A Survey on the Application of Neural Networks in the Safety Assessment of Oil and Gas Pipelines

Mohamed Layouni
ECE Department
Concordia University
Montreal, Quebec, Canada

Sofiène Tahar
ECE Department
Concordia University
Montreal, Quebec, Canada

Mohamed Salah Hamdi
Information Systems Department
Ahmed Bin Mohammed Military College,
Doha, Qatar

***Abstract*—Pipeline systems are an essential component of the oil and gas supply chain today. Although considered among the safest transportation methods, pipelines are still prone to failure due to corrosion and other types of defects. Such failures can lead to serious accidents resulting in big losses to life and the environment. It is therefore crucial for pipeline operators to reliably detect pipeline defects in an accurate and timely manner. Because of the size and complexity of pipeline systems, however, relying on human operators to perform the inspection is not possible. Automating the inspection process has been an important goal for the pipeline industry for a number of years. Significant progress has been made in that regard, and available techniques combine analytical modeling, numerical computations, and machine learning. This paper presents a survey of state-of-the-art methods used to assess the safety of the oil and gas pipelines. The paper explains the principles behind each method, highlights the setting where each method is most effective, and shows how several methods can be combined to achieve a higher level of accuracy.**

I. Introduction

Transporting oil and gas from remote extraction sites all the way to the consumer is a very delicate and important part of the energy business. The oil and gas industry uses a variety of means to transport its products (e.g., trains, ships, and trucks), but pipelines are thought to be the safest. In Canada for example, 97% of all Canadian natural gas and crude oil production is transported by pipeline [1].

Oil and gas pipelines do have some drawbacks, however. They do go, for example, through fragile ecosystems, or nearby urban areas, which makes any failure or fuel spillage extremely harmful. Moreover, oil and gas pipelines are very complex systems and their inspection is very difficult. They stretch very long distances, and comprise a large number of components, all of which can be a potential source of failure. Their inspection is not a small challenge, and requires highly sophisticated technology. In general terms, the inspection consists of analyzing scans of the pipeline walls, and recognizing certain patterns that would indicate the presence of defects. The vast majority of encountered inspection methods revolve around the use of Artificial Neural Networks (ANNs) [3], [4], [5], and machine learning techniques in general. ANNs are particularly suited for this type of applications thanks to their ability of learn and recognize patterns. In this paper we present a survey of ANN-based techniques used to assess the safety of oil and gas pipelines.

The remainder of this paper is organized as follows. In Section II, we describe the overall pipeline inspection process, and explain the main idea behind Magnetic-Flux-Leakage (MFL) inspection – currently the most common inspection method used in the industry. In Section III, we enumerate the four main classes of pipeline safety-related problems that appear in the machine-learning literature. Section IV is dedicated to one of the most challenging tasks in the pipeline inspection process, namely the sizing of defects from MFL sensor data. We survey a whole set of techniques used to that effect: analytical modeling, numerical methods, algebraic techniques, and show how they can be combined with ANNs to achieve an accurate and computationally-efficient solution. In Section V, we present a slightly different approach that combines analytical modeling and numerical techniques. This approach is less general than the one based on neural networks, but is more computationally efficient. Finally, in Section VI, we summarize our findings, and provide recommendations on the best approaches to assess the safety of oil and gas pipelines.

II. Preliminaries: Pipeline Systems and their Inspection

A. Pipeline Inspection: The Overall Process

The inspection process consists mainly of sending a robot inside the pipeline to scan its walls. The recorded sensor-data is then analyzed in order to detect signs of defects. The oil and gas industry has several scanning techniques at its disposal. These include Ultrasound, X-ray, and Magnetic Flux Leakage (MFL)-based techniques. MFL-scanning is by far the most widely used [1]. MFL-based inspection works as follows. A tool called Pipeline Inspection Gauge (PIG) is sent inside the pipeline. The tool magnetizes the walls of the pipeline, and measures any Magnetic Flux that may leak from the pipe wall. The presence of Magnetic Flux Leakage indicates a possible defect in the pipe wall. (In Section II-B, we explain the principle behind MFL-based inspection, and explain how the recorded MFL data can be used to detect defects and characterize their geometric properties.)

Once a pipeline defect is detected, one has to estimate its length and depth, which in turn will serve to determine its severity level, according to safety standards such as ASME B31G [2].

B. Magnetic-Flux-Leakage Inspection

In MFL-based pipeline inspection, the walls of the pipeline are first magnetized. If a crack or a metal defect is present on surface or inside the wall of a pipe, then a bulging of magnetic

978-1-4799-4509-2/14/$31.00 ©2014 IEEE

flux occurs at the surface of the pipe (See Fig. 1). This bulging of magnetic flux is called Magnetic Flux Leakage (MFL), and is due to the difference in magnetic permeability between the steel of the pipe wall and the air gap created by the defect. MFL-based inspection tools such as the one shown in Fig. 2 are equipped with sensors to measure this flux leakage. Fig. 3 shows a sample MFL diagram, along with pictures of metal defects and their corresponding MFL signatures.

