RECyT
Año 22 / Nº 34 / 2020 / 13–23

Predicción de fracasos en implantes dentales mediante la integración de múltiples clasificadores

Predicting dental implant failures by integrating multiple classifiers

Nancy B. Ganz[1, *], Alicia E. Ares[1], Horacio D. Kuna[2]

1- Instituto de Materiales de Misiones (IMAM), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Facultad de Ciencias Exactas Químicas y Naturales (FCEQyN), Universidad Nacional de Misiones (UNaM), Félix de Azara 1552, N3300LQH, Posadas, Misiones, Argentina.

2- Instituto de Investigación, Desarrollo e Innovación en Informática (IIDII), Facultad de Ciencias Exactas Químicas y Naturales (FCEQyN), Universidad Nacional de Misiones (UNaM), Félix de Azara 1552, N3300LQH, Posadas, Misiones, Argentina.

*E-mail: nancy.bea.ganz@gmail.com

Recibido: 10/12/2019; Aprobado: 29/05/2020

Resumen

El campo de la Ciencia de Datos ha tenido muchos avances respecto a la aplicación y desarrollo de técnicas en el sector de la salud. Estos avances se ven reflejados en la predicción de enfermedades, clasificación de imágenes, identificación y reducción de riesgos, así como muchos otros. Este trabajo tiene por objetivo investigar el beneficio de la utilización de múltiples algoritmos de clasificación, para la predicción de fracasos en Implantes Dentales de la provincia de Misiones, Argentina y proponer un procedimiento validado por expertos humanos. El modelo abarca la integración de varios tipos de clasificadores. La experimentación es realizada con cuatro conjuntos de datos, un conjunto de Implantes Dentales confeccionado para el estudio de caso, un conjunto generado artificialmente y otros dos conjuntos obtenidos de distintos repositorios de datos. Los resultados arrojados del enfoque propuesto sobre el conjunto de datos de Implantes Dentales, es validado con el desempeño en la clasificación por expertos humanos. Nuestro enfoque logra un porcentaje de acierto del 93% de casos correctamente identificados, mientras que los expertos humanos consiguen un 87% de precisión. En base a esto podemos alegar, que los sistemas de múltiple clasificadores son un buen enfoque para predecir fracasos en implantes dentales.

Palabras clave: Combinación de clasificadores, clasificación, aprendizaje automático, implantes dentales, predicción de fracasos.

Abstract

The field of data science has made many advances in the application and development of techniques in several aspects of the health sector, such as in disease prediction, image classification, risk identification and risk reduction. Based on this, the objectives of this work were to investigate the benefit of using multiple classification algorithms to predict dental implant failures in patients from Misiones province, Argentina, and to propose a procedure validated by human experts. The model used the integration of several types of classifiers.The experimentation was performed with four data sets: a data set of dental implants made for the case study, an artificially generated data set, and two other data sets obtained from different data repositories. The results of the approach proposed were validated by the performance in classification made by human experts. Our approach achieved a success rate of 93% of correctly identified cases, whereas human experts achieved 87% accuracy. Based on this, we can argue that multi-classifier systems are a good approach to predict dental implant failures.

Keywords: Combination of classifiers, classification, machine learning, dental implants, prediction of failures.

Introduction

In decision-making, the combination of classification models can be fundamental, because such a combination aims to obtain an appropriate solution for a particular problem. Individually, classification methods are based on different estimation concepts or procedures. Thus, by combining them in some way, it is possible to bring together the best properties of each of them and to combine the decisions obtained with the same or different base classifiers [1]. Combination methods are those in which, given a set of already trained classifiers, the results are combined in different ways to return a more precise value than that of the individual classifiers [2]. This integration is often more accurate, because training data may not provide enough information to choose a better classifier and, in this situation, the combination is the best option. Therefore, the combination may be equivalent to very complex decision trees [3].

Based on this, the aim of the present study was to evaluate the application of multiple classifiers for the prediction of cases of dental implant failure. The data set used was based on clinical histories of patients who had undergone surgical processes of dental implant placement in the province of Misiones, Argentina. The model proposed used the following classifiers: Random Forest (RF) [4], C-Support Vector (SVC) [5], K-Nearest Neighbors (KNN) [6], Multinomial Naive Bayes (MNB) [7] and Multi-layer Perceptron (MLP) [8]. The proposed integration of these classifiers aimed to combine the results of their predictions to determine the degree of class membership and to achieve greater accuracy than that achieved by each of the classifiers individually for the target class label (dental implant failure).

The contributions of this work include the proposed of an automatic learning model for the prediction of failure in dental implants, which is a little known field. Likewise, we demonstrated that multiple classifier systems can also be applied to the case study, as they allow achieving better classification performance than that performed by the human experts.

This section has presented the motivations of our work. The rest of the paper is structured as follows: Section 2 presents related work on the application of multiple classifier systems, section 3 describes in detail the integrated approach of multiple classifiers and an overview of each of the individual classifiers, section 4 presents the experimental results obtained, and section 5 summarizes the main conclusions drawn from this work and outlines future lines of research.

