International Journal of Pure and Applied Mathematics
Volume 119 No. 18 2018, 97-114
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue
http://www.acadpubl.eu/hub/

A Comparative Analysis of Machine Learning Techniques for Early Detection of Osteoporosis

[1]Zixian Wang, [2]Jiangnan Fu, [3]Rujun Yan, [4]Joon Soo Park and [5]Hong Gyung Kim

[1]Cornell University,
New York.
zixianwang8@gmail.com

[2]Dalian Maritime University,
Dalian, China.

[3]Wesleyan University,
Connecticut.

[4]Shanghai American School,
China.

[5]Beijing International Bilingual Academy,
Beijing, China.

Abstract

Osteoporosis is a condition in which the density of bone decreases and the body of the patient stops producing bone as required by a healthy individual. This is a serious public health concern because of its increasing frequency across many countries. This indicates an urgent need to predict the problem in its early stages to overcome limitations of existing clinical decision tools with low accuracy. This paper investigates various machine learning techniques for predicting osteoporosis. The paper explores and analyzes the use of the feature selection technique in combination with classification techniques. Classification techniques implemented in WEKA are employed to benchmark an osteoporosis data set. The results are computed using different testing options including 10-fold cross-validation, training sets, and the percentage split with and without the feature selection technique. The results are compared in terms of correctly classified instances, execution time, kappa statistics, and mean absolute values for experiments with and without the feature selection technique. The results suggest that IBK training set testing gives the best outcome without feature selection, whereas LMT, IBK, J48, SMO, JRip, and bagging methods give the best outcome with the feature selection technique.

1. Introduction

The human body is composed of the skeleton, which protects the body from injury. With support of muscles, bone facilitates the movement of the body.

Osteoporosis is a skeletal disorder that can lead to bone mass loss and diminished bone strength (as characterized by density and quality) and thus to an increased risk of fracture. Bone loss typically results from estrogen deficiency in postmenopausal women or from other age-related mechanisms such as secondary hyperparathyroidism and reduced mechanical loading [1]. There are mainly two types of osteoporosis. The first type is primary osteoporosis, which occurs from age as well as from reduced estrogen in women. The second type is secondary osteoporosis, which may occur in any age groups due to cancer, hormonal disorder, and certain types of medication. The most commonly affected parts of the body by osteoporosis include the spine, wrist, and hips.

Osteoporosis is a serious public health concern because of its increasing frequency across countries. Therefore, osteoporosis has become an essential health and economic index in most countries [2]. Osteoporosis prevention is complicated, but it holds some promise as the best way to reduce future fracture [3]. The social economic burden of osteoporosis is so great that its etiology, prevention, and treatment have become an urgent issue that needs to be addressed worldwide. Modeling the relationships between a disease and its potential risk factors is a crucial task in epidemiology and public health [4, 5]. Usually, numerous potential types of osteoporosis need to be considered simultaneously to assess disease determinants and predict its progression for the purpose of disease control or prevention.

More importantly, some common diseases may be clinically silent but can cause significant mortality and morbidity after onset. Unless prevented or treated in early stages, such diseases may affect the quality of life and increase healthcare cost burdens. With the success of risk factor analysis and disease prediction based on an intelligent computational model, unnecessary tests may be avoided. Useful information can help evaluate the risk of a disease, monitor its progression, and facilitate early prevention measures.

A number of epidemiological studies have developed clinical decision tools for osteoporosis risk assessment to select postmenopausal women for the measurement of bone mineral density. The purpose of such tools is to help estimate the risk of osteoporosis, not to diagnose osteoporosis. The osteoporosis self-assessment tool is a clinical decision tool based on a simple formula using age and body weight [6]. Although this tool uses only two factors to predict osteoporosis risk, it has been shown to have good sensitivity with an appropriate cutoff value [7]. However, the decision tool has the limitation of low accuracy for clinical use [8].

Machine learning is an area of artificial intelligence research using various methods for data classification. Several machine learning techniques have been applied in clinical settings to predict disease and have shown higher accuracy for diagnosis than classical methods [9]. Support vector machines (SVMs), random forests (RFs), and artificial neural networks (ANNs) have been widely used approaches in machine learning [9].

The present study develops prediction models for osteoporosis using various machine learning methods implemented in WEKA, including SVMs, RFs, and ANNs, among others.

