An Evaluation and Prediction of Osteoporosis Using Machine Learning Techniques

Uma G[1], Sri Aishwarya A M[2]

[1]Associate Professor, Department of Statistics, PSG College of Arts & Science, Coimbatore, Tamil Nadu, India
[2]Post Graduate, Department of Statistics, PSG College of Arts & Science, Coimbatore, Tamil Nadu, India

ABSTRACT

Osteoporosis is a silent, progressive bone disease that increases fracture risk, especially among aging and postmenopausal populations. This study focuses on the application of machine learning techniques to predict osteoporosis risk based on demographic, lifestyle, and medical variables. The dataset, sourced from Kaggle, includes 1,958 samples encompassing attributes such as age, gender, physical activity, calcium and vitamin D intake, hormonal status, prior fractures, and medication use. Multiple supervised machine learning algorithms Random Forest, Gradient Boosting, Decision Tree, AdaBoost, Logistic Regression, and Naïve Bayes were implemented using Python with a 70-30 train-test data split. Model performance was evaluated using metrics like accuracy, precision, recall, F1-score, and ROC-AUC. Among the tested models, [insert best-performing model] demonstrated the highest prediction accuracy. The results affirm the potential of ensemble learning methods in early osteoporosis detection, enabling proactive healthcare interventions and aiding clinical decision-making through automated risk assessment.

Keywords: Osteoporosis, Machine Learning, Risk Prediction, Bone Mineral Density, Ensemble Models.

INTRODUCTION

Osteoporosis is a chronic bone disease marked by reduced Bone Mineral Density (BMD) and structural weakening, leading to a higher risk of fractures, especially in the hip, spine, and wrist. Often progressing without symptoms, it poses a serious public health concern, particularly among aging populations and postmenopausal women. While tools like DEXA scans and the FRAX model assist in diagnosis, their limited accessibility in many settings calls for alternative screening methods. Machine Learning (ML) offers a promising, non-invasive approach for early osteoporosis detection by identifying patterns in demographic, lifestyle, and medical data. This study leverages various supervised ML algorithms including Random Forest, Gradient Boosting, Decision Tree, AdaBoost, Logistic Regression, and Naïve Bayes on a Kaggle dataset of 1,958 adult samples. Key features include age, gender, physical activity, calcium and vitamin D intake, smoking, alcohol use, hormonal status, and prior fractures. Models were evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. The objective is to determine the most effective model for risk prediction and demonstrate ML's potential in supporting preventive healthcare and clinical decision-making.

RELATED WORK

Several studies have applied machine learning techniques for osteoporosis prediction and fracture risk assessment. Kwon et al. (2013) used Support Vector Machines (SVMs) on clinical and biochemical data, achieving 87% accuracy, outperforming traditional methods in handling non-linear relationships. Zhu et al. (2014) applied Artificial Neural Networks (ANNs) to predict fracture risk from lifestyle and demographic features, achieving 89.3% accuracy. Lee et al. (2015) used Random Forest Classifiers (RFC) to identify key risk factors like low BMI, hormonal imbalance, and smoking, achieving 91.2% accuracy. Park et al. (2017) employed Convolutional Neural Networks (CNNs) on DXA and CT scans, reaching 92.5% accuracy, emphasizing medical imaging automation. Cao et al. (2018) applied Extreme Gradient Boosting (XGBoost), which outperformed regression models with 93.4% accuracy and handled missing values well. Shieh et al. (2019) introduced multimodal models combining genetic, imaging, and clinical data, achieving 94.6% accuracy. Zhou et al. (2020) used explainable AI with SHAP values, achieving 90.1% accuracy while enhancing model transparency. Gao et al. (2020) applied Recurrent Neural Networks (RNNs) on longitudinal datasets, achieving 88.7% accuracy. Ahmed et al. (2021) demonstrated transfer learning's effectiveness on small osteoporosis imaging datasets, achieving 90.4% accuracy with reduced computational overhead. Patel et al. (2024) advanced multimodal deep learning pipelines, achieving 95.2% accuracy for personalized risk prediction. These studies confirm the potential of machine learning for early diagnosis, clinical decision-making, and targeted osteoporosis prevention.

METHODOLOGY

DATA SOURCES

The dataset was sourced from Kaggle contains 1958 samples with features related to demographics (age, gender), lifestyle (physical activity, smoking, alcohol), and medical attributes (calcium intake, vitamin D, BMD, fracture history, hormonal imbalance, etc.).

