Politecnico di Torino

Master degree course in CCNE Communication and Computer Network Engineering

Master Degree Thesis

Investigation on optimization of some implanted antennas with machine learning

Supervisor

Prof. Ladislau Matekovits.

Candidate

Pouria Pazoki

Student Number: 249598

Accademic year 2021-2022

Summary

Development in implantable electronic devices have been made a considerable progress in healthcare and biomedical technologies. Improving the healthcare quality by the possibility of continuously monitoring the physiological signals of different parts of human body, providing the support to an organ and drug delivery to name a few, in addition to the continuous miniaturization of electrical sensors, advancement in wireless communication and digital signal processing, has entered biomedical technology into the new era. The main part of the implantable devices which provides the wireless communication is the antenna.

The performance of these antennas inside the body has faced with multiple challenges such as being in the lossy medium from radio propagation perspective, their size and material which should be compatible physically and chemically inside the human body.

This thesis is focused on designing and optimizing the implanted antennas with taking advantage of Machine Learning techniques.

To achieve this goal two different methods of machine learning which are regression and deep neural networks, known as DNNs, have been employed. Both methods are recommended as learning method for the optimization but for some challenges like overfitting and accuracy both have been considered. The reason of this decision is also discussed in the thesis.

Acknowledgments

I am very grateful to acknowledge and give very special thanks to my supervisor Professor Ladislau Matekovits to his support, guidance and supervision throughout my master's degree thesis.

My family deserves endless gratitude: my parents for their unconditional, unequivocal, and loving support. My Sister Dorsa and my brother Kasra for adding laugh and hope to my life and my wife Shahla who endured this long process with me, always offering support and love.

Contents

List of tables 8

List of figures 9

I. Comprehensive survey of implanted antennas 12

1.1 Introduction 14

1.2 Research Motivation 15

1.3 Thesis Outline 16

II. Literature Review of Antennas in Lossy media: 18

2.1 Introduction 19

2.2 The effect of human body on the performance of implanted antenna 19

2.3 Human body model 19

2.4 Design of the single antenna 21

2.4.1 Size of antenna 21

2.4.2 Frequency of design 21

2.4.3 Biological characteristics of the considered dielectrics 22

2.4.4 Antenna radiation performance 22

III. Machine learning 24

3.1 Introduction 25

3.1.1 Basics of machine learning 25

3.2 Regression Models 27

3.2.1 Linear Regression 27

3.2.2 Shrinkage methods 28

3.2.3 Regression Models Beyond Linearity 29

3.3 Classification 30

3.3.1 Logistic Regression 30

3.3.2 Linear Discriminant Analysis 31

3.3.3 Decision Trees 32

3.3.4 Support Vector Machines 34

3.4 Deep Learning 34

IV. Design and Optimization **37**

4.1 Introduction 38

4.2 Implanted Microstrip Patch Antenna Design 38

4.2.1 Design and optimization of a simple microstrip patch antenna 38

4.2.2 Design and Optimization of trapezoidal patch antenna 42

4.2.3 Design and Optimization of Cylindrical patch antenna 48

4.3 Conclusion 49

References **51**

List of tables

Table 2.1 Electrical data of bone 14

Table 2.2 Electrical data of fat, muscle, skin 14

Table 4.1 Materials and tissues parameters at frequency 2.45 GHz 47

List of figures

1.1 Overview of wearable and implantable devices for cardio vascular health management systems 14
1.2 Capsule endoscopy 15
1.3 data transmission from an implanted antenna inside the body 15
2.1 The body model 20
2.2 Geometry and dimensions [mm] of the optimized design 21
2.3 Matching Performance and 3D radiation pattern of the final structure considering a free space environment. 21
3.1 Deep Feedforward Neural Network with 3 layers 34

4.1 2D view of first antenna 38
4.2 3D view of first antenna 38
4.3 Simulated S11 of the patch antenna 39
4.4 Some examples of inputs for antenna dimension 40
4.5 Comparison of the all output with prediction 41

4.6 one layer geometry for trapezoidal patch antenna from two different views 42
4.7 Simulated S11 of the patch antenna 43
4.8 configuration of the network 43
4.9 properties of the network 44
4.10 Performance of the network 44
4.11 Neural network training state 44
4.12 Neural network error histogram 44
4.13 Neural network training regression 44
4.14 Optimised output of Neural Network at frequency 2.4 GHz 45
4.15 One of the simulated S11 at frequency 2.45 GHz 46
4.16 Optimised S11 by Neural Network at frequency 2.45 GHz 46
4.17 Z smith chart of optimised antenna at 2.45 GHz 46
4.18 Y smith chart of optimised antenna at 2.45 GHz 46
4.19 Multi-layer model considering the body tissues in trapezoidal patch antenna design 47
4.20 different views of Cylindrical patch antenna 48
4.21 S11 of the optimized model 48
4.22 Z smith chart of optimised antenna at 2.45 GHz 49
4.23 Y smith chart of optimised antenna at 2.45 GHz 49

I. Comprehensive survey of implanted antennas

1.1 Introduction

Recent developments in implantable electronics technology have created a unique opportunity to improve diagnostic and therapeutic procedures in medical practice. Potential applicability of biocompatible materials to clinical practices, including physiological sensing, drug delivery, cardiovascular monitoring, and brain stimulation has made a noteworthy impact on biomedical applications [1]. Implantable electronic devices with providing the possibilities like continuous monitoring some disease development, reduced the risk of some disease complications and healthcare cost in some cases, and increased the healthcare quality [2].

Medical telemetry which is known as biotelemetry takes advantage of telemetry in medicine and healthcare. Some capabilities like remote monitoring of various vital signals in ambulatory patients, real-time transmission of biological signals from a patient and continuously collect the bio-logical data, can strongly reduce hospitalization time, and consequently associated costs, with-out compromising the quality of the service [3].

In some cases, the required signal can be collected from on-body sensors, but sometimes the information should be collected from inside the body and transmitted outward. This type of information can be achieved from implanted antennas in location inside the body depending on the given telemetry application [3]. Figure 1.1 shows various examples of on-body sensors and implanted antennas inside the body.

There are some other implantable devices called “Capsule Endoscopy” that can be used for diagnosing digestive system disorders by taking the photographic images from the digestive system and transmitting to the outside receivers [5], as shown in figure 1.2.

Implanted antennas are the basic elements of any of such system, where communication between an implanted device in the human body with the outside receivers is required [3], as shown in figure 1.3.

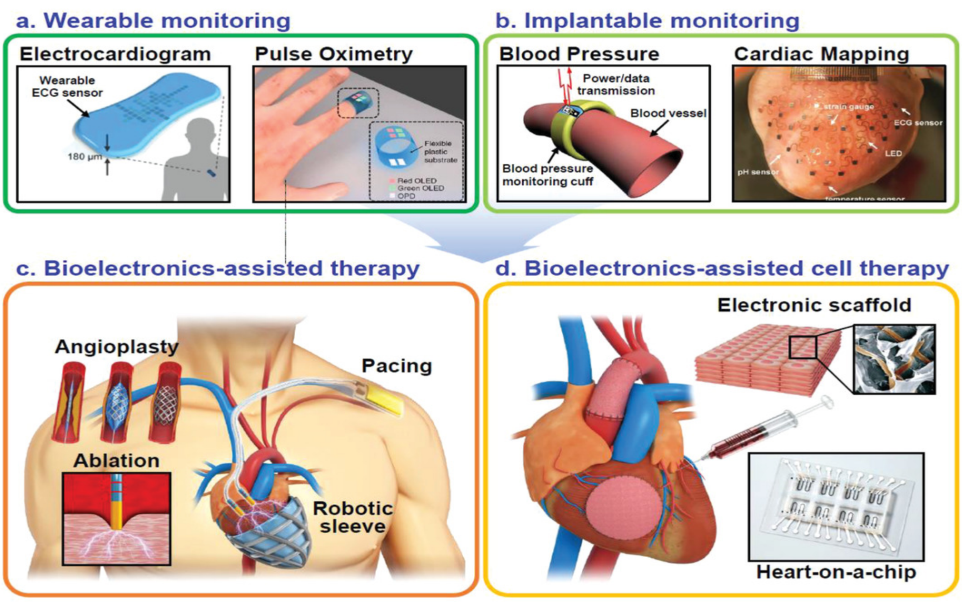


Figure 1.1: Overview of wearable and implantable devices for cardiovascular health management systems [4]

Transferring the physiological information such as blood pressure, cardiac beat, hyperthermia and so on, via biomedical wireless communication, are subject to two fundamental factors: time and precision. It should provide highly precise diagnostic information for prevention and/or timely treatment of diseases. The performance of the implanted antennas under the skin or on the body in terms of the validity of the transmitted information and its accuracy plays a vital role in the quality of the diagnosis and treatment [6].

Combination of wireless technology with bioelectronics has provided noninvasive control, biotelemetry, and wireless power transfer (WPT). The main features of antennas that should be considered are their radiation characteristics, materials, integration with rest of the electronics, and experimental setup [1].

The design of implanted antennas is subject to additional challenges that reflect both in the link budget, due to high loss, and corrosion in time, due to the harsh environment surrounding the transmitter [6]. Wireless implantable bioelectronics and accurate in vivo characterization of path loss require efficient and robust in-body antennas [7].

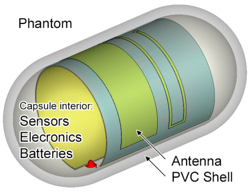


Figure 1.2: Capsule endoscopy [8]

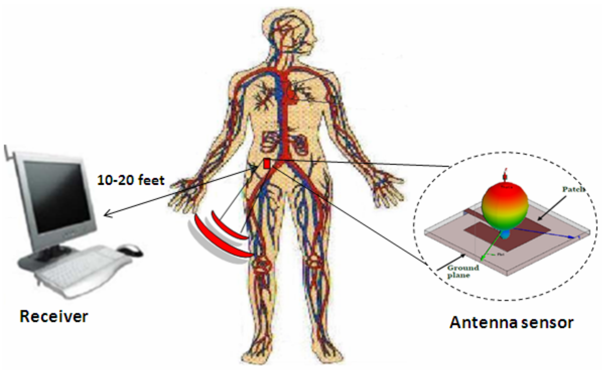


Figure 1.3: Data transmission from an implanted antenna inside the body [9]

1.2 Research Motivation

Implanted antennas play a pivotal role in medicine and healthcare by the capability of continuously and remote monitoring the patients' situation with capturing the interest signals from the body. Although, still there is a long way to achieve an ideally high performance antennas which transmit the signals with the high quality while they have been positioned inside the body which arises the additional challenges in these antennas' functionality.

This thesis is focused on design and optimization problem of the implanted antennas on bone structures by considering the dimensions in width, length and depth of antennas, different possible materials in designing the implanted antennas and the position of coaxial cable in them. Microstrip implanted antennas has been designed in CST Studio Suite. To solve the optimization problem, different regression methods of machine learning has been employed.

