

# Clinical Data Based Classification of Osteoporosis and Osteopenia Using Support Vector Machine

Ramesh.T<sup>a,1</sup> and Santhi.V<sup>b</sup>

<sup>a</sup> *Research Scholar, Dept of CSE, Vellore Institute of Technology, Vellore, India*

<sup>b</sup> *Professor, Dept of CSE, Vellore Institute of Technology, Vellore, India*

**Abstract.** In the current world millions of people are suffering with bone diseases such as osteoporosis and osteopenia. The early detection of osteoporosis and osteopenia disease is very important as it helps people to be cautious and get treated on time. Hence research on early detection of osteoporosis and osteopenia disease has gained importance across the world. In this paper, analysis of suitable kernel for Support Vector Machine (SVM) focussing on the classification of osteoporosis and osteopenia disease has been carried out and presented. The kernel functions considered includes polynomial, linear, RBF and Gaussian to find the optimal one for the classification of osteoporosis and osteopenia disease with improved accuracy.

**Keywords.** Osteoporosis; Osteopenia; Data Mining; Big Data; SVM; Health care

## 1. Introduction

Due to change in food habits, lack of physical activities and aging factor, people are often getting affected with diseases. In this way, the osteoporosis is considered as one of the critical diseases that affects bones severely, due to low Bone Mass Density (BMD) [1]. If this kind of disease is not diagnosed at the earlier stage, in due course of time low BMD may lead to bone fracture too. In general, all women after their menopause stage and all men who have crossed age of 50 will be affected with this low BMD due to the stoppage of increasing bone density level. As bones loses its density, chances of getting bone fracture increases, and become quite common among aged people. The bone density can be measured clinically and its value can be used as a measure for further treatment. The clinically measured BMD value is called T-Score [2], and it is based on bone mineral content that reveals out bone mineral density. In general, the value of T score goes down if the calcium and iron minerals content goes down in human bodies due to vitamin-D deficiency.

In literature, there are many classification techniques exists for the prediction of various disease in health care sector that include Support Vector Machine (SVM), Logistic Regression (LR), Artificial Neural Networks (ANN) etc. Support Vector Machines is considered as one of the efficient classification techniques among various learning methodologies. Basically SVM is designed for binary classification problems

<sup>1</sup>Santhi.V, Professor, Dept of CSE, Vellore Institute of Technology, Vellore, India.  
Email:vsanthi@vit.ac.in.

where decision boundaries can be used by splitting classes into two categories, but in recent days there is huge requirements for multi class classification in healthcare sector [3]. SVM can be applied for multi class classification by sub dividing multiple classes into two-class problems. As SVM keeps dividing data into two class problems in multiple times, it finds a new hyper planes which minimizes misclassification errors, and also based on the size of the data, the testing and classification time varies in SVM. Hence to minimize the testing time various kernel functions such as RBF, Polynomial etc. can be applied over SVM [4].

In SVM, positive and negative classes will be separated by their regions very effectively compared to other learning techniques. In healthcare SVM plays a vital role in predicting various diseases such as Alzheimer, where the estimation of life logging data need to be extracted in order to bring memory back to the Alzheimer patient. The same manner, SVM technique can be applied over other diseases in order to bring out confidential information for the prediction of diseases. In mobile healthcare systems a multicast routing can be applied using algorithms such as AdaBoost with SVM classifier, in order to enhance the dependability and scalability in a secured manner [5]. Apart from this Multi SVM would be much more effective in prediction of breast cancer and different types of cancer. In general SVM Technique is very simple as it prefers binary classification at multiple levels of SVM when it comes to multi class classification. In the process of predicting Osteoporosis and Osteopenia, there are various algorithms applied over radiographic images using Principal Component Analysis (PCA) and SVM etc.[6] [7]. In general bone mineral density plays a vital role in the prediction of osteoporosis and osteopenia diseases. The low bone mass density refers to osteopenia, whereas severe loss of bone mass density leads to osteoporosis. By extracting various features from radiographic images, the Osteoporosis and Osteopenia diseases can be classified by applying various statistical measurements over bone mineral density. The bone mineral density can be measured especially for women who cross post-menopausal stage at different states such as walk and rest in order to predict Osteoporosis based on T-score values by selecting relevant features using Principal Component Analysis (PCA) [8]. These predictions will help women to be cautious about decrease in bone mineral strength as it leads to Osteoporosis.