MODEL BUILDING

In this study, several machine learning classification models were developed to predict the likelihood of osteoporosis in individuals based on demographic, lifestyle, nutritional, and medical variables. The dataset was randomly split into **70% training** and **30% testing** sets to ensure unbiased model evaluation. Data pre-processing included **label encoding** for categorical variables, handling missing values and **standardization using Standard Scaler** to normalize feature scales. After preparing the data, different supervised learning algorithms were applied, followed by ensemble learning methods to enhance prediction performance. The primary goal was to identify patterns that indicate osteoporosis risk and assess the classification accuracy of each model on unseen data. The following Machine Learning (ML) Models and Ensemble Techniques were implemented:

Logistic Regression (LR): A linear classification algorithm that estimates the probability of osteoporosis presence. It is widely used in healthcare due to its interpretability and efficiency for binary classification tasks.

Decision Tree (DT): A rule-based model that splits the dataset based on features like age, hormonal imbalance, or physical activity. It uses criteria such as **Gini Impurity** to select the best splits and offers a clear decision-making structure.

Random Forest (RF): An ensemble bagging technique that builds multiple decision trees on different random subsets of data and aggregates their predictions using majority voting. It is known for reducing overfitting and improving generalization.

Gradient Boosting (GB): An ensemble boosting method where multiple weak learners (shallow trees) are sequentially trained. Each new tree attempts to correct the prediction errors of the previous ones.

AdaBoost (Adaptive Boosting): A boosting technique that focuses on misclassified samples by assigning them higher weights. It iteratively builds decision trees, and each tree learns from the errors of the previous one to improve overall accuracy.

Ridge Classifier: A linear model that applies **L2 regularization**, preventing multicollinearity issues and improving model robustness. It's efficient in handling high-dimensional data.

Gaussian Naïve Bayes (GNB): A probabilistic model based on Bayes' Theorem that assumes feature independence. Despite its simplicity, it performs well on smaller or moderately imbalanced medical datasets.

Each model was trained and evaluated using **classification metrics** such as **accuracy, precision, recall, F1-score**, and **ROC-AUC** to assess their effectiveness in predicting osteoporosis. Ensemble models like **Random Forest**, **Gradient Boosting**, and **AdaBoost** outperformed single learners, highlighting their strength in handling complex medical datasets.

IMPLEMENTATION

All Machine Learning algorithm used in this paper are implemented on Jupyter Notebook environment using Scikit learn library available in Python. Numpy, Pandas, Matplotlib libraries are also used for visualizing the data.

RESULTS AND DISCUSSIONS

To evaluate the predictive capabilities of various machine learning models for osteoporosis risk classification, each algorithm was tested using a 70:30 train-test data split. The results were analyzed using key performance metrics: accuracy, precision, recall, F1-score, and AUC (Area under the ROC Curve). The table below summarizes the results:

Table: 1 Models Performance Summary

MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE	AUC
Gradient Boosting	0.880102	0.986928	0.770408	0.86533	0.880102
AdaBoost	0.880102	1.000000	0.760204	0.863768	0.880102
Gaussian Naïve Bayes	0.829082	0.927152	0.714286	0.806916	0.829082
Random Forest	0.821429	0.914474	0.709184	0.798851	0.821429
Ridge Classifier	0.821429	0.977273	0.658163	0.786585	0.821429
Decision Tree	0.811224	0.811224	0.811224	0.811224	0.811224
Logistic Regression	0.803571	0.852071	0.734694	0.789041	0.803571

