Intelligent Microwave Detection of Surface and Sub-Surface Anomalies

by

Abdulbaset Ali

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctoral of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2017

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Samir Trabelsi Lead Scientist
Supervisor(s)	Omar Ramahi Professor
Internal Member	Amir Khandani Professor
Internal Member	Kshirasagar Naik Professor
Internal-external Member	Daniel Stashuk Professor

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

In what follows is a list of publications which I have co-authored and used their content in this dissertation. For each publication, I present a list of my contributions

The use of the content, from the listed publications, in this dissertation has been approved by all co-authors.

1. Ali, A.; Hu, B.; Ramahi, O. "Intelligent Detection of Cracks in Metallic Surfaces Using a Waveguide Sensor Loaded with Metamaterial Elements." *Sensors 2015*, 15, 11402-11416.

    - Developed the idea and performed majority of the experiment work
    - Analyzed the data
    - Developed the AI models
    - Wrote the article

2. Moomen, A.; Ali, A.; Ramahi, O.M. "Reducing Sweeping Frequencies in Microwave NDT Employing Machine Learning Feature Selection." *Sensors 2016*, 16, 559.

    - Performed the experiment work
    - Analyzed the data
    - Developed the majority of AI models
    - Wrote majority of the article

3. A. Ali, A. Albasir and O. M. Ramahi, "Microwave sensor for imaging corrosion under coatings utilizing pattern recognition," 2016 *IEEE International Symposium on Antennas and Propagation* (APSURSI), Fajardo, 2016, pp. 951-952.

    - Performed the simulation
    - Analyzed the data

- Developed majority of the AI models
- Wrote majority of the article

4. A. Ali, M. El Badawe and O. M. Ramahi, "Microwave Imaging of Subsurface Flaws in Coated Metallic Structures Using Complementary Split-Ring Resonators," in *IEEE Sensors Journal*, vol. 16, no. 18, pp. 6890-6898, Sept.15, 2016.

   - Developed the idea and performed the simulation
   - Performed majority of the experiment work
   - Analyzed the simulation and experimental data
   - Wrote majority of the article

5. Abdulbaset Ali and Omar Ramahi,"Microwave Imaging of Subsurface Defects in Coated Metallic Structures Using Small Ring Resonators," The NDT in Canada 2015 Conference, Edmonton 15-17 Jun 2015.

   - Developed the idea and performed the simulation
   - Analyzed the data
   - Wrote the article

Abstract

Microwave near-field testing is a promising nondestructive testing method because of its unique capability to interrogate metallic surfaces and multi-layer dielectric structures. Due to today's need for lighter, stronger, and more-durable materials, enhanced dielectrics are increasingly being used to replace or coat metals. Consequently, conventional testing methods, with their limited penetration, are no longer adequate, but microwave testing sensors transmit signals that can penetrate into dielectrics and so detect surface and subsurface anomalies.

Due to the growing use of microwave near-field sensors for different daily life applications, there is an ongoing need to improve their performance. Recently, artificial engineered electromagnetic materials (metamaterials) have been utilized to demonstrate strong localization and enhancement of electrical fields around sensing elements in order to improve probes sensitivities. Metamaterials are being used to enhance sensors design at the hardware level for better anomaly and flaw detection. Currently, microwave sensors are being used to capture large and complex information, but doing so requires better integration of signal processing methods. Implementing artificial intelligence algorithms to process information collected by microwave sensors can address the challenge associated with information complexity or obscure pattern changes.

To address this gap in microwave near-field evaluation, this study integrates machine learning techniques with microwave near-field testing. Machine learning is a subset of artificial intelligence that denotes a set of methods that can automatically detect patterns in data to build a learning model. The learned model is then used for decision making about unseen data. Employing machine learning techniques for building classification models, this work combines machine learning algorithms with microwave near-field testing. In particular, it aims to build machine learning models that enhance flaw and anomaly detection in microwave near-field testing. The trained machine models can be integrated or embedded in a portable device or rack mounted microwave near-field testing equipment. The value of this approach is confirmed through numerical simulations and laboratory measurements.

Acknowledgements

I would like to express my gratitude to my advisor, Dr. Omar Ramahi and his research group members during the period of 2012 to 2017.

I would like to thank my committee members Dr. Daniel Stashuk, Dr. Amir Khandani and Dr. Sagar Naik for serving in my examination committee and for their feedback. I would like to thank Dr. Samir Trabelsi for participating in my defence exam as the external faculty member.

I owe a debt of gratitude to my mother, siblings, and friends for their support thought out this academic journey.

I would like to express my gratitude to Mary Mcpherson for her help.

Thanks to the administrative staff, Research Infrastructure Support Group at the Electrical and Computer Engineering at the University of Waterloo for their help and support.

This research was supported by the Libyan Ministry of Higher Education.

بسم الله الرحمن الرحيم
و الحمد لله
و الصلاة و السلام على رسوله الكريم وآله و صحبه

In the name of Allah, the most Gracious, the most Merciful
and praise be to Allah
prayer and peace be upon the Messenger of Allah, his family, and his companions

Dedication

dedicated to my parents and siblings

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Microwaves refer to alternating current signals or electromagnetic waves with frequencies between 300 MHz and 30 GHz, with a corresponding wavelength range of 1000 –10 mm. The penetration of microwaves through conducting materials is small due to the skin depth effect; however, they have good penetration in nonmetallic materials. Microwave testing has the advantages of low cost, good penetration in nonmetallic materials, good resolution, and contact-less testing.

1.1 Motivation

During the last decade, the sensitivity of near-field microwave sensors has experienced unique enhancement due to the use of single and double negative materials, commonly referred to as metamaterial. This enhancement takes the form of higher sensitivity, improved resolution and/or lower operating frequency [1, 2, 3, 4, 5]. For instance, Figure 1.1 shows two scenarios of 2-dimensional microwave images generated by x-y scanning of a 1mm crack on a metallic surface using a waveguide sensor. Figure 1.1 (a) is the enhanced image obtained when the sensor is loaded with a single negative material. On the other

hand, Figure 1.1 (b) is the image obtained when the sensor is used directly, and so suffers from low lateral resolution [3]. In another example, loading near-field microwave sensors with metamaterial particles has led to sensor miniaturization, improving the quality factor, and increasing the sensors' sensitivity as in [6]. The use of metamaterial to enhance the sensitivity of near-field microwave sensors is performed in the hardware of the sensors by adding metamaterial particles or layers to confine the electric fields to a small region .

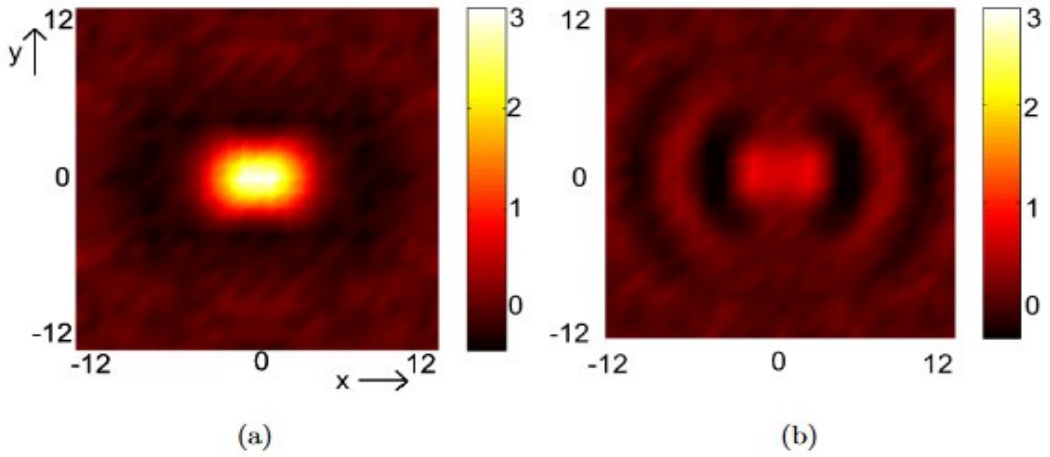


Figure 1.1: Image of 1 mm crack obtained by scanning the waveguide in x-y plane (a) with an 0.85 mm single negative layer having 0.2 loss tangent; (b) same scan without a single negative layer (reproduced with permission) [3].

The encouraging increase in the sensitivity of near-field microwave sensors can be further improved by adding additional enhancement at the signal processing level. Incorporating recent advances of machine learning for processing signals obtained from near-field microwave sensors leads to enhanced sensitivity through the detection of small changes and discovery of hidden patterns.

Based on reviewing recent work about near-field microwave testing, one can conclude the following: first, research interest has been partially focused on the integration of ma-

chine learning and signal processing with near-field microwave testing [7, 8, 9, 10, 11]. Furthermore, the research that utilized machine learning for near-field microwave testing available in the literature is generally directed towards ultra-wide band antennas for tissue imaging or material characterization, and less on anomaly detection. Second, machine learning implementation [7, 8, 9, 10, 11], is limited to one model, such as neural networks or support vector machines. On the other hand, utilizing advanced machine learning techniques like boosting, adaptive boosting or other classifier combination techniques mentioned in [12] can improve machine learning model accuracy and generalization.

Nowadays, due to advances in material science, lighter, stronger, and more durable dielectric materials can be used to replace or coat metals. These new materials require suitable testing approaches, as conventional methods may not be able to inspect these new materials [13]. In recent years, near-field microwave testing and evaluation attracted intense research because it has advantages over the conventional testing and evaluation methods such as eddy current, acoustic emission, and ultrasonic techniques [14, 15, 16]. Some of the advantages of the near-field microwave testing [13, 17] are:

1. Microwave sensors are sensitive to dielectric property variation or buried anomalies and defects.
2. Microwave sensors are highly sensitive to conductor surface properties such as cracks, corrosion, paint, coating , etc.
3. Microwave sensors operate over a large available bandwidth.
4. Near-field microwave measurement systems are relatively inexpensive and adaptable because they are:
    - Robust and repeatable;
    - Non-contact;
    - Compact and small;
    - Low power;

- One-sided/double sided sensing systems.

Signals captured by near-field microwave sensors are mainly processed based on human eye observations using vector network analyzers. Relying on human processing of near-field microwave captured signals greatly limits the capability of such testing due to the following:

1. Small frequency shifts and changes in amplitude or/and phase are usually if not always overlooked.
2. Microwave sensors available nowadays allow the capture of large amounts of information (datasets). However, large datasets are associated with complex and hidden information and require advanced signal processing.
3. Conventional vector network analyzers are needed to plot collected signals limits *in situ* testing.
4. The need for human observations and decisions prevents fully automated testing and evaluation and can be prone to human errors.
5. Building microwave images from collected datasets is not feasible based on human processing.

For instance, crack detection on metallic surfaces using a microwave waveguide sensor in a conventional scenario is based on plotting reflected data on a vector network analyzer screen, and then trained technicians check for changes in the reflection coefficient pattern, which would indicate the presence of a crack [5]. However, if the crack depth is not large, then observation of a change in the reflection coefficient becomes a challenge for the human eye. Thus, the pattern of the reflection coefficient from some cracked surfaces can hardly be distinguished from that in non-cracked surfaces. Figure 1.2 shows four measured reflection coefficients for cracked and non-cracked metallic surfaces. The difference between the first two plots is noticeable, whereas the variation between the last two is subtle and can easily

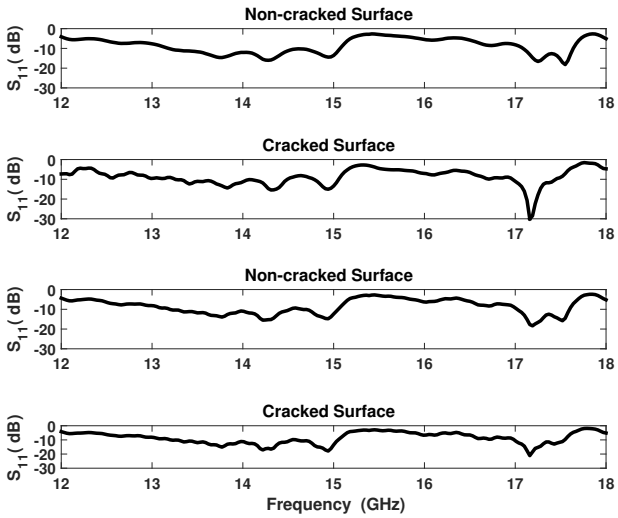


Figure 1.2: Reflection coefficient magnitude from different cracked and non-cracked metallic surfaces.

be overlooked. Therefore, there is a need for a mechanism that enhances testing and evaluation so as to overcome such limitations at the signal possessing level.

Based on the ability of machine learning to automatically detect patterns in data with high sensitivity, this work incorporates different machine learning algorithms for near-field microwave testing in order to address some of the current limitations. The learned machine models can be integrated or embedded in near-field microwave testing equipment, either portable devices for *in situ* testing or rack-mounted ones for laboratory evaluation.

1.2 Contribution

To enhance the performance of near-field microwave testing by addressing some of the issues mentioned in the motivation section, this work integrates machine learning techniques with near-field microwave testing. This integration was intended to contribute the following outcomes:

- Detection of surface and sub-surface anomalies and flaws with improved accuracy

in dielectric materials and metallic structures, using machine learning classification models.

- Construction of microwave images of objects under test for visual anomaly and flaw characterization using single metamaterial particle microwave sensor.
- Building of microwave images of objects under test using a novel pixel classification.
- Use of feature engineering (normalization, cleaning, selection or extraction) of captured signals to address issues arising from using different stand-off distances or background noise.
- Development of machine leaning agents to pave the road towards automating of the microwave testing process to save time and effort.

1.3 Thesis Outline

This thesis is structured as follows: Chapter 1 sheds light on the motivation and contribution of this work. Chapter 2 reviews the relevant literature and discusses some open issues about near-field microwave sensors. Furthermore, Chapter 2 provides background on machine learning, with a focus on supervised learning using neural networks, random forest and support vector machines. Chapter 3 presents experimental work for applying artificial intelligence to improve surface near-field microwave testing performance. Chapter 4 presents experimental measurements and numerical simulation work for implementing intelligent subsurface anomaly detection. Chapter 5 concludes the work and outlines remaining challenges and possible topics for future work.

Chapter 2

Near-Field Microwave Testing: Literature Review

2.1 Introduction

There are several well known non-destructive testing and evaluation (NDT&E) techniques, including acoustic emission testing [18], eddy current [19], pulsed eddy current [20], ultrasonic [21], and magnetic flux leakage testing [22]. Eddy current method is widely used for surface defect detection as it can examine large areas very quickly and does not require the use of coupling liquids. However, eddy current method has some crucial limitations; for instance, it can be used only for electrically conducting materials, the surface under test must be accessible, and it cannot detect defects with large stand-off distance (the distance between the sensor and the material under test). Ultrasonic inspection is limited by high attenuation in the material. Magnetic flux leakage can be used only for alloy and ferromagnetic materials [14].

Removal of surface coating, or using liquids to facilitate defect detection, is an undesirable procedure due to cost, time and environmental concerns [14]. Currently, there is a need for lighter and more durable dielectric materials to replace or coat metals to address

these cost or environmental requirements. These new materials require alternative testing approaches, as conventional non-destructive testing (NDT) methods may not be applicable due to attenuation and scattering caused by various layers, low electrical conductivity associated with the layers, and thin planar anomalies and imperfections that are commonly present in such structures [13]. Alternatively, microwave near field techniques are well suited for testing these structures since microwave and millimeter signals can interact with them and respond to multi-layer structures and any buried defects. For relatively low-loss materials, the penetration depth depends on the operation frequency and the complex permittivity $\epsilon = \epsilon^{'} + j\epsilon^{''}$. This depth is given by

$$\delta_d = \frac{\lambda\sqrt{\epsilon'}}{2\epsilon''}, \tag{2.1}$$

where λ is the operating-frequency wavelength [15].

Near-field (N-F) microwave testing methods are attracting an increasing research interest , as a result of their additional advantages such as remote detection, detection of filled and covered defects, estimation of the physical dimensions and orientation of defects, and ease of operation [23, 24]. For instance near-field open-ended rectangular waveguide technique has been used successfully for detecting surface cracks or disbonds within layered structures [25]. The waveguide sensing method operates by scanning metal surfaces (coated or uncoated) using an open-ended waveguide, while defects are detected by analyzing the reflection coefficient of the incident electric field. Rectangular waveguide sensors need to operate at relatively high frequency to detect small defects [15, 25]. However, with recent advances in metamaterials, near-field microwave sensors have been loaded with metamaterials' structures to enhance sensitivity, as explained next.

2.2 Metamaterial Enhanced Near-Field Microwave Sensors

Metamaterials are artificially engineered electromagnetic materials consisting of periodically arranged metallic elements which are ($\lambda/10$) or smaller of the wavelength of the incident electromagnetic wave in size [26]. These materials can manipulate electromagnetic waves in unconventional ways and demonstrate some unusual electromagnetic properties [26], such as backward propagation, reverse Dopler effect [27], negative refraction [28, 29], imaging beyond diffraction-limit [30, 31], and cloaking [31, 32]. The extraordinary properties of electromagnetic waves in metamaterials strongly depend on the geometry of metamaterial particles [32]. Pendry et al. [28] proposed metamaterials in 1999; later Smith et al. [29] demonstrated experimental work for building metamaterial structure. Since then, metamaterials have attracted increasing interest from various microwave research groups and physicists. As a result applications and implementation of metamaterials have been extended to include a wide range of the electromagnetic spectrum such as far-, mid-, and near-infrared regimes and even optical frequencies [26, 33]. Figure 2.1 shows a picture of the split ring resonator (SRR) stacked with electric LC (ELC) resonators to build up a negative index material sample [34].

An attractive property of metamaterials for NDT is their ability to demonstrate a strong localization and enhancement of fields so that they can be used to improve sensor sensitivity in detecting small anomalies or defects [26, 35]. Different novel or enhanced applications of metamaterials have been proposed recently. For example, Alu and Engheta have proposed a method of dielectric sensing by near-zero narrow waveguide channels [36], and Shreiber, Gupta and Cravey developed a novel microwave nondestructive evaluation sensor by using a metamaterial lens to detect material defects smaller than the operating wavelength [37]. Huang and Yang studied the performance of metamaterial sensors and they showed that the sensitivity and resolution of sensors can be greatly enhanced by metamaterials [26, 38]. Integration of metamaterials and near-field microwave sensors has opened more degrees of freedom in sensor design and sensitivity [26]. Thus, the inter-

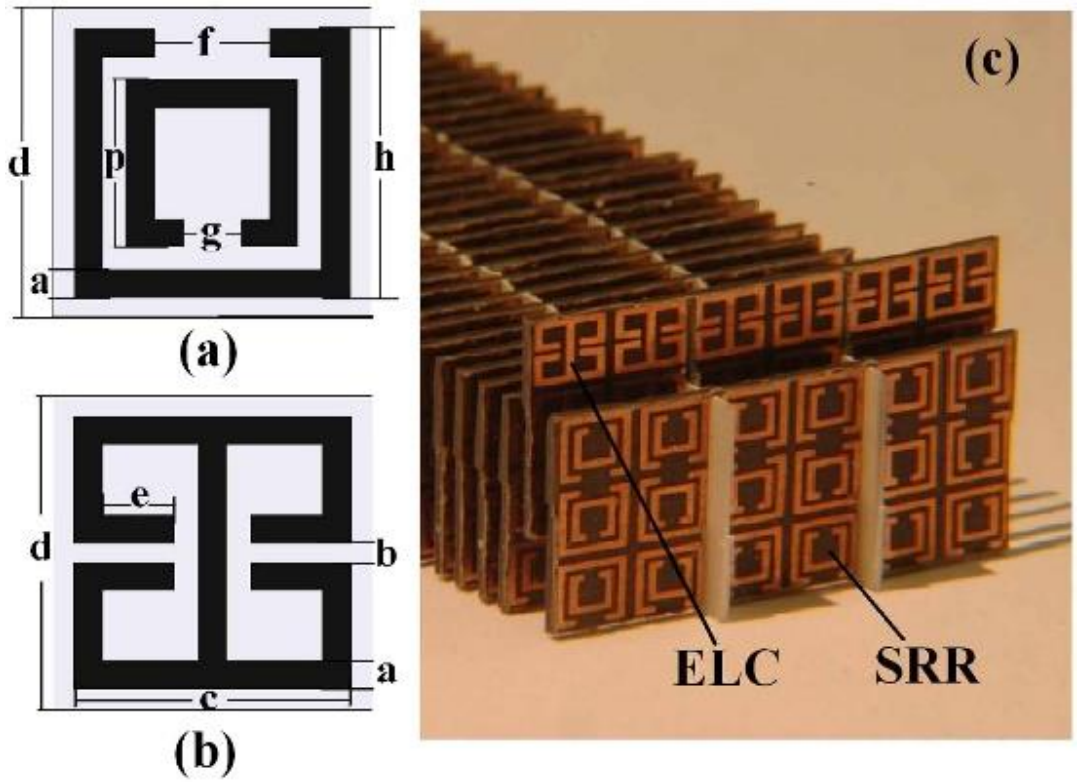


Figure 2.1: (a) split ring resonator dimensions: a=0.3mm, d=3.333mm, f=1.2mm, g=0.6mm, h=2.9mm, and p=1.8mm (b) electric LC resonator dimensions: a=0.3mm, b=0.2mm, e=0.75mm, c=2.9mm, and d=3.333mm. (c) Picture of the split ring resonator-ELC negative index material sample (reproduced with permission [34]).

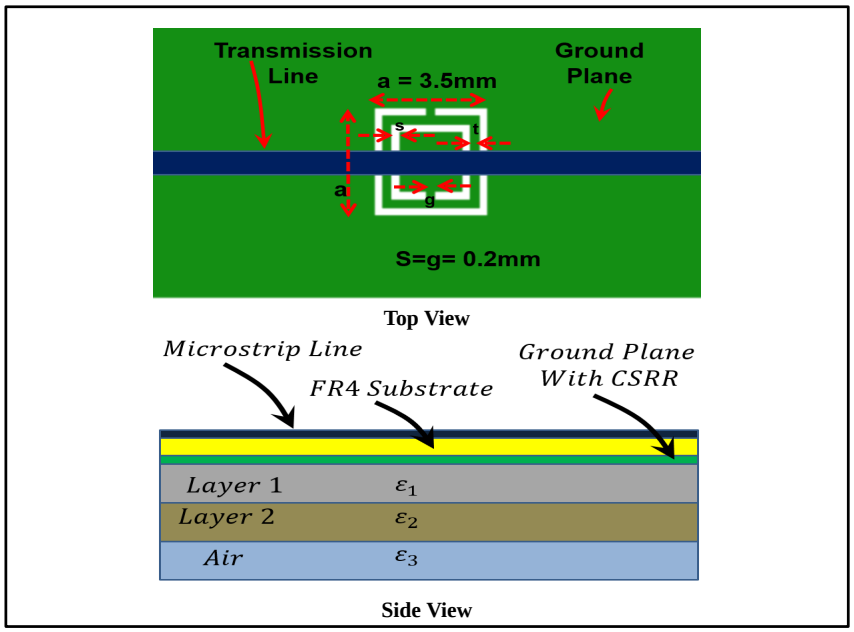


Figure 2.2: Complementary split ring resonator senor layout , and an illustration of the sensor for thickness and permittivity measurement in multi-layered dielectric structures.

disciplinary boundary between the metamaterial field and sensing technology has become a rich area for new technological and industrial developments. For instance, a near field microwave sensor based on metamaterial structure (shown in Figure 2.2) was used for thickness and permittivity measurement in multi-layered dielectric structures in the microwave range [39].

The following points explain how sensing devices can use metamaterials to detect small defects and anomalies [26]:

- Conventional sensors face a trade-off between the need to operate at low frequency to avoid background and substrate absorption, and the need to have a small layout for high resolution. In contrast, metamaterial-loaded sensors can operate at relatively low frequency while having a small layout and field confinement in small areas [26].

- Conventional sensing devices need to produce a strong and measurable readout signal

with a sharp resonant behavior to accurately track shifts in the transmission or reflection signals. Alternatively, metamaterial-loaded sensors, by design, have a higher Q factor than conventional sensors [26].

Based on the outcomes of the above mentioned studies and other similar works available in the literature it is clear that metamaterials have added a new dimension to near-field microwave sensors by improving their sensitivities and resolution to detect small or hidden flaws and anomalies. The encouraging enhancement of the sensitivity in near-field microwave sensors due to utilizing metamaterial particles has opened a door for the need for advanced signal processing techniques in near-field microwave testing. This need can be addressed by integrating advanced machine learning methods to process signals (data) obtained from near-field microwave sensors.

Sensor development depends on the following two main aspects: investigative instruments and data analysis [40] . The advanced investigative instruments available nowadays allow the capture of large amounts of data about materials or objects under test. However, the obtained data may contain complex information or small variations. That being the case, sensor systems increasingly depend on advanced data processing techniques to make decisions and predictions based on trained classification and regression models. Machine learning is becoming important for developing novel sensors and effectively helps with data analysis and mining to accomplish correct predictions and knowledge modeling [40]. In general, the machine learning process can be divided into three phases: data pre-processing, feature extraction and dimension reduction, and system modeling and tuning [40, 12]. The next section provides a background on machine learning.

2.3 Integration of Near Field Microwave Sensors and Machine Learning

Machine learning is a subset of artificial intelligence that denotes a set of methods that can automatically detect patterns in data, and then use the detected patterns to build a

model that can predict future data or perform other kinds of decision making under some uncertainty [41].

Machine learning is usually divided into three main types: supervised, unsupervised, and reinforcement learning. In **supervised learning**, the goal is to learn a mapping from inputs **(x)** to outputs (y), given a labeled set of input-output pairs: $D = \{(x_i, y_i)\}_{i=1}^{N}$. Here, D is called the training set, and N is the number of training examples. Each training input $\mathbf{x_i}$ is a d-dimensional vector. Elements of $\mathbf{x_i}$ vector are called features, variables or attributes. In general, $\mathbf{x_i}$ can be an image, a sentence, an email message, a time series, a molecular shape, a graph, etc. Similarly the form of the output or response variable y_i can be a categorical or nominal variable from some finite set, $y_i \in 1, ..., C$, or it can be a real-valued scalar. If y_i is categorical, the problem is known as classification or pattern recognition, and when y_i is real-valued, the problem is known as regression [41]. In **unsupervised learning** approach we are only given inputs $D = \{x_i\}_{i=1}^{N}$, and the goal is to find attractive patterns in the data, which is sometimes called knowledge discovery. This process is a much-less well-defined problem, since we are not told what kinds of patterns to look for, and there is no obvious error metric to use (unlike in supervised learning, where one can compare a prediction of y for a given x to the observed value). The third type of machine learning **reinforcement learning,** is useful for learning how to act or behave based on occasional reward or punishment signals [41].

Supervised learning

Supervised learning is the form of machine learning most widely used in practice and can be either classification or regression type. The coming sections explain both types and give examples from daily-life activities.

Classification

For any classification model, the goal is to learn a mapping from inputs $\mathbf{x}$ to outputs y, where $y \in 1, ..., C$, with C being the number of classes. If $C = 2$, this is called binary

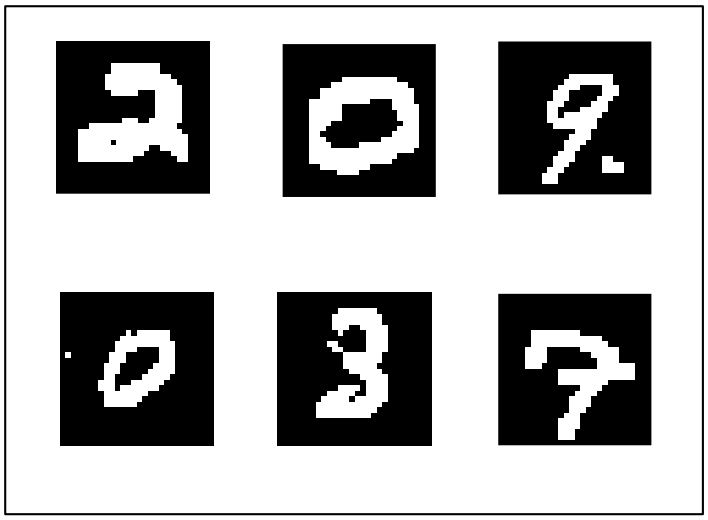


Figure 2.3: First 6 test Modified National Institute of Standards images.

classification, while if $C > 2$, this is called multiclass classification. Classification can be formalized as function approximation. By assuming $y = f(x)$ for some unknown function (f), the goal of learning is to estimate the function (f) given a labeled training set, and then to make predictions using $\hat{y} = \hat{f}(x)$. (The hat symbol is used to denote an estimate.) The final goal of a classifier is to maximize right predictions on new input data (data that the model has not seen during the training phase); this is called model generalization. Image classification for handwriting recognition is a daily life classification example, for instance, handwriting recognition is being used to classify a postal or ZIP code on a letter. One standard dataset used for training and testing handwriting recognition applications is known as Modified National Institute of Standards. This dataset contains gray scale images of the digits 0 to 9 written in different ways. Figure 2.3 shows the first six images of the test dataset.

