

Predicting Cell Viability from Titanium Surface Properties Using Machine Learning-Based Decision Tree Analysis

Mateus Luiz Gamba^{1*} , Fabiano Rodrigues Fernandes¹ , Victor Velho de Castro¹ ,
Leandro Krug Wives² , Célia de Fraga Malfatti¹ 

¹Universidade Federal do Rio Grande do Sul (UFRGS), Departamento de Engenharia, Laboratório de Pesquisa em Corrosão (LAPEC), Porto Alegre, Brasil.

²Universidade Federal do Rio Grande do Sul (UFRGS), Programa de Pós-graduação em Computação (PPGC), UFRGS, Porto Alegre, Brasil

Received: February 12, 2025; Revised: August 21, 2025; Accepted: September 21, 2025

This paper introduces a computational tool designed to assist in classifying and predicting *in vitro* cellular activity using a dataset derived from the roughness, wettability, and surface morphology of titanium dioxide (TiO₂) and titanium (Ti) surfaces. Numerous studies compare TiO₂/Ti surface treatments to enhance osteoblast cellular activity; however, critical gaps remain in understanding how surface properties influence cellular responses. This research compiles a dataset based on peer-reviewed scientific articles published on academic platforms, focusing on surface characteristics: roughness, contact angle, presence of nanostructures such as nanotubes, and the percentage gain in cellular viability of MC3T3-E1 osteoblasts obtained from MTT assays (3-(4, 5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide), relative to control samples. Using this data, an Index of Classification of Increased Cell Activity (ICICA) was developed to categorize cellular responses into three levels: low, medium and high. Using the constructed dataset, decision tree algorithms were applied to develop a model capable of predicting cellular viability. Among the tested algorithms, the Random Forest model demonstrates superior performance regarding accuracy and Kappa coefficient. The developed model provides valuable insights to guide the design of new surface treatments regarding surface properties, such as roughness, wettability, and morphology of pure Ti, aiming to improve the cellular viability.

Keywords: Surface Treatments, Osseointegration, Titanium, Machine Learning, Decision Trees.

1. Introduction

Due to its biocompatibility and osseointegration properties, titanium (Ti) has been successfully used in clinical applications, including dental and orthopedic implants. The high success rate of titanium implants can be attributed to their strong physical connection with the bone surface, alongside their excellent mechanical properties, low modulus of elasticity, and outstanding corrosion resistance¹⁻³.

Surface properties of Ti implants, including roughness, chemical composition, and wettability, can significantly influence cellular interactions. These interactions affect key cellular activities such as adhesion and proliferation, which are critical for successful implant integration and function³.

Ti implants have become a cornerstone in orthopedic, dental, and reconstructive therapies due to their effectiveness and essential role in patient care⁴. However, implant failures still occur in certain cases, particularly when the patient's bone quality is compromised³. Additionally, with the growing demand for implant treatments driven by an aging population, there is increasing pressure to improve the success rate, expand the range of applications, and

enhance the osseointegration process⁵. Therefore, efforts are ongoing to address these challenges and improve the long-term outcomes of Ti implant therapies^{5,6}.

Several articles have been published aiming to refine the properties of implant surfaces and modify their structure and composition to enhance their interaction with bone tissue, thus increasing cellular activity⁷. In any case, the effect of surface properties is unclear because the studies do not present an experimental standardization, and many times, different types of cells are used, making it difficult to compare these studies. These compromises understanding surface properties in the osseointegration process and comparing different surface performances.

With the advent and evolution of Artificial Intelligence (AI) algorithms, machines are able to address many challenges in materials engineering and biomaterials design⁸. In particular, Machine Learning (ML) can significantly accelerate the discovery of new materials and the maturation of innovative biomaterials^{9,10}.

One of the key advantages of machine learning (ML) models is their ability to learn complex functions without requiring prior knowledge of the problem¹¹. Among the various ML techniques, decision trees are particularly popular for classification tasks^{12,13}. A decision tree aims to construct a tree-like structure that

*e-mail: mateusgamba@gmail.com

Associate Editor: Ana Sofia de Oliveira.

Editor-in-Chief: Luiz Antonio Pessan.

makes accurate predictions and serves as an interpretive tool for distinguishing between different classes^{13,14}. This learning process is supervised, meaning the model learns from labeled examples. The dataset is typically divided into two subsets: a training set, used to build the decision tree, and a test set, used to evaluate the model's performance¹³.

Once the model is trained, it can process new data and make predictions based on the patterns learned during training¹⁴.

Several recent studies have introduced computational tools that integrate ML algorithms into biomaterials development. For instance, Ganz et al.¹⁵ proposed a computational tool that employs multiple feature selection methods and classification algorithms to improve the accuracy of dental implant failure predictions in the province of Misiones, Argentina. The tool was validated by human experts, who used two datasets: a case study dataset of dental implants and an artificially generated dataset. The proposed approach achieved a failure prediction accuracy of 79%, outperforming individual classifiers that reached a maximum of 72%.

Similarly, Liu et al.¹⁶ evaluated the accuracy of an AI-based convolutional neural network for detecting marginal bone loss in periapical radiographs. The model was trained on 1670 radiographic images, divided into training (n = 1370), validation (n = 150), and test (n = 150) datasets. The system was assessed using metrics such as sensitivity, specificity, and misdiagnosis rate, demonstrating moderate to substantial agreement with expert evaluations, and indicating its potential for clinical use.

Furthermore, ML tools have been used to predict the properties of bone regeneration materials, including biocompatibility, mechanical properties, toxicity, antibacterial characteristics, degradability, and osteogenic and angiogenic potential¹⁷.

Although many studies have investigated the influence of surface characteristics on cellular behavior, there is still a need to understand better how surface properties of Ti/TiO₂-based materials relate to cellular activity and viability. In this context, AI, particularly ML techniques such as decision tree algorithms, offers powerful tools to model complex relationships and predict biological responses based on material characteristics.

This research compiles a dataset based on peer-reviewed scientific articles published on Science Direct, PubMed, and Wiley Online Library platforms, focusing on surface characteristics: roughness, contact angle (used to assess wettability), presence of nanostructures such as nanotubes, and the % Gain in cellular viability of MC3T3-E1 osteoblasts obtained from MTT assays (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide), relative to control samples. Using this data, an Index of Classification of Increased Cell Activity (ICICA) was developed to categorize cellular responses into three levels: low, medium, and high.

Based on the constructed dataset, decision tree algorithms were applied to create a predictive model capable of classifying and estimating *in vitro* cellular activity from surface property data. In addition to serving as a predictive tool, this model also aids in interpreting the relationship between surface properties and cellular responses, and as a predictive model, enabling the estimation of cellular activity levels in new datasets.

2. Materials and Methods

2.1. Selection criteria for surface properties influencing cellular viability

The cells used in the dataset articles were from the MC3T3-E1 (rat cells) lineage. This cell lineage has been widely used to evaluate the effects of different topographies on implant surfaces – they have an expansion rate similar to human osteoblastic cells^{18,19}. This makes them a reliable model for studying the biological responses to different surface treatments *in vitro*.

Building on this, the relationship between surface properties, such as roughness and wettability, and cellular responses underscores the complexity of biomaterial performance. Incorporating MTT results alongside parameters like roughness and contact angle enhances the predictive capacity of decision models, allowing for a clearer understanding of how surface modification drives cell activity and improves biological outcomes.

Rougher surfaces tend to exhibit lower contact angles and higher surface energy values. Materials with superior surface energy (hydrophilic surfaces) promote greater osteoblast differentiation compared to those with lower surface energy. Additionally, surfaces that are stronger electron acceptors stimulate osteoblastic differentiation¹.

The contact angle values effectively represent the interactions between biomaterial surfaces and different biological fluids. As a result, the contact angle significantly influences the cell fixation process and plays an essential role in enhancing cell activity.

