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

Name:Syllas Otutey,Mary Nnipaa Meteku

course/ Instructor :SAT 5141/Guy Hembroff

Date:11/04/2025

Introduction

Osteoporosis affects over **200 million people worldwide** and causes more than **8.9 million fractures each year**, or **one fracture every three seconds**. These fractures lead to disability, loss of independence, and high healthcare costs. Although early detection is crucial, screening tools like DXA scans are often underused due to limited access and cost. Because osteoporosis risk is influenced by lifestyle, nutrition, hormones, and demographics, machine learning offers a promising way to identify at-risk individuals before serious fractures occur.

Problem Statement

Despite its global impact, osteoporosis often remains undiagnosed until a fracture occurs. Traditional screening tools cannot fully capture the complex combination of factors that contribute to risk. There is a clear need for accessible, accurate, and data-driven methods for early osteoporosis detection.

Research Objective

This study aims to develop a machine learning–based Clinical Decision Support System that predicts osteoporosis risk using demographic, lifestyle, nutritional, and clinical features. By comparing multiple algorithms and incorporating calibration, decision-curve analysis, and explainable AI, the goal is to identify a model that is accurate, interpretable, and clinically useful for early intervention.

Literature Review

Global Burden & Need for Early Detection

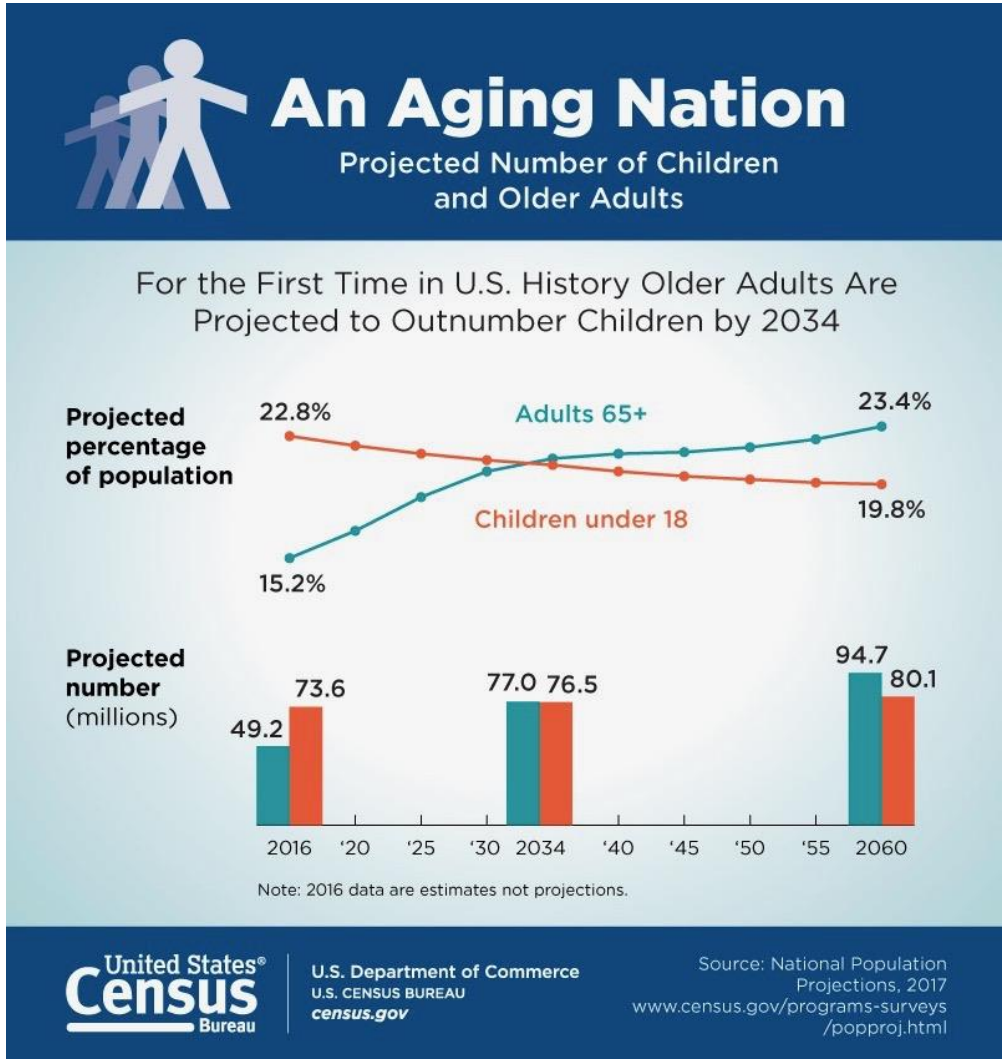
- Over **200 million people** worldwide are affected by osteoporosis.
- Causes **8.9 million fractures annually** one every three seconds.
- Early diagnosis remains low due to limited use of DXA scans and under-screening (Kanis et al., 2020).

Limitations of Traditional Risk Tools

- Tools like **FRAX** rely on linear clinical variables.
- Poor at capturing **nonlinear interactions** among age, hormones, nutrition, lifestyle, and comorbidities.
- Reduced predictive accuracy across diverse populations (Warriner & Saag, 2023).

Explainable AI (XAI) for Clinical Trust

- SHAP provides transparency by showing how each feature contributes to predictions.
- Supports clinician trust and helps validate biomedical relevance (Lundberg & Lee, 2017).



