Feature Selection Optimization for Breast Cancer Diagnosis*

Ana Rita Antunes[1 [0000−0003−4004−9901]], Marina A. Matos[1 [0000−0001−5387−8771]], Lino A. Costa[1 [0000−0003−4772−4404]], Ana Maria A. C. Rocha[1 [0000−0001−8679−2886]], and Ana Cristina Braga[1 [0000−0002−1991−9418]]

ALGORITMI Center, University of Minho, 4710-057 Braga, Portugal
id9069@alunos.uminho.pt, mmatos@algoritmi.uminho.pt, lac@dps.uminho.pt, arocha@dps.uminho.pt, acb@dps.uminho.pt

Abstract. Cancer is one of the leading causes of death in the world, which has increased over the past few years. This disease can be classified as benign or malignant. One of the first and most common cancers that appear in the human body is breast cancer, which, as the name implies, appears in the breast regardless of the person's gender. Machine learning has been widely used to assist in the diagnosis of breast cancer.
In this work, feature selection and multi-objective optimization are applied to the Breast Cancer Wisconsin Diagnostic dataset. It is intended to identify the most relevant characteristics to classify whether the diagnosis is benign or malignant. Two classifiers will be used in the feature selection task, one based on neural networks and the other on support vector machine. The objective functions to be used in the optimization process are to maximize sensitivity and specificity, simultaneously. A comparison was made between the techniques used and there was a better performance by neural networks.

Keywords: feature selection, optimization, neural network, support vector machine, breast cancer

1 Introduction

Cancer is a disease that has increased in recent years, being one of the main causes of death in the world. This is defined as a disease in which some cells in the body grow uncontrollably and can spread to other parts of the body. It starts anywhere in the body, where cells grow and multiply into new cells as needed by our body. When cells age or become damaged, they die and new cells are then formed into place. Sometimes, these cells are formed with damage, and in turn, they grow and spread throughout the body. That said, they can form tumors that can be benign or malignant.

* This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020 and also supported by grant number UI/BD/150936/2021.

Benign tumors have the characteristic of not spreading to different tissues of the body, growing and developing in a specific place. In contrast, malignant tumors can spread to different tissues, even if they started in a specific location [4]. Cancer changes the genetics of DNA through different causes.

According to the World Health Organization in 2019, cancer is the first or second leading cause of death for people under 70 in 61% of countries around the world. The statistical study presented in [27] indicates that 19.3 million new cases of cancer emerged in 2020. One of the oldest and most common cancers that can arise in the human body is breast cancer [5,26].

The breast cancer is considered one of several tumors responsible for the high number of deaths in the world, being more common in females. In 2020, female breast cancer caused 2.3 million new cases and it is predicted that by 2040 there will be an increase to 28.4 million new cases due to demographic and climate changes [27]. It is usually formed by a tumor that can be detected by a breast palpation (felt like a lump) or by an x-ray. Breast cancer can start in different parts of the breast. Then, it can spread to other parts of the body when cancer cells enter the blood or the lymphatic system (network of lymph vessels that connect the lymph nodes that contain the cells). When cancer cells appear, they make the nodes grow, detecting breast cancer.

Machine learning (ML) consists of an evolutionary set of computational algorithms designed to emulate human intelligence. The scientific development has spread the use of ML techniques in different fields, such as pattern recognition, computer vision, spaceship engineering, finance, entertainment and computational biology, and biomedical and medical applications. For example, ML has been a very rewarding method to increase the speed of the entire radiotherapy process, which is widely used in the treatment of cancer patients. Therefore, ML techniques have been used to optimize and automate these processes, and their ability to learn and generalize hidden tasks allows improvements in the safety and effectiveness of radiotherapy practice, leading to better results. Radiotherapy is an example among many [14].

Breast cancer has been extensively studied over the years. To better diagnose and predict the development of this cancer, several techniques and tools based on very powerful and advanced methods, such as machine learning algorithms, are currently used by medicine [6]. The comparison of ML algorithms is highly appreciated in order to obtain the best results, for example, in breast cancer diagnosis and prediction. Agarap A. in [2] made a comparison between six machine learning algorithms using the Breast Cancer Wisconsin Diagnostic database and concluded that all the applied algorithms achieved good results (they all exceeded 90% of test accuracy).

