Effect sizes as a statistical feature-selector-based learning to detect breast cancer

Nicolás Masino
Data Science and AI Laboratory, Data Science Department
Catholic University of Argentina (UCA), Argentina
nicolasmasino@uca.edu.ar

Antonio Quintero-Rincón
Data Science and AI Laboratory, Data Science Department
Computer Science Department
Catholic University of Argentina (UCA), Argentina
antonioquintero@uca.edu.ar

***Abstract*—Breast cancer detection is still an open research field, despite a tremendous effort devoted to work in this area. Effect size is a statistical concept that measures the strength of the relationship between two variables on a numeric scale. Feature selection is widely used to reduce the dimensionality of data by selecting only a subset of predictor variables to improve a learning model. In this work, an algorithm and experimental results demonstrate the feasibility of developing a statistical feature-selector-based learning tool capable of reducing the data dimensionality using parametric effect size measures from features extracted from cell nuclei images. The SVM classifier with a linear kernel as a learning tool achieved an accuracy of over 90%. These excellent results suggest that the effect size is within the standards of the feature-selector methods.**

***Index Terms*—Effect Size, Cohen's d, Standardized Mean Difference, Feature selection, Breast Cancer**

I. Introduction

Breast cancer is a disease frequently diagnosed in women in which abnormal breast cells grow uncontrollably until tumors form. According to the World Health Organization [1] in 2022, breast cancer was diagnosed to over 2.3 million women and 670,000 deaths globally. Cancer treatment is patient-specific through radiation therapy, medications, and surgery; and depends on the type of cancer and its spread in the body.

Effect size (EF) is an association-magnitude measure between two populations under research. EF has become more popular in recent years in meta-analyses of psychological, educational, and behavioral treatments [2]. EF can be estimated through parametric or non-parametric kernels. EF expresses across a numerical decision rule the practical significance or strength of a research outcome. This numerical scale yields an index that lets us know how meaningful the relationship between variables or the difference between groups is. It is common knowledge that statistical significance denoted by p-values affects an outcome, while effect sizes represent practical significance. A large effect size means that the research outcome has practical importance, while a small effect size indicates limited practical applications. The family of EF indices can be separated into two types of measures. The standardized differences between two groups and the correlation measure of effect size [3]. This work is focused on estimating the parametric standardized difference between two independent populations vs a dichotomous dependent variable.

979-8-3503-6593-1/24/$31.00 ©2024 IEEE

The well-known free Diagnostic Wisconsin Breast Cancer database hosted in the Machine Learning Repository at the University of California, Irvine [4] was considered for experimentation purposes. Its features describe the characteristics of the cell nuclei present in a digitized image of a fine needle aspirate (FNA) of a breast mass. For details, consult the Section II-A introduced below.

This work aims to propose a statistical feature-selector-based learning to detect breast cancer through effect size. The underlying idea of feature-selector-based learning lies in finding a good interaction between the features and selecting only the most relevant. In this context, a correct combination of features may provide higher predictive power and increase the precision of a learning model with minimal risk. Therefore, performing feature-selector-based learning allows for identifying an optimal feature combination and a dimensional data reduction to improve the predictive learning capacity. This process has several advantages. It can significantly reduce model training time and prediction speed during production deployment. Additionally, it makes the model less complex and easier to explain.

Feature-selection methods can be grouped into three approaches: Filter method, Wrapper method, and Embedded method [5], [6]. Filter methods are based on a relevance index based on correlation coefficients or test statistics. The methods used are Correlation-based filters, Relevance indices based on Distances between distributions, or based on Information Theory, Decision trees for ranking, or Reliability and Bias of relevance indices. Wrapper's methods utilize the performance of a learning machine trained using a given feature subset. The methods used are based on forward selection and backward elimination. Both methods use search strategies to explore the space of all possible feature combinations. Embedded methods incorporate feature subset generation and evaluation in the training algorithm [6], [7]. The methods used are based on forward selection and backward elimination with an optimization of scaling factors. Embedded and wrapper methods perform well for a given classifier. However, these approaches represent more computational complexity, above all the data dimension is high [5], [8], [9]. Note that, the proposed EF as statistical feature-selection-based learning fits the filter method. Many

well-known feature selection techniques are within the filter, wrapper, or embedded methods. That is why choosing a specific technique is not an easy task. It is essential to know the data dimension, the size of the data, and the computational cost that is required and accepted. In medical diagnosis scenarios, the most common techniques are Correlation-Based Feature Selection, Consistency-Based Filter, INTERACT, Information Gain, Relief, Recursive Feature Elimination, and Lasso Regularization [9]. Specifically in feature selection to identify relevant cell nuclei features, a variety of studies can be found to diagnose breast cancer, for example, in [10] three different techniques were proposed: minimum redundancy maximum relevance, Wilcoxon's rank-sum test, and Random Forest, in [11] by employing a Sequential Forward Selection (SBS), or across an embedded approach using SVM based on the F-score to predict it [12]. Another study from the same field compared Random Forest (RF) with other select feature procedures, like SVM-RF, RRF, SBS, and VarSelRF. Binary Random Forest Feature Selection method (BRFFS) was proposed to reduce the nuclear features and classify leukocytes [8]. For a comprehensive treatment of feature selection see [13], [14].