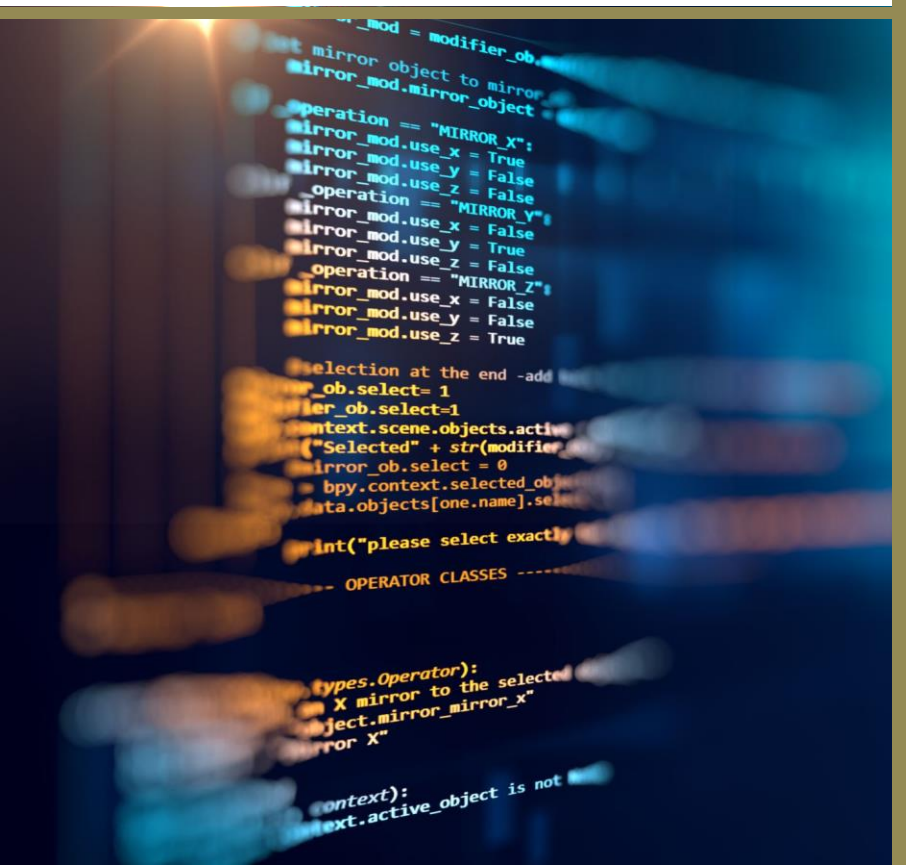
Dataset description

- Kaggle dataset: Lifestyle Factors Influencing Osteoporosis
- 1,958 adults, balanced classes
- Represents general population
- Useful for low-resource settings
- Link to dataset : https://www.kaggle.com/datasets/amitvkulkarni/lifestyle-factor s-influencing-osteoporosis

❖ Features include:

- ➢ Age, Weight, Calcium Intake
- ➢ Vitamin D, Physical Activity
- ➢ Hormonal Changes, Smoking
- ➢ Race, Family History
- ➢ Medical Conditions, Prior Fractures

Data Preprocessing and Feature Engineering



Data Preprocessing

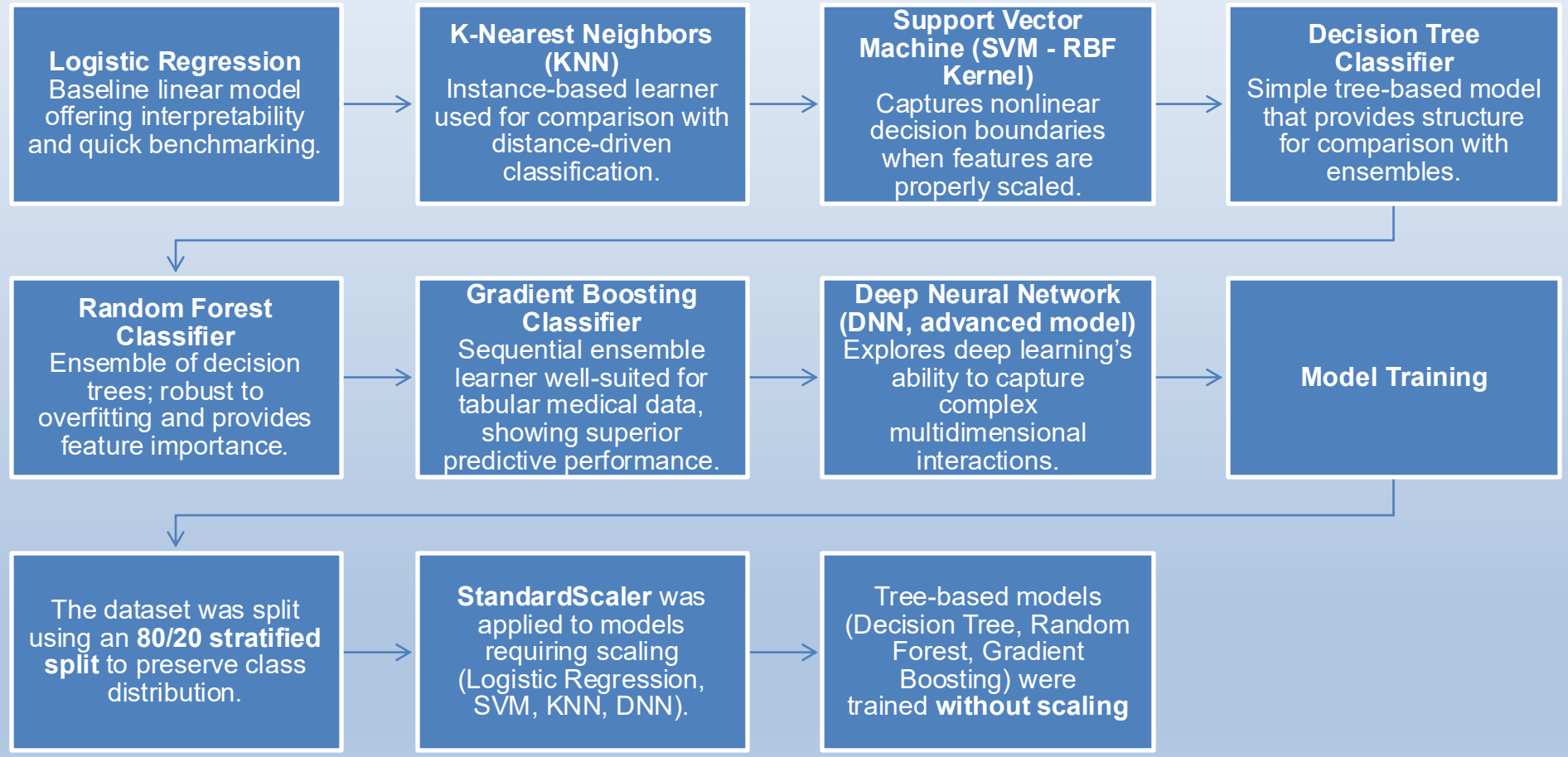
- Validated all 1,958 records and removed the non-informative *Id* column.
- Cleaning: Handling missing values and Removing outliers
- Normalization: Scaling numerical features for optimal model performance
- Categorical encoding: Converting categorical variables into numerical format using one-hot encoding

