- ORIGINAL ARTICLE -

Procedure to Improve the Accuracy of Dental Implant Failures by Data Science Techniques

Procedimiento para Mejorar la Precisión en el Acierto de los Fracasos en Implantes Dentales mediante Técnicas de Ciencia de Datos

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

[1] *Instituto de Materiales de Misiones (IMAM-CONICET), Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones. Posadas, Misiones, Argentina.*
{nancy.bea.ganz, a.e.ares}@gmail.com

[2] *Instituto de Investigación, Desarrollo e Innovación en Informática (IIDII), Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones. Posadas, Misiones, Argentina.*
hdkuna@gmail.com

Abstract

Nowadays, the prediction about dental implant failure is determined through clinical and radiological evaluation. For this reason, predictions are highly dependent on the Implantologists' experience. In addition, it is extremely crucial to detect in time if a dental implant is going to fail, due to time, cost, trauma to the patient, postoperative problems, among others. This paper proposes a procedure using multiple feature selection methods and classification algorithms to improve the accuracy of dental implant failures in the province of Misiones, Argentina, validated by human experts. The experimentation is performed with two data sets, a set of dental implants made for the case study and an artificially generated set. The proposed approach allows to know the most relevant features and improve the accuracy in the classification of the target class (dental implant failure), to avoid biasing the decision making based on the application and results of individual methods. The proposed approach achieves an accuracy of 79% of failures, while individual classifiers achieve a maximum of 72%.

Keywords: feature selection, classifier, ensemble, failure, dental implants.

Resumen

Hoy en día, la predicción del fracaso de un implante dental está determinado a través de una evaluación clínica y radiológica. Por esta razón, las predicciones dependen en gran medida de la experiencia del implantólogo. Además, es extremadamente crucial detectar a tiempo si un implante dental va a fallar, por cuestiones de tiempo, costo, traumas al paciente, problemas postoperatorios, entre otros. En este trabajo se propone un procedimiento mediante la utilización de múltiples métodos de selección de características y algoritmos de clasificación, para mejorar la precisión en el acierto de los fracasos en implantes dentales de la provincia de Misiones, Argentina validado por expertos humanos. 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. El procedimiento propuesto permitió conocer las características más relevantes y mejoró la precisión en la clasificación de la clase objetivo (fracaso del implante dental), permitiendo no sesgar la toma de decisión en base a la aplicación y resultados de método individuales. El procedimiento propuesto consigue una precisión del 79% de los fracasos, mientras que los clasificadores individuales alcanzan un máximo del 72%.

Palabras claves: selección de características, clasificación, integración, implantes dentales.

1. Introduction

Data science has made many advances in the development and application of techniques in several aspects of the health sector, such as in disease prediction, image classification, and decision support systems based on the analysis of related data sets, among many others. Despite the fact that there are works which use dental implant data sets to analyze the osseointegration process [1]–[5], they do not focus their attention on the biomaterial (implant surface treatment). This is the reason why this work jointly addresses the study of the characteristics of the implant itself, linked not only to the traits and health patients' conditions but also to those of the surgical processes.

It is extremely important to detect in time if an implant will fail not only because of time, but also because of costs, patient trauma, such as pain, medication that can interfere with bone marrow processes, among others. In addition, postoperative problems such as infections, loss of bone quality, lack of sensitivity due to anesthesia, pain, or damage to a nerve area. For these reasons, it is essential to detect in time if a surgical procedure for dental implant placement is unsuccessful [1], [4].

This paper presents an approach using multiple feature selection methods and classification algorithms with the aim of improving the classification efficiency of failures of a data set based on patients' clinical histories who have undergone surgical processes of dental implant placement in the province of Misiones, Argentina. This approach is also validated by human experts in this specific area, such as implantologists.

A combination of several feature selection methods is used in order to find the most relevant subset of characteristics by evaluating the quality of classification accuracy and the number of features selected on each method. The experimentation consisted in obtaining the values of importance of each characteristic according to the integration of the of feature selection methods: Information Gain [6], Gain Ratio [7], Random Forest importance [8], Relief [9] and Chi Squared [10]. The steps performed were basically three: the generation of the features subsets, the obtainment of the performance measures and the evaluation of those measures to contrast them with the proposed procedure.