The hypothesis to be assessed in this work is under the assumption that the behavior of a variable is different for each class. Specifically in breast cancer scenarios, the features should have different values for their classes, called malignant cancerous and benign non-cancerous samples. To assess this hypothesis, the parametric measure of effect size was proposed to quantify the difference in the distributions. Note that the techniques used in this study are all well-known in the scientific community. Still, to our knowledge, the effect size measurement has never been used as a statistical feature-selector-based learning scheme. This is the main contribution of this work.

The remainder of this paper is organized as follows. Section II presents the proposed method and provides an introduction of the dataset used in Section II-A, the effect size in Section II-C, the feature-selector-based decision rule in Section II-D, the feature-selector-based learning with a support vector machine in Section II-E, and the confidence intervals in section II-F. In Section III, the results are discussed and analyzed. Finally in Section IV conclusion, advantages, limitations, and future works are given.

II. Methodology

The pipeline proposal is as follows. Initial data with the features of cell nuclei present from digitized images obtained by FNA of a breast mass were given. Then, the parametric effect size measurement is computed for each sample to estimate a dimensional feature reduction from the original feature data. Feature-selector-based learning is calculated according to the values of a numerical decision rule. If the observed value of the effect size is greater than a numerical value, then it is decided if the feature is significant or not. The numerical value is chosen for the standardized effect size scale value, or estimating the mean of the rank of their values. Finally, with the new data arising from feature reduction, a classification-based learning scheme for detecting breast cancer is computed in such a way that can be utilized in prevalent clinical settings. The dataset, software, parametric effect size methods, feature selector-based decision rule, and feature selector-based learning are then presented.

A. Dataset

For experimentation purposes, the Diagnostic Wisconsin Breast Cancer Database from the University of California, Irvin, was considered [4], [15]. This dataset is based on image-processing techniques that use custom active contour models, known as snakes. Images close to the boundaries of a set of cell nuclei are selected using a fine needle aspiration slide. Thus, size, shape, and texture can be acquired precisely because the snakes are deformed to the exact shape of the nuclei. For each nucleus, 10 features called: radius, perimeter, area, texture, compactness, smoothness, concavity, concave points, symmetry, and fractal dimension, were estimated. Also, the mean value, the largest or worst value, and the standard error of each feature are found over the range of isolated cells. Different combinations of features were then tested to find those that best discriminated between benign and malignant samples, yielding 30 features in total related to mean texture, worst area, and worst smoothness, see Table II. The dataset contains 569 binary observations that were split into two groups. 212 malignant cancerous samples, and 357 benign non-cancerous samples. See [15] for a detailed description, and explanation of this dataset.

B. Software

The implementations were implemented through RStudio 2023.09.1+494 "Desert Sunflower" Release, using the Learning Statistics with R (LSR) library and MBESS Package.

C. Effect size

Let $\boldsymbol{M} \in \mathbb{R}^{F\times N}$ denote the matrix gathering with F raw features and N values extracted from cell nuclei images with two groups related to the malignant cancerous and benign non-cancerous samples. Let $M_i \in \mathbb{R}^{1\times N_{1,i}}$ the malignant cancerous sample vector, and $B_i \in \mathbb{R}^{1\times N_{2,i}}$ the benign non-cancerous sample vector for each i feature, with $1 \leq i \leq F$, $N_1 \leq N$, and $N_2 \leq N$ respectively. For a detailed explanation of the estimation of the effect size and its interpretation, we refer the reader to [2], [3], [16]

1) Cohen's d standardized effect size: Cohen's d standardized effect size [3] is defined as:

$$d = \frac{\overline{M}_i - \overline{B}_i}{\sqrt{\frac{(N_{1,i}-1)SD^2_{N_{1,i}}+(N_{2,i}-1)SD^2_{N_{2,i}}}{N_{1,i}+N_{2,i}-2}}} \tag{1}$$

where $\overline{\bullet}$ is the mean for each group, SD is the pooled standard deviation, $N_{1,i}$ and $N_{2,i}$ are the samples size for each i feature. Note that the standard deviation is estimated from the differences between each observation and the mean for the group. These differences are the sum of squares used to avoid the positive and negative values from canceling each

other out and summing. This value is divided by the number of observations minus one, which is Bessel's correction for bias in the population calculation variance based on the least squares estimate [17]. Finally, the square root is computed. Since Cohen's d expressed the effect size for t-test results in units of variability, two assumptions must be considered: the first is related to the normal assumption and the second is the assumption of homogeneity of variance.

2) Cohen's D: Cohen's D is the Cohen's d without homogeneity of variance assumption, also known as Welch's t-test. It can be formulated as:

$$D = \frac{\overline{M_i} - \overline{B_i}}{\sqrt{\frac{SD^2_{N_{1,i}} + SD^2_{N_{2,i}}}{2}}} \tag{2}$$

3) U measures: Let Φ be the standard normal cumulative distribution function. Cohen's U_3 represents the percentage of M_i is upper half of the cases of the B_i, is computed as:

$$U_3 = \Phi(d) \tag{3}$$

In the same way, Cohen's U_2 is defined as:

$$U_2 = \Phi\left(\frac{d}{2}\right) \tag{4}$$

Cohen's U_2 measures the percentage in B_i that exceeds the same percentage in M_i. Finally, Cohen's U_1 is the following non-overlapping percentage:

$$U_1 = \frac{2U_2 - 1}{U_2} \tag{5}$$

Cohen's U_1 is the amount of combined area not shared by the two population distributions, M_i and B_i [3].

