

Osteoporosis Detection and Classification of Femur X-ray Images Through Spectral Domain Analysis using Texture Features

Dhanyavathi A, Veena M B

Department of Electronics and Communications, BMS College of Engineering, Bengaluru, India
Visvesvaraya Technological University, Belagavi-590018, India

Abstract—Osteoporosis commonly diagnosed as a bone disorder that affects the significant portion of the population. The Dual X-ray Absorptiometry (DXA) is one of the most accepted standard methods of analyzing the bone disorder, but it is exorbitant. However X-ray is a cost effective, therefore the proposed work introduces a new technique to improve osteoporosis detection and classification of femur bone X-ray image. The spectral based sub band images texture features are used to analyze the Region Of Interest (ROI) femoral head trabecular bone. A spectral domain based on the Two-Dimensional Discrete Wavelet Transform (2D-DWT) is used to represent variations in finer details in the image. Trabecular femur bone texture is determined only by horizontal, vertical, and diagonal sub bands of DWT coefficients. The sub band images are further enhanced by applying the maximum response filter (MRF) at different scales, thereby enhancing the most significant responses. Consequently, the sum of the MRFs of different scale images is considered as the supervised database. To detect osteoporosis, the test and supervised images are analyzed to calculate two significant attributes such as Zero Mean Normalized Cross-Correlation (ZMNC) and Sum Squared Difference (SSD). Based on experimental results, the performance metrics measure is improved in all aspects over current methods.

Keywords—Classification; feature; femur; images; normal; osteopenia; osteoporosis; texture

I. INTRODUCTION

This The disease osteoporosis causes loss of bone density and increases the risk of fractures in millions of people worldwide [1][2]. Debilitating fractures can be effectively managed and prevented with early detection. Osteoporosis is often diagnosed with X-ray imaging, but traditional methods often rely on visual assessment, which may be subjective and subject to human errors [3]. A Convolution Neural Network (CNN) model was used to assess osteoporosis based on hip radiographs. An ensemble model with clinical covariates was also investigated [4]. From a single Dual-Energy X-Ray Absorptiometry (DXA) image of the proximal femur, reconstruct both the 3D bone shape and the Three Dimension Bone Mass Density (3D-BMD) distribution [5]. A set of Quantitative Computed Tomography (QCT) scans, a statistical model of the combined shape and BMD distribution is constructed to detect osteoporosis [6]. A method of estimating the apparent physical BMD of the proximal femur from CT images with good accuracy when evaluating post-menopausal

osteoporosis. The proximal femur radiographs were analyzed using Gabor filters, wavelet transformations, and fractal dimensions-based texture analysis methods to identify osteoporosis [7]. The volumetric estimation of femur bone based on an x-ray image using a computer-based algorithm to detect osteoporosis [8]. A comparison of BMD of CT scan and BMD revealed a difference of 4.53 percent in volume. An analysis was conducted to examine how they related with BMD and anthropometric factors such as height and weight [9]. In recent study, 34% of Indian women had osteoporosis and 20% had osteopenia, respectively. The study measured the energy of the proximal femur trabecular bone as a result of osteoporosis postmenopause using dual-tree complex wavelet transforms (DT-CWT) [10]. DT-CWT has been successfully used to analyze the trabecular pattern on the right proximal femur on radiographs. The Gabor filter was used to calculate features from the trabecular pattern recorded on proximal femur radiographs in the assessment of osteoporosis [11]. In order to justify the classification result, Singh indexes of trabecular pattern are used. An expert system designed for diagnosing osteoporosis based on measuring bone texture using fuzzy X-ray images [12]. A fuzzy X-ray imaging technique analyzes trabecular bone texture and thus calculates bone density by combining resolution enhancement algorithms and edge detection algorithms. Both algorithms are efficient at calculating disease severity. An image of a femur bone can be classified using morphometric features from the image segmented [13]. The femur bone structure is segmented using active contour method from a 2D X-ray radiograph and morphometric measurements are calculated in pixel values for head diameter, head height, neck diameter, and intertrochanteric distance in the proximal femur and neck. Several images of patients with osteoporotic, osteopenia, and normal conditions are collected using computer tomography (CT) [14]. MIMICS software is used to analyze the condition more effectively. Initially, the images must be imported into the MIMICS software for analysis. An effected, normal femur bone is generated in 3D by MIMICS software and osteopenia and normal patients can use the analysis as a precaution. Gradient Harmony Search (GHS) optimization based deep networks are used for classification [15]. A Harmonic Search (HS) algorithm is incorporated with a Gradient Descent (GD) algorithm is used to build GHS. Utilizing machine learning and image processing techniques to detect early-stage fractures caused by osteoporosis in femur image [16]. The Fracture Risk Assessment tool (FRAX) calculations were performed