Once the most important subset of features is found, several classification algorithms are applied to improve the accuracy of the label of interest (dental implant failure). The classifiers are used in: Random Forest [8], C-Support Vector [11], K-Nearest Neighbors [12], Multinomial Naive Bayes [13] and Multi-layer Perceptron [14]. The assemble consisted of applying weights to the classifiers and averaging their predictions.

The contribution of this work is an automatic learning approach for both, the selection of the most important features and the prediction of failures in dental implants. Furthermore, we demonstrate that multi-method systems can also be applied to the case study since they allow better performances than those achieved individually.

This section displays our inner motivations. The next parts of the paper are structured as follows: Section 2 presents works related to the prediction of dental implants, the application of feature selection methods and multiple classifier systems; section 3 thoroughly describes the approach proposed in this work; section 4 shows the experimental results obtained, and section 5 summarizes the main conclusions drawn from this work and outlines future research lines.

2. Related work

The section 2 presents works related to the prediction of dental implants (section 2.1), the application of feature selection methods (section 2.2), and multiple classifier systems (section 2.3).

2.1. Background on prediction of Dental Implants

Tamez *et al.* [1] show a statistical analysis to determine the factors that influence the success of dental implants placed in the Postgraduate Program of Prosthodontics and Implantology at the Universidad De La Salle Bajío, Mexico. Domínguez *et al.* [2] carried out a study to determine if there is a relationship between dental implant failures and systemic diseases (specifically osteoporosis, hypertension, diabetes and hypothyroidism) in a population of patients undergoing dental implant surgery at the San José hospital in Santiago, Chile. In the work of Oliveira *et al.* [3] they present a comparative analysis of three machine learning techniques: SVM [15], weighted SVM and a neural network [14] with selection parameter, for the prediction of the success of dental implants. The characteristics considered in that work were patient age, sex, implant type, implant position, surgical technique, indication of whether the patient was a smoker or not, and indication of whether the patient had a previous disease (diabetes or osteoporosis) or medical treatment (radiotherapy). The data set used consists of 157 cases, registered by a single surgeon of the Faculdade de Odontologia, Universidade Federal do Rio Grande do Sul, Brazil. Another work with the same characteristics is that of Moayeri *et al.* [4], where they present a combined predictive model to evaluate the success of dental implants. The classifiers used in this model are a decision tree (J48) [16], an SVM, a neural network, a k nearest neighbor and Naive Bayes [17]. To evaluate the effectiveness of the proposed algorithms, they use 224 cases of patients who had dental implants placed. This dataset belongs to the Dental School of Tehran University, Iran and contains different variables, which are gender, age, smoking, implant location, placement time, loading protocol, implant diameter and length, implant connection type, overdenture, and maxillary sinus elevation. In the work of Braga *et al.* [5] they propose a set of binary logistic models to evaluate the probability of success or non-success in the oral rehabilitation process, taking into account some genetic factors, individual habits and clinical and non-clinical factors. The study was conducted in a retrospective evaluation and consisted of 155 subjects undergoing oral rehabilitation in the northern region of Portugal.

Although these works use data sets of dental implants, they do not focus their attention to the biomaterial (implant surface treatment), which is why this work jointly addresses the study of the characteristics of the implant itself, linked to the traits and health conditions of the patients, as well as to the conditions of the surgical process, to improve the accuracy of dental implant failures.

2.2. Background in the use of feature selection methods

Many researchers have evaluated and compared different methods of feature selection to choose, classify and eliminate irrelevant features with the purpose of improving the results at the classification stage. Some previous works on this topic are going to be presented and discussed.

Several methods of feature selection for text classification were evaluated by Kou *et al.* [18]. Nonetheless, Information Gain, Gain Ratio, Gini Index [19], Chi-Squared, Mutual Information [20], among others, were chosen because of their variations in performance. Chaudhary *et al.* [21], on the other hand, evaluated the performance of two feature selection methods: Gain Ratio and Information Gain, with Naive Bayes [17], which were optimized in a mobile device and based on the following performance measures: accuracy, true positive rate and recall [22]. They concluded that Gain Ratio had a comparatively better performance than the other method.