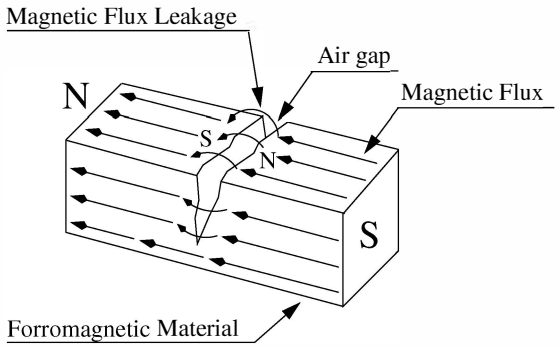

Fig. 1. Magnetic Flux Leakage around a metal loss feature

Fig. 2. MFL-based pipeline inspection tools (Source: Baker Hugues Inc., www.bakerhughes.com)

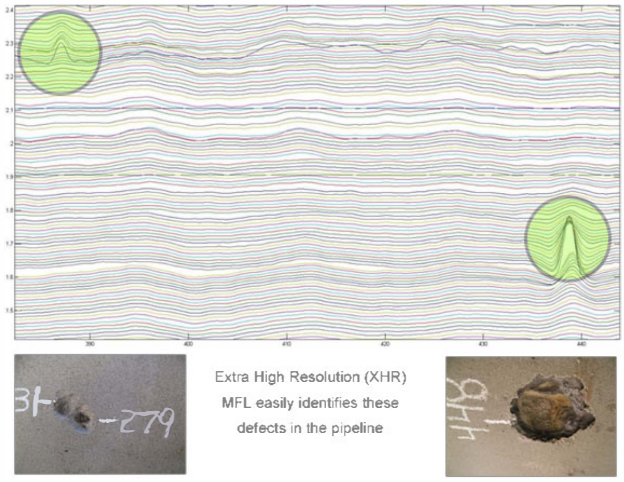

Fig. 3. Sample MFL measurements with corresponding metal loss defects (Source: Pure Technologies Ltd., www.puretechltd.com)

III. ANNs as a Tool for Reasoning about Pipeline Safety

ANNs [3], [4], [5] can be defined as a computational tool that mimics the ability of the human brain to recognize and predict patterns based on learning and recalling processes. It is an effective tool for reasoning and prediction because of its ability to learn from historical data. An ANN is made of a large number of simple processing units, the artificial neurons, that are randomly arranged and connected in different layers (input, hidden and output). Neural networks are especially well suited for solving problems for which no analytical solution is known, or in cases where the analytical solution is too expensive to compute. Sadiq, Kleiner, and Rajani [6] state that ANN is a modeling technique that is useful for applications where causal relationships among variables are unknown. Sawhney and Mund [7] add that ANN is especially useful to represent problems where solutions are not clearly articulated or where the relationships among inputs and outputs are not adequately identified.

Oil and gas pipelines are a notable example of systems that are too complex to model. Coming up with general analytical solutions to reason about their safety has been a real challenge. Neural networks, on the other hand, have been quite effective at solving similar complex problems, and a lot of research effort has been therefore dedicated to the use of neural networks in the assessment of pipeline safety. Below is a brief summary of the literature in this area.

Neural networks in pipeline safety: Four major classes of problems: The use of neural networks in the area of pipeline safety has been quite extensive [8], [9], [10], [11], [12], [13], [14], [15]. Their use spans a large number of applications and problems. The types of problems that were addressed by neural networks, however, can be reduced to **four major classes**:

1) Predicting the probability of failure of a pipeline
2) Predicting the causes of failure of a pipeline
3) Classification of metal defects on a pipeline
4) Detection and sizing of metal defects on a pipeline

In the next sections, we will introduce each one of these classes, and will illustrate it with a representative work from the literature. The subsequent sections of the paper will be dedicated to the fourth type of problems, since it is the most challenging of all four, and also because information about defect sizes is exactly what the pipeline inspection industry relies on to assess the severity of pipeline defects, and ensure the operational safety of pipelines.

A. Predicting the Probability of Failure of a Pipeline

The work in [8] uses *Probabilistic Neural Networks* [3] to predict the probability of failure of pipelines from material properties and dimensions of the pipeline, as well as statistical information on metal defects present in the pipeline. More precisely, the input parameters to the neural network are: the yield strength of the pipe, pipe wall thickness, pipe outer diameter, average crack depth, standard deviation of crack depth, average crack length, standard deviation of crack length, and operating pressure. The proposed PNN shows better precision than other types of neural networks (radial basis, back-propagation neural network, general regression neural network) and more efficiency (fewer iterations) than back-propagation neural networks.

Along the same lines, the authors in [13] use neural networks to predict the reliability of a buried network of

connected pipelines, when subjected to an earthquake. More precisely, following an earthquake, large lateral and axial movements of the soil can lead to a deformation of the walls of the buried pipes, and hence to a reduction of the hollow section of these pipes. This can result in a decrease of the maximum possible flow inside this network of pipes. The work in [13] tries, for different levels of operational performance (i.e., ratio between the maximum allowable flow in the pipeline before and after the earthquake), to estimate the probability of failure of the pipeline network, given input parameters such as:

- **The physical properties of the soil**: angle of internal friction of the soil, the soil's specific weight, the shear wave velocity in the soil, etc.
- **The dynamic properties of the earthquake**: vertical and horizontal soil displacements caused by the earthquake, etc.

B. Predicting the Causes of Failure of a Pipeline

Being able to identify factors that are likely to cause failures in a pipeline is a very valuable tool to the oil and gas industry. The work in [14] describes techniques to develop models that can predict the causes of oil and gas pipeline failures (e.g., mechanical, operational, corrosion, third party, and natural hazards) from data such as: (1) type of product carried by the pipe, (2) pipe location, (3) pipe age, (4) land use and (5) pipe diameter.

The paper develops models based on two different techniques, namely:

- **ANNs.** (Already introduced in Section III)
- **Regression analysis.** Regression analysis [17] is a statistical methodology that utilizes the relationship between two or more quantitative or qualitative variables to predict dependent variables from the independent variables.

The two types of models, obtained through regression analysis and ANNs, were developed based on historical data of pipeline accidents. The two models were able to predict possible causes of pipeline failures with an average validity of 90% for the regression model and 92% for the ANN model.

