

CLASSIFICAÇÃO DE DESCONTINUIDADES CRÍTICAS EM HASTES E ENGATES DE VAGÕES FERROVIÁRIOS COM AUXÍLIO DE REDES NEURAS ARTIFICIAIS *

Mateus Ciríaco Amaral¹
Marcelo Lucas Pereira Machado²

Resumo

A inspeção não-destrutiva por ultrassom é empregada no ramo ferroviário para inspecionar *in loco* as hastes e engates que ligam vagões entre si, a fim de detectar trincas críticas na composição durante manutenção preditiva. O ensaio requer cuidado pois não são detectadas somente trincas, mas também descontinuidades da natureza da peça. Ainda, uma oficina pode conter múltiplos inspetores, fator que introduz subjetividade ao processo. Uma Rede Neural Artificial pode auxiliar na identificação das descontinuidades críticas, dado um treinamento baseado em análises anteriores. O objetivo deste estudo foi elaborar um modelo programado em Python capaz de prever, por imagens obtidas pelo equipamento de ultrassom, a presença ou não de trincas críticas nos relatórios. A partir do processamento de 1222 imagens registradas pelo aparelho Phasor XS (General Electric), elaborou-se um banco de dados contendo valores de intensidade para cada relatório (variáveis de entrada), e valores binários relacionados a existência ou não de trincas críticas (variável de saída). Os dados registrados foram inseridos em uma rede perceptron de múltiplas camadas, de configuração 20-36-1 e ativações logísticas. Resultados parciais apontam um percentual de predições corretas de 97%, compatível com estudos similares, e métricas de suporte que validam o aprendizado do modelo.

Palavras-chave: Redes Neurais Artificiais; Ultrassom; Ensaio Não-Destrutivo.

CLASSIFICATION OF CRITICAL DISCONTINUITIES IN RAIL-WAGONS' COUPLINGS AND RODS WITH AID OF ARTIFICIAL NEURAL NETWORKS

Abstract

Non-destructive ultrasonic inspection is employed in the railroad industry to inspect *in loco* the rods and couplings that connect wagons to each other in order to detect critical cracks in the composition during predictive maintenance. The test requires care because not only cracks are detected, but also natural discontinuities of each part. Also, a repair workshop may contain multiple inspectors, a factor that introduces subjectivity to the process. An Artificial Neural Network can aid in the identification of critical discontinuities, given training based on previous analyzes. The aim of this study was to elaborate a Python-programmed model capable of predicting, using images obtained by the ultrasound equipment, the presence or not of critical cracks in each report. A database containing intensity values for each report (input variables), and binary values related to the existence or not of critical cracks (output variable) was elaborated from processing 1222 images generated by the Phasor XS (General Electric) equipment. The data recorded were inserted into a multi-layer perceptron network, of 20-36-1 configuration and logistic activations. Partial results point to a 97% prediction accuracy, value that is compatible with similar studies, and support-metrics that validate model learning.

Keywords: Artificial Neural Networks; Ultrasound; Non-Destructive Testing.

¹ Engenharia Metalúrgica e de Materiais, M.Sc., Programa de Pós-Graduação em Engenharia Metalúrgica e Materiais, IFES, Vitória, ES, Brasil.

² Engenharia Elétrica, D.Sc., Professor Titular do IFES, Programa de Pós-Graduação em Engenharia Metalúrgica e Materiais, IFES, Vitória, ES, Brasil.

1 INTRODUÇÃO

Durante o translado sob malha ferroviária, os vagões de um trem exercem esforços cíclicos de tração e compressão nas peças metálicas que os interligam. Tais peças, denominadas de hastes (rígidas, quando ligam vagões geminados) ou engates (acopláveis, quando ligam uma dupla de vagões à outra), são construídas em aço fundido e apresentam furos e encaixes em suas geometrias para alojar outros componentes e permitir suas fixações nos vagões [1].

Mesmo resistentes, as hastes e engates acabam sofrendo esforços suficientes ao longo do tempo para nuclear trincas por fadiga. Caso continuem empregadas em uma composição, essas trincas recém-formadas podem crescer em tamanho e atingir dimensões críticas, que por consequência ocasionariam na fratura do material.

Para evitar que uma dessas peças fraturem durante o funcionamento de um trem, podendo levar a composição a uma falha catastrófica, é comum que oficinas de manutenção de ativos ferroviários disponham de técnicas de inspeção preventivas e preditivas, a fim de capturar o crescimento de trincas e condenar qualquer material defeituoso antes que o mesmo se rompa durante atividade.

