Preliminary Study of a Mobile Microwave Breast Cancer Detection Device Using Machine Learning

by

Jorge A. Sacristán

A thesis submitted to
The Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements
of the degree of

Master of Science

Department of Physics and Astronomy
The University of Manitoba
Winnipeg, Manitoba, Canada
December 2017

Thesis advisor

Stephen Pistorius, PhD.

Author

Jorge A. Sacristán

Preliminary Study of a Mobile Microwave Breast Cancer Detection Device Using Machine Learning

Abstract

Current breast cancer screening, using X-ray mammography has various drawbacks. These include the use of ionising radiation, the need for breast compression, high cost, and the difficulty in implementing this technology in rural communities. The prototype of a portable screening device is presented; its design is aimed at addressing the aforementioned problems.

Eight machine learning classifiers have been trained to predict the presence or absence of tumour tissue from numerical breast models. Scattered electromagnetic field values have been used as inputs for the classifiers. The performances of the algorithms are presented and discussed, with two of the classifiers achieving metrics comparable to those obtained by current X-Ray mammography modalities [1].

Further improvements on classification features might be necessary to adapt the proposed approach to clinical practice. To this extent, the use of time-domain information arising from both breasts in the classification process will be considered in the future.

Contributions

The notable contributions of the author of this thesis are the following:

- Design and assembly of a portable Breast Microwave Sensing (BMS) system's preliminary prototype. A rotating chamber houses a transmitting antenna and 12 microwave sensors. This rotating device can interrogate a user's breast with a microwave pulse from five positions within a 48° arc.

- Implementation of a two-dimensional Method of Moments (MoM) electromagnetic forward solver with parallel processing capabilities. An Application Program Interface (API) was designed to compute the electromagnetic field scattered by arbitrary cross-sections of tissues with known permittivity values. A database of tissues and their dielectric properties was compiled by Hayden Banting under the author's supervision.

- Creation of five datasets comprised of numerical two-dimensional breast models. Each dataset contains data from the electromagnetic scattered field produced by the tissues of each breast model. One of the datasets includes homogeneous contents, the remaining four include Fibroglandular Tissue (FGT). Two of the datasets contain information from five positions of the rotating chamber, the remaining three use one position.

- Application of eight Machine Learning (ML) classifiers on the five numerical breast datasets. A comparison of the performances obtained by the classifiers is presented.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to begin by thanking my advisor, my committee, my parents, my significant other, and all the people who have supported me along the way.

To my parents, Alma and Luis, nothing would have been possible without your constant love and support.

Chapter 1

Introduction

A 2017 study attributes 13% of all cancer-related deaths in Canadian women during 2017 to breast cancer [2]. The same study claims that over 32,000 breast cancer deaths have been avoided in Canada between the years of 1987 and 2012, their findings suggest these deaths have been avoided due to the advances in breast cancer treatment as well as the availability of breast cancer screening technologies. Other studies have also shown a reduction in mortality associated with X-ray mammography breast cancer screening [3, 4]. However, in the last decade other authors have challenged these findings and have argued that the benefits associated with mammography are still inconclusive [5, 6].

The reported incidence of breast cancer is higher in the developed nations, while people in emerging economies have lower survival rates. For instance, the five-year survival rate for breast cancer is less than 50% in Gambia, Uganda, and Algeria, while it is nearly 90% in the United States [7]. In Manitoba, Canada, women from rural areas have a cancer mortality-to-incidence rate of 60%, while their urban counterparts

have a rate of 37% [8]. These statistics also suggest a link between survival rates and availability of breast cancer screening technologies.

Aside from the existence of divided opinions regarding the benefits of X-ray mammography, the technology itself has many drawbacks, some of them are: a) The use of ionising radiation, which can raise the risk of developing cancer. b) The need for breast compression, which many women find uncomfortable. c) High costs and the difficulty in bringing this technology to remote or rural communities. d) Relatively low sensitivity and specificity rates [9].

X-Ray mammography is not the only technology capable of detecting breast tumours, Ultrasound and Magnetic Resonance Imaging (MRI) can also fulfil this purpose. However, these approaches have their own problems, In the case of MRI these include higher costs and the subjection patients to a confined space, which might render MRI unsuitable for patients who suffer from claustrophobia. In the case of Ultrasound, disadvantages include low sensitivity rates might lead to unnecessary procedures such as biopsies, and the need of a highly skilled operator to detect a tumour.

It has been suggested that X-Ray mammography on its own, or even when combined with Ultrasound, is insufficient for the early diagnosis of breast cancer in the case of patients with increased risk due to family history of cancer [10].

In the last decades, there has been an increasing number of researchers using microwave technologies to address the shortcomings of current breast cancer screening modalities. A common approach is the one provided by Microwave Imaging (MWI), which produces images based on the scattering of electromagnetic pulses within the

microwave range.

The research presented in this thesis is part of an ongoing project that aims to improve the availability of breast cancer screening technologies by providing a portable Breast Microwave Sensing (BMS) system that is suited to the needs of low and middle-income countries and rural communities. This BMS system has been designed by borrowing different approaches and ideas from several other researchers, to maximise portability and minimise costs without compromising the diagnostic capabilities. This thesis evaluates the feasibility of such system by analysing numerical simulations of microwave scattering data from breast models inside the system. Instead of the more popular MWI approach, Machine Learning (ML) classifiers have been used to determine the presence or absence of breast tumours. A comparison between the performances of eight classifiers is provided, as well as future research directions.

The purpose of this thesis is to determine what is the most effective type of ML classifier that could be employed in the BMS system to determine the presence or absence of a tumour inside a patient's breast.

Chapter 2

Literature Review

Chapter 1 presented the relationship between breast cancer incidence and mortality rates, as well as the link between decreasing death rates and increasing breast cancer screening availability. Some of the drawbacks inherent to the most common screening technology, X-ray mammography, Ultrasound and Magnetic Resonance Imaging (MRI) were presented.

This chapter presents an introduction to Machine Learning (ML) in general, with instances of previous studies using ML for tumour detection being presented. The technical details of microwave-based tumour detecting prototypes, published by other research groups are also presented in this chapter. An overview of the computational simulations of microwave scattering in different materials used for this research is presented. Finally, the different approaches used in the breast tumour detection prototypes of other researchers are explored, with an emphasis on identifying technologies relevant to the design and implementation of a new portable Breast Microwave Sensing (BMS) system.

2.1 Machine Learning

In broad terms, Machine Learning (ML) involves the use and development of algorithms that can make predictions about future instances by learning patterns from previous "test" instances. In this context, the term "instances" also known as "samples", refer to the observation of a phenomenon that can have several measurable features associated with said sample [11].

Features typically have a numerical value, but other forms of features are possible, like a string of characters. In this thesis, the feature vectors consist of the microwave power values found at 12 sensor locations. These power values are obtained by simulating the electrical scattering produced by numerical breast models exposed to 5 signals with different frequencies in the 2.3 GHz to 6.5 GHz range. Further details on the feature vectors are presented in Chapter 3.

Supervised and unsupervised learning are the two prominent categories of ML. In the supervised scenario, a "label" associated with every sample is given to the classifier during the training phase, and new samples are assumed to belong to one of the possible labels that were provided to the classifier during training. In the case of unsupervised learning, the classifier clusters samples and discovers different classes of items based on the provided features. In this thesis, only supervised ML is used, the labels employed are related to the presence or absence of a tumour.

An introductory example of the type of data used in supervised ML, is the classical "Iris flower dataset" published by Fisher in 1936 ([12]), where samples consist of four features (sepal and petal's width and length) and the associated label correspond to the species of the flower (Iris setosa, Iris versicolor and Iris virginica). An ML

classifier can learn patterns produced by the combination of values that these four features have, and use those patterns to predict the species of a new given sample.

The following subsections serve as a short introduction to some of the most popular types of classifying algorithms found in the literature:

2.1.1 Logic-based Classifiers

Decision trees are non-parametric classifiers [13] based on Boolean logic; they operate by recursively dividing the feature space into regions that can be associated with a Boolean evaluation. The features of a sample are subjected to such evaluations to assign the sample to a class [11].

Due to the recursive nature of the trees, one evaluation can branch off to further evaluations until a leave of the tree is reached. The leaves of a tree represent the decision of assigning a sample to a specific class. A maximum depth is usually imposed during the training process to prevent the tree from recursively growing indefinitely.

The reader is invited to consult the "Tree-Based Methods" section in *The elements of statistical learning* ([14]) for a more technical breakdown of decision trees.

2.1.2 Statistical Classifiers

Linear Discriminant Analysis (LDA) is a probabilistic model that computes the probability of a sample belonging to a class; it finds the linear combination of features that separates most efficiently two or more classes [11]. This classifier assumes the classes to be normally distributed within the feature space. As the name suggests, the decision function of this classifier divides the feature space with a straight line in

two-dimensions or a hyperplane in higher dimensions. In the same vein, Quadratic Discriminant Analysis (QDA) is a technique that separates the feature space using a quadratic surface instead of a line.

The third technique in this family is a special case of QDA called the Naive Bayes classifier. If the features are assumed to be independent to one another in QDA, then the covariance matrices will be diagonal and the resulting classifier is equivalent to Naive Bayes. Even though this assumption of independence might be wrong in several real-world cases, Domingos and Pazzani found this classifier to be sometimes superior to other learning techniques even on datasets in which the features were strongly dependent on each other [15] [11].

2.1.3 Instance Based Learners

K-Nearest Neighbours (KNN) classifier is a type of non-parametric ML algorithm; it is based on the assumption that the features of a sample will have a small distance to the features of other samples of the same class [11]. In this technique, the distance between two samples can be defined as the Euclidean distance between their feature vectors. However, other definitions of this metric can be used. Some examples of alternative distances are: Minkowski, Manhattan, Chebyshev, and Canberra distances [11].

The performance of KNN depends on selecting a good value of 'K'. As the classifier's name suggests, this is the number of closest "neighbours" to which a new sample is compared [13]. A new sample is assigned the class of the neighbours with the closest distance. To avoid having a divided decision, it is common to set 'K' to an

odd number.

2.1.4 Margin Classifiers

Support Vector Machines (SVMs) in their current form were introduced in 1995 by Cortes and Vapnik [16]. This technique relies on the concept of a margin, a hyperplane in the feature space that divides samples classes from one another. The objective of a SVM is then to maximise the distance between the classes and the hyperplane [11]. If the classes of a dataset are linearly separable and a hyperplane is found, the specific samples that fall in the margin are called Support Vector Points (SVPs), these points are used to represent the solution while the rest of the samples are ignored.

In the cases where a hyperplane that perfectly separates the classes is not found, the introduction of a transformation kernel that maps feature vectors into higher-dimensional spaces can often help find a better solution [11]. Different kernels might rely on the selection of parameters that affect the performance of the classifier, for this reason, techniques such as Grid-Search Cross-Validation (GSCV) and Particle Swarm Optimization (PSO) are often employed to find parameters values that maximise performance for a given dataset [13]. Some popular choices for kernel functions are: a) Linear given by $K(x, x') = \langle x, x' \rangle$, b) Radial Basis Function (RBF) given by $K(x, x') = e^{-\gamma||x-x'||^2}$, and c) dth-degree polynomial given by $K(x, x') = (1 + \langle x, x' \rangle)^d$.

The book *The elements of statistical learning* by Friedman, Hastie, and Tibshirani provides a nice introduction to the concept of SVM in the "Support Vector Machines and Kernels" section. The reader is also encouraged to read the lecture notes by Ng

for a detailed crash-course on SVM [17].

2.1.5 Artificial Neural Networks

Artificial Neural Network (ANN) are algorithms loosely based on the structure of a human brain, where individual nodes (also called perceptrons) resemble the role of a neuron that is linked to other neurons within the network [13].

The simplest form of ANN is a single perceptron, this one-node classifier computes the weighed sum of inputs corresponding to the features of a sample, the output of the weighed sum is typically given to an activation function ϕ and its result (typically within the [-1, 1] range) determines the sample's class [11]. These perceptrons can be linked together to construct more complex structures according to the complexity of the dataset. A Multi Layer Perceptron (MLP) is a popular type of a Neural Network. It is comprised of an inner layer, zero or more hidden layers and an output layer of perceptrons. The choice of nomenclature for different Neural Networks is usually determined by the algorithm used in the training phase of the classifier, some common examples are Genetic Neural Network (GNN), Back Propagation Neural Network (BPNN), Radial Basis Function Neural Network (RBNN) [13].

Back-propagation is a popular technique to train a neural network by determining the weights linking one neuron with each of the neurons from the previous layers. This algorithm makes use of the gradient descent method to iteratively minimise the error in the output layer of the network. Since the derivative of the activation function is computed by gradient descent, the chosen activation function needs to be differentiable. A popular choice is the logistic function, given by $f(x) = \frac{1}{1+e^{-x}}$.

The section "Error Backpropagation" in the book *Pattern Recognition and Machine Learning* by Bishop contains a detailed breakdown of the steps involved in the back-propagation method for training neural networks [18].

2.1.6 Ensemble techniques in Machine Learning

The term "ensemble" in ML refers to the combination of predictions from multiple "base classifiers". A simple example of this scheme is a hard-voting ensemble, where a number of classifiers cast a vote to predict the class of a sample, the class with more votes is then returned as the result of the ensemble. Soft-voting is a similar approach that makes use of the predicted probability outputs of the classifiers. These predicted probabilities are then summed and the result of the ensemble is the class with a higher sum.

Random forests are an example of an ensemble ML method, they achieve classification by training several decision trees and making those trees with different structure cast a "vote" on every new sample. The class with the most votes is returned by the random forest as the prediction [13].

Boosting is another example of an ensemble method, it is designed to improve the performance of weak classifiers by sequentially addressing the shortcomings of the previous weak classifier. The final prediction, similarly to other ensemble techniques, is the weighed sum of the prediction of each individual weak classifier. AdaBoost (Adaptive Boosting) was introduced in 2003 by Schapire and it is a popular example of a boosting method for ML [19]. An intuitive of analogy between boosting and the use of human specialists is provided by Pareek and Ravikumar [20].

2.1.7 Machine Learning summary

The previous subsections have presented an introductory background on different supervised ML classifiers, their nomenclature, some of the methods used for their training as well as recommended literature sources for further reading. Eight of these classifiers were implemented and their results were analysed in the research leading to this thesis. For consistent column sizes in tables and software debugging, the implemented classifiers were given a 9-character identifier [shown between squared brackets]:

1. Gaussian Naive Bayes [**NaivBayes**]
2. K-Nearest Neighbours [**KNearestN**]
3. Decision Tree [**DecsnTree**]
4. Random Forest [**RndForest**]
5. AdaBoost classifier [**AdaBostC**]
6. Support Vector Machine (SVM) with a linear kernel [**LinearSVM**]
7. Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel [**RBFuncSVM**]
8. Multi Layer Perceptron (MLP) [**MLPClassf**]

Figure 2.1 shows a comparison of several of the classifiers presented in this chapter, including the eight selected ones.

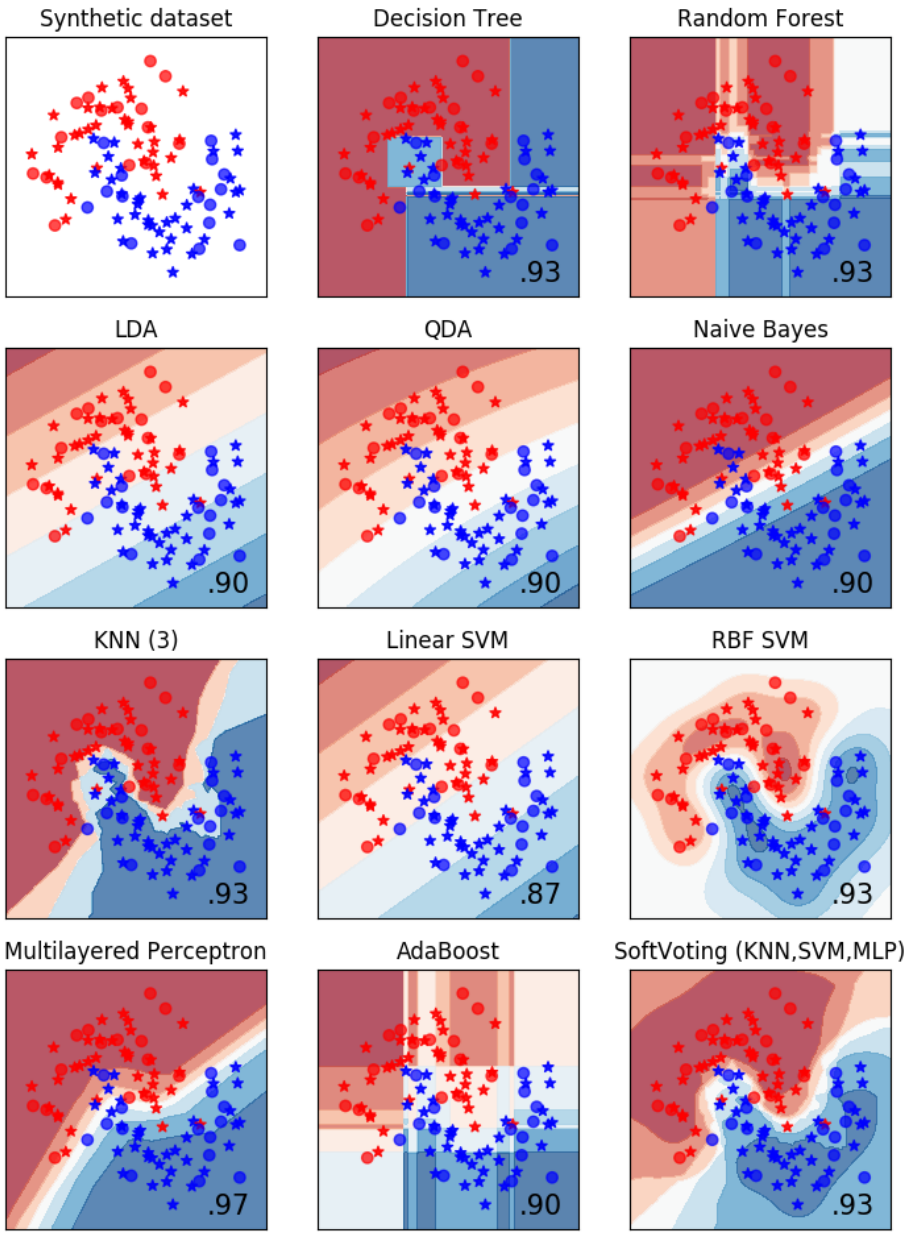


Figure 2.1: *Comparison of 11 classifiers on a 2-class synthetic dataset. The upper left tile shows two classes of data (red and blue) represented by two features* (x, y)*. Samples drawn as stars and circles represent the training and testing sets respectively. The darkness of the colours represents the classifier's confidence regarding the class of a sample. The score achieved by each classifier is shown in the lower right corner of each tile. Figure adapted from [21], the code that produces this image is provided in Appendix B*

2.2 Previous studies involving tumour detection using Machine Learning

The following researchers have introduced the use of supervised Machine Learning (ML) for tumour detection in several modalities. Some of these modalities go beyond the detection of a tumour presence and can assess if the tumour is benign or malignant as well as specifics of the tumour location.

- In 2010, AbdulSadda, Bouaynaya, and Iqbal [22] published a paper where they produced time-domain simulations of microwave scattering on 1025 homogeneous breast models. The backscattering data was pre-processed using Discrete Wavelet Transform (DWT) and then fed to a Genetic Neural Network (GNN). They were able to detect and classify breast tumours as well as determine location, size, depth and spatial orientation by optimizing the number of layers and the weights of the nodes in the neural network.

  Later, in 2013 AbdulSadda [23] published a paper based on the previous one, where he used Simultaneous Perturbation Neural Network (SPNN) instead of a GNN to determine the optimum structure and weights of the network.

- In 2012, Al-Badarneh, Najadat, and Alraziqi [24] published a paper presenting an automatic classification system for tumours in Magnetic Resonance Imaging (MRI). The system used 275 brain MRIs 181 of which, were abnormal while the 94 remaining ones were healthy. The images were 256x256 pixels in size, 278 features were extracted from the images and fed into a K-Nearest Neighbours (KNN) and an Artificial Neural Network (ANN). The author reported

diagnostic accuracies of 98.55 % for both the ANN and the KNN.

- In 2011, Conceição, O'Halloran, Glavin, and Jones published a paper in which she used Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Support Vector Machine (SVM) to classify homogeneous and heterogeneous breast phantoms [25].

  Later, in 2013, Conceição et al. [26] examined the introduction of fibroglandular clusters on breast models from 3D MRIs of breasts taken from the UWCEM Numerical Breast Phantom Repository [27, 28]. She introduced tumours of several shapes and sizes, modelling them using Gaussian Random Spheres (GRS) and employing permittivity values reported by Lazebnik et al. [29, 30]. Electromagnetic scattering was simulated using 3D Finite-Difference Time-Domain (FDTD) with a cubic grid of 0.5 mm resolution. For the classification aspect of the work, features were extracted using Principal Component Analysis (PCA) and a SVM was used as the classifier method.

2.3 Microwave data acquisition prototypes

Microwave based technologies have recently emerged as potential solutions to some of the shortcomings inherent to X-ray mammography. Some of the recently proposed systems based on microwave technology make use of Machine Learning (ML) classifiers to automatically detect the presence of a tumour inside a breast. The most common approach to use microwave technology, however, is still Microwave Imaging (MWI). There are currently two popular approaches to MWI being actively re-

searched. The first one, Microwave Tomography (MWT) consists on reconstructing a map of permittivity values corresponding to a cross-section of a target. This is done by solving non-linear and ill-posed inverse scattering problems [31]. The second approach to MWI is radar-based imaging, which uses the electromagnetic scattering to build a map of reflections from the dielectric contrasts of a target.

In this section, the work of other researchers in the field of tumour detection using microwave technologies is presented. The technical aspect of the technologies chosen by different research groups will be emphasised, namely the frequency ranges used, the type of antennas, the algorithms used as forward solvers, and the hardware employed.

- In 2005, Nilavalan et al. [9] presented a prototype consisting of 16 Ultrawideband (UWB) patch antennas in a flat array configuration. Their work started as a theoretical study employing Finite-Difference Time-Domain (FDTD) simulation models and eventually those results were confirmed with a physical prototype. The prototype acquired data from the antennas, by way of a 16 channel RF-switching matrix. Their array of antennas was used to scan phantoms made of synthetic biological material that included breast tissue, skin and tumour. In 2009, Klemm et al. [31] published a paper presenting a similar array of 16 antennas operating up to 3 GHz, configured in a hemispherical fashion.

- In 2009, Halter et al. [32] published a paper where they detailed a MWT system comprised of 16 antennas operating from 0.6 GHz to 1.7 GHz. A patient lies prone on a bed with one of her breasts inside an aperture, where the circular array of antennas surrounds the patient's breast. Each antenna broadcasts

a microwave signal while the remaining antennas act as sensors. The dielectric properties of the patient's breast is estimated using a 3D Finite Element Method (FEM) forward model coupled with an iterative regularized Newton-Raphson inversion routine. The array of antennas moves vertically to collect measurements at several planes from the chest wall to the nipple.

- In [33], Rubæk, Kim, and Meincke [33] published a paper describing an MWI system and its performance on simulated data. A Method of Moments (MoM) forward solver was used to find the solution of the forward scattering problem. 32 antennas oriented in a cylindrical shape are used to collect the data. The reconstructing method was based on a iterative Newton algorithm

- In 2000, Meaney et al. [34] published a paper where they reported a MWI system consisting of a 16 monopole antenna array operating in the 0.3 GHz to 1.0 GHz range. The data from the antennas are captured using a microwave switching receiver with a capacity of up to 32 channels. The antenna array is placed inside a tank filled with a saline solution that acts as a matching medium between the antennas and the breast skin. Their system has the capability of acquiring data at different heights by moving the antenna array using a hydraulic jack located below the array housing.

- In 2010, Ostadrahimi et al. [35] published a paper presenting a MWI system consisting of 24 modified Double Layer Vivaldi Antenna (DLVA) operating in the 3.1 GHz to 10.6 GHz range. The antennas were designed using the commercial FEM simulation package HFSS, and physical measurements were compared

after fabrication. A 2x24 mechanical switch network was used to connect the antennas to a network analyser. The array of antennas was mounted on a Plexiglas cylinder of 45 cm diameter, equally separated by 15°.

- In 2012, Bourqui, Sill, and Fear published a paper describing a breast MWI system. The data acquisition aspect of their system consists of a bed on which a patient lies prone and places one breast inside a 130 mm diameter opening. The breast is submerged in oil, which acts as a matching medium for the skin interface. A Balanced Antipodal Vivaldi Antenna with a Director (BAVA-D), operating from 2.4 GHz to 18 GHz, rotates inside the oil tank and around the patient's breast while taking microwave measurements at different points until completely encircling the breast. Measured data from this system was validated by comparing it to simulations using an FDTD solver that took into account the breast phantom, the BAVA-D, and the immersion liquid used in the tank [36].

- In 2013, Cao et al. [37] published a paper describing a MWI technique involving solid-state spintronic sensors. These sensors are a few millimetres in size and have the capability of rectifying microwave radiation of both E-field and H-field into a DC voltage. In this study, a standard X-band horn antenna of 8.0 GHz to 12.0 GHz was used to illuminate a target, while the reflected signals were captured with the spintronic sensors 15 cm away from the target. Additionally, this technology showed the capability of working in the near and far fields at microwave frequencies.

2.4 Background on Electromagnetic scattering

The design process of a microwave data acquisition system usually involves simulating the electromagnetic scattering phenomenon that takes place in the system's geometry. Forward solvers such as Finite-Difference Time-Domain (FDTD) ([9, 25, 36]) and Finite Element Method (FEM) ([32, 35]) have been implemented to carry out such simulations.

