vagão de minério GDE	41
Figura 11 – Perfil do trilho TR-68	42
Figura 12 – Carro Controle EM-80 - EFVM	43
Figura 13 – Dimensões do Carro Controle EM-80 - EFVM	43
Figura 14 – Localização do sistema de leitura do Carro Controle EM-80 - EFVM	44
Figura 15 – Sistema de leitura laser do Carro Controle EM-80 - EFVM.....	45
Figura 16 – Leitura da geometria do trilho na tela do computador de bordo do Carro Controle.....	46
Figura 17 – Leitura da geometria do trilho pelo Carro Controle EM-80	47
Figura 18 – Trecho da coletada de dados na EFVM	48
Figura 19 – Distribuição dos registros para o período de 2011 a 2017 trilho esquerdo – linha 1 – EFVM	50
Figura 20 – Distribuição dos registros referentes aos trechos em tangentes e curvas para o trilho esquerdo – linha 1 - EFVM	51
Figura 21 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2011 - linha 1 - EFVM	52
Figura 22 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2012 - linha 1 - EFVM.....	53
Figura 23 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2013 - linha 1 - EFVM.....	54

Figura 24 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2014 - linha 1 - EFVM	55
Figura 25 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2015 - linha 1 - EFVM	56
Figura 26 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2016 - linha 1 - EFVM	57
Figura 27 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2017 - linha 1 - EFVM	58
Figura 28 – Distribuição das categorias de estimação da vida útil remanescente para um horizonte de 6 anos e 8 meses	71
Figura 29 – Distribuição dos registros no conjunto de dados – categoria mês	73
Figura 30 – Distribuição dos registros no conjunto de dados – categoria trimestre ..	74
Figura 31 – Distribuição da categoria trimestre para os seus respectivos meses	75
Figura 32 – Distribuição dos registros no conjunto de dados – categoria semestre..	76
Figura 33 – Distribuição da categoria semestre para os seus respectivos meses	76
Figura 34 – Distribuição dos registros no conjunto de dados – categoria ano	77
Figura 35 – Distribuição da categoria ano para os seus respectivos meses	78
Figura 36 – RNA investigadas para estimação da vida útil remanescente de trilhos ferroviários	86
Figura 37 – Matriz de confusão da Configuração 39 - Semestre	96
Figura 38 – Matriz de confusão da Configuração 39 - Semestre	96
Figura 39 – Matriz de confusão da Configuração 54 - Ano	97
Figura 40 – Matriz de confusão da Configuração 54 - Ano	97
Figura 41 – Matriz de confusão da Configuração 61 - Semestre	103
Figura 42 – Matriz de confusão da Configuração 61 - Semestre	103
Figura 43 – Matriz de confusão da Configuração 64 – Ano	104
Figura 44 – Matriz de confusão da Configuração 64 – Ano	104
Figura 45 – Comparação das configurações com melhores desempenho da RNA e K-NN	106

LISTA DE TABELAS

Tabela 1 – Matriz de confusão	29
Tabela 2 – Raios e grau da curva - km 16+696 ao km 45+742 - EFVM.....	58
Tabela 3 – Peso da carga transportada sobre a linha 1 - EFVM.....	61
Tabela 4 – Quantidade de vagões que trafegaram sobre a linha 1 - EFVM.....	61
Tabela 5 – Peso total dos vagões que trafegaram na linha 1 - EFVM	62
Tabela 6 – Quantidade de locomotivas que trafegaram sobre a linha 1 – EFVM.....	62
Tabela 7 – Peso total das locomotivas que trafegaram sobre a linha 1 - EFVM	63
Tabela 8 – Peso do carregamento total trafegado sobre a linha 1 - EFVM.....	63
Tabela 9 – Conjunto de Dados.....	66
Tabela 10 – Conjunto de dados final da linha 1 - EFVM	67
Tabela 11 – Levantamento realizado para o trilho esquerdo da linha 1 – EFVM (35+737).....	69
Tabela 12 – Levantamento do peso médio trafegado da linha 1 – EFVM (35+737) .	70
Tabela 13 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – mês.....	82
Tabela 14 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – trimestre.....	82
Tabela 15 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – semestre.....	83
Tabela 16 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – ano	83
Tabela 17 – Configurações do k-NN para cada categoria de período.....	88
Tabela 18 – Resultados de treinamento/validação da categoria de período mês – grupo G1	90
Tabela 19 – Resultados de treinamento/validação da categoria de período trimestre – grupo G2	90
Tabela 20 – Resultados de treinamento/validação da categoria de período semestre – grupo G3	91
Tabela 21 – Resultados de treinamento/validação da categoria de período ano – grupo G4	91

Tabela 22 – Resultados das configurações da RNA na fase de treinamento/validação e teste	93
Tabela 23 – Resultados das configurações da RNA na fase de teste.....	95
Tabela 24 – Resultados das configurações do k-NN na fase de treinamento e validação	99
Tabela 25 – Resultados das configurações dos k-NN na fase de treinamento/validação e teste	100
Tabela 26 – Resultados das configurações do k-NN na fase de validação.....	102
Tabela 27 – Comparação das configurações da RNA com as configurações do k-NN para cada uma das categorias de períodos	106

LISTA DE SIGLAS

ABNT – Associação Brasileira de Normas Técnicas

AREMA – American Railway Engineering and Maintenance-of-Way Association

CCD – Charge-Coupled Device

EFVM – Estrada de Ferro Vitória a Minas

GPS – Sistema de Posicionamento Global

k-NN – K-Vizinhos mais próximos

MPL – Perceptron Multicamadas

NBR – Norma Brasileira Registrada

RAM – Memória de Acesso Aleatório

RNA – Rede Neural Artificial

SUMÁRIO

1	INTRODUÇÃO	13
1.1	OBJETIVOS	16
1.1.1	Objetivo Geral	16
1.1.2	Objetivos Específicos	17
1.2	JUSTIFICATIVA.....	17
1.3	ESTRUTURA DA PROPOSTA DE DISSERTAÇÃO	18
2	REFERENCIAL TEÓRICO	20
2.1	REDE NEURAL ARTIFICIAL	20
2.2	K-VIZINHOS MAIS PRÓXIMOS (K-NN)	30
2.3	REDES NEURAS ARTIFICIAIS APLICADAS EM FERROVIAS.....	31
3	DESCRIÇÃO E CONCEITUAÇÃO DO PROBLEMA	34
3.1	O PROBLEMA ANALISADO	38
3.2	DESCRIÇÃO DA FERROVIA ESTUDADA.....	40
3.3	VEÍCULO DE INSPEÇÃO FERROVIÁRIA - CARRO CONTROLE	42
4	LEVANTAMENTO DE DADOS	48
4.1	TRATAMENTO DOS DADOS	65
5	REDE NEURAL ARTIFICIAL DE CLASSIFICAÇÃO PARA A ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DO TRILHO FERROVIÁRIO.....	79
6	K-VIZINHOS MAIS PRÓXIMOS (K-NN) DE CLASSIFICAÇÃO PARA A ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DO TRILHO FERROVIÁRIO.....	87
7	RESULTADOS EXPERIMENTAIS	89
7.1	RESULTADOS E ANÁLISES DA REDE NEURAL ARTIFICIAL (RNA)	89
7.2	TESTE DA REDE NEURAL ARTIFICIAL (RNA).....	92
7.3	ANÁLISES E RESULTADOS PARA O K-VIZINHO MAIS PRÓXIMO (K-NN)	98

7.4	TESTE DO K-VIZINHOS MAIS PRÓXIMOS (K-NN)	99
7.5	COMPARAÇÃO ENTRE A REDE NEURAL ARTIFICIAL (RNA) E O K- VIZINHOS MAIS PRÓXIMOS (K-NN)	105
8	CONCLUSÕES	108
8.1	TRABALHOS FUTUROS.....	110
	REFERÊNCIAS.....	111

1 INTRODUÇÃO

Os trilhos de aço estão presentes nos sistemas ferroviários ao redor do mundo, desde seus primórdios. O transporte ferroviário tem uma característica que o diferencia dos demais modos de transporte, as rodas dos veículos ferroviários não entram em contato com o solo, tendo os trilhos a função de realizar o suporte, e ser o guia das rodas dos veículos ferroviários (HAY,1982; PROFILLIDIS, 2006; FAIZ, 2010; ELKHOURY et al., 2018). Pode-se afirmar que sem os trilhos não há o transporte ferroviário.

Os trilhos, portanto, exercem função essencial neste sistema de transporte e devem apresentar alto índice de confiabilidade e segurança (ZUMPANO e MEO, 2006). Devido às constantes passagens dos veículos ferroviários sobre os trilhos, gerando o contato entre as rodas das composições e os trilhos, estes apresentam desgastes em sua geometria, sendo obrigatória as suas substituições, para que a ferrovia continue em condições de operação com segurança (PROFILLIDIS, 2006; HAY,1982).

A determinação da vida útil remanescente de um trilho, ou seja, determinar o momento que os trilhos deverão ser substituídos, torna-se essencial para o modo ferroviário, e de suma importância para os operadores ferroviários. Uma vez que, determinar essa vida útil remanescente do trilho pode ser entendido não somente como uma problemática de ordem técnica e de planejamento para os operadores ferroviários, mas também de ordem econômica tendo em vista que os trilhos são os elementos de maior custo financeiro na manutenção na superestrutura ferroviária (ELKHOURY et al., 2018; FALAMARZI et al.,2018; GULER, JOVANOVIĆ e EVREN, 2011; PROFILLIDIS, 2006; ANDRADE e TEIXEIRA, 2012; FALAMARZI et al., 2018; AUDLEY e ANDREWS, 2013, HAY,1982).

Técnicas que possam auxiliar na estimação da vida útil remanescente dos trilhos são vitais, e de suma importância na gestão da manutenção da via permanente ferroviária (FALAMARZI et al., 2018; ANDRADE e TEIXEIRA, 2012). Assim, como hipótese de pesquisa, deseja-se saber se a utilização de algoritmos de aprendizado de máquina pode realizar a estimação da vida útil remanescente dos trilhos ferroviários. Uma vez que, apresentando os dados levantados desta ferrovia específica e sendo os

algoritmos de aprendizado de máquina treinado com estes dados, ou seja, com as características geométricas do trilho, da geometria da via férrea e carregamentos, os algoritmos de aprendizado de máquina poderá estimar a vida útil remanescente dos trilhos ferroviários, de maneira a auxiliar ao planejador na tomada de decisão na substituição destes trilhos na via férrea. Além disso, não foi identificado no campo de pesquisa, trabalhos que buscaram a realização da estimação da vida útil remanescentes de trilhos ferroviários utilizando algoritmos de aprendizado de máquina.

Essa dissertação propõe uma Rede Neural Artificial (RNA) e um algoritmo K-Vizinhos mais próximos (k-NN) que seja capaz de realizar a estimação da vida útil remanescente dos trilhos por meio de categorias de períodos e com base em bancos de dados históricos. Foram definidas quatro categorias: 1) mês; 2) trimestre; 3) semestre; e 4) ano. Essas categorias de vida útil remanescentes do trilho são obtidas dividindo-se a vida útil remanescente, representada por um número contínuo, por, respectivamente, 12, 4, 2 e 1. Assim, têm-se as categorias representadas por uma variável categórica. Essas variáveis categóricas foram denominadas “categorias de períodos de vida útil remanescente”.

Para esta dissertação foram desenvolvidas e testadas configurações das RNA de classificação, utilizando variáveis categóricas. As RNA, independentemente da categoria para a estimação da vida útil remanescente dos trilhos ferroviários, terão 5 variáveis na camada de entrada, ou seja, 5 neurônios na camada de entrada que são: 1) grau da curva; 2) trilho interno ou externo a curva; 3) largura do boleto; 4) altura do boleto; e 5) peso médio transportado.

A camada intermediária, independentemente da categoria para a estimação da vida útil remanescente dos trilhos ferroviários, foi avaliada variando-se a quantidade de camadas intermediárias e a quantidade de neurônios nestas camadas intermediárias. Testou-se configurações de RNA com 1, 2 e 3 camadas intermediárias. Estabeleceu-se que as 3 camadas sempre teriam a mesma quantidade de neurônios. Assim, testou-se camadas intermediárias com: 30, 50, 100, 200 e 400 neurônios para as configurações das RNA para as categorias trimestre, semestre e ano, e 100, 200 e 400 neurônios para as configurações das RNA para a categoria mês. Perfazendo um total de 54 configurações de RNA para serem avaliadas nessa dissertação.

A camada de saída representa a categoria de período da vida útil remanescente dos trilhos ferroviários. Sendo que o número de neurônios da camada de saída é dependente da categoria de período da vida útil remanescente dos trilhos. Assim para a categoria de período mês tem 80 neurônios, para a categoria de período trimestre tem 27 neurônios, para a categoria de período semestre tem 14 neurônios e para a categoria de período ano tem 7 neurônios.

A RNA deverá identificar com base nos 5 neurônios de entrada, qual a categoria de períodos de vida útil que o trilho se classifica. Da mesma maneira deverá ser procedido com o k-NN, que por intermédio das mesmas variáveis de entrada utilizadas para a RNA e com os valores de $k=5$, $k=7$ e $k=9$ deverá categorizar os períodos de vida útil que o trilho se classifica. Assim a utilização da RNA e do k-NN podem auxiliar o planejador da manutenção ferroviária a obter a estimativa da vida útil remanescente dos trilhos.

Utilizaram-se para o treinamento, teste e validação das configurações da RNA os dados levantados da Estrada de Ferro Vitória a Minas (EFVM), para um trecho de 29 km de ferrovia e durante um período de 6 anos e 8 meses. Os dados referentes a geometria dos trilhos e a geometria da via férrea foram levantados por meio de um veículo de inspeção ferroviária denominado Carro Controle. As toneladas médias trafegadas pela ferrovia foram obtidas por meio de relatórios de volumes trafegados da EFVM.

Cabe ressaltar que os dados utilizados na realização desta dissertação foram obtidos da empresa Vale S.A. por intermédio do convênio com a Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES), número 75528452/2016. Neste contrato, a empresa Vale S.A. autoriza a utilização de alguns dados operacionais para fins acadêmicos. A fim de viabilizar a publicação, os valores foram transformados por uma função não divulgada com o objetivo de manutenção do sigilo dos dados apresentados.

O conjunto de dados (*dataset*) para a realização desta pesquisa foi composto de 1.275.034 registros tratados para o período de 6 anos e 8 meses. Os 1.275.034 registros tratados foram separados em 2 partes, dos quais 1.100.000 registros foram utilizados para o treinamento e validação das configurações da RNA e do k-NN, e 175.034 registros foram utilizados para o teste das configurações da RNA e do k-NN.

A RNA e o k-NN foram implementados em Python, versão 3.7, no ambiente *Spyder* versão 3.2.6, onde utilizou-se a biblioteca *Scikit-Learn*, versão 0.20.3. Utilizou-se no modelo o neurônio artificial *Perceptron* e o algoritmo de *Backpropagation*, com a estrutura tipo *Perceptron Multicamadas* (MPL). Para o k-NN utilizou-se o classificador *K-Neighbors Classifier*, os valores utilizados para o reconhecimento do vizinho mais próximos, ou seja, os valores de k , utilizados nas configurações testadas do k-NN foram $k = 5$, $k = 7$ e $k = 9$. Aplicou-se a validação cruzada denominada de *k-fold* para divisão do conjunto de registros, sendo que nesta pesquisa foi utilizado $f=10$.

Para a validação das configurações da RNA e do k-NN, escolheu-se a configuração da RNA e do k-NN que apresentaram o melhor desempenho dentre as configurações investigadas, utilizando-se o valor da acurácia e do *F1-score*. Realizou-se uma análise comparativa dos resultados obtidos das configurações das RNA que apresentaram o melhor desempenho, com os resultados obtidos das configurações do k-NN que apresentaram o melhor desempenho, com o objeto de se verificar o desempenho das configurações das RNA investigadas.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Investigar algoritmos de aprendizado de máquina para estimação da vida útil remanescente dos trilhos ferroviários medida como categorias de períodos de estimação (mês, trimestre, semestre e ano), tomando como base a geometria do trilho, a geometria da via férrea e o carregamento total trafegado e a data de substituição dos trilhos.

1.1.2 Objetivos Específicos

Os objetivos específicos dessa dissertação de mestrado são:

- Investigar e analisar diferentes configurações da RNA e do k-NN e determinar para cada algoritmo, a configuração que obtém o melhor desempenho para estimação das categorias de períodos de vida útil remanescente de trilhos;
- Utilizar dados reais da Estrada de Ferro Vitória a Minas (EFVM) para avaliar as configurações propostas;
 - Realizar o levantamento e coleta dos dados das medições realizadas nos trilhos ferroviários por meio do Carro Controle para o período de 6 anos e 8 meses;
 - Realizar o levantamento das toneladas trafegadas para o período de 6 anos e 8 meses para a ferrovia em análise;
 - Realizar o tratamento de todos os dados levantados para o período de 6 anos e 8 meses;
- Realizar o treinamento, validação e teste das configurações da RNA para a estimação da vida útil remanescente dos trilhos na EFVM;
- Realizar o treinamento, validação e teste das configurações do k-NN para a estimação da vida útil remanescente dos trilhos na EFVM;
- Comparar os resultados das configurações da RNA com a do k-NN para analisar qual dos dois algoritmos de aprendizado de máquina teve melhor desempenho.

1.2 JUSTIFICATIVA

Na literatura científica, encontram-se vários estudos realizados e em desenvolvimentos referentes à utilização dos algoritmos de aprendizado de máquina para as mais diversas áreas do conhecimento, i.e. exatas, humanas e biológicas (JAIN et al., 2013; ADELI, 2001). Entretanto, para a análise das condições da estimação de vida útil remanescente em trilhos ferroviários, não foi identificada nenhuma pesquisa.

Logo, esta pesquisa visa preencher essa lacuna, tornando-se uma oportunidade para o desenvolvimento e avanço neste campo de pesquisa.

Identificaram-se algumas pesquisas desenvolvidas utilizando algoritmos de aprendizado de máquina, principalmente RNA para questões de verificações nos processos de predições industriais, como na fabricação de trilhos e rodas ferroviárias, e quanto às questões de falhas durante o processo de fabricação e laminação de trilhos ferroviários. Identificaram-se também outras pesquisas para análises das condições da geometria da linha ferroviária e do lastro ferroviário.

Essa dissertação se justifica por investigar técnicas de aprendizado de máquina capaz de realizar a estimação da categoria de período da vida útil remanescente de trilhos ferroviários, auxiliando o planejador da manutenção na tomada de decisões quanto a necessidade da substituição dos trilhos ferroviários. Conhecendo a categoria de período de estimação da vida útil remanescente do trilho ferroviário, será possível planejar a realização da substituição dos trilhos. Desta maneira, conseguirá realizar uma alocação orientada na aplicação dos recursos financeiros e humanos de forma estruturada em relação ao tempo e ao longo da ferrovia, mantendo-se como premissa fundamental a confiabilidade e a segurança operacional da ferrovia.

1.3 ESTRUTURA DA PROPOSTA DE DISSERTAÇÃO

Para atender aos objetivos desta dissertação, o conteúdo desenvolvido foi estruturado em 8 capítulos, incluindo este primeiro que é uma introdução ao problema, e os demais distribuídos da seguinte maneira:

No Capítulo 2 são apresentados os principais conceitos de RNA e do k-NN e o referencial teórico existente sobre a utilização das RNA para ferrovias.

O Capítulo 3 apresenta a revisão da literatura sobre os principais conceitos de trilhos ferroviários, o problema analisado, a descrição da ferrovia estudada e o veículo de inspeção ferroviária.

O Capítulo 4 apresenta o levantamento de dados históricos utilizados nesta dissertação, como o carregamento trafegado na ferrovia, o carregamento dos vagões vazios e das locomotivas, os dados referentes a geometria da linha ferroviária e das substituições dos trilhos e o tratamento destes dados para serem utilizados nesta dissertação.

O Capítulo 5 explora as configurações investigadas das RNA de classificação para a estimação da vida útil remanescente dos trilhos ferroviários, por meio de categorias de períodos.

O Capítulo 6 apresenta as configurações investigadas do k-NN de classificação para a estimação da vida útil remanescente dos trilhos ferroviários, por meio de categorias de períodos.

O Capítulo 7 apresenta os resultados experimentais e análises das configurações investigadas para as RNA e para o k-NN, e a comparação entre a RNA e o k-NN.

Por fim, no Capítulo 8 apresentam-se as conclusões desta dissertação e sugestões de trabalhos futuros a serem realizados.

2 REFERENCIAL TEÓRICO

Neste capítulo são apresentados os principais conceitos referente a Rede Neural Artificial (RNA) e o K-Vizinhos mais próximos (k-NN) e o referencial teórico existente sobre a utilização da RNA para ferrovias.

2.1 REDE NEURAL ARTIFICIAL

Nos últimos anos, houve um aumento crescente na utilização da RNA para os mais diversos campos de pesquisa, sendo que as RNA apresentam resultados satisfatórios quando envolvem conjuntos de dados, e problemas altamente complexos e mal definidos, onde os seres humanos provavelmente decidiriam por meio de uma análise com base intuitiva e sensitiva. Logo, as RNA podem ser utilizadas para resolverem funções que dependem de muitas entradas e incógnitas (KALOGIROU, 2001; ELKHOURY et al., 2018).

Pode-se dizer que as RNA possuem a capacidade de reproduzirem as habilidades cognitivas de especialistas humanos. Por isso, as RNA tornam-se atrativas para a resolução de problemas onde não há uma teoria subjacente e formal para a resolução destes. As RNA podem ser utilizadas como uma ferramenta para o auxílio na tomada de decisões das mais diversas áreas de estudo (FALAMARZI et al., 2018; ELKHOURY et al., 2018; CAO, 2015; BARAI, e PANDEY, 1995).

A principal propriedade e vantagem da RNA é a capacidade de aprender por meio de exemplos, ou seja, pelo treinamento implementado, o que às torna um método geral para construção de mapeamentos entre dois conjuntos de dados e por serem capazes de lidar com problemas não lineares. Outra característica é a habilidade de conseguir generalizar, pois, uma vez construído o mapeamento, a RNA obtida pode ser utilizada para classificar padrões desconhecidos e conseguir estimar condições futuras, por meio dos treinamentos realizados (KROHLING e BRAUN, 2016; ADELI, 2001; KALOGIROU, 2001).

As RNA são sistemas de processamentos paralelos de informações constituído pelas interconexões de pequenas unidades de processamento, os quais são denominados de neurônios artificiais. Os neurônios artificiais são unidades de processamento que têm o funcionamento inspirado nos neurônios biológicos, ou seja, os neurônios recebem sinais de entrada de outros neurônios e produzem um sinal de saída. Esses neurônios artificiais são conectados entre si por meio de pesos. Para a determinação desses pesos, as RNA utilizam o processo de aprendizagem indutiva, ou seja, por meio de exemplos e observações em conjunto de dados inseridos na RNA (PUTRA e SAFRILAH, 2017; ELKHOURY et al., 2018).

A estrutura básica da configuração de uma RNA consiste em geral nas seguintes divisões de camadas: a camada de entrada, a camada intermediária ou oculta e a camada de saída. Na camada de entrada os dados são apresentados para a RNA. Na camada intermediária ou oculta é onde realiza-se o processamento, por intermédio das conexões entre os neurônios, esta camada pode ser considerada como a extratora de características da RNA. Na camada intermediária é que se propicia a obtenção de soluções, ligada principalmente na quantidade de neurônios nesta camada, e na quantidade destas camadas. A camada de saída é onde se apresenta o resultado do processamento da RNA (CATTAN e MOHAMMADI, 1997).

As melhores configurações das RNA frequentemente são determinadas por meio de diferentes testes, em relação aos números de neurônios nas camadas intermediárias e nas quantidades de camadas intermediárias. É, portanto, comum testar e comparar as diferentes configurações das RNA para identificar a configuração que apresenta o melhor desempenho (WOLDEMARIAM, MURILLO-HOYOS e LABI, 2015; DUDA, HART e STORK, 2012). Assim, a configuração mais adequada para uma RNA depende da área de pesquisa e dos conjuntos de dados disponíveis a serem inseridos para a realização do treinamento, validação e teste da RNA (ELKHOURY et al., 2018).

Os dados a serem analisados são apresentados para os neurônios da camada de entrada. Assim, estes dados são carregados para os neurônios da camada intermediária para processá-los e, finalmente, eles se movem para os neurônios da camada de saída para produzirem os resultados. Os neurônios recebem a ativação ponderada de outros neurônios mediante suas conexões de entrada (PUTRA e SAFRILAH, 2017).

Uma estrutura de RNA comumente utilizada é o *Perceptron* Multicamadas (*Multi-layer Perceptron* - MLP). O MLP consiste em camadas de neurônios definidos em camadas de entrada, intermediária e de saída, e que estão totalmente conectados com as camadas adjacentes. Essas conexões são direcionadas apenas para direções à frente (*feed-forward*). A RNA do tipo MLP também são caracterizadas como sendo RNA supervisionadas, uma vez que devem ser fornecidos os dados de entradas e de saídas (WOLDEMARIAM, MURILLO-HOYOS e LABI, 2015).

As RNA definidas como MLP são treinadas de forma a estimar os pesos entre os neurônios de diferentes camadas. Este procedimento de estimativa de pesos se processa com pesos aleatórios iniciais, com o objetivo de minimizar o erro quadrático médio (WOLDEMARIAM, MURILLO-HOYOS e LABI, 2015).

Dentre os algoritmos de treinamento para as RNA existentes, o *Backpropagation* é o mais empregado. No campo da Engenharia Civil e da Engenharia Ferroviária, em geral, as aplicações das RNA baseiam-se principalmente no algoritmo de *Backpropagation* (KROHLING e BRAUN, 2016; ADELI, 2001; DUDA, HART e STORK, 2012).

O algoritmo *Backpropagation* trata-se de um treinamento supervisionado da RNA, tendo como objetivo minimizar o erro observado nas saídas da RNA, corrigindo os pesos da saída para a entrada. O erro do treinamento, pode ser entendido como sendo a diferença quadrática da soma das unidades de saída obtidas pelo treinamento, com a saída desejada (ADELI, 2001; WOLDEMARIAM, MURILLO-HOYOS e LABI, 2015, DUDA, HART e STORK, 2012). A regra de treinamento do algoritmo *Backpropagation* é baseada na descida de gradiente, onde os pesos são inicializados com valores aleatórios e são alterados na direção que conseguirão reduzir o erro (DUDA, HART e STORK, 2012).

O algoritmo *Backpropagation* têm a função de recursivamente, apresentar a RNA os padrões dos dados de treinamento, comparando os resultados obtidos das saídas processadas pela RNA, com as desejadas, e utilizando essas informações para ajustar os pesos da RNA (KROHLING e BRAUN, 2016). Assim, a camada intermediária recebe o somatório de valores de entrada, ponderados da camada de entrada. Depois de ponderá-los, a camada intermediária envia os valores de entrada para a camada de saída. Quando a camada de saída recebe os valores ponderados,

o valor da saída ponderada resultante é comparado com o valor desejado. Se o valor de saída for diferente do valor desejado, o erro é propagado de volta para a camada intermediária para correção, com o objetivo de ajustar os pesos da RNA, sendo que esse processo ocorre recursivamente até um critério de parada seja satisfeito (WOLDEMARIAM, MURILLO-HOYOS e LABI, 2015).

O treinamento da RNA, baseado na descida de gradiente, tem como critério de validação do treinamento e critério de parada, quando se atinge a um erro mínimo de validação estabelecido para a RNA (DUDA, HART e STORK, 2012). A função de ativação, pode ser entendido como o relacionamento que existe entre entradas e saídas de um neurônio e uma RNA (WOLDEMARIAM, MURILLO-HOYOS e LABI, 2015). Uma propriedade que é buscada e de suma importância na função de ativação, é a simplicidade computacional, onde procura-se uma função, cujo valor e derivada possam ser facilmente computados. Uma classe de função de ativação que possui essa propriedade é a sigmóide. A sigmóide é uma função contínua, diferenciável, não linear e saturante, sendo que admite um modelo linear se os pesos da RNA forem pequenos, conseguindo-se maximizar a transmissão de informações para recursos que são normalmente distribuídos (DUDA, HART e STORK, 2012). Na Figura 1 é apresentado o pseudocódigo de treinamento *Backpropagation* de uma RNA.

Figura 1 – Pseudocódigo da RNA em Backpropagation

```

1  Input: TamanhoProblema, InputParametros, MaxIteracoes, TaxaAprendizagem
2  Output: RNA
3  RNA = ConstruirCamadasRede()
4  PesosRNA = InicializarPesos(RNA, TamanhoProblema)
5  Para (i=1 até MaxIteracoes)
6      Parametros(i) = SelecionarInputParametros (InputParametros)
7      Saida(i) = ForwardPropagacao (Parametros(i), RNA)
8      BackwardPropagacaoErro (Parametros(i), Saida(i), RNA)
9      AtualizarPesos (Parametros(i), Saida(i), RNA, TaxaAprendizagem)
10 FimPara
11 Retonar (RNA)

```

Fonte: Mitchell (1997, adaptado)

Atualmente, a função de ativação *ReLU* é utilizada nas RNA. Sendo que a função *ReLU* não é linear, o que significa que pode facilmente copiar os erros para trás e ativar várias camadas de neurônios por esta função. A principal vantagem de usar a função *ReLU* sobre as outras funções de ativação é que ela não ativa todos os

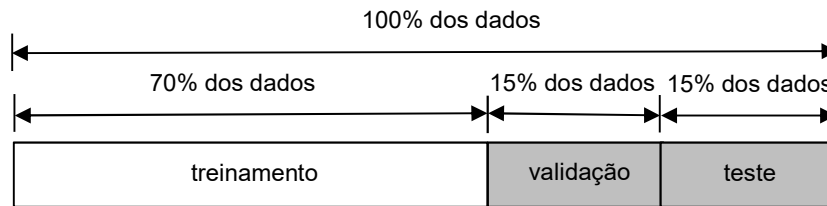
neurônios ao mesmo tempo. Um outro ponto, é que com a função *ReLU* consegue-se otimizar, uma vez que esta função é extremamente parecida com a função identidade, sendo assim, a função *ReLU* é melhor do que a função sigmóide (GÉRON, 2017).

O *solver Adam* utilizado em RNA refere-se a um otimizador baseado em gradiente estocástico. *Adam* é um algoritmo para otimização baseada em gradiente de primeira ordem de funções objetivas estocásticas, baseado em estimativas adaptativas de momentos de baixa ordem. O método é de simples implementação, e computacionalmente eficiente, necessitando de poucos requisitos de memória, é invariante para o reescalonamento diagonal dos gradientes e é adequado para problemas que apresentam grandes quantidades de dados ou de parâmetros (KINGMA e BA, 2014).

Para que se consiga gerar resultados satisfatórios em uma RNA, torna-se necessário um conjunto de dados devidamente tratados (PUTRA e SAFRILAH, 2017). Em geral, os desenvolvedores de RNA utilizam 70% dos dados para o treinamento das RNA, enquanto o restante, ou seja, os 30% são utilizados para a realização da validação e teste das RNA. Dos 30% dos dados separados para a realização da validação e teste das RNA, 15% são utilizados para validação e 15% são utilizados para o teste das RNA. A divisão desta maneira no conjunto de dados para a RNA produz impactos na precisão dos resultados (ELKHOURY et al., 2018).

A desvantagem da maneira da divisão dos dados em 70% dos dados para treinamento e 30% dos dados para validação e teste da RNA é que os resultados são altamente dependentes da escolha dos registros neste conjunto de dados para a formação dos conjuntos de treinamento, validação e teste da RNA. Isto se deve, uma vez que, pode haver registros importantes que ficaram definidos na parte de validação e que seriam representativos para a generalização da RNA se estivessem na parte de treinamento contribuindo assim para a RNA. Na Figura 2 apresenta-se o esquema da divisão dos dados por meio desse método.