U measures lie between the values 0 and 1. A value 0 means a full overlapping or null effect, while a value 1 means a large effect. Hence the variables whose U measures are closer to 1 are better at classifying among benign and malignant samples.

D. Feature-selector-based decision rule

The effect size value (or strength) between two samples, such as the cancerous or benign non-cancerous samples, can be interpreted through the numerical decision rule presented in Table I. Precisely, the observed value from this measure is the statistical value calculated that allows a reduction in the dimensionality of data.

TABLE I
DECISION RULES FOR ASSESSING THE STRENGTH OF THE EFFECT SIZE MEASUREMENT OF THE OBSERVED DIFFERENCE. μ IS THE MEAN.

Observed value for Cohen's d and Cohen's D	**Strength**
Effect size value < 0.2	Null
$0.2 \leq$ Effect size value < 0.5	Small
$0.5 \leq$ Effect size value ≤ 0.8	Medium
Effect size value $>\ 0.8$	Large
Observed value for Cohen's U_1 to Cohen's U_3	**Strength**
Effect size value $< \mu$(rank(Effect size values))	Null
Effect size value $\geq \mu$(rank(Effect size values))	Large

E. Feature-selector-based learning

Support Vector Machines (SVM) are based on finding a hyperplane that separates the classes under study at the same distance. SVM results in an optimization problem whose primary purpose is to find the largest number of margins between the points in space and the separating hyperplane. The points within these margins are called support vectors. Fig. 1 illustrates how a Support Vector Machine works.

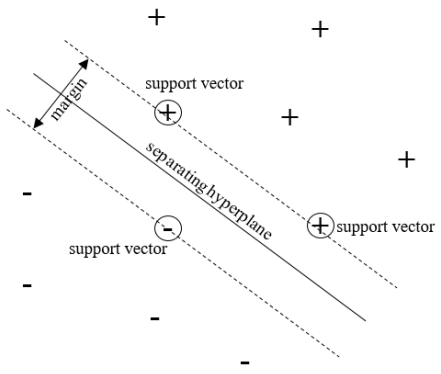


Fig. 1. Linear-SVM illustration

Let a set with N observations value in bi-dimensional space, with order pairs (x_1, x_2). Let a variable target with two-factor levels $Y_i \in \{+1, -1\}$. Then the separating hyperplane is defined as:

$$\vec{w}.\vec{x_i} + b = 0 \tag{6}$$

$$x_i = \vec{x_\perp} + r\frac{\vec{w}}{||\vec{w}||} \tag{7}$$

The expression (7) is the distance of a point to the decision boundary, where r is the distance of x_i from the decision boundary whose normal vector is $\vec{w}$, and $\vec{x_\perp}$ is the orthogonal projection of x_i onto this boundary. See [18], [19] for a comprehensive treatment of the properties of the SVM.

F. Confidence intervals (CIs)

CIs are estimated using the non-centrality parameter (NCP) method. This is a pivot method that finds the NCP of a non-central t, F, or χ^2 distribution that places the observed t, F, or χ^2 test statistic at the desired probability point of the distribution [20]. Furthermore, the NCP method is useful to find the confidence intervals for Cohen's d and Cohen's D, but not for the U measures. For these measures, the Bootstrap method is used [19], [21]. After estimating these confidence bounds on the NCP, they are converted into the effect size metric to yield a confidence interval for the effect size.

Some equations are necessary to express the relationship between effect size and NCP. Please note that equation (8) is given the Cohen expression (1) and the t-test expression defined in (10), therefore it is possible to convert the t-test statistic into a Cohen's d value.

$$d = t\sqrt{\frac{N_{1,i} + N_{2,i}}{N_{1,i}N_{2,i}}} \tag{8}$$

In such case, especially for the standardized mean difference, the population NCP for the two independent groups t-test is defined as:

$$\lambda = \frac{\mu_{1,i} - \mu_{2,i}}{\sigma_i \sqrt{\frac{1}{N_{1,i}} + \frac{1}{N_{2,i}}}} \tag{9}$$

Where $\mu_{1,i}$ and $\mu_{2,i}$ are the population means for the malignant and benign samples for each i feature. Since Cohen's d assumes homogeneous variances, both $\sigma^2_{1,i}$ and $\sigma^2_{2,i}$ are identical, so the populations mean the difference is divided by the product between σ_i and the square root. The population NCP is estimated as follows:

$$\hat{\lambda} = \frac{\overline{M_i} - \overline{B_i}}{SD_i \sqrt{\frac{1}{N_{1,i}} + \frac{1}{N_{2,i}}}} \tag{10}$$

Note that $\hat{\lambda}$ equals the observed t-test statistic. For the estimated non-centrality parameter, SD is the pooled deviation defined in (1). Given the NCP equations, it is possible to compute the confidence intervals for λ. Let p, the probability that the feature is contained within a random interval is $1-\alpha$, where α is the Type I error rate and $1-\alpha$ is the confidence level coverage, and let T be the standardized of the effect size. The confidence intervals are estimated using the following equation, with ν degrees of freedom.