1.3 Thesis Outline

This thesis is organized in 4 chapter. Following this introductory, the rest of this thesis is organised as follows.
In chapter 2 a literature review of the micro strip patch antenna design and challenges has been represented. Chapter 3 is a review on Machine Learning methods with focus on its application in optimisation problems.
Three different antennas has been designed and reported in chapter 4. In this chapter for each design the optimization methods have been applied to achieve the antennas with the best performance suitable to be implanted in human body around the bone.

II. Literature Review of Antennas in Lossy media:

2.1 Introduction

In this chapter some details of the implantable antennas performance and the challenges in designing these antennas and the impact of some important factors such as antenna's gain, bandwidth, the specific absorption rate (SAR) and size which play a vital role in the performance of implantable antennas, has been discussed.

Using the wireless communication in human body consists a wide area which these days named as Body Sensor Network (BSN), Body Area Network (BAN) or Body-Centric Wireless Communication (BCWC) and this is formally defined by IEEE 802.15 as, "a communication standard optimized for low power devices and operation on, in or around the human body (but not limited to humans) to serve a variety of applications including medical, consumer electronics, personal entertainment and other" [10].

There are numerous applications of implanted devices in the context of biotelemetry, i.e. biomedical wireless communication, to transfer physiological information such as blood pressure, cardiac beat, hyperthermia and so on [6]. Depending of the telemetry application the location of the implanted devices is determined. For instance, for a wireless RF powered brain machine interface application, antennas are situated in brain [11] and [12] or in some application as reported in [13] they can be implanted under the skin.

2.2 The effect of human body on the performance of implanted antenna

One of the main challenges in designing the implantable devices is that some of them should be placed deeply inside the body. These challenges are less severe for the antennas implanted under the skin. In [13] it is shown that the position of the implanted antenna and electrical properties of the insulating layer also affect the RF power reception and the properties of the tissues surrounding the implanted antenna can influence the electrical characteristics of the antennas.

In this thesis the main focus is on the implanted antennas on bone structures. These particular antennas are used to monitor the healing process of the broken bones.

In [14] to monitor the healing of the broken bone, a waveguide antenna with a circular polarisation which provides bio-medical wireless telecommunication is designed.

In order to investigate the functionality and fabrication challenges of the loop antennas, first the interesting part of human body should be modelled and the effect of relative permittivity of the bio-tissues on the implanted antenna resonance should be studied as it is investigated in[15] and [16]

2.3 Human body model

In [17] and [3] a cylindrical body model with the dimensions for an adult's leg has been proposed. The CAD geometry of this study as shown in figure 2.1, includes different tissues of body which are modeled as concentric cylinders. The thicknesses of the skin, fat and muscle are 2 mm, 4 mm, 70 nun, respectively. The radius of the bone is 30 mm. The electrical characteristics of different tissues has been presented in table 2.1 and table 2.2.

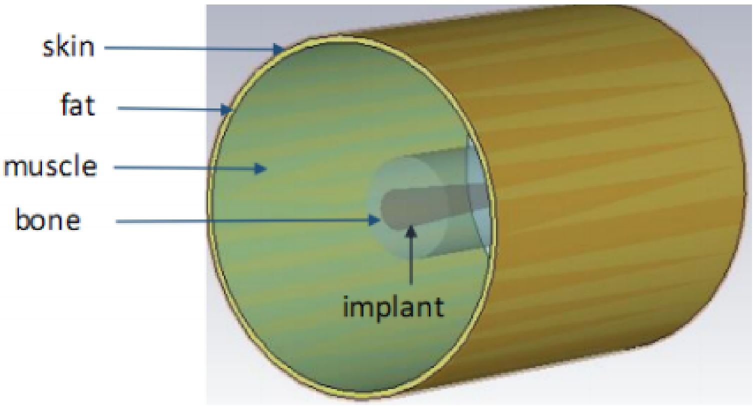


Figure 2.1: The body model [3]

Table 2.1: Electrical data of bone[3]

Biological tissue	Diffusivity (m^2/s)	Mass density (kg/m^3)
Bone	1.70478e-007	1850

Table 2.2: Electrical data of fat, muscle, skin[3]

Biological tissue	Permittivity (ε_r)	Conductivity (σ, S/m)
Fat	5.6	0.04
muscle	57.1	0.79
skin	46.7	0.69

In the model proposed in [17] part of the bone is replaced with a conical implant made up by a biocompatible metal.The implant has a length of L=150 mm and top and bottom radii of rt=12 mm and rb=4 mm. Although, because of the mechanical properties, high corrosion resistance and excellent biocompatibility of Titanium and its alloys, these material are used usually in orthopaedic field, the new researches demonstrate that Ta and Nb are non toxic elements and they reveal very good biocompatibility and high corrosion resistance as well as Ti. Because of these physical and chemical properties, they have been used as constituent elements of Ti alloys for such applications.

2.4 Design of the single antenna

2.4.1 Size of antenna

The size of antennas should be small enough in order to be physically implantable inside the body. On the other hand electromagnetic performance of antennas reduces with the size reduction. Thus, the new approach should be investigated to maintain a combination of small size antenna and an acceptable performance.

Different techniques have been employed to reduce the size of the antennas. One technique proposed in [18] to reduce the implantable antenna size is designing it in a folded or spiral shape . An example of Geometry and dimensions of the optimized spiral design and matching Performance and 3D radiation pattern of the final structure considering a free space environment are shown in figure 2.2 and 2.3. The other technique that helps to reduce the size of antenna is designing it at the higher frequency, for instance in 5.8GHz [19]. In some researches a substrate with a high dielectric constant is used to reduce the size of the antennas, while antenna with a high dielectric constant substrate is susceptible to surface wave excitation, which degrades the radiation pattern of the antenna [20, 21]

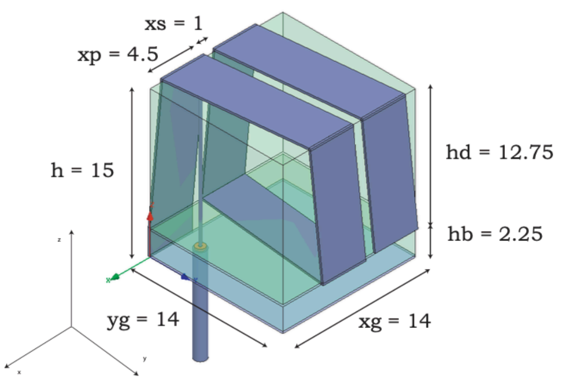


Figure 2.2 Geometry and dimensions [mm] of the optimized design [18]

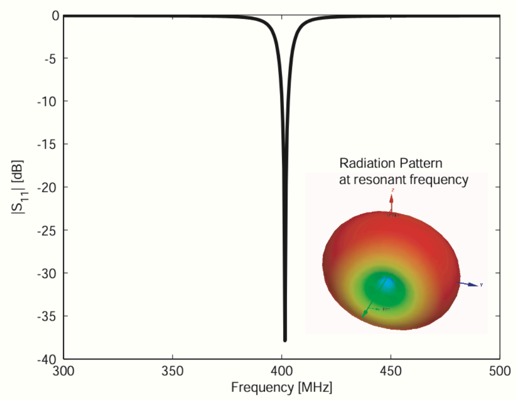


Figure 2.3 Matching Performance and 3D radiation pattern of the final structure considering a free space environment [18]

2.4.2 Frequency of design

The industrial, scientific, and medical radio band (ISM band) (433.1-434.8 MHz, 868-868.6 MHz, 902.8-928 MHz, and 2400-2500 MHz) refers to a group of radio bands or parts of the radio spectrum that are internationally reserved for the use of radio frequency (RF) energy intended for scientific, medical and industrial requirements rather than for communications. ISM bands are generally open frequency bands, which vary according to different regions and permits [22]. The optimised design from the frequency point of view should be at the center of the 2.4 - 2.5 GHz Industrial, Scientific and Medical (ISM) radio band, i.e., at 2.45 GHz.

2.4.3 Biological characteristics of the considered dielectrics

The main requirement for the bio-materials implanted inside the body is a proper mechanical behaviour and biocompatibility with human body. According to the specific applications, different biomaterials including metals and their alloys, ceramics, composites and synthetic or natural polymers, have been used in designing in antennas. One of the significant functionalities of the implanted biomaterials are the substitution of a missing body part and/or to provide support to the organs. To design the implantable devices for each purpose the physical and chemical properties of materials should be considered. For example polymeric materials is appropriate for low friction articulating surfaces because of their flexibility and stability. On the other hand, ceramic and composite types biomaterials are used in articulating surfaces in joints and in teeth as well as bone bonding surfaces in implants [3].

2.4.4 Antenna radiation performance

Due to the high path loss, the implanted antenna should have gain as high as possible in the desired direction to guarantee communication between the antenna and the external devices. In [6], the performances of the two antennas are investigated, focusing on the distribution of the electric field inside the body propagation. The attenuation of the radiated power by the antenna depends on both the length of the path and on the losses of the layers (materials/tissues) the electromagnetic field goes through when it propagates from the source outward the body.

III. Machine learning

3.1 Introduction

Optimization is one of the most important parts of machine learning which is growing because of facing more complex models and algorithms. Machine learning methods has been processed properly to solve the new challenges in case of optimization. In this part of thesis some of the most important and useful methods of machine leaning which are suitable for an optimization problem is clarified.

3.1.1 Basics of machine learning

The idea of changing the people's life by the intelligent machines was raised at in the late 1950s [24] The aim of this idea was based on machines who can learn from the early existence problem and try to solve the more complex version of these situation and give the better response in case of differences of variables. As soon as problems get involved with more and more amount of data ,machine learning methods became more popular between scientists and engineers. Here it has been provided the very first approaches in machine learning.

The first important concept in machine learning is the model which will give the best results as output by considering the certain amount of input. In other words, we will give the fix amount of data to our machine, it learns how to predict the output from input. Then if machine receives a different input, precise results in output are expected.

In this way machine learning is made in two major categories. First, supervised learning, for this kind of learning, there is an exact amount of the output for the specific input and predictions must be based on known sets of outputs and inputs and model must provide best function which is fitted to the sets of outputs and inputs. The accuracy of the model should be considered. There are several techniques to measure the prediction accuracy. The next step after making the model is measurement of the error. Then the distance between the exact amount and predicted amount which is provided by fitted function will be called as error. So based on the techniques of measuring errors and the threshold which confine the model to prevent to reach the risky predictions, the accuracy will be determined and tried to reduce the distance between real value and predicted value. For this measurement the specific loss function is considered:

$$\mathbb{E}_p[\mathcal{L}(f(x), y)] = \int \int p(x, y)\mathcal{L}(f(x), y)dxdy,$$

For this equation $p(x, y)$ is the probability of accuracy on the data (x, y) [25]. It should be considered $p(x, y)$ is not known. So, as far as the assumption is independent and identically distributed sample of data points (x_1, y_1) to (x_n, y_n) can be reduced by f^* :

$$f^* = \arg\ \min\ \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(f(x_i), y_i).$$

When the risk is minimized, there is one important issue in learning model which is called overfitting.