The paper computes the results of various machine learning techniques based on identified data sets in terms of the accuracy of correctly classified instances of the data set. The experimental results show the existence of some irrelevant and redundant features present in the data set. Therefore, the paper explores different feature selection techniques available in WEKA for selecting the most promising features of the benchmark data set and attempts to enhance the accuracy of results from shorter execution times. The results with and without the feature selection technique are compared to assess the benefits of using most promising features in terms of execution time and accuracy.

The rest of this paper is organized as follows: Section 2 describes machine learning techniques, followed by feature selection techniques in Section 3. Section 4 presents the state of art in the field of detection for the osteoporosis problem.

Section 5 presents the proposed framework for investigating machine learning techniques in combination with feature selection techniques. The results are presented in Section 6. Finally, Section 7 concludes with suggestions for future research.

2. Machine Learning Techniques

Machine learning is a data analytic technique that teaches computers to do what comes naturally to humans and animals: learning from experience. Machine learning techniques use computational methods to “learn” information directly from data without relying on a predetermined equation as a model. Algorithms adaptively improve their performance as the number of samples available for learning increases.

Machine learning uses two types of techniques (Fig. 1). Supervised learning trains a model using known input and output data so that it can predict future outputs, and unsupervised learning finds hidden patterns or intrinsic structures in input data.

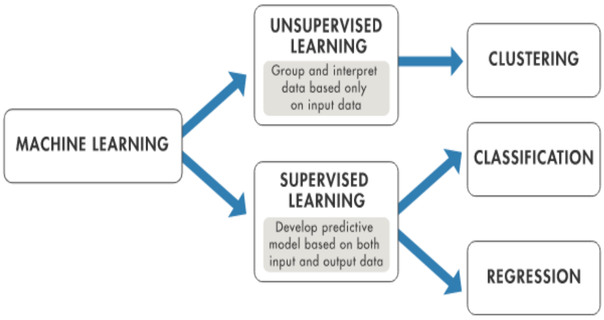

Fig. 1: Types of machine learning techniques

This paper focuses on a supervised set of machine learning techniques for classifying subjects into two classes (osteoporosis and non-osteoporosis) selected based on their popularity and usage frequency in biomedical engineering problems. The most popular machine learning techniques are briefly described in Table 1. There are many different machine learning techniques and feature selection techniques [10].

Table 1: Most Popular Machine Learning Techniques

Technique	Description
ZeroR	This is the simplest technique that ignores all predictors and relies only on the target. It predicts the majority category class and acts as the baseline performance for remaining classification techniques.
OneR	This works on the principle of selecting one rule corresponding to each predictor in the data set. This one rule is generated based on the assumption that there must be a smallest total error for that rule. A frequency table is used for this purpose. The output from the OneR method is comparatively equal to state-of-the-art classification techniques.
J48	This is based on a decision tree generated by C4.5. It provides many benefits such as the ability to deal with incomplete or noisy data, handle both discrete and continuous data, and expose the most inequitable features.
IBk	This is an instance-based technique where K is the number of neighbors under consideration. It generates the output that contains root relative squared error, relative absolute error, mean absolute error, Kappa statistic, and confusion matrix, among others.

Naive Bayes	This is a statistical method that helps classification based on supervised learning. It helps in understanding model uncertainty with the help of outcome probabilities. There are numerous applications of Naive Bayes, such as text classification, spam filtering, and hybrid recommender system by combining collaborative filtering with Naïve Bayes classifier.
SMO	Sequential minimal optimization (SMO) refers to an optimization algorithm used inside SVM implementation. SVMs work best for binary classification, although they can be extended to other classifications. It classifies instances based on hyper planes.
KStar	This is an instance-based classifier. An entropy-based distance function is used for its functioning that helps in determining a class of a test. The pre-classified database already trained helps in the classification of new data by comparing them.
JRip	FOIL's information gain is the reason for the efficiency of the Ripper algorithm. The disjunctive normal form is used in the RIPPER algorithm to represent every rule. All these rules help in training the data set and classify new data sets. The detection of malicious URLs and customer membership cards represent an application of JRip.
Bagging	This works as a meta-estimator in which the base classifier is fitted on any random subset of the original data set and then their cumulative prediction is made from individual predictions. Unstable learning algorithms such as neural networks and decision trees in which some minor changes can result in large deviations in predictions are very well classified using Bagging. Bagging is effective in "unstable"' learning algorithms where small changes in the training data set result in large changes in predictions.
LMT	The logistic model tree (LMT) is a supervised training algorithm that combines decision tree learning and logistic regression (LR) to produce efficient results. Multi-class target variables, binary, missing values, and nominal attributes can easily be handled by the LMT.

