UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Estudio de variables que Influyen en la aparición de periimplantitis

Trabajo Fin de Grado

Grado en Ciencia de Datos

AUTOR/A: Arteaga del Olmo, Ángela Huacong

Tutor/a: Conejero Casares, José Alberto

CURSO ACADÉMICO: 2023/2024

Resumen

Uno de los retos significativos a los que se enfrenta la odontología en la gestión de la salud bucodental, es la prevención y tratamiento de la periimplantitis, una complicación común en pacientes con implantes dentales. Este trabajo se enfoca en explorar y analizar las variables que influyen en la aparición de esta patología, con el objetivo de mejorar la comprensión y su manejo. Se llevará a cabo un análisis multivariante para evaluar la variabilidad de los datos recopilados de registros médicos y datos de seguimiento de pacientes. La integración de estos datos permitirá identificar patrones, correlaciones y factores de riesgo asociados con la periimplantitis. Se emplearán técnicas estadísticas y de análisis de datos para estudiar la relación entre dichas variables con el fin de contribuir a mejorar sus estrategias de prevención y tratamiento.

Keywords: Periimplantitis, análisis multivariante, implantes dentales.

Abstract

One of the significant challenges facing dentistry in managing oral health is the prevention and treatment of periimplantitis, a common complication in patients with dental implants. This work focuses on exploring and analyzing the variables that influence the onset of this condition, with the aim of enhancing understanding and management. A multivariate analysis will be conducted to assess the variability of data collected from medical records and patient follow-up data. The integration of these data will allow the identification of patterns, correlations, and risk factors associated with periimplantitis. Statistical and data analysis techniques will be employed to study the relationship between these variables in order to contribute to improving its prevention and treatment strategies.

Keywords: Periimplantitis, multisource analysis, dental implants.

Resum

Un dels reptes significatius als quals s'enfronta l'odontologia en la gestió de la salut bucodental és la prevenció i tractament de la periimplantitis, una complicació comuna en pacients amb implants dentals. Aquest treball es centra en explorar i analitzar les variables que influeixen en l'aparició d'aquesta patologia, amb l'objectiu de millorar la comprensió i el maneig. Es durà a terme un anàlisi multivariable per avaluar la variabilitat de les dades recopilades de registres mèdics i dades de seguiment de pacients. La integració d'aquestes dades permetrà identificar patrons, correlacions i factors de risc associats amb la periimplantitis. Es faran servir tècniques estadístiques i d'anàlisi de dades per estudiar la relació entre aquestes variables i així contribuir a millorar les seves estratègies de prevenció i tractament.

Keywords: Periimplantitis, anàlisi multisource, implants dentals.

Table of contents

Table of Figures

Introduction

In Spain, the average of teeth in young adults (35-44 years) is of 26 teeth and in older adults (65-74) is of 17 teeth, when in fact, the total number of teeth an adult could have is 32. As a consequence, each year between 1.2 and 1.4 million of implants are performed due to teeth loss (Consejo dentistas, 2021). Dental implants appear as a recurrent and effective solution.

Performing this surgery can lead to an infection which could cause periodontal disease provoquing several teeth health consequences, such as recurrent gum swelling, inflamation or bleeding, and in the most advanced stage bone loss. This condition is the primary indicator for peri-implantitis.

Peri-implantitis is a plaque-associated pathological condition occurring in tissues around dental implants, characterized by inflammation in the peri-implant mucosa and subsequent progressive los of supporting bone. (Schwartz & cols, 2017).

The main clinical characteristic of peri-implant mucositis is bleeding on gentle probing. Erythema, swelling, and/or suppuration may also be present. An increase in probing depth is often observed in the presence of peri-implant mucositis due to swelling or decrease in probing resistance. There is strong evidence from animal and human experimental studies that plaque is the etiological factor for peri-implant mucositis.(Schwarz et al,. 2017)

Consequently, dentrists are focusing on studying its treatment and causes due to its increasing prevalence and limited treatment efficacy. (Zhu & cols Y, 2024)

In clinical terms we asume there is Peri-implantitis by: (1) the presence of peri-implant signs of inflammation,(2) radiographic evidence of bone loss following initial healing, and (3) increasing probing depth as compared to probing depth values collected after placement of the prosthetic reconstruction. (Renvert S & cols, 2018)

Preventative measures are crucial in managing peri-implantitis. Good oral hygiene practices, regular dental check-ups, and professional cleanings are essential in preventing plaque buildup around dental implants. Recent advancements in treatment include laser therapy, the use of antimicrobial agents, and bone regeneration techniques. However, the success of these treatments varies, and research continues to seek more effective solutions.

The impact of peri-implantitis extends beyond dental health. It can affect the overall well-being of patients, leading to discomfort, reduced quality of life, and increased healthcare costs.

Understanding the risk factors, such as smoking, diabetes, and poor oral hygiene, is vital for both prevention and management.

In summary, while dental implants are a highly effective solution for tooth loss, the complications such as peri-implantitis create the need of further research and improved treatment strategies. Preventative care and early intervention remain key in ensuring the long-term success of dental implants.

Objectives

This project has three main goals:

1. Periimplantitis etiology is due to multiple factor. Nevertheless, research on this topic is still discovering new important features and triggering factors. Therefore, a multivariate analysis will be done focusing on the features that can influence on the origin and development of this pathology.
2. Identifying relevant variables for the odontologist that provided their data, which contains information about their type of implants.
3. Forecasting the onset of this illness by this studied features with different machine learning models, being able to compare their performance and predictive results.

Motivation

The decision to focus my thesis on periimplantitis emanates not only from an intellectual curiosity but also from a pragmatic recognition of the need for empirical insights within this domain. As I navigated through the landscape of dental literature, it became evident that periimplantitis represents a critical gap in our understanding of oral health. It also allows me to deepen in the field of medical data analytics.

Additionally, the possibility of being able to contribute to the prevention or prediction of a disease seems very apealing to me.

Moreover, the prospect of working at the intersection of medicine and data science really appeals my curiosity and profesional development, which may give me also a view about future long-term professional aspirations.

Value of the Project

This project aims to contribute to dentistry and society, particularly in addressing the common complication of peri-implantitis following dental implant procedures. It aims to deepen our understanding of the factors contributing to this condition, enabling early detection and personalized interventions. Especially, it aims to be useful for the Odontologists of the Universidad de Barcelona, as it contains some features related to their implant surgerys such as the implant type. Then, we aim to provide also specific information that can lead to better results for them. This can also be useful for this topic as most of them obtain different conclusions based on their samples and different focus.

By using machine learning techniques, the project endeavors to give more information about preventive dentistry by developing predictive models for identifying individuals at risk of peri-implantitis. This proactive approach has the potential to reduce some dental complications and enhance long-term implant success rates. Moreover, the project's interdisciplinary nature offers professional growth and innovation in both dental research and clinical practice, areas that I haven’t worked on before.

The Project will:

- Identify variables that allow us to predict peri-implantitis, contrasting this information with previous studies to validate information.
- Compare the performance of various machine learning models to determine the most effective predictive tools.
- Leverage the collaboration with the Universidad de Barcelona, which will provide the majority of the information and data, ensuring that the findings are robust and relevant to real-world clinical settings.

By integrating machine learning with clinical expertise, this project aims to contribute to the statements and findings in the management and research of peri-implantitis, with the purpose of contributing to improved patient outcomes and advancing this particular field of dentistry.

State of the art

Another study considered a similar approach to the subject very recently. This aimed to predict peri-implantitis by using machine learning techniques and clarify interactions between risk indicators. This study differs on the variables used and it perform a Chi-squared test. Nonetheless, the rest of the models will be compared to the ones they present. (Mameno et al., 2021)

It is cientifically evidenced that Diabetes and Periodontitis are related. There are studies that show after a treatment of periodontal disease an improvement in both clinical and immunological parameters of periodontitis and glycemic control in long-term diabetes (Muñoz-Corcuera et al, 2015)

According to Schwarz's group study, there is strong evidence that there is an increased risk of developing peri-implantitis in patients who have a history of severe periodontitis, poor plaque control, and no regular maintenance care after implant therapy.

Also, it came out that smoking and diabetes, which were potential risk indicators for peri-implantitis, are inconclusive. (Schwarz et al,, 2017)

Although a year later another study reported a medium-high level of evidence that smoking, diabetes and history or presence of periodontitis were identified as risk factors of peri-implantitis. As well as in a medium-high evidence that a patient's age, gender and maxillary implants are not related to peri-implantitis. (Dreyer, 2018)

Colliding also with another study's conclusions which found a statistically significant relationship between age and sex with the prevalence of peri-implantitis, communicating that as younger and male patients they tend to have a higher prevalence of peri-implantitis (Zegarra et al., 2018).

Also another group studied different factors that could predict periimplantitis concluding these variables could affect it: number of teeth, age, gender, periodontitis severity and years of implant service. (Zhu Y et al., 2024) Except the number of teeth and implant functionality time the rest variables will aso be studied in this Project.

A more specific study, evidenced that local factors such as accessibility for oral hygiene at the implant sites seems to be related to the presence or absence of peri-implantitis. (Serino et al., 2009)

Moreover, other features have been documented but with few evidence this are: implant position or Keratinized width tissue, which is often regarded to be ≥ 2mm. (Ravidà et al,, 2022).

Mantaining bone around dental implants is an important matter due to the direct relation of marginal bone loss around the implant to delayed implant failures. So different variables that could have biomechanical effects are important because of the contact between the bone and the implant that transfer occlousal loads directly to the bone. (Aparecido et al., 2021)

Less studies have reported a Marginal Bone loss in Inernal Conexion implants. For instance, one concluded that internal connection implants showed lower stress in the cortical bone only in models without bone loss, while external connection implants exhibited higher stress in the implants and screws under oblique loading. (Aparecido, 2021)

Another study related, concluded that IC implants showed a more favorable bone response regarding MBL in posterior areas without peri-implantitis or periodontal disease (Ju H etal., 20218). In our study we will analyse which conexion may be related to have Peri-implantitis.

The peri-implant keratinized mucosa is tightly attached to the underlying bone, serving as a functional barrier between the oral environment and dental implants. However, following tooth extraction, resorption of the surrounding bone and keratinized gingiva often occurs, potentially leading to a deficiency of keratinized mucosa during subsequent implant placement. Consequently, the presence of keratinized tissue around dental implants is increasingly recognized as essential for maintaining health and ensuring tissue stability. (Chiu Y, 2015).

Other studies stated that and inadequate Keratininzed Mucosa Width (KMW) (<2 mm) may be significant risk indicators for peri-implant disease and pain/discomfort during brushing. (Latimer J et al., 2021).

Moreover less than 2 mm of KMW was identified as a main risk factor for peri-implantitis in patients who do not regularly follow implant manteinance. (Figueredo et al, 2024). This was also afirmed in another study some six years before (Monje et al., 2018)

It has been reported that the choice of a shorter abutment, represented in our data as ‘Pilar’ may increase MBL, outlining that the keratinized tissue width is not the critical factor (Galindo-Moreno et al., 2014), which is also supported by another study due to not convincing and low level of evidence (Dreyer et al., 2018)

This non-determined statements will be analysed further in this study, aiming to obtain our conclusions about the importance of these features.

Regarding to the emergence angle of the (<30º or >30º) some evidence supported in a low to moderate certainity that platform-matched implants with an emergence angle of ≤30º may have positive effects on the peri-implant marginal bone level changes (Momen et al., 2023).

On the contrast, supporting this theory another study concluded that an emergence angle of >30 degrees is a significant risk indicator for Peri-implantitis. (katafuchi et al., 2017)

Another study suggested that placement of implants in undersized osteotomy sites will result in an increased remodeling of the cortical bone during the early healing process. (Abrahamsson et al., 2021)

Also a positive correlation between the insertion torque and the Implant Stability Quotient (ISQ) was found. (Pérez-Pevida et al., 2020). The ISQ-values are used as an indicator for mechanical implant stability (Gupta G, 2021).

An study stated that the macrogeometry of the implant and the drilling sequence have a significant effect on both primary stability values (ISQ and insertion torque). It was found that greater torque insertion values and ISQ were significantly higher for Coral Implant rather than ocean's (Pérez-Pevida et al., 2020).

This Figure 1 shows how the implant surrounding area is inflammated, after having done the implant surgery. Also the bone will progressively reabsorb.

Figure 1: Periimplantitis

As it has been exposed, there are no clear conclusions regarding to the risk factors of Peri-implantitis. Some studies even contradict the conclusions of the others.

It is important to contribute in this research field, as better understanding would make a relevant difference in the diagnosis and treatment. Doctors would be more informed and capable of leading their patients to a good recover or avoiding posible complications.

Methodology

A preliminary study of the categorical variables is made, showing the ones that appear significant, its relation with Peri-implantitis one to one. For this, contingency tables and the Chi-squared test is made. This will also allow to calculate the different probabilities for different combination of features.