Feature Engineering

Cleaned binary fields such as Family History, Prior Fractures, and Hormonal Changes for consistent encoding.

- Verified that the target variable “Osteoporosis” was encoded as 0 = No, 1 = Yes.
- Performed SHAP analysis on the Gradient Boosting model, confirming clinically consistent predictors (e.g., *Age, Calcium Intake, Vitamin D Intake, Hormonal Changes, Physical Activity*).

Model Selection and Training



Evaluation Metrics

Accuracy: Measures the overall correctness of the model in predicting both healthy and at-risk individuals.

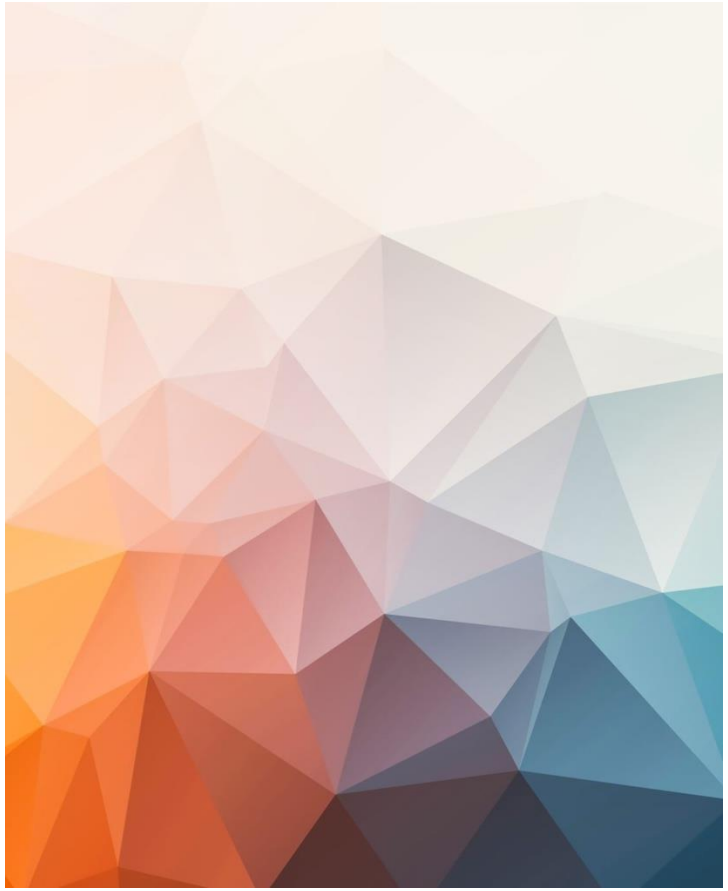
Precision: Ensures the model correctly identifies patients with osteoporosis **without over-predicting false positives**, which is important to avoid unnecessary clinical follow-ups.

Recall (Sensitivity): Captures the model's ability to detect actual osteoporosis cases, helping minimize missed diagnoses.

AUC–ROC: Evaluates how well the model distinguishes between individuals **with and without osteoporosis** across all classification thresholds.

Calibration Curve: Assesses the agreement between **predicted risk** and **actual observed outcomes**, ensuring probability estimates are reliable.

Decision Curve Analysis (DCA): Determines the **net clinical benefit** of using the model at different risk thresholds, showing whether predictions meaningfully improve clinical decision-making



```python
import pandas as pd # data processing
import seaborn as sns # Plots
import matplotlib.pyplot as plt # Plots
import warnings
from sklearn.isotonic import IsotonicRegression
# Data processing
from sklearn.model_selection import train_test_split, cross_val_score, GridSearchCV
from sklearn.preprocessing import LabelEncoder
# Evaluation metrics
from sklearn.metrics import accuracy_score, confusion_matrix, ConfusionMatrixDisplay, classification_report
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score
# ML models
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from xgboost import XGBClassifier

# Deep Learning
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
warnings.filterwarnings("ignore")
```

Loading of Data

```python
# Data from Kaggle
df = pd.read_csv('osteoporosis_clean.csv')

# First 5 rows
df.head()
```

```python
# Label encoding for categorical variables (exclude target)
le_dict = {}
target = "Osteoporosis"

for col in features_cat.columns:
    if col != target:
        le = LabelEncoder()
        df[col] = le.fit_transform(df[col].astype(str))
        le_dict[col] = le
```

