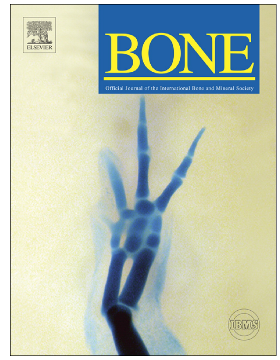
Accepted Manuscript




# Machine learning to predict the occurrence of bisphosphonate-related osteonecrosis of the jaw associated with dental extraction: A preliminary report

Dong Wook Kim, Hwiyoung Kim, Woong Nam, Hyung Jun Kim, In-Ho Cha

PII: S8756-3282(18)30174-1
DOI: doi:10.1016/j.bone.2018.04.020
Reference: BON 11631

To appear in: *Bone*

Received date: 4 January 2018
Revised date: 8 April 2018
Accepted date: 23 April 2018

Please cite this article as: Dong Wook Kim, Hwiyoung Kim, Woong Nam, Hyung Jun Kim, In-Ho Cha , Machine learning to predict the occurrence of bisphosphonate-related osteonecrosis of the jaw associated with dental extraction: A preliminary report. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Bon(2017), doi:10.1016/j.bone.2018.04.020

**Title Page**

**Article Type:** Original article

**Title:** Machine learning to predict the occurrence of bisphosphonate-related osteonecrosis of the jaw associated with dental extraction: a preliminary report

**Authors:** Dong Wook Kim[1], Hwiyoung Kim[2], Woong Nam[1,3], Hyung Jun Kim[1,3], In-Ho Cha[1,3,*]

**Affiliations:**
[1] Department of Oral & Maxillofacial Surgery, Yonsei University College of Dentistry, 50-1 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea
[2] Artificial Intelligence Research Institute, 22, Daewangpangyo-ro 712beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, 13488, Republic of Korea
[3] Oral Cancer Research Institute, Yonsei University College of Dentistry, 50-1 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea

**Authors' e-mail addresses:**
Dong Wook Kim: DWKIM617@gmail.com
Hwiyoung Kim: ASTARIA82@gmail.com
Woong Nam: OMSNAM@yuhs.ac
Hyung Jun Kim: KIMOMS@yuhs.ac
In-Ho Cha: CHA8764@yuhs.ac

**Corresponding author:**
Prof. In-Ho Cha, DDS, PhD
Department of Oral and Maxillofacial Surgery, Yonsei University College of Dentistry,
50-1 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea.
Tel.: +82 2 2228 3140, Fax: +82 2 2227 7825.
E-mail address: CHA8764@yuhs.ac

**Abstract word count**: 234 words

**Total word count**: 3685 words

**Total number of figures/tables:** 5 figures and 3 tables.

**Total number of supplementary materials:** 1 .pdf file, 2 .csv files.

**Number of references:** 29 papers

## ABSTRACT

***Introduction***: The aim of this study was to build and validate five types of machine learning models that can predict the occurrence of BRONJ associated with dental extraction in patients taking bisphosphonates for the management of osteoporosis.

***Patients & Methods***: A retrospective review of the medical records was conducted to obtain cases and controls for the study. Total 125 patients consisting of 41 cases and 84 controls were selected for the study. Five machine learning prediction algorithms including multivariable logistic regression model, decision tree, support vector machine, artificial neural network, and random forest were implemented. The outputs of these models were compared with each other and also with conventional methods, such as serum CTX level. Area under the receiver operating characteristic (ROC) curve (AUC) was used to compare the results.

***Results***: The performance of machine learning models was significantly superior to conventional statistical methods and single predictors. The random forest model yielded the best performance (AUC = 0.973), followed by artificial neural network (AUC = 0.915), support vector machine (AUC = 0.882), logistic regression (AUC = 0.844), decision tree (AUC = 0.821), drug holiday alone (AUC = 0.810), and CTX level alone (AUC = 0.630).

***Conclusions***: Machine learning methods showed superior performance in predicting BRONJ associated with dental extraction compared to conventional statistical methods using drug holiday and serum CTX level. Machine learning can thus be applied in a wide range of clinical studies.


**Keywords**: Osteoporosis; Osteonecrosis; Bisphosphonate-Associated Osteonecrosis of the Jaw; Artificial Intelligence; Machine Learning; Random Forest

# 1 INTRODUCTION

With the recent advent of Google's AlphaGo and IBM's Watson for Oncology, artificial intelligence and machine learning have begun to take center stage among emerging technologies. Machine learning is a branch of artificial intelligence that employs a variety of statistical and optimization techniques that allow computers to "learn" from past examples and to detect hard-to-discern patterns and relationships from complex data sets.[1] Due to its capability, machine learning has drawn interest as a tool for medical research.[1, 2] Machine learning is not new to medical research; according to the latest PubMed statistics, more than 16,000 papers have been published on the subject of machine learning in medicine, a majority of those concerned with detecting and classifying diseases.[1]

Bisphosphonates are commonly prescribed for the management of osteoporosis.[3, 4] They have been reported to reduce both vertebral and nonvertebral fractures in patients with osteoporosis.[3, 5] The wide usage of bisphosphonates in clinical practice has led to some reports of adverse effects, including osteonecrosis of the jaw (ONJ).[3, 6] Since the first case series reported by Ruggiero et al. 2004, various studies have reported on the pathophysiology, risk factors, and management of bisphosphonate-related osteonecrosis of the jaw (BRONJ).[6]

As tooth extraction is known to be one of the risk factors of BRONJ, its avoidance is recommended, as is conservative dental management.[7-9] However, in some situations, tooth extractions are immediately necessary due to discomfort in daily life, possibility of more severe infectious complications, or in some emergency. Discontinuation of bisphosphonate therapy, a so-called "drug holiday," is recommended before invasive dental surgery, including tooth extraction, to lower the risk of BRONJ.[8] Nevertheless, the optimal duration of the drug holiday before dental treatment is controversial. Some studies suggest that the drug holiday is unnecessary and that dental extraction can be done without resulting in BRONJ, [4, 10] while other studies recommend a 3-month drug holiday before dental extraction, suggesting that the longer the discontinuation, the lower the risk of BRONJ.[8, 11, 12]

Accurate prediction of a disease outcome is one of the most interesting and challenging tasks for clinicians.[2] Though machine learning has gained wide use in medical research, attempts to apply it in the prediction and prognosis of a disease are relatively recent, the majority involving detection and classification.[1] There are very few published studies applying machine learning to osteoporosis, and none applying it to BRONJ, to our knowledge.

The aim of this study was to build and validate five types of machine learning models designed to predict the occurrence of BRONJ associated with dental extraction in patients taking bisphosphonates for the management of osteoporosis. The performance of machine learning models were compared along with single predictors such as serum CTX level.