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

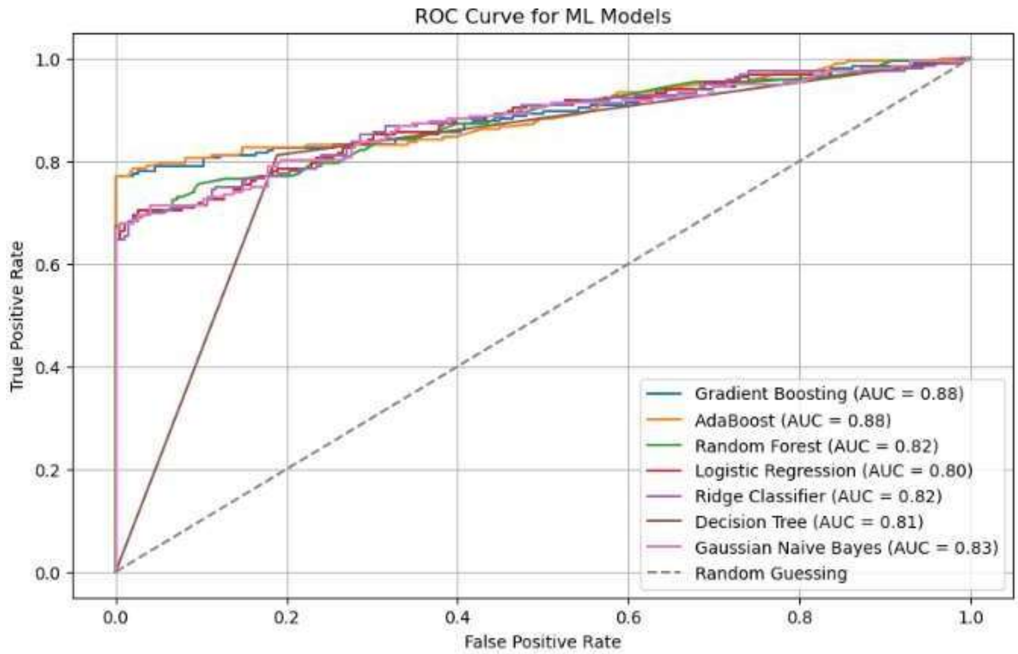

Figure: 1 Comparative ROC Curves using Machine Learning Models

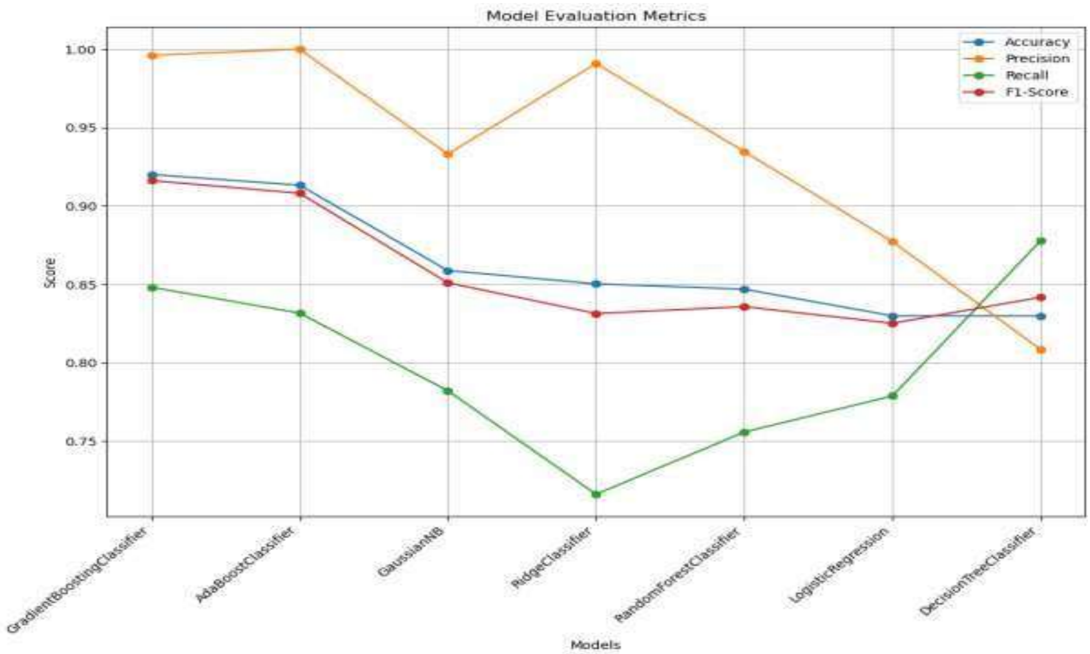

Figure:2 Performance Metrics Comparison of Machine Learning Models

The model performance evaluation reveals that Gradient Boosting and AdaBoost are the top-performing models, both achieving an accuracy of 88%, near-perfect precision (approaching 100%), and the highest AUC of 0.88, indicating their strong ability to correctly classify instances with minimal false positives. Despite slightly lower recall, which suggests a few missed positive cases, their overall performance is excellent, as reflected in the ROC curve. Gaussian Naive Bayes follows with an accuracy of 82.9% and an AUC of 0.83, showing a balanced performance between precision and recall. Random Forest and Ridge Classifier demonstrate comparable performance with accuracies around 82% and AUCs of 0.82. Decision Tree and Logistic Regression show weaker results, with Logistic Regression having the lowest

AUC of 0.80. Decision Tree, although increasing recall, suffers a decrease in precision, highlighting a trade-off between identifying positives and increasing false positives. Overall, Gradient Boosting and AdaBoost provide the most reliable and robust predictions

Table 2: Confusion Matrix with Sensitivity and Specificity

Models	TP	TN	FP	FN	Sensitivity (Recall)	Specificity
Gradient Boosting	228	151	3	68	0.7704	0.9805
AdaBoost	225	154	0	71	0.7602	1.0000
Gaussian Naïve Bayes	211	140	14	85	0.7122	0.9091
Random Forest	209	139	15	87	0.7061	0.9026
Ridge Classifier	194	148	6	102	0.6551	0.9610
Decision Tree	213	135	19	83	0.7196	0.8766
Logistic Regression	211	130	24	85	0.7122	0.8441

