

Research article

Mathematical modeling and machine learning for public health decision-making: the case of breast cancer in Benin

Cyrille Agossou¹, Mintodê Nicodème Atchadé^{1,2,3,*}, Aliou Moussa Djibril¹ and Svetlana Vladimirovna Kurisheva³

¹ National Higher School of Mathematics Genius and Modelization, National University of Sciences, Technologies, Engineering and Mathematics, Abomey, Benin Republic

² University of Abomey-Calavi/ International Chair in Mathematical Physics and Applications (ICMPA : UNESCO-Chair), 072 BP 50 Cotonou, Benin Republic

³ Saint-Petersburg State University of Economics, Department of Statistics and Econometrics, Russian Federation

* **Correspondence:** Email: nickpowerabc@gmail.com.

Abstract: Breast cancer is the most common type of cancer in women. Its mortality rate is high due to late detection and cardiotoxic effects of chemotherapy. In this work, we used the Support Vector Machine (SVM) method to classify tumors and proposed a new mathematical model of the patient dynamics of the breast cancer population. Numerical simulations were performed to study the behavior of the solutions around the equilibrium point. The findings revealed that the equilibrium point is stable regardless of the initial conditions. Moreover, this study will help public health decision-making as the results can be used to minimize the number of cardiotoxic patients and increase the number of recovered patients after chemotherapy.

Keywords: breast cancer; tumor classification; cardiotoxicity; machine learning; mathematical modeling; numerical simulations

1. Introduction

Breast cancer is a malignant tumor of mammary gland [1]. In other words, it is a cancer that originates in the cell units whose function is to secrete milk, the ductal-lobular units of the breast, mainly in women. Eight out of ten breast cancers occur after the age of 50.

According to Ly et al. [2], breast cancer is the most frequent cancer in women. As reported by the World Health Organization (WHO) [3], there were 2.3 million women diagnosed with breast cancer and 685 000 deaths in 2020. In addition, as of December 2020, there were 7.8 million women alive

who were diagnosed with breast cancer in the past 5 years. It is among the four most common cancers in the world and is the leading cause of death among women in developed countries and the second leading cause of death in developing countries [4]. Breast cancer affects more women than men. It occurs earlier in black women and is often diagnosed at a complicating stage. It is the second most common cancer after cervical cancer in women in sub-Saharan Africa [2].

In Benin, according to the registry of the National Hospital and University Center of Cotonou christened CNHU-HKM (Centre National Hospitalier Universitaire Hubert Koutoukou MAGA), breast cancer represents more than 32.5% [5] of registered patients and the mortality rate related to this disease is very high, particularly due to late diagnosis. Although screening for this disease is easy, its technique is still not well known by the Beninese population.

There is no single cause of breast cancer. Various factors, both genetic and environmental, can influence the onset and development of breast cancer. Scientists estimate that 5–10% of breast cancer cases are linked to inherited genetic mutations [6]. Age is another risk factor for developing the disease [7]. Other risk factors for breast cancer are being female, exposure to ionising radiation before the age of 35, obesity, nulliparity (absence of pregnancy), hormone replacement therapy after the menopause, having had a first breast cancer, absence of breastfeeding would also favor this cancer [8–10]. There are two methods of early detection: early diagnosis and screening. Timely diagnosis allows to detect the earliest signs and symptoms in the population and facilitate treatment, reduce the likelihood of cancer spread and increase the chances of a complete cure [11]. Screening is systematic in presumed asymptomatic populations, with the aim of identifying individuals with possible cancerous abnormalities in the mammary gland as early as possible [12]. Screening methods include mammography, which is the only one that has proven to be effective, and breast self-examination [13].

Through screening, some tumors can be discovered before symptoms of cancer appear. This is the purpose of screening: to detect cancer through a systematic examination before the first signs of cancer appear. This is essential because a cancer that is detected early will be better treated and the chance of recovery increases. For the detection of the tumor during diagnosis, several techniques are used and have already proven their effectiveness. To help doctors in this decision making, scientists have studied this issue and decided to use artificial intelligence (AI) for detection with a high accuracy. In this paper [14], it is about an AI that is superchampion in the detection of breast cancer. This proves the maximum reduction of diagnostic errors thanks to the use of Machine Learning. Others have also performed detection analyses using existing algorithms such as KNN (K-nearest neighbors), SVM, naive Bayes classification. Consequently they have performed tumor classification to help in decision making. According to [15], the KNN algorithm is a supervised learning method. It can be used for regression as well as for classification. According to [16], support vector machines (SVMs) are a set of supervised learning techniques and are particular linear classifiers which are based on the margin maximization principle. They performed structural risk minimization, which improves the complexity of the classifier with the aim of achieving excellent generalization performance. The SVM accomplishes the classification task by constructing, in a higher dimensional space, the hyperplane that optimally separates the data into two categories. It is mostly used in classification problems [17]. Thus, thanks to these existing algorithms, a classification with a percentage of accuracy going in the order of 96%-97% has been achieved. They used R and Python to perform the analysis. The highest accuracies were obtained with the naive Bayes classification algorithm. All this information is available on [18]. Principal component analysis was performed with a much greater focus on model accuracy.

Cancer stages determine cancer severity. The method used by physicians to describe the stage of cancer is the TNM (Tumor, Node, Metastasis) system. This system uses three criteria to determine the stage of cancer: tumor size, spread to lymph nodes and spread to other organs (metastasis) [19]. The process of curing breast cancer is easy if the cancer is detected at an early stage. The higher the stage, the lower the chances of cure. There are many treatments available to fight breast cancer: surgery, chemotherapy, radiotherapy, targeted treatments. The treatment's choice is made by a multidisciplinary team, depending on the stage of the breast cancer. In Benin, radiotherapy is not yet used, hormone therapy is targeted and excessively expensive, and finally chemotherapy is the most commonly used cancer treatment [5]. The latter is used to eliminate cancer cells or prevent cancer cells from receiving signals for cell division [20]. Chemotherapy is a cancer treatment that uses several drugs, unfortunately having adverse effects on the body of those with a health history [21]. This can lead to heart problems and even death.

Mathematical modeling can be used to study the dynamics of a disease. The history of mathematical modeling of cancer has been explained in numerous papers [22–24]. Mathematical modeling of cancer continues to develop to date. Dixit et al. [25] discussed the mathematical model of chemotherapy for cancer treatment. The model consists of tumor cells and energy specific to Adriamycin (chemotherapy drugs), while the energy of tumor cells depends on various tumor cells. Schättler et al. [26] proposed a mathematical model with minimal parameters for low-dose chemotherapy. They took into account angiogenic signals between tumors and the vasculature as well as tumor inhibitory effects from tumor-immune system interactions. Jordaõ et Tavares [27] constructed a deterministic mathematical model that is derived from biochemical models of a human cell in two distinct cases, for comparison: healthy cells and cancer cells. Mahlbacher et al. [28] used mathematical modeling for experimental investigation that is suggested to measure and predict interactions between tumor and immunity. Some studies specifically dealing with mathematical models of breast cancer are as follows. A mathematical model of breast cancer development, local treatment and recurrence, by Enderling et al. has been established in [29]. The breast cancer model uses intermediate cell clonal expansion growth rates and mutation rates as parameters and constructs two and six-step models to fit the age-specific incidence of breast cancer in the Surveillance, Epidemiology, and End Results Registry (SEER) by Zhang, X. et al. [30]. The cancer treatment model of radiation and chemotherapy was discussed by Liu et Yang [31]. The breast cancer model focuses on its heterogeneity and the role of mathematical modeling and simulation in revealing the underlying biophysical processes by Simmons et al. [32]. Segun Isaac Oke et al. [33] discussed about a mathematical model of breast cancer governed by a system of ordinary differential equations in the presence of chemotherapy treatment and ketogenic diet. In previous studies, Fathoni et al. [34] presented a mathematical model of breast cancer at the patient population level. This study deals with mathematical models at the population level of cancer patients. The model designed is a mathematical model of breast cancer staging analysis with side effects on the heart of chemotherapy patients. The model includes subpopulations of stage 1 and 2 (A), stage 3 (B), stage 4 (C), disease free (D) and cardiotoxic (E). In this study, given data from Benin, we constructed the model by separating subpopulation A into two subpopulations A_1 (stage 1) and A_2 (stage 2), making six subpopulations.

In this paper, we used the Support Vector Machine and improved its accuracy compared to previous works and thus achieved a better classification and output of decision probabilities. The classification prediction performance with SVM in some earlier studies has been presented in the discussion. In addition, we used the compartmental model applied to Benin data by building a new six-compartmental

model. The six subpopulations are modeled by forming a system of differential equations. Equilibrium and stability point analysis were used to determine the dynamics of the six populations as a function of time. Numerical simulations were carried out to visualize the dynamic analysis results.