C. Classification of Metal Defects on a Pipeline

The author in [18] presents a technique to recognize corrosion in oil and gas pipelines, and distinguish it from healthy steel and geometric defects (such as gouging, dents, cracks, etc.). The proposed technique works as follows. First, a model for each of the target classes is developed, namely "corrosion", "geometric defect", and "healthy steel". These models are obtained through the use of parametric modeling and optimization techniques [19]. The class models are then fed into a Fuzzy-Logic classifier, which uses them as a blueprint to detect and classify defects in ultrasonic pipeline scans. The author in [18] reports success rates in the range of 87% for correct corrosion classification, 100% for correct defect classification, and 94% for correct healthy metal detection. Misclassification, and false alarm rates, for defect or corrosion, were at 3% and 5.4%, respectively. According to [18], the level of performance achieved is in line with acceptable industry standards. It remains to be seen however if fuzzy-logic predictors can attain comparable levels of accuracy when considering the more challenging task of estimating the size of metal defects from acoustic or MFL signals.

D. Detection and Sizing of Metal Defects on a Pipeline: The Forward and Inverse Problems

The detection and sizing of pipeline defects is one of the most challenging and valuable tasks when it comes to assessing the operational safety of pipelines. The goal here is to answer the following two questions, just by analyzing MFL sensor data:

- Are there any metal defects on the pipeline? And what is their location?
- What is the topology of those defects (their shapes, lengths, widths, and depths)?

Once the dimensions of a defect are known, safety codes such as ASME B31G [2] are used to determine the severity of the defect, and decide on the urgency of performing reparations.

The problem of defect detection and sizing has been addressed by many researchers in the literature [11], [20], [16], [12]. The general approach has been to split the problem in two smaller sub-problems, and to solve them in sequence. The two sub-problems are namely:

- **The forward problem:** Consists of predicting the 3D representation of the MFL field generated by a known volumetric defect.
- **The inverse problem:** Consists of predicting the 3D location and geometric topology of a defect based on measured MFL sensor data.

The usual approach is to start by studying the forward problem, and then use the gained knowledge (i.e., the understanding of how MFL fields react to defects with different geometries) to solve the inverse problem.

More details on the defect detection and sizing problems are given in Sections IV and V, respectively.

IV. Defect Characterization using the ANN-based approach

Defect characterization consists of (1) detecting the presence of metal defects along the pipeline, (2) locating them, and (3) estimating their geometric shape and size. In this context, we assume a setting where an MFL-based inspection tool, such as the one shown on Fig. 2, travels inside a pipeline and measures MFL signals at the surface of the pipeline wall. For ANN-based defect characterization to work, one needs to:

1) Understand the general behavior of the MFL field.
2) Identify metrics that relate specific geometric properties of the defect to specific properties of the MFL field.
3) Using the above, infer defect topology from MFL measurements.

One of the most prominent approaches to handle defect characterization was put forward by Dutta et al. [11], [20],

[16], and consists of dividing the defect characterization problem into the **forward** and **inverse** sub-problems. They first start by deriving a solution to the forward problem, and then use it to solve the inverse problem. The next subsections describe how these two problems are handled. The approach in [11] is based essentially on ANNs, but uses the analytical modeling and FEM techniques as initial steps (1) to better understand the defect characterization problem, and (2) to generate training data for the neural networks.

A. Solving the Forward Problem

One of the first and obvious approaches that may come to mind when trying to solve the forward problem is to rely on experimental MFL sensor data measured for various defects with different geometric topologies. Unfortunately, this approach, however logical, is not always effective. This is due in part to the complexity of the sensing apparatus, and also to the noise that can affect the actual MFL signal. It is worth noting also that MFL sensor data is not easy to obtain from pipeline operators and inspection service providers due to intellectual property and regulatory constraints. The alternative to the experimental approach is based on analytical modeling and simulation.

Section IV-A1 describes an analytical model of Magnetic Flux Leakage based on the idea of magnetic dipoles [21], while Section IV-A2 presents a modeling approach based on the Finite Elements Method [22].

1) Dipole Modeling of Magnetic Flux Leakage: The analytical modeling of the forward problem is relatively well understood, and relies on the Maxwell equations of magnetism [21]. Dutta et al. [11], [20], [16] make a number of assumptions however in order to derive their model. A summary of those assumptions, as well as a description of the model setting are provided next.

a) Model Setting and Assumptions: The work in [11], [20], [16] considers a ferromagnetic specimen, for example a pipeline wall, with a surface breaking feature (e.g., corrosion, pitting, dent, etc.). The specimen is then subjected to a magnetic field using a device equipped with strong permanent magnets. The magnetic field results in a magnetic flux inside and outside the specimen, which can be quantified by measuring the density of magnetic flux. It is the pattern that the density of magnetic flux follows around the surface of the specimen that indicates the presence of metal loss features. More precisely, the presence of a metal defect results in a magnetic flux leakage (MFL). In [11], [20], [16], Dutta et al. present a model to predict MFL signals from known volumetric defects. Their model is based on Maxwell equations of magnetism [21], and makes the two following assumptions.