When the model is complex it may lead to a problem for the training sets and really fit to them on the exact value of the data and the model works perfect just on the training data. On the other hand, they are far from test sets . It should be considered that data could be involved with noisy values so it could be a cause of mistake in learning a model, so we can consider $f*$ as following:

$$f^* = \arg \min_{f \in F} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(f(x_i), y_i). \tag{1}$$

It should be considered the degree of complexity is not in our hand because most of the time it is related to the nature, so the training data must be big enough to avoid overfitting.

As it is mentioned before there is another type of machine learning which is unsupervised machine learning. In this kind of learning the responses are not clear so the aim is to find the features and the characteristics of the observations.

In opposite of supervised learning that is every set of data are clear when they have been predicted known response, in unsupervised learning could be hard to determine validation of extracted structure. In this thesis have been focused on the supervised learning because in each steps the sets of inputs and outputs are clear and we extract the data from some patch antenna the are known as a matter of material and size, the main issue is about finding the most optimized design of this implanted micro-strip antenna.

Before introducing the concepts of every kind of supervised machine learning it could be mentioned three aspects of relation between operation research and machine learning:

(a) machine learning will be used to manage the complex science problem (b) it is common to use machine learning for optimization problem (c) machine learning problems formulated as optimization problems.

3.2 Regression Models

3.2.1 Linear Regression

Linear regression is one of the most famous and common ways of supervised learning from the moment of existing the statistics. The main assumption is based on correlation between the correlated variables like feature measurements, predictors or something like input vector and uncorrelated variable such as real valued output and responses that can be defined by linear function (regression function) with acceptable accuracy score. Linear function models have a lot of benefits one of them is simplicity, it is easier to interpret machine learning in an understandable way for the people. In linear regression the goal is to find the best linear function f to represent best output f(x) for the input vector x of dimension p [26]:

$$f(x) = \beta_0 + x^\top \beta, \tag{2}$$

where $\beta_0 \in R$ is the intercept of the regression line and $\beta \in R^p$ is the vector of coefficients [23] corresponding to each of the input variables. Here the estimation of the regression parameters β_0 and β should be found, there should be a training set (x, y) where $x \in R^{n\times p}$ stand for n training inputs $x_1, \ldots, x_n$ and y which stands for n training outputs where each $x_i \in R^p$ is associated with the real-valued output y_i.

The purpose is to decline the empirical risk (1), To quantify via β_j the association between predictor x_j and the response, for every $j = 1,...,p$. The most used loss function for regression is the least squared estimate, where fitting a regression model decreases to minimizing the residual sum of squares (RSS) between the labels and the predicted outputs, such as

$$RSS(\beta) = \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2. \tag{3}$$

The least square estimate has the smallest variance between all of the linear unbiased estimations and it has a closed form solution. But , this selection is not always the best choice from fitting point of view, because it can announce a model with the low prediction accuracy score which it will be ended by a non interpretable model as a lot of non zero coefficients excist. There are other solution to avoid overfitting like Shrinkage methods and Forward/backward elimination[27].

It should be considered most of the time a clear set of data is not received. Information could be noisy and it may lead to the statistical inaccuracy. A model which consider the noise as a feature of linear regression problems is presented in [28], this model also considers the relation between the regularization and robustness to noise .The assumption is that noise will be changed in an uncertain set of U $\in R^{n\times p}$, so the learner should be adopted to the robust perspective:

$$\min_{\beta_0,\beta} \max_{\Delta \in \mathcal{U}} g(y - \beta_0 - (X + \Delta)\beta), \tag{4}$$

Here g is a convex function which calculates the residuals. Uncertainty characterization will affect straightly on complexity of the problem. To achieve a linear regression model with high level of quality, different features and aspects should be considered but most of the time these properties are against each other and they are not compatible together. A fitting procedure based on Mixed Integer Quadratic Programming (MIQP) in [29] is shown which considers sparsity robustness to noisy data, stability against outsider and etc. Mixed Integer Programming (MIP) models for regression which has been discussed in [30] can be observed. There is a problem about regression method which is the modelling of an assignment of data points to groups that have the same regression coefficients. In order to accelerate the fitting and get the better accountability of regression model, some actions such as eliminating the unrelated variables by feature selection techniques can be done. For instance, when some of the regression variables are highly related, it is reasonable to use this strategy of feature selection and keep the most correlated variables and eliminate others. This amount of multicollinearity can be detected by The Variance Influence Factor (VIF) [31]. To get this chosen feature in situations like this case, [32] announced a mixed integer semidefinite programming formulation to exclude multicollinearity by bounding the condition number. So for this attitude it is needed to solve just a optimizition problem, in the case in opposite of algorithm [29], [33] proposes a mixed integer quadratic optimization formulation with an upper bound on VIF. This method is a best valid statistical indicator for multicollinearity by considering the condition number.

3.2.2 Shrinkage methods

This methods also called regularization methods to achieve the value decreasing of the regression coefficients. Here the goal is to get a more accountable model and reach to the model with less relevant features at the cost of having the biased model. A famous shrinkage method is Ridge regression. In this formulated ridge regression 2-norm penalizition on the regression coefficients is added to the loss function like:

$$\mathcal{L}_{ridge}(\beta_0, \beta) = \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2, \tag{5}$$

where λ is the controller of the shrinkage magnitude.

There is also anther way which is called lasso regression where penalizes the 1-norm of the regression coefficient and tries to decreases the amount of following quantity as much as possible:

$$\mathcal{L}_{lasso}(\beta_0, \beta) = \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|. \tag{6}$$

If λ is big enough, The 1-norm punishment enforces some of the coefficients and estimates them equal to zero, when models are provided by the lasso it is easier to interpret then in comparison with Ridge regression. Both of these ways are the bigger class of regression approach which is called Sparse regression [34,29,35,36,30]. The formulated Sparse regression is the best subset of selection problem:

$$\min \frac{1}{2}\|y-\beta_0-X\beta\|_2^2 \tag{7}$$
$$\text{s.t } \|\beta\|_0 \leq k, \tag{8}$$
$$\beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^p, \tag{9}$$

K is an upper bound on the prediction numbers with a non-zero regression coefficient. Based on [35] the recents works show that by using optimization techniques the problem of choosing the best subsets for different values of p among thousands be solved. Specially with considering $s \in \{0,1\}^p$. So this action helps to convert the sparse regression problem to the MIQP like the following:

$$\min \frac{1}{2}\|y-\beta_0-X\beta\|_2^2 \tag{10}$$
$$\text{s.t } -Ms_j \leq \beta_j \leq Ms_j \quad \forall j=1,\dots,p, \tag{11}$$
$$\sum_{j=1}^{p} s_j \leq k, \tag{12}$$
$$\beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^p, \tag{13}$$
$$s \in \{0,1\}^p, \tag{14}$$

Here M is a large constant, $M \geq ||\beta||_\infty$. It is recommended to use Specially Ordered Sets Type I [37] because the crossing subsets are highly related to the amount of constant M and this constant effects on the MIQP.

As it was mentioned the most important part is the best choice of subset. It is visible in [38] the accuracy of the prediction for the best subset choice is correlated to the noise in the input data sets this issue makes it impossible to create an absolute relation between the lass regression and stepwise selection. So to decrease the amount of noise effect on the input data it is recommended to robust the model and avoid numerical issues.

3.2.3 Regression Models Beyond Linearity

After linear regression the next extension must be non-linear regression. In [27,39] non-linear regression models such as polynomial regression, exponential regression, step functions, regression splines, smoothing splines and local regression are presented.

On the other hand GAMs which stands for the Generalized Additive Models keep the additivity of original predictors also they maintain the relationship between every features and y responses by using the non linear function $f_j(X_j)$ like:

$$y = \beta_0 + \sum_{j=1}^{p} f_j(X_j). \tag{15}$$

GAMs may provide some flexibility and accuracy for the predictions with regard to the linear regression when GAMs keep the certain amount of the interpretable predictions. But, there is one limit about it by the additivity assumption of the features. If it is needed to increase the amount of flexibility, $X_i \times X_j$ should be the amount of predictors or take the advantages of non-parametric models , such as random forests and boosting. It is observed when the number of predictor are too much less than observation GAMs do not solve the problem properly[23].

There are also some other techniques which are represent the relation between linear and non linear models like [40] which shows us the purpose of fitting constrained version of GAMs or we can observed mix integer models in [41].

3.3 Classification

In this section one the most important applications of machine learning has been discussed.

Classification will be used when it is needed to determine membership of on data X. Where it is needed to clarify X_i is the member of class Y by considering the training data set off (X,Y). There are several methods of classification such as logistic regression, linear discriminant analysis, decision trees, and support vector machines which has been discussed briefly here.

3.3.1 Logistic Regression

Most of the time there is no functional relations between X and Y, for this issue it is possible to consider the probability distribution function $P(x, y)$ when the training set are independent from P.

There is a binary assumption for y which means Y can be 0 or 1. Logistic regression will determine the probability of membership for the data for two cases:

$$P(y=1|x,\beta_0,\beta) = h(x,\beta_0,\beta) = \frac{1}{1+e^{-(\beta_0+\beta^\top x)}},$$
$$P(y=0|x,\beta_0,\beta) = 1 - h(x,\beta_0,\beta).$$

The decision bound will be divine as $\beta_0 + \beta^\top x = 0$ where the bounds have

$$P(1|x,\beta_0,\beta) = P(0|x,\beta_0,\beta) = 0.5$$

when The parameters β0 and β are often get from maximum-likelihood estimation [42]

$$\max\ \Pi_{i=1}^{n} P(y_i|x_i,\beta_0,\beta) = \max\ \Pi_{i=1}^{n} (h(x_i,\beta_0,\beta))^{y_i}(1-h(x_i,\beta_0,\beta))^{1-y_i},$$

which is equal to

$$\min -\sum_{i=1}^{n}(y_i\log\, h(x_i, \beta_0, \beta) + (1 - y_i)\log(1 - h(x_i, \beta_0, \beta))).$$

There is an observation in problem (16) which says this problem is and differentiable so this makes it possible to apply Newton method which is the second order method and it would be the best case as a matter of global optimal solution.

Overfitting is the most important obstacle in the most of the machine learning methods. So to avoid overfitting and tune the logistic regression, it is reasonable to omit irrelevant subsets of x and just kept the relevant subsets [27]. For this case statistical approach like forward selection or backward elimination can be used. Interaction terms also can be added for the more complex models to prevent overfitting on the training data.

3.3.2 Linear Discriminant Analysis

LDA which stands for Linear Discriminant Analysis is the method of classification. It is useful for the data such as images because these kinds of data have a lot of features and the goal of this method is to reduce dimensions.