2. Materials and methods

2.1. Materials

2.1.1. Materials for tumor classification

In this work, it will be an analysis on the breast cancer dataset. The data used are diagnostic data from Wisconsin patients. We collected them on [18]. We used these data because we do not have data on these tumor characteristics from Benin patients.

The features are calculated from a scanned image of a fine needle aspiration (FNA) of a breast mass. They describe the characteristics of the cell nuclei present in the image. The three-dimensional space is the one described in *Robust linear programming discrimination of two linearly inseparable sets* [35]. This database is available on the UCI (University of California Irvine) Machine Learning Repository [18].

Attribute information:

1. Identification number
2. Diagnosis (M = malignant, B = benign)

Ten real-valued features are calculated for each cell nucleus: radius (average of distances between center and perimeter points), texture (standard deviation of gray scale values), perimeter, area, regularity (local variation in radius lengths), compactness ($perimeter^2/area - 1.0$), concavity (severity of concave portions of the contour), concave points (number of concave portions of the contour), symmetry, fractal dimension (approximation of coastline - 1).

The mean, standard error, least and greatest (average of the three largest values) of these features were calculated for each image, resulting in 30 features.

Missing attribute values : none

Class distribution: 357 benign, 212 malignant. The description of the variables is presented in Table 1.

The SVM analysis process is summarized in Figure 1.

2.1.2. Materials for modeling cardiotoxicity

This part of the work is based on the study "Management of Breast Cancer in Visceral Surgery of CNHU-HKM of Cotonou in Benin [5]. A complete census of breast cancer patients admitted to the university visceral surgery clinics A and B of CNHU-HKM during the study period was conducted. The sample was exhaustive, taking into account all breast cancer cases. Patients with pathologically confirmed breast cancer; advanced breast tumor with all clinical features of malignancy; or those who had a complete medical record were included in this study. Patients with benign breast tumors and patients with incomplete medical records were excluded.

Table 1. Description of variables used in SVM.

Variables	Description
Diagnosis	Diagnosis of breast tissues (M = malignant, B = benign)
Radius_mean	Mean of distances from center to points on the perimeter
Texture_mean	Standard deviation of gray-scale values
Perimeter_mean	Mean size of the core tumor
Area_mean	Average area of central tumor
Smoothness_mean	Mean of local variation in radius lengths
Compactness_mean	$\frac{(\text{mean of perimeter}^2)}{\text{area}-1,0}$
Concavity_mean	Mean of severity of concave portions of the contour
Concave points_mean	Mean of number of concave portions of the contour
Symmetry_mean	Mean of the symmetries with respect to the central tumor
Fractal_dimension_mean	Mean for "coastline approximation" - 1
Radius_se	Standard error for the mean of distances from center to points on the perimeter
Texture_se	Standard error for standard deviation of gray-scale values
Perimeter_se	Standard error for central tumor size
Area_se	Standard error for central tumor area
Smoothness_se	Standard error for local variation in radius lengths
Compactness_se	Standard error for $\frac{(\text{mean of perimeter}^2)}{\text{area}-1,0}$
Concavity_se	Standard error for severity of concave portions of the contour
Concave points_se	Standard error for number of concave portions of the contour
Symmetry_se	Standard error for symmetry values relative to the central tumor
Fractal_dimension_se	Standard error for "coastline approximation" - 1
Radius_worst	"Worst" or largest mean value for mean of distances from center to points on the perimeter
Texture_worst	"Worst" or largest mean value for standard deviation of gray-scale values
Perimeter_worst	"Worst" or largest mean for central tumor size
Area_worst	"Worst" or largest mean value for central tumor area
Smoothness_worst	"Worst" or largest mean value for local variation in radius lengths
Compactness_worst	"Worst" or largest mean for $\frac{(\text{mean of perimeter}^2)}{\text{area}-1,0}$
Concavity_worst	"Worst" or largest mean value for severity of concave portions of the contour
Concave points_worst	"Worst" or largest mean value for number of concave portions of the contour
Symmetry_worst	"Worst" or largest mean value for symmetries with respect to the central tumor
Fractal_dimension_worst	"Worst" or largest mean value for "coastline approximation" - 1

It is assumed that there were no healthy patients on first arrival at the hospital. At the time of the first medical record, patients are classified into subpopulations of stage 1, stage 2, stage 3 or stage 4. During the treatment process in the hospital, all cancer patients are assumed to receive chemotherapy treatment. Patients will experience changes in the severity of their cancer over time, so each sub-population will experience changes in numbers. There are patients whose disease worsens, others who recovered (disease free) and others who experienced cardiotoxic effects during chemotherapy. The changes in the population dynamics of these subpopulations are represented in a diagram called the compartment diagram (Figure 2).

The model is constructed from six compartments representing subpopulations of breast cancer patients. Each sub-population is represented by the variables A_1, A_2, B, C, D and E . Subpopulation A_1 represents patients with stage 1 cancer. Subpopulation A_2 represents patients with stage 2A and 2B cancer. Subpopulation B represents patients with stage 3A and 3B cancer. Subpopulation C represents patients with stage 4 cancer. Subpopulation D represents patients without disease after chemotherapy. In the absence of disease, the cancer is no longer visible on observation. Subpopulation E represents cancer patients with cardiotoxicity.

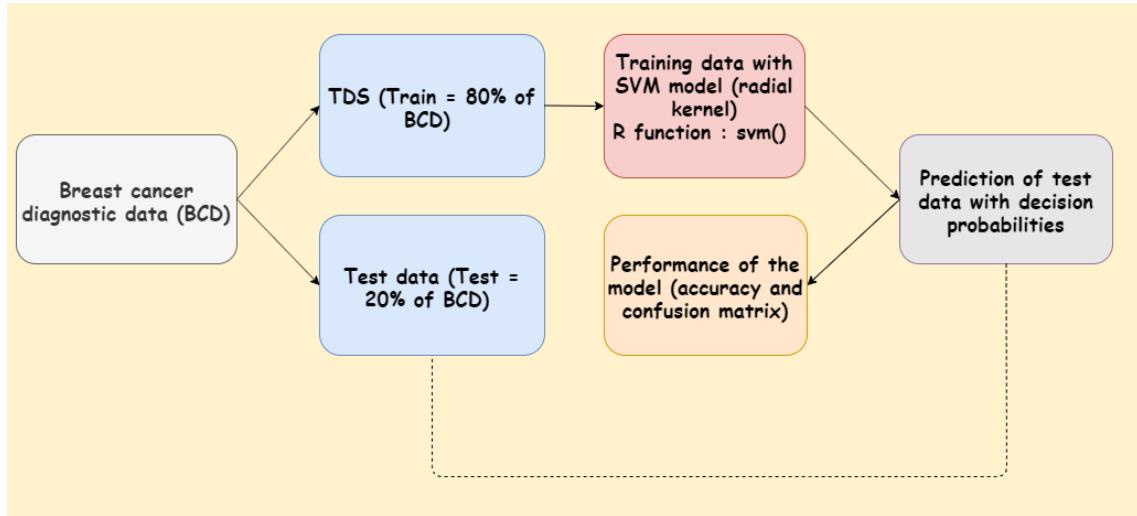


Figure 1. SVM Process diagram.

2.1.3. Methods for modeling cardiotoxicity

The rate of change of each subpopulation is expressed in a system of differential equations as in Eq 2.1. Be the set defined by $\Omega = \{(A_1, A_2, B, C, D, E) \mid A_1 > 0, A_2 > 0, B > 0, C > 0, D > 0, E > 0\}$. That is, all solutions satisfy $(A_1(t), A_2(t), B(t), C(t), D(t), E(t))$ for any t if $(A_1(0), A_2(0), B(0), C(0), D(0), E(0))$.