The work by Benbrahim et al. [6] presents a comparative study between eleven machine learning algorithms using the same database, whose main goal was to create two classifiers to define benign and malignant breast nodules. The neural network algorithm was the one that obtained the best result. In the work of Gupta et al. [15], an overview of the evolution of machine learning techniques in cancer diseases was performed, and learning algorithms were compared, such

as linear regression, random forest, multi-layer perceptron and decision trees. They concluded that the multi-layer perceptron algorithm performed better than other techniques.

In this work, feature selection optimization is performed to identify the most relevant features for the classification of the diagnosis of benign or malignant cancer, using the Breast Cancer Wisconsin Diagnostic dataset. The work is carried out by means of two classifiers, one based on neural networks and the other on support vector machine. In addition, a multi-objective optimization algorithm is performed to simultaneously optimize the sensitivity and specificity measures. Finally, a comparison of the performance of the techniques used is made.

This article is structured as follows. Section 2 presents a literature review, where some concepts about machine learning techniques, neural networks, support vector machine and feature selection are briefly explained. The dataset and the methods used in this work are described in Sect. 3 and Sect. 4, respectively. The results obtained and their discussion are presented in Sect. 5 and the main conclusions are reported in Sect. 6.

2 Literature Review

In this section, the concepts of machine learning techniques, neural networks, support vector machine and feature selection are presented.

2.1 Machine Learning Techniques

There is a wide range of machine learning algorithms, which fall into three categories: supervised learning, unsupervised learning and reinforcement learning.

The supervised learning algorithms start with the entry of labeled data to train the algorithm, that is, for each dataset provided, an output or solution is predicted based on the relationships and dependencies learned from the dataset [14]. This type of ML algorithm is applied to regression and classification problems.

The most popular supervised learning algorithms include: linear regression, logistic regression, decision tree, support vector machine, k-nearest neighbor and neural networks. Linear regression establishes a relationship between independent and dependent variables, adjusting the best line and using estimated real values. To predict the probability of occurring an event by a discrete estimated value is called a Logistic regression. When dividing data into two or more homogeneous sets, based on independent variables is defined as Decision tree. The Support vector machine algorithm aims to find a hyperplane in an n-dimensional space (n-the number of resources) that distinctly classifies the data points. The K-nearest neighbor is a simple algorithm that stores all available cases and classifies them in new cases considering the closest distance to their k neighbors. The last three algorithms mentioned earlier are used in classification problems [8]. A Neural network is said to learn supervised, if the desired output is already

known. Neural networks are trained by processing examples, each of which containing a known input and output, forming probability-weighted associations between the two, which are stored within the data structure of the net itself.

Unsupervised learning algorithms begin with a model that receives unlabeled data and discovers hidden patterns or data groupings without providing any guidance or instructions [14]. That said, the grouping of data is carried out in which the comparison is made, between them, by the model to guess the output. The unsupervised learning algorithms are applied in clustering and association problems. Some of the most common algorithms used in unsupervised learning include: Clustering, Anomaly detection, Neural networks and Approaches for learning latent variable models. Clustering is a technique for grouping datasets, so that the subset of data in the same group (called cluster) has similarities to each other than to other groups. A widely used clustering method is the K-means, which aims to partition a dataset (composed of data points) among k groups (clusters), allocating every data point to the nearest cluster. Anomaly detection, also known as outlier detection is the process of identifying extreme points or observations that are significantly deviating from the remaining data. Concerning the Neural networks for unsupervised learning, there are no target outputs. The units (weight values) of the neural network are "arranged" inside a certain range, depending on the given input values. The goal is to group similar units close together in certain areas of the value range.

Reinforcement learning algorithms are based on interaction with their environment, that is, an agent always has the initial and final states, where there can be several different paths to reach the final state [14], interacting with the environment by performing actions and learning with mistakes or rewards and does not use predefined data.

2.2 Neural Networks

Neural networks (NN) have been increasingly used to solve artificial intelligence problems. The diagnosis of breast cancer is an example where NN have been widely used [11]. Neural networks are machine learning techniques that simulate the learning mechanism in biological organisms and are networks of interconnected artificial neurons [3].

The main advantage of neural networks is that they are data-driven and do not require restrictive assumptions about the shape of the basic model [13]. In addition to this advantage, NN allow detecting complex nonlinear relationships between independent and dependent variables and to identify all possible interactions between predictor variables, having the ability to store information on the network (memory capacity). In addition, NN can work with incomplete knowledge and in parallel processing. However, NN also have disadvantages, such as hardware dependency, the lack of determination of the appropriate network structure, more computational resources requirements and limited ability to identify possible casual relationships [28].