The dataset (x,y) is given where each data sample $x_i \in R^p$ belongs to one of K classes such that if xi belongs to the k-th class then $y_i(K)$ is 1 where $y_i \in \{0,1\}^K$, the input data is partitioned into K groups then two new parameters should be consider as S_w and $S_b$where $S_b$stands for within-class scatter and S_w stands for between-class scatter:

$$S_w = \sum_{k=1}^{K} \sum_{x_i \in \pi_k} (x_i - \mu_k)(x_i - \mu_k)^\top, \tag{17}$$

$$S_b = \sum_{k=1}^{K} n_k(\mu_k - \mu)(\mu_k - \mu)^\top. \tag{18}$$

S_w measures the amount of data around class mean but S_b will evaluate the spread around the global mean. So for these data S_b and S_w will be considered as $\overline{S_w}$ and $\overline{S_b}$.

$$\overline{S}_w = \sum_{k=1}^{K} \sum_{q_i \in \pi_k} (q_i - \overline{\mu}_k)(q_i - \overline{\mu}_k)^\top = G^\top S_w G, \tag{19}$$

$$\overline{S}_b = \sum_{k=1}^{K} n_k (\overline{\mu}_k - \overline{\mu})(\overline{\mu}_k - \overline{\mu})^\top = G^\top S_b G. \tag{20}$$

The LDA optimization problem has conflict which is controversy between minimizing the within-class scatter and maximizing the between-class scatter. The optimal transformation G can be obtained by maximizing the Fisher criterion (the ratio of between-class to within-class scatters)

$$\max \frac{|G^T S_b G|}{|G^T S_w G|}. \tag{21}$$

there is also the other way of the formulization of LDA is mentioned in the [43].

3.3.3 Decision Trees

Decision tree is the most classic method for making the decisions. In this method by dividing the rules and organizing them to the tree data structure then method has been reached to the decision structure. Decision tree method is not parametric and this makes the model by the splitting the predictors to the sub-regions and they declare the result of the prediction by considering the statistical indicators like median and mode of the each part of the training data. Decision tree is really useful in case of facing regression and classification problems.

By top-down recursive splitting method dividing the training dataset to distinct and without overlapping area will be easier. It is started from the root(first node) and will include the whole dataset .

[44] proposes a mixed integer programming formulation with an exponential complexity in the depth of the tree to get the best decision tree as optimizarion point of view.there is a constant depth D, and maximum number of nodes is $T = 2^{D+1} - 1$ they are indexed by t = 1, . . . , T

The mixed integer programming formulation is:

$$
\begin{aligned}
\min\ & \frac{1}{\hat{L}} \sum_{t \in T_L} L_t + \alpha \sum_{t \in T_B} d_t \\
\text{s.t. } & L_t \geq N_t - N_{kt} - n(1 - c_{kt}), \quad \forall k = 1, \ldots, K,\ t \in T_L, && (22)\\
& 0 \leq L_t \leq N_t - N_{kt} + n c_{kt} \quad \forall k = 1, \ldots, K,\ t \in T_L, && (23)\\
& N_{kt} = \frac{1}{2} \sum_{i=1}^{n} (1 + Y_{ik}) z_{it}, \quad \forall k = 1, \ldots, K,\ t \in T_L, && (24)\\
& N_t = \sum_{i=1}^{n} z_{it} \quad \forall t \in T_L, && (25)\\
& && (26)\\
& \sum_{k=1}^{K} c_{kt} = l_t \quad \forall t \in T_L, \\
& && (27)\\
& \sum_{t \in T_L} z_{it} = 1 \quad \forall i = 1, \ldots, n, \\
& && (28)\\
& z_{it} \leq l_t \quad \forall i = 1, \ldots, n,\ t \in T_L, \\
& \sum_{i=1}^{n} z_{it} \geq N_{\min} l_t \quad \forall t \in T_L, && (29)\\
& && (30)\\
& a_m^\top (x_i + \epsilon) \leq b_m + (1 + \epsilon_{\max})(1 - z_{it}) \quad \forall i = 1, \ldots, n,\ t \in T_L,\ m \in A_L(t), \\
& && (31)\\
& a_m^\top x_i \geq b_m - (1 - z_{it}) \quad \forall i = 1, \ldots, n,\ t \in T_L,\ \forall m \in A_R(t), \\
& && (32)\\
& \sum_{j=1}^{p} a_{jt} = d_t \quad \forall t \in T_B, \\
& && (33)\\
& 0 \leq b_t \leq d_t \quad \forall t \in T_B, \\
& && (34)\\
& d_t \leq d_{p(t)} \quad \forall t \in T_B \setminus \{1\}, \\
& && (35)\\
& z_{it},\ l_t \in \{0, 1\} \quad \forall i = 1, \ldots, n,\ \forall t \in T_L, \\
& && (36)\\
& c_{kt} \in \{0, 1\} \quad \forall k = 1, \ldots, K,\ t \in T_L, \\
& a_{jt},\ d_t \in \{0, 1\} \quad \forall j = 1, \ldots, p,\ t \in T_B. && (37)\\
& && (38)
\end{aligned}
$$

In [23] the equations has been described according to the following explanations.

$\frac{1}{\hat{L}} \sum_{t \in T_L} L_t$ and $\sum_{t \in T_B} d_t$ in (22) are misclassification and the complexity of the decision tree respectively. α considered as a tuning paradigm where $\hat{L}$ stands for baseline loss obtained by predicting the most popular class between whole dataset. (23)–(24) try to set the misclassification loss L_t at leaf node t as $L_t = N_t - N_{kt}$ if node t is assigned label k (i.e $C_{kt} = 1$), when N_t is the total number of data points at leaf node t and N_{kt} is the total number of data points at node t whose true labels are k [23]. The calculating of $N_k t$ and N_t is imposed by (25) and (26), **sequentially**, where Y_{ik} is a parameter taking the value of 1 if data point i has a label k and −1 otherwise. Restriction on (27) show each leaf node that is used have to be assigned to a label k = 1 . . . K. Constraints (28) determine that each data point should be assigned to exactly one and only one leaf node. Constraints (29)–(30) show which data points can be assigned to a node only if that node is used and if a node is used then at least Nmin data points should be assigned to it. The splitting of the data points at each of the branch nodes is enforced by constraints (31)–(32) where $A_L(t)$ is the set of ancestors of t whose left branch has been followed on the path from the root node to node t. Similarly, $A_R(t)$ is the set of ancestors of t whose right branch has been followed on the path from the root node to node t. and max are small numbers to enforce the strict split $a^\top x$ < b at the left branch (see [28] for finding good values for and max). Constraints (33)–(34) indicate that the splits are restricted to a single variable with the option of not splitting a node (dt = 0). As enforced by constraints (35), if p(t), the parent of node t, does not apply a split then so is node t. Finally constraints (36)–(38) set the binary conditions.

3.3.4 Support Vector Machines

Another type of supervised learnings are SVMs that stands for Support Vector Machines. This method is based on statistical learning and is suitable for optimization problems. There are several types of SVM such as Hard Margin SVM , Soft-Margin SVM , Sparse SVM ,The Dual Problem and Kernel Tricks, Support Vector Regression ,Support Vector Ordinal Regression. In this thesis the details of the SVMs has not been described as it is not considered the case for the proposed optimization models.

3.4 Deep Learning

In this part basic concept of deep learning and neural network has been discussed. As it has been mentioned before DNNs and regression are the most important methods of machine learning during this thesis.

Neural networks with a just one layer with a finite number of units can represent any multi-component continuous function on a compact subset in R_n with arbitrary precision. Nevertheless, the computational complexity required for training Deep Neural Networks.

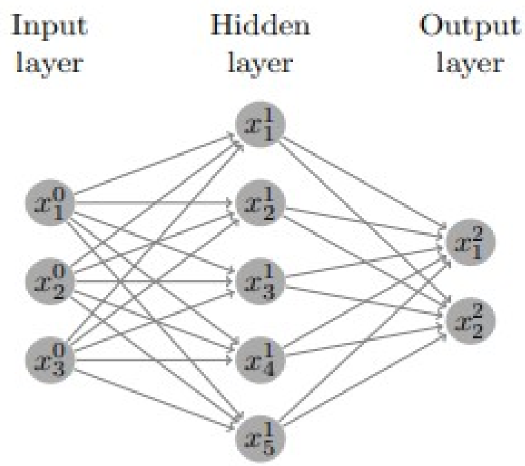


Figure 3.1: Deep Feedforward Neural Network with 3 layers.

Figure 3.1 is a full connected network where the first layer or input layer has three components and second layer or hidden layer has 5 components and the output layer has 2 components where each neurone in one layer is connected to all neurone in the next layer. For training these kinds of network it is necessary to determine the weight of each edge for several reasons, including the development of advanced processing units, namely GPUs, the advances in the efficiency of training algorithms such as back propagation, the establishment of proper initialisation parameters, and the massive collection of data enabled by new technologies in a variety of domains [23]. DNNs are compatible for the regression and this compatibility satisfies the requirements to get the most optimize patch antennas. The goal of this part is to declare the decision optimization problems associated with DNN architectures. To describe it easier, the notation for the common parameters is provided in Table 3.1.

$\{0, \ldots, L\}$	*layers* indices.
n^l	number of *units*, or *neurons*, in layer l.
σ	element-wise activation function.
$U(j, l)$	j-th unit of layer l.
$W^l \in \mathbb{R}^{n^l \times n^{l+1}}$	weight matrix for layer $l < L$.
$b^l \in \mathbb{R}^{n^l}$	bias vector for layer $l > 0$.
(X, y)	*training* dataset, with observations x_i and responses $y_i, i = 1, \ldots, n$.
x^l	output vector of layer l ($l = 0$ indicates *input feature* vector, $l > 0$ indicates *derived feature* vector).

Table 3.1: Notation for DNNs architecture[23]

The activation function shows which neurone should be activated and which one should not be activated in the output vector of the deep learning neural network it is possible to compute the propagated information by considering the weight of the edges and activation functions. It is reasonable to mention activation function is responsible for the capability of the DNNs to learn each complex models and determine the correlation between the input and output layers.

IV. Design and Optimization

4.1 Introduction

Chapter two outlined the characteristics of implanted antenna and presented the challenges of getting high performance implanted antenna, and in chapter three the different models of Machine Learning for optimisation purpose was presented. This chapter presents the performance optimization of implanted antenna using Machine Learning methods by investigating the different antennas designed in CST Studio Suite® which is a high-performance 3D EM analysis software package for designing, analyzing and optimizing electromagnetic (EM) components and systems.
This chapter proposes the design, characterisation and simulation of patch implanted antennas for body wireless communication, and optimisation with Machine Learning models.
The performance of antennas is monitored by calculating the S11 for all scenarios.
This chapter is organized as follows: Section 4.2 presents the design and optimization of first implanted antenna to test and examine the optimization model performance. In section 4.3 the implanted antenna has been designed in a trapezoidal shape and coaxial cable has been added to feed the antenna and the optimisation model developed by a larger data set of S11 parameters. In section 4.4 to have a more realistic scenario the antenna has been designed in a cylindrical shape, the suitable shape to be located around the bone in human body. Its dimensions has been optimized to get the best performance in frequency 2.45 GHZ. Finally, the conclusion is drawn in section 4.5.

