Research Article
Volume 1, Issue 1, 2025

InOsteoD: A Cost-effective Integrative Approach for Osteoporosis Detection

Parthajit Borah[1]*, Rachayita Bhardwaj[2†], Pankaj Barah[3†], Paramdeep Singh[4†], D.K. Bhattacharyya[1†]

[1] Department of Computer Science and Engineering, Tezpur University, Tezpur, 784028, Assam, India.
[2] Centre for Biomedical Engineering, Indian Institute of Technology Delhi, Delhi, 110016, New Delhi, India.
[3] Department Molecular Biology and Biotechnology, Tezpur University, Tezpur, 784028, Assam, India.
[4] All India Institute of Medical Sciences, Dabwali Rd, Lal Singh Nagar, Bathinda, AIIMS, Punjab 151001, Punjab, India.
* Correspondence: parthajit@tezu.ernet.in; †These authors contributed equally to this work.
Scopus Author ID 57220804586
Received: 01 December 2024; Accepted: 20 December 2024; Published: 21 January 2025

Abstract: Osteoporosis is a prevalent bone condition that typically affects older adults, occurring when bone mineral density and bone mass decline or when the bone's quality or structure deteriorates. It increases the risk of bone fractures and causes many complications for patients. X-ray is a low-cost medical imaging test that can be used for predicting Osteoporosis by extracting relevant and non-redundant features followed by appropriate downstream analysis. This paper presents a cost-effective integrative approach called InOsteoD that exploits both deep learning as well as tree-based ensemble learning to detect Osteoporosis from X-ray images with high precision. The performance of InOsteoD has been found satisfactory on real-life datasets.

Keywords: Osteoporosis; X-ray; Deep learning; features; classification

1. Introduction

Osteoporosis is a multifaceted condition marked by a gradual, asymptomatic decline in bone mass per unit volume. When bone mass falls below a critical threshold, maintaining structural integrity and mechanical support becomes challenging, leading to fractures with minimal pain. Common fracture sites include the vertebrae, humerus, pelvis, ribs, proximal femur, and distal radius[1]. Cortical and trabecular bones are both present in adult humans. At least 50% of the vertebrae are made up of the former, which provides stiffness, and the latter, which offers strength and elasticity (the axial skeleton)[1] [2]. Bone tissue, being metabolically active, gets replaced constantly by cellular activities that resorb (osteoclastic) and form (osteoblastic) bone [1] [3]. In normal adult bone, resorption and formation are in balance and the disruption in the balance in these processes can lead to increase or decrease in bone mass. Osteoporosis can further be categorized into primary and secondary osteoporosis. Primary osteoporosis, most common form of the disease includes postmenopausal osteoporosis (type I), which leads to loss of androgen and estrogen. Consequently, it results in 1) bone turnover increase, 2) bone resorption, and 3) loss of trabecular bone

Senile or Type II osteoporosis is distinguished based on its etiology. This type is visible with age-related bone loss in both male and female categories due to systemic senescence. It is mainly caused by 1) loss of cortical bone and 2) loss of stem-cell precursors [4].

Bone mass peaks between the ages of 25 and 30, then decline steadily at a rate of 0.2 to 0.5 percent per year until the age of 35 to 45 in both genders (men and women). In the eight

to ten years prior to and beyond menopause, women experience a 2–5% annual rate of bone loss. After then, both sexes experience bone loss at the same slow pace [1] [5].

Osteopenia is mostly identified by bone density reduction. It is not severe enough like osteoporosis [1][6].

Bone mineral density (BMD) tests, also known as T-scores[1], help in both identifying and treating the condition.

Commonly used ranges for T-score are given below:

- T-score ≥ -1: Normal.
- $-1 >$ T-score > -2.5 : Osteopenia.
- $-2.5 \geq$ T-score : Osteoporosis.

The risk of osteoporosis in someone with osteopenia is very significant [1].

Osteomalacia, in contrast, is a bone disorder marked by decreased mineralization in newly formed bone at sites of bone regeneration or turnover. In osteomalcia, the bone does not harden, in contrast to osteoporosis, where the bone weakens. Osteoporosis was formerly exclusively diagnosed in patients who had already experienced one or more fractures [1][7], but now it may be found by assessing bone mass using photon(single or dual) absorptiometry [1][8] or quantitative computed tomography[1][9].

1.1. Osteoporosis Occurences

Osteoporosis is recognized as a significant health concern, responsible for nearly 9 million fractures annually[2] [10]. This condition affects approximately 60% of males and 21% of females aged 50 years and older, impacting an estimated 500 million individuals worldwide (both genders)[2][11][12] [13] [14]. A survey conducted in the year 2000 across America and Europe revealed that osteoporosis-related fractures primarily included hip fractures (1.6 million cases), forearm fractures (1.7 million cases), and clinical vertebral fractures (1.4 million cases). Of these, 51% occurred in the Americas and Europe, followed by the Western Pacific and Southeast Asia regions[2] [10]. In 2019, an estimated 32 million individuals aged over 50 were diagnosed with osteoporosis in Europe (including the EU, Switzerland, and the UK), accounting for 5.6% of the region's population. This figure included approximately 25.5 million women and 6.5 million men[2] [15]. Projections indicate that by 2050, the global incidence of hip fractures will rise by 310% in men and 240% in women compared to 1990 rates[2] [16].

1.2. Diagnostic Methods

To detect osteoporosis, two types of methods are commonly used: morphometric and absorptiometric[17]. Morphometric methods involve quantitative and qualitative analysis from plain radiographs (X-ray images). The X-ray principle states that calcium concentration directly correlates with the X-ray energy absorption. Less energy is absorbed and more lucency (an area of low density) is visible on the radiograph if the bone mass is decreased. Some commonly used absorptiometric methods are discussed next.