2.3 Support Vector Machine

Support Vector Machine (SVM) is characterized by its simplicity and flexibility to deal with a wide variety of classification problems. The objective of the SVM is to find a hyperplane that classifies the data points [21].

In SVM, a data point is viewed as a n-dimensional vector (n features), and the goal is to separate such points with a (n-1)-dimensional hyperplane. This is a linear classifier. There are many hyperplanes that might classify the data, turjillo but the goal is to find the hyperplane that represents the largest separation between the two classes, maximizing the distance form the hyperplane to the nearest data point on each side. SVM can also be used for nonlinear separate functions through a kernel function.

SVM has been widely used in the research of brain diseases and in the diagnosis of cancerous diseases, such as in the psychiatric context in order to diagnose neurological diseases (Alzheimer, schizophrenia and depression). Neuroimaging analysis using SVM allows to study a large amount of cancer data, leading to the discovery of new drugs and a better understanding of genes originated in cancer [16]. In addition, SVM has also been used to breast cancer diagnosis and classification [22].

2.4 Feature Selection

Feature selection is the process of obtaining a subset of the original feature set, according to a certain criterion, which selects the most relevant features for classification. Feature selection plays a very important role in data compression and allows pre-processing of machine learning algorithms, in order to improve learning accuracy, reduce learning time and simplify results of learning [9]. Feature selection can be classified into three methods: filter, wrapper and embedded. The filter method aims to the selection of features based on a performance measure (e.g., correlation, Chi-square and Fisher score) in order to find the best subset of features. The wrapper method exploits the feature space to score subsets of resources according to their predictive power, optimizing the subsequent induction algorithm that uses the respective subset for classification [18]. This method requires more computation time than the filter method, but it achieves more accurate results. The embedded method selects the features in the training process, without splitting data into training and testing, aiming to reduce the computation time [9].

Feature selection has been a research topic used in many fields, such as image recognition, image recovery, text mining, intrusion detection, bioinformatics data analysis and fault diagnosis. Several studies with the breast cancer diagnosis have been carried out using SVM with feature selection [4,10,23].

3 Dataset

In this work, the Breast Cancer Wisconsin Diagnostic database[4] was selected from the UCI Machine Learning Repository [12]. This data considers information about the digitized image of fine-needle aspirates (FNA) for malignant and benign diagnoses [29]. There are thirty features divided into three indicators, such as the mean value, standard error and "worst" or largest value, which is the mean of the three largest values. Thereby, each cell nucleus has ten measures: radius (R), that is the mean of distances from center to points, texture (T), that is the standard deviation of gray-scale values, perimeter (P), area (A), smoothness (ST), the local variation in radius lengths. Compactness (CPT), expressed by $\frac{Perimeter^2}{area-1}$, concavity (CC), that represents the severity of concave portions of the contour, concave points (CP), that are the number of concave portions of contour, symmetry (SY) and fractal dimension (FD) that is expressed by "coastline approximation" − 1. For simplicity, these features will be referred as Feature_Name$_{mean}$, Feature_Name$_{se}$ and Feature_Name$_{wst}$, where *mean*, *se* and *wst* are the mean, standard error and worst values, respectively. There are information about 569 diagnoses, where 357 are benign and 212 are malignant as represented in the bar plot of Fig. 1.

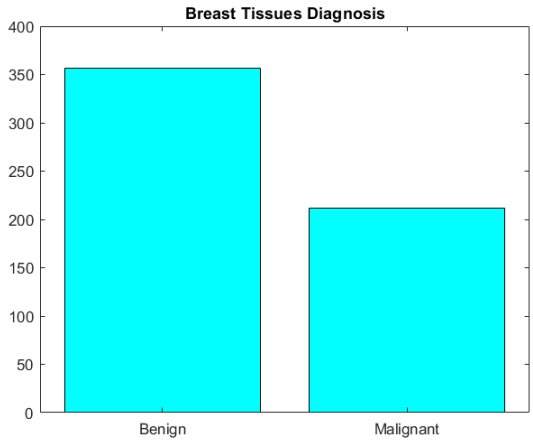

Fig. 1: Breast tissues diagnosis distribution.

In order to check if there are big differences between benign and malignant values, the minimum and maximum value for each feature per diagnosis were computed (Table 1). The features P$_{mean}$, P$_{se}$, P$_{wst}$, A$_{mean}$, A$_{se}$ and A$_{wst}$ have

[4] Available in: `https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)`, Accessed: 2021-05-07

the highest range difference values. These features may be one of the candidates for predicting cancer breast.

