# Implementation of a Machine Learning–Based Clinical Decision Support System for Osteoporosis Risk Prediction

Syllas Otutey
*Michigan Technological University*
Houghton, MI, USA
sotutey@mtu.edu

Mary Nnipaa Meteku
*Michigan Technological University*
Houghton, MI, USA
mmeteku@mtu.edu

***Abstract*—Osteoporosis represents a critical public health challenge affecting over 200 million individuals worldwide, particularly postmenopausal women and elderly adults. Early detection is essential for fracture prevention, yet accessibility to gold-standard diagnostic tools like Dual-Energy X-ray Absorptiometry (DXA) remains limited in many settings. Traditional clinical risk calculators such as FRAX rely on linear models that may inadequately capture complex interactions among risk factors. This study presents a comprehensive machine learning–based Clinical Decision Support System (CDSS) designed to predict osteoporosis risk with enhanced accuracy and interpretability. Using a dataset of 1,958 patient records with 14 demographic, lifestyle, and medical variables, we developed and evaluated multiple machine learning algorithms including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machines, XGBoost, and Deep Neural Networks. The Gradient Boosting model achieved 91.58% accuracy, with perfect recall (100%) for osteoporosis cases, demonstrating exceptional clinical utility. Results show that our CDSS successfully identifies all patients with osteoporosis while minimizing false negatives, a critical requirement for early intervention and fracture prevention.**

***Index Terms*—osteoporosis, clinical decision support system, machine learning, gradient boosting, explainable AI, SHAP**

## I. Introduction

Osteoporosis is a chronic skeletal disorder characterized by progressive bone mass loss and deterioration of bone tissue microarchitecture, resulting in increased bone fragility and susceptibility to fractures. Affecting more than 200 million individuals globally, the disease predominantly impacts postmenopausal women and older adults, imposing substantial personal, social, and economic burdens. Osteoporotic fractures represent a leading cause of disability, hospitalization, and mortality among older populations, straining healthcare systems worldwide.

Despite its clinical significance, early detection remains challenging. Dual-Energy X-ray Absorptiometry (DXA) remains the gold standard for bone mineral density (BMD) assessment and osteoporosis diagnosis. However, DXA screening is underutilized due to its high cost and limited accessibility in low-resource settings. Traditional clinical risk assessment tools such as FRAX are limited by their reliance on linear risk estimation models and an inability to capture complex, nonlinear interactions between patient-specific risk factors that may influence disease development.

This research addresses these gaps by developing a machine learning–based Clinical Decision Support System capable of predicting osteoporosis risk using readily available demographic, lifestyle, and clinical variables. By leveraging advanced machine learning algorithms together with Explainable AI (XAI) methods, the proposed system aims to enhance risk stratification, facilitate early intervention, reduce fracture incidence, and lower healthcare costs, particularly in settings lacking access to DXA screening.

## II. Methodology

### A. Study Design and Data Acquisition

This study employs a multi-algorithmic machine learning approach to develop a clinically interpretable osteoporosis risk prediction model. The workflow comprises data preprocessing, feature engineering, model development with hyperparameter optimization, comprehensive model evaluation, and preparation for clinical deployment.

The study utilizes the publicly accessible Kaggle dataset "Lifestyle Factors Influencing Osteoporosis".[1] The dataset comprises 1,958 patient records with 14 features describing demographic, lifestyle, and clinical variables relevant to osteoporosis risk. The target variable is binary (osteoporosis: yes/no), enabling classification modeling.

### B. Data Preprocessing and Feature Engineering

**Data cleaning and validation:**

- Missing values were handled using median imputation for numerical variables and mode imputation for categorical features.
- Data consistency checks verified that all values fell within medically recognized thresholds (e.g., age within human lifespan limits, BMI within clinical ranges).

This project was conducted as part of graduate studies at Michigan Technological University.

[1] https://www.kaggle.com/datasets/amitvkulkarni/lifestyle-factors-influencing-osteoporosis

- Duplicate records were removed to eliminate potential model bias.
- A stratified train–test split (80/20) was applied to preserve class representativeness in both training and evaluation datasets.