[1] https://www.myosteoteam.com/resources/types-of-osteoporosis

[2] https://bit.ly/4gIV1yR

(a) Radiographic Absorptiometry (RA): This method involves the use of a reference step wedge, typically made from hydroxyapatite or an aluminum alloy with an atomic number close to that of bone. The wedge is radiographed alongside specific bone sites, such as the metacarpals, radius, ulna, or femur. Bone density is determined by analyzing the bone and the reference wedge after processing the radiographic film. While RA is costeffective, fast, and simple to perform, it has limitations, including unstable X-ray sources, inconsistent beam quality, and issues related to processing and radiation scattering, all of which compromise image quality, accuracy, and precision. Additionally, due to its high dependence on operator skill, RA is not suitable for routine diagnostic applications. The emergence of nonradiographic absorptiometry techniques has largely rendered this method outdated.

(b) Radiogrammetry: Radiogrammetry primarily focuses on quantitative analysis of tubular bones, particularly cortical geometry. Cortical thickness serves as a key parameter for estimating bone mineral density (BMD). In digital radiogrammetry (DXR), high-performance computer systems process digitized images. This process includes identifying anatomical landmarks to automatically determine the ”Regions of Interest” (ROIs). The scanned image, with the overlaid ROIs, is displayed on the screen. Advanced algorithms, such as those based on active shape models (ASMs), are employed to detect and analyze the metacarpals[18].

(c) Single-energy Quantitative Computed Tomography (QCT) : This method utilizes readily available CT scanners and calibration phantoms. This method provides Satisfactory reproducibility, sensitivity and specificity. However, it also has certain drawbacks such as poor accuracy and relatively high radiation exposure. Accuracy can be raised using dual-energy QCT, albeit at the expense of subpar reproducibility and more radiation exposure.

(d) Single Photon Absorptiometry(SPA): Single Photon Absorptiometry is an effective method for estimating bone density. This quantitative analysis method can help in detecting Osteoporosis. It is based on the principle that the bone mineral content is proportional to the absorption of radioactive substances by bone tissue. This method, also known as gamma-ray absorption, calculates the attenuation degree of single energy gamma photon beam through bone tissue. Higher the attenuation degree, the higher is the absorption by bone minerals and the higher is the bone mineral content. SPA is suitable for bone mineral quantization. It is simple and consumes less radiation. It can help in early detection of trabeculae osteopenia. However, it is not free from limitations. SPA is not cost-effective, since it requires to replacing radioactive sources twice or thrice in a year. In addition to that it is not able to handle the spine and proximal femur effectively.

(e) Dual Photon Absorptiometry (DPA): This method is useful in diagnosing Osteoporosis. It can help measure the mineral content (density) of bone based on the absorption of energy (produce by DPA machines) by the soft tissues subtracted from the absorption of energy by the bone. A DPA machine produces two distinct energy levels i.e., 44 keV and 100 keV. DPA can give bone mineral content of spine more accurately towards estimation of BMD for central skeletal bones. Like the previous one, it is also not free from limitations. The four main sources of DPA that may cause errors in the BMD estimation are :(i) mechanical wears and tears, (ii) inappropriate patient placement, (iii) inaccurate metal placement in the abdominal area, and (iv) use of professionals for interpretation.

(f) Single Electron X-ray Absorptiometry (SEXA): It differs from SPA in its use of X-ray instead of photon energy. It is relatively compact and portable. It is more preferred to SPA in measuring heel, wrist and calcaneus BMD. *(g)* Double Electron X-ray Absorptiometry (DEXA): This technique employs dual-energy transmission scanning, where photons are emitted from an X-ray source. It offers improved reproducibility, reduced scanning duration, and produces images that are comparable in quality to standard skeletal radiographs. Both DPA and DEXA can be used to assess the mineral density of the proximal femur as well as the entire skeleton.

Although, DEXA has several merits, cost wise it is on the higher side, hence, not preferred by common patents. Despite having a good number of diagnosis methods for Osteoporosis detection, a cost-effective, yet a highly accurate and scalable method for the detection of such a multifactorial disease is a need of the hour.

In order to predict osteoporosis and its stages before it clinically appears as fractures, machine learning algorithms have recently been proven to be useful in evaluating plain radiographs by extracting their key features in conjunction with low-level picture content-based analysis. However, the cost-effective detection of osteoporosis in X-ray images with high accuracy is a challenging task. In this work, we present a cost-effective yet robust integrative approach (deep learning as well as tree-based ensemble learning) that can help detect osteoporosis from X-rays with high precision. Our approach has already been tested on other X-ray images and the performance has been found satisfactory.

2. Problem Definition

We define the detection of osteoporosis as a binary classification problem. For a given test X-ray image or its k-dimensional feature representation x^{test}_i , the method is to identify the class label *y*(normal or disease, i.e., osteoporosis) with high precision at a low cost with reference to a given set of training instances x^{train}_i which may be X-rays or their k-dimensional feature representations. Here, the training and testing images are views of the crucial parts (such as the Pelvis or knees) of human bodies of various age categories or the representation of the images in terms of optimal subsets of features. The output is a binary label $y \in \{0,1\}$, representing the absence (0) or presence (1) of osteoporosis. We aim to achieve a high detection accuracy (i.e., minimum false alarm) for any given input test instance in identifying their class labels at a low cost.

3. Proposed Method: InOsteoD

This section outlines the conceptual framework of InOsteoD, which employs an integrative approach for osteoporosis detection. The major components of InOsteoD are : (i) Pre-processing, (ii) Region of Interest (RoI) extraction,(iii) Feature Extraction),(iv) Feature Selection, (v) Consensus function and finetuning. Each of these components is described next. We aim to generate results (Class label for osteoporosis or normal) with high accuracy i.e., minimum false alarms. To achieve it, we incorporate a feedback analysis (post-alarm diagnosis) component which keeps provisions for fine-tuning the parameter(s) if the results are found not satisfactory.