Next, we'll train some predictive classification models:

- **Logistic Regression**: this supervised learning model is used to predict the probability of a binary outcome based on one or more predictor variables. It uses the logistic function to model the relationship between the predictors and the outcome, providing the probability that a given input point belongs to a certain class. This model estimates the coefficients of the input variables using maximum likelihood estimation.(Medium, 2023)
- **Support Vector Machines**: SVM is a supervised learning classifier that works by finding the hyperplane that best separates the classes in a high-dimensional space. For linearly separable data, SVM finds the hyperplane that maximizes the margin between the classes (IBM, 2023). For non-linearly separable data, SVM uses kernel functions (like polynomial or radial basis function kernels) to transform the data into a higher dimension where a hyperplane can be used to separate the classes.
- **Random Forest**: this is an ensemble supervised learning method that constructs a multitude of decision trees during training and outputs the mode of the classes for classification or the mean prediction for regression tasks. Each tree is trained on a random subset of the data, and the results are aggregated to improve accuracy and control overfitting. It is a robust model in terms of overfitting. (Medium, 2020)

It will provide a comprehensively evaluation of the performance and predictive power of various machine learning algorithms for predicting Peri-implantitis.

By comparing these models, we can identify which algorithm provides the highest metrics in predicting the onset of Periimplantitis.

The key metrics used for evaluation included (Medium, 2023):

- Accuracy: The proportion of correctly classified instances.

- Precision: The proportion of true positive instances among the instances predicted as positive.
- Recall: The proportion of true positive instances among all actual positive instances.
- F1 Score: The harmonic mean of precision and recall, providing a single metric that balances both concerns.

Logistic regression will offer a baseline for understanding linear relationships between the categorical predictors and the disease. At first an step-wise regression will be performed in order to later create a significant logistic regression model.

For SVM model, different types of kernel will be compared and after obtaining the best one, its performance and results will be analysed.

To perform the Random Forest, that will allow us to explore the impact of feature importance and the robustness of ensemble methods, this will be created, trained and hypertunned, maximizing its performance results.

By computing the feature importance of these models we can quantify the contribution of each feature to the overall prediction made by the model. Being able to indentify and ranking these contributions allow us to gain insights into which variables are most influential in determining the outcome.

The results from these models will guide us in understanding the underlying factors contributing to Periimplantitis, thus providing valuable insights for developing preventive strategies and improving patient outcomes.

This multifaceted approach ensures a thorough examination of the data, leveraging the strengths of each model to achieve the most reliable and actionable results.

In order to carry out the Project we used the following libraries:

- Pandas: for the treatment and use of the data
- Statsmodels: provided the functions for performing the logistic regression
- Sklearn: supplied the functions for the models, as well as other useful tasks such as the scaling, cross validation, the calculation of the metrics or dataset partition.
- Numpy: to calculate measures and treat with the arrays.
- Seaborn: provided the tools for data visualization.

- Matplotlib: another visialization library

Ethical and legal Analysis

The following data, was collected by two students from the Lifelong learning Master's degree certificate in oral surgery an orofacial implantology. It Previous information to the surgery was needed such as personal information, habits and medical background. On the other side, during the healing process the data was also collected in order to diagnose if the patient was healthy or with a peri-implantitis.

The protocol was be sent to the Ethics and Drug Research Committee (CEIm) of the *Hospital Odontològic de la Universitat de Barcelona* so that all the possible ethical implications of this study could be assessed. On one hand, the investigators undertake to respect the requirements stipulated in the Helsinki declaration of 1975 and revised in 2013, as well as to comply with the regulations of good clinical practice and the ICH for the conduct of the clinical trial. In addition, all participants signed a specific informed consent form and were be given a sheet with information about the study, stating that participation was voluntary and that the patient could withdraw from the study without having to justify his or her decision. Participants are able to contact one of the investigators at any time to ask any questions about the study or to report any unexpected event. On the other hand, the patient's privacy will be safeguarded, as well as the protection of his/her data in accordance with the current Personal Data Protection Law. In this sense, the participant will have access to the records generated during his/her participation in the study. Moreover, identifyers and personal data that could allow their identification was excluded from this study.

This protocol was approved in order to carry out investigations.

Data exploration and understanding

The dataset used in this study contains information from of 88 patients, totalling 213 implants and 44 features pertaining to patient and implant specifications. Notably, the dataset exhibits an inherent class imbaalnce, with 33 implants diagnosed with periimplantitis, while the majority, comprising 180 implants, were deemed healthy.

This imbalance presents a significant challenge in model training in evaluation, as it can lead to bias and suboptimal performance. Adressing it is crucial to ensure the future treatment in each model.

Categorical Variables

Sex

Medical gender of the patient. Its values are 0 if the patient is a man and 1 if it's a female

DM

Describes if the patient has diabetes. A metabolic diseases characterised by chronic hyperglycemia resulting from defects in insulin secretion, insulin action, or both. (Craig et al., 2009)

Tab

Represents if the patient is an smoker or not

Imp

In these surgeries we can differenciate between two different types of implants created by Avinent (Avinent Science and technology, 2024), an advanced implant system business from Spain:

Coral Implant

This is the first implant system developed by Avinent. The aim to be simple safe and predictible

Its Biomimetic contributes to the optimal preservation of the tissues, allowing also great asthetic results.

It also can be seen that the neck diameter is greater than that of the body.

Ocean Implant

This implant mantains characteristics from coral, but its polished Surface favors the tissue adaptation.

Figure 2: Ocean and Coral Implant

Perio

Periodontitis is considered a chronic localised infection of the oral cavity that can trigger inflammatory host immune responses at local and systemic levels, and can also be a source of bacteremia. (Muñoz-Corcuera et al,, 2015). This feature represents if the patient had periodontitis at any point of their lives.

Placa

The variable plaque has 4 different variables:

- Grade O. No plaque.

- Grade I. Thin film of plaque on the gingival margin, only recognizable by smear with the probe.

- Grade 2. Moderate plaque along the gingival margin; interdental spaces free; recognizable with the naked eye.

- Grade 3. Heavy plaque along the gingival margin;interdental spaces occupied by plaque (Bordoni etal, 1992).

It will have only two unique values: 0 and 1, effectively converting it into a binary variable, were grade 0 and1 were replaced by 0 and grade 2,3 by 1.

Sang

Bleeding on probing (BOP) this has set as indispensable diagnostic tool for evaluating periodontal disease activity, which is also used as diagnosis feature in our study. Nonetheless it was stated that “BOP should be used in addition to other parameters such as visual signs of inflammation, probing depth, and progressive bone loss before a peri-implant diagnosis is established”. (Dukka et al., 2020)

Sup

The presence and grade of supuration are associated with peri-implant bone loss, probing depth, and defect morphology in patients with peri-implantitis, which makes it a diagnostic criterium for Peri-implantitis(Monje et al., 2020).

Keratinized mucosa

This an oral mucosa that forms a physical barrier between the oral environment and the underlying connective tissues of the periodontium. This keratinized tissue has functions both for sensation and protection. (Brito et al., 2013)

Tprot

We can distinguish between different types of prothesis:

- Bridge: type of fixed prosthesis that is anchored to teeth or implants and replaces several dental crowns.
- Overdenture: is a removable prosthesis of the entire dental arch that can be attached to roots or implants (Llerandi i Béjar, 2024).
- Crown: is a covering that passes over an existing tooth and resembles a tooth (JDC, 2022).

AngProt

The emergence angle (EA) is defined as the angle of an implant restoration's transitional contour as determined by the relation of the abutment's surface to the implant body's long axis (Kou et al., 2023)

It can be appreciated in Figure 3 (Mattheos et al., 2021):

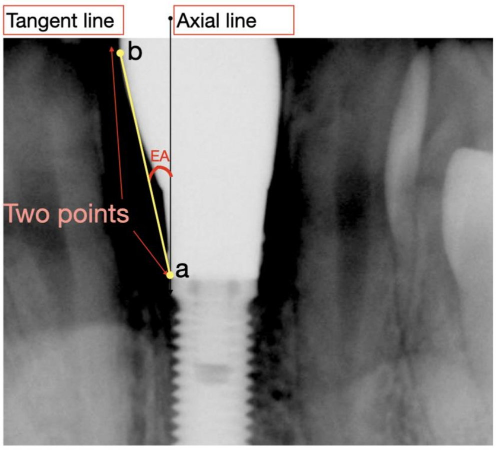


Figure 3: Emergence Angle

ROG/RTG

Dental surgical procedures that employ barrier membranes to facilitate the growth of new bone and gum tissue in areas where there is not enough bone or gum tissue for proper function, appearance, or prosthetic restoration are known as guided bone regeneration and guided tissue regeneration (Larsen P, 2024). Guided bone regeneration usually involves procedures to increase the volume or dimensions of the jawbone, while guided tissue regeneration focuses on utilizing bone grafts with barrier membranes to reconstruct small defects around dental implants (Kim et al., 2020)

Conex

This seems to be an important factor in modulating bone level changes in implant-supported reconstructions [4]. Marginal bone changes around implants with different connection types have been attributed to several etiological factors, such as biomechanical factors that increase the stress at marginal bone tissue and potentially contribute to alveolar bone resorption. (Aparecido et al., 2017)

This feature has three values:

- Conical implant–abutment connections are characterized by large clamping force, which is transformed from the large frictional resistance in the conical interface and helps 2-piece connections function as a single entity. (Yao et al., 2018) They are popular for their excellent connection stability, which is attributable to frictional resistance in the connection.
- External conexion implants(EC) were devoleped first and they are characterized by an external hexagon, this have been widely used for several decade. Although, the micromovements of the abutments due to their limited hexagonal height have remained a drawback
- Internal conexión implants (IC) with a conical internal self-locking system, have shown excellent mechanical stability and the ability to reduce stress on the marginal bone by transferring the exerted stress toward the apical area (Hyun Ju et al., 20218). They were designed to reduce the complications found in external connections.

Regarding to the Implant connection. On the other side, Internal conexion (IC) implants,

Internal conexion implants These have been claimed to be more mechanically stable, since the load is distributed deep within the implant, where engagement with a long internal wall shields the abutment screw (Gracis et al., 2012) (Aparecido et al., 2017)

There were just 3 values for the Conical implant so it was transformed with the value 0 that will refer to External and Conic conexión while 1 referes to Internal conexion.

This conversion into binary format simplifies the dataset, making it easier to interpret and analyze and required by the logistic regression.

Pilar

This feature refers to the Prothesis abutment. It can be of different heights:

- 1 mm
- 2 mm
- 3 mm
- 4 mm
- 5 mm

Indirectly it implies if we have a thick gum (>2mm) or not, which will be favorable for not having peri-implantitis.

Dx

This variable is the diagnosis. Patients can be classified as healthy, or diagnosed with mucosisits or periimplantitis. Whereas Peri-implantitis is an inflammatory process resulting in loss of supporting bone, peri-implant mucositis has been defined as a reversible inflammatory change of the peri-implant soft tissues without bone loss (Albrektsson & Isidor 1994)

defdiam and deflong

Implant length and diameter are usually selected based on bone availability.

The diameter can be narrow, regular or wide. Whereas the length could be short or regular. When there are narrower bone ridges, we usually use small diameter implants to avoid bone grafting procedures. The same applies to implant length (if there is little bone height). These variables do not usually have a major impact on marginal bone loss. Short and/or narrow implants have been shown to function virtually the same as implants of standard diameter and length.

These features were transformed as one-hot-encoded for using them in the ML models.

Defpilar

This features describe the use of an abutment or not. It is usually favorable (see the attached study by Galindo-Moreno), especially if it is more than 2mm in height.

We can see in Table 1 a view the categorical features.

Variables	**Name Variable**		**PI-Group**		**Non-PI-group**	
Categorical variables			**n**	**%**	**n**	**%**
Total			33		180	
Sex	Sex	Men	15	45,5	102	56,7
		Woman	18	54,5	78	43,3
Diabetes	DM	Yes	3	9,1	27	15
		No	30	90,9	153	85
Smoker	Tab	Yes	7	21,2	39	21,7
		no	26	78,8	141	78,3
Implant type	Imp	Coral:0	24	72,7	94	52,2
		Ocean: 1	9	27,3	86	47,8
History of Periodontitis	Perio	Yes	24	72,7	75	41,7
		No	9	27,3	105	58,3
Plaque Index	Placa	0	0	0	3	1,7
		1	8	24,2	51	28,3
		2	16	48,5	106	58,9
		3	9	27,3	20	11,1
BOP	Sang	Yes	32	97	89	49,4
		No	1	3	91	50,6
Supuration	Sup	Yes	3	9,1	1	0,6
		No	30	90,9	179	99,4
Keratinized Tissue W	KT	<2mm: 0	12	36,4	22	12,2
		>2mm: 1	21	63,6	158	87,8
Type protesis	Tprot	Corona:0	12	36,4	94	52,2
		Puentes: 1	17	51,5	65	36,1
		Sobredentadura: 2	4	12,1	21	11,7
		Híbrida:3	0	0	0	0
Protesis Angle	AngProt	>30º	6	18,2	6	3,3
		<30º	27	81,8	172	95,6
Guided bone regeneration/ Guided tisue regeneration	ROG/RTG	No	31	93,9	156	86,7
		Sí	2	6,1	24	13,3
Type of Conexión	Conex	External	15	45,5	53	29,4
		Internal	15	45,5	112	62,2
		Conic	3	9,1	15	8,3
Pillar height	Pilar	1mm	2	6,1	7	3,9
		2mm	2	6,1	31	17,2
		3mm	4	12,1	8	4,4
		4mm	1	3	29	16,1
		5mm	1	3	4	2,2
Diagnosis	Dx	Healthy	0	0	91	50,6
		Mucositis	0	0	89	49,4
		Periimplantitis	33	100	0	0
Diameter	defdiam	Slim	4	12,1	24	13,3
		Regular	25	75,8	120	66,7
		Wide	4	12,1	36	20
Longitud	deflong	Short	0	0	7	3,9
		Regular	33	100	173	96,1
Pillar	defpilar	No	23	69,7	101	56,1
		Yes	10	30,3	79	43,9

Table 1: Distribution categorical data

Moreover, we can observe in Figure 4 the distribution of the variables we aim to use as predictors for the models.