Figura 2 – Divisão dos dados para treinamento, validação e teste



Fonte: O autor (2019)

Para se evitar tais desvantagens, parte-se para a validação cruzada em RNA. A validação cruzada é uma técnica para avaliar a capacidade de generalização de uma RNA, ou seja, a capacidade de responder corretamente a dados que não foram utilizados no processo de treinamento. Essa técnica é empregada em problemas onde o objetivo da RNA é a classificação. Na validação cruzada, o conjunto de dados é dividido de forma aleatória em partes, estabelecendo-se partes para a realização do treinamento e partes para a realização da validação (SILVA, ALMEIDA e YAMAKAMI, 2012, KOHAVI et al., 1995).

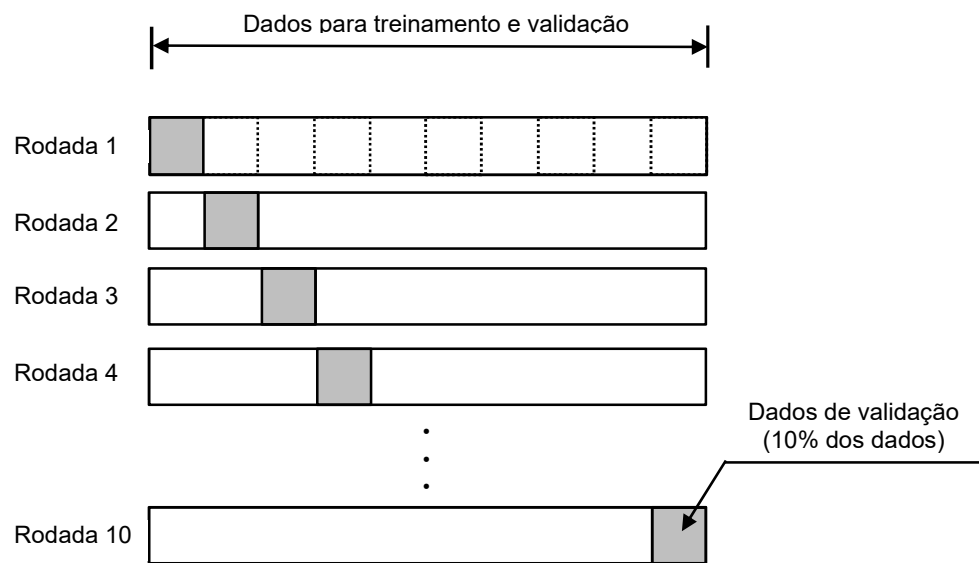
A validação cruzada pode ser utilizada como sendo um critério de parada para a RNA, uma vez que, o processo de treinamento é interrompido quando o erro da parte de validação começa a aumentar. Este critério de parada é realizado por meio do erro quadrático, sendo este o mais comum. Entretanto, este critério não garante uma interrupção do treinamento, sendo assim, necessário definir um limite de *épocas* ou um limite para o valor do erro ou do vetor gradiente (DUDA, HART e STORK, 2012). Convencionou-se chamar de uma época para a RNA a apresentação de toda parte de padrões do treinamento.

A utilização da validação cruzada também evita a ocorrência de *overfitting*, estimando o erro de generalização da RNA durante o processo de treinamento. O *overfitting* ocorre quando as RNA memorizam os padrões de treinamento, deixando de extrair características do modelo para se tornarem eficientes na generalização dos padrões, ou seja, para acertar padrões que as RNA não observaram durante o processo de treinamento (DUDA, HART e STORK, 2012; SILVA, ALMEIDA e YAMAKAMI, 2012).

No método de validação cruzada, uma das técnicas utilizadas é o *k-fold cross validation*, onde um conjunto de dados (*k-fold*), é dividido em várias partes de forma aleatória, sendo estas partes divididas em partes de treinamento e partes de validação

(f). Um classificador é aprendido usando $f-1$ partes e um valor de erro é calculado testando o classificador nas partes restantes. Este processo é realizado k vezes alternando de forma circular as partes estabelecidas para validação. Na Figura 3 apresenta-se o esquema da divisão dos dados por meio do método *k-fold cross validation*.

Figura 3 – Divisão dos dados por k-fold cross validation



Fonte: O autor (2019)

A escolha da divisão das partes (f) tendem a não perturbar os dados e resultam em maior variância. A divisão de partes de dados em $f=2$, não é praticada, uma vez que propicia em partes para o treinamento e partes para a validação muito pequenas em relação ao conjunto de dados para a realização do treinamento, impactando no resultado da RNA. A divisão do conjunto de dados de treinamento em partes maiores, como os de valores de $f=5$ ou $f=10$ são os que resultam em melhores valores, e são comumente utilizados, sendo que a utilização de $f=10$ é a divisão, em geral aplicada em conjunto de dados para treinamento, independentemente da quantidade de registros no conjunto banco de dados (MCLACHLAN, DO e AMBROISE, 2005; KIM, 2009; FUSHIKI, 2009).

A utilização dessa metodologia *k-fold* depende de dois fatores: do conjunto de dados para a realização do treinamento, validação e teste das RNA, e da quantidade de divisão de partes deste conjunto de dados utilizados para a realização do treinamento, validação e teste das RNA. É importante ressaltar que a quantidade de registros,

disponível no conjunto de dados de treinamento e validação seja múltipla da divisão em partes que se deseja realizar, para que se possa obter resultados satisfatórios nas RNA (RODRIGUEZ, PEREZ e LOZANO, 2010; GRIMM, MAZZA e DAVOUDZADEH, 2016).

Na validação cruzada utilizando-se o valor de $f=10$, os registros no conjunto de dados utilizados para treinamento e validações são seccionados aleatoriamente em 10 partes iguais. Destas partes, uma única parte é retida como validação e as outras 9 partes restantes são utilizadas como treinamento. O processo é repetido 10 vezes, com cada uma das 10 partes, alternando-se as partes em treinamento e validação. A metodologia do *k-fold* consegue predições com 90% dos dados, tornando as RNA mais suscetíveis de serem generalizadas para conjunto de dados completos (SILVA, ALMEIDA e YAMAKAMI, 2012).

O processo de treinamento das RNA, utilizando o *k-fold* ocorrem da seguinte forma: depois de ter havido o seccionamento do conjunto de dados em partes de treinamento e validação, como citado anteriormente. A RNA utiliza uma das partes a qual foi definida como parte de treinamento e a utilizam no processo de treinamento da RNA obtendo-se os pesos de toda a rede, da camada de entrada para a camada intermediária, da camada intermediária para a camada de saída, e gerando resultados para a RNA para esta parcela de dados utilizada. A parte seccionada e definida como parte de validação, é utilizada para avaliar o desempenho da RNA, ou seja, esta parcela dos dados é comparada com os resultados obtidos durante a fase de treinamento da RNA com os resultados esperados, os quais estão na parte de validação. Esta comparação entre os resultados obtidos e esperados da parcela (f) define a precisão, ou acurácia da RNA (WANG, 2015).

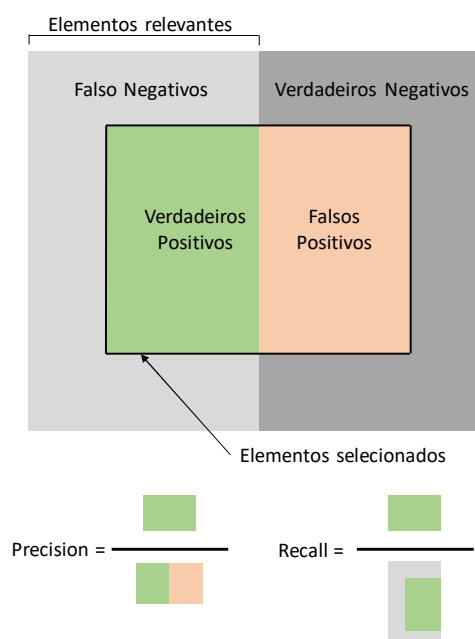
Este processo é repetido para cada uma das partes de treinamento e validação, sendo assim para um $f=10$ serão processados 10 treinamentos e validações, sendo que nesses 10 treinamentos e validações, cada uma destas partes de treinamento e validação obterá um resultado, ou seja, ter-se-á 10 resultados para a RNA.

A acurácia para uma RNA pode ser definida como a taxa de acerto global, ou seja, a proporção de predições corretas em relação a parte de treinamento e validação, sendo essa a medida de desempenho mais comum em RNA. Sendo que os valores para a

determinação das acurácias são obtidos por meio da matriz de confusão (SILVA, ALMEIDA e YAMAKAMI, 2012).

O *F1-score* pode ser interpretado como uma média harmônica dos valores *precision* e *recall*, sendo que o *F1-score* alcança seu melhor valor em 1 e o seu pior valor em 0. O *F1-score* é uma medida que considera tanto os valores da *precision* como o *recall* para calcular a pontuação. *Precision* é o número de resultados positivos corretos dividido pelo número de todos os resultados positivos devolvidos pelo classificador, e *recall* é o número de resultados positivos corretos dividido pelo o número de todos os resultados relevantes, ou seja, todos os resultados que deveriam ter sido identificados como positivos. No caso de multiclases, esta pode ser entendida como a média da pontuação *F1-score* de cada uma das classes com ponderação, dependendo dos parâmetros (SCIKIT-LEARN, 2011). A Figura 4 ilustra a obtenção dos valores da *precision* e *recall* para obtenção do valor de *F1-score*. As equações (1) e (2) são utilizadas para calcular, os valores da *precision* e *recall*, onde: v_p = verdadeiros positivos, ou seja, referentes ao número de registros corretamente classificados, f_p = falsos positivos, referentes ao número de registros que foram incorretamente classificados, v_n = verdadeiros negativos, referentes ao número de registros classificados corretamente e f_n = falsos negativos, referentes ao número de registros classificados incorretamente.

Figura 4 – Obtenção dos valores da precision e recall



Fonte: Wikipedia (2019, adaptado)

$$Precision = \frac{v_p}{v_p + f_p} \quad (1)$$

$$Recall = \frac{v_p}{v_p + f_n} \quad (2)$$

Com isto, as RNA obtêm a quantidade de resultados de acurácia e de *F1-score*, proporcionais as quantidades de partes divididas pelo *k-fold*, ou seja, ter-se-á um resultado de acurácia e um resultado de *F1-score* para cada uma das partes divididas pelo *k-fold*.

Pode-se afirmar que a parte de treinamento e validação que apresentou um resultado de acurácia média, assim como um resultado de *F1-score* médio dentre as demais partes de treinamento e validação, divididas pelo *k-fold*, esta é a parte que apresenta os melhores pesos para generalização. Assim, estes pesos que deverão ser armazenados para serem utilizados no processamento com outros conjuntos de dados para obtenção de novos resultados para a RNA. A Tabela 1 apresenta a distribuição da matriz de confusão em classes positivas e classes negativas. A acurácia (*Acc*) e o *F1-score* são calculados, respectivamente, pelas equações (3) e (4).

Tabela 1 – Matriz de confusão

	Classe positiva	Classe negativa
Estimação positiva	v_p	f_p
Estimação negativa	f_n	v_n

Fonte: O autor, (2019)

$$Acc = \frac{v_p + v_n}{total\ da\ amostra} \quad (3)$$

$$F1-score = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

Com base no exposto, pretende-se investigar os algoritmos de aprendizado de máquina, no caso, as RNA como ferramentas para serem utilizadas nas estimações da vida útil remanescente dos trilhos ferroviários. No subcapítulo 2.3 serão apresentadas as RNA que foram aplicadas em estimações em componentes e estruturas ferroviárias.

2.2 K-VIZINHOS MAIS PRÓXIMOS (K-NN)

O algoritmo k-Vizinhos mais próximos (k-NN) é um método não paramétrico usado para a classificação e para a regressão, devido às suas aplicações nas áreas de reconhecimento de padrões (YU, 2015). O k-NN é um algoritmo popular e simples devido à sua implementação, e propicia a resultados significativos no campo da classificação de padrões. Este método produz resultados competitivos em certos domínios, quando combinados com o conhecimento prévio de especialistas. Entretanto, este método apresenta uma fragilidade quando se atribui um valor fixo de *k*-vizinhos mais próximos para todos os conjuntos de treinamento e na medida de similaridade entre dois dados (ZHANG et al., 2017).

A regra do k-NN é classificar cada exemplo não rotulado pelo rótulo majoritário entre seus *k*-vizinhos mais próximos no conjunto de treinamento, ou seja, pela regra da maioria. Assim, seu desempenho depende crucialmente da métrica de distância utilizada para identificar os vizinhos mais próximos (YU et al., 2015; ZHANG et al., 2017).

A maioria dos classificadores do k-NN usam distâncias euclidianas simples para medir as diferenças entre as amostras, sendo que vários métodos foram propostos para abordar as questões da métrica de distância, como a distância de Mahalanobis, e distância de Minkowsky e suas variantes. Assim, conclui-se que para diferentes aplicações precisam-se de diferentes medidas de distância, sendo assim, a métrica de distância para a utilização do k-NN deve ser adaptada ao problema a ser resolvido (YU et al., 2015; ZHANG et al., 2017).

Os métodos de classificação k-NN que definem um valor fixo de k-vizinho mais próximo para todas as amostras de teste ou a realização de validação cruzada para estimar o valor k-vizinho mais próximo para cada ponto da amostra de teste, leva a uma baixa taxa de estimação em aplicações de classificação real, devido ao fato de que esses métodos não levam em consideração a distribuição das amostras (ZHANG et al., 2017).

Assim, o valor de k-vizinho mais próximo para o classificador K-NN deve ser adequado para os diferentes tipos de conjuntos de dados, sendo que seu desempenho será afetado em alguns conjuntos de dados, principalmente com o conjunto de dados com pequena quantidade de amostras para o treinamento e conjuntos de dados com desequilíbrio entre as classes de treinamento (YU et al., 2015). Na Figura 5 é apresentado o pseudocódigo do k-NN.

Figura 5 – Pseudocódigo do k-NN

1	Input: k , conjunto de dados
2	Para cada nova amostra
3	Calcular a distancia para todas as outras amostras
4	Determinar o conjunto das K 's distancias mais próximas
5	Rotula os mais representantes no conjunto dos K 's vizinhos mais próximos
6	Escolhe o rotulo com mais representantes
7	FimPara
8	Retornar: Conjunto de rotulos de classificação

Fonte: Fukunaga e Narendra (1975, adaptado)

2.3 REDES NEURAIAS ARTIFICIAIS APLICADAS EM FERROVIAS

A pesquisa para a realização deste trabalho limitou-se em publicações de artigos técnicos científicos para o período de 2007 até 2018, e em artigos publicados em congressos, livros e demais publicações científicas. O objetivo desse subcapítulo será apresentar de forma resumida os trabalhos que abordaram a utilização de RNA em ferrovias e que se aproximam do tema de pesquisa proposto nesse trabalho de dissertação.

Verificou-se a utilização da RNA com o objetivo de se modelar o comportamento mecânico do lastro ferroviário sob cargas estáticas. Os dados utilizados para o

desenvolvimento da RNA foram obtidos da literatura e, também, compreendeu uma série de 29 testes de compressão triaxial drenada. Foram estudados três tipos de materiais comumente usados para a execução de lastros em ferrovias, sendo eles: o basalto, a dolomita e o granito. As estimações do comportamento mecânico dos lastros ferroviários obtidas por meio das RNA foram comparadas com os resultados de testes experimentais e com os resultados obtidos por meio do modelo constitutivo do solo. Os resultados indicaram que a RNA proposta foi capaz de estimar o comportamento de tensão-deformação e as mudanças de volume do lastro ferroviário (SHAHIN e INDRARATNA, 2006).

Aplicou-se a RNA com o objetivo de se modelar a deterioração da geometria e da plataforma de linha férrea por meio de investigações de campo. Os dados foram obtidos em um período de 2 anos e em aproximadamente 180 km de linha férrea, sendo que a linha férrea foi dividida em segmentos para o desenvolvimento do estudo. Para cada segmento de linha férrea, foram levantados os dados referentes a estrutura da plataforma, as características do tráfego ferroviário, o perfil do trilho, os fatores ambientais, a geometria da linha e os dados de manutenção e da renovação da linha férrea. As RNA foram desenvolvidas com os principais parâmetros da geometria da estrutura da linha férrea e que produziam correlações significativas entre as variáveis. Os resultados obtidos por meio das RNA mostraram que estas, podem ser utilizadas como um método alternativo para se estimar a deterioração da plataforma e estrutura das linhas férreas (GULER, 2013).

As RNA foram utilizadas nos sistemas ferroviários para o problema de estimação das avarias do sistema ferroviário, priorizando a segurança ferroviária, uma vez que as formas tradicionais de se prever essa segurança, são muito dispendiosas em termos de tempo, tornando-as em determinadas circunstâncias ineficientes. Portanto, utilizaram-se as RNA para lidar com o problema de segurança. Os dados de entrada para as RNA foram as irregularidades no posicionamento dos trilhos ferroviários. Diferentes configurações das RNA foram desenvolvidas para a verificação de seus desempenhos, tanto em termos de erro quadrático médio, como de coeficiente de correlação. Essas verificações foram avaliadas a fim de se buscar a melhor configuração das RNA para a estimação da segurança ferroviária. Os resultados apontaram que as RNA conseguiram de forma satisfatória, realizar a estimação da

segurança ferroviária por meio das irregularidades no posicionamento dos trilhos ferroviários (NEFTI e OUSSALAH, 2004).

Desenvolveram-se RNA para a detecção e identificação de falhas em trilhos ferroviários, utilizando os dados ultrassônicos processados, obtidos por meio de leituras realizadas em um carro de inspeção para trilhos ferroviários. Os dados foram levantados de uma ferrovia de teste do serviço ferroviário de Sperry na cidade de Danbury, localizada no estado americano de Connecticut, no Condado de Fairfield. A ferrovia de teste apresentava 9 tipos de defeitos estabelecidos e conhecidos. Esses defeitos estabelecidos e conhecidos foram determinados por meio do manual de defeitos elaborado para a ferrovia de teste do serviço ferroviário Sperry. Elaborou-se 2 configurações de RNA para a verificação de seus desempenhos, sendo que os números de neurônios nas camadas intermediárias foram determinados durante a fase de treinamento e de forma iterativa. As configurações das RNA desenvolvidas conseguiram, por meio das leituras dos dados ultrassônicos, obtidas por meio de carro de inspeção, apresentar índices maiores que 80% na identificação dos defeitos nos trilhos ferroviários para os defeitos estabelecidos e conhecidos para a ferrovia de teste do serviço ferroviário Sperry. Com esse resultado, mostrou-se que as RNA desenvolvidas obtiveram sucessos na detecção e identificação de defeitos em trilhos ferroviários na ferrovia de teste do serviço ferroviário Sperry (CHOU e CLARK, 1999).

Foram desenvolvidas RNA para a avaliação das condições da qualidade das linhas ferroviárias, com o objetivo de analisar os trilhos, os dormentes, os fixadores e o lastro ferroviário. Os dados foram levantados em inspeções realizadas por intermédio do carro de inspeção em 2 linhas ferroviárias paralelas no centro do Irã, sendo uma linha ferroviária considerada como em boas condições de tráfego e outra linha ferroviária considerada como em condições precárias de tráfego. Definiu-se como resultados de saída para as RNA, 4 tipos de classes para as diferentes avaliações de qualidade das condições de tráfego da linha ferroviária: Linhas de classe A e B onde apresentam boas condições de tráfego, linhas de classe C com média condições de tráfego e classe D, linhas com condições precárias de tráfego. As RNA desenvolvidas obtiveram resultados satisfatórios na estimação das classificações das condições da qualidade da linha ferroviária (SADEGHI e ASKARINEJAD, 2012).

3 DESCRIÇÃO E CONCEITUAÇÃO DO PROBLEMA

Os trilhos possuem função primordial na ferrovia, pois suportam e guiam as rodas dos veículos ferroviários e devem apresentar continuidade ao longo da ferrovia. Essa continuidade nos trilhos ferroviários é que se constitui no caminho de rolamento ferroviário, por onde as rodas dos veículos ferroviários trafegam. O objetivo desse caminho de rolamento é guiar as composições ferroviárias.

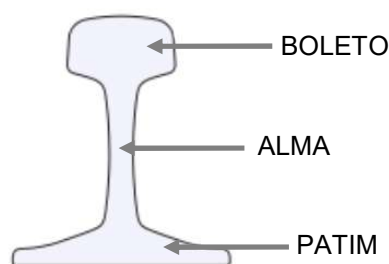
Os perfis dos trilhos têm sido constante objeto de pesquisa e melhorias desde o surgimento das estradas de ferro (ELKHOURY et al., 2018; PROFILLIDIS, 2006; HAY 1982; AUDLEY e ANDREWS, 2013).

No início do século XVIII, os trilhos utilizados nos sistemas ferroviários eram de ferro fundido, passando no começo do século XIX a serem utilizados os trilhos em aço em substituição aos trilhos em ferro fundido (CANNON et al., 2003). O aço é o material que oferece as melhores condições e vantagens para a fabricação dos trilhos, uma vez que, os trilhos devem possuir dureza, tenacidade, elasticidade e resistência à flexão para que possam suportar as cargas a que são submetidos pelos veículos ferroviários (HAY, 1982, PROFILLIDIS, 2006).

Os trilhos ferroviários fabricados em aço, apresentam as seguintes composições químicas: ferro; carbono; manganês; silício; e fósforo, sendo que o ferro está presente em 98% da composição (PROFILLIDIS, 2006; ROSA e RIBEIRO, 2017).

O perfil de trilho atualmente usado é do tipo Vignole que é constituído por três partes: 1) o boleto; 2) a alma; e 3) o patim (PROFILLIDIS, 2006; ROSA e RIBEIRO 2017; HAY, 1982). O boleto é a seção de maior massa, onde existe o contato direto com as rodas dos veículos ferroviários com os trilhos e, devido a esse contato, sofre o desgaste superficial, perdendo massa. A alma é a seção de ligação entre o boleto e o patim, não devendo ser excessivamente alta para que não comprometa a estabilidade do trilho. O patim é a parte mais larga do trilho onde é realizada a fixação do trilho com os dormentes (ELKHOURY et al., 2018; ROSA e RIBEIRO, 2017; PROFILLIDIS, 2006, HAY, 1982). A Figura 6 apresenta as partes de um trilho ferroviário; 1) o boleto; 2) a alma; e 3) o patim.

Figura 6 - Partes do trilho ferroviário tipo Vignole



Fonte: O autor (2019)

O modo ferroviário deve ser entendido como de suma importância no transporte de cargas para longas distâncias terrestres, sendo este, o sistema de transporte mais eficiente. Para tanto, este sistema de transporte deve possuir altos índices de confiabilidade. A confiabilidade dos trilhos ferroviários tem sido uma preocupação desde os primórdios e continua para os dias atuais para os operadores deste sistema de transporte. Para isto, os operadores deste sistema de transporte buscam mitigar e evitar falhas nos trilhos, uma vez que, a ocorrência destas falhas propicia a perda do material rodante, da carga transportada e de vidas humanas (GULER, 2013; SCHAFER e CHRISTOPHER, 2008; ZUMPANO e MEO, 2006).

Vários são os fatores que podem causar os desgastes excessivos nos boletos dos trilhos ferroviários, propiciando a ocorrência destas falhas, como o aumento de cargas por eixo, aumento do tráfego, aumento das velocidades das composições, curvatura dos trilhos, tipo de trilhos, perfil do trilho, elevações do trilho, superelevação da linha, soldagem no trilho, lubrificação no trilho, condições das rodas dos veículos ferroviários, abertura de bitolas, temperatura dentre outros (ELKHOURY et al. 2018; SCHAFER e CHRISTOPHER, 2008; ZUMPANO e MEO, 2006; AUDLEY e ANDREWS, 2013).

Logo, as falhas apresentadas nos trilhos ferroviários, devido a esses desgastes excessivos, não sendo estes trilhos substituídos no momento adequado, podem gerar a descarrilamentos, podendo assim, ocasionar a situações catastróficas para a sociedade e para o meio ambiente. Por isso, manter o alto índice de confiabilidade nos trilhos e definir o momento de sua substituição na via férrea, deve ser objeto de busca constante pelos operadores e engenheiros ferroviários (ELKHOURY et al., 2018; SCHAFER e CHRISTOPHER, 2008; ZUMPANO e MEO, 2006).

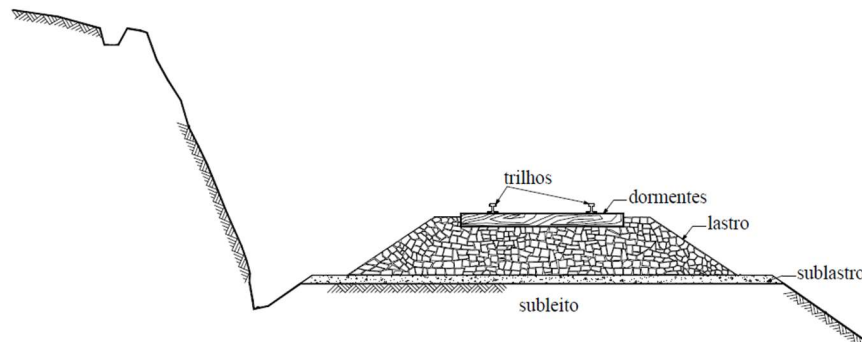
Para que se possa garantir essa confiabilidade nos trilhos ferroviários torna-se necessária a realização de inspeções e leituras nesses elementos, com o propósito de se obter e determinar o grau de desgaste na geometria nos trilhos ferroviários, devido a esta contínua passagem das rodas dos veículos ferroviários (PENG e JONES, 2013).

As ferrovias dispostas ao redor do mundo com grande demanda de carga e tráfego, possuem veículos ferroviários que realizam essas inspeções e leituras nos trilhos ferroviários. Esses veículos ferroviários realizam as leituras dos desgastes geométricos nos trilhos de maneira automatizada. Os veículos ferroviários utilizados para a realização dessas atividades são denominados de veículos de inspeção ferroviária ou carro controle. Cabe ressaltar, que estas inspeções e leituras na geometria dos trilhos ferroviários podem ser realizadas de maneira manual.

Essas inspeções e leituras realizadas nos trilhos ferroviários tem como finalidade formar e retroalimentar o conjunto de dados, com o intuito de se analisar os avanços dos desgastes dos trilhos ao longo do tempo. Com este conjunto de dados, e utilizando diferentes métodos de análises, busca-se a estimativa da vida útil remanescente dos trilhos, para se obter o momento de sua substituição na via férrea. Esse conjunto de dados, juntamente com os diferentes métodos de análises, auxiliam o planejador da programação da manutenção ferroviária, na tomada de decisão na substituição dos trilhos ferroviários. Assim, a atividade do planejador da programação da manutenção ferroviária é buscar substituir os trilhos desgastados na via férrea no tempo apropriado, e evitar que sejam realizadas substituições de trilhos desnecessariamente, mas, mantendo a confiabilidade e segurança deste sistema de transporte (ELKHOURY et al., 2018; FALAMARZI et al., 2018; SANTOS, TEIXEIRA e ANTUNES, 2015; LYNGBY, 2009; ZHANG, MURRAY e FERREIRA, 2000).

Essa busca constante pelos operadores ferroviários para estabelecer os momentos para a realização das substituições dos trilhos ferroviários, além de manter a confiabilidade e segurança deste sistema de transporte, deve-se também ao fato de que os trilhos são os elementos de maior custo na manutenção, dentre os elementos da superestrutura ferroviária. Para esta pesquisa a superestrutura ferroviária é interpretada, como sendo os trilhos, os dormentes, o lastro, o sublastro e os aparelhos de mudança de via. A Figura 7 apresenta a seção de uma superestrutura ferroviária.

Figura 7 – Seção da superestrutura ferroviária



Fonte: Rosa e Ribeiro (2017, adaptado)

Como citado anteriormente, a ocorrência destas falhas nos trilhos ferroviários, podem ocasionar situações catastróficas, incentivando assim, que os operadores ferroviários busquem maneiras de estimação da vida útil remanescente dos trilhos, de forma a evitar e mitigar estas ocorrências. Desse modo, justificam-se as pesquisas para se aprimorar e desenvolver ferramentas e técnicas de análises de maneira a auxiliar aos operadores ferroviários na tomada de decisões para estabelecer a realização das substituições dos trilhos na via férrea (HAY, 1982; ANDRADE e TEIXEIRA, 2012; FALAMARZI et al., 2018; AUDLEY e ANDREWS, 2013; SOLEIMANMEIGOUNI, AHMADI e KUMAR, 2016; SADEGHI e ASKARINEJAD, 2008).

Portanto, deve-se buscar uma ferramenta de estimação que consiga estabelecer o momento para a realização das substituições destes trilhos ferroviários na via férrea, ou seja, a estimação da vida útil remanescente desses trilhos. Esta ferramenta terá como objetivo auxiliar ao planejador da programação da manutenção ferroviária, a elaborar um planejamento estruturado, para a realização das substituições dos trilhos, assim como, na alocação dos recursos financeiros, de equipamentos e de pessoal necessários para a realização dessa atividade ao longo da via férrea.

As condições geométricas dos trilhos ferroviários, ou seja, os desgastes ocorridos nos trilhos são utilizados como sendo parâmetros para se diagnosticar e estabelecer o nível de segurança da ferrovia, bem como, apontar o momento da necessidade da substituição do trilho na via férrea (LIU, 2017; LI et al., 2015; LI e REN, 2012; ROSA e RIBEIRO, 2017; AREMA, 2013).

Entretanto, a dificuldade das equipes de planejamento da programação da manutenção ferroviária e das equipes de inspeção dos trilhos ferroviários, é estabelecer com base somente nas leituras dos desgastes da geometria dos trilhos ferroviários, e do peso médio trafegado, uma relação que se consiga estimar qual será a vida útil remanescente desses trilhos, com as práticas utilizadas e aplicadas atualmente pelas ferrovias.

Isto se deve, face a grande complexidade, diversidade e a variabilidade de fatores que ocasionam aos desgastes geométricos nos trilhos ferroviários. Assim, todos os métodos atualmente aplicados pelos operadores ferroviários não apresentam resultados satisfatórios na estimação da vida útil remanescente dos trilhos ferroviários. Sendo que muitos destes métodos foram desenvolvidos para ferrovias específicas, apresentando características específicas, condições de tráfegos específicos e muitas vezes diferentes da ferrovia onde está sendo aplicada esta metodologia, ou seja, não condizendo com as situações e as características desta ferrovia.

3.1 O PROBLEMA ANALISADO

As ferrovias, de forma geral, planejam a manutenção dos trilhos, com sua eventual substituição, em períodos do ano que podem ser medidos em: mês, trimestre, semestre ou ano. Assim, para fazer esse planejamento, elas necessitam calcular a vida útil remanescente dos trilhos em função destas quatro categorias de períodos. Vale mencionar que nas ferrovias, via de regra, as categorias de períodos mais utilizadas são o semestre e o ano.

Em geral a vida útil remanescente dos trilhos é calculada por um único número contínuo que representa quantos dias o trilho ainda poderá ser utilizado com segurança na ferrovia. Entretanto, para esta dissertação, a vida útil remanescente foi representada por meio de categorias de períodos de vida útil remanescente, refletindo como na prática as ferrovias planejam a vida útil remanescente dos trilhos.

Foram definidas quatro categorias de períodos de vida útil remanescente: 1) mês; 2) trimestre; 3) semestre; e 4) ano. Essas categorias de vida útil remanescentes do trilho

configurações dos algoritmos de aprendizado de máquina que apresentaram o melhor resultado e desempenho para cada uma das categorias analisadas.

3.2 DESCRIÇÃO DA FERROVIA ESTUDADA

A Estrada de Ferro Vitória a Minas (EFVM) interliga a Região Metropolitana de Vitória, no estado do Espírito Santo, a Belo Horizonte, capital do estado de Minas Gerais, em seus 905 km de percurso, sendo essa ferrovia de concessão da empresa VALE S.A. e realiza o transporte de cargas e passageiros.

A EFVM, teve seu início em fevereiro de 1902 quando o Governo Federal concedeu, por meio de um decreto lei, a criação da Companhia Estrada de Ferro Vitória a Minas, iniciando dessa forma a construção da ferrovia. Na década de 40 realizou-se a remodelação do trecho entre Vitória e Colatina/ES, sendo que na década de 1950 que a ferrovia se caracterizou pela introdução de novas melhorias. A duplicação da linha ferroviária aconteceu entre os anos de 1971 e 1977, tornando-se um marco na evolução da ferrovia EFVM.

