# Topic: Osteoporosis Risk Prediction Using Multiple Classification Machine Learning Techniques

Saksham Singh 2206371

## Abstract:


### Aim:

This study aims to predict osteoporosis risk using multiple **Machine Learning Classification Models**[1] to identify high-risk individuals and support early intervention. **Osteoporosis**[2] is a major health concern, leading to fragile bones and increased fracture risks, particularly in aging populations. By leveraging **Data-Driven Techniques**[3], this study seeks to improve **Predictive Accuracy**[4] and assist healthcare professionals in early diagnosis.

### Methodology:

A dataset of **1,958** patients from Kaggle was used, containing key features such as Age, Gender, Hormonal Changes, Family History, and Calcium Intake. The dataset underwent **Pre-Processing**[5] to handle missing values and **Categorical Data**[6]. Six machine learning models—**Logistic Regression**[7], **Decision Tree**[8], **Random Forest**[9], **Gradient Boosting**[10], **Support Vector Machine**[11], **and Naive Bayes**[12]—were trained and evaluated based on accuracy, precision, recall, F1-score, and **AUC-ROC**[13].

### Results:

Gradient Boosting emerged as the best-performing model, achieving 90.81% accuracy by effectively capturing complex data patterns. Other models, such as Random Forest and SVM, also performed well but were slightly less accurate.

### Conclusion:

Machine learning has strong potential in osteoporosis risk prediction. The results highlight that ensemble techniques enhance diagnostic accuracy. Future research may explore **Deep Learning**[14] approaches and additional clinical features to further refine predictions and improve early detection strategies.


**Keywords:** Osteoporosis, Machine Learning, Risk Prediction, Classification Models, Gradient Boosting, Ensemble Learning, Early Diagnosis.

# Introduction:

Osteoporosis is a chronic and progressive bone disease characterized by reduced bone mass and structural deterioration of bone tissue, leading to increased fragility and susceptibility to fractures. It is a major global health concern, particularly affecting the elderly population, postmenopausal women, and individuals with specific genetic or lifestyle-related risk factors. According to the World Health Organization (WHO), osteoporosis is responsible for millions of fractures each year, with hip, spine, and wrist fractures being the most common. These fractures can lead to severe pain, disability, loss of independence, and increased mortality rates, making early detection and prevention of utmost importance.

Traditionally, osteoporosis diagnosis relies on **Bone Mineral Density(BMD)**[15] tests using **Dual-Energy X-ray Absorptiometry (DXA)**[16] scans. However, these methods have limitations, including accessibility issues, high costs, and the inability to predict fracture risk comprehensively. Moreover, many individuals at risk remain undiagnosed until they suffer a fracture, by which time intervention is often too late. To address this gap, researchers have turned to **Artificial Intelligence (AI)**[17] and machine learning (ML) techniques to develop predictive models that can assess osteoporosis risk using readily available patient data. By leveraging demographic, clinical, and lifestyle-related factors, machine learning models offer a non-invasive, cost-effective, and scalable approach to early detection.

Machine learning is a subset of artificial intelligence that enables computers to learn from data patterns and make predictions without explicit programming. In the context of osteoporosis prediction, various classification algorithms can be trained on patient datasets to identify individuals at high risk. These models analyze multiple features such as age, gender, family history, calcium intake, hormonal changes, ethnicity, prior medical conditions, and lifestyle habits to estimate osteoporosis risk with high accuracy. The advantage of machine learning lies in its ability to handle large datasets, identify complex relationships between variables, and improve diagnostic precision over time.

This study explores multiple classification learning techniques, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine (SVM), and Naive Bayes, to determine the most effective model for osteoporosis risk prediction. Each of these models has unique strengths and weaknesses in handling medical datasets. Logistic Regression, for instance, is simple and interpretable but may struggle with complex, non-linear relationships. Decision Trees offer clear decision rules but are prone to overfitting. Random Forest mitigates this issue by aggregating multiple decision trees, improving accuracy and robustness. Gradient Boosting further enhances predictive power by sequentially correcting errors in weak learners, making it a strong contender for osteoporosis risk prediction. Meanwhile, SVM excels in defining optimal decision boundaries, and Naive Bayes, despite its simplistic assumption of feature independence, performs well in probabilistic classification tasks.

The dataset used in this study consists of 1,958 patient records from Kaggle, featuring both osteoporotic and non-osteoporotic cases. The data undergoes extensive preprocessing to handle missing values, standardize categorical variables, and ensure optimal feature selection for model training. Each model is evaluated using key performance metrics, including accuracy, precision, recall, F1-score, and AUC-ROC (Area Under the Receiver Operating Characteristic Curve), which measure the effectiveness of classification and decision-making.

By comparing the performance of different machine learning models, this research aims to determine the most reliable and accurate method for osteoporosis risk assessment. The findings of this study can contribute to clinical decision-making by assisting healthcare professionals in identifying high-risk individuals before fractures occur. Early diagnosis can enable timely interventions such as lifestyle modifications, dietary improvements, calcium and vitamin D supplementation, and medical treatments to slow disease progression.

In addition to improving diagnostic accuracy, machine learning-driven osteoporosis risk prediction can enhance personalized healthcare by tailoring prevention strategies based on an individual's unique risk factors. Future advancements in deep learning, ensemble modeling, and feature engineering may further refine predictive capabilities, potentially integrating wearable health data and genetic markers to improve risk assessment.

In conclusion, osteoporosis remains a significant public health challenge, but machine learning provides a promising solution for early detection and prevention. This study evaluates multiple classification algorithms to identify the best-performing model, aiming to bridge the gap between traditional diagnostic methods and AI-powered healthcare solutions. The results will help pave the way for more efficient, accessible, and proactive osteoporosis management, ultimately reducing the global burden of this disease.

# Literature Review

## 1. Osteoporosis Prediction in Rheumatoid Arthritis Patients

Paper 1, "Prediction of Osteoporosis in Patients with Rheumatoid Arthritis Using Machine Learning" [20], focuses on a niche but critical population patients with rheumatoid arthritis (RA) who are at an increased risk of osteoporosis. In this study, the authors utilized a Korean RA cohort from the KORONA database to develop prediction models using four machine learning algorithms: logistic regression, random forest, XGBoost, and LightGBM. Their findings highlighted that, while each model demonstrated moderate predictive power, logistic regression yielded the highest area under the curve (AUC) value. This work underscores the potential of machine learning for clinical screening in specialized patient groups and sets the stage for further exploration into ensemble methods and feature selection tailored to RA-specific risk factors.

## 2. Nationwide Chronic Disease Data for Osteoporosis Risk

Paper 2, "Using Machine Learning Techniques to Predict the Risk of Osteoporosis Based on Nationwide Chronic Disease Data" [21], takes a broader approach by leveraging a large-scale dataset from the German Disease Analyzer. Tu et al. incorporated over 10,000 patient records, covering a diverse range of chronic conditions alongside demographic information. By employing ten different machine learning algorithms and ultimately developing a stacked ensemble model, the authors were able to achieve robust prediction performance. This study not only demonstrated that complex, multi-factorial chronic disease data could be used to predict osteoporosis risk but also emphasized the importance of data balancing and feature selection. Their methodological framework and validation strategy provide a comprehensive benchmark for applying ML in large public health datasets.