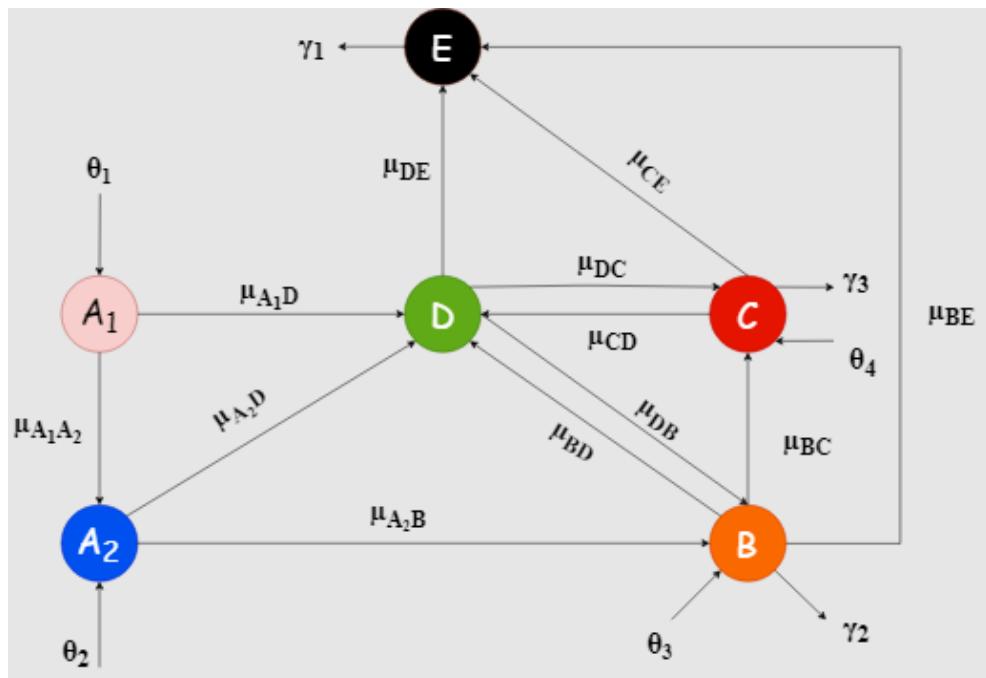


Figure 2. Compartment diagram.

2.2. Methods

2.2.1. Methods for tumor classification

Before performing the tumor classification, we will first look at the correlation between the variable Diagnosis and the other variables to see which ones are correlated with the latter to refine the model. Then we will have to use the SVM model to perform the classification. The SVM model will allow us to classify into two categories by drawing a hyper plane to separate the two classes.

SVM model

Consider the pattern classifier, that uses a hyperplane to separate two classes of patterns based on the given examples $\{(x_i, y_i)\}, i = 1, \dots, n$, where $(x_i, y_i) \in \mathbb{R}^n \times \{-1, 1\}$. SVM is a machine learning method that classifies binary groups by finding and using a class boundary [36]. The training data samples along the hyper planes near the class boundary are called support vectors, and the margin is the distance between the support vectors and the class boundary hyperplanes. The SVM is based on the concept of decision planes that define decision boundaries. A decision plane is one that separates the assets of objects having different class memberships. SVM is a useful technique for data classification. A classification task usually involves training and testing data. Each instance in the training set contains one “target value” (class label) and several “attributes” (features).

Given a training set of instance label pairs $(x_i, y_i), i = 1, \dots, n$, where $(x_i, y_i) \in \mathbb{R}^n \times \{-1, 1\}$, the SVM requires the solution of the following optimization problem:

$$\min_{\vec{w}, b} \frac{1}{2} \|\vec{w}\|^2 + c \sum_{i=1}^n \epsilon_i$$

subject to

$$\begin{cases} y_i(\vec{w} \cdot \phi(x_i) + b) > 1 - \epsilon_i \\ \text{for all } 1 \leq i \leq n, b \in \mathbb{R} \text{ and } \epsilon_i \geq 0 \end{cases}$$

Here training vectors x_i are mapped into a higher dimensional space by the function ϕ . Then SVM finds a linear separating hyper plane of the equation $\vec{w} \cdot \vec{x} + b = 0$; $\vec{w} \in \mathbb{R}^n, b \in \mathbb{R}$ (learns from the training data set and then applies its knowledge to classify the unknown data) with the maximal margin in this higher dimensional space and a penalty parameter of the error term ϵ . Furthermore, $k(x_i, x_j) = \phi(x_i) \phi(x_j)$ is called the kernel function. There are a number of kernels that can be used in SVM models. These include linear polynomial, Radial Basis Function (RBF) and sigmoid. In our case, we use RBF classifiers $k(x_i, x_j) = \exp(-\gamma|x_i - x_j|^2)$ from [36].

The SVM is the maximum margin hyper plane that lies in some space. The original SVM is a linear classifier. For SVMs, using the kernel trick makes the maximum margin hyper plane fit in a feature space. The feature space is a nonlinear map from the original input space, usually of much higher dimensionality than the original input space. In this way, nonlinear SVMs can be created. Support vector machines are an innovative approach to constructing learning machines that minimize the generalization error [36]. They are constructed by locating a set of planes that separate two or more classes of data. Constructing these planes, the SVM discovers the boundaries between the input classes; the elements of the input data that define these boundaries are called support vectors. The kernel is then modified in a data dependent way by using the obtained support vectors. The modified

kernel is used to get the final classifier.

Implementation in R software

The package that we used for implementing SVM algorithm in R is e1071. The function used is *svm()*. First, we create the training and testing data sets. The function used is *CreateDataPartition()*. We use 80% for training and 20% for test in our study. Second, we train the model on our training dataset. The target function is *svm()*. Third, we use our trained model to make a prediction based on our testing data. In R, we use the function *predict()*. The function *attr(pred_svm, "probabilities")* allows to output the probabilities. To see the performance, we compute the confusion matrix. The main SVM parameters used are in Tables 2, 3, 4.

Table 2. Description of variables of *CreateDataPartition()* function.

Parameters	Description	Values
<i>y</i>	A vector of outcomes	Diagnosis variable in patients data
Times	The number of partitions to create	1
<i>p</i>	The percentage of data that goes to training	0.8
list	Logical – should the results be in a list (TRUE) or a matrix with the number of rows equal to <i>p</i> *length(<i>y</i>) and times columns	False

Table 3. Description of variables of *svm()* function.

Parameters	Description	Values
train	The training set as a data.frame	train_data
labels	Class labels of the training set	
gamma	The gamma parameter (if a vector, cross-over validation is used to choose the best size)	$2^{(-3:3)}$: default in R
cost	The cost parameter (if a vector, cross-over validation is used to choose the best size)	$2^{(-3:3)}$: default in R
Kernel	Kernel type	Radial

Table 4. Description of variables of *predict()* function.

Parameters	Description	Values
Object	A model object for which prediction is desired	Test_data
Model	Model used	<i>svm()</i>
Probability	Allow to show probabilities of prediction of object	TRUE

$$\left\{ \begin{array}{l} \frac{dA_1}{dt} = \theta_1 - \mu_{A_1A_2}A_1 - \mu_{A_1D}A_1 \\ \frac{dA_2}{dt} = \theta_2 - \mu_{A_2B}A_2 - \mu_{A_2D}A_2 \\ \frac{dB}{dt} = \theta_3 - \mu_{A_2B}A_2 - \mu_{BC}B - \mu_{BD}B - \mu_{BE}B - \gamma_2B + \mu_{DB}D \\ \frac{dC}{dt} = \theta_4 + \mu_{BC}B - \mu_{CD}C - \mu_{CE}C - \gamma_3C + \mu_{DC}D \\ \frac{dD}{dt} = \mu_{A_1D}A_1 + \mu_{A_2D}A_2 + \mu_{BD}B + \mu_{CD}C - \mu_{DC}D - \mu_{DB}D - \mu_{DE}D \\ \frac{dE}{dt} = \mu_{BE}B + \mu_{CE}C + \mu_{DE}D - \mu_{DB}D - \gamma_1E \end{array} \right. \quad (2.1)$$

The parameters in Eq (2.1) are assumed to have positive values and are described in Table 5.

Patients who received treatment for the first time and were diagnosed with stage 1 cancer are included in subpopulation A_1 with rate θ_1 . Patients in subpopulation A_1 who underwent chemotherapy have two possibilities, namely recovery (disease-free) with rate μ_{A_1D} or worse with rate $\mu_{A_1A_2}$.

Patients whose tumor size already exceeds 2 cm but not more than 5 cm have been diagnosed with stage 2 cancer and are included in the A_2 subpopulation of θ_2 rate. Patients in subpopulation A_2 who have undergone chemotherapy have two possibilities, namely recovery (disease free) with rate μ_{A_2D} or worse with rate μ_{A_2B} .

The patients who were treated for the first time in the hospital were mostly suffering from stage 3 cancer, so they were grouped in subpopulation B with rate θ_3 . The axillary lymph nodes become increasingly enlarged and the cancer becomes inflammatory. Patients in subpopulation B may have died of cancer with a rate of γ_2 . Patients in this subpopulation, after undergoing chemotherapy, may become disease free with μ_{BD} and may also get worse with μ_{BC} . Subpopulation B with more intensive chemotherapy than subpopulations A_1 and A_2 may cause cardiotoxicity in patients with μ_{BE} .