Dos 905 km de extensão total, 601 km de extensão da EFVM estão em linha dupla. A realização das operações de transporte, ocorrem nas linhas duplas denominadas, linha 1 e linha 2, sendo que a linha 1 é a mais antiga. A Figura 9 apresenta o mapa geral de atuação da EFVM. Em média, são transportados pela ferrovia EFVM cerca de 145 milhões de toneladas de produtos por ano, por meio das linhas duplas, linha 1 e linha 2. O principal produto transportado na EFVM é o minério ferro, que corresponde cerca de 122 milhões de toneladas por ano, seguido pelos demais produtos que representam cerca de 23 milhões de toneladas por ano, como o carvão mineral, aço, ferro gusa, farelo de soja, derivados do petróleo, dentre outros.

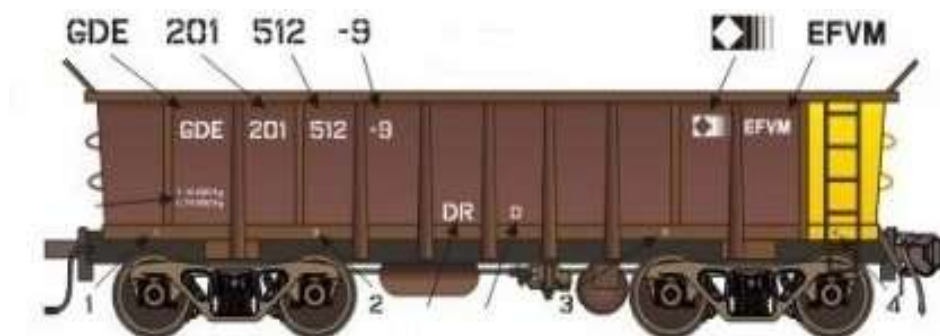
Figura 9 – Mapa geral do traçado da EFVM



Fonte: Programa de Parcerias de Investimento (2019)

A ferrovia EFVM foi concebida em bitola métrica e apresenta raios mínimos em seu traçado de 195 m, e rampas compensadas máximas de 1,5%, o que permite uma velocidade máxima de 65 km/h. O vagão utilizado para o transporte de minério de ferro utilizado na EFVM é o vagão do tipo GDE. O vagão do tipo GDE tem capacidade de carga de 110 toneladas brutas, o que corresponde a 27,5 toneladas brutas por eixo. A Figura 10 apresenta o vagão do tipo GDE utilizado na EFVM para o transporte de minério.

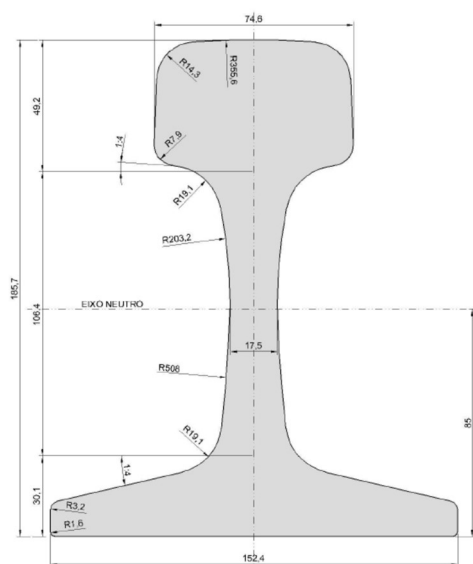
Figura 10 – Vagão de minério GDE



Fonte: Álbum de Vagões da EFVM (2005, p.21)

Os trilhos utilizados na EFVM são do tipo TR-68 definidos pela Norma NBR-7590 (ABNT, 2012) o que equivale ao trilho 136RE (AREMA, 2013). A EFVM possui aproximadamente 120 mil toneladas de trilhos assentados em sua via férrea. A Figura 11 apresenta as dimensões do trilho ferroviário TR-68.

Figura 11 – Perfil do trilho TR-68



Fonte: DNIT (2015, p. 14)

Para a realização das inspeções e leituras das geometrias dos trilhos ferroviários, visando as verificações dos desgastes nos trilhos ferroviários provocados devido à contínua passagem das rodas dos veículos ferroviários, a EFVM utiliza um veículo de inspeção ferroviária denominado de Carro Controle EM-80, que realiza estas atividades automaticamente. Este veículo de inspeção ferroviária denominado de Carro Controle EM-80 será explicado no subcapítulo a seguir.

3.3 VEÍCULO DE INSPEÇÃO FERROVIÁRIA - CARRO CONTROLE

O veículo de inspeção ferroviária utilizado na EFVM e denominado de Carro Controle EM-80, foi fabricado pela empresa Plasser & Theurer no ano de 2008, de maneira a atender as características e as necessidades operacionais da EFVM para a realização das inspeções e leituras ao longo da linha férrea. A Figura 12 apresenta o Carro Controle fabricado pela Plasser & Theurer para a EFVM.

Figura 12 – Carro Controle EM-80 - EFVM

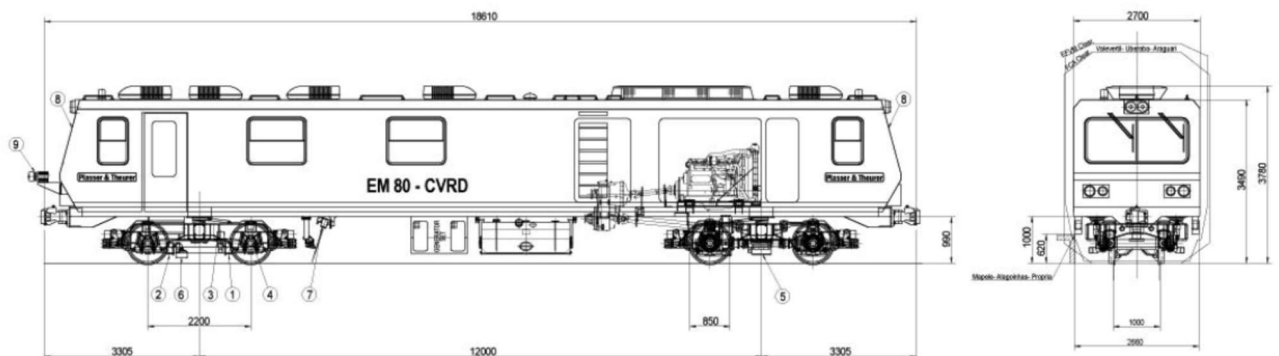


Fonte: Plasser & Theurer (2008, p. 10)

O Carro Controle EM-80 é utilizado para a realização de diversas atividades ao longo da linha férrea da EFVM, como na leitura da geometria dos trilhos ferroviários, leituras das superelevações da geometria da linha, levantamentos e leituras de desnivelamentos da linha, leituras de aberturas de bitolas, leituras de gabaritos verticais e horizontais em túneis, pontes e viadutos, dentre outros.

Este veículo de inspeção ferroviária apresenta as seguintes dimensões: comprimento de 18,61 m, largura de 2,70 m e altura de 3,78 m acima do topo do boleto do trilho. A Figura 13 apresenta as dimensões do Carro Controle EM-80 produzido para a EFVM.

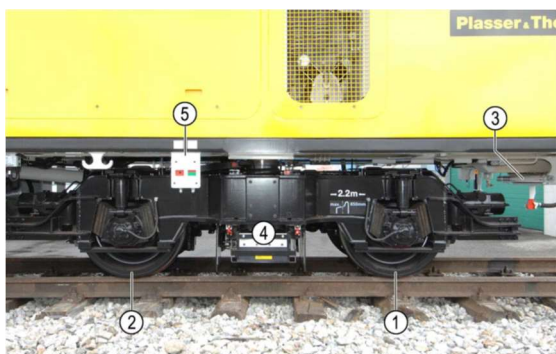
Figura 13 – Dimensões do Carro Controle EM-80 - EFVM



Fonte: Plasser & Theurer (2008, p. 121)

O Carro Controle EM-80 realiza as leituras da geometria dos trilhos ferroviários por meio de um sistema de equipamentos a laser e ópticos localizados em seus truques frontais, sendo estes posicionados no lado esquerdo e direito do veículo ferroviário, entre os eixos um e dois. A Figura 14 apresenta a localização dos dispositivos de leituras da geometria dos trilhos ferroviários.

Figura 14 – Localização do sistema de leitura do Carro Controle EM-80 - EFVM



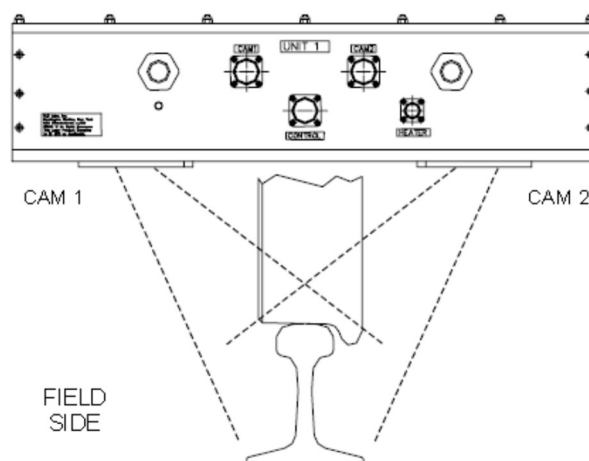
- 1 = Eixo1
- 2 = Eixo2
- 3 = Luz de trabalho do truque frontal Eixo1
- 4 = Sistema KLD: Caixa do laser, lado direito
- 5 = Indicador de status de freio de estacionamento
VERDE: Freio de estacionamento liberado
VERMELHO: Freio de estacionamento acionado

Fonte: Plasser & Theurer (2008, p. 25)

Este veículo de inspeção ferroviária dispõe de um sistema de leitura que consiste em um processamento óptico e a laser. O sistema de processamento óptico de imagens denominado ORIAN VII é usado para fornecer os dados dimensionais em tempo real da geometria dos trilhos ferroviários com alta resolução. Sendo esses dados processados em tempo real para avaliar e monitorar as dimensões da geometria dos trilhos ferroviários.

O sistema óptico utilizado é o sistema de visão de máquinas de alta velocidade. Este sistema óptico que realiza as leituras da geometria do trilho utilizando uma fonte de luz laser para iluminar o boleto do trilho e uma câmera de vídeo CCD de alta resolução que captura estas imagens do trilho. Cada cabeça leitora gera um plano momentâneo de luz laser que ilumina a seção transversal de cada um dos trilhos. O laser utilizado nas cabeças leitoras para gerar o plano momentâneo é registrado como dispositivos de classe 3B. As câmeras de vídeo CCD de alta resolução capturam essas imagens dos contornos dos trilhos iluminados. Estas imagens são transformadas em sinais eletrônicos, e enviados para o computador de bordo do Carro Controle EM-80, onde são processadas para serem obtidos os valores das leituras dos trilhos ferroviários. A Figura 15 apresenta o sistema de leitura laser utilizado pelo Carro Controle EM-80.

Figura 15 – Sistema de leitura laser do Carro Controle EM-80 - EFVM



Fonte: Plasser & Theurer (2008, p. 1006)

O Carro Controle EM-80, por meio de seus leitores ópticos, consegue identificar automaticamente o tipo de trilho a qual está se realizando as leituras, e dispõe de uma biblioteca com os padrões e as dimensões geométricas dos tipos de trilhos ferroviários normatizados e comumente utilizados em vias férreas. De qualquer forma, pode-se inserir novas geometrias de trilhos ferroviários com dimensões diferentes da apresentada na biblioteca disponibilizada pelo equipamento.

As leituras na geometria dos trilhos ferroviários são realizadas pelo Carro Controle a uma distância de 25 cm entre leituras ao longo do trilho ferroviário. Estas leituras são realizadas simultaneamente em ambos os trilhos, durante o tráfego do veículo de inspeção ferroviária sobre a via férrea.

O Carro Controle EM-80 dispõe de dois sistemas de controle para o posicionamento geográfico na via férrea, sendo um, o sistema de localização realizado pelo Sistema de Posicionamento Global (GPS) e o outro, o sistema de correção por meio de marcos geográfico disponibilizados ao longo da linha férrea da EFVM, denominados de estaqueamentos da linha.

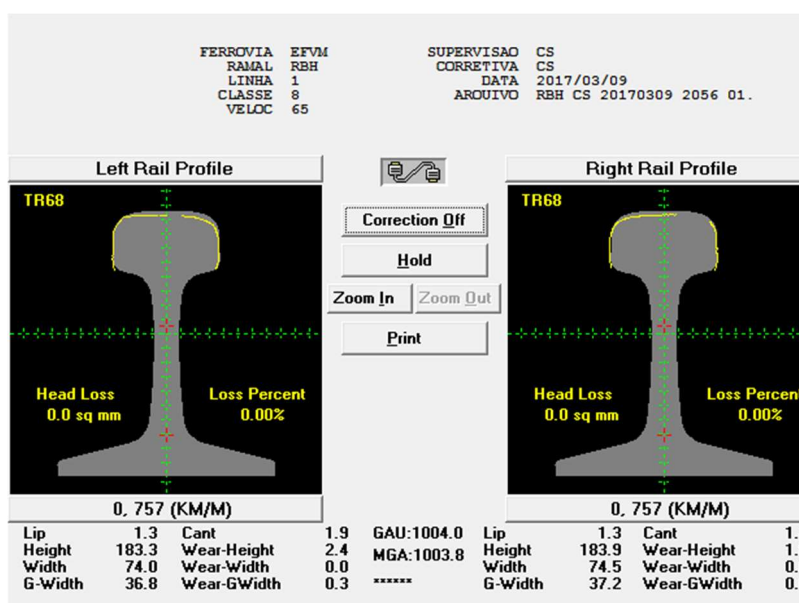
Este sistema de posicionamento por GPS e marcos geográfico estabelecem e confirmam que as leituras da geometria dos trilhos sejam realizadas continuamente nos mesmos pontos, de forma que se consiga realizar estas leituras continuamente nestes mesmos pontos. Com este posicionamento geográfico pode-se garantir que os registros apresentados nos conjuntos de dados apresentam continuidades de leituras nos mesmos pontos, sendo assim, possível avaliar e analisar a ocorrência de avanços

dos possíveis desgastes nos trilhos ferroviários para cada um destes pontos de leituras.

O operador do Carro Controle EM-80 por meio de um monitor instalado na cabine de controle, que se encontra interligado com o computador de bordo do veículo, consegue visualizar o processamento das leituras dos trilhos ferroviários em tempo real. A formatação da tela visualizada pelo operador do Carro Controle EM-80 apresenta a geometria do trilho implantado na linha, na cor cinza, denominado de gabarito do trilho e sobreposto a este, a leitura da geometria do trilho ferroviário naquele ponto, na cor amarela.

O Carro Controle EM-80 fornece juntamente com as leituras da geometria do trilho ferroviário a sua posição geográfica na via férrea, ou seja, em quais marcos geográfico da ferrovia estão sendo realizadas as leituras, como exposto anteriormente. A Figura 16 apresenta a diagramação da tela que o operador visualiza no monitor, assim como todas as informações disponibilizadas ao operador do Carro Controle EM-80 em tempo real.

Figura 16 – Leitura da geometria do trilho na tela do computador de bordo do Carro Controle

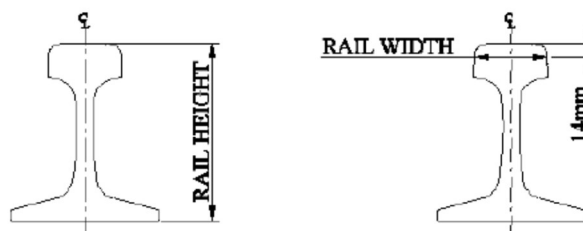


Fonte: Plasser &Theurer (2008, p. 1020)

Este veículo de inspeção ferroviária realiza as leituras da geometria dos trilhos ferroviários: na altura do trilho e na largura do boleto. Com base nestas duas leituras realizadas para esta seção do trilho, é possível analisar e verificar o avanço dos possíveis desgastes na geometria dos trilhos ferroviários.

A altura do trilho ferroviário no Carro Controle EM-80 é definido como *Height*, sendo medido a partir da base do trilho até o topo do trilho, em seu centro. A largura do boleto do trilho é definida como *Width*, sendo medido a uma distância de 14 mm abaixo do topo do trilho ferroviário em seu eixo. A Figura 17 apresenta o esquema de leituras da altura do trilho e da largura do boleto realizadas pelo Carro Controle EM-80.

Figura 17 – Leitura da geometria do trilho pelo Carro Controle EM-80



Fonte: Plasser &Theurer (2008, p. 179)

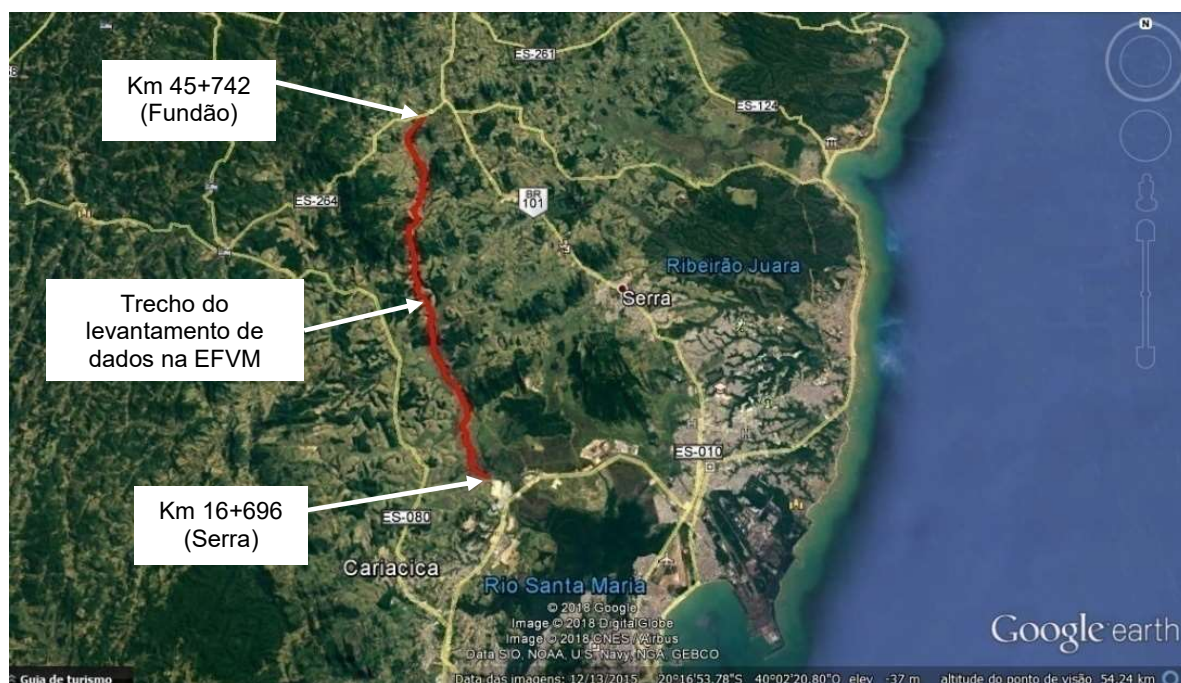
As leituras realizadas pelo Carro Controle EM-80 referentes a geometria dos trilhos ferroviários geram conjunto de dados. Esses conjuntos de dados gerados são compostos pelas leituras da altura do trilho, da largura do boleto e seu marco geográfico na via férrea, sendo estes dados utilizados na análise e avaliação dos possíveis avanços dos desgastes na geometria dos trilhos. No capítulo seguinte será apresentado a metodologia aplicada para a realização do levantamento de dados, a formação dos conjuntos de dados e seus respectivos tratamentos para que estes dados sejam aplicados nesta dissertação.

4 LEVANTAMENTO DE DADOS

O levantamento de dados para o desenvolvimento desta dissertação, ocorreu na linha 1 da EFVM, a partir do marco geográfico do km 16+696, no município de Serra até o km 45+742, no município de Fundão, o que representa a um trecho de 29 km de ferrovia, estando presentes neste trecho de 29 km, trechos em tangentes e curvas.

A Figura 18 apresenta a localização do trecho do km 16+696 até o km 45+745 onde foram levantados os dados para o desenvolvimento dessa dissertação.

Figura 18 – Trecho da coleta de dados na EFVM



Fonte: Google Earth, (2018)

Define-se como trilho direito e trilho esquerdo da linha 1, da seguinte forma: posicionando-se entre os dois trilhos e no sentido crescente aos marcos geográfico, Vitória para Belo Horizonte, tem-se do lado esquerdo o trilho esquerdo e do lado direito o trilho direito.

Para o desenvolvimento desta dissertação fez-se necessário o levantamento de dados referentes: 1) a geometria dos trilhos, sendo a altura do trilho, a largura do boleto e seu respectivo marco geográfico na linha; 2) os projetos geométricos da linha, para verificar os trechos em tangentes, curvas e os graus da curva e seu respectivo marco

geográfico na linha; 3) os levantamentos referentes ao carregamento total trafegado sobre os trilhos direito e esquerdo, sendo estes, o peso da carga transportada, o peso das locomotivas e o peso dos vagões vazios e 4) o levantamento das datas de quando ocorreram as substituições dos trilhos para o trecho de 29 km e seus respectivos marcos geográfico na linha. Sendo que todos estes dados foram levantados para o trecho de 29 km, do km 16+696 ao km 45+742.

Os dados da geometria do trilho, referentes as leituras realizadas na geometria dos trilhos ferroviários foram levantadas pela equipe de manutenção da via permanente da EFVM por intermédio do Carro Controle EM-80, citado anteriormente. Eles foram obtidos no período de abril de 2011 até novembro de 2017, o que corresponde a um período de 6 anos e 8 meses de dados levantados para os dois trilhos ferroviários.

As leituras da geometria dos trilhos realizados pelo Carro Controle EM-80 ocorrem a uma distância de 25 cm entre cada leitura ao longo do trilho, como citado anteriormente. Assim, um trilho que apresenta 12 m de comprimento obterá 48 leituras neste trilho.

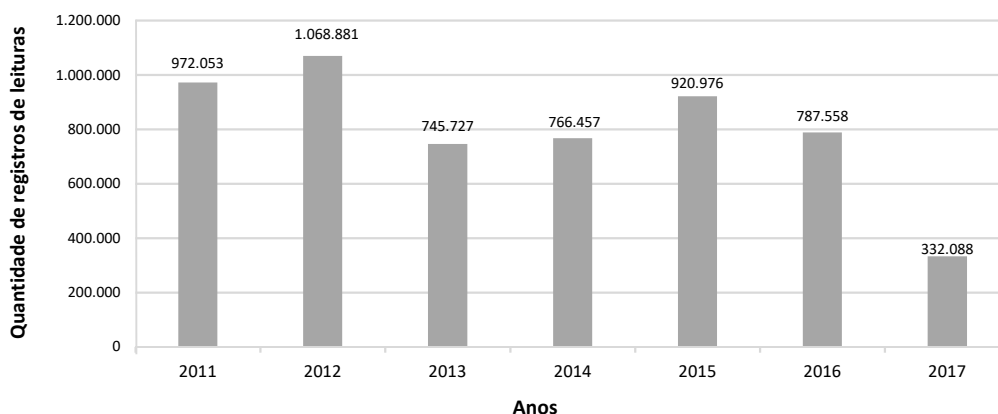
Estas leituras ocorrem simultaneamente em ambos os trilhos, esquerdo e direito. Entretanto, estes dados são apresentados separadamente, para cada um dos dois trilhos, gerando assim, conjunto de dados distintos para cada um destes dois trilhos, sendo incorporado a estas leituras nos trilhos os seus respectivos marcos geográfico.

Para o trecho de 29 km no período de 6 anos e 8 meses obtiveram 11.187.480 registros de leituras da geometria dos trilhos ferroviários. Nestes registros estão contemplados a altura do trilho, a largura do boleto do trilho e seu respectivo marco geográfico. Estes 11.187.480 registros referem-se aos trilhos esquerdo e direito, sendo que, para cada um dos dois trilhos foram levantados 5.593.740 registros da geometria do trilho, o que perfazem para os dois trilhos o valor de 11.187.480 registros de leituras.

Esse levantamento de 11.187.480 registros, referem-se aos valores extraídos diretamente das leituras realizadas pelo Carro Controle EM-80, para ambos os trilhos. A Figura 19 ilustra a distribuição dos 5.593.740 registros levantados em leituras realizadas para o trilho esquerdo da linha 1, observa-se que nos anos de 2011 a 2016 os valores se mantiveram equilibrados e seguiram aos planejamentos de leituras da

equipe de manutenção da EFVM, para o ano de 2017 obtiveram um número reduzido de leituras, havendo para este ano uma redução de leituras para este trecho, o que propiciou a um valor reduzido de amostras, distanciando-se dos valores apresentados para os anos de 2011 a 2016.

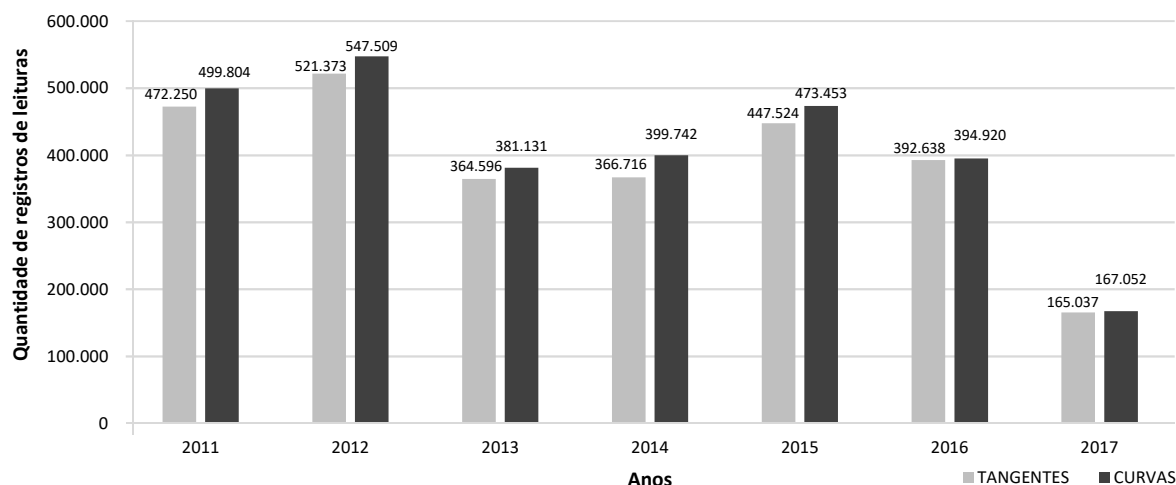
Figura 19 – Distribuição dos registros para o período de 2011 a 2017 trilho esquerdo – linha 1 – EFVM



Fonte: O autor (2019)

Esses registros ilustrados no gráfico da Figura 19 foram desmembrados para os trechos em tangentes e curvas, tendo como objetivo, analisar a sua distribuição e verificar se apresentavam uma distribuição equilibrada e proporcional entre as partes, ou seja, entre os trechos em tangentes e curvas. A Figura 20 ilustra o gráfico da distribuição dos registros referentes aos trechos em tangentes e curvas para o trilho esquerdo da linha 1 da EFVM ao longo do trecho de 29 km para cada um dos anos, perfazendo a um período de 6 anos e 8 meses. Cabe ressaltar que a distribuição dos registros para o trilho direito da linha 1 são os mesmos apresentados para o trilho esquerdo da linha 1, uma vez que os trilhos são paralelos. Da mesma forma, pode-se observar que no gráfico mostrado na Figura 20 os dados referentes a leitura da geometria dos trilhos para o ano de 2017 apresenta uma amostra reduzida em relação aos períodos de 2011 até 2016. Entretanto para os períodos de 2011 até 2017 os valores se mantiveram equilibrados entre estes anos e com uma distribuição equilibrada na quantidade de registros de leituras entre tangentes e curvas.

Figura 20 – Distribuição dos registros referentes aos trechos em tangentes e curvas para o trilho esquerdo – linha 1 - EFVM



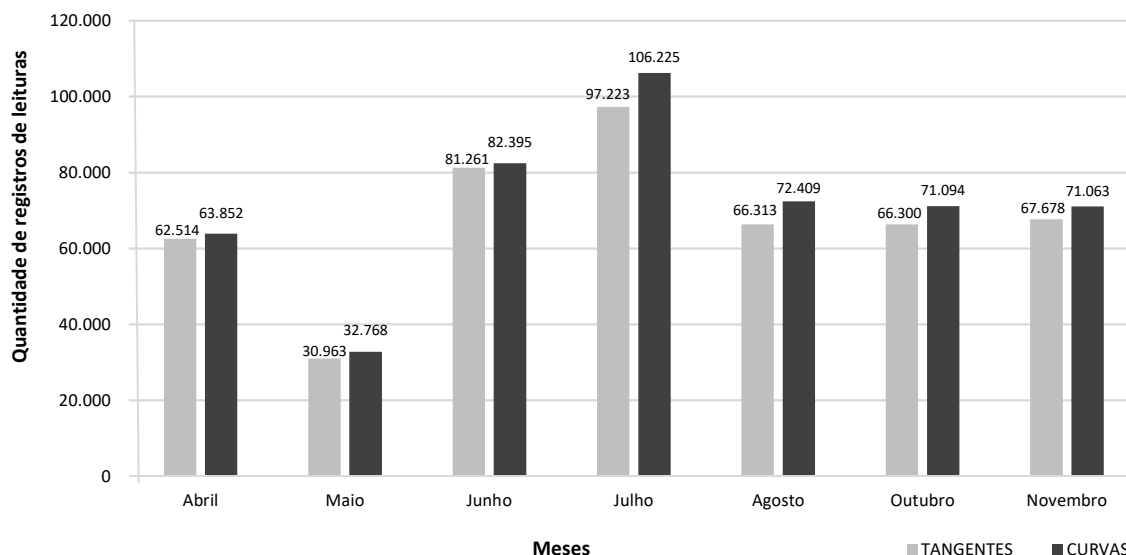
Fonte: O autor (2019)

Para se obter uma compreensão mais detalhada das leituras da geometria dos trilhos ferroviários levantados para cada um dos anos, serão apresentados a seguir as informações referentes a cada ano do levantamento das leituras, apresentando os meses que foram realizadas essas leituras e a quantidade de registros levantados para cada um destes meses e a sua distribuição entre tangentes e curvas para este período de 6 anos e 8 meses.

No ano de 2011, foram realizadas sete leituras da geometria do trilho esquerdo e direito da linha 1 da EFVM, sendo estas leituras realizadas nos seguintes meses: abril, maio, junho, julho, agosto, outubro e novembro. Nestas leituras foram levantados 972.053 registros para o trilho esquerdo, sendo que 472.250 registros são referentes a trechos em tangentes e 499.804 registros são referentes a trechos em curvas. Para o trilho direito, foram levantados 972.053 registros, sendo que 472.250 registros são referentes a trechos em tangentes e 499.804 registros são referentes a trechos em curvas. No gráfico da Figura 21 é ilustrada a distribuição dos registros levantados nas leituras realizadas para o ano de 2011 da geometria dos trilhos ferroviários para os trechos em tangentes e curvas para o trilho esquerdo da linha 1. O mês mais representativo para o ano de 2011 foi julho, seguindo pelos meses de junho, abril, agosto, outubro e novembro que se mantiveram de forma equilibrada na obtenção das leituras, o mês de maio foi o que apresentou a menor obtenção de leituras neste ano. Entretanto, pode-se verificar que a distribuição das quantidades de registros de

leituras entre tangentes e curvas se mantiveram equilibradas entre todos os meses para o ano de 2011.