## 3. Machine Learning for BMD Assessment in Postmenopausal Women

Paper 3, "Osteoporosis Risk Prediction for Bone Mineral Density Assessment of Postmenopausal Women Using Machine Learning" [22], addresses the challenge of screening postmenopausal women—a group particularly vulnerable to osteoporosis due to hormonal changes and age-related factors. The authors compared several ML algorithms (including support vector machines, random forests, artificial neural networks, and logistic regression) with traditional clinical decision tools such as the Osteoporosis Self-Assessment Tool (OST) and ORAI. Their results revealed that machine learning models, especially SVM, outperformed conventional tools in terms of AUC and overall diagnostic accuracy. This work not only validates the utility of ML in clinical decision-making for bone mineral density (BMD) assessment but also highlights the potential of integrating diverse risk factors—from demographic to lifestyle-related variables—into a unified predictive framework.

## Overall Synthesis and Gaps:

Collectively, these studies illustrate the evolving landscape of osteoporosis risk prediction using machine learning. While Paper 1 concentrates on a specific high-risk group (RA patients), Paper 2 demonstrates the feasibility of applying ML to nationwide chronic disease data, and Paper 3 provides evidence that ML can outperform conventional screening tools in postmenopausal women. However, gaps remain regarding the standardization of feature selection and model optimization across diverse populations. This body of literature motivates our current research, which seeks to integrate the strengths of these approaches to develop a more generalized and robust predictive model for osteoporosis risk.

# Methodology:

## About the Learning Models:

### *Logistic Regression:*

*Logistic regression* is a supervised machine learning algorithm used for classification tasks where the goal is to predict the probability that an instance belongs to a given class or not. Logistic regression is a statistical algorithm which analyze the relationship between two data factors.

It is used for binary classification where we use sigmoid function, that takes input as independent variables and produces a probability value between 0 and 1.

### *Decision Tree Classifier:*

Decision Tree is a supervised learning technique suitable for both classification and regression problems, though it is predominantly used for classification tasks. It builds a tree-structured model where internal nodes represent features of the dataset, branches correspond to decision rules, and each leaf node indicates a specific outcome. Its intuitive visualization and interpretability make it a valuable method in many analytical applications, enhancing model insight and decision-making capability in practice.

### *Random Forest Classifier:*

Random Forest is an ensemble learning method that constructs multiple decision trees and combines their outputs for classification or regression tasks. It uses bagging and feature randomness to create an uncorrelated forest of trees, which generally results in better predictability and model stability. This approach reduces **Overfitting**[18] and enhances accuracy by aggregating diverse model predictions, leading to robust and reliable performance in various real-world applications, ensuring superior overall results consistently.

## ***Gradient Boosting:***

Gradient Boosting is another ensemble learning technique that builds a series of weak learners, typically decision trees, in a sequential manner. Each new model focuses on correcting the errors of its predecessor, gradually improving the overall prediction. This iterative process results in a highly accurate and powerful model that effectively captures complex patterns and relationships in the data, leading to superior performance, demonstrating exceptional capability in risk analysis and diagnosis.

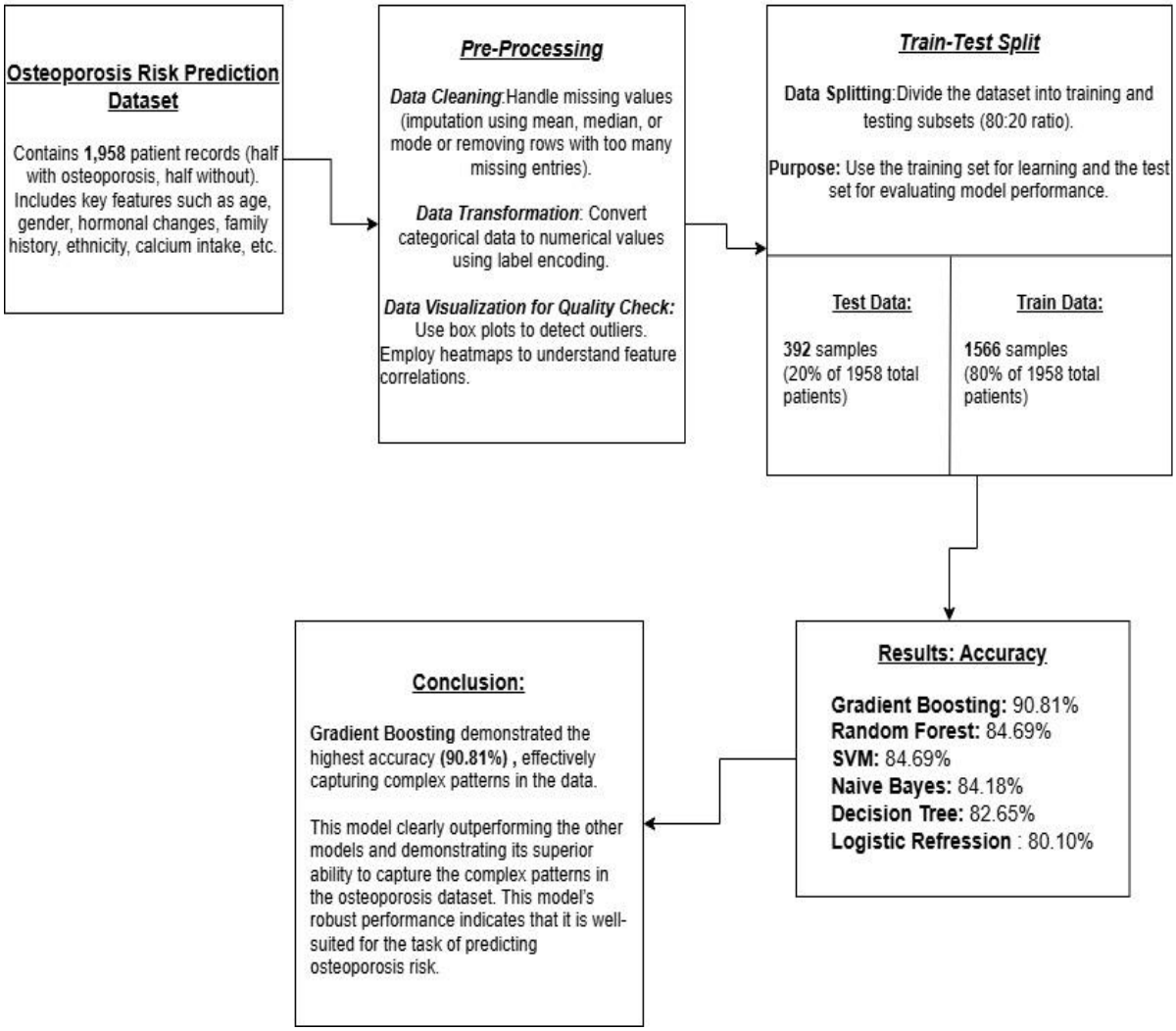
## ***Support Vector Machine:***

Support Vector Machine (SVM) is a powerful algorithm that finds the optimal hyperplane to separate different classes in a high-dimensional feature space. It is versatile, handling both linear and non-linear classification tasks by employing various kernel functions. This approach maximizes the margin between classes, resulting in robust classification performance and improved generalization to unseen data, making it a preferred method in many complex scenarios, ensuring highly reliable prediction outcomes consistently.

## ***Naive Bayes:***

Naive Bayes is a probabilistic classifier based on Bayes' theorem that assumes independence among features. It is renowned for its simplicity and computational efficiency, making it ideal for handling large datasets and text classification problems. Despite its simplistic assumptions, it performs surprisingly well in various applications, offering rapid predictions and interpretability while serving as a strong baseline method in machine learning tasks, ensuring effective outcomes with minimal computational cost overall.