# 2 MATERIAL & METHODS

## 2.1 Patients

This study included patients who were referred and treated in the Department of Oral & Maxillofacial Surgery, Yonsei University College of Dentistry, between January 2009 and March 2017. A retrospective review of the medical records was conducted to obtain cases and controls.

Patients under current or previous bisphosphonate treatment for management of osteoporosis who underwent dental extraction were included.

Patients who had received bisphosphonates as part of the cancer treatment protocol were excluded from the study. Patients with history of radiation therapy to the head and neck region were excluded. The BRONJ cases not associated with dental extraction were excluded. Patients with an incomplete medical record for analysis also had to be excluded.

BRONJ was diagnosed as defined by an AAOMS position paper, as “Current or previous treatment with bisphosphonates, exposed bone or bone that can be probed through intraoral or extraoral fistula in the maxillofacial region that has persisted for longer than eight weeks, without history of radiation therapy to the jaws.”[8, 13] Staging of BRONJ was also made according to the AAOMS position papers, as follows:[8, 13]

- Stage I: exposed necrotic bone in asymptomatic patients without evidence of infection
- Stage II: exposed necrotic bone with pain and/or signs of infection
- Stage III: exposed necrotic bone with pain and/or signs of infection and one or more of the following: exposed and necrotic bone extending beyond the region of alveolar bone, (i.e., inferior border and ramus in the mandible, maxillary sinus and zygoma in the maxilla) resulting in pathologic fracture, extraoral fistula, oral antral/oral nasal communication, or osteolysis extending to the inferior border of the mandible or the sinus floor.

After the above selection process, 41 cases and 84 controls were selected for the study. Among the BRONJ cases, only those that developed BRONJ at the area of the extraction socket and at a time relevant to the extraction procedure were considered cases. Comparison of parameters between cases and controls are shown in **Table 1**.

**Table 1: Comparison of parameters between cases and controls**

| BRONJ | Yes (Case) (n = 41) | No (Control) (n = 84) | p | | |
|---|---|---|---|---|---|
| Stage | | | | | |
| 1 | 5 (12.2%) | N/A | | | |
| 2 | 33 (80.5%) | N/A | | | |
| 3 | 3 (7.3%) | N/A | | | |
| Age (year) | | | | | |
| mean ± sd | 74.1 ± 7.8 | 71.0 ± 8.6 | | | |
| median [IQR] | 76.0 [71.0; 80.0] | 72.5 [65.5; 76.0] | 0.017 | * | Mann-Whitney U test |
| Gender | | | 0.683 | | Fisher's exact test |
| Female | 38 (92.7%) | 80 (95.2%) | | | |
| Male | 3 (7.3%) | 4 (4.8%) | | | |
| Type of bisphosphonates | | | 0.333 | | Fisher's exact test |
| Alendronate | 16 (39.0%) | 31 (36.9%) | | | |
| Ibandronate | 15 (36.6%) | 20 (23.8%) | | | |
| Risedronate | 9 (22.0%) | 27 (32.1%) | | | |
| Zoledronate | 1 (2.4%) | 6 (7.1%) | | | |
| Route of Administration | | | 0.134 | | Chi square test |
| IV | 4 (9.8%) | 19 (22.6%) | | | |
| PO | 37 (90.2%) | 65 (77.4%) | | | |
| Duration of administration (months) | | | | | |
| mean ± sd | 40.5 ± 32.7 | 37.4 ± 41.7 | | | |
| median [IQR] | 35.5 [14.2; 53.5] | 19.3 [6.8; 48.7] | 0.115 | | Mann-Whitney U test |
| shorter than 6 months | 4 (9.8%) | 18 (21.4%) | 0.06 | | Cochran-Armitage Trend test |
| 6 months ~ 12 months | 6 (14.6%) | 16 (19.0%) | | | |
| 12 months ~ 36 months | 12 (29.3%) | 21 (25.0%) | | | |
| 36 months or longer | 19 (46.3%) | 29 (34.5%) | | | |
| Drug holiday before dental extraction | | | | | |
| mean ± sd | 24.5 ± 65.3 | 111.6 ± 106.1 | | | |
| median [IQR] | 0.0 [0.0; 0.0] | 95.0 [14.0; 168.0] | < 0.001 | *** | Mann-Whitney U test |
| 3 months or shorter | 37 (90.2%) | 39 (46.4%) | < 0.001 | *** | Chi square test |
| longer than 3 months | 4 (9.8%) | 45 (53.6%) | | | |
| Area | | | 0.645 | | Fisher's exact test |
| Mandible Anterior | 3 (7.3%) | 5 (6.0%) | | | |
| Mandible Premolar | 4 (9.8%) | 9 (10.7%) | | | |
| Mandible Molar | 17 (41.5%) | 30 (35.7%) | | | |
| Maxilla Anterior | 2 (4.9%) | 13 (15.5%) | | | |
| Maxilla Premolar | 2 (4.9%) | 5 (6.0%) | | | |
| Maxilla Molar | 13 (31.7%) | 22 (26.2%) | | | |
| Concomitant corticosteroid use | | | 0.119 | | Fisher's exact test |
| Yes | 7 (17.1%) | 6 (7.1%) | | | |
| No | 34 (82.9%) | 78 (92.9%) | | | |
| Perioperative serum CTX level (pg/mL) | | | | | |
| mean ± sd | 220.5 ± 220.9 | 255.0 ± 117.4 | | | |
| median [IQR] | 163.0 [95.0; 254.0] | 235.0 [174.0; 317.5] | 0.003 | ** | Mann-Whitney U |
| lower than 150 pg/ml | 20 (48.8%) | 9 (10.7%) | < 0.001 | *** | Chi square test |
| 150 pg/ml or higher | 21 (51.2%) | 75 (89.3%) | | | |

Abbreviations; sd: standard deviation, IQR: Interquartile range
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

## 2.2 Machine learning methods

This study compared the outputs of five machine learning models, along with single factors including serum CTX level and drug holiday. Prior to formulating machine learning models, the data set was split into two mutually exclusive sets, training (70%) and testing (30%), a method known as the holdout method.[14] The training set was utilized to generate the prediction model, and the remaining 30% of the data (testing set) was employed to estimate the model's accuracy. The measure used to compare the performance of the proposed methods was the area under the Receiver Operating Characteristic (ROC) curve (AUC).

A brief description of each machine learning algorithm is provided below. Each algorithm is explained in clear detail. All machine learning models were implemented using the R programming language (R Core Team, Vienna, Austria, 2016).