Uma dessas técnicas é o uso de equipamentos de ultrassom na própria linha de manutenção das oficinas [2]. O técnico especializado faz a varredura de todas as hastes ou engates de um trem estacionado, para garantir que a composição não deixe a oficina com alguma trinca de tamanho crítico (trinca com tamanho a partir do qual seu crescimento não pode ser controlado, sendo assim risco de falha iminente). Caso necessário, a peça com o defeito interno é substituída por uma nova antes da liberação do trem.

Todavia, essa varredura consome uma certa quantidade de tempo e requer um nível elevado de cuidado do inspetor – nem sempre o equipamento de ultrassom facilita a visualização de uma trinca, que muitas vezes se confundem com outros ruídos (vide Seção 1.1). Além disso, um chão de fábrica que recebe muitos vagões para manutenção e opera 24 horas por dia requer o emprego de múltiplos inspetores: é possível que nem todos possuam o mesmo grau de instrução ou nível de cuidado, além de estarem expostos a diferentes fatores externos.

Observa-se que o procedimento de ensaio *in loco* das hastes e engates por ultrassom pode ser otimizado tanto em relação à velocidade e eficiência do processo, quanto em relação à redução do fator de subjetividade. Para isso, é possível implementar um modelo computacional que auxilie cada inspetor nas tomadas de decisão, reduzindo o tempo gasto com cada análise e possibilidades de erro humano. O presente trabalho sugere a criação de um modelo utilizando Redes Neurais Artificiais (RNAs) para que a mesma aprenda a classificar a existência ou não de trincas críticas a partir de um histórico de relatórios reais de ultrassom.

Redes neurais artificiais já foram implementadas para a otimização de processos em trabalhos similares: Nazário et al. (2009) aliaram RNAs e inspeção por ultrassom para detectar, com taxas de acerto em cerca de 95%, possíveis adulterações em leite bovino a partir da detecção dos diferentes teores de água e gordura contidos no leite, e correlação com a qualidade final do produto [3]. Marcomini (2017) também utilizou de um modelo em RNA para a classificação de nódulos em banco de imagens de resultados de ultrassonografia mamária, obtendo taxas de sucesso dentre 90% e 94,2% [4]. Costar et al. (2018) empregaram uma rede neural para a identificação do estágio de propagação de trincas por fadiga em um corpo rígido tubular, utilizando como variáveis dados extraídos por uma técnica similar de

inspeção (emissão acústica), obtendo sucesso de classificação de 82,9% [5]. Os três trabalhos foram motivadores para a execução deste estudo.

1.1 Ensaio Não-Destrutivo por Ultrassom

Os Ensaio Não-Destrutivos (ENDs) são técnicas utilizadas para inspeção de materiais que não alteram de forma permanente suas propriedades físicas, químicas, mecânicas ou dimensionais [6]. Estão entre as principais ferramentas de controle de qualidade em setores industriais, por permitir a manutenção de peças e equipamentos sem causar danos aos mesmos, muitas das vezes *in loco*.

Dentre os principais ENDs, encontra-se a técnica de ultrassom, baseada na reflexão de ondas acústicas pelo interior de uma peça. Tais ondas são emitidas por cristais que vibram a frequências superiores a 20 kHz, e se propagam por meios sólidos até serem transmitidas para fora do material ou refletidas por alguma descontinuidade – sejam pelas próprias interfaces do sistema (ecos de fundo), descontinuidades geométricas (e.g.: furos em peças metálicas) ou defeitos internos (e.g.: trincas).

Existem diversos métodos de varredura, o mais simples sendo conhecido como A-Scan [7], onde todo sinal refletido é captado por um elemento transdutor e exibido pelo aparelho de ultrassom como uma imagem correlacionando intensidade em função do tempo/profundidade (Figura 1):

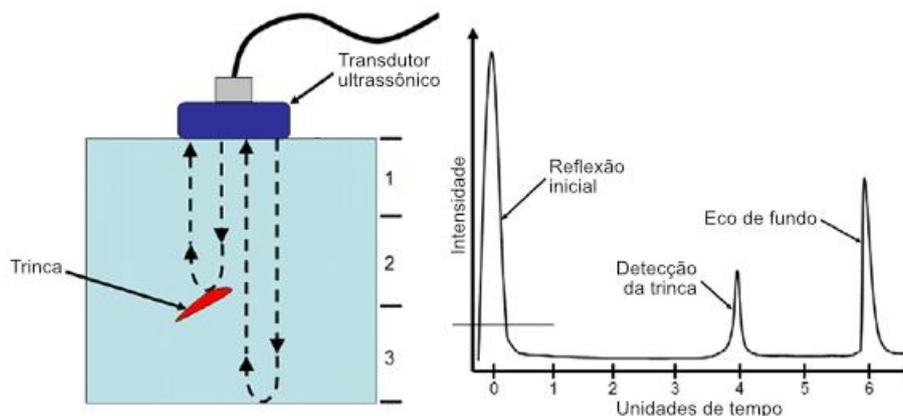


Figura 1. Representação do sinal A-Scan “pulso-eco”