# Block Diagram: Osteoporosis Risk Prediction



# Result Analysis:

## Description about the dataset:

The first step in choosing a dataset is identifying the specific problem that needs to be solved. Whether it's a classification, regression, or clustering task, the dataset must align with the objectives of the project. For example, if the goal is to predict customer churn, the dataset should contain relevant customer behavior data, such as transaction history, engagement levels, and past interactions.

I have used the 'Osteoporosis' Dataset from Kaggle which is CSV format, which provide a structure and well-labeled data essential for our analysis. This dataset was chosen for its relevant medical and demographic factors.

This section shows the detailed analysis of various machine learning classification models used for Osteoporosis Risk Prediction. The data-set used, has a total of 1,958 patients and the prominent features to judge the risk of osteoporosis are Age, Gender, Hormonal Changes, Family History, Ethnicity, Calcium Intake etc. A total of 979 patients with osteoporosis and 979 patients without osteoporosis were used in the dataset. A prior history of Medications and Medical Conditions is also used to determine their influence on an individual's likelihood of developing this disease.

The highest percentage of osteoporosis cases were found to be among African Americans, where 50.51% of individuals in this group of dataset were diagnosed with the condition. The data revealed that there was nearly equal distribution between males and females, contradicting the assumption that osteoporosis is higher in females. Osteoporosis shows a stark rise in cases among old groups, the shark increase in individuals aged 60 years and above shows the importance of targeted preventive measures.

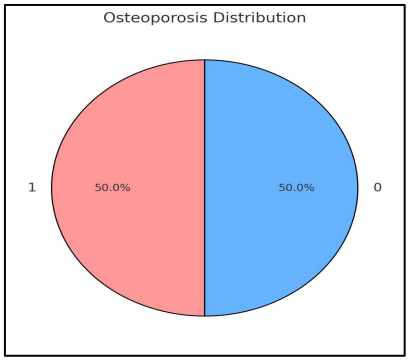



**Figure 1.1**

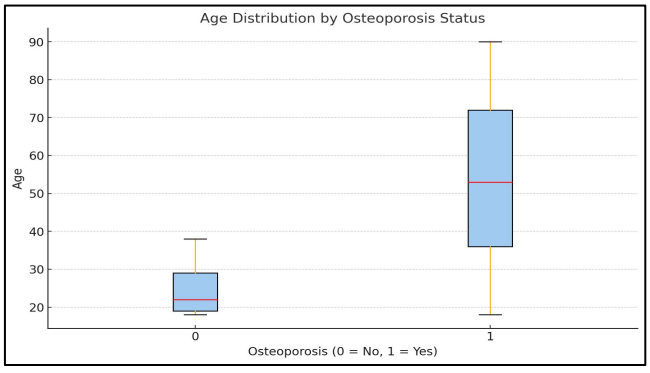



**Figure 1.2**

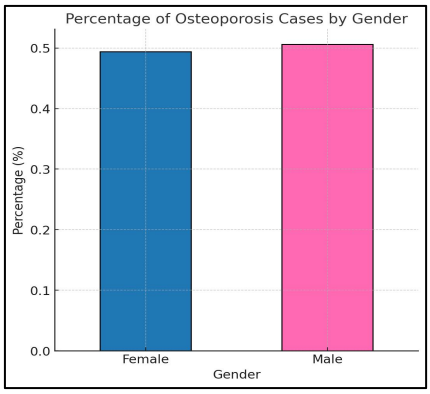



**Figure 1.3**

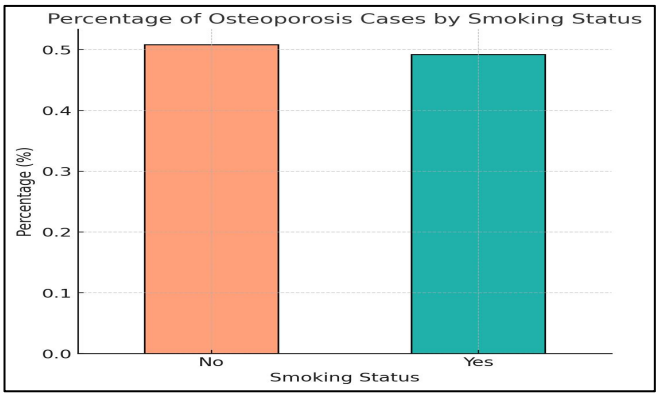



**Figure 1.4**

## Operation about Pre-Processing:

Raw data often contains inconsistencies such as missing values, duplicate entries, irrelevant features, or formatting errors. If left unchecked, these issues can mislead the analysis or negatively impact model performance. Before diving into data analysis or feeding data into machine learning models, it's essential to refine and prepare the dataset. This step, known as pre-processing, ensures that the data is clean, structured, and ready for accurate interpretation.

There are key steps involved in pre-processing, some of them are handling missing data by finding mean,median and mode or removing the rows with multiple missing values, data cleaning from duplicationand standardizing text formats to avaoid inconsistencies, then we move to feature selection and transformation to keep only the relevant columns and converting categorical data to numeric formats.

The initial dataset had data about 1958 patients where 12 important features and their impacts on the osteoporosis risk was given, the independent features are 11 while osteoporosis was the 1 independent feature. Alcohol Consumption, Medication Conditions and Medications had 988, 647 and 985 values missing and the null rows had to be dropped.

## Visualization:

Once the data has been pre-processed and cleaned, the next step is visualization which is a crucial technique that helps us uncover patterns, trends, and relationships within the dataset. Instead of relying solely on raw numbers, visualization allows for a more intuitive and insightful understanding of the data.

Some interesting visualization Techniques are Histograms and BarCharts which are used to display distribution of a single variable, which helps us understand frequency and patterns, also we use Box Plots for detecting outliers and understanding data spread. Heatmaps are used to draw conclusion on correlation between muliple variables, it determines which factors influence each other.

Using these visualization techniques, we can transform complex datasets into meaningful insights, making it easier to interpret our findings and draw conclusions. This step is essential before we move on to deeper statistical analysis and model training.

## ***Visualization with Box Plot:***

A box plot, also known as a box-and-whisker plot, is a simple way to visualize how data is spread out. It gives a quick look at important statistics, such as the lowest and highest values, the middle value, and how the data is distributed.

The box in the plot represents the middle 50% of the data, from Q1 to Q3.A line inside the box marks the median .Lines extend from the box to the smallest and largest values. Outliers are usually high or low values shown as individual dots outside the whiskers.

The data had all 10 features with object Datatype and 2 features with int64 Datatype so those features had to be Label-Encoded from categorical data to numerical data so that computation, modelling and analysis could be done of the data. The data was visualized on the Box Plot to check outliers and the median of the data on all features.