3. Feature Selection Techniques

Many applications in different fields have been observed to produce high-dimensional data. High-dimensional data have many features to be analyzed for further processing. However, the analysis of high-dimensional data suffers from a problem called the curse of dimensionality [11].

One way to solve this problem is to reduce dimensions of the data set to process without loss of significant information. Data dimensions can be reduced either by feature combination or selection. Feature combination transforms features either linearly or nonlinearly. Major methods employed for feature combination include principal component analysis (PCA), independent component analysis, and linear discriminant analysis, among others. Feature selection or reduction retains original features and selects a subset of features that can predict the target class variable with maximum classification accuracy [12]. Feature selection provides many benefits such as reducing training time, reducing overfitting, and improving accuracy.

The present study explores different feature selection techniques and identifies correlation-based feature selection as the most promising feature selection technique as per the state of art in the field. To select promising features in this set of experiments, this paper employs a correlation-based feature selection technique based on the greedy search method using the WEKA tool. See [13,14,15] for further details on feature selection.

The results are compared with and without the use of the feature selection technique in terms of training time and accuracy.

4. Literature Review

Several studies have proposed different machine learning techniques for detecting osteoporosis.

Chidozie et al. [16] proposed the use of supervised machine learning to forecast the risk of osteoporosis. They prepared a data set using questionnaires and advice from medical experts. They considered a total of 19 parameters for the person under observation, including gender, alcohol frequency, smoke frequency, a meal rich in calcium, and exercise, among others. They used multi-layer perception and naïve Bayes techniques for predicting osteoporosis by using the WEKA tool. They found that multi-layer perception provides better results. The proposed work can be directly combined with the existing healthcare system and can result in improved clinical decisions for the real-time assessment of remote data.

Iliou et al. [17] employed a combination of feature selection and machine learning techniques for the detection of osteoporosis. They used a data set composed of 3426 subjects (2343 healthy cases and 1083 pathological cases) and applied 20 machine learning techniques for the categorization of subjects into non-osteoporosis and osteoporosis with four diagnostic factors including weight, age, gender, and height. These classifiers were evaluated using the 10-fold cross-validation method. The results suggest the applicability of their method for accurately predicting osteoporosis.

Singh et al. [18] used associative classification for the prediction of heart disease. They employed algorithms such as the k-nearest neighbor, Aprior, OneR, ZeroR, FP-Growth, J48, and naïve Bayes on a data set accessed from the machine learning repository of the University of California, Irvine (UCI) and reported the accuracy of their results up to 99.19% by using the hybrid technique for classification associative rules.

Kim et al. [19] worked on the identification of the risk of osteoporosis in postmenopausal women by employing machine learning algorithms such as logistic regression (LR), SVMs, ANNs, and RFs using a data set from Korea National Health and Nutrition Surveys for Korean postmenopausal women. These learning models were compared with the osteoporosis self-assessment tool (OST), and the results showed 76.7% accuracy, 76% specificity, and 77.8% sensitivity.

Ciusdel et al. [20] developed a prediction system based on a deep learning model using convolutional networks. They validated their work using a synthetically developed data set using the finite element analysis (FEA) method. To ensure the accuracy of the proposed prediction model, they compared it with a physical computation method using some distinct test data set. They used already trained SVM models.

Wibawa et al. [21] employed a combination technique based on a classification technique and a feature selection method for diagnosing Parkinson's disease by examining symptoms of dysphonia. They used voice data as a parameter for diagnosing Parkinson's disease. They collected the data set from the UCI repository and normalized the data using the PCA, correlation-based feature selection (CFS), and wrapper. They employed four types of classification techniques, namely, the k-nearest neighbor (kNN), SVMs, multi-layer perceptron (MLP), and the Bayesian network, and reported 98.97% accuracy, 97.92% specificity, and 99.32% sensitivity of results.

Kavitha and Kannan [22] used feature selection and feature extraction methods for effective classification of heart diseases. They proposed a framework using the PCA and feature selection. They used a wrapper filter along with feature subset selection to select promising features and enhance accuracy.

These studies suggest that machine learning may provide important insights into data and help classify data into different classes. Previous results indicate that the machine learning technique may facilitate accurate classification results if employed appropriately in combination with the feature selection technique. Therefore, retaining the benefits of effective classification outcomes for machine learning techniques, the present paper employs a set of most popular machine learning techniques in combination with the feature selection technique to classify osteoporosis.