Nanotubes can significantly enhance adherence and osteoblastic activity at the biomaterial-tissue interface, thereby contributing to increased bone formation. However, the biological performance of nanotubes can vary and is not fully understood, as their geometry can differ considerably²⁰.

The MTT test is fundamental for assessing cellular viability. However, inconsistencies in methodological approaches hinder the comparison of results across different research groups. This variability leads to unreliable data regarding the methods utilized for the test, as highlighted by the authors²¹.

The incubation time of the MTT assay influences the results by affecting the extent of formazan production and, consequently, the measured cellular viability. Including incubation time as an input parameter enhances the robustness of predictive models by accounting for variations in metabolic activity arising from different experimental protocols.

2.2. Dataset construction process

Articles were selected and critically analyzed based on the inclusion criteria established in the initial phase of the study. To identify relevant literature, comprehensive searches were conducted using key terms such as biomaterials, titanium, surface properties, cellular activity, and MTT assays. These searches were primarily performed across reputable scientific databases, including ScienceDirect, PubMed, and Wiley Online Library. While some pertinent studies from other sources may not have been captured initially, the dataset remains under continuous review and is periodically updated to incorporate newly identified and relevant publications.

For the construction of the dataset, 15 articles covering 39 samples were selected. Table 1 presents the articles used,

their labels, and a brief description of the surface treatments utilized in each.

The dataset construction followed three main steps described below:

Table 1. Summary of surface treatment information regarding the selected articles.

Reference	Label	Description
1	Pure Ti (Control)	Polished by SiC papers in successive grade 220 to 800 grit.
	TN	Polished by SiC papers in successive grade 220 to 800 grit + TiO ₂ nanotube surfaces fabricated in a fluorinated glycerol solution.
	TN-CTAB	Polished by SiC papers in successive grade 220 to 800 grit + TiO ₂ nanotube surfaces fabricated in a fluorinated glycerol solution + fluorinated glycerol solution with 1 wt% anionic surfactant sodium dodecyl sulphate + fluorinated glycerol solution with 1 wt% cationic surfactant cetyl trimethyl ammonium bromide.
	TN-SDS	Polished by SiC papers in successive grade 220 to 800 grit + TiO ₂ nanotube surfaces fabricated in a fluorinated glycerol solution + fluorinated glycerol solution with 1 wt% anionic surfactant sodium dodecyl sulphate.
2	Flat Ti (Control)	Ti.
	Ti-30nm	Nanotubes were prepared in a 1:7 volumetric ratio of acetic acid to 0.5% w/v hydrofluoric acid in water at 5, 10, 15, and 20 V for 30 min + nanotube diameter ~30 nm.
	Ti-50nm	Nanotubes were prepared in a 1:7 volumetric ratio of acetic acid to 0.5% w/v hydrofluoric acid in water at 5, 10, 15, and 20 V for 30 min + nanotube diameter ~50 nm.
	Ti-70nm	Nanotubes were prepared in a 1:7 volumetric ratio of acetic acid to 0.5% w/v hydrofluoric acid in water at 5, 10, 15, and 20 V for 30 min + nanotube diameter ~70 nm.
	Ti-100nm	Nanotubes were prepared in a 1:7 volumetric ratio of acetic acid to 0.5% w/v hydrofluoric acid in water at 5, 10, 15, and 20 V for 30 min + nanotube diameter ~100 nm.
22	as-polished (Control)	Polished with no. 1000 SiC abrasive paper.
	anodic oxidized	Polished with no. 1000 SiC abrasive paper + Anodic oxidation (using a direct current voltage source + anodic electrode + graphite used as cathodic electrode + 1 M NaF solution was used as electrolyte + process was carried out under a constant voltage of 10 V at room temperature for 1 h.
	UV irradiated	Polished with no. 1000 SiC abrasive paper + Anodic oxidation (using a direct current voltage source + anodic electrode + graphite used as cathodic electrode + 1 M NaF solution was used as electrolyte + process was carried out under a constant voltage of 10 V at room temperature for 1 h + irradiated by UV light (125 W high-pressure mercury lamp).
23	P (Control)	Polished with paper of #400-1200.
	PN	Polished with paper of #400-1200 + Nanotubes formed by anodization.
	PNA	Polished with paper of #400-1200 + Nanotubes formed by anodization + alkali treatment (immersed in 5 M NaOH aqueous solution maintained at 60 °C for 24 h)
	PNAH	Polished with paper of #400-1200 + Nanotubes formed by anodization + alkali treatment (immersed in 5 M NaOH aqueous solution maintained at 60°C for 24h) + heat treatment (heated up to 550°C at a rate of 5°C/min in the electric furnace for achieving a high temperature, maintained at this temperature for 1 h.
24	Ti sheet (Control)	Annealed.
	Annealed 150nm	Annealed + chemically etched for 1 min in the mixture of HNO ₃ and HF + a two-electrode preparation system with a DC power supply was utilized to conduct the anodization process. A platinum electrode served as the cathode. The electrolyte consisted of 0.5 wt.% NH ₄ F and 10 vol.% H ₂ O in ethylene glycol + anodized at 60 V, 90 V, 120 V, 150 V and 180 V for 3 h + annealed at 500 °C for 3 h + nanotube diameter 150 nm.
	Annealed 260nm	Annealed + chemically etched for 1 min in the mixture of HNO ₃ and HF + a two-electrode preparation system with a DC power supply was utilized to conduct the anodization process. A platinum electrode served as the cathode. The electrolyte consisted of 0.5 wt.% NH ₄ F and 10 vol.% H ₂ O in ethylene glycol + anodized at 60 V, 90 V, 120 V, 150 V and 180 V for 3 h + annealed at 500 °C for 3 h + nanotube diameter 260 nm.
	Annealed 360nm	Annealed + chemically etched for 1 min in the mixture of HNO ₃ and HF + a two-electrode preparation system with a DC power supply was utilized to conduct the anodization process. A platinum electrode served as the cathode. The electrolyte consisted of 0.5 wt.% NH ₄ F and 10 vol.% H ₂ O in ethylene glycol + anodized at 60 V, 90 V, 120 V, 150 V and 180 V for 3 h + annealed at 500 °C for 3 h + nanotube diameter 360 nm.
	Annealed 470nm	Annealed + chemically etched for 1 min in the mixture of HNO ₃ and HF + a two-electrode preparation system with a DC power supply was utilized to conduct the anodization process. A platinum electrode served as the cathode. The electrolyte consisted of 0.5 wt.% NH ₄ F and 10 vol.% H ₂ O in ethylene glycol + anodized at 60 V, 90 V, 120 V, 150 V and 180 V for 3 h + annealed at 500 °C for 3 h + nanotube diameter 470 nm.

Table 1. Continued...