These forward solvers have advantages and disadvantages: FDTD and FEM require the solution domain to be truncated, while Method of Moments (MoM) does not [38]. MoM operates in the frequency domain with one frequency value point at a time, while FDTD and FEM do so in the time domain and can encompass several frequencies simultaneously. In this thesis, the choice of using MoM as a forward solving method was largely due to the familiarity of the author with this particular algorithm [39]. MoM has been previously used as a forward solver in 2D for arbitrary dielectric materials (Richmond [40]), and more specifically in 3D for breast cancer screening systems (Rubæk, Kim, and Meincke [33]).

$$\mathbf{X}\vec{Y} = \vec{Z} \tag{2.1}$$

$$\mathbf{C}\vec{E}^{tot} = \vec{E}^{inc} \tag{2.2}$$

At its core, MoM finds the unknown values of a column vector $\vec{Y}$ that satisfies Equation 2.1, where $\mathbf{X}$ is a known matrix , and $\vec{Z}$ is a column vector with known values. Specifically to solve the electromagnetic scattering problem, Richmond reformulated Equation 2.1 as Equation 2.2, where $\vec{E}^{inc}$ holds the values of the incident field at every cell within a scatterer in the grid of the solution domain. $\vec{E}^{tot}$ is the unknown

column vector of the values corresponding to the total field at every cell of the scatterer in the grid. **C** is a square coefficient matrix containing each cell's permittivity value (ϵ_r) as well as the distance values between every pair of cells within the scatterer ($\rho_{(m,n)}$). In this context, the scatterer is defined as any cell with permittivity value other than ϵ_0.

$$E^{inc}_{(m,n)} = \begin{cases} (\frac{j}{2})[\pi k a H_1^{(2)}(ka) - 2j], & \text{if } m = n \\ (\frac{j\pi ka}{2})[J_1(ka)H_0^{(2)}(k\rho_{(m,n)})], & \text{if } m \neq n \end{cases} \tag{2.3}$$

Equation 2.3 represents the incident field $E^{inc}_{(m,n)}$ produced by a point source located at grid-cell m, at the location of any cell within the scatterer (n), this type of point source corresponds to the two-dimensional representation of an isotropic antenna. The components of Equation 2.3 are as follows:

- j is the imaginary number $j = \sqrt{-1}$
- k corresponds to the "wavenumber" given by $k = \frac{\omega}{C_0} = \frac{2\pi f}{C_0}$, where C_0 represents the speed of light in vacuum, ω is the angular frequency and f is the frequency value in Hertz.
- a is the radius of a circle with an area equivalent to that of a grid-cell, given by $\frac{cell\_size}{\sqrt{\pi}}$
- $H_1^{(2)}$ is the Hankel function of the second kind of order 1.
- J_1 is the Bessel function of the first kind and order 1.
- $H_0^{(2)}$ is the Hankel function of the second kind of order 2.

- $\rho_{(m,n)}$ is the Euclidean distance between cell m and n.

Richmond computed the coefficient matrix $\mathbf{C}$ as presented by Equation 2.4 The notation in this formula is similar than the previous one, with the addition of ϵ_m and ϵ_n corresponding to the complex permittivity values of every pair of cells m and n within the scatterer.

$$C_{(m,n)} = \begin{cases} 1 + (\epsilon_m - 1)(\frac{j}{2})[\pi ka H_1^{(2)}(ka) - 2j], & \text{if } m = n \\ (\epsilon_n - 1)(\frac{j\pi ka}{2})[J_1(ka)H_0^{(2)}(k\rho_{(m,n)})], & \text{if } m \neq n \end{cases} \tag{2.4}$$

Finally, the total electrical field at the location of each cell within the scatterer $\vec{E}^{tot}$, due to the scattering of the incident field upon every cell of the grid, was computed by solving the System of Linear Algebraic Equations (SLAE) given by Equation 2.2. This section presented a general idea of the inner working of the MoM to solve the electromagnetic scattering problem as implemented by Richmond. The reader is encouraged to reference Richmond's paper [40] for further information regarding the nature of Equations 2.3 and 2.4.

2.5 Background on constituent breast tissues

The permittivity of a material is a physical property that measures its capacity to oppose an electrical field; vacuum possesses the lowest known permittivity value ($\epsilon_0 \approx 8.85 \times 10^{-12}$ Farads per metre). The relative permittivity of a material (relative to that of vacuum) varies as a function of frequency, moisture, temperature and pressure.

However, moisture, temperature and pressure can be assumed to be constant if the measurement takes place in a short period of time [41]. Air possesses a relative permittivity close to that of vacuum, to the point that its difference is often considered negligible [42].

The state-of-the-art method for modelling a material's relative permittivity is fitting previously measured values into a Cole-Cole relaxation model [43]. This equation, often referred to as "Debye relaxation" when the value of α is zero, is presented in equation 2.5 [30]:

$$\epsilon_r = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + (j\omega\tau)^{1-\alpha}} + \frac{\sigma_s}{j\omega\epsilon_0} \tag{2.5}$$

This is of importance, as there is a growing body of work reporting the values of the parameters in equation 2.5 corresponding to different materials. Having such parameters readily available, allows the modelling and simulation of the electromagnetic scattering phenomenon using the properties of known materials. The values for the tissues that constitute human breasts will be discussed in section 3.2.2.

2.6 Relevance of previous work

This chapter has presented some background on the subjects of Machine Learning (ML), Method of Moments (MoM) and its use as a forward solver for the microwave scattering problem. The books, theses and papers mentioned in this chapter are of importance to the design and development of a new Breast Microwave Sensing (BMS) system, this new apparatus was inspired by some of the technical aspects of other researchers' work. Section 3.1 will explore some of the technologies that are

compatible with the purposes of this research.

Chapter 3

Methods and materials

3.1 Prototype design

The purpose of this research is to examine the feasibility of a Breast Microwave Sensing (BMS) system. The design of such system, and the prototype that was built upon that design, had the following constraints:

1. The components that comprise the system should fit in a piece of luggage not bigger than 55 cm × 55 cm × 55 cm.

2. The overall system weight should not exceed 25 Kg.

3. Most breast sizes should be able to be examined by the system.

4. Trained health personnel should not be needed for the system's operation.

The first two constraints address the portability aspect of the system, as one of the design objectives is to provide an apparatus that is easy to move to remote places

by an individual. These constraints also limit the type of technologies available, as some of the systems presented in Chapter 2 make use of beds ([32, 34, 36]), microwave switches ([9, 34, 35]), and tanks for matching media ([34, 36]). Such approaches would conflict with the portable system paradigm, as they are normally big in size and/or weight.

On the other hand, techniques such as rotating an antenna around the breast (as proposed in [36]) and microwave sensors of only a few millimetres in size (as presented by [37]) are more compatible with the portability purpose of our system. Having these techniques in mind, a rotating chamber was designed to hold one transmitting antenna and multiple receiving microwave sensors. The rotating chamber can move within a 48° arc. As shown by [36], using multiple angles to interrogate the breast allowed the collection of data coming from tumours that might otherwise be hidden behind a Fibroglandular Tissue (FGT) patch from the antenna's perspective.

The system described in [36] used a 13 cm diameter aperture for the breast. However, a previous study found that for breasts with D-cup size, the mean breast diameter (± standard error) was 13.7 cm ± 0.2 cm and their mean length in pendant position was 9.7 cm ± 0.2 cm [44]. In the interest of accommodating a broader group of breast sizes (the third constraint of the list), a 15 cm diameter and 15 cm height were chosen for the breast aperture and chamber depth respectively.

The last item in the list of constraints eliminated the possibility of using a Microwave Imaging (MWI) or Microwave Tomography (MWT) approach, as both of these require the interpretation of an expert to give a diagnosis on the presence or absence of a tumour. This constraint was addressed in two ways: a) The use of an

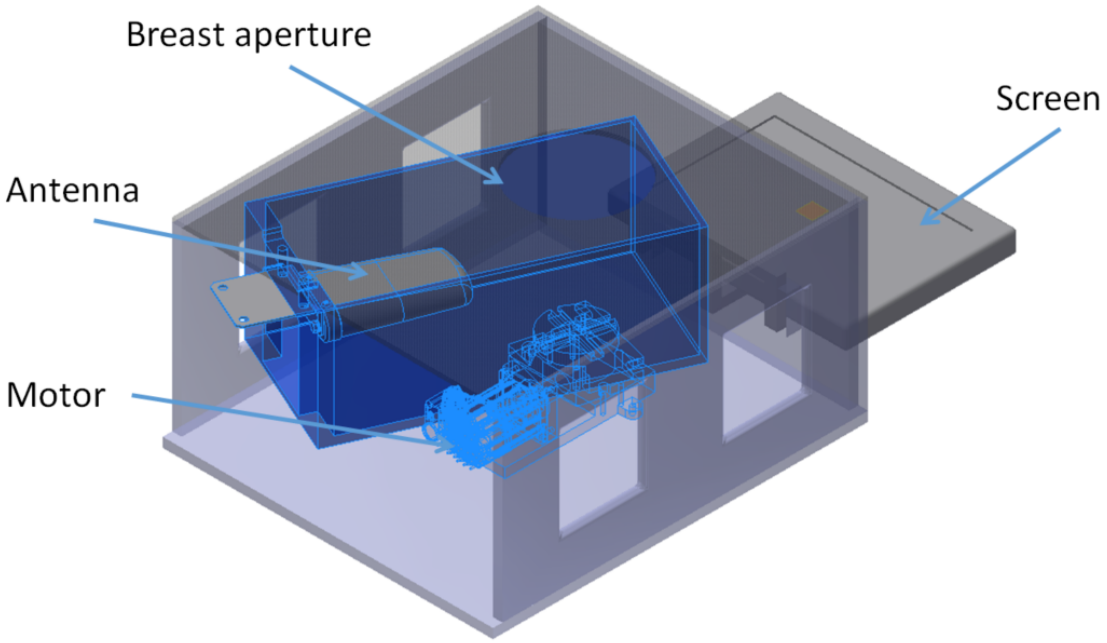


Figure 3.1: *Isometric CAD render of the BMS system. The servomotor enables the rotation of the chamber around the breast while the screen allows a user to follow instructions while a breast is placed in the aperture.*

automatic classification system using Machine Learning (ML) to determine the presence or absence of a tumour, and b) The use of on-screen instructions that guide the patient in using the system from the microwave scanning process until the classifier's prediction on the presence or absence of a tumour.

Figures 3.1 and 3.2 show Computer-Aided Design (CAD) renders of the system designed taking into consideration the constraints.

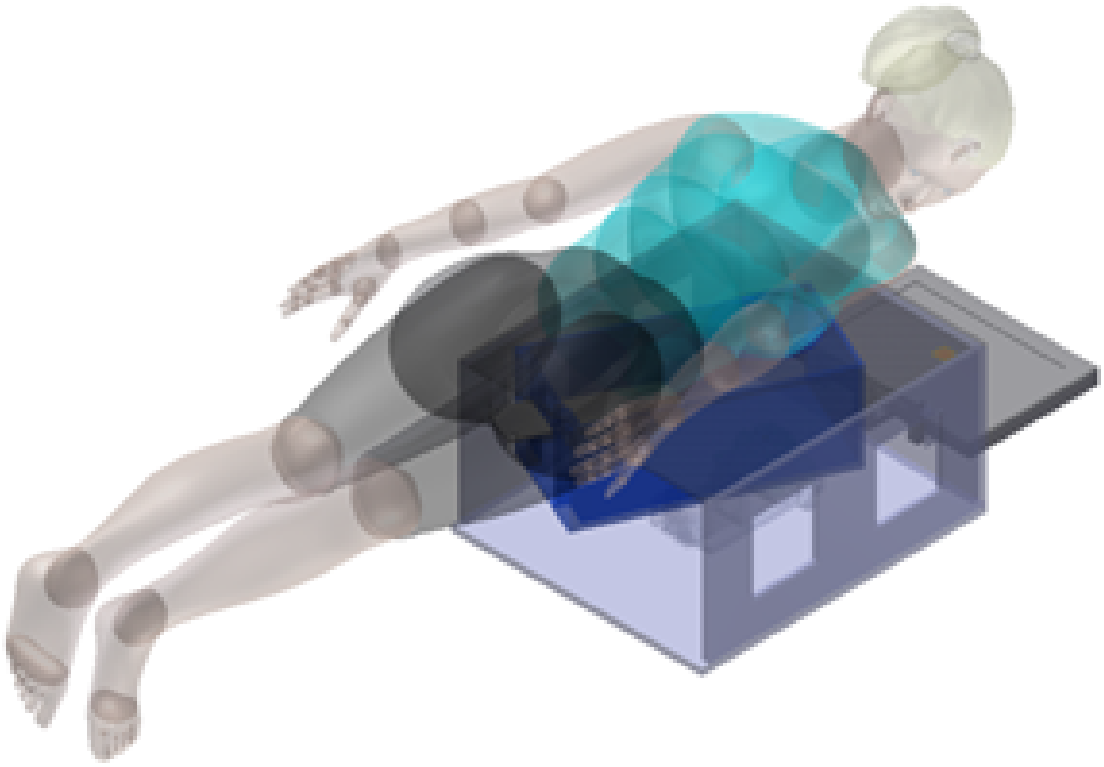


Figure 3.2: *CAD render of a user laying on top of the BMS prototype facing the screen while a breast is inserted in the aperture for microwave scanning. This design is based on several bed systems presented in Chapter 2. This approach requires the breast to be in a pendant state, similar to several bed-based systems presented in that chapter, without the need of having a full-sized bed.*

3.2 Electromagnetic scattering and simulation

3.2.1 Forward Solver Implementation

Section 2.4 laid out the background of Richmond's implementation of the Method of Moments (MoM) as a forward solver for the application of electromagnetic scat-

tering. The 2D algorithm presented by Richmond, has been re-implemented and the results provided by his original publication have been reproduced using the NumPy package for Python [45].

Additionally to the contribution made by Richmond, part of his algorithm has been reimplemented using multi-thread support, specifically, the construction of the coefficient **C** matrix given by Equation 3.1. The elements contained in the **C** matrix are independent to one another. Therefore, the order in which each element was computed had no bearing on any other element of the matrix, this made parallel computing of their values possible.

$$C_{(m,n)} = \begin{cases} 1 + (\epsilon_m - 1)(\frac{j}{2})[\pi ka H_1^{(2)}(ka) - 2j], & \text{if } m = n \\ (\epsilon_n - 1)(\frac{j\pi ka}{2})[J_1(ka)H_0^{(2)}(k\rho_{(m,n)})], & \text{if } m \neq n \end{cases} \tag{3.1}$$

A grid resolution of 2 mm was used for the implementation of MoM, the computational domain was partitioned into square cells. Consequently, the value of a circle's radius with equivalent area to that of a grid-cell, became $a = \frac{2mm}{\sqrt{\pi}}$.

$$A = (cell\_size)^2 = \pi a^2 \tag{3.2}$$

$$\therefore a = \frac{cell\_size}{\sqrt{\pi}} = \frac{2mm}{\sqrt{\pi}} \tag{3.3}$$

Similarly to the value of a, the wavenumber (k) needed to be given a value before the computation of the **C** matrix and the $\vec{E}^{inc}$ vector. Five different frequencies were used for the simulations (discussed in section 3.2.2). This required the computation of five different wavenumbers given by $k = \frac{2\pi f}{C_0}$ where C_0 is the speed of light in vacuum

and f is the frequency value in Hertz.

3.2.2 Breast tissues and experiment design

Section 3.1, showed the design of a Polyvinyl chloride (PVC) chamber where a transmitting antenna and several microwave sensors were fixed. In the absence of any external material inside this chamber, the microwave sensors could only detect the electromagnetic field produced by the transmitting antenna and scattered by the PVC walls of the chamber. However, as external material was introduced into the chamber, the electromagnetic field was perturbed, and the values logged by the sensors were affected. The degree to which electromagnetic field scattering occurs is caused by a difference between the relative permittivity ϵ_r of the introduced material and its surrounding medium, which in this case is air. The contrast in permittivity exhibited by healthy and malignant tissues is relevant for tumour detection purposes, as it produces electromagnetic scattering patterns that were analysed by various Machine Learning (ML) techniques discussed in section 3.3.

$$\epsilon_r = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + (j\omega\tau)^{1-\alpha}} + \frac{\sigma_s}{j\omega\epsilon_0} \tag{3.4}$$

Two-dimensional numerical models of breasts were simulated as collections of circular patches made out of dielectric materials with permittivity values previously reported by other researchers ([29, 46, 47, 41]). As discussed in section 2.5, the permittivity of tissue can be modelled as a Cole-Cole curve, given by Equation 3.4.

Table 3.1: *Cole-Cole parameters of the materials used for Electromagnetic scattering simulations. The values presented in this table were inserted into Equation 3.4 to obtain permittivity as a function of frequency for each tissue type in a breast model.*

Material	ϵ_s	ϵ_∞	α	$\tau(ps)$	σ_s	reference
Skin	36.33	4.0	0.0	6.9	1.4	[47]
Fatty tissue	10	7.0	0.0	7.0	0.15	[46]
Fibroglandular Tissue (FGT)	21.57	6.14	0.0	7.0	0.31	[46]
Malignant tumour	56.839	6.749	0.051	10.5	0.794	[29]

Figure 3.3 presents the permittivity values and their variation with frequencies between 1 GHz and 11 GHz. Electromagnetic scattering experiments were simulated taking the cross-section of a breast model into account, as well as the PVC walls that make up the chamber. Figure 3.4 shows a cross-section of the chamber and a model of a breast including malignant breast tissue, the details of the dimensions and positions of the different tissues that make up a breast are presented in Section 3.2.3.

The location of 12 solid state sensors and a transmitting antenna were fixed with respect to the chamber's PVC walls. The experiment simulations in this thesis assume the use of sensors smaller than the grid-size of 2 mm, similar to those previously used by [37]. The volume taken by the sensors is small enough that their contribution to the electromagnetic scattering is not considered as each sensor's cross-section would occupy less than one cell in the grid. Five frequencies were used: 2.3 GHz, 3.35 GHz, 4.4 GHz, 5.45 GHz and 6.5 GHz. These frequency points were chosen to cover the bandwidth supported by the antenna that was used in the physical prototype.

The PVC chamber was designed to be capable of rotating around the breast. As a result, information about the scattering patterns of a breast model, can be

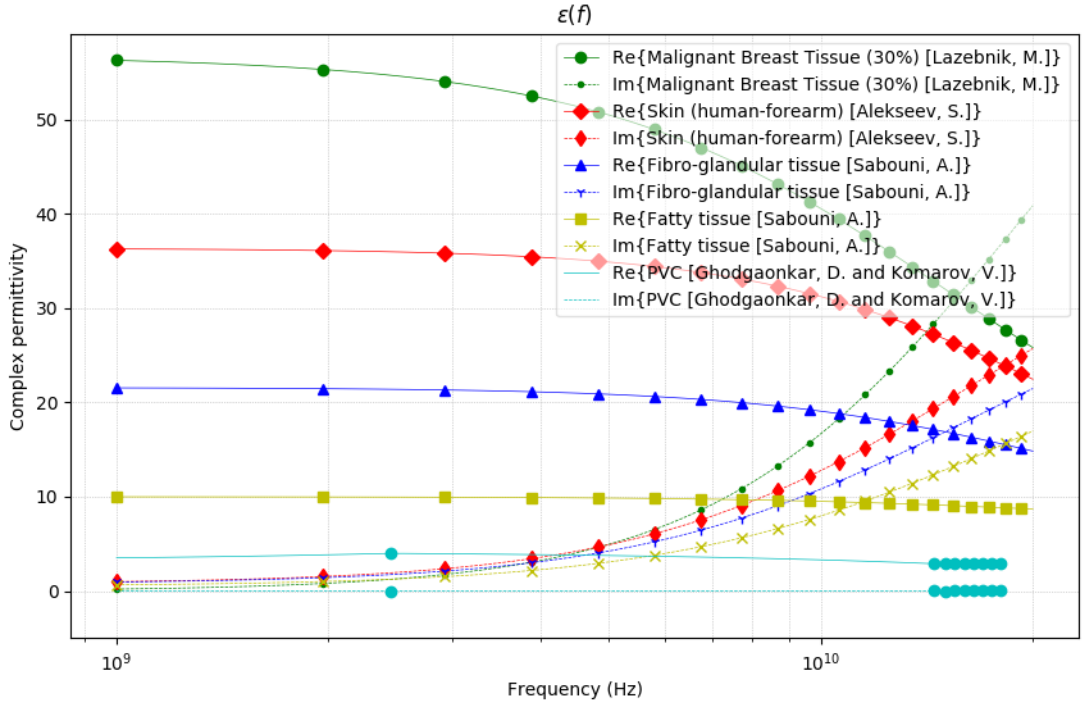


Figure 3.3: *Complex permittivity values as functions of frequency for the different tissues used to build breast models across five datasets. Solid lines represent the real part of the complex permittivity, while the dotted lines indicate the imaginary part. These curves were produced by plugging the Debye parameters from Table 3.1 into Equation 2.5. The Debye parameters corresponding to each tissue were taken from references between squared brackets in the legend [29, 47, 46, 48, 41].*

captured from several angles. For this reason, two types of experiment were carried out. The first type of experiment placed the chamber in the middle position of the 48° arc, and logged the electromagnetic scattering values from that position alone. The second type of experiment accounted for five angles within the 48° arc, capturing electromagnetic scattering information at each one of those angles. The combination

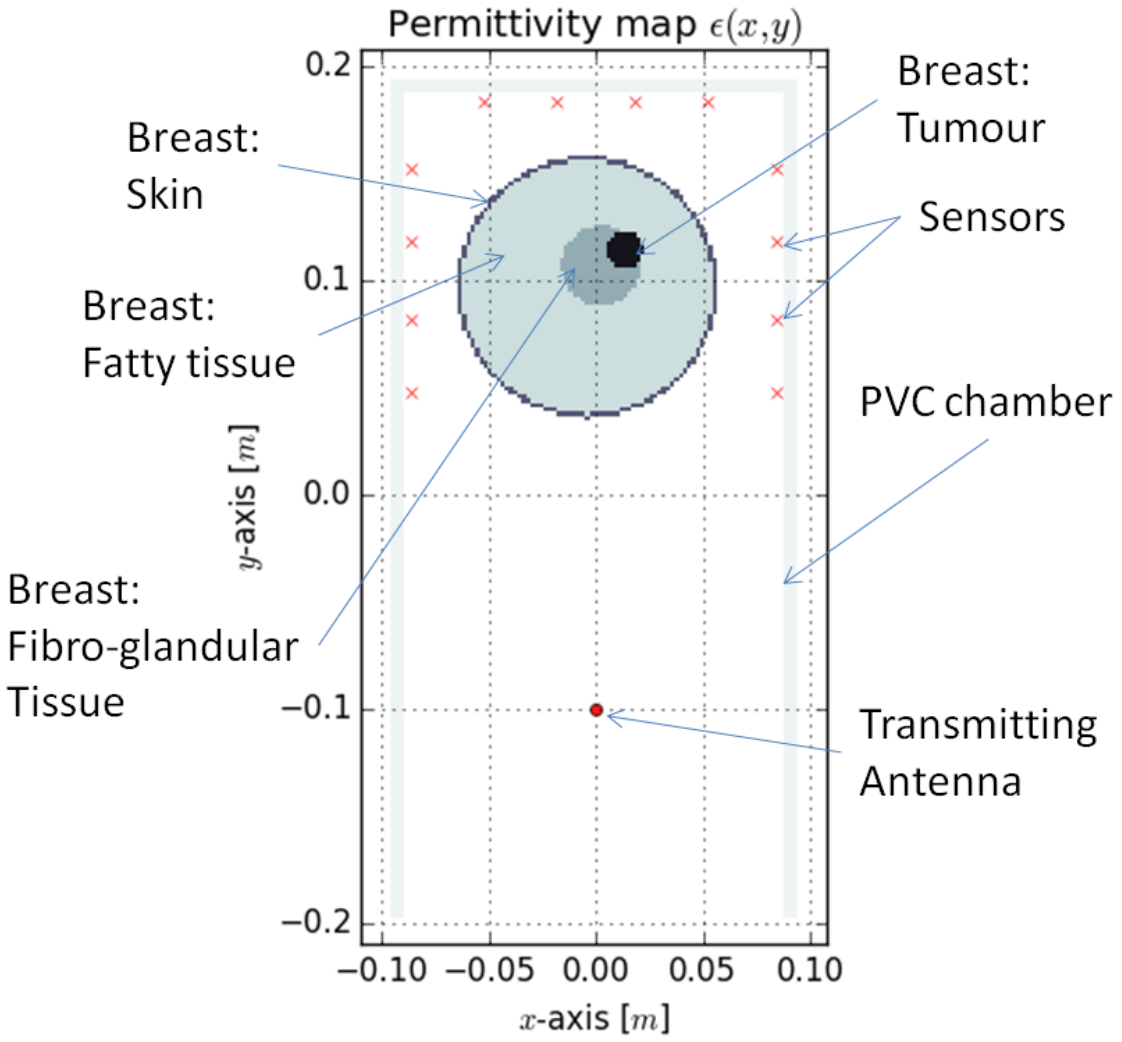


Figure 3.4: *Cross-section view of the BMS chamber simulation and a breast model consisting of skin, fatty tissue, FGT and a tumour. The spatial resolution of the models is 2mm. The permittivity values of the employed materials depend on the transmitting frequency of the antenna, these permittivity values have to be computed once for every frequency used in the experimental simulations.*

of 12 sensors and five frequency points yielded a vector of 60 values (referred to as "features" in ML terminology) in the case of non-rotating experiments, and 300 values in those that include five rotational positions. These 60 or 300 feature vectors represented a "sample" that was labelled as positive when a tumour was present in the model or negative otherwise.

A group of simulated experiments with similar characteristics in terms of the composition of their breast models and feature vector sizes, is called a dataset. Five different datasets were created for the analysis of this thesis. The first two datasets involve no rotation of the PVC chamber and the details of these datasets are presented in section 3.2.4. The three remaining datasets are derived from experiments involving five rotational positions of the chamber. However, one of these datasets discards information from four rotational positions. These datasets are presented in in section 3.2.5.