Gao *et al.* [23] used the SVM classifier supported by the Information Gain method to filter out irrelevant and redundant genes. Subsequently, they evaluated five data sets of cancer gene expression and selected a few genes. The chosen genes served as the basis for the classifier. The results demonstrated that, in comparison with other feature selection methods, the suggested combination achieved the best classification accuracy.

Phyu y Oo [24] proposed a feature selection algorithm based on the perspective of conditional Mutual Information [20]. These authors evaluated the effectiveness of the suggested algorithm by comparing it with other feature selection algorithms such as Information Gain, Symmetrical Uncertainty and Relief and used standard data sets from UC Iravine and Weka. After that, they evaluated the performance of the proposed algorithm by the classification accuracy of the Naive Bayes and J48 [16] classifiers and by the number of features selected. The authors concluded that, although some algorithms may further reduce the number of features, their accuracy in classification was not very good. In addition, they claimed that their algorithm selected as few features as possible and is more accurate in classifying several of the data sets used.

Finally, Peker *et al.* [25] used Minimum Redundancy Maximum Relevance [17] and Relief to select the feature set, using Random Forest, C4.5, SVM, Naive Bayes and two types of neural network. The authors found that the best results were achieved when using the subset of features obtained from the Relief algorithm with the Random Forest classifier.

The approach of these works allows to appreciate the vision of comparison and combination of several feature selection methods, which are based on different criteria, not only to improve classification accuracy, but also to ensure that the selected features are those that provide the greatest information gain to the problem.

2.3. Background in the integration of classifiers

In decision making, the combination of classification models can be crucial because such a combination is aimed at obtaining an appropriate solution to a particular problem. Particularly, classification methods are based on different concepts or estimation procedures. It is logical to try to bring together the best properties of each one by combining them in some way.

Several studies have evaluated the combination or integration of classifiers to improve the success rate or even to avoid biasing the decision on the results of a single classifier [26].

Miao *et al.* [27], improve the accuracy of genes identification by integration of SVM and Random Forest classifiers, applying Relief to select the most relevant characteristics. After training and predicting, the results were combined using a majority voting method [26]. The integration of the probabilities made it possible to obtain a higher accuracy than with individual classifiers. Similarly, Catal and Nengir [28] presented a model for the classification of feelings by integrating the Naive Bayes and SVM classifiers. For the integration of several predictions, the authors used the majority voting method and demonstrated that multiple rating systems improve accuracy. Another work with similar characteristics is that of Pandey and Taruna [29] who propose an integrated classifier by using a J48 and a KNN on a data set of academic performance among engineering students. In this model, each individual classifier generates its predictive value and these are integrated where the final class label is represented by the maximum of a subsequent probability. Ruano-Ordás *et al.* [30] propose a model to automatically determine the biological activity of molecules based on 2048 chemical substructures and 84 physicochemical properties. The authors performed the process in three stages: grouping of features, construction and optimization of hyper parameters of each classifier, and classification. They also used SVM with radial kernel, AdaBag [31] and rpart [32], and combined the individual results of each classification into a single result by using the majority voting method. In addition, Nweke *et al.* [33] presented a survey of the use of multiple classifier systems in the recognition

of human activity and health monitoring. These authors have also tried 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 and C4.5), KNN, Artificial Neural Network, Naive Bayes and Random Forest.
Inspired by the above ideas, the use of multiple classifiers is proposed for the case study.

3. Materials and methods

The section 3 thoroughly describes the approach proposed in this work (section 3.1), and the characteristics of the data sets used for the experimental evaluation are detailed (section 3.2).

3.1. Proposed approach

Figure 1 summarizes the steps of the proposed mechanism to select the most relevant features and improve the hit to failure of the *Dental Implant* data set.
The present procedure extends from a previous work [34], which does not contemplate the exhaustive search for the best features for the case study.