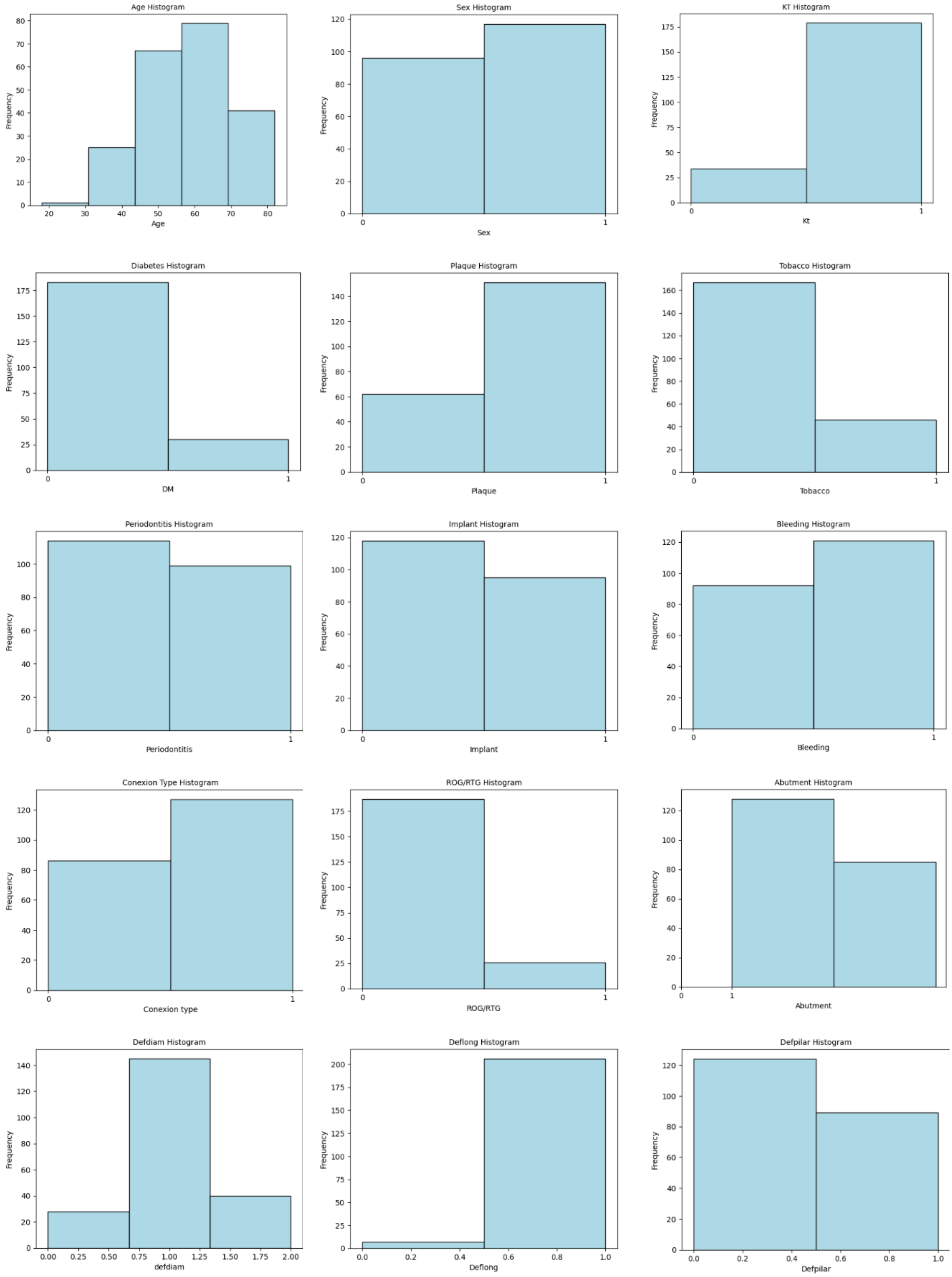


Figure 4: Graph distribution categorical variables

Numerical Variables

<u>Caso</u>

Each patient can have more than one implant, this feature counts the implants this patient has.

Edad
The age of the patient is represented by this variable.

Diam
The diameter of the impant in milimeters is collected by this variable.

Long
The longitud of the implant in milimeters is collected by this variable.

Vmant
The number of manteinance visits is collected by this feautre for each patient. Most of them had just one or none.

Bone loss
In order to diagnosing Peri-implantitis Marginal bone loss is used. According to some studies 2 mm as the maximum acceptable MBL after 1 year of loading for considering an implant to be a success (Misch et al. 2008).

The same los measure criterio was used in this study. The following variables determined this loss, in different parts of the implant.

Basal Marginal Bone Loss refers to the initial measurement, after the surgery, of bone loss around a dental implant after it has been placed

Final Marginal Bone Loss is the measurement of bone loss at later stage, which allowed to determine wether the bone loss was over the threshold (2mm), ergo determining peri-implantitis.

Figure 5 shows the bound surrounding sites, where bone loss was measured and stored in variables: mesial (MI) and distal (DI) which are aspects of the implant restoration (Kan et al., 2003)

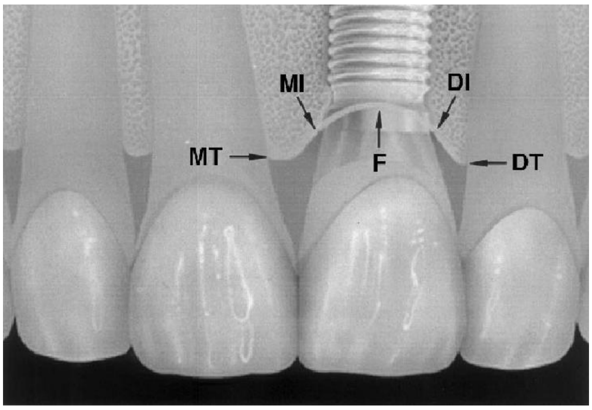


Figure 5: Bound surroundings

FollowUp
It is calculated the number of days between the date when the study is being conducted and the date when the surgery was performed. After this it is divided by 365.25 to convert this difference from days into years, including leap years.

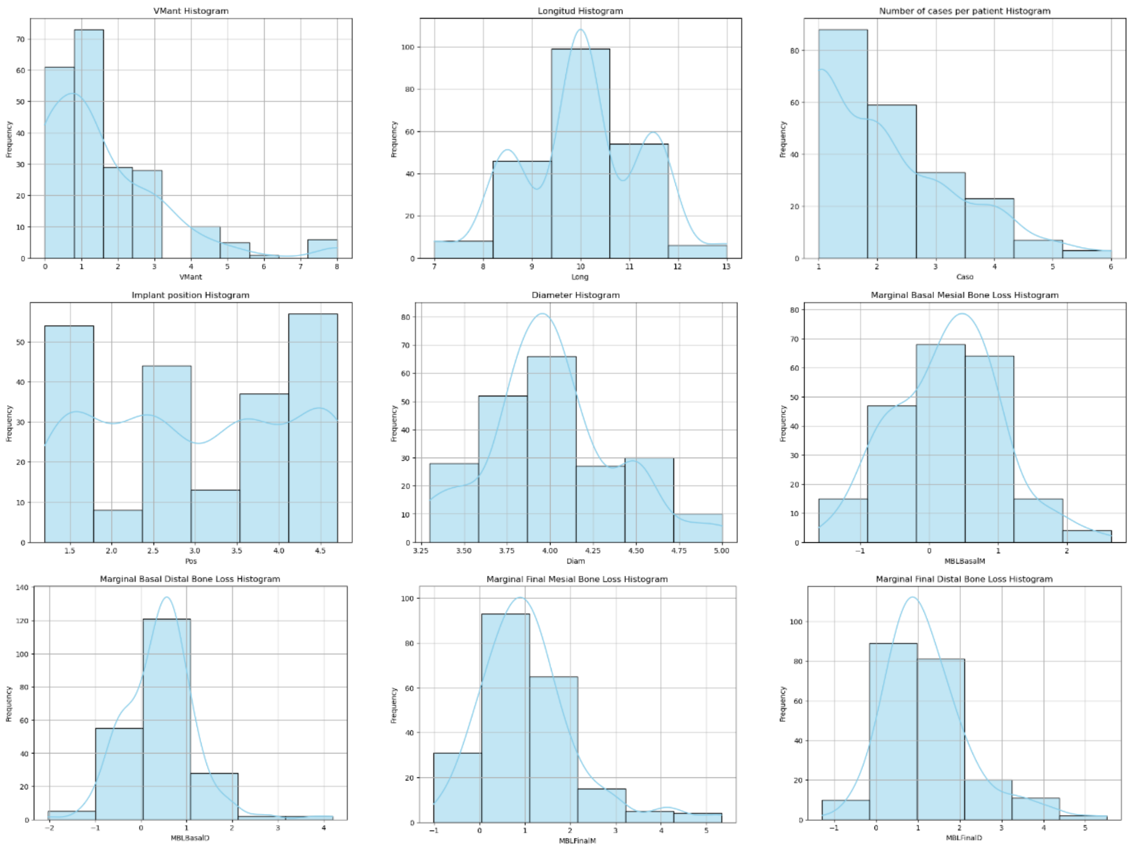
RatioMant
This is the number of mantainance visits divided by the years of follow-up.

We can also see in a view in Table 2 the numerical variables.

Numerical Variables			Mean	SD	min	max	Mean	SD	min	max
Implant number	Nimplant	(unit)	114,18182	41,681422	27	204	105,68333	64,62928	1	213
Case	Caso	(unit)	2,121212	1,408847	1	6	2,111111	1,200041	1	6
Age	Edad	(unit)	59,090909	11,930433	36	77	59,111111	11,532267	18	82
Diameter	Diam	(mm)	3,939394	0,323979	3,3	4,5	4,025556	0,374781	3,3	5
Longitud	Long	(mm)	10,454545	1,325021	8,5	13	9,958333	1,257242	7	13
Counts manteinance check	VMant	(unit)	2,333333	2,508319	0	8	1,427778	1,491243	0	8
Marginal Basal Mesial Bone Loss	MBLBasalM	(mm)	0,813636	0,944679	-1,45	2,64	0,169111	0,701591	-1,6	2,07
Marginal Basal Distal Bone Loss	MBLBasalD	(mm)	1,015758	1,051312	-1,31	4,2	0,325556	0,678747	-2,03	2,61
Marginal Final Mesial Bone Loss	MBLFinalM	(mm)	2,371212	1,553974	0	5,33	0,852833	0,777491	-1,01	2,97
Marginal Final Distal Bone Loss	MBLFinalD	(mm)	2,511515	1,326299	0	5,51	0,978611	0,767898	-1,29	3,65
Follow Up	FollowUp	(unit)	6,346919	2,18141	2,332649	9,199179	4,482881	1,877115	1,262149	9,845311
Manteinance Ratio	RatioMant	(unit)	0,396423	0,3646	0	1,222159	0,356427	0,350323	0	1,366155

Table 2: Distribution numerical data

And their distributions shown in Figure 6.