Figura 21 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2011 - linha 1 - EFVM

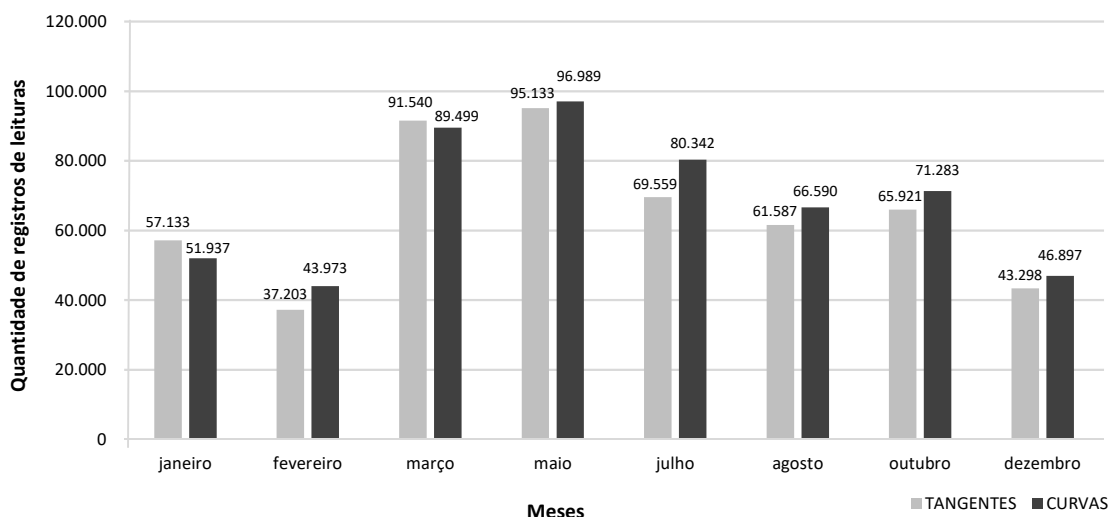


Fonte: O autor (2019)

Para o ano de 2012, foram realizadas oito leituras, para a geometria do trilho esquerdo e direito da linha 1 da EFVM, sendo estas leituras realizadas nos seguintes meses: janeiro, fevereiro, março, maio, julho, agosto, outubro e dezembro. Nestas leituras foram levantados 1.068.881 registros para o trilho esquerdo, sendo que 521.373 registros são referentes a trechos em tangentes e 547.509 registros são referentes a trechos em curvas. Para o trilho direito, foram levantados 1.068.881 registros, sendo que 521.373 registros são referentes a trechos em tangentes e 547.509 registros são referentes a trechos em curvas. No gráfico da Figura 22 é ilustrada a distribuição dos registros levantados nas leituras realizadas para o ano de 2012 da geometria dos trilhos ferroviários para os trechos em tangentes e curvas para o trilho esquerdo da linha 1. Os meses mais representativos para o ano de 2012 foram março e maio, seguindo pelos meses de julho, outubro, agosto que se mantiveram de forma equilibrada na obtenção das leituras, os meses de janeiro e dezembro tiveram uma redução na obtenção das leituras, mas se mantiveram de forma equilibrada, o mês de fevereiro foi o que apresentou a menor obtenção de leituras para o ano de 2012. Entretanto, pode-se verificar que a distribuição das quantidades de registros de

leituras entre tangentes e curvas para todos os meses se mantiveram equilibradas para o ano de 2012.

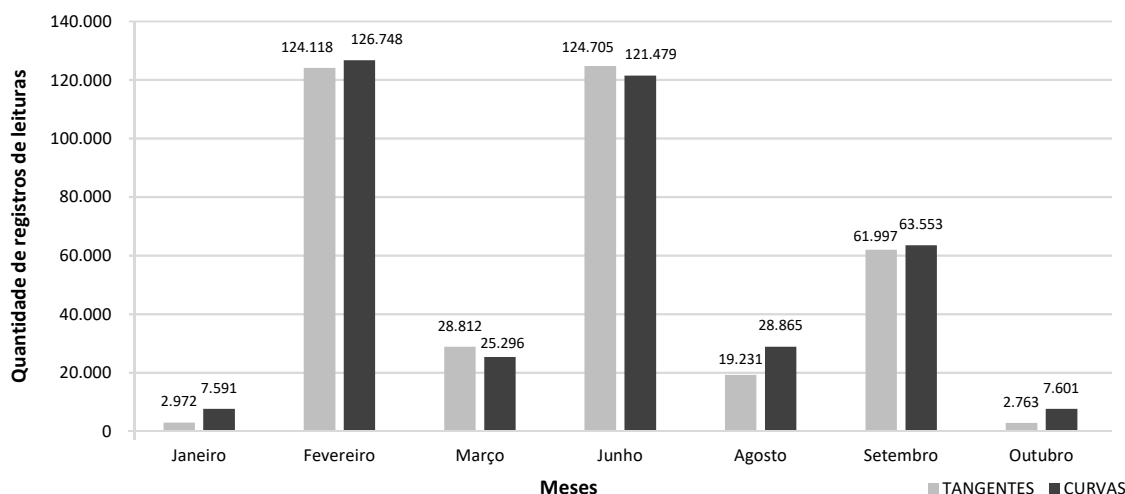
Figura 22 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2012 - linha 1 - EFVM



Fonte: O autor (2019)

Em 2013, foram realizadas sete leituras, para a geometria do trilho esquerdo e direito da linha 1 da EFVM, sendo estas leituras realizadas nos seguintes meses: janeiro, fevereiro, março, junho, agosto, setembro e outubro. Nestas leituras foram levantados 745.727 registros para o trilho esquerdo, sendo que 364.596 registros são referentes a trechos em tangentes e 381.131 registros são referentes a trechos em curvas. Para o trilho direito, foram levantados 745.727 registros, sendo que 364.596 registros são referentes a trechos em tangentes e 381.131 registros são referentes a trechos em curvas. No gráfico da Figura 23 é ilustrada a distribuição dos registros levantados nas leituras realizadas para o ano de 2013 da geometria dos trilhos ferroviários para os trechos em tangentes e curvas para o trilho esquerdo da linha 1. Os meses mais representativos para o ano de 2013 foram fevereiro e junho, sendo que o mês de setembro teve uma redução na obtenção das leituras em relação aos meses de fevereiro e junho. Os meses de janeiro, março, agosto e outubro foram os meses que apresentaram a menor obtenção de leituras para o ano de 2013. De qualquer forma, pode-se observar que a distribuição das quantidades de registros de leituras entre tangentes e curvas se mantiveram equilibradas entre todos os meses para o ano de 2013.

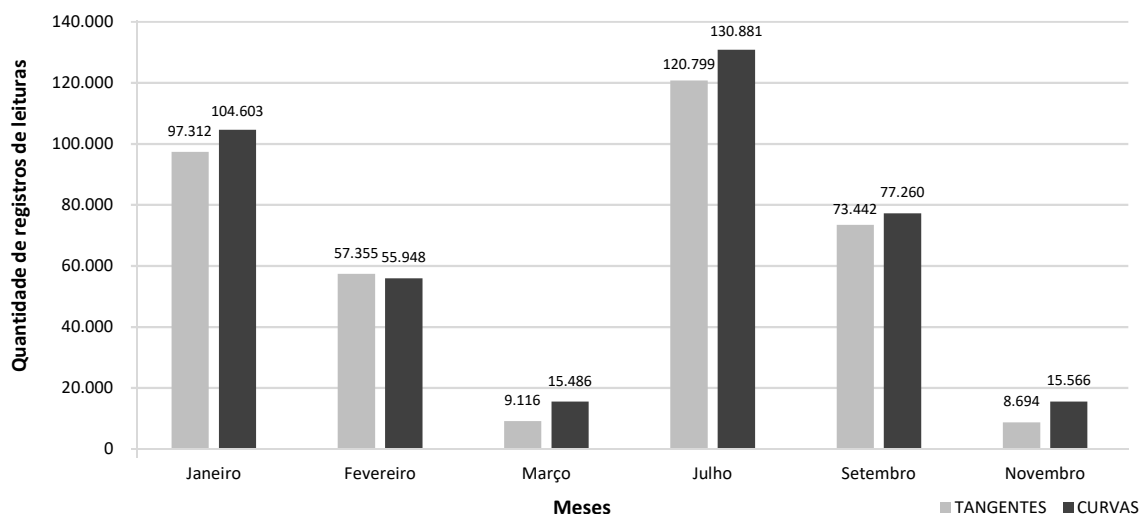
Figura 23 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2013 - linha 1 - EFVM



Fonte: O autor (2019)

Para o ano de 2014, foram realizadas seis leituras, para a geometria do trilho esquerdo e direito da linha 1 da EFVM, sendo estas leituras realizadas nos seguintes meses: janeiro, fevereiro, março, julho, setembro e novembro. Nestas leituras foram levantados 766.475 registros para o trilho esquerdo, sendo que 366.716 registros são referentes a trechos em tangentes e 399.742 registros são referentes a trechos em curvas. Para o trilho direito, foram levantados 766.475 registros, sendo que 366.716 registros são referentes a trechos em tangentes e 399.742 registros são referentes a trechos em curvas. No gráfico da Figura 24 é ilustrada a distribuição dos registros levantados nas leituras realizadas para o ano de 2014 da geometria dos trilhos ferroviários para os trechos em tangentes e curvas para o trilho esquerdo da linha 1. O mês mais representativo para o ano de 2014 foi julho, seguido pelos meses de janeiro, setembro e fevereiro. Os meses de março e novembro tiveram uma redução na obtenção das leituras em relação aos meses anteriores, e apresentaram a menor obtenção de leituras para o ano de 2014. Entretanto, pode-se verificar que a distribuição das quantidades de registros de leituras entre tangentes e curvas para todos os meses se mantiveram equilibradas para o ano de 2014.

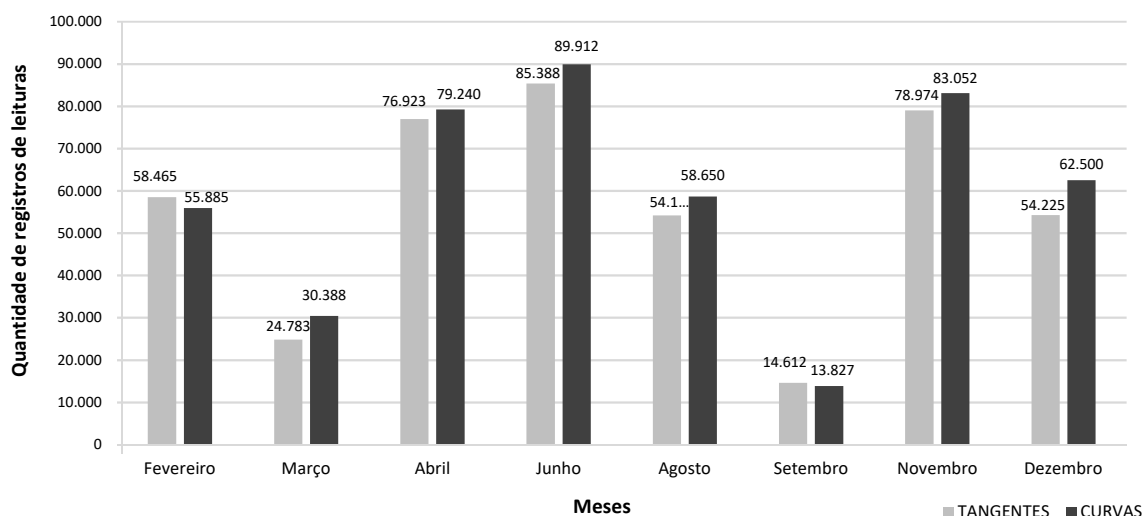
Figura 24 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2014 - linha 1 - EFVM



Fonte: O autor (2019)

No ano de 2015, foram realizadas oito leituras, para a geometria do trilho esquerdo e direito da linha 1 da EFVM, sendo estas leituras realizadas nos seguintes meses: fevereiro, março, abril, junho, agosto, setembro, novembro e dezembro. Nestas leituras foram levantados 920.976 registros para o trilho esquerdo, sendo que 447.524 registros são referentes a trechos em tangentes e 473.453 registros são referentes a trechos em curvas. Para o trilho direito, foram levantados 920.976 registros, sendo que 447.524 registros são referentes a trechos em tangentes e 473.453 registros são referentes a trechos em curvas. No gráfico da Figura 25 é ilustrada a distribuição dos registros levantados nas leituras realizadas para o ano de 2015 da geometria dos trilhos ferroviários para os trechos em tangentes e curvas para o trilho esquerdo da linha 1. O mês mais representativo para o ano de 2015 foi junho, seguido pelos meses de novembro, abril, dezembro e fevereiro. Os meses de março e setembro tiveram uma redução na obtenção das leituras em relação aos meses anteriores, sendo que o mês de setembro que apresentou a menor obtenção de leituras para o ano de 2015. Entretanto, pode-se verificar que a distribuição das quantidades de registros de leituras entre tangentes e curvas se mantiveram equilibradas entre todos os meses para o ano de 2015.

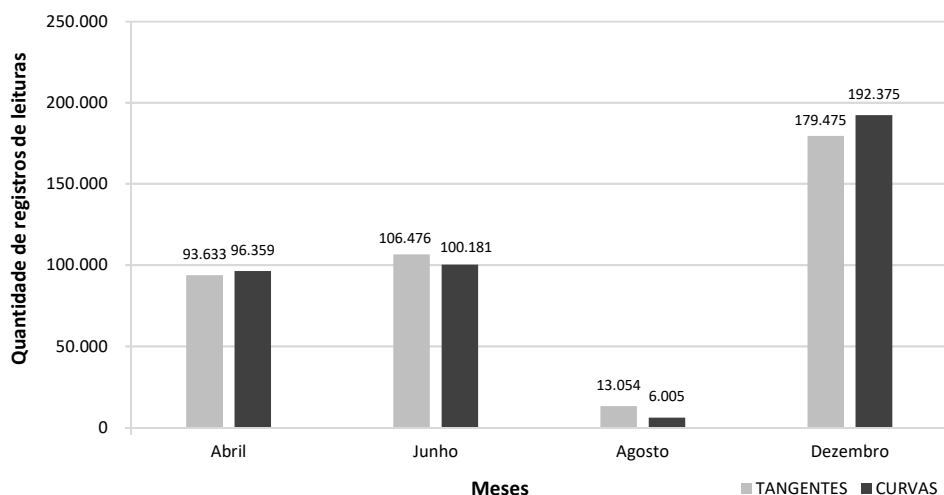
Figura 25 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2015 - linha 1 - EFVM



Fonte: O autor (2019)

Em 2016, foram realizadas quatro leituras, para a geometria do trilho esquerdo e direito da linha 1 da EFVM, sendo estas leituras realizadas nos seguintes meses: abril, junho, agosto e dezembro. Nestas leituras foram levantados 787.558 registros para o trilho esquerdo, sendo que 392.638 registros são referentes a trechos em tangentes e 394.920 registros são referentes a trechos em curvas. Para o trilho direito, foram levantados 787.558 registros, sendo que 392.638 registros são referentes a trechos em tangentes e 394.920 registros são referentes a trechos em curvas. No gráfico da Figura 26 é ilustrada a distribuição dos registros levantados nas leituras realizadas para o ano de 2016 da geometria dos trilhos ferroviários para os trechos em tangentes e curvas para o trilho esquerdo da linha 1. O mês mais representativo para o ano de 2016 foi dezembro, seguido pelos meses de junho e abril. O mês de agosto teve uma redução na obtenção das leituras em relação aos meses anteriores, e apresentou a menor obtenção de leituras para o ano de 2016. Outro ponto a ser observado para o ano de 2016 é que houve uma redução na quantidade de leituras em relação aos anos anteriores de 2011 até 2015. De qualquer forma, pode-se verificar que a distribuição das quantidades de registros de leituras entre tangentes e curvas se mantiveram equilibradas entre todos os meses para o ano de 2016.

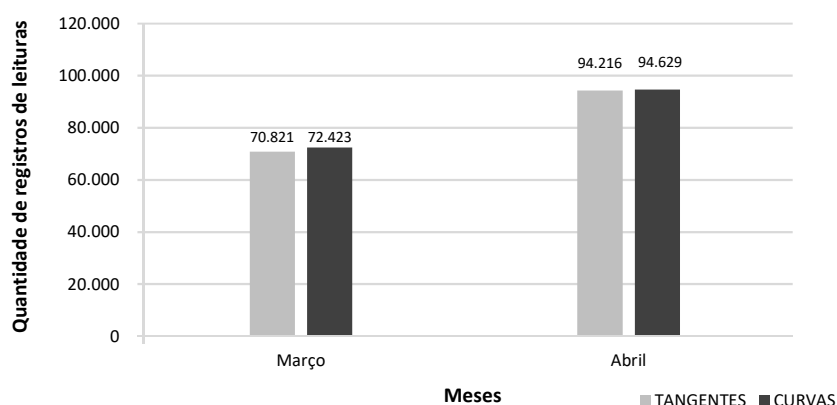
Figura 26 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2016 - linha 1 - EFVM



Fonte: O autor (2019)

Para o ano de 2017, até o período de levantamento de dados para a elaboração desta dissertação, tinham sido realizadas somente duas leituras, para a geometria do trilho esquerdo e direito da linha 1 da EFVM, sendo estas leituras realizadas nos seguintes meses: março e abril. Nestas leituras foram levantados 332.088 registros para o trilho esquerdo, sendo que 165.037 registros são referentes a trechos em tangentes e 167.052 registros são referentes a trechos em curvas. Para o trilho direito, foram levantados 332.088 registros, sendo que 165.037 registros são referentes a trechos em tangentes e 167.052 registros são referentes a trechos em curvas. No gráfico da Figura 27 é ilustrada a distribuição dos registros levantados nas leituras realizadas para o ano de 2017 da geometria dos trilhos ferroviários para os trechos em tangentes e curvas para o trilho esquerdo da linha 1. O mês mais representativo para o ano de 2017 foi abril, seguido pelo mês de março, para o ano de 2017. Houve somente a obtenção de duas leituras para este ano, podendo assim refletir nos resultados. De qualquer forma estes dados foram considerados no desenvolvimento do trabalho para a obtenção da vida útil remanescente dos trilhos. Pode-se verificar que a distribuição das quantidades de registros de leituras entre tangentes e curvas entre estes dois meses se mantiveram equilibradas para o ano de 2017.

Figura 27 – Distribuição dos registros referentes as leituras para o trilho esquerdo para o ano de 2017 - linha 1 - EFVM



Fonte: O autor (2019)

Os dados referentes a determinação de trechos em tangentes e curvas, graus da curva e seus respectivos marcos geográfico na linha 1 foram levantados por meio dos projetos geométricos da linha 1 da EFVM, sendo estes confirmados durante a execução das inspeções e leituras pelo Carro Controle EM-80. Cabe ressaltar que os graus da curva têm correlação com o raio, sendo usual sua utilização em projetos de traçados ferroviários, sendo que para tangentes, o grau da curva é zero. Na Tabela 2 são apresentados os raios e seus respectivos graus da curva para o trecho do km 16+696 ao km 45+742 da linha 1 da EFVM.

Tabela 2 – Raios e grau da curva - km 16+696 ao km 45+742 - EFVM

Raio (m)	Grau da curva (graus)
1.942,00	0,59
769,00	1,49
764,00	1,50
721,00	1,59
573,00	2,00
451,00	2,54
382,00	3,00
378,00	3,03
364,00	3,15
361,00	3,17
358,00	3,20
347,00	3,30
344,00	3,33
286,00	4,00

Fonte: O autor (2019)

Os dados referentes ao carregamento total trafegado sobre a linha 1 da EFVM foram levantados para o período de abril de 2011 até novembro de 2017, mesmo período do levantamento das leituras da geometria dos trilhos, perfazendo um período de 6 anos e 8 meses. Para este período de 6 anos e 8 meses, trafegou sobre a linha 1 da ferrovia EFVM o carregamento total de 733.190.887 toneladas, estando incluso neste valor os carregamentos referentes a carga transportada, o peso das locomotivas e o peso dos vagões. Os carregamentos referentes a carga transportada, o peso das locomotivas e o peso dos vagões, são apresentados nas Tabela 3, Tabela 5 e Tabela 7 e na Tabela 8 é apresentado o carregamento total trafegado sobre a linha 1 da EFVM. Cabe ressaltar que nestas tabelas são apresentados somente os valores referentes ao período de levantamento de dados para esta dissertação, abril de 2011 até novembro de 2017 sendo excluídos destas os demais meses e anos.

O peso referente a carga transportada sobre a linha 1 da ferrovia EFVM, para o período de 6 anos e 8 meses, foi de 480.965.763 toneladas. Este peso foi obtido por meio de registros levantados pelas áreas de operação e planejamento ferroviário da EFVM. A Tabela 3 apresenta os pesos referente a carga transportada pela EFVM sobre a linha 1 para cada mês e ano, para o período de 6 anos e 8 meses.

A quantidade de vagões e o tipo de vagões que trafegaram sobre a linha 1 para este período de 6 anos e 8 meses na EFVM, foram levantados por meio dos registros com a operação ferroviária da EFVM. Com estes dados levantados, obteve-se a parcela referente ao peso dos vagões vazios. A Tabela 4 apresenta a quantidade de vagões utilizados para transportar a carga apresentada na Tabela 3 sobre a linha 1 da EFVM para cada mês e ano, para o período de 6 anos e 8 meses.

Uma vez conhecidos os tipos de vagões e a quantidade de vagões utilizados na realização do transporte do peso da carga apresentada na Tabela 3, determinou-se o peso referentes a estes vagões vazios. Logo, determinou-se o peso destes vagões multiplicando-se a quantidade de vagões por seu respectivo peso, apresentado na especificação técnica do vagão. A Tabela 5 apresenta a parcela correspondente ao peso dos vagões vazios que trafegaram sobre a linha 1 da ferrovia EFVM, para cada mês e ano, para o período de 6 anos e 8 meses.

A EFVM forma trens com 168 vagões e 252 vagões, e adota que cada lote de 84 vagões necessita de uma locomotiva, assim para a formação de trens com 168 vagões

são necessárias duas locomotivas e para a formação de trens com 252 vagões são necessárias três locomotivas.

A quantidade de locomotivas, bem como o tipo de locomotivas necessárias para transportar o peso da carga apresentada na Tabela 3 foram levantados por meio dos registros com a operação ferroviária da EFVM. A Tabela 6 apresenta a quantidade de locomotivas utilizadas para transportar o peso da carga apresentada na Tabela 3 sobre a linha 1 da EFVM para cada mês e ano, para o período de 6 anos e 8 meses.

Da mesma maneira que foram obtidos os pesos para os vagões vazios, procedeu-se para obtenção dos pesos para as locomotivas. Conhecida a quantidade de locomotivas e o tipo de locomotivas utilizadas na realização do transporte do peso da carga apresentada na Tabela 3, determinou-se o peso destas locomotivas, multiplicando-se a quantidade de locomotivas por seu respectivo peso apresentado na especificação técnica de locomotivas. A EFVM utiliza locomotivas do tipo *Dash 9* para a realização do transporte ferroviário, esta locomotiva *Dash 9* possui o peso total de 160 toneladas. A Tabela 7 apresenta a parcela correspondente ao peso das locomotivas que trafegaram sobre a linha 1 da ferrovia EFVM, para cada mês e ano, para o período de 6 anos e 8 meses.

O carregamento total trafegado sobre a linha 1 da EFVM no período de abril de 2011 até novembro de 2017, corresponde ao somatório referente ao peso da carga transportada, ao peso das locomotivas e ao peso dos vagões vazios, perfazendo o valor total de 733.190.887 toneladas, apresentado anteriormente. A Tabela 8 apresenta o carregamento total por mês e ano trafegado na linha 1 da EFVM para o período de 6 anos e 8 meses.

Cabe ressaltar que o carregamento total de 733.190.887 toneladas se refere aos dois trilhos da linha 1 da EFVM. Assim, estabeleceu-se por hipótese que o carregamento total se distribui igualmente em 50% para cada um dos dois trilhos. Desta forma, o trilho esquerdo recebe a carga de 366.595.443 toneladas e o trilho direito recebe a carga de 366.595.443 toneladas, perfazendo o carregamento total de 733.190.887 toneladas trefegada sobre a linha 1 da EFVM para o período de 6 anos e 8 meses.

Tabela 3 – Peso da carga transportada sobre a linha 1 - EFVM

ANO	TOTAL (toneladas)	JANEIRO (toneladas)	FEVEREIRO (toneladas)	MARÇO (toneladas)	ABRIL (toneladas)	MAIO (toneladas)	JUNHO (toneladas)	JULHO (toneladas)	AGOSTO (toneladas)	SETEMBRO (toneladas)	OUTUBRO (toneladas)	NOVEMBRO (toneladas)	DEZEMBRO (toneladas)
2011	55.429.500	-----	-----	-----	5.798.000	6.140.500	6.197.000	6.548.500	6.558.000	6.378.000	6.354.000	6.063.500	5.392.000
2012	71.621.500	4.633.500	5.769.500	5.831.500	5.933.500	6.154.000	5.917.000	6.128.500	6.192.500	6.282.000	6.394.000	5.969.500	6.416.000
2013	69.839.700	5.013.200	5.121.500	5.510.500	5.863.500	6.038.000	5.938.500	6.390.500	6.330.500	6.345.000	6.561.500	6.218.000	4.509.000
2014	71.487.500	5.915.000	5.022.000	5.915.000	5.695.000	6.217.500	6.097.500	6.107.000	6.285.000	6.157.000	6.158.500	5.807.500	6.110.500
2015	74.576.500	5.960.500	5.347.000	5.841.000	5.834.000	6.594.500	6.223.000	6.711.000	6.446.000	6.617.000	6.642.500	5.784.500	6.575.500
2016	71.431.948	5.175.938	5.295.064	5.423.597	5.964.128	6.310.040	5.736.901	6.507.502	6.559.913	6.156.966	6.457.500	5.851.399	5.993.000
2017	66.579.115	6.029.193	5.581.758	6.209.802	5.448.252	6.224.148	6.321.334	6.206.261	5.998.873	6.127.839	6.401.252	6.030.400	-----
TOTAL	480.965.763	32.727.331	32.136.823	34.731.399	40.536.380	43.678.688	42.431.235	44.599.263	44.370.786	44.063.806	44.969.252	41.724.799	34.996.000

Fonte: O autor (2019)

Tabela 4 – Quantidade de vagões que trafegaram sobre a linha 1 - EFVM

ANO	TOTAL (unidade)	JANEIRO (unidade)	FEVEREIRO (unidade)	MARÇO (unidade)	ABRIL (unidade)	MAIO (unidade)	JUNHO (unidade)	JULHO (unidade)	AGOSTO (unidade)	SETEMBRO (unidade)	OUTUBRO (unidade)	NOVEMBRO (unidade)	DEZEMBRO (unidade)
2011	1.419.760	-----	-----	-----	148.569	157.329	158.729	167.698	167.995	163.373	162.741	155.270	138.056
2012	1.834.446	118.640	147.760	149.336	151.983	157.670	151.624	157.031	158.636	160.894	163.758	152.841	164.272
2013	1.788.635	127.795	131.230	141.246	150.216	154.676	152.124	163.720	162.210	162.515	168.073	159.263	115.568
2014	1.831.261	151.463	128.686	151.535	145.947	159.315	156.242	156.457	160.962	157.660	157.712	148.766	156.517
2015	1.910.283	152.585	136.984	149.732	149.448	168.865	159.482	171.942	165.072	169.421	170.035	148.237	168.479
2016	1.829.291	132.599	135.693	139.037	152.804	161.688	147.017	166.606	167.963	157.603	165.246	149.766	153.267
2017	1.704.840	154.234	142.856	159.051	139.634	159.395	161.887	158.921	153.643	156.905	163.908	154.406	-----
TOTAL	12.318.516	837.315	823.209	889.938	1.038.600	1.118.939	1.087.106	1.142.375	1.136.480	1.128.372	1.151.474	1.068.549	896.160

Fonte: O autor (2019)

Tabela 5 – Peso total dos vagões que trafegaram na linha 1 - EFVM

ANO	TOTAL (toneladas)	JANEIRO (toneladas)	FEVEREIRO (toneladas)	MARÇO (toneladas)	ABRIL (toneladas)	MAIO (toneladas)	JUNHO (toneladas)	JULHO (toneladas)	AGOSTO (toneladas)	SETEMBRO (toneladas)	OUTUBRO (toneladas)	NOVEMBRO (toneladas)	DEZEMBRO (toneladas)
2011	26.385.447	-----	-----	-----	2.773.229	2.933.550	2.950.013	3.109.704	3.126.011	3.037.956	3.022.613	2.877.583	2.554.787
2012	34.081.365	2.196.359	2.742.321	2.769.129	2.825.193	2.938.827	2.831.394	2.929.964	2.952.742	2.987.818	3.039.983	2.828.058	3.039.577
2013	33.195.038	2.250.637	2.448.668	2.645.601	2.797.604	2.878.657	2.830.486	3.049.658	3.027.085	3.019.250	3.125.250	2.959.097	2.163.043
2014	34.072.507	2.806.173	2.402.448	2.822.237	2.727.770	2.973.175	2.916.329	2.914.528	2.987.113	2.921.128	2.924.675	2.767.436	2.909.496
2015	35.520.992	2.818.288	2.551.160	2.807.444	2.780.830	3.129.192	2.981.663	3.205.093	3.060.753	3.135.324	3.138.800	2.769.996	3.142.450
2016	33.924.716	2.469.073	2.535.260	2.607.935	2.847.856	3.017.955	2.747.139	3.080.848	3.108.966	2.908.706	3.039.601	2.760.915	2.800.463
2017	31.581.217	2.826.443	2.631.686	2.954.699	2.611.998	2.956.347	3.003.220	2.944.392	2.853.216	2.905.341	3.035.509	2.858.368	-----
TOTAL	228.761.283	15.366.972	15.311.544	16.607.046	19.364.479	20.827.702	20.260.242	21.234.187	21.115.887	20.915.523	21.326.432	19.821.452	16.609.817

Fonte: O autor (2019)

Tabela 6 – Quantidade de locomotivas que trafegaram sobre a linha 1 – EFVM

ANO	TOTAL (unidade)	JANEIRO (unidade)	FEVEREIRO (unidade)	MARÇO (unidade)	ABRIL (unidade)	MAIO (unidade)	JUNHO (unidade)	JULHO (unidade)	AGOSTO (unidade)	SETEMBRO (unidade)	OUTUBRO (unidade)	NOVEMBRO (unidade)	DEZEMBRO (unidade)
2011	16.902	-----	-----	-----	1.769	1.873	1.890	1.996	2.000	1.945	1.937	1.848	1.644
2012	21.839	1.412	1.759	1.778	1.809	1.877	1.805	1.869	1.889	1.915	1.950	1.820	1.956
2013	21.293	1.521	1.562	1.682	1.788	1.841	1.811	1.949	1.931	1.935	2.001	1.896	1.376
2014	21.801	1.803	1.532	1.804	1.737	1.897	1.860	1.863	1.916	1.877	1.878	1.771	1.863
2015	22.741	1.816	1.631	1.783	1.779	2.010	1.899	2.047	1.965	2.017	2.024	1.765	2.006
2016	21.777	1.579	1.615	1.655	1.819	1.925	1.750	1.983	2.000	1.876	1.967	1.783	1.825
2017	20.296	1.836	1.701	1.893	1.662	1.898	1.927	1.892	1.829	1.868	1.951	1.838	-----
TOTAL	146.649	9.968	9.800	10.595	12.364	13.321	12.942	13.600	13.530	13.433	13.708	12.721	10.669

Fonte: O autor (2019)

Tabela 7 – Peso total das locomotivas que trafegaram sobre a linha 1 - EFVM

ANO	TOTAL (toneladas)	JANEIRO (toneladas)	FEVEREIRO (toneladas)	MARÇO (toneladas)	ABRIL (toneladas)	MAIO (toneladas)	JUNHO (toneladas)	JULHO (toneladas)	AGOSTO (toneladas)	SETEMBRO (toneladas)	OUTUBRO (toneladas)	NOVEMBRO (toneladas)	DEZEMBRO (toneladas)
2011	2.704.305	-----	-----	-----	282.988	299.675	302.342	319.425	319.990	311.188	309.983	295.752	262.964
2012	3.494.182	225.980	281.449	284.450	289.491	300.323	288.807	299.107	302.164	306.466	311.920	291.125	312.900
2013	3.406.923	243.418	249.961	269.041	286.125	294.621	289.760	311.847	308.971	309.552	320.140	303.358	220.129
2014	3.488.117	288.501	245.116	288.639	277.994	303.457	297.605	298.014	306.594	300.305	300.403	283.363	298.127
2015	3.638.635	290.639	260.921	285.204	284.662	321.649	303.776	327.508	314.423	322.707	323.877	282.357	320.913
2016	3.484.363	252.570	258.463	264.833	291.055	307.978	280.033	317.344	319.929	300.197	314.755	285.269	291.938
2017	3.247.315	293.779	272.106	302.954	265.970	303.609	308.356	302.707	292.654	298.866	312.206	294.107	-----
TOTAL	23.463.841	1.594.886	1.568.016	1.695.121	1.978.285	2.131.312	2.070.678	2.175.953	2.164.724	2.149.281	2.193.283	2.035.331	1.706.971

Fonte: O autor (2019)

Tabela 8 – Peso do carregamento total trafegado sobre a linha 1 - EFVM

ANO	TOTAL (toneladas)	JANEIRO (toneladas)	FEVEREIRO (toneladas)	MARÇO (toneladas)	ABRIL (toneladas)	MAIO (toneladas)	JUNHO (toneladas)	JULHO (toneladas)	AGOSTO (toneladas)	SETEMBRO (toneladas)	OUTUBRO (toneladas)	NOVEMBRO (toneladas)	DEZEMBRO (toneladas)
2011	84.519.252	-----	-----	-----	8.854.217	9.373.725	9.449.354	9.977.629	10.004.001	9.727.144	9.686.596	9.236.835	8.209.751
2012	109.197.048	7.055.840	8.793.270	8.885.079	9.048.184	9.393.150	9.037.200	9.357.572	9.447.406	9.576.284	9.745.903	9.088.682	9.768.477
2013	106.441.662	7.507.255	7.820.130	8.425.142	8.947.229	9.211.278	9.058.746	9.752.005	9.666.556	9.673.802	10.006.890	9.480.455	6.892.173
2014	109.048.124	9.009.673	7.669.565	9.025.876	8.700.763	9.494.132	9.311.433	9.319.542	9.578.706	9.378.433	9.383.578	8.858.299	9.318.123
2015	113.736.127	9.069.426	8.159.081	8.933.648	8.899.492	10.045.341	9.508.438	10.243.601	9.821.176	10.075.031	10.105.177	8.836.852	10.038.863
2016	108.841.027	7.897.581	8.088.788	8.296.365	9.103.039	9.635.973	8.764.072	9.905.694	9.988.807	9.365.869	9.811.856	8.897.583	9.085.401
2017	101.407.647	9.149.415	8.485.550	9.467.456	8.326.220	9.484.104	9.632.910	9.453.360	9.144.743	9.332.046	9.748.967	9.182.875	-----
TOTAL	733.190.887	49.689.190	49.016.383	53.033.566	61.879.144	66.637.702	64.762.155	68.009.403	67.651.397	67.128.609	68.488.968	63.581.582	53.312.788

Fonte: O autor (2019)

As datas referentes a substituição dos trilhos esquerdo e direito da linha 1 da EFVM, foram levantadas com a área de manutenção da via permanente da EFVM, a qual é responsável pela realização desta atividade. Assim, levantaram-se as datas em que os trilhos, esquerdo e direito na linha 1 para o trecho do km 16+696 ao km 45+742 foram substituídos, bem como os seus respectivos marcos geográfico. Para este trecho de 29 km as substituições dos trilhos ocorreram no período de maio de 2015 até novembro de 2017. Cabe ressaltar que a notação apresentada de km 16+696, o primeiro valor refere-se em quilômetros e os valores após o sinal de soma indicam os valores em metro, assim sendo, têm-se 16 quilômetros e seiscentos e noventa e seis metros.