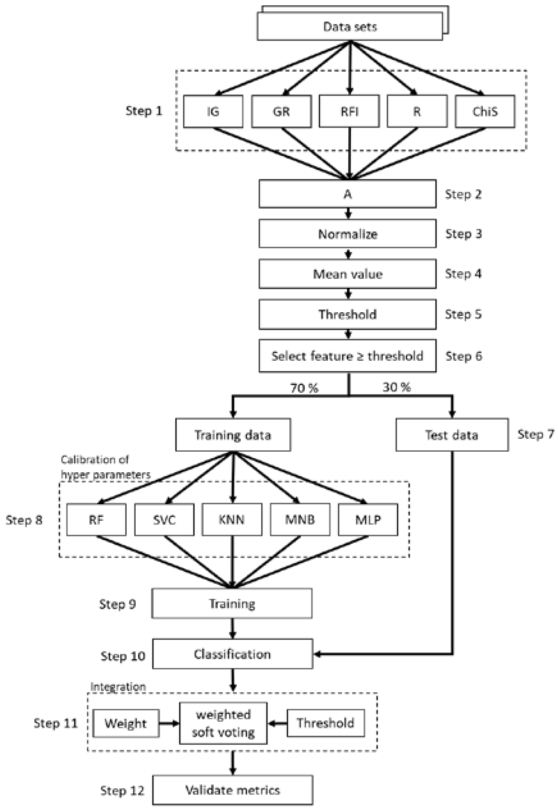

Fig. 1. Proposed approach. This representation summarizes the steps of the mechanism proposed in this work: to select the most relevant features through the Information Gain (IG), Gain Ratio (GR), Random Forest importance (RFI), Relief (R) and Chi Squared (ChiS) methods, as well as to improve the failure accuracy of the *Dental Implant* data set by integrating the predictions of Random Forest (RF), C-Support Vector (SVC), K-Nearest Neighbors (KNN), Multinomial Naive Bayes (MNB) and Multi-layer Perceptron (MLP) classifiers.

Each of the steps and specifications covered by the procedure outlined in Figure 1 are detailed below.
First read the data set and select the class or target feature for the prediction (class variable defining the successes and failures of dental implants).

Step 1. Obtain the feature subsets from feature selection methods: Information Gain (IG), Gain Ratio (GR), Random Forest importance (RFI), Relief (R) and Chi-Squared (ChiS).

Step 2. Make a matrix (A) that gathers the importance value obtained for each feature by the different methods. In other words, there will be five different possible values of importance for the same feature.

Step 3. Normalize the values. Due to the fact that the methods used are performed with different ranges this fundamental step is necessary (so as) to achieve an average value for each feature. For this purpose, the "*normalize"* function (Eq. (1)), which allows normalizing values on the basis of the minimum-maximum method. Minimal-maximum normalization regulates the features in a range [35] where min_A and max_A are the minimum and maximum values for feature A respectively. The minimum-maximum normalization maps a v_i value of A to v_i' in the $[new\_min_A, new\_max_A]$ range by means of:

$$v_i' = \frac{v_i - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A \quad \text{(Eq. 1)}$$

This standardization criterion was used because it allows preserving all the relationships of the original value data, i.e. it does not introduce any potential bias. In addition, it has shown to perform better in classification. The range used was [0,1].

Step 4. Obtain a mean value by each feature according to the values obtained by the different methods. The median [35] was used as a central tendency measure because the values of importance, which were given by the different methods, did not follow a normal distribution. If these values followed a normal distribution, the mean would be applied [35].

Step 5. Obtain a threshold. This threshold was determined by a grid search, using a test parameter