Simulating the electromagnetic scattering in the different models is a time-consuming process, as the contribution of each cell to the scattered field value of every other cell in the grid needs to be calculated. New breast models for each dataset were computed until adding samples resulted in less than a 1% improvement in either of the four classification metrics. Table 3.2 summarizes the modelled datasets in terms of their respective attributes and tissue contents.

Table 3.2: *Summary of characteristics of the modeled datasets*

Name	Rotation positions	Samples	Features	Samples/ features ratio	Material: Fat	Skin	Tumour	FGT
hom1b1p	1	1534	60	25.57	✓	✓	✓	-
het1b1p	1	2394	60	39.90	✓	✓	✓	✓
het1b5pv2Alpha	5†	7000	60	116.67	✓	✓	✓	✓
het1b5pv2Beta	1⋆	1400	60	23.33	✓	✓	✓	✓
het1b5pv2Gamma	5	1400	300	4.67	✓	✓	✓	✓

†: Each of the 5 positions are treated as independent samples in this modality.

⋆: Only 1 out of 5 positions was used in this modality.

3.2.3 Breast modelling using random variables

Breast models were built as collections of circular patches of different tissues with known permittivity values. To account for differences between breast models, random sizes and positions of the patches were used. The following limitations were imposed unto the random variables:

1. The overall breast radius (R_B) should not exceed the radius of the Breast Microwave Sensing (BMS) system's breast cavity (75 mm).

2. The breast centre could have an offset distance (O_B) with respect to the system's breast cavity centre, provided that this offset distance does not place the breast outside of the breast cavity.

3. The skin thickness (T_S) remains constant with a value of 2 millimetres.

4. FGT patches would not be outside of the skin radius. Therefore the FGT patch radius (R_{fgt}) could not exceed the radius of the fatty tissue ($R_B - T_S$).

5. The tumour tissue could not be outside of the skin radius.

6. The centre of a tumour and the FGT patch could have an offset distance with respect to the breast centre (O_T and O_{fgt} respectively).

Having considered the previous caveats, the first random variable that was determined for every breast model was the overall breast radius (R_B). R_B was given random values from a uniform distribution ranging from 50 mm to 75 mm to account for different breast sizes.

The second random variable to be assigned a value was the breast offset distance (O_B) with respect to the BMS system's breast cavity centre. O_B was sampled from a uniform distribution ranging from 0 mm (No offset) to a maximum of the remaining space between the breast and the system's cavity. The distance between the breast skin and the edge of the system's cavity depends on the breast radius, for this reason, the value of R_B was taken into consideration. An upper bound of (75 mm - R_B) ensured the breast was placed within the confines of the breast cavity.

A third random variable θ_B was sampled from a discrete uniform distribution ranging from 0° to 359°. θ_B was used to compute the x, y coordinates of the breast's centre point by following Equations 3.5 and 3.6.

$$x = O_B cos(\theta_B) \tag{3.5}$$

$$y = O_B sin(\theta_B) \tag{3.6}$$

At this point, a permittivity matrix associated with each breast model only contained values of the empty space (ϵ_0) and the PVC's permittivity corresponding to walls of the system's chamber. Once the breast radius was chosen and the x, y coordinates of the breast's centre were computed, the permittivity matrix was updated by overwriting the values of cells corresponding to the circle with centre (x, y) and radius R_B. The process of determining values for a radius, an offset distance and angle to compute x, y coordinates for a circle, was repeated for the rest of the circular patches that compose the breast model. In the case of the fatty tissue, the radius was given by $R_B - T_S$ to comply with the limitations outlined previously. Tables presented in sections 3.2.4 and 3.2.5 address the particular distribution ranges used for selecting position and sizes of the FGT patch and the tumour for the different datasets.

3.2.4 Datasets of a single antenna position

As presented in section 3.1, the transmitting antenna and the microwave sensors of the BMS system prototype are contained inside a PVC chamber. The chamber was designed to rotate around the breast cavity within an arc of 48°. The current section presents two datasets comprised of experiments that make use of data collected from a single position of the rotating chamber.

The first dataset was given the codename **hom1b1p**, indicating that the inside of the breast is homogeneous in the sense that it contains only adipose tissue. 50% of the models include a tumour with an uniform distribution of radius ranging from 2 mm to 10 mm. The lower limit of this range was chosen to match the grid resolution of the simulations: 2 mm. 1534 breast models were computed for this dataset.

The second dataset, called **het1b1p**, was produced similarly to the first one, with the additional presence of a circular FGT patch of a size such that the fibroglandular-to-fatty tissue ratio (also known as fibroglandular density) was uniformly distributed from 0 to 25 %. Breasts with fibroglandular densities in this range are known as Breast Imaging Reporting and Data System (BI-RADS) class 1, or "almost entirely fat" [49]. Equation 3.7 presents the equivalence of FGT density, where R_{fgt} is the radius of the FGT patch, and R_B is the breast radius.

$$FGTDensity = \frac{\pi(R_{fgt})^2}{\pi(R_B - T_S)^2} \leq 0.25 \tag{3.7}$$

$$R_{fgt} \leq \sqrt{0.25(R_B - T_S)^2} = 0.5(R_B - T_S) \tag{3.8}$$

Limiting the fibroglandular density to 25 % led to the upper limit of the uniform distribution of R_{fgt} being given by Equation 3.8. 2394 models were created for the het1b1p dataset using this upper limit. A summary of the different random variables employed in the creation of the single antenna position datasets is presented in Table 3.3.

Table 3.3: *Random variables used in the hom1b1p and het1b1p datasets.*

Variable	Minimum	Maximum	Description
R_B	50 mm	75 mm	Breast radius
O_B	0 mm	75 mm - R_B	Breast offset
R_{fgt}	0 mm	$0.5(R_B - T_S)$	FGT radius
O_{fgt}	0 mm	R_B - R_{fgt} - T_S	FGT offset
R_T	2 mm	10 mm	Tumour radius
O_T	0 mm	R_B - R_T - T_S	Tumour offset

3.2.5 Datasets of multiple antenna positions (het1b5pv2)

A third dataset **het1b5pv2** was created with two major differences from the previous one. The first difference consisted in changes of the uniform distributions of random variables. The second major difference consisted in the collection of electromagnetic data from five rotational positions of the transmitting antenna with respect to the breast cavity; each rotational position was separated by 12° within a 48° arc.

The adjustments made to the random variables were made to limit variation and edge cases, such as a breast having a FGT patch completely separated from the breast centre and contacting the skin tissue. The maximum values for the breast offset with respect to the centre of the chamber (O_B), and the offset of the FGT (O_{fgt}) were adjusted in this dataset to only allow 50% of what the previous datasets allowed. The tumour radius (R_T) minimum and maximum were given values of 4 mm and 12 mm respectively, allowing lager tumours. Finally, the tumour offset (O_T) ranges from 0 to the radius of the FGT (R_{fgt}).

Figure 3.5 shows an example of a breast model in this dataset along with its five positions. Table 3.4 presents the upper and lower limits of the uniform distributions used for this dataset. A total of 1400 samples were produced in this dataset.

Table 3.4: *Random variables used in the rotation-based datasets (alpha, beta, gamma).*

Variable	Minimum	Maximum	Description
R_B	50 mm	75 mm	Breast radius
O_B	0 mm	$\mathbf{0.5}(75mm - R_B)$	Breast offset
R_{fgt}	0 mm	$0.5(R_B - T_S)$	FGT radius
O_{fgt}	0 mm	$\mathbf{0.5}(R_B - R_{fgt} - T_S)$	FGT offset
R_T	**4 mm**	**12 mm**	Tumour radius
O_T	0 mm	$\boldsymbol{R_{fgt}}$	Tumour offset

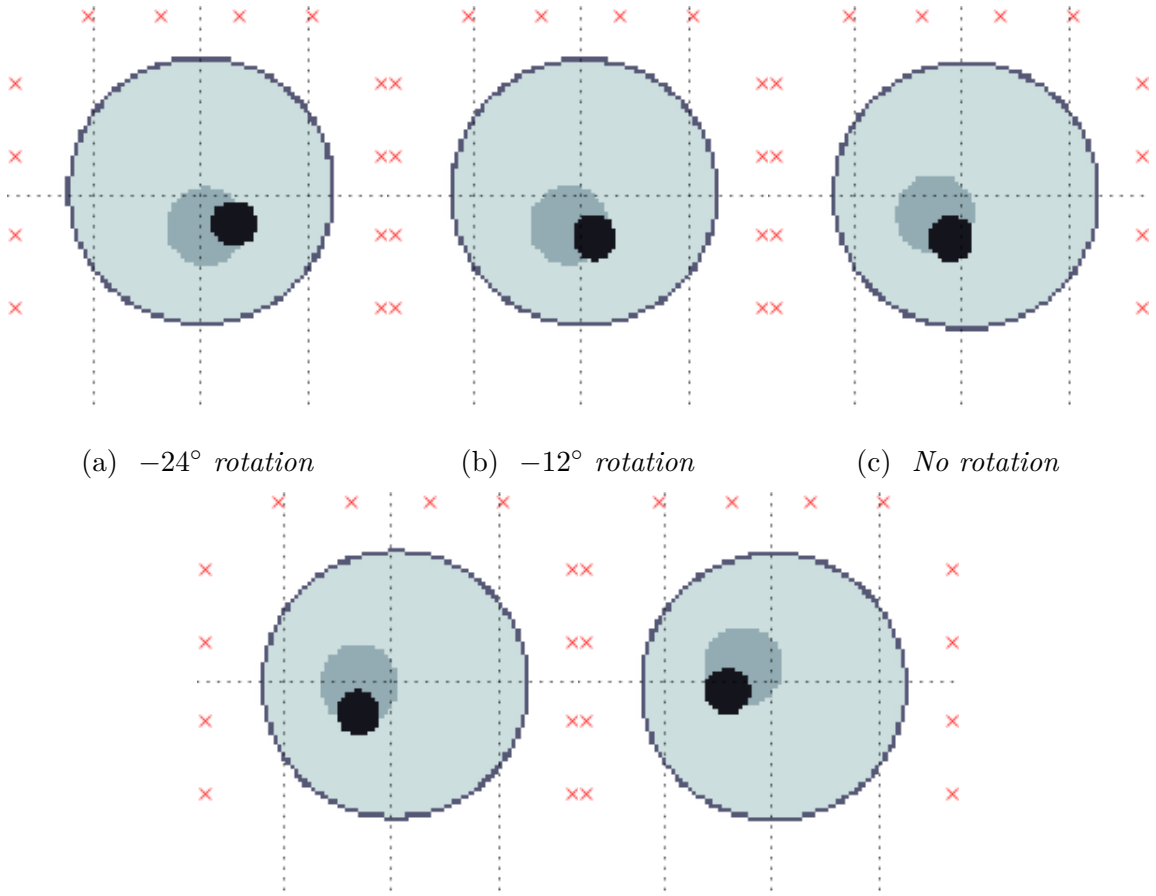
By having data captured at five different transmitting antenna positions, the feature vector size increased from the original 60 to 300 values. The increase in the number of features prolonged the time needed to compute each experiment, while decreased the samples-to-features ratio. This dataset was further divided into three different modalities that produced different ratios:

1. **Alpha**: This modality assumed each one of the five rotated positions represented a different experiment, effectively turning the feature vector size back to 60. A total of 7000 experiments are contained in this modality, bringing the samples-to-features ratio to 116.67.

2. **Beta**: The second modality discards four out of the five positions and all their data, only the middle position is used in this modality. Thus, the number of features is still 60 in this modality. However, the number of samples becomes 1400 due to the discarding of models. This brings the samples-to-features ratio to 23.33.

   This modality is similar to the het1b1p dataset in the sense that both contain fat, skin, FGT and tumours. However, the distributions of the tumour sizes and of the offset of FGT and tumorous tissue is different, as indicated by Tables 3.3 and 3.4.

3. **Gamma**: The third modality treated data from the five positions as a single sample, while this brought the number of features to 300 and kept the number of samples at 1400, it also brought down the samples-to-features ratio to 4.67.



(a) $-24°$ *rotation* (b) $-12°$ *rotation* (c) *No rotation*

(d) $12°$ *rotation* (e) $24°$ *rotation*

Figure 3.5: *Breast models corresponding to an experiment from the het1b5pv2 dataset. The sizes and positions of the circular patches were randomly determined once, then the whole breast model was rotated with respect to the chamber's centre into five different positions with an angular separation of 12 degrees. The antenna and the 12 sensors (represented with red crosses) remain static with respect to the chamber.*

3.3 Machine Learning

Eight classifiers from the scikit-learn python package [50, 51] were used to classify the samples from the five different datasets presented in the previous section. Each classifier was trained and applied a total of 100 times on each dataset, the resulting metrics were averaged and are presented in Chapter 4. Figure 3.6 shows an example of the four metrics resulting from running a classifier 100 times, the dashed lines represent the averaged score of each metric.

The four metrics employed to evaluate the performance of a classifier were the following:

1. Diagnostic accuracy: Measures the fraction of samples correctly classified, presented in Equation 3.9

2. True positive rate (also called **sensitivity** or recall): Measures the fraction of positive samples correctly classified, presented in Equation 3.10

3. True negative rate (also called **specificity**): Measures the fraction of negative samples correctly classified, presented in Equation 3.11

4. Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC).

$$DA = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.9}$$

$$TPR = \frac{TP}{TP + FN} \tag{3.10}$$

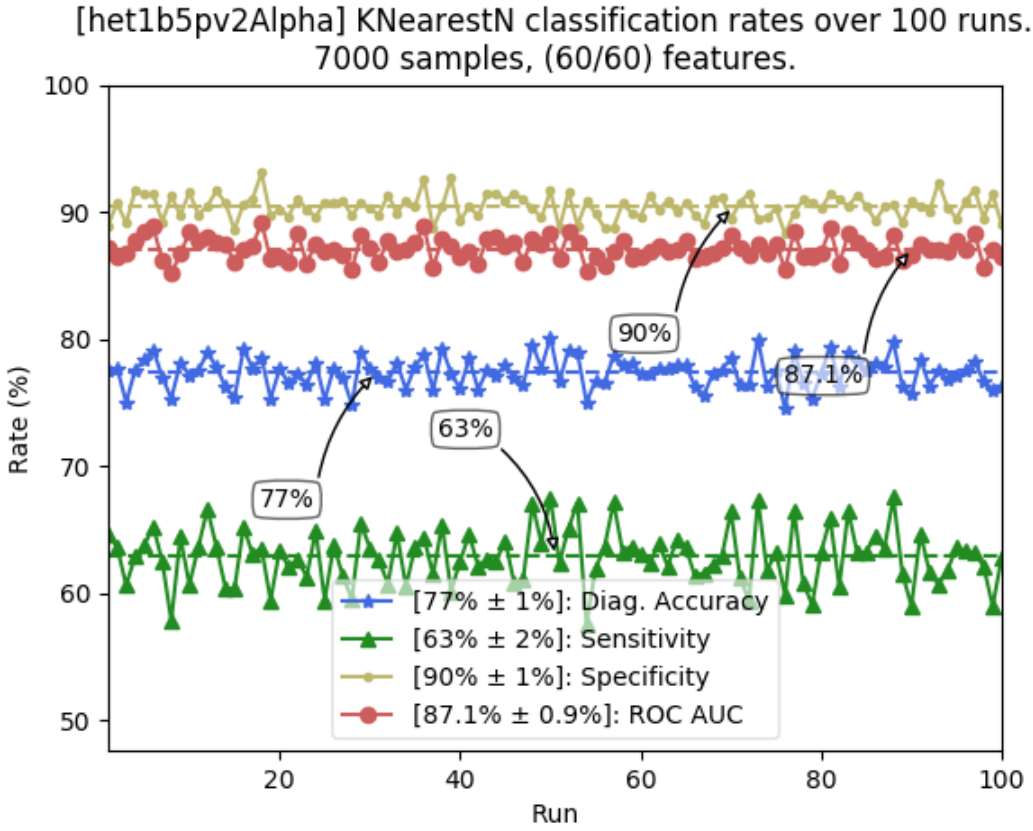


Figure 3.6: *Example of the four performance metrics obtained from running a classifier 100 runs on a dataset. This particular graph shows the KNearestN classifier on the het1b5pv2Alpha dataset.*

$$TNR = \frac{TN}{TN + FP} \tag{3.11}$$

The first three metrics were calculated using the number of samples that fell in the following categories:

- TP: Number of samples correctly classified as positives (true positives).
- TN: Number of samples correctly classified as negatives (true negatives).

- FP: Number of samples incorrectly classified as positives (false positives).
- FN: Number of samples incorrectly classified as negatives (false negatives).

These metrics are all important on their own, and the performance of the different classifiers in terms of those metrics can be found in Chapter 4. However, the goal of the training process was to maximise the value of the AUC, as ROC curves account for both sensitivity and specificity simultaneously. Figure 3.7 shows an example of a ROC curve, where the x-axis corresponds to the false positive rate (1 - specificity) and the y-axis corresponds to the true positive rate (sensitivity). ROC curves also have the property of accounting for skewed class distributions [52, 53, 54]. In general, striving for a maximal value of the AUC achieved high values across the other three metrics without compromising one or the other.

3.3.1 Classifier training process

A process was written to shuffle the samples from any dataset and divide them into a training set (80%) and a testing set (20%). Then, data from the training set was fed to a Principal Component Analysis (PCA) procedure, and its result was used to train the classifiers. Once the classifiers were trained, they classified data coming from the testing set. The training process in some of the classifiers included the use of a Grid-Search Cross-Validation (GSCV) procedure, where one or more parameters were given different values until a maximum performance was obtained. These steps were repeated for each dataset and each of the classifiers. Figure 3.8 presents a flow chart illustrating the steps involved in the training process.

The ML package scikit-learn provided a unified Application Program Interface

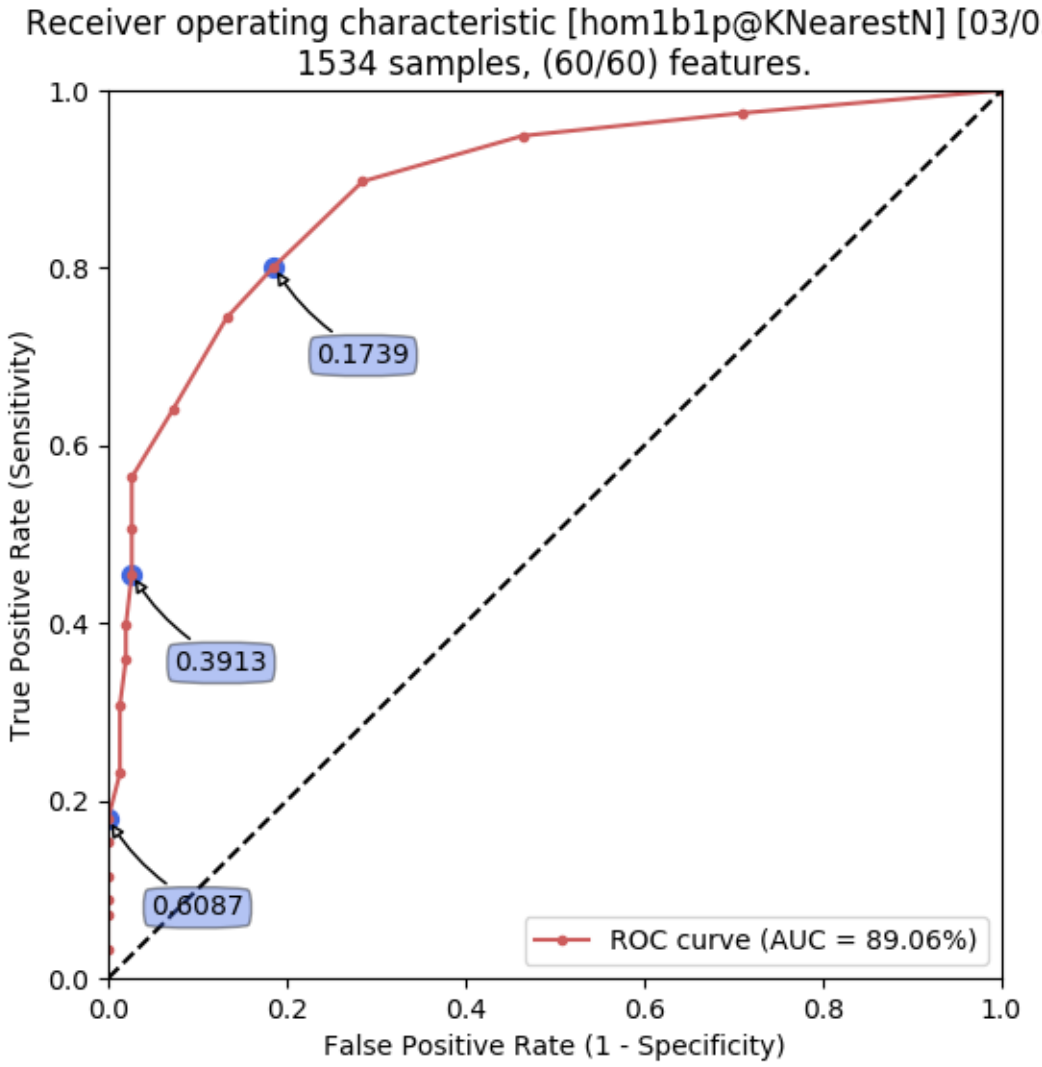


Figure 3.7: *Example of a ROC curve produced by a KNN classifier. The dashed line represents an AUC of 50% which is what can be expected from random guessing. The curved line represents the performance of this classifier. A perfect classifier would have the curved line covering the entire area, for that reason, high performance classifiers have the curved line close to the top left corner.*

(API) for the chosen machine learning methods [55], the flexibility of this package permitted the use of different classifiers in the same process. The classifiers selected from the scikit-learn package are presented in table 3.5, with code names included between parentheses. These code names are used to refer the classifiers in subsequent figures and tables.

3.3.2 Principal Component Analysis

PCA is a method that maps a dataset of M features into a new space of the same size, where the new features (called principal components) are orthogonal to each other. The principal components of the new space are presented in order of decreasing variance, and by only using the first few principal components out of the total M, a high percentage of the variance can be conserved, while the dimensionality can be reduced [56]. Dimensionality reduction can be important as it can have the desired effect of decreasing the computational time demanded by a classifier. An example of reducing dimensionality while keeping most of the variance is given in [57], where it was shown that applying PCA into the classical "Iris flower dataset" by Fisher [12], the first principal component accounted for 92.5% of the variance in the original dataset.

The PCA implementation used in this thesis had no dimensionality reduction, as it was found that this had a negative impact on classification performances. However, even without dimensionality reduction, the use of PCA helped raise the classification performances by transforming the original features into principal components ordered by decreasing variance.

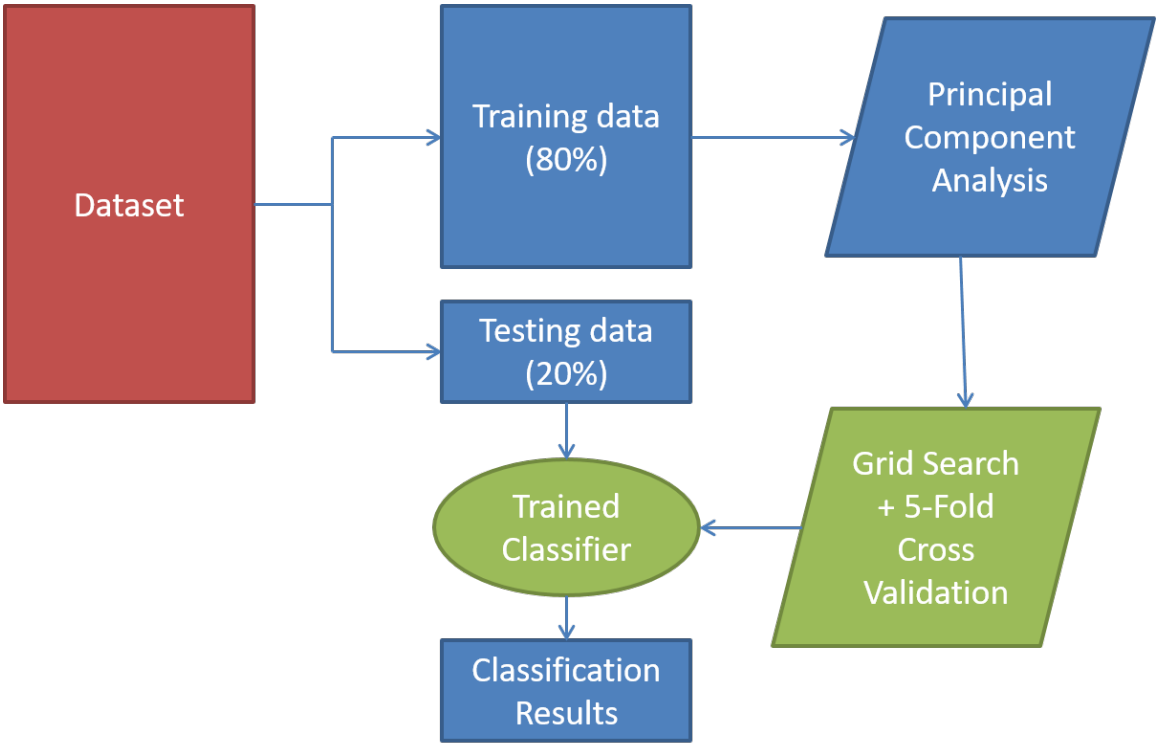


Figure 3.8: *Flow chart of the process developed to allow the use of several classifiers on a dataset. This process facilitated the comparison of performance metrics obtained by each classifier on a dataset. The different classifiers (presented as green blocks) were given as a list upon which the process iterates. A similar approach was implemented for datasets (presented as a red block), where a list of the different datasets was provided and every classifier will be trained and evaluated for a particular dataset. The blue blocks represent parts of the process that required no user input and remain the same across every combination of dataset-classifier. The parameters that GridSearch tries to optimize, vary according to each classifier, and the parameter spaces were passed along as options.*

3.3.3 Parameter optimization using Grid-Search Cross-Validation

Seven out of the eight selected classifiers required one or more parameter values to be specified in order to run. The choice of value for each classifier's parameter had an impact on the classification performances, and it was not known in advance which values would produce the best performances for a given dataset.