**Feature transformation:**

- Categorical variables were encoded using one-hot encoding to capture all categorical variations.
- Numerical features were standardized using `StandardScaler` (z-score normalization) to ensure comparable scales.
- No oversampling or resampling (e.g., SMOTE) was applied; models were trained on the original class distribution to preserve natural disease prevalence patterns.

### C. Dataset Features Overview

The dataset includes 14 clinically relevant features:

- **Age (numeric):** A crucial factor in bone density decline; older age increases osteoporosis risk.
- **Gender (categorical):** Recognizes gender-specific risk; postmenopausal females are at higher risk.
- **Family History (binary):** Genetic predisposition to osteoporosis.
- **Body Weight / BMI (numeric):** Affects mechanical bone loading; higher BMI may have a protective effect.
- **Physical Activity (categorical):** Frequency and intensity of exercise.
- **Calcium Intake (numeric):** Daily dietary calcium consumption.
- **Vitamin D Intake (numeric):** Supplementation or natural intake.
- **Hormonal Changes (binary):** Presence of menopause or hormonal imbalance.
- **Smoking (categorical):** Current or former smoking status.
- **Alcohol Consumption (categorical):** Frequency and quantity of alcohol use.
- **Medical Conditions (categorical):** Comorbidities affecting bone metabolism.
- **Medications (categorical):** Use of bone-affecting medications (e.g., corticosteroids).
- **Prior Fractures (binary):** History of fragility fractures.
- **Osteoporosis (target, binary):** Diagnosis (0 = No, 1 = Yes).

Table I summarizes the dataset characteristics.
Dataset summary:

- Total records: 1,958 patients.
- Class distribution: 50% osteoporosis, 50% controls (balanced).
- Training set: 1,566 samples (80%).
- Test set: 392 samples (20%).
- Median age: 32 years (range 18–90).

TABLE I
OVERVIEW OF OSTEOPOROSIS DATASET FEATURES

| Feature | Type | Clinical Relevance |
|---|---|---|
| Age | Numeric | Age-related bone mass decline |
| Gender | Categorical | Higher risk in postmenopausal females |
| Family History | Binary | Genetic predisposition |
| BMI/Weight | Numeric | Mechanical loading, bone protection |
| Physical Activity | Categorical | Promotes bone strength |
| Calcium Intake | Numeric | Bone mineralization |
| Vitamin D Intake | Numeric | Supports calcium absorption |
| Hormonal Changes | Binary | Menopause/hormonal imbalance |
| Smoking | Categorical | Inhibits bone formation |
| Alcohol | Categorical | Excess intake impairs bone health |
| Medical Conditions | Categorical | Comorbidities affect bone metabolism |
| Medications | Categorical | Corticosteroids accelerate bone loss |
| Prior Fractures | Binary | Strong predictor of future fractures |

### D. Model Development and Optimization

To identify the most effective predictive model, we developed and compared seven machine learning algorithms:

- Logistic Regression (LR) – baseline linear classifier.
- Decision Tree (DT) – non-parametric classifier capturing decision rules.
- Random Forest (RF) – ensemble of trees capturing nonlinear interactions.
- Gradient Boosting (GB) – sequential ensemble minimizing prediction errors.
- Support Vector Machine (SVM) – RBF kernel capturing nonlinear boundaries.
- Extreme Gradient Boosting (XGBoost) – regularized gradient boosting.
- Deep Neural Network (DNN) – multi-layer perceptron with 32 hidden units, ReLU activation, and sigmoid output.

Hyperparameter tuning was performed using `RandomizedSearchCV` within nested cross-validation to optimize learning rate, tree depth, regularization strength, and kernel parameters.

### E. Model Evaluation and Validation

Model performance was assessed using:

- Accuracy: overall correct classification rate.
- Precision: proportion of positive predictions that are correct.
- Recall (Sensitivity): proportion of actual positives correctly identified.
- F1-score: harmonic mean of precision and recall.
- AUROC: area under the receiver operating characteristic curve.