5. The Proposed Work

The main objective of this work is to investigate machine learning techniques in combination with feature selection techniques for the accurate detection of osteoporosis. The paper develops various prediction models using classification algorithms with different techniques offered by the WEKA tool and compares them in terms of correctly classified instances. The identified classification technique is applicable in the early diagnosis of osteoporosis.

5.1. The Proposed Framework

The proposed framework for developing prediction machine learning models and their comparison are depicted in Fig. 2.

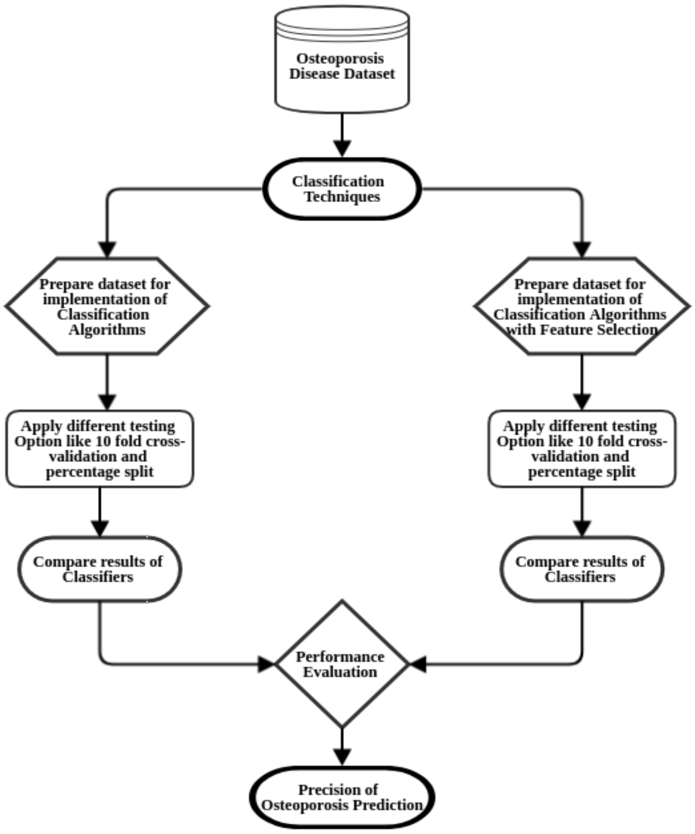

Fig. 2: Workflow of the proposed work

The important phases of the proposed framework are described as follows:

Stage 1. Dataset Selection Stage

This stage involved the selection of the benchmark data set. A data set consisting of 1000 patients with nine attributes were selected, including family history, blood calcium, vitamin D, thyroid, kidney function, and phosphorus. These features were the most significant features for predicting osteoporosis.

Stage 2. Pre-processing and Transformation Stage

This stage involved the preprocessing of the data set to remove noise and scale values of different features and transforming them into a format compatible with underlying classification techniques such as the CSV or the ARFF (attribute-relation file format).

Stage 3. Feature Selection Stage

This stage involved the selection of promising features using the WEKA-implemented feature selection technique. This helped to reduce the number of features, which in turn led the reduction of training time as well as prediction time. Correlation-based feature selection subset evaluator and greedy stepwise search methods were employed. Four features were selected for the analysis of the osteoporosis data set: family_history, blood_calcium, vitamin_D, and phosphorus.

Stage 4. Model Generation Stage

This stage consisted of model generation by selecting different options for training as well as testing for 10-fold cross-validation, a training set and percentage split using the WEKA tool for the osteoporosis data set. Two sets of experiments were conducted: with and without feature selection.

Stage 5. Performance Evaluation Stage

This stage consisted of analyzing results from different classification techniques, namely J48, ZeroR, IBk, SMO, KStar, JRip, Naive-Bayes, Bagging OneR, and LMT, for the classification of the osteoporosis data set in terms of corrected classified instances.

Stage 6. Precision of Diseases

The trained model of machine techniques was used to predict unknown instances into osteoporosis and non-osteoporosis classes.

In this set of experiments, the identified benchmark data set was divided into training and test data sets per different options available in WEKA.

5.2. Benchmark Data Set

In this set of experiments, a data set consisting of 1000 patients with nine attributes, including family history, blood calcium, vitamin D, thyroid, kidney function, and phosphorus, was selected (Table 2).