Patients who are being treated for the first time may also fall into subpopulation C because the cancer has metastasized to other parts of the body (distant metastases), such as bone, liver, lung or brain. This is also known as metastatic breast cancer. In this condition, it is unlikely that chemotherapy treatment can cure the cancer, so the rate towards a disease-free state with a rate of μ_{CD} is assumed to be the least compared to μ_{A_2D} and μ_{BD} . Conversely, the μ_{CE} rate towards cardiotoxicity is assumed to be of high value as the patient undergoes very intensive chemotherapy. This subpopulation also experienced cancer death with a γ_3 rate.

The absence of disease in subpopulation D can be increased in patients in subpopulations A_1 , A_2 , B and C . The condition of this disease benefit may last indefinitely or only briefly. If it lasts only for a while, then patients in subpopulation D may revert to subpopulations of B and C with their respective rates of μ_{DB} and μ_{DC} . A prolonged period of subpopulation D may also experience direct cardiotoxicity to a level of μ_{DE} . Patients with cardiotoxicity or in subpopulation E may experience cardiac death with γ_1 .

2.2.2. Object of merging the two methods

The SVM classification will allow decision makers to do better in the diagnosis of breast cancer. With this classification, we can determine if the tumor noticed is malignant or benign. When the tumor is malignant, a pathological examination is done to identify if the cells are cancerous or not. If these cells are identified as cancerous, the stage of cancer is detected and treatment is started. The treatment often used is chemotherapy which is the cause of cardiotoxicity in some patients due to their health history. The analysis of the population dynamics will allow us to understand the effect of this

chemotherapy on the heart of certain patients. Thus, the screening being negative, the decision makers would take adequate measures so that once the cancer is noticed and the stage of this cancer also, a health check-up would be requested from this patient to identify his history of disease and genetic diseases related to the heart. After this assessment, chemotherapy treatment can be started based on this assessment to avoid cardiotoxic effects that can lead to death.

Table 5. Parameters description for cardiotoxicity mathematical modeling.

Parameter	Description
θ_1	Number of new patients diagnosed with stage 1 cancer
θ_2	Number of new patients diagnosed with stage 2 cancer
θ_3	Number of new patients diagnosed with stage 3 cancer
θ_4	Number of new patients diagnosed with stage 4 cancer
$\mu_{A_1A_2}$	Passage rate from stage 1 to stage 2 (progressive disease)
μ_{A_1D}	Rate of stage 1 patients who achieve a complete response
μ_{A_2B}	Rate of passage from stage 2 to stage 3 (progressive disease)
μ_{A_2D}	Rate of stage 2 patients who experience a complete response
μ_{BC}	Increased rate of passage from stage 3 to stage 4 (progressive disease)
μ_{BD}	Rate of stage 3 patients who achieve a complete response
μ_{BE}	Rate of stage 3 cancer chemotherapy patients who experience cardiotoxic
μ_{CD}	Rate of stage 3 patients who experience a complete response
μ_{CE}	Rate of stage 4 cancer chemotherapy patients who experience cardiotoxic
μ_{DB}	Rate of disease-free patients who relapse back to stage 3
μ_{DC}	Rate of disease-free patients who relapse back to stage 4
μ_{DE}	Rate of disease-free patients who experience cardiotoxic
γ_1	Death rate of cardiotoxic patients
γ_2	Death rate of stage 3 cancer patients
γ_3	Death rate of stage 4 cancer patients

The purpose of merging these two methods is to make doctors think of creating a system based on the health assessment of the individual. After screening and noticing that the tumor is malignant and the cells of the tumor are cancerous, physicians will be able to prescribe an adequate chemotherapy treatment that will not cause a cardiotoxicity based on the health assessment of the individual. Thus, they will save time.

3. Results

3.1. Tumor classification

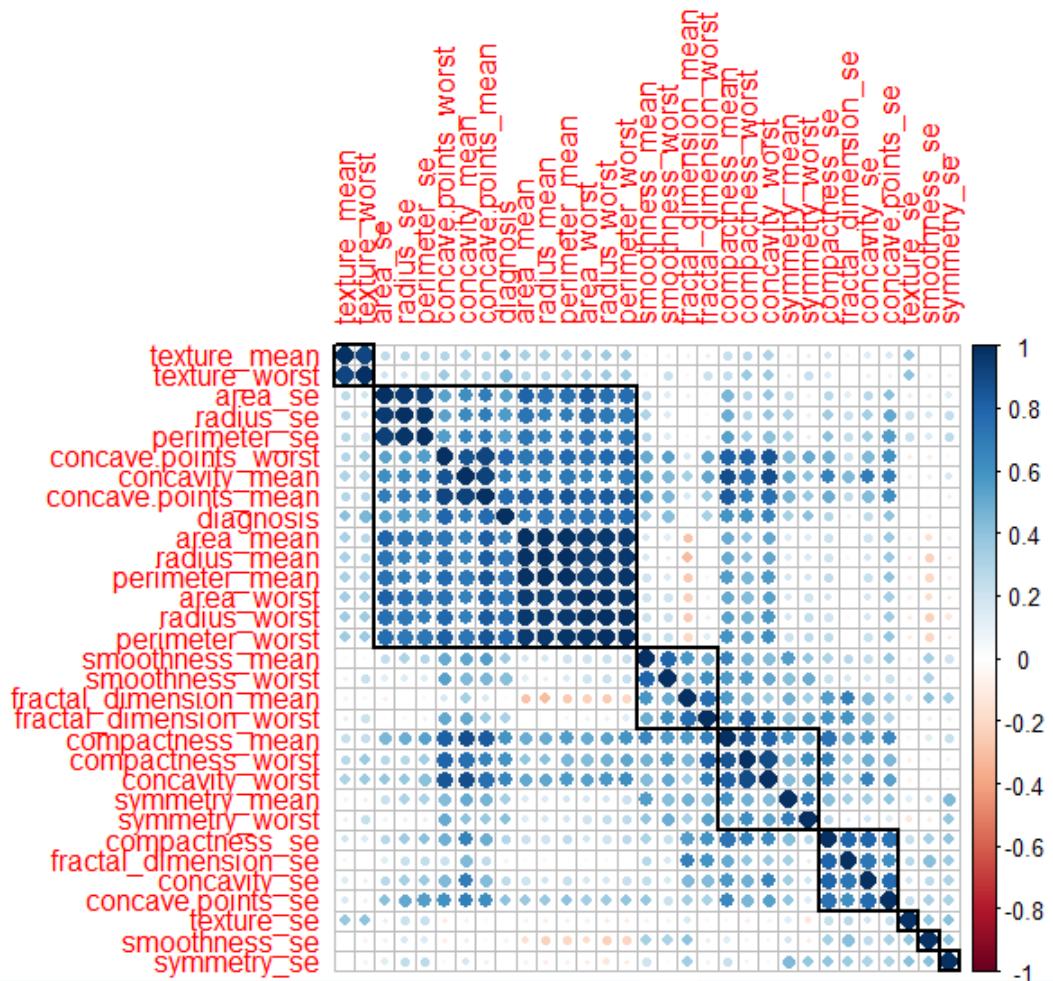


Figure 3. Correlation matrix.

From the correlation matrix (Figure 3), we notice that some variables are more correlated with the variable Diagnosis than others. We are going to run the learning test with a SVM model which has the particularity to draw a hyperplane that tries to divide the two categories of tumors as well as possible and make the classification possible. First, we did this on all variables and obtained the following results:

Confusion Matrix and Statistics		
Reference		
Prediction	B	M
B	69	0
M	2	42
Accuracy : 0.9823		
95% CI : (0.9375, 0.9978)		
No Information Rate : 0.6283		
P-Value [Acc > NIR] : <2e-16		
Sensitivity : 1.0000		
Specificity : 0.9718		
Pos Pred Value : 0.9545		
Neg Pred Value : 1.0000		
Prevalence : 0.3717		
Detection Rate : 0.3717		
Detection Prevalence : 0.3894		
Balanced Accuracy : 0.9859		

We note that the model can make this prediction with a confidence level of 98.23%. A doctor can with the help of the machine predict 98% that the tumor is malignant or benign. Now, let us put ourselves in the position of a doctor who has received a patient and wants to diagnose and has obtained the different characteristics of the tumor. Suppose that he gets the following characteristics:

Tumor characteristics 1

M B
0.0016686210 9.983314e-01

In this case, the doctor will be able to tell the patient that her tumor is 99.83% benign and that if she wishes to remove, it will be possible.

Tumor characteristics 2

M B
0.9998388423 1.611577e-04

Thus, in this case, he will be able to tell the patient 99.98% of the time that her tumor is malignant and that she should start treatment.

Tumor characteristics 3

M B
0.5131059859 4.868940e-01

In this case, the two probabilities are almost equal. The doctor will not be able to make decisions and refer to other diagnoses.