retrospectively on 560 volunteers (age at least 50 years) who underwent hip-spine X-rays, BMD scans and FRAX tool calculations [17]. To determine whether Cortical Thickness Index (CTI) and Canal Flare Index (CFI) are used to calculate neck BMD (nBMD) on anteroposterior radiographs, both indices were measured on anteroposterior radiographs. Based on characterization of BMD and trabecular bone microarchitecture, an automated approach was developed to predict biomechanical bone strength in proximal femur specimens [18]. As a diagnostic biomarker for osteoporosis diagnosis, tracking disease progression, and evaluating the response to therapeutic intervention, the automated and objective way trabecular bone microarchitecture is analyzed, and the subsequent reduction performance achieved suggest that it may be utilized in this manner. Two Deep learning Convolution Neural Networks (DCNNs) i.e., Alex Net and Google Net were trained on anteroposterior hip radiograph images to detect risk in the neck of femur [19]. Feasible element analysis (FEA) of the hip guided by high-resolution magnetic resonance imaging (MRI) has been developed to assess subject-specific bone strength [20]. If technique is further validated, management of hip fracture risks in the clinic may be useful. Researchers evaluated the involvement of spinal and hip flexion discordances in Korean patients with a typical femoral fractures and femur neck fractures [21]. Discordances might be affected by osteoporotic fracture locations. Using conventional radiographs obtained for various indications, a robust opportunistic screening tool for osteoporosis and fracture risk assessment was demonstrated to provide vertebral compression fractures detection, BMD estimation, and fracture risk estimation in a fully automated manner [22]. In assessing fracture risk, parameters and their interactions are analyzed [23]. In a multiple regression analysis, bone density and the loading directions in a sideways fall have been considered as independent variables along with the fracture risk index as a dependent variable. As independent variables, angle about the femoral neck axis at the coronal and transverse ends of the shaft was measured in both coronal and transverse planes. In the interaction analysis of parameters, bone density appears to have a greater effect on fracture risk. Although analyzing the current techniques for detecting osteoporosis in femur images helps to a definite conclusion, while several challenges remain, there is a need for the work to be advanced. To produce more comprehensive work and enhance system performance, a new approach of osteoporosis detection in femur image based on spectrum analysis is introduced in this work.

II. METHODOLOGY FOR PROPOSED WORK

The Fig. 1 illustrates the proposed work analysis is on the basis of spectral domain, 2D-DWT spectral domain analysis is introduced to analyze texture features of the X-ray of ROI femur images, offering valuable insights beyond the visual representation.

An analysis of the spectral features of image texture in two dimensions is carried out by one level decomposition of 2D-DWT [24][25]. Detecting osteoporosis involves two steps: the first is analyzing texture features on ROI images of femur bones and then deciding whether the given test image is normal (healthy bone) or abnormal (osteopenia or osteoporosis). By focusing on Horizontal Coefficients (HC), Vertical

Coefficients (VC), and Diagonal Coefficients (DC), the feature dimension in 2D-DWT can be reduced substantially. As a result, the proposed system model can detect bone diseases effectively with this reduced set of image texture. Ultimately, this work aims to obtain meaningful information from texture features by using symmetric wavelet family. With wavelet transformations, extracting texture information at different directions depending on how it is oriented. Further the image texture analysis of 2D-DWT is executed using MRFs at different scales. In order to classify normal or abnormal images, test and supervised images are matched based on attributes of ZMNCC and SSD of MRFs.

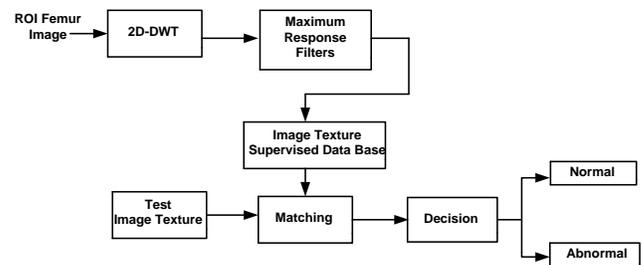


Fig. 1. Block schematic for the suggested work.

A. Two Dimension Discrete Wavelet Transform

The 2D-DWT is a versatile and powerful tool for analyzing and processing 2D data like images, and its significance be situated in its capability to provide a compact and meaningful representation that facilitates various texture analysis in spectral domain. There were four sub band images in a one level decomposed 2D-DWT is illustrated in Fig. 2. Rows of the input image matrix are first transformed with an LPF and HPF, then its columns are treated with an LPF and HPF to produce sub band images are like approximation coefficients i.e., AC and three detail sub band images like HC, VC, and DC correspond to horizontal, vertical, and diagonal coefficients, respectively and each sub band images size are quarter of the input image. All three detail coefficients are represented half-resolutions of the input image and edge variations are almost exclusively visible in these three images. These three sub band images are considered in this study because of their texture information gives significant variations of intensity for analyzing the trabecular micro architecture of the femur bone.

The Symlet-4 wavelet is the basis function used in this work, due to its higher-scale detail coefficients are captured as finer details and texture patterns, while lower-scale detail coefficients capture broader texture patterns. The Symlet-4 is a type of wavelet that is used due to its balance between compact support and frequency localization. The sym in Sym4 stands for symmetric, which means that it has a symmetric shape. It is like Daubechies wavelets but have slightly different properties. The significance of using the Symlet-4 wavelet for image texture analysis lies in its ability to capture both low-frequency and high-frequency information effectively.