GSCV is a method that given a grid of discrete parameter values, searches for those that achieve the highest classification results. The cross-validation aspect of GSCV prevents over-fitting by dividing the training set into K subsets (also known as "folds") and testing them individually against the rest of the subsets [58]. $K = 5$ was selected as the number of folds for cross-validation, this number was chosen as a balance between computational time and performance.

Figure 3.9 presents an example of a K-Nearest Neighbours classifier. Every sample was compared to a fixed number of neighbours before making a prediction. The number of neighbours is the only parameter that needs to be determined in the case of 'KNearestN'. The remainder of the graphs showing tuned parameters can be found in Appendix A

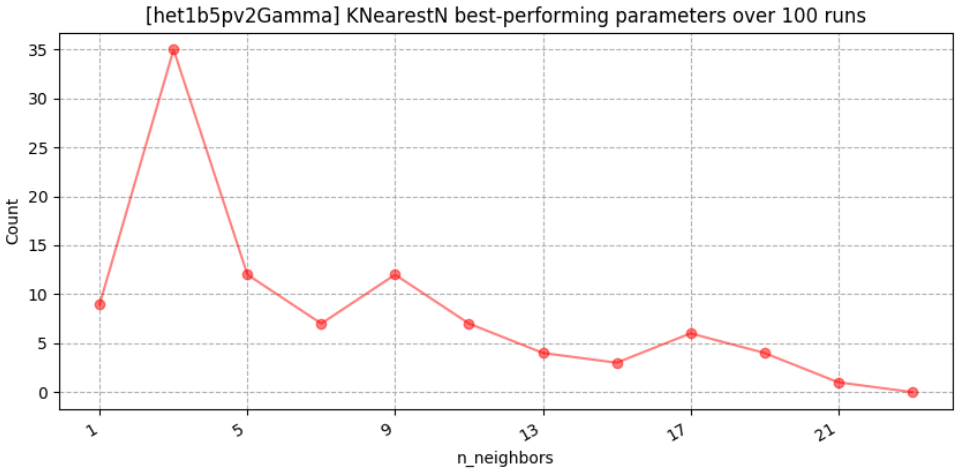


Figure 3.9: *Best performing parameter found by Grid-Search after 100 runs of the 'KNearestN' classifier on the 'het1b5pv2Gamma' dataset. Each point represents the number of times a parameter value was found to be the best performing one for a particular shuffle in the given dataset. In this case, n_neigbors = 3 was found to be the value that was repeated the most times throughout the 100 runs.*

Table 3.5: *Classifier algorithms and their parameters search space used by Grid-Search Cross-Validation*

Classifier	Parameter	Range of values
Gaussian Naive Bayes (NaivBayes)	-	-
K-Nearest neighbors (KNearestN)	n_neighbors	Odd integers from 1 to 23
Decision Tree (DecsnTree)	max_depth	Integer numbers from 1 to 11
Random Forest (RndForest)	max_depth n_estimators max_features	Integers from 2 to 10, inclusive. Integers from 2 to 17, inclusive. Integers from 2 to 10, inclusive.
Ada Boost (AdaBoostC)	n_estimators	Integers ranging from 20 to 49
SVM with linear kernel (LinearSVM)	C	3.33×10^{-7} to 1000, logarithmic
SVM with RBF kernel (RBFuncSVM)	C gamma	2.5×10^{-7} to 250, logarithmic 0 to 1000, logarithmic
Multi-layer Perceptron (MLPClassf)	hidden_layer_sizes max_iter tol	Fixed 4 layers: (400, 200, 100, 50 nodes) Fixed value of 500 iterations Fixed tolerance value of 1×10^{-5}

Chapter 4

Results

This chapter presents the training times and classification metrics achieved by each classifier on the different datasets. The elapsed times and metrics presented in this chapter have the form: $\mu \pm \sigma$, where μ stands for the average and σ stands for the standard deviation. The training process was run 100 times in a multi-user server running *Scientific Linux 6.1* with an *Intel® Xeon® X5680* processor (24 cores running at 3.33GHz) and 96 GB of installed Random Access Memory (RAM).

Table 4.1: *Average elapsed training times for the eight classifiers on the five datasets*

Classifier	Training time in seconds ($\mu \pm \sigma$)				
	hom1b1p	het1b1p	het1b5pv2Alpha	het1b5pv2Beta	het1b5pv2Gamma
AdaBoostC	4.2 ± 0.1	5.8 ± 0.2	16.1 ± 0.3	4.6 ± 0.2	4.8 ± 0.2
DecsnTree	1.14 ± 0.08	1.51 ± 0.09	2.0 ± 0.1	1.80 ± 0.07	2.01 ± 0.07
KNearestN	2.0 ± 0.2	3.1 ± 0.1	46 ± 4	2.4 ± 0.2	2.7 ± 0.2
LinearSVM	300 ± 90	800 ± 100	4000 ± 1000	800 ± 100	700 ± 200
MLPClassf	3.2 ± 0.3	5 ± 2	6 ± 2	2.9 ± 0.3	2.3 ± 0.4
NaivBayes	0.24 ± 0.04	0.24 ± 0.05	0.27 ± 0.06	0.29 ± 0.07	0.30 ± 0.09
RBFuncSVM	74.6 ± 0.4	203 ± 1	7000 ± 900	65.8 ± 0.6	66.7 ± 0.6
RndForest	21 ± 3	28 ± 4	44 ± 1	20 ± 3	21 ± 3

Table 4.1 shows the averaged time taken to train each one of the eight classifiers on the five different datasets. Out of the different classifiers used, "RBFuncSVM" and "MLPClassf" consistently showed either similar or higher performance metrics when compared to the rest of the classification algorithms. The performance of the different datasets, however, varies significantly.

The following sections show the performance results of the different classifiers used, represented by four metrics: diagnostic accuracy, true positive rate (sensitivity), true negative rate (specificity), and the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC). The classifiers are presented in descending order in terms of the AUC.

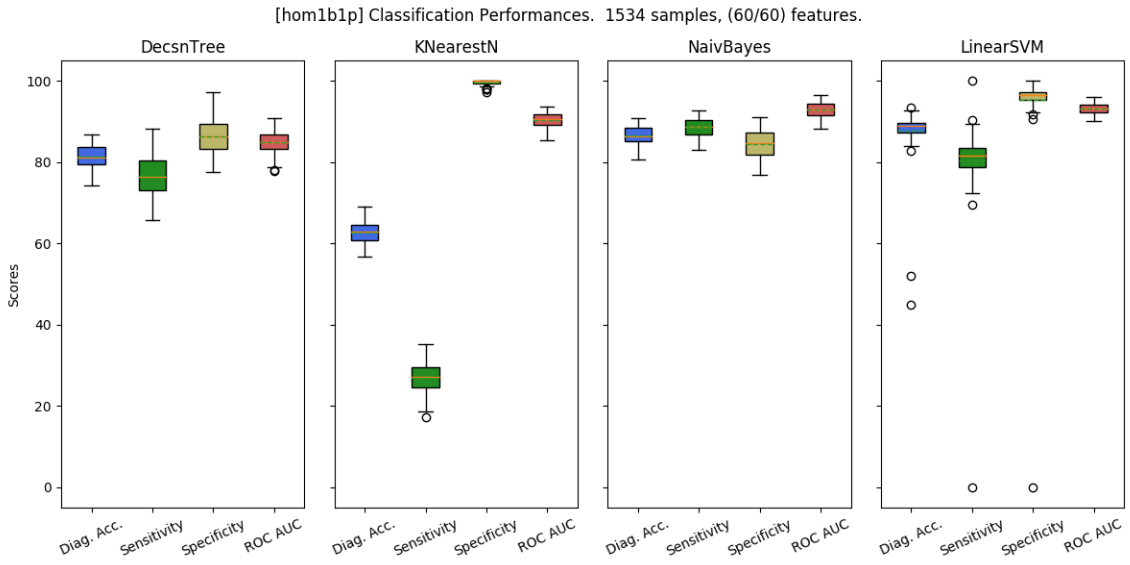
4.1 Classification results of the homogeneous dataset (hom1b1p)

This dataset contained 1534 samples and 60 features. The homogeneous dataset was the least complex in terms of its constituent materials as it did not contain fibroglandular patches. The AUC performances obtained for most of the classifier were higher than 90%. These comparatively higher results are consistent with what is expected from classifying a dataset of homogeneous breast models.

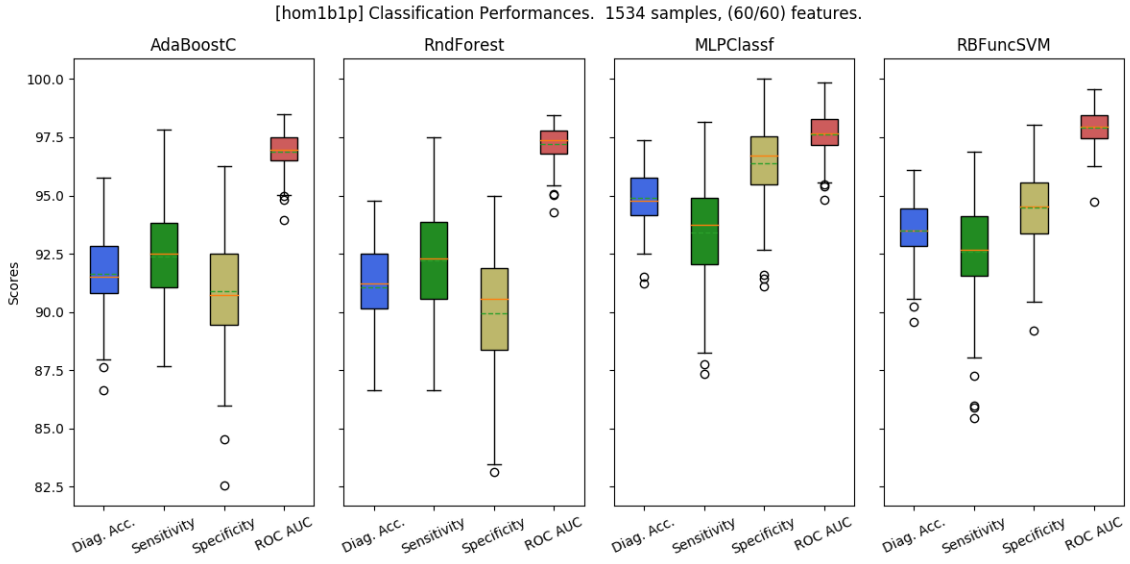
Table 4.2 and Figure 4.1 show the resulting metrics of the different classifiers on this dataset.

Table 4.2: *Classifier performance comparison for the hom1b1p dataset*

Classifier	Performance metric			
	Diag. Accuracy	Sensitivity	Specificity	ROC AUC
DecsnTree	81 % ± 3 %	77 % ± 5 %	86 % ± 4 %	85 % ± 3 %
KNearestN	63 % ± 3 %	27 % ± 4 %	99.5 % ± 0.6 %	90 % ± 2 %
NaivBayes	87 % ± 2 %	89 % ± 2 %	84 % ± 4 %	93 % ± 2 %
LinearSVM	88 % ± 6 %	81 % ± 9 %	95 % ± 10 %	93 % ± 1 %
AdaBoostC	92 % ± 2 %	92 % ± 2 %	91 % ± 3 %	96.9 % ± 0.9 %
RndForest	91 % ± 2 %	92 % ± 2 %	90 % ± 3 %	97.2 % ± 0.8 %
MLPClassf	95 % ± 1 %	93 % ± 2 %	96 % ± 2 %	97.6 % ± 0.9 %
RBFuncSVM	93 % ± 1 %	93 % ± 2 %	94 % ± 2 %	97.9 % ± 0.8 %



(a) *Classifiers with lowest AUC: 'DecsnTree', 'KNearestN', 'NaivBayes', 'LinearSVM'.*



(b) *Classifiers with highest AUC: 'AdaBoostC', 'RndForest', 'MLPClassf', 'RBFuncSVM'.*

Figure 4.1: *Performance comparison of eight classifiers on the hom1b1p dataset The classifiers are in ascending order by their AUC. For this dataset, 'RBFuncSVM' obtained the highest performance in terms of AUC*

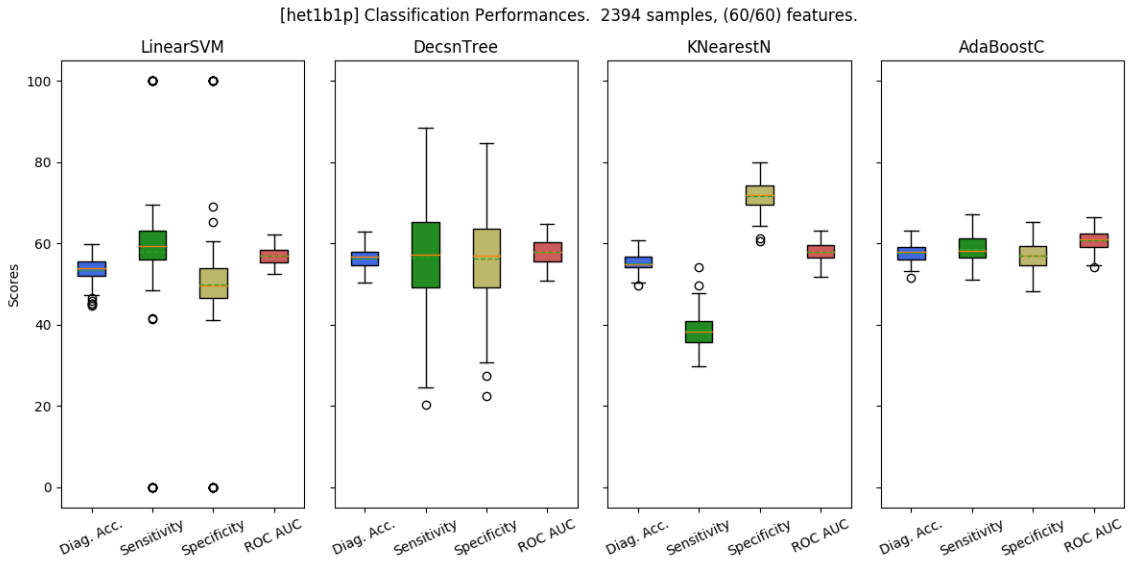
4.2 Classification results of the heterogeneous dataset (het1b1p)

This dataset contained 2394 samples and 60 features. Similarly to the homogeneous dataset, the two highest performing classifiers were 'MLPClassf' and "RBFuncSVM", with average AUC within 1% of each other. The three lowest performing classifiers were 'LinearSVM', 'DecsnTree' and 'KNearestN', just like in the previous dataset. However, the maximum performances in terms of AUC did not exceed 70% in this dataset. Lower performances with respect to the homogeneous dataset were expected due to the inclusion of fibroglandular patches.

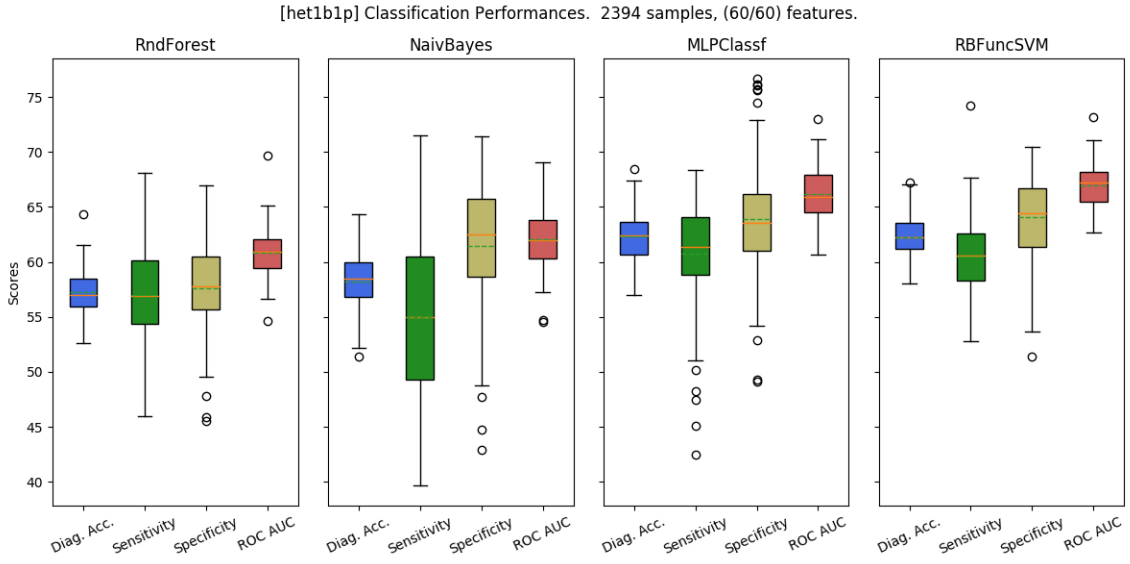
Table 4.3 and Figure 4.2 show the resulting metrics of the different classifiers on this dataset.

Table 4.3: *Classifier performance comparison for the het1b1p dataset*

Classifier	Performance metric			
	Diag. Accuracy	Sensitivity	Specificity	ROC AUC
LinearSVM	53 % ± 3 %	60 % ± 20 %	50 % ± 20 %	57 % ± 2 %
DecsnTree	56 % ± 3 %	60 % ± 10 %	60 % ± 10 %	58 % ± 3 %
KNearestN	55 % ± 2 %	39 % ± 4 %	72 % ± 4 %	58 % ± 3 %
AdaBoostC	58 % ± 2 %	59 % ± 3 %	57 % ± 3 %	61 % ± 3 %
RndForest	57 % ± 2 %	57 % ± 4 %	58 % ± 4 %	61 % ± 2 %
NaivBayes	58 % ± 2 %	55 % ± 7 %	61 % ± 6 %	62 % ± 3 %
MLPClassf	62 % ± 2 %	61 % ± 5 %	64 % ± 5 %	66 % ± 2 %
RBFuncSVM	62 % ± 2 %	61 % ± 4 %	64 % ± 4 %	67 % ± 2 %



(a) *Classifiers with lowest AUC: 'LinearSVM', 'DecsnTree', 'KNearestN', 'AdaBoostC'.*



(b) *Classifiers with highest AUC: 'RndForest', 'NaivBayes', 'MLPClassf', 'RBFuncSVM'.*

Figure 4.2: *Performance comparison of eight classifiers on the het1b1p dataset The classifiers are in ascending order by their AUC. For this dataset, 'RBFuncSVM' obtained the highest performance in terms of AUC*

4.3 Classification results of the heterogeneous dataset with rotation (het1b5pv2Alpha)

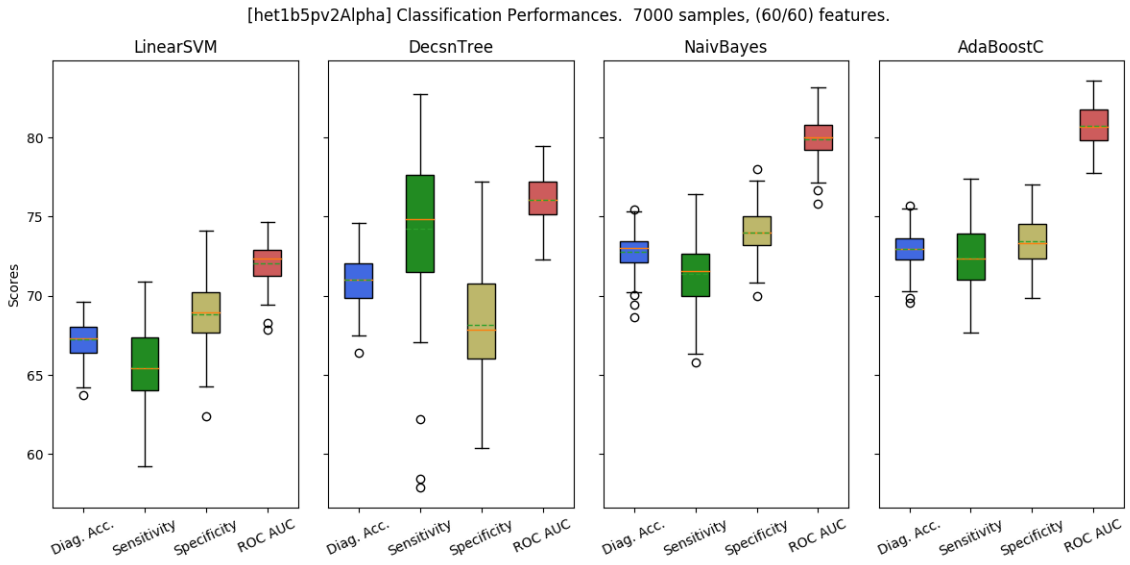
This dataset mode consisted of 7000 samples and 60 features. This modality achieved the highest AUC performance out of the three rotational setups. This higher performance is consistent with having a larger samples-to-features ratio, and this dataset modality has the highest ratio as presented in Table 3.2.

In terms of AUC, out of the eight classifiers used, 'KNearestN' ranked 3rd place in this dataset's modality, whereas the same classifier ranked 6th and 7th in the non-rotational datasets.

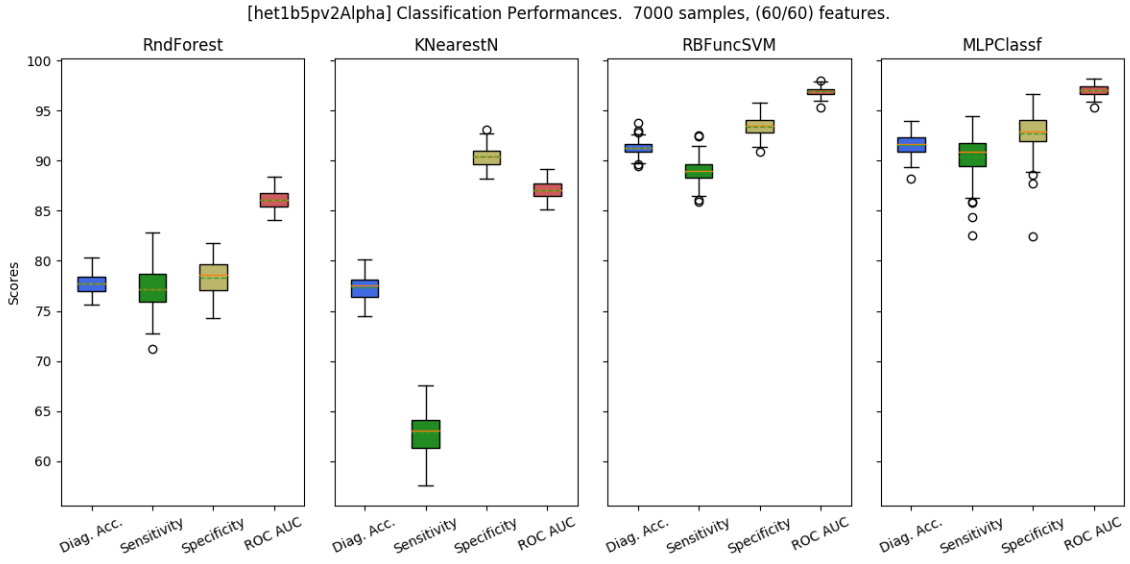
Table 4.4 and Figure 4.3 show the resulting metrics of the different classifiers on this dataset.

Table 4.4: *Classifier performance comparison for the het1b5pv2Alpha dataset*

Classifier	Performance metric			
	Diag. Accuracy	Sensitivity	Specificity	ROC AUC
LinearSVM	67 % ± 1 %	66 % ± 2 %	69 % ± 2 %	72 % ± 1 %
DecsnTree	71 % ± 2 %	74 % ± 4 %	68 % ± 4 %	76 % ± 1 %
NaivBayes	73 % ± 1 %	71 % ± 2 %	74 % ± 2 %	80 % ± 1 %
AdaBoostC	73 % ± 1 %	72 % ± 2 %	73 % ± 2 %	81 % ± 1 %
RndForest	78 % ± 1 %	77 % ± 2 %	78 % ± 2 %	86 % ± 1 %
KNearestN	77 % ± 1 %	63 % ± 2 %	90 % ± 1 %	87.1 % ± 0.9 %
RBFuncSVM	91.3 % ± 0.7 %	89 % ± 1 %	93 % ± 1 %	96.9 % ± 0.5 %
MLPClassf	92 % ± 1 %	90 % ± 2 %	93 % ± 2 %	97.0 % ± 0.5 %



(a) *Classifiers with lowest AUC: 'LinearSVM', 'DecsnTree', 'NaivBayes', 'AdaBoostC'.*



(b) *Classifiers with highest AUC: 'RndForest', 'KNearestN', 'RBFuncSVM', 'MLPClassf'.*

Figure 4.3: *Performance comparison of eight classifiers on the het1b5pv2Alpha dataset The classifiers are in ascending order by their AUC. For this dataset, 'MLPClassf' obtained the highest performance in terms of AUC*

4.4 Classification results of the heterogeneous dataset (het1b5pv2Beta)

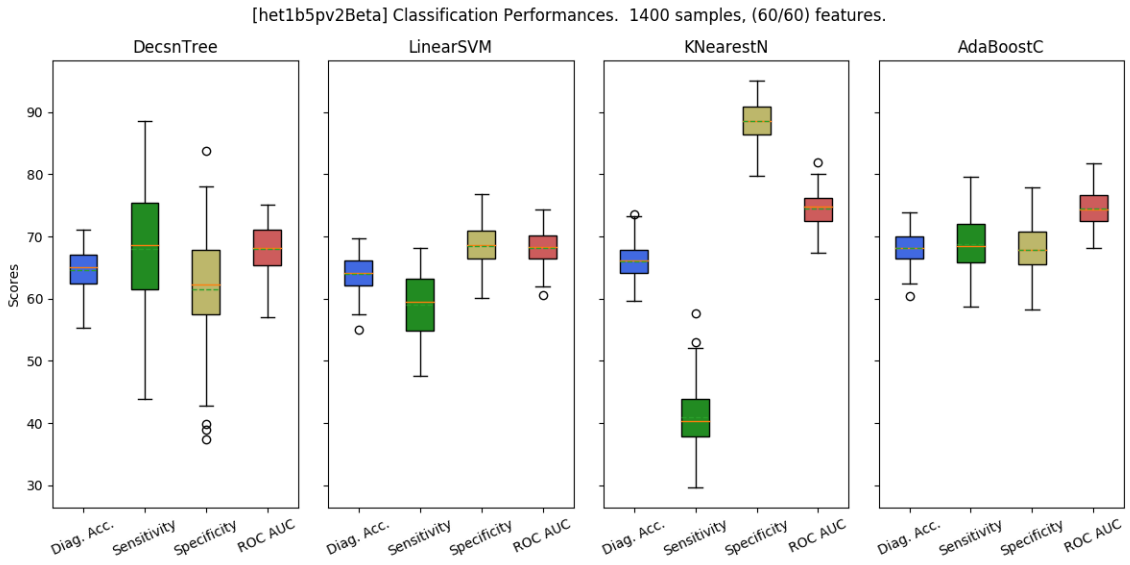
This dataset modality consisted of 1400 samples and 60 features. The tissues used here are the same as the ones used to build the heterogeneous dataset (het1b1p). However, the two datasets differ in the number of samples and the distribution of the random variables that determine tissue size and placement within the breast, as presented in Tables 3.3 and 3.4.