There are many types of classifiers, and one of their most important distinguishing characteristics is the number of parameters. Classifiers that have a fixed number of parameters are called parametric, whereas classifiers with parameters that grow with the dimension of training data are called non-parametric. Parametric classifiers have the advantage of often

being faster to use, but they have the disadvantage of making stronger assumptions about the nature of the data distributions. Non-parametric classifiers are more flexible, but often computationally intractable for large datasets. For example, Naive Bayes classifier is a simple probabilistic classifier. It is based on Bayes theorem with independence assumptions between predictors and easy to build. Naive Bayesian classifier is simple and useful in performance measurement of more-complex classifiers in working on large datasets. In other words, it can be used as a baseline for more-sophisticated classifiers performance [41, 42]. On the other hand, K-Nearest Neighbors is a non-parametric classifier. K-Nearest Neighbors simply looks at the K points in the training set that are nearest to a new test input [41]. Additionally, Support Vector Machine is a nonparametric model and considered to be among the top classifiers.

Regression

The second type of supervised learning is regression. It is similar to classification, except that response variable is continuous (a real-valued number). Figure 2.4(a) shows an illustration of a linear regression where a single real-valued input $xi \in \mathbb{R}$, are projected to a single real-valued response $yi \in \mathbb{R}$. Furthermore, Figure 2.4 (b) depicts a nonlinear regression model of the same data. Among examples of a regression problem are:

- Predicting tomorrow's stock market price given current and previous market conditions.
- Predicting the price of a house using area, age, district, etc.

In unsupervised learning, data only is given, without any labels. The goal is to discover useful clusters in the data. Unlike supervised learning, the model cannot be told what the desired output is for each input. Instead, an unsupervised model formalizes the task as a density estimation of the form $p(\mathbf{x}_i|\theta)$. Two differences distinguish unsupervised learning from supervised learning. First, unsupervised learning is unconditional density estimation; whereas supervised learning is conditional density estimation. Second, in unsupervised

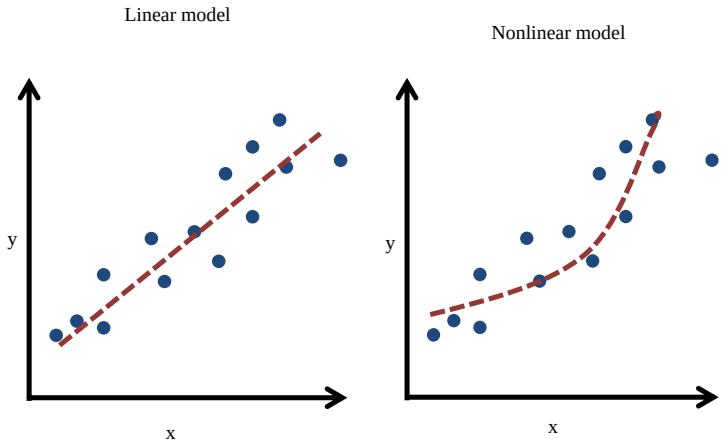


Figure 2.4: (a) Linear regression on 1-d data. (b) Same data with polynomial regression.

leaning multivariate probability models must be built. By contrast, in supervised learning, y_i is usually just a single variable that the model is trying to predict. For most supervised learning problems, univariate probability models (assuming input-dependent parameters) can be used. Unsupervised learning is arguably more typical of human learning. It is also more widely applicable than supervised learning, since it does not require a human expert to manually label the data. Labeled data is not only expensive to acquire, but it also contains relatively little information, certainly not enough to reliably estimate the parameters of complex models [41]. There follow two examples of unsupervised learning:

- Document clustering, the aim is to group similar documents. For example, news reports can be subdivided as those related to politics, sports, fashion, arts, and so on [43].
- Image compression. In this case, a clustering program assigns pixels with similar colors to be the same color to reduce image size [43].

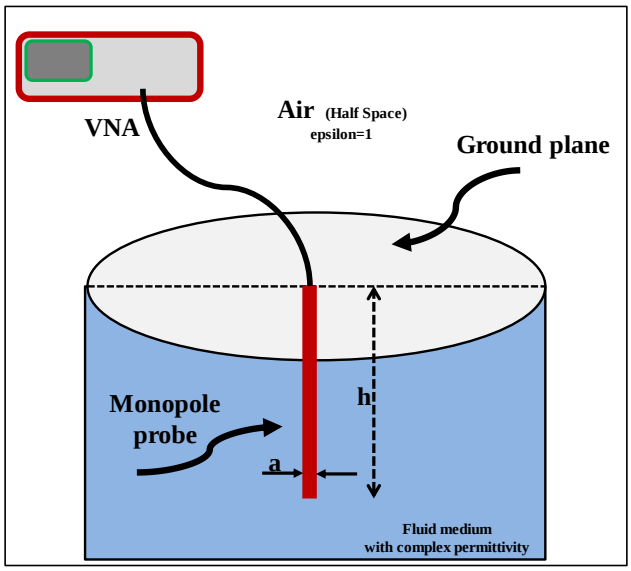


Figure 2.5: Monopole probe used for measurement of complex permittivity utilizing neural networks.

Earlier Works

Microwave near-field testing integrated with machine learning [7, 8, 9, 10] is mainly directed towards human tissue detection or material characterization. Figure 2.5 depicts the set-up of using monopole antenna used for measurement of complex permittivity using artificial neural networks. The configuration incorporates two networks, same in architecture, but differ in weight matrices and outputs. Outputs of both networks then combined together to reconstruct the complex permittivity of the material under test, as in Figure 2.6 [10].

Limited work orientated for integrating machine learning and near field microwave for NDT applications [11, 44, 45]; regardless of, the task of near field microwave sensors in NDT is distinguishing flawed or anomalous structures from healthy or normal structures (which perfectly suites machine learning classification.) Furthermore, machine learning implementation in near field microwave sensing as in [8, 9, 10], has been based on a single AI model. Utilizing advanced machine learning techniques such as boosting, adaptive boosting or other classifier combination techniques as in [12] can improve the accuracy

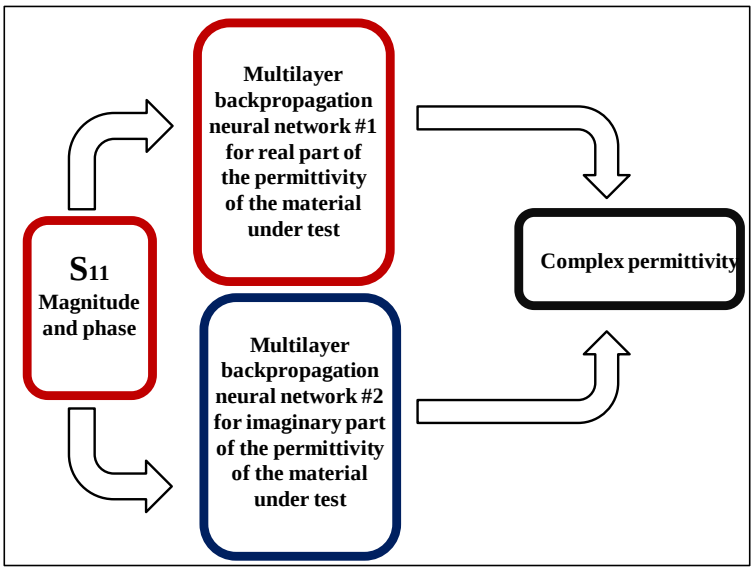


Figure 2.6: Two neural networks combined together to reconstruct the complex permittivity of the material under test.

and generalization of AI models. The implementation of machine learning in this thesis is focused on supervised algorithms of machine learning. Neural networks (NN), random forest (RF), and support vector machines (SVM) are the main algorithms implemented, because they can achieve higher out of sample accuracy rates than other classifiers [46, 45, 11]. Utilizing NN, RF, and SVM requires understating important concepts about them for better model training and parameter tuning. Sections 2.4, 2.5 and 2.6 summarize NN, RF and SVM algorithms as studied in [46, 47, 48].

2.4 Neural Networks

A basic neural network is as a series of functional transformations, where the basic function is a nonlinear function operates on a linear combination of the inputs. For M hidden unit constructing M linear combinations of the input variables $x_1, ..., x_N$ yields the activation a_j , which is given by

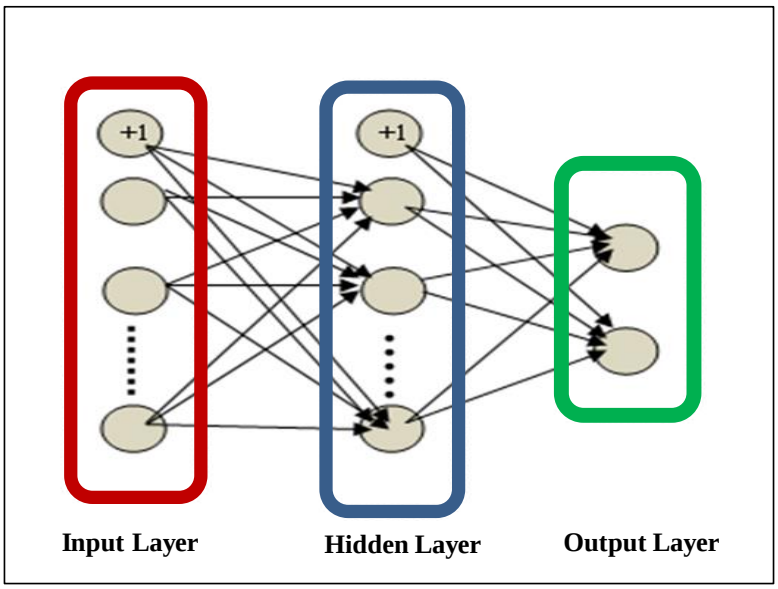


Figure 2.7: Three layer neural network configuration

$$a_j = \sum_{i=1}^{N} w_{ji}^{(1)} x_i + w_{j0}^{(1)} z_j = h(\ a_j) \tag{2.2}$$

where $j = 1, ..., M$ (total number of hidden units) The superscript (1) indicates that the corresponding parameters are in the first layer of the network. Parameters $w_{ji}^{(1)}$ refer to weights and $w_{j0}^{(1)}$ to biases. The activations are then transformed using a differentiable, nonlinear activation function $h(\theta)$, which is given by

$$z_j = h(\ a_j) \tag{2.3}$$

where z_j corresponds to the outputs of the hidden units.

A basic NN model consists of an input layer, a hidden layer, and an output layer as shown in Figure 2.7.

The nonlinear functions $h(\theta)$ are generally chosen to be sigmoid functions; then outputs of the hidden units are linearly combined to give output unit activations

$$a_{jk} = \sum_{i=1}^{N} w_{ki}^{(2)} z_j + w_{k0}^{(2)} \tag{2.4}$$

where $k = 1, ..., K$, and K is the total number of outputs or labels. The transformation in Eq. 2.4 corresponds to the 2^{nd} layer of the network, and the $w_{k0}^{(2)}$ are bias parameters. In the end, the output unit activations are transformed using an appropriate activation function to give a set of network outputs y_k. The selection of the activation function is controlled by the data nature and the assumed distribution of the labels. For multiple binary classification problems, each output unit activation is transformed using a logistic sigmoid function so that

$$y_{jk} = \sigma(a_k) \tag{2.5}$$

where

$$\sigma\,(\ a) = \frac{1}{1 + e^{-a}} \tag{2.6}$$

For multi-class problems, a soft-max activation function of the form Eq 2.7 is used.

$$p\ \ (c_k x) = \frac{e^{a_k}}{(\ 1 + \sum_j e^{a_j}\quad)} \tag{2.7}$$

One can combine equations to get the overall network function, for sigmoidal output unit activation functions, which takes the form

$$y_k\,(x, w) = \sigma\left(\sum_{j=1}^{M} w_{ki}^{(2)} h\left(\sum_{i=1}^{N} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right) \tag{2.8}$$

where, the set of all weight and bias parameters have been grouped together into a vector **w**. Hence, the NN model is a nonlinear function from a set of input variables x_i to a set of output variables y_k controlled by a vector **w** of adaptable parameters. The process

of evaluating Eq. 2.8 can then be interpreted as a forward propagation of information through the network.

In feed-forward architecture each hidden or output unit computes a function given by

$$z_k = h\left(\sum_j w_{kj} z_j\right) \tag{2.9}$$

where, the sum runs over all units that send connections to unit k (the a bias parameter is included in the summation). For a given set of values applied to the inputs of the network, successive application of Eq. 2.9 allows the activations of all units in the network to be computed including the output units.

Network Training

During the network training phase, a training data set $\{x_n\}$ of input vectors with a corresponding true label/target vector $\{t_n\}$,where $n = 1, ..., N$, , the goal is minimizing the following error function

$$E(\mathbf{W}) = \frac{1}{2}\sum_{n=1}^{N} \| y(\mathbf{x}_n, \mathbf{w}) - \mathbf{t}_n \ \|^2 \tag{2.10}$$

In the case of binary classification $t = 1$ denotes class $C1$ and $t = 0$ denotes class $C2$. So a single output unit whose activation function is a logistic sigmoid as in Eq. 2.11 can address the task .

$$y = \sigma\,(\ a) \equiv \frac{1}{1 + e^{-a}} \tag{2.11}$$

So any estimated label $y(\mathbf{x}, \mathbf{w})$ is bounded such that $0 \leq y(\mathbf{x}, \mathbf{w}) \leq 1$. One can interpret $y(\mathbf{x}, \mathbf{w})$ as the conditional probability $p(C1|x)$, with $p(C2|x)$ given by $1 - y(\mathbf{x}, \mathbf{w})$. The conditional distribution of labels given inputs is then a Bernoulli distribution of the form

$$P\,(t|\mathbf{x}, \mathbf{w}) = y(\mathbf{x}, \mathbf{w})^t \{1 - y(\mathbf{x}, \mathbf{w})\}^{1-t} \tag{2.12}$$

Considering a training set of independent examples/samples, then the error function (which is given by the negative log likelihood) can be treated as a cross-entropy error function of the form

$$E(\mathbf{w}) = -\sum_{n=1}^{N} \{t_n \ \ ln(y(\mathbf{x}_n\ , \mathbf{w})) + (1 - t_n \quad) ln(1 - y(\mathbf{x}_n\ , \mathbf{w})\} \tag{2.13}$$

Using the cross-entropy error function instead of the sum-of-squares for a classification problem leads to a faster training and enhanced generalization. In case of K separate binary classifications, then a network with K outputs can be used. Each of the K outputs has a logistic sigmoid activation function and each output is related with a binary class label $t_k \in 0, 1$, where $k = 1, ..., K$. Assuming that the class labels are independent then the conditional distribution of the labels is

$$P(t|\mathbf{x}, \mathbf{w}) = \prod_{k=1}^{K} y_k(\mathbf{x}, \mathbf{w})^{t_k}[1 - y_k(\mathbf{x}, \mathbf{w})]^{1-t_k} \tag{2.14}$$

Taking the negative logarithm of the corresponding likelihood function yields the following error function:

$$EE(\mathbf{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{K} \{t_{nk} \ \ ln(y_{nk} \ \ ) + (1 - t_{nk} \quad)ln(1 - y_{nk} \ \ )\} \tag{2.15}$$

where y_{nk} denotes $y_k(\mathbf{x}_n, \mathbf{w})$.

The derivative of the error function with respect to the activation for a particular output unit takes the form $(y_k - t_k)$ just as in a regression case. In multi-class classification problem each input is assigned to one of K mutually exclusive classes. Thus the binary labels $t_k \in 0, 1$ have a 1-of-K coding scheme indicating the class, and the network outputs are interpreted as $\mathbf{y_k}(\mathbf{x}, \mathbf{w}) = \mathbf{p}(\mathbf{t_k} = \mathbf{1}|\mathbf{x})$,with error function depicted by

$$E(\mathbf{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn} \quad ln(\ \ y_k(\mathbf{x}_n, \mathbf{w})\ ) \tag{2.16}$$

As no analytical solution to the equation $\nabla E(\mathbf{w}) = 0$, we go for numerical procedures. Optimization continuous nonlinear functions is a broadly studied. Most techniques start with some initial value $w(0)$ for the weight vector and then moving through weight space in a succession of steps of the form

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} + \Delta\mathbf{w}^{(\tau)} \tag{2.17}$$

here τ is the iteration step. Algorithms have different options for the weight vector update $\Delta\mathbf{w}(\tau)$. Gradient information is widely used which require the value of $\nabla E(\mathbf{w})$ to be evaluated at the new weight vector$\mathbf{w}(\tau + 1)$ after each update. Next section studies gradient descent optimization basics.

Gradient information optimization

In gradient information simplest approach is to select the weight update in Eq. 2.17 to make a small step in the direction of the negative gradient, such that

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \alpha\nabla E(\mathbf{w}^{(\tau)} \quad) \tag{2.18}$$

where the parameter ($\alpha > 0$) is the learning rate. After each update, the gradient is re-evaluated for the new weight vector in repeated manner. In this context the error function is defined with respect to a training set, thus each step needs that the entire training set to be processed in order to evaluate ∇E. The weight vector is moved in the direction of the greatest rate of decrease of the error function in each step. This approach is known as gradient descent or steepest descent. Techniques that handle the whole dataset at once are called batch methods. For batch optimization, there are more efficient techniques such as conjugate gradients which are much more robust and much faster than simple gradient descent. Unlike gradient descent, in each iteration conjugate gradient methods the error function always decreases, except the weight vector has reached a local or global minimum. It may be necessary to run a gradient-based algorithm multiple times using a

different randomly chosen starting point; to find a sufficiently good minimum. Then the performance compared using an independent validation set. The on-line version of gradient descent is a practical for training neural networks on large data sets. The Error functions based on maximum likelihood for a set of independent examples/samples represent a sum of terms, one for each data point given by

$$E\left(\mathbf{w}\right) = \sum_{n=1}^{N} E_n\left(\mathbf{w}\right) \tag{2.19}$$

The on-line gradient descent (also known as sequential gradient descent or stochastic gradient descent) updates to the weight vector based on one data point at a time, so that

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \alpha \nabla E_n(\mathbf{w}^{(\tau)} \quad) \tag{2.20}$$

The update process is repeated by cycling through the data either sequentially or randomly with replacement. One advantage of on-line methods over the batch methods is the efficient handling of redundancy in the data. Additionally the on-line gradient descent has high possibility of escaping local minima, since a stationary point with respect to the error function for the whole data set will generally not be a stationary point for each data point individually.

Error back propagation

In training a feed-forward neural network the target is to find an efficient technique to evaluate the gradient of the error function $E(\mathbf{w})$. Once the gradient of the error function $E(\mathbf{w})$ at the output layer has been computed, the gradient information is sent backward through the network (this process is known as **error back propagation**). The term back propagation is used to describe the training of a multilayer perceptron using gradient descent applied to a sum-of-squares error function. Generally training algorithms imply an iterative procedure for minimization of error functions, through weight adjustments. Hence, the derivatives of the error function with respect to the weights are needed. Later

the derivatives are used to compute needed adjustments of weights. Derivatives can be calculated using optimization methods. Gradient descent or conjugate gradients are some examples.

Evaluation of error-function derivatives

Considering a simple linear model in which the outputs y_k are linear combinations of the input variables x_i , such that

$$y_k = \sum_i \mathbf{w}_{ki} x_i \tag{2.21}$$

Hence sum of squares error function, for a particular input n, is:

$$E_n = \frac{1}{2} \sum_i (y_{nk} - t_{nk} \ )^{\ 2} \tag{2.22}$$

where $y_{nk} = y_k(\mathbf{x}_n, w)$ and the t_{nk} is the true label of n^{th} example. So the gradient of this error function with respect to a weight w_{ji} is:

$$\frac{\partial E_n}{\partial w_{ji}} = \frac{1}{2} \ (y_{nj} - t_{nj}) x_{ni} \tag{2.23}$$

Eq. 2.23 can be interpreted as a local computation involving the product of an error signal($y_{nj} t_{nj}$) of the link w_{ji} and the variable x_{ni} associated with the input end of the same link. Generally in feed-forward networks each unit computes a weighted sum of its inputs as:

$$a_j = \sum_i w_{ji} z_i \tag{2.24}$$

where z_i is the activation of a unit, or input, that sends a connection to unit j, and w_{ji} is the weight associated with that connection. By introducing an extra input unit with

activation fixed at (+1) the bias terms can be included in this sum. The sum in Eq. 2.24 is transformed by a nonlinear activation function $h(\theta)$ to give the activation z_j of unit j in the form

$$z_j = h(\ a_j) \tag{2.25}$$

Applying the related input vector to the network and computing the activations of all of the hidden and output units by successive application of Eq. 2.24 and Eq. 2.25 is called *forward propagation*. Forward propagation can be observed as a forward flow of information through the network.

Considering evaluation of the derivative of E_n with respect to a weight w_{ji}, the outputs of the various units will depend on the particular input example n. However, E_n depends on the weight w_{ji} only via the summed input a_j to unit j. Therefore one can apply the chain rule to obtain

$$\frac{\partial E_n}{\partial w_{ji}} = \frac{\partial E_n}{\partial a_j}\ \frac{\partial a_j}{\partial w_{ji}} \tag{2.26}$$

To simply the formulas the following notation is introduced

$$\delta_j \equiv \frac{\partial E_n}{\partial a_j} \tag{2.27}$$

where the δ 's are often referred to as errors.

By using Eq. 2.24, the following derivative is obtained

$$\frac{\partial a_j}{\partial w_{ji}} = z_i \tag{2.28}$$

Finally, substituting Eq. 2.27 and Eq.2.28 into Eq. 2.26, yields

$$\frac{\partial E_n}{\partial w_{ji}} = \delta_j\ z_i \tag{2.29}$$

Equation 2.29 indicates that the required derivative is calculated simply by multiplying the value of δ for the unit at the output end of the weight by the value of z for the unit at the input end of the weight ($z = 1$ in the case of a bias unit). Therefore, to evaluate the derivatives, one needs only to calculate the value of δ_j for each hidden and output unit in the network, and then apply Eq. 2.29 . The error for the output units is given by

$$\delta_k = y_k - t_k \tag{2.30}$$

To evaluate the δ's for hidden units, we apply the chain rule for partial derivatives

$$\delta_j \equiv \frac{\partial E_n}{\partial a_j} = \sum_k \frac{\partial E_n}{\partial a_k} \frac{\partial a_k}{\partial w a_j} \tag{2.31}$$

where the sum runs over all units k to which unit j sends connections. By substituting the definition of δ in Eq. 2.27 into Eq. 2.31 and using Eq. 2.24 and Eq. 2.25, the back propagation formula illustrated as

$$\delta_j = h'(a_j) \sum_k w_{kj} \delta_k \tag{2.32}$$

Equation2.32 shows that the value of δ for a specific hidden unit can be obtained by propagating the δ's backwards from units higher up in the network. In Eq. 2.32 the summation is taken over the first index on w_{kj} (related to the error information backward propagation); however, in the forward propagation Eq. 2.9 the summation is taken over the second index. Having the values of the δ's for the output units and recursively evaluating Eq. 2.32 the δ's for all of the hidden units in a feed-forward network can be calculated. the following steps summarize the back propagation process:

- Apply an input vector $\mathbf{x}_n$ to the input of the network then forward propagate through the network by Eq. 2.24 and Eq. 2.25 to evaluate the activations of all the hidden and output units.

- Compute the δ_k for all the output units using Eq. 2.30.
- Back propagate the δ's using Eq. 2.32 to find δ_j for the hidden units in the network.
- Use Eq. 2.29 to compute the required derivatives.

2.5 Support Vector Machines

Support vector machines (SVM) algorithm has become one of the top algorithms for solving problems in classification and regression. An important feature of support vector machines is that the choice of the model parameters corresponds to a convex optimization problem. The SVM is a decision machine and does not provide posterior probabilities.

Considering SVM for classification of a two-class problem using linear models of the following form:

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \tag{2.33}$$

where $\phi(x)$ denotes a fixed feature-space transformation , and b is the bias parameter.

Eq. 2.33 can be represented using a dual expression in terms of kernel functions, to avoid working explicitly in feature space (also known as kernel trick). Training with data set with $\mathbf{x}_1, ..., \mathbf{x}_N$, input vectors, and binary label vector of $\{t_1, ..., t_N\}$ where $t_n \in \{1, -1\}$, then new data points $\mathbf{x}$ are classified according to the sign of $y(\mathbf{x})$. In case of linearly separable data at least one choice of the parameters $\mathbf{w}$ and b exist such that Eq. 2.33 satisfies$y(\mathbf{x}_n) > 0$ for points having $t_n = +1$ and $y(\mathbf{x}_n) < 0$ for points having $t_n = -1$. That being the case $t_n y(\mathbf{x}_n) > 0$ for all training data points.

By implementing support vector machines the decision boundary between classes is selected to be the one for which the margin is maximized see Figure 2.8. Furthermore, the desirable solutions are in which all data points are correctly classified, so that $t_n y(\mathbf{x}_n) > 0$ for all n. Thus the distance of a point $\mathbf{x}_n$ to the decision surface is given by

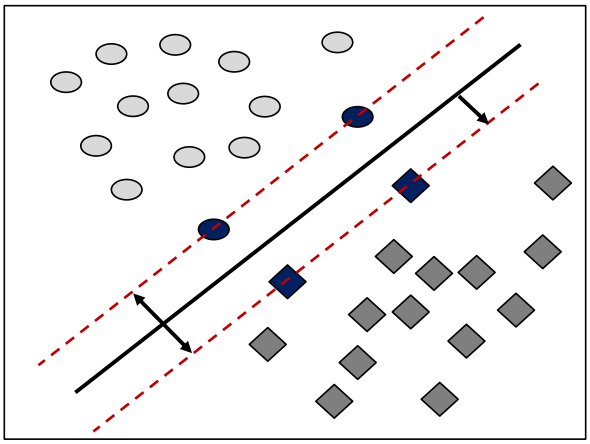


Figure 2.8: Support vectors based classification system, support vectors are the solid shapes and the decision boundary is the solid line.

$$\frac{t_n y\left(\mathbf{x}_n\right)}{\|\mathbf{w}\|}=\frac{t_n(\mathbf{w}^T \phi\left(\mathbf{x}_n\right)+b)}{\|\mathbf{w}\|} \tag{2.34}$$

The margin is the perpendicular distance to the closest point $\mathbf{x}_n$ from the data set, and the task is to optimize the parameters w and b to maximize this distance (margin). Therefore, the maximum margin solution is found by solving the following equation

$$arg\ max_{w,b}\{\ \frac{1}{\|\mathbf{w}\|}\{min_{\ n}\ \ \left[t_n(\mathbf{w}^T \phi\left(\mathbf{x}_n\right)+b)\right]\ \} \tag{2.35}$$

Because $\mathbf{w}$ does not depend on n the factor $1/\|\mathbf{w}\|$ has been taken outside the optimization over n . Handling this optimization problem directly is complex; however, converting it into an equivalent problem is easier to handle. Rescaling $\mathbf{w} \rightarrow k\mathbf{w}$ and $b \rightarrow kb$ does not change the distance from any point $\mathbf{x}_n$ to the decision surface and we have the following equation for the point that is closest to the surface.

$$t_n\left(\mathbf{w}^T \phi\left(\mathbf{x}_n\right)+b\right)=1 \tag{2.36}$$

In this scenario, all data points will satisfy the constraints given by

$$t_n \left(\mathbf{w}^T \phi\left(\mathbf{x}_n\right) + b\right) \geq 1, \quad n = 1, .., N \tag{2.37}$$

This gives the canonical representation of the decision hyper plane. For some data points the equality holds and they are called *active* constraints, however, the rest of the data points they are called *inactive*. All the time there will be at least one active constraint, as there will always be a closest point. There will be at least two active constraints once the margin has been maximized. This optimization requires maximizing $\|\mathbf{w}\|^{(-1)}$, which is equal to minimizing$\|\mathbf{w}\|^2$. Thus, we need to solve the following optimization problem subject to the constraints given by Eq. 2.37

$$\arg\min_{w,b} \frac{1}{2} \|\mathbf{w}\|^2 \tag{2.38}$$

This is a quadratic programming problem in which a quadratic function need to be minimized subject to a set of linear inequality constraints. The bias parameter b is determined implicitly via the constraints, due to changes to $\|\mathbf{w}\|$ be compensated by changes to b.By introducing Lagrange multipliers $a_n \geq 0$, with one multiplier a_n for each of the constraints in Eq. 2.37 this optimization can be handled

$$L\left(\mathbf{w}, b, \mathbf{a}\right) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n \left\{t_n \left(\mathbf{w}^T \phi\left(\mathbf{x}_n\right) + b\right) - 1\right\} \tag{2.39}$$

where $a = (a_1, ..., a_N)^T$. The minus sign in front of the Lagrange multiplier term, as the goal is minimizing with respect to $\mathbf{w}$ and b, and maximizing with respect to $\mathbf{a}$. By setting the derivatives of $L(\mathbf{w}, b, \mathbf{a})$ with respect to $\mathbf{w}$ and b equal to zero leads to the following two conditions

$$\mathbf{w} = \sum_{n=1}^{N} a_n \; t_n \phi\left(\mathbf{x}_n\right) \tag{2.40}$$

$$\mathbf{0} = \sum_{n=1}^{N} a_n \; t_n \tag{2.41}$$

Using these conditions we can eliminate w and b from $L(\mathbf{w}, b, \mathbf{a})$ to get dual representation of the maximum margin problem in following form, in which the task is maximizing with respect to **a** subject to the constraints in Eq. 2.43 and Eq. 2.44

$$L\left(\mathbf{a}\right) = \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m t_n a t_m \quad k(\mathbf{x}_n, \mathbf{x}_m) \tag{2.42}$$

$$a_n \geq 0, \quad n = 1, N, \tag{2.43}$$

$$\sum_{n=1}^{N} a_n \; t_n = 0 \tag{2.44}$$

kernel function in Eq. 2.42 is defined by $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}')$. Eq. 2.42 is a quadratic programming problem in which the goal is to optimize a quadratic function of **a** subject to a set of inequality constraints. Quadratic programming problems in M variables in common have solutions with computational complexity of $O(M^3)$. Making the use of dual formulation switches the original optimization problem of minimizing Eq. 2.38 over M variables , into the dual problemEq. 2.42, of N variables . For a fixed set of basis functions whose number M is smaller than the number N of data points (samples), the switch to the dual problem appears not attractive. Nonetheless, using it the model can be represented using kernels, thus the maximum margin classifier can be applied efficiently to feature spaces whose dimensionality exceeds the number of samples. Classifying new samples using the trained model, requires only evaluating the sign of $y(\mathbf{x})$ defined by Eq. 2.42. This can be expressed in terms of the parameters $\{a_n\}$ and the kernel function by substituting for $(\mathbf{w})$ using Eq. 2.40 to obtain

$$y\left(\mathbf{x}\right) = \sum_{n=1}^{N} a_n t_n k\left(\mathbf{x}, \mathbf{x}_m\right) + b \tag{2.45}$$

This constrained optimization satisfies the $Karush-Kuhn-Tucker(KKT)$ conditions, where the following three properties hold

$$a_n \geq 0 \tag{2.46}$$

$$t_n y\left(\mathbf{x}_n\right) - 1 \geq 0 \tag{2.47}$$

$$a_n\{t_n y\left(\mathbf{x}_n\right) - 1\} \geq 0 \tag{2.48}$$

Thus for each sample, either $a_n = 0$ or $t_n y(\mathbf{x}_n) = 1$. Any sample for which $a_n = 0$ will not contribute to the sum in Eq. 2.45 and it has no role about making predictions for new samples. Remaining samples are called *support vectors*, where they satisfy $t_n y(\mathbf{x}_n) = 1$. Therefore support vectors correspond to samples that lie on the maximum margin hyper planes in the feature space, as illustrated in Figure 2.8. After training the model, a significant proportion of the samples can be ignored and only the support vectors retained.