Com base na comparação de espectros de ultrassom entre peças adotadas como padrão e peças com algum tipo de descontinuidade, é possível correlacionar a presença de novos picos com descontinuidades inseridas nas mesmas, e daí se torna possível a detecção de trincas com o uso da técnica.

No entanto, embora apresente muitas vantagens em relação a outros tipos de ensaios, a inspeção por ultrassom requer grande conhecimento teórico e experiência por parte do operador do equipamento – ainda mais se tratando de situações onde não se tenham peças-padrão por perto, e de peças que apresentem descontinuidades geométricas além de possíveis trincas. Como exemplo, a inspeção em campo de uma haste ou engate requer o reconhecimento do furo na extremidade e atenção para a detecção de defeitos adicionais sem que se perca o furo de vista.

O uso de aparelhos de ultrassom modernos que apresentem novas tecnologias de transdutores (e.g.: múltiplos cristais em *phased array*), funções adicionais (como a captura de relatórios em campo e transmissão de dados em rede local) e portabilidade em geral (isto é, aparelhos com baterias melhores e dimensões reduzidas) tem aumentado a eficiência durante a inspeção preditiva de defeitos.

Conforme sugerido, eventualmente pode-se aliar tais equipamentos com um modelo baseado em redes neurais capaz de auxiliar o inspetor em campo, aumentando a eficiência do processo de inspeção.

1.2 Redes Neurais Artificiais

Uma rede neural artificial é um modelo matemático inspirado na estrutura neural de organismos inteligentes, capaz de aprender pelo exemplo [8]. Um dos primeiros e mais simples tipos de rede neural é conhecido como o *perceptron* [9], criado em 1956 por Frank Rosenblatt, capaz de realizar classificações binárias através de uma fronteira de decisão linear.

Como o *perceptron* sozinho se limita a problemas de classificação de soluções lineares, casos mais complexos requerem o uso de vários neurônios em paralelo, em várias camadas densamente ligadas – este tipo de rede em particular é denominado *Multi-Layer Perceptron* (MLP).

Os *perceptrons* de múltiplas camadas são geralmente aplicados em problemas de aprendizado supervisionado – isto é, as redes são treinadas em pares de entradas-saídas a fim de modelar correlações ou dependências entre os dados apresentados. O processo de propagação dos impulsos (dados) para a próxima camada é denominado *feedforward*, enquanto a correção de erros é dada de trás para frente, pelo ajuste dos pesos e vieses de cada neurônio (das camadas exceto a de entrada) em um processo denominado *backpropagation*.

1.2.1 Redes MLP em Ambiente de Programação Python

Com o interesse crescente do mercado pela área de aprendizado de máquina, a linguagem de programação Python tem sido muito utilizada por pesquisadores e cientistas de dados para a elaboração de algoritmos capazes de processar e interpretar tarefas envolvendo grandes quantidades de dados, como é o caso das aplicações de redes neurais artificiais.

A linguagem Python foi criada em 1991 por Guido van Rossum [10] para enfatizar o esforço do programador sobre o esforço computacional – e por ser uma linguagem de propósito geral, fácil legibilidade e alto nível de abstração, tem sido amplamente adotada por pesquisadores. Ainda, o Python apresenta grandes vantagens em relação a soluções especializadas como o MatLab ou Statistica, por ser uma linguagem gratuita, de código aberto e disponível em vários sistemas operacionais.

Existem diversos esforços comunitários na criação de bibliotecas (conjuntos de algoritmos prontos) que facilitem o uso da linguagem para determinadas aplicações. Em se tratando de aprendizado de máquina, uma biblioteca muito utilizada é o Scikit-Learn [11], por apresentar diversos algoritmos já pré-construídos, reduzindo ainda mais o tempo que um pesquisador levaria apurando um programa próprio, permitindo que o foco do usuário esteja apenas na modelagem do problema.