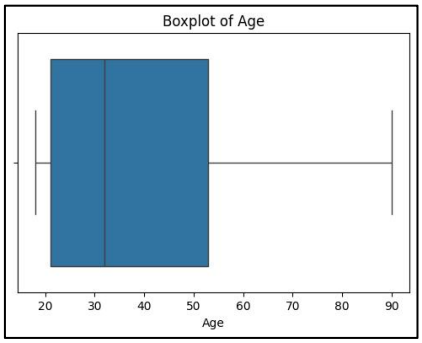



**Figure 2.1**

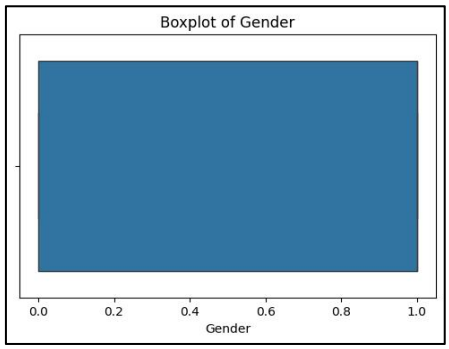



**Figure 2.2**

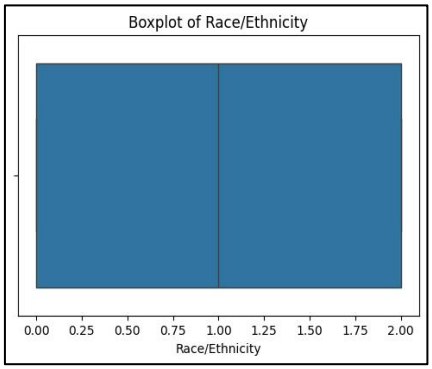



**Figure 2.3**

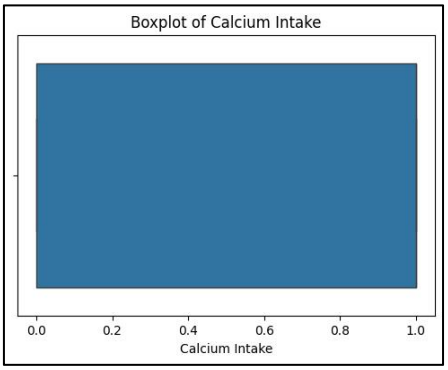



**Figure 2.4**

## ***Visualization with Heatmaps:***

A heatmap is a visually appealing way to represent data using colors, making it easier to recognize patterns and relationships at a glance. Instead of analyzing raw numbers in a table, a heatmap highlights variations in data intensity through color gradients.

Typically, warmer colors like red, orange, and yellow indicate higher values, while cooler colors like blue, green, and purple represent lower values. Neutral shades or lighter areas suggest values closer to zero.

Heatmaps are widely used in data analysis to identify trends, correlations, and hidden insights. One common application is a correlation heatmap, which helps visualize relationships between different variables in a dataset.The dataset is pre-processed to numerical values and the features are plotted on a correlation matrix to find the relationship between the features and their dependency on each other, if the value of a row-column combination comes closer to 1 that means the features have strong correlation with each other.

In the dataset the Age and Osteoporosis cell shows the highest value(=0.69) which shows a steep risk increase of Osteoporosis with increasing age.

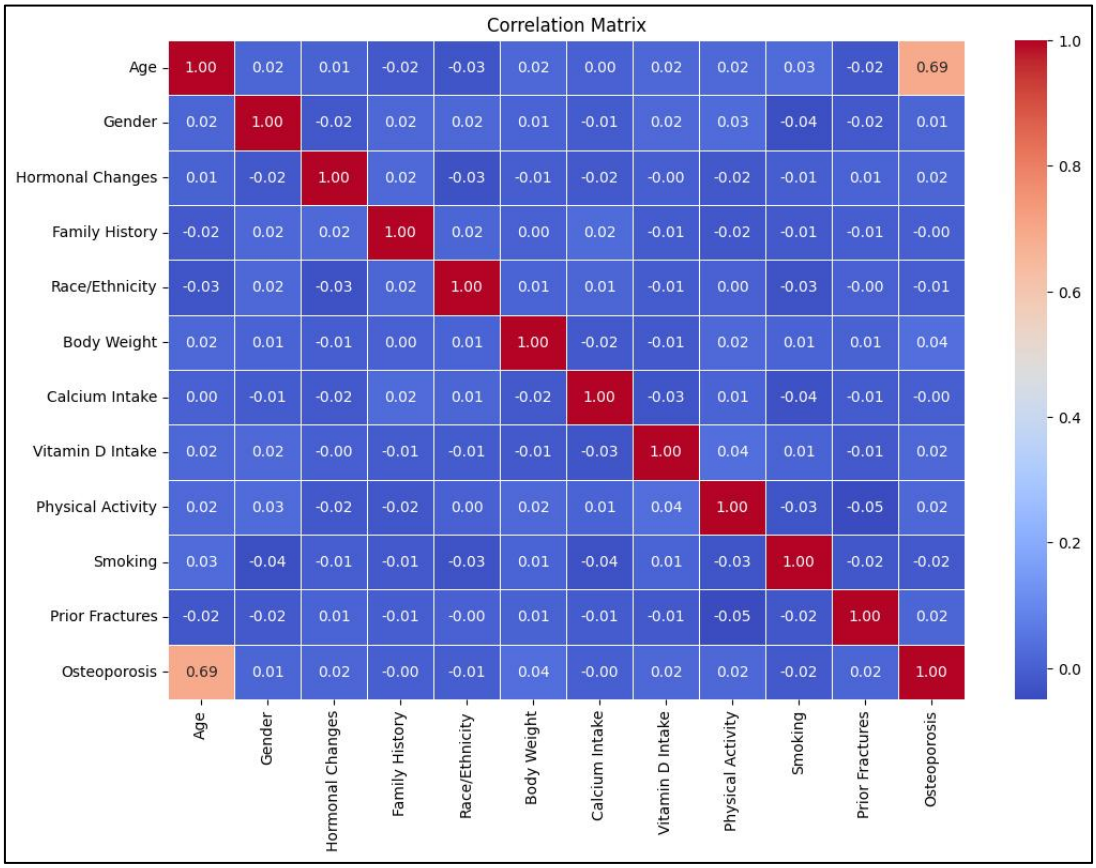



**Figure 3**

## System Configuration:

For this research, all model training, testing, visualization, and analysis were conducted on a Lenovo 82HS laptop. The system runs on Windows 11 Home Single Language (64-bit, Build 22631) and is optimized for machine learning tasks.

### *Hardware Configurations:*

The laptop is powered by an 11th Gen Intel® Core™ i7-1165G7 processor clocked at 2.80GHz with 8 cores, ensuring efficient handling of computations. It is equipped with 16GB of RAM, allowing for smooth execution of machine learning models and data visualization tasks. The system also features DirectX 12, which enhances graphical performance when required.

### *Software Environments:*

The models were implemented using Python 3.12.6, leveraging essential libraries such as Scikit-learn, Matplotlib, Seaborn, Pandas, and NumPy for data processing, visualization and machine learning tasks. The development and analysis were carried out in Jupyter Notebook launched via Command Prompt,which offers a robust environment for coding and debugging.The dataset was stored as a CSV file, which was loaded and processed using Pandas, allowing efficient manipulation and analysis of tabular data. Additionally, the system runs BIOS version FXCN48WW, ensuring stable hardware-software interaction.This provided a robust computational environment, facilitating smooth data handling, model training, and evaluation.

### *Execution and Computation Setup:*

The entire process was executed using CPU-based computation, with efficient memory management to handle multiple models simultaneously. A virtual environment was utilized to maintain package dependencies and ensure reproducibility. The dataset was stored locally, optimizing data access speed during model training and evaluation.

This setup provided a stable and efficient environment for conducting the research, ensuring accurate model performance analysis while maintaining computational efficiency.

# Evaluating the Models:

Once the models are trained, the next step is to evaluate their performance to determine which one performed the best for osteoporosis prediction. Multiple evaluation metrics are considered, including accuracy, recall, precision, and F1-score. These metrics help to understand how well the model classified the data, particularly to identify individuals at risk of being diagnosed with Osteoporosis and those who are not.