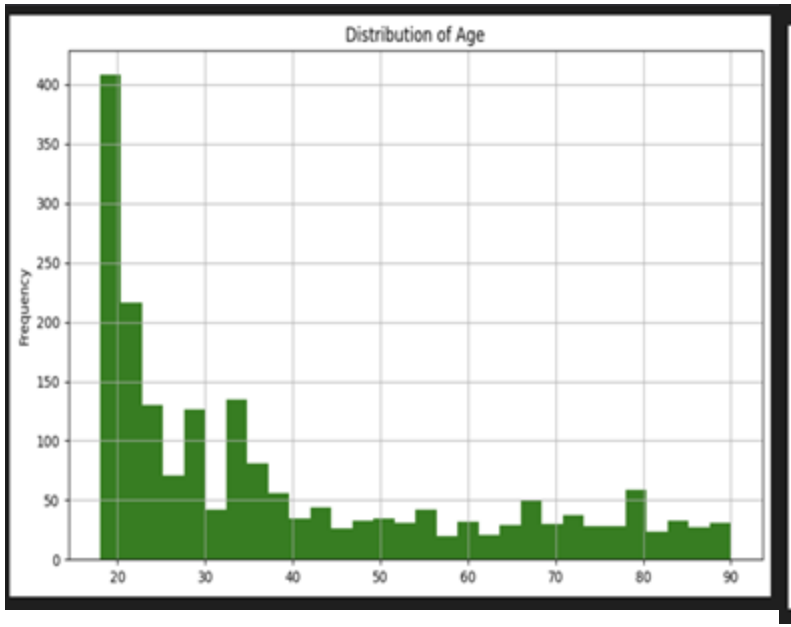
```python
# Split data into train & test sets
df_train, df_test = train_test_split(df, test_size= 0.2, random_state=24)
# Save labels and features seperately
train_labels = df_train['Osteoporosis']
train_features = df_train.drop(columns = ('Osteoporosis'))
# Repeat for test set
test_labels = df_test['Osteoporosis']
test_features = df_test.drop(columns = ('Osteoporosis'))
```

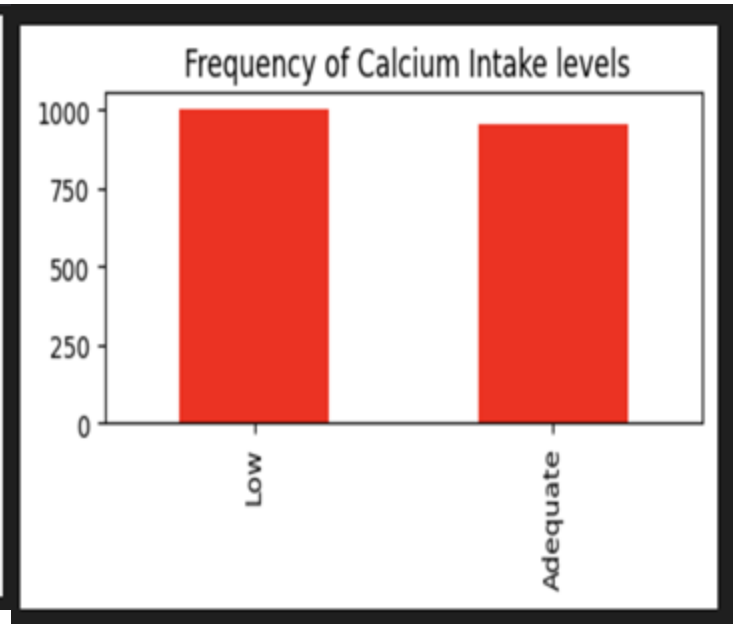
```python
# Models to compare
models = {
    "Logistic Regression" : LogisticRegression(max_iter=1000),
    "Decision Tree" : DecisionTreeClassifier(random_state = 24),
    "Random Forest" : RandomForestClassifier(random_state = 24),
    "Gradient Boosting" : GradientBoostingClassifier(random_state = 24),
    "SVM" : SVC(),
    "XGBoost" : XGBClassifier(random_state = 24),
    "DNN" : Sequential([
        Dense(32, activation='relu', input_shape=(train_features.shape[1],)),
        Dense(1, activation='sigmoid')
    ])
}
```

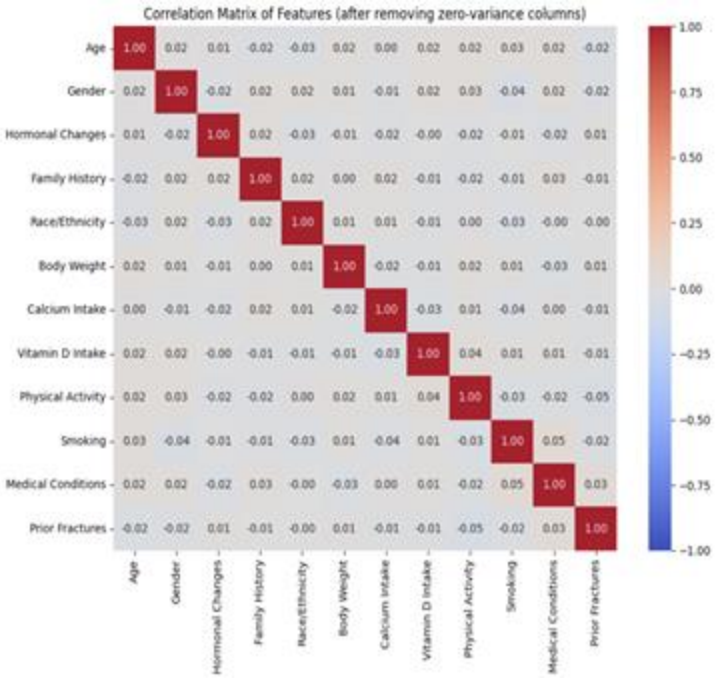
```python
# Results of each model trained to train set
results = []
for model_name, model in models.items():
    if isinstance(model, Sequential):  # Check if the model is a Keras Sequential model
        model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy']) # Compile the model
        model.fit(train_features, train_labels, epochs=20, verbose=0) # Train the DNN
        _, acc = model.evaluate(test_features, test_labels, verbose=0) # Evaluate the DNN
        acc *= 100
    else:
        model.fit(train_features, train_labels)
        y_preds = model.predict(test_features)
        acc = accuracy_score(test_labels, y_preds) * 100
    results.append((model_name, round(acc, 2)))
```

```python
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder

# Start from dataset
corr_df = df.copy()

# 1. Drop ID and target
cols_to_drop = []
for c in ["Id", "Osteoporosis"]:
    if c in corr_df.columns:
        cols_to_drop.append(c)

corr_df = corr_df.drop(columns=cols_to_drop, errors="ignore")

# Encode categorical columns before computing correlation
for col in corr_df.select_dtypes(include='object').columns:
    le = LabelEncoder()
    corr_df[col] = le.fit_transform(corr_df[col].astype(str))

# 2. Drop zero-variance columns (Alcohol Consumption, Medications, etc.)
zero_var_cols = [c for c in corr_df.columns if corr_df[c].nunique() <= 1]
print("Dropping zero-variance columns from correlation matrix:", zero_var_cols)

corr_df = corr_df.drop(columns=zero_var_cols, errors='ignore')

# 3. Compute correlation on remaining numeric features
corr = corr_df.corr()

plt.figure(figsize=(18, 8))
sns.heatmap(corr, annot=True, fmt=".2f", cmap="coolwarm", vmin=-1, vmax=1)
plt.title("Correlation Matrix of Features (after removing zero-variance columns)")
plt.tight_layout()
plt.show()
```

