Journal of Applied Research and Technology

www.jart.icat.unam.mx

Journal of Applied Research and Technology 21 (2023) 730-741

Original

Eddy currents assessment of rail cracks using artificial neural networks in a laboratory setup

M. Gutiérrez[a]* • J. Fava[a] • J. Vorobioff[b] • F. Checozzi[a] • M. Ruch[b] • T. Di Fiore[a]

[a]*Universidad Tecnológica Nacional, Facultad Regional Haedo, Buenos Aires, Argentina*
[b]*Comisión Nacional de Energía Atómica, Buenos Aires, Argentina*

Received 05 24 2021; accepted 01 13 2023
Available 10 31 2023

Abstract: Although flaws associated with rolling contact fatigue and the corresponding traffic induced damage, which are a cause of failure in railways, have been of great concern in railway system maintenance and safety strategies in many countries for at least two decades, this serious problem has not been yet adequately tackled in the Argentine railway system. The present upgrading activities undertaken in the Argentine railway system (in infrastructure and in rolling stock) are prompting the need for R&D in non-destructive testing techniques and procedures, to satisfy requirements of the new rolling stock and to ensure safe and economic operation of passengers and cargo. Rolling contact fatigue damage appears as surface and near surface defects and grows into cracks which in time will propagate along the running surface and through the cross-section. Eddy current testing is a very efficient in-service inspection method for this task, the near field technique being especially recommended for ferromagnetic components.

In the present paper, an artificial neural network method for automatic classification of flaws with lift-off compensation is presented and evaluated. The tests consist of the eddy current nondestructive evaluation of right-angle artificial cracks concerning the rolling surface of the railhead on a rail calibration coupon; the depth of the cracks studied ranged from 1 to 7 mm. The technique permitted to compensate the weakening of the signals caused by the lift-off effect, allowing signal cracks classification with lift-off variations of up to 5.4 mm. The effect of crack skewness on the eddy current signals is also studied. Because the rolling contact fatigue cracks penetrate the rail at oblique angles, (10° to 30° to the rolling surface), an additional uncertainty component is added to the experiments if calibration is made with a piece having perpendicular cracks. To estimate the additional uncertainty associated with the artificial neural network method presented here, further tests were made with a second calibration piece with cracks at 25° to the surface. Comparison of results showed that the peak-to-peak amplitudes for both types of cracks are not equivalent at all the tested depths.

Keywords: Eddy current testing, flaw evaluation, artificial neural network, signal processing, Railway infrastructure, head checks

*Corresponding author.
E-mail address: mgutierrez@frh.utn.edu.ar(M. Gutiérrez).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

1. Introduction

One of the main causes of in-service failure of railway tracks is the propagation of defects on the running surface originated by rolling contact fatigue (RCF) on the wheel/rail interface (Australian Rail Track Corporation LTD, 2006; Magel, 2011). RCF is a gradual damage process consisting in the formation and growth of cracks initiated on the running surface of the rail (or a few mm beneath the surface - subsurface damage) until it breaks. This process develops under the influence of cyclic traffic loads, transferred to the rail through the contact surface between wheel and rail. This effect is increased by eventual sliding of the wheel and the presence of liquids (water, lubricants, etc.) and grease (Magel, 2011). Therefore, the deformed surface layer of the rail hardens with respect to its original condition. Finally, when the material is fatigued, incipient flaws will appear and turn into cracks, which, as rolling cycles continue, will propagate along the running surface and through the cross-section. This is the typical growth of fatigue damage. The gauge corner of the outer rail in curves is particularly affected by this phenomenon because the wheel flange is pushed towards it. In the case of a straight length of rail, the contact zone is the running surface itself; here, although the stresses are different and smaller than in the curves, the phenomenon can also occur.

Head checks (HC, or just checking) are one of the most common RCF damage types. They appear on the railhead surface as a pattern of small cracks, as shown in Figure 1 (a). This figure shows incipient HC in railway tracks of the Sarmiento line near Haedo station, Buenos Aires, Argentina. It is known that HC extend towards the inside of the rail at an angle between 10° and 30° to the surface in the running direction (ARTC. LTD, 2006), with an average penetration angle of 25°, (Dey et al., 2014), Figure 1 (b) is a schematic of this skewed penetration. Besides, large amounts of HC appear distributed throughout the affected zone, Figure 1 (a), for hundreds of meters even kilometers of rail tracks. At advanced growth stages, they may cause scaling as large as 15-10 mm wide and 3 mm deep (ARTC. LTD, 2006); the most dangerous case is when HC propagate along the cross-section, which can result in fatigue break of the rail. The depth of HC (crack depth in Figure 1 (b)) cannot be estimated by visual inspection of the surface, nor by inspection with magnetic particles or dye penetrants; a more specific non-destructive testing (NDT) method is necessary.

To minimize operational risks, owners of rail infrastructure perform grinding processes as preventive maintenance against RCF, a problem that may be the cause of premature re-railing or of in-service rail failure and which is a serious concern in the maintenance of modern intensively used railway tracks. The importance of its detection and control received a strong impulse worldwide after the fatal accident in Hatfield, England, in 2000, where it was determined that the accident was caused by defects originated in RCF (Doherty et al., 2005). Flaws of this type have been observed in the main Argentine railway lines, and many more can appear when traffic frequency increases because of the updating of the rail infrastructure.

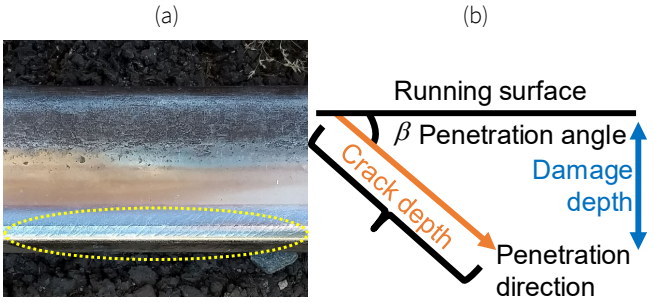

Figure 1. (a) Gauge corner head check, railway track at Haedo Station, Argentina. (b) Schematic of HC skewed penetration on the railhead.

In the context of in-service NDT of railway tracks, it is important to develop inspection techniques, which require short service interruptions. Ultrasonic testing (UT) on rails is used to inspect welds on tracks, and to detect cracks which had penetrated the railhead or have a considerable size in the web or foot, i.e., internal and/or volumetric defects (penetration greater than 4-5 mm). UT detection of surface cracks smaller than 4 mm is rather difficult, even in laboratory conditions and/or at low speeds (5-10 km/h). Although some high-speed UT techniques have been developed, (Heckel et al., 2009; Thomas et al., 2007) they are optimized for volumetric defects. Besides, inspection speed is limited by sound propagation velocity in the material and rail surface condition and by the need of a coupling medium (generally water).