Once the quadratic programming problem has been solved and a value for $\mathbf{a}$ has been found, the value of the threshold parameter b can be determined as any support vector $\mathbf{x}_n$ satisfies $t_n\ y(\mathbf{x}_n) = 1$. Using Eq. 2.45 this leads

$$t_n \left(\sum_{m \in S} a_m t_m k\left(\mathbf{x}_n, \mathbf{x}_m\right) + b \right) = 1 \tag{2.49}$$

where S denotes the set of indices of the support vectors. Solving Eq. 2.49 for b numerically gives a stable solution by first multiplying by t_n, making use of $t_2^n = 1$, and then averaging these equations over all support vectors and solving for b to obtain

$$b = \frac{1}{N_S} \sum_{n \in S} \left(t_n - \sum_{m \in S} a_m t_m k\left(\mathbf{x}_n, \mathbf{x}_m\right) \right) \tag{2.50}$$

where N_S is the total number of support vectors.

This ends the summary about NN and SVM adopted from [46].

2.6 Decision Trees and Random Forests

A decision tree is a hierarchical model for supervised learning in which the local region is identified in a sequence of recursive splits in a smaller number of steps. A decision tree is composed of internal decision nodes and terminal leaves. Each decision node implements a test function $f_m(\mathbf{x})$ with discrete outcomes labeling the branches. At each node, given an input, a test is applied and one of the branches is taken depending on the outcome. This process starts at the root and is repeated recursively until a leaf node is hit, at which point the value written in the leaf constitutes the output [47]. A decision tree is a nonparametric model, in fact we do not assume any parameters about class densities and the tree structure is not fixed but only the tree grows, branches and leaves are added. In a decision tree, a complex function is broken down into a series of simple decisions. Different decision tree methods assume different models for the test function $f_m(.)$, and the model class defines the shape of the discriminant regions. Each leaf node has an output, which in the case of classification is the class label and in regression is a numeric value. A leaf node defines a localized region in the input space where instances falling in this region have the same labels (in classification), or very similar numeric outputs (in regression). Decision trees have the advantage of interpret-ability. in fact , they can be converted to a set of IF-THEN rules that are easily understandable [47]

Random Forest

Random forests are a combination of tree predictors where each tree depends on the values of a random vector sampled independently using the same distribution for all trees in the forest. The generalization error converges to a limit for a large number of trees in the forest . Moreover, for forest of tree classifiers the generalization error depends on the strength of the individual trees in the forest and the correlation between them. Splitting features using a random selection of features at each node yields error rates that are more robust with respect to noise [48].

A random forest classifier consists of a collection of tree-structured classifiers

$h(\mathbf{x}, \mathbf{\Theta}_k), k = 1, ...$ where the $\mathbf{\Theta}_k$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input $\mathbf{x}$. Significant improvements in classification accuracy have been achieved from growing an ensemble of trees and letting them vote for the most popular class. In order to grow these ensembles, often random vectors are generated that govern the growth of each tree in the ensemble. To grow each tree of the ensemble a random selection is made from the samples in the training set. In random split selection $\mathbf{\Theta}$ consists of a number of independent random integers between 1 and K best splits. The type and dimensionality of $\mathbf{\Theta}$ depends on its use in tree construction. finally,after a large number of trees is generated, they vote for the most popular class [48].

2.7 Discussion

This Section provides an introduction to some key ideas in machine learning classification that can help with model selection, tuning, and evaluation. During the training of classifiers one needs to be careful to avoid overfitting the training data. That is, reducing training error by fitting the noise and minor fluctuations in the training data. Overfitted classifiers cannot generalize well during test stage and consequently have high test error rates [41]. For classifier evaluation the simplest form of the performance measure is error rate, that is, the percentage of number of examples the classifier miss classified to the total number of examples. The error rate and accuracy rate are complementaries and used interchangeably. The error rate is useful if the class distribution in the data is uniform. Otherwise we need to go for more advanced measures like the f-measure or confusion matrix which provides class by class details. As in the recent years more computational resources become more available classifier combination becomes affordable. Thus, instead of having one classifier to address the problem a group of base classifiers (usually weak learners) are used to make the final decision. Different methods have been developed to structure how the base classifiers are combined like boosting, adaptive boosting (Adaboost), bootstrap aggregating (Bagging) , etc.[12].

Due to their high accuracy rates NN, RF, and SVM are considered strong classifiers. Neural networks training is time consuming in order to tune the network parameters to avoid over-fitting the training data. However the testing phase of NN classifiers is faster than the training stage, because once the optimized weights are obtained testing becomes simply matrix multiplication (a feature can be implemented in hand-held testing devices). Random selection of features at each node in RF classification improves error rates, thus they are less vulnerable to noise. Moreover, SVM classifiers have the advantages of convex optimization problem and margin maximization. For training models purposes, one needs to have datasets with equal class distribution as possible. In case of skewed dataset advanced performance metrics are needed such as the confusion matrix and F-score. Moreover, using combined machine learning models is more reliable than depending on one model as the combined model consolidates the strong features of each individual classifier.

Chapter 3

Intelligent Microwave Detection of Surface Anomalies

3.1 Introduction

Microwave sensor performance at the hardware level has been implemented using metamaterials to make strong localization of electrical fields around the sensing element, leading to improved sensitivity and resolution of microwave near-field sensors [49, 38, 5]. One-dimensional or two-dimensional metamaterial particle arrangements enhance the sensors sensitivity to detect small anomalies [49, 38]. Thus, increasing use of near-field microwave sensors necessitates improving their performance, not only at the sensor design level but also at the signal post processing level. Due to the lack of advanced post-processing, small signal changes and variations captured by microwave sensors available nowadays are subject to numerous errors when measured with the human eye. In effect, the collected information (raw datasets) requires advanced signal processing to make accurate decisions. In some occasions, decisions are supercritical, for instance, decisions about the structural health of aircraft fuselages or structures of nuclear reactors, as discussed in Chapter 2, ML theories offer a natural framework for addressing damage detection. Hence, the next case studies present an implementation of artificial intelligent models for detecting sub-millimeter

cracks in metallic surfaces using a waveguide sensor loaded with split-ring resonators. Split-ring resonators, which are constituent elements used in a variety of metamaterial designs, were recently found to enhance the sensitivity of waveguide sensors [5]. The conventional method for sub-millimeter crack detection in metallic surfaces using waveguide sensors consists of observing any frequency shifts in the magnitude of the reflection coefficient ($|S_{11}|$) pattern. Additionally, some research has been carried out in which the frequency shift was changed to voltage levels[14]; these voltage levels were then used to detect cracks. Using the conventional approach minor shifts are not easily observable or detectable. The importance of employing Al models comes from their ability to detect small variations and to automate the surface scan and crack detection process. Furthermore, AI preprocessing methods such as feature normalization can significantly improve crack detection in terms of minimizing the effect of using different stand-off distances [40]. In the experimental studies presented here, supervised ML was used by labeling training datasets acquired from scans of cracked and non-cracked surfaces. Then ML classifiers were trained for pattern recognition to classify new surface scans obtained using the sensor, to validate the concept two case studies are presented. The methodology used in case study 1[1] is based on applying AI classification after implementing feature extraction, whereas the implemented methodology in case study 2[2] is based on applying AI classification after utilizing feature selection.

[1]The material presented in this case study is published in [11].

[2]The material presented in this case study is published in [45].

3.2 Detecting Cracks in Metallic Surfaces Using a Waveguide Sensor: Case Study 1

Sensor Model and Experimental Setup

The sensor used in this work for scanning metallic surfaces was an open-ended waveguide probe enhanced with an array of split-ring resonator (SRR) cells [5]. The waveguide operates at the Ku-band of 12-18 GHz and has a cross section of 15.8 mm by 7.9 mm, with a standard flange with dimensions of 33.30 mm by 33.30 mm. The experimental setup operates by scanning a metallic plate containing 0.5 mm surface cracks ranging in depth from 0.5 mm to 2.25 mm, with increments of 0.25 mm. Figure 3.1 below shows a diagram of a waveguide sensor scanning a metallic plate with cracks. The sensor is placed at a 0.5 mm stand-off distance, with the long dimension of the waveguide being parallel to the cracks.

Integrating a metamaterial layer between the waveguide and the surface under test enhances the sensor's sensitivity, as presented in [49]. Various shapes for metamaterial cells are available in the literature, for instance split rings [50], omega shaped [51], AV-shaped [52] and U-shaped [53] cells. However, for sub-millimeter crack detection a split ring represents an excellent choice due to the high electrical field confinement in the gap. When,utilizing the gap as the sensing element, a major frequency shift was observed if the field in the gap was perturbed by the existence of a crack [5].

A printed circuit board (PCB) with low loss (Rogers 4003) was used to fabricate the SRRs. Figure 3.2 (a) displays the front and back views of the PCB patch used at the open end of the waveguide. Figure 3.2 (b) shows a photograph of the waveguide sensor.

In a simplified microwave sensor configuration similar to the one used in this case, the sensor has only one port, which is connected to a vector network analyzer (VNA) as shown in Figure 3.3, where the sensor is scanning along the y-axis.

The VNA sends signals to the sensor at different frequencies in a sweep manner and collects the reflected signals (data) from the sensor. The information reflected back is

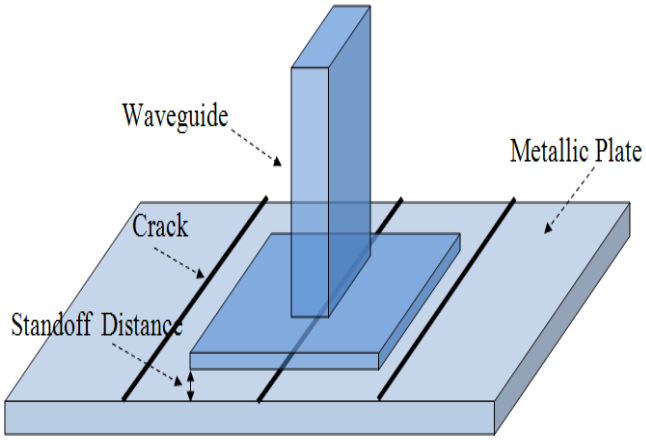


Figure 3.1: Schematic drawing of a waveguide sensor scanning a metallic plate with surface cracks.

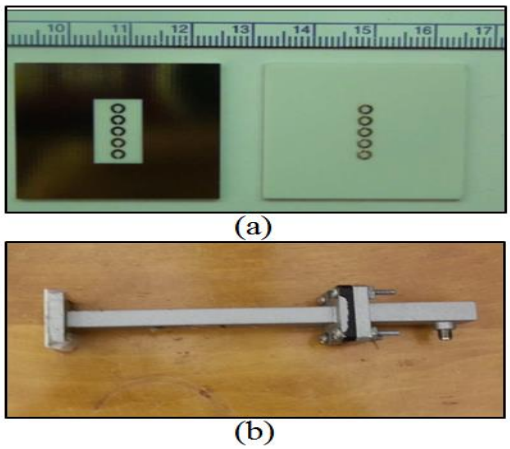


Figure 3.2: (**a**) Photographs of the front and back views of the split-ring resonator (SRR) array etched on a printed circuit board; (**b**) Photograph of the sensor.

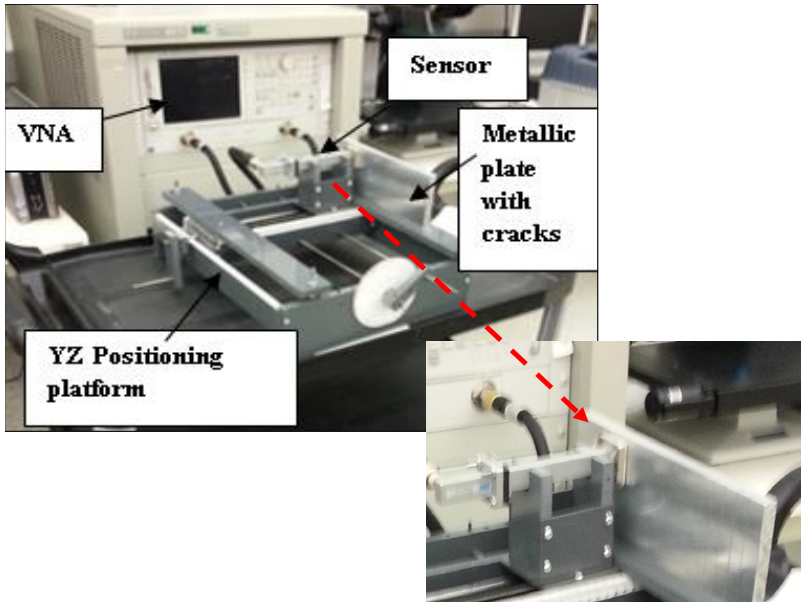


Figure 3.3: Photograph of the experimental configuration. Scanning along y-axis at 0.5 mm stand-off.

valuable since it reveals details about the sensor's environment. For instance, the reflected data from a non-cracked surface is different from that reflected from a cracked surface. Figure 3.4 illustrates the reflection coefficient S_{11} magnitude over the operating frequency range of the probe from cracked and non-cracked surfaces.

Based on the presence of a crack underneath the sensor, the magnitude of the reflection coefficient ($|S_{11}|$) experiences a frequency shift and level change. In other words, the S_{11} pattern changes according to the presence or absence of cracks. In this experiment, for each scan, the VNA was swept over a frequency range compatible with that of the sensor (12 to 18 GHz), with increments of 30 MHz. Then the data was collected and saved in an array format to be used later by the AI algorithms.

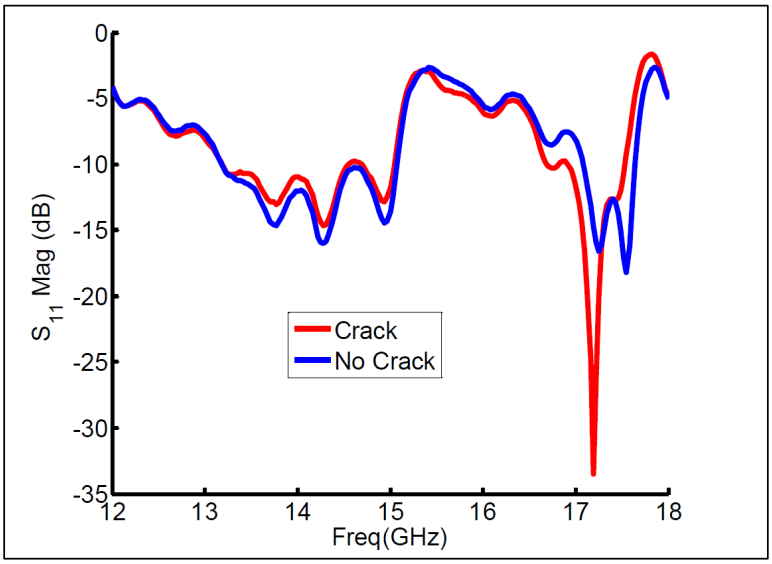


Figure 3.4: Reflection coefficient magnitude from cracked and un-cracked surfaces at 0.5 mm stand-off.

Data Collection and PCA Processing

The data were collected on three different days. On each day, the experiment configuration was rebuilt to ensure full reliability in the repeatability of experimental findings. Then the data were mixed randomly into one dataset, with 415 total samples (scans) and 201 features (frequency points from 12-18 GHz with 30 MHz increments). Additionally, the distribution of the samples was intended to be symmetric (207 samples with a crack and 208 with no crack). In view of the fact that the first class (cracked scans) and the second class (non-cracked scans) have approximately the same number of samples, the accuracy of the AI model is expected to give a realistic and accurate evaluation of the detection mechanism's performance.

Principal Component Analysis (PCA) was used for data pre-processing. PCA is one of the most widely used techniques for feature extraction and dimensionality reduction to build lower dimension datasets from higher ones. PCA finds a set of the most representative

orthogonal projection vectors, where the projected samples retain the most information about original samples. PCA also has the ability to remove correlations among variables and enhance the signal to noise ratio [54],[40]. Another effective feature extraction method is Independent Component Analysis (ICA), which maps input data onto basis vectors, that are as statistically independent as possible [54]. An important difference between PCA and ICA is related to the number of components used in each methodology. In PCA, this number can be determined by the variance criteria, but in ICA, there are no criteria for determining how many components represent the dynamic range of the data [55]. Linear Discrimination Analysis (LDA) is another widely used feature extraction method for sensor development [40]. Unlike PCA, LDA requires class information (labels) to find a set of vectors that maximizes between-class scatter while minimizing within-class scatter [54], [40]. Thus, PCA was implemented in this current study to build datasets of reduced dimensionality by extracting important features according to their variance contribution. PCA was applied after scaling the original dataset to zero mean and unity standard deviation (Z- scaling). Figure 3.5 below shows the PCA results of the first 13 principal components. The PCA analysis shown in Figure 3.5, makes it clear that the first two principal components contribute more than 75% of the variance. Moreover, the first eight and first thirteen principal components achieve more that 90% and 95% ,respectively, of the whole original dataset variance. Data visualization gives a good perspective about the data distribution and can help with selecting the learning algorithm. However, humans can handle only up to three dimensions. Therefore, one of the datasets built using PCA was chosen to be a two-dimensional (2-d).

To visualize the data distribution in two-dimensional space, the first principal component was plotted against the second component (Figure 3.6).The green triangles represent samples/scans for non-cracked surfaces, and the red circles correspond to samples/scans with cracks. One important observation about Figure 3.6 is that the green triangles are concentrated in one region of the plot, whereas the red circles are widely spread. These concentrations differ because scans for surfaces with no cracks have the same pattern,but scans for cracked surfaces have different patterns as a result of varying crack depths. From the PCA results, three datasets were built in total: Dataset 1: a matrix of 415 samples and

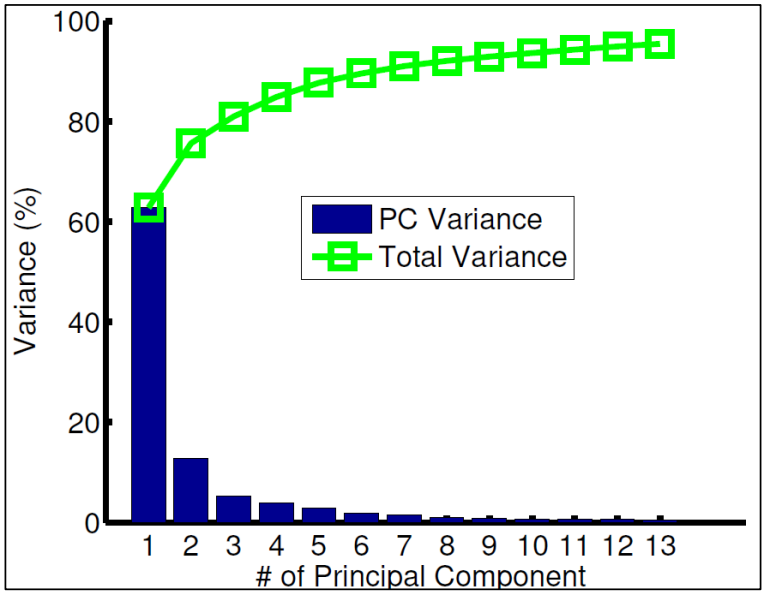


Figure 3.5: Variance cumulative sum for the first 13 principal components [11].

2 features (the first two principal components); Dataset 2: a matrix of 415 samples and 8 features (the first eight principal components); and Dataset 3: a matrix of 415 samples and 13 features (the first thirteen principal components that together attained more than 95% of the collected data variance).

Implemented Models

Each feature of the raw data collected by the sensor was normalized to its maximum value. The normalized dataset was then passed to the feature-extraction stage where PCA was applied to generate the three reduced datasets. After feature extraction, the resultant datasets were delivered to the classification stage, where two ANN and one SVM classifiers were implemented to build a combined AI model. Figure 3.7 provides a flow chart of the data preprocessing and the Al model architecture. An odd number of classifiers was chosen to avoid ties during the combination phase. These classifiers were combined using majority voting process that made the final decision (Figure 3.7).

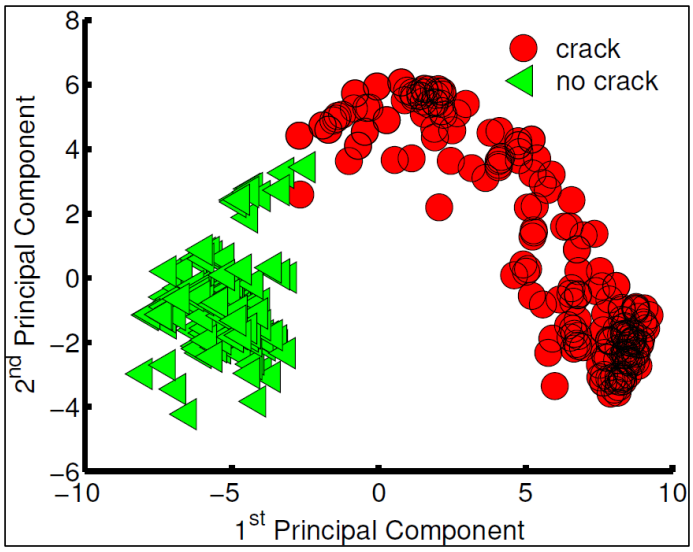


Figure 3.6: Plot of the first two principal components, which contributed more than 75% of the variance.

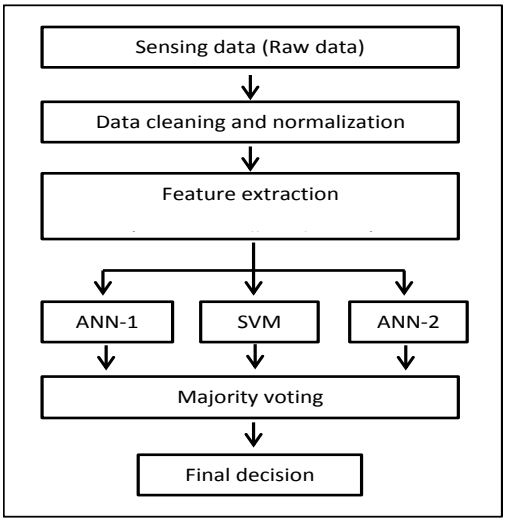


Figure 3.7: Implemented model architecture. Majority voting combination is used for final decision.

Table 3.1: Size of training ,validation and testing subsets.

Dataset	**Training** (# of Samples)	**Validation** (# of Samples)	**Testing** (# of Samples)
2-d	38	8	369
8-d	37	9	369
13-d	36	10	369

It is important to mention that each dataset had its own AI model because the model parameters were a function of the dataset dimensions. For example, the 2-d dataset needed 2 input units and one bias unit for the ANN input layer; whereas the 8 dimensional (8-d) dataset needed 8 input units and one bias unit.

To maintain a low testing error (out of-sample-error), each dataset was divided into training, validation, and testing subsets; the model parameters were then tuned based on minimizing the validation subset error rate. The sizes of training and validation subsets for each dataset were chosen according to [56]. Table 3.1 illustrates the size (number of samples) of the training, validation, and testing subsets for each dataset.

Artificial Neural Network Classifiers

The NN classifiers implemented in this work were based on a three-layer fully connected configuration, as shown in Figure 3.8, with a back-propagation algorithm and a log sigmoid function. In fact, different types of activation function can be used in NNs, such as linear, step, or Gaussian. In this work, a log sigmoid activation function has been selected as it has a convenient derivative (easy to compute from the sigmoid function itself), which is suitable when training networks using back-propagation algorithms[57]. The log sigmoid transfer always limits its output to the range between 0 and 1, as given by

$$g(z) = \frac{1}{1 + e^{-z}} \tag{3.1}$$

The derivative of the log sigmoid function is given by

$$g'(z) = \frac{1}{1+e^{-z}} * (1 - \frac{1}{1+e^{-z}}) \tag{3.2}$$

In the literature, the back-propagation (BP) algorithm is well studied for ANN training, and it has been used in various applications [58],[59]. However, it is subject to local convergence and slowness [60]. On the other hand, particle swarm optimization (PSO) has been gaining interest in recent years and shows good results as in [60]. The PSO algorithm shows faster converge during the initial stages of a global search; nevertheless around the global optimum, the search process is very slow. In contrast, the gradient descend method used in BP tends to achieve faster convergent speed around the global optimum [61]. More recent hybrid algorithms have been reported [61], that combine PSO and BP algorithms to unify the strong global searching ability of the former (PSO) and the strong local searching ability of the latter (BP).

In the current work, to help the back propagation algorithm avoid local minima, each NN classifier was started with random weights for each run. Furthermore, the number of hidden units was optimized based on the validation error (not on the training error) to avoid over-fitting as much as possible. In addition, a weight decay (regularization parameter) term was included in the algorithm as an additional measure to prevent or minimize over-fitting. The output layer has two units. The first unit gives the probability of a scan belonging to the first class (cracks), and the second unit gives the probability of a scan belonging to the second class (no cracks). At the end, the algorithm assigns the scan to the class with the higher probability.

Support Vector Machine Classifier

The support vector machine (SVM) algorithm has become one of the most-effective algorithms for solving problems in classification and regression. An important feature of SVM is that the determination of the model parameters corresponds to a convex optimization problem, thus any local solution is also a global optimum [46]. The SVM is a decision machine and therefore does not provide posterior probabilities.

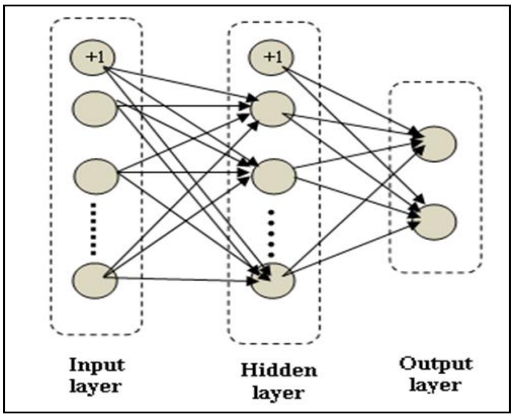


Figure 3.8: Three-layer neural network with fully connected configuration.

Considering SVM for classification of a two-class problem using linear models of the form:

$$y(x) = W^T \phi(x) + b \tag{3.3}$$

where $\phi(x)$ denotes a fixed feature-space transformation, and b is the bias parameter, Eq. 3.2 can be represented using a dual expression in terms of kernel functions to avoid working explicitly in the feature space (an approach as the kernel trick) [46]. By implementing support vector machines, the decision boundary between classes has a unique feature, as it is chosen to be the one for which the margin is maximized [12, 46]. The implemented SVM classifier in this work was based on a Matlab interface from Libsvm [62] The SVM classifier achieved high classification accuracy, mainly by optimizing the kernel type,degree,and termination criterion tolerance. To have a baseline for the performance of the proposed combined models, a Naive Bayes (NB) classifier was employed. NB and SVM and their variants are often used as baselines in classification tasks, such as in text classification [63]. NB and SVM performance varies significantly depending on the model variant, features, and dataset used [63],[64]. Since SVM was used in the combined models, the NB was considered as the baseline. The average accuracy rate was used as a performance measure by averaging over 1500 runs.

Table 3.2: First model using Dataset-1 (2-d) performance profile along with a comparison to Naive Bayes model.

Model	**Dataset Features**	**Average Accuracy (Standard Deviation)**		
		Training	**Validation**	**Testing**
Combined Model	1st and 2nd (PCA)	99.93% (0.5)	99.30% (2.9)	99.31% (0.5)
NB Model	1st and 2nd (PCA)	99.92% (0.47)	98.18% (4.99)	98.25% (1.64)

Table 3.3: Second model using Dataset-2 (8-d) performance profile along with a comparison to Naive Bayes model.