Related work

Several studies have evaluated the combination or integration of classifiers to improve the percentage of success or even not to bias the decision on the results of a single classifier [9]. Miao et al. [10], for example, proposed a procedure to improve the accuracy in the identification of genes by integrating the Support Vector Machines (SVM), RF, and Extreme Learning Machines (ELM) classifiers, by applying ReliefF [11] to select the most relevant characteristics of the data set. After training and prediction with the three classifiers, the authors combined the results through the majority voting method [9]. The integration of the predictions allowed them to obtain greater accuracy than with the individual classifiers. Similarly, Catal and Nengir [12] presented a model for the classification of feelings by combining the Naive Bayes, SVM and Bagging classifiers. For the integration of predictions, the authors used the majority voting method and demonstrated that multiple classifier systems improve accuracy. Another work of similar characteristics is that of Pandey and Taruna [13], who proposed an integrated classifier using a J48 Decision Tree, K-Nearest Neighbor and Aggregating One-Dependence Estimators (AODE), on a data set of academic performance of engineering students. In this model, each individual classifier generates its predictive value and these are integrated through the probability product, where the final class label is represented by the maximum of a subsequent probability. Yan et al. [14] also proposed the integration of the Naive Bayes, Decision Tree (ID3) and Maximum Entropy classifiers with the majority voting method for semantic dependency analysis in Chinese. In this model, each of the three classifiers is trained with the same training data. The approach proposed achieved 86% accuracy in experimentation, which, according to the authors, is promising for semantic dependency analysis in Chinese. Ruano-Ordás et al. [15]the amount of acquired knowledge about the design and synthesis of pharmaceutical agents and bioactive molecules (drugs proposed a model to automatically determine the biological activity of molecules based on 2048 chemical substructures (coded using binary values) and 84 physicochemical properties (coded using discrete and continuous values). The authors performed the process in three stages: grouping of characteristics, construction and optimization of hyper parameters of each classifier, and classification. They also used SVM with Radial Basis Function (RBF) kernel, AdaBag and rpart, and combined the individual results of each classification into a single result by using the majority voting method. In addition, Oliveira et al. [16] addressed the problem of pedestrian detection using the MLP and SVM classifiers. To combine the outputs of the classifiers, these authors used two types of fusion methods: the majority voting method and the diffuse integral. The authors demonstrated that the integration allows improving the percentage of success in the classification. Nweke et al. [17]ambient assisted living, activity of daily living (ADL presented a survey of the use of multiple classifier systems in the recognition of human activity and health monitoring. These authors also sought to reduce uncertainty and ambiguity by merging the results generated by different classification models. To this end, they addressed different design and fusion approaches with multiple classifiers, such as SVM, Decision Tree (ID3, J48, C4.5), K-Nearest Neighbor, Artificial Neural Network, Naive Bayes, and RF.

Based on all the above, here we propose an automatic learning procedure using multiple classifiers for a little known field, as is the case of dental implants, and validated it with the performance in classification by human experts.

The following section contains a detailed description of our proposal.

Materials and methods

This section presents the approach proposed, which consists of an automatic learning process (Fig. 1) to obtain the degree of belonging of the class attribute of the dental implant data set.

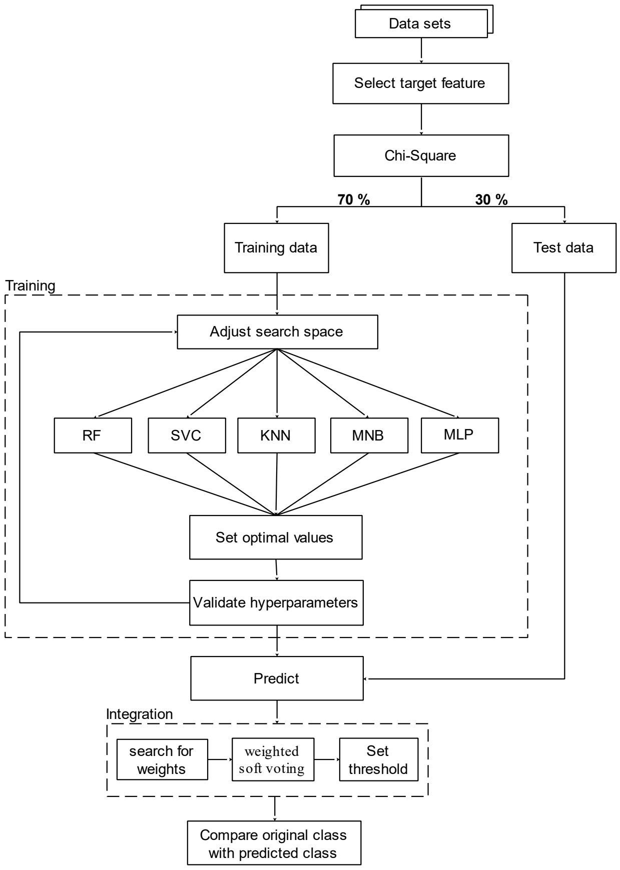


Fig. 1: Proposed approach. This representation summarizes the steps of the mechanism proposed in this work for the integration of the predictions of the following classifiers: Random Forest (RF), C-Support Vector (SVC), K-Nearest Neighbors (KNN), Multinomial Naive Bayes (MNB) and Multi-layer Perceptron (MLP).

Methods of feature selection

A significant step in automatic learning is the selection of features, as it eliminates irrelevant and redundant features, achieving reduced dimensionality and calculation requirements, as well as improving the performance of classifiers. Its purpose is to find an optimal subset of features that will provide good predictive results [11], [18], [19]. In general, feature selection methods can be divided into two categories: Filter and Wrapper. Filter methods use an approximate scale to rate a subset of characteristics and are considerably fast. Examples of filter methods include: Mutual Information [20], Correlation, Consistency, Gain Ratio [21], Information Gain [22], Symmetrical uncertainty [23], and Chi-Square [24]. Wrapper methods [25] first use an optimization algorithm in which several features are added or removed to form different subsets. These are slower than Filter methods. Examples of this type of method include: Sequential Forward Selection, Sequential Backward Selection, Bidirectional Search, and Relevance in Context [26].

In the present study, we used Chi-Square (^{X2}), which is a widely used method to select characteristics [27]–[31]. This method evaluates the value of a characteristic by calculating the statistical value of X^2 with respect to the class (equation (1)). It is given by:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{1}$$

where O_{ij} is the observed frequency and E_{ij} is the expected (theoretical) frequency. The higher the value of X^2, the greater the evidence of correlation between the two characteristics.

The cut-off criterion was the use of a level of significance, which in general is equal to 0.01, 0.05 or 0.10, but it can be any value between 0 and 1 [30], [32]. For this work, we proposed a significance level of $p \leq 0.05$ for all data sets.

Classification methods

An important step in this work was the search for the best individual classifiers for our case study. After researching the existing methods and taking into account the combination approach and the types of classifiers of the papers mentioned in the Related Work section, we propose the use of the following five classifiers: Random Forest (RF) [4], C-Support Vector (SVC) [5], K-Nearest Neighbors (KNN) [6], Multinomial Naive Bayes (MNB) [7] and Multi-layer Perceptron (MLP) [8]. In exploratory evaluations, these classifiers obtained the best performance in comparison with other explored combinations, which included different methods such as: Rpart, Ada, Gradient Boosting Machine (GBM) [3] and different Naive Bayes classifiers [7].