B. Two-Dimensional Maximum Response Filter

A two-dimensional maximum response filter (2D-MRF) is introduced in this work to extract fine and coarse detail information of HC, VC and DC and its kernel is based on a Gaussian filters [26][27] at different scales. Each pixel location

must be analyzed in order to obtain the response values, first convolve the image with the Gaussian filter at multiple scales. Thus, the filtered output is calculated by taking the maximum value from the responses for each pixel location.

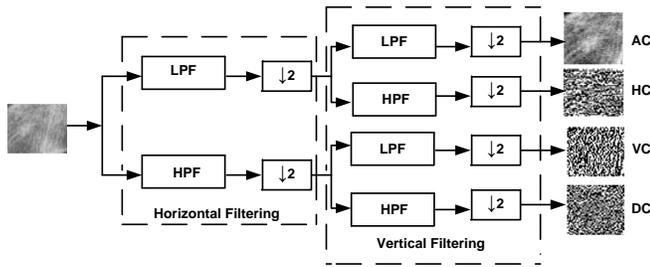


Fig. 2. First level 2D-DWT decomposition structure.

The MRF can help improve the robustness of feature extraction methods to noise and variations in the image. By emphasizing the most prominent responses and suppressing less significant ones, the filter helps in reducing the influence of noise or unwanted artifacts on the extracted features. The MRF can exhibit scale properties i.e., it can be effective in detecting the same pattern at different scales within the image. It helps in reducing the computational load by focusing on the most relevant features, this results in a faster processing and analysis of image datasets. Eq. (1) is the 2D-MRF using a Gaussian function, given an input sub band images $S_B(x, y)$, which are HC, VC and DC.

$$G(x, y) = \frac{1}{(2\pi\sigma^2)} e^{-\frac{(x^2+y^2)}{(2\sigma^2)}} \quad (1)$$

However, the Gaussian filter is characterized by a standard deviation, which is a scaling factor of the filter to compute the response values at each pixel location in the image i.e., where is the different scales and perform convolution operation ($*$) between the image and the Gaussian filter at multiple scales shown in (2).

$$R_V(x, y, S) = S_B(x, y) * G(x, y, S) \quad (2)$$

Each pixel location is evaluated at multiple scales after obtaining the response values, a maximum response filter selects the response value that is the highest i.e., from the set of responses for each pixel demonstrated in (3).

$$R_{max}(x, y) = \max(R_V(x, y, 1), R_V(x, y, 2), \dots, R_V(x, y, S)) \quad (3)$$

The total number of scales used in this work are two different sets i.e., [0.3, 0.6, 0.9] and [5, 10, 15] to get fine and coarse detailed texture information respectively. Finally, the filtered output $F_o(x, y)$ as in (4) represents the maximum pixel value at each location.

$$F_o(x, y) = R_{max}(x, y) \quad (4)$$

The MRFs process efficiently highlights the most salient features across different scales and emphasizes significant structures while suppressing less important ones. The scale

values determine the filter output i.e., Gaussian filter standard deviation (σ) can be adaptable to for fine and coarse image texture requirements. The total texture image $T_I(x, y)$ is a supervised image texture data in this proposed work, its value is calculated at each pixel location (x, y) in the image using the following algorithm.

Algorithm : Determination of fine and coarse texture image information

Step1: Fine texture analysis

$$\text{Find, } F_1(x, y) = R_{max}(x, y)$$

$$R_{max}(x, y) = \max(R_H(x, y, 0.3), R_V(x, y, 0.3), R_D(x, y, 0.3))$$

The $R_H(x, y, 0.3)$ is filter response due to HC input at 0.3 scale, the $R_V(x, y, 0.3)$ is filter response due to VC input at 0.3 scale and $R_D(x, y, 0.3)$ is filter response due to DC input at 0.3 scale.

Similarly, find $F_2(x, y) = R_{max}(x, y)$ at 0.6 scale

and $F_3(x, y) = R_{max}(x, y)$ at 0.9 scale.

Step 2: Determine fine texture sum i.e., $S_F(x, y)$

Step 3: Coarse texture analysis

Repeat Step 1: to find $F_4(x, y)$ at 5 scale,

$F_5(x, y)$ at 10 scale and $F_6(x, y)$ at 15 scale.

Step 4: Determine coarse texture sum $S_C(x, y)$

$$S_C(x, y) = F_4(x, y) + F_5(x, y) + F_6(x, y)$$

Step 5: Determine total texture image i.e., $T_I(x, y)$

$$T_I(x, y) = S_F(x, y) + S_C(x, y)$$

C. Matching and Decision process

A general match is based on finding test features that correspond to supervised databases [28]. In the matching process in which the test images and supervised images texture features are matching based on two attributes such as Zero Mean Normalized Cross-Correlation (ZNCC) and Sum of Squared Difference (SSD), which are obtained from total texture image i.e., The classification of the input test image is developed on having the highest possible value of ZNCC and lowest value of SSD, which are considered as the matching test classes. Pixels are compared based on their intensity values. These attributes help quantify the similarity between two images. Based on their intensity values, these attributes help quantify the similarity between two images.