1) **Small size defects:** Defects under considerations are small compared to the size of the magnetizing device, and the radius of the pipe. Therefore, it can be safely assumed that the applied magnetic field is uniform in the vicinity of the defect, including the defect cavity. In this case, the magnetizing device need not be part of the system being modeled.
2) **Static setting:** There is no relative motion between the applied magnetic field and the defect, and hence the behavior of the MFL field is quasi-static. In particular free currents that might be possibly induced by the magnet motion are assumed to be null.

b) Analytical Model: The model assumes a cylindrical defect of radius R and depth b, such as the one shown in Fig. 4. Let P denote a point in space above the surface of the ferromagnetic specimen, and let $\mathbf{r} = (x_0, y_0, z_0)$ denote its vector position in the Cartesian coordinate system X, Y, Z with orthonormal basis $(\mathbf{i}, \mathbf{j}, \mathbf{k})$. Assume the applied magnetic field is in the same direction as the Y axis. This will result in the walls of cylindrical defect to be polarized as north for $y < 0$, and south for $y > 0$. In Fig. 4, north poles are marked with a "+" sign, while south poles are marked with a "-". Let $\mathbf{s}$ denote the position vector of a point on the wall of the surface breaking defect, and let $dS(\mathbf{s})$ denote the surface area of the small region around that point (See Fig. 4 for an illustration). Conceptually speaking, one can think of the surface element $dS(\mathbf{s})$ as contributing a magnetic field $d\mathbf{H}(\mathbf{r})$ at $P(x_0, y_0, z_0)$, and that the net total magnetic field at point $P(x_0, y_0, z_0)$ is a summation of all elementary magnetic fields contributed by each of the surface elements found on the defect walls.

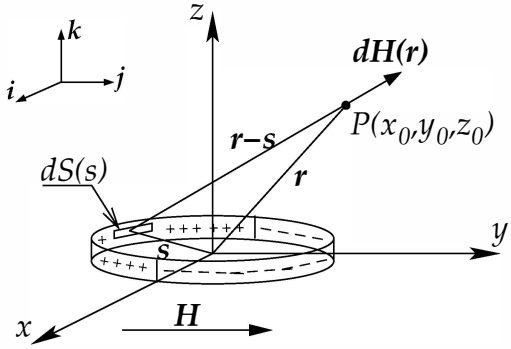

Fig. 4. Model of a cylindrical surface breaking defect [11], [20], [16]

As shown in [11], [20], [16], the density of magnetic flux leakage is given by:

$$\mathbf{B}_{\mathrm{MFL}}(\mathbf{r}) = \frac{\mu_0 M}{4\pi} \int_S \mathbf{n}(\mathbf{s}) \cdot \mathbf{j} \, \frac{\mathbf{r} - \mathbf{s}}{|\mathbf{r} - \mathbf{s}|^3} \, dS(\mathbf{s}) \qquad (1)$$

where μ_0 denotes the magnetic permeability of free space, M the magnetization inside the ferromagnetic material, and $\mathbf{n}(\mathbf{s})$ the surface normal vector at position $\mathbf{s}$.

Let θ denote the angle between $\mathbf{s}$ and unit vector $\mathbf{j}$ (as shown in Fig. 4), and let z_0 be the Z coordinate of surface element $dS(\mathbf{s})$. Then we have

$$\begin{aligned} \mathbf{t} &= \mathbf{r} - \mathbf{s} \\ &= t_x \, \mathbf{i} + t_y \, \mathbf{j} + t_z \, \mathbf{i} \\ &= (x_0 - R\cos\theta) \, \mathbf{i} + (y_0 + R\sin\theta) \, \mathbf{j} + (z_0 - z) \, \mathbf{i} \end{aligned}$$

Let t denote the distance between point P and surface element $dS(\mathbf{s})$

$$t = |\mathbf{t}| = (t_x^2 + t_y^2 + t_z^2)^{1/2}$$

The density of magnetic flux leakage is then given by

$$\mathbf{B}_{\mathrm{MFL}}(\mathbf{r}) = B_x(\mathbf{r})\,\mathbf{i} + B_y(\mathbf{r})\,\mathbf{j} + B_z(\mathbf{r})\,\mathbf{k} \tag{2}$$

where

$$B_x(\mathbf{r}) = \frac{\mu_0 MR}{4\pi}\int_0^{2\pi}\int_{-b}^{0}\frac{t_x}{t^3}\sin\theta\,dz\,d\theta \tag{3}$$

$$B_y(\mathbf{r}) = \frac{\mu_0 MR}{4\pi}\int_0^{2\pi}\int_{-b}^{0}\frac{t_y}{t^3}\sin\theta\,dz\,d\theta \tag{4}$$

$$B_z(\mathbf{r}) = \frac{\mu_0 MR}{4\pi}\int_0^{2\pi}\int_{-b}^{0}\frac{t_z}{t^3}\sin\theta\,dz\,d\theta \tag{5}$$

Equations (2), (3), (4), and (5) represent the analytical model of the Magnetic Flux Leakage density generated by a cylindrical surface breaking defect of radius R and depth b. $B_x(\mathbf{r})$, $B_y(\mathbf{r})$, and $B_z(\mathbf{r})$ are the **tangential**, **axial**, and **radial** components, respectively, of the MFL density field. Analytical models for other simple defect geometries are also available in the literature [12], [11], [20], [16].

2) MFL Modeling based on the Finite Element Method: Models of MFL fields generated by known defect topologies can also be derived by using the Finite Element Method (FEM) [22] to solve Maxwell's equations of magnetism [21]. The FEM-based approach is more general than Dipole Modeling (shown in the previous subsection) however, makes fewer simplifying assumptions, and can be applied to arbitrary defect shapes and sizes (albeit at an additional computational cost). For arbitrarily shaped, and large defects, the following assumptions are not valid anymore.

1) Uniformity of the applied magnetic field in the vicinity of the defect cavity.
2) Uniformity of magnetization on the surface of the defect cavity.
3) Uniformity of magnetic permeability inside ferromagnetic material.