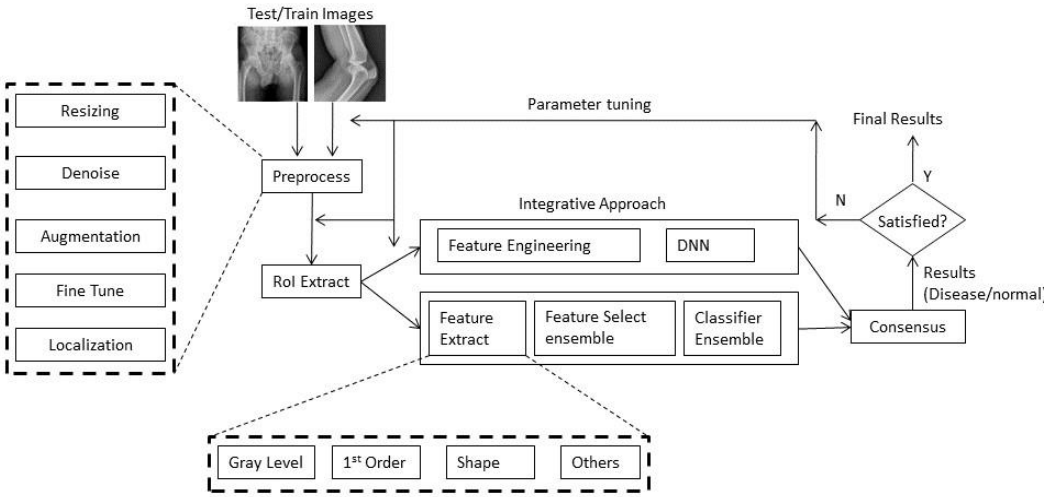

Figure 1. A Generic Framework of InOsteoD

3.1. Preprocessing

This component is responsible for transforming the input raw sample radiographs/X-rays into qualitative and ready-to-use images for subsequent processing.

3.2. Region of Interest(RoI) Identification

The samples (X-rays in DICOM format) used in our study do not have any RoI annotations. We extract the RoIs semiautomatically with the support of domain experts' knowledge. We consider the lower part of the pelvis region as potential ROIs. The method extracts the RoIs from all the given sample radiographs for subsequent downstream processing such as feature extraction, selection, and analysis. Appropriate RoI extraction can help (i) minimize the cost of processing and analysis significantly and (ii) achieve better accuracy. Prior to downstream analysis, RoIs extracted are validated with the support of domain experts (radiologists).

3.3. Feature Extraction

This component aims to extract salient features from the contents of the RoIs of sample images and transforms the images into feature vectors for easier handling without losing any information from the original data. Features may include form related to volume, shape, surface, density, intensity, texture, location, and relations with the surrounding area. We consider the radiomics features of X-rays to get additional insights into the images. The previous component i.e., preprocessing has an important role on the extraction of quality radiomics features. More focused and relevant region of interest identification can help extract relevant and crucial features significantly. A total of 120 radiomics features are extracted which can be further grouped into 8 categories.

The description of features is listed in Table 1. A conceptual framework of the feature extraction process is given in Figure 2. Each of the feature categories is described in brief, next.

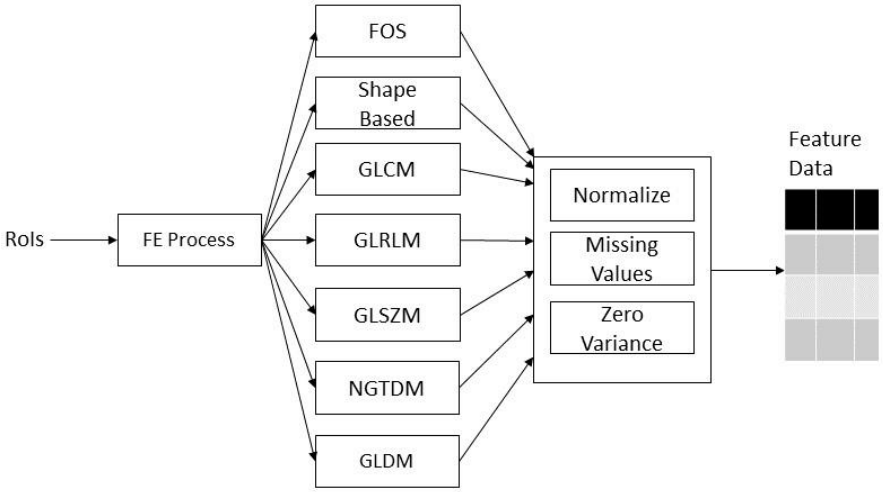

Figure 2. The conceptual framework of the feature extraction process

Table 1. Categories of Extracted Features

Feature Category	No. of Features
Intensity-Based Metrics (First-Order)	19
3D Shape-Based Features	16
2D Shape-Based Features	10
Texture Features (Co-occurrence Matrix)	24
Run-Length Features	16
Size Zone Features	16
Tone Difference Features	5
Dependency Features	14

(a) First-Order Statistics provide basic measures to describe the voxel intensity distribution within the specified region of interest.

(b) Shape-Based Features (2D and 3D) focus on the geometric attributes of the region of interest and the dimensions of the image. These features do not rely on the gray-level intensity distribution within the region.

(c) The Gray-Level Co-occurrence Matrix (GLCM) is a second-order statistical texture analysis method. It examines the spatial arrangement of pixel intensities to measure how often specific combinations of pixel values appear at a given spatial distance and direction.

(d) Advanced texture features are derived using the Gray-Level Run-Length Matrix (GLRLM). This method identifies ”gray-level runs,” which are sequences of pixels with the same intensity along a specific direction. The length of these runs is computed, representing the count of pixels in each sequence.

(e) The Gray-Level Size Zone Matrix (GLSZM) measures the size of uniform gray-level zones in an image, where each zone consists of connected voxels with identical intensity values.

(f) The Gray-Level Dependence Matrix (GLDM) assesses the degree of dependency among various gray levels in an image, providing insights into the image's homogeneity.

(g) The Neighboring Gray-Tone Difference Matrix (NGTDM) quantifies the intensity differences between a pixel and the average intensity of its neighboring pixels at a given distance.

3.3.1. Feature Data Preprocessing

After feature extraction, we transform the feature data into an unbiased ready-to-use form by applying appropriate pre-processing techniques such as normalization, missing value estimation, and zero variance elimination. The normalization technique brings down the feature values into a uniform and similar range. Zero variance feature elimination is applied to avoid redundancy. If any missing value occurs, it is imputed using the appropriate estimation technique.