Random Forest, which was introduced by Leo Breiman [4], is an increasingly popular learning algorithm based on decision trees, which enables fast training, excellent performance and great flexibility to handle all types of data [33], [34]. Among the main rules used to divide binary data is the Gini index (equation (2)):

$$\mu = \sum_{a=1}^{A} p_a(1 - p_a) \tag{2}$$

where A is the target class and p_a the proportion of the class sample. This index measures the impurity of the node and is the most used [33]–[37]. A small value of a indicates that the node contains predominantly single-class observations, i.e., it is a purity node with good separation between classes [36].

C-Support Vector is a type of support vector machine, which can incorporate different kernels [5]. It can be used for classification or regression [38], [39], and its operation consists in constructing a set of hyper-planes in a high dimensional space. The separation is measured as the distance between the hyper-planes and is called the functional margin. The larger the margin, the smaller the generalization error of the classifier. Examples of some kernels [40] (equation (3)) include:

linear:(x,x')

polynomial: $(\gamma\ (x,x') + r)^d$

and

RBF: exp $(-\gamma \parallel x, x' \parallel^2)$ (3) where $\gamma > 0$

K Nearest Neighbors is a type of learning based on instances or non-generalized learning [41]. This method searches, in a set D, the k neighbors q closest to the object p to be classified in D, and assigns the class label according to most of its neighbors (equation (4)), with *dist* $(p,q) \leq$ *dist* (p,o), that is:

$$\mathrm{KNN}_k\ (p) = \{q \mid \forall q \in D,\ dist\ (p,q) \leq dist\ (p,o)\} \qquad (4)$$

where *dist*(p,o) is the distance between p and the k-th object o. To actually contribute to the adjustment, both the optimal choice of the k-value and the distance to be used for the nearest neighbors are highly dependent on the data [6].

Naive Bayes is based on the principle of the Bayes theorem, which assumes that the input characteristics are independent of each other, called conditional independence (equation (5)). It is given by:

$$f_i(X) = \prod_{j=1}^{N} P(x_j|c_i)\, P(c_i) \qquad (5)$$

where $xj = (x_1, x_2, \dots , x_N)$ is the characteristic vector and c_i, with $i = 1, 2, \dots, N$, indicates possible class labels. The training phase consists in estimating the conditional probabilities $P(x_j \mid c_i)$ and the previous probabilities $P(ci)$ [7]. In this work, we applied a variant called Multinomial Naive Bayes (MNB), which supports categorical data and is mainly used for the classification of documents and texts [42]–[46] due to its simplicity, efficiency and effectiveness.

Multi-layer Perceptron is widely used due to its ability to use both linear and nonlinear applications [47]–[52]. It consists of an input layer, one or more hidden layers and an output layer. The number of neurons in the input layer corresponds to the number of characteristics, whereas the number of neurons in the output layer corresponds to the number of outputs. The connection between the neurons in the different layers is calculated using weights (equation (6)). Its training purpose is to find suitable values for the weights of the links between the neurons. The general output function and the error function are given by:

$$\begin{aligned} y_i &= f\left(\textstyle\sum_{i=1}^{N} w_{ji} x_i\right) \\ E &= \tfrac{1}{2} \textstyle\sum_i (d_i - y_i)^2 \end{aligned} \qquad (6)$$

where x_i are the input data, w_{ji} refers to the weight values, $f(\cdot)$ is the activation function, y_i is the network i-th output, and d_i is the expected i-th output [8].

Integration of the classifiers

To determine the final class label, we applied a weighted soft voting method [53], [54]. This rule allowed achieving the best predictive results for the case study. The integration of the predictions consisted in multiplying, for each tuple, the probability value of the target and non-target class, obtained by each classifier by the weight assigned to it. The weight was determined by means of a grid search using a test parameter w with values between 0 and 1. This search was subjected to a cross-validation of 10 iterations, in which the accuracy [55], [56] of each classifier for the class in question was measured, and the value of w that achieved the best accuracy was selected [15]–[17].

Once the weights were determined, the weighted soft voting method was applied [53], [54]. This method collects the predicted class probabilities for each classifier, multiplies them by the weight assigned to each classifier, and then averages them. The final class label is derived from the class label with the highest average probability (equation (7)), given by:

$$\hat{y} = \arg max_i \textstyle\sum_{j=1}^{m} w_j p_{ij} \qquad (7)$$

where p_{ij} is the probability predicted by the j-th classifier and *wj* is the weight assigned to the j-th classifier. This approach is only recommended if the classifiers are well calibrated.

In the present work, instead of using the maximum average, we applied a threshold [3], [16], because, in exploratory evaluations, it allowed us to achieve better results in the classification. This threshold was determined by a grid search using a test parameter μ with values between 0.1 and 0.5, with 0.1 increments in each test. The value of μ selected was the one that allowed obtaining the best classification result for all the data sets used.

Generation of artificial data

An artificial data set generated with the SMOTE algorithm was used for validation [57]. This algorithm generates new artificial tuples to balance the data sample based on the nearest neighbor rule, in which, to classify a new instance, the distance between each attribute of the new instance and the rest of the instances of the data set is calculated (equation (8)) and associated with the class of the nearest instance. Therefore, given $x_i, \bar{x} \in N_{min}$, this algorithm can be described as:

$$x_{syn} = x_i + (\bar{x} - x_i)\ .\times\ rand(0{,}1) \tag{8}$$

Here, x_i is the minority class sample to be oversampled, $\bar{x}$ is another minority sample that is generally selected from N_{min} near x_i, the expression .× represents the multiplication by element, and *rand*(0,1) indicates a random number in the interval (0,1).

This method, which is widely used to balance data sets [58]–[62], has the advantage of not losing information and of being able to repeat samples with noise. The method should be provided with the following inputs: Number of minority class samples *T*; Amount of SMOTE *N%*; and Number of nearest neighbors k, and should provide the following output: (*N/10*0) * *T* synthetic minority class samples [57].

Structure of the data set

The approach proposed was used to experiment with four data sets: case study set (i.e. a data set of actual dental implant cases), and three validation sets: a data set artificially generated with the Synthetic Minority Over-sampling Technique (SMOTE) [57] based on actual dental implant cases, and two other data sets obtained from the kaggle and OpenML (Heart Disease, Breast Cancer) repositories. Table 1 presents the summarized characteristics of these sets.

To perform a classification task, after selecting the most important characteristics of a data set, it is necessary to divide the data. A common strategy is to take all labeled data and divide them into training and evaluation subsets, usually with a proportion of 70 to 80% for training and 20 to 30% for evaluation or testing [29], [30], [34], [36], [42], [63]. This division will depend to a large extent on the total number of samples and the model to be trained [16], [64]–[67]. In our case, the data were randomly divided to preserve the distribution of both classes: 70% for training and 30% for evaluation [35], [45], [47], [49], [50], [68]–[70], ensuring that all cases were represented in both sets.

Table 1: Characteristics of the data sets used for the experimental evaluation. From left to right: names of the data sets, number of samples, number of attributes per tuple, number of characteristics selected by the Chi-Square method () and size of the training and test sets.

Data set	Sample	Feature		Training	Test
Dental Implants[1]	1165	33	17	815	350
Artificial[2]	1748	33	21	1223	525
Heart Disease[3]	303	13	10	212	91
Breast Cancer[4]	277	10	5	193	84

[1]***Dental Implants:*** this data set consisted of 1165 tuples of clinical histories of patients from Misiones Province, Argentina, undergoing surgical processes of placement of dental implants. It was made up of 32 categorical characteristics and an unbalanced binary class attribute (1009 cases labeled as success and 156 as failure).

[2]***Artificial:*** this data set consisted of an artificial set generated with the SMOTE algorithm, where, to obtain the artificial cases of the minority class, the input consisted of: *T* = 156 tuples; *SMOTE N%* = 250%; and *k* = 5, and, to generate the artificial cases of the majority class, the input consisted of: *T* = 1009 cases; *SMOTE N%* = 250%; and *k* = 5. For the latter, instead of taking the subset of tuples with the lowest index, the algorithm was modified so that it took the subset of the highest index, which corresponds to the cases of the success class. The procedure to generate the cases was the same as for the minority class. Finally, the cases generated for both classes were extracted and a new artificial data set was created with a distribution similar to that of the *Dental Implants* data set.

Table 2 presents the characteristics of the *Dental Implants* and *Artificial* data sets in more detail.

[3]***Heart Disease:*** this data set consisted of a total of 303 tuples with 12 categorical attributes and one binary class attribute. Each tuple represented the data obtained from a patient. The objective characteristic refers to the presence or absence of heart disease. It consisted of 138 cases with absence of the disease and 165 with presence of the disease. This set was extracted from the kaggle repository [71].

[4]***Breast Cancer:*** this data set contains breast cancer registries obtained at the Institute of Oncology of the University Medical Center in Ljubljana, Yugoslavia. It consists of 277 tuples with 9 categorical characteristics and a binary class attribute. The class attribute reflects cases of recurrence and non-recurrence to the disease. This set was extracted from the Open Machine Learning (OpenML) repository [72].

Table 2: Dimensions of the *Dental Implants* and *Artificial* data sets.

Dimensions	Description	Features
Patient Data	Features related to the antecedents and medical conditions of the patients at the time of the intervention.	Age range, gender, profession, social security, antecedent, smoking habit, alcoholism, periodontitis, toothless, med intake, and allergy.
Implant Data	Features related to the implant used by the implant specialist.	Surface treatment, design, length, diameter, connection, and origin.
Data of the Surgical Phase	Features related to the surgical intervention and improvement of the patient's bone bed.	Season, patient zone, register, dental piece, load protocol, exodontia, bone expansion, maxillary sinus elev, regeneration of hard tissues, regeneration of soft tissues, additional procedure, placement time, bone type, prosthetic indication, and surgical complication.
Data of the Post-operative Follow-up	Particularities of the outcome of the implant placement process, i.e. whether the tissue/implant osseointegration process was successful or not.	**Post-op follow-up.**

Training

To obtain a robust model and optimize the results of the classifiers, a grid search was carried out to adjust the hyper parameters [35], [39], [40], [43], [50]the quantitative effects of heat acclimation (HA. This search was performed with the training data from each of the data sets. For this process, we specified:

1. A search space: we defined value ranges for the hyper parameters and adjusted them according to the performance measurement.
2. An optimization or adjustment algorithm: we used the GridSearchCV method [73], which is the most expensive in terms of performance, but allows covering all the search space defined.
3. An evaluation method: we used cross-validation of 10 iterations as a resampling strategy.

A measure of performance: we used the equilibrium accuracy metrics, which is given by the true positives plus the true negatives divided by the totality of samples from the data set [74].

Table 3 shows the hyper parameters that were sought to be adjusted for each classifier on each data set and the search spaces defined for each parameter. The implementation uses the Python programming language with the Scikit-learn library [75].

Table 3: Hyper parameters and search ranges defined for the RF, SVC, KNN, MNB and MLP classifiers.

Classifiers	Hyper parameters	Search space
RF	n_estimators	range (1, 150)
	criterion	gini, entropy
	bootstrap	True, False
SVC	kernel	linear, rbf, poly
	C	range (1, 10)
	gamma	range (1, 10)
	degree	range (1, 10)
KNN	n_neighbors	range (1, 100)
	weights	uniform, distance
	p	manhattan, euclidean
MNB	alpha	[0, 0.1, 0.2, 0.3, ..., 0.9, 1]
	fit_prior	True, False
	class_prior	[0.5,0.5], [0.4,0.6], [0.6,0.4]
MLP	hidden_layer_sizes	range (1, 10)
	activation	logistic, tanh, relu
	alpha	[0.0001, 0.05]
	solver	lbfgs, sgd, adam
	learning_rate	constant, invscaling

Evaluation parameters

The parameters used to evaluate and compare the performance of the individual classifiers with the approach proposed were: true positive (TP), true negative (TN), false positive (FP), false negative (FN), sensitivity, specificity, accuracy and error [55], [56]. TP is the percentage of correctly classified observations of the target class; TN is the percentage of correctly classified observations of the non-target class; FP is the percentage of erroneously classified observations of the non-target class; FN is the percentage of erroneously classified observations of the target class; sensitivity (equation (9)) is the ability of the model to correctly classify the target samples; specificity (equation (10)) is the fraction of non-target samples classified as non-target samples by the model; accuracy (equation (11)) is the total proportion of instances correctly classified for both classes; and error (equation (12)) allows the total proportion of instances incorrectly classified for both classes to be measured.

$$ensitivity = \frac{TP}{TP + FN} \tag{9}$$

$$Specificity = \frac{TN}{TN + FP} \tag{10}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$Error = \frac{FP + FN}{TP + TN + FP + FN} \tag{12}$$

Human level classification performance

Human-level performance allows estimating an optimal error rate and corroborating the operation of the classification system. To evaluate the performance of the proposed approach on the *Dental Implants* data set, a comparison was made with human expert opinion. The evaluation was subject to classification by two experts in the area (selected from the Provincial Registry of Professionals who practice Maxillofacial Buco Surgery, Implantology, Periodontics and Tissue Manipulation), each of whom was provided with a random sample distinct from the 10% prevalence of cases. The cases were presented without the label so that the experts could classify them according to their experience, and in this way be able to contrast with the values found by our classification approach.