Using the FEM approach, one does not need to make any of the above assumptions. Maxwell's equations, in their general form, are fed into an FEM computation software, along with a representation of the topology of defects under consideration. For large defects, where the applied magnetic field in the defect cavity is not necessarily uniform, the FEM approach makes it possible to model the magnetizing device along with the pipeline and the defect feature. Several FEM computation software are used in practice. The COMSOL Multiphysics® simulation environment [23] and the ANSYS Maxwell® electromagnetic field simulation software [24] are two notable examples of such tools.

MFL field data simulated using the FEM approach is used both to better understand the MFL behavior with respect to different defect shapes, and also as a substitute to experimental sensor data when studying the inverse problem.

B. Solving the Inverse Problem

In [11], Dutta presents two approaches to solving the inverse problem. The first is based on linear algebra and works only for certain configurations of the inverse problem, while the second is more general and uses neural networks. Sections IV-B1 and IV-B3 are dedicated to describing each approach.

1) The Linear Algebra Approach: This approach consists of reducing the inverse problem to a matrix inversion problem. The reduction is obtained as follows. Assuming the setting in Fig 4, the MFL field density contributed by magnetized surface element $dS(\mathbf{s})$ at point $P(x_0, y_0, z_0)$ is given by

$$\begin{aligned} d\mathbf{B}_{\mathrm{MFL}}(\mathbf{r}) &= \frac{\mu_0 M}{4\pi}\mathbf{n}(\mathbf{s})\cdot\mathbf{j}\,\frac{\mathbf{r}-\mathbf{s}}{|\mathbf{r}-\mathbf{s}|^3}\,dS(\mathbf{s}) \\ &= \frac{\mu_0}{4\pi}\frac{\mathbf{r}-\mathbf{s}}{|\mathbf{r}-\mathbf{s}|^3}\,dp(\mathbf{s}) \end{aligned} \tag{6}$$

where $dp(\mathbf{s}) = M \times (\mathbf{n}(\mathbf{s})\cdot\mathbf{j}) \times dS(\mathbf{s})$ can be thought of as the magnetic charge of surface element $dS(\mathbf{s})$. It should be noted however that the concept of "magnetic charge" is just a convenient way to mathematically represent magnetization. Magnetic monopoles have not been shown to actually exist in nature.

It follows from Equation (6) that

$$\mathbf{B}_{\mathrm{MFL}}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int_S \frac{\mathbf{r}-\mathbf{s}}{|\mathbf{r}-\mathbf{s}|^3}\,dp(\mathbf{s}) \tag{7}$$

In practice, the position vector $\mathbf{r}$ in expression (7) is drawn from a discrete domain, called the **measurement domain**, since MFL measurements are made at discrete intervals along the pipeline. Furthermore, because integrals are computed as a discrete sum by numerical tools, position vector $\mathbf{s}$, specifying the position of surface element $dS(\mathbf{s})$, also needs to be discretized. The domain from which $\mathbf{s}$ is chosen is called the **defect domain**. Dutta [11] discretizes these two domains as shown in Fig. 5.

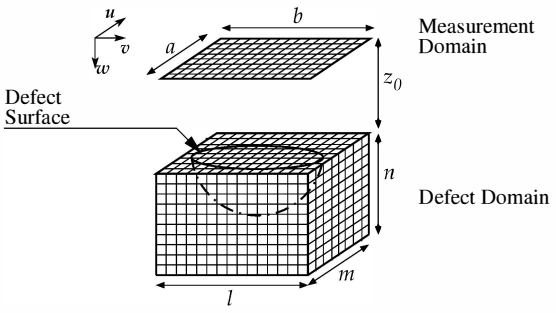

Fig. 5. Discretization of the measurement and defect domains [11]

Let M and N be the sizes of the measurement domain and the defect domain respectively. Using the notation in Fig. 5, it can be stated that $M = l \times m \times n$ and $N = a \times b$. Using the above, the discrete version of Equation (7) can be written as

$$\mathbf{B}_i(\mathbf{r}_i) = \frac{\mu_0}{4\pi} \sum_{j=1}^{N} \frac{\mathbf{r_i} - \mathbf{s_j}}{|\mathbf{r_i} - \mathbf{s_j}|^3} \, dp(\mathbf{s_j}), \qquad 1 \leq i \leq M \tag{8}$$

where $\mathbf{r_i}$ denotes the i^{th} position vector in the measurement domain, $\mathbf{s_j}$ the j^{th} position vector in the defect domain, $\mathbf{B}_i(\mathbf{r_i})$ the i^{th} measurement, and $dp(\mathbf{s_j})$ the magnetic charge at position vector $\mathbf{s_j}$.

For $1 \leq j \leq N$, if $dp(\mathbf{s_j}) \neq 0$ then surface element $dS(\mathbf{s_j})$ is part of the defect surface, and has magnetic charge equal to $dp(\mathbf{s_j})$. If $dp(\mathbf{s_j}) = 0$ then the point specified by position vector $\mathbf{s_j}$ is not part of the defect surface. In other words, identifying the surface of the defect boils down to computing $dp(\mathbf{s_j})$, for $1 \leq j \leq N$.