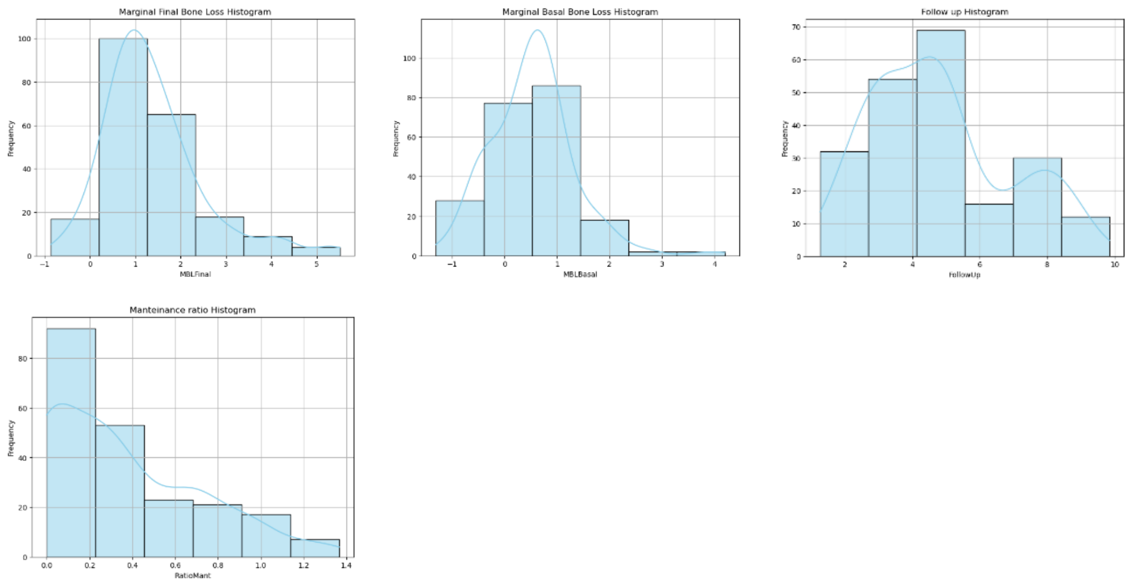


Figure 6: Graph distribution numerical variables

Imputation of missing data

Besides, an inputation of missing data before creating the models was done with Python. This was performed with Nearest Neighbours algorithm (kNN) , which is an efficient method to fill in missing data (Brownlee, 2020). Each missing value on a record is replaced by a value from related cases in the whole set of records that depends on the type of variable used: categorical missing values are replaced by the mode and quantitative ones are replaced by the mean.

The RMSE is a comprehensive evaluation metric for kNN regression models, it allows to choose the optimal value of K that minimizes prediction errors and ensures good generalization performance on unseen data. This was done by determining how the Root Mean Square Error varied among the number of neighbours A Lower RMSE values indicate that the model fits the data well and has more precise predictions, so after performing it, as it can be seen in Figure 7 the lowest value is by 8 neighbours, followed by 11 and 12. So the number of neighbors was fixed to 8.

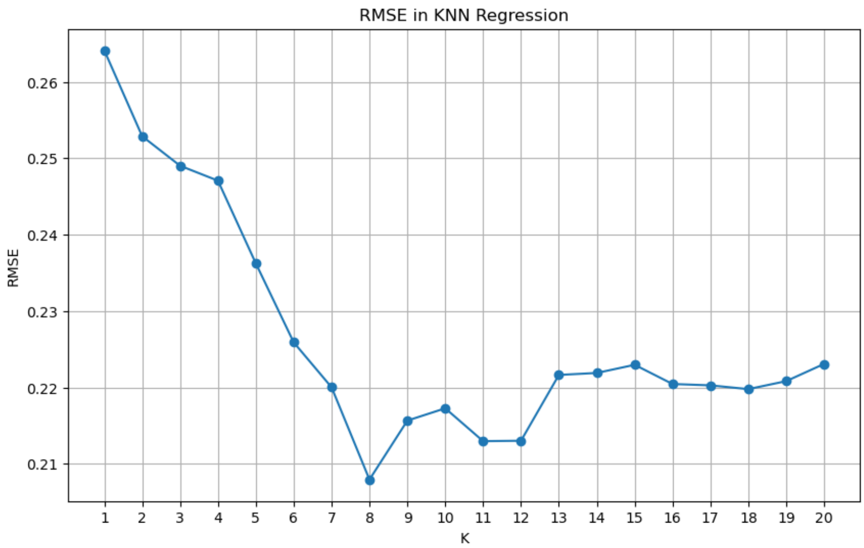


Figure 7: RMSE with kNN imputation missing data

Chi-squared Test

Subsequently, multiple chi-square test of independence were performed. This enable to examine the relation between each categorical variable and the Periimplantitis. Among 136 tests, only 6 of them were significant. So this relations were analysed, by calculating the proportions and the Relative Risk. The following conclusions exposed in Table 3 were made.

Feature	Levels	p-value	Interpretation
Implant	2	0.029	The risk of peri-implantitis with the Coral implant is 2.14 times higher than with the Ocean implant (p=0.029).
Periodontitis	2	0.001	Periodontitis multiplies by 3.07 the risk of peri-implantitis (p=0.001).
Bleeding	2	<0.001	Bleeding multiplies by 24-3 the risk of peri-implantitis (p<0.001).
KT	2	<0.001	The risk of peri-implantitis with KT<2mm is 3.00 times higher than with KT>2mm (p<0.001).
AngProt<30	2	0.001	The risk of peri-implantitis with AngProt>30 is 3.69 times higher than with AngProt<30 (p=0.001).
Type of conexion	2	0.071	The risk of Peri-implantitis with Non-Internal Connection is 1.77 times higher than with Internal Connection (p=0.071).

Table 3: Chi-squared test Categorical data

Predictive modelling

Logistic Regression Model

Model Setup

In order to predict Periimplantitis with the categorical variables, a logistic regresion model was created taking into account the categorical variables.

First, a backward elimination step-wise regression was performed. This consisted on creating regression models with all posible categorical variables which had a significant relation in the Chi-squared tests. Discarding one-by-one the features that were not significant.

After, a logistic regression model with only significant variables was performed with Python. It was taken into account the imbalanced by applying weights to the classes, including the parameter 'class_weight'= 'balanced', which adjusts the weight of each class inversely proportional to its frequency by following the next equation:

$$\text{weight}_i = \frac{n_{samples}}{n_{classes} x \; n_i}$$

Being:

- $\text{weight}_i =$ weight for class i.
- $n_{samples} =$ total number of samples.
- $n_{classes} =$ number of unique classes.
- $n_i =$ number of samples in class i.

Peri-implantitis Probabilities

Moreover, some interactions where considered due to a posible relation. The result was a model that allowed to obtaining different probabilities for the different combinations of features.

Implant	Periodont	KT>2mm	Conex	Probability of Peri-implantitis
Coral	Yes	<2mm	Non Internal	88.09%
Coral	No	<2mm	Non Internal	66.91%
Coral	Yes	<2mm	Internal	61.98%
Coral	Yes	>2mm	Non Internal	35.42%
Coral	No	<2mm	Internal	30.83%
Ocean	Yes	>2mm	Internal	21.90%
Ocean	Yes	>2mm	Non Internal	14.40%
Ocean	Yes	<2mm	Internal	14.02%
Coral	No	>2mm	Non Internal	13.04%
Coral	Yes	>2mm	Internal	10.78%
Ocean	Yes	<2mm	Non Internal	8.91%
Ocean	No	>2mm	Internal	7.12%
Ocean	No	>2mm	Non Internal	4.40%
Ocean	No	<2mm	Internal	4.27%
Coral	No	>2mm	Internal	3.20%
Ocean	No	<2mm	Non Internal	2.61%

Table 4: Feature probabilities for Peri-Implantitis

In Table 4 It can be observed that the coral implant (red) is associated to a greater extent with peri-implantitis (high positions in the graph, indicating high probability of peri-implantitis), especially when accompanied by non-internal connection and KT<2mm, since the interaction with these factors accentuates the effect of the coral implant.

It can be appreciated in Figure 8: Estimated probability of peri-implantitis for different feature combinations that the combination of KT<2mm and Non-Internal Connection aggravate the effect of the Coral Implant, this is the explanation of the two interactions. Note that if we change the first combination (the most dangerous with Risk=88.1%) to KT>2mm and Non Internal Connection, its risk decreases 9 positions (up to 10.8%).

The effect of Periodontitis is easily seen, since it is associated with high probabilities, while the combinations with lower probabilities are those that do not include periodontitis.

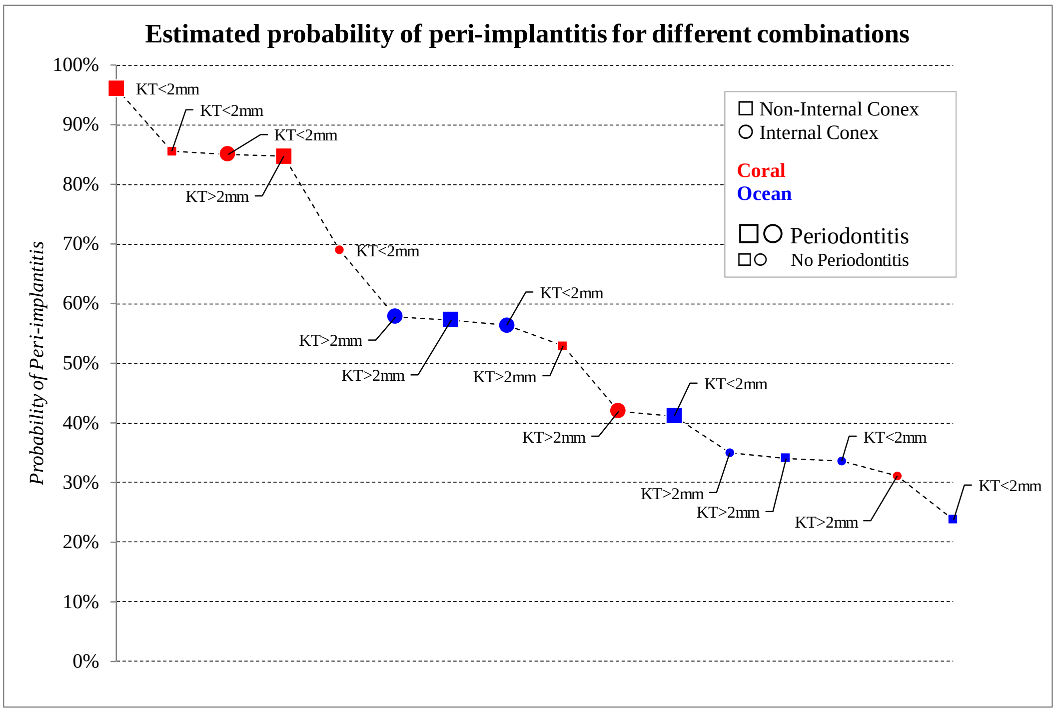


Figure 8: Estimated probability of peri-implantitis for different feature combinations

In summary, Non-Internal conexion seems to be the most problematic as well as a lower Keratinized Tissue, which matches with previous studies.

Coral implant with different combinations appear to have between a 90 and 30% probability of peri-implantitis, which shows a problem with this implant type that should be studied.

Nevertheless, Ocean also shows a lower-moderate probability with a KT>2mm, and with Internal the probability of Peri-implantitis is bigger than with Non-Internal. This can probably be realted to the interaction of type of Implant and type of conexión. The same happens with KT, higher than 2mm has bigger probability of Peri-implantitis than lower than 2mm. There is also an interaction between the Implant type and the KT, which needs to be studied as for Ocean we can see different criteria regarding to develop the disease.

Although the probability with Ocean implant is much smaller than Coral, so this doesn't have to mean that Coral and Ocean implant have different criteria for developing Peri-implantittis, but it can also mean other variables, such as the numerical ones, not considered in this model, influenced the appearance of the disease. This will be studied by the other models.

Regarding to feature importance, from Figure 9: Feature Importance in Logistic Regression modelwe can observe that the Implant type (*Impl*) is the most explanatory variable, which is followed by the interaction conexion and type of Implant, with a lower importance.

In comparison, Periodiontitis (*Perio)*, *KT* and Conextion type (*Conex)* show similar values, showing around half of the importance of Implant type.

It is noticeable that Conex* Impl importance is higher than *Conex* itself. It reveals that there is a big influence in choosing an implant and the conexion, both afecting Peri-implantitis more than only the conexion by itself.

Lastly, the interaction between KT and the Implant type value is less than half of the first importance feature. It is clear that even though both features may influence, the KT still being important not depending more on the implant type rather than being thicker.