Contudo, é importante entender o funcionamento destes módulos, mesmo que já prontos para a utilização. Ainda a respeito do Scikit-Learn, observa-se que o algoritmo para redes MLP calcula os erros da etapa de *backpropagation* por uma função denominada entropia cruzada, ou *log-loss*. A Equação 1 exemplifica como é feito o cálculo do erro (*loss*), para um caso simplificado com apenas duas classes:

$$loss = -(y * \log(p) + (1 - y) * \log(1 - p)) \quad (1)$$

Onde y seria a classe real de uma amostra e p a probabilidade prevista pelo sistema de que a amostra realmente pertença à classe y . Além disso, o otimizador da função de entropia é um algoritmo de descida de gradiente, que visa caminhar a função para o menor erro possível – em termos práticos, isto significa reduzir a entropia de informação da rede. O módulo Scikit-Learn apresenta vários tipos de otimizadores para seleção, sendo ADAM o padrão [12].

1.3 Medidas Estatísticas de Desempenho

Para avaliar o desempenho do modelo e estabelecer comparações entre estudos, é preciso definir critérios estatísticos de relevância. Para trabalhos envolvendo tarefas de classificação, uma série de métricas estatísticas são utilizadas, todas derivadas de uma estrutura denominada matriz de confusão [13]. Também, para trabalhos de aprendizado de máquina, um estimador simples conhecido como modelo *dummy* geralmente é empregado [14], fornecendo uma linha de base para comparações.

1.3.1 Matriz de Confusão e Métricas Relacionadas

Em problemas estatísticos envolvendo a simulação de classificações, é comum a utilização de uma tabela de layout específico que permite o cruzamento dos dados reais (isto é, conforme coletados) com os dados simulados (e.g.: dados sugeridos por um modelo). Esta tabela é conhecida no campo estatístico como matriz de confusão [13].

Uma matriz de confusão tem formato quadrado (como mesmo número de linhas e colunas), de dimensão é dada pelo número de classes do problema. Para tarefas envolvendo classificações binárias, por exemplo, a matriz apresenta duas linhas e duas colunas (Tabela 1):

Tabela 1. Matriz de confusão para classificações binárias

		Previsto	
		Negativo	Positivo
Real	Negativo	TN	FP
	Positivo	FN	TP

Para um caso específico de classificação binária, o cruzamento entre os dados gera quatro resultados distintos:

- os verdadeiros positivos (*true positives*, TP) – valores da classe real positiva, classificados corretamente pelo modelo;
- os verdadeiros negativos (*true negatives*, TN) – valores da classe real negativa, classificados corretamente pelo modelo;
- os falsos positivos (*false positives*, FP) – valores da classe real negativa, classificados erroneamente pelo modelo como positivos;
- os falsos negativos (*false negatives*, FN) – valores da classe real positiva, classificados erroneamente pelo modelo como negativos.

Tanto os falsos positivos quanto negativos são falhas de classificação do modelo, no entanto, cada valor representa um tipo de erro de gravidade diferente (dependendo da importância das classes negativa/positiva do estudo).

Dos quatro valores resultados da matriz de confusão, derivam-se algumas métricas relevantes de avaliação. A principal delas é denominada acurácia, e é simplesmente a taxa de acertos do modelo [15]. A acurácia (*accuracy*) é dada pela Equação 2:

$$accuracy = \frac{TN + TP}{(TN + FN + TP + FP)} \quad (2)$$

Porém, a acurácia sozinha não é uma boa métrica quando a distribuição dentre classes do problema é desbalanceada. De fato, quanto mais desbalanceada a relação entre as classes positiva/negativa (e.g.: apenas 1% das amostras se enquadram na classe positiva, e as outras 99% na classe negativa), menor a confiabilidade da acurácia (pois a taxa de acertos torna-se naturalmente maior) [16]. Logo, definem-se duas métricas de suporte para a acurácia que ignoram a classe de maior aparição nos bancos de dados. A primeira delas é denominada precisão (*precision*), e representa o percentual de relevância dentro dos valores sugeridos pelo modelo [15] (Equação 3):

$$precision = \frac{TP}{(TP + FP)} \quad (3)$$

A segunda métrica é conhecida como *recall*, e representa o percentual, de todos os valores relevantes na amostragem, dos valores que foram capturados pelo modelo [15], vide Equação 4:

$$recall = \frac{TP}{(FN + TP)} \quad (4)$$

Para uma rede neural artificial onde a classificação positiva é bem menos frequente, o uso das métricas auxiliares é de extrema importância. Ainda, cada uma das métricas calculadas apresentam significados diferentes variando-se o contexto do estudo, que precisa ser averiguado para que as sejam otimizadas as métricas de maior relevância.