Table 1: Range for each feature per diagnosis.

		Benign			Malignant		
Feature	**Type**	**Mean**	**SE**	**Worst**	**Mean**	**SE**	**Worst**
R	Min	6.981	0.112	7.930	10.950	0.194	12.840
	Max	17.850	0.881	19.820	28.110	2.873	36.040
T	Min	9.710	0.360	12.020	10.380	0.362	16.670
	Max	33.810	4.885	41.780	39.280	3.568	49.540
P	Min	43.790	0.757	50.410	71.900	1.334	85.100
	Max	114.600	5.118	127.100	188.500	21.980	251.200
A	Min	143.500	6.802	185.200	361.600	13.990	508.100
	Max	992.100	77.110	1210.000	2501.000	542.200	4254.000
ST	Min	0.053	0.002	0.071	0.074	0.003	0.088
	Max	0.163	0.022	0.201	0.145	0.031	0.223
CP	Min	0.019	0.002	0.027	0.046	0.008	0.051
	Max	0.224	0.106	0.585	0.345	0.135	1.058
CC	Min	0.000	0.000	0.000	0.024	0.011	0.024
	Max	0.411	0.396	1.252	0.427	0.144	1.170
CPT	Min	0.000	0.000	0.000	0.020	0.005	0.029
	Max	0.085	0.053	0.175	0.201	0.041	0.291
SY	Min	0.106	0.010	0.157	0.131	0.008	0.157
	Max	0.274	0.061	0.423	0.304	0.079	0.664
FD	Min	0.052	0.001	0.055	0.050	0.001	0.055
	Max	0.096	0.030	0.149	0.097	0.013	0.208

4 Methods

In this section, the performance measures used to evaluate the models quality are introduced. Thereafter, the feature selection approach describes the required optimization steps followed by the implementation details.

4.1 Performance Measures

The performance of the machine learning algorithms can be assessed using different performance measures. Therefore, correct and wrong classifications must me analyzed. Depending on the defined hypotheses, the outcome can result on

true positive, true negative, false negative, or false positive classification. These values are used to define the confusion matrix [25]. In this study, the defined hypotheses were H_0: the diagnosis is benign and H_1: the diagnosis is malignant. The confusion matrix is presented in Table 2. When a case is benign and is classified as benign, this corresponds to a correct classification and it is a true positive. In contrast, a false positive occurs when a benign case is classified as malignant. If a malignant case is correctly classified as malignant, it is a true negative. Although, if the true classification is malignant and it is classified as benign, it is a false negative.

Table 2: Confusion Matrix.

	Hypotheses	
Decision	**Benign**	**Malignant**
Classified as Benign	TP	FN
Classified as Malignant	FP	TN

Thereby, the sensitivity and specificity, also known as true positive rate and true negative rate, respectively, are two measures to evaluate the effectiveness of the model to classify positive and negative labels, i.e., the correct classification for benign and malignant diagnosis. So, the sensitivity and specificity can be expressed by Eq. (1) and (2), respectively [25].

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{1}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2}$$

Other commonly used measure is the area under the ROC curve, known as AUC that is an aggregate measure of the performance of the classifier. This metric is applied to assess the ability of predictive capacity of classifiers and can be expressed by Eq. (3) [17,25].

$$\text{AUC} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{3}$$

If AUC is near to one, then it is a reliable diagnostic [7]. On the other hand, if AUC is close to 0.5, the classifier performs like a random choice and it is non-informative classifier.

4.2 Feature Selection Approach

The feature selection optimization steps to diagnose the type of breast cancer are depicted in Fig. 2. The dataset includes the outcome of each case. Therefore, supervised machine learning classification algorithms are used.

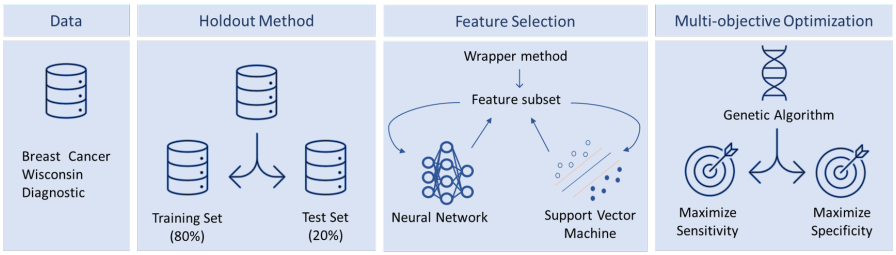

Fig. 2: Steps to achieve the proposed aims.