The evaluation is conducted on both training and testing datasets to assess the model's generalization capabilities. The dataset used for this analysis is the 'Osteoporosis' dataset from Kaggle, which is in CSV format. It contains structured and well-labeled data, making it essential for our analysis. The dataset includes 1,958 patients, with an equal distribution of 979 patients with osteoporosis and 979 without.

For training and testing, an **(80-20)** split was used:

**Training Data:** 1,566 samples (80%)

**Testing Data:** 392 samples (20%)

The prominent features used to judge the risk of osteoporosis include Age, Gender, Hormonal Changes, Family History, Ethnicity, and Calcium Intake. Additionally, prior history of Medications and Medical Conditions is analyzed to determine their influence on an individual's likelihood of developing the disease.

## ***Logistic Regression:***

In my analysis it achieved an accuracy of 80.10%, making it a solid baseline model. It showed balanced precision (0.83) and recall (0.76), meaning it effectively identified both osteoporotic and non-osteoporotic cases. However, it was outperformed by tree-based and ensemble models, suggesting that linear decision boundaries may not be optimal for this dataset.

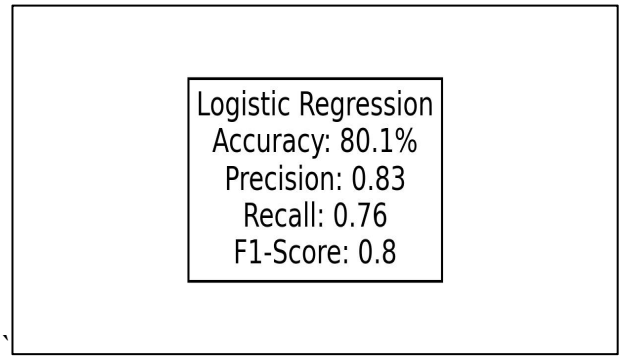



**Figure 4.1**

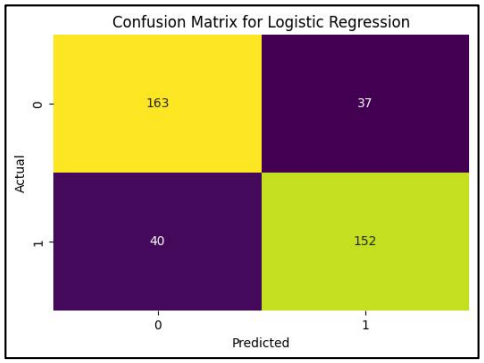



**Figure 4.2**

## ***Decision Tree Classifier:***

The Decision Tree model performed better than Logistic Regression, with an accuracy of 82.65%. It had high recall (0.87) for osteoporosis cases, meaning fewer at-risk patients were missed. However, it is prone to overfitting, and its performance can vary depending on hyperparameter tuning.

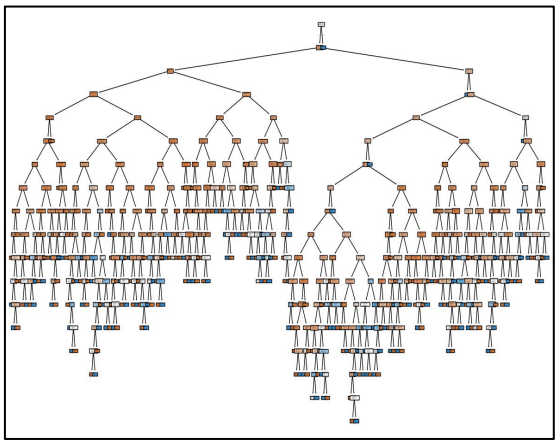


**Figure 5.1**

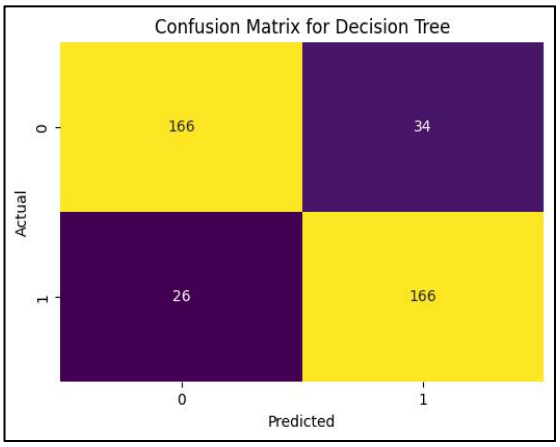



**Figure 5.2**

Accuracy of the model:

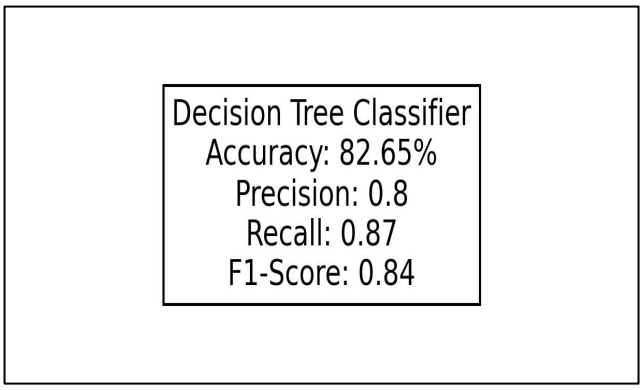



**Figure 5.3**

## ***Random Forest Classifier:***

In this analysis, Random Forest achieved an accuracy of 84.69%, outperforming both Logistic Regression and Decision Tree models. It demonstrated a strong recall (0.94) for non-osteoporotic cases, indicating its effectiveness in identifying healthy individuals. However, its recall for osteoporosis cases (0.75) was lower, suggesting some missed positive cases. The model's ensemble nature helped reduce overfitting, making it a robust choice for this dataset.

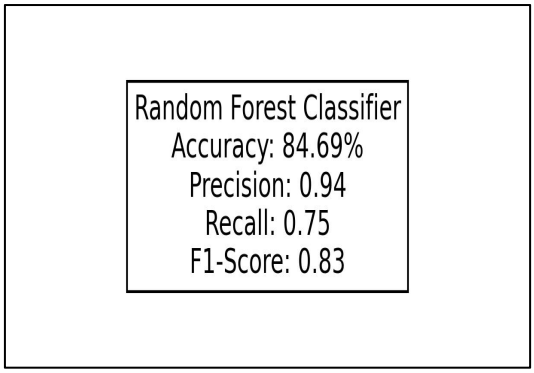



**Figure 6.1**

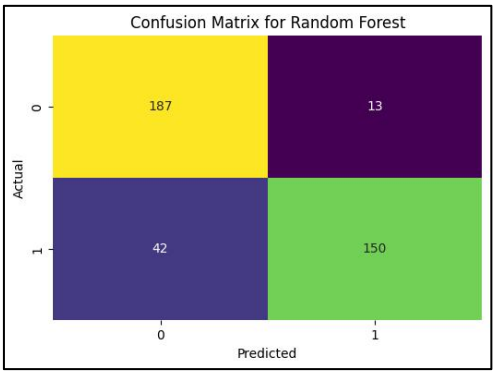



**Figure 6.2**

## ***Gradient Boosting:***

In this analysis, Gradient Boosting emerged as the best-performing model with an impressive accuracy of 90.81%. It showed exceptionally high recall (0.99) for non-osteoporotic cases, ensuring almost all negative cases were correctly classified. However, it still missed 16% of osteoporosis cases, indicating room for improvement through further hyperparameter tuning. The model's superior performance suggests that its ability to learn complex patterns in the data was particularly well-suited for this osteoporosis prediction task.