1.3.2 Modelo *Dummy* de Classificação

O classificador postigo (do inglês *Dummy Classifier*) é um modelo utilizado em aprendizado de máquina que realiza previsões completamente ignorando as variáveis de entrada e apenas seguindo regras simples [14]. Este classificador é considerado como uma linha de base para comparação com outros modelos, já que é teoricamente um classificador sem capacidade nenhuma de aprendizado.

A biblioteca Scikit-Learn apresenta modelos *dummy* que verificam as variáveis de saídas de um banco de dados e adotam estratégias diferentes para elaborar as classificações [12]:

- Um modelo de estratégia “estratificada” irá prever os resultados de forma aleatória, mas respeitando a distribuição de classificações do banco de dados inicial (e.g.: o modelo arrisca uma classificação positiva 80% das vezes, caso os dados de treinamento possuam 80% de classes positivas);
- Um modelo de estratégia “mais frequente” irá sempre arriscar a classificação que mais aparecer dentro do banco de dados utilizado no treinamento;
- Um modelo de estratégia “uniforme” irá sempre prever os resultados de forma aleatória, respeitando chances iguais (e.g.: em uma classificação binária, o modelo arrisca 50% das vezes cada uma das classes).

Como todas as estratégias independem das variáveis de entrada, os classificadores *dummy* servem como linha de base para melhorias verificadas em outros modelos, como as redes neurais artificiais, onde quantifica-se o efeito do aprendizado pelo exemplo no aumento da taxa de acertos. Estes estimadores ainda servem como

testes de sanidade para os estudos estatísticos – dependendo da estratégia, espera-se que os resultados de um classificador *dummy* aproximem-se de valores estatisticamente compatíveis, e qualquer fuga dos resultados esperados mostra que existe algum erro na implementação dos modelos.

2 DESENVOLVIMENTO

2.1 Metodologia

Relatórios de ultrassom gerados pelo equipamento Phasor XS da General Electric foram salvos e transferidos para um computador, onde foram manualmente classificados quanto à presença ou não de trincas consideradas de comprimento crítico ou acima. Ao todo, 1222 relatórios foram manualmente classificados, sendo 102 imagens de hastes/engates com trincas de tamanho crítico – o que constitui a classe positiva como sendo a menor população da amostra (8,347% do total). Os relatórios classificados foram então codificados em vetores numéricos contendo informações a respeito da intensidade de sinal na região do A-Scan (Figura 2). A codificação foi feita de forma automática por um *script* em Python de autoria própria.

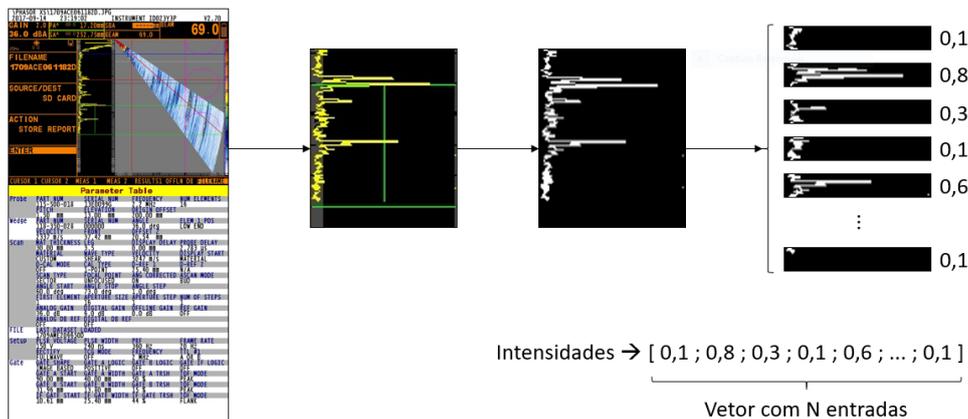


Figura 2. Roteiro de tratamento e conversão dos relatórios

A partir de testes iniciais, verificou-se a possibilidade de implementação da rede a partir de um vetor “entradas” de dimensão $N = 20$. Um número superior indica maior resolução para as entradas (o que pode ou não afetar no aprendizado da rede) a um custo extra no tempo de processamento.