### 2.2.1 Logistic regression (LR)

Logistic regression (LR), a common statistical method, was used to evaluate the relationship between categorical variables. It is widely applied in evaluating risk factors or predicting likelihoods of diseases in medical research. LR* uses a mathematical function which relates the dependent variable and the independent variable. The dependent variable is the likelihood of occurrence, and the independent variables are the features or attributes used to predict the occurrence. A multivariable logistic regression model was made with variable selection by backward elimination.

### 2.2.2 Decision tree (DT)

Decision tree (DT) learning is a supervised machine learning technique used to create a model that predicts the value of a target variable based on several input variables by repeated classification, also known as recursive partitioning.[15, 16] The flowchart of this model consists of node, branch, and leaf, resembling a tree structure (**Figure 1**). Ordering of each node is done using a mathematical approach called attribute selection. The measure for attribute selection is crucial for accuracy. The criteria used for the selection are *information gain* and *gini index*, which reflect the reduction in entropy due to sorting of the attribute. The entropy is the uncertainty of an attribute. By calculating the entropy measure of each attribute, the information gain and gini index can be calculated. An attribute with higher value should be placed as root and a branch with 0 entropy should be converted to a leaf node. A branch with entropy higher than 0 needs further splitting. The visualization of a decision tree offers interpretability (**Figure 1**). As this machine learning algorithm is simple, it yields relatively low prediction accuracy compared to other machine learning algorithms. The R package rpart was used in this study.

* Logistic regression

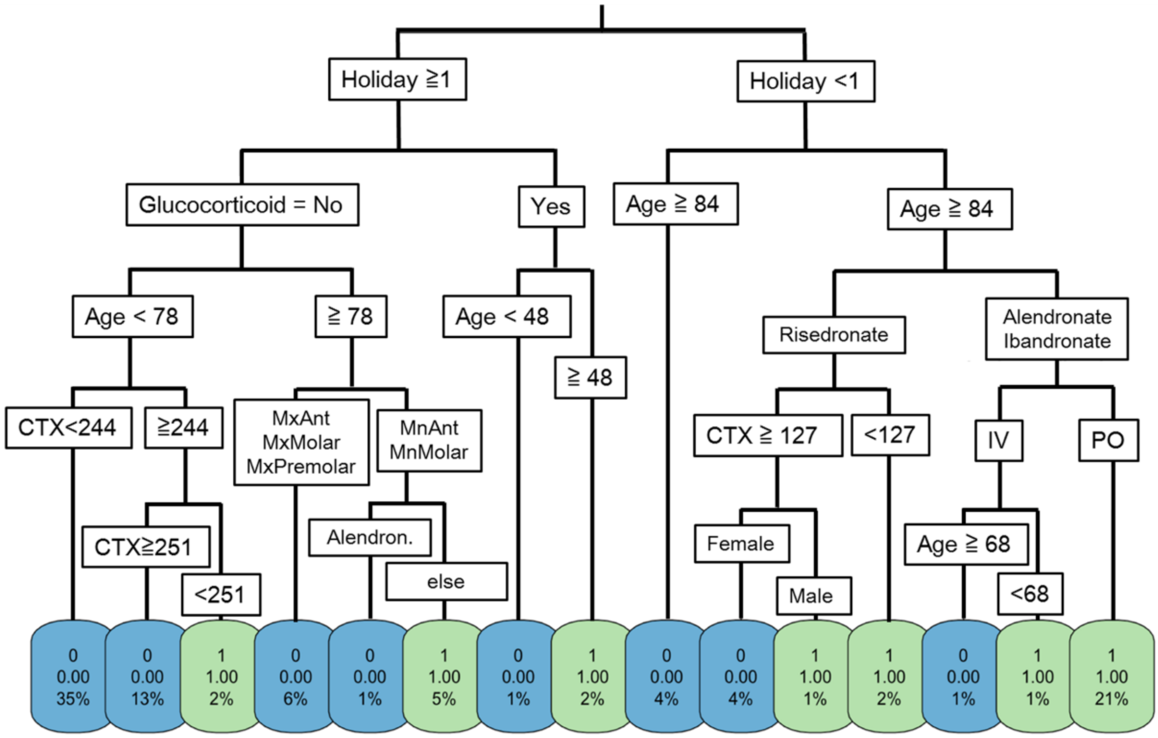



**Figure 1: Decision tree.** This model consists of node, branch, and leaf, resembling a tree structure. Here, the drug holiday is the root node. The position of an attribute is decided by calculating the reduction in entropy due to sorting of an attribute, also known as *information gain* and *gini index*. An attribute with higher value should be placed at the root and a branch with 0 entropy should be converted to a leaf node. A branch with entropy higher than 0 needs further splitting. The visualization of a decision tree offers interpretability. Due to its simplicity, however, its prediction accuracy is relatively low compared to other machine learning algorithms.

### 2.2.3 Support Vector Machine (SVM)

Support Vector Machines (SVMs) are based on the concept of a decision plane that separates a set of objects having different characteristics. An SVM* performs classification tasks by constructing hyperplanes that separate the subject of one class (or group) from another in a multidimensional space. (**Figure 2**) SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables.[1, 14, 17] SVMs are known to be effective when the number of features is large, even when the number of features exceeds the number of samples. However, with a greater number of samples, it performs relatively poorly. The R package e1071 was employed in this study.