The classification performances on both datasets are significantly different: even the smallest AUC value obtained in this dataset (68 % ± 4 %, by 'DecsnTree') is similar to the highest AUC value obtained in the heterogeneous one (67 % ± 2 % by 'RBFuncSVM').

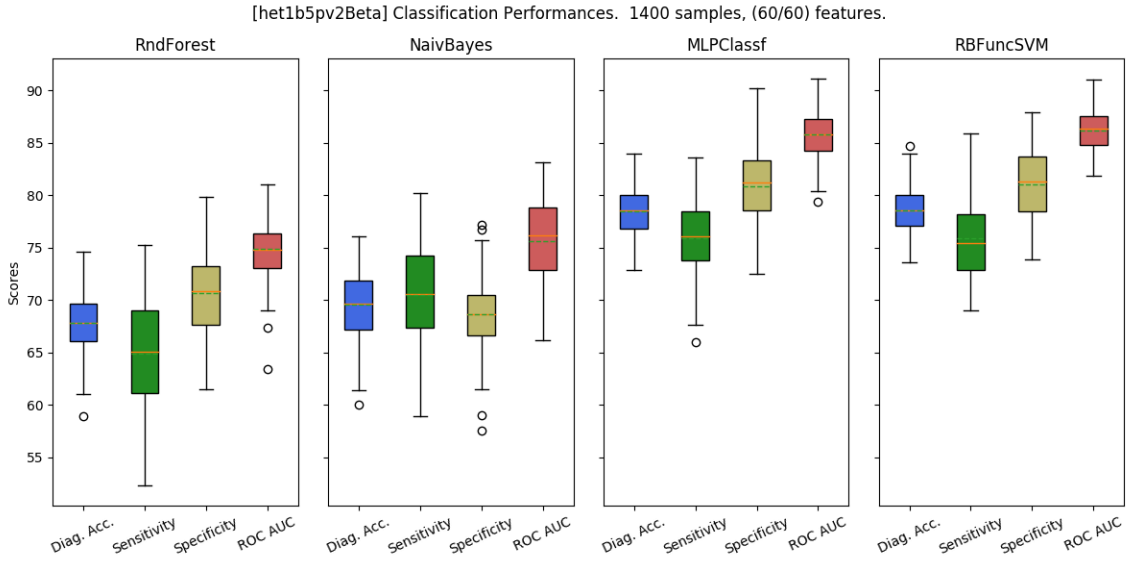
Table 4.5 and Figure 4.4 show the resulting metrics of the different classifiers on this dataset.

Table 4.5: *Classifier performance comparison for the het1b5pv2Beta dataset*

Classifier	Performance metric			
	Diag. Accuracy	Sensitivity	Specificity	ROC AUC
DecsnTree	65 % ± 3 %	68 % ± 10 %	61 % ± 9 %	68 % ± 4 %
LinearSVM	64 % ± 3 %	59 % ± 5 %	69 % ± 4 %	68 % ± 3 %
KNearestN	66 % ± 3 %	41 % ± 5 %	89 % ± 3 %	74 % ± 3 %
AdaBoostC	68 % ± 3 %	69 % ± 4 %	68 % ± 4 %	75 % ± 3 %
RndForest	68 % ± 3 %	65 % ± 6 %	71 % ± 4 %	75 % ± 3 %
NaivBayes	70 % ± 3 %	71 % ± 5 %	69 % ± 4 %	76 % ± 4 %
MLPClassf	79 % ± 2 %	76 % ± 4 %	81 % ± 3 %	86 % ± 2 %
RBFuncSVM	79 % ± 2 %	76 % ± 4 %	81 % ± 3 %	86 % ± 2 %



(a) *Classifiers with lowest AUC: 'DecsnTree', 'LinearSVM', 'KNearestN', 'AdaBoostC'.*



(b) *Classifiers with highest AUC: 'RndForest', 'NaivBayes', 'MLPClassf', 'RBFuncSVM'.*

Figure 4.4: *Performance comparison of eight classifiers on the het1b5pv2Beta dataset The classifiers are in ascending order by their AUC. For this dataset, 'RBFuncSVM' obtained the highest performance in terms of AUC*

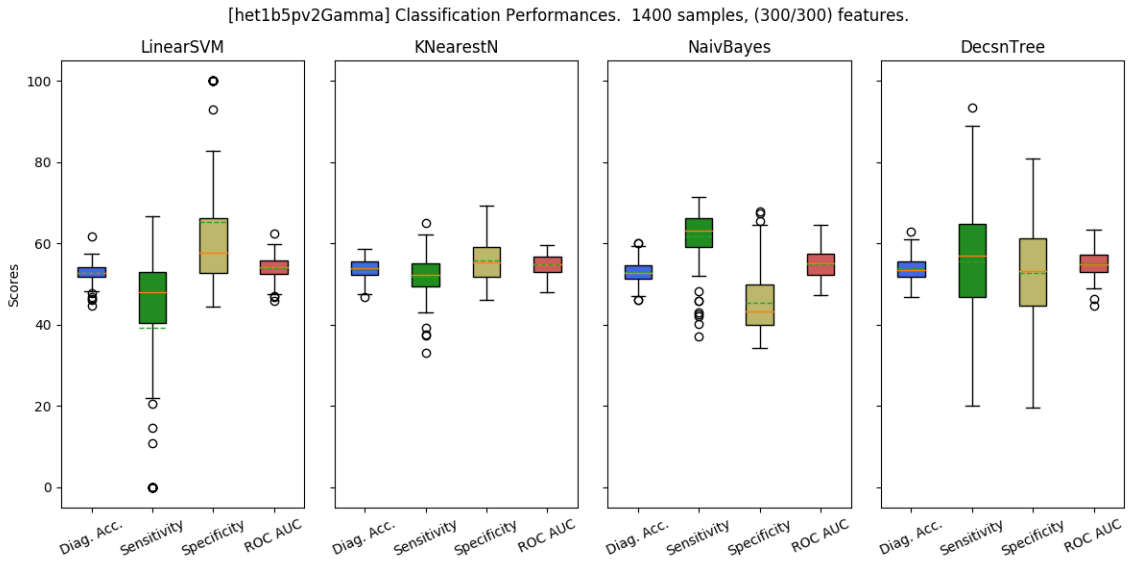
4.5 Classification results of the heterogeneous dataset with rotation (het1b5pv2Gamma)

This dataset mode consisted of 1400 samples and 300 features. The obtained classification performances for this dataset were the lowest of all 5 datasets. This dataset modality had the lowest samples-to-features ratio. The small ratio is consistent with low classification performances.

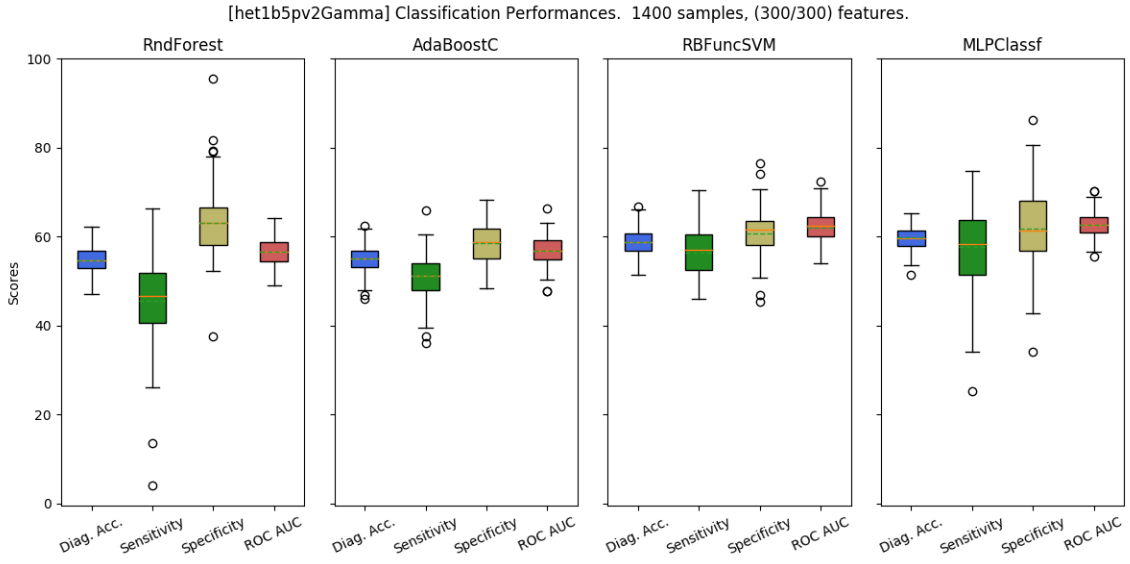
Table 4.6 and Figure 4.5 show the resulting metrics of the different classifiers on this dataset.

Table 4.6: *Classifier performance comparison for the het1b5pv2Gamma dataset*

Classifier	Performance metric			
	Diag. Accuracy	Sensitivity	Specificity	ROC AUC
LinearSVM	53 % ± 3 %	40 % ± 20 %	70 % ± 20 %	54 % ± 3 %
KNearestN	54 % ± 3 %	52 % ± 5 %	56 % ± 5 %	55 % ± 3 %
NaivBayes	53 % ± 3 %	62 % ± 7 %	45 % ± 8 %	55 % ± 4 %
DecsnTree	54 % ± 3 %	60 % ± 10 %	50 % ± 10 %	55 % ± 3 %
RndForest	55 % ± 3 %	45 % ± 10 %	63 % ± 8 %	57 % ± 3 %
AdaBoostC	55 % ± 3 %	51 % ± 5 %	59 % ± 4 %	57 % ± 3 %
RBFuncSVM	59 % ± 3 %	56 % ± 5 %	61 % ± 5 %	62 % ± 3 %
MLPClassf	60 % ± 3 %	58 % ± 10 %	62 % ± 8 %	63 % ± 3 %



(a) *Classifiers with lowest AUC: 'LinearSVM', 'KNearestN', 'NaivBayes', 'DecsnTree'.*



(b) *Classifiers with highest AUC: 'RndForest', 'AdaBoostC', 'RBFuncSVM', 'MLPClassf'.*

Figure 4.5: *Performance comparison of eight classifiers on the het1b5pv2Gamma dataset The classifiers are in ascending order by their AUC. For this dataset, 'MLPClassf' obtained the highest performance in terms of AUC*

Chapter 5

Discussion and conclusions

5.1 Summary of work

Chapter 1 presented the relationship between breast cancer screening availability and mortality figures related to breast cancer. The use of X-ray mammography as the primary screening technology and ultrasound and Magnetic Resonance Imaging (MRI) as secondary screening technologies was established, as well as their benefits and shortcomings. As a response to such drawbacks, the emerging microwave technologies have been presented, including the two most common variation of Microwave Imaging (MWI): Radar imaging and Microwave Tomography (MWT).

Chapter 2 laid out some of the characteristics of MWI systems, as well as the option of using Machine Learning (ML) classifiers on microwave data to determine the presence or absence of tumours. The various aspects of different microwave systems were explored in that chapter, highlighting the hardware and software that were used as inspiration for the preliminary design of a new portable Breast Microwave Sensing

(BMS) system.

Chapter 3 showed the steps taken in the design and simulation of a BMS system that aims to reduce the size of previous systems, while trying not to compromise in detection performance. That chapter also explored the process of producing datasets composed of breast models that included the statistical distribution of the tissue contents. The ML techniques applied to the simulated data obtained from numerical models of breasts inside the system's chamber were discussed as well.

Chapter 4 showed a comparison of the results obtained by using eight different ML classifiers. Two of them consistently outperformed the rest: a Multi Layer Perceptron (MLP) (MLPClassif) and a Support Vector Machine (SVM) (RBFuncSVM), despite one of them (MLPClassif) having had no optimization of its layer structure.

This chapter presents the conclusions as well as ideas and suggestions for the future of this research.

5.2 Discussion

Four out of the five datasets used in this thesis, used breast models which contained Fibroglandular Tissue (FGT). The fibroglandular densities of these breast models were limited to the Breast Imaging Reporting and Data System (BI-RADS) class 1. Currently, annual mammograms are recommended for women of 40 years of age and older. Breasts of BI-RADS class 1 are common in this age group, as there is an inverse relation between the fibroglandular densities and the age of a woman. However, future research might want to introduce breast models of BI-RADS classes 2 and higher. These models would account for breasts with higher fibroglandular

content, which are normally found in younger women.

5.3 Conclusions

The results presented in Chapter 4, suggest that the use of ML classifiers to detect the presence of tumours, by analysing microwave data is feasible. The classification results for the heterogeneous dataset with rotation (het1b5pv2Beta), show a ROC's AUC of 86% ± 2% for 'RBFuncSVM' and similarly, 86% ± 2% in the case of 'MLP-Classf'. These results are similar to those found in a 2012 study by Ying et al., where the reported AUC for X-ray mammography for 549 patients was of 88.6% ± 1.6%. Despite the use of synthetic data, these results are encouraging, especially when considering that there are several possible improvements to the simulations and pre-processing of the data.

The neural network (MLPClassf) was not subjected to Grid-Search Cross-Validation (GSCV) or any other means of optimization for its layer structure. It is also encouraging that despite this lack of optimization, MLPClassf consistently scored in the first or second place in terms of AUC when compared to the seven other classifiers. This suggests that an Artificial Neural Network (ANN) might be one of the most efficient classifiers to detect tumours from data in the format used by this research.

5.4 Suggestions for future work

5.4.1 Modifications to the prototype's design

The design of the Polyvinyl chloride (PVC) chamber used in this research, holds the transmitting antenna and 12 receiving solid-state microwave sensors at the same level from the chamber's floor. This design allows for convenient analysis of the cross-section corresponding to the height at which the solid-state sensors are located. However, this 2D approach poses the problem that a tumour might go undetected if it is far from the sensors' level. Some systems, as presented in Chapter 2, address this issue by using a hydraulic jack to move the sensors, covering different levels from the chest wall to the nipple. Nevertheless, a hydraulic jack could significantly increase the weight specifications of the system and could require a re-design of the device. Another option could be the addition of sensors located at different levels. Evaluating the benefits of several sensor layers might necessitate the introduction of a 3D forward-solver for the electromagnetic scattering. Further analysis of this issue needs to be carried out.

Another aspect of the design of the PVC chamber is the location of the 12 solid-state sensors. They have been located to surround the breast cavity without considering that certain regions of the chamber might contribute more useful information than the others. Future research could study the correlation between sensor position and classification performance, effectively finding if there are optimal sensor positions. The opposite might also be true; future research could find if there are sensors currently located in redundant regions and getting rid of such sensors could lower the

feature vector sizes while having little or no impact in classification performance.

5.4.2 Alternative forward-solver for the electromagnetic scattering simulation

This research relies exclusively on the Method of Moments (MoM) forward-solver of Electromagnetic scattering, which is a frequency-domain approach. However, an alternative simulation modality, such as the Finite-Difference Time-Domain (FDTD) method could bring its own set of benefits to this research. As shown in Figure 3.3, of all the tissues used for simulations, the skin is the healthy tissue with the highest relative permittivity value. By having time-domain data available, the reflections returning from the skin could be algorithmically removed. By removing the reflections from the skin, the classifiers could potentially increase their performance, as they would deal with information coming from higher contrasts of permittivity in the case where there is a tumour (permittivity of fat and FGT tissue against the permittivity of a tumour), and very little contrast when there is not one (permittivity of fat and FGT tissue).

In addition to the use of a time-domain forward solver, more realistic simulations could be carried out by taking into account the transmitting antenna's radiation pattern. The simulations performed in this thesis used a point source, this is equivalent to an isotropic antenna in 2D. The consideration of the physical antenna's radiation pattern would bring the simulations closer to the real-life data produced by the physical prototype.

This research made use of 2D numerical breasts modelled as circular patches of

tissues with known permittivity values. However, more realistic numerical breast models could be employed, even MRI-derived repositories such as the one provided by [27, 28] could make the simulations closer to reality. Using FDTD and MRI-derived numerical breasts could require higher computational power, this could be addressed by using hardware acceleration in the FDTD method, as presented by Wang et al. [60].

5.4.3 Classification process

As shown in the parameter graphs of 'RBFuncSVM' in Appendix A.5 (Figures A.21, A.22, A.23, A.24, and A.25), GSCV only found good-performing parameter combinations (candidates) in the upper middle area of the $gamma, C$ plane. It has been recommended in the literature (Hsu, Chang, Lin, et al. [58]) that a good approach to GSCV, consists of refining the search space to a finer grid after having identified a region containing good candidates. Future research could try and limit the search space to find potentially even better-performing parameters in less time.

Besides the use of Principal Component Analysis (PCA), there were no feature extraction or feature selection techniques applied to the data used in this research. There is a vast amount of literature that covers feature selection and extraction techniques for images, audio or text data. However, the type of information collected by the BMS system is neither, so a more specific approach to pre-processing the information coming from the sensors is needed. Focusing on feature engineering might have the biggest impact in a future research project that seeks to continue the work presented on this thesis, as having "good" features can significantly reduce the

need for large datasets.

A good first step to improve the feature engineering, could be to use log-magnitude and phase representation of the electric field as features, instead of the current approach, which is the magnitude of the complex energy values at the sensor locations. This log-magnitude and phase technique has been suggested by Meaney, Paulsen, Pogue, and Miga [61], with a significant improvement in image reconstruction for high permittivity contrast scenarios. The same approach has also been employed by Rubæk, Kim, and Meincke [33] with success; this suggests that, although the application has been to image reconstruction algorithms, classification of that type of data could potentially improve as well.

Although the choice of using the ML package scikit-learn is a good one, as the complexity of the data increases, Graphical Processing Unit (GPU) acceleration might become a necessity. Tensorflow ([62]) is a good example of a ML package that is compatible with GPU acceleration. The transition to such package should be relatively smooth, as python wrappers are available, this means that most of the current code could be adapted without having to write everything again.

Acronyms

ANN Artificial Neural Network. 9, 10, 14, 15, 63

API Application Program Interface. iii, 43

AUC Area Under the Curve. 41, 43, 50, 51, 53, 55, 57, 63

BAVA-D Balanced Antipodal Vivaldi Antenna with a Director. 18

BI-RADS Breast Imaging Reporting and Data System. 37, 62

BMS Breast Microwave Sensing. iii, 4, 5, 22, 24, 34–36, 61, 62, 66

BPNN Back Propagation Neural Network. 10

CAD Computer-Aided Design. 26

DLVA Double Layer Vivaldi Antenna. 17

DWT Discrete Wavelet Transform. 14

FDTD Finite-Difference Time-Domain. 15, 16, 18, 19, 65, 66

FEM Finite Element Method. 17, 19

FGT Fibroglandular Tissue. iii, 25, 30, 35–39, 62, 65

GNN Genetic Neural Network. 10, 14

GPU Graphical Processing Unit. 67

GRS Gaussian Random Spheres. 15

GSCV Grid-Search Cross-Validation. 9, 43, 47, 63, 66

KNN K-Nearest Neighbours. 8, 14, 15

LDA Linear Discriminant Analysis. 7, 15

ML Machine Learning. iii, 4–6, 8, 11, 14, 15, 22, 26, 29, 33, 43, 61–63, 67

MLP Multi Layer Perceptron. 10, 12, 62

MoM Method of Moments. iii, 17, 19, 21, 22, 27, 28, 65

MRI Magnetic Resonance Imaging. 3, 5, 14, 15, 61, 66

MWI Microwave Imaging. 3, 4, 15–18, 25, 61

MWT Microwave Tomography. 16, 25, 61

PCA Principal Component Analysis. 15, 43, 45, 66

PSO Particle Swarm Optimization. 9

PVC Polyvinyl chloride. 29, 30, 33, 36, 64

QDA Quadratic Discriminant Analysis. 7, 15

RBF Radial Basis Function. 9, 12

RBNN Radial Basis Function Neural Network. 10

ROC Receiver Operating Characteristic. 41, 43, 50, 63

SLAE System of Linear Algebraic Equations. 21

SPNN Simultaneous Perturbation Neural Network. 14

SVM Support Vector Machine. 8, 9, 12, 15, 62

SVP Support Vector Point. 9

UWB Ultrawideband. 16

Index of terms

Bibliography

[1] Elizabeth A. Rafferty et al. "Diagnostic Accuracy and Recall Rates for Digital Mammography and Digital Mammography Combined With One-View and Two-View Tomosynthesis: Results of an Enriched Reader Study". In: *American Journal of Roentgenology* 202.2 (Jan. 2014), pp. 273–281. ISSN: 0361-803X. DOI: 10.2214/AJR.13.11240. URL: http://dx.doi.org/10.2214/AJR.13.11240.

[2] Canadian Cancer Society's Advisory Committee on Cancer Statistics. *Canadian Cancer Statistics*. June 2017. URL: cancer.ca/Canadian-CancerStatistics-2017-EN.pdf (visited on 11/01/2017).

[3] G Van Schoor et al. "Increasingly strong reduction in breast cancer mortality due to screening". In: *British Journal of Cancer* 104.6 (2011), pp. 910–914.

[4] Jacques Fracheboud et al. "Decreased rates of advanced breast cancer due to mammography screening in The Netherlands". In: *British Journal of Cancer* 91.5 (2004), pp. 861–867.

[5] Anthony B Miller et al. "Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial". In: *Bmj* 348 (2014), g366.

[6] Joe B Harford et al. "Problem solving for breast health care delivery in low and middle resource countries (LMCs): consensus statement from the Breast Health Global Initiative". In: *The Breast* 20 (2011), S20–S29.

[7] American Cancer Society. "Cancer in Africa". In: (2011).

[8] Department of Epidemiology and CancerCare Manitoba Cancer Registry. *Cancer in Manitoba, 2012 Annual Statistical Report*. 2015.

[9] R Nilavalan et al. "Breast tumour detection using a flat 16 element array". In: *Proceedings of the 16th International Zurich Symposium on Electromagnetic Compatibility-Topical Meeting on Biomedical EMC, Zurich, Zwitzerland*. IEEE, 2005, pp. 81–84.

[10] Christiane K Kuhl et al. "Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer". In: *Journal of clinical oncology* 23.33 (2005), pp. 8469–8476.

[11] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. *Supervised machine learning: A review of classification techniques*. 2007.

[12] Ronald A Fisher. "The use of multiple measurements in taxonomic problems". In: *Annals of Human Genetics* 7.2 (1936), pp. 179–188.

[13] Amanpreet Singh, Narina Thakur, and Aakanksha Sharma. "A review of supervised machine learning algorithms". In: *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on*. IEEE. 2016, pp. 1310–1315.

[14] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Vol. 1. Springer series in statistics New York, 2001.

[15] Pedro Domingos and Michael Pazzani. "On the optimality of the simple Bayesian classifier under zero-one loss". In: *Machine learning* 29.2 (1997), pp. 103–130.

[16] Corinna Cortes and Vladimir Vapnik. "Support-vector networks". In: *Machine learning* 20.3 (1995), pp. 273–297.

[17] Andrew Ng. *CS229 Lecture notes*. 2000. URL: `http://cs229.stanford.edu/notes/cs229-notes3.pdf` (visited on 08/01/2017).

[18] Christopher M Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006. ISBN: 978-0-387-31073-2.

[19] Robert E Schapire. "The boosting approach to machine learning: An overview". In: *Nonlinear estimation and classification*. Springer, 2003, pp. 149–171.

[20] Harsh Pareek and Pradeep Ravikumar. "Human boosting". In: *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*. 2013, pp. 338–346.

[21] Gaël Varoquaux, Andreas Müller, and Jaques Grobler. *Classifier comparison*. 2017. URL: `http://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html`.

[22] Ahmad AbdulSadda, Nidhal Bouaynaya, and Kamran Iqbal. "Noninvasive breast tumor localization based on ultrawideband microwave backscatter". In: *Signal Processing and Communications (SPCOM), 2010 International Conference on*. IEEE. 2010, pp. 1–4.

[23] Ahmad AbdulSadda. "Breast Tumor Localization Using Simultaneous Perturbation Stochastic-Neural Algorithm". In: *Cancer and Oncology Research* 1.3 (2013), pp. 75–79.