Para a linha 1 no trecho de 29 km de ferrovia e para o período de maio de 2015 até novembro de 2017 foram substituídos 9.637 m de trilhos esquerdo e 8.656 m de trilhos direito. Assim, para o trecho de 29 km foram substituídos um total de 18.293 m de trilhos que se refere a soma da substituição dos trilhos esquerdo e direito.

Estes registros das substituições dos trilhos, foram ajustados de forma a poderem ser inseridos no conjunto de dados gerado das leituras da geometria dos trilhos, uma vez que, as leituras no trilho foram realizadas a cada 25 cm, assim o conjunto de dados referente a substituição dos trilhos teve que ser adequado a este intervalo de 25 cm e para os marcos geográfico das leituras da geometria do trilho. Logo, foram obtidos 38.548 registros de substituições para o trilho esquerdo e 34.624 registros de substituições para o trilho direito, perfazendo 73.172 registros de substituição de trilhos para o trecho de 29 km da linha 1.

Nestes registros de substituição dos trilhos são apresentados o ano da substituição do trilho; o trilho substituído, sendo este esquerdo ou direito; o marco geográfico inicial e final de sua substituição; o comprimento em metros de trilho substituído; e a data de sua substituição.

Para esta dissertação foram levantados os dados referentes: 1) a geometria dos trilhos, sendo a altura do trilho, a largura do boleto, e seus respectivos marcos geográfico na linha; 2) os projetos geométricos da linha, para se verificar os trechos em tangentes, curvas e os graus da curva e seus respectivos marcos geográfico na linha; 3) os levantamentos referentes ao carregamento total trafegados sobre os trilhos direito e esquerdo, sendo estes, o peso da carga transportada, o peso das locomotivas e o peso dos vagões vazios e 4) o levantamento das datas de quando ocorreram as substituições dos trilhos para o trecho de 29 km e seus respectivos marcos geográfico na linha.

Todos estes quatro conjuntos de dados levantados encontram-se em conjuntos de dados distintos, devendo todos estes dados serem tratados e agrupados para formar um único conjunto de dados, de maneira a ser utilizado nesta dissertação. A etapa seguinte contou com a realização do tratamento e agrupamento destes dados de maneira a gerar um único conjunto de dados.

4.1 TRATAMENTO DOS DADOS

Esta etapa contou com a realização do tratamento dos dados levantados e da realização do agrupamento dos diversos conjuntos de dados gerados, para a formação de um único conjunto de dados para ser utilizado nesta dissertação para o treinamento, validação e teste das configurações da RNA e do k-NN.

O conjunto de dados elaborado para ser utilizado nesta dissertação apresenta a seguinte estrutura: 1) grau da curva; 2) trilho interno ou externo a curva; 3) largura do boleto; 4) altura do boleto; 5) peso médio transportado e 6) vida útil remanescente do trilho ferroviário. A Tabela 9 apresenta a estrutura do conjunto de dados elaborado e utilizado nesta dissertação.

Tabela 9 – Conjunto de Dados

Linha 1 - Trilho Direito / Esquerdo - Sentido Vitória a Minas					
Dados de Entrada					Dados de Saída
Grau da Curva	Trilho externo ou interno a curva	Largura do boleto (medido) (mm)	Altura do trilho (medido) (mm)	Peso Médio Transportado (Toneladas) / mês	Vida útil remanescente do Trilho (nº de categorias)

Fonte: O autor (2019)

Foi necessário realizar o tratamento dos dados referente as leituras da geometria dos trilhos direito e esquerdo da linha 1 da EFVM, para que fossem eliminados destes os dados incorretos ou inconsistentes, sendo assim, o tratamento destes dados foi realizado em duas fases.

A primeira fase constou em analisar todo o conjunto de dados e excluir deste, todos os registros que apresentaram erros de leituras, sendo apresentado no conjunto de dados como N/D. Estes erros ocorrem em geral, por não ser possível durante a fase de leitura da geometria do trilho o reconhecimento da seção do trilho ou de realizar as leituras em toda a seção do trilho. Este impedimento pode ser ocasionado como por exemplo o excesso de brita ao lado do trilho que impossibilita o reconhecimento da seção do trilho e a realização das leituras, dentre outros fatores. Após a execução desta fase, o conjunto de dados inicialmente com 11.187.480 registros passou a ter 9.844.983 registros.

A segunda fase teve como objetivo verificar a consistência das leituras realizadas pelo Carro Controle EM-80. Esta análise foi realizada por meio da verificação nas leituras obtidas com base nas cronologias da execução destas inspeções ao longo do tempo. O objetivo era verificar se o desgaste da geometria do trilho em duas leituras consecutivas aumentou ou se manteve igual. Caso contrário, estas duas leituras deveriam ser descartadas, pois, o desgaste não pode diminuir quando analisadas duas leituras para o mesmo marco geográfico em ordem cronológica crescente. Então, o conjunto de dados após a execução da segunda fase passou a ter 8.368.236 registros.

Realizou-se o agrupamento deste conjunto de dados com o conjunto de dados da substituição dos trilhos. Este agrupamento entre os conjuntos de dados foi realizado utilizando como ponto comum de interligação os dados referentes aos marcos

geográfico. Após a realização deste agrupamento dos conjuntos de dados, todos os registros que não tinham data de substituição dos trilhos foram excluídos deste conjunto de dados pois, nestes casos, não seria possível definir a vida útil remanescente do trilho.

O conjunto de dados foi bastante reduzido, ficando com 1.275.034 registros. A Tabela 10 apresenta a quantidade de registros após a realização do agrupamento entre os conjuntos de dados para linha 1 da EFVM, para o trilho esquerdo e trilho direito para o trecho de 29 km, formando assim o conjunto de dados final para ser utilizado no treinamento, validação e teste da RNA.

Tabela 10 – Conjunto de dados final da linha 1 - EFVM

Trilho	Tangentes	Curvas internas	Curvas externas	Total
Esquerdo	210.446	178.137	366.854	755.437
Direito	124.805	224.575	170.217	519.597
Total	335.251	402.712	537.071	1.275.034

Fonte: O autor (2019)

Os dados referentes ao grau da curva foram inseridos neste conjunto de dados por meio dos levantamentos nos projetos geométricos da linha 1, sendo estes marcos geográfico confirmados durante a execução das inspeções e leituras realizadas pelo Carro Controle EM-80, como citado anteriormente.

As inclusões dos dados referentes aos trilhos internos ou externos as curvas, foram inseridos neste conjunto de dados com valores numéricos, 0, 1 e 2, sendo 0 para tangentes, 1 para trilhos internos à curva e 2 para trilhos externos à curva. Sendo que, trilhos internos ou externos às curvas referem-se à posição de implantação do trilho em relação ao raio da curva e no sentido do tráfego. O raio da curva é medido no eixo da linha, ou seja, entre os dois trilhos, assim, o trilho que estiver implantado apresentando um raio menor em relação ao raio do eixo da linha e no sentido do tráfego denomina-se de trilho interno à curva, e o trilho que apresenta um raio maior em relação ao raio do eixo da linha no sentido do tráfego denomina-se de trilho externo à curva.

Os períodos de vida útil remanescente para os trilhos ferroviários foram calculados para cada um dos segmentos de leituras a cada 25 cm ao longo do trilho com base

no mesmo marco geográfico, da seguinte maneira: a data do levantamento da primeira leitura da geometria deste trilho subtraindo-se com a data de substituição deste trilho, sendo que o valor resultante é denominado de vida útil remanescente do trilho ferroviário. Este cálculo foi realizado para cada uma destas seções de 25 cm do trilho, e para todos os registros presentes neste conjunto de dados. Como as leituras foram executadas de maneira cronológica, esses períodos de vida útil remanescente, sofrem variações a cada período na realização de novas leituras da geometria dos trilhos até a data da substituição do trilho.

Como exemplo para o trilho esquerdo no marco geográfico 35+737, onde 35 refere-se ao quilômetro e o valor após o sinal de soma refere-se em metro (trinta e cinco quilômetros e setecentos e trinta e sete metros) a primeira leitura levantada para a elaboração desta dissertação ocorreu em 05/04/2011 e ele foi substituído em 01/11/2017, portanto, o período de vida útil remanescente para este trilho para este intervalo de tempo foi de 80 meses de vida útil remanescente para este trilho. Esse mesmo trilho sendo inspecionado em 16/03/2012, apresenta uma vida útil remanescente para este intervalo de tempo de 69 meses e assim respectivamente para cada nova leitura. Esta metodologia foi aplicada para todos os registros do conjunto de dados para a determinação da vida útil remanescente dos dois trilhos ferroviários da linha 1 da EFVM para o trecho de 29 km.

A Tabela 11 apresenta como exemplo o levantamento realizado para o trilho esquerdo no marco geográfico 35+737 da linha 1 da EFVM. Pode-se observar na Tabela 11 que ao longo do tempo, a geometria do trilho sofreu desgastes na largura do boleto e na altura do trilho à medida que se aumentava o tempo do trilho instalado na via férrea, propiciando a redução da sua vida útil remanescente. Assim em 05/04/2011 data esta da realização da primeira leitura, o trilho apresentava 73,8 mm de largura do boleto e 184,2 mm de altura do trilho, em 17/04/2011 este trilho apresentou 68,2 mm de largura do boleto e 170,1 mm de altura do trilho, ocasionando uma redução de 5,6 mm de largura do boleto e 14,1 mm na altura do trilho para o período de 6 anos e 1 mês, sendo este trilho substituído em 01/11/2017 com 6 anos e 8 meses.

Tabela 11 – Levantamento realizado para o trilho esquerdo da linha 1 – EFVM (35+737)

Marco geográfico	Grau da curva	Trilho externo ou interno a curva	Largura do boleto (mm)	Altura do trilho (mm)	Data da leitura	Data da substituição do trilho	Vida útil remanescente (meses)
35+737	1,49	EXTERNO	73,8	184,2	05/04/2011	01/11/17	80
35+737	1,49	EXTERNO	73,8	184,0	24/06/2011	01/11/17	78
35+737	1,49	EXTERNO	73,7	183,9	22/07/2011	01/11/17	77
35+737	1,49	EXTERNO	71,6	183,1	25/08/2011	01/11/17	76
35+737	1,49	EXTERNO	71,4	182,5	10/10/2011	01/11/17	74
35+737	1,49	EXTERNO	71,4	182,1	17/11/2011	01/11/17	73
35+737	1,49	EXTERNO	71,2	181,3	16/03/2012	01/11/17	69
35+737	1,49	EXTERNO	71,0	180,6	18/05/2012	01/11/17	67
35+737	1,49	EXTERNO	71,0	180,4	31/08/2012	01/11/17	64
35+737	1,49	EXTERNO	70,9	179,8	27/12/2012	01/11/17	60
35+737	1,49	EXTERNO	70,7	179,3	04/02/2013	01/11/17	58
35+737	1,49	EXTERNO	70,5	178,9	03/06/2013	01/11/17	54
35+737	1,49	EXTERNO	70,3	178,4	30/09/2013	01/11/17	51
35+737	1,49	EXTERNO	70,1	178,1	09/01/2014	01/11/17	47
35+737	1,49	EXTERNO	69,9	177,7	17/02/2014	01/11/17	46
35+737	1,49	EXTERNO	69,7	177,4	07/07/2014	01/11/17	41
35+737	1,49	EXTERNO	69,5	177,0	01/09/2014	01/11/17	39
35+737	1,49	EXTERNO	69,5	176,8	09/02/2015	01/11/17	33
35+737	1,49	EXTERNO	69,5	176,2	14/04/2015	01/11/17	32
35+737	1,49	EXTERNO	69,5	175,6	09/06/2015	01/11/17	29
35+737	1,49	EXTERNO	69,5	175,3	09/06/2015	01/11/17	29
35+737	1,49	EXTERNO	69,5	174,7	30/10/2015	01/11/17	25
35+737	1,49	EXTERNO	69,5	174,2	17/12/2015	01/11/17	23
35+737	1,49	EXTERNO	69,5	173,6	05/04/2016	01/11/17	19
35+737	1,49	EXTERNO	69,4	172,8	09/06/2016	01/11/17	17
35+737	1,49	EXTERNO	69,2	172,1	12/12/2016	01/11/17	11
35+737	1,49	EXTERNO	69,0	171,5	08/12/2016	01/11/17	11
35+737	1,49	EXTERNO	68,5	171,1	06/03/2017	01/11/17	8
35+737	1,49	EXTERNO	68,2	170,1	17/04/2017	01/11/17	7

Fonte: O autor (2019)

Da mesma forma, inseriu-se neste conjunto de dados o peso médio transportado sobre o trilho para cada um dos registros dos marcos geográfico. Este valor foi obtido por meio do carregamento total trafegado sobre os trilhos, apresentado no subcapítulo anterior, e que representa o peso da carga transportada, o peso das locomotivas e o peso dos vagões vazio. O valor do peso médio transportado foi obtido dividindo-se o carregamento total trafegado de cada período pela vida útil remanescente do trilho para cada período.

Assim, como exemplo para o trilho esquerdo, no marco geográfico 35+737 da linha 1 da EFVM em 05/04/2011 o carregamento total trafegado foi de 366.595.443 toneladas e a vida útil remanescente do trilho foi de 80 meses, realizando-se a divisão entre estes valores, obtém-se o peso médio transportado de 4.571.016 toneladas/ mês, para este período de vida útil remanescente, sendo este o valor inserido no conjunto de dados. Esta metodologia foi aplicada para todos os registros do conjunto de dados

para a determinação do peso médio transportado e para cada período de vida útil remanescente.

A Tabela 12 apresenta como exemplo o levantamento realizado para o trilho esquerdo no marco geográfico 35+737 da linha 1 da EFVM, com os valores do carregamento total trafegado e o valor do peso médio transportado, obtido conforme apresentado anteriormente. Pode-se observar na Tabela 12 que à medida que a vida útil remanescente do trilho diminui, o peso médio transportado aumenta, assim esta variável é inversamente proporcional a vida útil do trilho. Logo, quanto maior a média da carga trafegada sobre o trilho menor será o período de vida útil remanescente do trilho.

Tabela 12 – Levantamento do peso médio trafegado da linha 1 – EFVM (35+737)

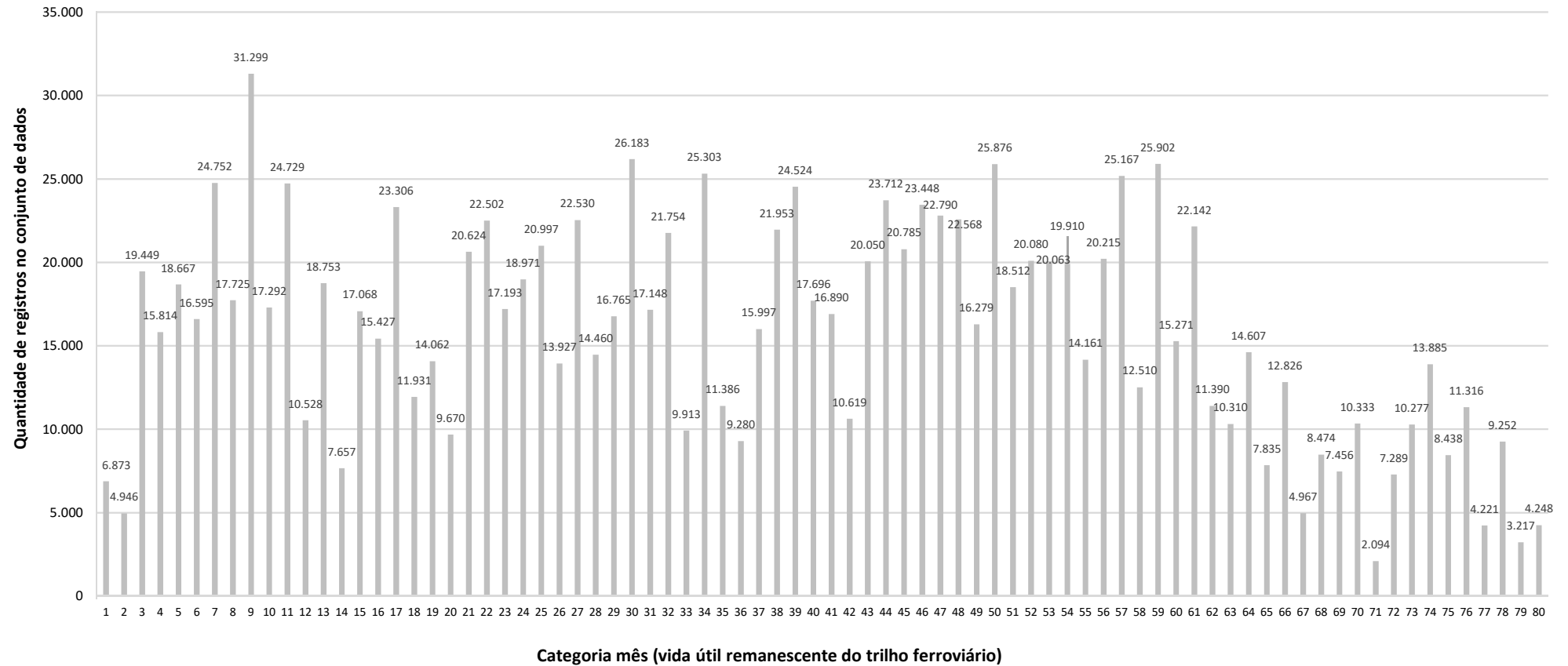
Marco geográfico	Grau da curva	Trilho externo ou interno a curva	Largura do boleto (mm)	Altura do trilho (mm)	Data da leitura	Data da substituição do trilho	carregamento total trafegado (toneladas)	Peso médio transportado (toneladas) / mês	Vida útil remanescente (meses)
35+737	1,49	EXTERNO	73,8	184,2	05/04/2011	01/11/17	366.595.443	4.571.016	80
35+737	1,49	EXTERNO	73,8	184,0	24/06/2011	01/11/17	357.481.472	4.573.324	78
35+737	1,49	EXTERNO	73,7	183,9	22/07/2011	01/11/17	352.756.795	4.571.362	77
35+737	1,49	EXTERNO	71,6	183,1	25/08/2011	01/11/17	347.767.981	4.567.881	76
35+737	1,49	EXTERNO	71,4	182,5	10/10/2011	01/11/17	337.902.408	4.560.086	74
35+737	1,49	EXTERNO	71,4	182,1	17/11/2011	01/11/17	333.059.110	4.558.291	73
35+737	1,49	EXTERNO	71,2	181,3	16/03/2012	01/11/17	316.411.262	4.583.456	69
35+737	1,49	EXTERNO	71,0	180,6	18/05/2012	01/11/17	307.444.631	4.588.726	67
35+737	1,49	EXTERNO	71,0	180,4	31/08/2012	01/11/17	293.550.670	4.591.512	64
35+737	1,49	EXTERNO	70,9	179,8	27/12/2012	01/11/17	274.621.532	4.587.219	60
35+737	1,49	EXTERNO	70,7	179,3	04/02/2013	01/11/17	265.983.666	4.601.794	58
35+737	1,49	EXTERNO	70,5	178,9	03/06/2013	01/11/17	248.781.777	4.624.197	54
35+737	1,49	EXTERNO	70,3	178,4	30/09/2013	01/11/17	234.543.122	4.623.058	51
35+737	1,49	EXTERNO	70,1	178,1	09/01/2014	01/11/17	216.516.462	4.639.638	47
35+737	1,49	EXTERNO	69,9	177,7	17/02/2014	01/11/17	212.011.626	4.645.982	46
35+737	1,49	EXTERNO	69,7	177,4	07/07/2014	01/11/17	189.910.741	4.673.767	41
35+737	1,49	EXTERNO	69,5	177,0	01/09/2014	01/11/17	180.461.617	4.679.212	39
35+737	1,49	EXTERNO	69,5	176,8	09/02/2015	01/11/17	157.457.687	4.704.911	33
35+737	1,49	EXTERNO	69,5	176,2	14/04/2015	01/11/17	148.911.323	4.727.344	32
35+737	1,49	EXTERNO	69,5	175,6	09/06/2015	01/11/17	139.438.906	4.732.090	29
35+737	1,49	EXTERNO	69,5	175,3	09/06/2015	01/11/17	139.438.906	4.732.090	29
35+737	1,49	EXTERNO	69,5	174,7	30/10/2015	01/11/17	119.614.783	4.709.243	25
35+737	1,49	EXTERNO	69,5	174,2	17/12/2015	01/11/17	110.143.769	4.713.713	23
35+737	1,49	EXTERNO	69,5	173,6	05/04/2016	01/11/17	92.982.971	4.817.771	19
35+737	1,49	EXTERNO	69,4	172,8	09/06/2016	01/11/17	83.613.465	4.842.479	17
35+737	1,49	EXTERNO	69,2	172,1	12/12/2016	01/11/17	55.246.524	4.947.450	11
35+737	1,49	EXTERNO	69,0	171,5	08/12/2016	01/11/17	55.246.524	4.947.450	11
35+737	1,49	EXTERNO	68,5	171,1	06/03/2017	01/11/17	41.886.341	5.128.940	8
35+737	1,49	EXTERNO	68,2	170,1	17/04/2017	01/11/17	37.152.613	5.208.310	7

Fonte: O autor (2019)

A obtenção da vida útil remanescente dos trilhos ferroviários foi calculada como sendo o período entre a data da leitura do Carro Controle EM-80 até a data de sua substituição, portanto uma variável contínua. Assim os 6 anos e 8 meses foram agrupados em 80 itens da categoria mês (6 anos e 8 meses vezes 12 meses/ano), 27 itens para a categoria trimestre (6 anos e 8 meses vezes 4 trimestres/ano), 14 itens

Como se pode notar na Figura 29, a distribuição encontra-se equilibrada a partir do item 3 até o item 66, entretanto, alguns itens dentro deste intervalo apresentam quantidades de registros inferiores a amostra do conjunto. A vida útil remanescente do trilho para os itens 1 e 2 assim como para os itens 67 ao 80 poderá propiciar resultados insatisfatórios devido apresentarem uma quantidade inferior de registros. Pode-se observar que as distribuições mais equilibradas ocorrem no intervalo entre os itens 43 a 61.

Figura 29 – Distribuição dos registros no conjunto de dados – categoria mês

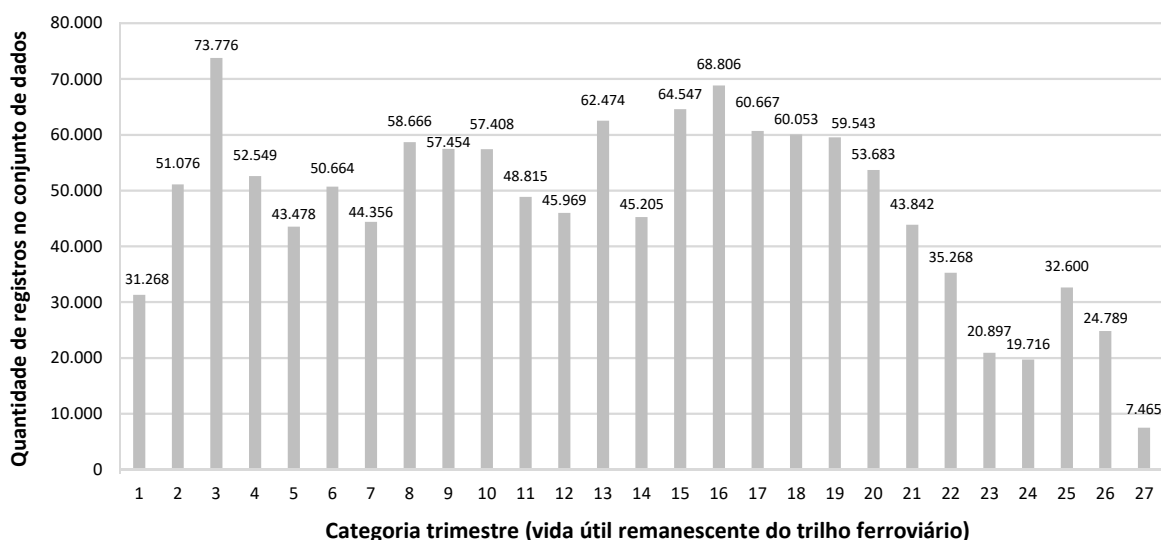


Fonte: O autor (2019)

Para a categoria trimestre, os 6 anos e 8 meses foram agrupados em 4 trimestres/ano. Assim, são apresentados 27 itens para a categoria trimestre, onde cada índice ou item da categoria trimestre representa 1 trimestre, portanto o índice 1 representa o primeiro trimestre, o índice 2 representa o segundo trimestre e assim sucessivamente até o índice 27 que representa o último trimestre para o período de 6 ano e 8 meses.

O gráfico da Figura 30 ilustra a categoria trimestre da vida útil remanescente para os trilhos da linha 1 da EFVM. Como se pode notar no gráfico da Figura 30, o comportamento da distribuição se manteve equilibrada a partir do item 4 até o item 22 para a categoria de período trimestre, período onde ocorreram as maiores quantidades de levantamento de dados e a realização das substituições dos trilhos da linha 1 da EFVM para o trecho de 29 km. Para os itens 23 até 27 apresentaram uma quantidade de registros inferiores a amostra do conjunto, que poderá propiciar resultados insatisfatórios para estes itens para esta categoria.

Figura 30 – Distribuição dos registros no conjunto de dados – categoria trimestre



Fonte: O autor (2019)

A Figura 31 apresenta a distribuição da categoria trimestre para os seus respectivos meses. Analisando a distribuição para a categoria trimestre, observa-se na Figura 31 que o item 27 é composto somente com 2 meses, o que representa um agrupamento inferior em relação aos demais conjuntos para a categoria trimestre. Por apresentar uma quantidade inferior de registros no item 27 isto poderá impactar nos resultados da RNA para este item na categoria trimestre. No gráfico da Figura 30 pode ser verificado que a quantidade de registros para este item da categoria trimestre

apresenta uma quantidade inferior de registros em relação aos demais itens para esta categoria.

Figura 31 – Distribuição da categoria trimestre para os seus respectivos meses

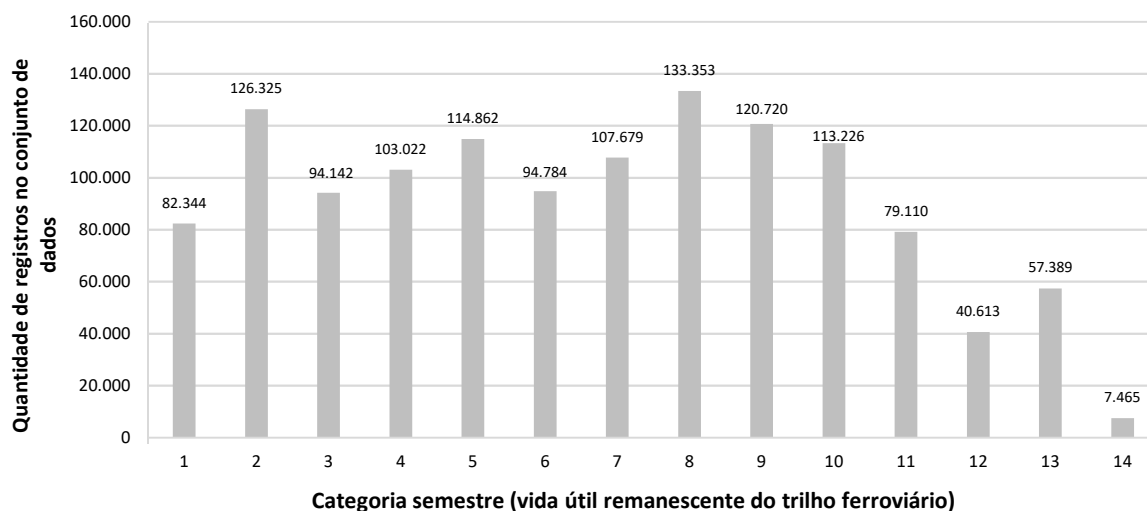
ANO	1º MÊS	2º MÊS	3º MÊS	4º MÊS	5º MÊS	6º MÊS	7º MÊS	8º MÊS	9º MÊS	10º MÊS	11º MÊS	12º MÊS
1	1		2		3		4		5		6	
2	5		6		7		8		9		10	
3	9		10		11		12		13		14	
4	13		14		15		16		17		18	
5	17		18		19		20		21		22	
6	21		22		23		24		25		26	
6 + 8 meses	25		26		27		28		29		30	

Fonte: O autor (2019)

Da mesma maneira que foi obtido para a categoria trimestre, foi realizada para a obtenção da categoria semestre, realizando-se por meio do agrupamento dos 6 anos e 8 meses para a categoria semestre, ou seja, por 2 semestres/ano. Assim são apresentados 14 itens para a categoria semestre, onde cada índice ou item da categoria semestre representa 1 semestre, portanto o índice 1 representa o primeiro semestre, o índice 2 representa o segundo semestre e assim sucessivamente até o índice 14 que representa o último semestre para o período de 6 anos e 8 meses.