Model Algorithm







Feature Patterns & Distributions

- Age distribution: Ages mostly fall within a younger to middle-aged range.

•Fewer older adults → explains lower osteoporosis prevalence.
•Useful for understanding population characteristics.

- Calcium intake: Intake varies widely across participants.

•Many individuals fall below recommended levels.
•Important modifiable factor influencing osteoporosis risk.

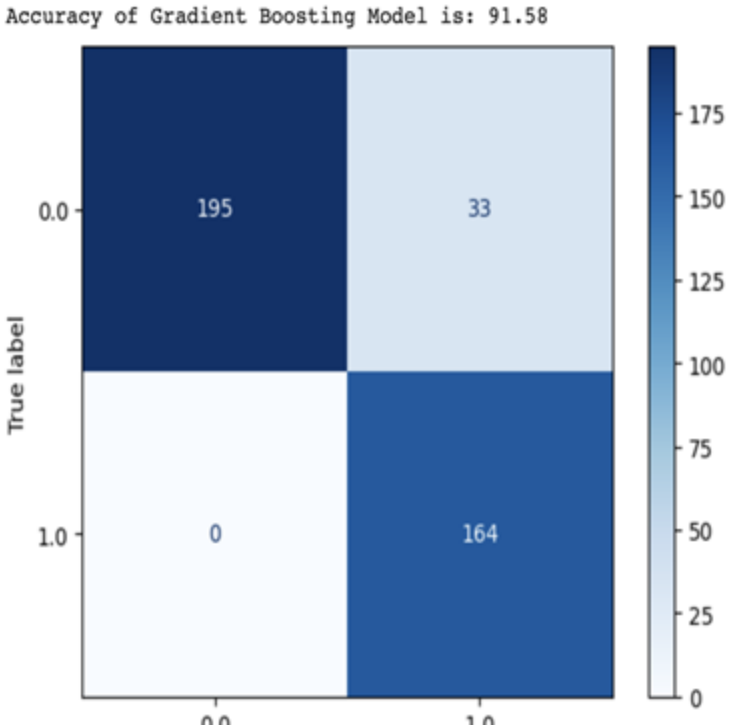
•Correlation Matrix: Features show weak–moderate correlations, supporting ML use.
•Age correlates most with osteoporosis risk.
• zero-variance features removed.

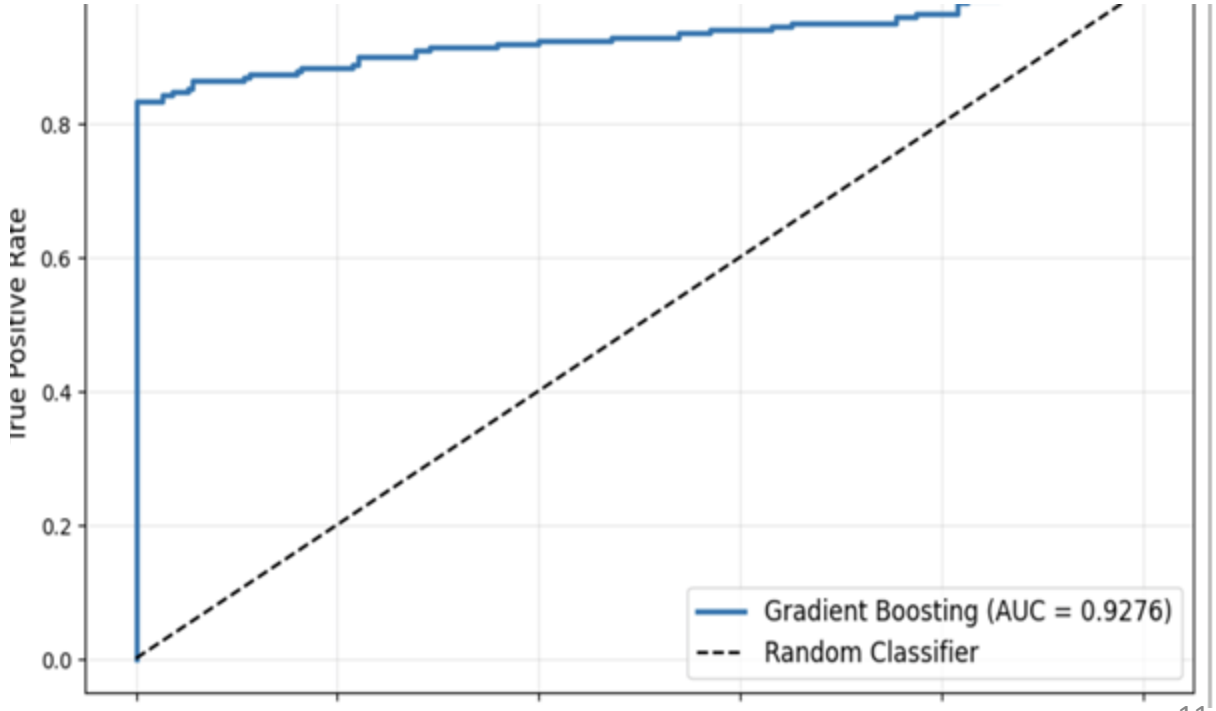
Model Performance Summary

	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	0.836700	0.827600	0.852800	0.840000	0.912300

Confusion Matrix & ROC Curve

- The confusion matrix shows that the Gradient Boosting model correctly classified most individuals, with relatively few false negatives.
- The ROC curve demonstrates strong discriminative ability with a high AUC score.
- This supports the model's reliability in distinguishing between high-risk and low-risk individuals.