[24] Amer Al-Badarneh, Hassan Najadat, and Ali M Alraziqi. "A classifier to detect tumor disease in MRI brain images". In: *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*. IEEE Computer Society. 2012, pp. 784–787.

[25] Raquel Cruz Conceição, Martin O'Halloran, Martin Glavin, and Edward Jones. "Effects of dielectric heterogeneity in the performance of breast tumour classifiers". In: *Progress In Electromagnetics Research M* 17 (2011), pp. 73–86.

[26] Raquel Cruz da Conceição et al. "Initial classification of breast tumour phantoms using a UWB radar prototype". In: *Electromagnetics in Advanced Applications (ICEAA), 2013 International Conference on*. IEEE. 2013, pp. 720–723.

[27] University of Wisconsin. *Computational Electromagnetics Laboratory (UWCEM)*. 2010. URL: `http://uwcem.ece.wisc.edu/`.

[28] Matthew J Burfeindt et al. "MRI-derived 3-D-printed breast phantom for microwave breast imaging validation". In: *IEEE Antennas and Wireless Propagation Letters* 11 (2012), pp. 1610–1613.

[29] Mariya Lazebnik et al. "A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries". In: *Physics in Medicine and Biology* 52.20 (2007), p. 6093. DOI: `10.1088/0031-9155/52/20/002`.

[30] Mariya Lazebnik et al. "A large-scale study of the ultrawideband microwave dielectric properties of normal breast tissue obtained from reduction surgeries". In: *Physics in Medicine and Biology* 52.10 (2007), p. 2637. DOI: `10.1088/0031-9155/52/10/001`.

[31] Maciej Klemm et al. "Radar-based breast cancer detection using a hemispherical antenna arrayExperimental results". In: *IEEE Transactions on Antennas and Propagation* 57.6 (2009), pp. 1692–1704.

[32] Ryan J Halter et al. "The correlation of in vivo and ex vivo tissue dielectric properties to validate electromagnetic breast imaging: initial clinical experience". In: *Physiological Measurement* 30.6 (2009), S121.

[33] Tonny Rubæk, Oleksiy S Kim, and Peter Meincke. "Computational validation of a 3-D microwave imaging system for breast-cancer screening". In: *IEEE Transactions on Antennas and Propagation* 57.7 (2009), pp. 2105–2115.

[34] Paul M Meaney et al. "A clinical prototype for active microwave imaging of the breast". In: *IEEE Transactions on Microwave Theory and Techniques* 48.11 (2000), pp. 1841–1853.

[35] Majid Ostadrahimi et al. "Investigating a double layer vivaldi antenna design for fixed array field measurement". In: *International Journal of Ultra Wideband Communications and Systems* 1.4 (2010), pp. 282–290.

[36] Jeremie Bourqui, Jeff M Sill, and Elise C Fear. "A prototype system for measuring microwave frequency reflections from the breast". In: *Journal of Biomedical Imaging* 2012 (2012), p. 9.

[37] Z.X. Cao et al. "Spintronic microwave imaging". In: *Applied Physics A* 111.2 (2013), pp. 329–337.

[38] Walton C Gibson. *The method of moments in electromagnetics*. Vol. 1. Chapman & Hall/CRC London, UK, 2008, p. 3. ISBN: 978-1-4200-6145-1.

[39] Jorge Sacristán, Bárbara Luz Oliveira, and Stephen Pistorius. "Classification of electromagnetic signals obtained from microwave scattering over healthy and tumorous breast models". In: *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. May 2016, pp. 1–5. DOI: 10.1109/CCECE.2016.7726761.

[40] Jack H. Richmond. "Scattering by a Dielectric Cylinder of Arbitrary Cross Section Shape." In: *IEEE Transactions on Antennas and Propagation* 13.3 (May 1965), pp. 334–341. ISSN: 0018-926X. DOI: 10.1109/TAP.1965.1138427.

[41] Vyacheslav V. Komarov. *Handbook of Dielectric and Thermal Properties of Materials at Microwave Frequencies*. Artech House, 2012.

[42] W. Eric Phillips. "The permittivity of air at a wavelength of 10 centimeters". In: *Proceedings of the IRE* 38.7 (1950), pp. 786–790.

[43] Deepak K Ghodgaonkar and Adib Bin Daud. "Calculation of Debye parameters of single Debye relaxation equation for human skin in vivo". In: *Telecommunication Technology, 2003. NCTT 2003 Proceedings. 4th National Conference on*. IEEE. 2003, pp. 71–74.

[44] Shih-Ying Huang et al. "The characterization of breast anatomical metrics using dedicated breast CT". In: *Medical Physics* 38.4 (2011), pp. 2180–2191.

[45] Stéfan van der Walt, S Chris Colbert, and Gael Varoquaux. "The NumPy array: a structure for efficient numerical computation". In: *Computing in Science & Engineering* 13.2 (2011), pp. 22–30.

[46] Abas Sabouni, Sima Noghanian, and Stephen Pistorius. "A global optimization technique for microwave imaging of the inhomogeneous and dispersive breast". In: *Canadian Journal of Electrical and Computer Engineering* 35.1 (2010), pp. 15–24.

[47] S.I. Alekseev and M.C. Ziskin. "Human skin permittivity determined by millimeter wave reflection measurements". In: *Bioelectromagnetics* 28.5 (2007), pp. 331–339.

[48] Deepak K Ghodgaonkar, Vasundara V Varadan, and Vijay K Varadan. "A free-space method for measurement of dielectric constants and loss tangents at microwave frequencies". In: *IEEE Transactions on Instrumentation and Measurement* 38.3 (1989), pp. 789–793.

[49] Laura Liberman and Jennifer H Menell. "Breast imaging reporting and data system (BI-RADS)". In: *Radiologic Clinics* 40.3 (2002), pp. 409–430.

[50] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[51] Lars Buitinck et al. "API design for machine learning software: experiences from the scikit-learn project". In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*. 2013, pp. 108–122.

[52] Tom Fawcett. "An introduction to ROC analysis". In: *Pattern Recognition Letters* 27.8 (2006), pp. 861–874.

[53] László A Jeni, Jeffrey F Cohn, and Fernando De La Torre. "Facing imbalanced data–Recommendations for the use of performance metrics". In: *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE. 2013, pp. 245–251.

[54] Maria Carolina Monard and Gustavo EAPA Batista. "Learmng with skewed class distrihutions". In: *Advances in Logic, Artificial Intelligence, and Robotics: LAPTEC* 85.2002 (2002), p. 173.

[55] Jake VanderPlas. *How Did Python Become A Data Science Powerhouse?* PyData Seattle. 2017. URL: `https://youtu.be/9by46AAqz70?t=1641`.

[56] Rasmus Bro and Age K Smilde. "Principal component analysis". In: *Analytical Methods* 6.9 (2014), pp. 2812–2831.

[57] Trent Hauck. *scikit-learn Cookbook*. Packt Publishing Ltd, 2014.

[58] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. *A practical guide to support vector classification*. 2003. URL: `https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf` (visited on 09/03/2017).

[59] Xuexiang Ying et al. "A Comparison of Mammography and Ultrasound in Women with Breast Disease: A Receiver Operating Characteristic Analysis". In: *Breast Journal* 18.2 (Mar. 2012), pp. 130–138. ISSN: 1075-122X.

[60] Xi-min Wang et al. "GPU-Accelerated Parallel Finite-Difference Time-Domain Method for Electromagnetic Waves Propagation in Unmagnetized Plasma Media". In: *arXiv preprint arXiv:1709.00821* (2017).

[61] Paul M Meaney, Keith D Paulsen, Brian W Pogue, and Michael I Miga. "Microwave image reconstruction utilizing log-magnitude and unwrapped phase to improve high-contrast object recovery". In: *IEEE Transactions on Medical Imaging* 20.2 (2001), pp. 104–116.

[62] Martn Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: `https://www.tensorflow.org/`.

Appendix A

Values found using Grid-Search

A.1 Parameters tuned for AdaBoost

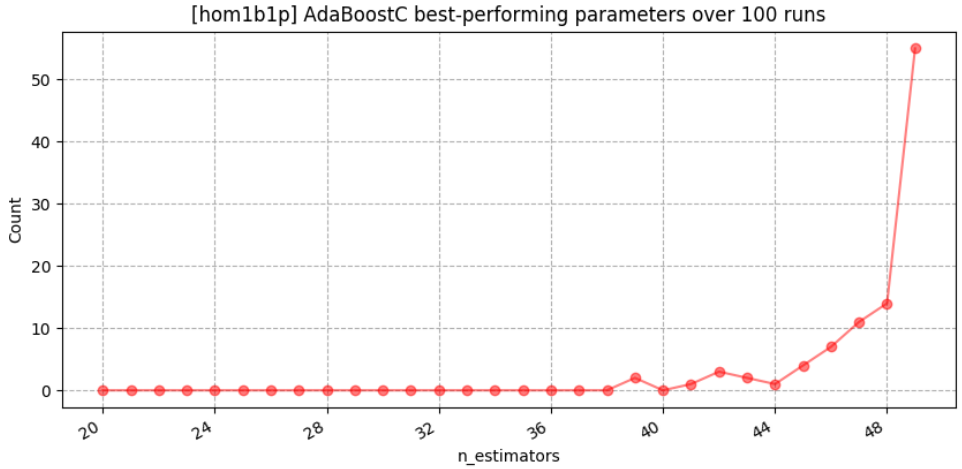


Figure A.1: *Grid-Search space showing the best-performing values for AdaBoostC's parameter (n_estimators) on the hom1b1p dataset.*

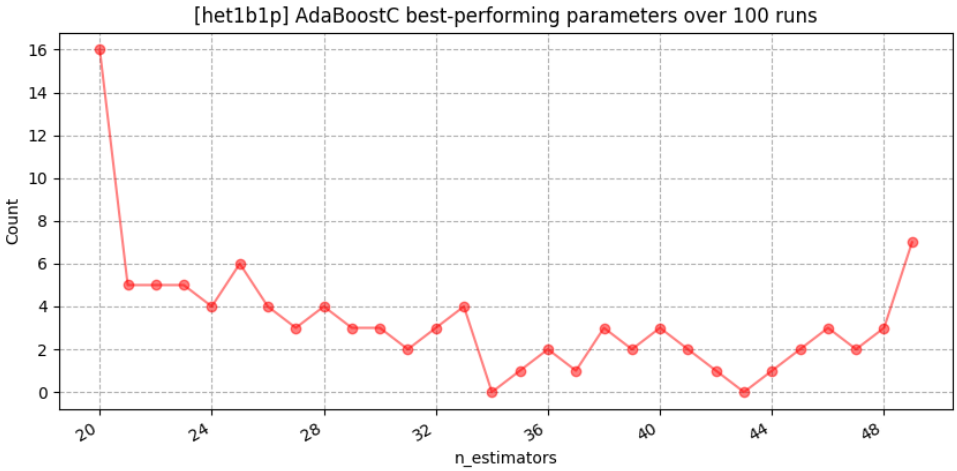


Figure A.2: *Grid-Search space showing the best-performing values for AdaBoostC's parameter (n_estimators) on the het1b1p dataset.*

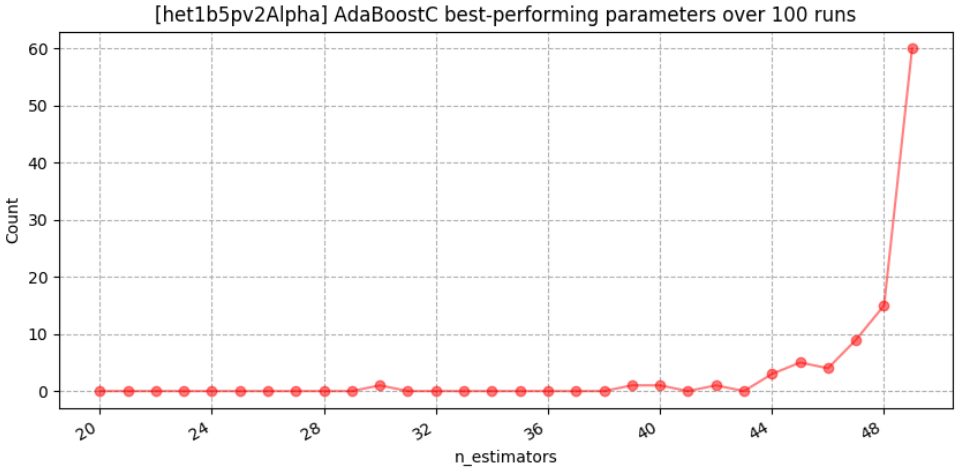


Figure A.3: *Grid-Search space showing the best-performing values for AdaBoostC's parameter (n_estimators) on the het1b5pv2Alpha dataset.*

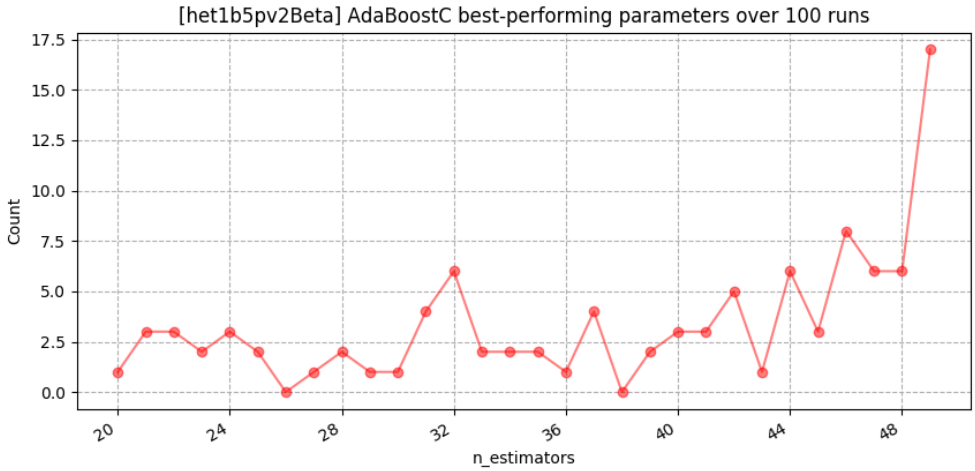


Figure A.4: *Grid-Search space showing the best-performing values for AdaBoostC's parameter (n_estimators) on the het1b5pv2Beta dataset.*

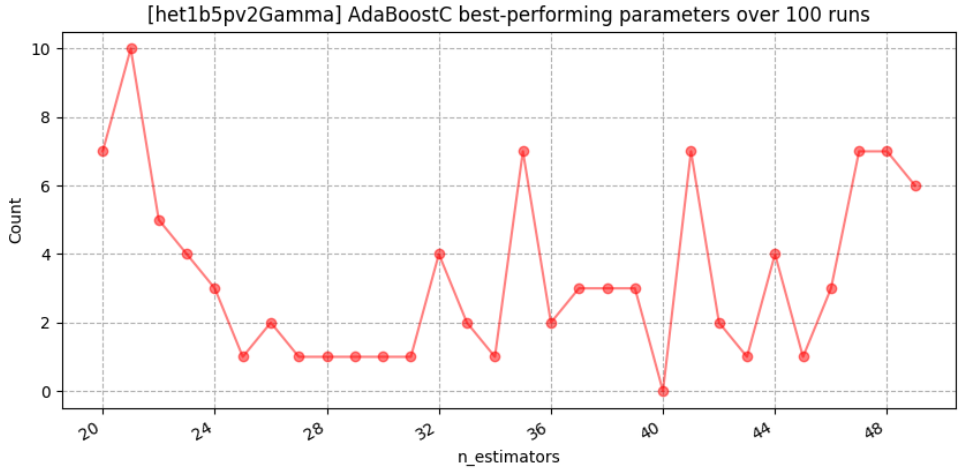


Figure A.5: *Grid-Search space showing the best-performing values for AdaBoostC's parameter (n_estimators) on the het1b5pv2Gamma dataset.*

A.2 Parameters tuned for DecsnTree

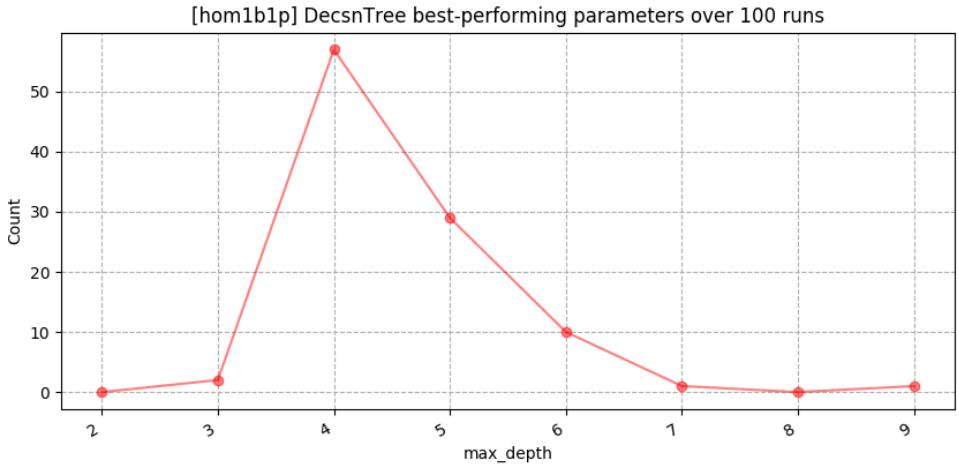


Figure A.6: *Grid-Search space showing the best-performing values for DecsnTree's parameter (max_depth) on the hom1b1p dataset.*

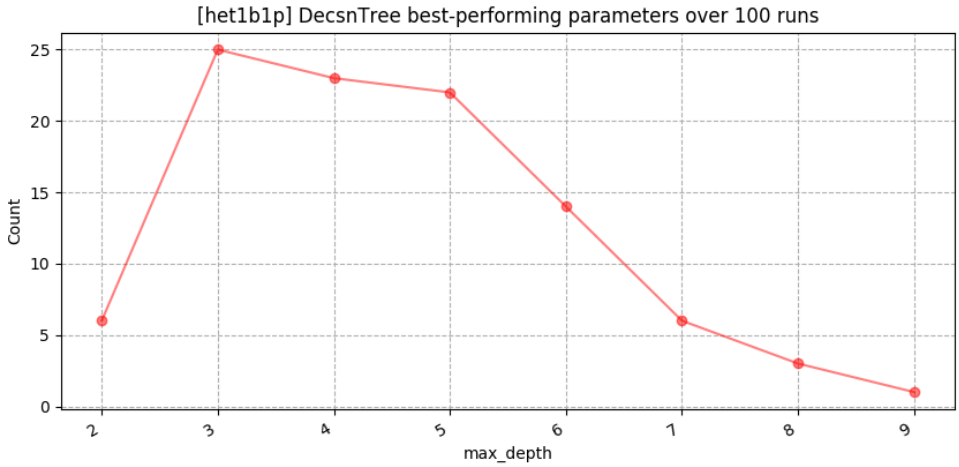


Figure A.7: *Grid-Search space showing the best-performing values for DecsnTree's parameter (max_depth) on the het1b1p dataset.*

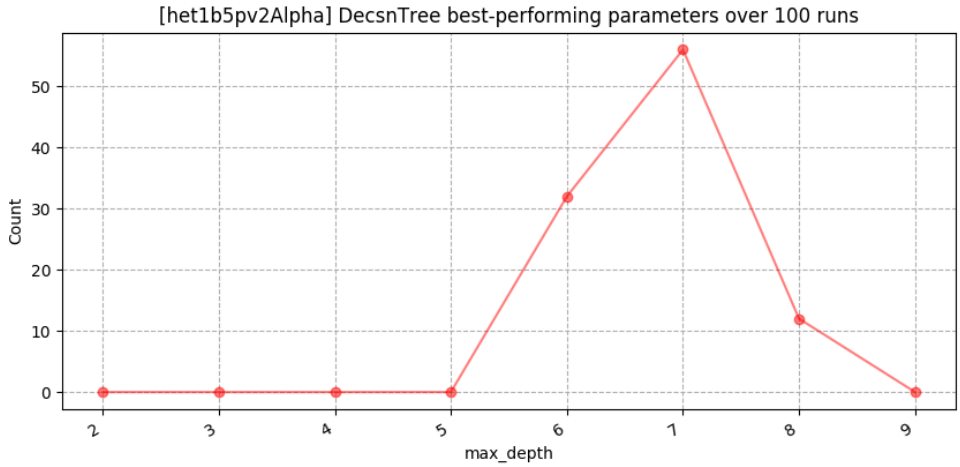


Figure A.8: *Grid-Search space showing the best-performing values for DecsnTree's parameter (max_depth) on the het1b5pv2Alpha dataset.*

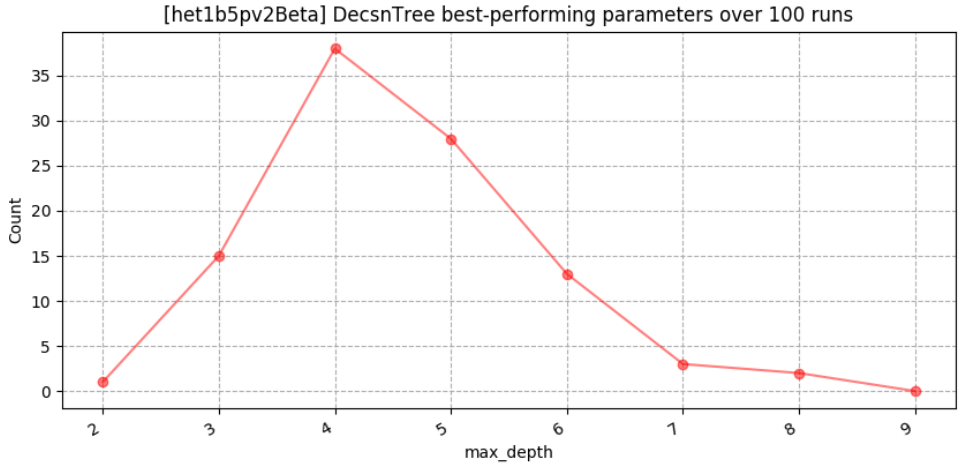


Figure A.9: *Grid-Search space showing the best-performing values for DecsnTree's parameter (max_depth) on the het1b5pv2Beta dataset.*

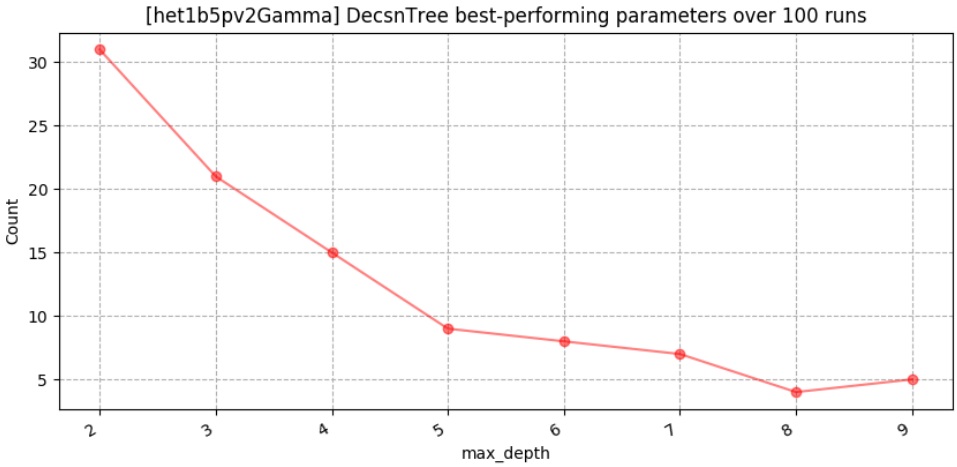


Figure A.10: *Grid-Search space showing the best-performing values for DecsnTree's parameter (max_depth) on the het1b5pv2Gamma dataset.*

A.3 Parameters tuned for KNearestN

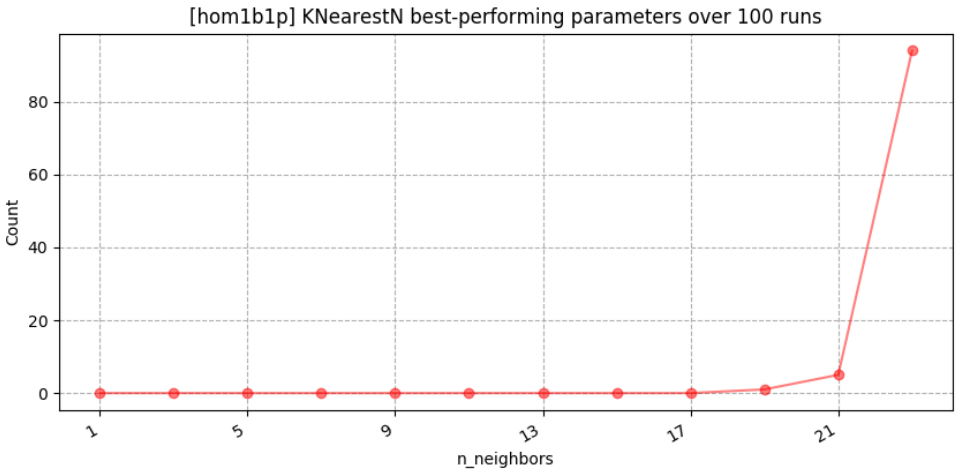


Figure A.11: *Grid-Search space showing the best-performing values for KNearestN's parameter (n_neighbors) on the hom1b1p dataset.*