4.2 Implanted Microstrip Patch Antenna Design

4.2.1 Design and optimization of a simple microstrip patch antenna

In this section a microstrip patch antenna is designed and investigated as the implant antenna. In the following every step of the design is elaborated.

Firstly, it is necessary to define the material as ground. In the first experience the material was Cooper. Dimension of subtract and transmission lines, dielectric height and the insets for this specific patch antennas has been defined. In this part this simple patch antenna was designed without using macros and parameters has been inserted manually. The aim was to find the most optimized point of this patch antenna by considering the center frequency at 2.45 GHz. In this case parameters inserted and as it was mentioned before the macros was not used. So 50 experiences are provided by changing three parameters which means there is three degrees of freedom. The next step was preparing the data as inputs for machine learning model. When all of 50 antennas have been provided, it has been tried to extract the optimized antenna. Figure 4.1 and 4.2 shows the Microstrip patch antenna designed for this experiments.

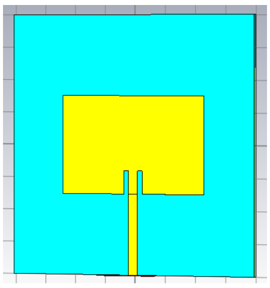


Figure 4.1: 2D view of first antenna

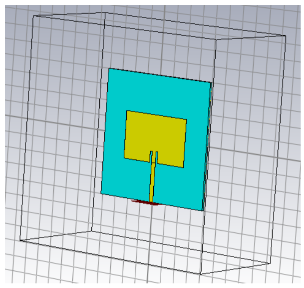


Figure 4.2: 3D view of first antenna

For all 50 experiments simulated S11 has been presented in figure 4.3 to monitor the results.

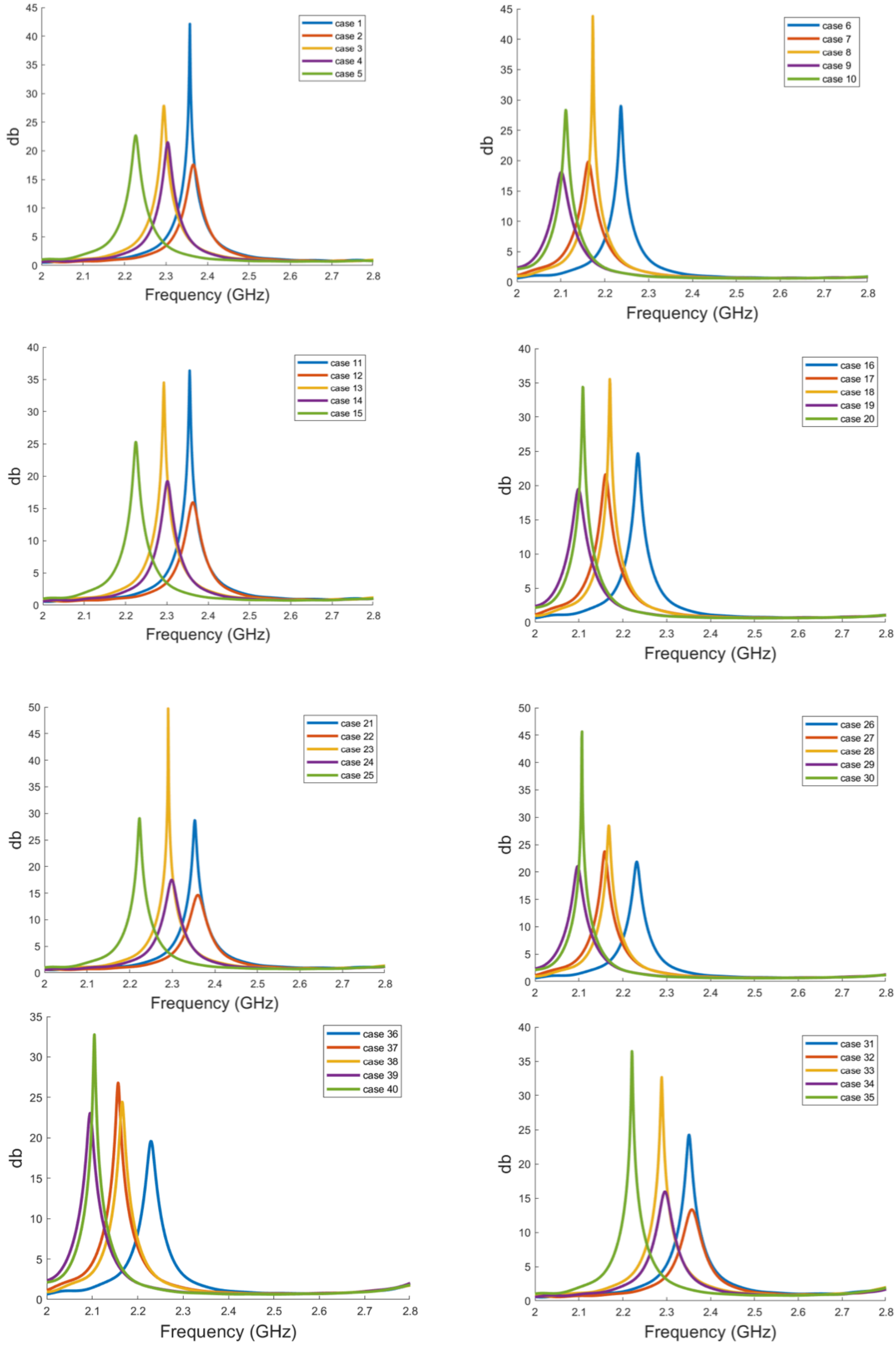
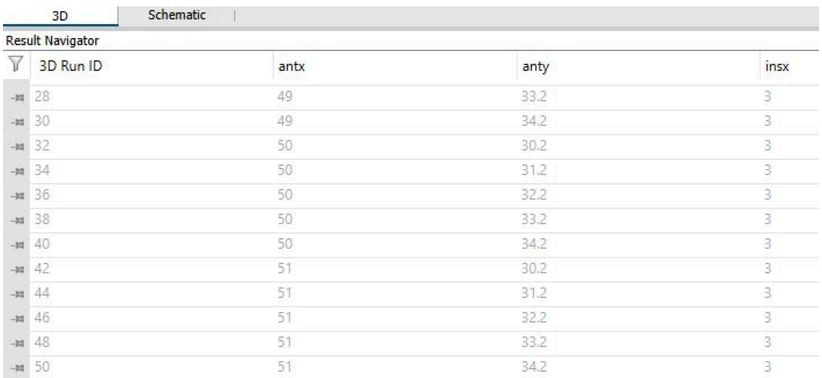


Figure 4.3: Simulated S11 of the patch antenna

Now there is a dataset here which is provided by these 50 experiences, and it has been tried to find the model by using all these 50 experiences and at the end compare the results for each one of the experiences.

In the following the procedure of first machine learning experience will be clarified. The results of our first experience provided in the following.

As it has been mentioned in the chapter three, two methods of machine learning is used in this thesis are regression and DNNs. For the first part of the experience the regression method with regression decision tree has been considered. As it mentioned before there are 3 degrees of freedom for this experiment first one is size of inset *"insx"* , the second one is the antenna's hight *"anty"* and the last one antenna's width *"antx"* . All the dimensions defined parametric, so it is possible to change them for each experience and it has been done for these 50 antenna as it is shown in figure 4.4.



3D Run ID	antx	anty	insx
28	49	33.2	3
30	49	34.2	3
32	50	30.2	3
34	50	31.2	3
36	50	32.2	3
38	50	33.2	3
40	50	34.2	3
42	51	30.2	3
44	51	31.2	3
46	51	32.2	3
48	51	33.2	3
50	51	34.2	3

Figure 4.4: Some examples of inputs for antenna dimension

All the exported data will the input of the regression learner in MATLAB software. As it is mentioned the regression tree method has been considered because Regression trees are easy to interpret, fast for fitting and prediction, and low on memory usage. To prevent overfitting, it has been considered to grow smaller trees with fewer larger leaves. With the Minimum leaf size setting, the leaf size can be controlled. To predict a response of a regression tree, it has to be followed the tree from the root (beginning) node down to a leaf node. The leaf node contains the value of the response.

Statistics and Machine Learning Toolbox™ trees are binary. Each step in a prediction involves checking the value of one predictor variable. For example, in figure 4.4 a simple regression tree is shown.

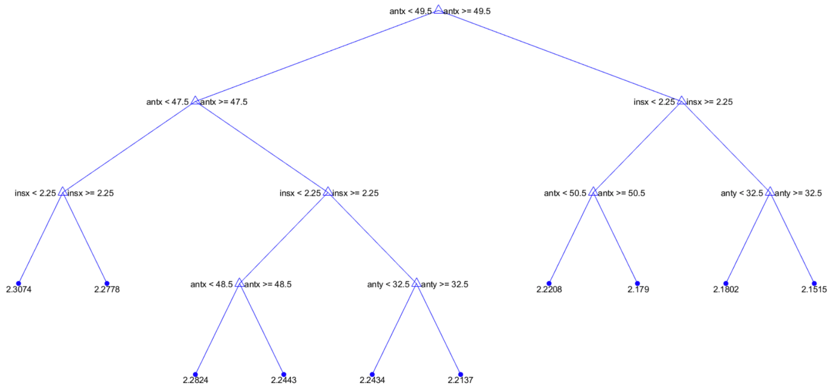


Figure 4.4: a simple decision tree

This tree predicts the response based on two predictors, x1 and x2. To make a prediction, it starts at the top node. At each node, it checks the values of the predictors to decide which branch to follow. When the branches reach a leaf node, the response is setted to the value corresponding to that node.

It is modeled with three different methods, which are Fine Tree, Medium tree, coarse tree. The model with the lowest RMSE is selected. Due to a small data set, prediction error is considerable.

Because of the obvious failure in this part DNNs has been chosen for the next part. On the other hand the model and parameters can be expandable through the macros and it is easy to get more and more data and find the best solution for the optimization.