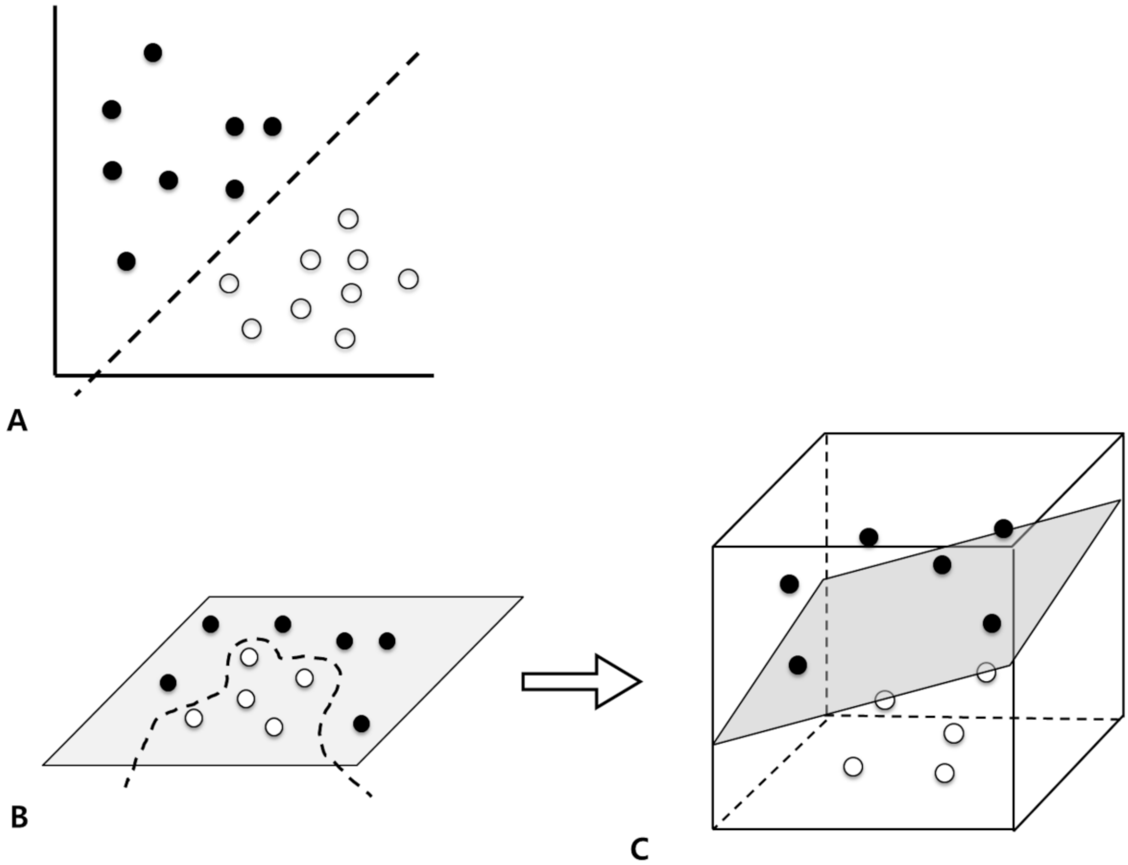



**Figure 2: Simplified illustration of SVM classification. A.** Dashed line will be the hyperplane dividing black filled dots and white filled dots. **B.** Samples are linearly separable. **C.** Transferring samples into higher dimension enables the separation with a hyperplane.

* Support Vector Machine

### 2.2.4 Artificial neural network (ANN)

An artificial neural network is a computing system inspired by neuronal networks of animal brain.[18] ANN*s consist of an input layer, one or more hidden layers, and an output layer. Each layer has neurons which are connected with those in adjacent layers. (**Figure 3**). Each neuron has its own weights in an initial state. When training data is entered in the input layer, the learning process begins. Data is passed to neurons in the next layer until they reach the output. The generated output is compared with the provided information and an error is generated. The error is back-propagated through the ANN, and the weights of the connections between the neurons are adjusted to decrease this error. Once this back-propagation reaches the input layer, another cycle of forward processing begins and reaches the output layer again. The weights are corrected for better accuracy throughout this recursive process. After the learning process, the ANN is presented with optimized weights for use on the test set which has not been seen by the ANN.[19] The R package neuralnet was implemented in this study (**Figure 3**).

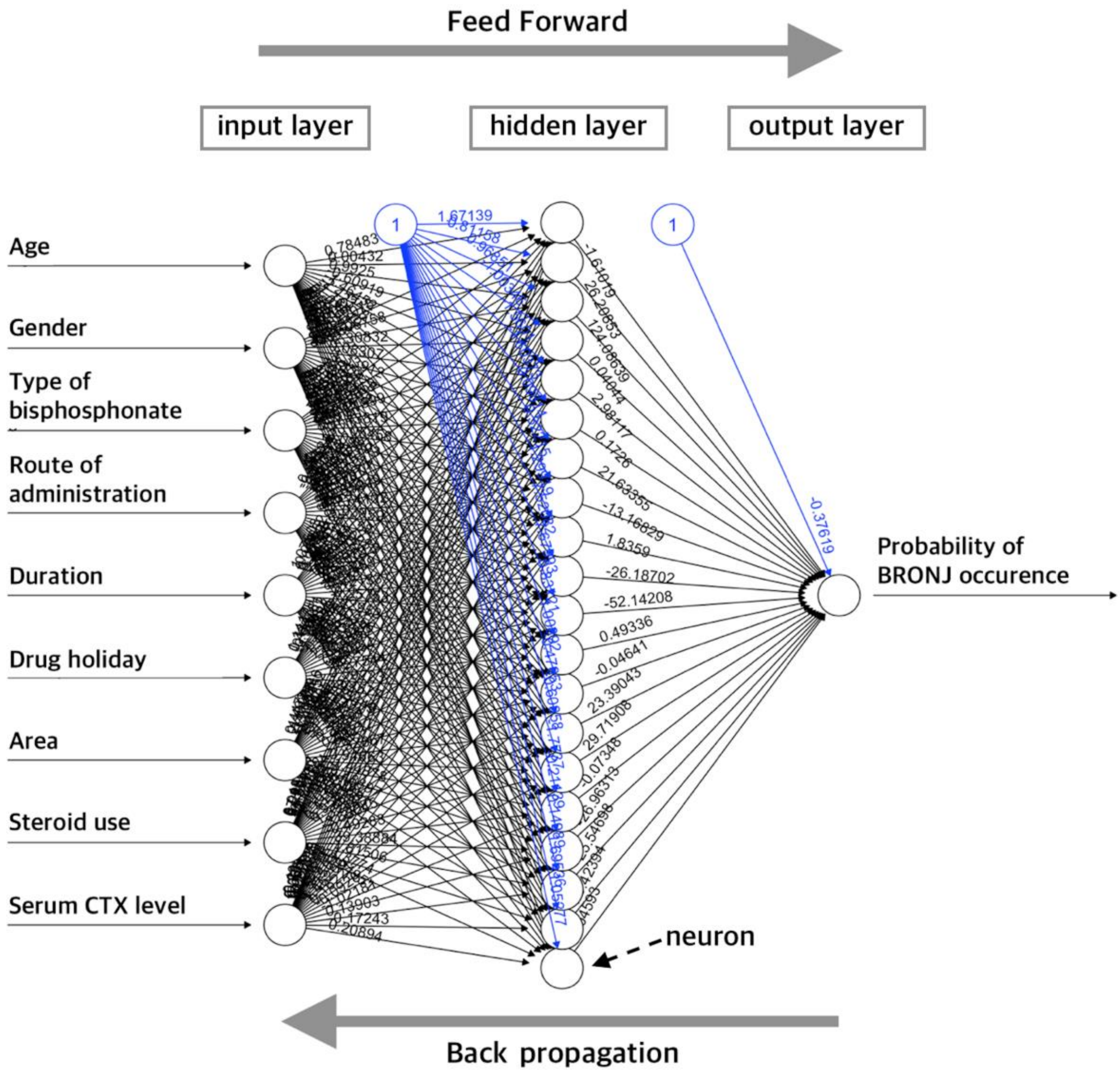



**Figure 3: Artificial neural network model.** ANNs consists of an input layer, one or more hidden layers, and an output layer. Each layer has neurons (dotted arrow) connected with

* Artificial Neural Network

consecutive layers. The black numbers associated with black arrows indicate the weights. The blue numbers associated with blue arrows are bias weights, which allow successful learning by shifting the sigmoid functions. This function is integrated in the R package neuralnet. After the training set is subjected to the feed forward process and back propagation, errors are generated and weights corrected based on the errors to build a more accurate ANN. After this learning process, the ANN is presented with optimized weights and applied to a test set previously unseen by the ANN.