$$p\left[t_{(\alpha/2;\nu)} \leq T \leq t_{(1-\alpha/2;\nu)}\right] = 1 - \alpha \tag{11}$$

For Cohen's d, T is given as:

$$T = \frac{(\overline{M_i} - \overline{B_i}) - (\mu_{1,i} - \mu_{2,i})}{SD_i \sqrt{\frac{1}{N_{1,i}} + \frac{1}{N_{2,i}}}} \tag{12}$$

For Cohen's D, T is given as:

$$T = \frac{(\overline{M_i} - \overline{B_i}) - (\mu_{1,i} - \mu_{2,i})}{\sqrt{\frac{SD^2_{1,i}}{N_{1,i}} + \frac{SD^2_{2,i}}{N_{2,i}}}} \tag{13}$$

To get the confidence intervals for Cohen's d, it is necessary to find the confidence intervals for λ, and then the bounds are transformed to the Cohen's d scale using (8). The CIs can therefore be written as:

$$p[\lambda_L \leq \lambda \leq \lambda_U] = 1 - \alpha \tag{14}$$

The value λ_L is found such that $p(\hat{\lambda}|\lambda_L) = \alpha/2$ and λ_U is found such that $p(\hat{\lambda}|\lambda_U) = 1 - \alpha/2$ with $\nu = N_{1,i} + N_{2,i} - 2$ degrees of freedom. Thus, the confidence intervals are finally defined as:

$$\left[\lambda_L \sqrt{\frac{N_{1,i} + N_{2,i}}{N_{1,i} N_{2,i}}}; \lambda_U \sqrt{\frac{N_{1,i} + N_{2,i}}{N_{1,i} N_{2,i}}}\right] \tag{15}$$

The Cohen's D procedure is similar, but using T based on Welch's t-test. Note that its standard deviation is a mean of the standard deviations of the benign and malignant groups, so the equation (8) should be written as:

$$d = \frac{t \sqrt{\frac{SD^2_{1,i}}{N_{1,i}} + \frac{SD^2_{2,i}}{N_{2,i}}}}{\sqrt{\frac{SD^2_{1,i} + SD^2_{2,i}}{2}}} \tag{16}$$

Moreover, the degrees of freedom are approximated by using the Welch-Satterthwaite equation [22]:

$$\nu = \frac{\left(\frac{SD^2_{1,i}}{N_{1,i}} + \frac{SD^2_{2,i}}{N_{2,i}}\right)^2}{\frac{\left(\frac{SD^2_{1,i}}{N_{1,i}}\right)^2}{N_{1,i}-1} + \frac{\left(\frac{SD^2_{2,i}}{N_{2,i}}\right)^2}{N_{2,i}-1}} \tag{17}$$

Confidence intervals for the U measures are estimated using the Bootstrap method. The main idea is to take B random samples from the original sample, just as samples are taken from the population. This allows us to create a Bootstrap distribution for $\hat{\theta}^*$ or a Bootstrap estimate that follows the sampling distribution for $\hat{\theta}$. Therefore, the goal is to estimate the distribution of $\frac{\hat{\theta}-\theta}{SD(\hat{\theta})}$ so that confidence intervals can be constructed, for this the equation (11) is used, but defining T as:

$$T^*_b = \frac{\hat{\theta}^*_b - \hat{\theta}}{SD(\hat{\theta}^*_b)} \tag{18}$$

where $\hat{\theta}^*_b$ is the value of $\hat{\theta}$ for the b^{th} bootstrap sample. Likewise, $\hat{\theta}$ is the effect size of interest. The below algorithm represents the Bootstrap procedure.

Algorithm 1 Confidence intervals for U measures.

Require: $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$ ▷ Data
Require: B_1 ▷ Bootstrap samples numbers for $\hat{\theta}^*$
Require: B_2 ▷ Bootstrap samples numbers for $SD(\hat{\theta}^*)$
Require: $1-\alpha$ ▷ Confidence level
  for $b = 1$ to B_1 **do**
    $\mathbf{X}^*_b = \{x^*_1, x^*_2, \ldots, x^*_n\}$ ▷ Generate B_1 Bootstrap sample from $\mathbf{X}$
    Compute $\hat{\theta}^*_b$
  end for
  for $b = 1$ to B_2 **do**
    $\mathbf{X}^*_b = \{x^*_1, x^*_2, \ldots, x^*_n\}$ ▷ Generate B_2 Bootstrap sample from $\mathbf{X}$
    Compute $SD(\hat{\theta}^*_b)$
  end for
  Sort Bootstrap values $\{\hat{\theta}^*_1, \hat{\theta}^*_2, \ldots, \hat{\theta}^*_B\}$ in ascending order
  Compute T^*_b and find the percentiles $\hat{t}^*_{\alpha/2}$ and $\hat{t}^*_{1-\alpha/2}$
    return $[\hat{\theta} - \hat{t}^*_{1-\alpha/2} SD(\hat{\theta}^*); \hat{\theta} - \hat{t}^*_{\alpha/2} SD(\hat{\theta}^*)]$

III. Results and Discussion

For illustration, Fig. 2 shows some feature examples from the dataset introduced in Section II-A. t-distributed Stochastic Neighbor Embedding (t-SNE) was used to yield the scatter plot between couples for malignant cancerous samples (red circles),

and benign non-cancerous samples (blue circles). Note that samples can be separated linearly, allowing the use of linear-methods-based learning schemes.