Experimental results

This section presents the results of applying the proposed approach to the four data sets.

As described in the section on materials and methods, the predictions were integrated through the weighted soft voting method. Using the threshold value that allowed obtaining the best classification accuracy to be obtained in each data set.

Table 4 shows the optimal values found in training for each of the classifiers with the training data for each data set. Table 5 presents the success percentages obtained by each classifier individually and the result of the proposed approach to the test data from the data sets used.

Table 4: Hyper parameters and optimal values found for the RF, SVC, KNN, MNB and MLP classifiers on the *Dental Implants, Artificial, Heart Disease* and *Breast Cancer* data sets.

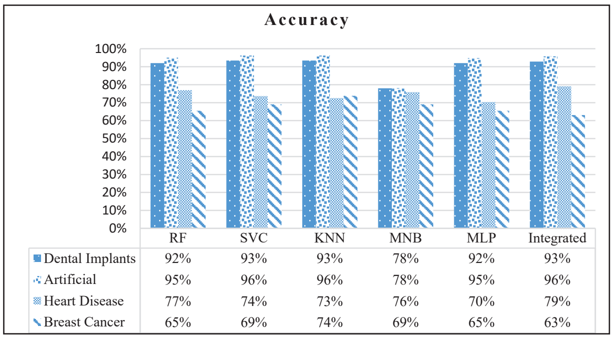
Classifiers	Hyper parameters	Optimal values			
		Dental Implants	Artificial	Heart Disease	Breast Cancer
RF	n_estimators	8	2	7	7
	criterion	entropy	entropy	gini	gini
	bootstrap	True	False	True	True
SVC	kernel	rbf	rbf	rbf	liner
	C	1	1	1	1
	gamma	1	1	1	1
	degree	0	0	0	0
KNN	n_neighbors	20	40	2	50
	weights	distance	distance	uniform	uniform
	p	euclidean	euclidean	manhattan	manhattan
MNB	alpha	1	0.7	0	0.2
	fit_prior	True	True	True	True
	class_prior	[0.6,0.4]	[0.6,0.4]	[0.6,0.4]	[0.6,0.4]
MLP	hidden_ layer_sizes	10	10	10	10
	activation	logistic	logistic	relu	logistic
	alpha	0.05	0.05	0.0001	0.0001
	solver	lbfgs	lbfgs	lbfgs	lbfgs
	learning_rate	constant	constant	constant	constant

Table 5: Efficiency in the success of the RF, SVC, KNN, MNB, and MLP classifiers and the proposed approach (Integrated) to the *Dental Implants, Artificial, Heart Disease* and *Breast Cancer* data sets.

Data sets	Classifiers	Target class	Non-target class
		Sensitivity	Specificity
Dental Implants	RF	59%	98%
	SVC	64%	**99%**
	KNN	64%	**99%**
	MNB	72%	79%
	MLP	66%	97%
	Integrated	**75%**	96%
Artificial	RF	81%	97%
	SVC	81%	**99%**
	KNN	81%	**99%**
	MNB	60%	81%
	MLP	82%	97%
	Integrated	**89%**	97%
Heart Disease	RF	81%	71%
	SVC	70%	**79%**
	KNN	70%	76%
	MNB	77%	74%
	MLP	72%	68%
	Integrated	**94%**	58%
Breast Cancer	RF	36%	78%
	SVC	36%	83%
	KNN	20%	**97%**
	MNB	52%	76%
	MLP	32%	80%
	Integrated	**60%**	64%

Table 5 shows that the SVC and KNN classifiers achieved the best performance over the non-target class for all data sets compared to the other classifiers, even exceeding the approach proposed in all cases. For the target class, it can be seen that the integration of the predictions of the five classifiers allowed achieving the highest success rate. For this class, it is also observed that the performance of the individual classifiers was varied. While the performance of the integration of the predictions was not the best option for the non-target class, it does not mean that it was the worst compared to the individual predictions. The integration of the probabilities for the target class was the best option, since it allowed obtaining the highest percentage of success.

The following graph (Fig. 2) presents the percentage of the accuracy metric achieved for each classifier and the proposed approach to the four data sets used.



	RF	SVC	KNN	MNB	MLP	Integrated
■ Dental Implants	92%	93%	93%	78%	92%	93%
⁘ Artificial	95%	96%	96%	78%	95%	96%
▨ Heart Disease	77%	74%	73%	76%	70%	79%
⟍ Breast Cancer	65%	69%	74%	69%	65%	63%

Fig. 2: Accuracy of the RF, SVC, KNN, MNB, and MLP classifiers and the proposed approach (Integrated) on the Dental Implants, Artificial, Heart Disease and Breast Cancer data sets.

Figure 2 shows that the SVC and KNN classifiers and the proposed approach showed the best performance on the *Dental Implants* and *Artificial* data sets. Also, the proposed approach achieved the best accuracy on the *Heart Disease* data set. The results on the *Breast Cancer* data set were not as good as with our model, although it was consistent in comparison with the results obtained with the other classifiers.

Finally, the results achieved with the proposed approach on the *Dental Implants* data set were compared with the accuracy achieved in classification by human experts (Fig. 3). Our model achieved 93% overall accuracy, with 7% error, whereas, on average, the classification made by the experts achieved a total accuracy of 87%, with an average error of 13% (Table 6).