Também em Python, um modelo de rede neural artificial foi programado, com auxílio da biblioteca Scikit-Learn. Para o modelo, escolheu-se uma rede MLP com os seguintes parâmetros internos:

- 20 neurônios de entrada (dimensão de cada vetor codificado) e 1 neurônio de saída (classificação da existência de trincas críticas – valor binário onde 0 indica negativo e 1 indica positivo)
- 36 neurônios para uma única camada escondida (quantidade definida por força bruta após testes iniciais)
- Funções logísticas de ativação para todas as camadas
- Taxa de aprendizagem dinâmica (a aprendizagem decai lentamente ao longo do treinamento, padrão do Scikit-Learn)
- Otimizador de pesos ADAM (também padrão do Scikit-Learn)

- f) 10000 épocas (iterações), ou parada antecipada caso não haja um decréscimo no erro de 10^{-7} entre iterações

Após treinada com os pares vetoriais de intensidade/classificação, a rede neural foi avaliada por uma matriz de confusão em relação à presença das trincas. Nesta etapa, o objetivo foi interpretar o que significam, na prática, as métricas importantes de acurácia, precisão e *recall*. O modelo como um todo também foi comparado com os trabalhos relacionados [3,4,5] e estimadores *dummy* quanto à taxa de acertos.

2.2 Resultados e Discussões

Após 5 rotinas de treinamento da rede MLP, adotou-se o modelo que obteve o melhor percentual de acertos na classificação, conforme indicado na Tabela 2. Observa-se aqui que a taxa de acertos varia levemente a cada procedimento de treino, influenciada principalmente pela inicialização aleatória dos pesos e vieses.

Tabela 2. Redes 20-36-1 em 5 treinamentos consecutivos

Rede	1	2	3	4	5
Acurácia (%)	97,3813	95,1718	97,1358	96,9722	97,2995
Épocas	2195	682	2483	1907	1798
Log-loss	0,0656	0,1138	0,0629	0,0713	0,0693

A partir da maior acurácia, selecionou-se a Rede 1 e elaborou-se uma matriz de confusão entre a classificação real e a classificação sugerida pela rede (Tabela 3).

Tabela 3. Matriz de confusão para a Rede 1

		Previsto	
		Negativo	Positivo
Real	Negativo	1103	17
	Positivo	15	87

Obs.: A classificação positiva se refere à existência de trincas críticas

As métricas de acurácia, precisão e recall (vide Seção 1.3.1) podem ser calculadas através do número de amostras em cada célula da Tabela 2. O modelo selecionado apresenta as seguintes métricas (Tabela 4):

Tabela 4. Métricas de avaliação do desempenho

	Acurácia	Precisão	Recall
Rede 1	97,3813	83,6538	85,2941

Obs.: Todos os valores percentuais (%)

Analisando na prática o que representam os resultados numéricos obtidos:

- Uma acurácia de 97% representa a taxa de acertos do modelo filtrado. O resultado indica que a cada 100 vagões, mais de 97 serão classificados de maneira correta. Reforça-se aqui que essa métrica é a única que leva em consideração os verdadeiros negativos – e como a essa população compõe a maior parte da amostragem, é esperado que a mesma contribua significativamente para o percentual alto.
- Uma precisão de 83% mostra o percentual de relevância das instâncias positivas classificadas pela rede. O número indica que a cada 100 casos classificados de modo positivo, 83 realmente serão trincas críticas. Nesse contexto, talvez mais interessante seja a avaliação inversa: dentre 100

trincas, 17 talvez sejam falsos positivos – um erro que representaria um pequeno gasto com reposição de material metálico de forma desnecessária.

- c) Uma taxa de *recall* de 85% mostra o percentual classificado pela rede de todos os resultados relevantes da amostragem. Isso diz que, a cada 100 trincas críticas existentes em uma amostra de vagões, a rede é capaz de detectar 85. Também aqui, o inverso tem maior importância: após a classificação, possivelmente 15 peças trincadas continuariam na composição – um erro de maior gravidade por apresentar maiores riscos em potencial.