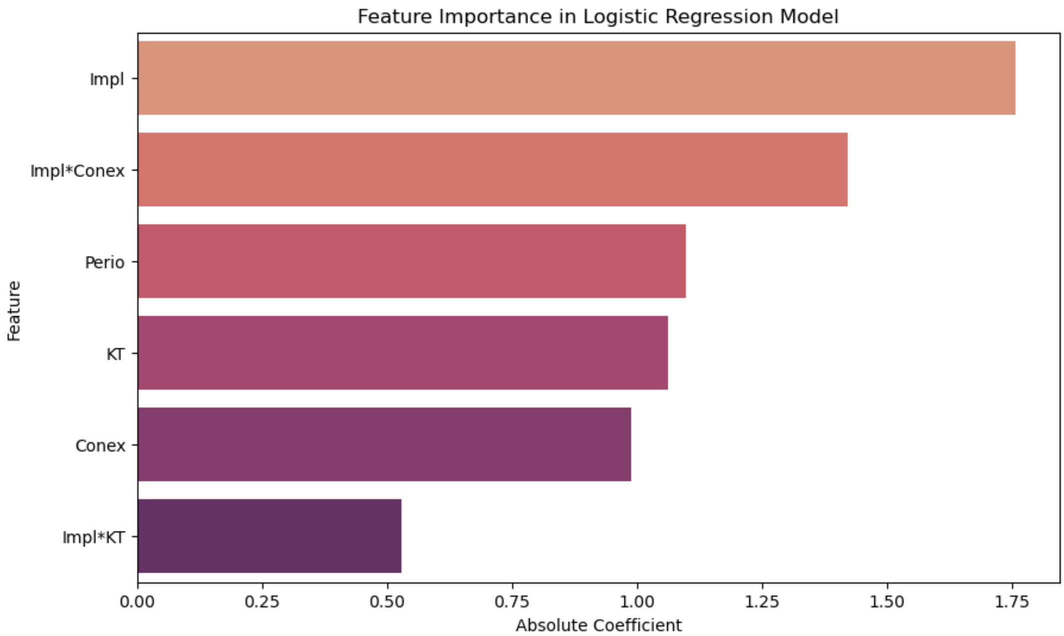


Figure 9: Feature Importance in Logistic Regression model

Support vector machines

Model Setup: comparison of different Kernels

SVMs are particularly effective in high-dimensional spaces, which make them very appropiate for our data set. Moreover, their ability to create robust and accurate models aligns well with the objectives of this research, as we look forward to develop a reliable predictive model for Periimplantitis.

SVM can use different kernels, this is function is to receive data as input and transform it into the desired form. We compared the linear, radial basis function and polynomial with 3rd degree. The results are in Figure 10: Performance comparison of Support Vector Machines Kernels, the polynomial kernel of third degree, was chosen due to it's performance results compared to other degrees.

Each model was trained 100 times, with the purpose of providing a better understanding of the models' performance variability.

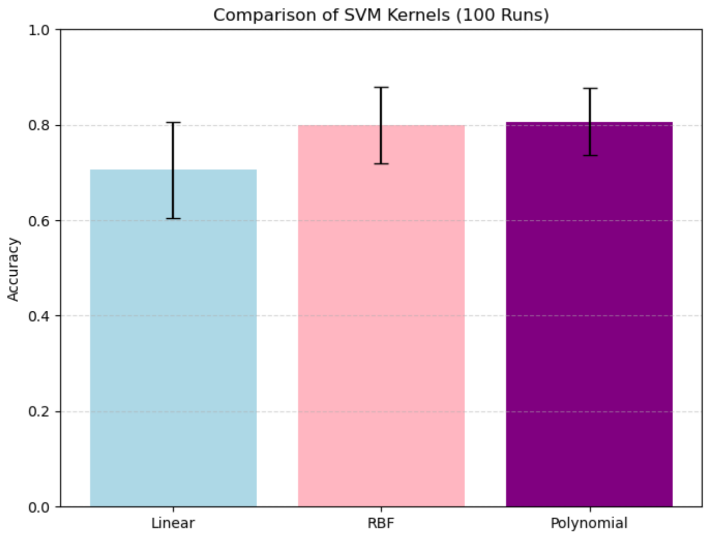


Figure 10: Performance comparison of Support Vector Machines Kernels

In order to reduce the variance of the model and to avoid overfitting, a cross validation with 5-folds was made and we obtained the scoring metrics to evaluate the performance of the model.

Confusion Matrix

The confusion matrix represented in Figure 11: SVM confusion matrix for the model reveals several key insights into its performance. The model correctly identified 35 negative cases (True Negatives) and 4 positive cases (True Positives), indicating that it has a reasonable ability to classify both the absence and presence of peri-implantitis accurately.

However, the model did incorrectly classify 2 negative cases as positive (False Positives = 2), which suggests that while it is fairly specific, there is still some risk of false alarms. Moreover, the model incorrectly identified 2 positive cases as negative (False Negatives = 2), meaning it missed some cases of peri-implantitis.

The balance between True Positives and False Negatives indicates that while the model performs reasonably well at detecting the presence of the condition, there is still room for improvement in identifying all positive cases. The presence of False Positives shows that some patients without peri-implantitis are flagged incorrectly, which could lead to unnecessary concern or further testing.

Overall, the model's performance suggests it is fairly specific but has limitations in sensitivity. Efforts to improve the model should focus on reducing both False Negatives and False Positives to enhance its reliability in clinical settings.

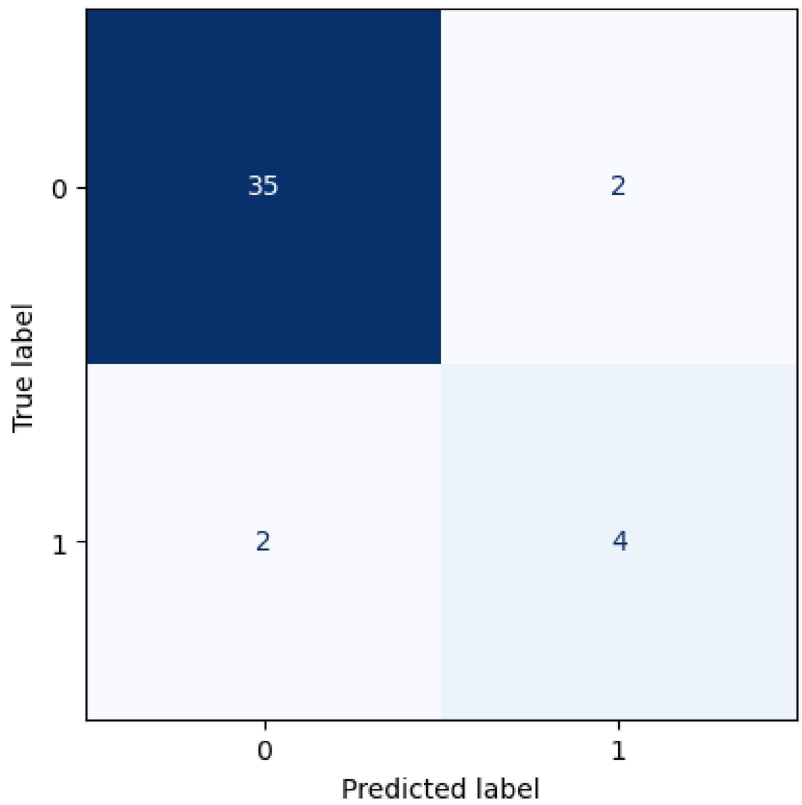


Figure 11: SVM confusion matrix

ROC-curve

By computing the ROC-curve we can observe in Figure 12: SVM ROC Curve that there is a 97% chance that the model will correctly distinguish between a randomly chosen positive instance and a randomly chosen negative instance. It also suggests that the model achieves high sensitivity (True Positive Rate) and high specificity (True Negative Rate) across different threshold levels.

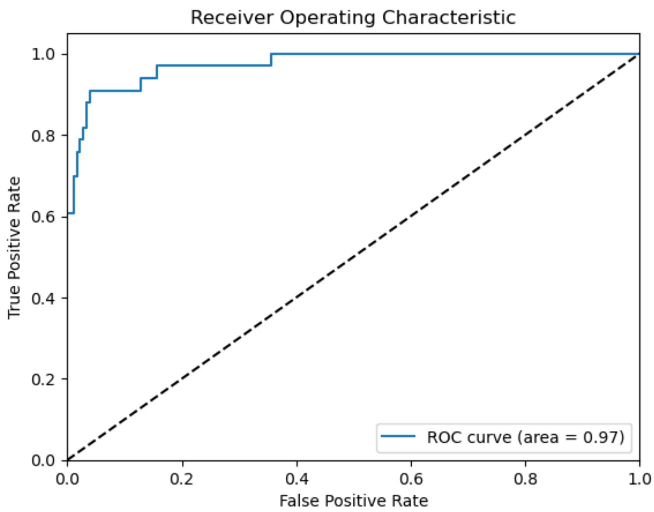


Figure 12: SVM ROC Curve

Precision-recall

Additionally, in Figure 13, by obtaining the Precision-recall curve we can observe that the model effectively manages the trade-off between precision and recall despite the imbalance, maintaining a high precision while also achieving a high recall. This means that among the instances predicted as positive, a large proportion are true positives (high precision), and among the actual positive instances, a large proportion are correctly identified (high recall).

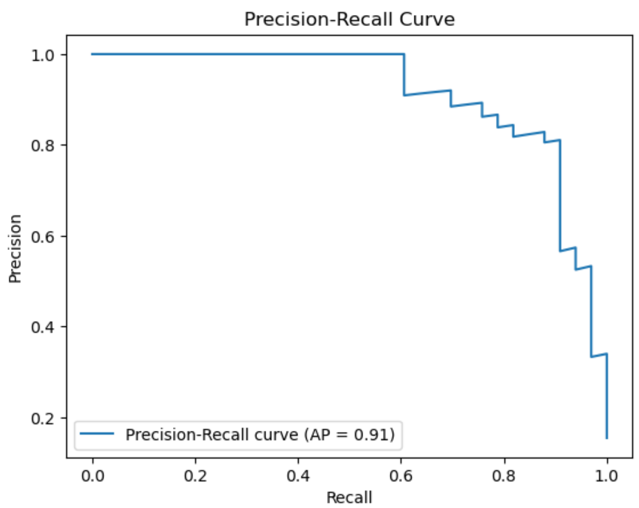


Figure 13: Precision-recall curve for SVM model

The following Figure 14 displays the importance of each feature in the trained SVM model. The importance is determined by the absolute values of the feature coefficients. As shown, *VMant* has the highest importance, indicating that doing medical checks has the most significant impact on the model's predictions. Followed by KT, which as we exposed, is a very discused feature and it is supported by this model. Also FollowUp appears thrid, expresses the follow up time, which is not related to the mantainance visits. So it makes sense that the most time it goes by, more probable would be developing the disease. But it is not taking into account any medical checks .

Compared to to the features of the logistic regression we can find some impact on predicting Peri-implantitis in the Number of implants per person (Caso) which appeared to be almost as important as FollowUp , this seems interesting as having more implants could influence the appearance of Peri-implantitis.

Moreover in less degree we have Tabaquism (*Tab*), which was also discused to be related to Peri-implantitis. Also having a Pilar (*defpilar_1*) or not (*defpilar_0*) shows to be relevant, the gender (*Sex*) and the emergence angle (AngProt). And some technical implant details are also important, such as the dimeter (*Diam)* of the implant that follows *Caso*.

Conversely, the position of the implant (*Pos*) and the longitud (*Long*) has the smallest impact.

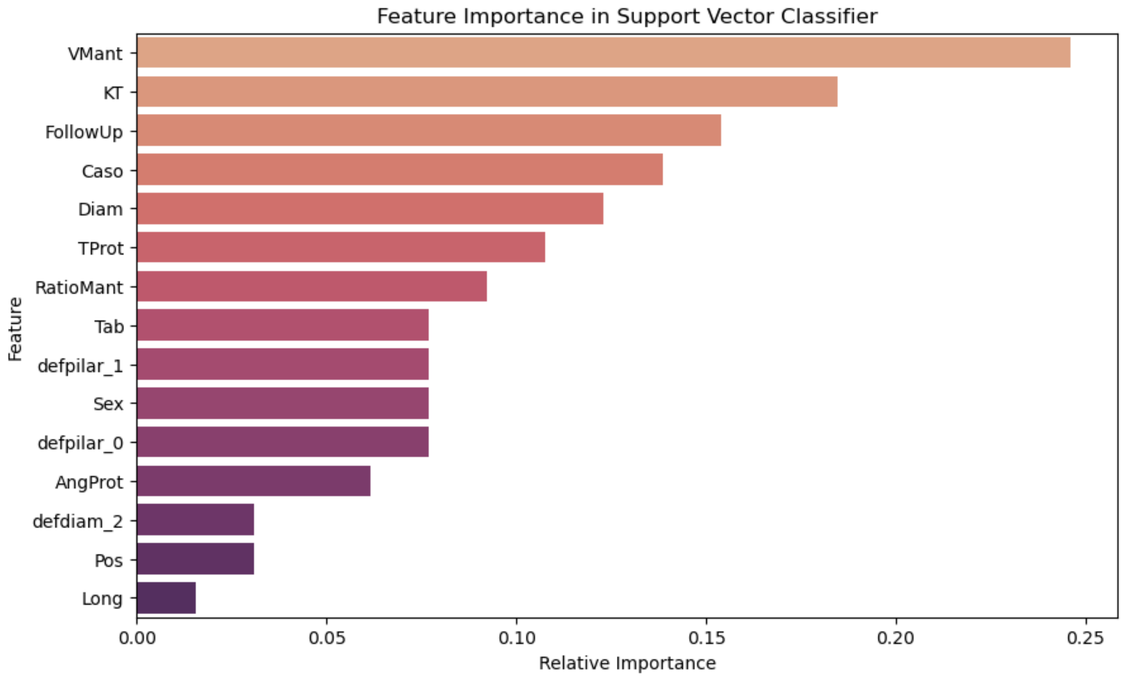


Figure 14: Feature Importance in Support Vector Machine Model

Random Forest

Initial Model Training

Afterwards, a Random Foest model was performed. The initial step in our modeling process was to train a RF model with a set of default hyperparameters. This provided us with a baseline performance metric, which we aimed to improve through systematic hyperparameter optimization.