Eddy-currents have constituted an electromagnetic efficient in-service inspection method for this task for several years (Heckel et al., 2009; Pohl et al., 2009; Thomas et al., 2007); but inspection of fatigue flaws is difficult, because of their small size and surface or subsurface location, and because rails are ferromagnetic (Wright, 2015). Another limitation of eddy current testing (ET) is the variation in the lift-off signal (lift-off is the separation between the sensor and the component under inspection). Despite all this, ET in rails is used for the early detection of RCF type surface defects and the evaluation of their growth. This cannot be achieved with UT, which nevertheless constitutes a powerful method for the detection of other types of defects.

Another electromagnetic method investigated for assessing surface damage on railway tracks is alternating current field measurement (ACFM) (Muñoz et al., 2013; Papaelias & Lugg, 2012); both cited papers use commercial sensors to induce an alternate current in the conductor and

measure the magnetic field components Bx and Bz on the same spot. The authors claim that the signal to noise ratio decreases with the increase of lift-off and given that the same sensor includes the inductor and the measuring pieces, the lift-off variations might lead to erroneous evaluations of defect depth. A method to compensate for these variations would be necessary. Besides, signals from clusters of defects cannot be separated, not even at very low testing speeds, because the ACFM signals are affected by the number of defects and the spacing of the defects in the cluster, (Papaelias & Lugg, 2012).

In Rowshandel et al. (2011) the authors present an automated cart with a robotic arm and an ACFM sensor to inspect the rail autonomously at a maximum speed of 10 km/h and a lift-off between 0 and 4 mm. The goal was to investigate the influence of a change in the lift-off on the magnetic field detected by the ACFM sensor. They made laboratory tests on a piece of rail with 8 mm long gauge corner cracks. They find that the Bx component of the field (which is used to evaluate crack depth) strongly decreases with increasing lift-off, hence, they suggest tests should be made with the smallest possible lift-off, which must be kept constant.

The authors also find some sensor orientation (scanning direction) related signal degradation, which they attribute to geometrical effects of rail profile and to material inhomogeneity due to hardening produced by rail-wheel contact. This orientation dependence is a consequence of the fact that ACFM sensors' response is best when the induced current is perpendicular to the crack. To circumvent this question, the authors propose to make an analysis of the Bz component (related to defect surface length) by means of multiple scans until the most favorable orientation is determined, or to use a sensor array. In Rowshandel et al. (2018), the authors apply artificial neural network (ANN) to size RCF cracks in clusters with ACFM measurements.

They work on a dataset simulated using the finite element (FEM) method and constructed with parameters observed in cracks within clusters, to study the influence of these parameters in the ACFM signals. This dataset is used to train and validate the ANN. The following parameters of the cracks in the clusters were considered in the simulations: surface length, crack spacing, number of cracks in the cluster, crack aspect ratio (for elliptical cracks), penetration angle (β, Figure 1 (b)); crack opening: 0.5 mm. In some of the clusters, all the cracks had the same depth, while in other clusters, the center crack was deeper; in the latter, depth relationship between this deeper crack at the center and the shallower cracks was another important parameter.

The output of the ANN trained with these simulated data is "crack depth". The authors compare the crack depth predictions obtained from two different inputs, namely, inputs from the FEM simulations and those from laboratory ACFM experiments on ad hoc constructed clusters of flaws, similar to the simulated ones. The experimental data were obtained manually with an equipment like that used in (Muñoz et al., 2013; Papaelias & Lugg, 2012). The error is about 10 % for both datasets and most of the studied clusters; no procedure to compensate lift-off variations is described. In conclusion, although with this technique it is possible to assess crack depth, a priori information on crack parameters - the input to the model - is necessary.

In Wilson et al. (2011) applied pulsed eddy current combined with infrared imaging to visualize RCF cracks in clusters on a piece of rail in a laboratory environment. This method is currently known as pulsed eddy current thermography (PECT), inductive thermography or electromagnetic thermography. The objective is the characterization of skewed RCF defects on the railhead, analogous to those in Figure 1, studying the distribution of heat around such a defect. In Zhu et al. (2020) a novel laboratory research on the application of PECT to the detection and characterization of fatigue RCF cracks is presented. The authors extract eight time and space characteristics from the PECT signals from real and ad hoc (realistic?) defects and study them for the evaluation of defect penetration angles.

This procedure has several steps: first crack shape is reconstructed, then the basic characteristics for the analysis are calculated, and the penetration angles of the ad hoc and real defects are estimated. They propose the most appropriate experimental conditions and characteristics for their analysis, i.e., those with the best correlation with the defect skewness. They point out that further research is necessary to characterize other RCF crack parameters, such as surface length, angle at the surface, crack depth and damage depth (for these last two see Figure 1 (b)).

PECT shows a good potential for quantification of the geometry of subsurface cracks with complex patterns within the clusters (like those in RCF cracks), such as penetration angle and depth. However, its use in an operative railway environment can be hampered by the fragility of the imaging equipment and speed restrictions for the measurements, setting aside the need to evaluate the effect of lift-off variations on the heating mechanisms. PECT could be a good complementary method for a deeper analysis of a region previously identified with another method, such as ACFM, ET or UT.

Some ways for the minimization of the effect of lift-off during field application of the methods discussed above could be: mounting the device on a test cart or the use of a guided damped system of sensors and head on the rail, such as that presented in (Heckel et al., 2009; Thomas et al., 2007). It must also be borne in mind that a satisfactory control of lift-off also depends on the surface condition of the rail (corrugations, profile variations, scaling, etc.). The practical implementation

of any of these methods should include a device for the measurement of lift-off during the inspection, to compensate for possible variations. It must be also born in mind that for a reliable in-service application of any electromagnetic method in a noisy environment, like in service inspection of rail-tracks, unwanted noise must be filtered out.

The basis of ET method is that a coil (or several depending on the technique) powered by an alternating current (AC) induces eddy currents which flow in a thin skin near the surface of the inspected conductor. Reflected impedance due to induced currents is measured through the same coil or another set of coils. This impedance depends on frequency, material electromagnetic properties (magnetic permeability and electric conductivity), microstructure, component and coil geometry, lift-off and component surface conditions (rugosity, corrosion, etc.). Conventional ET equipment uses the AC bridge technique to measure probe's impedance variations (Wright, 2015) produced by discontinuities. Testing proceeds in this way: reference signals in the AC bridge are balanced with coils over a zone with no discontinuities (flaws, edges, etc.), then the surface is inspected. If the probe goes over a discontinuity in the component, it produces an unbalance on the bridge that is registered as a signal. To be able to recognize and evaluate those signals, signals from ad hoc defects on calibration blocks are previously registered. That means that the ET is a comparative method, as are other NDT methods. The usual ET calibration blocks with cracks perpendicular to the inspection surface are not the best representation of a HC crack, Figure 1. Besides, due to the large number of parameters that may affect the indications in an electromagnetic test for clusters of cracks (crack size, spacing and density in a cluster, penetration angle, etc.) as compared to single cracks, calibration curves for single cracks might not be adequate for the assessment of cracks in clusters. However, before using a new technique for clusters, it is necessary to know its performance for individual cracks.