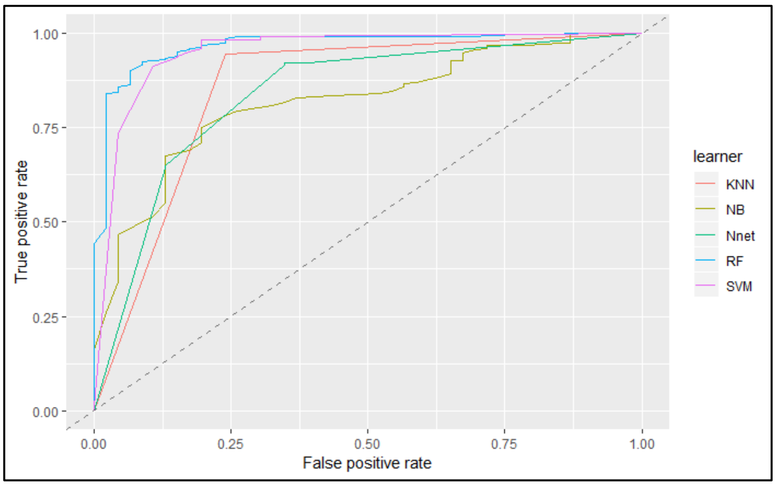

Fig. 2. ROC Curve. Performance of K-Nearest Neighbors (KNN), Naive Bayes (NB), Neural network (Nnet), Random Forest (RF) and Support Vector Machine (SVM) classifiers on the *Dental Implants* data set.

with values between 0.1 and 1 and with 0.1 increments in each test. This search was subjected to a ten-interaction cross validation. This test is performed on the values obtained from step 4. The chosen threshold value was the one that allowed obtaining the best accuracy in the classification with Random Forest.

A Random Forest classifier was used to search for thresholds, as it provided the best performance compared with other classifiers. To determine the performance, several types of classifiers were examined (with calibration) and the area under the curve (auc) was calculated. This is reflected by the ROC curve in Figure 2, which summarizes the performance of the KNN (auc 0.85), NB (auc 0.82), Nnet (auc 0.85), Random Forest (auc 0.97) and SVM (auc 0.95) classifiers on the *Dental Implant* data set (without feature selection). This behavior is equivalently repeated for the data set Artificial used in the experimental validation.

Step 6. Select the features that meet the condition of being equal to or greater than the threshold obtained in step 5.

Step 7. To perform a classification task, after the selection of the most important features of a set, it is necessary to split the data. For the case study, the data was randomly divided to preserve the distribution of both classes into: 70 % for training and 30 % for evaluation [36]–[39], ensuring that all cases are represented in both sets.

Step 8. An important step in any classification task is the search for the best individual classifiers for the case study. After examining the types of classifiers used in the articles surveyed in section *2.3 Background in the integration of classifiers*, we propose the use of the following classifiers: Random Forest (RF), C-Support Vector (SVC), K Nearest Neighbors (KNN), Multinomial Naive Bayes (MNB) and Multi-layer Perceptron (MLP). More than one classifier is used in order not to decline the decision based on the results of only one.

To obtain a robust model and optimize the results of the classifiers, a grid search was carried out to adjust the hyper parameters [36], [39], [40]. This search was performed with the training data from each of the data sets. For this process, we specified:

1. A <u>search space</u>, defining ranges of values for each of the hyper parameters and adjusting them according to the selected performance measure.
2. An <u>optimization or adjustment algorithm</u>, the grid search method [41] was used. Although it is the most costly in terms of performance, it allows covering the entire defined search space.
3. An <u>evaluation method</u>, as a resampling strategy a cross validation of 10 iterations was used.
4. A <u>measure of performance</u>, 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 [22].

The parameters were established based on those used in the literature surveyed in section 2.3 Background in the integration of classifiers and taking into account the parameterization recommended by sciki-learn for the adjustment of an estimator [41]. In conjunction, to choose the values of the search spaces we used the database of the mlrHyperopt tool [42], this tool includes default search space for the most common machine learning methods, it also offers a web server for sharing, uploading and downloading improved search space.

Table 1 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.

Table 1. Hyper parameters and search ranges defined for the Random Forest (RF), C-Support Vector (SVC), K-Nearest Neighbors (KNN), Multinomial Naive Bayes (MNB) and Multi-layer Perceptron (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*

Step 9. Perform the training of each classifier with the optimal values found for each hyper parameter in step 8. Classifier coefficients are estimated with the training set.

Step 10. Perform prediction with the test set data.