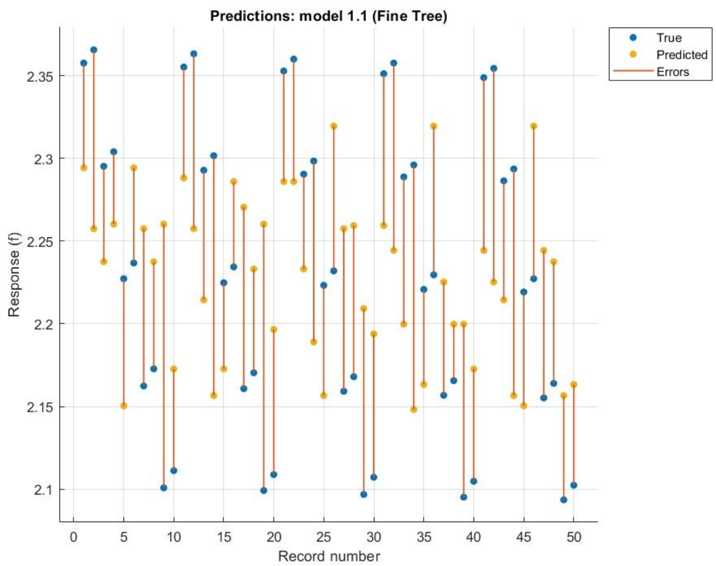


Figure 4.5: Comparison of the all output with prediction

4.2.2 Design and Optimization of trapezoidal patch antenna

In this geometry of antenna, the new microstrip patch antenna was designed to be compatible with the human body. The target is to get best performance at 2.45GHz. Firstly the one-layer microstrip Patch antenna has been designed and in figure 4.6 the antenna's geometry is indicated from different views.

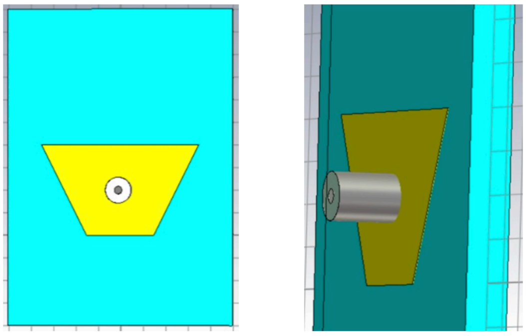


Figure 4.6 : one layer geometry for trapezoidal patch antenna from two different views

It is necessary to mention each step of this design. For this geometry unlike the last simple geometry which has been discussed before, there is a coaxial cable to feed the patch antenna, so it is necessary for the optimization to find the best location of the coaxial cable. The other important difference is the shape of microstrip antenna which in this section has been designed in a trapezoidal shape. In fact, this is the first step of designing a cylindrical antenna which is suitable to be implanted around the bone in human body. The cylindrical design will be described in section 4.2.3.

To design a microstrip patch antenna with an optimised performance which is the aim of this thesis, a suitable data set is required. Using macros in CST, a data set of 599 antenna is prepared by considering the width, length, thickness and coaxial cable location as parameters. In the following the macros code has been used for this iteration is presented.

```
' macro2

Sub Main ()
    StoreParameter("a", 30)
    StoreParameter("b", 70)
    StoreParameter("h", 40)
    StoreParameter("thickness", 10)
    StoreParameter("ccx", 0)
    StoreParameter("ccy", 20)
    Rebuild
    save

End Sub
```

It is observable there are 599 times of the experiments with macros which produces 599 simulated S11. Figure 4.7 shows the simulated S11 in an antenna as an example of designed antennas.

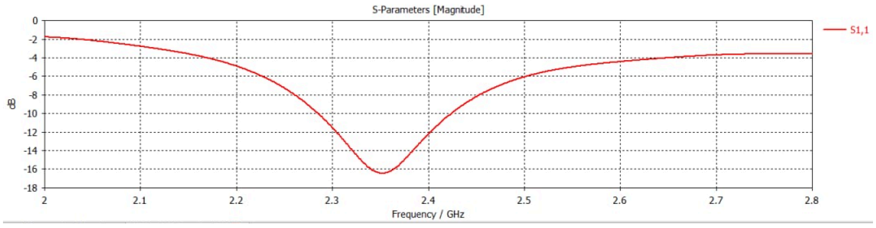


Figure 4.7: Simulated S11 of the patch antenna

Neural network has been used as optimization tool. In order to train validate and test the model, the data set is divided as following. 70 percent of the data set has been considered as training data, 15 percent has been considered for validation and 15 percent for testing the model. The configuration of the network will be clarified in the following lines:

To enter four degree of freedom into the system, these four parameters has been considered as input matrixes: `inputs=[a;b;h;ccy];`

The definition of the output: `output=[p];`
To make 10 nodes for each of the three layers: `net=fitnet([10,10,10]);`
In figure 4.8 the configuration of the network is shown.

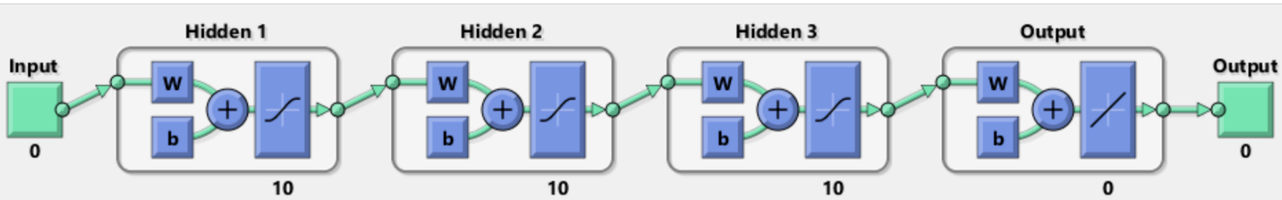


Figure 4.8 : configuration of the network

As it has been mentioned the data set will be divided to training set, validation and test set using following code. The specified parameters to the network has been presented in figure 4.9. In figure 4.10 the result of training performance is shown.

```
net.divideParam.trainRatio=70/100;
net.divideParam.valRatio=15/100;
net.divideParam.testRatio=15/100;
[net,tr]=train(net,inputs,output)
```

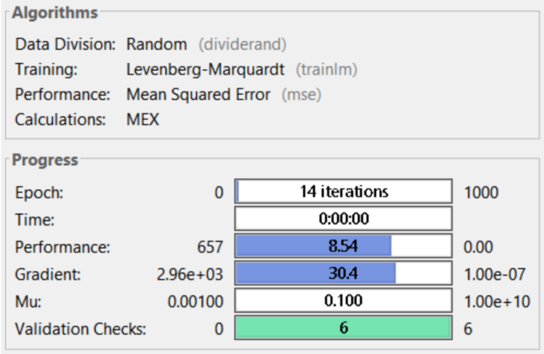


Figure 4.9: properties of the network

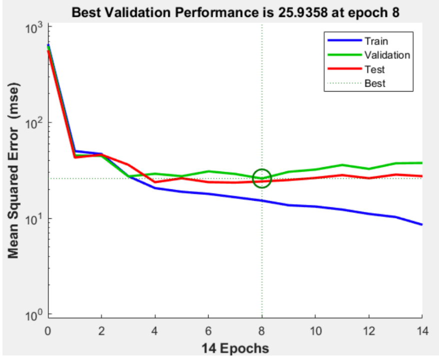


Figure 4.10: Performance of the network

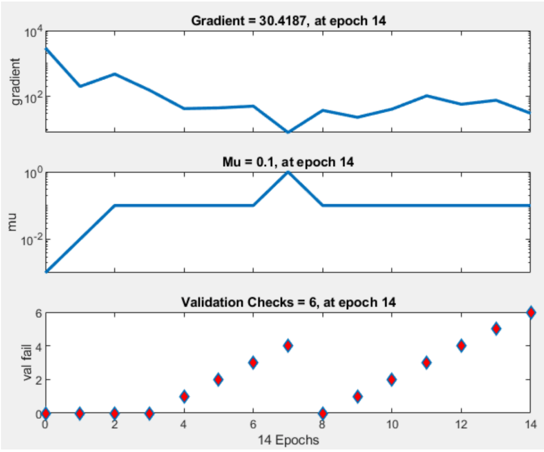


Figure 4.11: Neural network training state

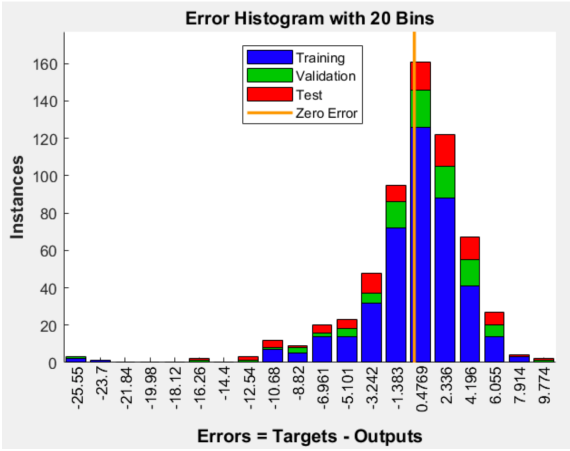


Figure 4.12: Neural network error histogram

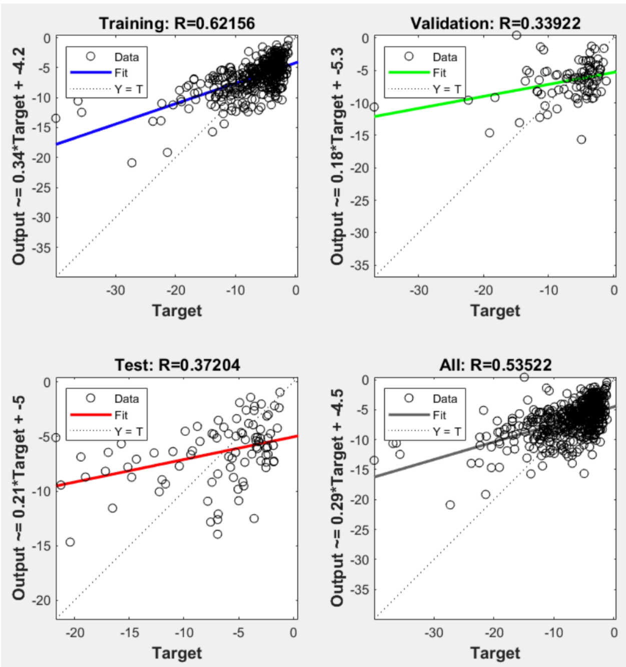


Figure 4.13: Neural network training regression

With the following code it is possible to reach to the best result for the optimization, as it has been discussed the aim of the thesis is to find the most optimized dimensions and the location of the coaxial cable to feed the antenna. After finishing all steps of machine learning final result has been cleared as following.

"a" width parameter :70 mm

"b" length parameter:110 mm

"h" Height parameter :40 mm

"ccy" stand for coaxial cable location: 13

```
for a2=10:100
    for b2=10:190
        for h2=30:70
            for ccy2=10:h-10
                f3(j)=2.4;
                a3(j)=a2;
                b3(j)=b2;
                h3(j)=h2;
                ccy3(j)=ccy2;
                j=j+1;

            end
        end
    end
end
test1=net([a3;b3;h3;ccy3]);
d=find(test1==min(test1));
test2=test1(d)
a4(:)=a3(d)
b4(:)=b3(d)
h4(:)=h3(d)
ccy4(:)=ccy3(d)
```

Figure 4.14 shows the output of configured neural network which works correctly at frequency 2.4 GHz.