Model	**Dataset Features**	**Average Accuracy (Standard Deviation)**		
		Training	**Validation**	**Testing**
Combined Model	First Eight (PCA)	100% (0)	99.50% (2.2)	99.60% (0.65)
NB Model	First Eight (PCA)	99.98% (0.23)	95.77% (7.28)	95.84% (3.2)

Results and Discussion

The first model was implemented using Dataset-1 (2-d). Each classifier was optimized individually to maximize the validation accuracy. NN model optimization was performed by tuning the number of units in the hidden layer and the regularization parameter. Furthermore, to prevent the models from over-fitting, the number of hidden units was kept as small as possible. During SVM optimization, the model showed that linear kernels achieved higher validation accuracy rates than non-linear kernels. The models implemented using Dataset-2 (8-d) and Dataset-3 (13-d) were similarly optimized to maximize validation accuracy. Tables 3.2, 3.3, and 3.4 give the average training, validation, and testing accuracy rates of each combined model. The standard deviation was also reported to indicate the accuracy rate distribution.

For Dataset-1 (2-d dataset), the training accuracy rates of the first proposed combined model and the NB were similar, and as high as 99%. However, the combined model slightly

Table 3.4: Third model using Dataset-3 (13-d) performance profile along with a comparison to Naive Bayes model.

Model	**Dataset Features**	**Average Accuracy (Standard Deviation)**		
		Training	**Validation**	**Testing**
Combined Model	First Thirteen (PCA)	100% (0)	99.65% (1.9)	99.62% (0.65)
NB Model	First Thirteen (PCA)	99.98% (2.1)	94.3% (8.21)	94.48% (3.78)

out performed its NB model counterpart, by 1%, during the validation and testing. In the case of Dataset-2 (8-d dataset), the gap between the training accuracy rates of the second proposed model and the NB model was very small, and a rate of 100% was achievable by the proposed model. However,gaps between the validation and test rates of the second proposed combined model and their NB model equivalent rates were 3.8% and 3.7%, respectively (Table 3.3).

Results of the Dataset-3 (13-d dataset), shown in Table 3.4, revealed that the training accuracy of the third combined model and the NB model were 100% and 99.9%, respectively. Nevertheless,the proposed model achieved higher validation and test rates than its NB equivalent, and the out- performance gaps were increased compared to Dataset-2. More specifically,the difference between the third model validation and test accuracy rates and their NB model equivalents was larger than 5% . Overall,testing of all proposed models revealed a better than 99% accuracy rate for crack classification. A visual explanation of the learned 2-d model is given by Figure 3.9, where the decision boundary (blue line) is plotted. This boundary separates scans with cracks (red circles) from scans with no cracks (green triangles). The learned model was relatively linear in general. The figure has two misclassified samples.However, avoiding perfect classification (over-fitting) during training is recommended for better out-of-sample generalization.

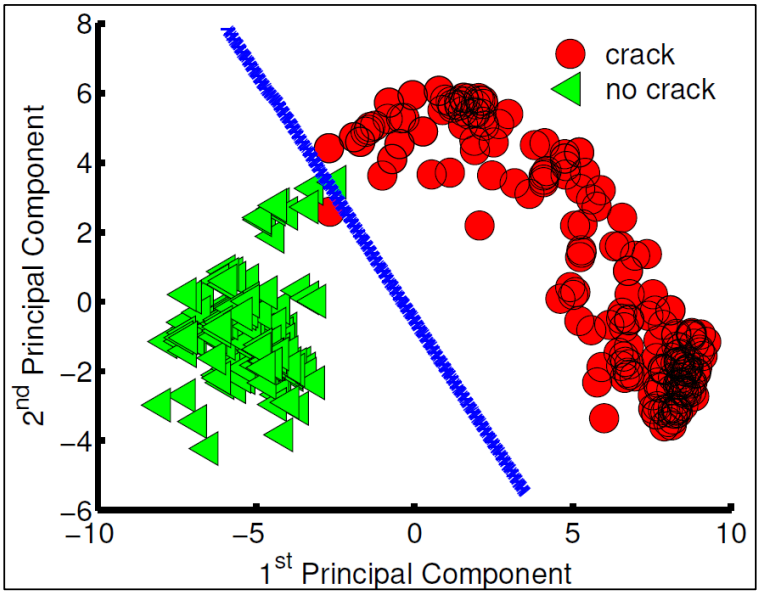


Figure 3.9: The combined model decision boundary (2-d Principal Component Analysis (PCA) dataset).

Conclusions

The case study presented in this section has demonstrated an implementation of combined AI models on different datasets obtained from a microwave waveguide sensor for sub-millimeter crack detection on metallic surfaces. The sensitivity of the waveguide sensor was enhanced with metamaterial particles. First PCA was applied as a feature-extraction technique to obtain a general view of the data in a 2-d space. Additionally, PCA was used to build additional reduced datasets of 8-d and 13-d with more than 90% of the variance of the original dataset. Then, for each dataset, a combined AI model was developed, composed of two neural network classifiers and one support vector machine classifier. The results using the three models clearly validate the learning feasibility. A testing accuracy rate of more than 99% was achieved. In addition, a base line comparison to a Naive Bayes classifier was implemented as a performance measure.

3.3 Reducing Sweeping Frequencies in Microwave Detection: Case Study 2

Introduction

Different microwave NDT sensors proposed in earlier works operate over large frequency ranges with fine sweeping increments. For instance, metal defect sizing and detection under thick coating using microwaves from 8.2 – 12.4 GHz was reported in [15]. In [5], a waveguide probe was used for crack detection in metallic surfaces with an operating frequency range of 12 –18 GHz. In [65] detection of cracks in non-metallic materials using a microwave resonator was implemented by sweeping over a frequency range larger than 1 GHz. More recently, non-invasive measurements of complex permittivity based on sweeping the operating frequency of a microwave sensor from 1.7 – 2.7 GHz was reported in [66]. This case study addresses the need to reduce the number of sweeping frequencies during microwave NDT to build efficient and cost effective sensors. It does so by implementing ML feature selection methods. This approach, in turn, leads to a reduction in the detection circuit complexity and size of acquired data, thereby, reducing the cost of the testing equipment, especially for hand-held devices. Employing feature selection methods to discover the most influential features among the full feature data set leads to minimizing the number of testing frequency points needed to perform NTD diagnostics. After selecting a subset of the important features, future structural health tests can be conducted using the selected features incorporated with supervised learning techniques for defect and damage detection using classification models. As a validation study for the concept of reducing sweeping frequencies using feature selection methods, the microwave waveguide sensor used in Case Study 1 was used here, and results were compared. The machine leaning implementation in case study 1 was based on feature extraction using principal component analysis (PCA). Unlike feature selection, feature extraction methods such as PCA extract the important features by mapping the original data set to a lower space. Features in the lower space do not correspond to the sweeping frequencies.

Feature Selection Techniques

Feature selection is the process of selecting a subset of relevant features for building learning models. When irrelevant features are eliminated from the original data set, the prediction accuracy of the models can be improved [67]. Moreover, some features can be redundant, and have high correlations. Thus, selecting un-noisy and uncorrelated features can increase classifier performance. Feature selection techniques can be classified into two main categories: filters and wrapper methods [68, 69].

Filter methods select a subset of features as a pre-processing step, independently of the learning algorithms of the classifiers. Filter methods are based only on general characteristics of the data such as the correlation between variable to be predicted. Filter methods suppress the least interesting variables and they work independent from data labels. These methods are particularly efficient in computation time and robust to overfitting.

Wrapper methods utilize the classifiers' performance to select feature subsets. They evaluate subsets of variables, thus allowing the possible interactions between variables to be detected. Wrapper models give better results compared to filter models, however, they are computationally expensive and classifier specific.

Filter-based feature selection techniques are adopted in this study based on their independence from the learning process. Four different filter-based feature selection and ranking techniques are investigated in order to identify the most important features, namely information gain, gain ration, and Relief. The following subsections present an overview of these techniques.

Information Gain (IG)

Information gain is the expected reduction in entropy obtained by partitioning the features according to a given aspect. The entropy characterizes the uncertainty associated with a random collection of features; measuring the impurity or disorder of the data set. Information gain measures the amount of information in bits about the class prediction based on a feature and the corresponding class distribution [70, 71]. The entropy of the data set as a

whole and for each class is needed To select the important features. For a given set of data D containing c different values (features), the entropy can be calculated as follows[70, 71]:

$$Entropy(D) = -\sum_{i=1}^{c} P(c_i) \times \log_2 \left(P(c_i)\right) \tag{3.4}$$

Where $P(c_i)$ is the probability of getting the i^{th} feature randomly selected from the data set D. If we make attribute A_i, with the v features, the root of the tree, this will partition dataset D into v subsets $D1, D2, ..., Dv$. The expected entropy if A_i is used:

$$Entropy_{Ai}(D) = -\sum_{i=1}^{v} \frac{\mid D_i \mid}{\mid D \mid} \times Entropy(D_i) \tag{3.5}$$

Information gained by selecting attribute A to branch or to partition the data set is:

$$Gain(D, A_i) = Entropy(D) - Entropy_{Ai}(D) \tag{3.6}$$

Equation 3.6 is used to select those features with the highest gain.

Gain Ratio (GR)

Gain ratio is a modified information gain method that prevents bias in information gain. Eq. 3.7 represents the information generated by splitting the training data set D into v partitions corresponding to v outcomes of a test on the feature A. Then the gain ratio is a normalized information gain as in Eq. 3.8. GR takes number and size of branches into account when choosing an feature. The feature with the highest gain ratio is selected as the splitting feature. [70, 72].

$$SplitInfo_A(D) = -\sum_{i=1}^{v} \frac{\mid D_i \mid}{\mid D \mid} \times \log_2 \frac{\mid D_i \mid}{\mid D \mid} \tag{3.7}$$

$$GainRatio(D, A_i) = Gain(D, A_i) - SplitInfo_A(D) \tag{3.8}$$

Relief

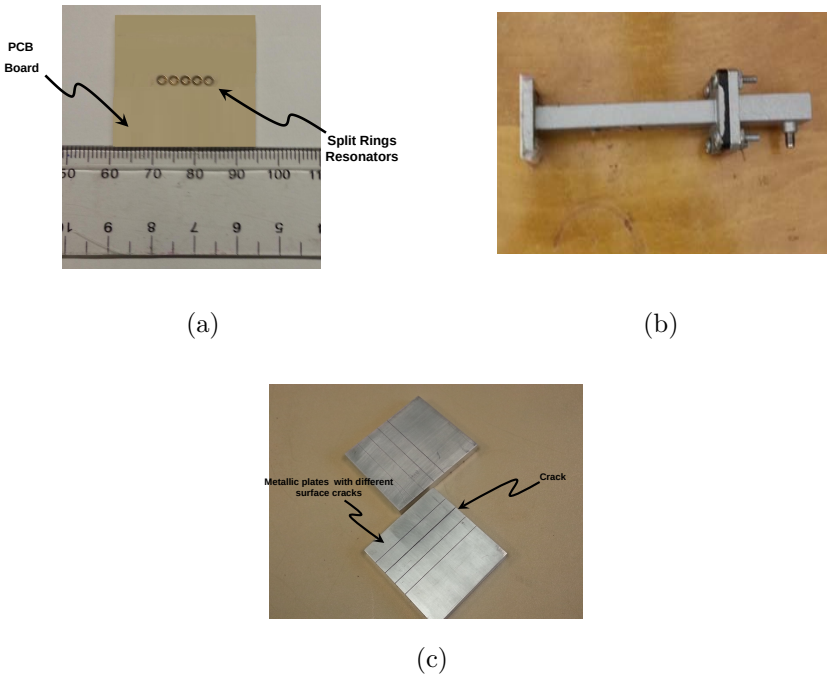
Relief is an instance-based algorithm that ranks on features by finding a relevance weight for each feature [73, 74]. The weight for a each feature indicates its ability to distinguish between classes. For each chosen example (observation), the nearest instance of the same class (nearest hit) and opposite class (nearest miss) are found. The feature's weight is then updated according to how well its values distinguish the chosen instance from its nearest hit and nearest miss as in Eq. 3.9. A feature receives a high weight if it differentiates between observations from different classes and has the same value for observations of the same class. Relief randomly samples observations from the training data [74].

$$W_X^{new} = W_X^{old} - \frac{diff(X, R, H)^2}{m} + \frac{diff(X, R, M)^2}{m} \tag{3.9}$$

Where W_X is the weight for feature X, R is a randomly sampled observation, H is the nearest hit, M is the nearest miss, and m is the number of randomly sampled observations. The function $diff$ calculates the difference between two observations for a given feature. For nominal attributes it is defined as either 1 (the value of the feature differs between the two observations) or 0 (the feature has the same value in both observations), while for continuous attributes the difference is the actual difference normalized to the interval [0; 1]. Dividing by m guarantees that all weights are in the interval [-1, 1] [74].

Experimental Setup

A measured data set is comprised of real measurements of metallic plate scans, obtained using a waveguide sensor loaded with metamaterial particles [5]. The metamaterial particles are arranged in an array as shown in Figure 3.10 (a). The waveguide was operated at the Ku-band, and has a cross section of 15.8 mm by 7.9 mm, with a standard flange with dimensions of 33.30 mm by 33.30 mm. Figures 3.10 (b) and (c) show the sensor and the metallic plates with different cracks, respectively. The surface testing was performed by scanning a metallic plate containing multiple 0.5 mm surface cracks ranging in depth from 0.5 mm to 2.25 mm (cracks are made using a milling machine). The sensor was



(a) (b)

(c)

Figure 3.10: Photographs of the sensor. (a) Used waveguide sensor side view. (b) PCB board with split ring resonators. (c) A photograph of metallic plates with different cracks.

connected to a vector network analyzer (VNA) in a one-port configuration, as illustrated in Figure 3.11. The sensing mechanism consists of the VNA sending signals to the sensor at different frequencies in a sweep manner and collecting the reflected signals (data) from the sensor while the sensor scans the surface of the metallic plate under test at 0.5 mm stand-off distance. The VNA was swept over a frequency range of (12 to 18 GHz) with increments of 30 MHz to measure the reflection at 201 frequency points (features) for each scanned position.

Data Set Descriptions

Signals reflected back from the sensor contain information about the health of the scanned metallic plate. For example, the pattern of the signal reflected from a non-cracked (healthy)

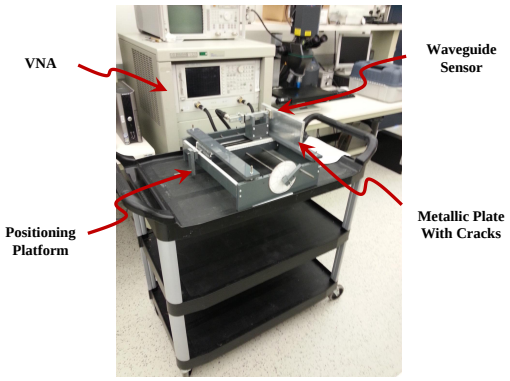


Figure 3.11: Testing configuration.

surface is different from one reflected from a cracked (unhealthy) surface. Figure 3.12 illustrates the reflection coefficient magnitude over the operating frequency range for cracked and non-cracked surfaces.

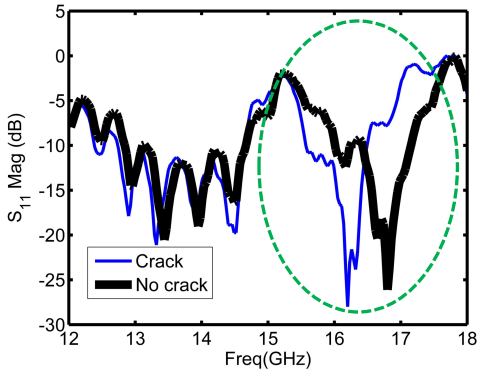


Figure 3.12: Reflection coefficient magnitude plots from healthy and unhealthy metallic surfaces.

As Figure 3.12, depicts the sensor experienced a shift when it encountered a crack, as highlighted by the circle in Figure 3.12. The highlighted part of Figure 3.12 is the region around the resonance frequency of the sensor and it is the main informative part in the plot about the structural health. Figure 3.13 shows a closer illustration of the reflection

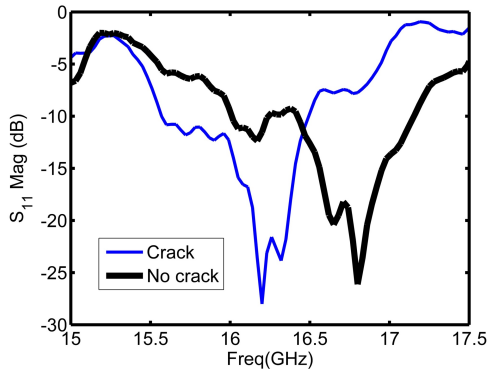


Figure 3.13: Reflection coefficient magnitude plot around the resonance frequency of the sensors.

coefficient magnitude around the resonance frequency of the sensor. The range from 15.6 to 17.2 GHz is the range of interest for the coming implemented feature selection algorithms. In total, there were 53 frequency points (features) in this range starting from the 122^{nd} feature to the 174^{th} feature in the initial data set. The objective of the feature selection algorithms implemented in coming sections is reducing the number of sweeping frequency points by selecting the main informative frequency points among these 53. Different classifiers were trained and tested for crack detection based on reduced data sets. The class distribution of the measured observations was symmetric (180 observations with a crack and 180 with no crack) to avoid an unbalanced class scenario.

Pre-Processing

Input normalization data is very important when dealing with parameters of different units and scales. Therefore, all parameters should have the same scale for a fair comparison between them. In the absence of normalization, features with large values have a greater influence on the cost function, which must be considered when the classifier is designed. The data set understudy was normalized using min-max normalization which restricts the

values of all features within predetermined ranges. The general formula is given as:

$$x' = \frac{x - min(x)}{max(x) - min(x)} * (b - a) + a \tag{3.10}$$

where x is the original value, x' is the normalized value, a and b are the min and max values of the new scale.

Results and Discussion

This section presents the outputs of the implemented feature selection techniques as well as their affect on the performance of the built classifiers.

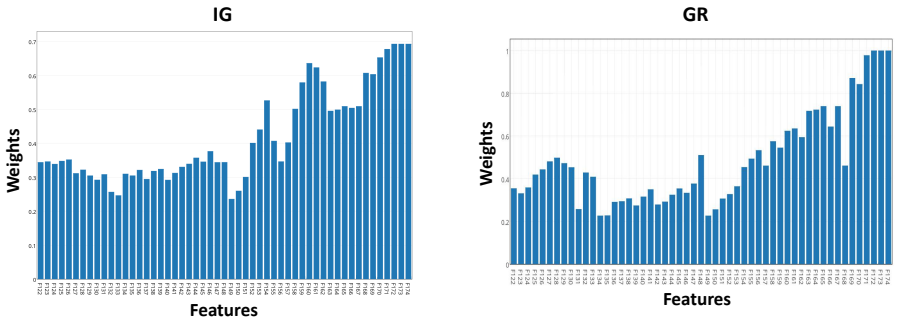
Feature Selection Results

In this study, information gain, gain ratio, and Relief algorithms were implemented using the R-project [75] to select the top five important features among the set of 53 features in the vicinity of the sensor's resonance. The feature weights after these algorithms were applied are shown in Figures 3.14 (a) to 3.14 (c). In these plots the higher the weight is, the higher the feature's importance. As illustrated, weights using IG and GR algorithms have a common trend. However, Relief algorithm selection was different compared to the rest of the algorithms.

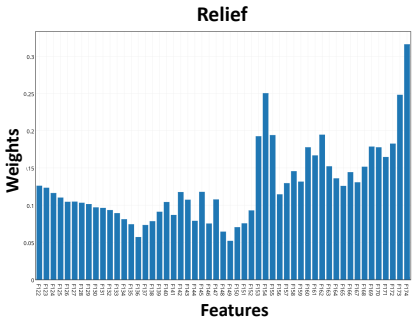
Table 3.5 summarizes the outcomes of the feature selection algorithms. GR selection is similar to IG with a difference only in the fifth feature. However, the ordering of feature when using Relief method was obviously different from that of the other methods, as depicted in Table 3.5.

Classification Implementation and Results

Deciding which classification algorithm to select for evaluating and classifying the data set is one of the challenges in machine learning research. Predictive accuracy has often been used as one of the evaluation criteria for the predictive performance of classification or data



(a) (b)



(c)

Figure 3.14: Weight plots of features for implemented algorithms. (a) Weights vs. features using Information Gain. (b) Weights vs. features using Gain Ratio. (c) Weights vs. features using Relief.

Table 3.5: Top 5 important feature using implemented feature selection algorithms

Algorithm	1^{st} Feature	2^{nd} Feature	3^{rd} Feature	4^{th} Feature	5^{th} Feature
IG	172	173	174	171	170
GR	172	173	174	171	169
Relief	174	154	173	155	153

Table 3.6: Classification average accuracy and standard deviation of 10 folds for KNN, NN, RF, and SVM classifiers using data sets of top 5-features of IG, GR, and Relief algorithms

Model	Acc (Top 5 IG)	Acc (Top 5 GR)	Acc (Top 5 Relief)
KNN	0.9963 (1.16%)	0.9964 (1.13%)	1.00 (0%)
RF	0.9963 (1.16%)	0.9964 (1.13%)	0.9963 (1.16%)
NN	0.9976 (1.16%)	0.9964 (1.13%)	0.9976 (1.16%)
SVM	1.000 (0%)	1.00 (0%)	1.00 (0%)

mining algorithms. To overcome this issue, we have conducted experiments using different classification algorithms, and used the classifier's predictive accuracy on the experimental data set as the evaluation criterion.

Defect (crack) detection using reduced data sets has been evaluated using KNN, RF, NN, and SVM classifier models. The classifiers were tuned using grid search and cross-validation. Three levels of search grid where used to tune the parameters of the implemented models. SVM models were tuned in terms of the polynomial kernel degree and the regularization constant. NN models were tuned in terms of the number of hidden units and weight decay. Three odd levels (to avoid ties) of nearest neighbors were used for tuning KNN classifiers. The average classification accuracy was used as a criterion for model selection. Because the feature selection models returned the the top five features, 5-dimensional data sets were used to build the classifiers. Furthermore, additional 2-dimensional data sets were used for defect detection, based on the first two important features. Results have revealed high classification accuracy rates. Table 3.6 and Table 3.7 summarize the 5-dimensional and 2-dimensional classification models respectively. In total, 360 samples (observations) were used as follows: 270 observation for training using 10-fold cross validation and 90 observations held for unseen testing. More details about designing the training and test sets can be found in [76, 56].

The average training accuracy rates were higher than 0.995 for all models. Furthermore, the classification results indicate that the SVM outperformed the rest of the implemented models for all data sets, as it scored 100% accuracy rate for all data sets. The strong

Table 3.7: Classification average accuracy and standard deviation of 10 folds for KNN, NN, RF, and SVM classifiers using first and second important features of IG, GR, and relief algorithms

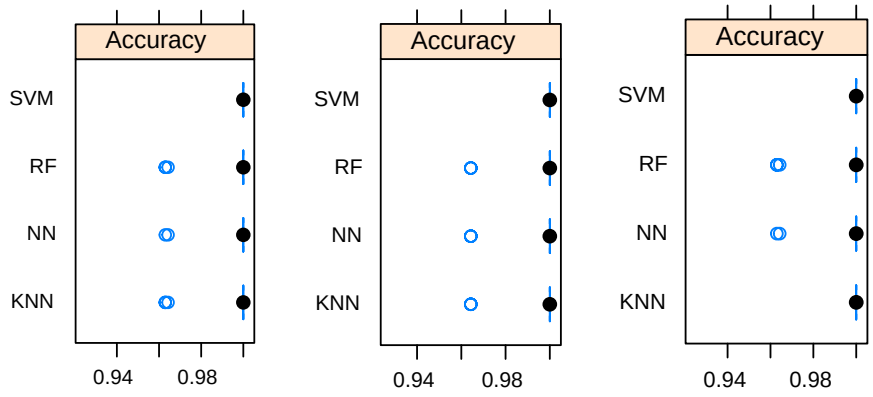
Model	Acc (Top 2 IG)	Acc (Top 2 GR)	Acc (Top 2 relief)
KNN	0.9964 (1.13%)	0.9964 (1.13%)	1.00 (0%)
RF	0.9964 (1.13%)	0.9964 (1.13%)	0.9963 (1.56%)
NN	0.9964 (1.13%)	0.9976 (1.13%)	0.9988 (1.17%)
SVM	1.000 (0%)	1.00 (0%)	1.00 (0%)

performance of SVM can be explained by the fact that the SVM measures the complexity of the hypotheses based on the margin with which it separates the dataset, not the number of features used, which in turn leads to better generalization compared to other algorithms[77].

Considering the relationship between the data sets and classification models employed, the data sets obtained by relief feature selection led to better accuracy, as the KNN classifier reached an accuracy rate of 100% only when it was working on data sets generated by the Relief algorithm. Accuracy variation, shown in Figure 3.15 and Figure 3.16 indicate that the metallic plate surface can be tested with only two frequencies, and an accuracy rate of 100% is achievable using SVM.

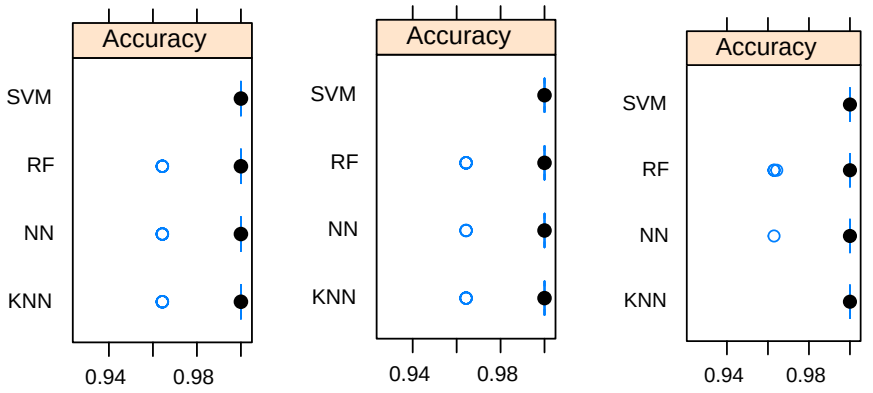
The performance of the classifiers implemented, based on the data set obtained using Relief feature selection technique was studied further in terms sensitivity and specificity, and results were reported as Kappa density [78] plots as in Figure 3.17 and Figure 3.18 for 5-dimensional and 2-dimensional data sets respectively. The kappa plots show that SVM and KNN outperformed RF and NN classifiers. Configurations of the classifiers operated on the data sets selected using Relief feature selection method are listed in Table 3.8.

Due to SVM classifier's strong performance in terms of the accuracy and kappa measures with all data sets, it has been selected as the final model among all implemented models. The 2-dimensional SVM model has been tested using unseen data, and the decision boundary learned is plotted in Figure 3.19. As observed from Figure 3.19 the accuracy rate of 100% was achieved.



(a) (b) (c)

Figure 3.15: Box-and-whisker diagrams indicating the accuracy variation for the implemented classifiers. (a) Based on the 5 top features using IG. (b) Based on the 5 top features using GR. (c) Based on the 5 top features using Relief.

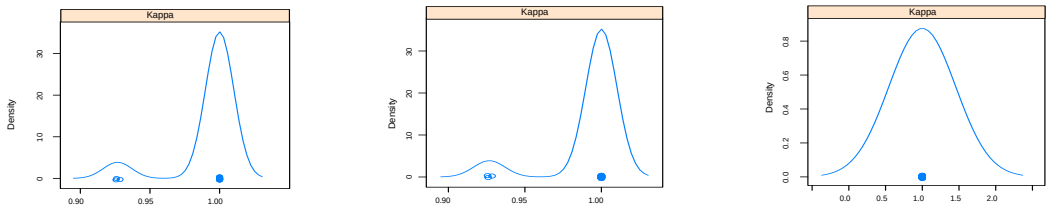


(a) (b) (c)

Figure 3.16: Box-and-whisker diagrams indicating the accuracy variation for the implemented classifiers using 2- dimensional data sets. (a) Using IG based data set. (b) Using GR based data set. (c) Using relief based data set.

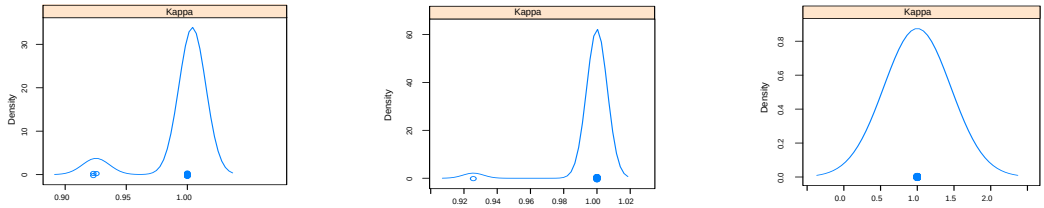
Table 3.8: Configurations of the classifiers built using the data sets selected using relief feature selection

	2-dimnesional,Models	5-dimnesional,Models
SVM	degree (1) regularization cost (0.25)	degree (3) regularization cost (0.5)
RF	variable per level (2)	variable per level (2)
NN	hidden units (3) weight decay (0)	hidden unit (5) weight decay (10^{-4})
KNN	k (5)	k (9)



(a) (b) (c)

Figure 3.17: Kappa density plots for the implemented classifiers using 5-dimensional data set using relief feature selection. (a) RF. (b) NN. (c) SVM [45].



(a) (b) (c)

Figure 3.18: Kappa density plots for the implemented classifiers using 2-dimensional data set using relief feature selection. (a) RF. (b) NN. (c) SVM.