Step 11. Integrating predictions. After examining and evaluating the different techniques used in the works cited in section *2.3 Background in the integration of classifiers*, the weighted soft voting method is applied [43], [44] in order to integrate the results of several classifiers and determine the final class label.
Therefore, the integration of the predictions consisted in multiplying, for each tuple, the probability value of the 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 [45] of each classifier was measured, and the value of w that achieved the best accuracy was selected [30], [33], [46].
Once the weights were determined, the weighted soft voting method was applied [43], [44]. 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 (Eq. (2)), given by:

$$\hat{y} = \arg max_i \sum_{j=1}^{m} w_j p_{ij} \quad \text{(Eq. 2)}$$

The place where p_{ij} indicates the probability predicted by the j-th classifier and w_j 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 [46], [47], because, in exploratory evaluations, it allowed us to achieve better classification results. 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 on each test. The μ value was selected as the one that allowed obtaining the best classification result for all the data sets used.
This step is performed with the predictions achieved with the test set.

Step 12. Validate the accuracy and error of each classifier with those found in the integration of step 11. In addition, analyze in particular the Sensitivity and Specificity metrics to corroborate whether the integration of the probabilities of the individual classifiers improves the accuracy of the diagonal of the confusion matrix. Taking into account that the aim is to optimize the accuracy of the minority class (Sensitivity), without neglecting the accuracy of the majority class (Specificity). Together, the error rate in the accuracy of both class labels is evaluated.

The implementation of the procedure up to and including step 6, which covers feature selection, is done with the R tool together with the mlr package[1]. The second part corresponding to classification (from step 7 to step 12) is done with the Pyhton tool and the scikit-learn library[2]. The source code of the complete procedure is hosted in a GitHub repository [3].

3.2. 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) [48] based on actual dental implant cases. Table 2 presents the summarized characteristics of these sets.

[1]mlr. Available in https://mlr.mlr-org.com/. (Consulted the 17/06/2021).
[2]Scikit-learn. Available in https://scikit-learn.org/stable/. (Consulted the 17/06/2021).
[3]Source code and data sets. Available in https://github.com/nancyganz/Tesis. (Consulted the 17/06/2021).

Table 2. 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 features selected by the proposed feature selection procedure and size of the training and test sets.

Characteristics	*Dental Implants*[1]	*Artificial*[2]
Sample	1165	1748
Feature	33	33
Selected features	16	20
Training	815	1223
Test	350	525

[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 3 presents the characteristics of the *Dental Implants* and *Artificial* data sets in more detail.

4. Results

Table 4 lists the features selected by the proposed approach for the *Dental Implants* and *Artificial dataset*.

Table 5 presents the rates obtained by each classifier and that obtained from the proposed approach on the test data of the used data sets. In this table, it is also observed the SVC and KNN classifiers achieved the best performance over the non-target class for the two data sets compared to the other classifiers, and even exceed the approach proposed. For the target class, it is appreciated that the integration of the predictions of the five classifiers allowed to achieve the highest percentage of correctness. 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 accuracy.

Finally, the results achieved with the proposed approach on the *Dental Implants* data set were compared with the accuracy achieved in the classification by human experts (Table 6). These were selected from the Provincial Registry of Professionals who practice Maxillofacial Buco Surgery, Implantology, Periodontics and Tissue Manipulation.

The evaluation consisted of a ranking by four experts in the area, each of whom was provided with a random sample distinct from the 10% prevalence of cases. The cases were presented without label so that the experts could classify them according to

Table 3. 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, occupation, social security, medical 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, hard tissues regeneration, soft tissues regeneration, 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**.

Table 4. Features selected by the proposed approach for the *Dental Implant* and *Artificial* datasets.

Data sets	Features
Dental Implants	Med intake, occupation, medical antecedent, surface treatment, dental piece, bone type, length, season, placement time, age range, diameter, connection, periodontitis, surgical complication, register, and soft tissues regeneration.
Artificial	Occupation, surface treatment, med intake, bone type, medical antecedent, dental piece, season, placement time, length, age range, periodontitis, diameter, gender, connection, soft tissues regeneration, load protocol, surgical complication, bone expansion, additional procedure, and register.

Table 5. Efficiency in performance of the Random Forest (RF), C-Support Vector (SVC), K-Nearest Neighbors (KNN), Multinomial Naive Bayes (MNB) and Multi-layer Perceptron (MLP) classifiers and the proposed approach (PA) to the *Dental Implants* and *Artificial* data sets.