3.4. Feature Selection

This component is responsible for selecting an optimal subset of relevant and interesting features from the extracted feature data. An optimal subset of relevant and non-redundant features can help improve classification accuracy and computational efficiency significantly. In the past two decades, a good number of feature selection techniques have been introduced [19] [20]. However,none of these are free from biasness. To overcome it, we use an ensemble feature selection technique [29][30] with three base feature selectors namely CMIM (Conditional Mutual Information Maximization) [21], MIFS (Mutual Information based Feature Selection) [22], and MRMR (Maximum Relevance — Minimum Redundancy). These three ranker algorithms are considered due to their consistently good performances in our previous work [23] [24]. The individual feature subsets are aggregated into one merged feature subset using an ensemble scheme which is depicted in figure 3. The list of selected features is listed in Table 3. To identify an optimal subset of features, we apply recursive elimination on the merged list of features given by our ensemble method. The feature elimination process is initiated with the lowest rank feature in the merged list and continues the process until there is no improvement in terms of classification accuracy. For example, if the merged list of ranked features initially given by the ensemble method includes n features such as $\{f_1,f_2,...,f_n\}$, where f_1 is the highly relevant (with the highest rank) and f_n is the lowest rank feature. In the elimination process, we initiate the process with f_n, and discard f_n iff it leads to improved performance with the rest (n-1) features and subsequently with fn−1 and so forth.

Table 2. Results of feature selection

CMIM	**MIFS**	**MRMR**	**Aggregated List**
No. of optimal features selected: 2			No. of Aggregated features: 3
1. Diagnostics Image original Mean 2. Original shape Maximum 2D Diameter Column	1. Diagnostics Image original Mean 2. Diagnostics Image original Minimum	1. Diagnostics Image original Mean 2. Diagnostics Image original Minimum	1. Diagnostics Image original Mean 2. Original shape Maximum 2D Diameter Column 3. Diagnostics Image original Minimum

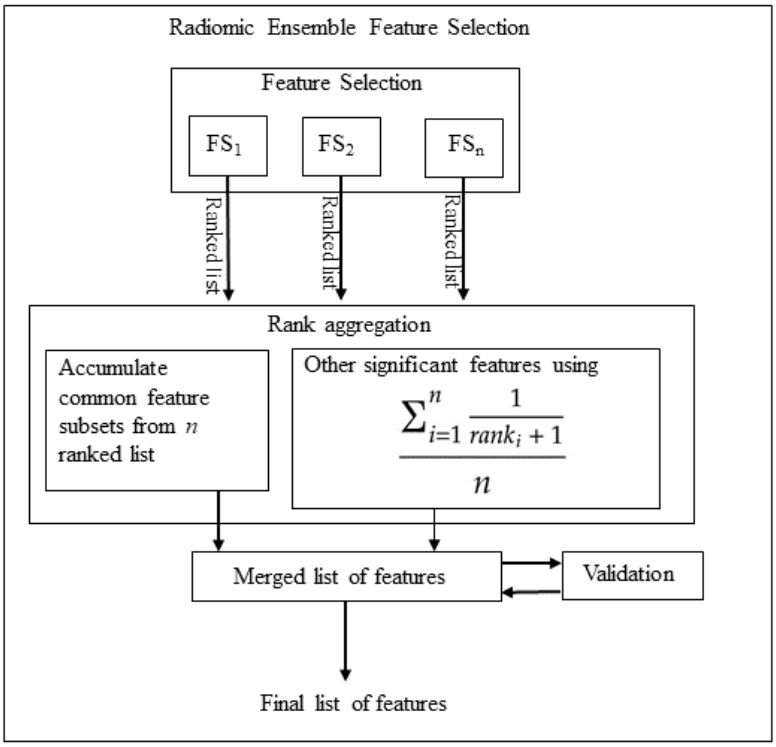

Figure. 3 The generic framework of the ensemble feature selection

Proposition 1 *The feature selected by InOsteoD is optimal for best possible accuray*

Proof: Let us consider a given sample feature dataset of X-ray images with n instances and each instance is of m-dimensions. Let us assume that the no. of dimensions selected by InOsteoD for best possible accuracy is k which is not optimal. Now, k features are obtained after (i) aggregation and (ii) recursive elimination. So, with (k+1) features there are no chances of significant accuracy improvement. Again, with (k-1) features, there is a high possibility of performance degradation in terms of classification accuracy.

Hence, the assumption is false and hence the proof.

3.5. Machine Learning Method Selection

In this study, different traditional machine-learning methods are considered to evaluate their performance on the generated feature dataset. Although, in the past decades, the world has seen a large number of supervised methods [25] [26] for the classification of test instances with respect to a given set of training instances, however, most classifiers fail to perform consistently across the domains, especially when adequate number of samples for training is not available. To address this issue, we consider an ensemble of five tree-based classifiers such as Random forest, Gradient boosting, Adaboost, XGBoost, and Extra Tree which have been found to perform significantly well in comparison to their other counterparts for numeric feature data. We consider the individual outcomes of these classifier and combine them using an appropriate consensus function toward the generation of unbiased results for better accuracy. A conceptual framework of this approach is depicted in Figure 7. The consensus function used by InOsteoD is based on weighted majority rule. These five classifiers have already been

evaluated on large number of benchmark datasets available in the UCIML [3] repository and based on their performance (consistent winning), these classifiers have been assigned weightage. While individual weights are aggregated, the consensus function uses the weights.

Proposition 2 *The performance of tree based ensemble cannot be less than individual classifier.*

Proof: Let the best possible classification accuracy given by these five classifiers be *x*%. Now, our tree ensemble is based on weighted majority based consensus function, in other words, while deciding the final class label for a given test instance, it gives high priority to the decisions given by the classifiers which have already been established as superior in comparison to others. Hence, the performance given by the ensemble cannot be less or inferior than those consistently well performing individual classifiers, hence the assumption is false and hence the proof.