Table 2: Attributes

Attributes	Description/Values
Gender	Having value one (1) for “Male” and value zero (0) for “Female”
Age	Range 21-102
Family History	Yes, No
Blood Calcium	Normal, Deficient
Vitamin D	Normal, Deficient
Thyroid	Yes, No
Kidney Function	Normal, Abnormal
Phosphorus	Normal, Deficient
Class Attribute (Disease)	Yes, No

5.3. Experimental Platform Setup

Two sets of experiments were conducted for different classification techniques with and without the feature selection technique. Table 3 shows the platform used.

Table 3: Platform setup

Parameter Name	Description
Processor	Intel® Core™ i5-4200M CPU @ 2.50GHz × 4
Operating System	Linux
RAM	8GB
Cache	L1d cache: 32K L1i cache: 32K L2 cache: 256K L3 cache: 3072K

6. Results & Discussion

Two sets of experiments were conducted using the identified benchmark data set with different classification techniques implemented in WEKA. The results were computed with and without the feature selection technique. Tables 2 and 3 (without feature selection) and Tables 4 and 5 (with feature selection) show the results for correctly classified instances, training time, kappa statistics, and mean absolute error.

Table 4: Comparison of different classifiers (without feature selection) using testing options in terms of correctly classified instances

Classifiers	Correctly classified instances using 10-fold cross-validation (%)	Correctly classified instances using training set (%)	Correctly classified instances using percentage split (%)
ZeroR	83.30	83.29	80.36
OneR	82.89	83.30	80.36
J48	86.23	88.56	82.73
IBk	82.89	98.38	81.84
Naive bayes	85.43	87.34	85.12
SMO	86.23	87.35	86.61
KStar	82.99	96.05	82.14
JRip	86.63	88.26	86.31
Bagging	85.75	92.61	85.71
LMT	87.14	86.94	86.30

The results in Table 4 show that the highest accuracy value was 87.14 for the LMT classifier with the 10-fold cross-validation testing method and 98.38 for the IBk classifier with the training set testing method. In the case of the percentage split testing method, the highest accuracy was 86.31 for the JRip classifier. The highest accuracy reported for different data sets was 98.38%.

Fig. 3 graphically shows the results in Table 2. It can be seen that the highest value was achieved for the IBk classifier with the training set testing option.

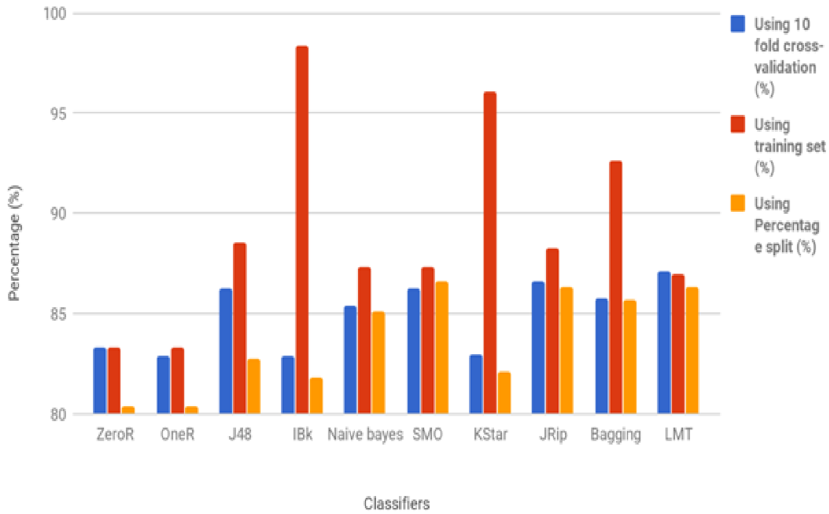

Fig. 3: Comparison of different classifiers (without feature selection) using testing options

Table 5: Comparison of different classifiers (without feature selection) using testing options in terms of execution time

Classifiers	Execution time using 10-fold cross-validation (in sec)	Execution time using training set (in sec)	Execution time using percentage split (in sec)
ZeroR	0.001	0.02	0.001
OneR	0.001	0.01	0.001
J48	0.05	0.01	0.01
IBk	0.001	0.10	0.1
Naive bayes	0.01	0.02	0.001
SMO	0.08	0.04	0.11
KStar	0.001	1.39	0.33
JRip	0.04	0.05	0.15
Bagging	0.02	0.07	0.14
LMT	0.46	0.20	0.19

Table 5 compares three testing options, namely 10-fold cross-validation, training set use, and percentage split, for execution time for different classifiers without using the feature selection technique. The minimum 0.001 sec was observed for ZeroR, OneR, IBk, and Kstar classifiers with the 10-fold cross-validation testing option, and the maximum time was observed for Kstar with the training set testing option.