Consequently, an eddy current laboratory prototype applying the near-field technique in the differential send-receive mode was developed. The following objectives were established in this work:

a) to analyze the prediction of an ANN for the detection and automatic evaluation of signals from railway cracks, also allowing for lift-off signal compensation.

b) to evaluate through digital signal processing, the uncertainty involved when calibrating an eddy-current test to inspect HC type notches with perpendicular crack calibration blocks.

The ANN structure used in objective a), for signal sorting and lift-off compensation, is a feedforward multilayer perceptron (MLP) with only one hidden layer. To get the optimum number of neurons in the hidden layer, their number was set from 2 to 15, the ANN was then trained to get the hit rate for each configuration; being optimal an ANN with 8 neurons in the hidden layer. For objective b), the assessing uncertainty from flaws at different penetration angles was accomplished by signal analysis of two different rail calibration blocks: the cracks in one of the blocks are perpendicular concerning the rolling surface of the railhead, while those in the other block penetrate the material at an angle of 25°. As with the ACFM method, in the send-receive technique studied in the present work, an AC current fed into the sending coil, thus assuring a constant incident field.

2. Equipment, materials, and methods

2.1. Experimental set up

The send-receive near-field differential ET technique was used to detect the distortion of the eddy current flow produced by defects. An inspection head with three cylindrical coils was constructed: the sending coil (emitter, Le) in the central position, the two receiving coils (receivers, L1 and L2) symmetrically located and connected in the differential mode, Figure 2. The emitter is fed by a sinusoidal wave Vref, and the receiver coils are connected to two branches of the AC bridge; voltage variations over VL1 and VL2 affect the characteristics of Vdif and Vtotal during the scans, Figure 3. The voltage value at Vdif will change when the receiver coils reach a heterogeneous region.

The value Vdif depends on lift-off; variations of the latter during the tests will affect the defect characterization process. The receiver probes connected in the differential mode will not affect the bridge balance by homogeneous changes in the lift-off under all the inspection-head; but the amplitude Vdif will be affected in the vicinity of a discontinuity. The effects produced on Vdif by those changes in lift-off can be evaluated and compensated by using the amplified voltage Vtotal = VL1 + VL2 (see Figure 3). Vtotal is a compensation parameter, by means of which the amplitude Vdif may be correctly re-established in places where there is a lift-off variation.

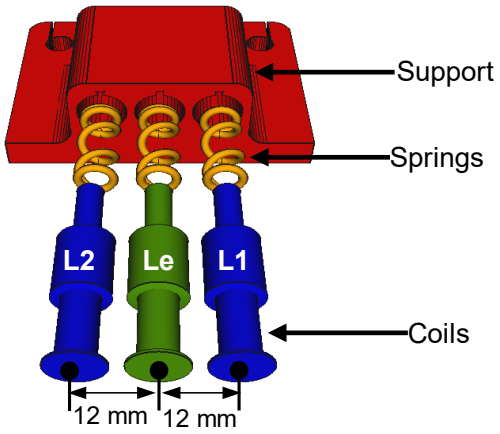

Figure 2. Inspection head in the ET send-receive method.

Tests were made with cylindrical coils wound on plastic acrylonitrile butadiene styrene (ABS) cores with inner radius 4 mm, lift-off 0.4 mm, height 15 mm and 400 turns of #36 AWG copper wire; in order to test if our goals of comparing the signals of two different standards, compensating the lift-off and sorting cracks signals, can be achieved.

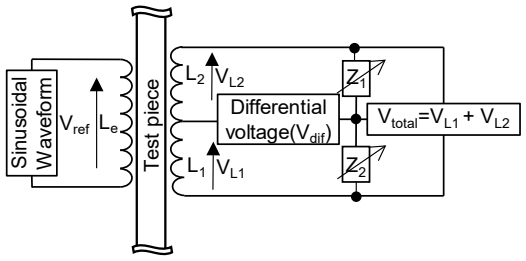

Figure 3. ET equipment. Differential mode, send/receive technique.

The inspection of the test blocks was made with an x-y positioner with a minimum displacement of the order of 0.01 mm. A specially designed spring-loaded coil holder assured a good coupling of the inductors to the blocks as shown in Figure 2. An electronic circuit was designed to control the current through the probe Le, adjust the balance of the AC bridge (Z1 and Z2 impedances) and adequate the voltages Vtotal and Vdif. These signals (Vtotal(t), Vdif(t), and Vref(t)) were sampled during the scanning. The test frequency was set at f=8 kHz (ωref=ωdif=ωtotal, ω=2πf), because at this frequency, all cracks were correctly separated. The differential voltage Vdif of the AC bridge is amplified with monolithic instrumentation amplifier 100 times greater than the input voltage, then this signal was digitized, and processed by the algorithms in the inspection device. A digital lock-in amplifier algorithm demodulates the Vdif and Vtotal signals with the reference Vref input, Figure. 4. For example, the output signal of the multiplier for the case ωdif=ωref, has frequency components at ωref−ωdif = 0 and ωref+ωdif = 2ωref. Both components have the signal of interest Vout, and in this work was implemented an infinite impulse response digital second order low-pass filter (LPF) design by Butterworth method that cancels out the 2ωref component completely, Figure 4. The output of the digital lock-in amplifier for the Vtotal input was called Vadd.

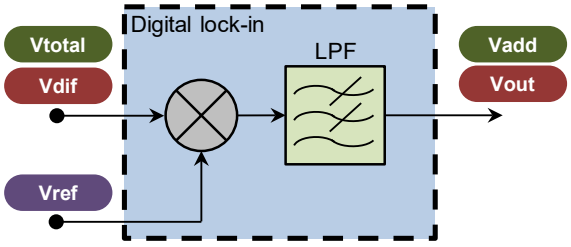

Figure 4. Block scheme of the digital lock-in amplifier.

2.2. Calibration blocks

Two test blocks were constructed on pieces of railway track: cracks were made by electrical discharge machining (EDM) with 0.20 mm diameter wire, the distance between cracks is 80 mm and their depths are 1, 3, 5 and 7 mm respectively, Figure 5.

The cracks in one of the blocks are perpendicular to the rolling surface Figure 5 (a); while those in the other block penetrate the material at an angle of 25°, Figure 5 (b), to simulate the mean penetrating angle of the HC as in Figure 1.

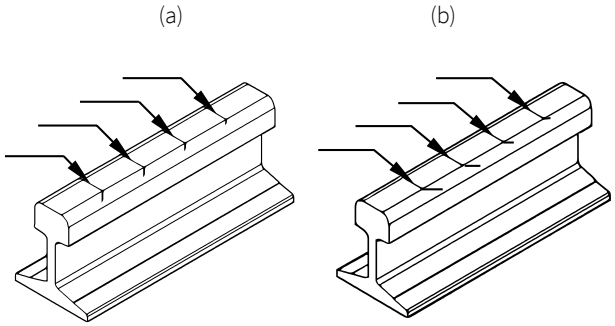

Figure 5. (a) "Perpendicular block" with perpendicular cracks. (b) "Skewed block" with skewed cracks, β=25° in Figure 1.