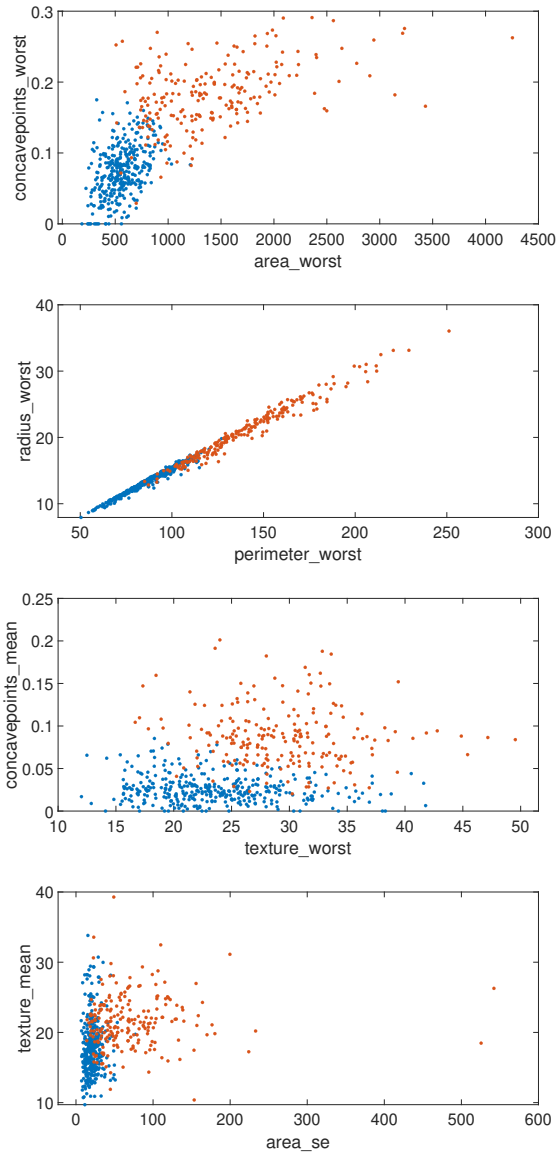
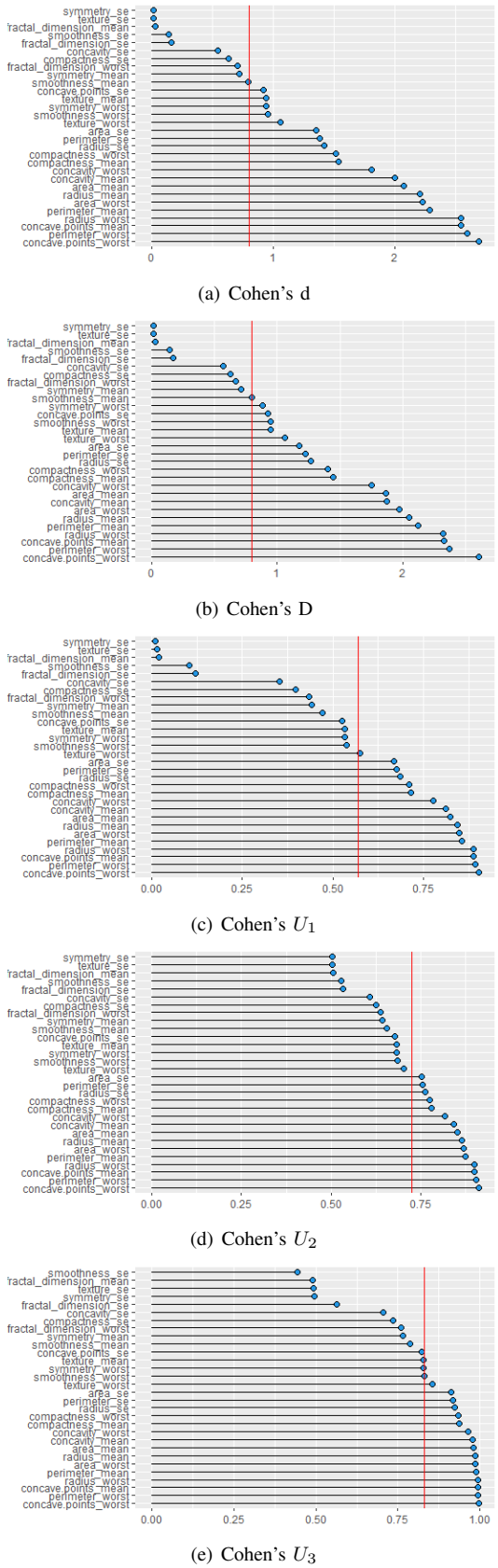


Fig. 2. t-SNE scatter plot examples of some initial features for malignant cancerous samples (red circles), and benign non-cancerous samples (blue circles). (a) Area worst vs. Concave points worst y (b) Perimeter worst vs. Radius worst (c) Texture worst vs. Concave point mean. y (d) Area see vs. Texture mean.