O gráfico da Figura 32 ilustra a categoria semestre da vida útil remanescente para os trilhos da linha 1 da EFVM. Como se pode observar no gráfico ilustrado na Figura 32, o comportamento da distribuição se manteve equilibrada a partir do item 1 até o item 11 para a categoria de período semestre, período onde ocorreram as maiores quantidades de levantamento de dados e a realização das substituições dos trilhos da linha 1 da EFVM para o trecho de 29 km. Para os itens 12 até 14, estes apresentam uma quantidade de registros inferiores a amostra do conjunto, que poderá propiciar resultados insatisfatórios para estes itens para esta categoria.

Figura 32 – Distribuição dos registros no conjunto de dados – categoria semestre



Fonte: O autor (2019)

A Figura 33 apresenta a distribuição dos semestres e os seus respectivos meses. Analisando a distribuição para a categoria semestre, observa-se na Figura 33 que o item 14 é composto somente com 2 meses, o que representa um agrupamento inferior em relação aos demais conjuntos para a categoria semestre. Por apresentar uma quantidade inferior de registros no item 14 isto poderá impactar nos resultados da RNA para este item na categoria trimestre. No gráfico ilustrado na Figura 32 pode ser verificado que a quantidade de registros para este item da categoria semestre apresenta uma quantidade inferior de registros em relação aos demais itens para esta categoria.

Figura 33 – Distribuição da categoria semestre para os seus respectivos meses

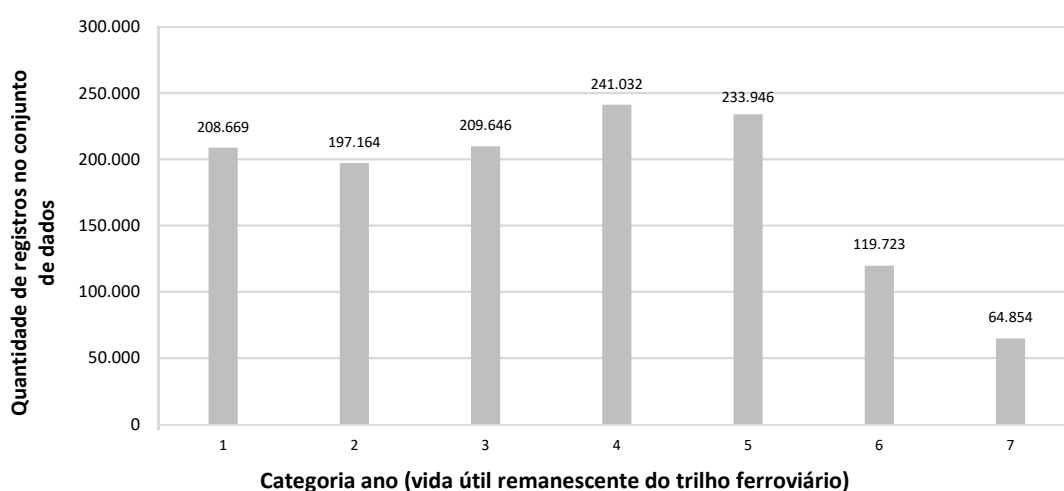
ANO	1º MÊS	2º MÊS	3º MÊS	4º MÊS	5º MÊS	6º MÊS	7º MÊS	8º MÊS	9º MÊS	10º MÊS	11º MÊS	12º MÊS
1	1						2					
2	3						4					
3	5						6					
4	7						8					
5	9						10					
6	11						12					
6 + 8 meses	13						14					

Fonte: O autor (2019)

Para a categoria ano, os 6 anos e 8 meses foram agrupados por 1 ano/ano. Assim são apresentados os 7 itens da categoria ano. Assim, cada índice ou item da categoria ano representa 1 ano, o índice 1 representa um ano, o índice 2 representa dois anos, o índice 3 representa três anos e assim sucessivamente até o índice 7 que representa o último ano, para o período de 6 ano e 8 meses.

O gráfico da Figura 34 ilustra a categoria ano da vida útil remanescente para os trilhos da linha 1 da EFVM. Como se pode observar no gráfico ilustrado na Figura 34, o comportamento da distribuição se manteve equilibrada a partir do item 1 até o item 5 para a categoria de período ano, período onde ocorreram as maiores quantidades de levantamento de dados e a realização das substituições dos trilhos da linha 1 da EFVM para o trecho de 29 km. Para os itens 6 e 7 estes apresentam uma quantidade de registros inferiores a amostra do conjunto, principalmente para o item 7 que poderá propiciar resultados insatisfatórios para este item para esta categoria.

Figura 34 – Distribuição dos registros no conjunto de dados – categoria ano



Fonte: O autor (2019)

A Figura 35 apresenta a distribuição da categoria ano para os seus respectivos meses. Analisando a distribuição para a categoria ano, observa-se na Figura 35 que o item 7, é composto somente com 8 meses, o que representa um agrupamento inferior em relação aos demais conjuntos para a categoria ano. Por apresentar uma quantidade inferior de registros no item 7 isto poderá impactar nos resultados da RNA para a categoria ano. No gráfico ilustrado na Figura 34 pode ser verificado que a quantidade de registros para este item da categoria ano apresenta uma quantidade inferior de registros em relação aos demais itens para esta categoria.

Figura 35 – Distribuição da categoria ano para os seus respectivos meses

ANO	1º MÊS	2º MÊS	3º MÊS	4º MÊS	5º MÊS	6º MÊS	7º MÊS	8º MÊS	9º MÊS	10º MÊS	11º MÊS	12º MÊS
1							1					
2							2					
3							3					
4							4					
5							5					
6							6					
6 + 8meses				7								

Fonte: O autor (2019)

Portanto, com as categorias de períodos definidas em: 1) mês; 2) trimestre; 3) semestre e 4) ano para a estimação da vida útil remanescente dos trilhos ferroviários, parte-se para o processamento das configurações da RNA e do k-NN investigados para cada uma destas categorias de períodos. Nos capítulos seguintes serão apresentadas as configurações da RNA e do k-NN testados para esta dissertação.

5 REDE NEURAL ARTIFICIAL DE CLASSIFICAÇÃO PARA A ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DO TRILHO FERROVIÁRIO

Para essa dissertação foram investigadas e testadas várias configurações para definir qual a configuração da RNA de classificação, que melhor realiza a estimação das categorias de período da vida útil remanescente dos trilhos ferroviários. Para essas configurações foi utilizando o neurônio artificial *Perceptron* e o algoritmo de treinamento *Backpropagation*. Utilizou nessa dissertação a estrutura de MLP devido ao fato de que em trabalhos desenvolvidos no campo da estimação em diversas áreas, e que utilizaram este tipo de estrutura para a classificação, obtiveram resultados satisfatórios (WOLDEMARIAM, MURILLO-HOYOS e LABI, 2015). Assim, para esta dissertação utilizou-se este tipo de estrutura para as configurações da RNA.

Para a simulação da RNA foi usado *Python* versão 3.7 no ambiente *Spyder* versão 3.2.6. A biblioteca para o desenvolvimento das configurações da RNA foi o *Scikit-Learn* versão 0.20.3 que é um código aberto para a linguagem de programação *Python*.

As configurações das RNA investigadas, tiveram como objetivo a estimação das categorias de períodos da vida útil remanescente em trilhos ferroviários. Assim sendo, foram testadas e analisadas as seguintes categorias de períodos de vida útil remanescente: 1) mês; 2) trimestre; 3) semestre; e 4) ano.

A categoria de período mês e trimestre, foram testadas, mas, geralmente as ferrovias realizam seus planejamentos da substituição dos trilhos em períodos maiores, como semestre e ano, adotando-se o período de categoria ano para esses planejamentos.

A RNA, independentemente da categoria para a estimação da vida útil remanescente dos trilhos ferroviários, tiveram 5 variáveis na camada de entrada, ou seja, 5 neurônios na camada de entrada que foram: 1) grau da curva, sendo estes dados inseridos em graus; 2) trilho interno ou externo a curva, inseridos em valores adimensionais, sendo 0 para a tangente, 1 para curva interna e 2 para curva externa; 3) largura do boleto, inserido em milímetros; 4) altura do boleto, inserido em milímetros; e 5) peso médio

transportado, inserido em toneladas. Todos os dados utilizados na camada de entrada foram levantados, por meio de base de dados históricos da substituição dos trilhos e de dados históricos das leituras realizadas pelo Carro Controle EM-80 da ferrovia EFVM para o período de 6 anos e 8 meses, citados anteriormente.

Devido as entradas dos dados apresentarem grandezas numéricas distintas, ou seja, valores apresentados em centenas e dezenas que referem-se aos desgastes dos trilhos e valores apresentados na classe de milhões que referem-se ao peso médio transportado, foi necessária a normalização destes valores. Esta normalização dos valores teve como objetivo a realização de uma aproximação numérica entre os dados de entrada. Então, para que fossem evitados resultados insatisfatórios, aplicou-se em todos os dados de entrada a função *scale* do *Scikit-Learn*. A função *scale* do *Scikit-Learn* tem como objetivo centralizar os valores em zero ou com uma variação na mesma ordem (SCIKIT-LEARN, 2019).

A camada intermediária, independentemente da categoria para a estimação da vida útil remanescente dos trilhos ferroviários, foi testada variando-se a quantidade de camadas intermediária e a quantidade de neurônios nestas camadas intermediária de forma iterativa. Assim, foram testadas e analisadas configurações com 1, 2 e 3 camadas intermediárias. Estabeleceu-se que a RNA teria no máximo 3 camadas e que em todas as camadas sempre teriam a mesma quantidade de neurônios. Logo, testou-se as camadas intermediária com: 30, 50, 100, 200 e 400 neurônios para as configurações da RNA para as categorias trimestre, semestre e ano, e 100, 200 e 400 neurônios para as configurações da RNA para a categoria mês. Perfazendo um total de 54 configurações de RNA para serem analisadas e avaliadas nessa dissertação. Os acréscimos nas quantidades de camadas intermediárias, e na quantidade de neurônios nestas camadas intermediárias, foi realizada para se conseguir obter uma configuração da RNA que apresentasse o melhor desempenho para a categorização de períodos de vida útil remanescente dos trilhos ferroviários.

Após os testes preliminares, detectou-se que aumentando a quantidade de camadas intermediárias além de 3 camadas com 400 neurônios, não foi obtida melhora no valor da acurácia e do *F1-score* e, em contrapartida, o tempo de processamento aumentava significativamente.

A menor camada intermediária testada para a estimação das categorias de período trimestral, semestral e ano tem 30 neurônios, e para a categoria de período mês tem 100 neurônios, uma vez que, a camada intermediária deve ser igual ou maior que a quantidade de neurônios apresentados para a camada de saída. Os demais valores para o tamanho da camada intermediária foram determinados de forma aleatória.

A quantidade de neurônios na camada de saída, representa a categoria de período da vida útil remanescente dos trilhos ferroviários, que depende da categoria do período: 1) mês; 2) trimestre; 3) semestre; e 4) ano. Assim, a quantidade de neurônios para a camada de saída sofreu variação, dependendo da categoria do período, sendo que a categoria de período mês possui 80 neurônios, a categoria trimestre possui 27 neurônios, a categoria semestre possui 14 neurônios e a categoria ano possui 7 neurônios.

Para as RNA, adotou-se a validação cruzada *k-fold*, sendo utilizado $k=10$, onde os registros do conjunto de dados para o treinamento e validações foram seccionados aleatoriamente em 10 partes iguais. Destas partes, uma única parte ficou retida como validação e as outras 9 partes restantes foram utilizadas como treinamento. O processo foi repetido 10 vezes em cada uma das 10 partes, alternando-se as partes em treinamento e validação. Assim, cada uma destas partes de treinamento e validação obteve um resultado da acurácia e de *F1-score*, portanto, para cada configuração obteve-se 10 resultados de acurácia e de *F1-score*.

Foram testadas e analisadas as 54 configurações da RNA, que são apresentadas em grupos e subgrupos, sendo que os grupos se referem as configurações elaboradas para categorizar o período de vida útil dos trilhos ferroviários em: 1) mês (grupo G1); 2) trimestre (grupo G2); 3) semestres (grupo G3) e 4) ano (grupo G4). Os subgrupos referem-se à quantidade de camadas intermediária e a quantidade de neurônios nestas camadas para as configurações do mesmo grupo. As Tabela 13, Tabela 14, Tabela 15 e Tabela 16 apresentam as configurações das RNA investigadas para cada categoria de períodos.

Tabela 13 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – mês

Grupo	Subgrupo	Configuração	Quantidades de Neurônios na Camada de Entrada	Quantidades de Camadas Intermediárias	Quantidades de Neurônios nas Camadas Intermediárias	Quantidades de Neurônios na Camada de Saída
G1	G1.1	1	5	1	100	80
		2		2	100	
		3		3	100	
	G1.2	4	5	1	200	80
		5		2	200	
		6		3	200	
	G1.3	7	5	1	400	80
		8		2	400	
		9		3	400	

Fonte: O autor (2019)

Tabela 14 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – trimestre

Grupo	Subgrupo	Configuração	Quantidades de Neurônios na Camada de Entrada	Quantidades de Camadas Intermediárias	Quantidades de Neurônios nas Camadas Intermediárias	Quantidades de Neurônios na Camada de Saída
G2	G2.1	10	5	1	30	27
		11		2	30	
		12		3	30	
	G2.2	13	5	1	50	27
		14		2	50	
		15		3	50	
	G2.3	16	5	1	100	27
		17		2	100	
		18		3	100	
	G2.4	19	5	1	200	27
		20		2	200	
		21		3	200	
	G2.5	22	5	1	400	27
		23		2	400	
		24		3	400	

Fonte: O autor (2019)

Tabela 15 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – semestre

Grupo	Subgrupo	Configuração	Quantidades de Neurônios na Camada de Entrada	Quantidades de Camadas Intermediárias	Quantidades de Neurônios nas Camadas Intermediárias	Quantidades de Neurônios na Camada de Saída
G3	G3.1	25	5	1	30	14
		26		2	30	
		27		3	30	
	G3.2	28	5	1	50	14
		29		2	50	
		30		3	50	
	G3.3	31	5	1	100	14
		32		2	100	
		33		3	100	
	G3.4	34	5	1	200	14
		35		2	200	
		36		3	200	
	G3.5	37	5	1	400	14
		38		2	400	
		39		3	400	

Fonte: O autor (2019)

Tabela 16 – Configurações da RNA para a categoria de período da vida útil remanescente dos trilhos ferroviários – ano

Grupo	Subgrupo	Configuração	Quantidades de Neurônios na Camada de Entrada	Quantidades de Camadas Intermediárias	Quantidades de Neurônios nas Camadas Intermediárias	Quantidades de Neurônios na Camada de Saída
G4	G4.1	40	5	1	30	7
		41		2	30	
		42		3	30	
	G4.2	43	5	1	50	7
		44		2	50	
		45		3	50	
	G4.3	46	5	1	100	7
		47		2	100	
		48		3	100	
	G4.4	49	5	1	200	7
		50		2	200	
		51		3	200	
	G4.5	52	5	1	400	7
		53		2	400	
		54		3	400	

Fonte: O autor (2019)

Para a realização do treinamento, validação e teste das configurações da RNA foram utilizados a quantidade de 1.275.034 registros. Sendo que 1.100.000 registros foram utilizados para a realização do treinamento e validação das configurações da RNA, uma vez que, na validação cruzada *k-fold*, o valor adotado foi de $f=10$. Logo, essa quantidade deve ser uma divisão múltipla do valor de f para que se possa obter resultados satisfatórios na RNA.

O restante, 175.034 registros foram utilizados no teste das configurações da RNA escolhidas, para cada um dos períodos de categoria, 1) mês; 2) trimestre; 3) semestre e 4) ano.

A função de ativação utilizada para a RNA foi a função *ReLU*, o *solver* utilizado para a RNA foi o *Adam*, a quantidade máxima de iterações foi de 500, a tolerância para a otimização foi de 1×10^{-7} e a programação da taxa de aprendizado para atualizações de peso foi de 0,01.

Para escolher a configuração da RNA que apresentou o melhor desempenho no treinamento e validação, e que será utilizada na fase de teste, para cada um dos períodos de categorias, seguiu 2 etapas.

A 1ª etapa, foi realizada comparando os resultados das acurácias e dos *F1-score* entre as diversas configurações validadas. As configurações que apresentaram os melhores resultados nos valores da acurácia e *F1-score*, para cada categoria de período, esta foi a configuração escolhida para ser testada. Assim, comparou-se a configuração composta com 1 camada intermediária com 30 neurônios nesta camada, com a configuração composta com 2 camadas intermediárias com 30 neurônios nesta camada e assim respectivamente, dentre todas as configurações, entre as diversas camadas intermediárias e quantidades de neurônios nestas camadas. Desta maneira, buscou-se as configurações que apresentaram o melhor desempenho para categorizar cada um dos períodos: mês, trimestre; semestre; e ano.

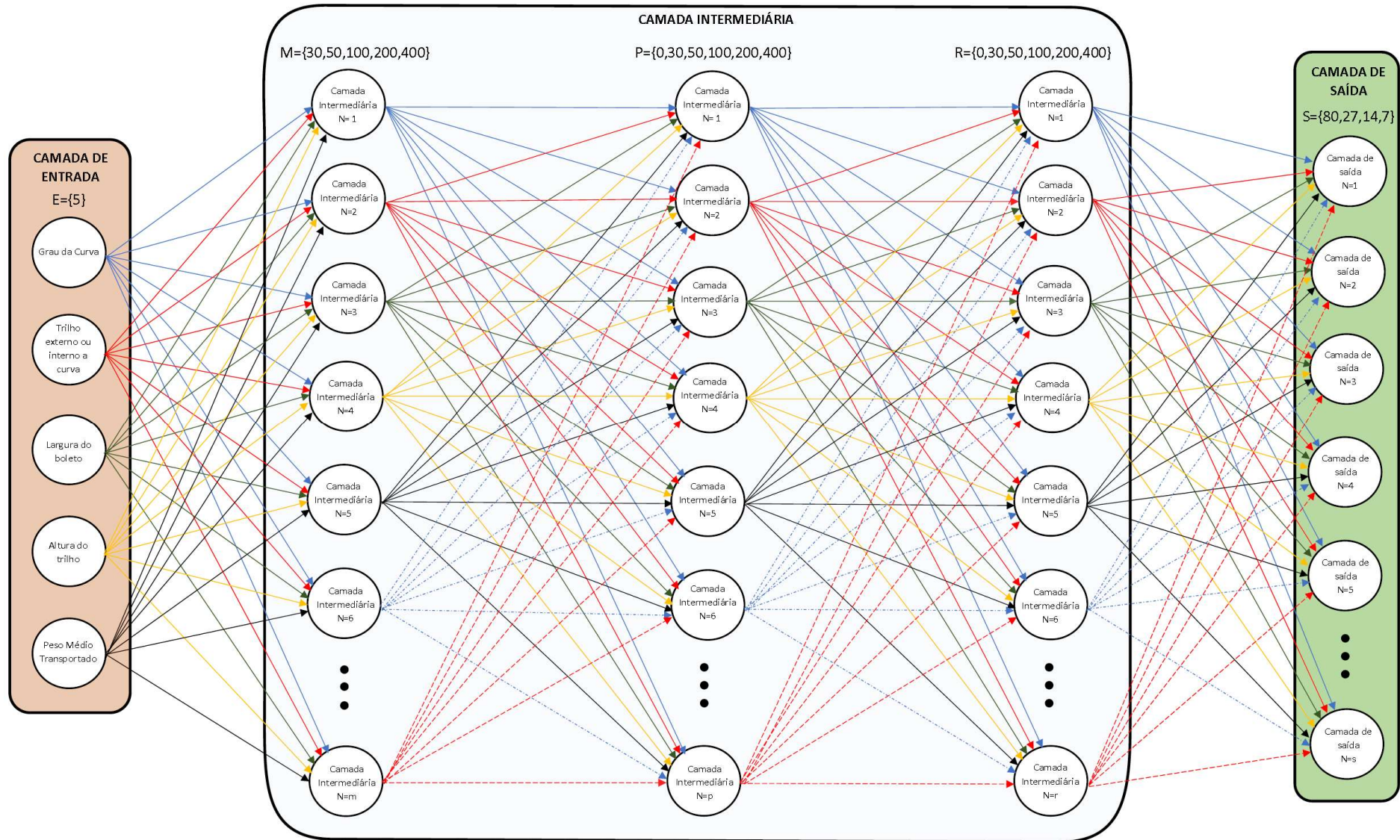
Escolhidas as configurações que apresentaram os melhores resultados no valor da acurácia e *F1-score*, para categorizar cada um dos períodos: mês, trimestre, semestre e ano, foi realizada a 2ª etapa, a qual buscou-se nestas configurações dentre as 10 partes seccionadas pelo *k-fold* para a realização do treinamento e validação, a parte que apresentou um valor médio na acurácia e no *F1-score*. Esta parte é que dispõem

dos melhores pesos para generalização da RNA. Desta forma, esta parte foi utilizada para o teste das configurações da RNA para categorizar os períodos da vida útil remanescente em trilhos ferroviários em categorias de períodos: mês, trimestre, semestre e ano.

Portanto, definidos quais seriam as variáveis utilizadas para a camada de entrada, a quantidade de camadas intermediárias e suas quantidades de neurônios e a quantidade de neurônios na camada de saída, representando as categorias de período da vida útil remanescente dos trilhos ferroviários, foi realizada as simulações.

A Figura 36 apresenta a RNA investigada para categorizar o período da vida útil remanescente dos trilhos ferroviários. Cabe ressaltar que em configurações que apresentam somente uma camada intermediária, as demais camadas 2 e 3 não existem, assim os valores zeros apresentados nos conjuntos P e R representam a inexistência da camada.

Figura 36 – RNA investigadas para estimação da vida útil remanescente de trilhos ferroviários



Fonte: O autor (2019)

6 K-VIZINHOS MAIS PRÓXIMOS (K-NN) DE CLASSIFICAÇÃO PARA A ESTIMAÇÃO DA VIDA ÚTIL REMANESCENTE DO TRILHO FERROVIÁRIO

Para se avaliar a qualidade dos resultados obtidos nas configurações da RNA utilizou-se o método de classificação K-Vizinhos mais próximos (k-NN). O objetivo de se utilizar o k-NN era de se verificar quais seriam os resultados obtidos com o mesmo conjunto de dados utilizados nas configurações da RNA no treinamento, validação e teste.

O algoritmo k-NN usou a implementação em *Python* versão 3.7 no ambiente *Spyder* versão 3.2.6. A biblioteca para o desenvolvimento do k-NN foi o *Scikit-Learn* versão 0.20.3 que é um código aberto para a linguagem de programação *Python*.

Da mesma forma que para as configurações da RNA de classificação para a estimação das categorias de períodos da vida útil. O k-NN teve como objetivo a realização da classificação na estimação das categorias de períodos: 1) mês, 2) trimestre; 3) semestres e 4) ano.

Para o algoritmo do k-NN, que implementa o aprendizado baseado nos vizinhos mais próximos de cada ponto da análise, sendo o valor de k um valor inteiro definido pelo programador. A classificação dos vizinhos mais próximos é uma técnica difundida, entretanto, esta técnica é dependente da escolha do valor de k -vizinhos mais próximos, e dos dados disponíveis para obtenção de resultados satisfatórios. Os valores do k -vizinhos mais próximos utilizados para o k-NN foram $k = 5$, $k = 7$ e $k = 9$.

Foram testadas e analisadas configurações do k-NN com $k= 5$, $k=7$ e $k=9$ para cada categoria de períodos: 1) mês; 2) trimestre; 3) semestre e 4) ano. Perfazendo um total de 12 configurações do k-NN.

Utilizou-se a função *scale* do *Scikit-Learn* com o objetivo de centralizar os valores em zero ou com uma variação na mesma ordem, da mesma forma realizada para as configurações da RNA.

Para as configurações do k-NN, adotou-se a validação cruzada, *k-fold*, sendo utilizado $f=10$, onde os registros do conjunto de dados para o treinamento e validação foram seccionados aleatoriamente em 10 partes iguais. Destas partes, uma única parte ficou retida como validação e as outras 9 partes restantes foram utilizadas como treinamento. O processo foi repetido 10 vezes em cada uma das 10 partes, alternando-se as partes em treinamento e validação. Assim, cada uma destas partes de treinamento e validação obteve um resultado da acurácia e de *F1-score*, portanto, para cada configuração obteve-se 10 resultados de acurácia e de *F1-score*.

Foi utilizado para treinamento e validação das configurações do k-NN a quantidade de 1.100.000 registros, os mesmos utilizados para o treinamento e validações da RNA, e para o teste das configurações do k-NN a quantidade de 175.034 registros, os mesmos utilizados para o teste da RNA.

O k-NN utilizou os seguintes parâmetros, sendo a função de definição do peso uniforme, o algoritmo usado para calcular os vizinhos mais próximos foi o padrão definido pelo *Scikit-Learn* que foi o auto, e a métrica padrão utilizada foi a métrica euclidiana padrão.

Para se analisar o desempenho das configurações do k-NN para cada uma das categorias de períodos: mês, trimestre, semestre e ano utilizou-se os valores médios dos resultados obtidos para as acurácias e para os *F1-score*, da mesma forma como foram realizadas para a RNA. A Tabela 17 apresenta as configurações do k-NN investigados para cada uma das categorias de períodos.

Tabela 17 – Configurações do k-NN para cada categoria de período

Configuração	Categoria de Período	K	Configuração	Categoria de Período	K
55	Mês	5	61	Semestre	5
56		7	62		7
57		9	63		9
58	Trimestre	5	64	Ano	5
59		7	65		7
60		9	66		9

Fonte: O autor (2019)

7 RESULTADOS EXPERIMENTAIS

Este capítulo apresenta os resultados experimentais obtidos para as configurações da RNA e para as configurações do k-NN, bem como uma análise comparativa destes valores obtidos para cada uma destas configurações. Além disso, é apresentada uma análise geral de todos os resultados experimentais obtidos.

7.1 RESULTADOS E ANÁLISES DA REDE NEURAL ARTIFICIAL (RNA)

Com as configurações da RNA e os parâmetros de processamento definidos, e com o conjunto de dados devidamente tratado, realizou-se o processamento destas configurações para cada categoria de períodos. Foi utilizado para o processamento de todas as configurações de RNA e do k-NN um computador com processador Intel i7 com 16 GB de memória RAM.

Para o processamento do treinamento e validação da RNA foram utilizados a quantidade de 1.100.000 registros. Do processamento destes registros para cada uma das configurações da RNA foram obtidos os valores dos resultados das acurácias e dos *F1-score*.

Os resultados das configurações são apresentados em grupos e subgrupos, sendo que os grupos referentes as configurações elaboradas para categorizar o período de vida útil dos trilhos ferroviários em: 1) mês (grupo G1); 2) trimestre (grupo G2); 3) semestres (grupo G3) e 4) ano (grupo G4). Os subgrupos referentes à quantidade de camadas intermediária e a quantidade de neurônios nestas camadas para as configurações do mesmo grupo. Os grupos e subgrupos com seus respectivos resultados, são apresentados nas Tabela 18, Tabela 19, Tabela 20 e Tabela 21. Nestas tabelas são mostrados os grupos e subgrupos, os nomes das configurações, a quantidade de camadas intermediárias e quantidade de neurônios nestas camadas, resultados obtidos para a maior acurácia, a menor acurácia e a acurácia média, assim como os resultados obtidos para o maior *F1-score*, o menor *F1-score* e o *F1-score*

médio e o tempo de processamento para a realização do treinamento e validação com a divisão do $f=10$.

Com base nestas tabelas, comparou-se as configurações de cada grupo para verificar os desempenhos obtidos nos valores das acurácias e dos *F1-score*. Tomando como base os resultados apresentados nas Tabela 18, Tabela 19, Tabela 20 e Tabela 21, respectivamente, pode se concluir que as Configurações 9, 24, 39 e 54, foram as que apresentaram os melhores desempenhos para os grupos G1, G2, G3 e G4 respectivamente.

Tabela 18 – Resultados de treinamento/validação da categoria de período mês – grupo G1

Grupo	Subgrupo	Configuração	Quantidades de Camadas Intermediárias	Quantidades de Neurônios nas Camadas Intermediárias	Maior acurácia	Menor acurácia	Média das acurácias	Maior F1-score	Menor F1-score	Média F1-score	Tempo de Processamento para 10 k-folds (min)
G1	G1.1	1	1	100	0,48535	0,42990	0,46238	0,48793	0,41730	0,45774	26,11
		2	2	100	0,69856	0,57871	0,65103	0,69181	0,53758	0,63510	98,22
		3	3	100	0,73595	0,60423	0,69198	0,71883	0,57192	0,67018	109,36
	G1.2	4	1	200	0,55423	0,50352	0,52693	0,55128	0,49622	0,52174	80,11
		5	2	200	0,75458	0,58106	0,69354	0,74766	0,53193	0,67209	247,05
		6	3	200	0,78743	0,62311	0,72005	0,76687	0,59835	0,69870	314,35
	G1.3	7	1	400	0,60459	0,54442	0,58297	0,59466	0,53336	0,57304	265,14
		8	2	400	0,76113	0,61413	0,71714	0,75217	0,57411	0,69798	590,02
		9	3	400	0,79485	0,57710	0,72244	0,77314	0,52728	0,69657	607,18

Fonte: O autor (2019)

Tabela 19 – Resultados de treinamento/validação da categoria de período trimestre – grupo G2

Grupo	Subgrupo	Configuração	Quantidades de Camadas Intermediárias	Quantidades de Neurônios nas Camadas Intermediárias	Maior acurácia	Menor acurácia	Média das acurácias	Maior F1-score	Menor F1-score	Média F1-score	Tempo de Processamento para 10 k-folds (min)
G2	G2.1	10	1	30	0,47398	0,42575	0,44609	0,47288	0,41602	0,43903	10,25
		11	2	30	0,60995	0,53358	0,57429	0,60372	0,53156	0,56817	18,12
		12	3	30	0,64694	0,56739	0,61959	0,63942	0,53677	0,60516	23,35
	G2.2	13	1	50	0,50346	0,45182	0,48352	0,49726	0,44666	0,48057	11,05
		14	2	50	0,67933	0,59975	0,62966	0,67330	0,57476	0,62269	25,12
		15	3	50	0,73572	0,65755	0,69774	0,72998	0,64058	0,68815	40,45
	G2.3	16	1	100	0,58233	0,54200	0,55591	0,58112	0,53515	0,55242	15,25
		17	2	100	0,73416	0,66031	0,70593	0,72541	0,64686	0,69619	68,04
		18	3	100	0,78824	0,68346	0,75320	0,77719	0,67050	0,73962	105,45
	G2.4	19	1	200	0,62558	0,57306	0,60342	0,61927	0,56512	0,59760	28,23
		20	2	200	0,76654	0,66859	0,74557	0,76152	0,61985	0,72924	208,56
		21	3	200	0,81038	0,70685	0,77128	0,79640	0,67862	0,75585	185,24
	G2.5	22	1	400	0,65284	0,61652	0,63806	0,64740	0,61238	0,63031	80,33
		23	2	400	0,76259	0,70175	0,74627	0,75376	0,66265	0,73377	460,12
		24	3	400	0,81870	0,70196	0,77762	0,80540	0,80540	0,75937	678,28

Fonte: O autor (2019)

Tabela 20 – Resultados de treinamento/validação da categoria de período semestre – grupo G3

Grupo	Subgrupo	Configuração	Quantidades de Camadas Intermediárias	Quantidades de Neurônios nas Camadas Intermediárias	Maior acurácia	Menor acurácia	Média das acurácias	Maior F1-score	Menor F1-score	Média F1-score	Tempo de Processamento para 10 k-folds (min)
G3	G3.1	25	1	30	0,59597	0,54555	0,57132	0,59534	0,55196	0,56881	8,25
		26	2	30	0,71113	0,65708	0,67974	0,71433	0,65035	0,67459	12,13
		27	3	30	0,74851	0,69465	0,72521	0,74602	0,68606	0,71953	16,47
	G3.2	28	1	50	0,63978	0,58201	0,61320	0,64088	0,58464	0,61201	8,34
		29	2	50	0,76074	0,71823	0,73770	0,75710	0,70984	0,73528	23,45
		30	3	50	0,79274	0,74513	0,77270	0,78837	0,73346	0,76617	32,22
	G3.3	31	1	100	0,68115	0,63999	0,67046	0,68256	0,63832	0,66811	11,13
		32	2	100	0,79243	0,74360	0,77393	0,78956	0,73133	0,76902	47,06
		33	3	100	0,81554	0,75727	0,79864	0,80612	0,73629	0,79071	73,15
	G3.4	34	1	200	0,71938	0,67856	0,69794	0,71899	0,67517	0,69755	24,33
		35	2	200	0,82417	0,76264	0,79520	0,81944	0,75515	0,78923	130,10
		36	3	200	0,84021	0,79091	0,82080	0,83272	0,77124	0,81256	174,08
	G3.5	37	1	400	0,72241	0,69204	0,70998	0,72397	0,68489	0,70960	47,34
		38	2	400	0,83183	0,77302	0,80534	0,82728	0,76721	0,80046	419,08
		39	3	400	0,87352	0,79410	0,84284	0,84851	0,74971	0,81482	619,02

Fonte: O autor (2019)

Tabela 21 – Resultados de treinamento/validação da categoria de período ano – grupo G4

Grupo	Subgrupo	Configuração	Quantidades de Camadas Intermediárias	Quantidades de Neurônios nas Camadas Intermediárias	Maior acurácia	Menor acurácia	Média das acurácias	Maior F1-score	Menor F1-score	Média F1-score	Tempo de Processamento para 10 k-folds (min)
G4	G4.1	40	1	30	0,76321	0,71716	0,74047	0,73449	0,64753	0,69733	6,15
		41	2	30	0,84012	0,80849	0,82416	0,80991	0,76530	0,78826	14,23
		42	3	30	0,86939	0,82038	0,84974	0,82868	0,77328	0,79992	24,37
	G4.2	43	1	50	0,79286	0,75403	0,77181	0,76200	0,70027	0,73167	8,02
		44	2	50	0,87406	0,83233	0,85678	0,84787	0,80425	0,82264	20,15
		45	3	50	0,89257	0,84447	0,87929	0,85427	0,80521	0,83731	27,45
	G4.3	46	1	100	0,81664	0,80162	0,80880	0,78569	0,74258	0,76910	10,28
		47	2	100	0,89181	0,87017	0,88384	0,86202	0,83607	0,84854	52,45
		48	3	100	0,90837	0,86973	0,89226	0,87393	0,82944	0,85384	58,23
	G4.4	49	1	200	0,82952	0,80846	0,81981	0,79927	0,75588	0,78052	19,34
		50	2	200	0,90010	0,87965	0,88765	0,86654	0,84123	0,85321	128,11
		51	3	200	0,91535	0,86846	0,89789	0,88056	0,83403	0,86109	163,48
	G4.5	52	1	400	0,83660	0,81282	0,82338	0,80264	0,78290	0,79114	30,52
		53	2	400	0,89827	0,87327	0,88545	0,86908	0,82063	0,84667	381,12
		54	3	400	0,90259	0,85661	0,88961	0,89292	0,83774	0,86741	553,13

Fonte: O autor (2019)

Independentemente do período de categoria da vida útil remanescente, observou-se que as configurações que apresentaram o melhor desempenho, ou seja, que apresentaram a melhor acurácia e melhor *F1-score*, foram as configurações compostas de 3 camadas intermediária com 400 neurônios. Dentre estas configurações, destaca-se a Configuração 54 que apresentou o melhor desempenho para a estimação da categoria de período da vida útil remanescente do trilho ferroviário.