#### 2.2.5 Random Forest (RF)

Random forest (RF) is a tree-based machine learning algorithm, also known as the ensemble method, which creates subsets of decision trees, then combines their output to improve the performance of the algorithm.[15] Combining weak decision tree outputs yields highly accurate results. The subsets are created by N times of a random sampling process, also known as bootstrap aggregation. In this study, N was set to 500 by calculating the votes of the subsets. Overfitting, the phenomenon of a model being overly well-fitted to the training set and thus poorly fitted to the test set, does not occur in the random forest model. Random forest can be used in both classification and regression tasks. The R package randomForest was employed in the study.

### 2.3 Statistical analysis

Comparison of parameters between cases and controls were assessed using the chi square test, Fisher's exact test, and the Cochran-Armitage trend test for categorical variables. Student's t-test and the Mann-Whitney rank sum test were used for continuous variables. Statistical analyses were performed using the R programming language (R Core Team, Vienna, Austria, 2016). $p < 0.05$ was considered significant.

### 2.4 Ethical approval

This study was approved by the Institutional Review Board (IRB) of Yonsei University Dental Hospital (Approval number: 2-2017-0036). Written or verbal informed consent was not obtained from any participants because the IRB waived the need for individual informed consent, as this study had a non-interventional retrospective design and all data were analyzed anonymously.

## 3 RESULTS

The performance of machine learning models was superior to conventional statistical methods using single predictors. The RF[*] model performed best (AUC = 0.973), followed by ANN[†] (AUC = 0.915), SVM[‡] (AUC = 0.882), LR[§] (AUC = 0.844), DT[**] (AUC = 0.821), drug holiday alone (AUC = 0.810), and CTX level alone (AUC = 0.630) (**Figure 4, Table 2**). Serum CTX level alone scored the lowest among all evaluated models. The cutoff value of serum CTX level was 148 pg/mL, the sensitivity and specificity at this cutoff being 50.0% and 87.0%, respectively. The performance of drug holiday alone was superior to serum CTX

[*] Random Forest
[†] Artificial Neural Network
[‡] Support Vector Machine
[§] Logistic Regression
[**] Decision Tree

level, but inferior to the other 4 machine learning models. The cutoff value of drug holiday was 2 days, the sensitivity and specificity at this cutoff being 73.3% and 87.0%, respectively.

Sensitivity and specificity of the best performing model, RF, was 100% and 83.3%, respectively, at its optimal cutoff. The sensitivity and specificity of each model at its optimal cutoff is listed in **Table 3**.

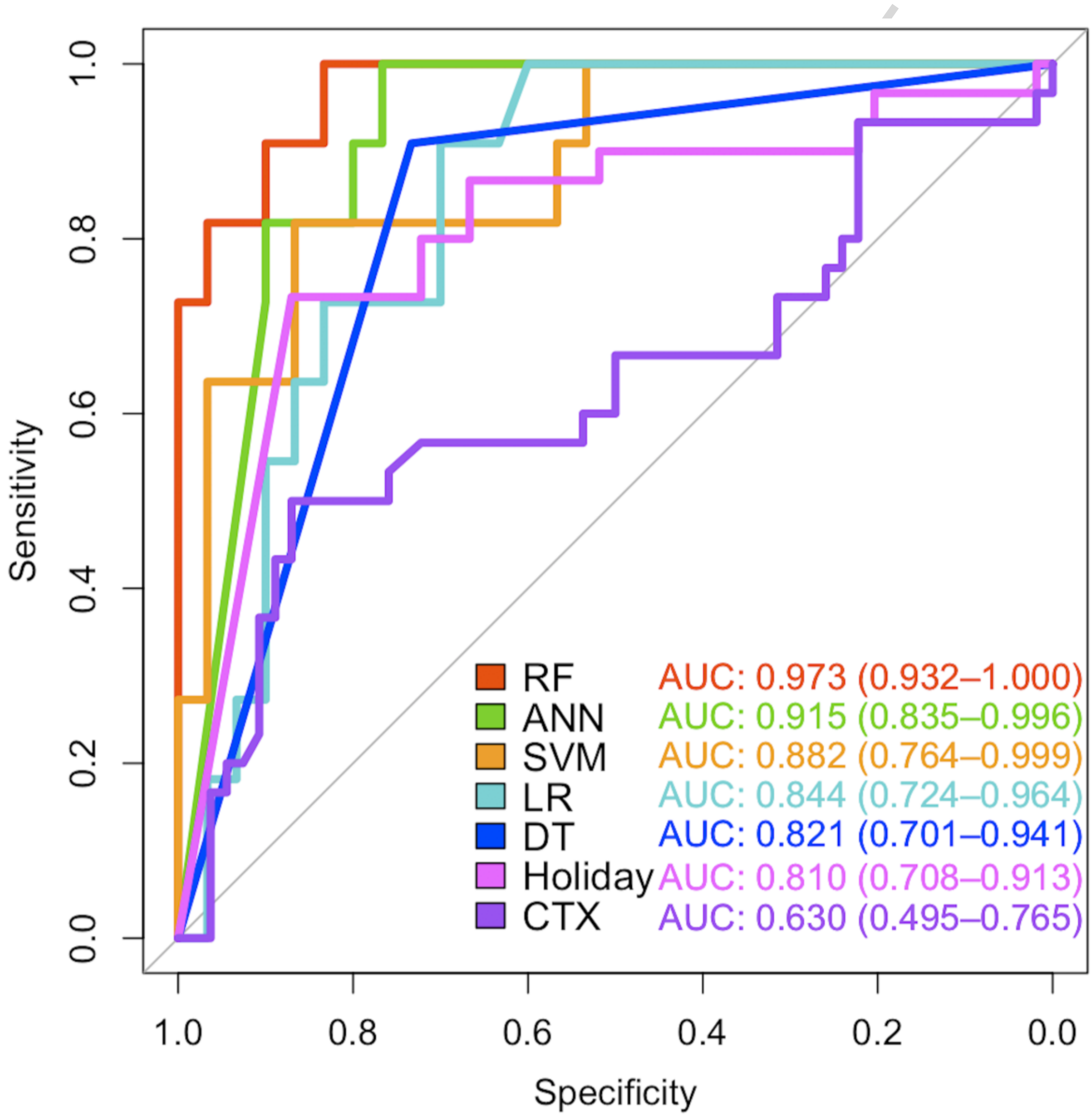



**Figure 4: ROC & AUC of machine learning methods.** The RF model performed best (AUC = 0.973), followed by ANN (AUC = 0.915), SVM (AUC = 0.882), LR (AUC = 0.844), DT (AUC = 0.821), drug holiday alone (AUC = 0.810), and CTX level alone (AUC = 0.630). Sensitivity and specificity of the best performing model, RF, was 100% and 83.3%, respectively, at its optimal cutoff. The sensitivity and specificity of each model at its optimal cutoff is listed in **Table 3**.