Reference	Label	Description
25	NP (Control)	Sandblasted + acid etched.
	P2	Sandblasted + acid-etched + non-thermal atmospheric pressure plasma jet treatment 2 minutes.
	P10	Sandblasted + acid-etched + non-thermal atmospheric pressure plasma jet treatment 10 minutes.
26	Ti-control (Control)	Polished with 800# SiC paper.
	SLA	Polished with 800# SiC paper + sandblasted with large-grit corundum (200–500 µm) and then acid etched in 67% HCl/H ₂ SO ₄ (1:1) at 80 °C for 10 min.
	ASLA	Polished with 800# SiC paper + sandblasted with large-grit corundum (200–500 µm) and then acid etched in 67% HCl/H ₂ SO ₄ (1:1) at 80 °C for 10 min + alkali and heat treatment.
	HSLA	Polished with 800# SiC paper + sandblasted with large-grit corundum (200–500 µm) and then acid etched in 67% HCl/H ₂ SO ₄ (1:1) at 80 °C for 10 min + H ₂ O ₂ /HCl and heat treatment.
27	Ti-S (Control)	Ti untreated.
	Group I	Treated in a furnace at 300°C for 30 min.
	Group II	Treated at 500°C for 30 min.
	Group III	Treated at 750°C for 30 min.
28	RBM (Control)	Grit-blasted using resorbable blast media.
	P	Wet-abraded to 1200 grit SiC abrasive paper + hydrothermally treated using a 2% (w/w) H ₃ PO ₄ solution at 180 °C for 6 h in a Teflon-lined reactor.
	SrP	Wet-abraded to 1200 grit SiC abrasive paper + hydrothermally treated using a 2% (w/w) H ₃ PO ₄ solution at 180°C for 6 h in a Teflon-lined reactor + SrCl ₂ was added to 2% (w/w) H ₃ PO ₄ . To produce a crystalline phosphate- and strontium-incorporated oxide layer, NaOH was added as a mineralizer, and hydrothermal treatment was conducted at the same reaction conditions to produce a P surface.
29	Cp-Ti	Polished Ti.
	N-Ti	Cp-Ti treated directly with anodic oxidation.
	MN-Ti	N-Ti annealed at 450 °C for 2 h.
20	Ti	Commercial pure titanium.
	Anodized Ti	Anodized Ti-Cp coating was formed at 270 V for 60 seconds.
30	Ti	Comercial pure titanium.
	TN	nanotube arrays as an implant surface layer via anodization.
	AuNP-TN	Au nanoparticles were uniformly loaded into TN arrays by using a simple electrochemical deposition method.
31	ATi (Control)	Titanium treated with NaOH (hydroxylated) – control
	pS	ATi grafted with polysulfobetaine (superhydrophilic zwitterion)
	pB	ATi grafted with poly(butyl acrylate) (hydrophobic)
	pSB	ATi grafted with a copolymer of sulfobetaine and butyl acrylate (moderate hydrophilicity)
32	Ti (Control)	Pure titanium, only polished
	AH-Ti	Titanium treated with 5M NaOH at 80°C for 24 h (alkali-heat treatment)
	AH-Ti/Sr30	AH-Ti + SrTiO ₃ sputtering for 30 minutes
	AH-Ti/Sr90	AH-Ti + SrTiO ₃ sputtering for 90 minutes
	AH-Ti/Sr150	AH-Ti + SrTiO ₃ sputtering for 150 minutes
	Bare Ti (Control)	Uncoated titanium discs used as control samples.
	BG+CIPRO/PMMA	Titanium discs coated with a layer of PMMA and an additional layer containing bioglass and ciprofloxacin, applied using MAPLE technique.

2.2.1. Identification of key surface features

Initially, the main superficial characteristics influencing cellular activity and viability on TiO₂/Ti surfaces were identified. Different authors reported on the influence of roughness, wettability (via contact angle), and the presence of TiO₂ nanotubes characteristics on the cellular activity process^{1,2,20,22–26,28–30,33,34}.

2.2.2. Data extraction and structuring

Quantitative data were extracted from the selected articles and organized into a structured dataset. The following attributes were defined as columns:

- Nanotubes: Presence or absence of TiO₂ nanotube structures on the surface;

- Roughness: Average surface roughness (Ra), expressed in micrometers (µm), representing the micrometric topography of the material.
- Contact angles: Values measured using water droplets, representing the surface’s wettability characteristics;
- Incubation time: Duration, in days, of the incubation period applied during the MTT assay, reflecting its influence on formazan production and cellular viability measurements;
- MTT: Quantitative result from the MTT assay, reflecting the metabolic activity and viability of MC3T3-E1 osteoblasts after exposure to different surface treatments;

- Gain: percentage of gain in cellular viability of *MC3T3-E1* osteoblasts obtained from MTT assays (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide), relative to control samples;
- Index classification increased cell activity (ICICA): the output label of the dataset, with the classification of increased cell activity index. This attribute is obtained from using the *t*-distribution on the % increase in cellular activity compared with the control. Defined as low, medium, or high.

2.3. Data standardization and preprocessing

To ensure consistency and comparability across the dataset, standardization procedures were applied to key variables. The roughness values were standardized by converting all measurements to micrometric units. Additionally, the MTT assay results were normalized by calculating the percentage gain relative to each study's control sample. Finally, the Index Classification of Increased Cell Activity (ICICA) was established to convert the numerical % Gain into a nominal classification, enabling a more effective application of decision tree models.

The roughness involves interactions in micrometric and nanometric scales. However, the decision tree algorithm is a nonparametric model, and there are no presumed relations between the dependent and independent variables. The nanometric roughness values were converted to the micrometric measure unity so the algorithm could consider them the same dimension³⁵.

By comparing the collected articles, it is possible to observe that the MTT answers are similar. However, they do not follow a pattern like in the P2 sample²⁵ case (sandblasted, acid-etched, non-thermal atmospheric pressure plasma jet treatment, 2 minutes), which had an activity result of 143.34 in the MTT test. The analysis was conducted on the third day after the beginning of cultivation (the time when the cells were growing) and concluded as favorable by the author. The P sample²⁸ (hydrothermally treated) induced a 461.53 response in activity. The analysis that was performed with the same test and time and the author's conclusion were unfavorable.

The authors' lack of standardization regarding the data representation form (percentage of cell viability) and quantity of cells makes the comparative analysis much more complex. The suggested solution for this situation was to obtain the sample gain percentage of surface treatments over the control sample defined by the authors. Thus, it is possible to achieve a percentage gain comparison.

The percentage gain was defined by the following Equation (1):

$$\%Gain = \left[\frac{MTTresponse_{sample} * 100}{MTTresponse_{control}} \right] - 100 \quad (1)$$

To exemplify the % Gain calculation, two different sample groups were selected and compared with their respective controls, as shown in Table 2.

The first group, from Lee et al.²⁵, composed of the control sample NP and the treated sample P2, which exhibited a cell viability gain of 43.34% relative to sample NP. The second group, from Park et al.²⁸, consists of the control sample RBM and the treated sample P, presenting a gain

Table 2. MTT test gain comparison.

Samples	MTT assay	Cell viability (% Gain ^{a)})
NP (Control) ²⁵	100	-
P2 ²⁵	143.34	43.34
RBM (Control) ²⁸	434.61	-
P ²⁸	461.53	06.19

^{a)} Parameter defined in the Equation 1;

of 6.19% compared to RBM. The acronyms NP, P2, RBM, and P correspond to the sample labels as defined by the respective authors. Treatment details for each sample are summarized in Table 1.

An important consideration in this study was the duration of cell culture before performing the MTT test. The research conducted by Neupane et al.¹ (Table 1) measured cell viability on the culture's first, third, and fifth days. In contrast, another study²³ (Table 1) collected data on the second and fifth days.

Since the studies reported MTT responses at different incubation times, it was not possible to standardize all measurements to a single day. Therefore, an additional input column, "Incubation time" (in days), was introduced into the dataset to account for this variability.

Although the dataset was initially obtained from 15 articles comprising 39 samples, the inclusion of the "Incubation time" variable expanded the dataset to a total of 94 rows, as each sample could have multiple measurements corresponding to different incubation periods.

To make the evaluation criteria of cell activity more suitable for use with a decision tree and not depending on the authors' conclusions, the *t*-distribution was used to determine the ICICA from the % Gain.

This study used the *t*-distribution to find the confidence interval relative to the % Gain. The average value calculated from the 94 collected rows is 27.52, and its standard deviation is 34.17.

It is worth noting that the standard deviation of %Gain (34.17) is higher than the average %Gain (27.52), which indicates considerable variability in the data. This occurs because some treated samples exhibited lower MTT responses than their respective controls, resulting in negative %Gain values. Such variability is expected given the diverse treatment conditions, incubation times, and sample types across the collected studies.