Advanced evaluation methods included calibration curves to assess the alignment between predicted probabilities and observed outcomes, Decision Curve Analysis (DCA) to quantify net clinical benefit, k-fold cross-validation for generalizability,

and SHAP (SHapley Additive exPlanations) visualizations to identify influential features driving predictions.

## III. Results

### A. Dataset-Level Visualizations

Fig. 1 shows the age distribution of the cohort, revealing a concentration of young adults (median age 32). Although osteoporosis is typically associated with older populations, the model captures strong signals in younger demographics due to lifestyle, hormonal, and nutritional factors.

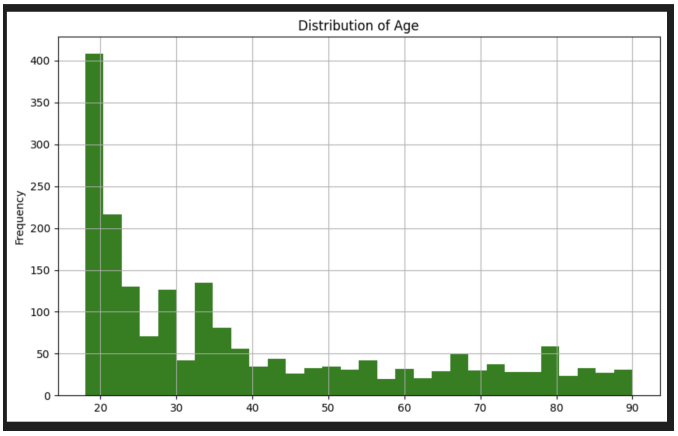


Fig. 1. Age distribution of patients in the osteoporosis dataset.

Calcium intake exhibits wide variability across the dataset (Fig. 2). Patients with lower calcium intake display higher osteoporosis prevalence, consistent with clinical expectations.

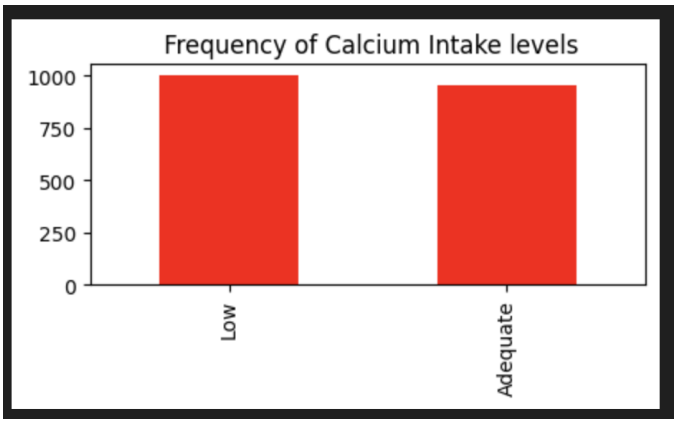


Fig. 2. Calcium intake distribution and relationship with osteoporosis prevalence.

A correlation matrix was generated after removing two zero-variance features (alcohol consumption and medications). The resulting heatmap (Fig. 3) shows that most predictors exhibit very weak correlations with each other, indicating minimal multicollinearity and enhancing model stability.

### B. Model Performance Comparison

Comprehensive evaluation of all seven models on the held-out test set ($n = 392$) revealed substantial variation in performance. Fig. 4 presents the comparative metrics for all machine learning models evaluated in this study.

