Diagnostic and Interventional Imaging (2020) **101**, 599—610

ORIGINAL ARTICLE / *Musculoskeletal imaging*

Radiomics for classification of bone mineral loss: A machine learning study

S. Rastegar[a,b], M. Vaziri[c,d], Y. Qasempour[c,d], M.R. Akhash[c,d], N. Abdalvand[i], I. Shiri[e], H. Abdollahi[c,d,*], H. Zaidi[e,f,g,h]

[a] *Student Research Committee, School of Paramedical Sciences, Rafsanjan University of Medical Sciences, Rafsanjan, Iran*
[b] *Department of Radiology Technology, School of Paramedical Sciences, Rafsanjan University of Medical Sciences, Rafsanjan, Iran*
[c] *Student Research Committee, Faculty of Allied Medicine, Kerman University of Medical Sciences, Kerman, Iran*
[d] *Department of Radiologic Sciences, Medical Physics, Faculty of Allied Medicine, Kerman University of Medical Sciences, Kerman, Iran*
[e] *Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva 4, Switzerland*
[f] *Geneva University Neurocenter, Geneva University, CH-1205 Geneva, Switzerland*
[g] *Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, 9713 GZ Groningen, The Netherlands*
[h] *Department of Nuclear Medicine, University of Southern Denmark, 5230 Odense, Denmark*
[i] *Department of Medical Physics, School of Medicine, Iran University of Medical Sciences, Tehran, Iran*

KEYWORDS
Bone mineral densitometry (BMD);
Radiomics;
Machine learning;
Osteoporosis;
Classification

Abstract
Purpose: The purpose of this study was to develop predictive models to classify osteoporosis, osteopenia and normal patients using radiomics and machine learning approaches.
Materials and methods: A total of 147 patients were included in this retrospective single-center study. There were 12 men and 135 women with a mean age of 56.88 ± 10.6 (SD) years (range: 28—87 years). For each patient, seven regions including four lumbar and three femoral including trochanteric, intertrochanteric and neck were segmented on bone mineral densitometry images and 54 texture features were extracted from the regions. The performance of four feature selection methods, including classifier attribute evaluation (CLAE), one rule attribute

Abbreviations: AUC, Area under the curve; BMD, Bone mineral densitometry; CLAE, Classifier attribute evaluation; GRAE, Gain ratio attribute evaluation; KN, K-nearest neighbor; LB, Logit-boost; ML, Machine learning; ORAE, One rule attribute evaluation; PRCA, Principal component analysis; RC, Random committee; RF, Random forest; ROC, Receiver operator characteristics.

* Corresponding author. Haftbagh Alavi Highway, Kerman University of Medical Sciences Campus, Department of Radiologic Sciences and Medical Physics, Faculty of Allied Medicine, Kerman University of Medical Sciences, Kerman, Iran.
E-mail address: hamid_rbp@yahoo.com (H. Abdollahi).

https://doi.org/10.1016/j.diii.2020.01.008

evaluation (ORAE), gain ratio attribute evaluation (GRAE) and principal components analysis (PRCA) along with four classification methods, including random forest (RF), random committee (RC), K-nearest neighbor (KN) and logit-boost (LB) were evaluated. Four classification categories, including osteopenia vs. normal, osteoporosis vs. normal, osteopenia vs. osteoporosis and osteoporosis + osteopenia vs. osteoporosis were examined for the defined seven regions. The classification model performances were evaluated using the area under the receiver operator characteristic curve (AUC).

Results: The AUC values ranged from 0.50 to 0.78. The combination of methods RF + CLAE, RF + ORAE and RC + ORAE yielded highest performance (AUC = 0.78) in discriminating between osteoporosis and normal state in the trochanteric region. The combinations of RF + PRCA and LB + PRCA had the highest performance (AUC = 0.76) in discriminating between osteoporosis and normal state in the neck region.

Conclusion: The machine learning radiomic approach can be considered as a new method for bone mineral deficiency disease classification using bone mineral densitometry image features.

Introduction

Bone mineral loss, in terms of osteoporosis and osteopenia especially in the vertebral and hip regions is associated with several disorders, mortality and morbidity [1,2]. Bone mineral density (BMD), measured by dual energy x-ray absorptiometry (DEXA), is considered as the gold standard and is a well-established approach for osteoporosis diagnosis and prognosis [3]. BMD images are usually obtained in the femur and in the lumbar spine from anterior-posterior projections and the resulting images are converted into several parameters, including 'T-scores' and 'Z-scores' by the BMD system software. These metrics constitute the basis for interpretation of the BMD results [4].

As a new advanced image processing approach, radiomics, has been studied for biomarker discovery through extraction of high throughput features from medical images [5,6]. In the radiomics approach, imaging features are extracted from the segmented regions and after selection are considered for model development. Recently, a comprehensive data compilation reporting a wealth of information on the use of radiomic features for cancer detection, diagnosis; prognosis and treatment response prediction has been reported [7—9].

Radiomics is commonly used in the context of clinical oncology, yet the number of studies using this approach focusing on bone diseases is limited. However, a considerable body of literature has been published on the use of textural features extracted from radiography (digital or conventional), computed tomography (CT) and magnetic resonance imaging (MRI) for osteoporosis detection, diagnosis, assessment and automatic bone disorder classification [10—16]. Yet, reports on the use of BMD radiomics for bone disease diagnosis and prognosis is still lacking.

Thanks to the availability of big datasets and computational power, a number of artificial intelligent (AI)-based solutions have been developed during the last decade and successfully implemented for medical imaging applications [17,18]. Machine learning (ML) is a subset of AI, providing a promising tool for the development of diagnostic, prognostic and predictive modelling tools from multimodality medical image data [19,20]. The combination of image-derived features and ML algorithms are used to build more accurate models in the era of precision medicine [21]. It is hypothesized that conventional diagnostic methods of many diseases could be replaced by ML radiomics in the future. A number of studies have indicated that ML tools are promising for predicting fragility fracture risk in people at risk of osteoporosis [22—24].

The purpose of this study was to develop predictive models to classify osteoporosis, osteopenia and normal patients using radiomics and machine learning (ML) approaches.

Material and methods

Human subjects, BMD imaging and reporting

All procedures performed in this study using human participants were in accordance with the ethical standards of the institution and the 1964 Helsinki declaration and its later amendments. Due to the retrospective nature of the study, the need for informed consent was waived. In this single-center analysis, data of patients referred for bone mineral densitometry from January 2018 to April 2019 were initially analyzed for further inclusion. The inclusion criteria were the availability of a high quality BMD and its report. The exclusion criteria were:

- history of hematological disorder;
- history of metabolic disorders;
- documented diabetes;
- bone metastases;
- poor quality of BMD.