Logo, fica claro que para o emprego de uma RNA para esta aplicação em específico, é necessário otimizar o modelo visando principalmente um aumento da taxa de *recall*. No entanto, ainda neste momento a rede se desempenha de forma satisfatória, o que fica visível ao comparar a acurácia deste trabalho com outros estudos de caso. A Figura 3 compara a taxa de acertos entre os estudos relacionados – além de compará-los com os classificadores *dummy*, definidos como linha de base:

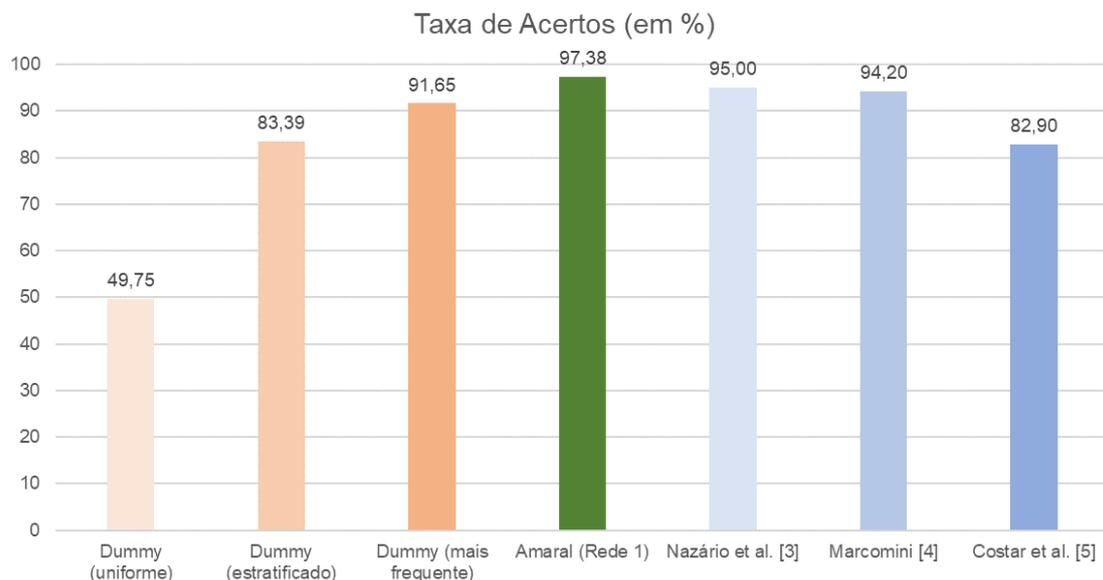


Figura 3. Comparação entre estimadores *dummy* e modelos baseados em RNAs

Observa-se que os percentuais de acerto dos classificadores postíços batem com o esperado estatisticamente, isto é:

- Para a estratégia “uniforme”, uma acurácia próxima 50% já era esperada. Na prática, esse valor corresponde, dado um número infinito de amostras, a taxa de acertos de um indivíduo que não possui treinamento nenhum e chuta uma classificação binária ao acaso;
- Para a estratégia “estratificada”, o valor encontrado se aproxima do estimado estatisticamente considerando o somatório dos quadrados das probabilidades $((91,65\%)^2 + (8,35\%)^2)$. Isso corresponde a taxa de acertos de um indivíduo que ainda não possui conhecimento, mas sabe que do todo apenas uma minoria (8,35%) apresentaria uma trinca crítica, arriscando uma classificação de maior frequência 91,65% das vezes;
- Para a estratégia “mais frequente”, o valor encontrado é igual à frequência da classe negativa, que aparece 91,65% das vezes. O resultado corresponde a um indivíduo que não classifica aleatoriamente, mas sempre responde com a classificação que mais apareceu na amostragem.

Assim, é possível verificar que a implementação estatística da avaliação de desempenho está adequada. Também, visualiza-se o quão importante é o processo de treinamento para o aumento na acurácia dos modelos capazes de adquirir conhecimento, tomando como exemplo os estudos relacionados.

Os resultados encontrados pela Rede 1, elaborada com a finalidade de auxiliar na classificação de trincas críticas, são validados por serem compatíveis com os valores encontrados nas literaturas que descrevem a aplicação dessas redes neurais como novas tecnologias.

É importante notar que métricas próximas a 100% são ideais, logo, nem sempre possíveis. No entanto, a maneira de aplicação da tecnologia garante que modelos sejam vantajosos a partir do momento em que suas métricas de desempenho superem o desempenho atual – isto é, uma rede neural com taxa de recall maior do que o recall do procedimento de inspeção manual já atinge seu objetivo. Além disso, a finalidade para um sistema desse tipo permite uma certa folga no desempenho do modelo: por exemplo, uma rede que auxilie tomada de decisão não precisa do mesmo desempenho do que uma rede que tome decisão de forma autônoma.

Este trabalho é resultado parcial de uma dissertação de Mestrado envolvendo a otimização dos parâmetros internos de uma rede MLP, desenvolvida especificamente para a aplicação na classificação de trincas críticas de hastes e engates de vagões ferroviários – o estudo completo está previsto para meados de 2020.