- Zero Mean Normalized Cross-Correlation (ZNCC): Using the ZNCC, the cross-correlation between two images can be normalized. An image's similarity can be measured with it, where a value close to 1 indicates high similarity, while a value close to -1 indicates high dissimilarity.

Eq. (5) is the ZNCC between two images supervised i.e. test given by:

$$ZNCC(S_I, T) = \frac{\sum [(S_I(i, j) - \mu_{(S_I)}) (T(i, j) - \mu_{(T)})]}{\sqrt{\sum (S_I(i, j) - \mu_{(S_I)})^2} \sqrt{\sum (T(i, j) - \mu_{(T)})^2}} \quad (5)$$

Where: $S_I(i, j)$ is the intensity value of supervised image S_I at position (i, j) . The $T(i, j)$ is the intensity value of test image T at position (i, j) , The $\mu_{(S_I)}$ is the mean intensity value of image. The is the mean intensity value of image T , the \sum denotes the summation over all pixel positions (i, j) in the images.

- Sum of Squared Difference (SSD): It measures the difference between corresponding intensity of the supervised image and the test image by summarizing their squares. Using the (6), it measures the degree of two images diverge from one another.

$$SSD(S_I, T) = \sum (S_I(i, j) - T(i, j))^2 \quad (6)$$

where, the summation is performed over all pixel positions in the images. In SSD case, test and supervised images are more similar when the value is lower. Based on SSD and ZNCC, the test input ROI image class is identified by reflecting on $D1 = \text{Max}\{ZCCC\}$ and $D2 = \text{min}\{SSD\}$, if both D1 and D2 conditions are satisfied, that class is reflected as prediction class. The three classes are employed in this work as abnormal labeled as class-1 (osteopenia), class-2 (osteoporosis) and class-3 (normal).

III. PERFORMANCE METRICS

The confusion matrix provides valuable insights into the model performance by quantifying different types of classifications [29]. The Table I displays the general confusion matrix used in the field of medicine to categorize the patient's positive and negative health conditions. From the confusion matrix, calculate various evaluation metrics, such as accuracy, precision, recall, and F1 score, which help assess the model effectiveness in different aspects.

TABLE I. GENERAL REPRESENTATION OF CONFUSION MATRIX

	Predicted Negative	Predicted Positive
Actual Negative	TN (True Negative)	FP (False Positive)
Actual Positive	FN (False Negative)	TP (True Positive)

- 1) *True positive (TP)*: Indicates that, number of samples were positive predictions made correctly.
- 2) *True negative (TN)*: Number of negatively predicted samples that were correct.
- 3) *False positive (FP)*: Incorrectly predicted positive samples are measured by this metric.
- 4) *False negative (FN)*: Number of negative samples predicted incorrectly.

In accordance with the confusion matrix, here are various evaluation metrics: Model accuracy measures how accurate the predictions as in (7),

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (7)$$

Precision (Positive Predictive Value, PPV), estimates the positive samples predicted correctly by the model using (8),

$$Precision = \frac{TP}{(TP + FP)} \quad (8)$$

Recall (Sensitivity, True Positive Rate, TPR) measures the model ability to correctly identify positive samples among all actual positive samples as in (9),

$$Recall = \frac{TP}{(TP + FN)} \quad (9)$$

Specificity (True Negative Rate, TNR) measures the model's ability to correctly identify negative samples among all actual negative samples measured in (10),

$$Specificity = \frac{TN}{(TN + FP)} \quad (10)$$

The F1 score represents in (11) is a balanced measure of the model's performance by combining precision and recall together.

$$F1 \text{ score} = \frac{2(Precision * Recall)}{(Precision + Recall)} \quad (11)$$

Analyzing the performance of a model using these metrics provides valuable insight into different aspects. A high accuracy indicates overall good performance, it indicates good class predictions when precision and recall are high. As it incorporates both precision and recall, the F1 score is particularly useful when both are equally important. Interpretation of these evaluation metrics must consider the context. If there is an imbalanced dataset, accuracy might not be a reliable measure, and other metrics like precision-recall curve or area under the receiver operating characteristic (AU-ROC-) curve might be more informative [30]. This curve plots between two parameters: True Positive Rate (TPR), False Positive Rate (FPR).

IV. DATASET DESCRIPTION

Total 51 X-ray femur images were collected with the focal distance was set at 0.812 m. The X-ray parameters were 75-80 kV and 80 mAs for all patients. This study used images supplied by a reputed Bangalore hospital, Karnataka state, India. The image data consist of 2D radiographic images in JPEG format of size 2140×1760. In that 23 are normal, 10 are osteopenia and 18 are osteoporosis femur images. Table II listed the region of interested (ROI) femur bone of left and right of size 170×114, the total 102 ROI images including both left and right, which are considered to verify the proposed system experimentally.