3.6. Deep Learning Model

In addition to the tree based ensemble approach which operates on the optimal numeric feature representations of the RoIs of sample X-Ray images, we also use state of the art machine learning technique i.e., deep learning model in our integrative approach for osteoporosis prediction. In the deep learning-based neural network, Convolutional Neural Networks (CNN) are found to be very effective on image data. In CNN, several models have been introduced such as Resnet, Densenet, Alexnet, etc. These models are capable of handling voluminous data. However, depending on the availability of (i) quality and

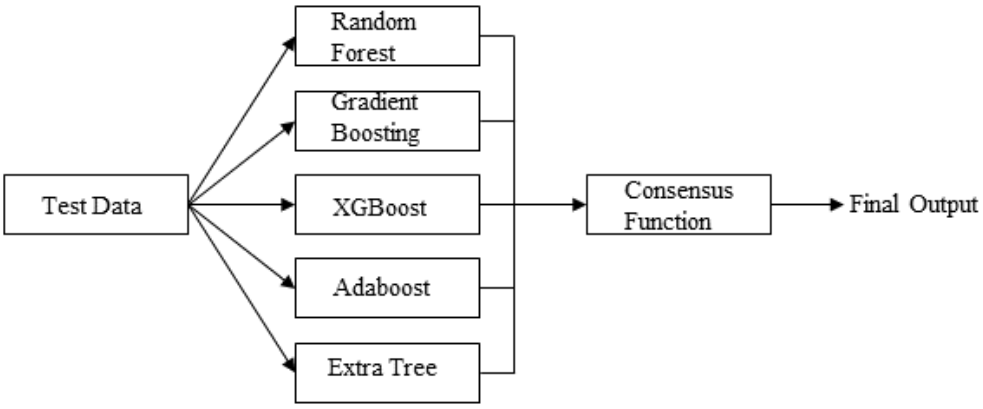

Figure 4. A conceptual framework of the tree ensemble approach

(ii) adequate number of samples, the performances of these models may vary. Therefore, we consider eight CNN models in our experimental research to justify the effectiveness of these models in the present problem in terms accuracy, specificity, and sensitivity.

To achieve cost-effective classification, we employ a transfer learning approach, initializing the network architectures with weights from a model pretrained on ImageNet. In

[3] https://shorturl.at/Vc1PO

our method, the pre-trained model's fully connected layer is replaced with a new one customized for our specific task. Two neurons make up the last layer, which represents the application's two target classes: osteoporosis and normal. The random weights of this recently introduced layer must be given appropriate weights. In order to maintain the layer weights from the pre-trained network, the network optimizer adjusts the weights of the recently added layer by freezing the prior layers. In order to acquire the optimal range of learning rate without participating in extensive experimentation, we also decide on a suitable learning rate prior to training utilizing a technique known as cyclical learning rates. Next, with all previous layers frozen, we train the newly inserted layer at random for one epoch, because the freshly added layer contains random weights. The layers are then unfrozen and trained for the specified number of epochs. The different architectures used in this work are listed in Table 4. A conceptual framework of a deep CNN model is shown in Figure 5. Transfer learning is used to fine-tune all of the models so they can accurately differentiate osteoporosis samples from normal or healthy samples.

The results generated against each test sample by the deep CNN are fed to the consensus function like the results given by the tree ensemble component. For the consensus component, there can be four distinct cases such as (i) case 1, when both the distinct components agrees on disease class i.e., '1' and the final output will also be '1'. (ii) case 2, when both agrees on healthy class, i.e., both generate '0' and the final output will also be 0. However, in case 3('0' & '1') and case 4('1' & '0'), where the components disagree, we need to decide the priority for each of these components. We consider radiomic feature-based

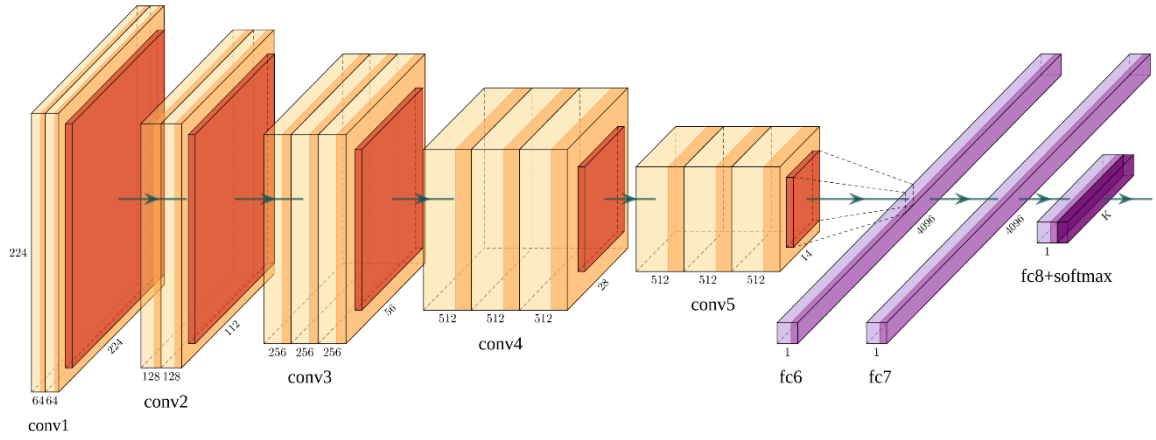

Figure 5. A conceptual framework of a deep CNN model

tree ensemble component as a high priority in the present experimental setup due to the significantly better performance shown by this component and also since with the available number of samples, the deep learning component may sometime generate false alarm. However, with a larger number of samples, it does not arise.

Our method is significant in view of the following points.

(a) X-Ray imaging is less costly than DEXA based method.

(b) Radiomic features are highly expressive and helpful in the effective discrimination of disease samples.

(c) Access to DEXA facilities is often limited, requiring patients to travel to distant referral centers.

(d) DEXA assesses bones and muscles but may be affected by the presence of fat. Furthermore, as a 2D projection method, it cannot account for bone geometry, microstructure, or size comprehensively[27].

On the other hand, changes in plain radiographs or X-rays are observable only after at least 30-50% bone loss.

3.7. Performance Evaluation

We conducted experiments in two phases to assess the performance of our method. In first phase, we used synthetic and benchmark datasets available at UCI ML repository[4], whereas in the second phase, we used real life samples obtained from AIIMS Bhatinda. However, in this paper, we focus our experimentation and result analysis centered around the later sample data. To validate the performance, we used three commonly used statistical measures. The next section present the dataset used and the measures used for validation.

3.7.1. Datasets Used

To understand the effectiveness of the method, we use pelvic X-ray image dataset provided by AIIMS Bhatinda. The images in the dataset belong to two classes i.e., normal and Osteoporosis. Prior to entering them into the network,

Table 3. Description of the dataset

Datasets	No of samples	Format	Annotations	Type
Original	705	DICOM	No	Gray

The images for the deep CNN component are shrunk to lower dimensions and then randomly cropped to either their full height or width for each epoch. In our scenario, an image interpolation method known as bilinear non-adaptive interpolation is used for image resizing. Then, while maintaining the original data's class labels, we employ data augmentation techniques to construct transformed versions of the raw data. A total of six data augmentation techniques are applied for creating variability in the dataset. The descriptions of the datasets before and after augmentation are tabulated in Table 3.

For the radiomic feature dataset, we extracted the features reported in Table1 for all the sample images. The extracted features are preprocessed using appropriate normalization technique followed by zero variance feature elimination to minimize redundancy. The preprocessed feature dataset is reported in Table 2.

3.7.2. Statistical Measure

To assess the performance of the machine learning algorithms, we consider three statistical measures namely accuracy, specificity, and sensitivity.

[4] https://shorturl.at/Vc1PO

(a) Accuracy: A metric for evaluating classification models. It defines how many predictions are correct out of the total predictions.

(b) Specificity: Specificity or True Negative Rate is where the model classifies data as a negative class given the data actually have the negative class. It quantifies the avoidance of false positives.

(c) Sensitivity: Sensitivity or True Positive Rate is where the model classifies data as positive given the actual class label as positive. Sensitivity quantifies the avoidance of false negatives.

3.7.3. Consultation with Domain Expert

In order to deal with the suspicious outcomes, we take help of the domain expert for further analysis. Based on the report of the domain expert, we further tune the models accordingly if necessary.

4. Performance Analysis

The method has been implemented in python using a Dell Precision 7810 workstation comprising 16 cores, 128GB RAM, NVIDIA Quadro GPU with 16GB VRAM, and Ubuntu OS. Preprocessing carried out on the datasets, and performance achieved are discussed next.

4.1. Results and Discussion

We carried out experiment in three phases. In phase 1, we implemented deep CNN models and tested them with both real-life and benchmark datasets. Phase 2 deals with tree based ensemble implementation and validation and finally, in phase 3, we implement the integrative approach and its valiation.

PyTorch 1.9.0 is used to implement each and every deep learning model. The weights of the network, particularly the recently added layers, are updated during the training phase using the cross-entropy loss function and the Adam optimization algorithm using the following parameter values: beta 1 = 0.98, beta 2 = 0.979, and epsilon = 1e-5.

During the training phase, to attain a learning rate of 0.003, the cyclical learning rate is employed. A batch size of 64 is applied throughout all of the tests that have been run. We divide the data into an 80:20 proportion for training and validation. After carefully training the models for 20 epochs, the weights achieving the highest validation accuracy are preserved. The performance of each model employed in the experiment are shown in Table 4 in terms of accuracy, sensitivity, and specificity. Resnet 50, which boasts an accuracy rate of 78.3%, provides the best performance. In order to understand the model decision on osteoporosis, a generalization of the class activation map called Grad-CAM [28] is used. It visualizes the significant information used by the deep learning model for the identification of osteoporosis. A coarse localization map is generated by incorporating the gradients of the target variable (osteoporosis) into the final convolutional layer, highlighting key areas in the image relevant for prediction. By identifying the unimportant areas that the model focuses on, the performance of the model is enhanced based on the study of the Grad-CAM.

In phase 2, for the radiomics feature data-based Implementation, five classifiers are used. All the classifiers are implemented in python framework. It can be noted that the

XGBoost classifier gives the best classification accuracy of 85.3%. Table 5 shows each model's performance in the experiment in terms of accuracy, sensitivity, and specificity.

During the phase 3 implementation, if both the components do not show agreements in terms of decisions (i.e., 0-0 or 1-1), we give priority to the decision given by the component which has been found to give consistently winning performance in the recent past. In the present scenario, we give high priority to the tree-based ensemble due to its superior performance. The best possible accuracy obtained in this phase is 85.3% which is similar to the phase 2 results. However, the results may improve with improved number of samples.

Table 4. Deep CNN performance on Osteoporosis image datasets

Model	Accuracy	Sensitivity	Specificity
ResNet18	77.1	77.8	77.2
ResNet34	78.1	78.7	78.4
ResNet50	**78.3**	**78.9**	**78.6**
DenseNet121	76	76.6	76.12
DenseNet161	76.4	77.1	76.6
DenseNet169	76.4	77.12	76.58
AlexNet	75	75.7	75.2
VGG	72	72.6	72.1

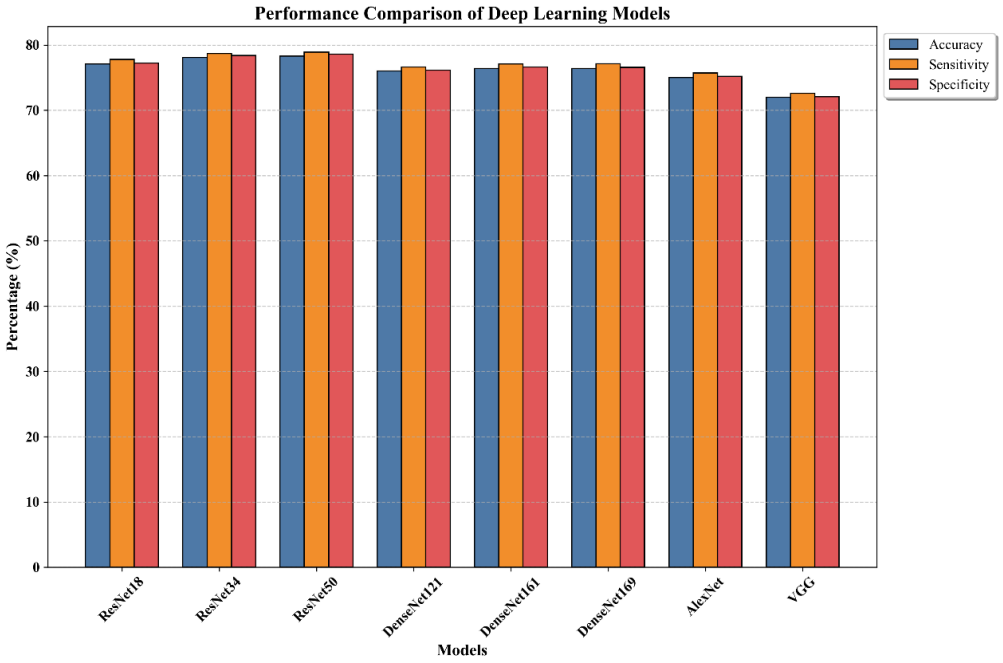

Figure 6. Performance Comparison of Deep Learning Models

Table 5. Ensemble classifier performance on Osteoporosis feature datasets

Model	Accuracy	Sensitivity	Specificity
Random Forest	83.1	83.7	83.1
Gradient Boosting	82.1	82.6	82.3
AdaBoost	83.3	84.1	83.6
XGBoost	**85.3**	**86**	**85.7**
Extra Tree	81.6	81.8	82

5. Limitations and Challenges

One major limitation is the issue of class imbalance, particularly for normal samples and osteoporosis class samples. This is because collecting such samples is difficult, and they are not publicly available. Privacy concerns also make it harder to access patient data. However, these challenges could be addressed by increasing efforts to collect diverse and balanced datasets while ensuring patient privacy. With access to more comprehensive data, the method could be significantly enhanced, improving its accuracy and reliability for broader applications.

6. Benefits of Multidisciplinary Research

Multidisciplinary research bridges diverse fields, combining the strengths of medical imaging, machine learning, and clinical expertise to address complex health challenges like osteoporosis. By integrating deep learning with Tree-based ensemble methods, this approach enhances diagnostic accuracy while remaining cost-effective. Such collaboration fosters innovative solutions, leveraging advancements in technology to improve patient outcomes and contribute to the broader field of healthcare.

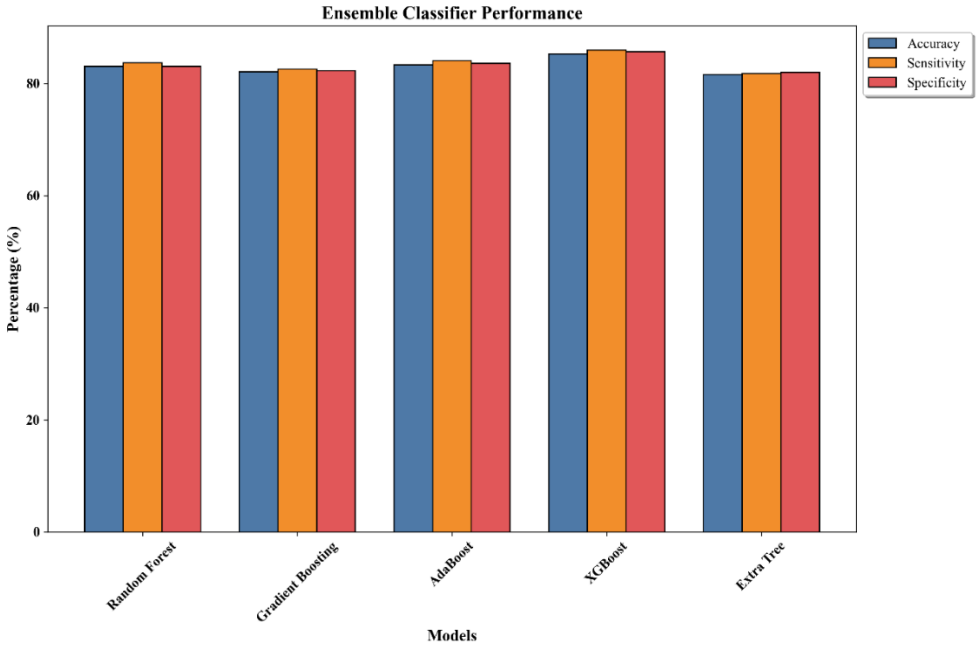

Figure 7. Performance Comparison of Ensemble Learning Models

7. Conclusions

A cost-effective, yet reliable method for Osteoporosis detection has been presented. To generate unbiased classification results, we exploit two proven machine learning approaches viz, deep learning, and tree-based ensemble learning on a relevant and optimal subset of features. The results obtained from these two components are combined using an effective consensus function to produce the final result. The proposed method is cost-effective and can be operated on low-cost handheld devices such as smartphones. The accuracy of the proposed method relies on having sufficient training samples. For these, there are enormous scopes for faster implementation of InOsteoD using a distributed approach in various phases such as preprocessing, feature extraction, feature selection, and classification.

Multidisciplinary Domains

This research covers the domains: (a) Medical Science, (b) Machine Learning, (c) Statistics.

Funding

This research received no external funding.