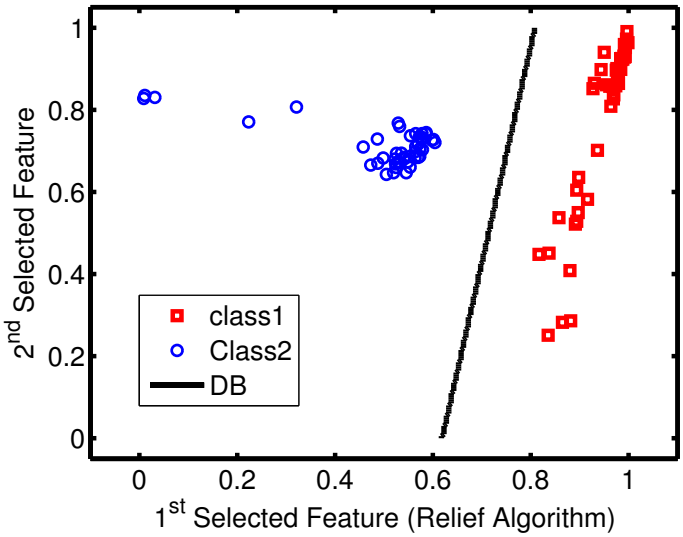


Figure 3.19: SVM decision boundary (DB) plot and unseen data from both classes.

Conclusions

This work has demonstrated the employment of ML feature selection to reduce sweeping frequencies in NDT. The experimental measurements required to test metallic surface hasa been reduced between five and two only. The resultant data sets have been evaluated using classification models including, RF, NN and SVM. The accuracy rates for all implemented classifiers' were higher than 0.995. Furthermore, based on the implemented classifiers' performance, the Relief selection algorithm was more effective than IG and GR.

Considering the training and testing classification accuracy rates achieved by the SVM classifier working on the 2d dimensional data set selected using Relief algorithm, the waveguide sensor can operate only at two frequencies (16.59 and 17.19 GHz), and a classification accuracy rate of 100% is achievable.

3.4 Discussion

Both case studies have shown that classification results were higher then 99% for out-of-sample scans (unseen data). As a result of implemented NN and SVM classifiers in case study 1 classification of any new samples is performed by operating the parameters of learned models on the new dataset (due to the fact that NN and SVM are model based). This means that, once the NN has been built (weights have been determined), classification of new data is merely matrix multiplication, a characteristic that is attractive for hand-held test equipment. Moreover, the proposed AI models decisions are made without any need for plotting the captured signals; hence, they can be embedded in in-field crack detection devices.

Outcomes of case study 2 have shown that utilizing feature selection for minimizing sweeping frequencies eliminates irrelevant features were from the original data set. Subsequently, selecting un-noisy and uncorrelated features leads to higher prediction accuracy than in case study 1. Moreover, case study 2 results have shown that the SVM classifier out-performed the NN and the RF classifiers. Furthermore, feature selection algorithms have indicated that the metallic surface can be tested using only two frequencies without compromising the classification accuracy, as illustrated in Table 3.7. As a major cost in developing detection systems is the frequency bandwidth; thus build scanning systems with only two instead of many operating frequencies leads to significant reduction in the electronic circuitry for portable detection systems and to a significant enhancement in time efficiency.

In brief, case sturdies 1 and 2 have demonstrated the practical feasibility of intelligent crack detection in metallic surfaces using AI models and a waveguide sensor. The proposed AI models were able to classify cracks that can easily be overlooked by the human eye. Consequently, this work can be generalized for different types of damage such as, corrosion or precursor pitting.

Chapter 4

Intelligent Near-Field Microwave Detection of Subsurface Anomalies

4.1 Introduction

Among[1] various non-destructive testing techniques, microwave testing is suitable for detecting buried anomalies in dielectric structures or hidden defects under coating or paint, as microwaves can penetrate coating layers and interrogate the metallic surfaces underneath [79, 13, 15]. Microwave-based detection modalities have gained strong interest in recent years because they do not require mechanical contact or mechanical penetration into the specimen under test [79, 13]. Near-field microwave-based sensors, in particular, can detect small flaws that are much smaller than the operating wavelength in free space [80]. In recent years, metamaterial and electrically small resonators have been used as sensors in the microwave frequency regime [49, 35, 38, 5]. Metamaterial cells such as complementary split-ring resonators (CSRR) have been used to detect cracks in metallic and non-metallic material and to gauge the thickness of substrates [81, 39, 65, 66].

[1] Part of the material presented in this chapter is published in [44],[89], and [97] and part is in preparation for publication.

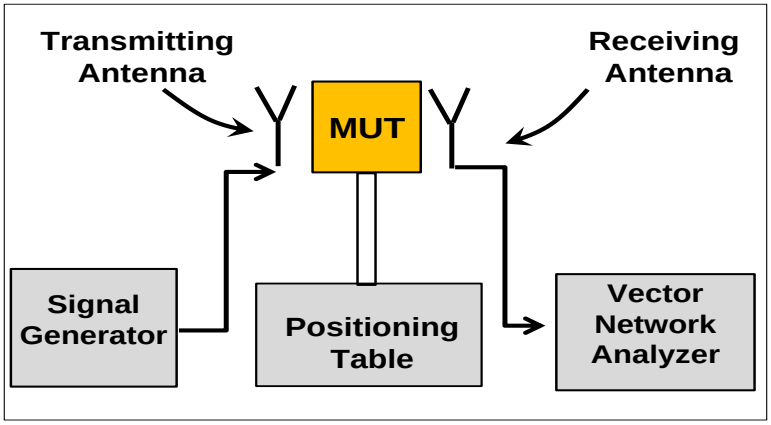


Figure 4.1: Typical microwave testing system configuration.

In general, microwave testing systems are comprised of transmission and receiving units that surround the material under test (MUT) as shown in Figure 4.1. However, in certain applications, only one side of the MUT is accessible. To address this challenge, one-sided sensors are used. For instance, open-ended waveguides, coaxial cables, and loaded transmission lines with split-ring resonators have been used for surface and subsurface detection [80, 49, 82, 83].

The use of near-field microwave sensors for defect detection is becoming increasingly attractive because results can be presented as images [80, 15, 84]. Mapping microwave sensing results as two-dimensional (2d) or three-dimensional (3d) images is useful in readily determining the size, shape, orientation and location of surface or buried anomalies or flaws. Conventional lenses are built from positive index materials and require curved surfaces to focus electromagnet waves on the image plane. However, finer details of the image are contained in the evanescent spectrum, which quickly decays before reaching the image plane [85]. The absence of evanescent waves restricts focusing with conventional lenses to a resolution on the order of half the operating wavelength. Enhancing the image resolution requires the contribution of the evanescent spectrum. Metamaterials have been used to enable sub-wavelength resolution [85, 86] and defects in certain structures can be imaged at high resolution by microwave sensors utilizing the evanescent spectrum [87].

Imaging systems using open-ended waveguides that operate at low frequency have a lower lateral resolution due to the large cross-sections of the waveguides [15]. To improve the lateral resolution waveguides need to work at high frequency [88]. Coaxial cable sensors have higher resolution than waveguides; however, coaxial cable sensors are sensitive to the stand-off distance and are time consuming to use [82]. Transmission lines loaded with CSRRs provide a practical trade-off between lateral resolution and imaging time. In addition, open-ended waveguide and coaxial cable sensors are one-port systems that can provide reflection coefficients only. In contrast, transmission line sensors loaded with CSRRs provide both transmission and reflection coefficients, thereby facilitating construction of images from transmission or reflection data. By making the sensors' rings (sensing elements) very small compared to the operating wavelength, the sensors become electrically small resonators. These sensors are one-sided, low profile and able to provide higher lateral resolution at lower frequency than waveguide sensors. The majority of microwave testing and imaging studies have focused on magnitude information [39, 81, 6]. Fewer studies [25, 83] have considered phase information. The reflection or transmission coefficient phases provide critical information as their changes show discontinuity patterns due to the existence of defects within the sensor's ambient region. In certain applications, this leads to higher detection sensitivity or sharper images than in cases when the magnitudes are used. Thus, the phase data of S_{11} and S_{21} are considered here.

In our recent work, CSRR based sensor was introduced as a viable sensor for imaging metallic surfaces [89]. This chapter presents comprehensive simulation and experimental studies of the CSRR sensor, showing that it can lead to practical and inexpensive imaging modality for anomaly detection dielectric and coated metallic structures. The magnitude and phase information were investigated to build two dimensional images and surface plots with the end goal of detecting surface flaws. The imaging resolution is largely dependent on the sensing element of the CSRR, which is the gap in the ring, not the arm length of the CSRR. The CSRRs gap in this study is very small compared to the size of the defect. The smaller the gap and scanning step are, the higher the resolution. A schematic diagram describing the imaging system is shown in Figure 4.2

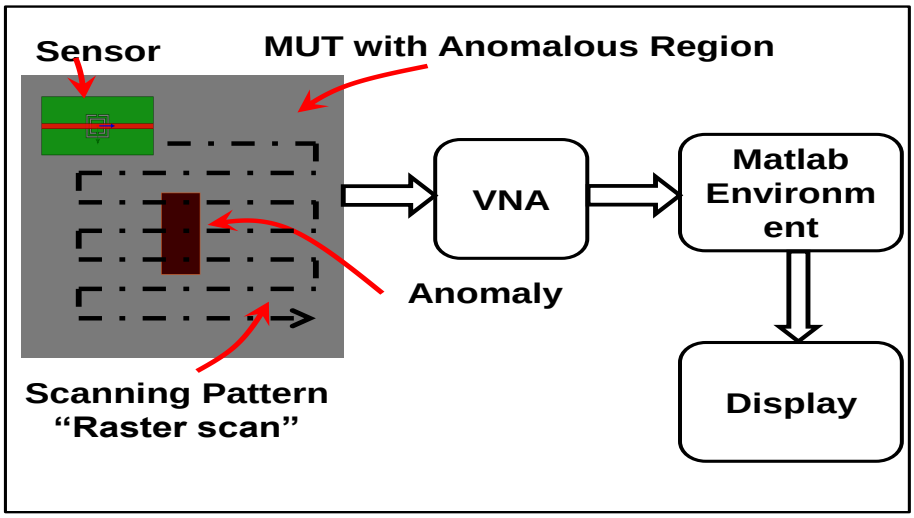


Figure 4.2: A schematic diagram of the imaging model.

4.2 Sensor Design and Operation

The employed sensor design is based on a microstrip transmission line with a complementary split-ring resonator (CSRR) etched in the ground plane of a printed circuit board (PCB). The transmission line is used to excite the CSRR structure, which acts as an electrically-small resonator whose dimensions were smaller than the operating wavelength. The sensing region is comprised of two co-centered split-square rings forming an inductor-capacitor (L-C) structure that resonates at a certain frequency. The CSRR cell is hosted on a t = 0.735 mm thick Rogers substrate with a dielectric constant of $\epsilon_r = 10.2$ and a loss tangent of $tan\delta = 0.002$. The geometric dimensions of the cell were chosen to maintain small sensor dimensions at a low frequency range (S-band). Figure 4.3 shows the sensor's layout. Figure 4.4 depicts the 3-d structure of the sensor, where the red line is the transmission line on top of the substrate and the green region is the ground plane on the bottom of the PCB. The CSRR has an arm length of 4 mm, gap of 0.3 mm, distance between inner and outer rings of 0.3 mm, and copper thickness of 35 μm.

The sensor's resonance frequency can be perturbed by changing the surrounding envi-

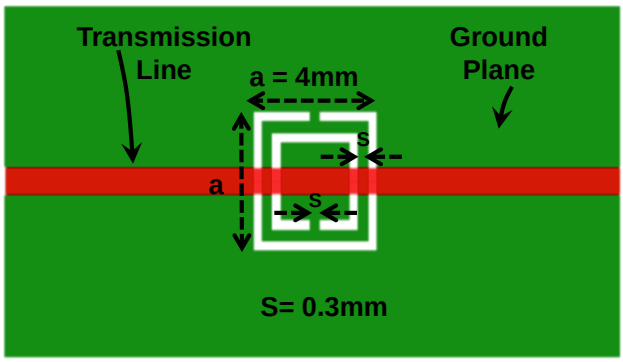


Figure 4.3: Schematic of sensor layout showing transmission line and etched CSRR in the ground plane.

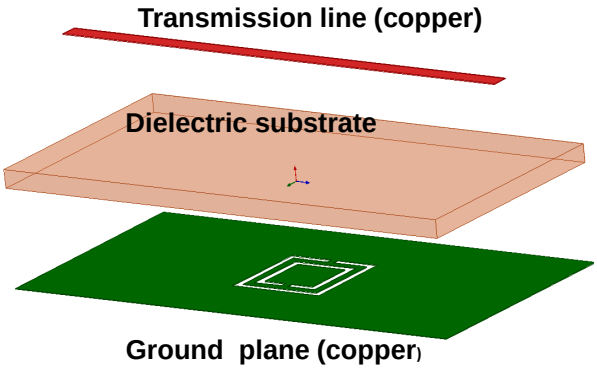


Figure 4.4: The sensor's structure.

ronment of the L-C circuit (the sensing element) [81, 39].

At resonance, the electric field energy and the magnetic field energy stored in the resonating structure are equal to each other. However, when an interacting material perturbs

the field distribution, the resonance frequency changes. The change in the resonance frequency is related to the interacting material properties and according to prior CSRR sensor studies, [90, 91, 39] this relationship formula can be presented by

$$\frac{\bigtriangledown f_r}{f_r} = \frac{\int_v (\bigtriangledown \epsilon \mathbf{E_1} \cdot \mathbf{E_0} + \bigtriangledown \mu \mathbf{H_1} \cdot \mathbf{H_0}) dv}{\int_v (\epsilon_0 |E_0|^2 + \mu_0 |H_0|^2) dv}, \tag{4.1}$$

where $\bigtriangledown f_r$ is the shift in the resonance frequency f_r, $\bigtriangledown \epsilon$, and $\bigtriangledown \mu$ are the changes in the permittivity and permeability, respectively, and v is the perturbed volume. $\mathbf{E_0}$ and $\mathbf{H_0}$ are the field distributions without the perturbation, and $\mathbf{E_1}$ and $\mathbf{H_1}$ are the field distributions with the perturbation. If the perturbation is small, then the field distributions are assumed to be unchanged. Thus, Eq. 4.1 can be simplified to

$$\frac{\bigtriangledown f_r}{f_r} = \frac{\int_v (\bigtriangledown \epsilon |E_0|^2 + \bigtriangledown \mu |H_0|^2) dv}{\int_v (\epsilon_0 |E_0|^2 + \mu_0 |H_0|^2) dv} \tag{4.2}$$

When the sensor scans a coated metallic plate or dielectric structure, it resonates at a certain design frequency. In this work, the resonance frequency was chosen at the S band, but other frequencies can be used. The reason for choosing the S band was to build a sensor that operates at low frequency with a sensing region of less than 5mm $\times$ 5mm to achieve high lateral resolution. The initial design of the resonator was proposed in [92] and studied and used for sensing in [81, 39, 65]. In this work, the resonator was tuned to operate at 3.8 GHz by fixing the gap and the separation between the rings and altering the rings' arms using the driven mode solver of ANSYS® High Frequency Structure Simulator (HFSS™). When the sensor encounters an anomalous region, its resonance frequency shifts. More precisely, when encountering a flaw such as corrosion, the transmission (S_{21}) and reflection (S_{11}) coefficients experience frequency and amplitude changes due to changes in the material's composition, shape or both. Figure 4.5 shows the $|S_{21}|$ plots for healthy and corroded surfaces. Figure 4.6 illustrates the simulated sensor's transmission coefficient ($|S_{21}|$) when the sensor is interacting with a healthy two-layer dielectric structure and when a buried anomaly was introduced in the second layer of the same structure. The following

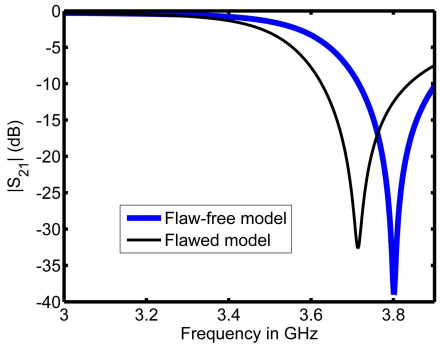


Figure 4.5: Magnitude of the transmission coefficient as a function of frequency.

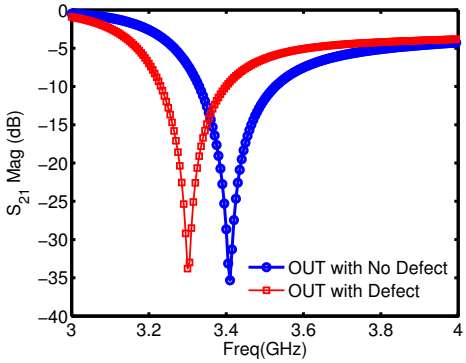


Figure 4.6: Transmission coefficient magnitude plots for anomaly-free and anomalous regions in a dielectric plate.

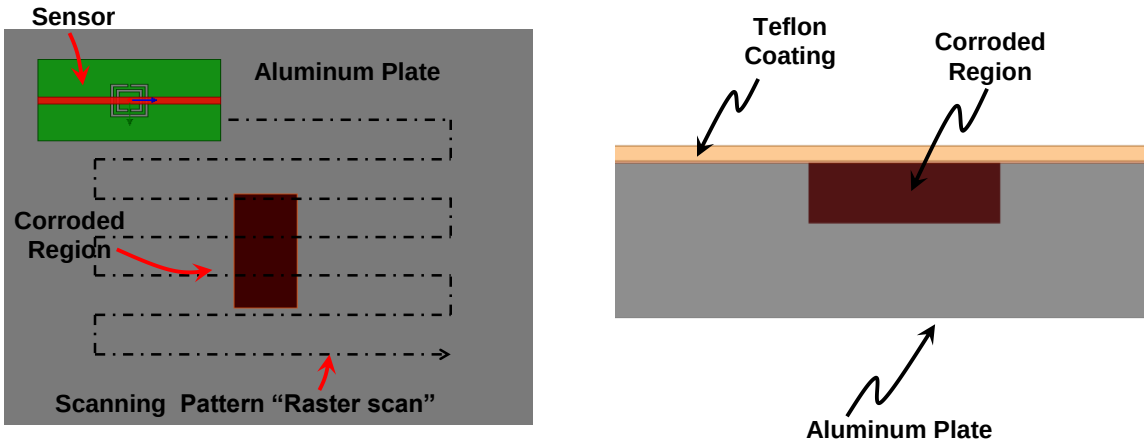
case studies present the ability of the CSRR sensor for detecting hidden anomalies in metallic and dielectric structures.

4.3 Imaging of Subsurface Anomalies using Single Metamaterial Resonator in Coated Metallic Structures: Case Study 3

In this study, two models of Teflon-coated aluminum plates were simulated. The first model had only one corroded region, whereas the second had two corroded regions with different depths. A raster scan around the flaw region(s), shown in Figure 4.7a, was used to image the MUTs. The resultant scattering matrix was later transferred to a Matlab environment for post-processing and image construction.

Simulation results and discussion

All simulation results presented in this work were obtained using ANSYS® HFSS™ [93].The first simulated MUT was an aluminum plate of 40 mm × 40 mm and a thickness of 5 mm, with a corroded region of 15 mm × 7.5 mm and depth of 2 mm, under a Teflon sheet of 50 μm. The corrosion parameters ($\epsilon_r = 8.42 - j1.03$) used were those reported in [25]. The Teflon cover is used to emulate paint layers that prevent the efficacy of visual inspection. The sensor scanned the MUT with a stand-off distance (air gap between the sensor and the imaged surface) of 0.5 mm, covering an area of 23 mm × 23 mm, in steps of 1mm in a raster scan pattern to construct an image of 529 pixels. Figures 4.7a and 4.7b show schematics for the first MUT's top and side views and the scanning procedure. For each pixel(scanning position) the frequency was swept over a 2-4 GHz frequency range in increments of 5 MHz. The results of the scattering matrix were transferred to a Matlab environment for post-processing. The sharpest images for the MUT were obtained near the higher end of the operating frequency range within the sensor's resonance frequency. A constructed image at 3.75 GHz using the magnitude of S_{21} for this simulation setup is shown in Figure 4.8a. The blue rectangular region in Figure 4.8a clearly shows the flaw region, and the yellow and red backgrounds correspond to the rest of the healthy scanned area. The magnitude of S_{21} at the center of the corroded region drops to -35 dB. However,

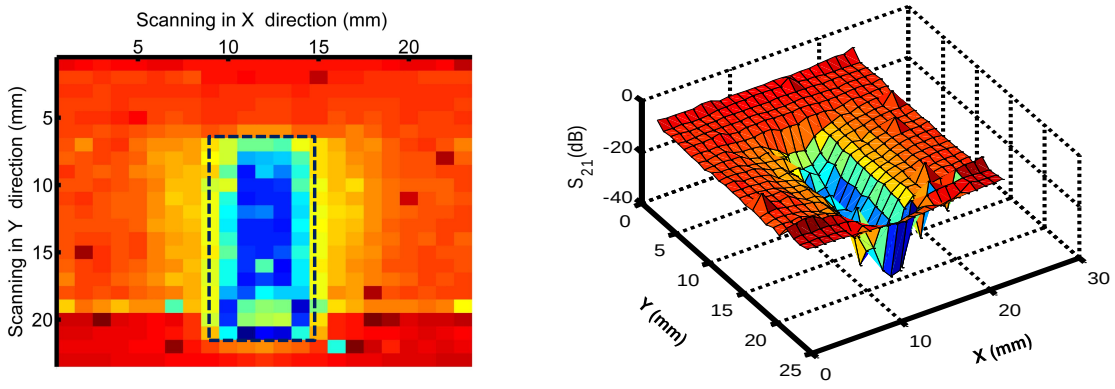


(a) (b)

Figure 4.7: (a) Raster scan around a corroded region in an aluminum plate coated with a Teflon layer of 50μm, and (b) Cross section view of the MUT.

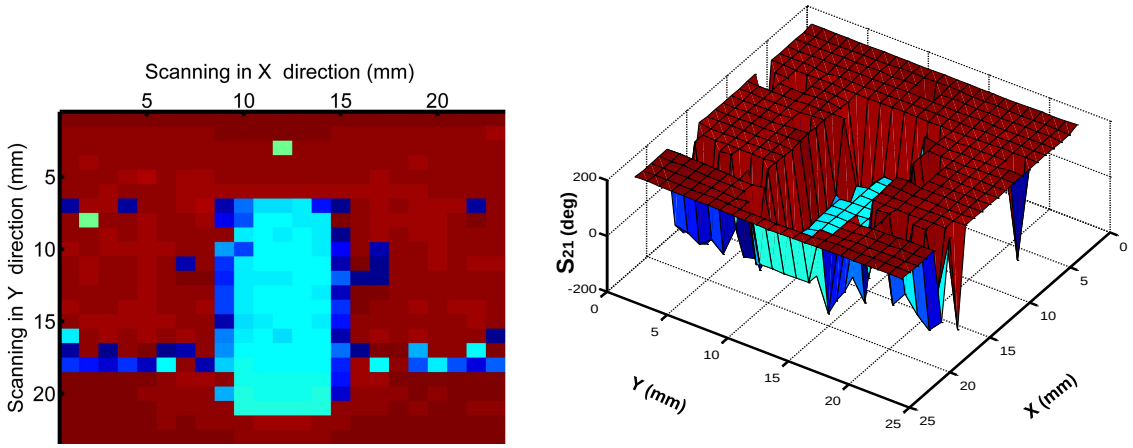
as the sensor leaves the flaw region, the magnitude of S_{21} starts to increase, to reach -15 dB around the corroded region (the yellow area in the plot) and increases to -5 dB when the sensor is far away from the defect (the red area in the plot). Comparing the flaws dimensions in the image to its real dimensions (shown in the dashed rectangle in Figure 4.8a) confirms that the constructed image not only detects the flaw but also characterizes it well in terms of shape and size. A second representation of the results is given as a surface plot in Figure 4.8b, which is more suitable for illustrating the flaw position and depth. This illustration represents the magnitude of S_{21} in dB of each 1mm $\times$ 1 mm of the scanned area. The plot of S_{21} magnitude sharply declines at the corroded region to form a cavity (in the surface plot) that corresponds to the presence of corrosion (the blue part of the surface plot).

So far, we have used the magnitude of S_{21} to build images of the corroded regions. Next, we produce images based on the phase of S_{21} and images based on the magnitude and phase of S_{11}. These images are shown in Figures 4.9a-4.11b. We observe that the transmission coefficient data reveals sharper images than the refection coefficient data. In



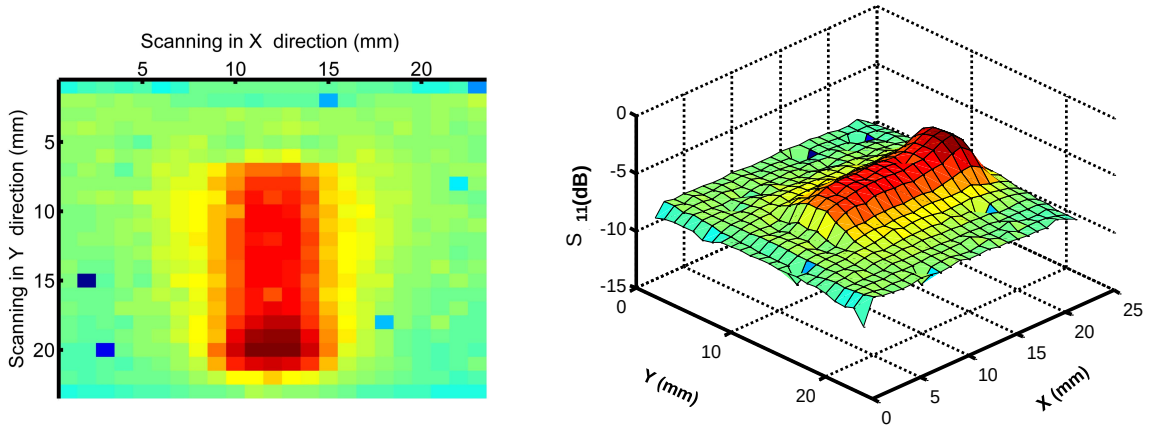
(a) (b)

Figure 4.8: Images of an aluminum plate with a corroded region using the magnitude of S_{21}. (a) Scaled image, and (b) Surface plot over the scanned area of 23 mm $\times$ 23 mm.



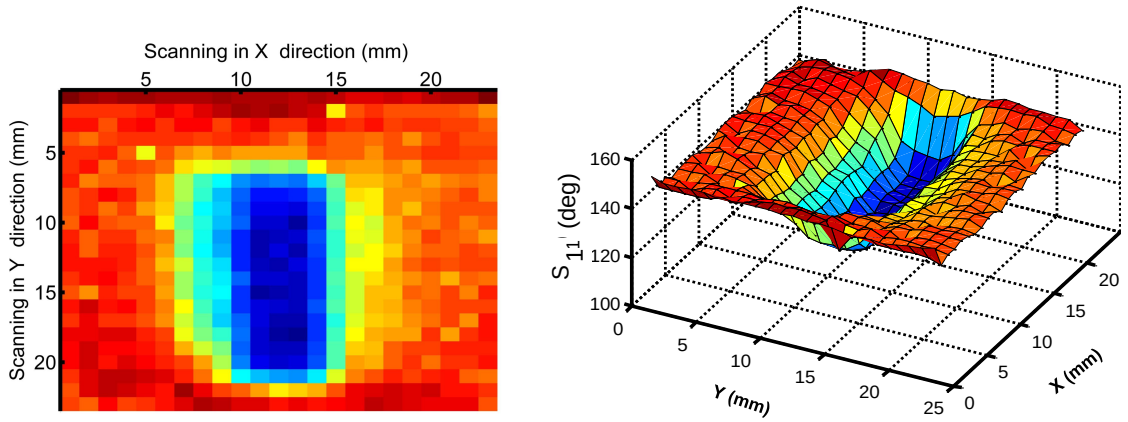
(a) (b)

Figure 4.9: Images of an aluminum plate with a corroded region using the phase of S_{21}. ((a) Scaled image, and (b) Surface plot over the scanned area.



(a) (b)

Figure 4.10: Images of an aluminum plate with a corroded region using the magnitude of S_{11}.(a) Scaled image, and (b) Surface plot over the scanned area.



(a) (b)

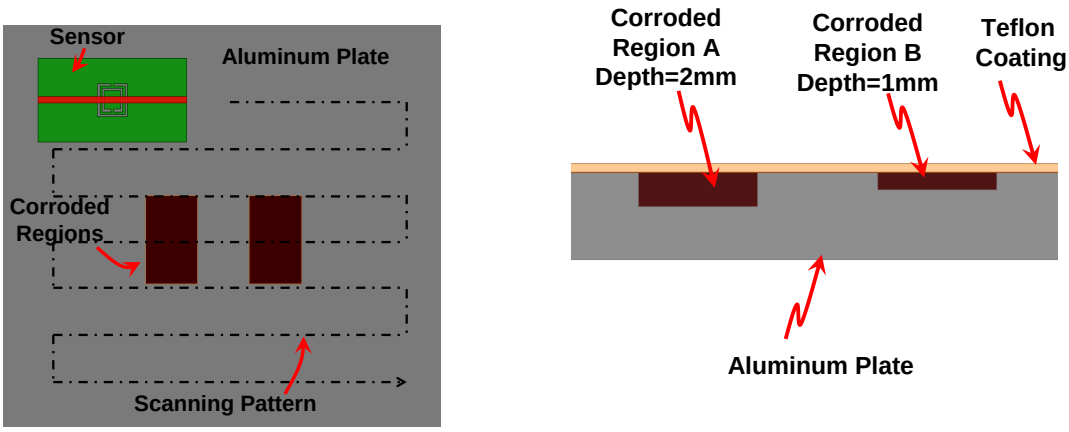
Figure 4.11: Images of an aluminum plate with a corroded region using the phase of S_{11}. (a) Scaled image, and (b) Surface plot over the scanned area.

fact, the image generated using the phase of S_{21} (Figure 4.9a is the least-blurry among the constructed image set, with the defect and its edges clearly detected and depicted as the light blue rectangular region in the plot. This high resolution in phase images is correlated to the high contrast between the S_{21} phase response at the flaw region and the rest of the MUT. In fact, the phase experienced discontinuity behavior when the sensor reached the flaw region, and it dropped from 180 to below -40 degrees, as shown in Figure 4.9b.