These cases are rare. With this model, the diagnosis will be made with a saving of time. According to the correlation matrix 3, let us try to select the variables that best explain the variable Diagnosis. The results obtained from this model are as follows:

Confusion Matrix and Statistics		
Reference		
	Prediction	B M
B	48	21
M	23	21
Accuracy : 0.6106		
95% CI : (0.5144, 0.7009)		
No Information Rate : 0.6283		
P-Value [Acc > NIR] : 0.6889		
Sensitivity : 0.5000		
Specificity : 0.6761		
Pos Pred Value : 0.4773		
Neg Pred Value : 0.6957		
Prevalence : 0.3717		
Detection Rate : 0.1858		
Detection Prevalence : 0.3894		
Balanced Accuracy : 0.5880		
'Positive' Class : M		

It can be seen that the percentage accuracy decreases. Thus keeping the variables would increase the accuracy of the model. Therefore it is better to keep the data so that the classification is better. We could also try to do a treatment of these data to look for a way to increase the percentage of accuracy while reducing the number of variables. With the tumor determined to be malignant, we can move on to the second part of the analysis of the effect of chemotherapy on the patient's heart by cancer stage.

3.2. Modeling of cardiotoxicity

The dynamic analysis using the system of Eq (2.1) includes the calculation of equilibrium points and stability analysis. The study of equilibrium and stability points were used to determine the dynamics of the six subpopulations over time. In the stability analysis of the equilibrium point, the Routh Hurwitz criterion is used. The Routh Hurwitz criterion determines the conditions for the polynomial roots of the left half plane (LHP) and cannot be used directly to study the stability of discrete time systems. The characteristic equation of this system is :

$$a_n \lambda^n + a_{n-1} \lambda^{n-1} + a_{n-2} \lambda^{n-2} + \cdots + a_1 \lambda + a_0 = 0 \quad (3.1)$$

For a polynomial of degree n , the Rough Hurwitz table is a matrix C with $n + 1$ rows and at least $\frac{n+1}{2}$ columns.

$$\begin{bmatrix} \lambda^n & C_{1,1} & C_{1,2} & C_{1,3} & \cdots \\ \lambda^{n-1} & C_{2,1} & C_{2,2} & C_{2,3} & \cdots \\ \lambda^{n-2} & C_{3,1} & C_{3,2} & C_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \\ \lambda^0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (3.2)$$

The first line of the table, indexed by λ^n , contains the coefficients a_n, a_{n-2}, \dots i.e., $C_{1,j} = a_{n+2-2j}$.

The second row of the table, indexed by λ^{n-1} , contains the coefficients $a_{n-1}, a_{n-3}, \dots, a_0$ i.e., $C_{1,j} = a_{n+1-2j}, j = 1, \dots$

For the line i , indexed by λ^{n-i} , the elements satisfy the following recurrent relation:

$$C_{i+1,j} = \frac{C_{i-1,j+1}C_{i,1} - C_{i-1,1}C_{i,j+1}}{C_{i,1}} \quad (3.3)$$

Necessary condition: all coefficients are positive or have the same sign. If one of the coefficients is zero or negative, and at least one other coefficient is positive, there is at least one imaginary root with a positive real part.

Sufficient condition: the system is stable when all $C_{i,1}$ in the rough table are positive or of the same sign. If there is a change of sign, the system is unstable.

The second aspect of the Rough Hurwitz criterion is that the real parts of the solutions obtained must all be less than 0.

Theorem: The equilibrium point $X^* = (X_1^*, X_2^*, \dots, X_n^*)$ is said to be asymptotically stable if the real parts of the eigenvalues are strictly negative.

X^* asymptotically stable $\Leftrightarrow \operatorname{Re}(\lambda_i) < 0 \forall i$.

Equilibrium point: Assume that:

$$\left\{ \begin{array}{l} \frac{dA_1}{dt} = 0 \\ \frac{dA_2}{dt} = 0 \\ \frac{dB}{dt} = 0 \\ \frac{dC}{dt} = 0 \\ \frac{dB}{dt} = 0 \\ \frac{dE}{dt} = 0 \end{array} \right. \quad (3.4)$$

It is assumed that the equilibrium point exists. From the system of equations, we obtain an equilibrium point, which is $(A_1^*, A_2^*, B^*, C^*, D^*, E^*)$, with $A_1^*, A_2^*, B^*, C^*, D^*$ and E^* given in the equations below. To obtain the equilibrium point, the following correspondence table (Table 6) has been made to shorten the parameters.

Table 6. Parameters correspondences.

Parameter	Correspondence
$\mu_{A_1A_2}$	a
μ_{A_1D}	b
μ_{A_2B}	c
μ_{A_2D}	d
μ_{BC}	e
μ_{BD}	f
μ_{BE}	g
μ_{CD}	h
μ_{CE}	i
μ_{DB}	j
μ_{DC}	k
μ_{DE}	l
$a + b$	$k1$
$c + d$	$k2$
$e + f + g + \gamma_2$	$k3$
$h + i + \gamma_3$	$k4$
$k + j + l$	$k5$
γ_1	$k6$
$\frac{c}{k2}$	m
$\frac{b}{k1}$	n
$\frac{d}{k2}$	o
$-\theta_3 + m\theta_2$	s
$-n\theta_1 - o\theta_2$	t

After various calculations, we obtain the equilibrium point:

$$\left\{
 \begin{aligned}
 A_1^* &= \frac{\theta_1}{k1} \\
 A_2^* &= \frac{\theta_2}{k2} \\
 B^* &= -\frac{h \times k \times s - j \times k4 \times t - k4 \times k5 \times s + h \times j \times \theta_4}{f \times j \times k4 + h \times k \times k3 - k3 \times k4 \times k5 + h \times j \times e} \\
 C^* &= \frac{f \times k \times s + k \times k3 \times t + f \times j \times \theta_4 - k3 \times k5 \times \theta_4 + j \times t \times e + k5 \times s \times e}{f \times j \times k4 + h \times k \times k3 - k3 \times k4 \times k5 + h \times j \times e} \\
 D^* &= \frac{f \times k4 \times s + k3 \times k4 \times t - h \times k3 \times \theta_4 + h \times s \times e}{f \times j \times k4 + h \times k \times k3 - k3 \times k4 \times k5 + h \times j \times e} \\
 E^* &= \frac{k6 \times (f \times j \times k4 + h \times k \times k3 - k3 \times k4 \times k5 + h \times j \times e) + g \times j \times k4 \times t + i \times k \times k3 \times t + g \times k4 \times k5 \times s + l \times k3 \times k4 \times t + f \times i \times j \times \theta_4 - g \times h \times j \times \theta_4}{h \times l \times k3 \times \theta_4 - i \times k3 \times k5 \times \theta_4} \\
 &\quad - \frac{k6 \times (f \times j \times k4 + h \times k \times k3 - k3 \times k4 \times k5 + h \times j \times e)}{k6 \times (f \times j \times k4 + h \times k \times k3 - k3 \times k4 \times k5 + h \times j \times e)}
 \end{aligned} \right. \quad (3.5)$$

Stability

Next, we look for the stability of the equilibrium point. To determine the stability of the equilibrium point, the equation is first given in matrix form.

$$\mathcal{F} = \mathcal{K}\mathcal{G} + \Theta \quad (3.6)$$

With :

$$\mathcal{F} = \begin{bmatrix} \dot{A}_1 \\ \dot{A}_2 \\ \dot{B} \\ \dot{C} \\ \dot{D} \\ \dot{E} \end{bmatrix}; \quad \mathcal{K} = \begin{bmatrix} -k1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -k2 & 0 & 0 & 0 & 0 \\ 0 & -\mu_{A_2B} & k3 & 0 & \mu_{DB} & 0 \\ 0 & 0 & \mu_{BC} & -k4 & \mu_{DC} & 0 \\ \mu_{A_1D} & \mu_{A_2D} & \mu_{BD} & \mu_{CD} & -k5 & 0 \\ 0 & 0 & \mu_{BE} & \mu_{CE} & \mu_{DE} & -k6 \end{bmatrix}; \quad \mathcal{G} = \begin{bmatrix} A_1 \\ A_2 \\ B \\ C \\ D \\ E \end{bmatrix} \text{ and } \Theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} \dot{A}_1 \\ \dot{A}_2 \\ \dot{B} \\ \dot{C} \\ \dot{D} \\ \dot{E} \end{bmatrix} = \begin{bmatrix} -k1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -k2 & 0 & 0 & 0 & 0 \\ 0 & -\mu_{A_2B} & k3 & 0 & \mu_{DB} & 0 \\ 0 & 0 & \mu_{BC} & -k4 & \mu_{DC} & 0 \\ \mu_{A_1D} & \mu_{A_2D} & \mu_{BD} & \mu_{CD} & -k5 & 0 \\ 0 & 0 & \mu_{BE} & \mu_{CE} & \mu_{DE} & -k6 \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ B \\ C \\ D \\ E \end{bmatrix} + \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ 0 \\ 0 \end{bmatrix} \quad (3.7)$$

We will use the Rough-Hurwitz criterion to determine the stability of the system. The matrix \mathcal{K} is a matrix of coefficients. The characteristic equation is calculated using $\det(\mathcal{K} - \lambda I)$. We have:

$$\begin{vmatrix} -k1 - \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & -k2 - \lambda & 0 & 0 & 0 & 0 \\ 0 & -\mu_{A_2B} & k3 - \lambda & 0 & \mu_{DB} & 0 \\ 0 & 0 & \mu_{BC} & -k4 - \lambda & \mu_{DC} & 0 \\ \mu_{A_1D} & \mu_{A_2D} & \mu_{BD} & \mu_{CD} & -k5 - \lambda & 0 \\ 0 & 0 & \mu_{BE} & \mu_{CE} & \mu_{DE} & -k6 - \lambda \end{vmatrix} \quad (3.8)$$

The resulting polynomial is tabulated as follows (Table 7):

Table 7. Resulting characteristic polynomial.