Hyperparameter Tuning with GridSearchCV

To optimize the performance of our Random Forest model, we used GridSearchCV, a comprehensive method for hyperparameter tuning that exhaustively searches through a specified parameter grid. This process helps in identifying the best combination of hyperparameters for the model After conducting the grid search, the best hyperparameters identified were:

- max_depth: 6. This limit avoids the model does not become too complex.
- max_features: 'log2'. This means the number of features considered for splitting at each node is the logarithm (base 2) of the total number of features
- max_leaf_nodes: 9. This restriction can reduce overfitting
- n_estimators: 100. The more number of trees the more it can increase the model's robustness and accuracy.

Retraining

With the optimal hyperparameters identified, we retrained the Random Forest model. This involved fitting the model to the training data using the best hyperparameters obtained from the grid search.

Evaluation with Stratified 5-Fold Cross-Validation

In order to evaluate robustly the performance of the retrained model, we employed stratified 5-fold cross-validation. This technique involves dividing the dataset into five folds, ensuring that each fold maintains the same proportion of class labels as the original dataset.The model is trained on four folds and validated on the remaining fold, and this process is repeated three times, the whole process will be repeated 15 times. Stratified cross-validation helps in providing a more accurate and reliable assessment of the model's performance, particularly for imbalanced datasets.

In Figure 15 we compare both model's by its metrics:

- **Accuracy:** The first model is slightly more accurate, with less variability.
- **Precision:** The first model performs better in terms of precision, meaning it's more reliable in its positive predictions, although with high variability.
- **Recall:** The second model performs slightly better in recall, meaning it's a bit better at identifying all positive cases.
- **F1 Score:** Both models are quite similar in terms of the F1 score, with the first model being marginally better.

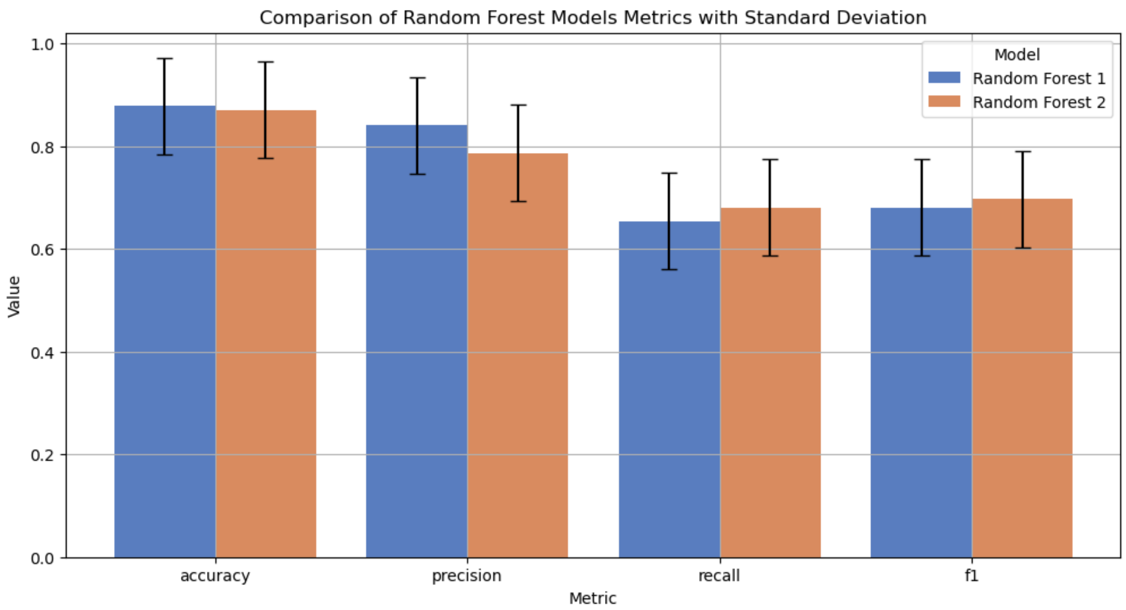


Figure 15: Random Forest models comparison with their deviation

While the first model shows slightly better accuracy and precision, the second model improves on recall. The F1 scores are very similar, indicating a trade-off between precision and recall. Depending on the specific application and whether false positives or false negatives are more critical, one model may be preferred over the other. The hyper-tuning of the second model has led to a slight improvement in recall at the expense of some accuracy and precision.

Decision tree

To further illustrate the structure and decision-making process of the Random Forest model, we can see in Figure 16 graphical representation of one of the individual decision trees within the ensemble.

In the decision tree chart, we can analyse the different values showed:

1. Each of the internal nodes has a decision rule that is shown in the first line that splits the data.
2. Gini, referred to as Gini ratio, measures the impurity of the node. You can say a node is pure when all of its records belong to the same class, such nodes known as the leaf node.
3. Samples refers to the number of data points (instances) that reached a particular node during the training process
4. Values is the proportion of weighted samples reaching this node for each output and class.
5. Class is useful in the leaf node understanding by the rules the path that lead to the final predictions, if they have or not Peri-Implantitis

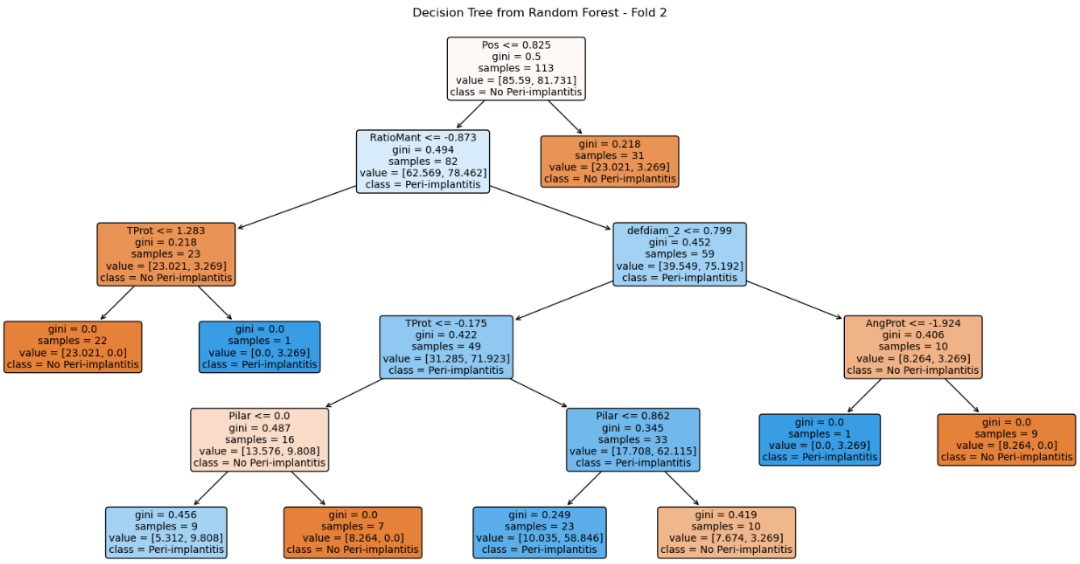


Figure 16: First decision tree Random Forest model

According to Figure 17: Feature importance for Random Forest Model, *FollowUp* shows up again as the most relevant feature. But as it was mentioned, it made sense as it was the tracking time.

In contrast with the SVM, *Pos*, which was almost the lastest important feature, in this model is very relevant, taking the second place. This feature was also outlined to affect to the disease, but few evidence was found.

Pilar height seems also useful aswell as the Age. These variables were low or not considered in the other models, which make it very intersting and also fit coincide with the studies exposed before.

It is also important to outline that the mantainance ratio and visits keep being critical predictive features, although a bit less than in the previous model.

Perio, which was quite relevant in logistic regresion, is nearly as important as *VMant*, and it is followed by *KT* which appears in every feature importance graph.

Additionally, some features such as *Diam* and *Long* appear to influence, but three times less relevant than *FollowUp*.

Features like *Sex, defpilar, AngProt*, Impl, *Conex, Tab, TProt* and *Caso* appear again*,* manifesting some importance but very limited for this model to predict Peri-implantits.

Previous features like *defdiam, deflong, ROG/RTG, Placa and DM* were not considered in the other models, even though for this one has a little contribution.

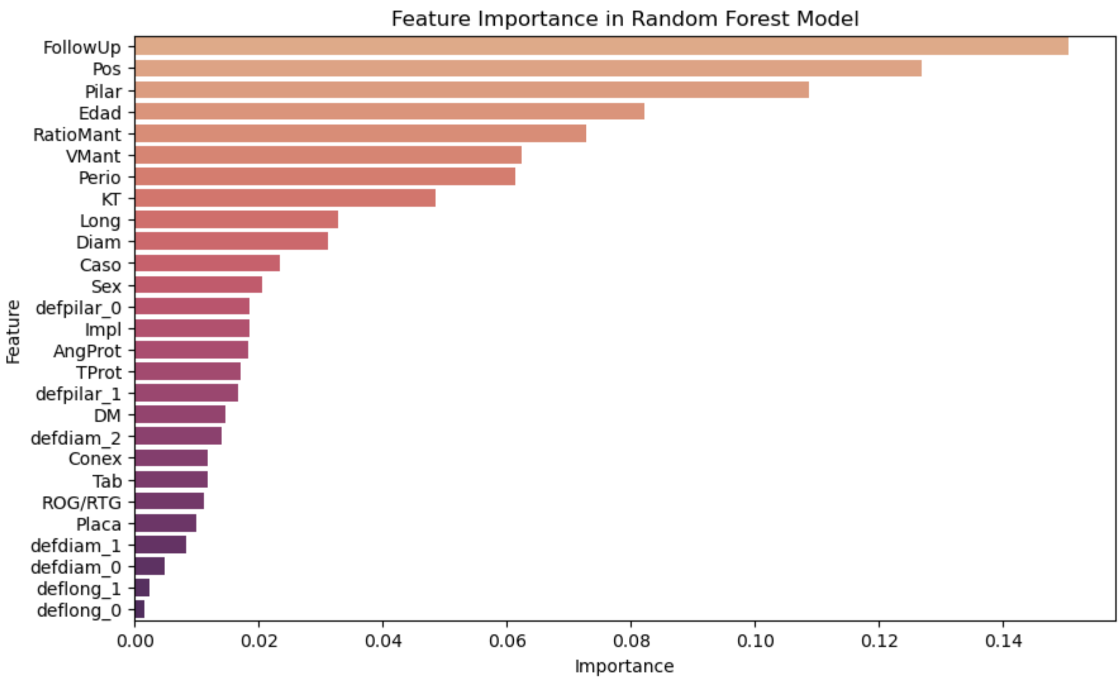


Figure 17: Feature importance for Random Forest Model

Confusion matrix

The confusion matrix represented in Figure 18 reveals several key insights into its performance. The model correctly identified 36 negative cases (True Negatives) and 1 positive case (True Positive), indicating that it has a strong ability to correctly classify the absence of peri-implantitis.

Notably, the model did not incorrectly classify any negative cases as positive (False Positives = 0), which suggests that it is highly specific and effectively minimizes the risk of false alarms. This is a positive aspect as it means the model is reliable in ensuring that patients without peri-implantitis are not subjected to unnecessary concern or further testing.

However, the model incorrectly identified 5 positive cases as negative (False Negatives = 5). This indicates that the model misses several cases of peri-implantitis, which is a significant limitation as these missed cases could go untreated.

The high number of True Negatives and the absence of False Positives show that the model excels in identifying patients without the condition. However, the imbalance between True Positives and False Negatives reveals a critical area for improvement: the model's sensitivity. Increasing sensitivity would help in detecting more cases of peri-implantitis and reducing the number of missed diagnoses.

Overall, while the model is highly specific and reliable in ruling out peri-implantitis in patients who do not have it, there is a significant need to enhance its sensitivity to ensure that it can identify and treat all positive cases effectively.