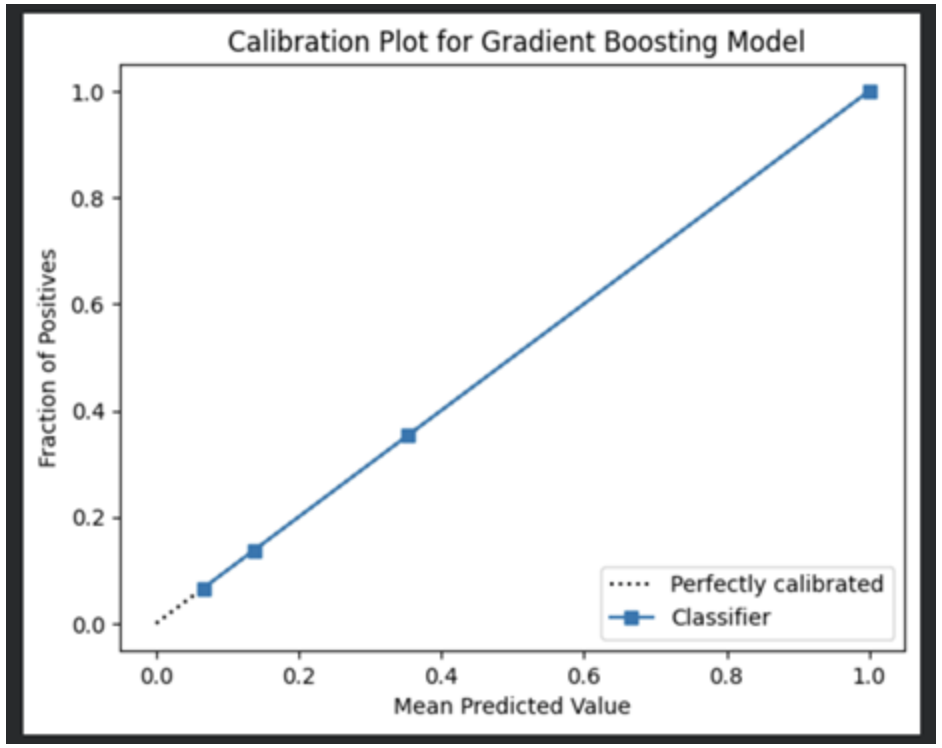


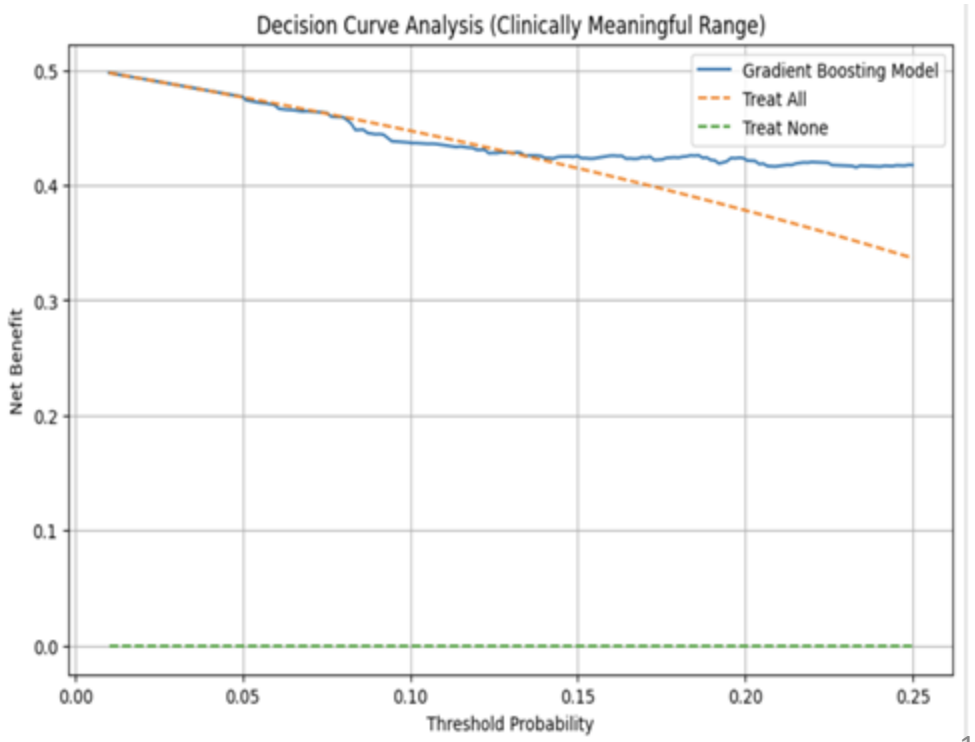


Calibration & DCA Curve

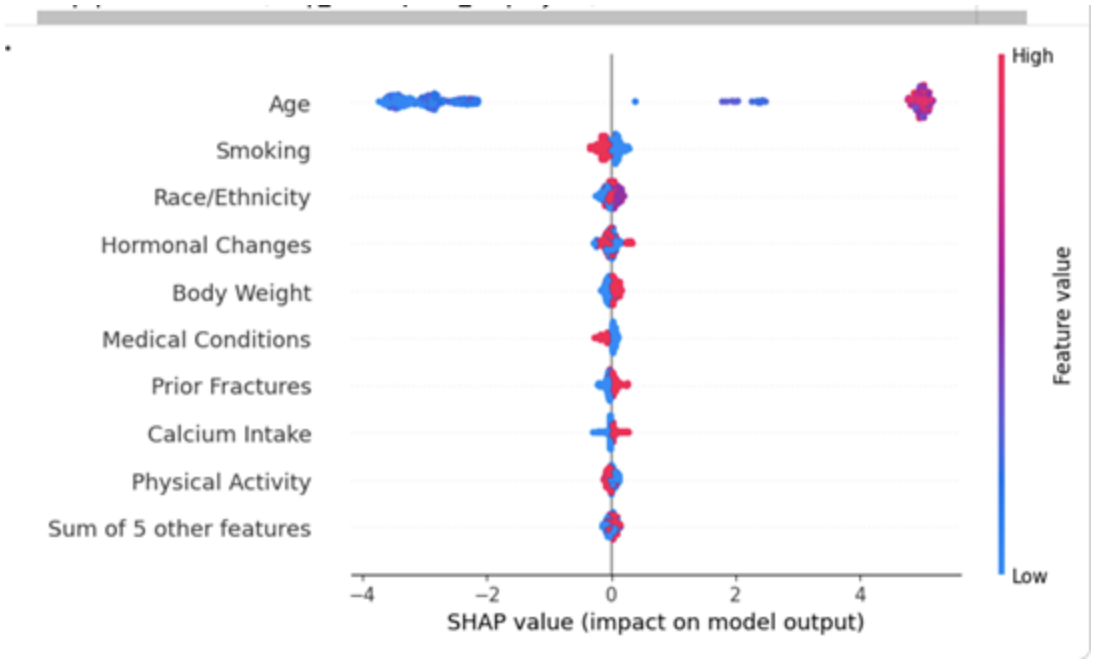
The calibration curve shows that predicted probabilities align well with actual outcomes essential for clinical risk prediction.

Decision Curve Analysis demonstrates that using the model leads to a higher net clinical benefit