Monomes	Coefficients	Separator
λ^6	1	+
λ^5	$k1 + k2 + k3 + k4 + k5 + k6$	+
λ^4	$-\mu_{BD} \times \mu_{DB} - \mu_{CD} \times \mu_{DC} + k1 \times k2 + k3 \times k4 + k3 \times k5 + k3 \times k6 + k4 \times k5 + k4 \times k6 + k5 \times k6 + (k1 + k2) \times (k3 + k4 + k5 + k6)$	+
λ^3	$(k1 + k2) \times (-\mu_{BD} \times \mu_{DB} - \mu_{CD} \times \mu_{DC} + k3 \times k4 + k3 \times k5 + k3 \times k6 + k4 \times k5 + k4 \times k6 + k5 \times k6) + k1 \times k2 \times (k3 + k4 + k5 + k6) - \mu_{BC} \times \mu_{CD} \times \mu_{DB} - \mu_{BD} \times \mu_{DB} \times k4 - \mu_{BD} \times \mu_{DB} \times k6 - \mu_{CD} \times \mu_{DC} \times k3 - \mu_{CD} \times \mu_{DC} \times k6 + k3 \times k4 \times k5 + k3 \times k4 \times k6 + k3 \times k5 \times k6 + k4 \times k5 \times k6$	-
λ^2	$(k1 + k2) \times (\mu_{BC} \times \mu_{CD} \times \mu_{DB} + \mu_{BD} \times \mu_{DB} \times k4 + \mu_{BD} \times \mu_{DB} \times k6 + \mu_{CD} \times \mu_{DC} \times k3 + \mu_{CD} \times \mu_{DC} \times k6 - k3 \times k4 \times k5 - k3 \times k4 \times k6 - k3 \times k5 \times k6 - k4 \times k5 \times k6) - k1 \times k2 \times (-\mu_{BD} \times \mu_{DB} - \mu_{CD} \times \mu_{DC} + k3 \times k4 + k3 \times k5 + k3 \times k6 + k4 \times k5 + k4 \times k6 + k5 \times k6) + \mu_{BC} \times \mu_{CD} \times \mu_{DB} \times k6 + \mu_{BD} \times \mu_{DB} \times k4 \times k6 + \mu_{CD} \times \mu_{DC} \times k3 \times k6 - k3 \times k4 \times k5 \times k6$	-
λ^1	$(k1 + k2) \times (\mu_{BC} \times \mu_{CD} \times \mu_{DB} \times k6 + \mu_{BD} \times \mu_{DB} \times k4 \times k6 + \mu_{CD} \times \mu_{DC} \times k3 \times k6 - k3 \times k4 \times k5 \times k6) + k1 \times k2 \times (\mu_{BC} \times \mu_{CD} \times \mu_{DB} + \mu_{BD} \times \mu_{DB} \times k4 + \mu_{BD} \times \mu_{DB} \times k6 + \mu_{CD} \times \mu_{DC} \times k3 + \mu_{CD} \times \mu_{DC} \times k6 - k3 \times k4 \times k5 - k3 \times k4 \times k6 - k3 \times k5 \times k6 - k4 \times k5 \times k6)$	-
λ^0	$k1 \times k2 \times (\mu_{BC} \times \mu_{CD} \times \mu_{DB} \times k6 + \mu_{BD} \times \mu_{DB} \times k4 \times k6 + \mu_{CD} \times \mu_{DC} \times k3 \times k6 - k3 \times k4 \times k5 \times k6)$	

We will use the second aspect of the Rough Hurwitz criterion which is that the real parts of the eigenvalues of \mathcal{K} are all less than 0. Thus, we set this obtained polynomial equal to 0 and determine the eigenvalues

$$\det(\mathcal{K} - \lambda I) = 0 \quad (3.9)$$

We noticed after calculation that the real part of the eigenvalues is always less than 0. We deduce that the system is stable.

4. Numerical simulations

In this section, we verify the results of the analysis with numerical simulations. We performed them using the values in Table 8.

Based on these parameters, we obtained the following equilibrium point: $A_1^* = 14.15, A_2^* = 32.77, B^* = 21.47, C^* = 34.07, D^* = 37.09, E^* = 69.48$. We will run the simulations with the initial values: $A_1(0) = 20, A_2(0) = 14, B(0) = 30, C(0) = 20, D(0) = 10, E(0) = 10$.

Table 8. Parameters values.

Parameter	Values	Sources
θ_1	15	[5]
θ_2	39	[5]
θ_3	43	[5]
θ_4	12	[5]
$\mu_{A_1A_2}$	0.46	Assumed
μ_{A_1D}	0.60	Assumed
μ_{A_2B}	0.56	[34]
μ_{A_2D}	0.63	[34]
μ_{BC}	0.62	[34]
μ_{BD}	0.35	[34]
μ_{BE}	0.30	[34]
μ_{CD}	0.10	[34]
μ_{CE}	0.30	[34]
μ_{DB}	0.36	[34]
μ_{DC}	0.42	[34]
μ_{DE}	0.30	[34]
γ_1	0.4	[34]
γ_2	0.5	[34]
γ_3	0.8	[34]

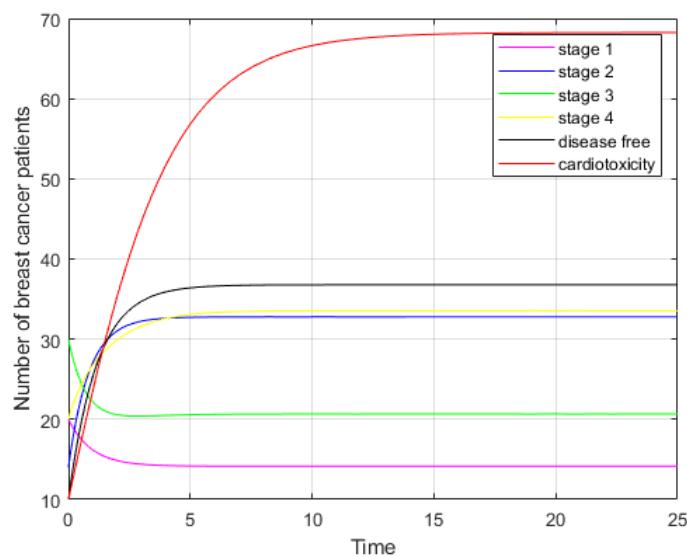
**Figure 4.** Simulation of the model with initial values.

Figure 4 shows that the steady state conditions occur from the 12th time period. The stage 1 sub-population of the initial condition with 20 patients dropped to 14 patients in the steady state conditions.

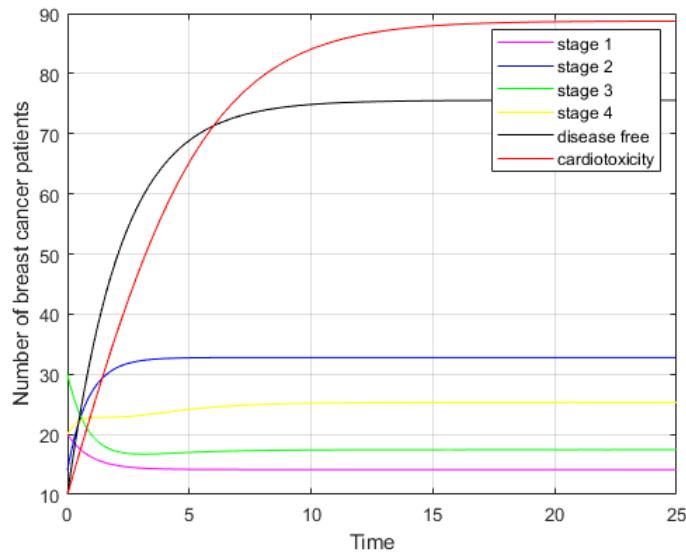


Figure 5. Simulation of the model with $\mu_{DB} = 0.1$ et $\mu_{DC} = 0.1$.