Breast Cancer Wisconsin Diagnostic dataset has thirty features. Firstly, data were split into training and test sets using the holdout method. This cross-validation method prevents the classification models from becoming overfitted with respect to the dataset. Eighty percent of the dataset is randomly selected as training set and the remaining is used as test set.

In order to select the features that provide best classification performance, the wrapper method was used. Two classification algorithms are considered and compared. A feedforward neural network and a linear SVM were used as classifiers to predict the diagnosis with the test set.

A multi-objective approach was performed to maximize sensitivity and specificity, simultaneously. This optimization approach is based on a multi-objective genetic algorithm used to identify the best subset of features capable of predicting whether the diagnosis is benign or malignant. Different trade-offs between sensitivity and specificity are sought by the optimization algorithm. In this algorithm, a population of potential solutions evolves along several generations. Each solution indicates different combinations of features to be selected. The fitness of each solution is assessed by the performance of the classification system. The best solutions have a higher probability of being selected to generate offspring by crossover and mutation. The outcome of this optimization process is a set of non-dominated solutions that represent different trade-offs between sensitivity and specificity. These solutions are incomparable and their images define a Pareto front.

4.3 Implementation details

The feature selection approach was implemented in MATLAB® [19] using Global Optimization and Statistics and Machine Learning toolboxes.

Firstly, for the classification of the Breast Cancer using feedforward NN the `feedforwardnet` function was used. After several experiments a network considering three hidden layers with eight, four, and two neurons was performed. The number of epochs was set to 750. In the feedforward NN, the training set ratio was specified to 1 and the validation and testing set ratio were defined as 0 in order to maintain the same training set during the optimization process.

For the classification of the Breast Cancer through the linear SVM, the `fitcsvm` function was used. Note that the default parameters were applied, thus the linear mode was employed.

The feature selection optimization was performed using the `gamultiobj` function. This function implements a multi-objective genetic algorithm that is a variant of the elitist NSGA-II [19]. In the multi-objective optimization process, the standard options were used, except the adaptable feasible mutation and 100 individuals as population size.

5 Computational Experiments

In this section, feature selection optimization results are presented, where the dataset used is to diagnose the type of breast cancer. A multi-objective optimization algorithm is applied to obtain solutions with good classification performance. The results achieved are described and discussed.

5.1 Experimental Results

The solutions obtained for feedforward NN and linear SVM are represented graphically in the Pareto front (Fig. 3). For linear SVM and feedforward NN, five (blue dots) and two (red dots) non-dominated solutions were found, respectively. In spite of the larger number of solutions found for linear SVM, all of them are dominated by the two obtained by NN. Thus, the neural network performed better than SVM in both sensitivity and specificity.

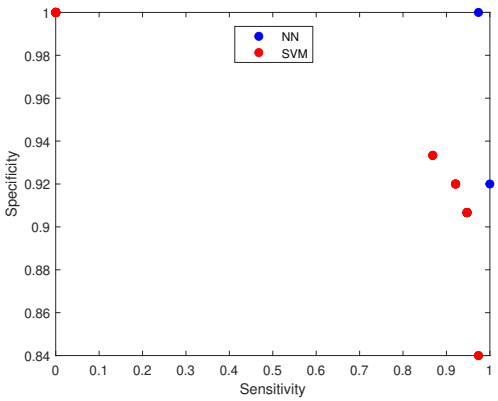

Fig. 3: Pareto front using feedforward NN and linear SVM: Sensitivity against Specificity.

The identification of the features that correspond to each non-dominated solution is relevant since can help to identify the most important features and

the existence of associations between them. Moreover, AUC allows to identify the best performing solution aggregating the measures of sensitivity and specificity. Note that optimizing AUC does not allow to obtain different trade-offs between this two measures.

Table 3 presents the features selected according to the classifier, the sensitivity, specificity, and AUC values. One interesting remark is that the solutions with the smallest number of features are from the neural network classifier, that retained twelve and fifteen features. Conversely, using SVM as classifier, the number of features are twenty, twenty-one, twenty-two and twenty-four. When comparing the solutions found, P_{mean}, P_{wst}, A_{wst}, CV_{mean}, and SY_{mean} are the features in common. On the other hand, CC_{mean}, FD_{mean}, FD_{se} and CP_{se} features are not used in any solution. Hence, these features are not important to classify if the diagnosis is benign or malignant. Therefore, the model with the best performance is the feedforward NN with fifteen features (NN_2 model), since it has the highest AUC value.