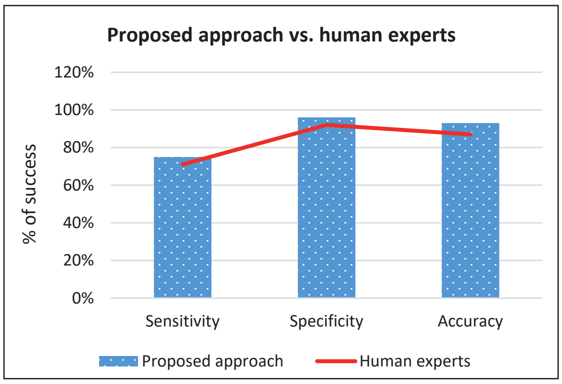


Fig. 3: Values of the Sensitivity, Specificity and Accuracy metrics achieved by the proposed approach compared to the classification made by the human experts.

Table 6: Comparison of the evaluation parameters achieved by the proposed approach and the classification of the experts on the Dental Implants data set.

Model	Sensitivity	Specificity	Accuracy	Error
Proposed approach	75%	96%	93%	7%
Human experts	71%	92%	87%	13%

Discussion

The purpose of this work was to apply multiple classifiers to increase the successful classification of the failures of a data set of clinical records of patients who had undergone surgical processes of placement of dental implants in the Province of Misiones, Argentina. We demonstrated that, in this field, it is better to integrate the predictions of the classifiers, so as not to bias the decision on a single outcome. Likewise, using integrated predictions allows knowing different points of view or results for the same case, since the use of more than one classifier allows assuring a more precise label or classification assignment.

The proposed approach was also validated with an artificial data set generated for the case study and two other test data sets. By experimenting on the original dental implant data set, the proposed approach achieved the best success rate of the target class, compared to the performance of individual classifiers and the classification by the human experts.

The experts in oral pathologies and complex rehabilitation in oral implantology consulted agreed and remarked that, in this field of study, it is less delicate to label a case as a failure than to label it as a success when it was an eventual failure. As a result, each classifier achieved up to 72% success of the target class, whereas the human expert achieved up to 71% success, whereas the proposed approach allowed reaching 75% of cases correctly identified as failures.

The SVC and KNN classifiers achieved the best performance over the non-target class for all the data sets compared to the other classifiers, even exceeding the proposed approach. For the target class, the proposed approach allowed achieving the highest success rate and lowest error rate for all cases.

Conclusions and future work

This work allowed studying the application of multiple classifiers to a little known field. We proposed an automatic learning model to improve the performance of prediction of failure in dental implants. According to the experimental results, the multiple classifiers approach can also be applied to the prediction of dental implant failures. Based on the results of the classification by the human experts, we can say that our approach allowed achieving a superior classification performance. Therefore, we have succeeded in proposing a knowledge extraction procedure validated by human experts in a little known field.

Finally, as future work, we propose validating the proposed approach with other data sets in the area of health or medicine. In addition, we propose including or extending the classifiers used, to assess the possibility of adjusting the success rate of both classes. Finally, we also propose extending the survey of cases of clinical histories of dental implants to other parts of the country as well as to other countries.

Acknowledgements

We are grateful to the National Council for Scientific and Technical Research of Argentina (CONICET) for funding through a doctoral scholarship, granted by Resolution D Nº 4869, and to the implant specialists who collaborated in the construction of the dental implant database of this research.

References

1. **Y. Lu,** *"Knowledge integration in a multiple classifier system,"* Appl. Intell., vol. 6, no. 2, pp. 75–86, 1996.
2. **L. I. Kuncheva,** *"Combining Pattern Classifiers: Methods and Algorithms,"* in Combining Pattern Classifiers, 2nd ed., John Wiley & Sons, 2014, pp. 290–325.
3. **M. Mohandes, M. Deriche, and S. O. Aliyu,** *"Classifiers Combination Techniques: A Comprehensive Review,"* IEEE Access, vol. 6, pp. 19626–19639, 2018.
4. **L. Breiman,** *"Random Forest,"* Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
5. **C. Chang and C. Lin,** *"LIBSVM : A Library for Support Vector Machines,"* ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 1–39, 2011.
6. **N. S. Altman,** *"An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression,"* Am. Stat., vol. 46, no. 3, pp. 175–185, 1992.
7. **[C. D. Manning, P. Raghavan, and H. Schutze,** *"Text classification and Naive Bayes,"* in Introduction to Information Retrieval, Cambridge University Press, 2009, pp.

253–287.