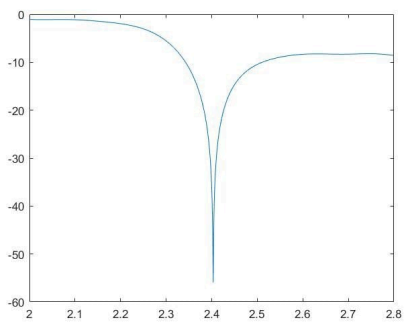


Figure 4.14: Optimised output of Neural Network at frequency 2.4 GHz

Figure 4.15 shows S11 parameter of a simulated antenna with the resonance frequency at 2.45 GHz. The result of the optimised antenna using neural network is shown as S11 parameter in figure 4.16. The impedance and admittance view of optimised antenna's smith chart is shown in figure 4.17 and 4.18.

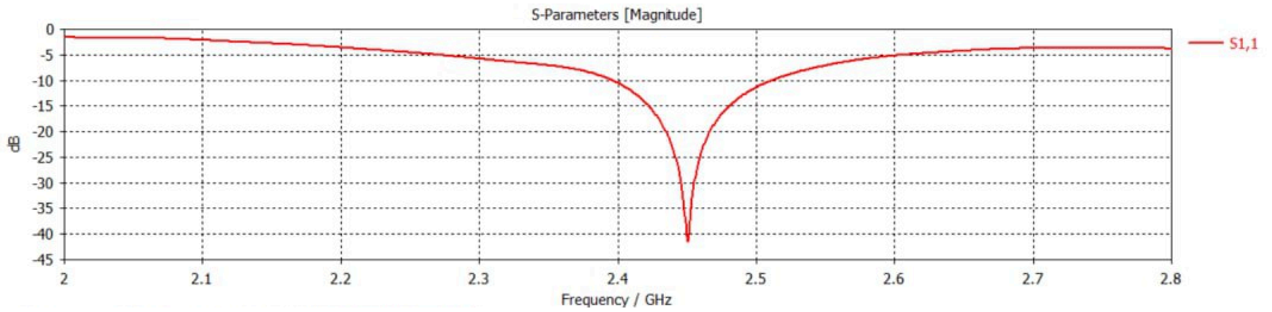


Figure 4.15: One of the simulated S11 at frequency 2.45 GHz

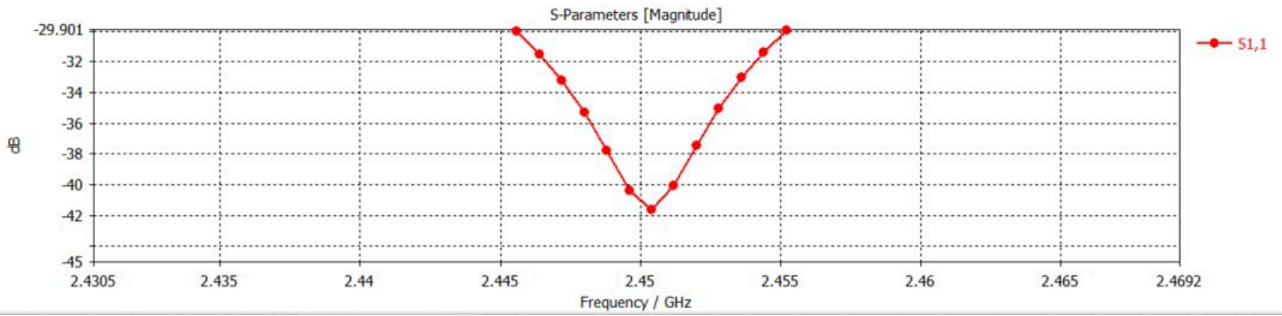


Figure 4.16: Optimised S11 by Neural Network at frequency 2.45 GHz

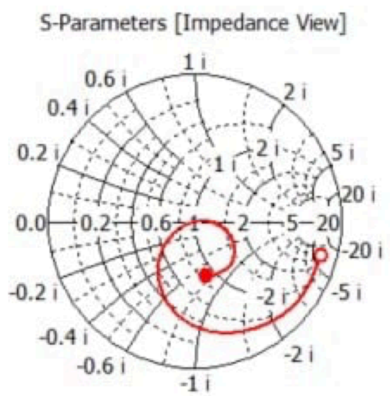


Figure 4.17 : Z smith chart of optimised antenna at 2.45 GHz

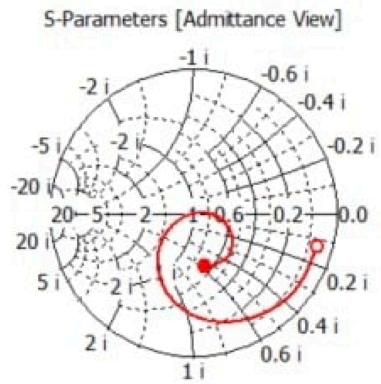


Figure 4.18: Y smith chart of optimised antenna at 2.45 GHz

In order to design a patch antenna compatible with human body, in this simulation the skin, fat, muscle and bone effects should be considered. First, the flat version of this geometry has been designed and the materials and tissues parameters are based on [6] which is reported in table 4.1 . In the following there are the different views of the geometry in figure 4.19.

Table 4.1 Materials and tissues parameters at frequency 2.45 GHz

	ε'	ε''	$\tan\delta$
muscle	57.1	5.68	0.0995
fat	5.56	.288	0.0514
skin	46.7	4.9611	.106

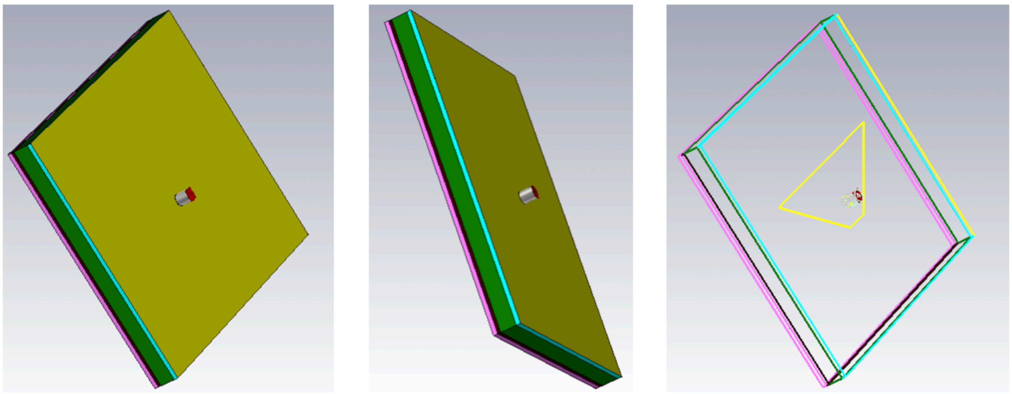


Figure 4.19: Multi-layer model considering the body tissues in trapezoidal patch antenna design

4.2.3 Design and Optimization of Cylindrical patch antenna

For this simulation the antenna has been designed in a cylindrical shape, suitable to be implanted on a bone in human body. The characteristics of the body tissues has been considered base on [6] which it considers the muscle (of thickness thmuscle = 15mm), fat (thfat = 3mm) and skin (thskin = 4mm). The central conductive lossy metal cylinder of radius R = 30 mm has been considered from a bio-compatible metal.

In this model the shape of the design has been changed to curvy from a flat antenna, but the other steps has been applied same as the previous model and it is possible to report the best optimization for this antenna base on the data which has been extracted from the simulation as following values:

“a” width parameter : 26 mm

“b” length parameter: 24 mm

“h” Height parameter : 18.5 mm

“ccy” stand for coaxial cable location: 10

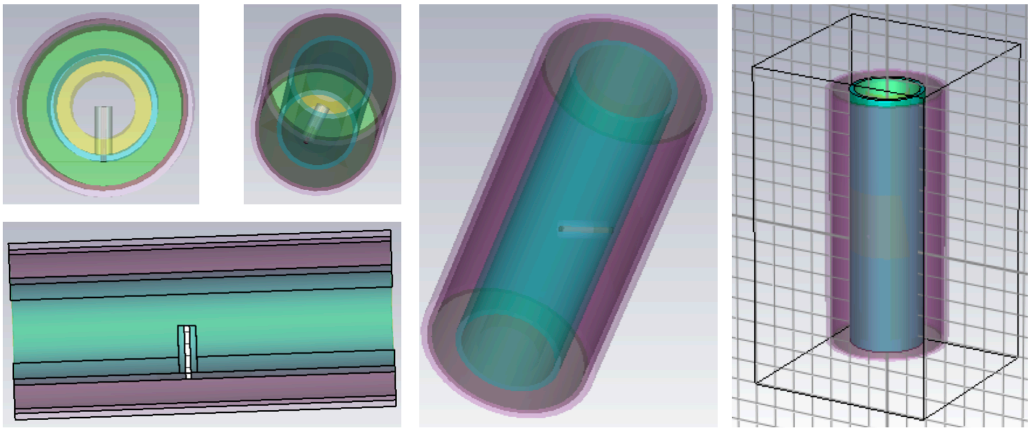


Figure 4.20: different views of Cylindrical patch antenna

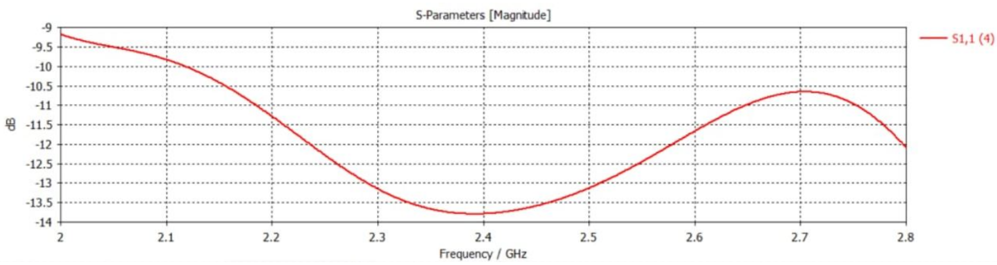


Figure 4.21 : S11 of the optimized model

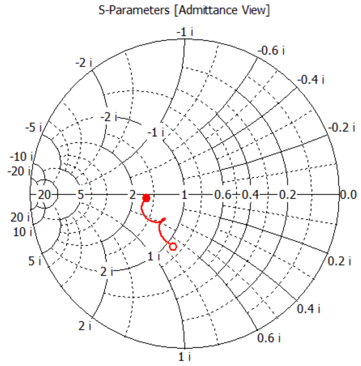


Figure 4.22: Z smith chart of
optimised antenna at 2.45 GHz

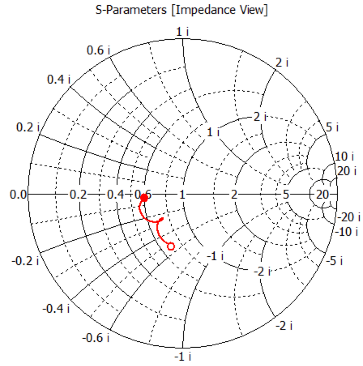


Figure 4.23 : Y smith chart of
optimised antenna at 2.45 GHz

4.3 Conclusion

This thesis presented a new optimization and design for the implantable antennas for biomedical applications. In particular, it was highlighted a new ways to produce high performance implanted antennas. A review of the implantable devices design, challenges and applications were presented. Comprehensive survey for the machine learning methods in solving the optimization problems presented.
It is demonstrated that one to the main challenges in antennas performance is their size. Three models for implantable patch antenna is proposed and in each step tried to achieve the more realistic model for the antenna. To justify the results, S11 parameter and Smith Chart for each antenna was monitored. To optimise the size of the antennas the machine learning methods is employed. After an unsuccessful attempt with decision tree, Neural Network was chosen as a main method for training a model. Finally the optimised size of antenna was introduced.
this study investigated to determine novel approaches to improve implantable antennas performance.