The study population included 147 patients with a mean age of 56.88 ± 10.60 (standard deviation [SD]) years (range:

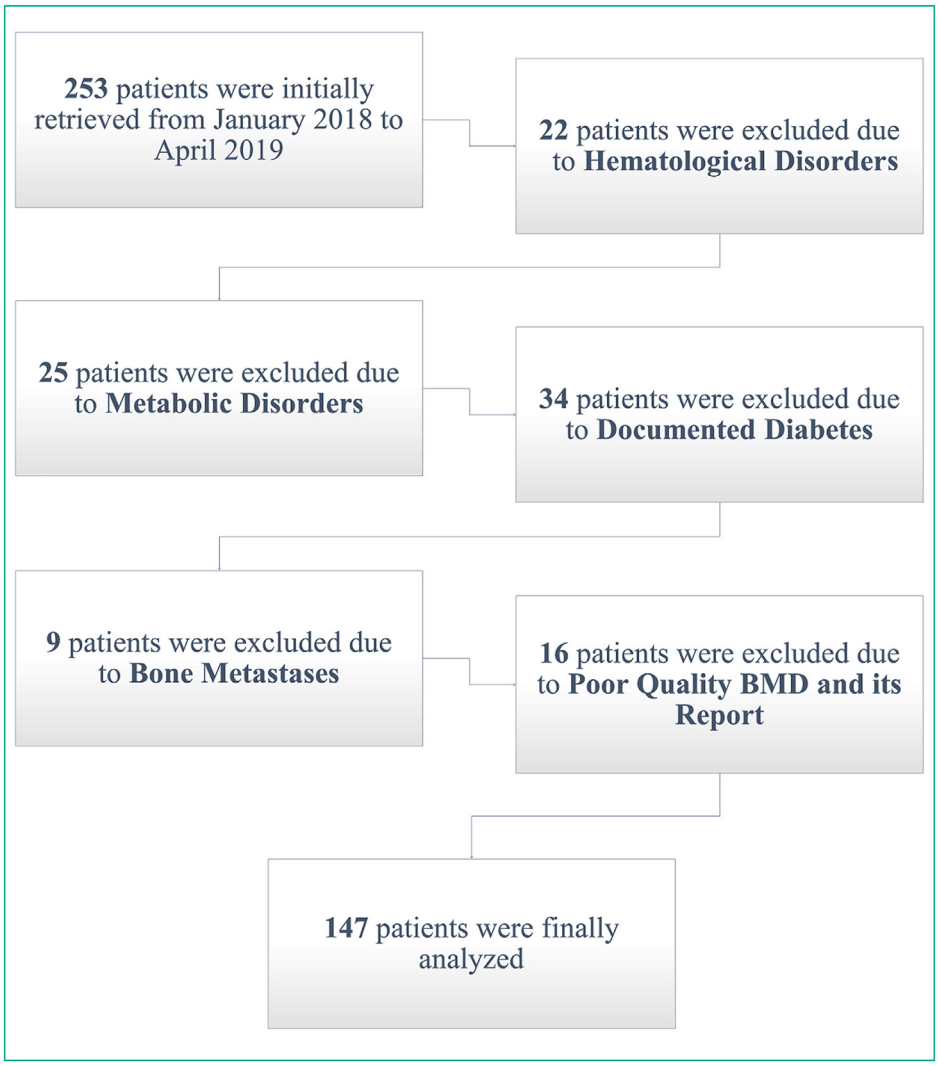

Figure 1. Flowchart shows patient selection process. BMD indicates bone mineral density.

28–87 years). There were 12 men with a mean age of 67.9 ± 10.3 (SD) years (range: 55–80 years) and 135 women with a mean age of 55.9 ± 7.43 (SD) years (range: 28–67 years). Fig. 1 portrays the patient selection flowchart. Fig. 2 illustrates the different steps of the methodology followed in this work.

The BMD images and reports of the included patients were analyzed and archived for further assessment. All BMD examinations were performed on Lexxos DEXA system (DMS, Digital 2D Densitometer) with 4.7 mm Al filter, 75 kVp and 7 mAs. Prior to the examinations, the standard quality assurance procedures, including consistency, calibration and electronic as well as mechanical checks were regularly carried out.

Based on the T-score obtained from the BMD analysis report, the patients were divided into three groups, including: 1) Osteoporosis, 2) Osteopenia and 3) Normal. Our grouping was based on the WHO criteria, which defined osteoporosis as T-score < −2.5, osteopenia as −2.5 < T-score < −1 and normal as T-score > −1.

Feature extraction

For all patients, BMD images were exported from the system in DICOM format and were inputted to the texture analysis software for feature extraction. To extract texture features, seven related regions of interest (ROIs) including four lumbar (L1–L4), and three femoral including trochanteric, intertrochanteric and neck were segmented on the BMD image for each patient (Fig. 3). All ROIs were drawn manually by a ten years experienced radiologist (A. A), with expertise in bone disease analysis and BMD reporting. The size of each ROI was different depending on the region where the ROI was drawn, with a mean size of 2 ± 0.77 (SD) cm (range: 1–3 cm). The size of one pixel was 1 mm^2.

A total of 54 texture features were extracted from each ROI. These were 12 gray level co-occurrence matrix (GLCM) features, 20 gray level run length matrix (GLRLM) features, 8 intensity histogram features, 5 auto-regressive model (ARM) features, 5 gradient features and 4 wavelet features. More details on the features are shown in Table 1.