Data sets	Classifiers	Sensitivity	Specificity	Accuracy	Error
Dental Implants	RF	59 %	98 %	92%	8%
	SVC	64 %	99 %	93%	7%
	KNN	64 %	99 %	93%	7%
	MNB	72 %	79 %	78%	22%
	MLP	66 %	97 %	92%	8%
	PA	79 %	96 %	93%	7%
Artificial	RF	81 %	97 %	95%	5%
	SVC	81 %	99 %	96%	4%
	KNN	81 %	99 %	96%	4%
	MNB	60 %	81 %	78%	22%
	MLP	82 %	97 %	95%	5%
	PA	90 %	97 %	96%	4%

their experience, and make a contrast with the values found by our classification approach.

Our model achieved 93% overall accuracy with 7% error on average, whereas, the classification made by the experts achieved a total accuracy of 87%, with an average error of 13% (Table 6).

Table 6. Comparison of the evaluation parameters achieved by the proposed approach (PA) and the classification of the human experts (Experts) on the *Dental Implants* data set.

Metrics	PA	Experts
Sensitivity	79%	71%
Specificity	96%	92%
Accuracy	93%	87%
Error	7%	13%

5. Conclusions and future work

This work has allowed the study and application of multiple feature selection methods and classification algorithms to a domain of little knowledge.

According to the experimental results, the multi-method approach can also be applied to the prediction of dental implant failures.

Based on the classification results by human experts, we can say that our approach has enabled us to achieve superior classification performance. Therefore, we have succeeded in proposing a knowledge extraction procedure validated by human experts.

Finally, it is proposed as a future work to develop and implement a decision support system using Fuzzy Logic, in order to provide implantologists specialists with an new alternative when analyzing irregular or complex situations. Based on certain specifications, it may also allow the specialists to model and analyze in advance, what could be the postoperative result of the surgical intervention of

the dental implant. At the same time, this work paves the way to evaluate the benefit and social impact on dental specialists in the Northeast of Argentina by providing with a virtual assistant to help them evaluate and determine the specific conditions of the patient and the most appropriate technique to use in each particular case.
We propose as future work to include an ablation study comparing our approach to the study of no feature selection and individual feature selection algorithms, similar to the approach taken in this paper to compare individual models versus a weighted ensemble. Also, add to the work information about the total execution time, including the tuning time of each algorithm.
Thus, we propose as future work to contrast the results found through some statistical test (such as Friedman's hypothesis test or a Benchmarking to compare the performance of classifiers).

Competing interests

The authors have declared that no competing interests exist.

Authors' contribution

All authors contributed to the development of the work, read and approved the final manuscript.

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 for this research.

References

[1] J. E. B. Tamez, F. N. Zilli, L. A. Fandiño, and J. M. Guizar, "Factores relacionados con el éxito o el fracaso de los implantes dentales colocados en la especialidad de Prostodoncia e Implantología en la Universidad de La Salle Bajío," *Revi. Esp. Cirugía Oral y Maxilofac.*, vol. 286, pp. 1–9, 2016.

[2] J. Domínguez, J. Acuña, M. Rojas, J. Bahamondes, and S. Matus, "Study of association between systemic diseases and dental implant failure," *Rev. Clínica Periodoncia, Implantol. y Rehabil. Oral*, vol. 6, no. 1, pp. 9–13, 2013.

[3] A. L. I. Oliveira, C. Baldisserotto, and J. Baldisserotto, "A comparative study on machine learning techniques for prediction of success of dental implants," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3789 LNAI, pp. 939–948, 2005.

[4] R. S. Moayeri, M. Khalili, and M. Nazari, "A Hybrid Method to Predict Success of Dental Implants," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 5, pp. 1–6, 2016.

[5] A. C. Braga, P. Vaz, J. C. Sampaio-Fernandes, A. Felino, and M. P. Tavares, "Decision model to predict the implant success," *Proc. 12th Int. Conf. Comput. Sci. Its Appl.*, vol. 7333 LNCS, no. PART 1, pp. 665–674, 2012.

[6] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[7] J. R. Quinlan, "Induction of Decision Trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.