**Table 2: Statistical significance of the difference between the areas under ROC curves.** DeLong's test and Bootstrap test were used.

| | ANN | SVM | LR | DT | Holiday only | CTX only |
|---|---|---|---|---|---|---|
| | p | p | p | p | p | P |
| Random Forest (RF) | 0.045* | 0.051 | 0.011* | 0.001** | 0.003** | 0.000*** |
| Artificial Neural Network (ANN) | | 0.311 | 0.109 | 0.059 | 0.057 | 0.000*** |
| Support Vector Machine (SVM) | | | 0.280 | 0.166 | 0.180 | 0.003** |
| Logistic Regression (LR) | | | | 0.356 | 0.332 | 0.008** |
| Decision Tree (DT) | | | | | 0.447 | 0.018* |
| Holiday only | | | | | | 0.990 |

$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

**Table 3: Sensitivity and specificity of each model at optimal cutoff point***

| Model | Sensitivity | Specificity | Optimal cutoff of probability |
|---|---|---|---|
| Random forest (RF) | 100.0% | 83.3% | 0.474 |
| Artificial Neural network (ANN) | 100.0% | 76.7% | 0.999 |
| Support Vector Machine (SVM) | 81.8% | 86.7% | 0.409 |
| Decision Tree (DT) | 90.9% | 73.3% | 1.000 |
| Logistic Regression (LR) | 90.9% | 70.0% | 0.214 |
| Drug holiday alone | 73.3% | 87.0% | 0.547 |
| Serum CTX level alone | 50.0% | 87.0% | 0.370 |

*Optimal cutoff was considered the point maximizing the sum of sensitivity and specificity

## 4 DISCUSSION

Though there are studies suggesting dental extraction can be safely done regardless of bisphosphonate use, bisphosphonates for the treatment of osteoporosis have been found to show a risk of BRONJ development, although lower than that for multiple myeloma and bone metastases of malignancies (0.01–0.04% versus 0.8–12%, respectively).[4, 8-10, 20-22]

Some studies insist that local infection rather than surgical trauma by means of extraction is the main trigger.[4, 10] A combination of perioperative antibiotic prophylaxis, atraumatic surgery, smoothening of sharp bony edges, and plastic wound closure is also emphasized for prevention of BRONJ.[4, 10]

Accurately predicting the risk of BRONJ prior to the dental extraction can aid in deciding whether or not to perform extraction, and may also prevent BRONJ. There have been previous attempts to assess the risk of BRONJ using biomarkers, such as serum C-terminal telopeptide of type I collagen (CTX).[12] However, results of other clinical studies have been controversial.[23-26] A recent position paper published by the American Association of Oral and Maxillofacial Surgeons (AAOMS) stated that there was no evidence to support the use of currently proposed markers. As BRONJ seems to be a multifactorial disease, a single predictive factor may not suffice to accurately predict its occurrence.[13, 24] Various factors including duration of administration, drug holiday, corticosteroid use, type and dose of bisphosphonate, and route of administration might need to be considered together as a complex.

The present study adopted machine learning techniques to predict the risk of BRONJ, which to our knowledge has not been attempted before. This study has shown the superior performance of several machine learning methods, including RF, ANN, and SVM, compared to conventional statistical methods such as multivariable LR or single predictors including serum CTX level. Accurate prediction can benefit both patients and clinicians.

Machine learning algorithms do not provide information on how long the drug holiday should be, what the critical level of serum CTX is, or what type of bisphosphonate has higher risk. This reflects the black-box nature of machine learning methods.[27] The rationale for the outputs generated is inscrutable not only to physicians but also to the engineers who develop the algorithms.[27] There is ongoing research to achieve this interpretability. Though not offering full interpretability, the random forest model can provide the relative importance of the variables by means of the gini index, as shown in **Figure 5**.

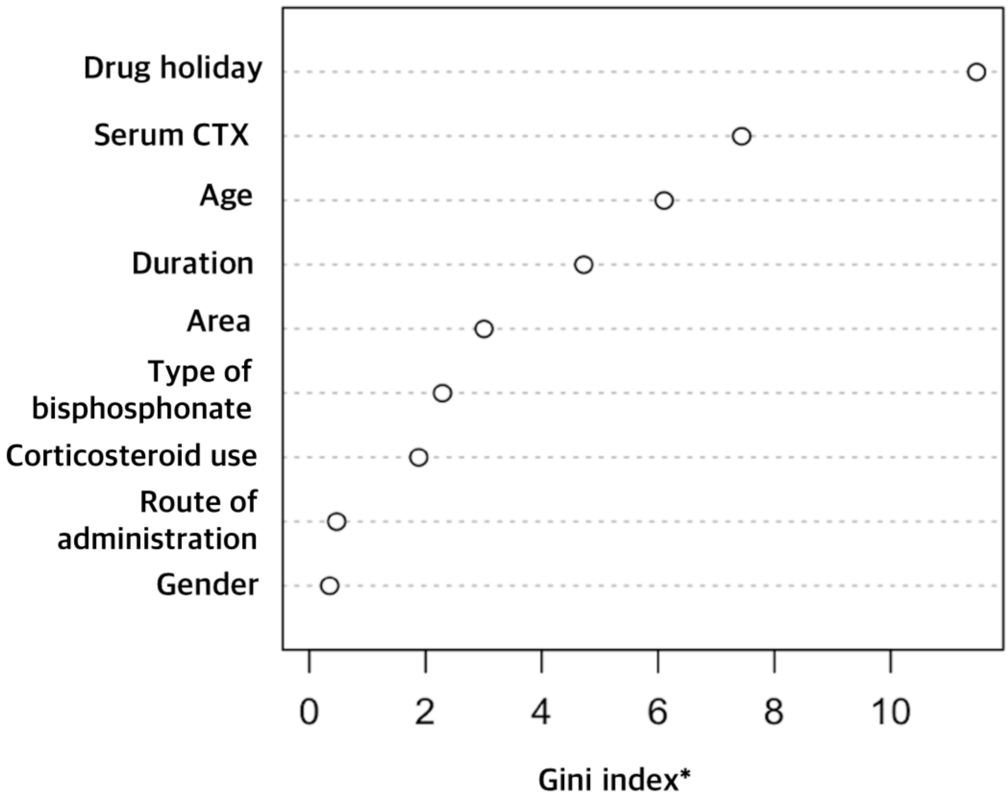



**Figure 5: Variable importance plot of random forest model.** The plot shows relative importance of the variables in random forest model. *Gini index reflects the reduction in entropy, which is the uncertainty, due to sorting of the attribute. An attribute with higher value should be placed as root and a branch with 0 entropy should be converted to a leaf node. A branch with entropy higher than 0 needs further splitting