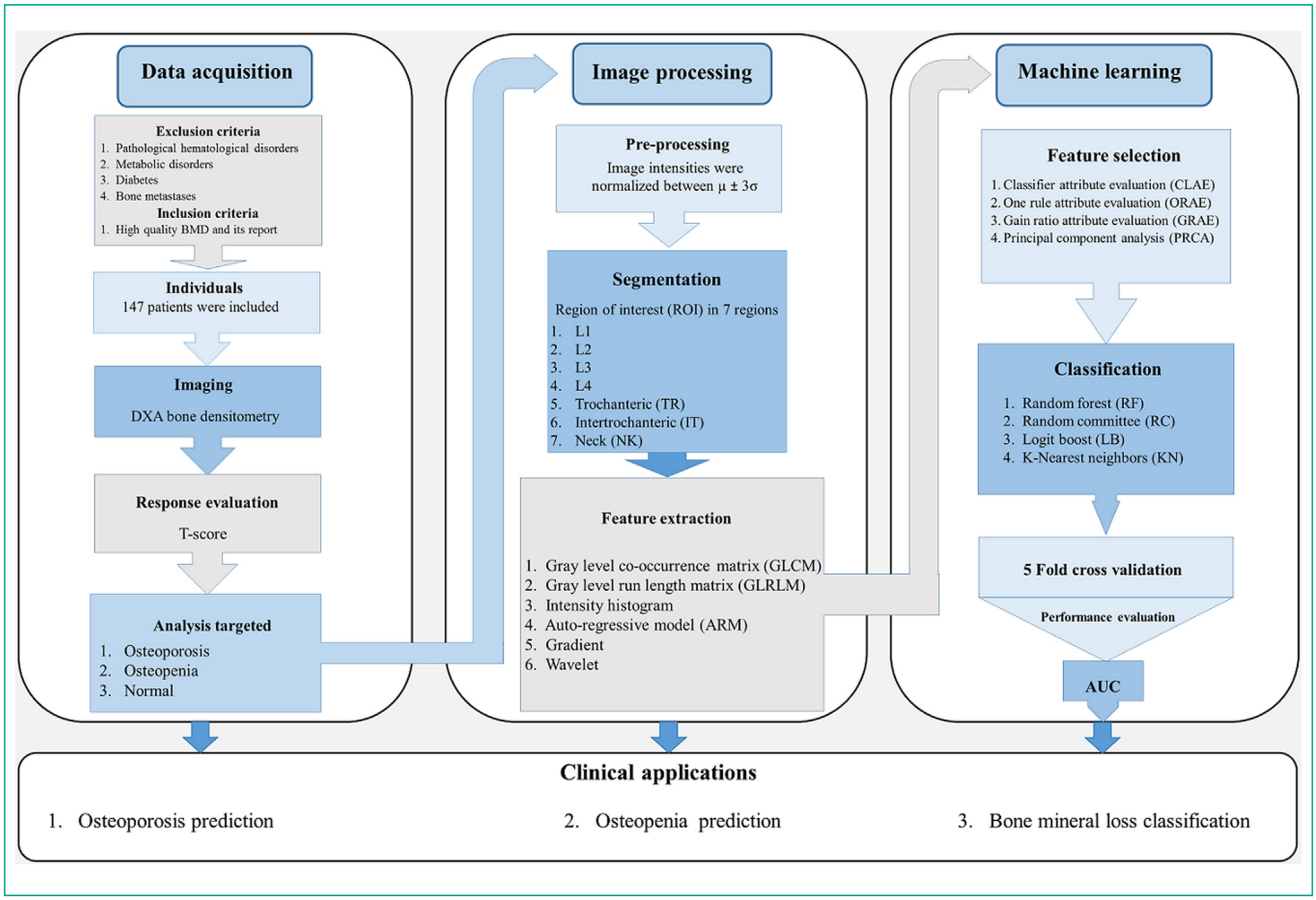

Figure 2. Diagram shows structured methodology adopted in this study protocol.

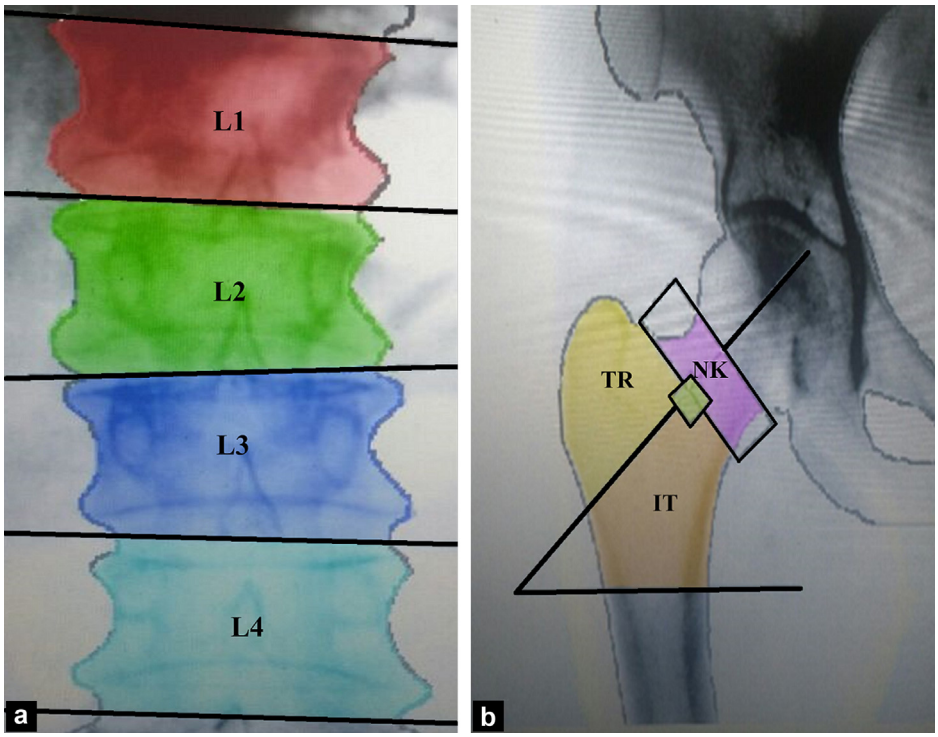

Figure 3. Regions of interest (ROIs) in *a*. Lumbar regions (L1, L2, L3 and L4) and *b*. Femoral regions, including trochanteric (TR), intertrochanteric (IT) and neck (NK) regions.

Table 1 Radiomics texture features.

Feature set	Texture features, definition and abbreviations
Intensity histogram	Mean (histogram's mean)
	Variance (histogram's variance)
	Skewness (histogram's skewness)
	Kurtosis (histogram's kurtosis)
	Perc.01% (1% percentile)
	Perc.10% (10% percentile)
	Perc.50% (50% percentile)
	Perc.90% (90% percentile)
	Perc.99% (99% percentile)
Gradient	GrMean (absolute gradient mean)
	GrVariance (absolute gradient variance)
	GrSkewness (absolute gradient skewness)
	GrKurtosis (absolute gradient kurtosis)
	GrNonZeros (percentage of pixels with nonzero gradient)
Gray level run-length matrix (GLRLM)	RLNonUni (run length non-uniformity)
	GLevNonU (grey level non-uniformity)
	LngREmph (long run emphasis)
	ShrtREmp (short run emphasis)
	Fraction (fraction of image in runs)
Gray level co-occurrence matrix (GLCM)	AngScMom (angular second moment)
	Contrast (contrast)
	Correlat (correlation)
	SumOfSqs (sum of squares)
	InvDfMom (inverse difference moment)
	SumAverg (sum average)
	SumVarnc (sum variance)
	SumEntrp (sum entropy); Entropy (entropy)
	DifVarnc (difference variance)
	DifEntrp (difference entropy)
	Features are computed for 5-pixel distance (1, 2, 3, 4, 5) and for 4 various directions (horizontal, 45 degrees, vertical, 135 degrees)
Autoregressive model	Teta1 (parameter $\theta 1$)
	Teta2 (parameter $\theta 2$)
	Teta3 (parameter $\theta 3$)
	Teta4 (parameter $\theta 4$)
	Sigma (parameter σ)
Wavelet transform	WavEn (wavelet energy) feature is computed at 5 scales within four frequency bands
	Low-pass filtering in both directions (LL) assessed the lowest frequencies
	Low-pass filtering followed by high-pass filtering (LH) assessed horizontal edges
	High-pass filtering followed by low-pass filtering (HL) assessed vertical edges
	High-pass filtering in both directions (HH) assessed diagonal details

Pre-processing

Prior to feature extraction, data pre-processing was carried out to reduce noise, increase sensitivity and normalize the intensities across all the patients. To this end, all image intensities were normalized between $\mu \pm 3\sigma$, where μ is the mean value of gray-levels inside the ROIs and σ is the standard deviation.

Classification and feature selection

Overall, four machine learning methods were used in the classification analysis. They included random forest (RF), random committee (RC), logit boost (LB), and K-nearest neighbors (KNN). For feature selection, four methods including classifier attribute evaluation (CLAE), one rule attribute evaluation (ORAE), gain ratio attribute evaluation (GRAE) and principal components analysis (PRCA) were applied. The acronyms related to the feature selection methods are defined in Table 2.

Statistical analysis

To build classification (prediction) models, the patients were labeled into four label categories:

- osteopenia vs. normal;
- osteoporosis vs. normal;
- osteopenia vs. osteoporosis;

Table 2 Machine learning algorithms for feature selection and classification.

Machine learning technique	Abbreviation	Application/Description [Based on the WEKA software]
Random forest	RF	Classification. Class for constructing a forest of random trees
Random committee	RC	Classification. Class for building an ensemble of randomizable base classifiers
K-nearest neighbor	KN	Classification. K-nearest neighbor's classifier
Logit-boost	LB	Classification. Class for performing additive logistic regression
Classifier attribute evaluation	CLAE	Feature selection. Evaluates the worth of an attribute by using a user-specified classifier.
One rule attribute evaluation	ORAE	Feature selection. Evaluates the worth of an attribute by using the OneR classifier
Gain ratio attribute evaluation	GRAE	Feature selection. Evaluates the worth of an attribute by measuring the gain ratio with respect to the class
Principal component's analysis	PRCA	Feature selection. Performs a principal components analysis and transformation of the data

- osteopenia/osteoporosis vs. normal.

In this work, we applied the cross-combination performance of the feature selection and classification methods on the features extracted from each ROI (7 ROIs) and compared them. The classifiers were trained using the 10-fold cross-validation method and their performance evaluated using the area under the receiver operator characteristic (ROC) curve (AUC).

For univariate analysis, the top ranked features were selected and compared between four label categories. Paired Student *t*-test was used to search for statistically significant differences between these features. The statistical significance level was set at $P < 0.05$. All statistical analyses were performed using GraphPad Prism 8.0 (GraphPad Software, Inc.).

Results

Patient characteristics

The mean height and weight of male patients was 164 ± 10.51 (SD) cm (range: 153–190 cm) and 72 ± 12.74 (SD) kg (range: 50–95 kg). The mean height and weight of female patients was 154.45 ± 9.58 (SD) cm (range: 141–168 cm) and 70.91 ± 14.08 (SD) kg (range: 40–158 kg). Based on the T-score results, 7 men (58%) developed osteoporosis, 2 men (17%) developed osteopenia and 3 men (25%) were normal. For women, 20 subjects (15%) developed osteoporosis, 58 subjects (43%) developed osteopenia and 57 subjects (42%) were normal.

Radiomic features

All radiomic feature values, including mean, SD and range based on the feature set are given in Table 3. As can be seen, there were large differences in the range of radiomic features.

Classification

We used the AUC to quantify the classification performance of different crossed feature selection and classification methods. In the present study, we examined 16 combinations of feature selection and classification methods for each region, resulting in 112 combinations for each label and a total of 448 combinations ($16 \times 7 \times 4$).

Tables 4–7 report the AUC values of all the crossed methods. In brief, we observed that the AUC values for all combinations ranged from 0.50 to 0.78.

Osteopenia vs. normal classification

The results of this classification are shown in Table 4. For L1 region, we observed that the combination methods LB + GRAE and LB + CLAE (AUC = 0.68) had the highest classification performance followed by LB + ORAE (AUC = 0.67) and LB + PRCA (AUC= 0.66). For L2 region, LB + ORAE (AUC = 0.57) had the highest performance and the other combinations had an AUC ranging from 0.50 to 0.55. For L3 region, the highest AUC was 0.53 (combinations of KN + GRAE/CLAE/ORAE). For L4 region, the highest AUC was 0.59 (combinations of LB+ PRCA/ORAE). For trochanteric region, we observed that the combination methods RF + PRCA (AUC = 0.66) had the highest classification performance, followed by RF + PRCA and KN + PRCA (AUC = 0.64) and RF + ORAE/CLAE and RC + GRAE (AUC, 0.63). For intertrochanteric region, the combinations RF + ORAE/GRAE/PRCA and KN + PRCA with an AUC of 0.61 had the highest performance. For the neck region, RF + PRCA with an AUC of 0.57 had the highest performance.

Osteoporosis vs. normal classification

The results of this classification are shown in Table 5. For L1 region, we observed that the combination methods RF + PRCA and RC + PRCA had the highest classification performance (AUC = 0.61) followed by KN + ORAE/CLAE/GRAE (AUC = 0.59). For L2 region, RC + PRCA had the highest performance (AUC = 0.66) followed by RC + GRAE (AUC = 0.64), LB + PRCA (AUC = 0.63) and RC + PRCA (AUC = 0.60). For L3 region, all combinations had an AUC of 0.50. For L4 region, the highest AUC was 0.59 (combinations of RF/RC+ PRCA). For trochanteric region, we observed that the combination methods RF + ORAE/CLAE and RC + ORAE had the highest classification performance (AUC = 0.78) followed by RF + PRCA and KN + PRCA (AUC, 0.64) and RF + ORAE/CLAE

Table 3 Radiomic feature values (Mean, standard deviation (SD) and range).

GLCM	Mean ± SD [range]	GLRLM	Mean ± SD [range]	Intensity histogram	Mean ± SD [range]
AngScMom	1.29 ± 5.62 [0.0067–46.67]	135dr_Fraction	12.67 ± 53.75 [0.54–282]	Mean	140.47 ± 48.43 [75.97–508]
Contrast	80.67 ± 102.77 [0.13–555.48]	135dr_GLevNonU	20.43 ± 6.36 [2.08–39.21]	Perc.01	61.54 ± 15.22 [30–103]
Correlat	1.12 ± 3.84 [0.044–26.98]	135dr_LngREmph	2.68 ± 2.47 [0.77–20.86]	Perc.10	84.94 ± 24.16 [30–136]
DifEntrp	20.53 ± 93.27 [0.78–492.06]	135dr_RLNonUni	145.91 ± 48.27 [13.50–223.41]	Perc.50	145.18 ± 26.64 [78–202]
DifVarnc	47.39 ± 61.59 [0.82–316.72]	135dr_ShrtREmp	1.18 ± 2.58 [0.61–20.55]	Perc.90	155.66 ± 27.14 [88–210]
Entropy	2.63 ± 3.05 [1.41–24.64]	45dgr_Fraction	7.21 ± 27.86 [0.58–158.91]	Perc.99	185.71 ± 91.08 [92–614]
InvDfMom	2.81 ± 10.79 [0.22–10.79]	45dgr_GLevNonU	20.24 ± 6.08 [2.15–35.33]	Skewness	1.14 ± 0.33 [0.054–1.88]
Kurtosis	3.07 ± 11.87 [0.003–76]	45dgr_LngREmph	2.7 ± 2.56 [0.73–22.66]	Variance	922.98 ± 1034.62 [0.61–5460.18]
SumAverg	67.27 ± 13.33 [19.53–96.50]	45dgr_RLNonUni	142.9 ± 45.52 [10.14–226.83]	**Auto–regressive model**	**Mean ± SD [range]**
SumEntrp	1.54 ± 1.35 [1.078–14.47]	45dgr_ShrtREmp	1.18 ± 2.66 [0.63–46.67]	Sigma	449.12 ± 2296.12 [0.43–1,6308.13]
SumOfSqs	49.43 ± 58.76 [0.32–308.50]	Horzl_Fraction	7.41 ± 29.17 [0.53–174.78]	Teta 1	0.59 ± 2.64 [0.018–21.57]
SumVarnc	116.46 ± 132.49 [1.10–725.63]	Horzl_GLevNonU	18.99 ± 5.83 [2.06–34.11]	Teta 2	0.53 ± 2.66 [0.007–22.25]
Gradient	**Mean ± SD [range]**	Horzl_LngREmph	2.84 ± 2.86 [0.77–26.68]	Teta 3	0.52 ± 2.67 [0.0002–22.92]
GrKurtosis	1.11 ± 2.62 [0.024–19.56]	Horzl_RLNonUni	146.32 ± 51.75 [3.41–229.43]	Teta 4	0.88 ± 2.63 [0.0064–23.59]
GrMean	2.51 ± 2.61 [0.91–19.60]	Horzl_ShrtREmp	1.19 ± 2.93 [0.56–26.27]	**Wavelet**	**Mean ± SD [range]**
GrNonZeros	12.23 ± 52.06 [0.40–271]	Vertl_Fraction	6.91 ± 26.89 [0.47–164.95]	WavEnHH	552.69 ± 2431.88 [9.40–1,7523.72]
GrSkewness	1.28 ± 2.55 [0.17–18.98]	Vertl_GLevNonU	18.49 ± 5.64 [2.073–32.84]	WavEnHL	85.85 ± 102.7 [9.81–586.02]
GrVariance	6.71 ± 7.88 [0.63–39.89]	Vertl_LngREmph	2.99 ± 2.72 [0.77–24.67]	WavEnLH	67 ± 81.43 [9.02–444.53]
		Vertl_RLNonUni	142.92 ± 51.67 [6.77–228.71]	WavEnLL	1,6831.09 ± 7386.73 [10.61–3,2118.36]
		Vertl_ShrtREmp	1.18 ± 2.78 [0.52–24.27]		

SD: standard deviation; AngScMom: Angular second Moment; Correlat: Correlation; DifEntrp: Difference entropy; DifVarnc: Difference variance; Horzl: Horizontal; Vertl: Vertical; InvDf-Mom: Inverse difference momentum; SumAverg: Sum of average; SumEntrp: Sum of entropy; SumOfSqs: Sum of squares; SumVarnc: Sum of variance; Gr: Gradient; 135dr: 135 Degree; 45dr; 45 Degree; GLevNonU: Gray level non-uniformity; LngREmph: Long run emphasis; RLNonUni: Run length non-uniformity; ShrtREmp: Short run emphasis; Perc: Percentile; Wav: Wavelet; En: Energy; H: High; L: Low.

Table 4 Results for classification performance (AUC) of feature selection (in columns) and classification (in rows) methods for L1, L2, L3 and L4 regions, trochanteric, intertrochanteric and neck regions to classify osteopenia from normal patients.

Osteopenia vs. normal				
	L1			
	ORAE	CLAE	GRAE	PRCA
RF	0.64	0.61	0.61	0.62
RC	0.62	0.61	0.61	0.61
LB	0.67	0.68	0.68	0.66
KN	0.60	0.60	0.60	0.60
	L2			
	ORAE	CLAE	GRAE	PRCA
RF	0.53	0.55	0.55	0.50
RC	0.50	0.50	0.54	0.50
LB	0.57	0.55	0.55	0.50
KN	0.50	0.50	0.50	0.50
	L3			
	ORAE	CLAE	GRAE	PRCA
RF	0.50	0.50	0.50	0.50
RC	0.50	0.50	0.50	0.50
LB	0.50	0.50	0.50	0.50
KN	0.53	0.53	0.53	0.50
	L4			
	ORAE	CLAE	GRAE	PRCA
RF	0.50	0.50	0.50	0.50
RC	0.50	0.50	0.50	0.50
LB	0.56	0.50	0.50	0.56
KN	0.52	0.50	0.50	0.52
	Trochanteric			
	ORAE	CLAE	GRAE	PRCA
RF	0.63	0.63	0.62	0.66
RC	0.59	0.57	0.63	0.64
LB	0.60	0.60	0.6	0.57
KN	0.54	0.54	0.54	0.64
	Inter trochanteric			
	ORAE	CLAE	GRAE	PRCA
RF	0.61	0.56	0.61	0.61
RC	0.59	0.58	0.57	0.56
LB	0.54	0.53	0.53	0.55
KN	0.57	0.57	0.57	0.61
	Neck			
	ORAE	CLAE	GRAE	PRCA
RF	0.53	0.53	0.53	0.57
RC	0.53	0.52	0.52	0.6
LB	0.5	0.5	0.5	0.5
KN	0.53	0.55	0.55	0.56

RF: Random forest; RC: Random committee; KN: K-nearest neighbor; LB: Logit-boost; CLAE: Classifier attribute evaluation; ORAE: One rule attribute evaluation; GRAE: Gain ratio attribute evaluation; PRCA: Principal component's analysis.

Table 5 Results for classification performance (AUC) of feature selection (in columns) and classification (in rows) methods for L1, L2, L3 and L4 regions, trochanteric, intertrochanteric and neck regions to classify osteoporosis from normal patients.

Osteoporosis vs; normal				
	L1			
	ORAE	CLAE	GRAE	PRCA
RF	0.50	0.50	0.50	0.61
RC	0.50	0.50	0.50	0.61
LB	0.50	0.50	0.50	0.56
KN	0.59	0.59	0.59	0.55
	L2			
	ORAE	CLAE	GRAE	PRCA
RF	0.52	0.54	0.57	0.6
RC	0.53	0.5	0.64	0.66
LB	0.56	0.51	0.56	0.63
KN	0.53	0.53	0.53	0.5
	L3			
	ORAE	CLAE	GRAE	PRCA
RF	0.50	0.50	0.50	0.50
RC	0.50	0.50	0.50	0.50
LB	0.50	0.50	0.50	0.50
KN	0.50	0.50	0.50	0.50
	L4			
	ORAE	CLAE	GRAE	PRCA
RF	0.55	0.55	0.53	0.59
RC	0.52	0.51	0.53	0.59
LB	0.52	0.52	0.52	0.5
KN	0.56	0.56	0.56	0.51
	Trochanteric			
	ORAE	CLAE	GRAE	PRCA
RF	0.78	0.78	0.76	0.67
RC	0.78	0.77	0.74	0.68
LB	0.68	0.67	0.67	0.62
KN	0.56	0.56	0.56	0.59
	Inter trochanteric			
	ORAE	CLAE	GRAE	PRCA
RF	0.66	0.67	0.67	0.64
RC	0.57	0.61	0.61	0.64
LB	0.51	0.51	0.5	0.61
KN	0.51	0.51	0.51	0.51
	Neck			
	ORAE	CLAE	GRAE	PRCA
RF	0.50	0.50	0.50	0.76
RC	0.50	0.50	0.50	0.69
LB	0.50	0.50	0.50	0.76
KN	0.51	0.51	0.50	0.50

RF: Random forest; RC: Random committee; KN: K-nearest neighbor; LB: Logit-boost; CLAE: Classifier attribute evaluation; ORAE: One rule attribute evaluation; GRAE: Gain ratio attribute evaluation; PRCA: Principal component's analysis.

and RC + CLAE (AUC, 0.77), RF + GRAE (AUC = 0.76) and RC + GRAE (AUC = 0.74). For intertrochanteric region, the combinations of RF + CLAE/GRAE and RF + ORAE with an AUC of 0.67 and 0.66 had the highest performances, respectively. For the neck region, the combination methods RF/RC/LB + PRCA with AUCs of 0.76, 0.69 and 0.76 had the highest performance. Finally, all combination results were compared using box-plot analysis and TR combinations had the highest performance.

Table 6 Results for classification performance (AUC) of feature selection (in columns) and classification (in rows) methods for L1, L2, L3 and L4 regions, trochanteric, intertrochanteric and neck regions to classify osteopenia from osteoporosis patients.

Osteopenia vs. osteoporosis				
	L1			
	ORAE	CLAE	GRAE	PRCA
RF	0.54	0.53	0.53	0.5
RC	0.5	0.51	0.51	0.5
LB	0.5	0.5	0.5	0.57
KN	0.51	0.52	0.52	0.54
	L2			
	ORAE	CLAE	GRAE	PRCA
RF	0.5	0.5	0.5	0.5
RC	0.5	0.5	0.5	0.5
LB	0.5	0.5	0.5	0.5
KN	0.5	0.5	0.5	0.5
	L3			
	ORAE	CLAE	GRAE	PRCA
RF	0.51	0.53	0.53	0.57
RC	0.57	0.51	0.51	0.56
LB	0.52	0.53	0.53	0.57
KN	0.62	0.62	0.62	0.55
	L4			
	ORAE	CLAE	GRAE	PRCA
RF	0.5	0.5	0.5	0.5
RC	0.53	0.53	0.5	0.5
LB	0.5	0.5	0.5	0.52
KN	0.5	0.5	0.5	0.5
	Trochanteric			
	ORAE	CLAE	GRAE	PRCA
RF	0.5	0.5	0.5	0.5
RC	0.5	0.5	0.5	0.5
LB	0.5	0.5	0.5	0.52
KN	0.5	0.5	0.5	0.5
	Inter trochanteric			
	ORAE	CLAE	GRAE	PRCA
RF	0.5	0.54	0.54	0.64
RC	0.52	0.59	0.59	0.61
LB	0.5	0.56	0.56	0.64
KN	0.5	0.6	0.6	0.58
	Neck			
	ORAE	CLAE	GRAE	PRCA
RF	0.5	0.5	0.5	0.64
RC	0.5	0.5	0.5	0.6
LB	0.54	0.54	0.51	0.55
KN	0.58	0.58	0.58	0.57

RF: Random forest; RC: Random committee; KN: K-nearest neighbor; LB: Logit-boost; CLAE: Classifier attribute evaluation; ORAE: One rule attribute evaluation; GRAE: Gain ratio attribute evaluation; PRCA: Principal component's analysis.

Table 7 Results for classification performance (AUC) of feature selection (in columns) and classification (in rows) methods for L1, L2, L3 and L4 regions, trochanteric, intertrochanteric and neck regions to classify Osteoporosis/Osteoporosis from normal patients.

Osteopenia osteoporosis vs. normal				
	L1			
	ORAE	CLAE	GRAE	PRCA
RF	0.5	0.5	0.5	0.51
RC	0.5	0.51	0.51	0.51
LB	0.61	0.59	0.59	0.5
KN	0.55	0.55	0.55	0.52
	L2			
	ORAE	CLAE	GRAE	PRCA
RF	0.57	0.6	0.6	0.52
RC	0.59	0.6	0.6	0.5
LB	0.58	0.58	0.58	0.5
KN	0.5	0.5	0.5	0.5
	L3			
	ORAE	CLAE	GRAE	PRCA
RF	0.5	0.59	0.5	0.5
RC	0.5	0.58	0.5	0.5
LB	0.5	0.59	0.5	0.5
KN	0.5	0.5	0.5	0.5
	L4			
	ORAE	CLAE	GRAE	PRCA
RF	0.5	0.56	0.51	0.53
RC	0.5	0.54	0.56	0.5
LB	0.57	0.57	0.54	0.53
KN	0.51	0.51	0.51	0.51
	Trochanteric			
	ORAE	CLAE	GRAE	PRCA
RF	0.69	0.66	0.65	0.65
RC	0.66	0.68	0.66	0.6
LB	0.65	0.65	0.65	0.63
KN	0.58	0.58	0.58	0.58
	Inter trochanteric			
	ORAE	CLAE	GRAE	PRCA
RF	0.63	0.65	0.66	0.57
RC	0.64	0.65	0.6	0.6
LB	0.62	0.62	0.62	0.54
KN	0.54	0.55	0.54	0.54
	Neck			
	ORAE	CLAE	GRAE	PRCA
RF	0.55	0.55	0.55	0.59
RC	0.56	0.57	0.57	0.61
LB	0.5	0.5	0.5	0.58
KN	0.56	0.57	0.57	0.54

RF: Random forest; RC: Random committee; KN: K-nearest neighbor; LB: Logit-boost; CLAE: Classifier attribute evaluation; ORAE: One rule attribute evaluation; GRAE: Gain ratio attribute evaluation; PRCA: Principal component's analysis.

Osteopenia vs. osteoporosis

The results of this classification are shown in Table 6. For L1 region, the highest AUC was 0.57 (combinations of LB+ PRCA). For L2 region, all combinations had an AUC of 0.50. For L3 region, the highest AUC was 0.62 (combinations of KN + GRAE/CLAE/ORAE). For L4 region, the highest AUC was 0.53 (combination of RC+ ORAE/CLAE). For trochanteric region, we observed that the combination LB + PRCA (AUC = 0.52) had the highest classification performance. For intertrochanteric region, the combination RF/LB + PRCA

with an AUC of 0.64 had the highest performance followed by RC + PRCA (AUC = 0.61) and KN + CLAE/GRAE (AUC = 0.60). For NK region, RF + PRCA with an AUC of 0.64 had the highest performance.

Osteopenia/osteoporosis vs. normal classification

The results of this classification are shown in Table 7. For L1 region, we observed that the combination LB + ORAE had the highest classification performance (AUC = 0.61) whereas for L2 region, RF/RC + CLAE/GRAE had the highest performance (AUC = 0.60). For L3 region, the highest AUC was 0.59 (combination of RF/RB + CLAE) while for L4 region, the highest AUC was 0.57 (combination of LB+ CLAE/ORAE). For TR region, we observed that the combination RF + ORAE had the highest classification performance (AUC = 0.69) followed by RF + CLAE, RC + ORAE and RC + GRAE (AUC = 0.66). For intertrochanteric region, the combination of RF + GRAE with an AUC of 0.66 had the highest performance followed by RF/RC + CLAE with an AUC of 0.65. For the neck region, RC + PRCA with an AUC of 0.61 had the highest performance. Finally, all combination results were compared using box-plot analysis and IT combinations had the highest performance.

Selected features

From each group, common top ranked features were selected and are showed in Fig. 4. Top rank means a feature having the highest correlation with the defined labels (classifications). Correlation were selected for osteopenia vs normal classification, *WavEnLL* (Wavelet Energy Low Low), for osteoporosis vs normal classification, *GrMean* (Gradient Mean), for osteopenia vs osteoporosis, *Vertl_RLNonUni* (Vertical Run Length Non Uniformity) and for osteopenia/osteoporosis vs normal classification. There were no significant differences between these features ($P > 0.05$).

Discussion

Bone mineral deficiency diseases, including osteoporosis and osteopenia are common metabolic disorders in a wide range of population. Previous studies have indicated that fractures due to osteoporosis, responsible for high rates of morbidity and mortality, decrease quality of life and increase healthcare costs [25—27]. In this context, early diagnosis and prediction of such disorders will result in better quality of life and lower costs.

In the present study, we developed classification models using radiomic texture features extracted from BMD images and several machine learning algorithms. We demonstrated that textural features extracted from BMD images could be used as new biomarkers to predict osteoporosis and osteopenia. Although our classification models performances are not high enough, ranging between 0.50 and 0.78, they are feasible, easy to use and cost-effective for bone mass disorders classification.

In comparison to BMD obtained by DEXA, radiomics brings additional and complementary information. Radiomics evaluate bone texture features, which are measures of bone heterogeneity whilst BMD is a measure of bone strength. It is a ratio of bone mass to the cross-sectional area of bone. In addition, as stated by Bolotin et al. *''the DXA-derived BMD value does not correctly represent the areal density*

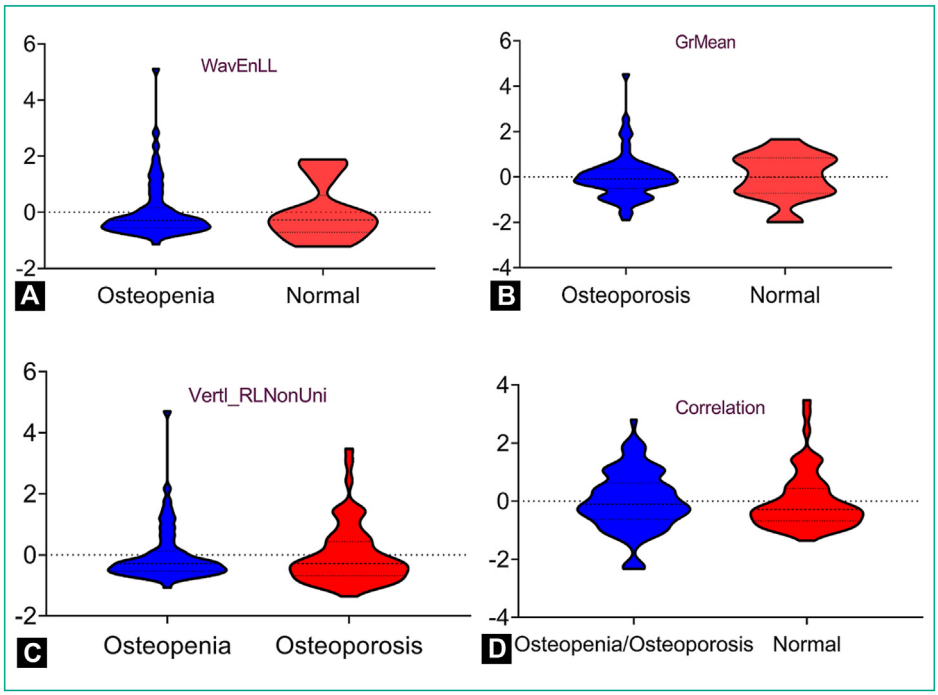

Figure 4. Top ranked features for classification of (A) osteopenia from normal patients, (B) osteoporosis from normal patients, (C) osteoporosis from osteopenia and (D) osteopenia/osteoporosis from normal patients. The abscissa displays the normalized value of the radiomic features while the ordinate shows the classification. WavEnLL = Wavelet energy low low; GrMean = Gradient of mean; Vertl_RLNonUni = Vertical run length non-uniformity.

of bone mineral material, as it is contaminated by sizable, unavoidable, inextricable, independent soft tissue contributions'' [28]. On the other hand, radiomic texture features extracted from medical bone images contain meaningful information, which could be used to predict bone mineral disorders more accurately [29]. In addition, previous studies indicated that the combination of texture analysis and bone mineral density improves the prediction of fracture load in human femurs [30].

In this work, several machine learning approaches were examined from different perspectives. Previous radiomic studies have revealed that it is necessary to evaluate and compare different feature selection and classification methods for the successful realization of radiomics-based prognostic/predictive analyses. In our previous studies, we investigated 84 cross-combinations of feature selection (7 algorithms) and classification methods (12 algorithms) for radiomics-based prediction of radiotherapy induced rectal toxicity in prostate cancer patients [31]. Parmer et al. applied fourteen-feature selection and twelve classification methods in terms of performance and stability for predicting overall survival in patients with lung cancer [32]. Zhang et al. studied 54 cross-combinations of six-feature selection and nine classification methods for radiomics-based survival prediction of patients with advanced nasopharyngeal carcinoma [33]. Hawkins et al. compared four different feature selection and classification methods for CT-based survival prediction of patients with non-small cell lung cancer [34].

Although our results are clear and significant, this study has a number of limitations. First, the number of included patients is relatively small. The sample size is a challenging issue in radiomics-based analysis for predictive modelling. A larger sample size would have undoubtedly resulted in more accurate models. Second, we used only BMD images for radiomic-based analysis. Further investigation involving other imaging modalities are needed to verify our results. Third, we compared radiomic results to fracture risk prediction because several patients present with osteoporotic fracture with a normal T-score. For example further studies are needed to compare radiomic results between one group with osteoporotic fracture and another group without osteoporotic fracture. Fourth, we suggest high reproducible, repeatable and robust radiomic features for such studies as suggested by various studies due to radiomics feature changes over image acquisition, reconstruction, segmentation and data modelling [35—37]. Finally, although we compared different ML approaches, we suggest to devise a unified algorithm to classify patients into osteoporosis, osteopenia and normal. In a clinical setting, this issue will help produce more accurate predictive models.

In conclusion, we analyzed BMD radiomics to classify bone mineral loss. We found that BMD radiomics combined with several machine learning algorithms could predict osteoporosis and osteopenia. It should be noted that this is a pilot study reporting on the feasibility of using radiomics for the assessment of bone quality on DXA images.

Human and animal rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans.

Informed consent and patient details

The authors declare that this report does not contain any personal information that could lead to the identification of the patient(s).

The authors declare that they obtained a written informed consent from the patients and/or volunteers included in the article. The authors also confirm that the personal details of the patients and/or volunteers have been removed.

Funding

This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Credit author statement

Sajjad Rastegar: methodology and software.

Mohammad Vaziri: data curation.

Younes Qasempour: visualization and investigation.

Mohammad Rashid Akhash: resources.

Neda Abdalvand: writing original draft preparation.

Isaac Shiri: software and validation.

Hamid Abdollahi: supervision, conceptualization, methodology, writing, reviewing and editing.

Disclosure of interest

The authors declare that they have no competing interest.

References

[1] Haider IT, Lobos SM, Simonian N, Schnitzer TJ, Edwards WB. Bone fragility after spinal cord injury: reductions in stiffness and bone mineral at the distal femur and proximal tibia as a function of time. Osteoporos Int 2018;29:2703—15.

[2] Kim J, Kim SW, Lee SY, Kim TH, Jung JH. Bone mineral density in osteoporotic patients with pyogenic vertebral osteomyelitis: effect of early versus late treatment for osteoporosis. Osteoporos Int 2018;29:2761—70.

[3] Kanis JA. Diagnosis of osteoporosis and assessment of fracture risk. Lancet 2002;359:1929—36.

[4] Blake GM, Fogelman I. Role of dual-energy X-ray absorptiometry in the diagnosis and treatment of osteoporosis. J Clin Densitom 2007;10:102—10.

[5] Abdollahi H, Mostafaei S, Cheraghi S, Shiri I, Rabi Mahdavi S, Kazemnejad A. Cochlea CT radiomics predicts chemoradiotherapy induced sensorineural hearing loss in head and neck

cancer patients: A machine learning and multi-variable modelling study. Phys Med 2018;45:192—7.
[6] Abdollahi H, Shiri I, Heydari M. Medical imaging technologists in radiomics era: an Alice in wonderland problem. Iran J Public Health 2019;48:184—6.
[7] Nazari M, Shiri I, Hajianfar G, Oveisi N, Abdollahi H, Deevband MR, et al. Non-invasive Fuhrman grading of clear cell renal cell carcinoma using computed tomography radiomics features and machine learning; 2019 [arXiv pre print arXiv: 190912286].
[8] Abdollahi H, Tanha K, Mofid B, Razzaghdoust A, Saadipoor A, Khalafi L, et al. MRI radiomic analysis of IMRT-induced bladder wall changes in prostate cancer patients: a relationship with radiation dose and toxicity. J Med Imaging Radiat Sci 2019;50:252—60.
[9] Abdollahi H. Radiotherapy dose painting by circadian rhythm based radiomics. Med Hypoth 2019;133:109415.
[10] Burian E, Subburaj K, Mookiah MRK, Rohrmeier A, Hedderich DM, Dieckmeyer M, et al. Texture analysis of vertebral bone marrow using chemical shift encoding-based water-fat MRI: a feasibility study. Osteoporos Int 2019;30:1265—74.
[11] de Sa Cavalcante D, da Silva Castro MG, Quidute ARP, Martins MRA, Cid A, de Barros Silva PG, et al. Evaluation of bone texture imaging parameters on panoramic radiographs of patients with Sheehan's syndrome: a STROBE-compliant case-control study. Osteoporos Int 2019;30:2257—69.
[12] Kawashima Y, Fujita A, Buch K, Li B, Qureshi MM, Chapman MN, et al. Using texture analysis of head CT images to differentiate osteoporosis from normal bone density. Eur J Radiol 2019;116:212—8.
[13] Panahi N, Soltani A, Ghasem-Zadeh A, Shafiee G, Heshmat R, Razi F, et al. Associations between the lipid profile and the lumbar spine bone mineral density and trabecular bone score in elderly Iranian individuals participating in the Bushehr Elderly Health Program: a population-based study. Arch Osteoporos 2019;14, 52.1-11.
[14] Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol Biol Phys 2018;102:1143—58.
[15] Valentinitsch A, Trebeschi S, Kaesmacher J, Lorenz C, Loffler MT, Zimmer C, et al. Opportunistic osteoporosis screening in multi-detector CT images via local classification of textures. Osteoporos Int 2019;30:1275—85.
[16] White R, Krueger D, De Guio F, Michelet F, Hans D, Anderson P, et al. An exploratory study of the texture research investigational platform (TRIP) to evaluate bone texture score of distal femur DXA scans — A TBS-Based Approach. J Clin Densitom 2019, http://dx.doi.org/10.1016/j.jocd.2019.06.004.
[17] SFR-IA Group; CERF; French Radiology Community. Artificial intelligence and medical imaging 2018: French Radiology Community white paper. Diagn Interv Imaging 2018;99: 727—42.
[18] Couteaux V, Si-Mohamed S, Nempont O, Lefevre T, Popoff A, Pizaine G, et al. Automatic knee meniscus tear detection and orientation classification with Mask-RCNN. Diagn Interv Imaging 2019;100:235—42.
[19] Awaysheh A, Wilcke J, Elvinger F, Rees L, Fan W, Zimmerman KL. Review of medical decision support and machine-learning methods. Vet Pathol 2019;56:512—25.
[20] Roblot V, Giret Y, Bou Antoun M, Morillot C, Chassin X, et al. Artificial intelligence to diagnose meniscus tears on MRI. Diagn Interv Imaging 2019;100:243—9.
[21] Abdollahi H, Mofid B, Shiri I, Razzaghdoust A, Saadipoor A, Mahdavi A, et al. Machine learning-based radiomic models to predict intensity-modulated radiation therapy response, Gleason score and stage in prostate cancer. Radiol Med 2019;124:555—67.
[22] Ferizi U, Besser H, Hysi P, Jacobs J, Rajapakse CS, Chen C, et al. Artificial intelligence applied to osteoporosis: a performance comparison of machine learning algorithms in predicting fragility fractures from MRI data. J Magn Reson Imaging 2019;49:1029—38.
[23] Ferizi U, Honig S, Chang G. Artificial intelligence, osteoporosis and fragility fractures. Curr Opin Rheumatol 2019;31:368—75.
[24] Halldorsson BV, Bjornsson AH, Gudmundsson HT, Birgisson EO, Ludviksson BR, Gudbjornsson B. A clinical decision support system for the diagnosis, fracture risks and treatment of osteoporosis. Comput Math Methods Med 2015;2015:189769.
[25] Kharroubi A, Saba E, Ghannam I, Darwish H. Evaluation of the validity of osteoporosis and fracture risk assessment tools (IOF One Minute Test, SCORE, and FRAX) in postmenopausal Palestinian women. Arch Osteoporos 2017;12:6.
[26] Conradie M, Conradie MM, Kidd M, Hough S. Bone density in black and white South African women: contribution of ethnicity, body weight and lifestyle. Arch Osteoporos 2014;9:193.
[27] Conradie M, Conradie MM, Scher AT, Kidd M, Hough S. Vertebral fracture prevalence in black and white South African women. Arch Osteoporos 2015;10:203.
[28] Bolotin H. DXA in vivo BMD methodology: an erroneous and misleading research and clinical gauge of bone mineral status, bone fragility, and bone remodelling. Bone 2007;41:138—54.
[29] Hirvasniemi J, Gielis WP, Arbabi S, Agricola R, van Spil WE, Arbabi V, et al. Bone texture analysis for prediction of incident radiographic hip osteoarthritis using machine learning: data from the Cohort Hip and Cohort Knee (CHECK) study. Osteoarthr Cartil 2019;27:906—14.
[30] Le Corroller T, Halgrin J, Pithioux M, Guenoun D, Chabrand P, Champsaur P. Combination of texture analysis and bone mineral density improves the prediction of fracture load in human femurs. Osteoporosis Int 2012;23:163—9.
[31] Abdollahi H, Mahdavi SR, Mofid B, Bakhshandeh M, Razzaghdoust A, Saadipoor A, et al. Rectal wall MRI radiomics in prostate cancer patients: prediction of and correlation with early rectal toxicity. Int J Radiat Biol 2018;94:829—37.
[32] Parmar C, Leijenaar RT, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, et al. Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer. Sci Rep 2015;5:11044.
[33] Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. Cancer Lett 2017;403:21—7.
[34] Hawkins J, Balagurunathan Y, Gu Y, Kumar V, Basu S, Hall LO, et al. Predicting outcomes of nonsmall cell lung cancer using CT image features. IEEE Access 2014;25:1418—26.
[35] Saeedi E, Dezhkam A, Beigi J, Rastegar S, Yousefi Z, Mehdipour LA, et al. Radiomic feature robustness and reproducibility in quantitative bone radiography: a study on radiologic parameter changes. J Clin Densitom 2019;22:203—13.
[36] Shiri I, Rahmim A, Ghaffarian P, Geramifar P, Abdollahi H, Bitarafan-Rajabi A. The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies. Eur Radiol 2017;27:4498—509.
[37] Shiri I, Abdollahi H, Shaysteh S, Rabi Mahdavi S. Test-retest reproducibility and robustness analysis of recurrent glioblastoma MRI radiomics texture features. Iran J Radiol 2017. Special iss:e48035.