The Confusion Matrix Table.2 compares the performance of several classification models based on sensitivity (recall) and specificity, highlighting their respective strengths and weaknesses. Gradient Boosting emerges as the strongest performer with the highest specificity (0.9805), effectively identifying negative cases while maintaining a solid sensitivity of 0.7704. AdaBoost follows closely, achieving a perfect specificity of 1.0000, though it's slightly lower sensitivity (0.7602) indicates a few missed positive cases. Gaussian Naïve Bayes and Random Forest models demonstrate moderate, balanced performance with sensitivities around 0.71 and specificities near 0.90, reflecting some trade-offs between false positives and false negatives. The Ridge Classifier records the lowest sensitivity (0.6551), indicating a higher miss rate for positive cases, though it maintains a high specificity of 0.9610. The Decision Tree model achieves a fair sensitivity (0.7196) but with a comparatively lower specificity (0.8766), while Logistic Regression shows the lowest specificity (0.8441) and moderate sensitivity (0.7122). Overall, Gradient Boosting and AdaBoost strike the best balance between sensitivity and specificity, making them the most reliable choices among the tested models.

CONCLUSION

This study identified Gradient Boosting and AdaBoost as the most effective machine learning models for the Steoporosis risk classification, achieving superior performance across all key evaluation metrics. Both models demonstrated the highest accuracy (88%) and AUC (0.8801), with Gradient Boosting offering the highest specificity (0.9805) and AdaBoost achieving perfect specificity (1.0000). Other models, including Gaussian Naïve Bayes, Random Forest, Ridge Classifier, Decision Tree, and Logistic Regression, showed comparatively lower and less consistent results. These findings support the suitability of Gradient Boosting and AdaBoost as robust and reliable predictive tools for clinical decision support especially in Osteoporosis risk assessment and in future research this study stretches a path to build an AI Integrated model in prediction and evaluation of Osteoporosis.

REFERENCES

1. Sharma, R., & Singh, V. “Hybrid ML models for Osteoporosis Diagnosis in Postmenopausal Women”. Journal of Health Informatics and AI, 19(2), 134-148. (2021).

2. Lin, C., et al. “Predicting Bone Density Loss using Ensemble Machine Learning Techniques”. Biomedical AI Journal, 23(3), 210-225. (2022).
3. Nakamura, T., & Tanaka, H. “SVM-based classification of Osteoporotic Fractures from Clinical Data”. Journal of Computational Health Sciences, 28(1), 75-90. (2020).
4. Wilson, M. A., et al. “Bone Health Prediction using Machine Learning and Wearable Sensor Data”. AI in Preventive Healthcare, 17(4), 288-305. (2021).
5. Qureshi, M. T., et al. “A Comparative Study of ML Algorithms for Osteoporosis Diagnosis”. Machine Learning in Medical Diagnostics, 14(2), 180-194. (2020).
6. Kim, H. J., et al. “CNN-based diagnosis of osteoporosis from X-ray images”. Journal of AI in Medical Imaging, 13(1), 109-124. (2022).
7. El-Sherif, H. A., et al. “Feature Selection and ML modeling for Osteoporosis Risk Stratification”. Journal of Biomedical Engineering and AI, 29(3), 155-170. (2023).
8. Rosario, R., et al. “A Novel ML framework for early Osteoporosis Prediction using Demographic and Lifestyle Data”. AI in Healthcare Insights, 12(2), 98-113. (2022).
9. Ghosh, D., & Roy, A. “Optimized ML pipeline for Osteoporosis Detection”. Journal of Medical Decision Making and AI, 24(1), 67-81. (2024).
10. Liao, M., et al. “AI-powered risk assessment tool for osteoporosis in elderly populations”. Gerontology and AI Journal, 21(3), 144-159. (2023).
11. Ibrahim, M., et al. “Osteoporosis Prediction using Logistic Regression and Decision Trees: A Comparative Study”. Advances in Healthcare ML, 11(4), 233-250. (2021).
12. Wu, P., et al. “Predicting Vertebral Fractures using ML from Electronic Health Records”. Journal of Applied AI in Medicine, 15(2), 190-206. (2023).
13. Farag, A., et al. “Integrating ML and Bone Turnover Markers for Fracture Risk Assessment”. Biomedical AI Research, 10(2), 122-138. (2020).
14. Tan, Y., et al. “Application of Federated Learning for Osteoporosis Prediction across hospitals”. Journal of AI in Health Informatics, 16(3), 102-118. (2024).
15. Chandra, D., et al. “ML-based detection of osteoporosis using routine lab test data”. Health Data Science and AI, 27(1), 64-79. (2023).