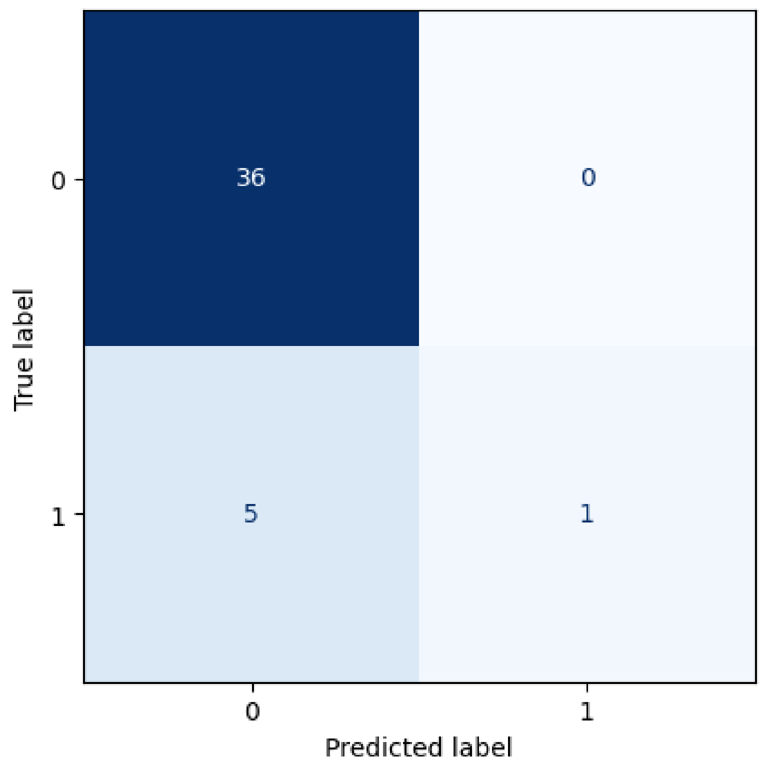


Figure 18: Confusion matrix for Random Forest model

Conclusions

Performance Metrics Comparison

The metrics performance from the different models is ilustrated in Figure 19. We proceed to obtain our conclusions from it:

1. SVM and Random Forest models demonstrate higher **accuracy** (82.7% and 87%) compared to logistic regression (63.9%), which indicats a better overall prediction capability.
2. **Precision** metrics vary across models, with SVM and Random Forest showing similar precision scores (73.6% and 78.7%), suggesting their reliability in correctly identifying positive cases. Logistic regression performs slightly lower in precision (55.7%).
3. **Recall** scores for SVM (61.5%) and Random Forest (61.5%) are comparable, indicating their ability to identify positive cases effectively. Logistic regression shows a lower recall rate (61.3%), potentially missing more positive cases.
4. **F1** scores are similar between SVM and Random Forest (62.0% and 69.7%), balancing precision and recall effectively. Logistic regression exhibits a lower F1 score (51.9%), reflecting a trade-off between precision and recall.

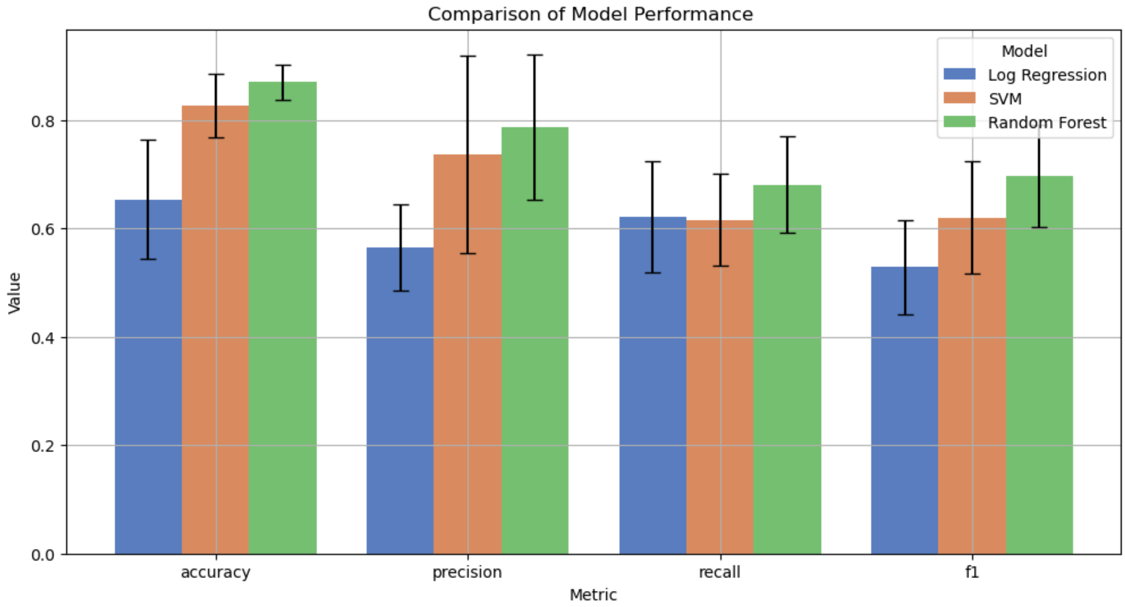


Figure 19: Models' Performance comparison

In comparing our study to the research conducted by Mameno et al. in 2021, oth their results shown in Figure 20: Models' Performance comparison Mameno's study several key points emerge, highlighting both the strengths and areas for improvement in our approach. They evaluated 254 implants with a balanced dataset with 4 years function, taking into account other variables:

- Fixation method: screw or cement
- PCR(Polymerase chain reaction): it's a technique that allows scientists to take a very small amount of DNA and amplify it, being able to identify a genetic change that can cause the disease
- Functional time of the implant: was also stated to be important for the development of Peri-implantitis.

It is important to outlin that they didn't consider any mantainance metrics or follow up, which where the ones that mostly contributed to our research.

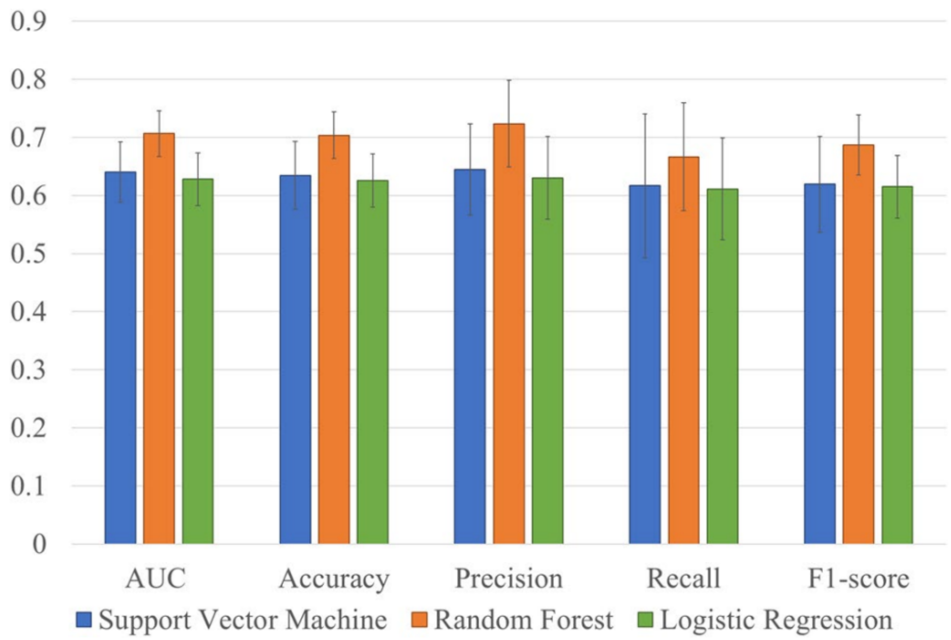


Figure 20: Models' Performance comparison Mameno's study

1. **Accuracy**:
    - Both studies achieved the same accuracy of 0.70.
2. **Precision**:
    - Our model demonstrates a superior precision of 0.79 compared to theirs of 0.72. Higher precision means our model has a lower rate of false positives. This improvement suggests that our model is better at correctly identifying patients who actually have peri-implantitis.
3. **Recall**:
    - Our model shows a lower recall of 0.61 compared to theirs of 0.66. While our model is less effective at capturing all actual cases of peri-implantitis, it still maintains a reasonable recall rate.
4. **F1 Score**:
    - The F1 score, which balances precision and recall, is slightly higher in our model (0.70) compared to theirs (0.69). This indicates that despite the lower recall, the overall balance between precision and recall is marginally better in our model. This slight edge in F1 score shows our model's effectiveness in managing the trade-off between precision and recall.

Feature Comparison across models

Based on the analysis of logistic regression, SVM, and Random Forest models we could notice that Random forest is one that gets more features, having a total of 27 variables, whereas Logistic regresion had 6 relevant features and SVM 15.

Certain features can be outlined as they consistently appear as relevant for predicting periimplantitis:

1. **FollowUp**: This feature consistently ranks highest across all models. The duration since implant placement strongly influences periimplantitis risk, suggesting that regular follow-up plays a critical role in monitoring and early detection.
2. **Keratinized Tissue (*KT*)**: KT thickness consistently appears as a significant predictor across models. Its role in maintaining periimplant health underscores its importance in risk assessment.
3. **Periodontitis (*Perio*)**: Periodontal health, indicated by the presence of periodontitis, consistently affects periimplantitis risk. This underscores the systemic nature of periimplantitis and its association with pre-existing periodontal conditions.
4. **Implant Type (*Impl*)** and **Connection Type (*Conex*)**: Variations in implant and connection types influence periimplantitis risk across models, though their significance varies slightly. Internal vs. Non-Internal connection types and specific implant designs impact outcomes.
5. **Maintenance Factor**: Variable related to maintenance, *VMant* in SVM and Random Forest, highlight the ongoing care and monitoring essential for periimplantitis prevention.
6. **Other Factors**: *Age*, smoking status (*Tab*), and anatomical considerations (e.g., pilar height, angle of prosthesis) also contribute to predicting periimplantitis, though to a varying extent depending on the model.

General conclusions

After having done a variate analysis of the variables, focusing on achieving objective 1, It was obtained that there can be some handicaps regarding to the implant type. The coral implant revealed to increase the probability of having Peri-implantitis. This may be due to the reason could be the macrogeometry of the implant and its higher insertion torque. The importance of this feature was also supported by the other models

Additionally, the implant type interacts with other two variables, the KT and the type of conexion. Supported by others studies, Non-internal conexion seemed more perjudicial, as well as a KT<2mm. Both features are relevant among all the models.

Moreover, aiming Objective 2, we came by with several key variables, that significantly influence the risk of periimplantitis. Besides the ones mentioned, we can highlight periodontal health (presence of periodontitis), consistently rank highest across all models as critical predictors and backed up by other researchers, and confirm the importance of the mantainance and checks after the surgery, which play a big role in the development of the disease. Technical features such as the Diameter, Longitud or the Position are also variables to take into account. Sex as well can be related to Peri-implantitis, but in contrast to previous studies Age didn't seem to be very significant. Tabaquism is also relevant, at least in the RF model, but Diabetes is not considered in any model. A different variable arise to be important, *Caso*, having more implants is connected to developing Peri-implantitis.

Finally, achieving our objective 3 of forecasting Peri-implantitis, our models, concretely the Random forest, show a very good performance for predicting Peri-implantitis, providing a robust tool for early detection and intervention.

Future work and Improvements

After performing various models to predict peri-implantitis, there are several areas for future work and potential improvements that could enhance the effectiveness and accuracy this study.

First, a significant area of improvement is the incorporation of Partial Least Squares Discriminant Analysis. This was performed after the other models so its usefulnes due to the dimensionality reduction was very little and wasn't taken into account. That's why doing it before would give the best outcomes.

Also it can be considered including additional figures, such as the number of teeth or the dental implants functional time. This features have been proved to be relevant. So it can lead to better results.

Another interesting point would be refining more the models. RF was already hypertunned but for the rest even if have shown promising results, further tuning and refinement could be beneficial. This includes exploring more sophisticated hyperparameter tuning techniques and possibly combining different modeling approaches to create an ensemble model that might offer better performance. This could be a PLS with SVM for instance.

Even if the the imbalanced data was handled, we can also consider further to introduce techniques such as SMOTE or ADASYN or other resampling methods.

For better robustness and ensuring generalizabilty of the results a good option would be having larger and more diverse datasets to be able to train and validate the models with them.

Additionally the exploration of new ML algorithms such as neural networks could give more information and provide different insights and performance.

Legacy

The variables and risk factors for peri-implantitis have been studied, yet no single factor or combination of factors can predict the condition with 100% certainty. This inherent uncertainty underscores the importance of leveraging advanced computational techniques like ML to improve prediction accuracy and aid in early diagnosis and intervention.

During the course of this research, we ensured the ethical handling of data, particularly sensitive patient information. Data anonymization was applied to protect patient identities, in compliance with privacy regulations. This involved good practice on collecting the data by the Universidad de Barcelona and the removal of personal identifiers to ensure individual patients could not be traced back from the dataset.

The preliminary results of the ML models developed in this study indicate a reasonable ability to predict peri-implantitis, but also highlight areas for improvement, particularly in increasing sensitivity and reducing false negatives. These findings are encouraging and suggest that with further refinement and larger datasets, ML models could become a more valuable tool in the clinical management of peri-implantitis.

In conclusion, while the prediction of peri-implantitis remains complex, the application of ML models represents a significant step forward. This Project has been a good starting point, as the previous research about its prediction was for this one, in order to succesfully predict this disease. For doctors or researchers from Avinent this can open another line of investigation regarding to their implants, so that they can study the technical structure of their implants and how it affects to the development or even inducement of this disease.