A confidence level of 95% was used. It was used to establish a confidence interval for defining the ICICA, resulting in a range between 20.53% and 34.52%. Based on this, the samples were classified into the following cell activity levels:

- **Low:** when % Gain is less than 20.53%.
- **Medium:** when % Gain is between 20.53% and 34.52%.
- **High:** when % Gain exceeds 34.52%.

2.4. Data cleaning

In datasets, the impact of any noise point on model performance is particularly significant³⁶.

To enhance its quality, noise points were removed using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm⁷.

DBSCAN is a classical density-based clustering algorithm suitable for datasets that contain non-spherical clusters^{37,38}. It identifies high-density regions in the dataset as clusters, while low-density objects are categorized as noise^{37,38}. The algorithm utilizes two key parameters: the Neighborhood Radius (Eps) and the Minimum Number of Points Required in the Neighborhood (MinPts)⁷.

For this study, MinPts was defined as 3, meaning that a cluster will only be formed if it has at least three neighboring points. The Eps parameter was set to 0.4, representing the radius of the neighborhood. A total of 10 noise points were removed, resulting in 84 rows.

2.5. Dataset for model execution

After the dataset preparation process, which included the definition of input features and the response attribute, as well as data preprocessing, normalization, and noise reduction, the final dataset was structured in ARFF (Attribute-Relation File Format) for execution in the Waikato Environment for Knowledge Analysis (Weka). The complete dataset is presented in Table 3.

The negative values in the % Gain column can be attributed to the fact that the surface treatment does not enhance cell activity compared to the control sample.

Table 3 presents the % Gain to demonstrate its relationship with cell activity. However, this column was excluded from the dataset (Table 3) used by the decision tree models, since the result is represented by the ICICA column. The % Gain is only used in the calculation process to obtain ICICA, but it is not included as an input feature in the model.

2.6. Weka (Waikato Environment for Knowledge Analysis)

For this study, Weka was utilized to implement the decision tree algorithms. The software provided the necessary tools to apply various decision tree models, which were essential for analyzing the dataset and predicting cellular activity based on the surface characteristics. Weka's functionality allowed for an efficient execution of these algorithms, enabling the development of a robust model to categorize cell responses based on the defined input attributes³⁹.

For algorithm execution via its graphical interface, data must be provided in the Attribute-Relation File Format (ARFF), a structured text format that can also be generated through database queries.

2.7. Decision tree algorithms applied

In this study, four decision tree-based algorithms were applied: Random Forest, NBTree, J48, and SimpleCart. Each algorithm presents distinct characteristics in handling classification tasks.

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification tasks. Its robustness against overfitting and ability to handle high-dimensional data make it a widely used algorithm in biomedical research⁴⁰.

NBTree integrates decision trees with probabilistic reasoning by combining the Naive Bayes classifier at the leaves of the tree. This hybrid approach often enhances predictive accuracy, especially in datasets with interdependent attributes⁴¹.

The J48 algorithm classifies data through a hierarchical tree structure, identifying key attributes during the process. It uses entropy to measure data impurity, supporting both classification and feature selection. This makes J48 a valuable tool for predictive modeling⁴².

The SimpleCart algorithm builds predictive models by recursively splitting the dataset into binary branches, aiming to improve classification accuracy. It can generate either classification or regression trees, using measures like entropy to assess data impurity⁴³.

2.8. Evaluation of the decision tree classification

To estimate and examine the quality of the decision tree qualification, the tenfold method was used to ensure that the decision trees were capable of generalizing beyond the training data^{44,45}.

The assessment rating divides the dataset randomly into a training set and a test set. Nine parts of the dataset were used in the training set, while one part was reserved for the test set. This process was repeated 10 times, and the evaluation results were averaged to provide estimates. The decision to repeat the procedure 10 times was based on different studies in machine learning, which have confirmed that this number is adequate for reliable results^{44,45}.

The performance evaluation of a decision tree uses the following metrics^{44,45}:

- Accuracy: the percentage of correctly classified instances.
- Kappa: used to evaluate the precision of a classifier concerning known validation data. It provides a better precision measure of the classifiers than precision.

In addition to these metrics, the Weka software allows testing a trained model on other datasets, generating the *Predictions on test data* output. This output contains two key columns: *Predicted*, which shows the class label assigned by the model to each instance, and *Prediction*, which indicates the probability or confidence score associated with that classification.

3. Results and Discussions

To explore the predictive potential of machine learning in analyzing biomaterial surface properties, four decision tree algorithms were trained and validated using the Weka. The goal was to develop models capable of supporting the design of new surface treatments based on key attributes: Nanotubes, Roughness, Contact Angles, and ICICA.

Random Forest, NBTree, J48, and SimpleCart algorithms were applied, with their performance evaluated through the Kappa coefficient and classification accuracy. The comparative results are summarized in Table 4, ranking the models from highest to lowest accuracy.

The Kappa coefficient is used to assess the quality of the results, and its interpretation is presented in Table 5 below^{46,47}.

The quantitative classification of the Kappa coefficient, determined using Random Forest, stands at 43.97%, marking it as the best result achieved. According to the classification criteria, this value represents the upper limit of "Modest" agreement, narrowly missing the threshold for "Moderate" agreement, which begins at 41%^{46,47}.

Although it falls short of the "Moderate" category, the result suggests that the model achieves a reasonable level of

Table 3. Dataset with the used values.

Nanotubes	Roughness Ra (μm)	Contact angle ($^{\circ}$)	Incubation time (Days)	%Gain ^{b)}	ICICA
yes	0.22	22.40	1	18	low
yes	0.17	29.80	1	7	low
yes	0.13	32.30	1	4	low
yes	0.22	22.40	3	39	high
yes	0.17	29.80	3	15	low
yes	0.13	32.30	3	27	medium
yes	0.22	22.40	5	27	medium
yes	0.17	29.80	5	16	low
yes	0.13	32.30	5	16	low
yes	0.013	11.00	1	67	high
yes	0.013	9.00	1	33	high
yes	0.013	7.00	1	33	medium
yes	0.013	4.00	1	33	medium
yes	0.013	11.00	2	-11	low
yes	0.013	9.00	2	100	high
yes	0.013	7.00	2	67	high
yes	0.013	4.00	2	78	high
yes	0.2	32.00	1	18	medium
yes	0.20	32.00	4	56	high
yes	0.22	26.00	4	31	medium
yes	0.20	32.00	7	72	high
yes	0.22	26.00	7	88	high
yes	0.14	30.41	2	24	medium
yes	0.36	18.28	2	-89	low
yes	0.14	30.41	5	-11	low
yes	0.36	18.28	5	-21	low
yes	0.44	5.95	5	-17	low
yes	0.22	3.00	1	50	high
yes	0.26	2.00	1	63	high
yes	0.30	3.00	1	113	high
yes	0.25	7.00	1	100	high
yes	0.22	3.00	2	56	high
yes	0.26	2.00	2	33	high
yes	0.30	3.00	2	67	high
yes	0.30	7.00	2	111	high
yes	0.22	3.00	7	44	high
yes	0.26	2.00	7	50	high
yes	0.23	3.00	7	39	high
yes	0.25	7.00	7	83	high
no	1.88	0	3	43	high
no	1.88	0	3	15	low
no	1.88	0	7	90	high
no	1.88	0	7	79	high
no	2.85	82	1	2	low
no	2.63	5.00	1	4	low
no	2.5	5.00	1	6	low
no	2.85	82	3	-3	low
no	2.63	5.00	3	2	low
no	2.5	5.00	3	3	low
no	2.85	82	7	18	low
no	2.5	5.00	7	31	medium
no	0.18	63.00	1	-24	low
no	0.21	55.00	1	-14	low

^{b)} Column included for illustrative purposes only. The %Gain values are shown to facilitate interpretation of the ICICA; they were not used as input features in the training of the decision tree models.

Table 3. Continued...

Nanotubes	Roughness Ra (μm)	Contact angle (°)	Incubation time (Days)	%Gain ^{b)}	ICICA
no	1.18	8.20	1	33	high
no	1.07	8.00	1	63	high
no	1.18	8.20	2	14	low
no	1.07	8.00	2	29	medium
no	1.18	8.20	3	34	high
no	1.07	8.00	3	57	high
yes	1.67	51	1	10	low
yes	4.11	17	1	19	low
yes	1.67	51	4	11	low
yes	4.11	17	4	19	low
yes	1.67	51	7	27	medium
yes	4.11	17	7	35	high
no	0.28	65.75	1	10	low
no	0.28	65.75	3	7	low
yes	0.15	30.00	2	8	low
yes	0.48	10.80	4	38	high
no	0.13	15.50	1	29	medium
no	0.14	30.90	1	43	high
no	0.15	21.80	1	57	high
no	0.13	15.50	4	-5	low
no	0.14	30.90	4	7	low
no	0.15	21.80	4	9	low
no	0.13	15.50	7	-6	low
no	0.14	30.90	7	17	low
no	0.15	21.80	7	8	low
no	0.05	9.00	4	-28	low
no	0.045	105.00	4	-8	low
no	0.045	9.00	7	-17	low
no	0.040	105.00	7	-5	low
no	0.040	94.00	7	5	low
no	0.044	94.00	7	-14	low
no	0.30	25.80	1	-5	low

^{b)} Column included for illustrative purposes only. The %Gain values are shown to facilitate interpretation of the ICICA; they were not used as input features in the training of the decision tree models.

Table 4. Evaluation comparison regarding the decision trees algorithms.

Rank	Method	Accuracy	Kappa
1°	Random Forest	67.05%	43.97%
2°	J48	64.70%	37.13%
3°	SimpleCart	63.52%	35.61%
4°	NBTree	63.52%	32.3%

agreement, demonstrating potential for reliable predictive performance.

To enhance the clarity and reproducibility of the method, the performance metrics of the Random Forest model, which achieved the highest accuracy among the tested algorithms, are presented in Tables 6 and 7.

The performance metrics reported in Table 6 correspond to the default output generated by the Weka software for classification tasks. While measures such as Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error, and Root Relative Squared Error are traditionally applied in regression problems, Weka calculates them based on the probabilistic outputs of classification models. In our case,

Table 5. Classification quality associated to the values of the Kappa coefficient.

Kappa coefficient	Quality
≤0	Poor
10% - 20%	Negligible
21% - 40%	Modest
41% - 60%	Moderate
61% - 80%	Substantial
81% - 100%	Almost perfect

they are presented solely for completeness, as they were not used in model selection or interpretation.

Table 7 presents the key performance metrics for each classification: Precision (proportion of true positives among predicted positives), Recall (proportion of true positives among actual positives), and F-Measure (harmonic mean of precision and recall).

Additionally, a calibration curve was included to illustrate how the model's predicted probabilities compare with the actual outcomes. Figure 1 shows this visualization, providing further insight into the reliability and consistency of the Random Forest model's predictions. This complements the quantitative performance metrics and enhances the interpretability of the results.

The calibration curve shows that, since the lines for the classes are below the dashed reference line, the model is considered overconfident in its predictions, indicating a tendency to assign probabilities higher than those actually observed⁴⁸.

To further validate these results, additional articles not included in the dataset were utilized as proof of concept (Table 3).

Table 6. Performance metrics of the Random Forest model.

Metric	Value
Correctly Classified Instances	67.05%
Kappa Statistic	0.4397
Mean Absolute Error	0.2747
Root Mean Squared Error	0.3965
Relative Absolute Error	67.94%
Root Relative Squared Error	88.23%
Total Instances	84

Table 7. Detailed Accuracy by Class.

Classification	Precision	Recall	F-Measure
low	0.771	0.860	0.813
medium	0.200	0.154	0.174
high	0.667	0.621	0.643

Table 8. Input values for validation.

Reference	Sample	Nanotubes	Roughness Ra (μm)	Contact angle ($^\circ$)	Incubation time (Days)	Cell viability	ICICA
⁴⁹	TNT	yes	0.06282	5	3	-10.26	low
⁴⁹	TNT-Ta1	yes	0.109	5.6	3	-2.56	low
⁴⁹	TNT-Ta3	yes	0.135	6.8	3	5.13	low
⁴⁹	TNT-Ta5	yes	0.158	5.9	3	10.26	low
⁴⁹	TNT-Ta7	yes	0.179	6.6	3	7.69	low
⁴⁹	TNT	yes	0.06282	5	7	-1.37	low
⁴⁹	TNT-Ta1	yes	0.109	5.6	7	6.85	low
⁴⁹	TNT-Ta3	yes	0.135	6.8	7	9.59	low
⁴⁹	TNT-Ta5	yes	0.158	5.9	7	17.81	low
⁴⁹	TNT-Ta7	yes	0.179	6.6	7	12.33	low
⁵⁰	NT/CaP	yes	0.317	35.3	1	116.2	high
⁵⁰	NT/CaP+SrR	yes	0.317	15.4	1	116.9	high
⁵⁰	NT/SrR	yes	0.317	8.7	1	145.6	high
⁵⁰	NT/CaP	yes	0.317	35.3	2	396.5	high
⁵⁰	NT/CaP+SrR	yes	0.317	15.4	2	254.6	high
⁵⁰	NT/SrR	yes	0.317	8.7	2	330.2	high

The paper Zhan et al.⁴⁹ and Weitzel et al.⁵⁰ were used to validate the reliability of the decision tree. The relevant values for the input attributes are presented in Table 8.

As shown in Table 8, all treatments resulted in relatively low cell viability gains, ranging from -10.26% to a maximum of 396.5% compared to the control sample. Although the study concludes that the treatment demonstrates promising bioactivity, these viability gains remain modest when compared to the results reported in other studies.

This sample was applied to Random Forest, NBTree, J48, and SimpleCart algorithms to validate the results. The following results and predictions were obtained, as shown in Table 9.

In Table 9, it is possible to observe that each algorithm has two relevant columns: Predicted and Prediction. The Predicted column indicates the class label assigned by the algorithm to each instance, representing its classification decision. The Prediction column, on the other hand, shows the confidence or probability associated with that prediction, providing an estimate of how certain the algorithm is about its classification.

The results obtained by the algorithms do not fully align with those reported by Zhan et al.⁴⁹, as the percentage gain was relatively low compared to the samples in the training dataset. However, when considering the validation data from Weitzel et al.⁵⁰, some consistent patterns emerged. For instance, the NT/SrR (Day 1) sample was classified as *high* by Random Forest, SimpleCart, and NBTree, while the NT/SrR (Day 2) sample was also predicted as *high* by SimpleCart and NBTree. This indicates that, despite the overall lower gain, certain samples still supported the reliability of the decision tree in capturing biologically relevant variations.

Table 9. Results and predictions obtained using the validation dataset.