Fig. 3 graphically shows the results in Table 3, including the maximum and minimum execution times for different classifiers.

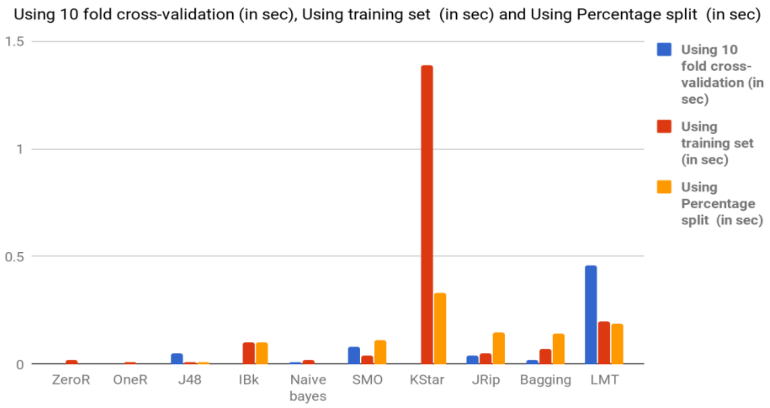

Fig. 3: Comparisons using 10-fold cross validation, training set use, and percentage split

Table 6: Comparison of different classifiers based on feature selection for different testing options

Classifiers	Correctly classified instances using 10-fold cross-validation (%)	Correctly classified instances using training set (%)	Correctly classified instances using percentage split (%)
ZeroR	83.30	83.30*	80.36
OneR	83.29*	83.30	80.36
J48	86.94*	86.94^	82.44^
IBk	86.13*	86.94^	82.44*
Naive Bayes	85.32^	86.74^	82.44^
SMO	86.94*	86.94^	82.44^
KStar	85.43*	85.43^	82.44*
JRip	86.94*	86.94^	86.31
Bagging	86.94*	86.94^	86.31*
LMT	86.63^	86.94	82.44^

Table 4 compares three testing options, namely 10-fold cross-validation, training set use, and percentage split, for correctly classified instances using the feature selection technique. In In Table 4, values labeled as "*" indicate correctly classified instances by classifiers in combination with the feature selection technique, whereas values labeled as "^" indicate a decrease in correctly classified instances by the classifier in combination with the feature selection technique relative to the corresponding values in Table 2. With the use of the feature selection technique, the highest accuracy was 86.94 for J48, SMO, JRip, and bagging with the 10-fold cross-validation testing method, and the same value of 86.94 was achieved for LMT, IBk, J48, SMO, JRip, and bagging classifiers with the training set testing method. In the case of the percentage split testing method, the highest accuracy was 86.31 for the bagging classifier. The highest accuracy was 86.94 based on the comparison of all testing options using the feature selection technique.

Fig. 4 shows the results in Table 4. The highest value was achieved using IBk, J48, JRip, bagging, SMO, and LMT classifiers with a different set testing option.

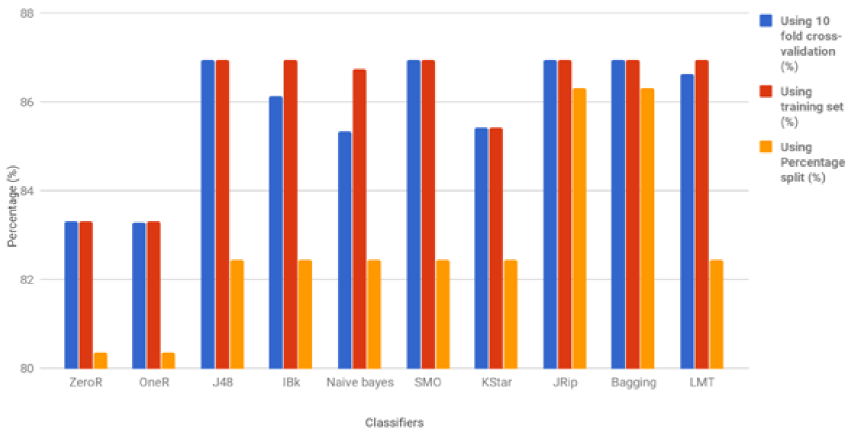

Fig. 4: Comparison of different classifiers based on feature selection

Table 7: Comparison of different classifiers for execution time based on feature selection