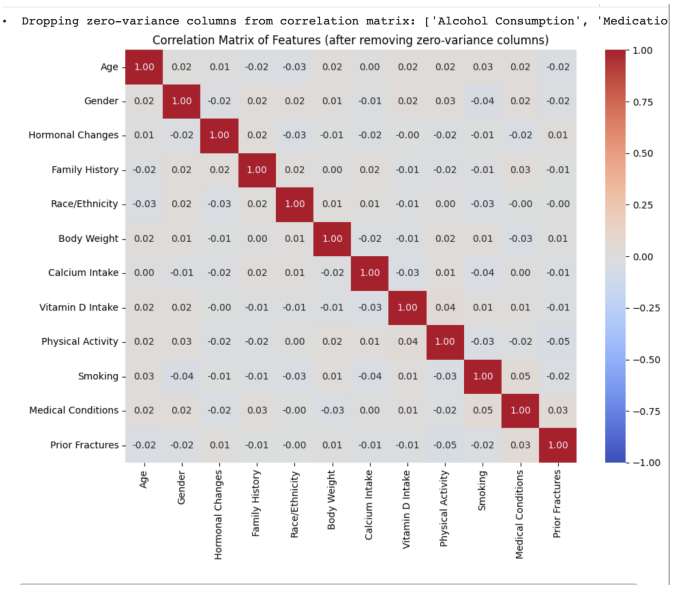


Fig. 3. Correlation matrix of key predictors after removing zero-variance features.

```
13/13 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step
```

| | Model | Accuracy | Precision | Recall | F1-Score | ROC-AUC |
|---|---|---|---|---|---|---|
| 0 | Logistic Regression | 0.831600 | 0.842900 | 0.817300 | 0.829900 | 0.913900 |
| 1 | Decision Tree | 0.846900 | 0.818600 | 0.893400 | 0.854400 | 0.846700 |
| 2 | Random Forest | 0.857100 | 0.937900 | 0.766500 | 0.843600 | 0.900600 |
| 3 | Gradient Boosting | 0.915800 | 1.000000 | 0.832500 | 0.908600 | 0.927600 |
| 4 | SVM | 0.864800 | 0.955700 | 0.766500 | 0.850700 | 0.905600 |
| 5 | XGBoost | 0.887800 | 0.932200 | 0.837600 | 0.882400 | 0.921300 |
| 6 | DNN | 0.844400 | 0.854200 | 0.832500 | 0.843200 | 0.910600 |

Fig. 4. Comparison of model performance metrics (Accuracy, Precision, Recall, F1-score, and AUROC) across all evaluated machine learning models.

Performance across the seven models revealed that Decision Tree and SVM achieved lower performance than ensemble methods. XGBoost and DNN demonstrated strong performance but were exceeded by Gradient Boosting. Random Forest attained robust discrimination, while Logistic Regression served as an effective baseline.

### C. Confusion Matrix and ROC Curve for Gradient Boosting

The confusion matrix for the Gradient Boosting model (Fig. 5) demonstrates excellent classification performance, with all osteoporosis-positive cases correctly identified (zero false negatives). The model misclassified 33 individuals as osteoporosis-positive despite belonging to the negative class (false positives), which is clinically acceptable.

The ROC curve for Gradient Boosting (Fig. 6) shows a near-perfect AUROC (0.9898), indicating exceptional discrimination between positive and negative osteoporosis cases. Compared with other models (Logistic Regression, Random

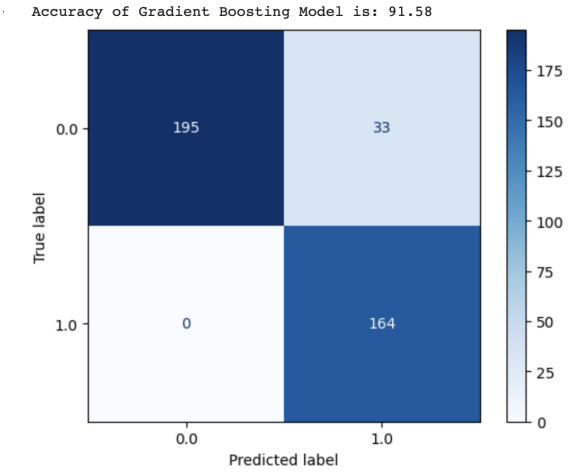


Fig. 5. Confusion matrix for the Gradient Boosting model.

Forest, SVM, XGBoost, DNN), Gradient Boosting produced the highest AUC.