across a wide range of thresholds compared to “treat all” or “treat none” strategies.





SHAP Summary Plot

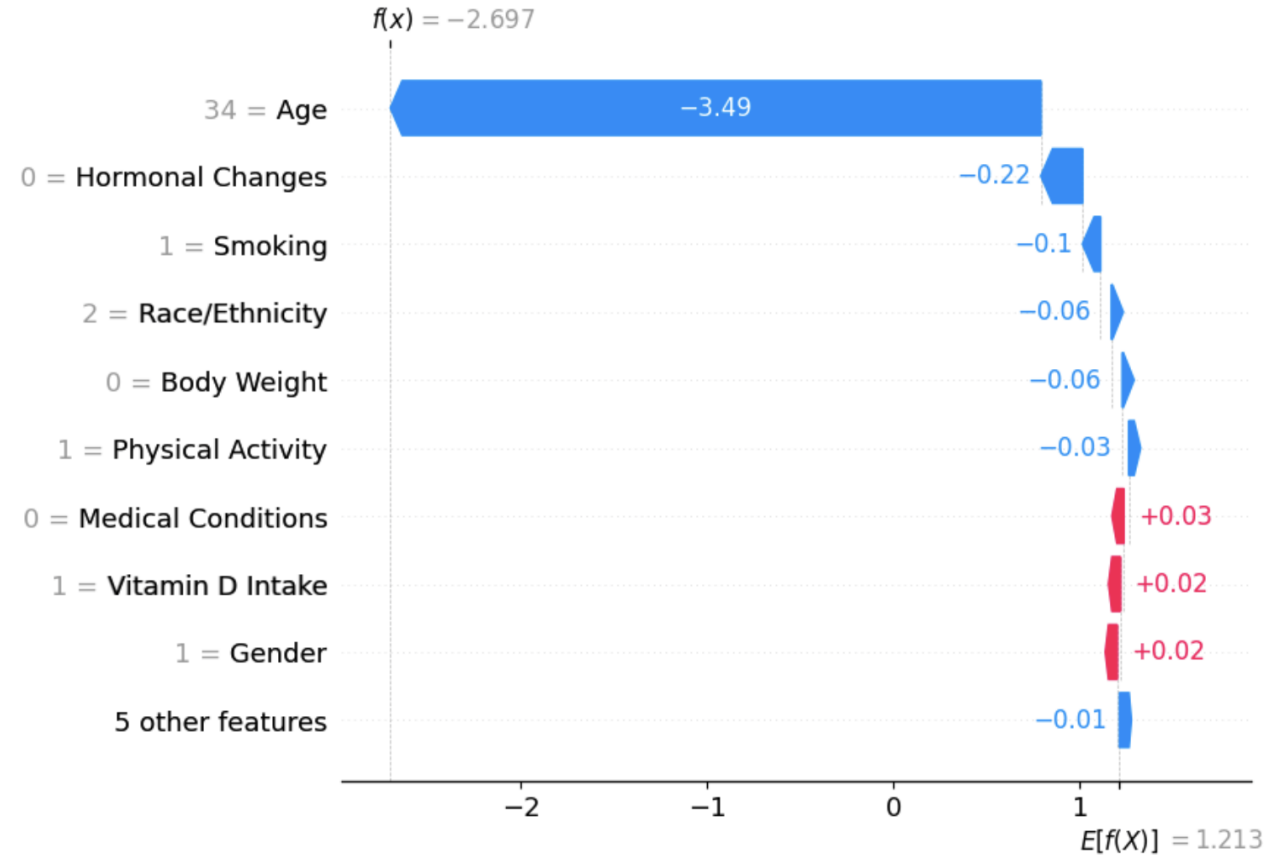


- SHAP global explanations showed that age, calcium intake, vitamin D intake, and hormonal changes were the strongest predictors consistent with biomedical knowledge.
- This confirms model reliability and interpretability.