TABLE II. DESCRIPTION OF DATA SET

Total ROI images: (23×2) + (10×2) + (18×2)=102					
Normal femur bone		Abnormal femur bone			
		Osteopenia femur		Osteoporosis femur	
Left	Right	Left	Right	Left	Right
23	23	10	10	18	18

A. Experimental Dataset Division

The proposed model can be experimentally tested by dividing the 102 total ROI of the image dataset into different cross-folding schemes, in the following Table III. A comparison between this and a previously conducted study on cross-folding methods reveals how well the proposed model performs.

TABLE III. EXPERIMENTAL DIVISION OF DATASET

Data - folding	Total number of samples of ROI images = 102					
	Supervised set			Test set		
	Normal	Abnormal		Normal	Abnormal	
	Osteopenia	Osteoporosis		Osteopenia	Osteoporosis	
Two-fold	12+12	5+5	9+9	11+11	5+5	9+9
	24+10+18=52			22+10+18=50		
Three-fold	16+16	7+7	12+12	7+7	3+3	6+6
	32+14+24=70			14+6+12=32		
Four-fold	18+18	8+8	14+14	5+5	2+2	4+4
	36+16+28=80			10+4+8=22		

V. RESULTS AND DISCUSSION

The input image of femur is a very low-quality image because the X-ray image of inner trabecular bones micro-architecture is not visible to distinguish between healthy bone and osteoporotic bone. Because of its versatility, DWT has become a very useful method for the process of decomposing an image into several resolutions through wavelet decomposition. By concentrating the wavelet energy in time and keeping its periodic properties, wavelets can simultaneously analyze both time and frequency of pixels intensity in the image. By decomposing a digital image into different sub bands, the 2D-DWT can resolve frequencies more precisely and time resolutions more coarsely at lower frequencies. Fig. 3 shows 2D-DWT output AC, HC, VC and DC. The approximation coefficient is same as the input image, Horizontal coefficient sub band is giving the horizontal information of texture features, vertical coefficient sub band gives the vertical information of texture features similarly the diagonal sub band gives the diagonal texture features. Only HC, VC and DC are texture information is sufficient to analyze the texture features of inner trabecular bones micro-architecture of femur bone.

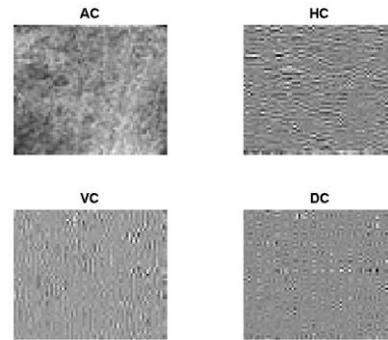


Fig. 3. First level 2D-DWT decomposition structure.

2D-DWT followed by Maximum Response Filter (MRF) based sub band image texture analysis is a technique used in order to extract meaningful texture features from an image as shown in Fig. 4, Multi-resolution analysis. In 2D-DWT, the image is decomposed into multiple frequency bands, representing different levels of detail or textures. This multi-resolution property is well-suited for texture analysis as textures often exhibit varying degrees of complexity at different scales. By applying the MRF at two sets of scale as [0.3, 0.6, 0.9] and [5, 10, 15] and filter size is technique after the 2D-DWT of HC, VC and DC, the most significant texture information from different sub bands can be extracted. MRF highlights regions with the maximum texture response, enhancing the representation of dominant texture patterns.

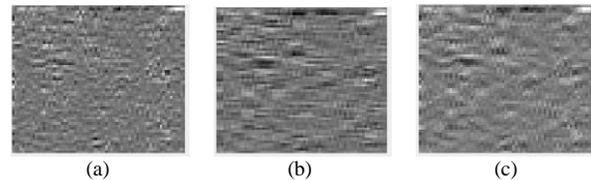


Fig. 4. First image texture (a) sum of MRFs scale at [0.3, 0.6, 0.9] (b) sum of MRFs scale at [5, 10, 15] (c) sum image of (a) and (b).

The matching is taking place between the supervised image database and test images based on SSD and ZNCC attributes and then decision is made on these attributes to decide whether the test ROI is abnormal (osteoporosis or osteopenia) or normal bone based on maximum ZNCC and minimum SDD.

According to the Table IV, different data cross-folding methods result in different confusion matrices. By applying cross-validation techniques, whether the model will perform well on unseen data and avoid over fitting or under fitting issues. As a result of supervised dataset for four folding, that is, 80 samples and the test dataset is 22 samples, the method achieves better results for four folding because of a greater number of supervised datasets. The system will have a larger number of texture patterns to decide whether an image is normal or abnormal, however osteoporosis and osteopenia bones both are included as abnormal in this proposed work. Actual positive sample (abnormal) set in test data set is 12 (osteoporosis and osteopenia) and actual negative sample set is 10 (normal). In the case of four folding, the model gives TP=12 and FP=0, which gives better results, however in normal case gives TN=9 out of 10 and FN=1.