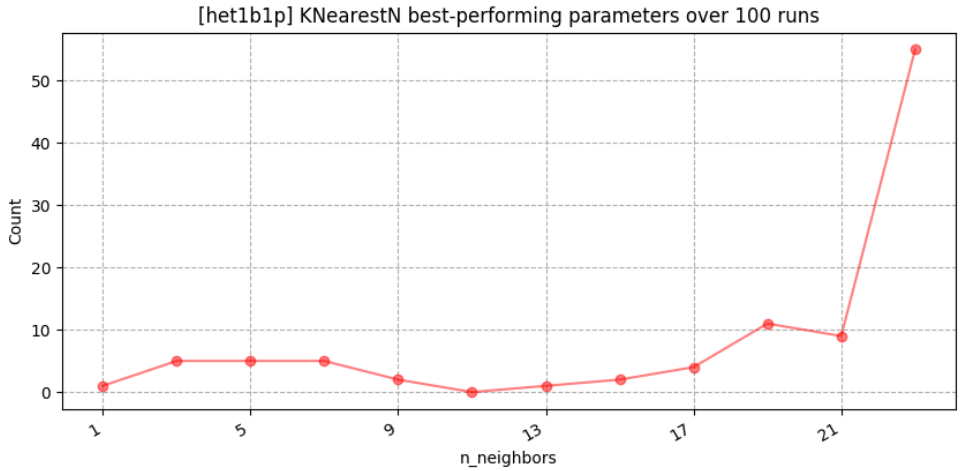


Figure A.12: *Grid-Search space showing the best-performing values for KNearestN's parameter (n_neighbors) on the het1b1p dataset.*

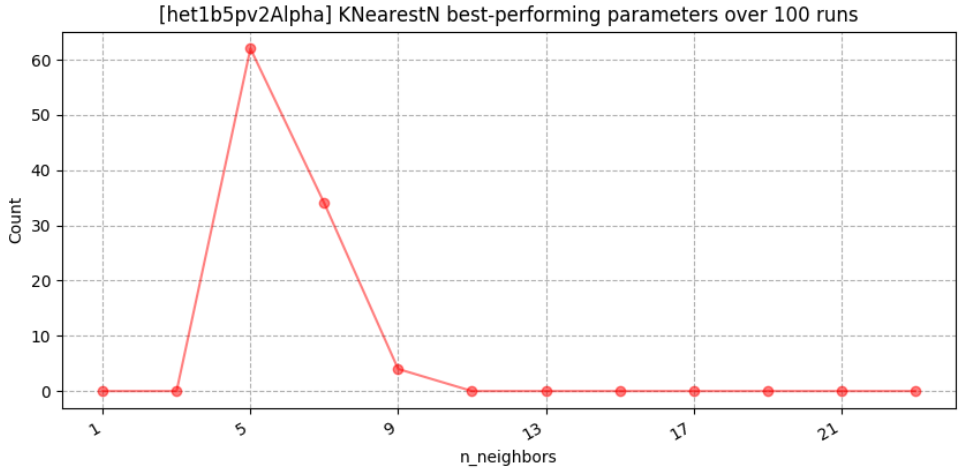


Figure A.13: *Grid-Search space showing the best-performing values for KNearestN's parameter (n_neighbors) on the het1b5pv2Alpha dataset.*

Figure A.14: *Grid-Search space showing the best-performing values for KNearestN's parameter (n_neighbors) on the het1b5pv2Beta dataset.*

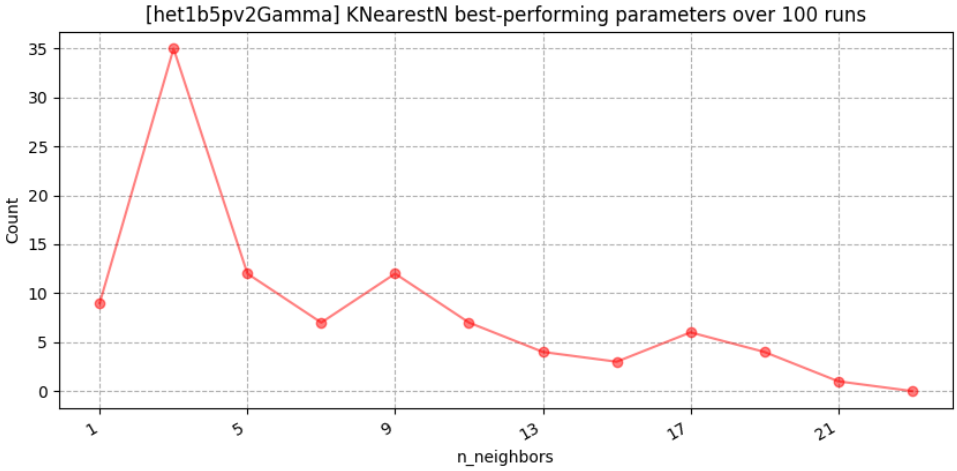


Figure A.15: *Grid-Search space showing the best-performing values for KNearestN's parameter (n_neighbors) on the het1b5pv2Gamma dataset.*

A.4 Parameters tuned for LinearSVM

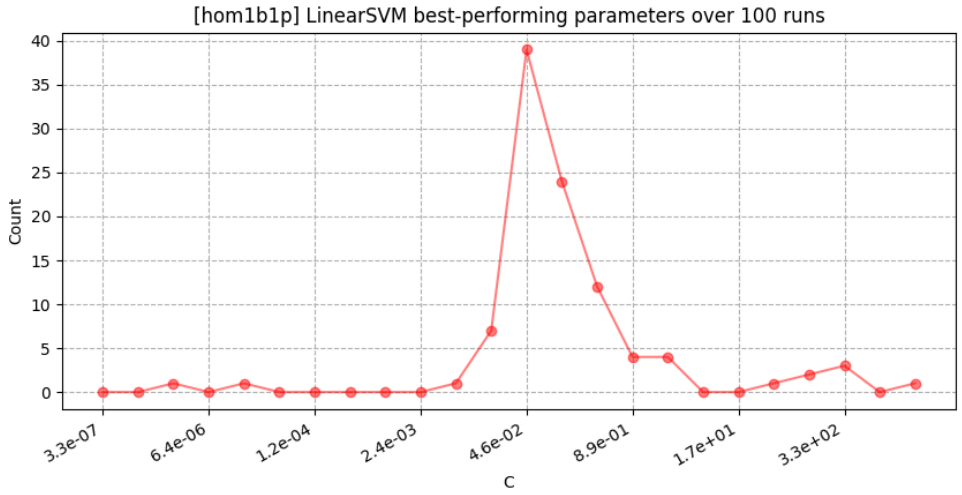


Figure A.16: *Grid-Search space showing the best-performing values for LinearSVM's parameter (C) on the hom1b1p dataset.*

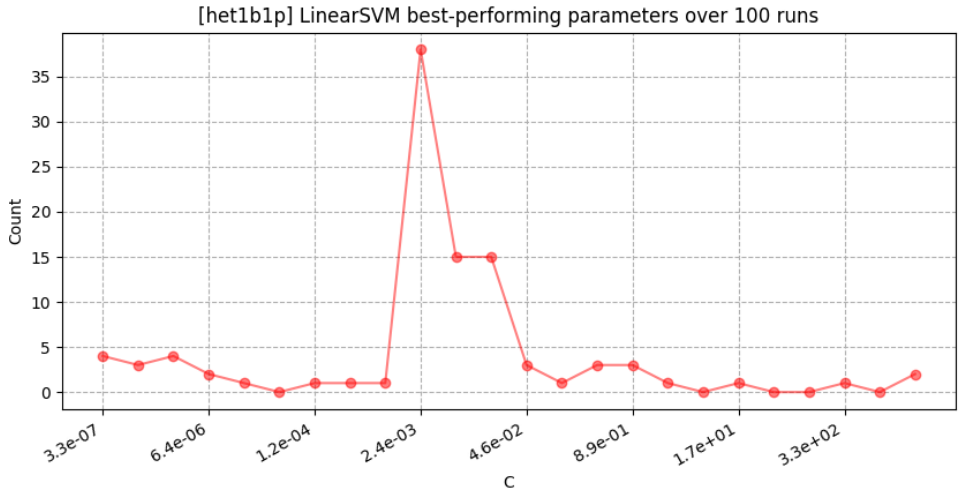


Figure A.17: *Grid-Search space showing the best-performing values for LinearSVM's parameter (C) on the het1b1p dataset.*

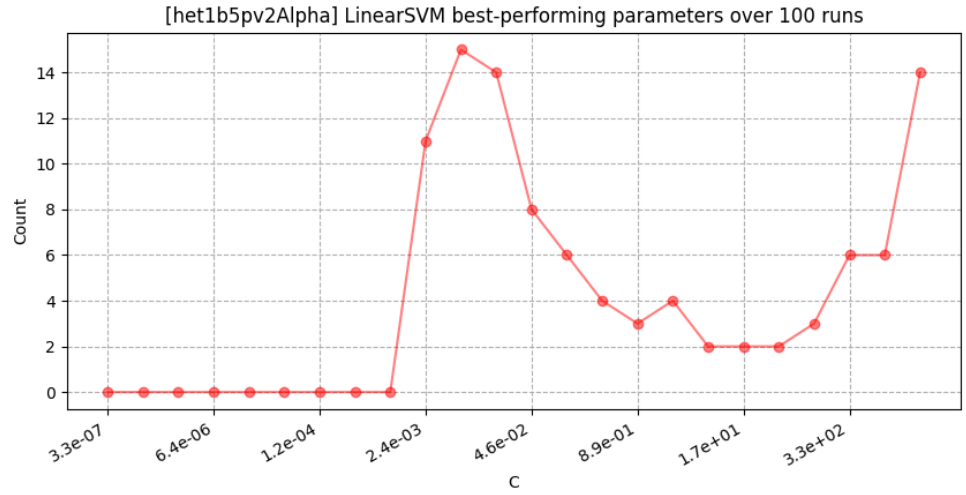


Figure A.18: *Grid-Search space showing the best-performing values for LinearSVM's parameter (C) on the het1b5pv2Alpha dataset.*

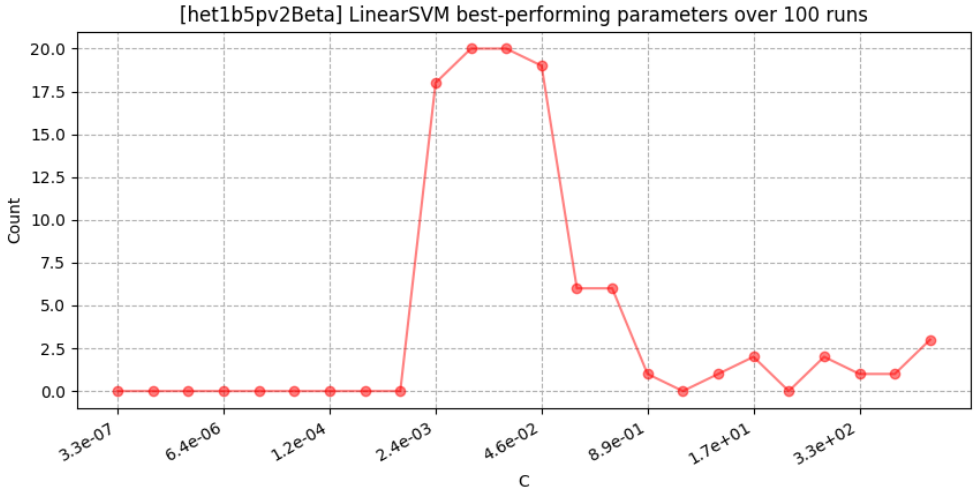


Figure A.19: *Grid-Search space showing the best-performing values for LinearSVM's parameter (C) on the het1b5pv2Beta dataset.*

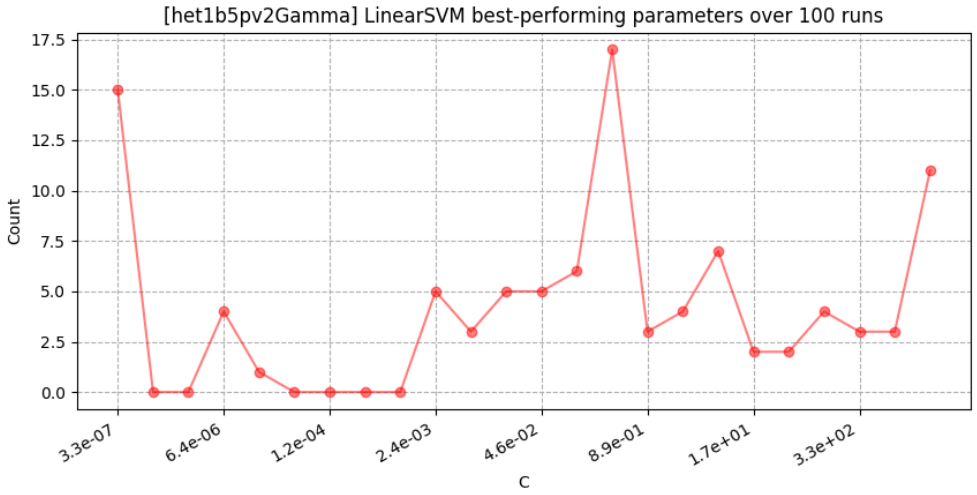


Figure A.20: *Grid-Search space showing the best-performing values for LinearSVM's parameter (C) on the het1b5pv2Gamma dataset.*

A.5 Parameters tuned for RBFuncSVM

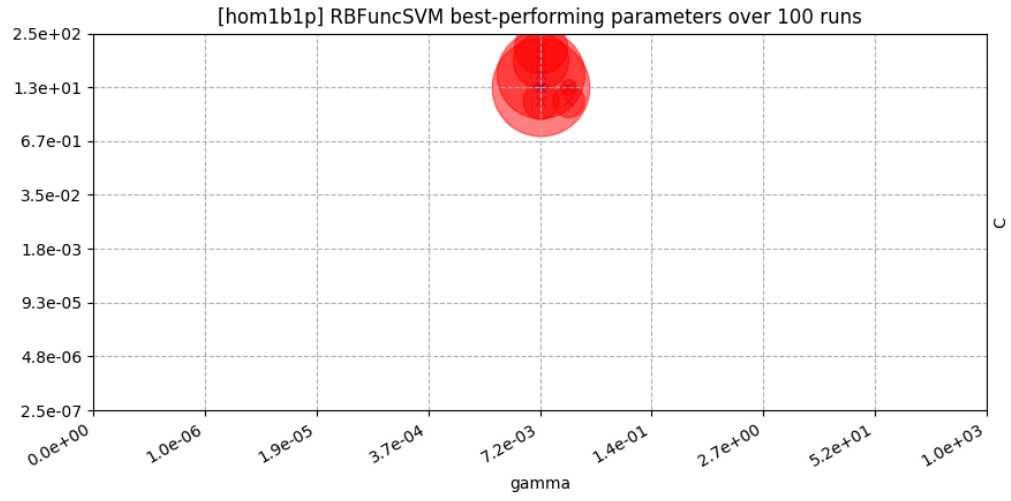


Figure A.21: *Grid-Search space showing the best-performing values for RBFuncSVM's parameters (gamma, C) on the hom1b1p dataset.*

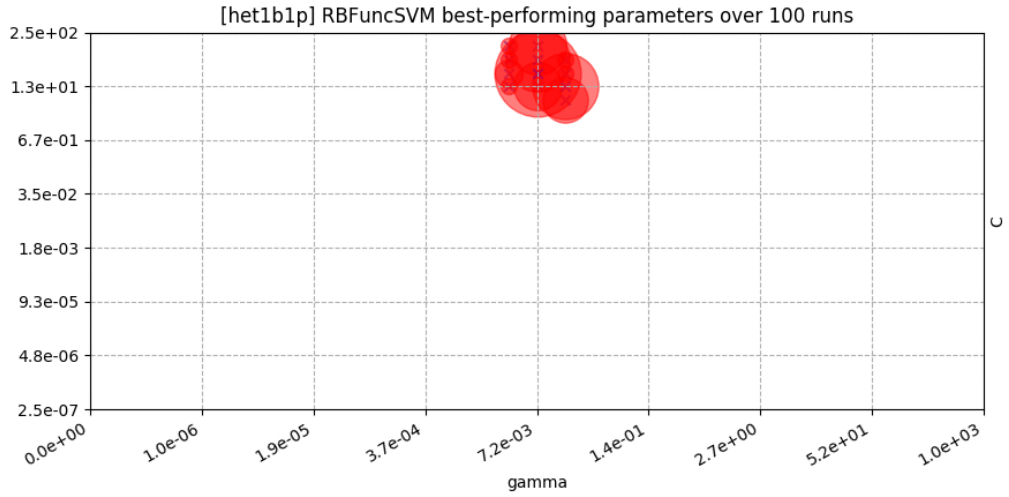


Figure A.22: *Grid-Search space showing the best-performing values for RBFuncSVM's parameters (gamma, C) on the het1b1p dataset.*

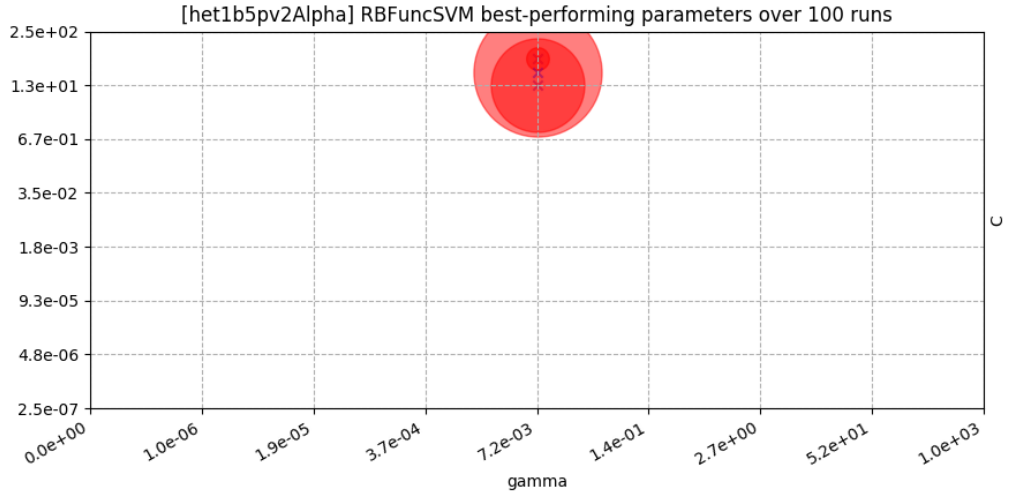


Figure A.23: *Grid-Search space showing the best-performing values for RBFuncSVM's parameters (gamma, C) on the het1b5pv2Alpha dataset.*

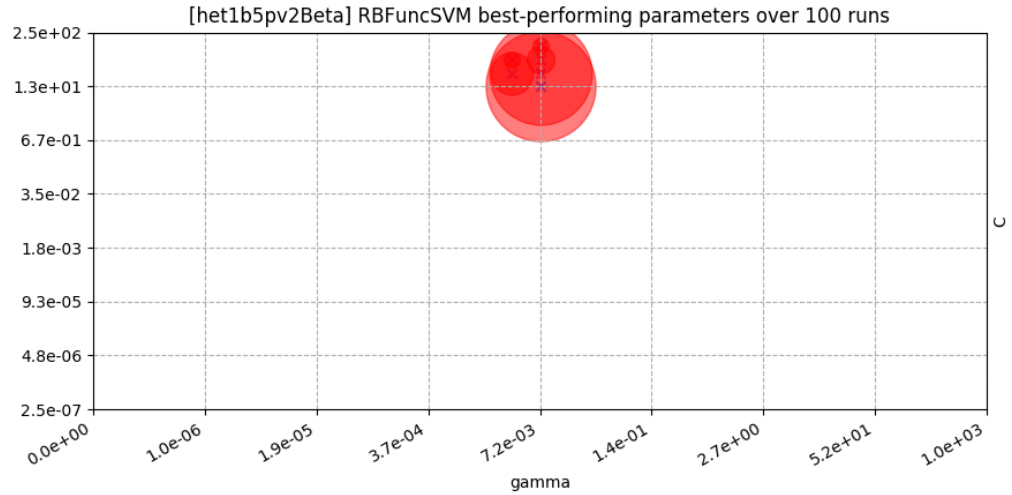


Figure A.24: *Grid-Search space showing the best-performing values for RBFuncSVM's parameters (gamma, C) on the het1b5pv2Beta dataset.*

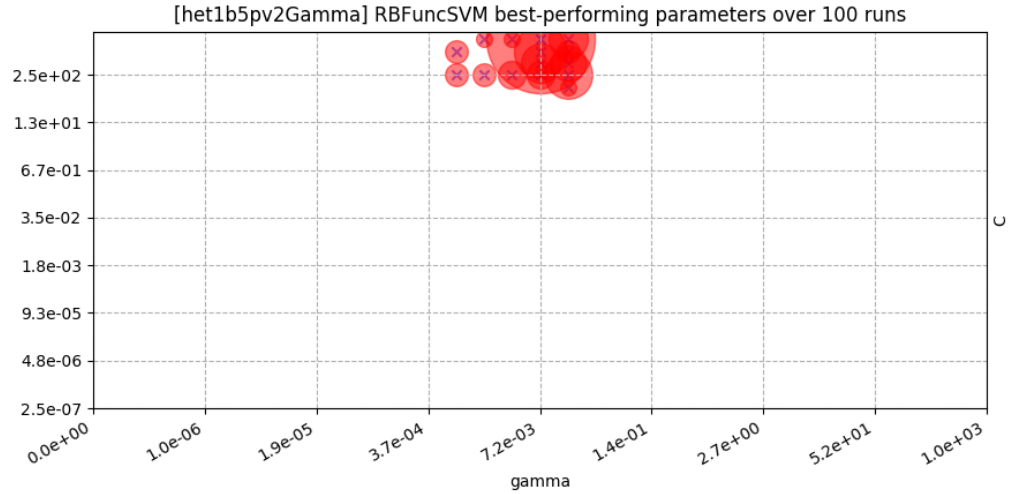


Figure A.25: *Grid-Search space showing the best-performing values for RBFuncSVM's parameters (gamma, C) on the het1b5pv2Gamma dataset.*

A.6 Parameters tuned for RndForest

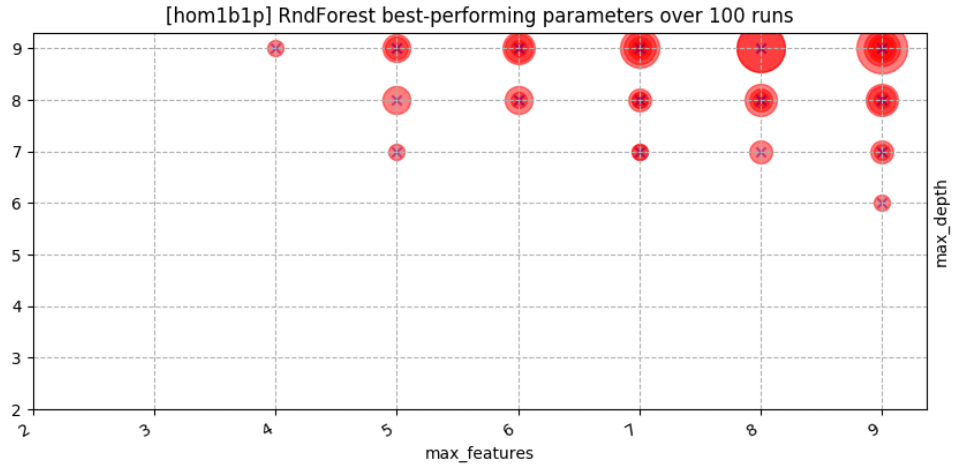


Figure A.26: *Grid-Search space showing the best-performing values for RndForest's parameters (max_features, max_depth) on the hom1b1p dataset.*

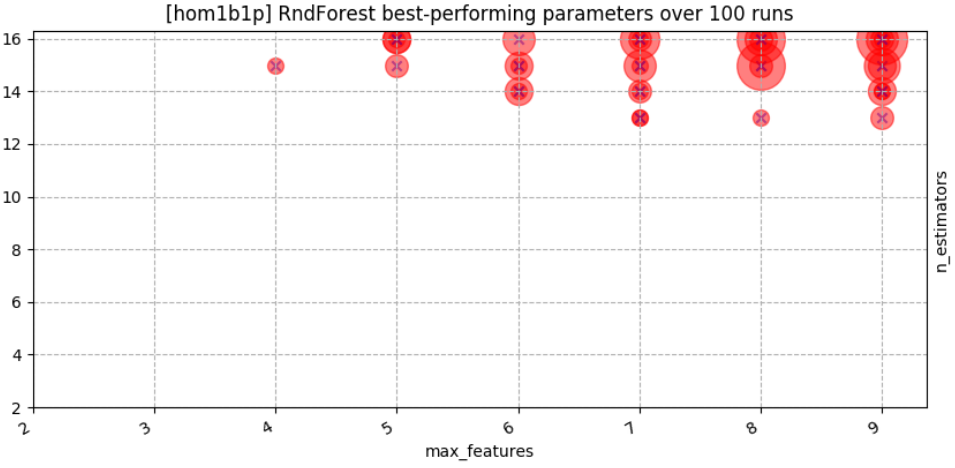


Figure A.27: *Grid-Search space showing the best-performing values for RndForest's parameters (max_features, n_estimators) on the hom1b1p dataset.*

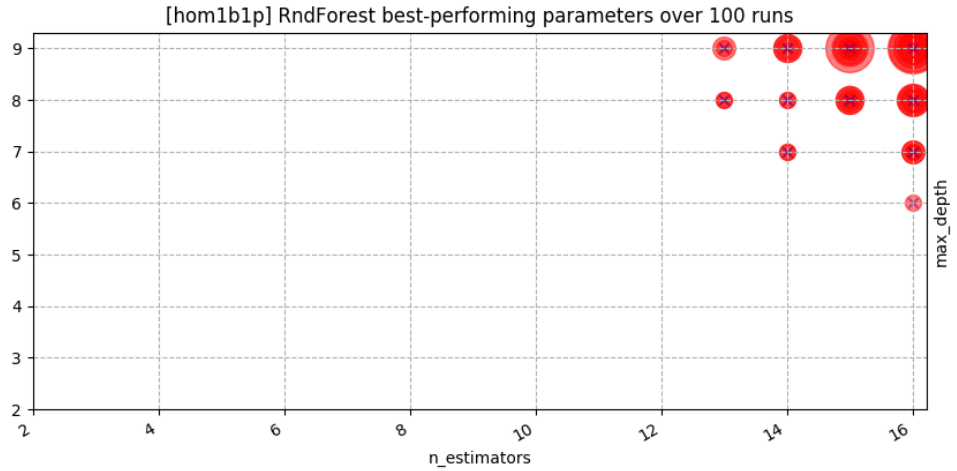


Figure A.28: *Grid-Search space showing the best-performing values for RndForest's parameters (n_estimators, max_depth) on the hom1b1p dataset.*