A Configuração 54 referente a categoria ano, composta com 3 camadas com 400 neurônios foi a que apresentou o melhor resultado de acurácia e de *F1-score* entre todas. Em função da categoria ano apresentar o menor número de itens, sete, tem-se uma maior probabilidade de acerto na estimação para cada item. Em contrapartida, para a categoria mês, que apresenta o maior número de itens, 80, percebeu-se que o desempenho foi inferior ao da categoria ano. Analogamente, esta mesma análise pode ser feita para trimestre e semestre.

Os pesos referentes aos neurônios da camada intermediária para as Configurações 9, 24, 39 e 54 foram salvos em arquivos, que serão lidos na fase de teste. Nas Configurações 9, 24, 39 e 54, buscou-se dentre as 10 partes seccionadas pelo *k-fold* na fase de treinamento e validação, a parte que apresentou o valor médio de acurácia e o valor médio de *F1-score*. Isto se deve ao fato que os valores médios tendem a representar uma maior capacidade de generalização da RNA. Assim, utilizou-se os pesos desta parte para o teste destas configurações da RNA para categorizar o período da vida útil remanescente nos trilhos. Assim, no subcapítulo seguinte será apresentado o procedimento de teste das configurações da RNA para cada categoria de período.

7.2 TESTE DA REDE NEURAL ARTIFICIAL (RNA)

O teste das Configurações 9, 24, 39 e 54, respectivamente, mês, trimestre, semestre e ano, foi realizada por meio de processamento com o conjunto de dados de teste composto com 175.034 registros utilizando os pesos salvos da configuração da RNA salvos durante a fase de treinamento e validação. Ressalta-se que os registros do conjunto de dados de teste não foram apresentados para as configurações da RNA durante a fase de treinamento e validação. O teste foi realizado analisando os valores da acurácia e de *F1-score* que indicam o desempenho destas configurações para a estimação da vida útil remanescente dos trilhos ferroviários por meio de categorias de períodos.

A Tabela 22 apresenta as configurações escolhidas para serem utilizadas para o teste, e os valores das acurácias e do *F1-score* obtidos na fase de treinamento e validação

e os resultados obtidos na fase de teste para cada uma das categorias de período: mês, trimestre, semestre e ano, respectivamente, para as Configurações 9, 24, 39 e 54. Nesta tabela, pode-se observar que os resultados da acurácia e do *F1-score* para o conjunto de dados de teste apresentaram valores de acurácia e de *F1-score* melhores que os resultados obtidos na fase de treinamento e validação, indicando assim que estas configurações possuem a capacidade de generalização para a estimação com novos registros.

Outro ponto que pode ser observado é que todas as quatro configurações respectivamente, apresentaram na fase de teste acurácia entre 0,76951 e 0,88936 e *F1-score* entre 0,75312 e 0,87983, o que representa um valor bom considerando que se está realizando a estimação para um período de 6 anos e 8 meses.

Tabela 22 – Resultados das configurações da RNA na fase de treinamento/validação e teste

Configuração	Categoria de Período	Acurácia (treinamento/validação)	F1-score (treinamento/validação)	Acurácia (teste)	F1-score (teste)
9	Mês	0,69198	0,67018	0,76951	0,75312
24	Trimestre	0,77762	0,75937	0,81458	0,80117
39	Semestre	0,84284	0,81482	0,84961	0,82510
54	Ano	0,88961	0,86741	0,88936	0,87983

Fonte: O autor (2019)

Com base na Tabela 23 pode-se perceber que os acertos totais foram: 134.690 para a categoria mês, 76,95% do total da amostra de teste; 142.579 para a categoria trimestre, 81,46% do total da amostra de teste; 148.710 para a categoria semestre, 84,96% do total da amostra de teste; e 155.669 para a categoria ano, 88,94% do total da amostra de teste. Esses valores mostram um nível de acerto bem razoável considerando-se uma estimação para 6 anos e 8 meses.

Para corroborar com a análise do parágrafo anterior, analisando-se as matrizes de confusão das configurações apresentadas na Figura 37, Figura 38, Figura 39 e Figura 40 que representam respectivamente as categorias de período: semestre e ano, observa-se que os resultados foram concentrados na diagonal principal, o que é o mais desejado, pois a diagonal representa a quantidade de acertos totais da estimação. Decidiu-se não apresentar a matriz de confusão referente a categoria mês

e trimestre pois elas possuem 80 itens e 27 itens nas duas dimensões respectivamente, o que impossibilita a sua visualização.

Os valores acima desta diagonal representam o falso positivo, ou seja, a estimação indica uma substituição do trilho após a categoria de período que ele efetivamente deveria ter sido substituído. Este tipo de erro pode ocasionar a possibilidade de quebra do trilho. Esta é uma situação perigosa para a ferrovia, pois pode levar a acidentes ferroviários caso não ocorra a substituição do trilho na via férrea. Os valores abaixo desta diagonal representam o falso negativo, ou seja, a estimação indica a necessidade da substituição do trilho antes da categoria de período que ele efetivamente deveria ter sido substituído. Este tipo de erro pode ocasionar um impacto financeiro pelas trocas antecipadas dos trilhos.

Em relação aos falsos positivos, tem-se para cada categoria os seguintes valores: 20.927 para a categoria mês, 11,96% do total da amostra de teste; 17.876 para a categoria trimestre, 10,21% do total da amostra de teste; 15.435 para a categoria semestre, 8,82% do total da amostra de teste; e 12.017 para a categoria ano, 6,87% do total da amostra de teste. Conforme explicado anteriormente, apesar do falso positivo ser um valor indesejável, pois pode acarretar eventuais acidentes, a RNA teve um percentual baixo, abaixo de 11,96% em relação a amostra de teste.

Sobre os falsos negativos, obteve-se para cada categoria os seguintes valores: 19.417 para a categoria mês, 11,09% do total da amostra; 14.579 para a categoria trimestre, 8,33% do total da amostra, 10.889 para a categoria semestre, 6,22% do total da amostra e 7.348 para o período ano, 4,20% do total da amostra. Conforme explicado anteriormente, apesar do falso negativo poder levar a substituição do trilho, o não é um problema em si muito grave, os valores de estimação da RNA apresentaram um percentual baixo, abaixo de 11,09% em relação a amostra de teste.

No entanto, quando se analisa as categorias semestre e ano, mais utilizadas no planejamento da manutenção ferroviária, percebe-se que os valores dos acertos são mais altos 84,96% e 88,94%, respectivamente. Em relação aos falsos positivos para estas duas categorias, a RNA obteve percentuais de 8,82% e 6,87% respectivamente, que são valores baixos para uma estimação para 6 anos e 8 meses. Os valores dos falsos negativos para as duas categorias também foram baixos, respectivamente, 6,22% e 4,20%.

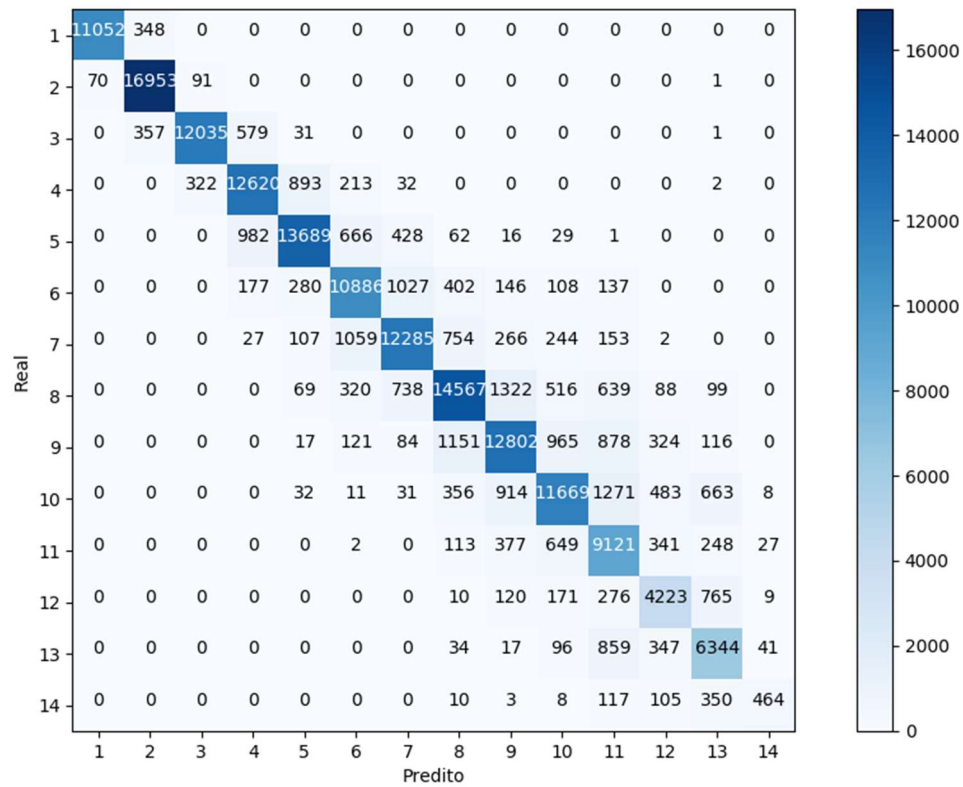
O agrupamento dos dados para categoria de período ano proporciona um aumento na quantidade de acertos totais para a RNA, propiciando uma redução para os erros de falsos positivos e negativos. Outro ponto a ser observado nestas matrizes de confusão é que nos itens 26 e 27 para o período trimestre, o item 14 para o período semestre e o item 7 para o período ano respectivamente, apresentam valores reduzidos de acertos, isto pode ser explicado devido apresentarem uma quantidade de registros inferiores para o treinamento, validação e teste em relação aos outros itens de período para cada categoria de período.

Tabela 23 – Resultados das configurações da RNA na fase de teste

Configuração	Categoria de Período	Verdadeiros Positivos (teste)	Falso Positivo (teste)	Falso Negativo (teste)	Acertos (teste) (%)	Falso Positivo (teste) (%)	Falso Negativo (teste) (%)
9	Mês	134.690	20.927	19.417	76,95%	11,96%	11,09%
24	Trimestre	142.579	17.876	14.579	81,46%	10,21%	8,33%
39	Semestre	148.710	15.435	10.889	84,96%	8,82%	6,22%
54	Ano	155.669	12.017	7.348	88,94%	6,87%	4,20%

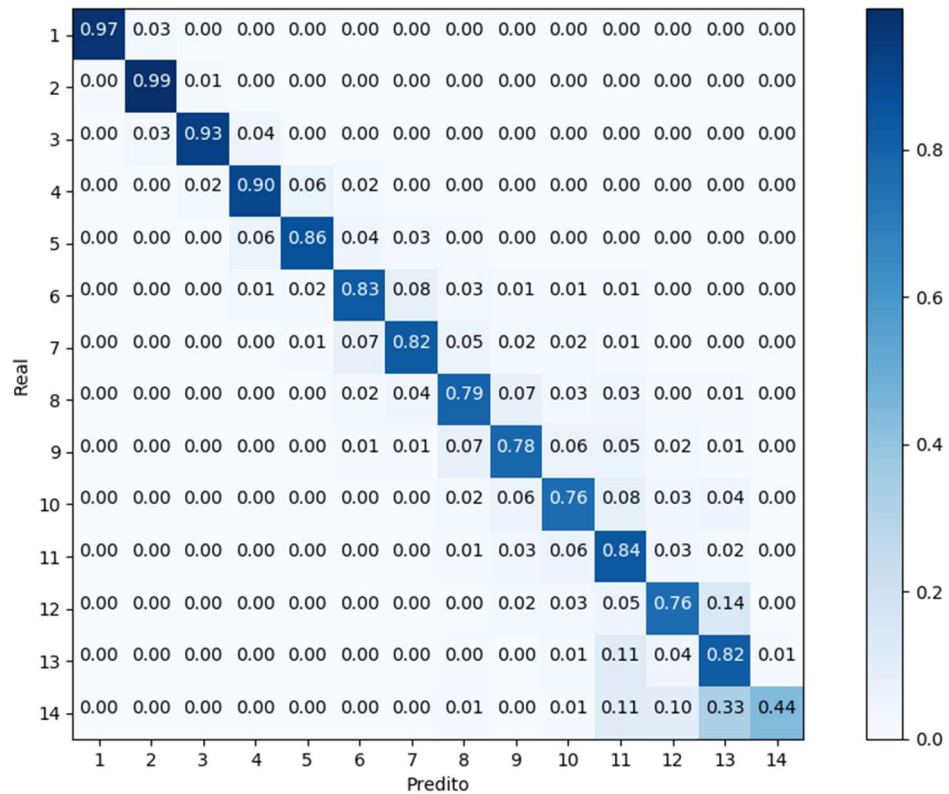
Fonte: O autor (2019)

Figura 37 – Matriz de confusão da Configuração 39 - Semestre



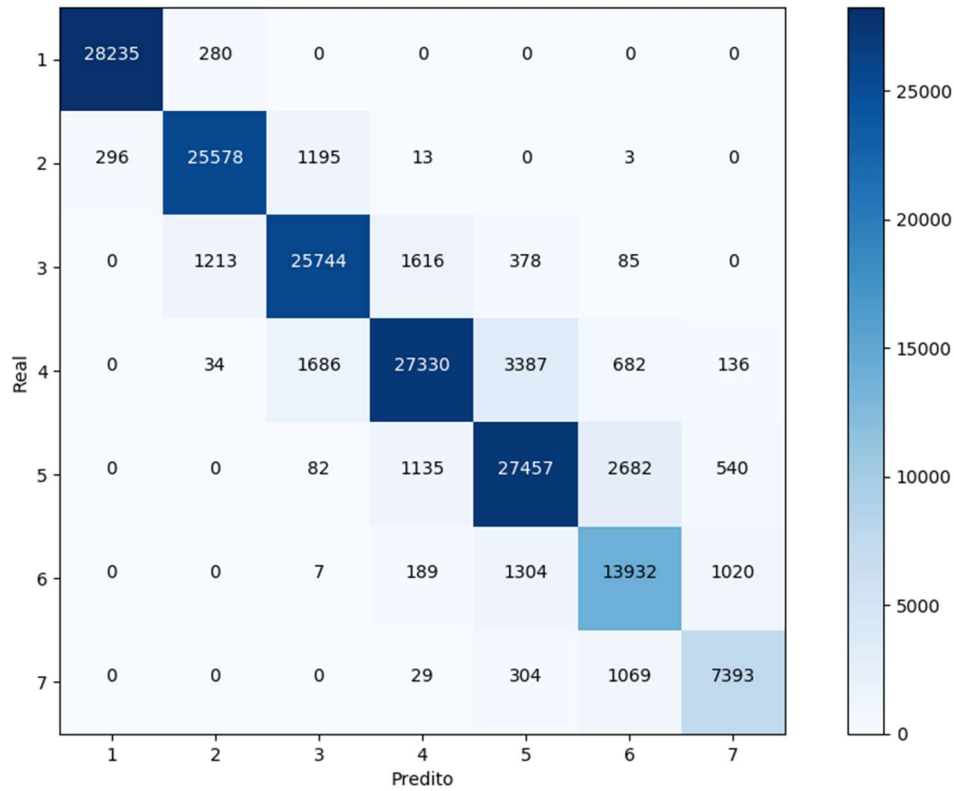
Fonte: O autor (2019)

Figura 38 – Matriz de confusão da Configuração 39 - Semestre



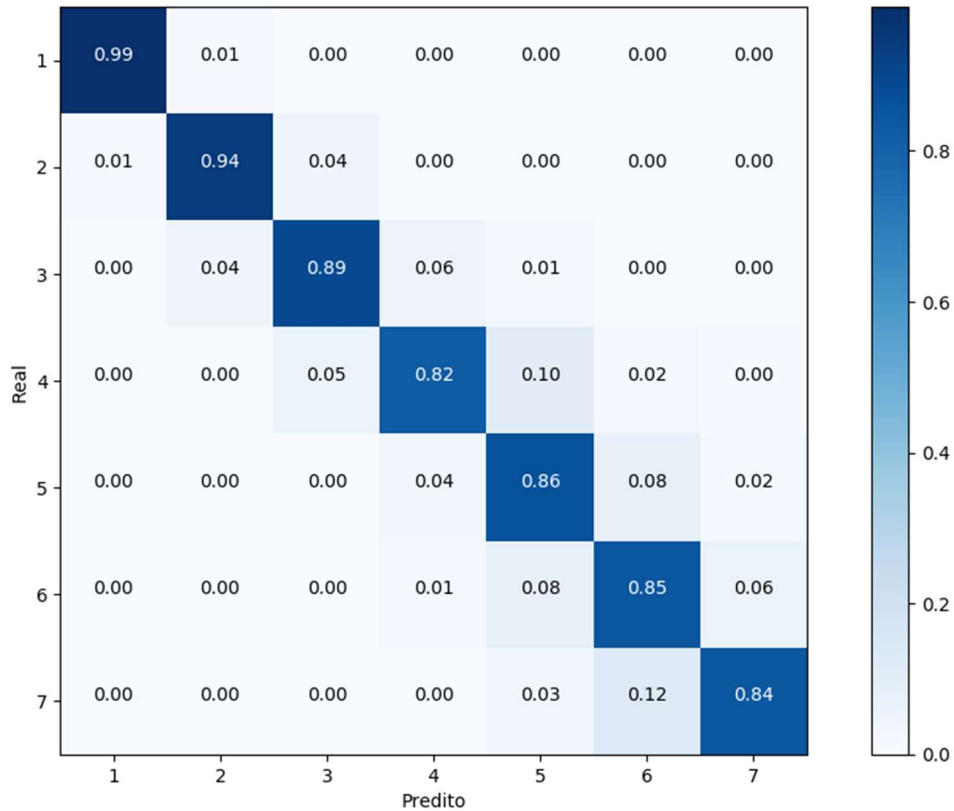
Fonte: O autor (2019)

Figura 39 – Matriz de confusão da Configuração 54 - Ano



Fonte: O autor (2019)

Figura 40 – Matriz de confusão da Configuração 54 - Ano



Fonte: O autor (2019)

7.3 ANÁLISES E RESULTADOS PARA O K-VIZINHO MAIS PRÓXIMO (K-NN)

Com as configurações do k-NN e os parâmetros de processamento definidos, e com o mesmo conjunto de dados utilizado para treinar a RNA, realizou-se o processamento destas configurações para cada categoria de períodos. Para o processamento do treinamento e validação do k-NN foram utilizados a quantidade de 1.100.000 registros. Do processamento destes registros para cada uma das configurações do k-NN foram obtidos os valores dos resultados das acurácias e dos *F1-score*.

Os pesos referentes as arestas das Configurações 55 a 66 foram salvos em arquivos, para serem lidos na fase de teste. Nas Configurações 55 a 66, buscou-se dentre as 10 partes seccionadas pelo *k-fold* na fase de treinamento e validação, a parte que apresentou o valor médio de acurácia e o valor médio de *F1-score*. Assim utilizou-se os pesos desta parte para o teste destas configurações do k-NN para categorizar o período da vida útil remanescente nos trilhos. Na Tabela 24 são apresentados os resultados das Configurações 55 a 66. As colunas desta tabela são: configurações, períodos de categoria, valores utilizados de *k-vizinho* mais próximo para rotulação do vizinho mais próximo e os valores obtidos para os resultados da acurácia e *F1-score* na fase de treinamento e validação.

Com base na Tabela 24, comparou-se as configurações do k-NN para verificar para cada categoria de período os desempenhos obtidos nos valores das acurácias e dos *F1-score*. Pode-se concluir que as configurações que apresentaram os melhores desempenhos foram aquelas que utilizaram $k = 5$, independentemente da categoria de período. Observa-se ainda que quanto maior for o valor de k , menor será o valor da acurácia e do *F1-score*. A explicação para esta situação é que o aumento do campo de atuação, tamanho do valor do *k-vizinho* mais próximo, para a rotulação do vizinho mais próximo, representa uma maior dificuldade para a classificação destes vizinhos mais próximos.

Tabela 24 – Resultados das configurações do k-NN na fase de treinamento e validação

Configuração	Categoria de Período	K	Maior acurácia	Menor acurácia	Média das acurácias	Maior F1-score	Menor F1-score	Média F1-score	Tempo de Processamento para 10 k-folds (min)
55	Mês	5	0,93978	0,86466	0,91624	0,94031	0,85602	0,91402	3,20
56		7	0,91833	0,84727	0,89621	0,91930	0,83878	0,89373	3,35
57		9	0,87443	0,81290	0,85527	0,87219	0,80304	0,85046	3,45
58	Trimestre	5	0,95125	0,88965	0,93180	0,95169	0,88107	0,92969	3,15
59		7	0,93305	0,87405	0,91444	0,93388	0,86587	0,91230	3,30
60		9	0,89564	0,84421	0,87901	0,89431	0,83469	0,87478	3,41
61	Semestre	5	0,96656	0,92602	0,95390	0,96790	0,91375	0,94901	3,12
62		7	0,95292	0,91413	0,94054	0,95352	0,90184	0,93535	3,27
63		9	0,92391	0,89106	0,91359	0,92172	0,87682	0,90568	3,39
64	Ano	5	0,97600	0,94585	0,96676	0,97668	0,93857	0,96413	3,05
65		7	0,96607	0,93696	0,95702	0,96617	0,92951	0,95405	3,24
66		9	0,94570	0,92111	0,93799	0,94292	0,91165	0,93262	3,36

Fonte: O autor (2019)

Os pesos das arestas obtidos para as Configurações 55, 58, 61 e 64, com o valor de $k=5$, na fase de treinamento e validação foram salvos em arquivo, para serem lidos na fase de teste. Nas configurações do k-NN, buscou-se dentre as 10 partes seccionadas pelo *k-fold* na fase de treinamento e validação, a parte que apresentou o valor médio de acurácia e o valor médio de *F1-score*. Assim utilizou-se os pesos desta parte para o teste destas configurações do k-NN para categorizar o período da vida útil remanescente nos trilhos. No subcapítulo seguinte será apresentado o procedimento de teste das configurações do k-NN para cada uma das categorias de períodos.

7.4 TESTE DO K-VIZINHOS MAIS PRÓXIMOS (K-NN)

O teste das Configurações 55, 58, 61 e 64, respectivamente, mês, trimestre, semestre e ano, foi realizada por meio de processamento com o conjunto de dados de teste composto com 175.034 registros utilizando os pesos da configuração do k-NN salvos durante a fase de treinamento e validação. Ressalta-se que os registros do conjunto de dados de teste não foram apresentados para as configurações do k-NN durante a fase de treinamento e validação. O teste foi realizado analisando os valores da acurácia e de *F1-score* que indicam o desempenho destas configurações para a

estimação da vida útil remanescente dos trilhos ferroviários por meio de categorias de períodos.

A Tabela 25 apresenta as configurações escolhidas para serem utilizadas para o teste, e os valores das acurácias e do *F1-score* obtidos na fase de treinamento e validação e os resultados obtidos na fase de teste para cada uma das categorias de período: mês, trimestre, semestre e ano, respectivamente, Configurações 55, 58, 61 e 64. Nesta tabela, pode-se observar que os resultados da acurácia e do *F1-score* para o conjunto de dados de teste apresentaram valores de acurácia e de *F1-score* melhores que os resultados obtidos na fase de treinamento e validação, indicando assim que estas configurações possuem a capacidade de generalização para a estimação de novos registros.

Outro ponto que pode ser observado é que todas as quatro configurações respectivamente, apresentaram na fase de teste acurácia entre 0,93274 e 0,97427 e *F1-score* entre 0,93276 e 0,97329, o que representa um valor bom considerando que se está realizando a estimação para um período de 6 anos e 8 meses.

Tabela 25 – Resultados das configurações dos k-NN na fase de treinamento/validação e teste

Configuração	Categoria de Período	Acurácia (treinamento /validação)	F1-score (treinamento / validação)	Acurácia (teste)	F1-score (teste)
55	Mês	0,91624	0,91402	0,93274	0,93276
58	Trimestre	0,93180	0,92969	0,94538	0,94579
61	Semestre	0,95390	0,94901	0,96424	0,96387
64	Ano	0,96676	0,96413	0,97427	0,97329

Fonte: O autor (2019)

Com base na Tabela 26 pode-se perceber que os acertos totais foram: 163.262 para a categoria mês, 93,27% do total da amostra de teste; 165.473 para a categoria trimestre, 94,54% do total da amostra de teste; 168.775 para a categoria semestre, 96,42% do total da amostra de teste; e 170.530 para a categoria ano, 97,43% do total da amostra de teste. Esses valores mostram um nível de acerto bom em se considerando uma estimação para 6 anos e 8 meses.

Para corroborar com a análise do parágrafo anterior, analisando-se as matrizes de confusão das configurações apresentadas na Figura 41 a Figura 44 que representam

respectivamente as categorias de período: semestre e ano, observa-se que os resultados foram concentrados na diagonal principal, o que é o mais desejado, pois a diagonal representa a quantidade de acertos totais da estimação. Decidiu-se não apresentar a matriz de confusão referente a categoria mês e trimestre, pois elas possuem 80 itens e 27 itens nas duas dimensões respectivamente, o que impossibilita a sua visualização.

Os valores acima desta diagonal representam o falso positivo, ou seja, a estimação indica uma substituição do trilho após a categoria de período que ele efetivamente deveria ter sido substituído. Este tipo de erro pode ocasionar a possibilidade de quebra do trilho. Esta é uma situação perigosa para a ferrovia, pois pode levar a acidentes ferroviários caso não ocorra a substituição do trilho na via férrea. Os valores abaixo desta diagonal representam o falso negativo, ou seja, a estimação indica a necessidade da substituição do trilho antes da categoria de período que ele efetivamente deveria ter sido substituído. Este tipo de erro pode ocasionar um impacto financeiro pelas trocas antecipadas dos trilhos.

Em relação aos falsos positivos, tem-se para cada categoria os seguintes valores: 5.497 para a categoria mês, 3,14% do total da amostra de teste; 4.546 para a categoria trimestre, 2,60% do total da amostra de teste; 3.037 para a categoria semestre, 1,74% do total da amostra de teste; e 2.292 para a categoria ano, 1,31% do total da amostra de teste. Conforme explicado anteriormente, apesar do falso positivo ser um valor indesejável, pois pode acarretar eventuais acidentes o k-NN obteve um percentual baixo, abaixo de 3,14% em relação a amostra de teste.

Sobre os falsos negativos, obteve-se para cada categoria os seguintes valores: 6.275 para a categoria mês, 3,59% do total da amostra; 5.015 para a categoria trimestre, 2,87% do total da amostra, 3.222 para a categoria semestre, 1,84% do total da amostra e 2.212 para o período ano, 1,26% do total da amostra. Conforme explicado anteriormente, apesar do falso negativo poder levar a subutilização do trilho, não é um problema em si muito grave, os valores de estimação do k-NN apresentaram um percentual baixo, abaixo de 3,59% em relação a amostra de teste.

Como na RNA, para o k-NN, os valores das categorias semestre e ano, mais utilizadas no planejamento da manutenção ferroviária, apresentam valores dos acertos mais altos com 96,42% e 97,43%, respectivamente. Em relação aos falsos positivos para

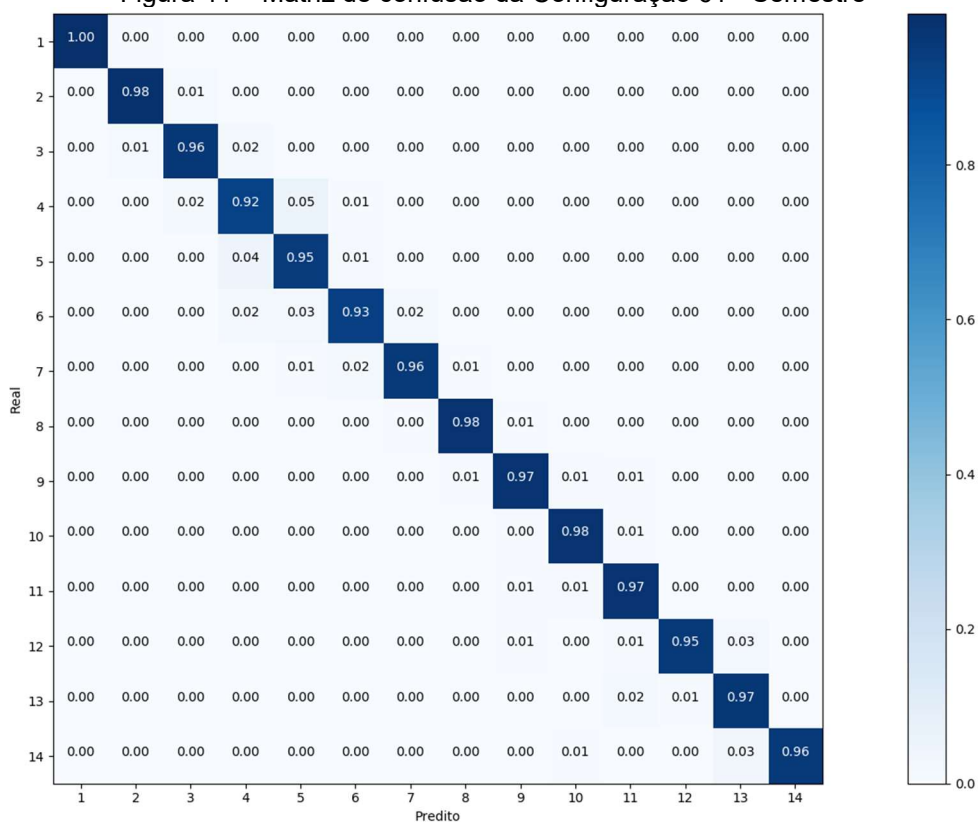
estas duas categorias, o k-NN obteve percentuais de 1,74% e 1,31% respectivamente, que são valores baixos para uma estimaco para 6 anos e 8 meses. Os valores dos falsos negativos para as duas categorias tambm foram baixos, respectivamente, 1,84% e 1,26%.