Sample	Random Forest		J48		SimpleCart		NBTree	
	Predicted	Prediction	Predicted	Prediction	Predicted	Prediction	Predicted	Prediction
TNT (Day 3)	high	0.44	high	0.85	high	0.85	high	0.704
TNT-Ta1 (Day 3)	high	0.4	high	0.85	high	0.85	high	0.704
TNT-Ta3 (Day 3)	high	0.55	high	0.85	high	0.85	high	0.704
TNT-Ta5 (Day 3)	high	0.51	high	0.85	high	0.85	high	0.704
TNT-Ta7 (Day 3)	high	0.59	high	0.85	high	0.85	high	0.704
TNT (Day 7)	high	0.5	high	0.85	high	0.85	high	0.86
TNT-Ta1 (Day 7)	high	0.47	high	0.85	high	0.85	high	0.86
TNT-Ta3 (Day 7)	high	0.62	high	0.85	high	0.85	high	0.86
TNT-Ta5 (Day 7)	high	0.55	high	0.85	high	0.85	high	0.86
TNT-Ta7 (Day 7)	high	0.67	high	0.85	high	0.85	high	0.86
NT/CaP (Day 1)	low	0.91	low	0.74	low	0.74	low	0.59
NT/CaP+SrR (Day 1)	low	0.8	low	0.74	low	0.74	low	0.59
NT/SrR (Day 1)	high	0.47	low	1	high	0.85	high	0.704
NT/CaP (Day 2)	low	0.79	low	0.74	low	0.74	low	0.59
NT/CaP+SrR (Day 2)	low	0.83	low	0.74	low	0.74	low	0.59
NT/SrR (Day 2)	low	0.46	low	1	high	0.85	high	0.704

Considering only the Random Forest algorithm, it is important to highlight that the Prediction values were low, indicating that the model assigned conservative probabilities to its predictions. This observation suggests that, although the calibration curve (Figure 1) indicated an overconfident behavior, as the lines were positioned below the reference line, in this specific case, Random Forest did not exhibit such behavior substantially. Therefore, it cannot be characterized as excessively confident, since its probabilistic outputs did not significantly overestimate the probability of correct classifications.

Decision tree algorithms are widely acknowledged for their effectiveness in predictive modeling within the realm of machine learning, especially for classification tasks. However, accurately predicting surface treatments designed to enhance cellular activity remains challenging. Although the Random Forest algorithm demonstrates superior accuracy and Kappa coefficients compared to SimpleCart and J48, the slight differences in prediction performance may be attributed to overfitting.

The Random Forest algorithm employs multiple decision trees, with each tree contributing a weighted classification⁵¹. This approach effectively handles categorical data and accommodates a large number of features. However, when applied to small datasets, it may result in overfitting and a struggle to generalize well to new data⁵².

In contrast, the J48 and SimpleCart algorithms are simpler models. These models typically exhibit lower variance, making them less susceptible to noise within the data⁵³. They generate a single decision tree that represents the entire dataset, making it easier to understand for humans.

On the other hand, Random Forest does not produce a single tree. Instead, it combines the results of multiple trees, improving both accuracy and robustness. Unlike SimpleCart, the predictions cannot be interpreted directly through a visual representation of the trees. To obtain predictions, it is necessary to input data into the model and use the tool to process the results.

J48, having achieved the second-highest accuracy, just after Random Forest, can produce a single, interpretable decision tree. This characteristic allows direct visualization

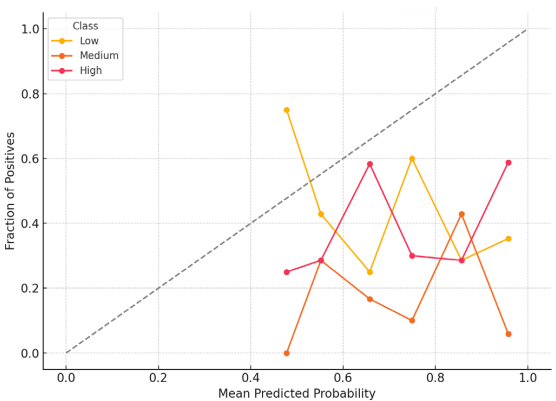


Figure 1. Calibration curve of the Random Forest model.

of how feature values contribute to the classification outcomes, providing a practical tool for interpreting model behavior. As shown in Figure 2, this representation supports the identification of property values associated with low, medium, or high ICICA levels, the interpretability of the tree may assist in the planning and refinement of future experimental designs.

In the presented tree, the paths that resulted in the “high” outcome occurred when the Contact angle is less than or equal to 11, Roughness is less than or equal to 1.88, and Nanotubes are present. Within this condition, if the Roughness is also less than or equal to 0.28, the result is “high”. Additionally, when Nanotubes are absent, if the Roughness is greater than 0.477, the result is also “high”.

As previously mentioned, the roughness of an implant surface is related to the contact angle (superficial wettability) and plays a crucial role in cell adhesion and tissue integration. This roughness influences the healing period and affects the activity of cells interacting with the implant, ultimately impacting the growth of bone cells⁵⁴.

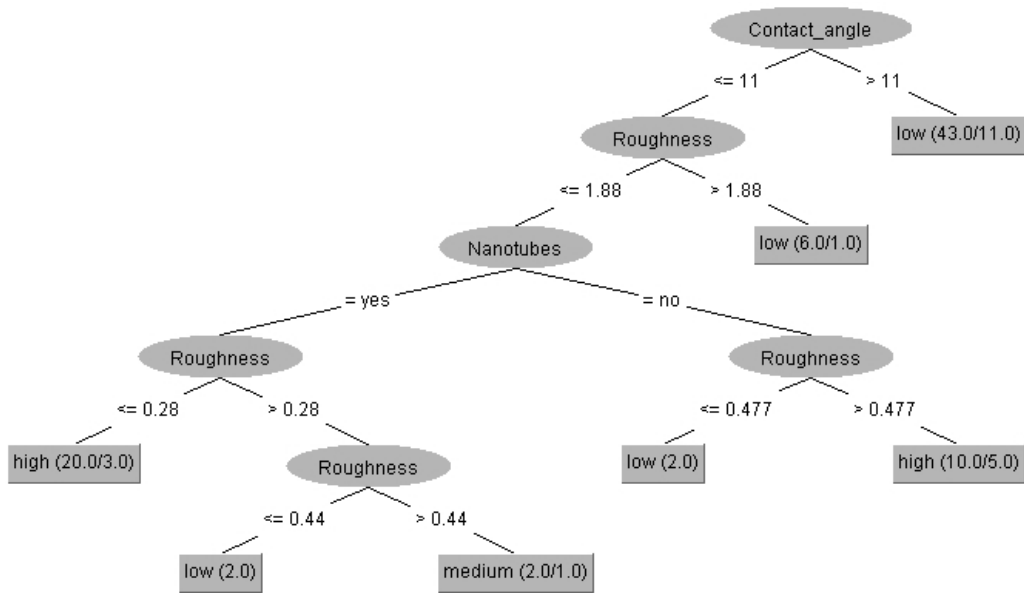


Figure 2. Decision tree generated by J48 algorithm.

This study examines the properties of roughness and contact angle; however, it is essential to recognize that material composition also plays a crucial role in influencing cellular activity^{1,23,26,28,33}. Surface treatments can modify the chemical composition of materials, directly affecting how cells interact with them.

The Kappa coefficient of 43.97% obtained in this research highlights the necessity for further analysis to achieve more accurate results. This may involve gathering additional articles that align with the defined dataset or utilizing alternative classification methods for knowledge extraction.

Surface treatment studies are concerned with possible alterations in cell metabolism and, consequently, alterations in adherence, growth, proliferation, and cell activity. The proposal was to build a data model to predict answers independently whether the model is close to reality or not¹².

Surface treatments present significant research challenges that extend beyond medicine and materials engineering. They also impose considerable demands on computational tools and mathematical modeling.

One of the primary difficulties lies in gathering a substantial amount of data for the dataset (Table 3) from published articles poses challenges due to the lack of standardization in the timing of analyses, data representation (percentage of cell viability), and cell quantities among various authors concerning MTT test outcomes.

These challenges highlight how each characteristic of these treatments is governed by variables that can be influenced by a range of factors. As a result, the field necessitates the development of advanced computational methods that integrate critical, meaningful guidelines to better capture and predict these complex interactions¹².

4. Conclusions

- The results, based on surface characterization of TiO₂/Ti substrates using osteoblastic MC3T3-E1

cells, demonstrate that this approach is effective in identifying surface regions that enhance cellular activity. Consequently, this enables the design of implant surfaces with tailored roughness, wettability, and morphology suited to specific biomedical applications.