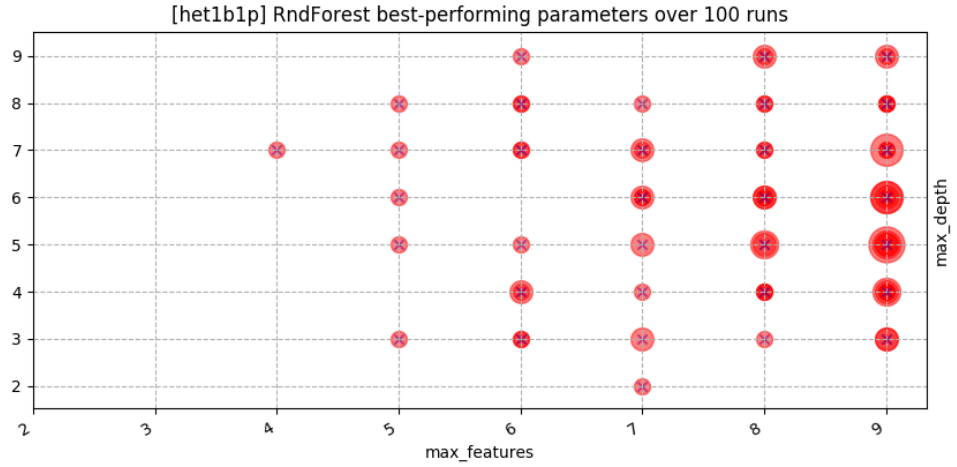


Figure A.29: *Grid-Search space showing the best-performing values for RndForest's parameters (max_features, max_depth) on the het1b1p dataset.*

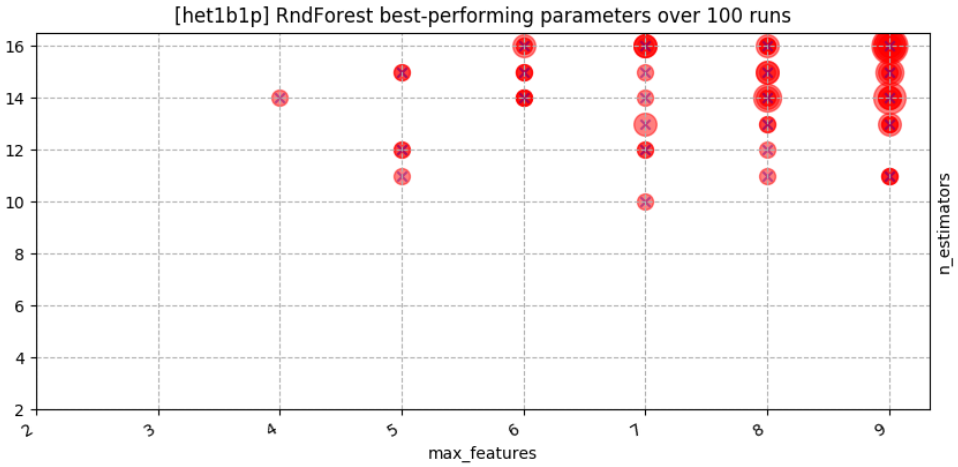


Figure A.30: *Grid-Search space showing the best-performing values for RndForest's parameters (max_features, n_estimators) on the het1b1p dataset.*

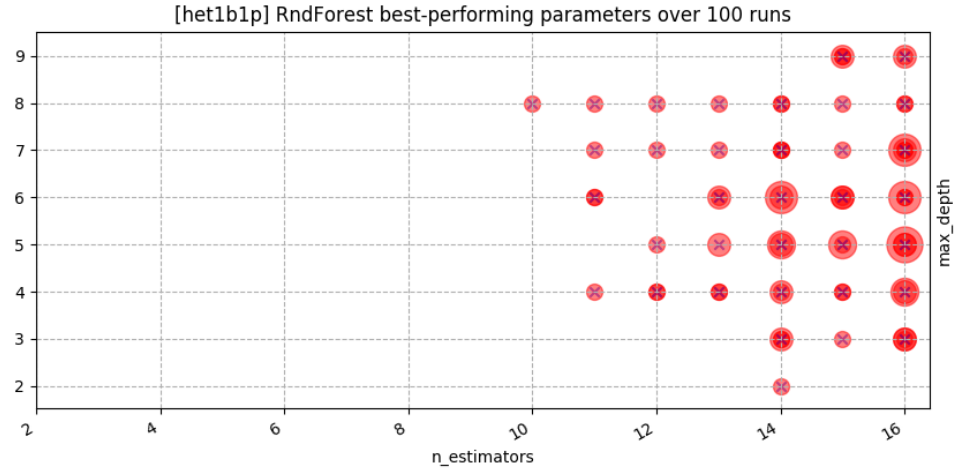


Figure A.31: *Grid-Search space showing the best-performing values for RndForest's parameters (n_estimators, max_depth) on the het1b1p dataset.*

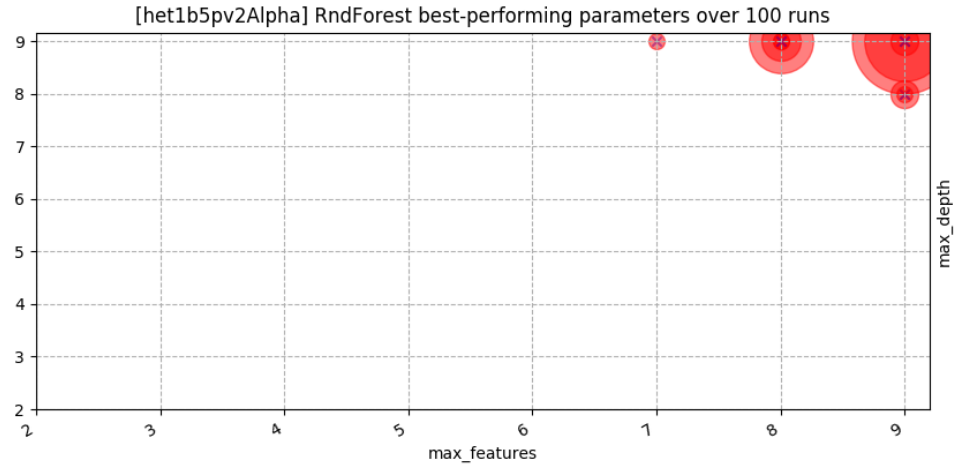


Figure A.32: *Grid-Search space showing the best-performing values for RndForest's parameters (max_features, max_depth) on the het1b5pv2Alpha dataset.*

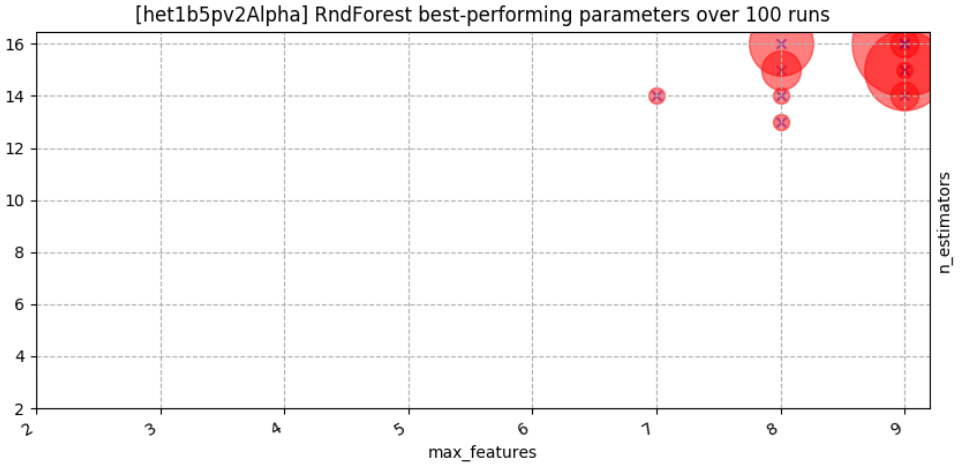


Figure A.33: *Grid-Search space showing the best-performing values for RndForest's parameters (max_features, n_estimators) on the het1b5pv2Alpha dataset.*

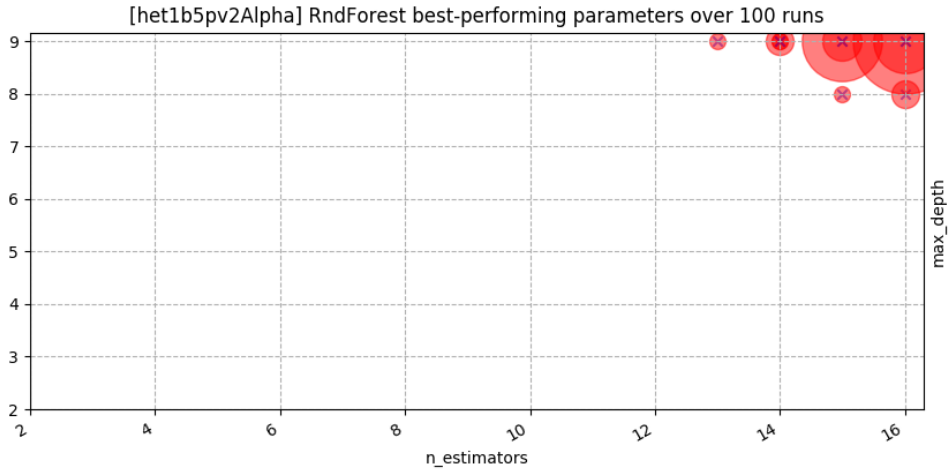


Figure A.34: *Grid-Search space showing the best-performing values for RndForest's parameters (n_estimators, max_depth) on the het1b5pv2Alpha dataset.*

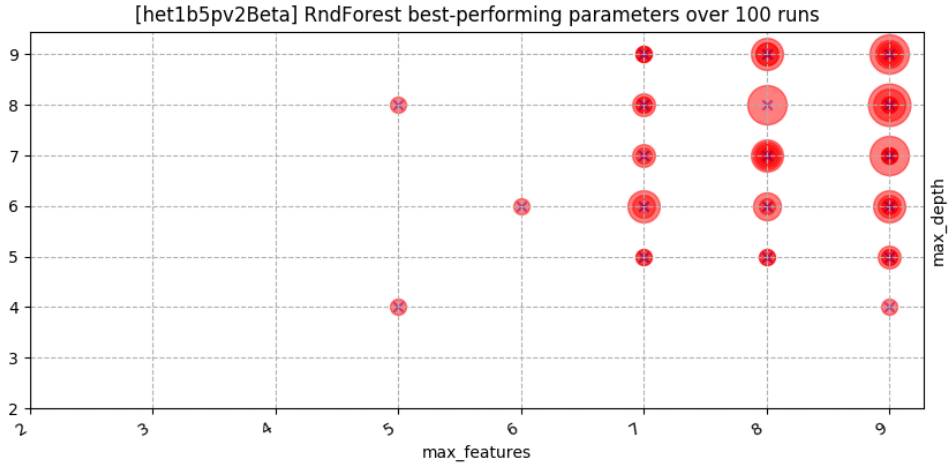


Figure A.35: *Grid-Search space showing the best-performing values for RndForest's parameters (max_features, max_depth) on the het1b5pv2Beta dataset.*

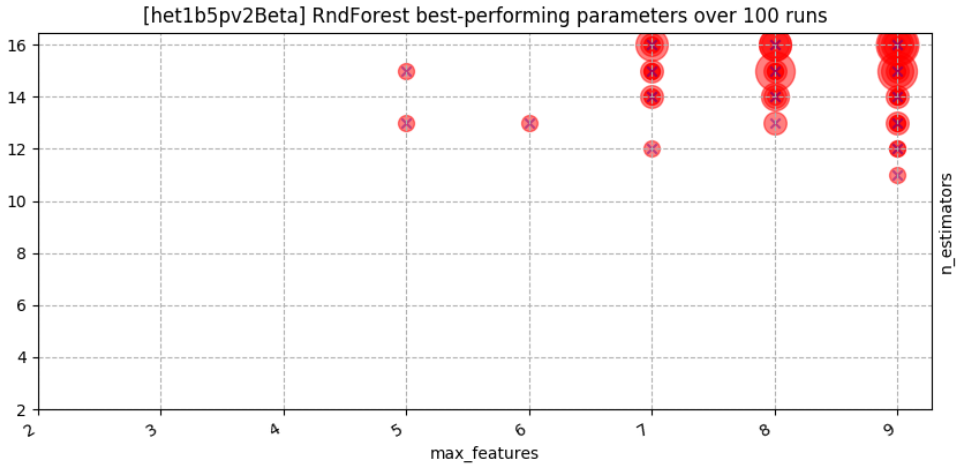


Figure A.36: *Grid-Search space showing the best-performing values for RndForest's parameters (max_features, n_estimators) on the het1b5pv2Beta dataset.*

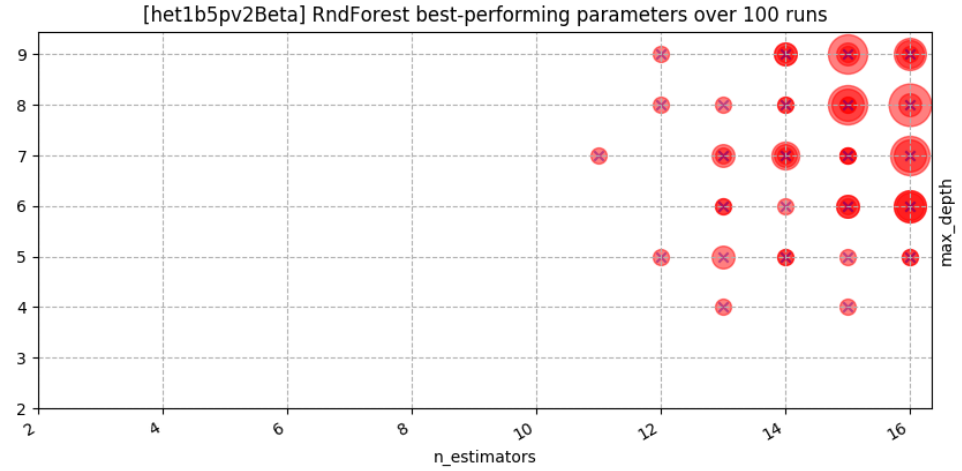


Figure A.37: *Grid-Search space showing the best-performing values for RndForest's parameters (n_estimators, max_depth) on the het1b5pv2Beta dataset.*

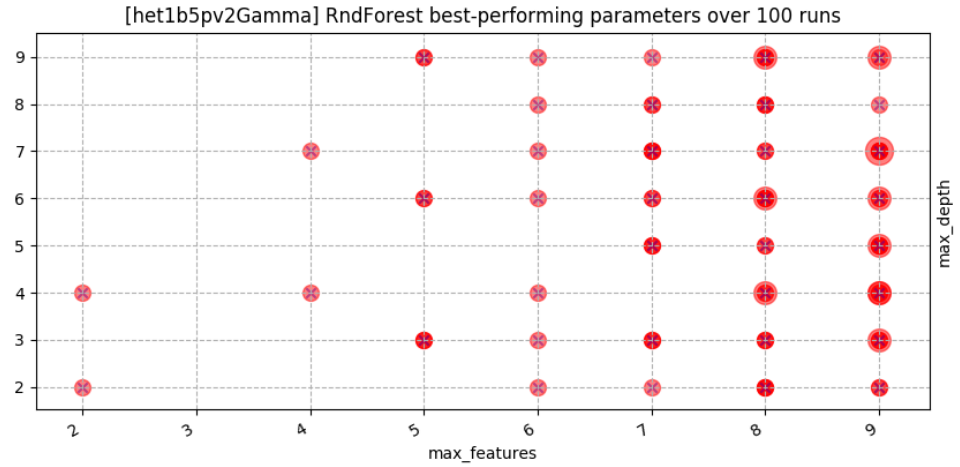


Figure A.38: *Grid-Search space showing the best-performing values for RndForest's parameters (max_features, max_depth) on the het1b5pv2Gamma dataset.*

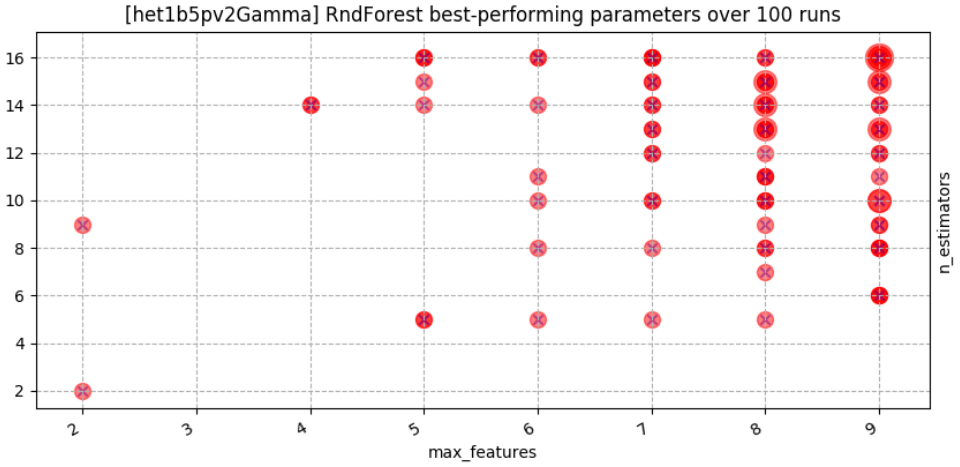


Figure A.39: *Grid-Search space showing the best-performing values for RndForest's parameters (max_features, n_estimators) on the het1b5pv2Gamma dataset.*

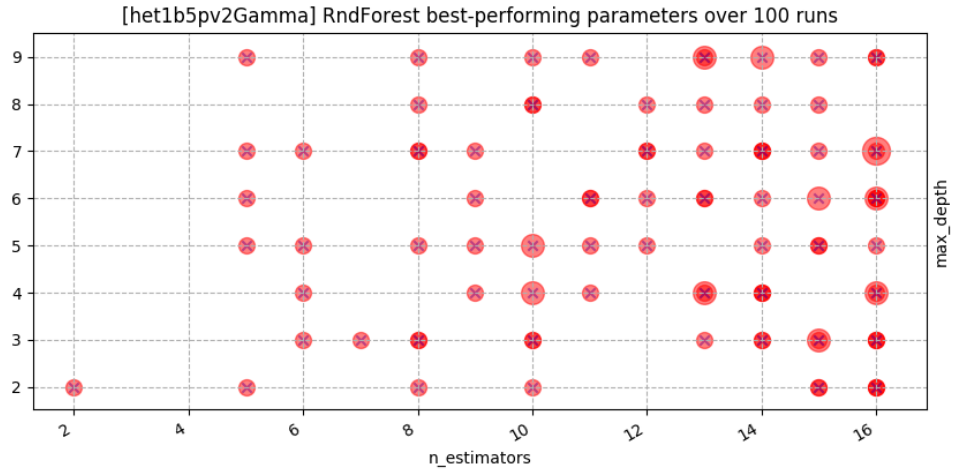


Figure A.40: *Grid-Search space showing the best-performing values for RndForest's parameters (n_estimators, max_depth) on the het1b5pv2Gamma dataset.*

Appendix B

BSD3-License code used to produce figure 2.1

Python code used to produce comparison graphics from Figure 2.1

```
#!/usr/bin/python
# -*- coding: utf-8 -*-

# Code source: Gaël Varoquaux
#              Andreas Müller
# Modified for documentation by Jaques Grobler
# Modified for inclusion into a thesis by Jorge Sacristan
# License: BSD 3 clause

import numpy as np
import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.datasets import make_moons, make_circles, make_classification
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis,
     QuadraticDiscriminantAnalysis
from sklearn.neural_network import MLPClassifier
from sklearn.ensemble import VotingClassifier


my_knn = KNeighborsClassifier(3)
my_svm = SVC(kernel='rbf', gamma=2, C=1, probability=True)
my_mlp = MLPClassifier(hidden_layer_sizes=(12, 4), max_iter=500, verbose=False, tol=1e-5,
     activation='relu')

classifiers=[("Decision Tree",DecisionTreeClassifier(max_depth=5)),
             ("Random Forest",RandomForestClassifier(max_depth=5,n_estimators=10,max_features=1)),
             ("LDA",          LinearDiscriminantAnalysis()),
             ("QDA",          QuadraticDiscriminantAnalysis()),
             ("Naive Bayes",  QuadraticDiscriminantAnalysis()),
             ("KNN (3)",      my_knn),
             ("Linear SVM",   SVC(kernel="linear", C=0.025, probability=True)),
             ("RBF SVM",      my_svm),
             ("Multilayered Perceptron", my_mlp),
             ("AdaBoost",     AdaBoostClassifier()),
             ("SoftVoting (KNN,SVM,MLP)",VotingClassifier(estimators=[('knn', my_knn),
                                                                      ('svc', my_svm),
                                                                      ('knn', my_mlp),],
                                                          voting='soft', weights=[1, 1, 1,],)),]

X, y = make_classification(n_features=2, n_redundant=0, n_informative=2, random_state=1,
                           n_clusters_per_class=1)
rng = np.random.RandomState(2)
X += 2 * rng.uniform(size=X.shape)
linearly_separable = (X, y)

```

```
ds_names = ["Moons dataset", "Circles dataset", "Linearly separable dataset"]
datasets = [make_moons(noise=0.3, random_state=0),
            make_circles(noise=0.2, factor=0.5, random_state=1),
            linearly_separable]

# Aesthetic options
marker_train = "*"
marker_test = "o"
alpha_train = 0.9
alpha_test = 0.7
alpha_cm = 0.7
my_fig_size = (7, 9.5)
grid_resolution = 0.02  # step size in the mesh
fig_rows = int(np.ceil(np.sqrt(len(classifiers))))
fig_cols = int(np.ceil(len(classifiers) / fig_rows))

for ds_idx in range(len(datasets)):
    # Peprocess dataset, split into training and test part
    ds = datasets[ds_idx]
    X, y = ds
    X = StandardScaler().fit_transform(X)
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)

    x_min, x_max = X[:, 0].min() - .5, X[:, 0].max() + .5
    y_min, y_max = X[:, 1].min() - .5, X[:, 1].max() + .5
    meshgrid_x, meshgrid_y = np.meshgrid(np.arange(x_min, x_max, grid_resolution),
                                         np.arange(y_min, y_max, grid_resolution))

    # just plot the dataset first
    my_colourmap = plt.cm.RdBu
    cm_bright = ListedColormap(['#FF0000', '#0000FF'])
    fig, axes = plt.subplots(fig_rows, fig_cols, figsize=my_fig_size)

    # Plot the training points
    axes[0, 0].scatter(X_train[:, 0], X_train[:, 1], c=y_train, cmap=cm_bright,
                       alpha=alpha_train, marker=marker_train)

    # and testing points
    axes[0, 0].scatter(X_test[:, 0], X_test[:, 1], c=y_test, cmap=cm_bright,
                       alpha=alpha_test, marker=marker_test)

    axes[0, 0].set_title("Synthetic dataset")

    # iterate over classifiers
    ax_index_row = 0
    ax_index_col = 1

    # for name, clf in zip(clf_names, clf_objects):
    for name, clf in classifiers:
        clf.fit(X_train, y_train)
        score = clf.score(X_test, y_test)

        # Plot the decision boundary. For that, we will assign a color to each
        # point in the mesh [x_min, m_max]x[y_min, y_max].
        if hasattr(clf, "decision_function"):
            Z = clf.decision_function(np.c_[meshgrid_x.ravel(), meshgrid_y.ravel()])
        else:
            Z = clf.predict_proba(np.c_[meshgrid_x.ravel(), meshgrid_y.ravel()])[:, 1]

        # Put the result into a color plot
        Z = Z.reshape(meshgrid_x.shape)
        axes[ax_index_row, ax_index_col].contourf(meshgrid_x, meshgrid_y, Z,
                                                  cmap=my_colourmap, alpha=alpha_cm)

        # Plot also the training points
        axes[ax_index_row, ax_index_col].scatter(X_train[:, 0], X_train[:, 1], c=y_train,
                                                 cmap=cm_bright, alpha=alpha_train, marker=
    marker_train)

        # and testing points
        axes[ax_index_row, ax_index_col].scatter(X_test[:, 0], X_test[:, 1], c=y_test,
                                                 cmap=cm_bright, alpha=alpha_test, marker=
    marker_test)

        axes[ax_index_row, ax_index_col].set_title(name)
        axes[ax_index_row, ax_index_col].text(meshgrid_x.max() - .3,
                                              meshgrid_y.min() + .3,
                                              ('%.2f' % score).lstrip('0'),
                                              size=15, horizontalalignment='right')

        if ax_index_col + 1 < fig_cols:
            ax_index_col += 1
        else:
            ax_index_col = 0
            ax_index_row += 1

```

```
135     # Aesthetic setup
136     for ax_row in axes:
137         for ax in ax_row:
138             ax.set_xticks(())
139             ax.set_yticks(())
140             ax.set_xlim(meshgrid_x.min(), meshgrid_x.max())
141             ax.set_ylim(meshgrid_y.min(), meshgrid_y.max())
142             # ax.axis('off')
143 
144     fig.subplots_adjust(left=0.02, right=0.98, bottom=0.01, top=0.97)
145 
146 plt.show()
```