TABLE IV. CONFUSION MATRIX FOR DIFFERENT FOLDING

Data Folding	TP	FP	TN	FN
Two-fold	11	3	4	4
Three-fold	12	2	5	3
Four-fold	12	0	9	1

Using graphical representation of non-normalized confusion matrix, system performance can easily assess at a momentary look as shown in Fig. 5 for four folding data. The diagonal of the confusion matrix (from top-left to bottom-right) represents correct predictions, while off-diagonal elements indicate misclassifications. This graphical representation helps in understanding the model strengths and weaknesses in terms of classifying different instances. The observation of four folding data from Table IV gives the TP=4+8=12 i.e., abnormal bone = Osteopenia (class-1) + Osteoporosis (class-2) and TN=10 i.e., normal bone (class-3). The system predicts better results with false negative was the only one sample, which is actually normal (healthy) bone showing it as abnormal.

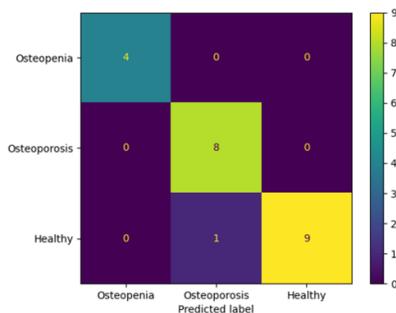


Fig. 5. Graphical representation confusion matrix for four folding data.

On all metrics measures, four folding achieves better results than the other folding, as revealed in Table V. Model performance metrics provide quantitative measures to evaluate how well a machine learning model is performing on a dataset. Metrics like accuracy, precision, recall, and other characteristics are essential for evaluating model. The choice of data folding (i.e., cross-validation method) can impact the way these metrics are computed.

TABLE V. MODEL PERFORMANCE METRICS FOR DIFFERENT DATA FOLDING

Data Folding	Precision	Recall	Specificity	F1 Score	Accuracy
Two-fold	78.57%	73.33%	57.14%	75.85%	68.18 %
Three-fold	85.71%	80.00%	71.42%	82.75%	77.27%
Four-fold	100%	93.33%	100%	96.54%	95.45%

Testing of the proposed model includes a general classifier technique in machine learning [31]: K-Nearest Neighbor (KNN), Discriminant Analysis (DA), Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF). For each classifier and proposed method, total twenty-one experiments were conducted including two, three, and four data folding. Comparing the system accuracy

for different classifiers is an essential step in selecting the most suitable classifier for a specific task. The accuracy metric provides an overall measure of how well the classifier performs in terms of correctly classifying instances. The Table VI shows that the proposed model is more accurate than other classification techniques, due to the effectiveness of the texture analysis method, so the system could be able to distinguish between normal and abnormal.

TABLE VI. FOUR-FOLD OF DATA COMPARISON OF SYSTEM ACCURACY FOR DIFFERENT CLASSIFIERS

Classifiers	Accuracy		
	Two-fold	Three-fold	Four-fold
KNN	58%	59%	61 %
DA	62%	64%	73 %
NB	60%	64%	66%
DT	67%	67%	68%
SVM	74%	75%	75%
RF	68%	68%	68%
Proposed method	68.18 %	77.27%	95.45%

The Table VII indicates that the proposed method for texture analysis performed better than other classification techniques, allowing the system to distinguish normal from abnormal behavior based on the performance evolution results.

TABLE VII. FOUR-FOLD COMPARISON OF PERFORMANCE EVALUATIONS FOR DIFFERENT CLASSIFIERS

Classifiers	Precision	Recall	Specificity	F1 Score	AU-ROC value
KNN	60 %	72 %	57%	65%	0.6128
DA	65%	72 %	60 %	68%	0.6789
NB	54%	60 %	67%	66%	0.6534
DT	60 %	60 %	60 %	60 %	0.5934
SVM	94 %	91%	93%	94%	0.9489
RF	92 %	92%	93 %	94 %	0.9393
Proposed method	100%	93.33%	95.45%	96.54%	0.9659

The An AU-ROC (Area Under the Receiver Operating Characteristic) plot is a graphical representation used to compare the performance of different classifiers in classification tasks. When comparing different classifiers using AU-ROC plots, the classifier with the highest AU-ROC value is generally considered as best performance. AU-ROC plot of the proposed system with two classification techniques is revealed in Fig. 6, in which the proposed occupies a larger area under the curve than the other two.

It is critical to compare the performance of different methods for determining osteoporosis in the femur bone when analyzing image data. Various methods or algorithms are employed to solve a particular problem, and comparing their performance helps in selecting the most effective one. As an illustration of the system performance for the different methods is as revealed in Table VIII, nevertheless proposed method performs better when data is folded four times.

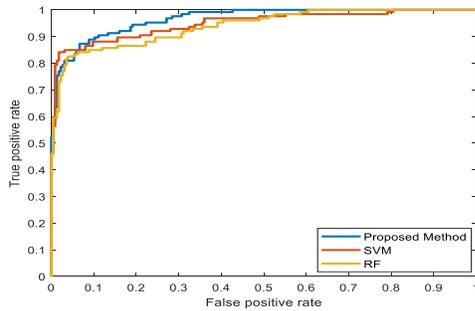


Fig. 6. An AU-ROC plots.