For the sake of a planar surface, a few mm of the rail surface of both blocks was ground flat before performing the EDM process.

2.3. Signal sorting and lift-off compensation by ANN

2.3.1. Data acquisition and pre-processing

The "perpendicular block" Figure 5 (a) was used for signal sorting and lift-off compensation by ANN, for the 9 lift-off values presented in Figure 6: 0.4 mm (the intrinsic lift-off of the sensor) to 5.4 mm, with 0.5 and 1 mm step. The change to a step of 0.5 mm in the range between 2.4 and 5.4 mm allows for a more precise characterization of lift-off where the changes in Vadd are more gradual, more values being available in this case. The scan for each lift-off value was repeated three times, in order to have more sample vectors with possible fluctuations of the differential voltage Vdif.

As previously mentioned in 2.1, Vdif is amplified, digitized, and processed by the electronics and software in the inspection device, and the signal called Vout is obtained. This signal was normalized to get Vnorm so that the data vector has the same baseline before entering the ANN. Here, normalization consists of the subtraction of a baseline Vout[0], the initial value of the vector signal Vout. So, the Vout vector values were updated using Eq. 1.

$$Vnorm = Vout - Vout[0] \quad (1)$$

As an example, Figure 6 presents the output signals from the 7 mm crack in the perpendicular block at all the available lift-off values. Signals from other cracks exhibited similar waveforms, though each had singular characteristics from the particular crack.

To correlate the voltage Vadd with lift-off, (see Figure 4), measurements were made with the lift-off shown in Figure 6, between the coils and the perpendicular block surface on a flaw-free region, Figure 7. In this way Vadd was generated for the different lift-off values; the voltage value at Vadd were obtained with the method described in (Gutiérrez, Fava et al., 2018).

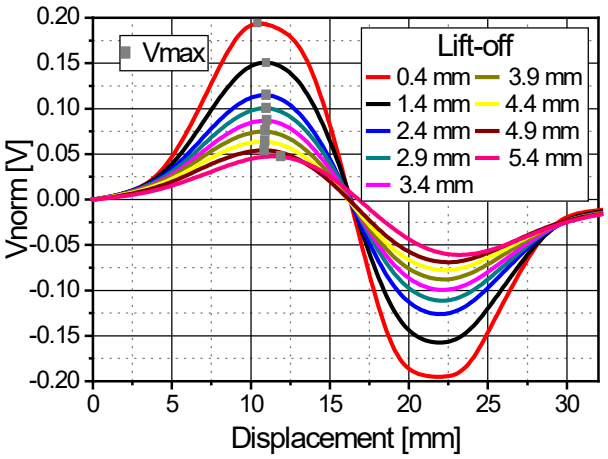

Figure 6. 7 mm deep crack Vnorm.

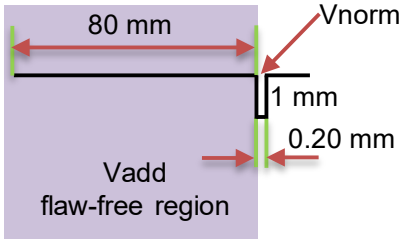

Figure 7. Location of Vadd and Vnorm at the 1 mm deep flaw in the perpendicular block.

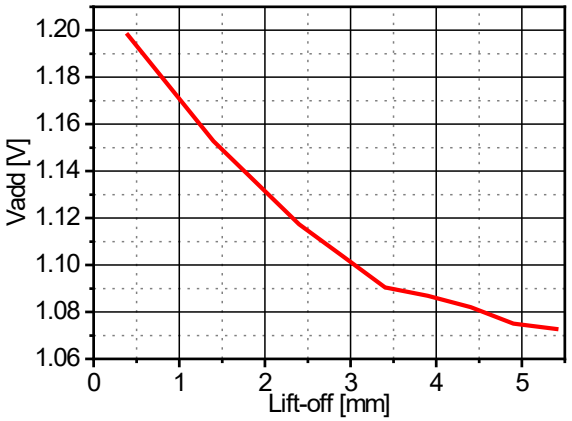

Figure 8. Vadd as a function of lift-off.

This procedure allows obtaining Vadd value immediately before detecting a flaw at Vnorm as shown in Figure 7. The results of Vadd are shown in Figure 8.

2.3.2. Structure and training of the ANN

An ANN was used in the present work for flaw signal classification and lift-off compensation. This is an improvement compared to the curve fitting technique applied in (Gutiérrez, Fava et al., 2018). The ANN method tested here was applied to the perpendicular block in Figure 5 (a). As it was mentioned, Vout is function of lift-off; variation of this voltage during the inspection will affect the correct evaluation of flaw depth. In this work, the input data fed into the ANN are the maximum values of voltage Vnorm, called Vmax (grey markings in Figure 6) and the effective voltage values at Vadd, which are classified according to the crack depths. The input data chosen for this work are the same parameters used in the application (Gutiérrez, Di Fiore, et al., 2018). Unlike the curve fitting method used in (Gutiérrez, Fava et al., 2018), this ANN does not require function parameters that best fit the initial values. Applying ANN to the ET signals, it is possible to classify by flaw depth with another set of input data; for example, in (Wrzuszczak & Wrzuszczak, 2005) amplitude, phase and frequency are used.

The structure used here is a feedforward MLP with only one hidden layer and one output, as shown in Figure 9.

The net input to the jth hidden layer neuron, z_j^{hdn} is defined in Eq. 2.

$$z_j^{hdn} = b_j^{hdn} + \sum_{i=1}^{2} v_{j,i} x_i \,, \mathrm{j} = 0, \ldots, \mathrm{N} \qquad (2)$$

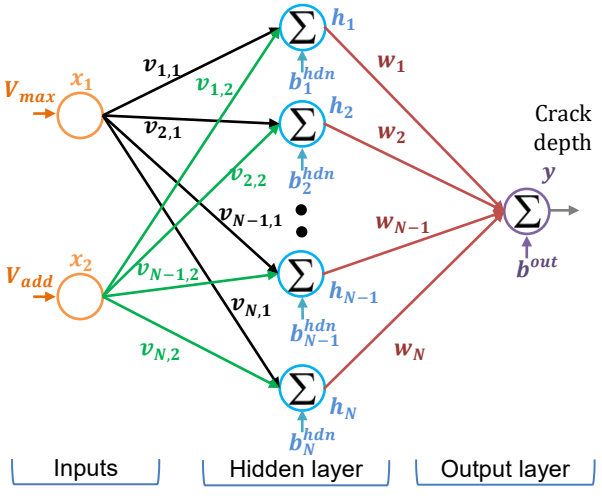

Figure 9. Feedforward ANN. In this work it was found that N = 8 obtains the best results.