Conflicts of Interest

Conflict of Interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] Council, N.R.; Diet and Health: Implications for Reducing Chronic Disease Risk; 1989.
[2] Arnaud, C.D. The calcitropic hormones and metabolic bone disease. *Basic Clin. Endocrinol.* **1986**, *202–271*.
[3] Baron, R. Lymphocytes, macrophages, and the regulation of bone remodeling. *Bone Miner. Res.* **1984**, *2*, 175–243.
[4] Dobbs, M.B.; Buckwalter, J.; Saltzman, C. Osteoporosis: the increasing role of the orthopaedist. *Iowa Orthop. J.* **1999**, *19*, 43.
[5] Odell, W.; Swerdloff, R. Male hypogonadism. *West. J. Med.* **1976**, *124*(6), 446.
[6] Alshehri, M. Osteoporosis, osteopenia, and osteomalacia. *Indo Am. J. Pharm. Sci.* **2019**, *06*(02), 3356–3361, https://doi.org/10.5281/zenodo.2560732.
[7] Jacoby, I.; Clark, S.M. Direct mailing as a means of disseminating NIH consensus statements: a comparison with current techniques. *JAMA* **1986**, *255*(10), 1328–1330.
[8] Mazess, R.B.; Barden, H. Single-and dual-photon absorptiometry for bone measurement in osteoporosis. *Osteoporos. Update* **1987**, *73–80*.
[9] Genant, H.K.; Block, J.E.; Steiger, P.; Glueer, C.-C.; Smith, R. Quantitative computed tomography in the assessment of osteoporosis. *Semin. Nucl. Med.* **1987**, *17*, 316–333.
[10] Johnell, O.; Kanis, J. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporos. Int.* **2006**, *17*(12), 1726–1733. https://doi.org/10.1007/s00198-006-0172-4
[11] Kanis, J.; Johnell, O.; Oden, A.; Sernbo, I.; Redlund-Johnell, I.; Dawson, A.; De Laet, C.; Jonsson, B. Long-term risk of osteoporotic fracture in Malmö. *Osteoporos. Int.* **2000**, *11*(8), 669–674. https://doi.org/10.1007/s001980070064
[12] Melton III, L.J.; Atkinson, E.J.; O'Connor, M.K.; O'Fallon, W.M.; Riggs, B.L. Bone density and fracture risk in men. *J. Bone Miner. Res.* **1998**, *13*(12), 1915–1923. https://doi.org/10.1359/jbmr.1998.13.12.1915
[13] Melton III, L.J.; Chrischilles, E.A.; Cooper, C.; Lane, A.W.; Riggs, B.L. Perspective: How many women have osteoporosis? *J. Bone Miner. Res.* **1992**, *7*(9), 1005–1010. https://doi.org/10.1002/jbmr.5650070902
[14] Curtis, E.M.; van der Velde, R.; Moon, R.J.; van den Bergh, J.P.; Geusens, P.; de Vries, F.; van Staa, T.P.; Cooper, C.; Harvey, N.C. Epidemiology of fractures in the United Kingdom 1988–2012: variation with age, sex, geography, ethnicity, and socioeconomic status. *Bone* **2016**, *87*, 19–26. https://doi.org/10.1016/j.bone.2016.03.006
[15] Kanis, J.A.; Norton, N.; Harvey, N.C.; Jacobson, T.; Johansson, H.; Lorentzon, M.; McCloskey, E.V.; Willers, C.; Borgström, F. SCOPE 2021: A new scorecard for osteoporosis in Europe. *Arch. Osteoporos.* **2021**, *16*(1), 1–82. https://doi.org/10.1007/s11657-020-00871-9
[16] Gullberg, B.; Johnell, O.; Kanis, J. World-wide projections for hip fracture. *Osteoporos. Int.* **1997**, *7*(5), 407–413. https://doi.org/10.1007/pl00004148
[17] Oei, L.; Koromani, F.; Rivadeneira, F.; Zillikens, M.C.; Oei, E.H. Quantitative imaging methods in osteoporosis. *Quant. Imaging Med. Surg.* **2016**, *6*(6), 680. https://doi.org/10.21037/qims.2016.12.13

[18] Cootes, T.F.; Taylor, C.J. Active shape models—'smart snakes.' In *BMVC92*; Springer **1992**; pp. 266–275.

[19] Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature selection: A data perspective. *ACM Comput. Surv. (CSUR)* **2017**, *50*(6), 1–45. https://doi.org/10.1145/3136625

[20] Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*(1), 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024

[21] Fleuret, F. Fast binary feature selection with conditional mutual information. *J. Mach. Learn. Res.* **2004**, *5*(Nov), 1531–1555.

[22] Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*(8), 1226–1238. https://doi.org/10.1109/tpami.2005.159

[23] Sarmah, U.; Bhattacharyya, D. Cost-effective detection of cyber-physical system attacks. In *Adv. Mach. Learn. Big Data Anal.*; Springer **2022**; pp. 33–69. http://dx.doi.org/10.1007/978-981-16-8930-7_2

[24] Borah, P.; Bhattacharyya, D.K.; Kalita, J.K. Cost-effective method for ransomware detection: An ensemble approach. In *Int. Conf. Distrib. Comput. Internet Technol.*; Springer **2021**; pp. 203–219. http://dx.doi.org/10.1007/978-3-030-65621-8_13

[25] Osisanwo, F.; Akinsola, J.; Awodele, O.; Hinmikaiye, J.; Olakanmi, O.; Akinjobi, J. Supervised machine learning algorithms: Classification and comparison. *Int. J. Comput. Trends Technol. (IJCTT)* **2017**, *48*(3), 128–138. https://doi.org/ 10.14445/22312803/IJCTT-V48P126

[26] Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised machine learning: A review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* **2007**, *160*(1), 3–24.

[27] Ferizi, U.; Honig, S.; Chang, G. Artificial intelligence, osteoporosis, and fragility fractures. *Curr. Opin. Rheumatol.* **2019**, *31*(4), 368. https://doi.org/10.1097/bor.0000000000000607

[28] Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proc. IEEE Int. Conf. Comput. Vis.* **2017**; pp. 618–626. https://link.springer.com/article/10.1007/s11263-019-01228-7

[29] Sarmah, U.; Bhattacharyya, D.K. Cost-effective detection of cyber-physical system attacks. In *Adv. Mach. Learn. Big Data Anal.*; Springer: Singapore **2022**; pp. 33–69. https://doi.org/10.1007/978-981-16-8930-7_2

[30] Sarmah, U.; Borah, P.; Bhattacharyya, D.K. Ensemble learning methods: An empirical study. *SN Comput. Sci.* **2024**, *5*(7), 924. https://doi.org/10.1007/s42979-024-03252-y