TABLE VIII. PERFORMANCE COMPARISON WITH DIFFERENT METHODS

Ref. No.	Specificity	F1 Score	Accuracy	Sensitivity	AU-ROC
[4]	88.24%	89.43%	88.50%	88.19%	89.01%
[15]	93.70%	93.39%	93.39%	93.90%	92.37%
[17]	84.20%	85.39%	91.39%	92.80%	86.50%
[22]	94.19%	90.28%	91.27%	80.29%	93.43%
Proposed Method	95.45%	96.54%	95.45%	93.33%	96.59%

VI. CONCLUSION

The spectral based analysis of texture features of ROI femur X-ray images produces very good results because the 2D-DWT gives pixel intensity variation at different scale. Therefore, the texture features are helpful in obtaining the significant texture using MRFs. The various scales in image texture benefit from the MRFs, contributing to its effectiveness and the capability to process fine and coarse information efficiently. By calculating the two attributes as ZNCC and SSD of the test and supervised images in the matching process helps to classify the test image. The best matching between the test and the supervised database is considered as the maximum correlation and minimum sum difference which is exactly the test predicted classes. The proposed work achieves better osteoporosis detection with less error.

In future the evaluation can be made in spatial domain so that the system can be in contrast with proposed one however requiring extra database to boost the system performance. The femur X-ray images must be derived from a customary public dataset to make this work useful in medical disciplines for early detection of osteoporosis. A generalized X-ray image dataset for the femur is not available to make the proposed work more significant for detecting osteoporosis.

REFERENCES

[1] LeBoff, M. S., S. L. Greenspan, K. L. Insogna, E. M. Lewiecki, K. G. Saag, A. J. Singer, and E. S. Siris. "The clinician's guide to prevention and treatment of osteoporosis." *Osteoporosis international* 33, no. 10 (2022): 2049-2102.

[2] Shen, Y., X. Huang, and J. Wu. "The global burden of osteoporosis, low bone mass, and its related fracture in 204 countries and territories, 1990–2019." *Front Endocrinol (Lausanne)* 13: 882241. (2022).

[3] Tarantino, Umberto, Giovanni Iolascon, Luisella Cianferotti, Laura Masi, Gemma Marcucci, Francesca Giusti, Francesca Marini et al. "Clinical guidelines for the prevention and treatment of osteoporosis:

summary statements and recommendations from the Italian Society for Orthopaedics and Traumatology." *Journal of orthopaedics and traumatology* 18, no. 1 (2017): 3-36.

[4] Yamamoto, N., S. Sukegawa, A. Kitamura, R. Goto, T. Noda, K. Nakano, K. Takabatake et al. "Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates." *Biomolecules*. 2020; 10 (11): 1534.

[5] Whitmarsh, Tristan, Ludovic Humbert, Mathieu De Craene, Luis M. Del Rio Barquero, and Alejandro F. Frangi. "Reconstructing the 3D shape and bone mineral density distribution of the proximal femur from dual-energy X-ray absorptiometry." *IEEE transactions on medical imaging* 30, no. 12 (2011): 2101-2114.

[6] Vijay, A., N. Shankar, C. Aroba Sahaya Liges, and M. Anburajan. "Evaluation of osteoporosis using CT image of proximal femur compared with dual energy X-ray absorptiometry (DXA) as the standard." In 2011 3rd International Conference on Electronics Computer Technology, vol. 3, pp. 334-338. IEEE, 2011.

[7] Pramudito, J. T., S. Soegijoko, T. R. Mengko, F. I. Muchtadi, and R. G. Wachjudi. "Trabecular pattern analysis of proximal femur radiographs for osteoporosis detection." *Journal of Biomedical & Pharmaceutical Engineering* 1, no. 1 (2007): 45-51.

[8] Supaporn, Kiattisin, and Chamnongthai Kosin. "Femur bone volumetric estimation from a single X-ray image for osteoporosis diagnosis." In 2006 International Symposium on Communications and Information Technologies, pp. 1149-1152. IEEE, 2006.

[9] Shankar, N., V. Sathagirivasan, A. Vijay, K. Kirthika, and M. Anburajan. "Evaluation of osteoporosis using radiographic hip geometry, compared with dual energy X-ray absorptiometry (DXA) as the standard." In 2010 International Conference on Systems in Medicine and Biology, pp. 259-264. IEEE, 2010.

[10] Sathagirivasan V., and M. Anburajan. "Analysis of trabecular proximal femur bone in diagnosing osteoporosis using digital x-ray: A comparison with DXA." In 2011 International Conference on Recent Trends in Information Technology (ICRITIT), pp. 869-874. IEEE, 2011.

[11] Mengko, Tati Rajab, and J. Tjandra Pramudito. "Implementation of Gabor filter to texture analysis of radiographs in the assessment of osteoporosis." In Asia-Pacific Conference on Circuits and Systems, vol. 2, pp. 251-254. IEEE, 2002.

[12] Reshmalakshmi, Chandrasekharan, and M. Sasikumar. "Fuzzy inference system for osteoporosis detection." In 2016 IEEE Global Humanitarian Technology Conference (GHTC), pp. 675-681. IEEE, 2016.