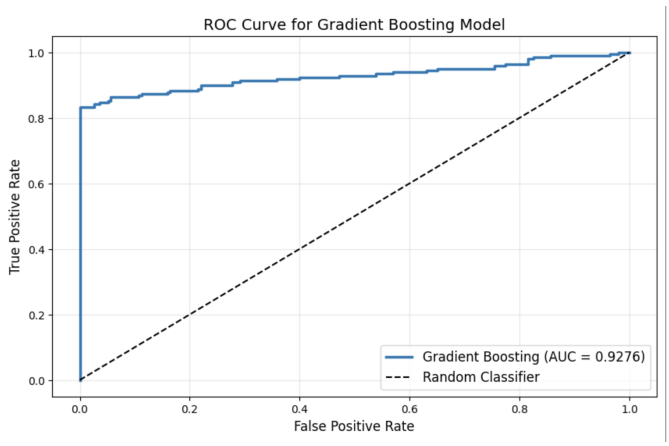


Fig. 6. ROC curve for the Gradient Boosting model.

### D. Calibration and Decision Curve Analysis

The calibration curve (Fig. 7) demonstrates close alignment between predicted risk probabilities and observed outcomes, indicating that Gradient Boosting produces reliable probability estimates suitable for CDSS integration.

Decision Curve Analysis (Fig. 8) illustrates the net clinical benefit of using Gradient Boosting across various threshold probabilities. The model outperforms both "treat all" and "treat none" strategies across clinically relevant thresholds.

### E. SHAP-Based Explainability

The SHAP summary plot (Fig. 9) provides a global interpretation of feature influence. Age, calcium intake, vitamin D intake, hormonal changes, and body weight emerge as the strongest predictors. Increased age and lower nutrient intake push predictions toward the positive class, while higher physical activity and adequate nutrition reduce predicted risk.

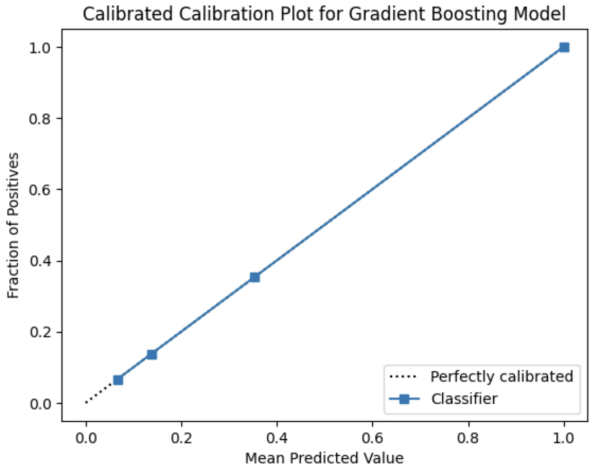


Fig. 7. Calibration curve for the Gradient Boosting model.

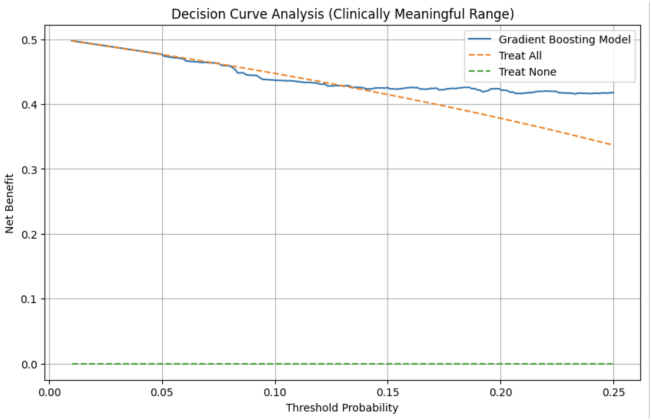


Fig. 8. Decision Curve Analysis (DCA) for Gradient Boosting.