Images reconstructed using the magnitude of the reflection coefficient are depicted in Figures 4.10a and 4.10b, with the red region representing the flawed area and the green background representing the healthy area of the MUT. Figures 4.11a and 4.11b show the 2d image and surface plot reconstructed from the phase of the reflection coefficient, with the blue colored regions showing the defect region in the MUT and the red colored regions showing the healthy areas of the MUT. The surface plot in Figure 4.10b depicts the amplitude of the reflection coefficient S_{11} as a red peak; however, the defect is projected as a dip in the phase surface plot of the reflection coefficient in Figure 4.11b. The contrast between flawed and healthy regions is higher in images constructed using S_{21} than in images constructed using S_{11}.

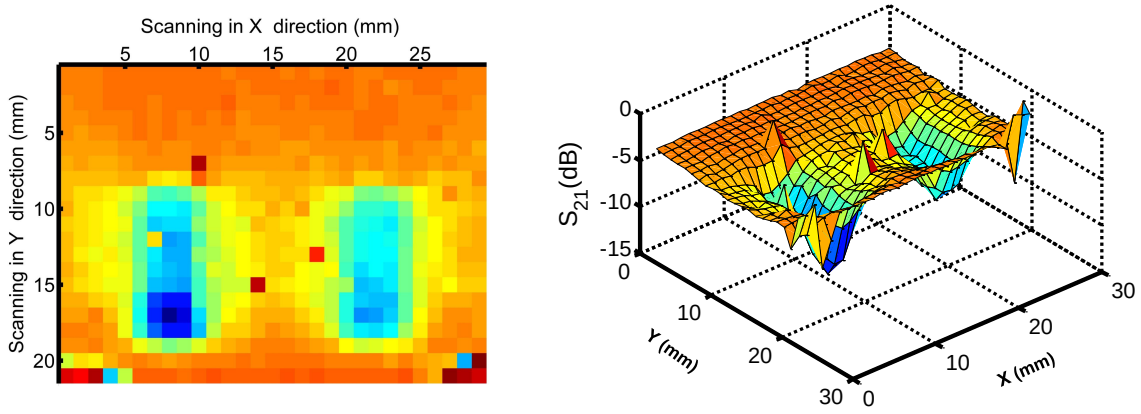
For the second test case, we considered a Teflon-coated aluminum plate with two corroded regions. The first flaw was a corroded region having an area of 10 mm $\times$ 7 mm and 2 mm depth; and the second flaw was a corroded region having an area of 10 mm $\times$ 7 mm and 1mm depth, as shown in Figure 4.12a. The separation distance between the regions was 7 mm. The sensor scanned the MUT with a stand-off distance of 0.5mm in steps of 1mm in a raster scan pattern constructing an image of 609 pixels for a scanned area of 29 mm $\times$ 21 mm. Scaled 2d images for the second simulation setup are given in Figures 4.13a, 4.14a, 4.15a, and 4.16a. These images clearly show the flaw regions, those with deeper flaws appearing darker than other regions. However, the image constructed using the phase ofS_{21} showed some noise as the sensor approached the edges of the MUT (Figure 4.14b).

For better flaw depth characterization, the results were depicted as surface plots, shown in Figures 4.13b, 4.14b, 4.15b, and 4.16b. These plots show the effect of the difference in



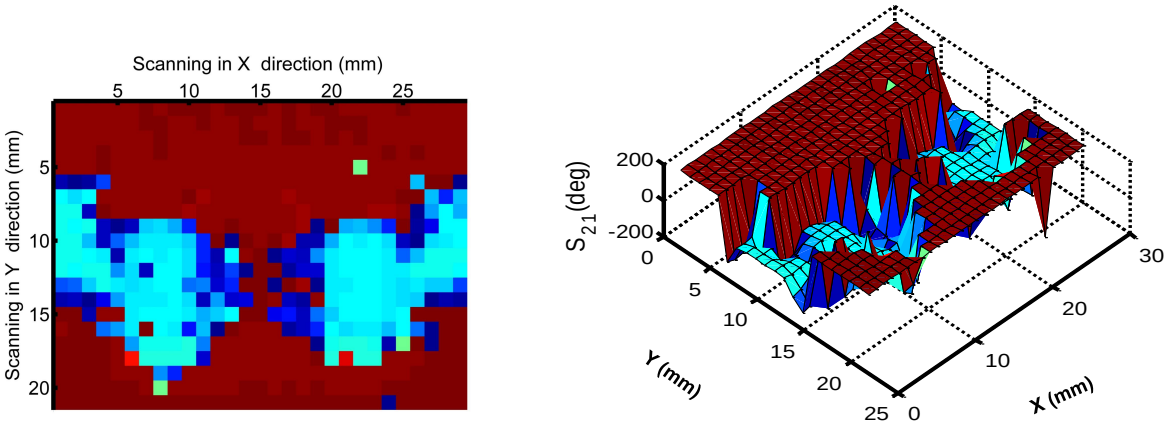
(a) (b)

Figure 4.12: (a) Raster scan around two corroded regions in aluminum plate coated with Teflon layer of 50μm, and (b) cross section view of MUT.



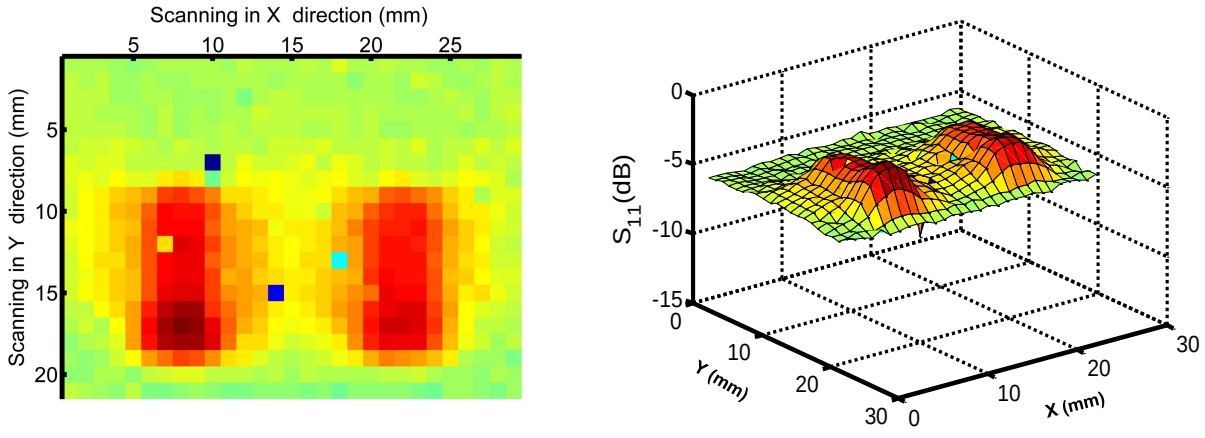
(a) (b)

Figure 4.13: Images of an aluminum plate with two corroded regions, constructed from the magnitude of S_{21}.(a) Scaled image, and (b) Surface plot over the scanned area.



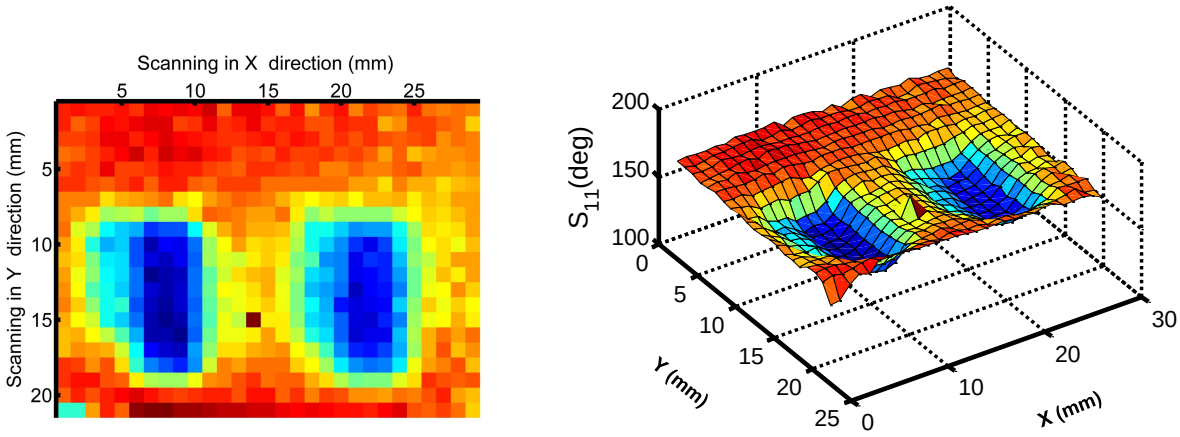
(a) (b)

Figure 4.14: Images of an aluminum plate with two corroded regions, constructed from the phase of S_{21}. (a) Scaled image, and (b) Surface plot over the scanned area.



(a) (b)

Figure 4.15: Images of an aluminum plate with two corroded regions, constructed from the magnitude of S_{11}. (a) Scaled image, and (b) Surface plot over the scanned area



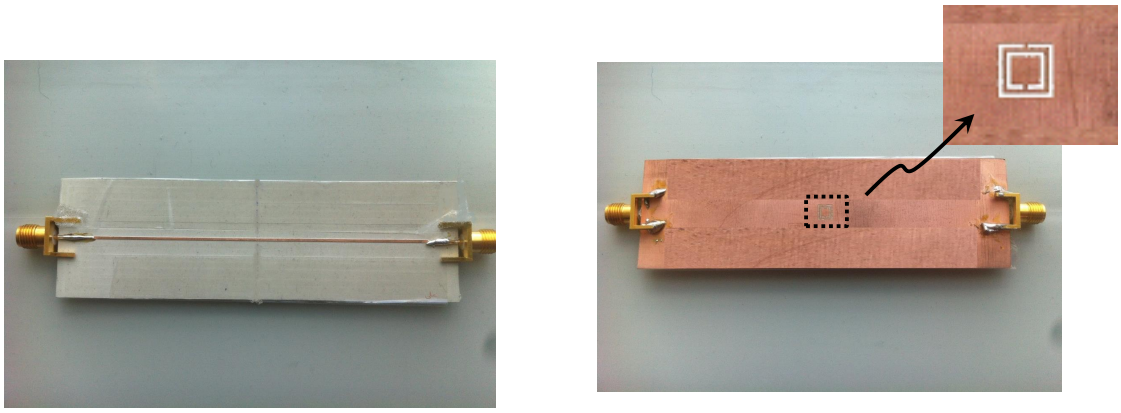
(a) (b)

Figure 4.16: Images of an aluminum plate with two corroded regions, constructed from the phase of S_{11}. (a) Scaled image, and (b) Surface plot over the scanned area

the flaws' depths more distinctly, with the 2mm deep corroded region showing a deeper dip than the 1mm deep region (Figure 4.13b). The separation distance between the flaw regions is depicted clearly in the 2d figures and even more distinctly in the surface plots, where it is shown as a flat region similar to the rest of the healthy part of the MUT.

Fabrication and experimental set-up

The CSRR sensor was fabricated using an RO6010 printed circuit board with a dielectric constant of $\epsilon_r = 10.2$ and a thickness of 0.64mm; Figure 4.17a and Figure 4.17b show top and bottom views of the fabricated sensor, respectively. An aluminum plate of 100 mm $\times$ 40 mm and a thickness of 12mm, with two rectangular dents, has been fabricated. The dents were filled with corrosion powder. The aluminum plate was covered using Teflon tape of a thickness of 80 μm (multi-purpose Teflon tape was used). The first dent (left-hand side) was 20 mm $\times$ 10 mm, with a depth of 2 mm , while the second one was 20 mm $\times$ 10 mm, with a depth of 1mm, and the separation distance between the corroded regions was



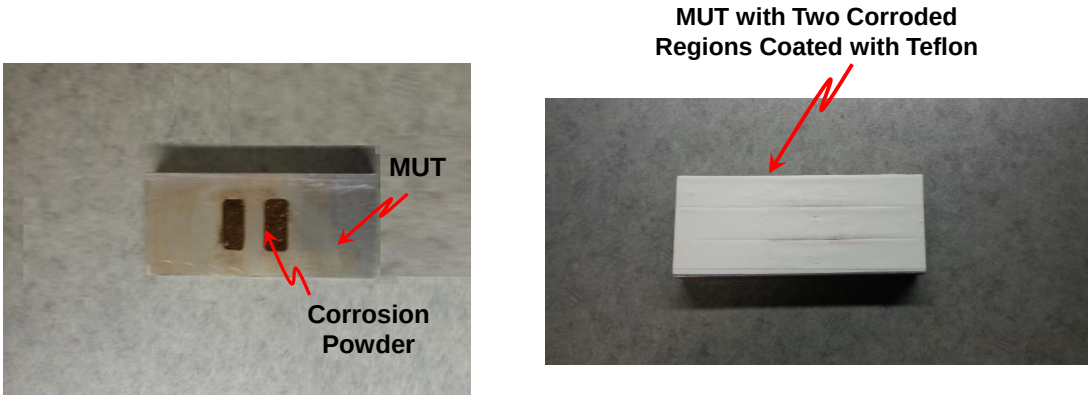
(a) (b)

Figure 4.17: Fabricated sensor photograph: (a) Top view, (b) Bottom view

10 mm. Figure 4.18a and Figure 4.18b show the MUT preparation process.

Figure 4.19 shows the magnitude of the transmission coefficient S_{21} for healthy and corroded aluminum plates. The S_{21} plots in Figure 4.19 indicate good agreement between experimental and simulation outcomes, demonstrated by the clear shift (S_{21}) experienced when the sensor was facing a corroded region. However, there was a small frequency difference of 50 MHz between the experimental and simulation results, as the thickness of the Teflon coating in the simulation is 50 μm, while the thickness of the commercial Teflon tape used in the experiment was was 80 μm.

Then the sensor was used to image the MUT manually in a raster scan with 3 mm steps in the X and Y directions and a stand-off distance of 0.5 mm to construct an image of 204 pixels (an area of 50 mm $\times$ 36 mm) around the corroded regions. In this set-up, a vector network analyzer (VNA) was used to obtain S-matrix parameters over a frequency range of 2 to 4.5 GHz .



(a) (b)

Figure 4.18: MUT preparation process: (a) MUT with two areas filled with corrosion powder as a flaw, and (b) Teflon coated MUT.

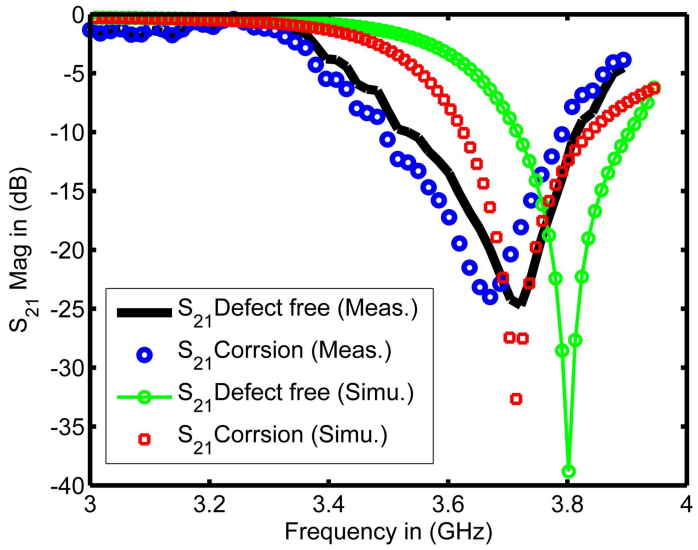


Figure 4.19: Comparison between simulated and measured $|(S_{21})|$ curves of the CSRR sensor.

Experimental results and discussion

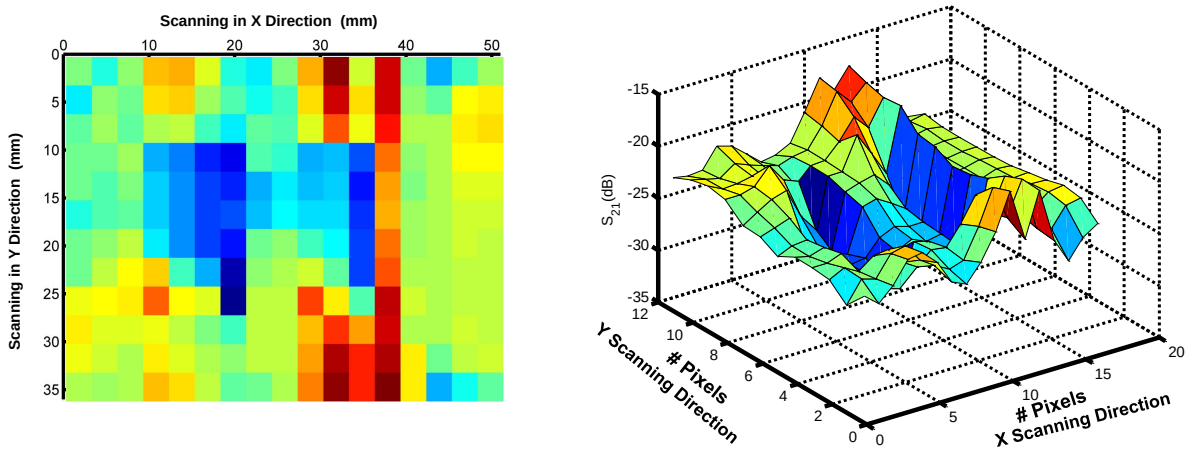
The experimental data was transferred to a Matlab environment for digital signal processing. The magnitude and phase of transmission and reflection coefficients were examined and processed to construct 2d images and surface plots for the MUT, as illustrated in the following subsections.

Images Based on the Transmission Coefficient

The constructed images based on the measured S_{21} revealed clear defect detection and localization using magnitude or phase information. As the simulation results showed, the 2d images constructed using phase data were sharper than images constructed using magnitude data. In particular, the edges of the flaw were exceptionally depicted when using the phase information, as illustrated in Figure 4.20a and Figure 4.21a. In fact, when the sensors encountered corrosion, the phase of S_{21} changed by more than 10°; however, the magnitude of S_{21} changed by less than -7 dB. Also, we observe from Figure 4.20a and Figure 4.21a that the corroded region with 2mm depth is depicted with a sharper image than the region with 1 mm depth. Furthermore, in the surface plots, both flaw areas were clearly shown as blue cavities (Figure 4.20b and Figure 4.21b). Although the scan was performed manually and with a scanning step three times larger than the simulation scanning step, the surface plots showed the separation distance between the flaw areas clearly, indicating good lateral resolution of the CSRR sensors.

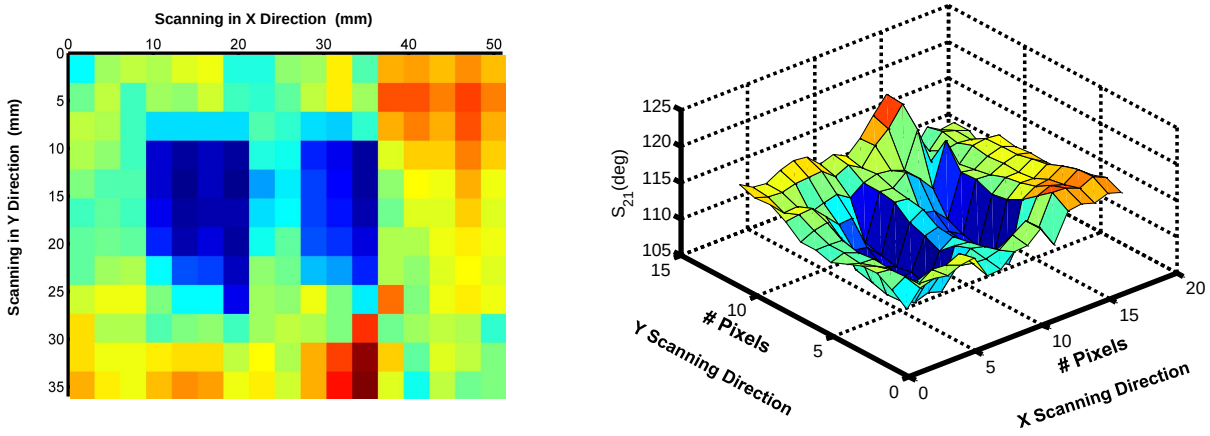
Images Based on the Reflection Coefficient

The constructed images based on the measured S_{11} presented in Figures 4.22a, 4.22b, 4.23a, and 4.23b indicated the defect regions. The image constructed from the phase data, however, is more informative and less blurry than the images constructed from the magnitude data. We also observed that the corroded region with 1 mm depth is depicted more clear in the image constructed from the phase data of S_{11} than in the magnitude image. In the constructed surface plots, the separation distance between the flaw regions was represented



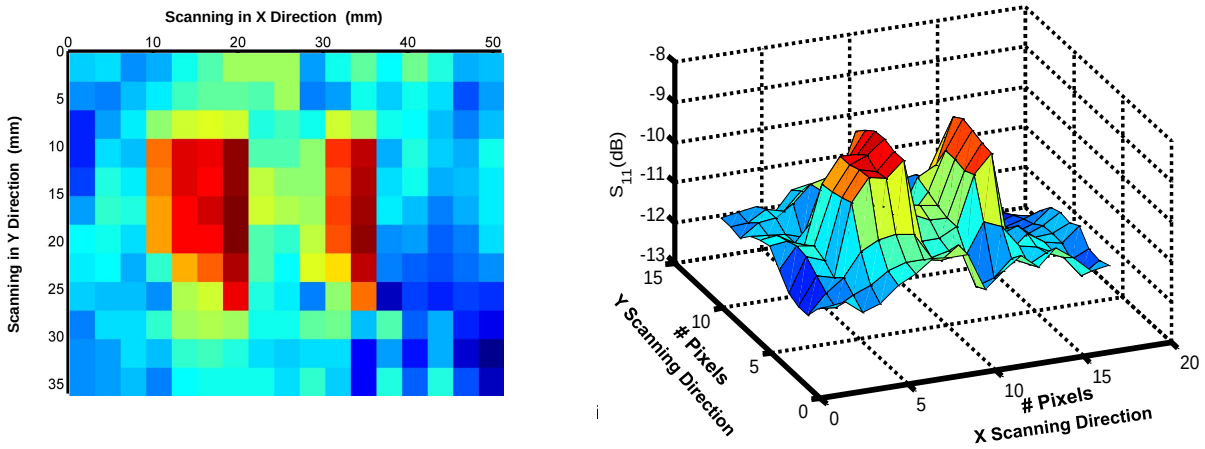
(a) (b)

Figure 4.20: Images of an aluminum plate with two corroded regions, constructed using measurements of the magnitude of S_{21}. (a) Scaled image, and (b) Surface plot over the scanned area of 50 mm $\times$ 40 mm



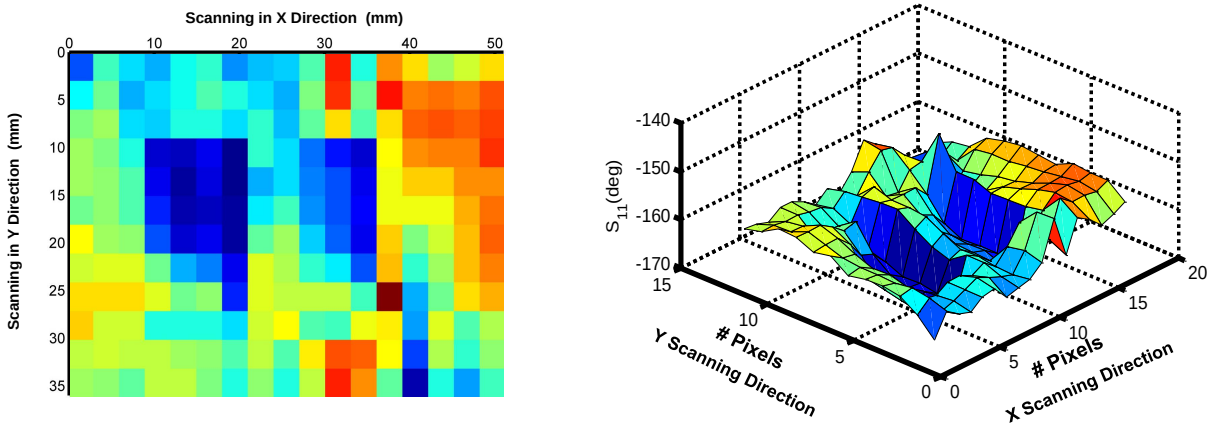
(a) (b)

Figure 4.21: Images of an aluminum plate with two corroded regions, constructed using measurements of the phase of S_{21}. (a) Scaled image, and (b) Surface plot over the scanned area



(a) (b)

Figure 4.22: Images of an aluminum plate with two corroded regions, using measurements of the magnitude of S_{11}. (a) Scaled image, and (b) Surface plot over the scanned area



(a) (b)

Figure 4.23: Images of an aluminum plate with two corroded regions, using measurements of the phase of S_{11}. (a) Scaled image, and (b) Surface plot over the scanned area

as a flat area between the two peaks when the magnitude data was used. For the phase measurements, the separation distance between the flawed regions was visible as the flat area between the two cavities.

Comparing the experimental results revealed that the images constructed using S_{21} were sharper than the images constructed using S_{11}. Similar to the simulation results, images generated using the phase information of S_{21} revealed more accurate information about the location of the defects than those obtained using the magnitude information.

Conclusion

This work has demonstrated numerically and experimentally the ability of electrically-small microwave ring resonators to image flaws in coated metallic structures. Results have shown good images and high lateral resolution for corroded regions in aluminum plates coated with Teflon sheets. In this study, the transmission coefficient phase has achieved the highest resolution and best edge detection among the scattering matrix parameters. Phase-based images of S_{21} and S_{11} were observed to be sharper than magnitude-based images, due to the rapid changes in the phase when the sensor encountered corroded regions. The CSRR sensor provided good flaw depth characterization as depicted by the surface plots where the dynamic range of the magnitude or phase change is correlated to flaw depth. Based on the numerical analysis performed in this study, the lateral resolution in the experimental results can be enhanced further by reducing the scanning steps.

Compared with sweep frequency microwave waveguide imaging [15, 88], microwave CSRR based imaging has the advantage of operating at low frequency with enhanced resolution utilizing evanescent waves. The proposed system operated in the S-band (2 GHz to 4 GHz) to image a coated corrosion of area 7.5 mm $\times$ 15 mm while waveguide imaging methods operated in the K band (18 GHz to 26.5 GHz) to image the coated corrosion of an area 30 mm $\times$ 30 mm.

4.4 Imaging Corrosion Under Coatings: An Artificial Intelligence Approach: Case Study 4

Introduction

The corrosion of metallic structure is considered to be a major problem facing engineers today. As infrastructure ages, maintenance tasks become more challenging technically and financially [94]. The problem of metallic-structure corrosion, for instance, may lead to greater consequences than simple metal loss [95]. In some cases, corrosion related issues may result in injuries, fatalities, environmental hazards, and reduced infrastructure performance. To give an idea on the direct cost of corrosion in the U.S., for example, the National Association of Corrosion Engineers International (NACE) estimated that the cost will exceed $ 1 trillion in 2013 [96].

Several studies have focused on Non-Destructive Testing (NDT). These NDT approaches are based on ultrasound, eddy currents, magnetic particle testing, dye penetrant, and visual testing. However, these technologies face challenges in detecting corrosion under non-metallic coatings. As coatings increase the stand-off distance between the sensor and target surface, the difficulty of detecting underneath corrosion increases [13]. Therefore, due to the good penetration property of electromagnetic waves in dielectrics, microwave sensing has been applied to detect corrosion under coatings[89, 13]. Motivated by the arguments above, the idea of having a system that combines artificial intelligence (AI) and microwave sensing has emerged and been tested on a simulated study of a near-field printed circuit board (PCB) based sensor. The sensor's design is studied at the beginning of this chapter. The current case study is based on using PCB sensors to test metallic plates with defects and defect-free. The obtained data were then processed using ML models to classify any defects in the plates. Finally, results were presented as binary images.

Simulation Set-up and Data Generation

Building an AI classifier model requires training and tuning using labeled data first; then it becomes ready for classification of unlabeled data during the test phase.