The stage 2 subpopulation of the initial condition with 14 patients rose to 33 patients in the steady state conditions. Similarly, the stage 3 subpopulation of the original condition with 30 patients dropped to 22 in the steady state. The stage 4 subpopulation increased from 20 to 37 and the disease-free subpopulation from 10 to 37 under steady state conditions. The cardiotoxic subpopulation experienced a significant increase in the initial number of 10 patients, rising dramatically to 69 patients under steady state conditions. The results of this simulation are considered as the initial simulation.

For the second simulation, we will try to lower the relapse rates at stages 3 and 4 μ_{DB}, μ_{DC} .

It can be seen from Figure 5 that the cardiotoxic sub-population is increasing as well as the cured sub-population. The results show that the disease-free sub-populations increased significantly by 76 patients. Although it managed to add people who recovered, the cardiotoxic subpopulation also increased significantly. The cardiotoxic sub-population reached 89 patients in stable condition. The subpopulations of patients in stages 1 and 2 are relatively identical to the initial simulation. The subpopulations of stage 3 and 4 patients decreased by 4 and 9 respectively.

Let's try to lower the cardiotoxicity rates μ_{BE}, μ_{CE} and μ_{DE} to 0.1.

Figure 6 shows that the cardiotoxic population decreases significantly to 31 patients. Another positive feature observed is that the disease-free population increased slightly to 49 patients. For the stage 3 sub-population, it has increased slightly from the initial simulation, and the stage 4 sub-population has also increased to 49 patients. For the other sub-populations, it is relatively the same as in the initial simulation.

5. Discussion

The principal target of this study is to use machine learning techniques and mathematical modeling for decision-making regarding breast cancer.

In the results, we noticed that using a SVM model, we obtain the probabilities allowing the attending

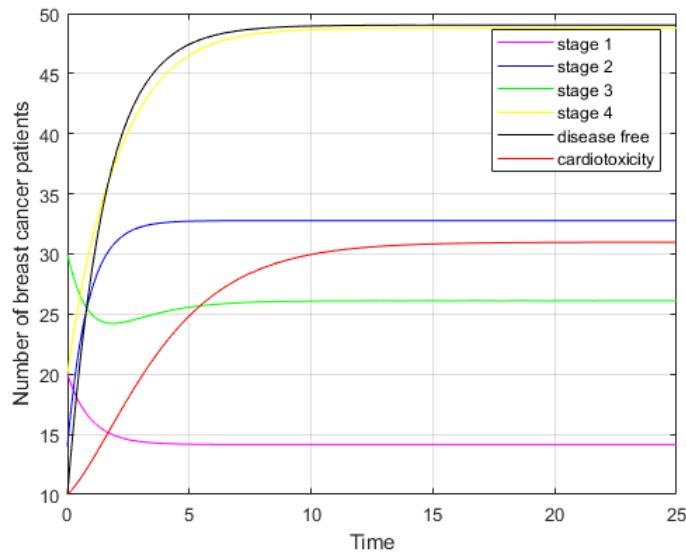


Figure 6. Simulation of the model with $\mu_{BE} = 0.1$, $\mu_{CE} = 0.1$ et $\mu_{DE} = 0.1$.

doctor to know whether the tumor is malignant or benign. Moreover, the proposed mathematical model shows the behavior of the number of patients in each subpopulation at time t . We obtained a stable equilibrium point. Our findings related to SVM and mathematical model are quite similar to the ones in [37, 38]. Table 9 shows the classification prediction performance with SVM in some previous studies. Modeling in medicine is an important tool in the planning and evaluation of interventions [39]. A recent study [40] found that all breast cancer risk factors can be regrouped in two categories. The first includes factors such as age, sex, race and genetic makeup. The second one is composed of diet and life style.

Compartmental models have essential insights regarding the transmission dynamics of communicable diseases [41]. Amani et al. in their work [42] have shown the influence of mathematical models on public health policy. Hsiu-Chuan Wei has studied breast tumor growth using mathematical modeling, which governs MCF-7 cell growth with interaction among different tumor cells [38]. Mkango et al. in their recent publication [43] using mathematical modeling investigated the effects of the combined radiotherapy and chemotherapy as a way to treat breast cancer in Tanzania.

The breast cancer in Benin is often deadly due to late diagnosis, and it is an evidence of the findings in this work because the new mathematical model we proposed can help to minimize the number of cardiotoxic patients and increase the number of recovered after chemotherapy.

In addition, this work outputs an explanation of how the combination of SVM and compartmental models contributes to optimal decision-making regarding breast cancer. This research is the first one that computed SVM and compartmental model to check whether a tumor is malignant or benign, and studied the dynamic of breast cancer patients. The findings are rather important for researchers in the domains of medicine, public health, mathematics and statistics. The limitations of this study are related to the availability of data on the tumor characteristics of breast cancer in Benin. We think that this information will give stronger pieces of evidence.

Table 9. SVM Accuracy comparison with previous studies.

Accuracy	Source
98 %	[44]
97 %	[45]
92 %	[46]
>98 %	Our work

6. Conclusions

In sum, the development of our work was done in two main parts.

First, we performed a classification of malignant and benign tumors to help in the diagnosis of breast cancer patients. We used an SVM model and obtained probabilities allowing the doctor to declare whether the tumor is malignant or benign at a certain percentage.

In a second step, a mathematical model was built with six variables and 19 parameters. The model was built according to medical phenomena concerning the cardiotoxicity of chemotherapy in breast cancer patients. The model consists of four subpopulations of breast cancer patients by stage, a disease-free subpopulation and a cardiotoxic subpopulation. A dynamic analysis is performed to determine the dynamics of the number of patients in each subpopulation at time t . The result of the dynamic analysis is a stable equilibrium point. Numerical simulations were performed to verify the behavior of the solutions around the equilibrium point. Based on the results of the simulations, it can be concluded that if all parameters are assumed to be constant, the state of the population will be stable at a given time, regardless of the initial conditions. This shows that the equilibrium point of the system was found to be stable without conditions. By reducing the relapse rate, an unexpected result is obtained, namely an increase in the toxic subpopulations. Better results are obtained by reducing the rate of cardiotoxicity. Under these conditions, the number of disease-free subpopulations increases and the number of cardiotoxic subpopulations decreases considerably. These simulation results provide a practical solution to minimise the number of cardiotoxic patients and increase the number of patients who recover or have a complete response after chemotherapy. The solution is to reduce the rate of cardiotoxicity in stage 3 and stage 4 subpopulations.

Acknowledgments

The authors are grateful to the lecturers at the International Chair in Mathematical Physics and Applications (ICMPA-UNESCO Chair), who were quite helpful in the improvement of this paper. We thank Mr. Yves Morel Sokadjo, who helped to proof-read this paper. Furthermore, we are thankful to Professor Mahouton Norbert Hounkonnou for numerous helpful discussions concerning this work and suggestions regarding the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Qu'est-ce que le cancer du sein, 2020, available from: <https://www.lillyoncologie.fr/cancer-du-sein/definition>, (accessed on: 21-07-2021).
2. M. Ly, M. Antoine, F. André, P. Callard, J. F. Bernaudin, D. A. Diallo, Le cancer du sein chez la femme de l'Afrique sub-saharienne : état actuel des connaissances, *Bulletin du Cancer*, **98** (2011), 797–806. doi: 10.1684/bdc.2011.1392.
3. Breast cancer, 2021, available from: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>, (accessed on: 10-06-2021).
4. N. Azamjah, Y. Soltan-Zadeh, F. Zayeri, Global trend of breast cancer mortality rate: a 25-year study, *Asian Pac. J. Cancer Prev.*, **20**, (2019), 2015–2020. doi: 10.31557/APJCP.2019.20.7.2015
5. D. G. Gbessi, I. Lawani, C. Tawo-Nounagnon, F. Dossou, Y. Souaïbou, D. Mehinto, et al., Management of Breast Cancer in Visceral Surgery of CNHU-HKM of Cotonou in Benin, *Surg. Sci.*, **07** (2016), 170–176. doi: 10.4236/ss.2016.73022.
6. G. Deloeuvre, Comprendre le cancer, 2018, available from: <https://urlz.fr/guSV>, (accessed on: 10-06-2021).
7. Cancers du sein: les facteurs de risque, 2021, available from: <https://www.fondation-arc.org/cancer/cancer-sein/facteurs-risque-cancer>, (accessed on: 20-06-2021).
8. X. Gruffat, Cancer du sein : Résumé sur le cancer du sein, 2021, available from: <https://www.creapharma.ch/cancer-du-sein.htm>, (accessed on: 10-03-2021).
9. Cancer du sein, 2019, available from: <https://www.cancer-environnement.fr/144-Cancer-du-sein.ce.aspx>, (accessed on: 19-03-2021).
10. A. Felman, What to know about breast cancer, 2019, available from: <https://www.medicalnewstoday.com/articles/37136>, (accessed on: 19-03-2021).
11. Cancer du sein, 2018, available from: <https://ressourcessante.salutbonjour.ca/condition/getcondition/cancer-du-sein>, (accessed on: 10-04-2021).
12. Le dépistage du cancer du sein en 10 questions, 2020, available from: https://www.doctissimo.fr/html/dossiers/cancer_sein/articles/13830-depistage-organise-questions-reponses.htm, accessed on: 05-04-2021.
13. Cancer du sein: Examens, 2018, available from: <https://urlz.fr/gWmD>, (accessed on: 15-04-2021).
14. B. Chevalier, Une IA superchampionne de détection du cancer du sein, 2020, available from: <https://www.adentis.fr/une-ia-superchampionne-de-detection-du-cancer-du-sein/>, (accessed on: 10-03-2021).
15. Y. Benzaki, Introduction à l'algorithme K Nearest Neighbors (K-NN), 2018, available from: <https://mrmint.fr/introduction-k-nearest-neighbors>, (accessed on: 10-03-2021).
16. M. Adankon, M. Cheriet, Support Vector Machine, *Encyclopedia of Biometrics*, Springer US, Boston, MA, (2015).