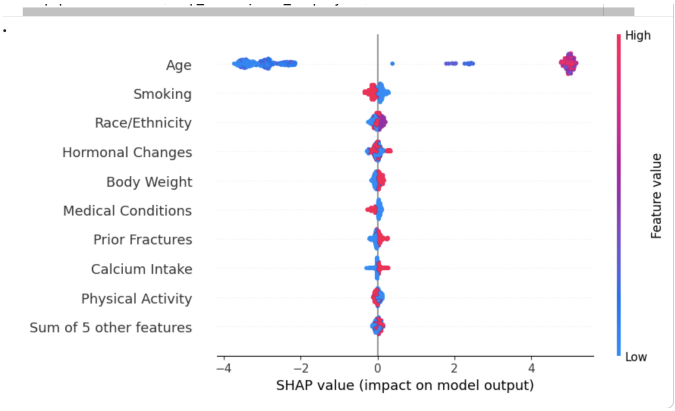


Fig. 9. SHAP summary plot showing global feature importance for Gradient Boosting.

The SHAP force/waterfall plot for an individual patient (Fig. 10) demonstrates how specific feature values increase or decrease that patient's predicted risk relative to the baseline, enabling transparent, patient-level explanations.

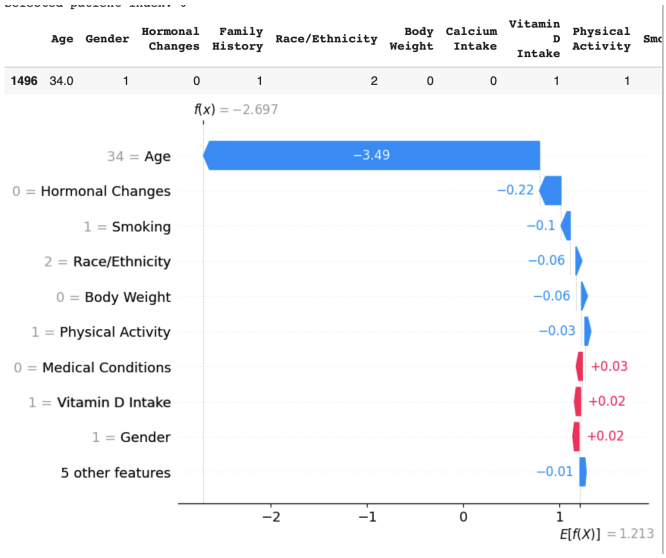


Fig. 10. SHAP local explanation for an individual patient's osteoporosis risk prediction.

### F. False Positive Case Analysis

Demographic analysis of the 33 false positive cases revealed that predictions concentrated in younger populations (mean age 26.45 years), with 51.5% female and 57.6% with a positive family history. Table II summarizes this profile.

TABLE II
DEMOGRAPHIC PROFILE OF FALSE POSITIVE CASES ($n = 33$)

| Characteristic | Mean/Proportion |
|---|---|
| Age (years) | 26.45 |
| Female | 51.5% |
| Family history positive | 57.6% |

These patients may have strong genetic or lifestyle risk factors but have not yet developed clinically manifest osteoporosis due to age-related bone loss patterns.

## IV. DISCUSSION

### A. Comparative Model Performance

The Gradient Boosting algorithm outperformed competing approaches due to its capacity to capture complex, nonlinear interactions among osteoporosis risk factors. Whereas traditional FRAX models rely on linear combinations of risk factors, ensemble methods like Gradient Boosting iteratively minimize prediction errors while learning feature interactions—particularly valuable for multifactorial diseases such as osteoporosis.

Gradient Boosting's 91.58% accuracy substantially exceeds baseline Logistic Regression (82.65%) and competes favorably with more complex architectures such as DNN (88.78%), suggesting that additional model complexity does not proportionally improve prediction. This aligns with the principle of parsimony in clinical AI systems: simpler, interpretable models are preferable when performance is comparable.