Classifiers	Execution time using 10-fold cross-validation (in sec)	Execution time using training set (in sec)	Execution time using percentage split (in sec)
ZeroR	0.001	0.01*	0.001
OneR	0.001	0.001*	0.001
J48	0.03*	0.01	0.001*
IBk	0.001	0.19^	0.06*
Naive bayes	0.001*	0.001*	0.001
SMO	0.05*	0.07^	0.06*
KStar	0.001	0.59*	0.15*
JRip	0.01*	0.03*	0.04*
Bagging	0.13^	0.04*	0.09*
LMT	0.38*	0.14*	0.14*

Table 7 compares three testing options, namely 10-fold cross-validation, training set use, and percentage split, for execution time for different classifiers using the feature selection technique. The minimum time of 0.001 sec was obtained for ZeroR, OneR, IBk, Kstar, and naive Bayes classifiers with the 10-fold cross-validation testing option, and the maximum time of 0.59 seconds was obtained for Kstar with the training set testing option.

Fig. 5 graphically shows the results in Table 5. The maximum and minimum execution times were observed for different classifiers using the feature selection technique.

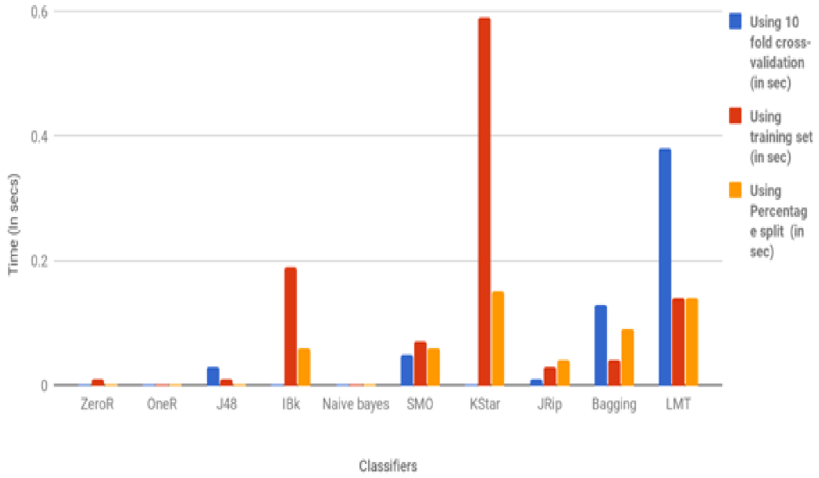

Fig. 5. Maximum and minimum execution times for different classifiers

Table 8. Comparison of different classifiers for the osteoporosis data set based on different parameters using 10-fold cross-validation

Classifiers	Correctly classified instances (%)	Incorrectly classified instances (%)	Kappa statistic	Mean absolute error
ZeroR	83.30	16.70	0.0	0.2787
OneR	82.89	17.10	-0.008	0.1711
J48	86.23	13.76	0.5529	0.1483
IBk	82.89	17.10	0.3715	0.1714
Naive Bayes	85.43	14.57	0.4861	0.1503
SMO	86.23	13.76	0.5644	0.1377
KStar	82.99	17.01	0.3375	0.1875
JRip	86.63	13.36	0.6294	0.1493
Bagging	85.75	14.27	0.5096	0.1471
LMT	87.14	12.85	0.6058	0.1506

Table 8 compares different classifiers in terms of incorrectly and correctly classified instances, kappa statistic, and mean absolute value. According to the results, LMT was a better classifier than other representative classifiers for the benchmark data set of osteoporosis in terms of correctly classified instances.

7. Conclusions and Future Research

This paper investigates different machine learning techniques for predicting osteoporosis. The paper explores and analyzes the effects of using the feature selection technique in combination with the classification technique. This paper employs classification techniques for benchmark data sets for osteoporosis using the WEKA tool and its testing options, namely 10-cross validation, training set use, and percentage split methods, in combination and without using the feature selection technique. The results were compared in terms of correctly classified instances, execution time, kappa statistics, and mean absolute value for experiments with and without the feature selection technique. According to the results, the best outcome was observed for the IBk training set testing without feature selection, whereas LMT, IBK, J48, SMO, JRip, and bagging methods gave the best results with feature selection.

Future research should analyze different supervised and unsupervised machine learning techniques with additional performance metrics for osteoporosis detection.

References

[1] P. Sambrook and C. Cooper, “Osteoporosis”, Lancet, 2006.

[2] Taylor, B. C., Schreiner, P. J. and Stone, K. L. (2004). Long-term prediction of incident hip fracture risk in elderly white women: study of osteoporotic fractures. American Journal of Geriatr Soc 52: 1479–1486.