8. B. Irie and Sei Miyake, *"Capabilities of Three-layered Perceptrons,"* IEEE nternational Conf. Neural Networks, pp. 641–648, 1988.
9. J. Fierrez, A. Morales, R. Vera-Rodriguez, and D. Camacho, "*Multiple classifiers in biometrics. part 1: Fundamentals and review,"* Inf. Fusion, vol. 44, no. December 2017, pp. 57–64, 2018.
10. Y. Miao, H. Jiang, H. Liu, and Y. dong Yao, *"An Alzheimers disease related genes identification method based on multiple classifier integration,"* Comput. Methods Programs Biomed., vol. 150, pp. 107–115, 2017.
11. G. Chandrashekar and F. Sahin, "*A survey on feature selection methods,"* Comput. Electr. Eng., vol. 40, no. 1, pp. 16–28, 2014.
12. C. Catal and M. Nangir, *"A sentiment classification model based on multiple classifiers,"* Appl. Soft Comput. J., vol. 50, pp. 135–141, 2017.
13. M. Pandey and S. Taruna, *"Towards the integration of multiple classifier pertaining to the Student's performance prediction,"* Perspect. Sci., vol. 8, pp. 364–366, 2016.
14. J. Yan, D. B. Bracewell, F. Ren, and S. Kuroiwa, *"Integration of Multiple Classifiers for Chinese Semantic Dependency Analysis,"* Electron. Notes Theor. Comput. Sci., vol. 225, no. C, pp. 457–468, 2009.
15. D. Ruano-Ordás, I. Yevseyeva, V. B. Fernandes, J. R. Méndez, and M. T. M. Emmerich, *"Improving the drug discovery process by using multiple classifier systems,"* Expert Syst. Appl., vol. 121, pp. 292–303, 2019.
16. L. Oliveira, U. Nunes, and P. Peixoto, *"On Exploration of Classifier Ensemble Synergism in Pedestrian Detection,"* IEEE Trans. Intell. Transp. Syst., vol. 11, no. 1, pp. 16–27, 2010.
17. H. F. Nweke, Y. W. Teh, G. Mujtaba, and M. A. Al-garadi, *"Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions,"* Inf. Fusion, vol. 46, no. June 2018, pp. 147–170, 2019.
18. U. M. Khaire and R. Dhanalakshmi, *"Stability of feature selection algorithm: A review,"* J. King Saud Univ. - Comput. Inf. Sci., 2019.
19. R. Zhang, F. Nie, X. Li, and X. Wei, *"Feature selection with multi-view data: A survey,"* Inf. Fusion, vol. 50, no. May 2018, pp. 158–167, 2019.
20. M. Bennasar, Y. Hicks, and R. Setchi, *"Feature selection using Joint Mutual Information Maximisation,"* Expert Syst. Appl., vol. 42, no. 22, pp. 8520–8532, 2015.
21. J. R. Quinlan, *"Induction of Decision Trees,"* Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.
22. C. E. Shannon, *"A Mathematical Theory of Communication,"* Bell Syst. Tech. J., vol. 27, no. 3, pp. 379–423, 1948.
23. S. S. Shreem, S. Abdullah, and M. Z. A. Nazri, *"Hybrid feature selection algorithm using symmetrical uncertainty and a harmony search algorithm,"* Int. J. Syst. Sci., vol. 47, no. 6, pp. 1312–1329, 2016.
24. K. Pearson, *"On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling,"* London, Edinburgh, Dublin Philos. Mag. J. Sci., vol. 50, no. 302, pp. 157–175, 1900.
25. R. Kohavi and G. H. John, "*Wrappers for feature subset selection,"* Artif. Intell., vol. 97, no. 1–2, pp. 273–324, 1997.
26. I. Guyon and A. Elisseeff, *"An Introduction to Variable and Feature Selection,"* J. Mach. Learn. Res., vol. 3, pp. 1157–1182, 2003.
27. Y. Yang and J. O. Pedersen, *"A Comparative Study on Feature Selection in Text Categorization,"* Proc. 14th Int. Conf. Mach. Learn., pp. 412–420, 1997.
28. X. Jin, A. Xu, R. Bie, and P. Guo, *"Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles,"* Data Min. Biomed. Appl., vol. 3916, pp. 106–115, 2006.
29. M. Moran and G. Gordon, *"Curious Feature Selection,"* Inf. Sci. (Ny)., vol. 485, pp. 42–54, 2019.
30. I. Sumaiya Thaseen and C. Aswani Kumar, *"Intrusion detection model using fusion of chi-square feature selection and multi class SVM,"* J. King Saud Univ. - Comput. Inf. Sci., vol. 29, no. 4, pp. 462–472, 2017.
31. H. Alshalabi, S. Tiun, N. Omar, and M. Albared, "*Experiments on the Use of Feature Selection and Machine Learning Methods in Automatic Malay Text Categorization,"* Procedia Technol., vol. 11, pp. 748–754, 2013.
32. J. Mielniczuk and P. Teisseyre, *"Stopping rules for mutual information-based feature selection,"* Neurocomputing, vol. 358, pp. 255–274, 2019.
33. G. Biau and E. Scornet, *"A random forest guided tour,"* Test, vol. 25, no. 2, pp. 197–227, 2016.
34. A. Verikas, A. Gelzinis, and M. Bacauskiene, *"Mining data with random forests: A survey and results of new tests,"* Pattern Recognit., vol. 44, no. 2, pp. 330–349, 2011.
35. D. Chong, N. Zhu, W. Luo, and X. Pan, *"Human thermal risk prediction in indoor hyperthermal environments based on random forest,"* Sustain. Cities Soc., vol. 49, no. April, p. 101595, 2019.
36. D. S. Cao, J. H. Huang, Y. Z. Liang, Q. S. Xu, and L. X. Zhang, *"Tree-based ensemble methods and their applications in analytical chemistry,"* Trends Anal. Chem., vol. 40, no. 2, pp. 158–167, 2012.
37. F. B. de Santana, W. Borges Neto, and R. J. Poppi, *"Random forest as one-class classifier and infrared spectroscopy for food adulteration detection,"* Food Chem., vol. 293, no. July 2018, pp. 323–332, 2019.
38. J. Liu and E. Zio, *"Integration of feature vector selection and support vector machine for classification of imbalanced data,"* Appl. Soft Comput. J., vol. 75, pp. 702–711, 2019.
39. J. Chorowski, J. Wang, and J. M. Zurada, *"Review and per-*

formance comparison of SVM- and ELM-based classifiers," Neurocomputing, vol. 128, pp. 507–516, 2014.

40. **J. Novakovic and A. Veljovic**, *"C-support vector classification: Selection of kernel and parameters in medical diagnosis,"* IEEE 9th Int. Symp. Intell. Syst. Informatics, pp. 465–470, 2011.

41. **M. Cover T and E. Hart P**, *"Nearest Neighbor Pattern Classification,"* IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, 1967.

42. **G. Singh, B. Kumar, L. Gaur, and A. Tyagi**, *"Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification,"* 2019 Int. Conf. Autom. Comput. Technol. Manag., pp. 593–596, 2019.

43. **S. Xu**, *"Bayesian Naïve Bayes classifiers to text classification,"* J. Inf. Sci., vol. 44, no. 1, pp. 48–59, 2018.

44. **M. Abbas, K. Ali Memon, A. Aleem Jamali, S. Memon, and A. Ahmed**, *"Multinomial Naive Bayes Classification Model for Sentiment Analysis,"* IJCSNS Int. J. Comput. Sci. Netw. Secur., vol. 19, no. 3, pp. 62–67, 2019.

45. **G. Isabelle, W. Maharani, and I. Asror**, *"Analysis on Opinion Mining Using Combining Lexicon-Based Method and Multinomial Naïve Bayes,"* 2018 Int. Conf. Ind. Enterp. Syst. Eng. (ICoIESE 2018), vol. 2, no. IcoIESE 2018, pp. 214–219, 2019.

46. **Y. Pan, H. Gao, H. Lin, Z. Liu, L. Tang, and S. Li**, *"Identification of bacteriophage virion proteins using multinomial Naïve bayes with g-gap feature tree,"* Int. J. Mol. Sci., vol. 19, no. 6, 2018.