[13] Bobby, T. Christy. "Estimation of femur morphometric features for CBIR application." In 2017 Third International Conference on Biosignals, Images and Instrumentation (ICBSII), pp. 1-5. IEEE, 2017.

[14] Sahiti Lahari, M., and Anburajan M. Vijay. "Finite Element Analysis of Femur in the Evaluation of Osteoporosis." *Electronics Computer Technology (ICECT)*. In 2011 3rd International Conference on. IEEE Conference Proceedings, vol. 3, pp. 415-419. 2011.

[15] Shankar, N., S. Sathish Babu, and C. Viswanathan. "Femur bone volumetric estimation for osteoporosis classification using optimization-based deep belief network in X-ray images." *The Computer Journal* 62, no. 11 (2019): 1656-1670.

[16] Raghavendra Chinchansoor, Dr. Subhangi DC. "Machine Learning Technique for Osteoporosis Caused Bone Fracture Detection in Femur Bones from 2D X-Ray Images". © 2018 JETIR December 2018, Volume 5, Issue 12

[17] Nguyen, B.N.T.; Hoshino, H.; Togawa, D.; Matsuyama, Y. Cortical thickness index of the proximal femur: A radiographic parameter for preliminary assessment of bone mineral density and osteoporosis status in the age 50 years and over population. *CiOs Clin. Orthop. Surg.* 2018, 10, 149–156. [CrossRef] [PubMed]

[18] C. C. Yang, M. B. Nagarajan, M. B. Huber, J. C. Gamio, J. S. Bauer et al., "Improving bone strength prediction in human proximal femur specimens through geometrical characterization of trabecular bone microarchitecture and support vector regression," *Journal of Electronic Imaging*, vol. 23, no. 1, pp. 13013, 2014.

[19] Adams, M.; Chen, W.; Holdorf, D.; McCusker, M.W.; Howe, P.D.L.; Gaillard, F. Computer vs. human: Deep learning versus perceptual

- training for the detection of neck of femur fractures. *J. Med. Imaging Radiat. Oncol.* 2019, 63, 27–32. [CrossRef]
- [20] G. Chang, A. H. Cho, H. Rusinek, S. Honig, A. Mikheev et al., "Measurement reproducibility of magnetic resonance imaging-based finite element analysis of proximal femur microarchitecture for in vivo assessment of bone strength," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 28, no. 4, pp. 407–412, 2015.
- [21] Yoon, Byung-Ho, Jang-Won Park, Chan Woo Lee, and Young Do Koh. "Different Pattern of T-Score Discordance between Patients with Atypical Femoral Fracture and Femur Neck Fracture." *Journal of Bone Metabolism* 30, no. 1 (2023): 87.
- [22] Hsieh, C. I., K. Zheng, C. Lin, L. Mei, L. Lu, W. Li, F. P. Chen et al. "Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. *Nat Commun* 12: 5472." (2021).
- [23] Awal, Rabina, and Tanvir R. Faisal. "Multiple regression analysis of hip fracture risk assessment via finite element analysis." *Journal of Engineering and Science in Medical Diagnostics and Therapy* 4, no. 1 (2021): 011006.
- [24] Goel, Akash. "Discrete wavelet transform (DWT) with two channel filter bank and decoding in image texture analysis." *Int. J. Sci. Res* 3, no. 4 (2014): 391-397.
- [25] Fortuna-Cervantes, Juan Manuel, Marco Tulio Ramírez-Torres, Marcela Mejía-Carlos, José Salomé Murguía, José Martínez-Carranza, Carlos Soubervielle-Montalvo, and César Arturo Guerra-García. "Texture and Materials Image Classification Based on Wavelet Pooling Layer in CNN." *Applied Sciences* 12, no. 7 (2022): 3592.
- [26] Litimco, C.E.O., Villanueva, M.G.A., Yecla, N.G., Soriano, M.N. and Naval, P.C., 2013, March. Coral Identification Information System. In *IEEE International Underwater Technology Symposium (UT)* (pp. 1-6). IEEE 2013.
- [27] Soni, Pramod Kumar, Navin Rajpal, and Rajesh Mehta. "Road centerline extraction from VHR images using SVM and multi-scale maximum response filter." *Journal of the Indian Society of Remote Sensing* 49 (2021): 1519-1532.
- [28] Angeles, Maria Del Pilar, and Carlos G. Ortiz-Monreal. An attribute-based classification by threshold to enhance the data matching process. *Journal of applied research and technology* 17, no. 4 (2019): 272-284. .
- [29] quality sea-ice surface reconstruction from aerial images." *Remote Sensing* 11, no. 9 (2019): 1055. Vujović, Ž. "Classification model evaluation metrics." *International Journal of Advanced Computer Science and Applications* 12, no. 6 (2021): 599-606.
- [30] Sarker, Iqbal H. "Machine learning: Algorithms, real-world applications and research directions." *SN computer science* 2, no. 3 (2021): 160.
- [31] Sankara Subbu, Ramesh. "Brief Study of Classification Algorithms in Machine Learning." (2017).