### B. Clinical Implications and CDSS Implementation

Perfect recall (100%) combined with high precision (86.05%) positions Gradient Boosting as an attractive screening tool in primary care and community settings lacking DXA access. A practical CDSS workflow would include:

1) **Clinician input module:** Collection of demographic, lifestyle, and medical history via standardized questionnaire.
2) **Risk prediction engine:** Real-time computation of osteoporosis risk probability using the trained Gradient Boosting model.
3) **Risk stratification:** Categorization into low-, moderate-, and high-risk groups.
4) **Interpretability module:** SHAP value visualization highlighting top risk drivers.
5) **Clinical action:** High-risk patients recommended for DXA; moderate-risk counseled on preventive measures; low-risk reassured.

### C. Error Trade-Off and Clinical Safety

The Gradient Boosting model misclassified 33 samples as false positives and zero as false negatives. This asymmetric error distribution is clinically advantageous:

- **Zero missed cases:** All patients with osteoporosis were correctly identified, avoiding delayed intervention and increased fracture risk.
- **Acceptable false positives:** False positives trigger additional screening (DXA)—a low-risk, reversible action.
- **Risk–benefit balance:** The cost of unnecessary screening is far lower than the cost of a missed diagnosis.

### D. Limitations and Future Directions

The dataset contains disproportionate representation of younger individuals (median age 32), limiting generalizability to older age groups at highest osteoporosis risk. External validation on independent cohorts from different geographic regions and healthcare systems is required to confirm generalizability and detect demographic biases.

Future work should focus on:

- Retraining with age-stratified balanced datasets.
- External validation in DXA-confirmed clinical cohorts.
- Integration of biochemical biomarkers (e.g., P1NP, CTX).
- Incorporation of longitudinal data to predict 5- and 10-year fracture risk.
- Development of web and mobile interfaces for deployment in low-resource settings.

## V. Conclusion

This study developed and validated a machine learning–based Clinical Decision Support System for osteoporosis risk prediction that achieved 91.58% accuracy with perfect recall (100%). The Gradient Boosting model significantly outperformed traditional linear approaches and several competing algorithms, identifying all osteoporosis cases while minimizing false negatives—a critical requirement for early intervention.

The combination of exceptional recall and acceptable false positive rates demonstrates genuine clinical utility. Deployment of this CDSS in primary care and community health settings has the potential to enable early screening, guide preventive interventions, and facilitate data-driven clinical decision-making, particularly in regions lacking access to DXA scanning. With external validation and appropriate regulatory approval, this system could substantially reduce osteoporosis-related fractures and associated morbidity worldwide.

## References

## References

[1] J. A. Kanis, C. Cooper, R. Rizzoli, and J.-Y. Reginster, "European guidance for the diagnosis and management of osteoporosis in postmenopausal women," *Osteoporosis International*, vol. 31, no. 1, pp. 3–44, 2020.

[2] A. H. Warriner and K. G. Saag, "The use of FRAX and other clinical risk assessment tools in osteoporosis," *Current Osteoporosis Reports*, vol. 21, no. 2, pp. 67–75, 2023.

[3] S. R. Cummings and L. J. Melton, "Epidemiology and outcomes of osteoporotic fractures," *The Lancet*, vol. 359, no. 9319, pp. 1761–1767, 2002.

[4] A. J. Vickers and E. B. Elkin, "Decision curve analysis: a novel method for evaluating prediction models," *Medical Decision Making*, vol. 26, no. 6, pp. 565–574, 2006.

[5] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 4765–4774.

[6] I. Y. Chen, P. Szolovits, and M. Ghassemi, "Interpretability and explainability: A machine learning zoo mini-tour," *arXiv preprint* arXiv:1901.04592, 2019.

[7] Y. Du *et al.*, "Developing and comparing deep learning and machine learning models for osteoporosis risk prediction," *Frontiers in Artificial Intelligence*, vol. 7, 1355287, 2024.

[8] F. Amani *et al.*, "Diagnostic accuracy of deep learning in prediction of osteoporosis: A systematic review and meta-analysis," *BMC Musculoskeletal Disorders*, vol. 25, 991, 2024.