Training Dataset

The method used to collect the training dataset is based on the following simulated models using ANSYS®-HFSS™: (1) A Teflon coated metallic plate was scanned using a raster scan 0.5 μm stand-off distance. The metallic plate under test has one defected region of 15 mm $\times$ 7.5 mm, with a 2 mm depth, as Figure 4.24 illustrates. (2) Another Teflon coated metallic plate with two corroded regions was scanned at a 0.5 μm stand-off distance . The corroded regions were 15 mm $\times$ 7.5 mm each. However, the defects were designed to have different depths (1mm and 2mm) as in Figure 4.25. Including the sensor's response from different corroded region depths helps the learned classifier to generalize better for different test scenarios. Collected sensor's responses (samples) were labeled according to the health status (a positive label means defect and a negative label means no defect). The AI model training and evaluation was performed using the R project environment [75].

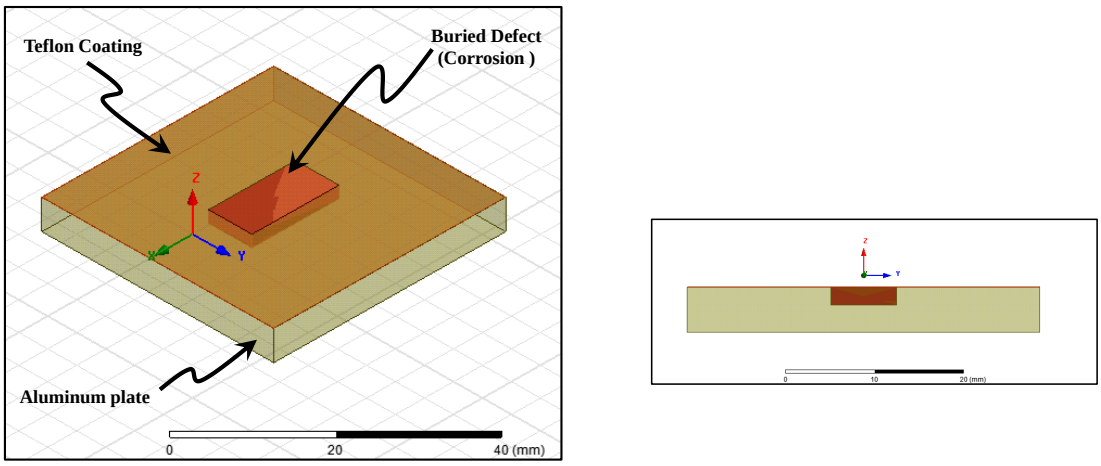
Support Vector Machine Classifier

The support vector machine (SVM) algorithm has unique property, where it maximizes the margin between classes. SVM is a binary classifier by default, which suites the case under study. SVM model parameters correspond to a convex optimization problem, thus any local solution is also a global optimum [46]. Considering SVM for classification of a two-class problem using linear models of the form:

$$y(x) = W^T \phi(x) + b \tag{4.3}$$

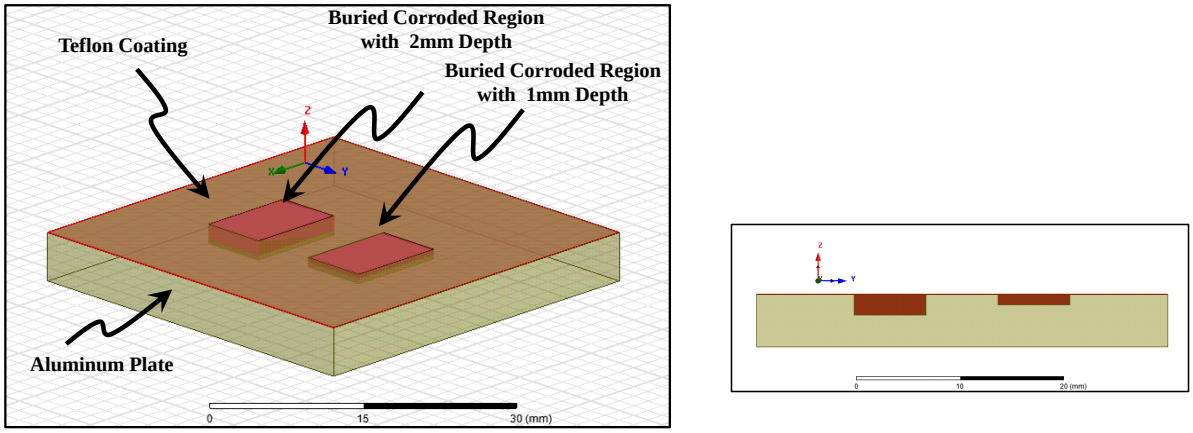
where $\phi(x)$ denotes a fixed feature-space transformation, and b is the bias parameter.

The SVM classifier achieved high classification accuracy, mainly by optimizing the kernel type, degree, and termination criterion tolerance as indicated in Figure 4.27. During



(a) (b)

Figure 4.24: Metallic plate with corroded region coated with 50μm Teflon layer (a) top view (b) side view



(a) (b)

Figure 4.25: Metallic plate with two corroded regions coated with 50μm Teflon layer (a) top view (b) side view

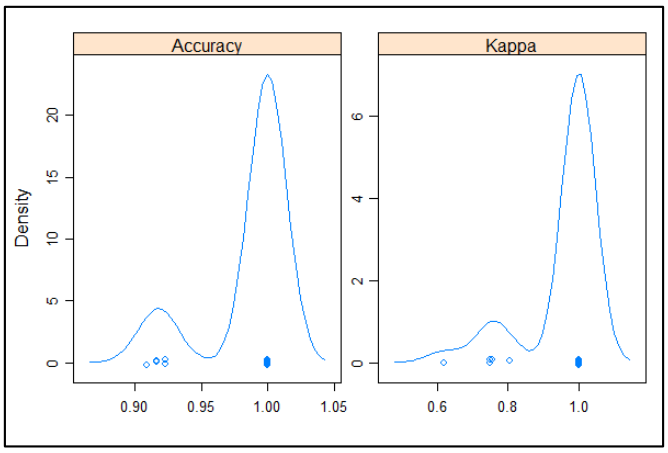


Figure 4.26: Accuracy and kappa density distributions for three repeated 10 fold cross-validation

SVM optimization, the model showed that a polynomial kernel of degree 3 achieved higher validation accuracy rates than other polynomial kernels of degree 1 or 2. The accuracy and kappa density distributions of the selected model are shown in Figure 4.26.

AI Model Testing and Results

Three test scenarios were simulated to validate the effectiveness of the trained (learned) AI model. Figure 4.28 shows the shape of the defected structure used to test the learned model. The test procedures were carried at 250 μm , 625 μm, and 750 μm stand-off distances to evaluate the ability of the built classifier to operate at different stand-off distances

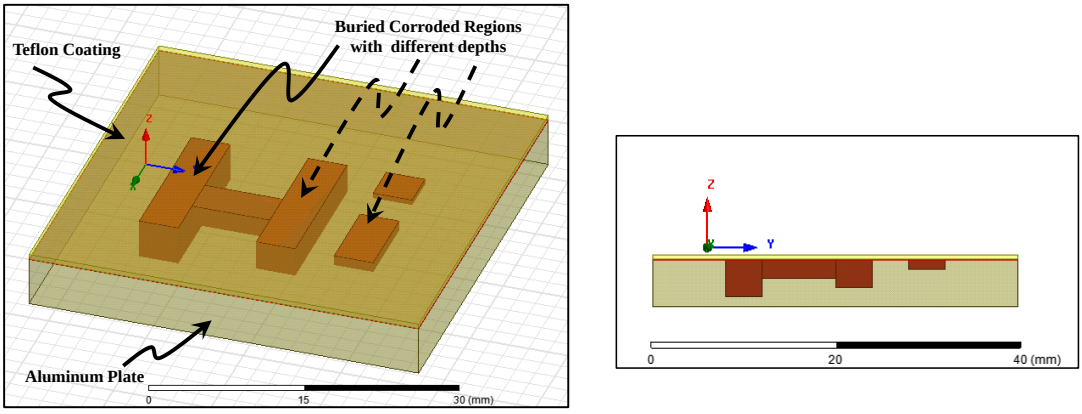
The MUT used to test the learned classifier was a metallic plate with corrosion defects arranged in an HI shape. The depth of defect arms were 0.5 mm, 1 mm, 2 mm , 3 mm, and 4 mm. The images shown in Figure 4.29 were constructed from the built SVM classifier's outputs after it operated on the testing MUT (Figure 4.28) at different stand-off distances. Each image shown in Figure 4.29 is 56 pixels (8×7). Although, all training samples were collected at 0.5 mm, the classifier achieved 100 % accuracy at 250 mm (see Figure 4.29 (a)) and 96 % at 625 μm and 750 μm (only two pixels were misclassified), as illustrated in Figure 4.29 (b) and Figure 4.29 (c). To emphasize the ability of the SVM classifier to maximize the separation margin between classes (a feature that suites the case under

Support Vector Machines with Polynomial Kernel
2 classes: 'CR', 'NC'
Resampling: Cross-Validated (10 fold, repeated 3 times)
Resampling results across tuning parameters

degree	scale	C	Accuracy	Kappa
1	0.001	0.25	0.9093046	0.7061828
1	0.001	0.50	0.9230186	0.7331740
1	0.001	1.00	0.9393357	0.7739919
1	0.010	0.25	0.9421134	0.7814919
1	0.010	0.50	0.9472416	0.7978441
1	0.010	1.00	0.9588578	0.8522981
1	0.100	0.25	0.9530497	0.8271544
1	0.100	0.50	0.9509518	0.8310294
1	0.100	1.00	0.9618493	0.8724616
2	0.001	0.25	0.9225524	0.7317157
2	0.001	0.50	0.9421134	0.7814919
2	0.001	1.00	0.9421134	0.7814919
2	0.010	0.25	0.9704351	0.9044295
2	0.010	0.50	0.9618493	0.8690447
2	0.010	1.00	0.9618493	0.8685715
2	0.100	0.25	0.9598096	0.8097016
2	0.100	0.50	0.9811577	0.9291802
2	0.100	1.00	0.9702214	0.8751843
3	0.001	0.25	0.9393357	0.7739919
3	0.001	0.50	0.9421134	0.7814919
3	0.001	1.00	0.9479215	0.8066355
3	0.010	0.25	0.9506993	0.8191355
3	0.010	0.50	0.9644522	0.8748263
3	**0.010**	**1.00**	**0.9862859**	**0.9560459**
3	0.100	0.25	0.9199883	0.6536786
3	0.100	0.50	0.9393357	0.7417908
3	0.100	1.00	0.9311383	0.7078317

Accuracy was used to select the optimal model using the largest value.
The final values used for the model were degree = 3, scale = 0.01 and C = 1.

Figure 4.27: SVM model resampling results for 10 fold cross-validation



(a) (b)

Figure 4.28: Testing MUT with HI shape corrosion (a)Top view (b) Front view

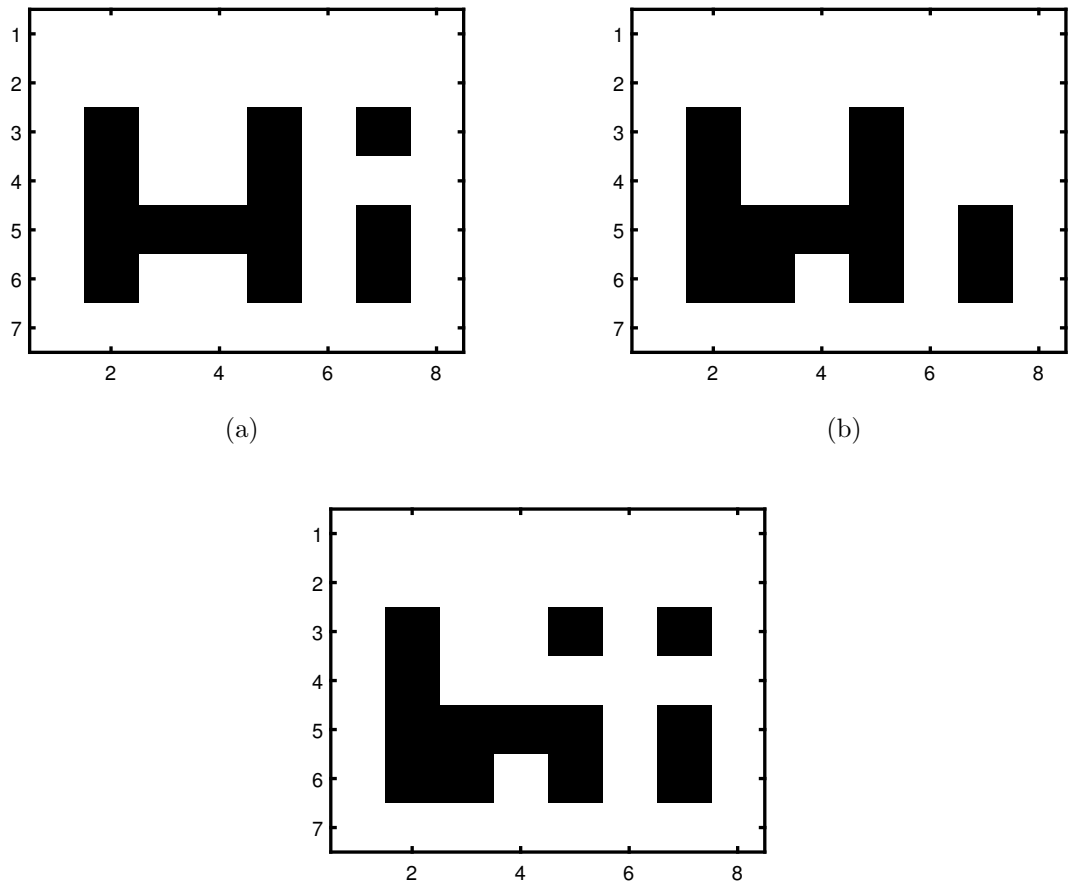
Table 4.1: My caption

	SVM Accuracy	RF Accuracy
Testing MUT at 250 μm stand-off distance	100%	92.8%
Testing MUT at 625 μm stand-off distance	96.5%	83.9%
Testing MUT at 750 μm stand-off distance	96.4%	87.5%

study), RF, NN, and KNN classifiers have been trained on the same training data and tested as was as the SVM. Results showed that the RF model accuracy rates were higher than the NN and KNN accuracy rates. However, It is clear from the constructed images in Figures 4.30 (a), (b), and (c) that the RF model was not able to generalize as the SVM did (see Figure 4.29) after variations in defect depth and changes in stand-off distances were introduced. Table 4.1 summaries the testing accuracy rates obtained from the trained SVM and RF classifiers for the testing MUT shown in the Figure 4.28 .

Conclusion

This work has numerically demonstrated a corrosion imaging system that utilizes a microwave sensor and Al model. The AI model was implemented using SVM because it maximizes the margin between classes and can thus generalize effectively with variations in defect depths and changes in standoff distances. The trained model has been trained for only two defect depths (1 mm and 2mm) and one stand-off distance of 500 μm. However, the trained AI model has shown high classification accuracy rates even for varying defect depths and stand-off distances.

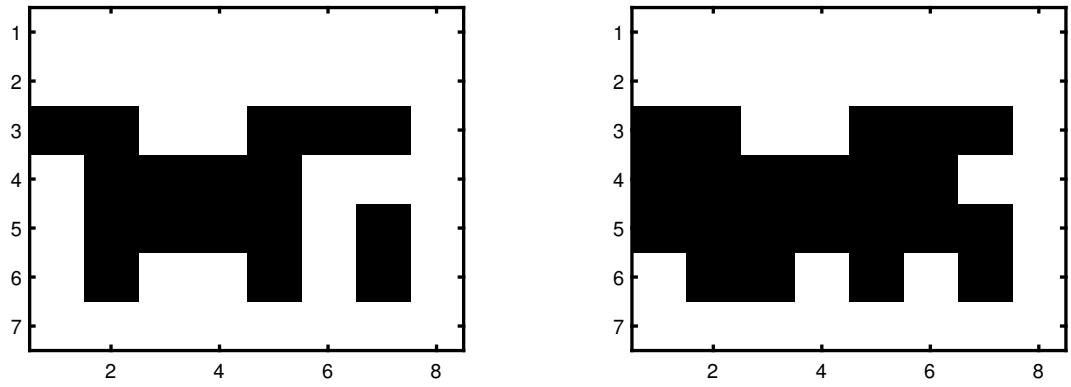


(a)

(b)

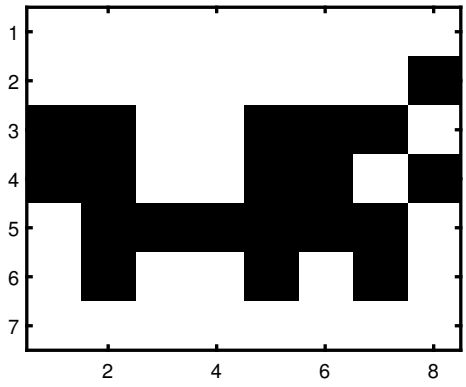
(c)

Figure 4.29: Constructed images of the testing MUT using pixel classification results obtained from the trained SVM classifier (a) Testing MUT image when the sensor operates at 250 μm stand-off distance. (b) Testing MUT image when the sensor operates at 625 μm stand-off distance. (c) Testing MUT image when the sensor operates at 750 μm stand-off distance.



(a)

(b)



(c)

Figure 4.30: Constructed images of the testing MUT using pixel classification results obtained from the trained RF classifier (a) Testing MUT image when the sensor operates at 250 μm stand-off distance. (b) Testing MUT image when the sensor operates at 625 μm stand-off distance. (c) Testing MUT image when the sensor operates at 750 μm stand-off distance.

4.5 Revealing Buried Anomalies in Multi-Layered Dielectric Structures: Case Study 5

Microwave imaging has gained increasing interest in recent years due to its ability to reveal hidden or buried objects [13]. Additionally, near field microwave sensors have the capability to detect small anomalies [80]. In different studies, artificially engineered electromagnetic materials (metamaterials) have been implemented to demonstrate a strong localization and enhancement of electrical fields around the sensing element in order to improve the microwave near-field sensor's sensitivity and resolution [49, 5, 38]. Microwave testing can be two sided or one sided based on the application and the material under test (MUT). In one-sided testing one port or multi-port configurations can be used. Reflection coefficient information is the only available information obtained with one-port configurations. In contrast, transmission and reflection coefficient information are available with multi-port configurations.

Different microwave imaging techniques using one-port systems have been studied including open-ended waveguides [13, 25, 15] and coaxial cables [82]. Coated metallic surfaces imaging using a two-port one-sided modality has been reported based on transmission lines loaded with small resonators using printed circuit board technology (PCB) [89, 97, 44]. This study explores the use of an imaging dielectric in which transmission lines are loaded with small resonators. By designing the resonators to be electrically-small compared to the operating wavelength, the sensor is expected to provide high lateral resolution.

Sensor Design and Testing Procedure

The sensor design adopted here is based on a microstrip transmission line with a complementary split-ring resonator (CSRR) etched in the ground plane of a PCB, similar to the approach in [89]. The transmission line is used to excite the CSRR structure that acts as an electrically-small resonator. Figure 4.31 shows the sensor's layout, where the red line is

the transmission line (copper) on top of a dielectric substrate, and the olive green region is the ground plane (copper) on the bottom of the PCB. The sensing region is comprised of two co-centered split square rings, which form an inductor-capacitor (L-C) structure that resonates at a certain frequency in air (no MUT loaded). If the sensor is loaded with dielectric MUT, the resonance frequency changes based on the dielectric constant of the MUT . Furthermore, if there is a change in the dielectric constant of the MUT due to anomaly, then the sensors' resonance frequency shifts. The sensor's transmission coefficient magnitude responses for anomaly-free and anomalous regions in a dielectric material are shown in Figure 4.32 for a MUT with dielectric constant of 2.33 and an anomalous region with dielectric constant of 10.2 . It can be noticed that there are two distinctive behaviors, one belonging to an anomalous scanned region (the red line with circular markers in the figure) and the other to an anomaly-free region (the blue line with square markers). The sensor's scattering information was recorded by scanning the frequency over the range of 3-4GHz with 5MHz increments.

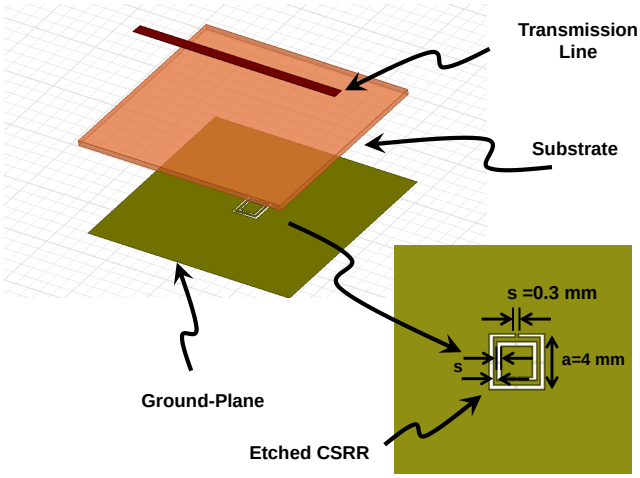


Figure 4.31: Schematic for the sensor's layout: the two co-centered split square rings in ground plane act as electrically small resonator.

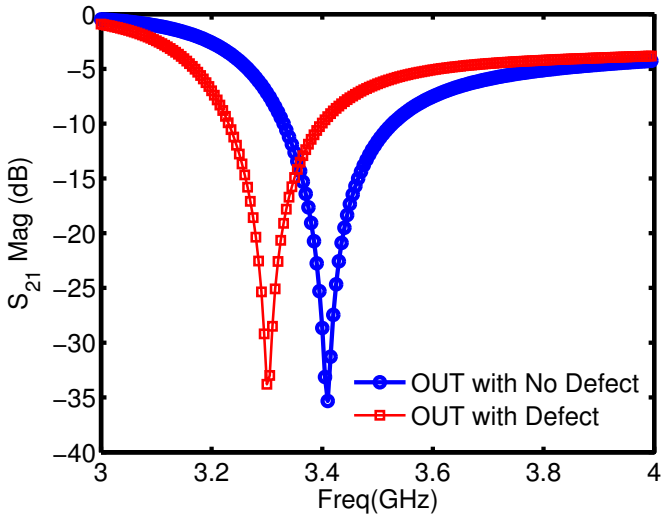


Figure 4.32: Transmission coefficient magnitude plots for anomaly-free and anomalous regions in a dielectric structure

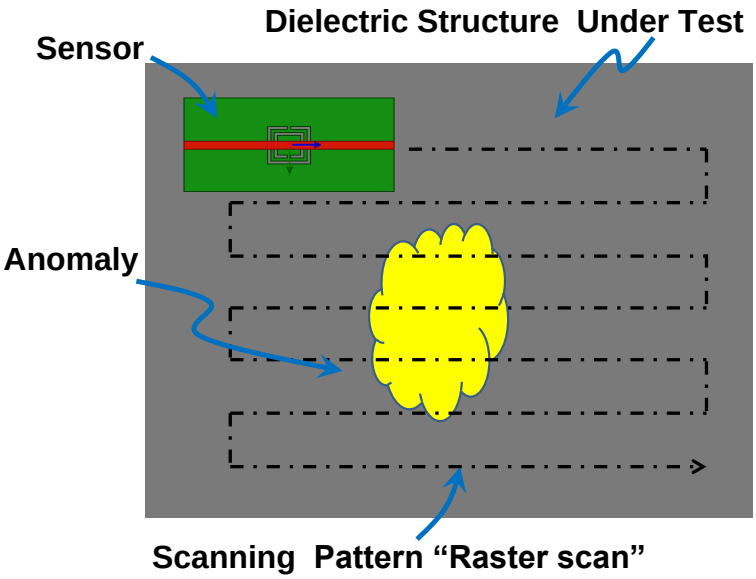


Figure 4.33: Raster scan around an anomalous region in a dielectric structure

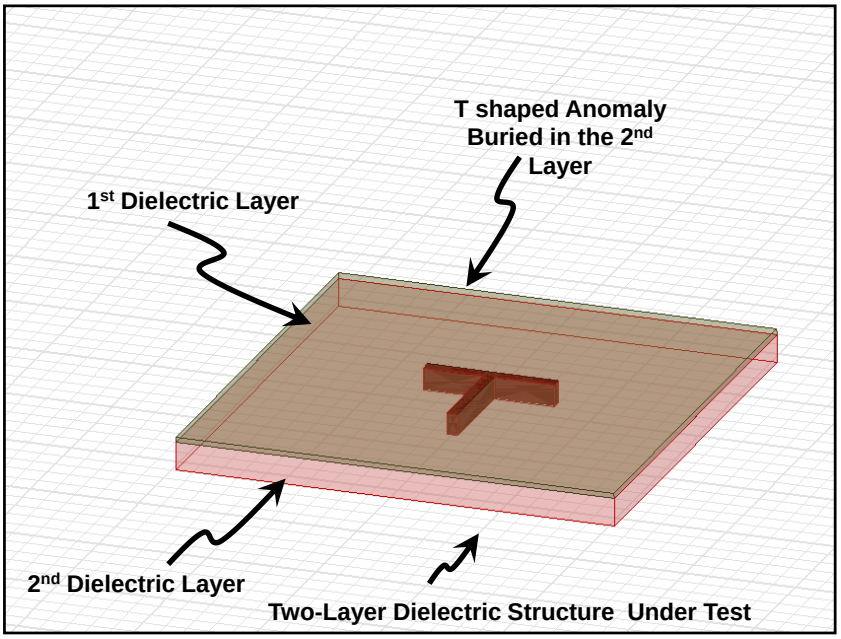
Image reconstruction for the MUT in this study is based on implementing the sensor in a raster scan around the regions under test, as shown in Figure 4.33. The resultant scattering matrix was later transferred to a Matlab environment for post-processing.

Studied Modalities

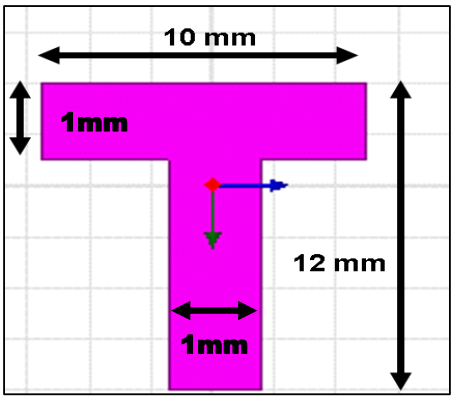
To study the ability of the N-F microwave CSRR-sensor for dielectric structures imaging, two dielectric structures with buried anomalies were studied. The goal of the study is to reconstruct images to reveal the buried anomalies. The first simulated MUT was a two layer dielectric structure. The second simulated MUT was a multi-layer structure built using three dielectric layers of different dielectric constants. The goal of this study is to evaluate the sensor's lateral resolution through the anomaly shape characterization. Simulation results presented in this work were obtained using ANSYS® HFSS™ [93].

Two Layer Dielectric Structure with Buried Anomalies

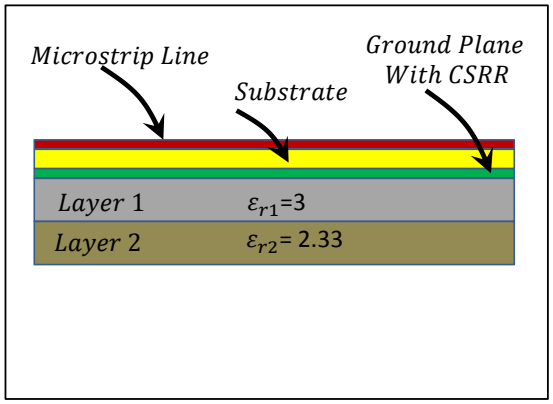
The structure under test is shown in Figure 4.34(a). The first layer is a dielectric plate of 40 mm × 40 mm and a thickness of 0.3 mm with a dielectric constant of 3 covering a second dielectric layer of 40 mm × 40mm and a thickness of 3 mm with a dielectric constant of 2.33. A T-shaped anomaly with a dielectric constant of 10.2 is buried in the second layer of the two-layer dielectric MUT, as depicted in Figure 4.34(b). Figure 4.34 (c) shows the structure front view and how the sensor is used to scan the MUT. The sensor scanned the region around the T-shaped anomaly covering an area of 25 mm × 25 mm, in steps of 1 mm to construct an image of 625 pixels. A reconstructed image using the magnitude of $|S_{21}|$ for the MUT is plotted in Figure 4.35. The red region in Figure 4.35 shows the shape of the anomaly, whereas the blue region corresponds to the rest of the healthy scanned area.



(a)



(b)



(c)

Figure 4.34: Subsurface anomaly imaging in dielectric structure (a) Two-layer dielectric structure under test with T shaped anomaly buried in the second layer. (b) Front view of structure under test. (c) T-shaped anomaly dimensions.