Fig. 3 and Table II show the effect size values observed for each feature using the different measures introduced in Section II-C and the feature-selector-based decision rule, Section II-D. Note that in Fig. 3 the Cohen's d and Cohen's D results are based on the decision rule when the effect size value is greater than 0.8, while Cohen's U_1 to Cohen's U_3 results, the decision rule is based on the mean of its rank. A red vertical line in the figures shows the decision rule threshold. Decision rule results in Tables III and confidence intervals results in Table IV complement the observed results of Fig. 3.



(a) Cohen's d

(b) Cohen's D

(c) Cohen's U_1

(d) Cohen's U_2

(e) Cohen's U_3

Fig. 3. Decision rule threshold results are illustrated by a red line for the features under study. (a) Cohen's d: 0.8. y (b) Cohen's D: 0.8. (c) Cohen's U_1: 0.5. (d) Cohen's U_2: 0.7.(e) Cohen's U_3: 0.8.

In Table II, the Feature column contains the variables and the Effect Sizes columns show the more significant feature for each effect size measure based on the decision rule. The feature-selector is highlighted with a black x and the common features for all effect sizes measure with a blue x.

TABLE II
EFFECT SIZE FEATURE-SELECTOR-BASED LEARNING RESULTS. IN BLUE, COMMON FEATURES FOR ALL EFFECT SIZE MEASURES ARE REPORTED.

Features	Effect Sizes					Common features
	Cohen's d	Cohen's D	Cohen's U_1	Cohen's U_2	Cohen's U_3	
Radius mean	x	x	x	x	x	x
Perimeter mean	x	x	x	x	x	x
Area mean	x	x	x	x	x	x
Texture mean	x	x				
Compactness mean	x	x	x	x	x	x
Smoothness mean						
Concavity mean	x	x	x	x	x	x
Concave points mean	x	x	x	x	x	x
Symmetry mean						
Fractal dimension mean						
Radius se	x	x	x	x	x	x
Perimeter se	x	x	x	x	x	x
Area se	x	x	x	x	x	x
Texture se						
Compactness se						
Smoothness se						
Concavity se						
Concave points se	x	x			x	
Symmetry se						
Fractal dimension se						
Radius worst	x	x	x	x	x	x
Perimeter worst	x	x	x	x	x	x
Area worst	x	x	x	x	x	x
Texture worst	x	x	x		x	
Compactness worst	x	x	x	x	x	x
Smoothness worst	x	x			x	
Concavity worst	x	x	x	x	x	x
Concave points worst	x	x	x	x	x	x
Symmetry worst	x	x				
Fractal dimension worst						

For feature-selector-based learning, a classical SVM with a linear kernel was used as a learner tool with 10-fold cross-validation. The experiment was repeated 20 times randomly to assess consistency in the detection. The True Positive Rate or Recall or Sensitivity (TPR), False Positive Rate (FPS) or false alarm rate, Accuracy (ACC), and Area Under the Curve (AUC) classification performance metrics yield interesting results in detecting abnormalities in breast cancer. The metrics used in this study show excellent results, over 90% for all the effect sizes studied, except for Cohen's U1 with an accuracy of 61.18%, but an acceptable AUC. Also, all common features for all the effect sizes were tested yielding an excellent performance. Additionally, the feature-selector filter method Relief [23] was tested for comparison with the effect size measures used, see Table V. It is important to highlight that the complexity of effect sizes is linear $\mathcal{O}(f.n)$, while the Relief is quadratic $\mathcal{O}(f.n^2)$, where f is the number of features and n is the number of instances [24]. These excellent results suggest that the effect size proposed is within the standards of the feature-selector methods.

TABLE III
RANKS OBSERVED, MEANS, AND DECISION RULES

Effect Size	Mean	Rank	Decision Rule
Cohen's d	1.29	$[0, 2.7]$	0.8
Cohen's D	1.20	$[0, 2.6]$	0.8
Cohen's U_1	0.56	$[0, 0.9]$	0.5
Cohen's U_2	0.74	$[0.5, 0.95]$	0.7
Cohen's U_3	0.84	$[0.3, 1]$	0.8

TABLE IV
CONFIDENCE INTERVALS FOR THE EFFECT SIZES STUDIED.