47. **K. Bhattacharjee and M. Pant**, *"Hybrid Particle Swarm Optimization-Genetic Algorithm trained Multi-Layer Perceptron for Classification of Human Glioma from Molecular Brain Neoplasia Data,"* Cogn. Syst. Res., vol. 58, pp. 173–194, 2019.

48. **T. Zarei and R. Behyad**, *"Predicting the water production of a solar seawater greenhouse desalination unit using multi-layer perceptron model,"* Sol. Energy, vol. 177, no. October 2018, pp. 595–603, 2019.

49. **S. Naeem, S. Shahhosseini, and A. Ghaemi**, *"Simulation of CO2 capture using sodium hydroxide solid sorbent in a fluidized bed reactor by a multi-layer perceptron neural network,"* J. Nat. Gas Sci. Eng., vol. 31, pp. 305–312, 2016.

50. **B. T. Pham, M. D. Nguyen, K. T. T. Bui, I. Prakash, K. Chapi, and D. T. Bui**, *"A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil,"* Catena, vol. 173, no. September 2018, pp. 302–311, 2019.

51. **Y. S. Kong, S. Abdullah, D. Schramm, M. Z. Omar, and S. M. Haris**, *"Optimization of spring fatigue life prediction model for vehicle ride using hybrid multi-layer perceptron artificial neural networks,"* Mech. Syst. Signal Process., vol. 122, pp. 597–621, 2019.

52. **J. Chaki, N. Dey, L. Moraru, and F. Shi**, *"Fragmented plant leaf recognition: Bag-of-features, fuzzy-color and edge-texture histogram descriptors with multi-layer perceptron,"* Optik (Stuttg)., vol. 181, no. December 2018, pp. 639–650, 2019.

53. **X. Fan and H. Shin**, *"Road vanishing point detection using weber adaptive local filter and salient-block-wise weighted soft voting,"* IET Comput. Vis., vol. 10, no. 6, pp. 503–512, 2016.

54. **L. N. Eeti and K. M. Buddhiraju**, "*A modified class-specific weighted soft voting for bagging ensemble,*" Int. Geosci. Remote Sens. Symp., vol. November, pp. 2622–2625, 2016.

55. **R. Susmaga**, *"Confusion Matrix Visualization,"* in Intelligent Information Processing and Web Mining, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 107–116.

56. **H. He and E. A. Garcia**, *"Learning from imbalanced data,"* IEEE Trans. Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284, 2009.

57. **N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer**, *"SMOTE: Synthetic Minority Over-sampling Technique,"* J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.

58. **G. Douzas, F. Bacao, and F. Last**, *"Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE,"* Inf. Sci. (Ny)., vol. 465, pp. 1–20, 2018.

59. **D. Elreedy and A. F. Atiya**, "*A Comprehensive Analysis of Synthetic Minority Oversampling TEchnique (SMOTE) for Handling Class Imbalance,"* Inf. Sci. (Ny)., vol. 505, pp. 32–64, 2019.

60. **S. Susan and A. Kumar**, "*SSO Maj -SMOTE-SSO Min : Three-step intelligent pruning of majority and minority samples for learning from imbalanced datasets,"* Appl. Soft Comput. J., vol. 78, pp. 141–149, 2019.

61. **M. Gao, X. Hong, S. Chen, and C. J. Harris**, *"A combined SMOTE and PSO based RBF classifier for two-class imbalanced problems,"* Neurocomputing, vol. 74, no. 17, pp. 3456–3466, 2011.

62. **J. Sun, H. Li, H. Fujita, B. Fu, and W. Ai**, *"Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting,"* Inf. Fusion, vol. 54, no. July 2019, pp. 128–144, 2019.

63. **B. Richhariya and M. Tanveer**, *"EEG signal classification using universum support vector machine,"* Expert Syst. Appl., vol. 106, pp. 169–182, 2018.

64. **M. M. Rahman, B. C. Desai, and P. Bhattacharya**, *"Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion,"* Comput. Med. Imaging Graph., vol. 32, no. 2, pp. 95–108, 2008.

65. **H. Heo, H. Park, N. Kim, and J. Lee**, *"Prediction of credit delinquents using locally transductive multi-layer perceptron,"* Neurocomputing, vol. 73, no. 1–3, pp. 169–175, 2009.

66. **X. Fan, L. Wang, and S. Li**, *"Predicting chaotic coal prices*

using a multi-layer perceptron network model," Resour. Policy, vol. 50, pp. 86–92, 2016.

67. **Y. Quan, Y. Xu, Y. Sun, and Y. Huang,** *"Supervised dictionary learning with multiple classifier integration,"* Pattern Recognit., vol. 55, pp. 247–260, 2016.

68. **V. Gholami, K. W. Chau, F. Fadaee, J. Torkaman, and A. Ghaffari,** *"Modeling of groundwater level fluctuations using dendrochronology in alluvial aquifers,"* J. Hydrol., vol. 529, no. March 2019, pp. 1060–1069, 2015.

69. **W. Chen et al.,** *"Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China,"* Sci. Total Environ., vol. 626, pp. 1121–1135, 2018.

70. **W. Chen et al.,** *"GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models,"* Sci. Total Environ., vol. 634, pp. 853–867, 2018.

71. **Hungarian Institute of Cardiology. Budapest, S. University Hospital, Zurich, S. University Hospital, Basel, and V.A. Medical Center, "***Heart Disease,"* kaggle, 2019. [Online]. Available: https://www.kaggle.com/ronitf/heart-disease-uci/version/1#_=_. [Accessed: 15-Nov-2019].

72. **University Medical Centre and Y. Institute of Oncology, Ljubljana,** *"Breast Cancer,"* OpenML, 2014. [Online]. Available: https://www.openml.org/d/13. [Accessed: 15-Nov-2019].

73. **sciki-learn,** *"Tuning the hyper-parameters of an estimator,"* 2019. [Online]. Available: https://scikit-learn.org/stable/modules/grid_search.html#grid-search. [Accessed: 15-Nov-2019].

74. **M. Sokolova and G. Lapalme,** *"A systematic analysis of performance measures for classification tasks,"* Inf. Process. Manag., vol. 45, no. 4, pp. 427–437, 2009.

75. **scikit-learn,** *"scikit-learn: Machine Learning in Python,"* 2019. [Online]. Available: https://scikit-learn.org/stable/. [Accessed: 04-Jul-2019].