Where $v_{j,i}$ represents the weight between the ith input and the jth hidden layer neuron, b_j^{hdn} is the bias for the jth hidden

layer neuron and x_i is the ith input to the neural network. For the network used in this work $x_1 = v_{max}$ and $x_2 = v_{add}$ as shown in Figure 6 and Figure 8. As it is known, (Hertz et al., 1991), for writing Eq. 2 it is necessary to define an auxiliary variable $x_0 = 1$; then the bias can be written: $b_j^{hdn} = v_{j,0}.x_0$.

h_j is the output of the activation function φ_{hdn} for the jth hidden layer neuron, Eq. 3.

$$h_j = \varphi_{hdn}\left(z_j^{hdn}\right) \tag{3}$$

In a similar way, the net input to the output neuron, z^{out}, is defined in Eq. 4.

$$z^{out} = b^{out} + \sum_{j=1}^{N} w_j \ h_j \tag{4}$$

Where w_j represents the weight between the jth hidden neuron output and the output neuron, b^{out} is the bias for the output neuron and $b^{out} = w_0 \ .x_0$.

The crack depth estimation y is obtained after applying the activation function φ_{out}, Eq. 5, as it is seen in Figure 9:

$$y = \varphi_{out}\left(z^{out}\right) \tag{5}$$

Because the ANN will be used for a regression application, a linear activation function was chosen for the output neuron. This choice gives the output a $[-\infty;\infty]$ range. The matrix equations system for this network will be Eqs. 6, 7 and 8:

$$\boldsymbol{x} = \begin{bmatrix} v_{max} \\ v_{add} \end{bmatrix} \tag{6}$$

$$\boldsymbol{h} = \varphi_{hdn}(\boldsymbol{V}\boldsymbol{x} + \boldsymbol{b}^{hdn}) \tag{7}$$

$$y = \boldsymbol{w}\boldsymbol{h} + b^{out} \tag{8}$$

The process of training a neural network involves tuning the values of the weights and biases of the network to optimize network performance by least square's function fitting (appropriated for a regression application, because it avoids learning slowdowns with linear output neurons). Performance is measured by the mean squared error of a set of M samples as a performance function or loss function $\mathcal{L}$, Eq. 9:

$$\begin{aligned} \mathcal{L}(\boldsymbol{p}) &= \frac{1}{M}\sum_{m=0}^{M-1}[t(\boldsymbol{x}_m) - y(\boldsymbol{p}, \boldsymbol{x}_m)]^2 \\ &= \frac{1}{M}\sum_{m=0}^{M-1}[e(\boldsymbol{p}, \boldsymbol{x}_m)]^2 \\ &= \frac{1}{M}\boldsymbol{e}(\boldsymbol{p})^T\boldsymbol{e}(\boldsymbol{p}) \end{aligned} \tag{9}$$

Where $\boldsymbol{p}$ is the network parameter vector (it contains every network weight $v_{j,i}$ and w_j , and every network bias b_j^{hdn} and b^{out}), $\boldsymbol{x}_m$ is the input voltage vector for the m-th sample, $\boldsymbol{t}$ is the network target function (in this case, crack depth), and $\boldsymbol{e}$ is the error vector containing all sample errors.

The Levenberg-Marquardt (LM) algorithm was used to calculate network parameters which inherits the convergence capability of gradient descent while benefitting partially with the speed of Gauss-Newton optimization (Croeze et al., 2012; Ozyildirim & Kiran, 2020). This method is very efficient when training networks which have up to a few hundred weights and if memory storage requirements are relatively less of a concern. While its speed per step is inferior to stochastic gradient descent (and other techniques), it has better convergence properties which makes it better suited for the relatively less complex neural networks used in this work. When using gradient descent, network parameters follow the steepest "downhill" path for $\mathcal{L}$, and hopefully converge to the desired values. This algorithm guarantees some degree of convergence for a small enough step (to avoid issues with steep descents derailing the path towards the solution) at the cost of speed. It is expected initial steps are steeper than those taken when approaching convergence. To find a minimum employing this method, network parameters are changed repeatedly by a step proportional to the negative of the gradient of the performance function, Eq. 10:

$$\boldsymbol{p}_{n+1} = \boldsymbol{p}_n - \mu\nabla\mathcal{L}(\boldsymbol{p}_n) \tag{10}$$

where $\boldsymbol{p}_n$ are the values network parameters take at the n-th gradient descent step, $\nabla\mathcal{L}(\boldsymbol{p}_n)$ is the gradient of the performance function $\mathcal{L}$ evaluated at the n-th step and μ is a parameter that controls step size.

Gauss-Newton assumes that the performance function is approximately quadratic near the optimal parameters and tries to minimize this approximation instead. This algorithm is fast in regions when this assumption holds true, but it may not converge far from there. For a sufficiently small distance $\boldsymbol{\Delta p}$ from optimal network parameters ($\boldsymbol{p}_{opt}$), $\mathcal{L}$ follows the quadratic assumption quite closely, as it is seen in Eq. 11:

$$\begin{aligned} &\mathcal{L}(\boldsymbol{p}_{opt} + \Delta\boldsymbol{p}) \approx \\ &\approx \frac{1}{M}\sum_{m=0}^{M-1}\left[y(\boldsymbol{p}_{opt}, \boldsymbol{x}_m) + \boldsymbol{J}_y(\boldsymbol{p}_{opt})\Delta\boldsymbol{p} - t(\boldsymbol{x}_m)\right]^2 \end{aligned} \tag{11}$$

Where $\boldsymbol{J}_y(\boldsymbol{p}_{opt})$ is the Jacobian matrix of the network function fit $\boldsymbol{y}$ evaluated at $\boldsymbol{p}_{opt}$.

Network parameters in this method are changed by steps calculated, using the Hessian of the performance function evaluated at the $\boldsymbol{n}$ iteration of Gauss-Newton, $\boldsymbol{H}_{\mathcal{L}}(\boldsymbol{p}_n)$, which arises naturally from studying a general second order expansion for $\mathcal{L}$:

$$\boldsymbol{p}_{n+1} = \boldsymbol{p}_n - \boldsymbol{H}_{\mathcal{L}}^{-1}(\boldsymbol{p}_n)\nabla\mathcal{L}(\boldsymbol{p}_n) \tag{12}$$

It can be shown that the gradient and Hessian can be expressed in terms of the Jacobian matrix of the error function $\boldsymbol{J_e(p)}$, (Croeze et al., 2012):

$$\nabla\mathcal{L}(\boldsymbol{p}) = \frac{2}{M}\boldsymbol{J_e^T(p)e(p)} \tag{13}$$

$$\boldsymbol{H_{\mathcal{L}}(p)} = \frac{2}{M}[\boldsymbol{J_e^T(p)J_e(p) + S(p)}] \tag{14}$$

It is assumed that second order derivative terms are negligible when the algorithm converges towards optimal parameters ($\boldsymbol{S(p)} \approx \boldsymbol{0}$), (Croeze et al., 2012). So, combining Eq. 12 with Eqs. 13 and 14 the update of network parameters can be rewritten as:

$$\boldsymbol{p_{n+1} = p_n - [J_e^T(p_n)J_e(p_n)]^{-1}J_e^T(p_n)e(p_n)} \tag{15}$$

Similarly, the Eq. 10 can be rewritten as:

$$\boldsymbol{p_{n+1} = p_n - \left(\frac{1}{\lambda}I\right)J_e^T(p_n)e(p_n)} \tag{16}$$

where $\lambda = 1/\mu$ is the damping parameter and $\boldsymbol{I}$ is the identity matrix.