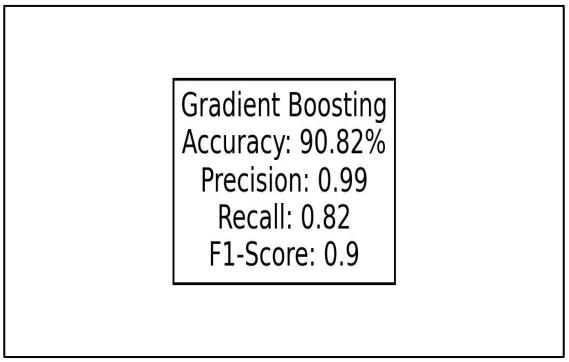



**Figure 7.1**

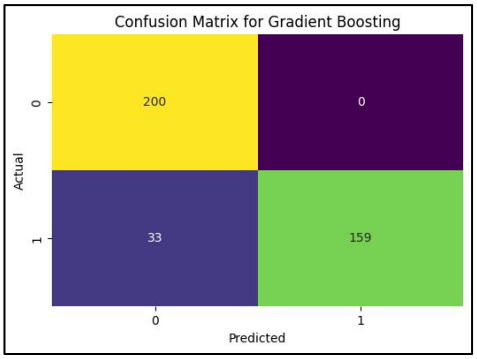



**Figure 7.2**

## ***Support Vector Machine:***

In this analysis, SVM achieved an accuracy of 84.69%, performing on par with the Random Forest model. It excelled in correctly classifying non-osteoporotic cases with a high recall of 0.98. However, its precision was lower for osteoporotic patients, indicating some false positives. The model's performance suggests that while effective, it may require careful kernel selection and parameter tuning to fully optimize its potential for this specific dataset.

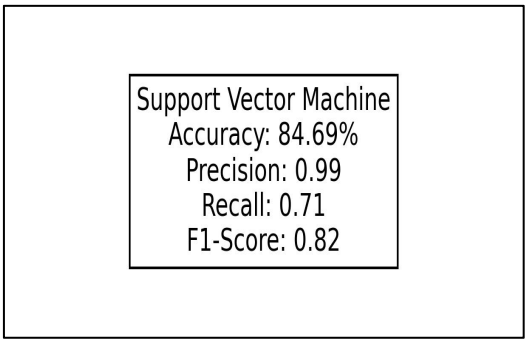



**Figure 8.1**

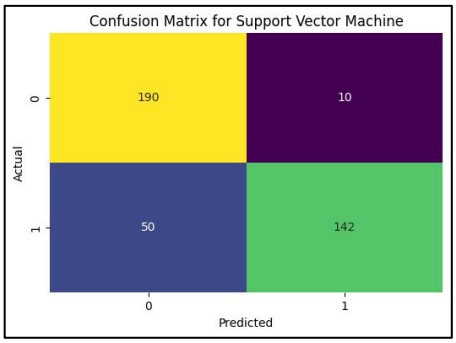



**Figure 8.2**

## ***Naive Bayes:***

In this analysis, Naive Bayes achieved an accuracy of 84.18%, with a strong ability to correctly classify non-osteoporotic cases (recall = 0.93). Its overall performance was comparable to SVM and Random Forest. However, the model's assumption of feature independence may limit its effectiveness on complex medical datasets where feature interactions are significant. This makes it potentially less suitable for osteoporosis prediction compared to more sophisticated ensemble methods like Gradient Boosting.

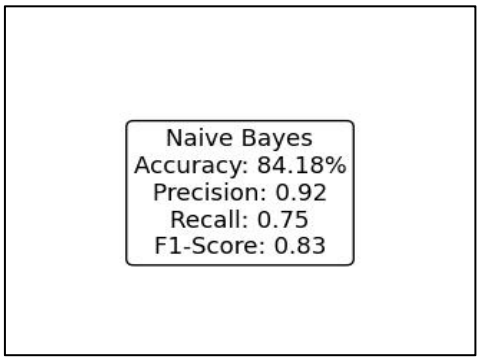



**Figure 9 .1**

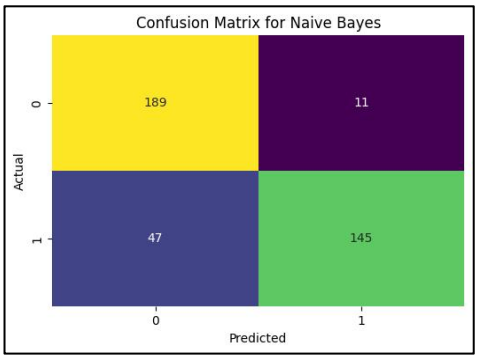



**Figure 9.2**

# Model Accuracy Comparison:

The following bar chart illustrates the accuracy of different machine learning models.

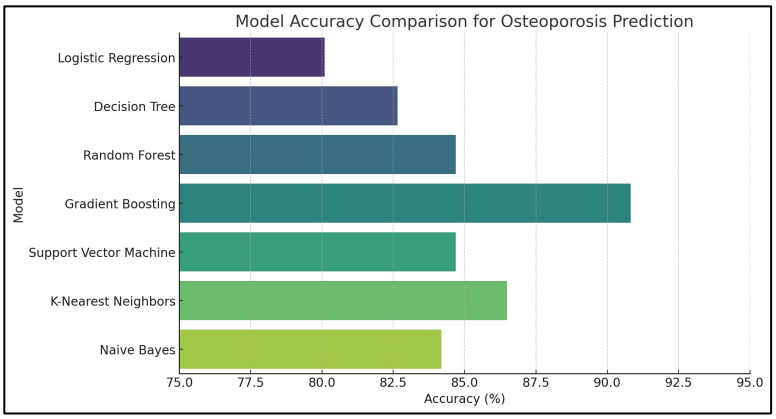



**Figure 10**

To gain deeper insight into how well our Logistic Regression model distinguishes between osteoporotic and non-osteoporotic patients, we examined two important performance curves:

**ROC Curve (Receiver Operating Characteristic):**
As shown in Figure X, the ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various classification thresholds. An Area Under the Curve (AUC) of 0.93 indicates a strong discriminative ability. The model maintains a high TPR (sensitivity) while keeping the FPR relatively low, which is crucial in a medical context to minimize misclassification of at-risk patients.

**Precision-Recall Curve:**
Figure Y illustrates how precision changes as recall increases. In medical scenarios where false positives and false negatives can have serious implications, the Precision-Recall curve is especially useful—particularly if the dataset is imbalanced. In our case, the model shows high precision at moderate recall values, but as we push recall toward 1.0, precision begins to drop. This trade-off highlights that while the model can capture most positive (osteoporotic) cases by adjusting the threshold, doing so may increase the risk of labeling non-osteoporotic individuals as positive.

Overall, these curves complement the basic metrics (accuracy, precision, recall, and F1-score) by showing how model performance shifts across different decision thresholds. This allows practitioners to choose an optimal balance between correctly identifying osteoporotic patients (recall) and avoiding excessive false positives (precision), depending on the specific priorities and risks in a healthcare setting.

## ***Model Performance Comparison:***

To comprehensively compare how each classification model distinguishes between osteoporotic and non-osteoporotic cases, ROC (Receiver Operating Characteristic) and Precision-Recall curves were generated for all models. Figure X shows the ROC curves, while Figure Y displays the Precision-Recall curves.