Tabela 26 – Resultados das configuraes do k-NN na fase de validao

Configurao	Categoria de Perodo	Acurcia	F1-score	Acertos	Acertos (%)	Falso Positivo	Falso Positivo (%)	Falso Negativo	Falso Negativo (%)
55	Ms	0,93274	0,93276	163.262	93,27	5.497	3,14	6.275	3,59
58	Trimestre	0,94538	0,94579	165.473	94,54	4.546	2,60	5.015	2,87
61	Semestre	0,96424	0,96387	168.775	96,42	3.037	1,74	3.222	1,84
64	Ano	0,97427	0,97329	170.530	97,43	2.292	1,31	2.212	1,26

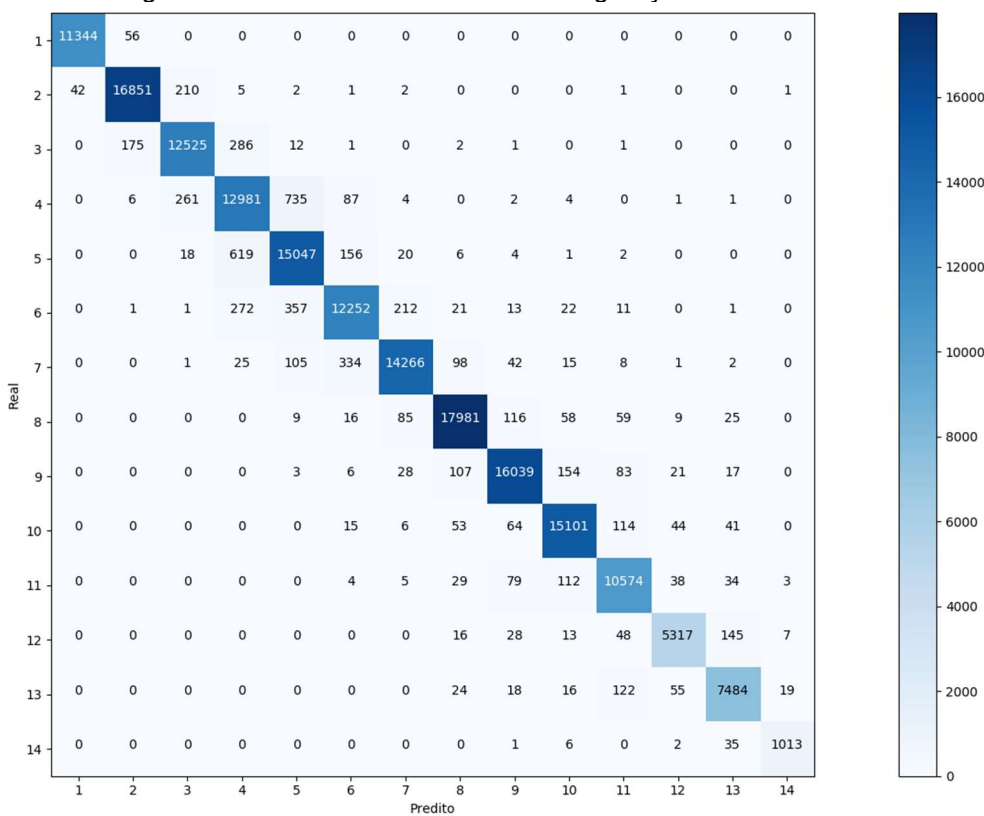
Fonte: O autor (2019)

Figura 41 – Matriz de confusão da Configuração 61 - Semestre



Fonte: O autor (2019)

Figura 42 – Matriz de confusão da Configuração 61 - Semestre



Fonte: O autor (2019)

Figura 43 – Matriz de confusão da Configuração 64 – Ano

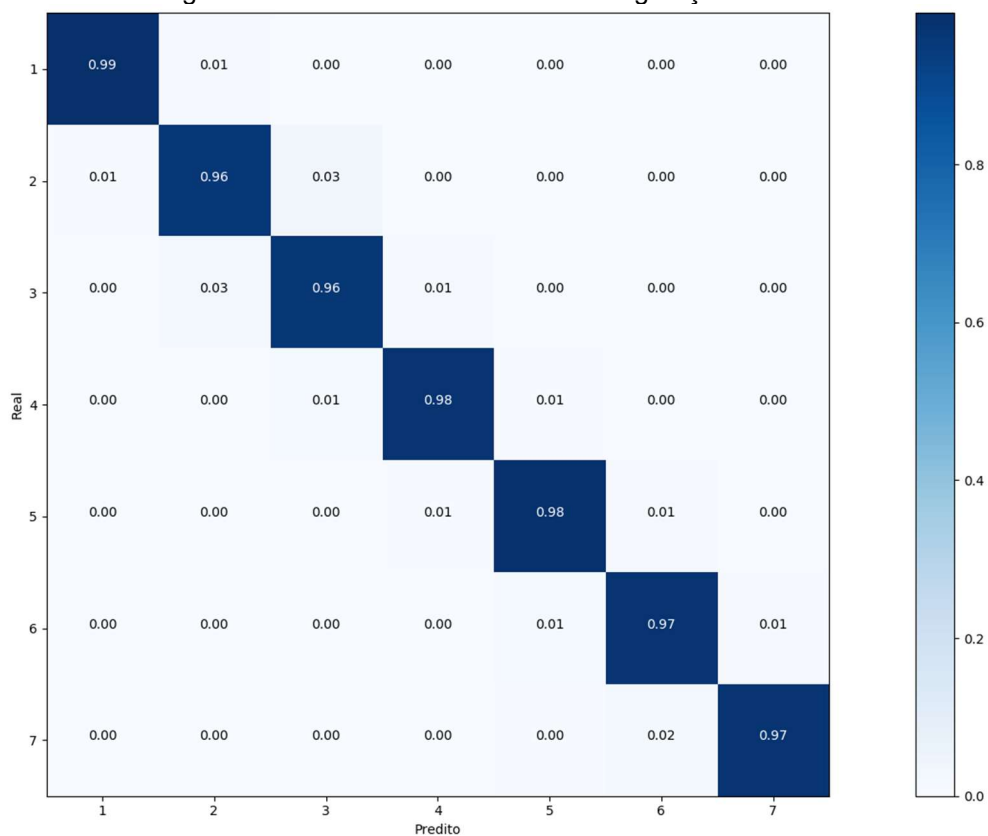
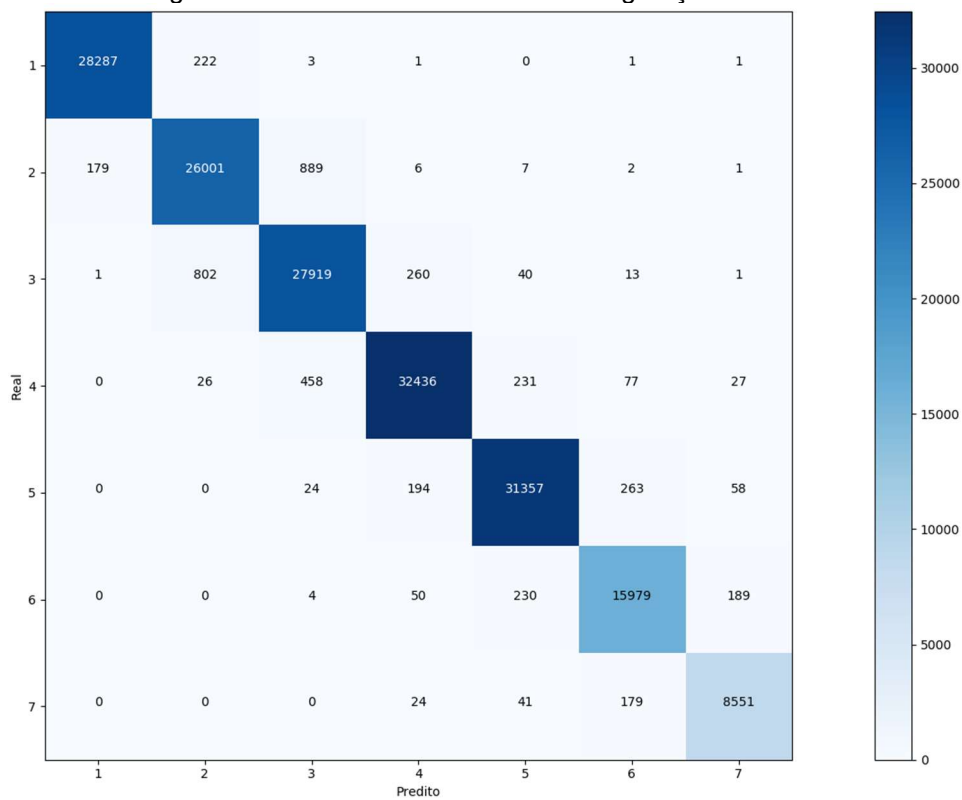


Figura 44 – Matriz de confusão da Configuração 64 – Ano



7.5 COMPARAÇÃO ENTRE A REDE NEURAL ARTIFICIAL (RNA) E O K-VIZINHOS MAIS PRÓXIMOS (K-NN)

Com base na Tabela 27 e no gráfico ilustrado na Figura 45, pode-se perceber que para todas as comparações entre as configurações da RNA e do k-NN (9 - 55, 24 - 58, 39 - 61 e 54 - 64), os valores de acurácia e *F1-score* para o k-NN são sempre melhores que os valores encontrados para a RNA. O ganho de acurácia e de acertos do k-NN em relação a RNA foi de 9,55%, 13,49%, 16,06% e 21,21%, respectivamente, ano, semestre, trimestre e mês. O ganho do *F1-score* do k-NN em relação a RNA foi de 10,62%, 16,82%, 18,05% e 23,85%, respectivamente, ano, semestre, trimestre e mês.

Em relação aos falsos positivos, o k-NN apresentou valores bem melhores que a RNA. O ganho em termos de redução de falso positivos do k-NN em relação a RNA foi de 80,93%, 80,32%, 74,57% e 73,73%, respectivamente, ano, semestre, trimestre e mês. Em relação aos falsos negativos, o k-NN também apresentou valores bem melhores que a RNA. O ganho em termos de redução de falso negativos do k-NN em relação a RNA foi de 69,90%, 70,41%, 65,60% e 67,68%, respectivamente, ano, semestre, trimestre e mês.

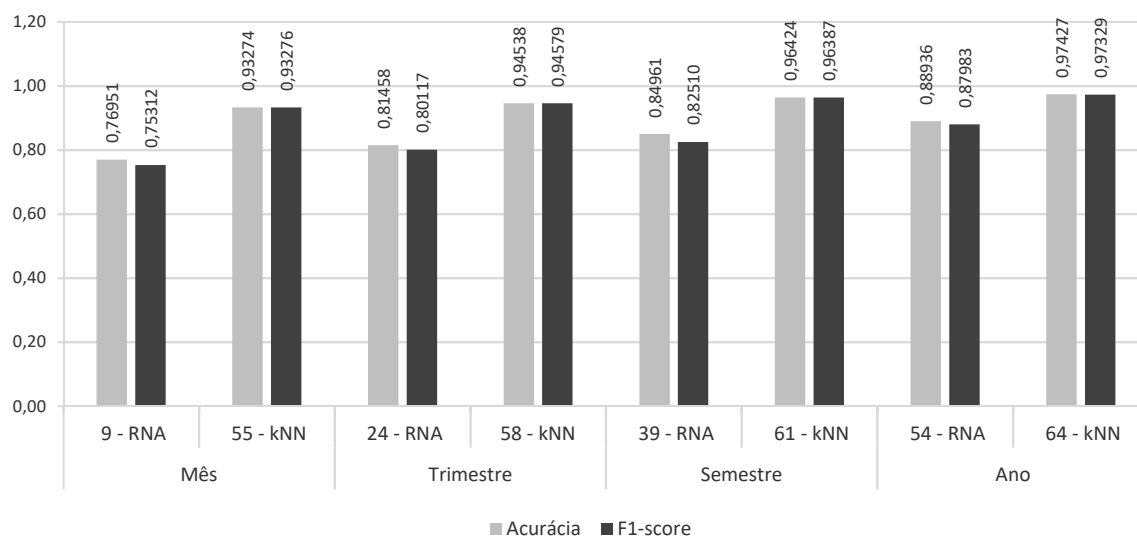
Tabela 27 – Comparação das configurações da RNA com as configurações do k-NN para cada uma das categorias de períodos

Categoria de Período	Configuração	Acurácia	F1-score	Acertos	Acertos (%)	Falso Positivo	Falso Positivo (%)	Falso Negativo	Falso Negativo (%)
Mês	9 - RNA	0,76951	0,75312	134.690	76,95	20.927	11,96	19.417	11,09
	55 - kNN	0,93274	0,93276	163.262	93,27	5.497	3,14	6.275	3,59
	Ganho (%) kNN / RNA	21,21	23,85	21,21		73,73		67,68	
Trimestre	24 - RNA	0,81458	0,80117	142.579	81,46	17.876	10,21	14.579	8,33
	58 - kNN	0,94538	0,94579	165.473	94,54	4.546	2,60	5.015	2,87
	Ganho (%) kNN / RNA	16,06	18,05	16,06		74,57		65,60	
Semestre	39 - RNA	0,84961	0,82510	148.710	84,96	15.435	8,82	10.889	6,22
	61 - kNN	0,96424	0,96387	168.775	96,42	3.037	1,74	3.222	1,84
	Ganho (%) kNN / RNA	13,49	16,82	13,49		80,32		70,41	
Ano	54 - RNA	0,88936	0,87983	155.669	88,94	12.017	6,87	7.348	4,20
	64 - kNN	0,97427	0,97329	170.530	97,43	2.292	1,31	2.212	1,26
	Ganho (%) kNN / RNA	9,55	10,62	9,55		80,93		69,90	

Fonte: O autor (2019)

O gráfico ilustrado na Figura 45 apresenta a comparação dos melhores resultados obtidos para a acurácia e para o *F1-score* para as configurações da RNA, Configurações 9, 24, 39 e 54 com as configurações do k-NN, Configurações 55, 58, 61 e 64.

Figura 45 – Comparação das configurações com melhores desempenho da RNA e K-NN



Fonte: O autor (2019)

Com base nos resultados e análises realizadas para as configurações investigadas de RNA e do k-NN para a estimação da vida útil remanescente dos trilhos ferroviários, pode-se afirmar que ambos algoritmos podem ser utilizados como uma ferramenta de auxílio ao planejador de manutenção da via férrea quanto a substituição dos trilhos ferroviários.

8 CONCLUSÕES

Esta dissertação teve como objetivo investigar dois algoritmos de aprendizado de máquina, a RNA e o k-NN, para realizar a estimação da vida útil remanescente dos trilhos ferroviários, com base nos dados históricos da substituição dos trilhos e de dados históricos das leituras realizadas pelo Carro Controle EM-80 na ferrovia EFVM para o período de 6 anos e 8 meses. Para a realização do treinamento, validações e teste das configurações da RNA e do k-NN foram utilizados 1.275.034 registros. Sendo que 1.100.000 registros foram utilizados para a realização do treinamento e validação e 175.034 registros foram utilizados para o teste das configurações da RNA e do k-NN. Para ambos algoritmos, foi utilizada a validação cruzada *k-fold*, sendo adotado o valor de $f=10$.

Foram testadas 54 configurações de RNA, que variavam a quantidade de camadas intermediárias e a quantidade de neurônios nestas camadas intermediária. Testou-se configurações com 1, 2 e 3 camadas intermediárias e com 30, 50, 100, 200 e 400 neurônios. A camada de saída representa a categoria da vida útil remanescente dos trilhos ferroviários. A camada de saída teve para cada categoria de períodos as seguintes quantidades de neurônios: 80 neurônios para a categoria mês, 27 neurônios para a categoria trimestre; 14 neurônios para a categoria semestre; e 7 neurônios para a categoria ano.

Para essas configurações foi utilizando o neurônio artificial *Perceptron* e o algoritmo de treinamento *Backpropagation*. Utilizou-se nessa dissertação para as configurações, a estrutura MLP. A função de ativação utilizada para as configurações foi a função *ReLU*, o solver utilizado para as configurações das RNA foi o Adam, a quantidade máxima de iterações foi de 500, sendo a tolerância para a otimização de 1×10^{-7} e a taxa de aprendizado para atualizações de peso foi de 0,01.

As configurações da RNA que alcançaram os melhores resultados foram as que tinham 3 camadas intermediárias com 400 neurônios cada. Para a categoria mês, a acurácia foi igual a 0,76951 e o *F1-score* igual a 0,75312. Para a categoria trimestre, a acurácia foi igual a 0,81458 e o *F1-score* igual a 0,80117. Para a categoria semestre, a acurácia foi igual a 0,84961 e o *F1-score* igual a 0,82510. Para a categoria ano, a

acurácia foi igual a 0,88936 e o *F1-score* igual a 0,87983. Percebe-se pelos resultados que categorias que consideram períodos maiores conseguem atingir valores melhores, ou seja, a precisão da estimação para o período anual, consegue ser mais assertivo.

Para o algoritmo k-NN, foram elaboradas 12 configurações variando os valores de $k=5, 7$ e 9 . As configurações do k-NN que alcançaram os melhores resultados foram as que tinham $k=5$. Para a categoria mês, a acurácia foi igual a 0,93274 e o *F1-score* igual a 0,93276. Para a categoria trimestre, a acurácia foi igual a 0,94538 e o *F1-score* igual a 0,94579. Para a categoria semestre, a acurácia foi igual a 0,96424 e o *F1-score* igual a 0,96387. Para a categoria ano, a acurácia foi igual a 0,97427 e o *F1-score* igual a 0,97329. Percebe-se pelos resultados do k-NN que, assim como na RNA, as categorias que consideram períodos maiores conseguem atingir valores melhores, ou seja, a precisão da estimação para o período anual, consegue ser mais assertivo.

Para se avaliar a qualidade dos resultados obtidos pelos algoritmos de aprendizado de máquina, comparou-se as métricas da RNA com as métricas do k-NN. Os valores obtidos para as configurações do k-NN foram melhores do que os obtidos para a RNA para todas as categorias de períodos. O ganho de acurácia e de acertos do k-NN em relação a RNA foi de 9,55%, 13,49%, 16,06% e 21,21%, respectivamente, para as categorias ano, semestre, trimestre e mês. O ganho do *F1-score* do k-NN em relação a RNA foi de 10,62%, 16,82%, 18,05% e 23,85%, respectivamente, para as categorias ano, semestre, trimestre e mês.

Com base nos resultados e análises realizadas para as configurações investigadas de RNA e do k-NN para a estimação da vida útil remanescente dos trilhos ferroviários, pode-se concluir que ambos algoritmos podem ser utilizados como uma ferramenta de auxílio ao planejamento da manutenção da via férrea quanto a substituição dos trilhos ferroviários.

8.1 TRABALHOS FUTUROS

Propõem-se para trabalhos futuros a realização de comparações com outros métodos de classificação, como *Random Forest*, *XGBoost*, dentre outros. Sugere-se ainda a utilização de outras métricas, como por exemplo, a Curva ROC. Sugere-se utilizar técnicas de *Grid Search* para avaliação de todos os parâmetros dos algoritmos de aprendizado de máquina testados. Um outro segmento de pesquisa que pode ser explorado, é tratar o mesmo problema desta dissertação por algoritmos de regressão.

REFERÊNCIAS

ADELI, Hojjat. Neural Networks in Civil Engineering: 1989-2000. **Computer-aided Civil and Infrastructure Engineering**, v. 16, n. 2, p.126-142, mar. 2001.

ABNT - ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 7590: Trilhos Vignole - Requisitos**. Rio de Janeiro. 2012.

ANDRADE, António Ramos; TEIXEIRA, Paulo Fonseca. A Bayesian model to assess rail track geometry degradation through its lifecycle. **Research in Transportation Economics**, v. 36, n. 1, p.1-8, set. 2012.

ANDREWS, John. A modelling approach to railway track asset management. **Proceedings of The Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit**, v. 227, n. 1, p.56-73, jul. 2012.

AREMA, L. M. D. American railway engineering and maintenance-of-way association. **Manual for railway engineering**, 2013.

AUDLEY, M; ANDREWS, Jd. The effects of tamping on railway track geometry degradation. **Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit**, v. 227, n. 4, p.376-391, mar. 2013.

BARAI, S. V.; PANDEY, P. C. Performance of the generalized delta rule in structural damage detection. **Engineering Applications of Artificial Intelligence**, v. 8, n. 2, p.211-221, abr. 1995.

CANNON, D. F. et al. Rail defects: an overview. **Fatigue & Fracture of Engineering Materials & Structures**, v. 26, n. 10, p. 865-886, 2003.

CAO, M. S. et al. Neural network ensemble-based parameter sensitivity analysis in civil engineering systems. **Neural Computing and Applications**, v. 28, n. 7, p.1583-1590, 15 dez. 2015.

CATTAN, Jacques; MOHAMMADI, Jamshid. Analysis of Bridge Condition Rating Data Using Neural Networks. **Computer aided Civil and Infrastructure Engineering**, v. 12, n. 6, p.419-429, nov. 1997.

CENTRO OESTE – BRASIL. **Trem de Passageiros Estrada de Ferro Vitória a Minas**. Disponível em: <http://vfco.brazilia.jor.br/Carros/efvm/trem-Vitoria-Belo-Horizonte-estacoes.shtml>. Acessado em 17 de fevereiro de 2019.

CHOU, Jung-Huai; CLARK, Robin. Application of neural networks to the inspection of Railroad rail. **Review of Progress in Quantitative Nondestructive Evaluation**, v. 18, p. 2121-2128, 1999.

DNIT- Departamento Nacional de Infraestrutura de Transportes, **PIM Procedimento de Inspeção de Material Trilho**, 44p. 2015.

DUDA, Richard O.; HART, Peter E.; STORK, David G.; **Pattern Classification**. John Wiley & Sons, 2012, 738p.

ELKHOURY, Najwa; HITIHAMILLAGE, Lalith; MORIDPOUR, Sara; ROBERT, Dilan
Degradation Prediction of Rail Tracks: A Review of the Existing Literature, **The Open Transportation Journal**, v.12, n. 1, p.88-104, fev. 2018.

FAIZ, Rizwan Bin. **An empirical rail track degradation model based on predictive analysis of Rail profile and track geometry**, Department of Computer Science Loughborough University, 178p. abr. 2010.

FALAMARZI, Amir; MORIDPOUR, Sara; NAZEM, Majidreza; HESAMI, Reyhaneh.
Rail degradation prediction models for tram system: Melbourne case study, **Journal of Advanced Transportation**, v. 2018, p.1-8, 2018.

FUKUNAGA, K. NARENDRA, P. M. A branch and bound algorithm for computing k-nearest neighbors. **IEEE Transactions on Computers**, v. 100, n. 7, p. 750-753, 1975

FUSHIKI, Tadayoshi. Estimation of prediction error by using K-fold cross-validation. **Statistics and Computing**, v. 21, n. 2, p.137-146, out. 2009. Springer Nature.

GÉRON, Aurélien. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. " **O'Reilly Media**, Inc.", 2017., 760p.

GRIMM, Kevin J.; MAZZA, Gina L.; DAVOUDZADEH, Pega. Model Selection in Finite Mixture Models: A k-Fold Cross-Validation Approach. **Structural Equation Modeling: A Multidisciplinary Journal**, v. 24, n. 2, p.246-256, dez. 2016.

GULER, Hakan. Prediction of railway track geometry deterioration using artificial neural networks: a case study for Turkish state railways. **Structure and Infrastructure Engineering**, v. 10, n. 5, p.614-626, maio 2014.

GULER, Hakan; JOVANOVIĆ, Stanislav; EVREN, Gungor. Modelling railway track geometry deterioration. **Proceedings of the Institution of Civil Engineers - Transport**, v. 164, n. 2, p.65-75, maio 2011.

GULER, Hakan. Decision Support System for Railway Track Maintenance and Renewal Management. **Journal of Computing in Civil Engineering**, v. 27, n. 3, p.292-306, maio 2013.

HASAN, M.S. **Deterioration prediction of concrete bridge components using artificial intelligence and stochastic methods**. RMIT University, mar. 2015.

HAY, William W. **Railroad engineering**. John Wiley & Sons, 2. ed. A Wiley Interscience publication. 1982. 758p.

JAIN, Lakhmi C. et al. A review of online learning in supervised neural networks. **Neural Computing and Applications**, v. 25, n. 3-4, p.491-509, dez. 2013.

KALOGIROU, Soteris. A. Artificial neural networks in renewable energy systems applications: a review. **Renewable and Sustainable Energy Reviews**, v. 5, n. 4, p.373-401, dez. 2001.

KIM, Ji-hyun. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. **Computational Statistics & Data Analysis**, v. 53, n. 11, p.3735-3745, set. 2009.

KINGMA, Diederik P.; BA, Jimmy. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.

KOHAVI, Ron et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. **Ijcai**. 1995. p. 1137-1145.

KROHLING, Renato A. BRAUN, Humberto W. **Redes Neurais Otimizadas por Meio de Enxame de Partículas com Métodos de Saltos Aplicadas em Aproximação de Funções e Predição de Séries Temporais**. Brasil, 2016, 19p.

LI, Qingyong et al. Rail Inspection Meets Big Data: Methods and Trends. **18th International Conference on Network based Information Systems**, p.302-308, set. 2015.

LI, Qingyong; REN, Shengwei. A Real-Time Visual Inspection System for Discrete Surface Defects of Rail Heads. **IEEE Transactions on Instrumentation and Measurement**, v. 61, n. 8, p.2189-2199, ago. 2012.

LIU, Xiang. Optimizing rail defect inspection frequency to reduce the risk of hazardous materials transportation by rail. **Journal of Loss Prevention in the Process Industries**, v. 48, p.151-161, jul. 2017.

LYNGBY, Narve. Railway Track Degradation: Shape and Influencing Factors. **International Journal of Performability Engineering**, v. 5, p177-186, n. 2, jan. 2009.

MCLACHLAN, Geoffrey; DO, Kim-Anh; AMBROISE, Christophe. **Analyzing microarray gene expression data**. John Wiley & Sons, 2005, 349p.

MITCHELL, Tom M. Artificial neural networks. **Machine learning**, v. 45, p. 81-127, 1997.

NEFTI, S.; OUSSALAH, M. A neural network approach for railway safety prediction. **IEEE - International Conference on Systems, Man and Cybernetics**, p.3915 - 3920, out. 2005.

PENG, D.; JONES, R. Lock-in thermographic inspection of squats on rail steel head. **Infrared Physics & Technology**, v. 57, p.89-95, mar. 2013.

PLASSER & THEURER. Carro Controle EM-80 CVRD - Brasil - **Manual de Operação BA 4992 - Parte 2- Fotografias &Figuras**, Áustria, 2008, 122p.

PLASSER & THEURER. **VALE (CVRD) Measuring Systems Manual**, Áustria, 2008, 1314p.

PODOFILLINI, Luca; ZIO, Enrico; VATN, Jørn. Risk-informed optimisation of railway tracks inspection and maintenance procedures. **Reliability Engineering & System Safety**, v. 91, n. 1, p.20-35, jan. 2006.

PROFILLIDIS, V.A. **Railway Management and Engineering**. 3. ed. United State Ashgate Publishing Company, 2006. 469p.

Programa de Parcerias de Investimento. **EFVM – Estrada de Ferro Vitória a Minas Gerais**. Disponível em: <https://www.ppi.gov.br/efvm-estrada-de-ferro-vitoria-a-minas>. Acessado em 18 de fevereiro de 2019.

PUTRA, J C P; SAFRILAH. Application of Artificial Neural Network to Predict the use of Runway at Juanda International Airport. **IOP Conference Series: Materials Science and Engineering**, v. 209, p.012107-012115, jun. 2017.

RODRIGUEZ, J.d.; PEREZ, A.; LOZANO, J.A.. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 32, n. 3, p.569-575, mar. 2010.

ROSA, Rodrigo de Alvarega, RIBEIRO, Rômulo Castello Henrique. **Estradas de Ferro - Projeto, Especificação & Construção**. Editora EDUFES, 2017, 352p.

SADEGHI, Javad; ASKARINEJAD, Hossein. Application of neural networks in evaluation of railway track quality condition. **Journal of Mechanical Science and Technology**, v. 26, n. 1, p.113-122, jan. 2012.

SADEGHI, Javad; ASKARINEJAD, Hossein. Development of improved railway track degradation models. **Structure and Infrastructure Engineering**, v. 6, n. 6, p.675-688, jul. 2008.

SANTOS, Rui; TEIXEIRA, Paulo Fonseca; ANTUNES, Antonio Pais. Planning and scheduling efficient heavy rail track maintenance through a decision rules Model. **Research in Transportation Economics**, v. 54, p.20-32, dez. 2015.

SCHAFER, D. H.; CHRISTOPHER, P.L. Barkan. A prediction model for broken rails and an analysis of their economic impact. Proc., **American Railway Engineering and Maintenance of Way Association (AREMA)**, Annual Conference. 2008.

SCIKIT-LEARN. **Machine Learning in Python**, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

SHAHIN, Mohamed A.; INDRARATNA, Buddhima. Modeling the mechanical behavior of railway ballast using artificial neural networks. **Canadian Geotechnical Journal**, v. 43, n. 11, p.1144-1152, nov. 2006.

SILVA, Renato M.; ALMEIDA, Tiago A.; YAMAKAMI, Akebo. Análise de desempenho de redes neurais artificiais para classificação automática de web spam. **Revista Brasileira de Computação Aplicada**, v. 4, n. 2, p. 42-57, 2012.

SOLEIMANMEIGOUNI, Iman; AHMADI, Alireza; KUMAR, Uday. Track geometry degradation and maintenance modelling: A review. **Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit**, v.232, n. 1, p.73-102, jul. 2016.

WANG, Liang-jie et al. A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. **Geosciences Journal**, v. 20, n. 1, p.117-136, maio 2015.

WIKIPÉDIA. **Precision and Recall**. Wikimedia Foundation, 2017. Disponível em: https://pt.wikipedia.org/w/index.php?title=Precis%C3%A3o_e_revoca%C3%A7%C3%A3o&oldid=49368588. Acessado em 18 de fevereiro de 2019.

WOLDEMARIAM, Wubeshet; MURILLO-HOYOS, Jackeline; LABI, Samuel. Estimating annual maintenance expenditures for infrastructure: Artificial neural network approach. **Journal of Infrastructure Systems**, v. 22, n. 2, p. 04015025, 2015.

YU, Zhiwen et al. Hybrid k-Nearest Neighbor Classifier. **IEEE Transactions on Cybernetics**, v. 46, n. 6, p. 1263-1275, 2015.

ZHANG, Y. J.; MURRAY, M. H; FERREIRA. L. Modelling Rail Track Performance: An Integrated Approach. **Transport Journal**, 187-194, 2000.

ZHANG, Shichao et al. Learning k for k-NN Classification. **ACM Transactions on Intelligent Systems and Technology (TIST)**, v. 8, n. 3, p. 43, 2017.

ZUMPANO, G.; MEO, M. A new damage detection technique based on wave propagation for rails. **International Journal of Solids and Structures**, v. 43, n. 5, p.1023-1046, mar. 2006.