LM gets a best of both worlds solution by combining Eqs. 15 and 16:

$$\boldsymbol{p_{n+1} = p_n - [J_e^T(p_n)J_e(p_n) + \lambda I]^{-1}J_e^T(p_n)e(p_n)} \tag{17}$$

When λ is large, LM will mostly follow the direction of steepest descent. Conversely, when λ is small LM will follow Gauss-Newton and converge faster towards the desired parameters.

Typical backpropagation algorithms for neural networks deal with Jacobian coefficient calculations efficiently by avoiding explicit numeric derivatives with a smart use of the chain rule and network parameters (Hagan & Menhaj, 1994). Errors for a specific parameter estimation are calculated, and then they are feedback for coefficient calculation.

In this work, the LM algorithm strategy followed was the following:

1. λ is set initially to 0.001, initial parameters are calculated.

2. Calculate $\boldsymbol{e(p_n), J_e(p_n)}$ by backpropagation and after that $\boldsymbol{p_{n+1}}$ using Eq. (17) and $\boldsymbol{\mathcal{L}(p_{n+1})}$ using Eq. 9.

3. If $\boldsymbol{\mathcal{L}(p_{n+1}) < \mathcal{L}(p_n)}$, decrease λ ($\lambda \rightarrow \boldsymbol{0.1\lambda}$) and set $\boldsymbol{p_n \rightarrow p_{n+1}}$, otherwise increase λ ($\lambda \rightarrow \boldsymbol{10\ \lambda}$).

4. Return to step 2 unless convergence has been achieved (if $\|\nabla\mathcal{L}(\boldsymbol{p}_n)\| < \|\nabla\mathcal{L}_{min}\| = 10^{-10}$, $n > n_{max} = 200$), the number of validation failures is more than 1000 or execution time is more than 60 s.

2.3.3. Determination of the number of neurons in the hidden layer

In order to get an adequate model to generalize the problem without overfitting or underfitting, an algorithm was designed to determine the optimal number of neurons in the hidden layer. The algorithm uses cross-validation, setting the number of samples for training, validation and test to 70%, 15% and 15%, respectively. The criterion to stop training was the correct classification of all training and validation samples or maximum processing time of 60 seconds.

Once the ANN has been trained, the hit rate is analyzed as a function of the number of neurons in the hidden layer and the percentage of correctly classified samples. The depth of a flaw was considered correctly classified if it fell within a nominal flaw depth 10% error. To get the optimum number of neurons in the hidden layer, their number was varied from 2 to 15; the ANN is trained to get the hit rate for each configuration. For the sake of statistics, the training of each ANN was repeated 10 times, using the activation functions hyperbolic tangent ($\varphi_{hdn} = \tanh$) and logarithmic ($\varphi_{hdn} = \log$). For the output layer, the linear activation function ($\varphi_{out} = \mathrm{lin}$) was selected.

2.3.4. Bootstrapping for ANN validation

The performance of the ANN was analyzed using the bootstrap method, which resamples the dataset, (Hall, 2013). This was prompted by the fact that in the training step, in which descent gradient algorithms are used, they might get trapped in a local minimum, resulting in inadequate learning, (Marquardt, 1963). The use of a bootstrap scheme enables better prediction hits for each trained ANN. In the present work, the bootstrap technique with balanced resampling was used. That technique generates a new dataset that contains B consecutive copies of the original dataset of size M, then BxM random permutations of the samples are made in this new dataset. The first M elements of the randomized vector are taken for the first bootstrap sample, the next M elements correspond to the second bootstrap sample, and so forth. B new datasets are produced, and each sample will appear exactly B times, considering all the generated datasets. The described method is shown in Figure 10; hence it is possible to generate B models of ANN, corresponding to the B datasets generated. In (Efron & Tibshirani, 1994) it is recommended that B be chosen between 50 to 200 to get a good estimate of the standard error; in the present work, the evaluation was made with B=200.

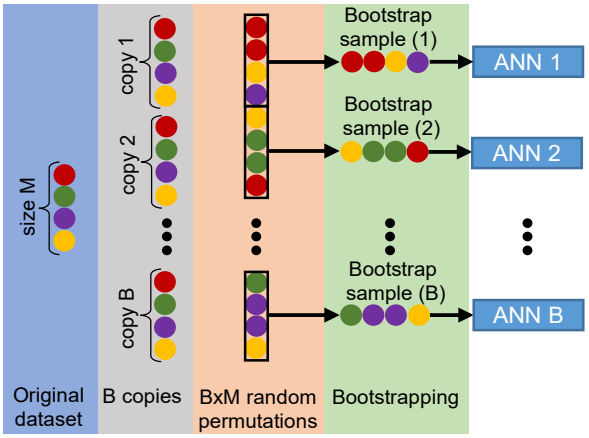

Figure 10. Bootstrap technique with balanced resampling.

2.3.5. Training scheme

The training of an ANN can give different classification results, depending on the initial conditions. To get a global overview of the behavior of the ANN for the set of collected samples, the bootstrap training scheme shown in Figure 11 was applied. For different ANN configurations, several dataset permutations are used, taking a part of the sample vector of Vmax (the maximum value of voltage Vnorm) and the corresponding Vadd. Then bootstrap with balanced resampling is applied to generate the B datasets input to the ANN. The training algorithm will stop only when the criterion mentioned in 2.3.3 is met. The results of the training were evaluated with the root mean squared error (RMSE) for each bootstrap model.

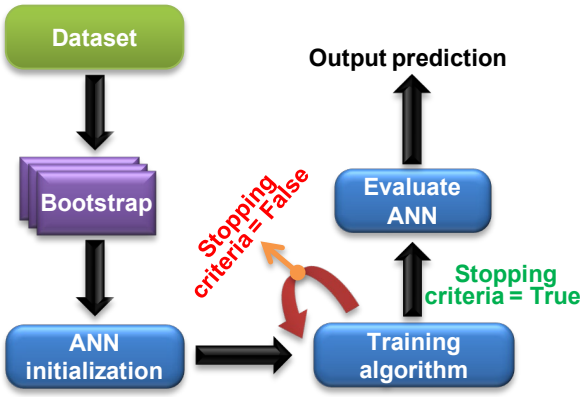

Figure 11. Training scheme for the ANN using a bootstrap resampling technique.