Features	Confidence Interval				
	Cohen's d	Cohen's D	Cohen's U_1	Cohen's U_2	Cohen's U_3
Radius mean	$[1.99, 2.41]$	$[1.81, 2.31]$	$[0.81, 0.87]$	$[0.84, 0.88]$	$[0.97, 0.99]$
Texture mean	$[0.76, 1.12]$	$[0.77, 1.13]$	$[0.45, 0.59]$	$[0.64, 0.71]$	$[0.77, 0.86]$
Perimeter mean	$[2.07, 2.5]$	$[1.88, 2.39]$	$[0.82, 0.88]$	$[0.85, 0.89]$	$[0.98, 0.99]$
Area mean	$[1.86, 2.28]$	$[1.62, 2.12]$	$[0.78, 0.85]$	$[0.82, 0.87]$	$[0.96, 0.98]$
Smoothness mean	$[0.61, 0.96]$	$[0.62, 0.97]$	$[0.38, 0.64]$	$[0.62, 0.68]$	$[0.73, 0.83]$
Compactness mean	$[1.34, 1.72]$	$[1.24, 1.66]$	$[0.66, 0.75]$	$[0.74, 0.8]$	$[0.91, 0.95]$
Concavity mean	$[1.79, 2.2]$	$[1.64, 2.12]$	$[0.77, 0.84]$	$[0.81, 0.86]$	$[0.96, 0.98]$
Concave points mean	$[2.31, 2.76]$	$[2.07, 2.61]$	$[0.85, 0.9]$	$[0.87, 0.91]$	$[0.98, 0.99]$
Symmetry mean	$[0.54, 0.89]$	$[0.53, 0.89]$	$[0.35, 0.51]$	$[0.6, 0.67]$	$[0.7, 0.81]$
Fractal dimension mean	$[-0.14, 0.19]$	$[-0.14, 0.19]$	$[0, 0.14]$	$[0.5, 0.53]$	$[0.42, 0.55]$
Radius se	$[1.23, 1.61]$	$[1.05, 1.49]$	$[0.63, 0.73]$	$[0.73, 0.78]$	$[0.89, 0.94]$
Texture se	$[-0.15, 0.18]$	$[-0.14, 0.18]$	$[0, 0.13]$	$[0.5, 0.53]$	$[0.42, 0.56]$
Perimeter se	$[1.19, 1.56]$	$[1.01, 1.45]$	$[0.61, 0.72]$	$[0.72, 0.78]$	$[0.88, 0.94]$
Area se	$[1.16, 1.54]$	$[0.96, 1.4]$	$[0.61, 0.71]$	$[0.71, 0.77]$	$[0.87, 0.93]$
Smoothness se	$[-0.03, 0.3]$	$[-0.02, 0.3]$	$[0, 0.21]$	$[0.5, 0.56]$	$[0.37, 0.51]$
Compactness se	$[0.45, 0.8]$	$[0.44, 0.8]$	$[0.3, 0.47]$	$[0.59, 0.65]$	$[0.67, 0.78]$
Concavity se	$[0.36, 0.71]$	$[0.4, 0.73]$	$[0.25, 0.43]$	$[0.57, 0.63]$	$[0.64, 0.76]$
Concave points se	$[0.74, 1.1]$	$[0.75, 1.11]$	$[0.44, 0.58]$	$[0.64, 0.7]$	$[0.77, 0.86]$
Symmetry se	$[-0.15, 0.18]$	$[-0.16, 0.19]$	$[0, 0.13]$	$[0.5, 0.53]$	$[0.42, 0.56]$
Fractal dimension se	$[-0.008, 0.33]$	$[-0.006, 0.33]$	$[0, 0.23]$	$[0.5, 0.56]$	$[0.49, 0.62]$
Radius worst	$[2.31, 2.76]$	$[2.07, 2.61]$	$[0.85, 0.9]$	$[0.87, 0.91]$	$[0.98, 0.99]$
Texture worst	$[0.87, 1.24]$	$[0.88, 1.25]$	$[0.5, 0.63]$	$[0.66, 0.73]$	$[0.81, 0.89]$
Perimeter worst	$[2.37, 2.82]$	$[2.11, 2.66]$	$[0.86, 0.91]$	$[0.88, 0.92]$	$[0.991, 0.997]$
Area worst	$[2.01, 2.44]$	$[1.72, 2.24]$	$[0.81, 0.87]$	$[0.84, 0.88]$	$[0.97, 0.99]$
Smoothness worst	$[0.78, 1.13[$	$[0.76, 1.13]$	$[0.46, 0.6]$	$[0.65, 0.71]$	$[0.78, 0.87]$
Compactness worst	$[1.32, 1.7]$	$[1.19, 1.62]$	$[0.65, 0.75]$	$[0.74, 0.8]$	$[0.9, 0.95]$
Concavity worst	$[1.61, 2.01]$	$[1.54, 1.98]$	$[0.73, 0.81]$	$[0.78, 0.84]$	$[0.94, 0.97]$
Concave points worst	$[2.46, 2.92]$	$[2.35, 2.87]$	$[0.87, 0.92]$	$[0.89, 0.92]$	$[0.993, 0.998]$
Symmetry worst	$[0.76, 1.12]$	$[0.69, 1.08]$	$[0.45, 0.59]$	$[0.64, 0.71]$	$[0.77, 0.86]$
Fractal dimension worst	$[0.53, 0.88]$	$[0.48, 0.85]$	$[0.34, 0.5]$	$[0.6, 0.67]$	$[0.7, 0.81]$

TABLE V
TRUE POSITIVE RATE OR RECALL (TPR), FALSE POSITIVE RATE (FPS), ACCURACY (ACC), AND AREA UNDER THE CURVE (AUC) CLASSIFICATION PERFORMANCE METRICS WITH THIS DATASET

Features	***TPR***	***FPS***	***ACC***	***AUC***
Cohen's d	97.20	92.06	95.29	0.98
Cohen's D	90.65	90.48	90.59	0.98
Cohen's U_1	72.90	41.27	61.18	0.88
Cohen's U_2	98.13	84.13	92.94	0.98
Cohen's U_3	96.47	99.07	92.06	0.98
Common features	94.39	93.65	94.12	0.97
Relief	100	95.24	98.24	0.99