17. S. Ray, Understanding Support Vector Machine(SVM) algorithm from examples (along with code), 2017, available from: <https://www.analyticsvidhya.com/blog/2017/09/understanding-support-vector-machine-example-code/>, (accessed on: 17-06-2021).

18. kaggle, Breast Cancer Wisconsin (Diagnostic) Data Set, 2016, available from: <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>, (accessed on: 12-03-2021).

19. Société canadienne du cancer, 2015, Cancer du sein : Comprendre le diagnostic, available from: <https://www.cancer.ca>, (accessed on: 16-03-2021).

20. Chimiotherapie, 2017, available from: <https://www.cancer.be/les-cancers/traitements/chimiotherapy>, (accessed on: 05-03-2021).

21. Chemotherapy, 2021, available from: <https://www.breastcancerfoundation.org.nz/breast-cancer/treatment-options/chemotherapy>, (accessed on: 05-03-2021).

22. H. M. Byrne, Dissecting cancer through mathematics: From the cell to the animal model, *Nat. Rev. Cancer*, **10** (2010), 221–230. doi: 10.1038/nrc2808.

23. P. Armitage, R. Doll, The age distribution of cancer and a multi-stage theory of carcinogenesis, *British journal of cancer*, **8** (1954), 1. doi: 10.1038/bjc.1954.1.

24. T. Alarcon, H. Byrne, P. Maini, Towards whole-organ modelling of tumour growth, *Prog. Biophys. Mol. Biol.*, **85** (2004), 451–472. doi: 10.1016/j.pbiomolbio.2004.02.004.

25. D. S. Dixit, D. Kumar, S. Kumar, R. Johri, A mathematical model of chemotherapy for tumor treatment, *Adv. Appl. Math. Biosci.*, **3** (2012), 1–10.

26. H. Schattler, U. Ledzewicz, B. Amini, Dynamical properties of a minimally parameterized mathematical model for metronomic chemotherapy, *Math. Biol.*, **72** (2016), 1255–1280. doi: 10.1007/s00285-015-0907-y.

27. G. Jordao, J. N. Tavares, Mathematical models in cancer therapy, *BioSyst.*, **162** (2016), 12–23. doi: 10.1016/j.biosystems.2017.08.007.

28. G. E. Mahlbacher, K. C. Reihmer, H. B. Frieboes, Mathematical modeling of tumor-immune cell interactions, *J. Theor. Biol.*, **469** (2019), 47–60. doi: 10.1016/j.jtbi.2019.03.002.

29. H. Enderling, M. A. Chaplain, A. R. Anderson, J. S. Vaidya, A mathematical model of breast cancer development, local treatment and recurrence, *J. Theor. Biol.*, **246** (2006), 245–259. doi: 10.1016/j.jtbi.2006.12.010.

30. X. Zhang, Y. Fang, Y. Zhao, W. Zheng, Mathematical modeling the pathway of human breast cancer, *Math. Biosci.*, **253** (2014), 25–29. doi: 10.1016/j.mbs.2014.03.011.

31. Z. Liu, C. Yang, A mathematical model of cancer treatment by radiotherapy followed by chemotherapy, *Math. Comput. Simul.*, **124** (2016), 1–15. doi: 10.1016/j.matcom.2015.12.007.

32. A. Simmons, P. M. Burrage, D. V.Jr. Nicolau, S. R. Lakhani, K. Burrage, Environmental factors in breast cancer invasion: a mathematical modelling review, *Pathology*, **49** (2017), 172–180. doi: 10.1016/j.pathol.2016.11.004.

33. S. I. Oke, M. B. Matadi, S. S. Xulu, Optimal Control Analysis of a Mathematical Model for Breast Cancer, *Math. Comput. Appl.*, **23** (2018), 21. doi: 10.3390/mca23020021.

34. M. I. A. Fathoni, Gunardi, F. A. Kusumo, S. H. Hutajulu, Mathematical model analysis of breast cancer stages with side effects on heart in chemotherapy patients, *AIP Conf. Proc.*, **2192** (2019), 060007. doi: 10.1063/1.5139153.

35. K. P. Bennett, O. L. Mangasarian, Robust Linear Programming Discrimination of Two Linearly Inseparable Sets, *Optim. Methods. Softw.*, (1992), 23–34. doi: 10.1080/10556789208805504.

36. Y. I. A. Rejani, S. T. Selvi, Early Detection of Breast Cancer using SVM Classifier Technique, *Int. J. Comput. Sci. Eng.*, **1** (2009), 127–130.

37. Y. Khourdifi, M. Bahaj, Applying Best Machine Learning Algorithms for Breast Cancer Prediction and Classification, *2018 Int. Conf. Electron. Control Optim. Comput. Sci.*, (2018), 1–5. doi: 10.1109/ICECOCS.2018.8610632.

38. W. H. Chuan, Mathematical modeling of er-positive breast cancer treatment with azd9496 and palbociclib, *AIMS Math.*, **5** (2020), 3446–3455. doi: 10.3934/math.2020223.

39. G. Vanagas, T. Krilavičius, K. L. Man, Mathematical Modeling and Models for Optimal Decision-Making in Health Care, *Comput. Math. Methods. Med.*, **2019** (2019), 2945021. doi: 10.1155/2019/2945021.

40. M. Kamińska, T. Ciszewski, K. Łopacka-Szatan, P. Miotła Paweł, E. Starosławska, Breast cancer risk factors, *Prz. Menopauzalny*, **14** (2015), 196–202. doi: 10.5114/pm.2015.54346.

41. A. L. Parks, B. Walker, W. Pettey, J. Benuzillo, P. Gesteland, J. Grant, et al., Interactive agent based modeling of public health decision-making, *AMIA Annu. Symp. Proc.*, **2009** (2009), 504–508.

42. A. Alahmadi, S. Belet, A. Black, D. Cromer, J. A. Flegg, T. House, et al., Influencing public health policy with data-informed mathematical models of infectious diseases: Recent developments and new challenges, *Epidemics*, **32** (2020), 100393. doi: 10.1016/j.epidem.2020.100393.

43. S. B. Mkango, N. Shaban, E. Mureithi, T. Ngoma, Dynamics of Breast Cancer under Different Rates of Chemoradiotherapy, *Comput. Math. Methods. Med.*, **2019** (2019), 5216346. doi: 10.1155/2019/5216346.

44. M. Divyavani, G. Kalpana, An analysis on SVM & ANN using breast cancer dataset, *Aegaeum J.*, **8** (2021), 369–379.

45. S. M. Basha, D. S. Rajput, N. Ch. S. N. Iyengar, R. D. Caytiles, A novel approach to perform analysis and prediction on breast cancer dataset using R, *Int. J. Grid Distrib. Comput.*, **11** (2018), 41–54. doi: 10.14257/ijgdc.2018.11.2.05.

46. M. Mir, P. Boccia, Case Study : Breast Cancer Classification Using a Support Vector Machine, <https://urlz.fr/gWn9>, (2021), (accessed on: 05-03-2021).



AIMS Press

© 2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)