There are studies using machine learning to predict disease outcome with comparable sample sizes. Shipp et al. conducted a study predicting diffuse large B-cell lymphoma outcome with sample size of 77, consisting of 58 cases and 19 controls.[28] Kan et al.

conducted a study predicting lymph node metastasis in esophageal squamous cell carcinoma with sample size of 28 patients consisting 17 cases and 11 controls.[19] Lee et al. conducted a study predicting occurrence of ventricular tachycardia using sample size of 104 patients consisting of 52 cases and 52 controls.[29] Nevertheless, larger sample size may enable building a more accurate model. This study does not claim that the prediction model built here is perfect and complete. Rather, it shows that machine learning can be utilized to predict disease occurrence with noteworthy performance.

This study was limited to patients who received bisphosphonates for the treatment of osteoporosis. Patients who took bisphosphonates for the treatment of cancer metastasis were not taken into account. This study was also limited to BRONJ that occurred in association with dental extraction. Adding additional parameters such as cause of dental extraction, oral hygiene, or specific gene expression may enable better performing models. It is interesting that considerable accuracy was achieved without incorporating such variables in building the machine learning models. While the results are remarkable, because of the limited number and characteristics of the patients, we entitled this study as preliminary.

For clinicians who are unfamiliar with this approach, it may seem too complex for employment in medical research; such was initially the case with the authors. Solid advances in machine learning techniques have led to an expanded scope of application, beyond engineering. It seems plausible that clinicians will be implementing machine learning in daily practice. In this study, only a modest amount of computer programming experience allowed the implementation of machine learning via open-source R packages (R Core Team, Vienna, Austria, 2016). The full source code used in this machine learning study is attached as supplementary material and shared so the readers can try it as well (**Supplementary material 1-3**). The supplementary material includes instructions on data entry and interpreting the prediction results.

## 5 CONCLUSION

Machine learning methods RF[*], ANN[†], and SVM[‡] showed superior performance compared to conventional statistical methods in predicting BRONJ associated with dental extraction in patients taking bisphosphonates for osteoporosis. Machine learning can be applied to a wide range of clinical studies.

## CONFLICT OF INTEREST

The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.


## ACKNOWLEDGEMENTS

This study was supported by the Yonsei University College of Dentistry Fund (6-2017-0014).


[*] Random Forest
[†] Artificial Neural Network
[‡] Support Vector Machine

**REFERENCES**

[1] J.A. Cruz, D.S. Wishart, Applications of machine learning in cancer prediction and prognosis., Cancer Inform 2 (2007) 59-77.

[2] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction., Comput Struct Biotechnol J 13 (2015) 8-17. doi: 10.1016/j.csbj.2014.11.005

[3] M. McClung, S.T. Harris, P.D. Miller, D.C. Bauer, K.S. Davison, L. Dian, D.A. Hanley, D.L. Kendler, C.K. Yuen, E.M. Lewiecki, Bisphosphonate therapy for osteoporosis: benefits, risks, and drug holiday., Am. J. Med. 126(1) (2013) 13-20. doi: 10.1016/j.amjmed.2012.06.023

[4] M. Mozzati, V. Arata, G. Gallesio, Tooth extraction in osteoporotic patients taking oral bisphosphonates, Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 24(5) (2013) 1707-1712. doi: 10.1007/s00198-012-2239-8

[5] P.D. Delmas, Treatment of postmenopausal osteoporosis., Lancet 359(9322) (2002) 2018-2026. doi: 10.1016/S0140-6736(02)08827-X

[6] S.L. Ruggiero, B. Mehrotra, T.J. Rosenberg, S.L. Engroff, Osteonecrosis of the jaws associated with the use of bisphosphonates: a review of 63 cases, Journal of oral and maxillofacial surgery : official journal of the American Association of Oral and Maxillofacial Surgeons 62(5) (2004) 527-34. doi: 10.1016/j.joms.2004.02.004

[7] W.-S. Choi, J.-I. Lee, H.-J. Yoon, C.-K. Min, S.H. Lee, Medication-related osteonecrosis of the jaw: a preliminary retrospective study of 130 patients with multiple myeloma., Maxillofac Plast Reconstr Surg 39(1) (2017) 1. doi: 10.1186/s40902-016-0099-4

[8] S.L. Ruggiero, T.B. Dodson, L.A. Assael, R. Landesberg, R.E. Marx, B. Mehrotra, A.A.o.O.a.M. Surgeons, American Association of Oral and Maxillofacial Surgeons position paper on bisphosphonate-related osteonecrosis of the jaws--2009 update., J. Oral Maxillofac. Surg. 67(5 Suppl) (2009) 2-12. doi: 10.1016/j.joms.2009.01.009

[9] P. Vescovi, G. Campisi, V. Fusco, G. Mergoni, M. Manfredi, E. Merigo, L. Solazzo, M. Gabriele, G.M. Gaeta, G. Favia, F. Peluso, G. Colella, Surgery-triggered and non surgery-triggered Bisphosphonate-related Osteonecrosis of the Jaws (BRONJ): A retrospective analysis of 567 cases in an Italian multicenter study., Oral oncology 47(3) (2011) 191-194. doi: 10.1016/j.oraloncology.2010.11.007

[10] S. Otto, M. Tröltzsch, V. Jambrovic, S. Panya, F. Probst, O. Ristow, M. Ehrenfeld, C. Pautke, Tooth extraction in patients receiving oral or intravenous bisphosphonate administration: A trigger for BRONJ development?, Journal of cranio-maxillo-facial surgery : official publication of the European Association for Cranio-Maxillo-Facial Surgery 43(6) (2015) 847-854. doi: 10.1016/j.jcms.2015.03.039