Nevertheless continued research in this area is essential to refine these models, validate their accuracy, and ultimately integrate them into clinical practice. By the usage of different data and training new models, we can enhance our understanding and management of peri-implantitis, leading to better patient outcomes.

Related to Degree Courses

The following courses during the degree have provided the theoretical and practical foundation necessary to carry out this project:

Fundamentals of Programming and Programming: The Programming was the major tool and mean to develop the Project. The exploratory analysis and machine learning models were programmed in Python.

Statistical Models for Decision Making I and II: These courses build on "Exploratory Data Analysis." They provide the basic concepts of some of the models used in the project.

Descriptive and Predictive Models I and II: Some models used were taught in these courses (SVC, logistic regresión, random forest)

Exploratory Data Analysis: Understanding the data at hand is essential for any project. Learning how to analyze and comprehend the data we are working with to extract useful information.

Visualization: A visual tool is fundamental for the understanding of the results. Visualization techniques were taught in this course.

Project I, II, and III: During three years we learn the steps to prepare us for data science projects, giving us also experience with analytical and technical skills.

Bibliography

- Abrahamsson I, Carcuac O, Berglundh T (2021) *Influence of implant geometry and osteotomy design on early bone healing: A pre-clinical in vivo study*
- Albrektsson, T. & Isidor, F. (1994) Consensus report of session IV. In: N. P. Lang & T. Karring (eds). Proceedings of the First European Workshop on Periodontology, pp. 365–369. London: Quintessence.
- Aparecido A, Ramos F, Yoshito P, Takanori D, Silva R, Marcela de Luna J, Piza E (2021), *Effect of bone quality and bone loss level around internal and external connection implants: A finite element analysis study. The Journal of Prosthetic Dentistry; Volume 125, Issue 1.*
- Aparecido C, Ramos F, Estevam A, Bonfante A, Ferreira J and Piza E (2017) *Comparison of external and internal implant-abutment connections for implant supported prostheses. A systematic review and meta-analysis*
- Avinent Science and technology (2024) *Implante Biomimetic CORAL* retrieved from https://implantsystem.avinent.com/biomimetic-implants/biomimetic-coral/
- Avinent Science and technology (2024*) Implante Biomimetic OCEAN* retrieved from https://implantsystem.avinent.com/biomimetic-implants/biomimetic-ocean/
- Bordoni N, Doño R, Miraschi C (1992) *Preconc. Organización Panamericana de la Salud.*
- *Brito C, Tenenbaum H, Wong B, Schmitt C, Nogueira-Filho G (2013) Is keratinized mucosa indispensable to maintain peri-implant health? A systematic review of the literatura*
- Brownlee J (2020) *kNN Imputation for Missing Values in Machine Learning.* Retrieved from https://machinelearningmastery.com/knn-imputation-for-missing-values-in-machine-learning/
- Chiu Y, Lee S, Lin Y, Lai Y (2015*) Significance of the width of keratinized mucosa on peri-implant health*
- Clinica Sicilia (2023) *¿Cuáles son los principales síntomas de la periimplantitis?* Retrieved from https://clinicasicilia.es/actualidad/periimplantitis-sintomas/
- Consejo dentistas (2021), *En España se colocan entre 1.2 y 1.4 millones de implantes al año*. Retrieved from https://consejodentistas.es/en-espana-se-colocan-entre-1-2-y-1-4-millones-de-implantes-al-ano/#:~:text=Madrid%2C%2028%20de%20octubre%20de%202021.&text=La%20media%20de%20dientes%20en,de%20implantes%20por%20p%C3%A9rdida%20dentaria.

- Craig M, Hattersley A, Donaghue K (2009) *Definition, epidemiology and classification of diabetes in children and adolescents 10 (Suppl. 12)*
- Dreyer H, Grischkel J, Tiede C, Eberhard J, Schweiter A, Toikkanen S, Glöckner S, Krause G, Stiesch M (2018*) Epidemiology and risk factors of peri-implantitis: A systematic review*
- Dukka H, Saleh M, Ravidà A, Greenwell H and Wang H (2021) *Is bleeding on probing a reliable clinical indicator of peri-implant diseases?*
- Figueredo R, Ruiz-Romero V, Toledano-Serrbona J, Abdelazim Y, Camps-Font O, Salazar-Salazar Y, Plana-Soler A, Subirà-Piarrafé C, Valmaseda-Castellón E (2024) *Peri implantitis in patients without regular supportive therapy: Prevalence and risk indicators*
- Galindo-Moreno P, León-Cano A, Ortega-Oller I, Monje A, Suárez F, O´Valle F, Spinato S, and Catena A (2014*) Prosthetic Abutment Height is a Key Factor in Peri-implant Marginal Bone Loss*
- Gharpure A, Latimer J, Aljofi F, Kahng J, Daubert D (2021), *Role of thin gingival phenotype and inadequate keratinized mucosa width (<2 mm) as risk indicators for peri-implantitis and peri-implant mucositis*
- Gracis, S., Michalakis, K., Vigolo, P., Vult von Steyern, P., Zwahlen, M. & Sailer, I. (2012) *Internal vs. external connections for abutments/reconstructions: a systematic review. Clinical Oral Implants Research 23: 202–216.*
- Gupta G (2021*) Implant Stability Quotient (ISQ): A Reliable Guide for Implant Treatment*
- IBM (2023) *What are SVMs?.* Retrieved from https://www.ibm.com/topics/support-vector-machine#:~:text=A%20support%20vector%20machine%20(SVM,in%20an%20N%2Ddimensional%20space.
- Janda M and Matheos N (2024*) Prosthetic design and choice of components for maintenance of optimal peri-implant health: a comprehensive review*
- Jefferson Dental & Orthodontics (JDC) (2022) *Implante dental vs corona: pros y contras.* Retrieved from https://www.jeffersondentalclinics.com/es/blog/implante-dental-vs-corona-pros-y-contras
- Ju H, Kim D, Kim S, Koo K, Kim T, Seol Y, Lee Y, Ku Y, Rhyu I (2018) *Comparison of marginal bone loss between internal- and external-connection dental implants in posterior areas without periodontal or peri-implant disease*
- Kan J, Runcharassaeng K, Umezu K, Kois J. (2003) *Dimensions of Peri-Implant Mucosa: An Evaluation of Maxillary Anterior Single Implants in Human*

- Katafuchi M, Weinstein B, Leroux B, Chen Y, Daubert D (2017*) Restoration contour is a risk indicator for peri-implantitis: A cross-sectional radiographic analysis*
- Kim Y and Ku J (2020) *Guided bone regeneration*
- Koo KT, Lee EJ, Kim JY, Seol YJ, Han JS, Kim TI, et al. (2012) *The effect of internal versus external abutment connection modes on crestal bone changes around dental implants: a radiographic analysis. J Periodontol. 2012;83:1104–1109*
- Kou Y, Li Q and Tang Z (2023) *Prosthetic emergence angle in different implant sites and their correlation with marginal bone loss: A retrospective study*
- Larsen P, Ghali GE (2004*). Peterson's Principals of Oral and Maxillofacial Surgery*
- Llerandi i Béjar, clínica dental (2024) *Implantes, puentes y sobredentadura*. Retrieved from https://www.llerandibejar.com/blog/implantes-dentales/implantes-puentes-y-sobredentadura/
- Mameno T, Wada M, Nozaki K, Takahashi T, Tsujioka Y, Akema S, Hasegawa D, Ikebe K (2021), *Predictive modeling for Peri-implantitis by using machine learning techniques*
- Medium (2020) Random Forest classification. Retrieved from https://medium.com/@pratishtha013/random-forest-classification-26d63cc74c3a
- Medium (2023) *Different Metrices in Machine Learning for Measuring performance of Classification Algorithms.* Retrieved from https://medium.com/@sachinsoni600517/different-metrices-in-machine-learning-for-measuring-performance-of-classification-algorithms-509e55c0a451
- Medium (2023) *Understanding Logistic Regression: Predicting Your Exam Outcomes. Retrieved* from https://medium.com/@glennlenormand/understanding-logistic-regression-predicting-your-exam-outcomes-6052e7705f48
- Misch, C.E., Perel, M.L., Wang, H.L., Sammartino, G., Galindo-Moreno, P., Trisi, P., Steigmann, M., Rebaudi, A., Palti, A., Pikos, M.A., Schwartz-Arad, D., Choukroun, J., Gutierrez-Perez, J.L., Marenzi, G. & Valavanis, D.K. (2008) *Implant success, survival, and failure: the international congress of oral implantologists (icoi) pisa consensus conference. Implant Dentistry 17: 5–15.*
- Moje A, Blasi G (2018) *Significance of keratinized mucosa/gingiva on peri-implant and adjacent periodontal conditions in erratic maintenance compliers*
- Monje A, Vera M, Muñoz-Sanz A, Wang H, Nart J (2020) *Suppuration as diagnostic criterium of peri-implantitis*
- Momen A, DclintDent, Shah M, Tawe-Smith A, et al. (2023) *Influence of implant restorative emergence angle and contour on peri-implant marginal bone loss: A systematic review and meta-analysis*

- Muñoz-Corcuera M, Bascones-Martínez A , Bascones-Iludain J (2015), *Diabetes and periodontitis: A bidirectional relationship Diabetes y periodontitis: una relación bidireccional*
- Pérez-Pevida D, Cherro R, Camps O, Piqué N (2020) *Effects of Drilling Protocol and Bone Density on the Stability of Implants According to Different Macrogeometries of the Implant Used: Results of an In Vitro Study*
- Ravidà A, Arena C, Tattan M, Troiano G, Caponio V… (2022), *The role of keratinized mucosa width as a risk factor for peri-implant disease: A systematic review, meta-analysis, and trial sequential analysis*
- Renvert S, Persson G, Pirih F, Camargo P (2018), *Peri-implant health, peri-implant mucositis, and peri-implantitis: Case definitions and diagnostic considerations*
- Schwarz F, Berglundh T, Chen S, Avila-Ortiz G, Blanco J, Renvert S, Lambert F…(2017) *Peri-implant diseases and conditions: Consensus report of workgroup 4 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions*
- Serino G, Ström C (2009) *Peri-implantitis in partially edentulous patients: association with inadequate plaque control*
- Yao K, Chen C, Cheng C, Fang H, Huang C, Kao H and Hsu M (2018*) Optimization of the Conical Angle Design in Conical Implant–Abutment Connections: A Pilot Study Based on the Finite Element Method*
- Zegarra P, Anyela M (2018), *Prevalencia de periimplantitis y mucositis en pacientes portadores de implantes en la clínica odontológica Medicdent, Arequipa 201*
- Zhu Y, Lu H, Yang S, Liu Y, Zhu Y, Li P, De Waal Y, Visser A, Tjakkes G, Li A, Xu S (2024) *Predictive Factors for the treatment success of peri-implantitis: a protocol for a prospective cohort study*

-

Appendix

Degree of relation ODS with the project

Objetivos de Desarrollo Sostenibles	High	Medium	Low	Not applicable
ODS 1. **No poverty**				X
ODS 2. **Zero hunger**				X
ODS 3. **Good health and well-being**	X			
ODS 4. **Quality education.**				X
ODS 5. **Gender equality**				X
ODS 6. **Clean wáter and sanitation**				X
ODS 7. **Affordable and clean energy**				X
ODS 8. **Decent work and economic growth**				X
ODS 9. **Industry, Innovation and infrastructure**	X			
ODS 10. **Reduced inequalities**				X
ODS 11. **Sustainable cities and communities**	X			
ODS 12. **Responsable consumption and production**				X
ODS 13. **Climate action**				X
ODS 14. **Life below water**				X
ODS 15. **Life on land**				X
ODS 16. **Peace, justice and strong institutions**				X
ODS 17. **Partnerships for the goals**	X			

This project contributes to SDG 3: Good Health and Well-being, by deepen into the dental health. Concretely, by creating predictive models for peri-implantitis, a critical issue in oral health. It is also important to adress periimplantitis which not only enhances dental health but also allows people who have this condition to get informed and be more cautious about hygiene or bad habits that could worsen this condition. It also targets indirectly to promote well-being for all ages, especially adults.

We can also consider it is focused towards SDG 9: Industry, Innovation, and Infrastructure, as it aims leverages cutting-edge machine learning techniques and healthcare technologies to enhance diagnostic accuracy and treatment efficacy for periimplantitis. By advancing dental care infrastructure and embracing technological innovations, we're laying the groundwork for more resilient healthcare systems and sustainable practices.

Moreover, by focusing on sustainable solutions and efficient implants in dental healthcare, it supports SDG 11's (Sustainable Cities and Communities) objectives to make cities and human settlements inclusive, safe, resilient, and sustainable.

This project also supports SDG 17: Partnerships for the Goals. Collaboration between different specialist was needed: odontologists from the Universidad de Barcelona and researchers were vital figures in order to get a full understanding on the matter and clear objectives. Knowleadge from diverse contexts and professions gave different perspectives.