[3] Kanis, J. A. and Johnell, O. (2005). Requirements for DXA for the management of osteoporosis in Europe. Osteoporosis International 16: 229–238.

[4] Moudani, W., Shahin, A., Chakik, F. and Mora-Camino, F. (2011a). Dynamic Rough Sets Features Reduction. International Journal of Computer Science and Information Security 9 (4): 1–12.

[5] Kanis, J. A., Johansson, H. and Johnell, O. (2005a). Alcohol intake as a risk factor for fracture. Osteoporosis International 16: 737–742.

[6] L. K. Koh, W. B. Sedrine, T. P. Torralba, A. Kung, S. Fujiwara, S. P. Chan, Q. R. Huang, R. Rajatanavin, K. S. Tsai, H. M. Park, and J. Y. Reginster, “A simple tool to identify Asian women at increased risk of osteoporosis,” Osteoporos Int, vol. 12, pp. 699-705, Sep 2001.

[7] F. Richy, M. Gourlay, P. D. Ross, S. S. Sen, L. Radican, F. D. Ceulaer, W. B. Sedrine, O. Ethgen, O. Bruyere, and J. Y. Reginster, "Validation and comparative evaluation of the osteoporosis self-assessment tool (OST) in a Caucasian population from Belgium," QJM, vol. 97, pp. 39-46, Jan 2004.

[8] L. G. Raisz, "Screening for osteoporosis," N Engl J Med, vol. 353, pp. 164-171, Jul 2005.

[9] C. H. Hsieh, R. H. Lu, N. H. Lee, W. T. Chiu, M. H. Hsu, and Y. C. Li, "Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks," Surgery, vol. 149, pp. 87-93, Jan 2011.

[10] Witten, I. H., Frank, E., Trigg, L. E., Hall, M. A., Holmes, G., & Cunningham, S. J. (1999). Weka: Practical machine learning tools and techniques with Java implementations.

[11] R. Bellman, Adaptive Control Processes: A Guided Tour, Princeton University Press, Princeton, 1961.

[12] R. Kohavi, G.H. John, Wrappers for featue subset selection, Artificial Intelligence, 1(2) (1997) 273–324.

[13] Kumar, G., & Kumar, K. (2012). An information theoretic approach for feature selection. Security and Communication Networks, 5(2), 178-185.

[14] Kumar, G., & Kumar, K. (2011, May). A novel evaluation function for feature selection based upon information theory. In Electrical and Computer Engineering (CCECE), 2011 24th Canadian Conference on (pp. 000395-000399). IEEE.

[15] Kumar, G., Kumar, K., & Sachdeva, M. (2010). An empirical comparative analysis of feature reduction methods for intrusion detection. International Journal of Information and Telecommunication Technology, 1(1), 44-51.

[16] Chidozie, E. N., Dooshima, M. P., Ademola, B. J., Komolafe, F., & Adebayo, I. P. (2017). Osteoporosis Risk Predictive Model Using Supervised Machine Learning Algorithms. Science Research, 5(6), 78-87.

[17] Iliou, T., Anagnostopoulos, C. N., & Anastassopoulos, G. (2014). Osteoporosis detection using machine learning techniques and feature selection. International Journal on Artificial Intelligence Tools, 23(05), 1450014.

[18] Singh, J., Kamra, A., & Singh, H. (2016, October). Prediction of heart diseases using associative classification. In Wireless Networks and Embedded Systems (WECON), 2016 5th International Conference on (pp. 1-7). IEEE.

[19] Kim, S. K., Yoo, T. K., & Kim, D. W. (2013, July). Osteoporosis risk prediction using machine learning and conventional methods. In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE (pp. 188-191). IEEE.

[20] Ciusdel, C. F., Vizitiu, A., Moldoveanu, F., Suciu, C., & Itu, L. M. (2017, July). Towards deep learning based estimation of fracture risk in osteoporosis patients. In Telecommunications and Signal Processing (TSP), 2017 40th International Conference on (pp. 659-665). IEEE.

[21] Wibawa, M. S., Nugroho, H. A., & Setiawan, N. A. (2015, October). Performance evaluation of combined feature selection and classification methods in diagnosing parkinson disease based on voice feature. In Science in Information Technology (ICSITech), 2015 International Conference on (pp. 126-131). IEEE.

[22] Kavitha, R., & Kannan, E. (2016, February). An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining. In Emerging Trends in Engineering, Technology and Science (ICETETS), International Conference on (pp. 1-5). IEEE.