Section 2 discusses about related works carried out on Osteoporosis and Osteopenia disease classifications. Section 3 includes methodology which deals with multi-level SVM in the process of prediction of osteoporosis and osteopenia and also description about the datasets being used to analyze classification approaches. Section 4 deals with experimental setup and obtained results.

## 2. Related works

In this section, a detailed of study of literature has been carried out, related to classification of Osteoporosis and similar type of diseases in healthcare using machine learning approaches such as support vector Machine (SVM) etc. Madhumita Kathuria et.al have proposed Wireless Body Area Networks (WBAN) for taking decision on patient's health related data which is versatile in nature. If data are heterogeneous in nature, classifying those data with high accuracy a major challenging as it contains more sensitive information. In order to overcome this situation, Support Vector Machines for multi class classification would be very appropriate accurate packet classification as it deals with various types packets based on priority. The SVM packet classification can

be done at central controller effectively. Here throughput of Wireless body Area Network system is less as packet delivery rate is low [9]. Alanazi et.al have proposed a predictive model in order to categorize various types of diseases using multi class classification to achieve better performance using the traumatic brain injury (TBI) datasets [3]. In general the accurate prediction refers to identification of diseases from historical data and it will help doctors to provide effective treatment as it will be aimed at multi class classification rather than binary classification. But this predictive model could be suitable for TBI outcomes and not for other diseases. Ilaria Bortone et.al have proposed a supervised approach to predicate postmenopausal women's BMD status based on alimental habits, lifestyle and precedent fractures [14]. In this approach, a detailed analysis of static and dynamic baropodometry will cull the pertinent features through the Principal component analysis (PCA). These two supervised classifiers, such as static and dynamic baropodometry, have proved to be a promising tool for screening both bone and muscle function declines. This approach can only deal with BMD rather than predicting the diagnosis of Osteoporosis. Ramkumar et.al have proposed a model aimed at the classification of Osteoporosis and Osteopenia by making use of radiographic bone images [8]. In this model, through the segmentation of abnormal areas of images, bone disorder can be predicted. In this approach, text-based features are extracted using Gray-Level Co-occurrence Matrix (GLCM) in segmentation. The usage of radiographic images can easily observe the changes in the structure of bone due to bone disorder. In order to reduce the size of the features in bone images Principal Component Analysis can be applied. Finally Multi class Support Vector Machine will classify the bone disorders using feature vectors. Devikanniga et.al have proposed a hybrid classifier model using Monarch butterfly optimization algorithm through Artificial Neural Network (ANN) classifier for early diagnosis that lead to the prevention of Osteoporosis [13]. This hybrid classifier can easily distinguish Osteoporotic patient and healthy person based on the values of BMD. This model has been applied on femoral neck and lumbar spine datasets for predicting the accuracy of diagnosing Osteoporosis. This hybrid classifier model is not suitable for high dimensional medical data classification. Victor Blanco et.al have proposed a novel SVM-based approach towards constructing multi-class classifiers in terms of arranging hyper planes using mixed integer programming by adopting suitable kernel [14]. It can be used to handle multi class classification by extending binary SVM classifiers. In this approach, it finds a polyhedral partitioning of the space feature and assigning classes to partition the cells by maximizing the separation between classes and minimizing two intuitive misclassification errors. But the computational burden is high for complex mixed integer programs. Yassine Naseer et.al focuses on diagnosing Osteoporosis from bone X-ray images by extracting features from low level images using deep learning approach [25]. The major challenge in diagnosing Osteoporosis was addressed here such as distinguishing a person who is osteoporotic and healthy from X-ray images. In order to tackle this issue, the proposed technique performs pre-processing in series of steps to upgrade the contrast of the image, image region with sliding window activity, feature extraction followed by stacked sparse Autoencoder and finally pooling activity followed by classification step using SVM classifier. Based on the literature survey, it is identified that many works have been carried out to diagnose Osteoporosis using radiographic images rather than bone mineral density values. It is observed that binary classification using SVM have been carried out in diagnosing Osteoporosis but not on Osteopenia. Also Multi level classification in diagnosing Osteoporosis and Osteopenia is lagging. All these research gaps are addressed in this paper.

### 3. Materials and Methods

Dataset has been gathered from the Disease Control and Prevention (CDC) centre which has created dataset through National Health and Nutrition Examination Survey (NHANES III) as a part of National Center of Health Statistics (NCHS). The data are retrieved from two file types, they are Demographic Data File (DDF) and Examination Data File (EDF). The required and relevant attributes are retrieved from these above said files. World Health Organization (WHO) and IOF suggest the NHANES database values as the standard reference group for classifying the said disease using BMD. The WHO suggests that non-Hispanic white women age in the range between 20 and 29 years have participated in NHANES III (1988-1994) as the reference group for the computation of T-score in the femoral neck and lumbar spine. The reference group data of NHANES is utilized as a base to find the T-score using BMD values [9] [10].

**Table 1.** WHO definition of osteoporosis based on BMD measurements from DXA

Definition	Bone Mass Density Measurement	T-Score Range
Normal	$\mu: \sigma < 1$	$\geq -1$
Osteopenia	$\mu: 1 \leq \sigma < 2.5$	$-1 < -2.5$
Osteoporosis	$\mu: \sigma \geq 2.5$	$\leq -2.5$

In Table 1, the category of disease is presented with the corresponding T score values. In osteoporosis classification, samples with T-score values above -1 are labeled as non-osteoporotic and normal category. Similarly, if calculated T score value lies between -1 and -2.5, then it is classified as osteopenia, and if the T-score value less than or equal to -2.5 are considered as osteoporotic. [11] [12]. The dataset has been analyzed by medical experts and it has 10 input variables such as weight, standard height, body mass index, gender, race, age in years, femoral neck BMD, femoral neck BMC, femoral neck area and T-Score values [13].

**Table 2.** Description of output attributes

Attribute	Name	Label For SVM	Description
T-Score	Class	0	Normal Health Class
		1	Osteopenia Class
		2	Osteoporosis Class

The data set considered for study from NHANES is pre-processed before performing classifications. The pre-processing steps include data integration, data reduction, data cleaning, data transformation and data splitting. There are more than 100 irrelevant attributes have been found and removed in data reduction phase. The feature set is retained with 12 attributes which are needed for diagnosing osteoporosis and osteopenia. Min-max normalization method is adopted to normalize all the attributes as a part of data transformation process. Finally in data splitting process, the data set has been divided into two category with respect to age, then the T score measurements has been applied only if age value is above 50. Table 2 shows the expected outcome of multi-level SVM.

### 3.1 Multi – level SVM

The Support Vector Machine (SVM) classifier is a kind of machine learning algorithm that attempts to find an optimal hyperplane with maximum margin. This algorithm separates the linearly separable data samples into two classes [14]. If the data is non-linearly distinguishable, then SVM maps the data into high-dimensional feature space and performs the classification. where ‘W’ is the normal vector that represents the angle or orientation of the hyper plane in m- dimensional space (synonymously it can be called as width of the margin), ‘X’ represents the input vector and ‘b’ represents the bias or threshold that represents the position or the distance of the hyper plane from the origin as shown in Figure 1. In multi-level SVM, there can be two or more binary SVM’s can be used to classify data set into many classes. In this paper, 3 input data set has been divided into three classes namely class 0, Class 1 and class 2 that represents people with normal health, people affected with osteopenia and people affected with osteoporosis [15]. Initially in binary SVM, the data set has been classified into normal and diseased categories. The diseased categories can further be classified into osteopenia and osteoporosis using binary SVM.

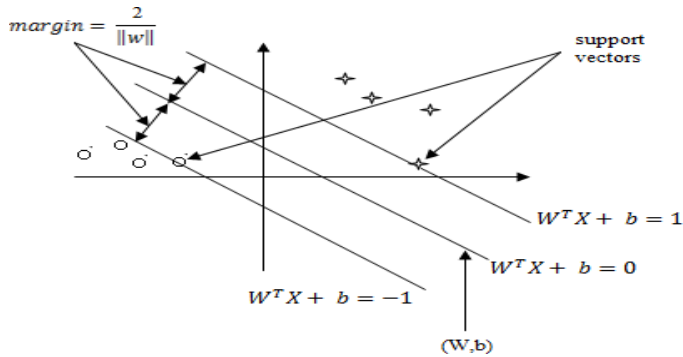


Figure 1. Working Model of SVM

In general SVM algorithms use different types of kernels in order to transform the input into required form; they are called as linear, polynomial, RBF and Gaussian. The detailed discussion on each kernel is presented below.

### 3.2 Polynomial Kernel

Polynomial kernel plays a vital role in learning of non-linear models. For identifying the similarity of vectors in training samples of kernelized models including Support Vector Machines, they make use of the polynomial kernel function. Learning of non-linear models can be done using polynomial kernel which works in feature space with polynomials of original variables and identifying the similarity of vectors. For degree-n polynomials, the polynomial kernel (2) is characterized where n represents order of the polynomial, a and b are vectors in the input space, i.e. vectors of features compared from training or test samples then  $c \geq 0$  is a parameter compromising the impact of higher-order versus lower-order terms in the polynomial.

### 3.3 Gaussian Kernel

Here  $\alpha$  chooses the width of the Gaussian and it is a function of Standard deviation and the square of Variance (3).

### 3.4 Radial Basis Function (RBF) Kernel

The RBF kernel is represented a RBF kernel decreases by distance in the range between zero (within limit) and one during estimation, and the feature space has an endless dimensions of this kernel.

### 3.5 Linear Kernel

Linear kernel can be defined by adding the inner product  $\langle a, b \rangle$  and the constant  $c$  which is optional. This particular kernel used by the kernel algorithms is similar to their non-kernel counterparts.

### 3.6 Exponential Kernel

This particular kernel is almost similar to Gaussian Kernel, without square of the norm. It is also a RBF kernel.

### 3.7 Laplace RBF Kernel

The Laplace Kernel is identical to the exponential kernel, aside from being less delicate for changes in the sigma boundary. Being same, it is also a RBF kernel. Note that the perceptions made about the sigma boundary for the Gaussian kernel can also apply to the Exponential in addition to Laplacian kernels.

### 3.8 Sigmoid Kernel

The sigmoid function can be used for artificial neurons as an activation function as Neural Network field produces sigmoid kernel. The sigmoid kernel is also known as the Hyperbolic Tangent Kernel, and also called as Multi-Layer Perceptron Kernel. As the sigmoid kernel evolves from neural network theory, it is playing significant role for Support Vector Machines. It is almost similar to perceptron neural network.

### 3.9 Anova Kernel

The Anova kernel is quite similar to RBF kernel, just like Laplacian kernel in addition to Gaussian kernels. In solving multidimensional regression problems this kernel plays a significant role compared to other kernels.

### 3.10 Proposed System Flow Diagram

The working model for the classification of osteoporosis and osteopenia using Multi SVM is shown in Figure 2. The required features would be extracted from input data set

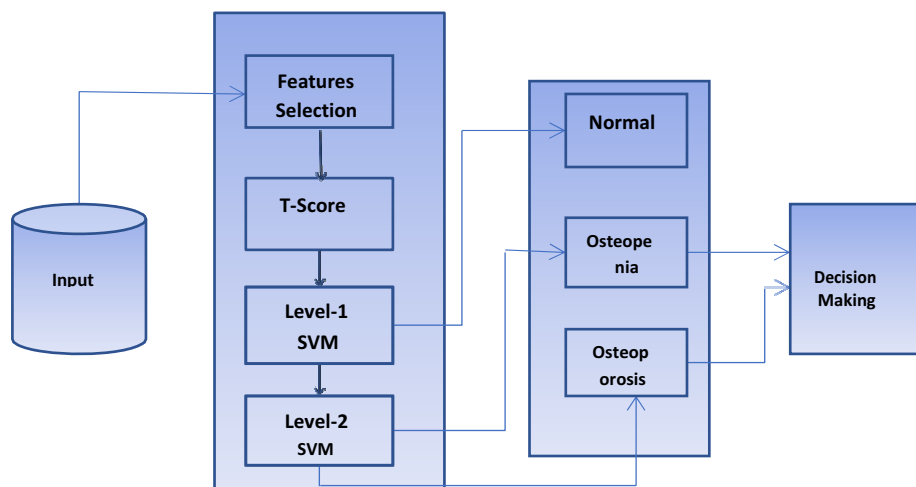


Figure 2. System Flow Diagram

and it would be used for T-Score calculations. The calculated values are given as inputs to SVM classify the diseases. In the first level of SVM, binary classification is carried out to diagnose a particular person to decide upon whether the person is affected with disease or not. If that person has disease then it will get into second level of SVM, where Input: Here the input has been taken from open source health related organization, where datasets are freely available as mentioned above. Here more than 100 input attributes are taken for processing such as Weight, Standing Height, Body Mass Index, Gender, Race, Age, and Femoral neck BMD etc. Later all these input attributes will go for features selection as it is very important to get the relevant attributes which are needed to calculate T-Score values. T-Score Calculation: In general T-Score (10) values need to be calculated using BMD values such as Here for each patient T-Score can be calculated using respective BMD value, mean and Standard Deviation of all BMD values which are associated to it.

#### 4. Experimental Results

The experiments are conducted in MATLAB using Multi SVM with different kernels such as Polynomial, Linear, RBF, and Gaussian. The femoral datasets has been considered to classify diseases into osteoporosis and osteopenia, and sample data set is shown in Table 3. The comparisons of kernels have been carried out and obtained results are shown in Table 4. The parameters considered are accuracy, sensitivity and specificity. The accuracy obtained using polynomial kernel is 93.20%, which is less compared to other kernels. Similarly the kernel which has given the highest accuracy is Gaussian kernel with the accuracy of 97.30% as shown in Figure 3.

Like accuracy, other parameters such as sensitivity of each kernel are calculated and it is found that Gaussian kernel has got more sensitivity of 96.30%. At last when it comes to Specificity, it is observed that polynomial kernel has lowest compared to other kernels, but at the same time Gaussian kernel has more sensitivity about 97.70%.

**Table 3.** Sample Data Set with extracted features

Weight (kg)	Standing Height (cm)	Body Mass Index (kg/m <sup>2</sup> )	Gender	Race/Ethnicity - Recode	Age in years	Femoral neck BMD	Femoral neck BMC	Femoral neck area	T-score	class
56.8	167.8	20.17	1	1	50	1.014	5.55	5.47	0.50	
65	156.6	26.51	1	1	50	0.934	4.67	5	-0.10	
87.4	171.7	29.65	1	1	50	0.931	5	5.37	-0.10	
108.8	174.2	35.85	1	1	50	0.725	4.26	5.88	-1.81	
75.5	169.8	26.19	1	1	50	0.779	3.92	5.04	-1.41	
81.4	173.2	27.13	1	1	50	0.813	4.03	4.96	-1.11	
51.2	164.9	18.83	1	1	50	0.949	4.95	5.22	0.00	
86.5	167.6	30.79	1	1	50	0.917	4.77	5.2	-0.20	
99.2	178.9	30.99	1	1	50	0.813	4.97	6.11	-1.11	
84.6	176.5	27.16	1	1	50	1.009	5.55	5.5	0.50	
111.9	190.7	30.77	1	1	50	0.976	5.99	6.14	0.20	
65.7	186.7	18.85	1	1	50	0.892	5.61	6.29	-0.50	

Like Accuracy, other parameters such as Sensitivity of each kernel are calculated and it is found that Gaussian kernel has got more Sensitivity of 96.30%. At last when it comes to Specificity, it is observed that polynomial kernel has lowest compared to other kernels, but at the same time Gaussian kernel has more sensitivity about 97.70%.

**Table 4.** Comparison of kernels with qualitative parameters

	Accuracy	Sensitivity	Specificity
Polynomial	93.20%	93%	93.20%
Linear	94.40%	94.40%	94.50%
RBF	95.30%	95.30%	95.70%
Gaussian	97.30%	96.30%	97.70%
Exponential	92.30%	92.50%	92.30%
Laplace RBF	91.30%	91.40%	91.10%
Sigmoid	95.90%	95.40%	96.20%
Anova	90.60%	89.90%	90.30%

In this proposed multi class SVM based works; it is observed that the Gaussian kernel has given better results compared with other kernels such as polynomial, Linear and RBF in terms of accuracy, sensitivity and specificity. In the process of classifying osteoporosis and osteopenia using Multi SVM with the help of various kernel functions and it is found that Gaussian kernel provides better result compared to other kernels as shown in Table 4. Here the performance measurements such as accuracy, sensitivity and specificity are analyzed with all the kernels as mentioned in the Table 4.

## 5. Conclusion

In this paper, analysis of the kernel function for the classification of diseases Osteoporosis and Osteopenia using multi class SVM has been carried out. A list of kernels considered for the experiment includes linear, polynomial, RBF, exponential, sigmoid, Laplace RBF, Anova and Gaussian function. The three output classes are labelled as 0, 1 and 2, where class 0 represent people not affected with disease and



normal, class 1 represent people are affected with Osteopenia and class 2 represent Osteoporosis patients. The evaluated results show that the Gaussian kernel is giving better accuracy compared to other kernels in SVM. This proposed research work will be extended by cascading multi class SVM with other techniques for better performance in future.

## References

- [1] Sayantan Bhattacharya, Ayush Goyal et.al. Computer Based Automatic Detection and Classification of Osteoporosis in Bone Radiographs. ©2019 IEEE. Rice AS, Farquhar-Smith WP, Bridges D, Brooks JW. Canabinoids and pain. In: Dostorovsky JO, Carr DB, Koltzenburg M, editors. Proceedings of the 10th World Congress on Pain; 2002 Aug 17-22; San Diego, CA. Seattle (WA): IASP Press; c2003. p. 437-68.
- [2] Devikanniga, R. Joshua Samuel Raj. Classification of osteoporosis by artificial neural network based on monarch butterfly optimization algorithm. Vol. 5, pp. 70–75.
- [3] Alanazi, A. H. Abdullah et.al. Accurate and dynamic predictive model for better prediction in medicine and healthcare. © Springer; July 2017.
- [4] Fereshteh Falah Chamasemani et.al. Multi-class Support Vector Machine (SVM) classifiers – An Application in Hypothyroid detection and Classification. © 2011 IEEE.
- [5] Maryam Sadat Hejazi, S. A. R. Al-Haddad et.al, Multiclass Support Vector Machines for Classification of ECG Data with Missing Values. 29:660–674; © 2015 Taylor & Francis Group.
- [6] Ying Wang, Enrico Coiera et.al. Using multiclass classification to automate the identification of patient safety incident reports by type and severity, BMC Medical Informatics and Decision Making (2017).
- [7] Saphthagirivasan V., Anburajan M, Diagnosis of osteoporosis by extraction of trabecular features from hip radiographs using support vector machine: an investigation panorama with DXA, Comput. Biol. Med., 2013, 43, pp. 1910–1919, (doi: 10.1016/j.combiomed.2013.09.002).
- [8] Ramkumar S, Malathi R. An Automatic Bone Disorder Classification Using Hybrid Texture Feature Extraction with Bone Mineral Density. 19 (12), 3517-3523; Asian Pacific Journal of Cancer Prevention, Vol 19 (2018).
- [9] Madhumita Katuria and Sapna Gambhir. A novel Optimization Model for Efficient Packet Classification in WBAN. International Journal of Energy, Information and Communications, Vol.7, Issue 4 (2016), pp.1-10.
- [10] Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD, 2005.
- [11] S. Bhattacharya, D. Nair, A. Bhan, A. Goyal, Computer based automatic detection and classification of osteoporosis in bone radiographs, in: Proc. 6th Int. Conf. On Signal Processing and Integrated Networks (SPIN). IEEE, Noida, India, 2019, pp. 1047–1052.
- [12] Harrar, L. Hamami, S. Akkoul, E. Lespessailles, R. Jennane. Osteoporosis assessment using Multilayer Perceptron neural networks. in: 2012 3rd International Conference on Image Processing Theory, Tools and Applications, Istanbul, 2012, pp. 217–221.
- [13] Devikanniga. Diagnosis of osteoporosis using intelligence of optimized extreme learning machine with improved artificial algae algorithm. International Journal of Intelligent Networks 1 (2020) .
- [14] Ilaria Bortone, Gianpaolo Francesco Trotta et.al., Supervised Approach to Classify the Status of Bone Mineral Density in Post-Menopausal Women through Static and Dynamic Baropodometry. 978-1- 5090-6014; ©2018 IEEE.
- [15] Victor Blanco, Alberto Japón et.al., Optimal arrangements of hyperplanes for SVM-based multiclass classification. 2019 © Springer;
- [16] Felicia et.al., The clinician's guide to prevention and treatment of osteoporosis. Osteoporos, Int. 25 (2014) 2359–2381.
- [17] Iliou, C.-N. Anagnostopoulos, G. Anastassopoulos. Osteoporosis detection using machine learning techniques and feature selection. Int. J. Artif. Intell. Tool. 23, (2014) 1450014, 9 pages.
- [18] Muller, K.-R., S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf. 2001. An introduction to kernel-based learning algorithms. IEEE Transactions on Neural Networks 12(2):181–201. doi: 10.1109/72.914517.
- [19] Devlin et.al, diagnosing osteoporosis by using dental panoramic radiographs: the OSTEODENT project. Oral Surg., Oral Med., Oral Pathol., Oral Radiol. Endod., 2017, 104, (6), pp. 821–828,