[11] T. Hasegawa, S. Ri, M. Umeda, H. Komatsubara, M. Kobayashi, T. Shigeta, I. Yoshitomi, H. Ikeda, Y. Shibuya, I. Asahina, T. Komori, The observational study of delayed

wound healing after tooth extraction in patients receiving oral bisphosphonate therapy., Journal of cranio-maxillo-facial surgery : official publication of the European Association for Cranio-Maxillo-Facial Surgery 41(7) (2013) 558-563. doi: 10.1016/j.jcms.2012.11.023

[12] R.E. Marx, J.E. Cillo, J.J. Ulloa, Oral bisphosphonate-induced osteonecrosis: risk factors, prediction of risk using serum CTX testing, prevention, and treatment., J. Oral Maxillofac. Surg. 65(12) (2007) 2397-2410. doi: 10.1016/j.joms.2007.08.003

[13] S.L. Ruggiero, T.B. Dodson, J. Fantasia, R. Goodday, T. Aghaloo, B. Mehrotra, F. ORyan, A.A.o.O.a.M. Surgeons, American Association of Oral and Maxillofacial Surgeons position paper on medication-related osteonecrosis of the jaw--2014 update., J. Oral Maxillofac. Surg. 72(10) (2014) 1938-1956. doi: 10.1016/j.joms.2014.04.031

[14] W. Kim, K.S. Kim, J.E. Lee, D.-Y. Noh, S.-W. Kim, Y.S. Jung, M.Y. Park, R.W. Park, Development of Novel Breast Cancer Recurrence Prediction Model Using Support Vector Machine, J Breast Cancer 15(2) (2012) 230-9. doi: 10.4048/jbc.2012.15.2.230

[15] L. Breiman, Random Forests, Mach Learn 45(1) (2001) 5-32. doi: 10.1023/A:1010933404324.

[16] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and regression trees, Taylor & Francis, 1984.

[17] S.-W. Chang, S. Abdul-Kareem, A.F. Merican, R.B. Zain, Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods., BMC Bioinformatics 14(1) (2013) 170. doi: 10.1186/1471-2105-14-170

[18] W.-T. Tseng, W.-F. Chiang, S.-Y. Liu, J. Roan, C.-N. Lin, The application of data mining techniques to oral cancer prognosis., J Med Syst 39(5) (2015) 59. doi: 10.1007/s10916-015-0241-3

[19] T. Kan, Y. Shimada, F. Sato, T. Ito, K. Kondo, G. Watanabe, M. Maeda, S. Yamasaki, S.J. Meltzer, M. Imamura, Prediction of lymph node metastasis with use of artificial neural networks based on gene expression profiles in esophageal squamous cell carcinoma, Annals of surgical oncology 11(12) (2004) 1070-8. doi: 10.1245/ASO.2004.03.007

[20] L.A. Assael, Oral bisphosphonates as a cause of bisphosphonate-related osteonecrosis of the jaws: clinical findings, assessment of risks, and preventive strategies., J. Oral Maxillofac. Surg. 67(5 Suppl) (2009) 35-43. doi: 10.1016/j.joms.2009.01.003

[21] H.-J. Kim, T.-J. Park, K.-M. Ahn, Bisphosphonate-related osteonecrosis of the jaw in metastatic breast cancer patients: a review of 25 cases., Maxillofac Plast Reconstr Surg 38(1) (2016) 6. doi: 10.1186/s40902-016-0052-6

[22] S. Otto, M.H. Abu-Id, S. Fedele, P.H. Warnke, S.T. Becker, A. Kolk, T. Mücke, G. Mast, R. Köhnke, E. Volkmer, F. Haasters, O. Lieger, T. Iizuka, S. Porter, G. Campisi, G. Colella, O. Ploder, A. Neff, J. Wiltfang, M. Ehrenfeld, T. Kreusch, K.-D. Wolff, S.R. Stürzenbaum, M. Schieker, C. Pautke, Osteoporosis and bisphosphonates-related osteonecrosis of the jaw: not just a sporadic coincidence--a multi-centre study., Journal of cranio-maxillo-facial surgery : official publication of the European Association for Cranio-Maxillo-Facial Surgery 39(4) (2011) 272-277. doi: 10.1016/j.jcms.2010.05.009

[23] H.W. Jang, J.-W. Kim, I.-H. Cha, Development of animal model for Bisphosphonates-related osteonecrosis of the jaw (BRONJ). Maxillofac Plast Reconstr Surg 37(1) (2015) 18. doi: 10.1186/s40902-015-0020-6

[24] J.-W. Kim, I.-H. Cha, S.-J. Kim, M.-R. Kim, Biomarkers for Bisphosphonate-Related Osteonecrosis of the Jaw., Clinical Implant Dentistry and Related Research 18(2) (2016) 281-291. doi: 10.1111/cid.12297

[25] J.-W. Kim, K.-A. Kong, S.-J. Kim, S.-K. Choi, I.-H. Cha, M.-R. Kim, Prospective biomarker evaluation in patients with osteonecrosis of the jaw who received bisphosphonates., Bone 57(1) (2013) 201-205. doi: 10.1016/j.bone.2013.08.005

[26] T.S. Lazarovici, S. Mesilaty-Gross, I. Vered, C. Pariente, H. Kanety, N. Givol, R. Yahalom, S. Taicher, N. Yarom, Serologic bone markers for predicting development of osteonecrosis of the jaw in patients receiving bisphosphonates, Journal of oral and maxillofacial surgery : official journal of the American Association of Oral and Maxillofacial Surgeons 68(9) (2010) 2241-7. doi: 10.1016/j.joms.2010.05.043

[27] F. Cabitza, R. Rasoini, G.F. Gensini, Unintended Consequences of Machine Learning in Medicine, JAMA : the journal of the American Medical Association 318(6) (2017) 517-518. doi: 10.1001/jama.2017.7797

[28] M.A. Shipp, K.N. Ross, P. Tamayo, A.P. Weng, J.L. Kutok, R.C. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, G.S. Pinkus, T.S. Ray, M.A. Koval, K.W. Last, A. Norton, T.A. Lister, J. Mesirov, D.S. Neuberg, E.S. Lander, J.C. Aster, T.R. Golub, Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning, Nature medicine 8(1) (2002) 68-74. doi: 10.1038/nm0102-68

[29] H. Lee, S.Y. Shin, M. Seo, G.B. Nam, S. Joo, Prediction of Ventricular Tachycardia One Hour before Occurrence Using Artificial Neural Networks, Sci Rep 6 (2016) 32390. doi: 10.1038/srep32390

**HIGHLIGHTS**

- Machine learning models that can predict the occurrence of BRONJ associated with dental extraction was built and validated.
- The performance of machine learning models was superior to conventional methods such as the length of drug holiday and serum CTX levels.
- Random forest model yielded the best performance (AUC = 0.973).
- Drug holiday alone (AUC = 0.810), and CTX level alone (AUC = 0.630) were the least accurate predictors.