IV. CONCLUSIONS

In this work, a statistical feature-selector-based learning tool based on effect sizes was proposed to detect abnormalities in breast cancer from features extracted from cell nuclei images. Five parametric methods based on Cohen's d, Cohen's D, Cohen's U_1, Cohen's U_2, and Cohen's U_3 were used as feature-selectors to estimate a dimensional reduction of the data based on a numeric decision rule. To assess the potentiality of the tool proposed a classical SVM with a linear kernel was used as a learner classifier with a 10-fold cross-validation. The experiment was repeated 20 times randomly to assess consistency in detecting abnormalities in breast cancer. The True Positive Rate or Recall or Sensitivity (TPR), False Positive Rate (FPS) or false alarm rate, Accuracy (ACC), and Area Under the Curve (AUC) were used as metric performance achieving excellent results, over 90% for all the effect sizes measures studied, except for Cohen's U1 with accuracy over 60%, but an acceptable AUC. These excellent results suggest that the effect size of statistical feature-selector-based learning to detect breast cancer is within the standards of the feature-selector methods. A notable advantage of using effect size as feature-selection-based learning is the lower computational complexity and versatility. The effect size measures could be treated as a filter method, so the complexity is lower than

other methods when the data dimension is high because they are estimated directly from the data and are independent of the sample size. The main limitation lies in the statistical significance. It can be misleading if it is influenced by sample size, since increasing the sample size may increase the probability of finding a statistically significant effect. Future work will focus on a comprehensive evaluation of the proposed approach with parametric and non-parametric effect size measures as feature-selection-based learning and on deriving instances of the method with other datasets tailored for specific medical applications in detecting abnormalities in breast cancer.

REFERENCES

[1] "World health organization: Breast cancer," https://www.who.int/news-room/fact-sheets/detail/breast-cancer, 2024, accessed: 2024-05-05.

[2] G. Cumming and R. Calin-Jageman, *Introduction to the New Statistics. Estimation, Open Science, and Beyond*. Routledge, 2024.

[3] J. Cohen, *Statistical power analysis for the behavioral sciences*. Routledge, 2013.

[4] W. Wolberg, O. Mangasarian, N. Street, and W. Street, "Breast Cancer Wisconsin (Diagnostic)," UCI Machine Learning Repository, 1995, doi: https://doi.org/10.24432/C5DW2B.

[5] Y. Li, T. Li, and H. Liu, "Recent advances in feature selection and its applications," *Knowledge and Information Systems*, vol. 53, pp. 551–577, 2017.

[6] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM computing surveys (CSUR)*, vol. 50, no. 6, pp. 1–45, 2017.

[7] J. Miao and L. Niu, "A survey on feature selection," *Procedia Computer Science*, vol. 91, pp. 919–926, 2016, promoting Business Analytics and Quantitative Management of Technology: 4th International Conference on Information Technology and Quantitative Management (ITQM 2016).

[8] M. Saraswat and K. V. Arya, "Feature selection and classification of leukocytes using Random Forest," *Medical & Biological Engineering & Computing*, vol. 52, no. 12, pp. 1041–1052, 2014.

[9] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Computers in Biology and Medicine*, vol. 112, p. 103375, 2019.

[10] C. Lu, D. Romo-Bucheli, X. Wang, A. Janowczyk, S. Ganesan, H. Gilmore, D. Rimm, and A. Madabhushi, "Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers," *Laboratory Investigation*, vol. 98, no. 11, pp. 1438–1448, 2018.

[11] A. Osareh and B. Shadgar, "Machine learning techniques to diagnose breast cancer," in *2010 5th International Symposium on Health Informatics and Bioinformatics*, 2010, pp. 114–120.

[12] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Systems with Applications*, vol. 36, no. 2, Part 2, pp. 3240–3247, 2009.

[13] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction Foundations and Applications*. Springer, 2006.

[14] A. Haque, *Feature Engineering and Selection for Explainable Models*. Lulu.com, 2023.

[15] W. N. Street, W. Wolberg, and O. Mangasarian, "Nuclear feature extraction for breast tumor diagnosis," *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 1905, pp. 861–870, 1993.

[16] G. Cumming, *Understanding the New Statistics Effect Sizes, Confidence Intervals, and Meta-Analysis*. Routledge, 2012.

[17] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs," *Frontiers in psychology*, vol. 4, p. 863, 2013.

[18] K. P. Murphy, *Probabilistic Machine Learning An Introduction*. The MIT Press, 2022.

[19] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. Springer, 2023.

[20] K. Kelley, "Confidence intervals for standardized effect sizes: Theory, application, and implementation," *Journal of Statistical Software*, vol. 20, no. 8, p. 1–24, 2007.

[21] J. Hardin, "Bootstrapping," https://st47s.com/Math154/Notes/boot.html, 2024, accesed: 2024-08-06.

[22] N. A. Ahad and S. S. S. Yahaya, "Sensitivity analysis of Welch's t-test," vol. 1605, no. 1, 2014, pp. 888–893.

[23] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Machine Learning Proceedings 1992*, D. Sleeman and P. Edwards, Eds. Morgan Kaufmann, 1992, pp. 249–256.

[24] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of Biomedical Informatics*, vol. 85, pp. 189–203, 2018.