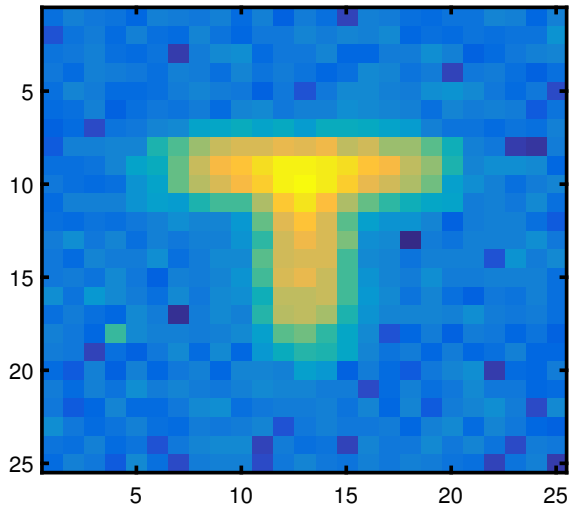
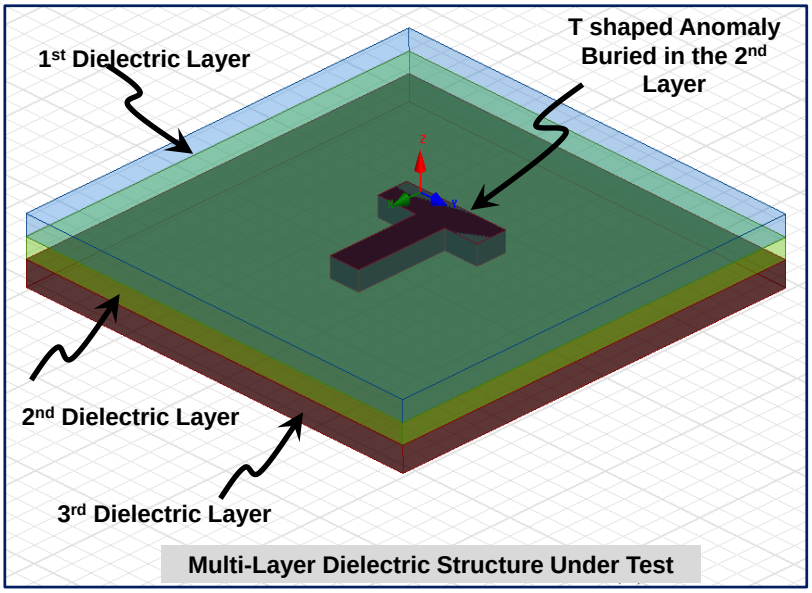


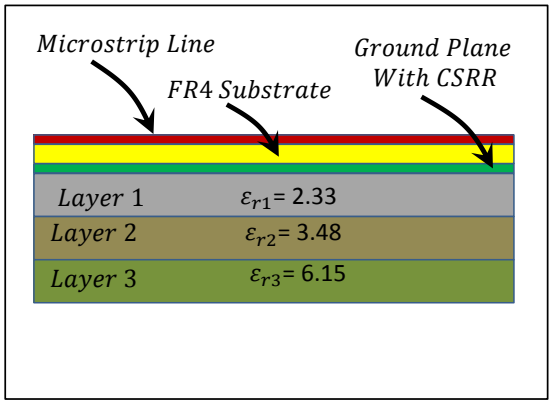
Figure 4.35: Reconstructed image using transmission coefficient magnitude for a two-layer dielectric structure under test with T shaped anomaly buried in the second layer

Multi-Layer Dielectric Structure with Buried Anomalies

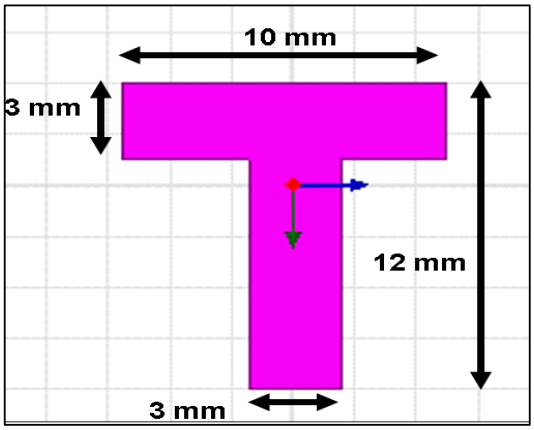
The second simulated MUT was a multi-layer structure consisting of three different dielectric layers with a size of 40 mm × 40 mm each and a thickness of 1.5 mm for each of the first two layers and 2 mm for the third layer. The top and front views of the structure are illustrated in Figure 4.36(a) and (b), where a T-shaped anomaly with a dielectric constant of 10.2 is buried in the second layer. Figure 4.36(b) shows the dimensions of the buried anomaly. The sensor scanned the region around the T-shaped anomaly covering an area of 25 mm × 25 mm, in steps of 1.5 mm, to construct an image of 289 pixels. A reconstructed image using the magnitude of $|S_{21}|$ for the MUT is plotted in Figure 4.37. The yellow region in Figure 4.37 shows the shape of the anomaly, whereas the blue region corresponds to the rest of the healthy scanned area.



(a)



(b)



(c)

Figure 4.36: Subsurface anomaly imaging in dielectric structure (a) Multi-layer dielectric structure under test with T shaped anomaly buried in the second layer. (b) Front view of structure under test. (c) T-shaped anomaly dimensions.

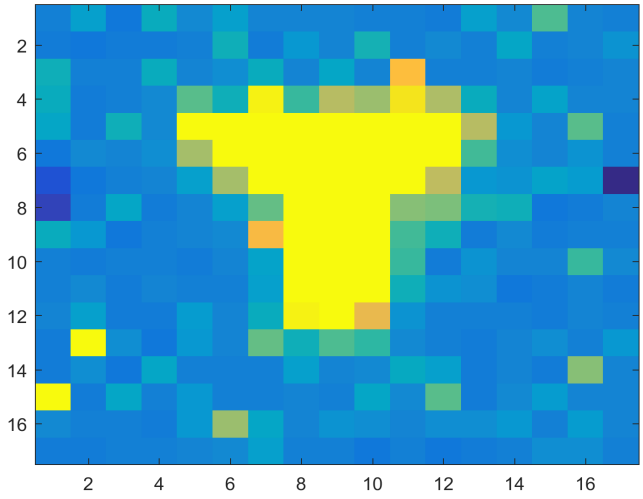


Figure 4.37: Reconstructed image using transmission coefficient magnitude for a multi-layer dielectric structure under test with T-shaped anomaly buried in the second layer

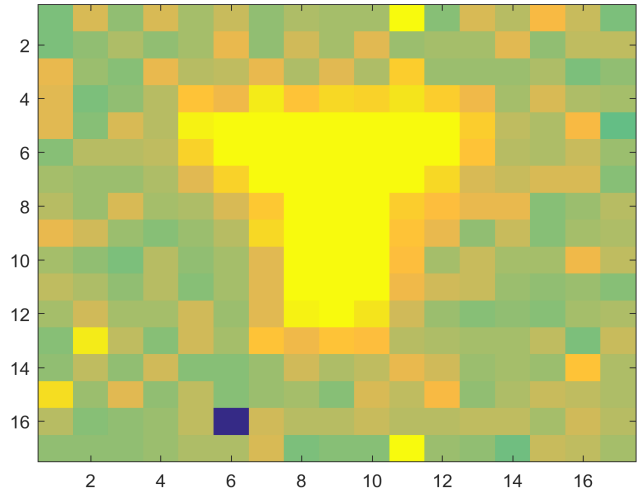


Figure 4.38: Reconstructed image using transmission coefficient phase for a multi-layer dielectric structure under test with T- shaped anomaly buried in the second layer

Conclusion

The presented case study has demonstrated numerically the ability of a microwave sensor based on electrically-small ring resonators to reveal buried anomalies in two-layer and multi-layer dielectric structures. The resultant images have revealed the buried anoma-

lies with high lateral resolution at frequency of 3.5 GHz, which is significantly lower the frequency needed with other near field microwave sensing techniques based on waveguides and horn antennas [88, 98]. The buried anomalies' shapes were clearly captured using magnitude and phase informations of the transmission coefficient. The Outcomes of this study are applicable to composite materials testing and evaluation.

4.6 Discussion

In this chapter, a PCB transmission line loaded with a single metamaterial particle etched in the ground plane has been used to reveal anomalies coated metallics as well as multilayer dielectric structures. The transmission coefficient magnitude and phase information have revealed images with high resolution.

Compared to microwave waveguide imaging [15, 88], microwave CSRR based imaging has the advantage of operating at a low frequency with enhanced resolution utilizing evanescent waves. The proposed system operates in the S-band to detect an area of coated corrosion of 7.5 mm $\times$ 15 mm, while waveguide imaging methods operate in higher frequency bands, such as the K band to visualize an area of coated corrosion of 30 mm $\times$ 30 mm [88, 25]. In case study 4 an ML classifier was trained to classify data obtained from the CSRR sensors and so build binary images for the MUT. For the images reconstructed using the classification results, all swept frequencies contributed to the classification decision for each pixel. The ML implementation based on the SVM classifier has showed significant tolerance for changes in the stand-off distance without compromising the classification accuracy rate. Furthermore, buried anomalies in dielectrics were clearly captured using magnitude and phase informations of the transmission coefficient of the proposed sensor.

Outcomes of these studies are applicable to testing and evaluation of coated or painted metallic structures and dielectric structures such as composite materials. Future work might take the direction of correlating the sensors response to the physical flaw depths using curve fitting or regression models.

Chapter 5

Conclusion and Future work

This research has introduced a novel approach to near-field microwave NDT by integrating machine learning techniques and image construction. Currently, near-field (N-F) microwave NDT depends heavily on the conventional vector network analyzers to plot collected signals. Additionally, trained technicians are needed to observe plotted signals and to make decisions about the MUT health condition. Furthermore, building microwave images from collected datasets is not feasible based on human processing.

The integration of ML with N-F microwave sensors presented in this thesis resolves the above mentioned open issues in N-F microwave testing. Conducted case studies results have shown that ML integration enhances sensing sensitivity for detecting small surface and subsurface anomalies in metallic and dielectric structures. The machine learning modalities developed eliminate the need for signal plotting screens, making the trained models suitable for hand-held *in situ* testing and automated NDT, which save time and effort. Furthermore, machine learning feature selection has been utilized to select top important frequencies for performing NTD testing at the selected frequencies instead of operating at a large range of frequencies, thereby reducing the hardware circuity, which leads to lower costs.

Construction of microwave images for dielectric and metallic MUTs using single metamaterial particle microwave sensor was achieved with higher lateral resolution and lower frequency compared to other available sensors. The images have depicted the buried anomalies clearly due to the small footprint of the sensing element of the CSRR sensor when compared to waveguide sensors. Pixel classification was applied in case study 4 to build binary images based on trained ML models. Images of MUT with higher than 96% accuracy rates were obtained at different stand-off distances. The results from the conducted case studies have shown that the SVM classifier outperformed the NN and RF classifiers due to its unique feature of maximizing the separation margin between classes.

Future Work

The ideas studied in this thesis can be improved further along following directions:

- Implementing simulation modalities and experimental studies to use machine learning regression for characterizing anatomies and flaws.
- Ingratiating machine learning with different microwave sensor types to build a system based on sensors fusion in order to collect uncorrelated signals and features.
- Implementing advanced methods of digital signal processing for constructed images to achieve:
  - Detecting anomalies using image classification methods.
  - Anomalous regions characterization using edge detection techniques.

References

[1] M. S. Boybay and O. M. Ramahi, "Near field probes using double and single negative media," *NATO Adv. Res. Workshop*, pp. 725–731, 2008.

[2] M. S. Boybay and O. M. Ramahi, "Near-field probes using double and single negative media," *Physical Review E*, vol. 79, no. 1, 2009.

[3] M. S. Boybay, "Sensitivity Enhancement of Near Field Probes Using Negative Materials," Ph.D. dissertation, 2009.

[4] M. Puentes, C. Weiss, M. Schüssler, and R. Jakoby, "Sensor array based on split ring resonators for analysis of organic tissues." IEEE, jun 2011, pp. 1–4.

[5] B. Hu, Z. Ren, M. S. Boybay, and O. M. Ramahi, "Waveguide Probe Loaded With Split-Ring Resonators for Crack Detection in Metallic Surfaces," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 62, no. 4, pp. 871–878, apr 2014.

[6] T. Yun and S. Lim, "High-q and miniaturized complementary split ring resonator-loaded substrate integrated waveguide microwave sensor for crack detection in metallic materials," *Sensors and Actuators A: Physical*, vol. 214, pp. 25–30, 2014.

[7] D. Byrne, M. O'Halloran, M. Glavin, and E. Jones, "Breast cancer detection based on differential ultrawideband microwave radar," *Progress In Electromagnetics Research M*, vol. 20, pp. 231–242, 2011.

[8] D. Byrne, M. O'Halloran, E. Jones, and M. Glavin, "Support vector machine-based ultrawideband breast cancer detection system," *Journal of Electromagnetic Waves and Applications*, vol. 25, no. 13, pp. 1807–1816, 2011.

[9] S. A. AlShehri and S. Khatun, "UWB Imaging for Breastr Cancer Detection Using Neural Network," *Progress In Electromagnetics Research C*, vol. 7, pp. 79–93, 2009.

[10] A. Hasan and A. F. Peterson, "Measurement of Complex Permittivity using Artificial Neural Networks," *IEEE Antennas and Propagation Magazine*, vol. 53, no. 1, pp. 200–203, feb 2011.

[11] A. Ali, B. Hu, and O. Ramahi, "Intelligent detection of cracks in metallic surfaces using a waveguide sensor loaded with metamaterial elements," *Sensors (Switzerland)*, vol. 15, no. 5, pp. 11 402–11 416, 2015.

[12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2001.

[13] S. Kharkovsky and R. Zoughi, "Microwave and millimeter wave nondestructive testing and evaluation-Overview and recent advances," *Instrumentation & Measurement Magazine*, 2007.

[14] R. Zoughi and S. Kharkovsky, "Microwave and millimetre wave sensors for crack detection," *Fatigue & Fracture of Engineering Materials & Structures*, vol. 31, no. 8, pp. 695–713, 2008.

[15] H. Zhang, B. Gao, G. Y. Tian, W. L. Woo, and L. Bai, "Metal defects sizing and detection under thick coating using microwave NDT," *NDT & E International*, vol. 60, pp. 52–61, dec 2013.

[16] A. McClanahan, S. Kharkovsky, A. R. Maxon, R. Zoughi, and D. D. Palmer, "Depth Evaluation of Shallow Surface Cracks in Metals Using Rectangular Waveguides at Millimeter-Wave Frequencies," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 6, pp. 1693–1704, jun 2010.

[17] Testing and Applied Microwave Nondestructive lab, "Microwave and Millimeter Wave NDT & E Principles, Methods and Applications."

[18] V. I. Ivanov, "Acoustic emission: some problems, tasks and solutions," *NDT international*, vol. 17, no. 6, pp. 323–328, 1984.

[19] B. Helifa, A. Oulhadj, A. Benbelghit, I. K. Lefkaier, F. Boubenider, and D. Boutassouna, "Detection and measurement of surface cracks in ferromagnetic materials using eddy current testing," *NDT & E International*, vol. 39, no. 5, pp. 384–390, 2006.

[20] L. Bai, G. Y. Tian, A. Simm, S. Tian, and Y. Cheng, "Fast crack profile reconstruction using pulsed eddy current signals," *NDT & E International*, vol. 54, pp. 37–44, 2013.

[21] B. W. Drinkwater and P. D. Wilcox, "Ultrasonic arrays for non-destructive evaluation: A review," *Ndt & E International*, vol. 39, no. 7, pp. 525–541, 2006.

[22] Y. Li, J. Wilson, and G. Y. Tian, "Experiment and simulation study of 3D magnetic field sensing for magnetic flux leakage defect characterisation," *NDT & E International*, vol. 40, no. 2, pp. 179–184, 2007.

[23] M. Sayar, D. Seo, and K. Ogawa, "Non-destructive microwave detection of layer thickness in degraded thermal barrier coatings using K-and W-band frequency range," *NDT & E International*, vol. 42, no. 5, pp. 398–403, 2009.

[24] C. Huber, H. Abiri, S. I. Ganchev, and R. Zoughi, "Modeling of surface hairline-crack detection in metals under coatings using an open-ended rectangular waveguide," *IEEE Transactions on Microwave Theory and Techniques*, vol. 45, no. 11, pp. 2049–2057, nov 1997.

[25] N. N. Qaddoumi, W. M. Saleh, and M. Abou-Khousa, "Innovative Near-Field Microwave Nondestructive Testing of Corroded Metallic Structures Utilizing Open-Ended Rectangular Waveguide Probes," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 5, pp. 1961–1966, oct 2007.

[26] T. Chen, S. Li, and H. Sun, "Metamaterials Application in Sensing," *Sensors*, vol. 12, no. 3, pp. 2742–2765, feb 2012.

[27] V. G. Veselago, "The electrodynamics of substances with simultaneously negative values of Epsilon and Mu," *Soviet physics uspekhi*, vol. 10, no. 4, p. 509, 1968.

[28] J. B. Pendry, A. J. Holden, D. J. Robbins, and W. J. Stewart, "Magnetism from conductors and enhanced nonlinear phenomena," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 11, pp. 2075–2084, nov 1999.

[29] R. A. Shelby, D. R. Smith, and S. Schultz, "Experimental verification of a negative index of refraction," *science*, vol. 292, no. 5514, pp. 77–79, 2001.

[30] J. B. Pendry, "Negative refraction makes a perfect lens," *Physical review letters*, vol. 85, no. 18, p. 3966, 2000.

[31] D. Schurig, J. J. Mock, B. J. Justice, S. A. Cummer, J. B. Pendry, A. F. Starr, and D. R. Smith, "Metamaterial electromagnetic cloak at microwave frequencies," *Science*, no. 5801, pp. 977–980.

[32] J. Pendry, "Photonics: Metamaterials in the sunshine," *Nature Materials*, vol. 5, no. 8, pp. 599–600, 2006.

[33] J. Valentine, S. Zhang, T. Zentgraf, E. Ulin-Avila, D. A. Genov, G. Bartal, and X. Zhang, "Three-dimensional optical metamaterial with a negative refractive index," *Nature*, vol. 455, no. 7211, pp. 376–379, sep 2008.

[34] R. Liu, A. Degiron, J. J. Mock, and D. R. Smith, "Negative index material composed of electric and magnetic resonators," *Applied Physics Letters*, vol. 90, no. 26, p. 263504, 2007.

[35] Z. Jakšić, S. Vuković, J. Matovic, and D. Tanasković, "Negative Refractive Index Metasurfaces for Enhanced Biosensing," *Materials*, vol. 4, no. 1, pp. 1–36, dec 2010.

[36] A. Alu and N. Engheta, "Dielectric sensing in -near-zero narrow waveguide channels," *Physical Review B*, vol. 78, no. 4, p. 045102, 2008.

[37] D. Shreiber, M. Gupta, and R. Cravey, "Comparative study of 1-D and 2-D metamaterial lens for microwave nondestructive evaluation of dielectric materials," *Sensors and Actuators A: Physical*, vol. 165, no. 2, pp. 256–260, feb 2011.

[38] M. Huang and J. Yang, *Microwave sensor using metamaterials*. INTECH Open Access Publisher, 2011.

[39] C.-S. Lee and C.-L. Yang, "Thickness and permittivity measurement in multi-layered dielectric structures using complementary split-ring resonators," *Sensors Journal, IEEE*, vol. 14, no. 3, pp. 695–700, March 2014.

[40] W. Zhao, A. Bhushan, A. D. Santamaria, M. G. Simon, and C. E. Davis, "Machine Learning: A Crucial Tool for Sensor Design," *Algorithms*, vol. 1, no. 2, pp. 130–152, dec 2008.

[41] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[42] S. Sayad, "Naive Bayesian," 2014-02-09, http://www.saedsayad.com/naive_bayesian.htm.

[43] E. Alpaydin, "Introduction to Machine Learning," 2010.

[44] A. Ali, A. Albasir, and O. M. Ramahi, "Microwave sensor for imaging corrosion under coatings utilizing pattern recognition," in *2016 IEEE International Symposium on Antennas and Propagation (APSURSI)*, June 2016, pp. 951–952.

[45] A. Moomen, A. Ali, and O. M. Ramahi, "Reducing sweeping frequencies in microwave ndt employing machine learning feature selection," *Sensors*, vol. 16, no. 4, 2016.

[46] C. M. Bishop, "Pattern Recognition," *Machine Learning*, 2006.

[47] E. Alpaydin, "Introduction to machine learning," 2014.

[48] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[49] M. S. Boybay and O. M. Ramahi, "Waveguide probes using single negative media," *Microwave and Wireless Components Letters, IEEE*, vol. 19, no. 10, pp. 641–643, Oct 2009.

[50] I. Bulu, H. Caglayan, K. Aydin, and E. Ozbay, "Compact size highly directive antennas based on the srr metamaterial medium," *New Journal of Physics*, vol. 7, no. 1, p. 223, 2005.

[51] N. Engheta, S. R. Nelatury, and A. Hoorfar, "The role of geometry of inclusions in forming metamaterials with negative permittivity and permeability," *Proc. 2002 URSI General Assembly*, 2002.

[52] K. Inamdar, Y. Kosta, and S. Patnaik, "Study of a new metamaterial particle of 'av' shape," in *Emerging Technology Trends in Electronics, Communication and Networking (ET2ECN), 2014 2nd International Conference on*. IEEE, 2014, pp. 1–4.

[53] M. Mutlu, A. E. Akosman, A. E. Serebryannikov, and E. Ozbay, "Asymmetric chiral metamaterial circular polarizer based on four u-shaped split ring resonators," *Optics letters*, vol. 36, no. 9, pp. 1653–1655, 2011.

[54] K. Delac, M. Grgic, and S. Grgic, "A comparative study of pca, ica, and lda," pp. 99–106, 2005.

[55] D. A. Tibaduiza Burgos, L. E. Mujica Delgado, M. Anaya, J. Rodellar Benedé, A. Güemes Gordo *et al.*, "Principal component analysis vs. independent component analysis for damage detection," 2013, http://www.ndt.net/article/ewshm2012/papers/fr1d4.pdf.

[56] I. Guyon, "A scaling law for the validation-set training-set size ratio," *AT&T Bell Laboratories*, pp. 1–11, 1997.

[57] C. Özkan and F. S. Erbek, "The comparison of activation functions for multispectral landsat tm image classification," *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 11, pp. 1225–1234, 2003.

[58] J. Yuan and S. Yu, "Privacy preserving back-propagation neural network learning made practical with cloud computing." *Parallel and Distributed Systems, IEEE Transactions on*, vol. 25, no. 1, pp. 212–221, 2014.

[59] R. Taormina, K.-W. Chau, and R. Sethi, "Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the venice lagoon," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 8, pp. 1670–1676, 2012.

[60] K. Chau, "Application of a pso-based neural network in analysis of outcomes of construction claims," *Automation in Construction*, vol. 16, no. 5, pp. 642–646, 2007.

[61] J.-R. Zhang, J. Zhang, T.-M. Lok, and M. R. Lyu, "A hybrid particle swarm optimization-back-propagation algorithm for feedforward neural network training," *Applied Mathematics and Computation*, vol. 185, no. 2, pp. 1026 – 1037, 2007, special Issue on Intelligent Computing Theory and Methodology.

[62] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[63] J. D. M. Rennie, L. Shih, J. Teevan, and D. R. Karger, "Tackling the poor assumptions of naive bayes text classifiers," in *In Proceedings of the Twentieth International Conference on Machine Learning*, 2003, pp. 616–623.

[64] S. Wang and C. D. Manning, "Baselines and bigrams: Simple, good sentiment and topic classification," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*. Association for Computational Linguistics, Jeju, Korea 8-14 Jul 2012, pp. 90–94.

[65] A. Albishi and O. M. Ramahi, "Detection of surface and subsurface cracks in metallic and non-metallic materials using a complementary split-ring resonator," *Sensors*, vol. 14, no. 10, pp. 19 354–19 370, 2014.

[66] M. A. H. Ansari, A. K. Jha, and M. J. Akhtar, "Design and application of the csrr-based planar sensor for noninvasive measurement of complex permittivity," *IEEE Sensors Journal*, vol. 15, no. 12, pp. 7181–7189, Dec 2015.

[67] H. Wang, T. Khoshgoftaar, K. Gao, and N. Seliya, "High-dimensional software engineering data and feature selection," in *Tools with Artificial Intelligence, 2009. ICTAI '09. 21st International Conference on*, Nov 2009, pp. 83–90.

[68] G. H. John, R. Kohavi, K. Pfleger *et al.*, "Irrelevant features and the subset selection problem," in *Machine Learning: Proceedings of the Eleventh International Conference*, 1994, pp. 121–129.

[69] N. Sánchez-Maroño, A. Alonso-Betanzos, and M. Tombilla-Sanromán, "Filter methods for feature selection–a comparative study," in *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, 2007, pp. 178–187.

[70] L. Rokach, *Data mining with decision trees: theory and applications*. World scientific, 2007.

[71] D. Roobaert, G. Karakoulas, and N. V. Chawla, "Information gain, correlation and support vector machines," in *Feature extraction*. Springer, 2006, pp. 463–470.

[72] A. G. Karegowda, A. Manjunath, and M. Jayaram, "Comparative study of attribute selection using gain ratio and correlation based feature selection," *International Journal of Information Technology and Knowledge Management*, vol. 2, no. 2, pp. 271–277, 2010.

[73] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowledge and information systems*, vol. 34, no. 3, pp. 483–519, 2013.

[74] M. A. Hall, "Correlation-based feature selection for machine learning," *PhD thesis, The University of Waikato*, 1999.

[75] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015.

[76] I. Guyon, J. Makhoul, R. Schwartz, and V. Vapnik, "What size test set gives good error rate estimates?" *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 1, pp. 52–64, 1998.

[77] T. Joachims, *Text categorization with support vector machines: Learning with many relevant features*. Springer, 1998.

[78] M. Feuerman and A. Miller, "The kappa statistic as a function of sensitivity and specificity," *International Journal of Mathematical Education in Science and Technology*, vol. 36, no. 5, pp. 517–527, 2005.

[79] R. Zoughi, *Microwave Non-Destructive Testing and Evaluation Principles*. Springer Science & Business Media, 2000, vol. 4.

[80] M. Ghasr, S. Kharkovsky, R. Zoughi, and R. Austin, "Comparison of near-field millimeter-wave probes for detecting corrosion precursor pitting under paint," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 4, pp. 1497–1504, Aug 2005.

[81] M. S. Boybay and O. M. Ramahi, "Non-Destructive Thickness Measurement Using Quasi-Static Resonators," *IEEE Microwave and Wireless Components Letters*, vol. 23, no. 4, pp. 217–219, apr 2013.

[82] S. Sun, D. J. Pommerenke, J. L. Drewniak, G. Chen, L. Xue, M. Brower, M. Y. Koledintseva *et al.*, "A novel trd - based coaxial cable sensor for crack/strain sensing in reinforced concrete structures," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 8, pp. 2714–2725, Aug 2009.

[83] Z. Ren, M. S. Boybay, and O. M. Ramahi, "Near-field subsurface detection in lossy media using single split resonator probe," in *Wireless Sensing, Local Positioning, and*

RFID, 2009. IMWS 2009. IEEE MTT-S International Microwave Workshop on, Sept 2009.

[84] Y. Deng and X. Liu, "Electromagnetic imaging methods for nondestructive evaluation applications," *Sensors*, vol. 11, no. 12, pp. 11 774–11 808, 2011.

[85] K. Aydin, I. Bulu, and E. Ozbay, "Subwavelength resolution with a negative-index metamaterial superlens," *Applied physics letters*, vol. 90, no. 25, p. 254102, 2007.

[86] A. K. Iyer and G. V. Eleftheriades, "Free-space imaging beyond the diffraction limit using a veselago-pendry transmission-line metamaterial superlens," *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 6, pp. 1720–1727, June 2009.

[87] M. Tabib-Azar, P. Pathak, G. Ponchak, and S. LeClair, "Nondestructive superresolution imaging of defects and nonuniformities in metals, semiconductors, dielectrics, composites, and plants using evanescent microwaves," *Review of Scientific Instruments*, vol. 70, no. 6, pp. 2783–2792, 1999.

[88] H. Zhang, Y. He, B. Gao, G. Y. Tian, L. Xu, and R. Wu, "Evaluation of atmospheric corrosion on coated steel using k -band sweep frequency microwave imaging," *IEEE Sensors Journal*, vol. 16, no. 9, pp. 3025–3033, May 2016.

[89] A. M. Ali and O. M. Ramahi, "Microwave imaging of subsurface defects in coated metallic structures using small ring resonators," *NDT in Canada 2015 Conference, Edmonton, 16-17 Jul 2015*.

[90] M. S. Boybay and O. M. Ramahi, "Material characterization using complementary split-ring resonators," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 11, pp. 3039–3046, Nov 2012.

[91] M. P. Vargas, *Planar Metamaterial Based Microwave Sensor Arrays for Biomedical Analysis and Treatment*. Springer Science & Business Media, 2014.

[92] F. Falcone, T. Lopetegi, M. A. G. Laso, J. D. Baena, J. Bonache, M. Beruete, R. Marqués, F. Martín, and M. Sorolla, "Babinet principle applied to the design of metasurfaces and metamaterials," *Phys. Rev. Lett.*, vol. 93, p. 197401, Nov 2004.

[93] *ANSYS HFSS Version 15.0.0, http://www.ansys.com.*

[94] R. W. Revie and H. H. Uhlig, *Corrosion and Corrosion Control: An Introduction to Corrosion Science and Engineering.* John Wiley & Sons, Inc., 2008.

[95] U. DOT and F. FHWA, "Status of the nations highway bridges, and transit: Conditions & performance," in *Report to Congress*, 2006.

[96] N. A. of Corrosion Engineers, *Study on Corrosion Cost and Preventive Strategies*, 2013.

[97] A. Ali, M. E. Badawe, and O. M. Ramahi, "Microwave imaging of subsurface flaws in coated metallic structures using complementary split-ring resonators," *IEEE Sensors Journal*, vol. 16, no. 18, pp. 6890–6898, Sept 2016.

[98] S.-H. Yang, K.-B. Kim, H. G. Oh, and J.-S. Kang, "Non-contact detection of impact damage in cfrp composites using millimeter-wave reflection and considering carbon fiber direction," *NDT & E International*, vol. 57, pp. 45 – 51, 2013.