3 CONCLUSÃO

Os resultados parciais obtidos indicam que o modelo elaborado por redes neurais artificiais é passível de implementação como ferramenta de auxílio ou até mesmo tomada de decisão durante ensaios em campo de ultrassom para a detecção de descontinuidades críticas.

A rede MLP foi capaz de aprender, a partir de 1.222 relatórios de ultrassom extraídos diretamente do equipamento, nuances no sinal A-Scan que permitem classificar corretamente uma peça como condenada por trinca ou não.

Os parâmetros até então adotados já são suficientes para obter-se uma taxa de acertos de 97%, comparável ao encontrado em trabalhos anteriores. A tendência dos resultados finais é obter uma taxa de acerto similar ou superior ao mesmo tempo em que a taxa de recall seja melhorada.

Em uma possível implementação em campo, um modelo de RNA que auxilie tomada de decisão em tempo real trará benefícios a indústria, como o aumento da velocidade de inspeção e redução de erro humano por reduzir a subjetividade. Eventualmente, a implementação de uma tecnologia como a descrita também possibilita a automatização do serviço de inspeção.

Agradecimentos

Os autores agradecem ao CNPq, à CAPES e à FAPES pelo fomento da pesquisa por meio do edital CAPES/FAPES PROCAP Mestrado 01/2018, e ao pessoal do PROPEMM-IFES pela possibilidade do estudo.

REFERÊNCIAS

- 1 Silva PRF da, Kina EJ, Bergantini M. Trainee corporativo VLI vagões. Vale S.A., editor. Belo Horizonte: Vale S.A.; 2013.
- 2 Stein RM. Estudo da técnica de ensaio não-destrutivo ultrassom phased-array [Internet]. Universidade Federal do Espírito Santo; 2017. Disponível em: http://www.engenhariamecanica.ufes.br/sites/engenhariamecanica.ufes.br/files/field/anelxo/11._pg_final_rafael_mattos_stein_2.pdf
- 3 Nazário SLS, Buiochi F, Kitano C, Isepon J dos S, Adamowski JC, Higuti RT. Characterization of bovine milk using ultrasound and artificial neural networks. Rev Control Automação. 2009;20(4):627–36.
- 4 Marcomini KD. Aplicação de modelos de redes neurais artificiais na segmentação e classificação de nódulos em imagens de ultrassonografia de mama. Universidade de São Paulo; 2017.
- 5 Costar LR, Marnet LR, Pinto CFC, Nosedá F, Calôba LP, Soares SD. Identificação do estágio de propagação de descontinuidade em um corpo rígido tubular utilizando emissão acústica e redes neurais. In: 22o Automação & TI, ABM Week. São Paulo; 2018.
- 6 Nogueira R. Métodos de ensaios não-destrutivos mais usuais [Internet]. 2016 [cited 2019 May 28]. Disponível em: <https://bcend.com.br/metodos-de-ensaios-nao-destrutivos-mais-usuais/>
- 7 Diederichs R. Visualization of test results via imaging methods [Internet]. 1999 [cited 2019 May 28]. Disponível em: <https://www.ndt.net/article/pow1297/schmid/schmid6.htm#471>
- 8 Instituto de Ciências Matemáticas e Computação - USP. Redes neurais artificiais [Internet]. [cited 2019 May 28]. Disponível em: <http://conteudo.icmc.usp.br/pessoas/andre/research/neural/>
- 9 Rosenblatt F. The Perceptron: a perceiving and recognizing automaton. Buffalo, New York; 1957.
- 10 Python Software Foundation. About Python [Internet]. [cited 2019 May 28]. Disponível em: <https://www.python.org/about/>
- 11 Pedregosa F, Weiss R, Brucher M. Scikit-Learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
- 12 Scikit-learn_Developers. scikit-learn user guide Release 0.20.2. 2018; Disponível em: https://scikit-learn.org/stable/_downloads/scikit-learn-docs.pdf
- 13 Powers DMW. Evaluation: from precision, recall and F-factor to ROC, informedness, markedness & correlation. Adelaide; 2007
- 14 Scikit-learn_Developers. Dummy estimators [Internet]. [cited 2019 May 28]. Disponível em: https://scikit-learn.org/stable/modules/model_evaluation.html#dummy-estimators
- 15 Joshi R. Accuracy, precision, recall & F1-score [Internet]. 2016 [cited 2019 May 28]. Disponível em: <https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/>
- 16 Shung KP. Accuracy, precision, recall or F1? [Internet]. 2018 [cited 2019 May 28]. Disponível em: <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>