Figure 4 shows the ROC curve for all the solutions. Again, it can be seen that NN_1 and NN_2 models are better than all linear SVM models. NN_1 is better than NN_2 in terms of sensitivity and the opposite in terms of specificity. Thus, if it is preferable to maximize the true positive rate, NN_1 model should be used. If the goal is to minimize false negative rate, NN_2 is preferable. The SVM_1 model is useless since performs like a random classification. All linear SVM models are worse the feedforward NN models.

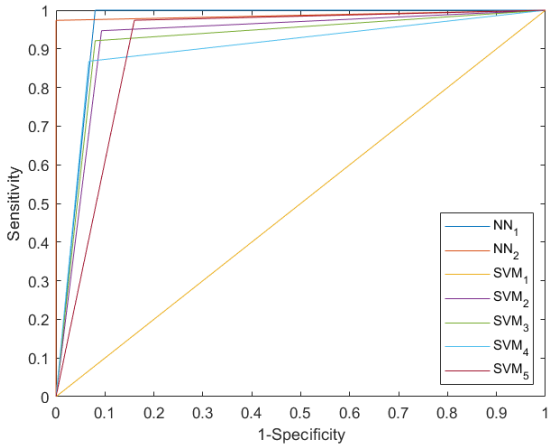

Fig. 4: ROC Curve.

Table 3: Optimization results for feedforward NN and linear SVM methods.

Model	**Feature**	**Sensitivity**	**Specificity**	AUC
NN_1	P_{mean}, P_{wst}, CPT_{mean}, CPT_{wst} SY_{mean}, SY_{se}, T_{se}, T_{wst}, CP_{se}, R_{wst}, A_{wst}, CC_{wst}	100.00	92.00	96.00
NN_2	R_{mean}, R_{wst}, T_{mean}, T_{wst}, P_{mean}, P_{wst}, CPT_{mean}, CPT_{wst}, SY_{mean}, SY_{se}, CP_{se}, A_{wst}, ST_{wst}, CC_{wst}, FD_{wst}	97.37	100	98.70
SVM_1	R_{mean}, R_{wst}, T_{mean}, T_{se}, P_{mean}, P_{se}, P_{wst}, A_{se}, A_{wst}, ST_{mean}, ST_{se}, ST_{wst}, CC_{mean}, CC_{se}, CPT_{mean}, SY_{mean}, SY_{wst}, FD_{mean}, FD_{se}, FD_{wst}	0	100.00	50.00
SVM_2	R_{mean}, R_{se}, R_{wst}, T_{mean}, T_{se}, T_{wst}, P_{mean}, P_{se}, P_{wst}, A_{mean}, A_{se}, A_{wst}, ST_{mean}, ST_{se}, CP_{mean}, CP_{se}, CP_{wst}, CC_{se}, SY_{mean}, SY_{wst}, FD_{se}, FD_{wst}	94.70	90.70	92.70
SVM_3	R_{mean}, R_{se}, T_{mean}, T_{se}, T_{wst}, P_{mean}, P_{se}, P_{wst}, A_{mean}, A_{se}, A_{wst}, ST_{mean}, ST_{se}, CP_{mean}, CP_{se}, CP_{wst}, CC_{mean}, CC_{se}, CPT_{mean}, SY_{mean}, FD_{wst}	92.10	92.00	92.70
SVM_4	R_{mean}, R_{se}, R_{wst}, T_{mean}, T_{se}, T_{wst}, P_{mean}, P_{se}, P_{wst}, A_{se}, A_{wst}, ST_{mean}, ST_{se}, CC_{se}, CC_{wst}, CPT_{mean}, SY_{mean}, SY_{se}, SY_{wst}, FD_{mean}, FD_{se}	86.80	93.30	90.00
SVM_5	R_{mean}, R_{se}, T_{mean}, T_{se}, T_{wst}, P_{mean}, P_{se}, P_{wst}, A_{mean}, A_{se}, A_{wst}, ST_{mean}, ST_{wst}, CP_{mean}, CP_{se}, CC_{se}, CPT_{mean},, CPT_{se}, SY_{mean}, SY_{se}, SY_{wst}, FD_{mean}, FD_{se}, FD_{wst}	97.40	84.00	90.70

5.2 Discussion of Results

This research intended to identify the relevant features to predict if a breast cancer is benign or malignant. A multi-objective optimization was performed to select the features that allow to maximize sensitivity and specificity. The best solution for this problem was obtained using a feedforward NN (NN_2), where sensitivity is 97.37%, specificity is 100%, and AUC is 98.70%. This means that the model was capable to predict all malignant diagnosis correctly. These results are very promising, since they are better than other studies summarized in Table 4 [1,20,24].