2.4. Evaluation of flaw signals at different penetration angles

As it was mentioned in the introduction, normally, ET calibration blocks are made with cracks perpendicular to the inspection surface; consequently, they are not the best representation of a HC crack, Figure 1. So, with the aim of assessing the uncertainty involved when calibrating skewed crack inspections with a perpendicular crack pattern, both blocks in Figure 5 were evaluated at the same frequency and scan speed than for lift-off compensation and signal sorting by ANN. The same 9 lift-off were applied too: 0.4 mm (sensor intrinsic lift-off) to 5.4 mm. The signals from the perpendicular block and the skewed block (at 25°) were analyzed and compared first. Then with the maximum values of Vout (called Vs-k from now on) of the perpendicular block signals, a calibration curve for flaw depth was adjusted. Last, this curve was used to assess the crack depth at 25°, taking as input their Vs-k. This would give an error in the estimation if flaws skewed to the rail are assessed employing a perpendicular calibration block.

3. Results and discussion

3.1. ANN Signal sorting and lift-off compensation

3.1.1. Classification hit rates

The results of the hit rates as a function of the number of neurons in the hidden layer for both activation functions evaluated are shown in Figure 12. At the beginning of the training with 2 neurons, the hit rate is low, and increases with the number of neurons. Then for 8 to 15 neurons, the hit rate is almost constant, and quite similar for both activation functions. Hence, an ANN with N=8 neurons in the hidden layer with the activation function $\varphi_{hdn} = log$ and for output layer $\varphi_{out} = lin$ was selected.

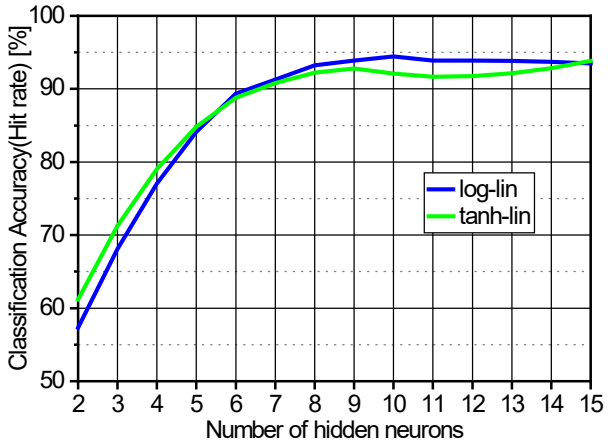

Figure 12. Classification accuracy measure.

3.1.2. Classification analysis

Table 1 presents the results of the classification for the bootstrap sample of size 200 used. The first column lists nominal flaw depths; the second column lists the average of 200 resampling estimates for each flaw. The third column shows the average of the root mean square error (RMSE) of 200 estimations. The decimals in Table 1 are only presented to

show the calculation output. Classification method uncertainty was chosen, for the conditions in the present work, as one half the average values of RMSE in Table 1; this is shown in Table 2.

Table 1. Bootstrap with balanced resampling.

Nominal crack depth (mm)	Mean estimated crack depth (mm)	Mean RMSE (mm)
1	1.007	0.096
3	3.006	0.123
5	5.030	0.157
7	6.965	0.159

Table 2. Uncertainty of the method.

Nominal crack depth (mm)	Mean estimated crack depth (mm)	Uncertainty of method (mm)
1	1.01	0.05
3	3.01	0.06
5	5.03	0.08
7	6.96	0.08

3.2. Evaluation of flaw signals at different penetration angles

The amplitudes of the output signals Vout from both blocks are superimposed in Figure 13. If the maximums Vs-k are considered, only 3 mm crack signals are similar; the signal amplitude for a 1 mm deep perpendicular crack is twice the value of a skewed one. For 5 and 7 mm cracks, the signals of the skewed flaws are 12-13% less than perpendicular flaws.

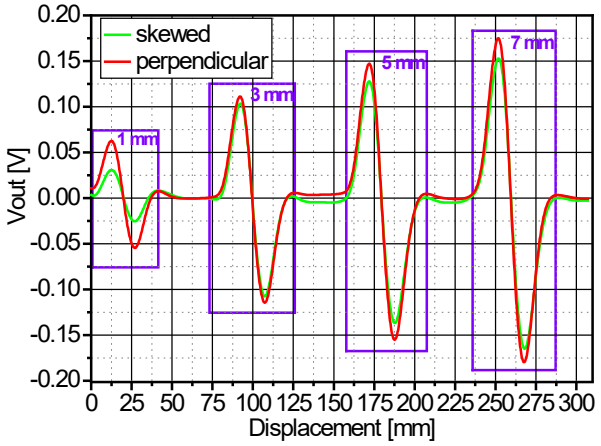

Figure 13. Rail test block output signals measured with an eddy current lab prototype.

Currently, calibration blocks for eddy currents evaluation of cracks are made with slots perpendicular to the block surface, (Pohl et al., 2009). Hence if a perpendicular block, such as that in Figure 5 (a), were used for calibration to assess skewed HC type damage, Figure 5 (b), the results in Figure 13 could be used to estimate the error. Therefore, a "piecewise linear calibration curve" was constructed using perpendicular block Vs-k signals, considering the crack depth as a function of Vs-k from the red curve in Figure 13; the graph of this calibration curve is given in Figure 14.

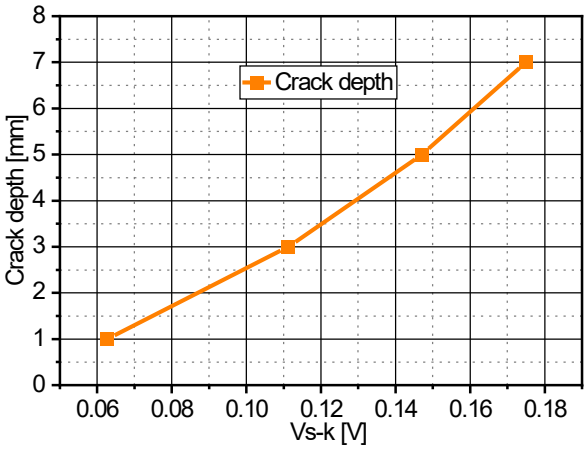

Figure 14. Segmented linear calibration curve: Crack depth as a function of Vs-k from the perpendicular block – red curve in Figure 13.

To construct the segmented curve in Figure 14, linear relationships were considered between pairs of consecutive cracks (in terms of depth), i.e., between 1-3 mm, 3-5 mm, 5-7 mm, represented by the segments joining the corresponding points in this figure. With these three linear relationships and taking the corresponding Vs-k values from the 25° skewed block (maxima of the green curve in Figure 13) as input, the depth of the skewed cracks was evaluated, Table 3. It must be first noted that the skewed 1 mm crack is out of the application range of Figure 14. If a perpendicular block with cracks smaller than 1 mm were not available, an extrapolation should be made towards smaller Vs-k values. Therefore, Figure 14 and Table 3 can be used to estimate the uncertainty component originated when the ANN method presented here is used to evaluate type HC cracks, when the system is calibrated or trained with standards with perpendicular cracks.