References

[1] Khan, Ahsan Noor, Young-ok Cha, Henry Giddens, and Yang Hao. "Recent advances in organ specific wireless bioelectronic devices: Perspective on biotelemetry and power transfer using antenna systems." *Engineering* (2022).

[2] Andre, David, and Donna L. Wolf. "Recent advances in free-living physical activity monitoring: a review." *Journal of diabetes science and technology* 1, no. 5 (2007): 760-767.

[3] Matekovits, Ladislau, Jie Huang, Ildiko Peter, and Karu P. Esselle. "Mutual coupling reduction between implanted microstrip antennas on a cylindrical bio-metallic ground plane." *IEEE Access* 5 (2017): 8804-8811.

[4] Hong, Yongseok Joseph, Hyoyoung Jeong, Kyoung Won Cho, Nanshu Lu, and Dae-Hyeong Kim. "Wearable and implantable devices for cardiovascular healthcare: from monitoring to therapy based on flexible and stretchable electronics." *Advanced Functional Materials* 29, no. 19 (2019): 1808247.

[5] Sidhu, R., D. S. Sanders, A. J. Morris, and M. E. McAlindon. "Guidelines on small bowel enteroscopy and capsule endoscopy in adults." *Gut* 57, no. 1 (2008): 125-136.

[6] Matekovits, Ladislau, Yalin Su, and Ildiko Peter. "On the radiation mechanism of implanted antennas with large conformal ground plane." *IET Microwaves, Antennas & Propagation* 11, no. 12 (2017): 1765-1769.

[7] Nikolayev, Denys, Wout Joseph, Anja Skrivervik, Maxim Zhadobov, Luc Martens, and Ronan Sauleau. "Dielectric-loaded conformal microstrip antennas for versatile in-body applications." *IEEE Antennas and Wireless Propagation Letters* 18, no. 12 (2019): 2686-2690.

[8] Available at http://zhadobov.fr/index.php/project/body-centric-wireless-communications/

[9] Afroz, S., Stephen W. Thomas, Gokhan Mumcu, and S. E. Saddow. "Implantable SiC based RF antenna biosensor for continuous glucose monitoring." In *SENSORS, 2013 IEEE*, pp. 1-4. IEEE, 2013.

[10] Gutierrez, Jose A., Marco Naeve, Ed Callaway, Monique Bourgeois, Vinay Mitter, and Bob Heile. "IEEE 802.15. 4: a developing standard for low-power low-cost wireless personal area networks." *IEEE network* 15, no. 5 (2001): 12-19.

[11] Zhao, Yujuan, Robert Rennaker, Chris Hutchens, and Tamer S. Ibrahim. "Simulation of implantable miniaturized antenna for brain machine interface applications." In *Proc. Adv. CEM Appl. II ACES conf.*, pp. 1-6. 2012.

[12] Shokry, M. A., and A. M. M. A. Allam. "Implanted antenna in brain." In *2013 Loughborough Antennas & Propagation Conference (LAPC)*, pp. 630-633. IEEE, 2013.

[13] Islam, Md Shahidul, Karu P. Esselle, David Bull, and Paul M. Pilowsky. "Converting a wireless biotelemetry system to an implantable system through antenna redesign." *IEEE Transactions on Microwave Theory and Techniques* 62, no. 9 (2014): 1890-1897.

[14] Zakavi, Parisa, Nemai Chandra Karmakar, and Ian Griggs. "Wireless orthopedic pin for bone healing and growth: Antenna development." *IEEE transactions on antennas and propagation* 58, no. 12 (2010): 4069-4074.

[15] Yazdandoost, Kamya Yekeh, and Ryu Miura. "Miniaturized UWB implantable antenna for brain-machine-interface." In *2015 9th European Conference on Antennas and Propagation (EuCAP)*, pp. 1-5. IEEE, 2015.

[16] Surapan, Saharat, Supakit Kawdungta, Hung-Chi Yang, and Chih-Kuang Wu. "Design of dual band implantable antenna for biomedical applications." In *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 1-4. IEEE, 2016.

[17] Li, Jingjing, Ildiko Peter, and Ladislau Matekovits. "Circularly polarized implanted antenna with conical Bio-Metallic ground plane." In *2017 13th IASTED International Conference on Biomedical Engineering (BioMed)*, pp. 265-269. IEEE, 2017.

[18] Abadia, Javier, Francesco Merli, J. F. Zürcher, Juan R. Mosig, and Anja K. Skrivervik. "3D-spiral small antenna design and realization for biomedical telemetry in the MICS band." *Radioengineering* 18, no. ARTICLE (2009): 359-367.

[19] Huang, Wei, and Ahmed A. Kishk. "Embedded spiral microstrip implantable antenna." *International Journal of Antennas and Propagation* 2011 (2011).

[20] Sánchez-Fernández, C. J., O. Quevedo-Teruel, J. Requena-Carrión, L. Inclán-Sánchez, and Eva Rajo-Iglesias. "Dual-band microstrip patch antenna based on short-circuited ring and spiral resonators for implantable medical devices." *IET microwaves, antennas & propagation* 4, no. 8 (2010): 1048-1055.

[21] Merli, Francesco. *Implantable antennas for biomedical applications*. No. THESIS. EPFL, 2011.

[22] available at https://life.itu.int/radioclub/rr/art1.pdf

[23] Gambella, Claudio, Bissan Ghaddar, and Joe Naoum-Sawaya. "Optimization problems for machine learning: A survey." *European Journal of Operational Research* 290, no. 3 (2021): 807-828.

[24] Solomonoff, Raymond J. "An inductive inference machine." In *IRE Convention Record, Section on Information Theory*, vol. 2, pp. 56-62. New York: Institute of Radio Engineers, 1957.

[25] Vapnik, Vladimir. *The nature of statistical learning theory*. Springer science & business media, 1999.

[26] Doshi-Velez, Finale, and Been Kim. "Towards a rigorous science of interpretable machine learning." *arXiv preprint arXiv:1702.08608* (2017).

[27] Friedman, J., T. Hastie, and R. Tibshirani. "The Elements of Statistical Learning. Volume 1 Springer; New York, NY, USA: 2001."

[28] Bertsimas, Dimitris, and Martin S. Copenhaver. "Characterization of the equivalence of robustification and regularization in linear and matrix regression." *European Journal of Operational Research* 270, no. 3 (2018): 931-942.

[29] Bertsimas, Dimitris, and Angela King. "OR forum—an algorithmic approach to linear regression." *Operations Research* 64, no. 1 (2016): 2-16.

[30] Bertsimas, Dimitris, and Romy Shioda. "Classification and regression via integer optimization." *Operations research* 55, no. 2 (2007): 252-271.

[31] Hadi, Ali S., and Samprit Chatterjee. *Regression analysis by example*. John Wiley & Sons, 2015.

[32] Tamura, Ryuta, Ken Kobayashi, Yuichi Takano, Ryuhei Miyashiro, Kazuhide Nakata, and Tomomi Matsui. "Best subset selection for eliminating multicollinearity." *Journal of the Operations Research Society of Japan* 60, no. 3 (2017): 321-336.

[33] Tamura, Ryuta, Ken Kobayashi, Yuichi Takano, Ryuhei Miyashiro, Kazuhide Nakata, and Tomomi Matsui. "Mixed integer quadratic optimization formulations for eliminating multicollinearity based on variance inflation factor." *Journal of Global Optimization* 73, no. 2 (2019): 431-446.

[34] Bertsimas, Dimitris, and Nathan Kallus. "From predictive to prescriptive analytics." *Management Science* 66, no. 3 (2020): 1025-1044.

[35] Bertsimas, Dimitris, Angela King, and Rahul Mazumder. "Best subset selection via a modern optimization lens." *The annals of statistics* 44, no. 2 (2016): 813-852.

[36] Bertsimas, Dimitris, and Bart Van Parys. "Sparse high-dimensional regression: Exact scalable algorithms and phase transitions." *The Annals of Statistics* 48, no. 1 (2020): 300-323.

[37] Claassen, G. D. H., and Th HB Hendriks. "An application of special ordered sets to a periodic milk collection problem." *European Journal of Operational Research* 180, no. 2 (2007): 754-769.

[38] Hastie, Trevor, Robert Tibshirani, and Ryan J. Tibshirani. "Extended comparisons of best subset selection, forward stepwise selection, and the lasso." *arXiv preprint arXiv:1707.08692* (2017).

[39] James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*. Vol. 112. New York: springer, 2013.

[40] Chouldechova, Alexandra, and Trevor Hastie. "Generalized additive model selection." *arXiv preprint arXiv:1506.03850* (2015).

[41] Toriello, Alejandro, and Juan Pablo Vielma. "Fitting piecewise linear continuous functions." *European Journal of Operational Research* 219, no. 1 (2012): 86-95.

[42] Dreiseitl, Stephan, and Lucila Ohno-Machado. "Logistic regression and artificial neural network classification models: a methodology review." *Journal of biomedical informatics* 35, no. 5-6 (2002): 352-359.

[43] Chen, Xiaobo, Jian Yang, David Zhang, and Jun Liang. "Complete large margin linear discriminant analysis using mathematical programming approach." *Pattern recognition* 46, no. 6 (2013): 1579-1594.

[44] Bertsimas, Dimitris, and Jack Dunn. "Optimal classification trees." *Machine Learning* 106, no. 7 (2017): 1039-1082.

[45] Izenman, Alan Julian. "Modern multivariate statistical techniques." *Regression, classification and manifold learning* 10 (2008): 978-0.