- Among the algorithms evaluated, the Random Forest model exhibited the highest performance, achieving superior accuracy and Kappa coefficient values compared to other models.
- Expanding the dataset could further enhance algorithm performance and potentially influence the predicted outcomes. The Random Forest model achieved the highest predictive accuracy, despite the limited dataset.
- While the developed model demonstrates significant potential, several limitations must be acknowledged. These include the relatively small and homogeneous dataset, which may constrain the generalizability of the predictive outcomes. Furthermore, improved standardization of the assays used to characterize the TiO₂/Ti surfaces would enhance the reproducibility and reliability of the findings.
- A classification model based on decision tree analysis was developed to predict cellular activity using nanotube structure, contact angle, surface roughness, and incubation time as input features. This model offers a valuable foundation for guiding future research into the development of advanced surface modifications for biomedical implant applications.

5. Acknowledgments

The authors are grateful for the financial support of CAPES (PROEX 88881.844968/2023-01), V. V. de Castro and C. F. Malfatti thank the National Council for Scientific and Technological Development CNPq (Grant 171719/2023-9 and Grant 313493/2023-5).

6. References

- Neupane MP, Park IS, Bae TS, Yi HK, Watari F, Lee MH. Biocompatibility of TiO₂ 2 nanotubes fabricated on Ti using different surfactant additives in electrolyte. *Mater Chem Phys*. 2012;134(1):536-41. <http://doi.org/10.1016/j.matchemphys.2012.03.029>.
- Brammer KS, Oh S, Cobb CJ, Bjursten LM, Heyde H, Jin S. Improved bone-forming functionality on diameter-controlled TiO₂ nanotube surface. *Acta Biomater*. 2009;5(8):3215-23. <http://doi.org/10.1016/j.actbio.2009.05.008>.
- Deng Z, Ma J, Yin B, Li W, Liu J, Yang J, et al. Surface characteristics of and in vitro behavior of osteoblast-like cells on titanium with nanotopography prepared by high-energy shot peening. *Int J Nanomedicine*. 2014;5565. <http://doi.org/10.2147/IJN.S71625>.
- Vasconcellos LMRD, Oliveira MVD, Graça MLDA, Vasconcellos LGOD, Carvalho YR, Cairo CAA. Porous titanium scaffolds produced by powder metallurgy for biomedical applications. *Mater Res*. 2008;11(3):275-80. <http://doi.org/10.1590/S1516-14392008000300008>.
- Aita H, Hori N, Takeuchi M, Suzuki T, Yamada M, Anpo M, et al. The effect of ultraviolet functionalization of titanium on integration with bone. *Biomaterials*. 2009;30(6):1015-25. <http://doi.org/10.1016/j.biomaterials.2008.11.004>.
- Araújo TG, Moreira CS, Neme RA, Luan H, Bertolini M. Long-term implant maintenance: a systematic review of home and professional care strategies in supportive implant therapy. *Braz Dent J*. 2024;35:e24-6178. <http://doi.org/10.1590/0103-6440202406178>.
- Jiang P, Lin L, Zhang F, Dong X, Ren L, Lin C. Electrochemical construction of micro-nano spongelike structure on titanium substrate for enhancing corrosion resistance and bioactivity. *Electrochim Acta*. 2013;107:16-25. <http://doi.org/10.1016/j.electacta.2013.05.120>.
- Goswami L, Deka MK, Roy M. Artificial intelligence in material engineering: a review on applications of artificial intelligence in material engineering. *Adv Eng Mater*. 2023;25(13):2300104. <http://doi.org/10.1002/adem.202300104>.
- Suwardi A, Wang FK, Xue K, Han M-Y, Teo P, Wang P, et al. Machine learning-driven biomaterials evolution. *Adv Mater*. 2022;34(1):2102703. <http://doi.org/10.1002/adma.202102703>.
- Costache AD, Ghosh J, Knight DD, Kohn J. Computational methods for the development of polymeric biomaterials. *Adv Eng Mater*. 2010;12(1-2). <http://doi.org/10.1002/adem.200980020>.
- Choi AH. Bone remodeling and osseointegration of implants. Singapore: Springer; 2023. Artificial intelligence, machine learning, and neural network; p. 83-96. http://doi.org/10.1007/978-981-99-1425-8_7.
- Czajkowski M, Grześ M, Kretowski M. Multi-test decision tree and its application to microarray data classification. *Artif Intell Med*. 2014;61(1):35-44. <http://doi.org/10.1016/j.artmed.2014.01.005>.
- Ali MM, Paul BK, Ahmed K, Bui FM, Quinn JMW, Moni MA. Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. *Comput Biol Med*. 2021;136:104672. <http://doi.org/10.1016/j.combiomed.2021.104672>.
- Aviad B, Roy G. Classification by clustering decision tree-like classifier based on adjusted clusters. *Expert Syst Appl*. 2011;38(7):8220-8. <http://doi.org/10.1016/j.eswa.2011.01.001>.
- Ganz NB, Ares AE, Kuna HD. Procedure to improve the accuracy of dental implant failures by data science techniques. *J Comput Sci Technol*. 2021;21(2):e13. <http://doi.org/10.24215/16666038.21.e13>.
- Liu M, Wang S, Chen H, Liu Y. A pilot study of a deep learning approach to detect marginal bone loss around implants. *BMC Oral Health*. 2022;22(1):11. <http://doi.org/10.1186/s12903-021-02035-8>.
- Fan J, Xu J, Wen X, Sun L, Xiu Y, Zhang Z, et al. The future of bone regeneration: artificial intelligence in biomaterials discovery. *Mater Today Commun*. 2024;40:109982. <http://doi.org/10.1016/j.mtcomm.2024.109982>.
- Bächle M, Kohal RJ. A systematic review of the influence of different titanium surfaces on proliferation, differentiation and protein synthesis of osteoblast-like MG63 cells. *Clin Oral Implants Res*. 2004;15(6):683-92. <http://doi.org/10.1111/j.1600-0501.2004.01054.x>.
- Czekanska E, Stoddart MJ, Richards RG, Hayes JS. In search of an osteoblast cell model for in vitro research. *Eur Cell Mater*. 2012;24:1-17. <http://doi.org/10.22203/eCM.v024a01>.
- Kim K, Lee B-A, Piao X-H, Chung H-J, Kim Y-J. Surface characteristics and bioactivity of an anodized titanium surface. *J Periodontal Implant Sci*. 2013;43(4):198. <http://doi.org/10.5051/jpis.2013.43.4.198>.
- Wang H, Cheng H, Wang F, Wei D, Wang X. An improved 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) reduction assay for evaluating the viability of *Escherichia coli* cells. *J Microbiol Methods*. 2010;82(3):330-3. <http://doi.org/10.1016/j.jmimet.2010.06.014>.
- Li B, Li Y, Li J, Fu X, Li C, Wang H, et al. Improvement of biological properties of titanium by anodic oxidation and ultraviolet irradiation. *Appl Surf Sci*. 2014;307:202-8. <http://doi.org/10.1016/j.apsusc.2014.04.015>.
- Kim SY, Kim Y, Park I, Jin G, Bae T, Lee M. Effect of alkali and heat treatments for bioactivity of TiO₂ nanotubes. *Appl Surf Sci*. 2014;321:412-9. <http://doi.org/10.1016/j.apsusc.2014.09.177>.
- Zhang R, Wu H, Ni J, Zhao C, Chen Y, Zheng C, et al. Guided proliferation and bone-forming functionality on highly ordered large diameter TiO₂ nanotube arrays. *Mater Sci Eng C*. 2015;53:272-9. <http://doi.org/10.1016/j.msec.2015.04.046>.
- Lee EJ, Kwon J-S, Uhm S-H, Song D-H, Kim YH, Choi EH, et al. The effects of non-thermal atmospheric pressure plasma jet on cellular activity at SLA-treated titanium surfaces. *Curr Appl Phys*. 2013;13:S36-41. <http://doi.org/10.1016/j.cap.2012.12.023>.
- Zhang EW, Wang YB, Shuai KG, Gao F, Bai YJ, Cheng Y, et al. In vitro and in vivo evaluation of SLA titanium surfaces with further alkali or hydrogen peroxide and heat treatment. *Biomed Mater*. 2011;6(2):025001. <http://doi.org/10.1088/1748-6041/6/2/025001>.
- Lee YJ, Cui D-Z, Jeon H-R, Chung H-J, Park Y-J, Kim O-S, et al. Surface characteristics of thermally treated titanium surfaces. *J Periodontal Implant Sci*. 2012;42(3):81. <http://doi.org/10.5051/jpis.2012.42.3.81>.
- Park JW, Kim Y-J, Jang J-H. Enhanced osteoblast response to hydrophilic strontium and/or phosphate ions-incorporated titanium oxide surfaces. *Clin Oral Implants Res*. 2010;21(4):398-408. <http://doi.org/10.1111/j.1600-0501.2009.01863.x>.
- Wang G, Wan Y, Ren B, Liu Z. Bioactivity of micropatterned TiO₂ nanotubes fabricated by micro-milling and anodic oxidation. *Mater Sci Eng C*. 2019;95:114-21. <http://doi.org/10.1016/j.msec.2018.10.068>.
- Wang C, Bai Y, Bai Y, Gao J, Ma W. Enhancement of corrosion resistance and bioactivity of titanium by Au nanoparticle-loaded TiO₂ nanotube layer. *Surf Coat Tech*. 2016;286:327-34. <http://doi.org/10.1016/j.surfcoat.2015.12.051>.
- Zhu B, Jia E, Zhang Q, Zhang Y, Zhou H, Tan Y, et al. Titanium surface-grafted zwitterionic polymers with an anti-polyelectrolyte effect enhances osteogenesis. *Colloids Surf B Biointerfaces*. 2023;226:113293. <http://doi.org/10.1016/j.colsurf.2023.113293>.
- Negut I, Ristoscu C, Tozar T, Dinu M, Parau AC, Grumezescu V, et al. Implant surfaces containing bioglasses and ciprofloxacin as platforms for bone repair and improved resistance to microbial colonization. *Pharmaceutics*. 2022;14(6):1175. <http://doi.org/10.3390/pharmaceutics14061175>.
- Zhou R, Wei D, Yang H, Feng W, Cheng S, Li B, et al. MC3T3-E1 cell response of amorphous phase/TiO₂ nanocrystal composite coating prepared by microarc oxidation on titanium. *Mater Sci Eng C*. 2014;39:186-95. <http://doi.org/10.1016/j.msec.2014.03.006>.
- Zhan J, Li L, Yao L, Cao Z, Lou W, Zhang J, et al. Evaluation of sustained drug release performance and osteoinduction of magnetron-sputtered tantalum-coated titanium dioxide nanotubes. *RSC Advances*. 2024;14(6):3698-711. <http://doi.org/10.1039/D3RA08769G>.