More precisely, let $\mathbf{G}$ be an M by N matrix, such that for $1 \leq i \leq M$, and $1 \leq j \leq N$, $\mathbf{G}_{i,j} = \frac{\mu_0}{4\pi} \frac{\mathbf{r_i} - \mathbf{s_j}}{|\mathbf{r_i} - \mathbf{s_j}|^3}$. Let $\mathbf{B}$ and $\mathbf{D}$ denote the vectors of size M and N, respectively, containing $\mathbf{B}_i(\mathbf{r}_i)$ for $1 \leq i \leq M$, and $dp(\mathbf{s_j})$ for $1 \leq j \leq N$. Equation (8) can then be written as

$$[\mathbf{B}]_M = [\mathbf{G}^T]_{N \times M} \cdot [\mathbf{D}]_N \tag{9}$$

Solving the inverse problem boils down to finding a vector $[\mathbf{D}]_N$ that satisfies Equation (9). Usually, the size of the defect domain is greater than that of the measurement domain ($N \gg M$). Therefore, for a given measurement vector $[\mathbf{B}]_M$, there may exist several defect shape vectors $[\mathbf{D}]_N$ that satisfy Equation (9). In order to find the actual defect shape, or at least to zero in on the most plausible one, Dutta [11] proposes using *prior knowledge*, about the instance of the problem at hand, as an additional constraint along with Equation (9). This extra knowledge may consist of information about the context of the application, the defect space, or parameters, or any other information that is not already included in matrix $\mathbf{G}$ from Equation (9).

A note on computational complexity: Finding a solution to the inverse problem using the linear algebra approach reduces to solving the linear system in Equation (9). Since matrix $\mathbf{G}$ in Equation (9) is not square, it is not directly invertible. A *generalized inverse* may exist however. Let $\mathbf{G}^*$ denote the generalized inverse of matrix $\mathbf{G}$, if it exists. Then we have

$$[\mathbf{G}^*]_{N \times M} \cdot [\mathbf{G}]_{M \times N} = [I]_{N \times N} \tag{10}$$

where $[I]_{N \times N}$ is the the identity matrix of size N. Matrix $\mathbf{G}^*$ can be obtained using the Singular Value Decomposition (SVD) method [25] for example. The defect geometry is then given by

$$[\mathbf{D}]_N = [\mathbf{G}^{*T}]_{M \times N} \cdot [\mathbf{B}]_M \tag{11}$$

The cost of computing the generalized inverse and recovering the defect vector $\mathbf{D}$ is in the order of $O(N^3 + MN^2) = O(N^3)$, assuming that $M \ll N$. Consequently, for large values of N, the computational cost of the linear algebra method can become prohibitively high. The next section will explore a more efficient approach based on neural networks.

2) Assessment of Previous Approaches: All methods presented so far have limitations when it comes to solving the inverse problem. The analytical method of Section IV-A1, for instance, expresses the MFL field as the value of an integral that depends on the geometry of the defect. More precisely, the MFL field is modeled by an expression of the form

$$\mathbf{B} = \int_S \mathbf{F}(\mathbf{s}) \, dS(\mathbf{s}) \tag{12}$$

where S denotes the surface of the defect, $\mathbf{s}$ the position vector of surface element dS, and $\mathbf{F}$ a function of $\mathbf{s}$. Solving the inverse problem reduces to finding a surface S that satisfies Equation (12) for the measured MFL field $\mathbf{B}$. However, because of the non-injective nature of integrals, there may exist several surfaces, $S_1, S_2, \cdots, S_N$, that satisfy Equation (12) for the observed MFL field. Since it is not possible to come up with a unique solution to Equation (12), one cannot rely solely on the analytical method to solve the inverse problem.

The same limitation, namely of the existence of multiple solutions, applies to the FEM and linear algebra methods (Sections IV-A2 and IV-B1, respectively) which makes them ineffective at solving the inverse problem. In addition, all three methods have high computational costs. The linear algebra method has a computational cost in the order of $O(N^3)$, where N is the size of the defect vector. Similarly, the FEM and analytical methods need to be run several times in the forward direction, on different defect geometries, until the simulated MFL signal coincides with the measured one. Therefore their complexity is at least proportional to the size of the defect space.

3) The Neural Networks Approach:

a) General Overview: Several authors have tried solving the inverse problem using neural networks in a supervised learning setting [3]. The general approach in supervised learning works as follows. In the training phase, the neural network is given a pair of inputs and corresponding outputs. During this phase, with each learning iteration, the neural network adjusts its internal parameters in such a way that it will be able to predict the outputs of entries it has already seen with minimal errors. The underlying assumption is that by the end of the training phase, the neural network will be able to extrapolate knowledge it has already acquired, and correctly guess the outputs corresponding to non-previously seen inputs.

b) Challenges Posed to the Direct Method: Applying neural networks directly to individual data records works effectively for so many applications (e.g., prediction of home prices from size and location), but not for recovering defect geometry from MFL signals. One of the main reasons for that, is that a defect geometry (i.e., the expected output of the neural network) is determined not by individual MFL measurements, but rather by a set of consecutive measurements. In addition, the size of such sets of consecutive measurements, which corresponds in reality to the length of a defect, is variable, and cannot be pre-determined.

c) Feature Extraction as an Enabling Step: To overcome the challenges posed by the direct method, many authors [9], [10], [11], [15] have added an extra step to preprocess the MFL data describing metal features, before feeding it to the neural network. The goal of this step is to derive a compact representation, that can accurately describe the MFL signals being analyzed. This preprocessing step is referred to as **Feature Extraction** in the literature [9], [10], [11], [15], and represents an essential building block in the defect sizing process. In the case of MFL signals analysis, components such as *normalized signal amplitude*, *signal width*, *signal slope*, and *average frequency content* proved to be among the best features to describe the data.

d) Mechanics of the Defect Sizing using Neural Networks: Fig. 6 shows the overall procedure used to predict defect geometries or classes from MFL signals.