Table 3. Evaluation of 25° cracks after a perpendicular crack calibration.

Nominal crack depth at 25° (mm)	1	3	5	7
Assessed crack depth at 25° (mm)	-	2.7	3.9	5.4

4. Conclusions

A lift-off compensation ET technique applying ANN was evaluated in the scanning and evaluation of a rail block with perpendicular cracks. It was possible to restore the signals decrease for the lift-off effect, allowing signal cracks classification. This technique represents an improvement with respect to the curve fitting technique presented in previous works by the authors, (Gutiérrez, Fava et al., 2018). Besides, it can be used as a complement to other EM techniques, such as ACFM or ECPT, or even to other methods such as UT. As is well known the complementarity of NDT produces a synergetic effect and is thus a desired way to obtain the best inspection results.

To contribute to the evaluation of the uncertainty of the studied technique, the signals obtained from a railway block with perpendicular cracks were compared to those obtained from another railway calibration block with cracks penetrating the material at an angle of 25º. Both blocks have four flaws with the same depth of crack (1, 3, 5 and 7 mm). Only the signals for the 3 mm cracks turned out to be similar, the signal amplitude for the 1 mm deep perpendicular crack was twice the value of that for the skewed defect and the signals for the 5 and 7 mm cracks were 12-13% less than those for the perpendicular flaws. Thus, an evaluation of cracks at 25º after a perpendicular crack calibration could give similar mismatches. These conclusions will be important when evaluating type HC cracks with perpendicular calibration blocks or with ANN trained with such blocks.

As future developments, reduce coil and test head size; locate three coils in order to inspect the full rolling surface and the gauge corner; apply this technique in the evaluation of clusters of cracks; study new inductor designs which would allow at least the evaluation of the deepest crack; study the influence of orientation and surface length of the cracks.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgements

Comisión Nacional de Energía Atómica. Facultad Regional Haedo, Universidad Tecnológica Nacional.

Funding

The authors received no specific funding for this work.

References

ARTC, LTD. (2006). Australian Rail Track Corporation, LTD. Some Rail Defects, their Characteristics, Causes and Control. RC, 2400, 48-54.

Croeze, A., Pittman, L., & Reynolds, W. (2012). Solving nonlinear least-squares problems with the Gauss-Newton and Levenberg-Marquardt methods. Portable Document format.

Dey, A., Hintze, H., & Reinhardt, J. (2014). Operation of railway maintenance machines with integrated eddy current technique–an overview of the new requirements in Germany. In Proceedings of the 11th European Conference on Non-Destructive Testing, Prague, Czech Republic.

Doherty, A., Clark, S., Care, R., & Dembowsky, M. (2005). Why rails crack. *Ingenia (23)*, 23-28.

Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press.

Gutierrez, M., Fava, J., Di Fiore, T., Ruch, M., Romero, R., & Vorobioff, J. (2018). Development of a differential test device for eddy current rail inspection. In 12th European Conference on Non-Destructive Testing (12th ECNDT), Gothenburg, Sweden (pp. 11-15).

Gutierrez, M., Di Fiore, T., Vorobioff, J., & Fava, J. O. (2018). Reconocimiento de defectos en rieles ferroviarios por corrientes inducidas aplicando redes neuronales. In Edición 2018. Congreso de Microelectrónica Aplicada. Catamarca, Argentina.

Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE transactions on Neural Networks*, *5*(6), 989-993. https://doi.org/10.1109/72.329697

Hall, P. (2013). *The bootstrap and Edgeworth expansion*. Springer Science & Busine. https://doi.org/10.1007/978-1-4612-4384-7

Heckel, T., Thomas, H. M., Kreutzbruck, M., & Rühe, S. (2009). High speed non-destructive rail testing with advanced ultrasound and eddy-current testing techniques. NDTIP Proceedings, Prague.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). Introduction to the theory of neural computation.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, *11*(2), 431-441. https://doi.org/10.1137/0111030

Magel, E. E. (2011). Rolling Contact Fatigue: A Comprehensive Review. U.S. Department of Transportation. *Federal Railroad Administration*. https://doi.org/10.4224/23000318

Munoz, J. C., Márquez, F. G., & Papaelias, M. (2013). Railroad inspection based on ACFM employing a non-uniform B-spline approach. *Mechanical Systems and Signal Processing*, *40*(2), 605-617. https://doi.org/10.1016/j.ymssp.2013.05.004

Ozyildirim, B. M., & Kiran, M. (2020). Do optimization methods in deep learning applications matter?. *arXiv preprint.* https://doi.org/10.48550/arXiv.2002.12642

Papaelias, M. P., & Lugg, M. (2012). Detection and evaluation of rail surface defects using alternating current field measurement techniques. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *226*(5), 530-541. https://doi.org/10.1177/0954409712444840

Pohl, R., Thomas, H. M., & Casperson, R. (2009). Mögliche Fehlerquellen und deren Einflüsse bei der Risstiefenbestimmung mit Wirbelstrom. In DGZfP-Jahrestagung 2009 (No. DGZfP-BB 115 (Di. 2. B. 2), pp. 1-11).

Rowshandel, H., Papaelias, M., Roberts, C., & Davis, C. (2011). Development of autonomous ACFM rail inspection techniques. *Insight-Non-Destructive Testing and Condition Monitoring*, *53*(2), 85-89. https://doi.org/10.1784/insi.2011.53.2.85

Rowshandel, H., Nicholson, G. L., Shen, J. L., & Davis, C. L. (2018). Characterisation of clustered cracks using an ACFM sensor and application of an artificial neural network. *NDT & E International*, *98*, 80-88. https://doi.org/10.1016/j.ndteint.2018.04.007

Thomas, H. M., Heckel, T., & Hanspach, G. (2007). Advantage of a combined ultrasonic and eddy current examination for railway inspection trains. *Insight-Non-Destructive Testing and Condition Monitoring*, *49*(6), 341-344. https://doi.org/10.1784/insi.2007.49.6.341

Wilson, J., Tian, G., Mukriz, I., & Almond, D. (2011). PEC thermography for imaging multiple cracks from rolling contact fatigue. *Ndt & E International*, *44*(6), 505-512. https://doi.org/10.1016/j.ndteint.2011.05.004

Wright, M. (2015). Eddy current testing technology, 2nd ed., Eclipse Scientific, Ontario, 2015. ISBN: 978-0-9917095-6-4

Wrzuszczak, M., & Wrzuszczak, J. (2005). Eddy current flaw detection with neural network applications. *Measurement*, *38*(2), 132-136. https://doi.org/10.1016/j.measurement.2005.04.004

Zhu, J., Withers, P. J., Wu, J., Liu, F., Yi, Q., Wang, Z., & Tian, G. Y. (2020). Characterization of rolling contact fatigue cracks in rails by eddy current pulsed thermography. *IEEE Transactions on Industrial Informatics*, *17*(4), 2307-2315. https://doi.org/10.1109/TII.2020.3003335