SHAP Force Plot

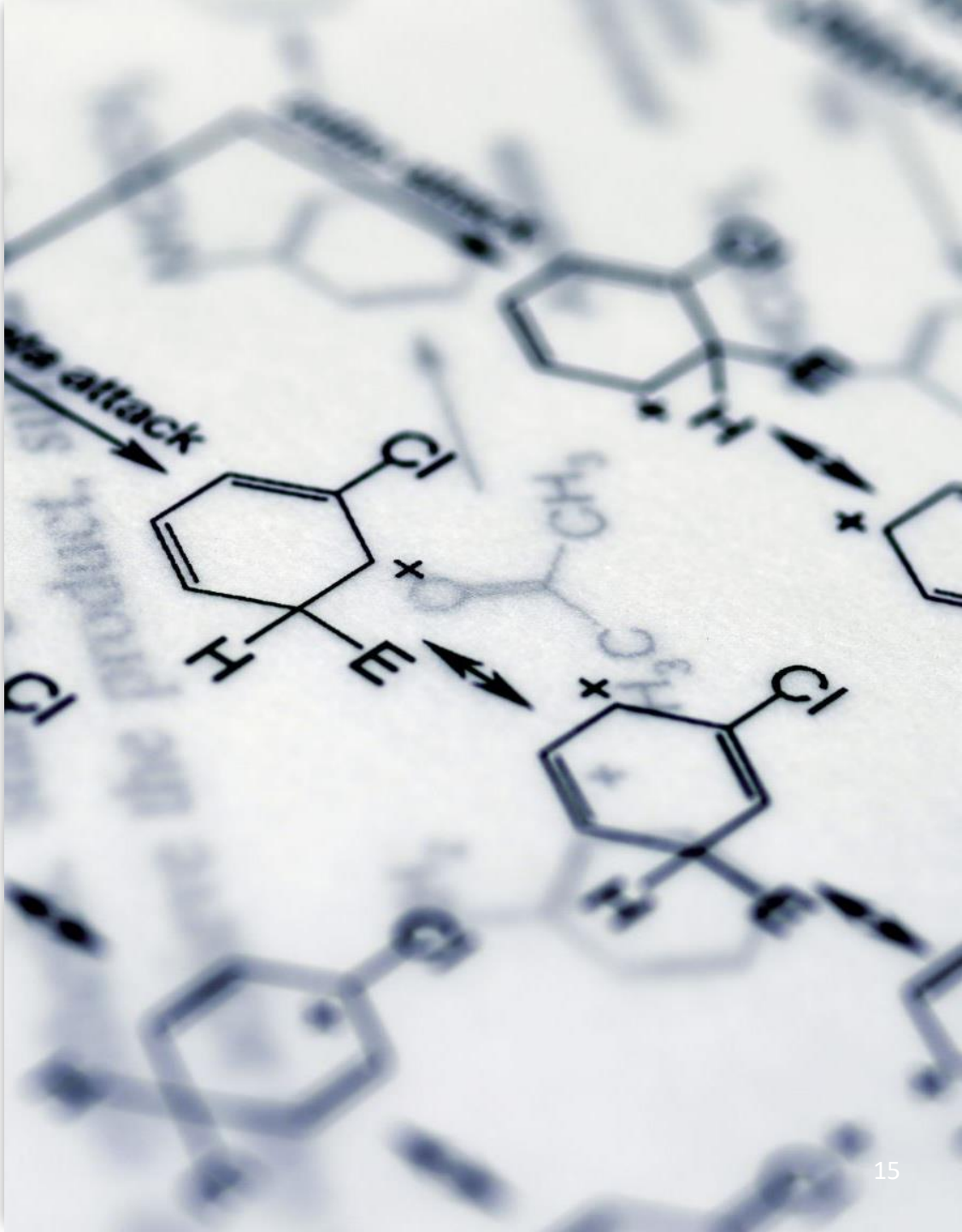
- The force plot explains a single patient's prediction.
  It highlights which factors pushed the model toward "high risk" versus "low risk," making the output clinically actionable.

	Age	Gender	Hormonal Changes	Family History	Race/Ethnicity	Body Weight	Calcium Intake	Vitamin D Intake	Physical Activity	Sm
1496	34.0	1	0	1	2	0	0	1	1	



Discussion

- The findings align strongly with established risk factors in clinical literature.
- Age remains the dominant predictor.
- Lifestyle factors calcium intake , vitamin D intake , physical activity independently influence risk.
- Strong calibration and DCA confirm that the model provides reliable probability estimates and clinical utility.
- Overall, the CDSS has strong potential for early intervention and personalized screening.



Limitations

Self-reported lifestyle and dietary data may introduce recall bias and reduce clinical precision.
Dataset includes a younger population, limiting generalizability to older adults who are at highest osteoporosis risk.
Lacks Bone Mineral Density (BMD) measurements, laboratory biomarkers, and imaging data key clinical indicators for diagnosis.
Future work should incorporate longitudinal data, multimodal clinical data, and external validation across diverse populations.

Conclusion

We successfully developed an interpretable, accurate ML-based CDSS for osteoporosis prediction.

Gradient Boosting emerged as the best-performing model.

This work forms the foundation for future integration into EHRs or mobile apps to enable widespread early screening and prevention.

References

Kanis, J. A., Cooper, C., Rizzoli, R., & Reginster, J.-Y. (2020).
European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporosis International*, 31(1), 3–44.

Warriner, A. H., & Saag, K. G. (2023).
The use of FRAX and other clinical risk assessment tools in osteoporosis. *Current Osteoporosis Reports*, 21(2), 67–75.

Vickers, A. J., & Elkin, E. B. (2006).
Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making*, 26(6), 565–574.

Lundberg, S. M., & Lee, S.-I. (2017).
A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS).*

Kaggle Dataset (2024).
Lifestyle Factors Influencing Osteoporosis.
Retrieved from: *https://www.kaggle.com/datasets/amitvkulkarni/lifestyle-factors-influencing-osteoporosis*

Compston, J., Cooper, A., & Cooper, C. (2019).
UK clinical guideline for the prevention and treatment of osteoporosis. *Archives of Osteoporosis*, 14(1), 1–24.

Leslie, W. D., & Majumdar, S. R. (2019).
Bone mineral density testing and fracture risk assessment. *Current Opinion in Rheumatology*, 31(4), 314–321.