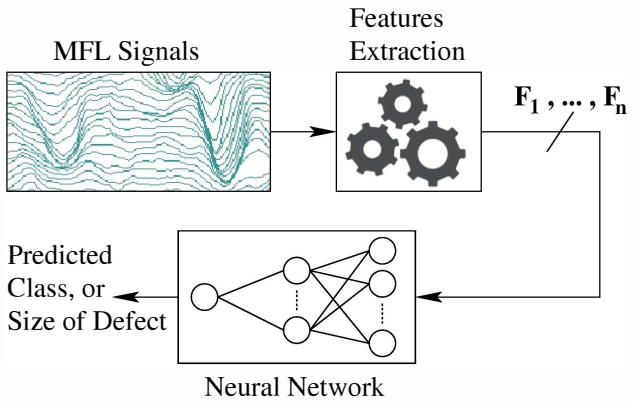

Fig. 6. Predicting defect class and size from MFL signals

The MFL signals at the beginning of the process can be either real experimental measurements, or simulated data using the FEM method or the analytical model. The MFL signals are then fed into a feature extraction engine, which processes them and extracts a compact representation that best describes those signals. The extracted representation, or set of features, is then passed on as an input to the neural network. As in any supervised learning setting, the neural network is provided both with inputs and the proper output (i.e., the right defect class or geometry) during the training phase. Later on in the production phase, the neural network is only provided with features extracted from MFL signals, and it is left to predict the corresponding output on its own.

e) Note on the Performance of Combining Feature Extraction and Neural Networks: The approach of combining feature extraction and neural networks to predict defect classes and geometries appears numerous times in the literature [9], [10], [15], [11], [16]. Most authors who used this method have reported high accuracy rates [10], [15], [16]. For example, the authors in [10], [15] report classification success rates in the 85% to 91% range. Similar accuracy levels are achieved for the defect sizing problem [16], [11]. Across the board, performance levels are influenced by the quality of the extracted features, and in some cases also, by the internal structure of the neural network (e.g., the number of neurons in the hidden layer, and the type of neural network being used: Backpropagation, Radial Basis, Wavelet Basis Function neural networks, etc.)

V. Defect Sizing using the Space Mapping Approach

The authors in [12] propose a slightly different approach to solve the defect sizing problem. The proposed approach is simpler than the one based on neural networks, but less general (i.e., works only for specific geometric shapes). The general idea is as follows. First, compute a coarse approximation of the defect geometry using the analytical model. Then, use a mapping to obtain a more accurate approximation, that is as precise as the solution that would have been obtained using the FEM. The computational cost of this approach, however, is much lower than that of the FEM, and achieves comparable levels of accuracy.

A. Computing the Space Mapping

Computing the space mapping is a one-time procedure, comparable to the initial training phase in neural networks. The space mapping is obtained as follows. First a hypothesis is made about the geometry of the defect (e.g., a rectangular defect with staircase-like slopes). Then the corresponding analytical model is derived. This will result in a parametrized model that depends on the features of the geometric hypothesis. Next, for each experimental MFL measurement available, we do the following steps:

1) Perform an exhaustive search on the defect geometry space, to find the optimal geometric configuration that would minimize the distance between the experimental MFL measurements and the analytical model.
2) Since the analytical model is not sufficiently accurate, repeat the search for the optimal defect geometry again, but using the Finite Elements Method.

All geometric configurations obtained through the analytical model and the FEM are recorded. Next, a mapping is established between the two sets. The result is a space mapping between solutions obtained using the analytical model and the FEM method. Fig. 7 summarizes the overall process.

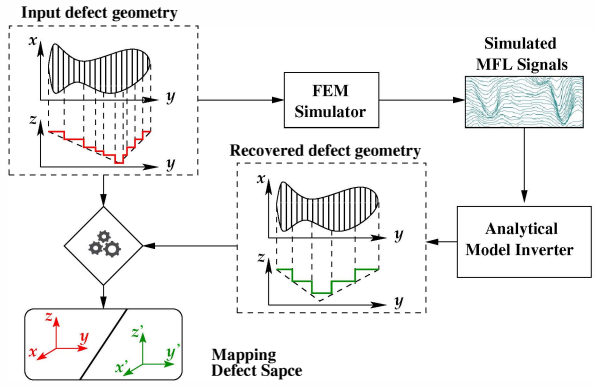

Fig. 7. Defect Space Mapping from analytical modeling and FEM simulation

The Finite Elements Method is more accurate than the one based on the analytical model, but computationally more expensive. The above space mapping makes it possible to obtain an accurate estimate of defect geometries without having to pay the high computational cost of the FEM method. This is done by using the less computationally expensive analytical

model to obtain a first estimate of the defect geometry. The first estimate is then fed to the mapping to obtain a more accurate one. The latter is shown to be as accurate as a solution computed directly through the FEM method [12]. Using this approach the authors in [12] report significant savings in CPU time compared to the Finite Elements Method (10 to 15 times faster in some cases), and an average error as low as 5%.

VI. Summary

The paper presents a survey of machine learning techniques used to assess the safety of oil and gas pipelines. Following a review of the literature, we note four major classes of problems where these techniques are used: (1) the prediction of probability of failure of a pipeline, (2) the prediction of causes of failure, (3) the classification of metal defects, and (4) the detection and sizing of metal defects.

The fourth class is by far the most challenging. A major portion of the paper is therefore dedicated to it. We have first presented the Magnetic Flux Leakage (MFL)-based inspection, and the physical principle behind it. Next, we have explained how MFL measurements can be used to detect and estimate the size of metal loss defects. We then split the defect sizing problem into two smaller, more manageable subproblems: the forward problem and the inverse problem. We have first presented techniques to solve the forward problem, and highlighted their advantages and limitations. The presented techniques are namely the analytical modeling, and the Finite Elements Method. Next, we have considered the inverse problem, and presented two approaches to solve it. The first, based on linear algebra, is conceptually simple but has a high computational cost and a low success rate. The second, based on neural networks, leverages knowledge from the analytical model and FEM-based simulations, and is able to achieve high levels of accuracy. The neural network approach requires some pre-processing work initially to extract features and train the network.