In Table 4, the different models proposed in other related works and the best models obtained using the proposed multi-objective optimization approach (MO_{ap}) are presented. For each model, the sensitivity, specificity and AUC are given. Taking into account the sensitivity and specificity values of each model, only the SVM (Linear) proposed by [20], NN_1 and NN_2 are non-dominated solutions, i.e., optimal trade-offs between sensitivity and specificity. In spite of SVM

(Linear), SVM (Radial Basis Function) and SVM (Polynomial) having AUC values superior to NN_1, only the SVM (Linear) model is an optimal trade-off since the others are dominated by NN_2. So, only the SVM (Linear) model has a performance very close to the neural networks classification models (NN_1 and NN_2) obtained using the MO_{ap} proposed method. NN_2 model has the highest AUC (98.70), followed by the SVM (Linear) model (98.50). In the SVM (Linear) model, seventeen features are retained [20], while NN_1 and NN_2 require twelve and fifteen features, respectively. Therefore, the NN_2 model has a smaller number of features and a higher AUC value when compared with the SVM (Linear) model. The NN_1 model has the smallest number of features. Thus, NN_2 emerges as the model with best overall performance. The most relevant features to diagnose benign or malignant breast cancer are : R_{mean}, R_{wst}, T_{mean}, T_{wst}, P_{mean}, P_{wst}, CPT_{mean}, CPT_{wst}, SY_{mean}, SY_{se}, CP_{se}, A_{wst}, ST_{wst}, CC_{wst} and FD_{wst}.

Table 4: Results for related works.

Ref	**Model**	**Sensitivity** %	**Specificity** %	**AUC** %
[24]	LR	90.06	99.89	95.25
	KNN	90.09	94.70	92.39
	SVM	88.20	84.90	86.55
[1]	PSO	33.30	92.90	63.10
	GA	31.00	92.80	61.90
	ANN	33.00	96.30	64.65
[20]	SVM (Linear)	98.00	99.00	98.50
	SVM (Radial Basis Function)	96.00	99.00	97.50
	SVM (Polynomial)	97.00	97.00	97.00
	SVM (Sigmoid)	60.00	54.00	57.00
MO_{ap}	NN_1	100.00	92.00	96.00
	NN_2	97.37	100.00	98.70
	SVM_2	94.70	90.70	92.70
	SVM_3	92.10	92.00	92.70

LR: Logistic Regression, KNN: k-Nearest Neighbor, PSO: Particle Swarm Optimization
ANN: Artificial Neural Network, MO_{ap}: our MO approach

6 Conclusions

Cancer is a disease that has been on the rise over the past few years. Breast cancer is one of the most fatal causes and also the most successful cure. This has been extensively studied over the years, in order to improve the diagnosis and predict its development. Thus, several techniques and tools based on very powerful and advanced methods, such as machine learning algorithms, have been studied.

In this work, a feature selection was optimized in order to identify the most relevant resources for the classification of the diagnosis of benign or malignant cancer, using the Breast Cancer Wisconsin Diagnostic dataset. Two classifiers were applied, a feedforward neural network and a linear support vector machine, and a multi-objective optimization algorithm to simultaneously optimize the sensitivity and specificity measures. Finally, the performance of the techniques used was compared.

In this study, two and five non-dominated solutions emerged from feedforward neural networks and linear support vector machine, respectively. The fewest features were obtained using the feedforward neural networks. Moreover, the results for this classifier performed better, since the AUC values are the highest. Thereby, the best solution (NN_2) has sensitivity and specificity values equal to 97.37% and 100%, respectively. Consequently, the AUC value achieved was 98.70% and the number of selected features was fifteen, that is quite promising when comparing with other related works.