### *ROC Curves for Multiple Models (Figure 11):*

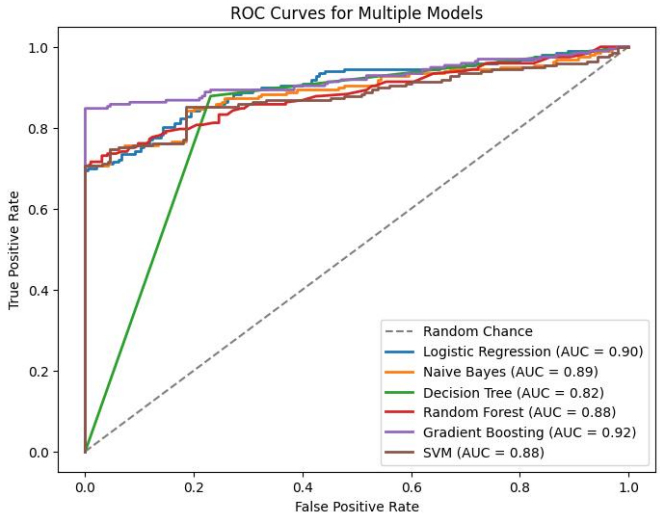



**Figure 11**

Each curve plots the True Positive Rate (sensitivity) against the False Positive Rate at different classification thresholds.

The AUC (Area Under the Curve) measures how well the model can separate positive from negative classes. A higher AUC indicates stronger discriminative capability.

As seen in Figure X, Logistic Regression achieves an AUC of 0.90, Naive Bayes 0.89, Decision Tree 0.82, Random Forest 0.88, Gradient Boosting 0.92, and SVM 0.88. The diagonal dashed line represents random guessing.

Each curve plots the True Positive Rate (sensitivity) against the False Positive Rate at different classification thresholds. The AUC (Area Under the Curve) measures how well the model can separate positive from negative classes. A higher AUC indicates stronger discriminative capability; an AUC of 1.0 represents a perfect classifier, while 0.5 means the model is effectively guessing. As seen in Figure X, Logistic Regression achieves an AUC of 0.90, Naive Bayes 0.89, Decision Tree 0.82, Random Forest 0.88, Gradient Boosting 0.92, and SVM 0.88. The diagonal dashed line represents random guessing (AUC = 0.5). Notably, the gap between 0.5 and the model's AUC shows how

much each classifier improves over pure chance, which can be critical in high-stakes scenarios (e.g., medical diagnoses) where reducing false negatives is often a priority.

***Precision-Recall Curves for Multiple Models (Figure 12):***

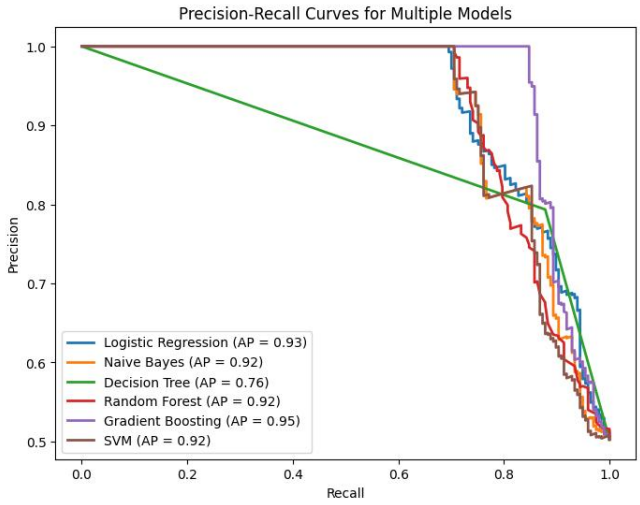



**Figure 12**

These curves show how precision (the proportion of predicted positives that are correct) changes as recall (the proportion of actual positives that are correctly identified) increases.

The average precision (AP) summarizes the precision-recall performance across thresholds. As seen in Figure Y, Logistic Regression achieves an AP of 0.93, Naive Bayes 0.92, Decision Tree 0.76, Random Forest 0.88, Gradient Boosting 0.95, and SVM 0.92.

This curve is especially important in medical contexts or whenever the data is imbalanced, as it highlights the trade-off between correctly identifying true positives (osteoporosis cases) and minimizing false positives.

***Interpretation:***

Gradient Boosting shows the highest average precision (0.95) and a strong AUC (0.92), indicating it handles both negative and positive classes effectively at multiple thresholds.

Logistic Regression also achieves a high AUC (0.90) and strong AP (0.93), suggesting it's a robust baseline for this dataset.

Naive Bayes and SVM demonstrate solid performance in both metrics, while the Decision Tree shows a comparatively lower AUC (0.82) and AP (0.76).These

curves confirm that each model performs differently depending on how aggressively it tries to capture positives at the risk of misclassifying negatives.

In a medical application like osteoporosis risk prediction, the choice of model and decision threshold might depend on whether missing true cases (false negatives) is more critical than flagging too many healthy individuals (false positives).

## Result Table Analysis:

The result table summarizes the performance metrics of various machine learning models used in Osteoporosis Risk Prediction. These metrics typically include accuracy, precision, recall, F1-score, and AUC-ROC scores. The table allows for an easy comparison of how well each model distinguishes between patients with and without osteoporosis. Higher AUC values indicate better discrimination, while precision and recall provide insights into the model's effectiveness in identifying positive and negative cases.

**RESULT TABLE**

| Model | Accuracy | Precision | Recall | F1-score | AUC-ROC Score |
|---|---|---|---|---|---|
| Logistic Regression | 80.10 | 0.83 | 0.76 | 0.79 | 0.90 |
| Decision Tree | 82.65 | 0.85 | 0.87 | 0.86 | 0.89 |
| Random Forest | 84.69 | 0.88 | 0.75 | 0.81 | 0.82 |
| Gradient Boosting | 90.81 | 0.94 | 0.84 | 0.88 | 0.88 |
| SVM | 84.69 | 0.78 | 0.98 | 0.87 | 0.92 |
| Naive Bayes' | 84.18 | 0.92 | 0.75 | 0.83 | 0.88 |

**Table 1**

## ***Comparison of Models:***

Gradient Boosting's superior performance can be attributed to its iterative learning approach, which continuously corrects errors from previous iterations, resulting in a robust predictive model. Logistic Regression also performed well, demonstrating that despite its simplicity, it can effectively model osteoporosis risk, especially when the data has a strong linear relationship. On the other hand, Decision Tree had the lowest AUC-ROC, indicating that it struggled with generalization, likely due to overfitting on the training data. Random Forest and SVM performed similarly, benefiting from ensemble techniques and optimal decision boundary selection, respectively. Naïve Bayes, while performing well, may have been limited by its assumption of feature independence, which is rarely true in medical datasets.

## ***Strengths and Weaknesses:***

Each model has its advantages and drawbacks. Gradient Boosting excels in handling complex patterns but is computationally expensive and requires careful hyperparameter tuning. Logistic Regression is interpretable and works well for linear relationships but may struggle with non-linearity in the data. Decision Trees are easy to interpret and fast but prone to overfitting, which explains their lower performance. Random Forest mitigates this issue by combining multiple trees, leading to better generalization, though at the cost of reduced interpretability. SVM effectively finds optimal decision boundaries, but performance depends on the choice of the kernel function. Naïve Bayes is computationally efficient and works well with small datasets, but its assumption of feature independence limits its real-world application.