In addition to the neural network-based techniques, we have also presented an approach based on space mappings. This approach combines analytical modeling and FEM simulations. The space mapping approach achieves high levels of accuracy similar to those of the Finite Elements Method, but at a much lower computational cost. Space mapping, however, works only for simple defect geometries. In cases where prior knowledge about the defect shape is available, and where the defect shape is approximable by a simple geometry, it might be appropriate to use the space mapping technique. In all other cases, however, it is better to rely on neural networks. The neural network approach is more general, and can be more convenient to work with in a real-world setting.

Acknowledgment

This publication was made possible by NPRP grant #[5-813-1-134] from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

[1] Canadian Energy Pipeline Association, "Facts about the use of energy pipelines in Canada," http://www.cepa.com/library/factoids, retrieved July 2014.

[2] American Society of Mechanical Engineers, "ASME B31G –Manual for Determining Remaining Strength of Corroded Pipelines," 1991.

[3] S. O. Haykin, *Neural Networks and Learning Machines*. Prentice Hall, 2008.

[4] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2010.

[5] G. Montavon, G. Orr, and K.-R. Müller, *Neural Networks: Tricks of the Trade*, ser. Lecture Notes in Computer Science - Computer State-of-the-Art Surveys. Springer, 2012, vol. 7700.

[6] R. Sadiq, Y. Kleiner, and B. Rajani, "Aggregative risk analysis for water quality failure in distribution networks," *Journal of Water Supply: Research & Technology AQUA*, vol. 53, no. 4, pp. 241–261, 2004.

[7] A. Sawhney and A. Mund, "Adaptive Probabilistic Neural Network-based Crane Type Selection System," *Journal of Construction Engineering and Management*, vol. 128, no. 3, pp. 265–273, 2002.

[8] S. K. Sinha and M. D. Pandey, "Probabilistic Neural Network for Reliability Assessment of Oil and Gas Pipelines," *Computer-Aided Civil and Infrastructure Engineering*, vol. 17, no. 5, pp. 320–329, 2002.

[9] R. Perazzo, A. Pignotti, S. Reich, and P. Stickar, "Feature extraction in mfl signals of machined defects in steel tubes," *AIP Conference Proceedings*, vol. 557, no. 1, pp. 619–626, 2001.

[10] A. R. Ramirez, N. Pearson, and J. Mason, "Automatic classification of defects with the review of an appropriate feature extraction," in *Electromagnetic Nondestructive Evaluation (XI)*, vol. 31. IOS Press, 2008, pp. 148–153.

[11] S. M. Dutta, "Magnetic Flux Leakage Sensing: The Forward and Inverse Problems," Ph.D. dissertation, Rice University, USA, 2008.

[12] M. Ravan, R. Amineh, S. Koziel, N. Nikolova, and J. Reilly, "Sizing of 3-D Arbitrary Defects Using Magnetic Flux Leakage Measurements," *IEEE Transactions on Magnetics*, vol. 46, no. 4, pp. 1024–1033, 2010.

[13] M. T. Roudsari and M. Hosseini, "Using Neural Network for Reliability Assessment of Buried Steel Pipeline Networks Subjected to Earthquake Wave Propagation," *Journal of Applied Sciences*, vol. 11, pp. 3233–3246, Dec. 2011.

[14] A. Senouci, M. Elabbasy, E. Elwakil, B. Abdrabou, and T. Zayed, "A model for predicting failure of oil pipelines," *Structure and Infrastructure Engineering*, pp. 1–13, 2013.

[15] D. Zhang, M. Zhao, Z. Zhou, and S. Pan, "Characterization of Wire Rope Defects with Gray Level Co-occurrence Matrix of Magnetic Flux Leakage Images," *Journal of Nondestructive Evaluation*, vol. 32, no. 1, pp. 37–43, 2013, http://dx.doi.org/10.1007/s10921-012-0156-6.

[16] S. Dutta, F. Ghorbel, and R. Stanley, "Simulation and Analysis of 3-D Magnetic Flux Leakage," *IEEE Transactions on Magnetics*, vol. 45, no. 4, pp. 1966–1972, 2009.

[17] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006.

[18] U. A. Qidwai, "Autonomous corrosion detection in gas pipelines: a hybrid-fuzzy classifier approach using ultrasonic nondestructive evaluation protocols," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 56, no. 12, pp. 2650–2665, 2009.

[19] M. Bettayeb, A. Yamani, U. Qidwai, and C. Chen, "H_∞ Deconvolution of Ultrasonic Non Destructive Testing Signals for Defect Identification," in *Review of Progress in Quantitative Nondestructive Evaluation*. Springer, 1999, pp. 719–725.

[20] S. Dutta, F. Ghorbel, and R. Stanley, "Dipole Modeling of Magnetic Flux Leakage," *IEEE Transactions on Magnetics*, vol. 45, no. 4, pp. 1959–1965, 2009.

[21] A. Aharoni, *Introduction to the Theory of Ferromagnetism*. Oxford Univ. Press, 1996.

[22] G. Strang and G. J. Fix, *An Analysis of the Finite Element Method*. Wellesley-Cambridge, 2008.

[23] COMSOL Inc., "COMSOL Multiphysics®," 2013, http://www.comsol.com/products/multiphysics/.

[24] ANSYS Inc., "ANSYS Maxwell® Electromagnetic Field Simulation Software," 2013, http://www.ansys.com/.

[25] G. Golub and W. Kahan, "Calculating the Singular Values and Pseudo-Inverse of a Matrix," *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, vol. 2, no. 2, pp. 205–224, 1965, http://epubs.siam.org/doi/abs/10.1137/0702016.