[8] L. Breiman, "Random Forest," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[9] K. Kira and L. A. Rendell, "A Practical Approach to Feature Selection," *Mach. Learn. Proc.*, pp. 249–256, 1992.

[10] 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.

[11] 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.

[12] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *Am. Stat.*, vol. 46, no. 3, pp. 175–185, 1992.

[13] 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.

[14] B. Irie and Sei Miyake, "Capabilities of Three-layered Perceptrons," *IEEE nternational Conf. Neural Networks*, pp. 641–648, 1988.

[15] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[16] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: A review," *Procedia Manuf.*, vol. 35, pp. 698–703, 2019.

[17] M. Togaçar, B. Ergen, and Z. Cömert, "Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks," *Biocybern. Biomed. Eng.*, vol. 40, no. 1, pp. 23–39, 2020.

[18] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi, "Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods," *Appl. Soft Comput. J.*, vol. 86, p. 105836, 2020.

[19] G. Louppe, L. Wehenkel, A. Sutera, and P.

Geurts, "Understanding variable importances in forests of randomized trees," *Adv. Neural Inf. Process. Syst. 26*, pp. 431–439, 2013.

[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] A. Chaudhary, S. Kolhe, and Rajkamal, "Performance Evaluation of feature selection methods for Mobile devices," *J. Eng. Res. Appl.*, vol. 3, no. 6, pp. 587–594, 2013.

[22] 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.

[23] L. Gao, M. Ye, X. Lu, and D. Huang, "Hybrid Method Based on Information Gain and Support Vector Machine for Gene Selection in Cancer Classification," *Genomics, Proteomics Bioinforma.*, vol. 15, no. 6, pp. 389–395, 2017.

[24] T. Z. Phyu and N. N. Oo, "Performance Comparison of Feature Selection Methods," *MATEC Web Conf.*, vol. 42, p. 06002, 2016.

[25] M. Peker, A. Arslan, B. Sen, F. V. Celebi, and A. But, "A novel hybrid method for determining the depth of anesthesia level: Combining ReliefF feature selection and random forest algorithm (ReliefF+RF)," *Int. Symp. Innov. Intell. Syst. Appl.*, 2015.

[26] 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.

[27] 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.

[28] C. Catal and M. Nangir, "A sentiment classification model based on multiple classifiers," *Appl. Soft Comput. J.*, vol. 50, pp. 135–141, 2017.

[29] 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.

[30] 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.

[31] E. Alfaro, M. Gamez, and N. García, "adabag : An R Package for Classification with Boosting and Bagging," *J. Stat. Softw.*, vol. 54, no. 2, 2013.

[32] T. Therneau, B. Atkinson, and B. Ripley, "rpart: Recursive Partitioning and Regression Trees," *R Packag. version*, 2019.

[33] 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.

[34] N. B. Ganz, A. E. Ares, and H. D. Kuna, "Predicting dental implant failures by integrating multiple classifiers," *Rev. Cienc. y Tecnol.*, vol. 34, pp. 13–23, 2020.

[35] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2012.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

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

[41] sciki-learn, "Tuning the hyper-parameters of an estimator," 2019. [Online]. Available: https://scikit-learn.org/stable/modules/grid_search.html#grid-search. [Accessed: 30-Jul-2020].

[42] Jakob Richter, "mlrHyperopt," 2019. [Online]. Available: https://jakob-r.de/mlrHyperopt/index.html. [Accessed: 29-Jun-2021].

[43] 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.

[44] 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.

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

[46] 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.

[47] M. Mohandes, M. Deriche, and S. O. Aliyu, "Classifiers Combination Techniques: A Comprehensive Review," *IEEE Access*, vol. 6, pp. 19626–19639, 2018.

[48] 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.

Citation: N.B. Ganz, A.E. Ares and H.D. Kuna. *Procedure to Improve the Accuracy of Dental Implant Failures by Data Science Techniques.* Journal of Computer Science & Technology, vol. 21, no. 2, pp. 146-156, 2021.
DOI: 10.24215/16666038.21.e13
Received: April 22, 2021 **Accepted:** August 19, 2021.