35. Kashani AT, Mohaymany AS. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. *Saf Sci*. 2011;49(10):1314-20. <http://doi.org/10.1016/j.ssci.2011.04.019>.
36. Pi Q, Li R, Han B, Yang K, Hu Y, Shi Y, et al. Predicting the porosity of as-built additive manufactured samples based on machine learning method for small datasets. *Opt Laser Technol*. 2024;177:111203. <http://doi.org/10.1016/j.optlastec.2024.111203>.
37. Cheng D, Zhang C, Li Y, Xia S, Wang G, Huang J, et al. GB-DBSCAN: A fast granular-ball based DBSCAN clustering algorithm. *Inf Sci*. 2024;674:120731. <http://doi.org/10.1016/j.ins.2024.120731>.
38. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Simoudis E, Han J, Fayyad U, editors. *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Washington: AAAI Press; 1996. p. 226-23.
39. Witten IH, Frank E, Hall MA. *Data mining: practical machine learning tools and techniques*. Amsterdam: Elsevier; 2011. Chapter 10, Introduction to Weka; p. 403-406. <http://doi.org/10.1016/B978-0-12-374856-0.00010-9>.
40. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. <http://doi.org/10.1023/A:1010933404324>.
41. Zhou X, Xu X, Zhang J, Wang L, Wang D, Zhang P. Fault diagnosis of silage harvester based on a modified random forest. *Inf Process Agric*. 2023;10(3):301-11. <http://doi.org/10.1016/j.inpa.2022.02.005>.
42. Tambake N, Deshmukh B, Patange A. Development of a low cost data acquisition system and training of J48 algorithm for classifying faults in cutting tool. *Mater Today Proc*. 2023;72:1061-7. <http://doi.org/10.1016/j.matpr.2022.09.163>.
43. Mangal N, Ramesh Nachiappan M, Elangovan M, Sugumaran V. Fault diagnosis of a single point cutting tool using statistical features by simple CART classifier. *Indian J Sci Technol*. 2016;9(33). <http://doi.org/10.17485/ijst/2016/v9i33/101339>.
44. Mestizo Gutiérrez SL, Herrera Rivero M, Cruz Ramírez N, Hernández E, Aranda-Abreu GE. Decision trees for the analysis of genes involved in Alzheimer's disease pathology. *J Theor Biol*. 2014;357:21-5. <http://doi.org/10.1016/j.jtbi.2014.05.002>.
45. Kirchner K, Tölle K-H, Krieter J. Optimisation of the decision tree technique applied to simulated sow herd datasets. *Comput Electron Agric*. 2006;50(1):15-24. <http://doi.org/10.1016/j.compag.2005.07.002>.
46. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159. <http://doi.org/10.2307/2529310>.
47. Saha S, Sarkar D, Mondal P. Assessing and mapping soil erosion risk zone in Ratlam District, central India. *Regional Sustainability*. 2022;3(4):373-90. <http://doi.org/10.1016/j.regsus.2022.11.005>.
48. Guo C, Pleiss G, Sun Y, Weinberger KQ. On Calibration of Modern Neural Networks. In: Precup D, The YW, editors. *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*. New York: JMLR.org; 2017. p. 1321-1330.
49. Zhan J, Li L, Yao L, Cao Z, Lou W, Zhang J, et al. Evaluation of sustained drug release performance and osteoinduction of magnetron-sputtered tantalum-coated titanium dioxide nanotubes. *RSC Advances*. 2024;14(6):3698-711. <http://doi.org/10.1039/D3RA08769G>.
50. Weitzel APR, Almeida TC, Mendonça R, Camarano DM, Azzi PC, Vieira GM, et al. Chemical modification of nanotubular Ti surfaces with calcium phosphate and strontium ranelate for biomedical applications. *Mater Chem Phys*. 2024;316:129122. <http://doi.org/10.1016/j.matchemphys.2024.129122>.
51. Tahsin MS, Abdullah S, Al Karim M, Ahmed MU, Tafannum F, Ara MY. A comparative study on data mining models for weather forecasting: a case study on Chittagong, Bangladesh. *Nat Hazards Rev*. 2024;4(2):295-303. <http://doi.org/10.1016/j.nhres.2023.12.014>.
52. Luan J, Zhang C, Xu B, Xue Y, Ren Y. The predictive performances of random forest models with limited sample size and different species traits. *Fish Res*. 2020;227:105534. <http://doi.org/10.1016/j.fishres.2020.105534>.
53. Klusowski J. Sparse learning with CART. *Adv Neural Inf Process Syst*. 2020;33:11612-22.
54. Györgyey Á, Ungvári K, Kecskeméti G, Kopniczky J, Hopp B, Oszkó A, et al. Attachment and proliferation of human osteoblast-like cells (MG-63) on laser-ablated titanium implant material. *Mater Sci Eng C*. 2013;33(7):4251-9. <http://doi.org/10.1016/j.msec.2013.06.020>.

Data Availability

The dataset was presented in the manuscript.