References

1. Aalaei, S., Shahraki, H., Rowhanimanesh, A., Eslami, S.: Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets. Iranian journal of basic medical sciences **19**(5), 476 (2016)
2. Agarap, A.F.M.: On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In: Proceedings of the 2nd international conference on machine learning and soft computing. pp. 5–9 (2018)
3. Aggarwal, C.C.: An introduction to neural networks. In: Neural Networks and Deep Learning, pp. 1–52. Springer (2018)
4. Akay, M.F.: Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications **36**(2), 3240–3247 (2009)
5. Akram, M., Iqbal, M., Daniyal, M., Khan, A.U.: Awareness and current knowledge of breast cancer. Biological research **50**(1), 1–23 (2017)
6. Benbrahim, H., Hachimi, H., Amine, A.: Comparative study of machine learning algorithms using the breast cancer dataset. In: International Conference on Advanced Intelligent Systems for Sustainable Development. pp. 83–91. Springer (2019)
7. Bhardwaj, A., Tiwari, A.: Breast cancer diagnosis using genetically optimized neural network model. Expert Systems with Applications **42**(10), 4611–4620 (2015)
8. Bonaccorso, G.: Machine Learning Algorithms: Popular algorithms for data science and machine learning. Packt Publishing Ltd (2018)
9. Cai, J., Luo, J., Wang, S., Yang, S.: Feature selection in machine learning: A new perspective. Neurocomputing **300**, 70–79 (2018)
10. Chen, H.L., Yang, B., Liu, J., Liu, D.Y.: A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. Expert systems with applications **38**(7), 9014–9022 (2011)
11. Chougrad, H., Zouaki, H., Alheyane, O.: Deep convolutional neural networks for breast cancer screening. Computer methods and programs in biomedicine **157**, 19–30 (2018)
12. Dua, D., Graff, C.: UCI machine learning repository (2019), `http://archive.ics.uci.edu/ml`

13. Dumitru, C., Maria, V.: Advantages and disadvantages of using neural networks for predictions. Ovidius University Annals, Series Economic Sciences **13**(1) (2013)
14. El Naqa, I., Murphy, M.J.: What is machine learning? In: machine learning in radiation oncology, pp. 3–11. Springer (2015)
15. Gupta, M., Gupta, B.: A comparative study of breast cancer diagnosis using supervised machine learning techniques. In: 2018 second international conference on computing methodologies and communication (ICCMC). pp. 997–1002. IEEE (2018)
16. Huang, S., Cai, N., Pacheco, P.P., Narrandes, S., Wang, Y., Xu, W.: Applications of support vector machine (svm) learning in cancer genomics. Cancer Genomics-Proteomics **15**(1), 41–51 (2018)
17. Janssens, A.C.J., Martens, F.K.: Reflection on modern methods: revisiting the area under the roc curve. International journal of epidemiology **49**(4), 1397–1403 (2020)
18. Maldonado, S., Weber, R.: A wrapper method for feature selection using support vector machines. Information Sciences **179**(13), 2208–2217 (2009)
19. MATLAB: version 9.9.0.1495850 (R2020b). The MathWorks Inc., Natick, Massachusetts (2020)
20. Memon, M.H., Li, J.P., Haq, A.U., Memon, M.H., Zhou, W.: Breast cancer detection in the iot health environment using modified recursive feature selection. wireless communications and mobile computing **2019** (2019)
21. Pisner, D.A., Schnyer, D.M.: Support vector machine. In: Machine Learning, pp. 101–121. Elsevier (2020)
22. Rejani, Y., Selvi, S.T.: Early detection of breast cancer using svm classifier technique. arXiv preprint arXiv:0912.2314 (2009)
23. Salama, G.I., Abdelhalim, M., Zeid, M.A.e.: Breast cancer diagnosis on three different datasets using multi-classifiers. Breast Cancer (WDBC) **32**(569), 2 (2012)
24. Sharma, A., Kulshrestha, S., Daniel, S.: Machine learning approaches for breast cancer diagnosis and prognosis. In: 2017 International conference on soft computing and its engineering applications (icSoftComp). pp. 1–5. IEEE (2017)
25. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Information processing & management **45**(4), 427–437 (2009)
26. Sun, Y.S., Zhao, Z., Yang, Z.N., Xu, F., Lu, H.J., Zhu, Z.Y., Shi, W., Jiang, J., Yao, P.P., Zhu, H.P.: Risk factors and preventions of breast cancer. International journal of biological sciences **13**(11), 1387 (2017)
27. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **71**(3), 209–249 (2021)
28. Tu, J.V.: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of clinical epidemiology **49**(11), 1225–1231 (1996)
29. Wolberg, W.H., Mangasarian, O.L.: Multisurface method of pattern separation for medical diagnosis applied to breast cytology. Proceedings of the national academy of sciences **87**(23), 9193–9196 (1990)