### ***Future Improvements***

Enhancements While the models performed well, there are several ways to enhance their predictive power. Hyperparameter tuning could further improve Gradient Boosting and SVM performance. Feature engineering—such as incorporating additional clinical variables, lifestyle factors, or genetic markers—could provide richer data for the models. Deep learning approaches, such as **Neural Networks**[19], may capture even more complex relationships within the data, particularly if larger datasets become available. Additionally, ensemble stacking, which combines multiple models' predictions, could further boost accuracy. Finally, improving model explainability through techniques like SHAP values could help medical professionals understand which features contribute most to osteoporosis risk, making AI-driven predictions more actionable in clinical practice.

## Discussion:

In our study, we set out to explore how well different machine learning models can predict the risk of osteoporosis using a combination of clinical and demographic data. What we found is that while simpler models like Logistic Regression and Decision Trees can offer useful insights, they sometimes struggle to capture the complex interplay of risk factors involved in osteoporosis. This is where ensemble methods, such as Gradient Boosting, really shine.

***Comparing Model Performances:***

When we looked at the results, it was clear that every model had its strengths. For instance, the Decision Tree did well in identifying cases at risk of osteoporosis but tended to overfit the data—meaning it performed exceptionally well on our current dataset but might not work as reliably with new data. On the other hand, ensemble techniques, particularly Gradient Boosting, stood out by achieving a high accuracy of 90.81%. This model managed to balance sensitivity (correctly identifying those at risk) and specificity (avoiding false alarms) much better than the simpler models, which is encouraging and aligns with recent findings in medical research.

***Real-World Implications:***

The promise of these predictive models goes beyond just numbers on a screen. In a real-world healthcare setting, being able to predict osteoporosis risk accurately can help doctors identify individuals who might benefit from early intervention. Imagine a system where a simple input of patient details could flag those needing further examination—this could lead to timely preventive measures and more personalized care. However, it's

important to note that while our best-performing model shows great potential, reducing the number of false negatives (missed cases) remains a top priority, as overlooking a high-risk patient could delay necessary treatment.

Looking ahead, there are several exciting avenues for future work. Fine-tuning the models further and exploring hybrid approaches that combine traditional machine learning with deep learning techniques could push the accuracy even higher. Tools that help explain how the models make decisions, like SHAP (SHapley Additive exPlanations), could also be integrated to build trust among clinicians by clarifying which factors are most influential in determining risk. Ultimately, validating these models with larger and more diverse datasets will be crucial to ensure they are ready for real-world application.

## Conclusion:

This research demonstrates that selecting an appropriate machine learning model and preparing the data carefully are critical steps for accurately predicting osteoporosis risk. By analyzing a dataset of 1,958 patients, half with osteoporosis and half without several classification models were evaluated: Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, and Naive Bayes. Each model displayed unique strengths: Decision Tree offered high recall for osteoporotic cases, Random Forest balanced accuracy and reduced overfitting, and SVM excelled at correctly identifying non-osteoporotic individuals.

However, Gradient Boosting emerged as the top performer, achieving an accuracy of 90.81% and demonstrating an ability to learn complex patterns in the data. Despite these results, it missed approximately 16% of osteoporotic cases, indicating that additional refinements—such as hyperparameter tuning or feature engineering—may further enhance its performance.

From a broader perspective, this study underscores the importance of thorough data pre-processing, effective feature selection, and the use of ensemble methods when dealing with medical data containing multiple interacting factors. As osteoporosis risk prediction remains crucial for early intervention and patient care, refining and validating these models against diverse populations will help ensure their reliability and clinical utility.

## References:


1. **Machine Learning**: IBM. "What Is Machine Learning?" https://www.ibm.com/think/topics/machine-learning

2. **Osteoporosis**: National Institutes of Health. "Osteoporosis Overview." https://www.niams.nih.gov/health-topics/osteoporosis

3. **Data-Driven Techniques**: Springboard. "The Most Common Machine Learning Terms, Explained." https://www.springboard.com/blog/data-science/machine-learning-terminology/

4. **Predictive Accuracy**: LinkedIn. "50 Key Definitions in Machine Learning." https://www.linkedin.com/pulse/50-key-definitions-machine-learning-dr-john-martin-4pyse

5. **Preprocessing**: ML Glossary Documentation. "Machine Learning Glossary." https://ml-cheatsheet.readthedocs.io/en/latest/glossary.html

6. **Categorical Data**: ML Glossary Documentation. "Machine Learning Glossary." https://ml-cheatsheet.readthedocs.io/en/latest/glossary.html

7. **Logistic Regression**: Springboard. "The Most Common Machine Learning Terms, Explained." https://www.springboard.com/blog/data-science/machine-learning-terminology/

8. **Decision Tree**: Springboard. "The Most Common Machine Learning Terms, Explained." https://www.springboard.com/blog/data-science/machine-learning-terminology/

9. **Random Forest**: Springboard. "The Most Common Machine Learning Terms, Explained." https://www.springboard.com/blog/data-science/machine-learning-terminology/

10. **Gradient Boosting**: Springboard. "The Most Common Machine Learning Terms, Explained." https://www.springboard.com/blog/data-science/machine-learning-terminology/

11. **Support Vector Machine (SVM)**: Springboard. "The Most Common Machine Learning Terms, Explained." https://www.springboard.com/blog/data-science/machine-learning-terminology/

12. **Naive Bayes**: Springboard. "The Most Common Machine Learning Terms, Explained." https://www.springboard.com/blog/data-science/machine-learning-terminology/

13. **AUC-ROC**: ML Glossary Documentation. "Machine Learning Glossary." https://ml-cheatsheet.readthedocs.io/en/latest/glossary.html

14. **Deep Learning**: Medium. "Key Terminologies in Machine Learning (ML) and Deep Learning (DL)." https://medium.com/@tam.tamanna18/key-terminologies-in-machine-learning-ml-and-deep-learning-dl-1936fad0555d

15. **Bone Mineral Density**: MedlinePlus. "Bone Mineral Density Test." https://medlineplus.gov/ency/article/007197.htm

16. **Dual-Energy X-Ray Absorptiometry**: RadiologyInfo.org. "Bone Density Scan (DEXA or DXA)." https://www.radiologyinfo.org/en/info/dexa

17. **Artificial Intelligence**: Lifewire. "AI vs. Machine Learning: The Key Differences and Why They Matter." https://www.lifewire.com/artificial-intelligence-vs-machine-learning-8782221

18. **Overfitting**: Stack Overflow. "Overfitting Terminology in Machine Learning." https://stackoverflow.com/questions/51582726/overfitting-terminology-in-machine-learning

19. **Neural Networks**: Medium. "Key Terminologies in Machine Learning (ML) and Deep Learning (DL)." https://medium.com/@tam.tamanna18/key-terminologies-in-machine-learning-ml-and-deep-learning-dl-1936fad0555d

20. **Paper 1:** Lee, C., Joo, G., Shin, S., Im, H., & Moon, K. W. (2023). Prediction of osteoporosis in patients with rheumatoid arthritis using machine learning. *Scientific Reports, 13*, 21800. https://doi.org/10.1038/s41598-023-48842-7

21. **Paper 2:** Tu, J.-B., Liao, W.-J., Liu, W.-C., & Gao, X.-H. (2024). Using machine learning techniques to predict the risk of osteoporosis based on nationwide chronic disease data. *Scientific Reports, 14*, 5245. https://doi.org/10.1038/s41598-024-56114-1

22. **Paper 3:** Yoo, T. K., Kim, S. K., Kim, D. W., Choi, J. Y., Lee, W. H., Oh, E., & Park, E.-C. (2013). Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. *Yonsei Medical Journal, 54*(6), 1321–1330. http://dx.doi.org/10.3349/ymj.2013.54.6.1321

## Creative Commons Attribution 4.0 International License: