Healthcare Sensor Data Analytics: Prediction of Osteoporosis Disease using Classification Tree Algorithms

T. Mathankumar
Ph.D. Research Scholar
Department of Computer Science
Bharathiar University, Coimbatore

S. Vijayarani, PhD
Associate Professor
Department of Computer Science
Bharathiar University, Coimbatore

ABSTRACT
Healthcare is the organized provision of medical services to preserve or improve physical and mental health. Enhance the well-being of the individual and the community, this involves identifying, treating, preventing, and managing illnesses. Analyzing health-related data for better decision-making, efficiency, and results is known as healthcare analytics. Healthcare sensor data analytics, in particular, focuses on using data from medical sensors to improve clinical decision-making and patient illness monitoring. Many sensor-based devices are used to collect the patient's health information. They can be processed using effective algorithms that help physicians get insights and make better treatment decisions. This article aims to predict osteoporosis disease with the collected data from sensor devices. The ultra-sonometer device is used to collect bone mineral density data. A bone's strength and fracture risk is indicated by its mineral content, mostly calcium, in a specific volume of bone measured by bone mineral density (BMD). This work focused on predicting osteoporosis disease using the Random Forest, Decision Tree, and ID3 classification algorithms. Execution time and classification accuracy are used to assess these algorithms' performance. According to experimental outcomes, the Decision Tree classifier is the most efficient technique as it gets the maximum classification accuracy on the other hand, the ID3 and Random Forest classifiers show the fastest execution times.

Keywords
Classification, Bone Mineral Density, Sensor, Osteoporosis, Decision Tree, Random Forest, ID3

1. INTRODUCTION
Researchers in the healthcare sector have a more challenging task: making predictions about illnesses based on a substantial medical database. Analytics for medical sensor data is becoming more and more crucial for the healthcare sector. Healthcare analytics is the methodical evaluation of historical and present data to obtain insights, make conclusions, and aid decision-making. Healthcare analytics uses statistical methodologies, machine learning algorithms, and artificial intelligence models to investigate data and discover significant patterns. The insights gained from this approach can help to increase operational efficiency, optimize healthcare delivery, and improve patient outcomes [1]. Healthcare sensor data analytics is the process of evaluating data obtained from various sensors and wearable devices used in medical settings. These sensors track vital signs, activity levels, and other health indicators, delivering real-time information on patients' ailments. By evaluating this data, healthcare practitioners may identify trends, anticipate prospective health concerns, and personalize treatments to individual requirements, eventually improving patient care and results [2].

The quantity of minerals (mostly calcium) contained in a given volume of bone is measured by the term Bone Mineral Density or BMD. It assists in determining fracture risk and bone strength. A procedure known as a DEXA scan (Dual-Energy X-ray Absorptiometry) is commonly used to assess BMD. At present many sensor-based devices are used for finding the BMD. The popular devices are Peripheral Dexa, Quantitative Computed Tomography, Ultra-Sonometer, and Single Energy X-ray Absorptiometry. In this work, we used the data generated by the Ultra-Sonometer Device for analysis. In this work, Osteoporosis disease has to be identified and classified into Mild, Normal, and Severe classes. Osteoporosis is a medical condition characterized by weakening bones, making them brittle and more prone to fracture. It occurs when the body loses excessive bone mass, creates inadequate bone, or both. This sickness is typically silent, progressing slowly over time with no symptoms until a bone fracture. Osteoporosis is more common in the elderly, particularly in women after menopause, and it may be treated with lifestyle changes, medications, and other approaches that strengthen bones and prevent fractures. However, BMD only accounts for 60%-80% of bone strength, and other skeletal features influence bone strength and fracture risk. Some of these skeletal features may be examined using sophisticated imaging modalities, although they are costlier and less widely available than basic DXA. As a result, implementing a widely available, noninvasive approach that enhances fracture-risk prediction beyond the capabilities of conventional DXA scans and clinical risk factors into clinical practice has been a considerable challenge [4]. Osteoporosis is a common skeletal disease characterized by reduced bone mineral density (BMD) and a higher risk of low-energy fracture. In the United States, the incidence of osteoporosis in elderly people is estimated to be 10.3%, whereas the prevalence of low bone mass is 43.9%. In China, the prevalence of osteoporosis in elderly people is estimated to be 15.7%, with a large increase as the population ages. Despite their significant impact on human health, there is presently a scarcity of very effective osteoporosis treatments with no negative side effects [5]. This research uses Decision Tree, Random Forest, and ID3 classifier algorithms to predict osteoporosis illness [6] [11].

The remaining portions of the paper are organized as follows. Related works are discussed in section 2. The proposed methodology is given in section 3. Section 4 analyzes the experimental results. Section 5 gives conclusions.

2. LITERATURE SURVEY

Wilson Ong [7] has discussed ways to enhance the diagnosis and risk assessment. This systematic review emphasized the possibility of combining artificial intelligence (AI) with CT scans for osteoporosis screening and categorization. The review addressed model constraints resulting from limited, single-center studies, technological issues, and the necessity of large-scale, multi-center research to attain clinical applicability. AI incorporates clinical and radiomics aspects that may also improve diagnosis precision, albeit clinical translation requires further study. Despite the scoping review's limitations, this paper offered a useful starting point for further research on AI-assisted osteoporosis screening.

Kaname Miura, et al., [8] have shown that using machine learning algorithms in conjunction with optical bone densitometry (OBD) as a screening tool for osteoporosis is feasible. Using age, weight, and OBD measurement, a ridge regression model was created to predict the t-score. The findings suggested that the model is a viable substitute for more traditional techniques and is sufficiently dependable for osteoporosis screening.

Rahul Paul, et al., [9] have proposed a method for classifying osteoporosis using X-ray images, a challenging task due to the visual similarity between healthy and osteoporotic images. Traditional features (GLCM, LBP, RLM) have been used for classification, but recent advances favor convolutional neural networks (CNNs) for feature extraction and classification. Given the limited training data available, a transfer learning approach using pre-trained CNNs on ImageNet was implemented. Three approaches were explored: classification using traditional features, deep features, and a combination of both. The best training accuracy of 79.31% (AUC 0.85) was achieved using deep features selected with a symmetric uncertainty ranking algorithm and a random forest classifier. Combining traditional and deep features yielded an accuracy of 75.86% (AUC 0.789). However, on the blind test set, using only deep features resulted in a lower accuracy of 44.82%. Future work will be focused on enhancing deep feature extraction and further tuning the CNN for improved classification performance on unseen data.

Nader Salari, et al., [10] have predicted meta-analysis and systematic review evaluated the prevalence of osteoporosis worldwide to guide health policy. A global osteoporosis prevalence of 18.3% was determined by analyzing data from 86 research with over 103 million individuals. Women had higher rates of osteoporosis (23.1%) than males (11.7%). The highest regional incidence, 39.5%, was found in Africa. To lower the risks associated with osteoporosis, their findings emphasized the importance of improved healthcare planning, treatment resources, and preventative measures. Osteoporosis is a major global health problem, especially in Africa and Europe.

Ferdoush HS, et al., [11] have discussed osteoporosis as a progressive skeletal disease characterized by reduced bone mass and tissue deterioration, often without symptoms until a fracture occurs. Its prevalence significantly increases with age, and DEXA scans are the gold standard for diagnosis. The proposed treatment includes lifestyle changes like proper nutrition and exercise, alongside fall prevention measures. Adequate calcium and vitamin D intake is essential. Predicted pharmacological options consist of bisphosphonates (e.g., Alendronate, Risedronate) and newer biologics like Denosumab, which are expected to help manage bone loss more effectively. Preventive measures should start early and involve maintaining bone health through regular physical activity and a balanced diet.

3. METHODOLOGY

3.1 Ultra-Sonometer Device

An ultra-sonometer with a quantitative ultrasound (QUS) measures the stiffness index (SI) at the heel for bone mineral density (BMD) evaluation. For the best ultrasound transmission, the patient's foot was placed on the apparatus while they were sitting, with the heel submerged in water. The apparatus used broadband ultrasound attenuation (BUA) and speed of sound (SOS) measurements to compute SI using the following formula:

$$SI=[(0.28\times SOS)+(0.67\times BUA)]-420$$

To provide T-scores for comparison to a healthy reference, SI was adjusted to set a young adult's value at 100. To evaluate bone health and fracture risk, SI is standardized to a T-score, and findings are compared to those of young individuals in good health [12].

3.2 Dataset

The bone mineral density real-time dataset has been collected from the Ultra-Sonometer sensor device. This dataset consists of three thousand and eight instances and 8 attributes. Attributes are patient ID, age, gender, height, weight, T-score, BMI, BMD, and scan site.

3.3 Data Preprocessing

Preprocessing of the osteoporosis dataset, which included characteristics like Patient ID, Age, Scan Site, BMI, BMD, Gender, Height, Weight, and T-Score, was done utilizing three crucial techniques:

3.3.1 Data cleaning

Mean or median imputation was used to address missing values in numerical fields (such as age, BMI, and BMD) while mode imputation was used for categorical variables (such as gender). To preserve the originality of the data, duplicates were found and eliminated using the patient ID.

3.3.2 Outlier detection

The Z-score and Interquartile Range (IQR) techniques were used to identify outliers. The dataset quality was maintained by removing or adjusting values that were outside the ±3 Z-score range or those that were more than $Q1-1.5\times IQR$ and $Q3+1.5\times IQR$.

3.3.3 Normalization

Numerical characteristics (Age, BMI, BMD, etc.) were subjected to min-max scaling to keep them within a 0–1 range. Furthermore, Z-score normalization was used to normalize characteristics with non-Gaussian distributions [3].

3.4 Proposed System

Figure 1 shows the proposed system architecture

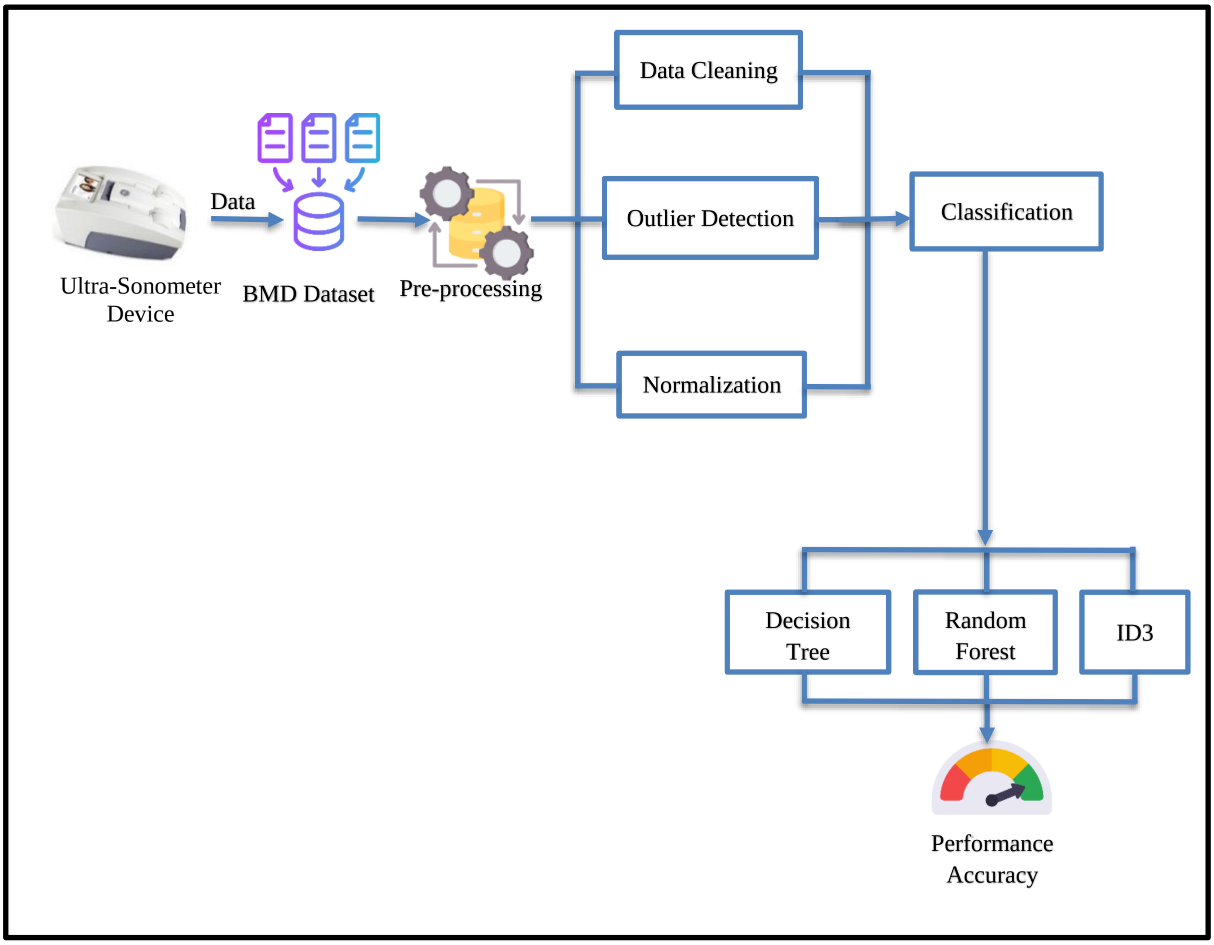

Figure 1: System Architecture

3.5 Classification Algorithm

A classification algorithm is a kind of machine learning algorithm that uses characteristics to group incoming data into one or more specified classes or categories. It is applied to jobs like spam identification, picture recognition, and medical diagnosis when the result is a distinct label.

3.5.1 Osteoporosis

Bone Mineral Density measures the concentration of minerals particularly calcium in bones. This measurement is expressed in grams per square centimeter(g/cm2) and helps to determine bone strength. Osteoporosis is characterized by reduced bone density and deteriorated bone structure. Which can be assessed by bone mineral density measurements. BMD is determining bone health and identifying stages of osteoporosis. Osteoporosis is classified into three categories based on the t-score: mild, severe, and normal. The T-Score value range is between -1.0 to -2.5 considered as the class mild. Although not yet recognized as osteoporosis, this category indicates decreased bone density. The T-Score value range is less than or equal to -2.5 considered as the class severe. It shows significant bone density loss and a higher risk of fractures. If the class is normal, then the T-score value ranges greater than or equal to -1.0. that indicates the bone density is good.

The dataset was divided into training and testing sets to evaluate model performance effectively, 70% split was used for the training model, while the remaining 30% was reserved for testing. This ensures that the model learns from a substantial portion of the data while retaining an adequate amount for the evaluation [5].

3.5.2 Decision Tree

A decision tree is a machine learning technique that divides data into subsets according to feature values and then utilizes a tree-like model of decisions and their potential outcomes to categorize the data. By categorizing people into risk groups, a decision Tree may forecast the chance of getting osteoporosis by analyzing patient data such as patient ID, age, weight, height, scan site, score, BMI, and bone mineral density.

Pseudocode

1. Start
2. Load Dataset - Features: [Feature1, Feature2, ..., Feature N] - Target: [Osteoporosis (Yes/No)]
3. Split Dataset - Divide the dataset into Training Set and Testing Set
4. Initialize Decision Tree - Create an empty tree
5. Train Decision Tree - For each feature in the dataset: - Compute the best split based on a criterion (e.g., Gini Index, Information Gain) - Create a decision node based on the best split - Recursively repeat the process for each child node
6. Test Decision Tree - For each sample in the Testing Set: - Traverse the tree based on the

feature values of the sample - Predict the target value
7. Evaluate Model - Calculate performance metrics
8. Output Results - Print performance metrics
9. End

3.5.3 Random Forest

An ensemble learning technique called a Random Forest builds many decision trees during training and combines their results to increase precision and reduce overfitting. To improve the accuracy and dependability in predicting the likelihood of developing osteoporosis, a Random Forest can combine forecasts from several decision trees using patient data such as PID, age, weight, height, scan site, score, and bone mineral density.

Pseudocode

1. Start
2. Load Dataset - Features: [Feature1, Feature2, ..., Feature N] - Target: [Osteoporosis (Yes/No)]
3. Split Dataset - Divide the dataset into Training Set and Testing Set
4. Initialize Random Forest with specified trees
   random_forest_model = RandomForestClassifier(n_estimators=100)
5. Train model on training data
   random_forest_model. Train (training_data, target=T-Score)
6. Predict osteoporosis risk on test data
   predictions_rf = random_forest_model. predict(test_data)
7. Evaluate model performance
   evaluate (predictions_rf, test_data_labels)
8. Output
9. End

3.5.4 ID3

In the ID3 (Iterative Dichotomiser 3) approach, the feature that maximizes information gain at each node is used to generate decision trees. The dataset is divided according to the property that yields the most reduction in entropy. By methodically selecting the most important elements, ID3 can build a decision tree for osteoporosis prediction using patient characteristics like age, weight, height, scan site, score, and bone mineral density. This tree can then be used to identify individuals who are more likely to develop the condition.

Pseudocode

1. Start
2. Load Dataset - Features: [Feature1, Feature2, ..., FeatureN] - Target: [Osteoporosis (Yes/No)]
3. Split Dataset - Divide the dataset into Training Set and Testing Set
4. Initialize ID3 model
   id3_model = ID3Classifier ()
5. Train model on training data
   id3_model. train (training_data, target=T-Score)
6. Predict osteoporosis risk on test data
   predictions_id3 = id3_model. predict (test data)
7. Evaluate model performance
   Evaluate (predictions_id3, test_data_labels)
8. output
9. End

4. EXPERIMENTS RESULTS

The osteoporosis disease was predicted by using the tree categorization algorithms, Decision Tree, Random Forest, and ID3. The performances of these algorithms are examined in this section. The following system specification was used for the execution of the algorithms. AMD pro-a4-4350b processor; 5 compute cores (2 CPU + 3 GPU); 2.50 GHZ is the base clock speed; 4 GB of RAM; 500 GB HDD is the hard drive; Windows 10 Pro are the operating systems and Python 3.8. software is used. The accuracy metric for the ID3 Tree, Random Forest, and Decision Tree classification algorithms is given in Table 1 and the same is presented in Figure 3. According to experimental results, the Decision Tree performs better than the ID3 Algorithm and Random Forest. Figure 2 shows the ROC curve.

All three algorithms Decision Tree, Random Forest, and ID3 achieved perfect classification results, with 100% of instances correctly classified. Each algorithm also had a True Positive Rate, Precision, and F-Measure of 1.00, indicating its flawless predictive capabilities. Figure 3 Shows the Accuracy Measure.

Table 1: Accuracy Measure for Osteoporosis Dataset

Algorithm	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)	TP Rate	Precision	F Measure
Decision Tree	100.00	0.00	1.00	1.00	1.00
Random Forest	100.00	0.00	1.00	1.00	1.00
ID3	100.00	0.00	1.00	1.00	1.00

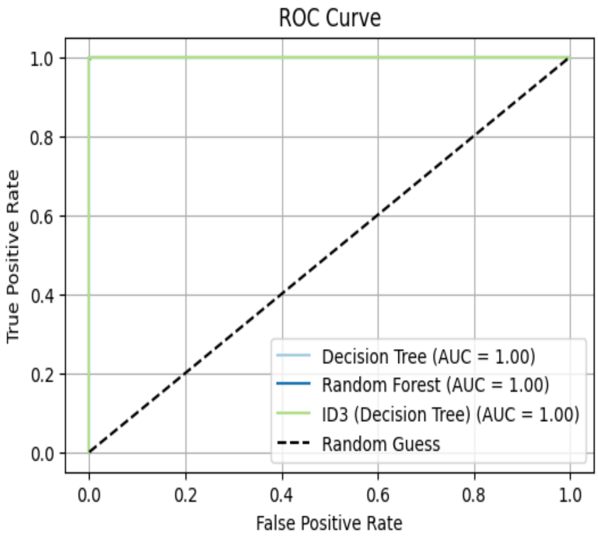

Figure 2: Roc Curve

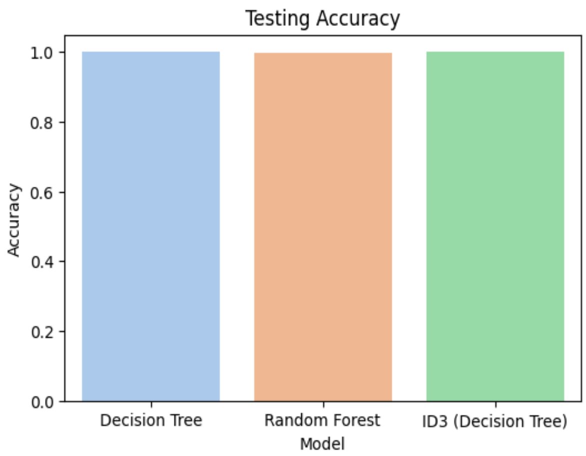

Figure 3: Accuracy Measure

The execution time required for the classification algorithm is given in Table 2 and its pictorial illustration is given in Figure 4. Decision Tree requires a longer execution time than Random Forest and ID3.

Table 2: Execution time Analysis for Osteoporosis Dataset

Algorithm	Execution Time in ms
Decision Tree	1.78
Random Forest	18.04
ID3	2.95

The analysis indicates that the Decision Tree algorithm is the fastest algorithm because the execution time required is 1.78 milliseconds, followed by ID3 at 2.95 milliseconds, and Random Forest, which takes the longest at 18.04 milliseconds. Figure 4 shows the Time execution time analysis.

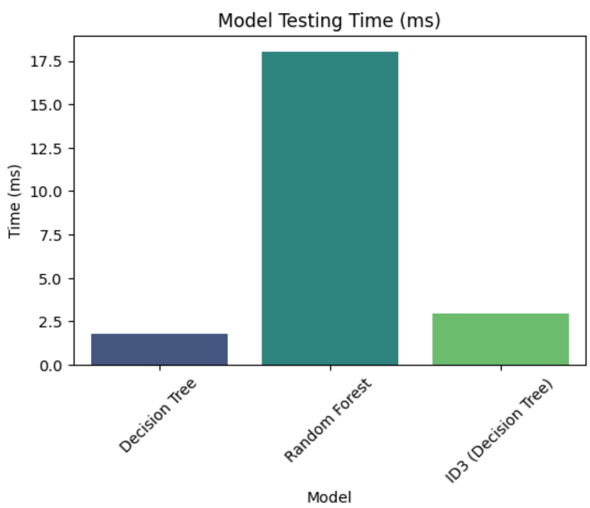

Figure 4: Execution Time Analysis

Table 3 provides information about the number of persons affected with the three levels of osteoporosis disease

Table 3: Classification of Osteoporosis Disease

Osteoporosis	Decision Tree	Random Forest	ID3
Mild	530	530	530
Normal	328	328	328
Severe	45	45	45

The results of the classification show that all three algorithms Decision Tree, Random Forest, and ID3 produced identical outcomes for the distribution of osteoporosis severity levels: categorizing 530 cases as Mild, 328 as Normal, and 45 as Severe.

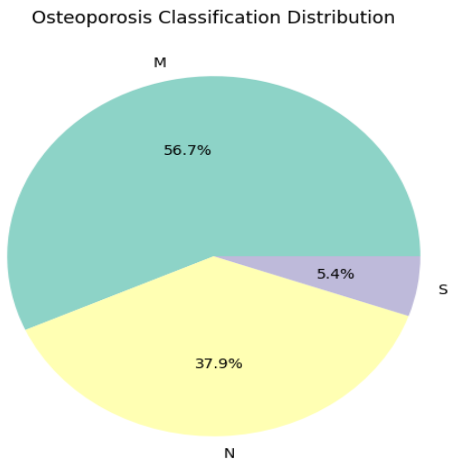

Figure 5: Osteoporosis Classification

Figure 5 shows how osteoporosis is categorized using methods like ID3, Random Forest, and Decision Trees. After examining the data, Decision Tree outperforms Random Forest and, ID3 as the overall best classification algorithm.

5. CONCLUSION

In the healthcare industry, the approach known as classification is mostly employed for illness prediction. To classify and predict osteoporosis, this work employed classification algorithms such as ID3, Random Forest, and Decision Tree. These algorithms' performances are compared based on execution time and classification accuracy. The Decision Tree classifier is regarded as the best method due to its maximum classification accuracy, Conversely, when evaluating the execution time, the Random Forest and ID3 classifiers need the shortest execution time. Future work can combine learning with real-time data to enhance prediction accuracy. Including diverse datasets will improve generalizability and could transform osteoporosis and treatment in telemedicine.

6. REFERENCES

[1] Alshymaa Alghamdi, "Healthcare Analytics: A Comprehensive Review", Engineering, Technology & Applied Science Research, Vol. 11, No.1, 2021,6650-6655.

[2] Michael J. Ward, "Applications of business analytics in healthcare", Volume 57, Issue 5, September–October 2014, Pages 571 582, https://doi.org/10.1016/j.bushor.2014.06.003

[3] EfatJabarpour, Amin Abedini, AbbasaliKeshtkar, "Osteoporosis Risk Prediction Using Data Mining Algorithms", Journal of Community Health Research, 2020, 9(2): 69 - 80.

[4] Barbara C. Silva, Susan B. Broy, Stephanie Boutroy, "Fracture Risk Prediction by Non-BMD DXA Measures: the 2015 ISCD Official Positions Part 2: Trabecular Bone Score", Volume 18, Issue 3, July 2015, Pages 309-330, https://doi.org/10.1016/j.jocd.2015.06.008,

[5] Xiang-He Menga, Xiang-Ding Chena, Jonathan Greenbaumb," Integration of summary data from GWAS and eQTL studies identified novel causal BMD genes with functional predictions" Accepted 11 May 2018, https://doi.org/10.1016/j.bone.2018.05.012.

[6] KottaimalaiRamaraj, "Sensors for Bone Mineral Density Measurement to Identify the Level of Osteoporosis: A Study", 20-22 January 2022, 10.1109/ICSSIT53264.2022.9716536.

[7] Wilson Ong, et al., Artificial Intelligence Applications for osteoporosisis classification using computed tomography, Bioengineering 2023, 10, 1364. 23 November 2023, https://doi.og/10.3390/bioengineering10121364.

[8] Kaname Miura, et al, "Osteoporosis prediction using machine-learned optical bone densitometry data", 10 March 2023, Biomedical Engineering Society, annals of biomedical engineering (2024) 52:396-405.

[9] Rahul Paul, et al., Make Your Bone Great Again: A study on Osteoporosis Classification, doi: 10.4550/arxiv.1707.05385,29september 2018.

[10] Nader Salari, et al., The global prevalence of osteoporosis in the world: a comprehensive systematic review and meta-analysis, Journal of Orthopaedic Surgery and Research, https://doi.og/10.1186/s13018-021-02772-0. 2021.

[11] Ferdou HS, et al, Osteoporosis: A review, Bird medical journal -June 2016, https:// www.researchgate.net/publication/304495902.

[12] Vishwanath SB, Et al., Correlation of periodontal status bone mineral density in postmenopausal women: A digital radiographic and quantitative ultrasound study, Indian journal of dental Research, 22(2), Doi:10.4103/0970-9290.84303.