



Integrating deep learning and machine learning for improved CKD-related cortical bone assessment in HRpQCT images: A pilot study

Youngjun Lee^a, Wikum R. Bandara^a, Sangjun Park^b, Miran Lee^c, Choongboem Seo^d, Sunwoo Yang^e, Kenneth J. Lim^f, Sharon M. Moe^f, Stuart J. Warden^g, Rachel K. Surowiec^{a,*}

^a Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, United States of America

^b Department of Radiation Therapy, Samsung Medical Center, The Sungkyunkwan University of Korea, South Korea

^c Department of Electronic Engineering, Sogang University, Cardiac Sonographer at the Department of Total Healthcare Center, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, South Korea

^d Department of Radiology, Seoul National University Bundang Hospital, The Seoul National University of Korea, South Korea

^e Department of Radiological Technology, Shingu University, South Korea

^f Division of Nephrology and Hypertension, Indiana University School of Medicine, Indianapolis, IN, United States of America

^g Department of Physical Therapy, School of Health and Human Sciences, Indiana University Indianapolis, Indianapolis, IN, United States of America

ARTICLE INFO

Keywords:

High resolution peripheral quantitative computed tomography
Chronic kidney disease
Deep learning
Machine learning
Texture analysis
Segmentation
Classification

ABSTRACT

High resolution peripheral quantitative computed tomography (HRpQCT) offers detailed bone geometry and microarchitecture assessment, including cortical porosity, but assessing chronic kidney disease (CKD) bone images remains challenging. This proof-of-concept study merges deep learning and machine learning to 1) improve automatic segmentation, particularly in cases with severe cortical porosity and trabeculated endosteal surfaces, and 2) maximize image information using machine learning feature extraction to classify CKD-related skeletal abnormalities, surpassing conventional DXA and CT measures.

We included 30 individuals (20 non-CKD, 10 stage 3 to 5D CKD) who underwent HRpQCT of the distal and diaphyseal radius and tibia and contributed data to develop and validate four different AI models for each anatomical site. Manually annotated cortical bone was used to train each segmentation deep-learning model. Textural features were extracted via Gray-Level Co-occurrence Matrix (GLCM) and classified as CKD or non-CKD using XGBoost with each segmentation model. For comparison, manufacturer-supplied segmentation was used to extract cortical geometry, microarchitecture, and finite element analysis (FEA) outcomes. Model performance was confirmed using the test dataset and a separate independent validation cohort which included HRpQCT imaging from 42 additional individuals (18 non-CKD, 24 CKD stage 5D).

For segmentation, the diaphyseal location showed strong performance on test datasets, with Mean IoUs of 0.96 and 0.95, and accuracies of 0.97 for both radius and tibia sites in CKD. Model 4 developed from the diaphyseal tibia region excelled in classifying test and independent validation datasets, achieving F1 scores of 0.99 and 0.96, AUCs of 0.99 and 0.94, sensitivities of 0.99, and specificities of 0.99 and 0.92. No single parameter, including BMD and cortical porosity, among conventional CT outcomes consistently differentiated CKD from non-CKD across all anatomical sites.

Integrating HRpQCT with deep and machine learning, this innovative approach enables precise automatic segmentation of severely deteriorated endocortical surfaces and enhances sensitivity to CKD-related cortical bone changes compared to standard DXA and HRpQCT outcomes.

1. Introduction

Unlike many other skeletal diseases, chronic kidney disease (CKD)

primarily impacts cortical bone through the formation/expansion of cortical pores (pathological holes inside the cortex) (Nickolas et al., 2008), cortical thinning, and trabecularization of endocortical bone

* Corresponding author.

E-mail addresses: lee4731@purdue.edu (Y. Lee), wranasin@purdue.edu (W.R. Bandara), sj92.park@samsung.com (S. Park), mr119.lee@samsung.com (M. Lee), 20391@snuh.org (C. Seo), kjlim@iu.edu (K.J. Lim), stwarden@iu.edu (S.J. Warden), rsurowie@purdue.edu (R.K. Surowiec).

<https://doi.org/10.1016/j.bonr.2024.101821>

Received 19 October 2024; Received in revised form 18 November 2024; Accepted 24 December 2024

Available online 26 December 2024

2352-1872/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

surfaces (Nickolas et al., 2010). Importantly, these key features of CKD-related cortical bone loss negatively impact mechanical properties essential for fracture resistance (Newman et al., 2014). Dual-energy X-ray absorptiometry (DXA) remains the standard clinical imaging assessment for predicting fracture risk. Still, DXA often fails to adequately discriminate fracture risk in CKD, even in cases of normal or high bone mineral density (BMD) (Piraino et al., 1988; Urena et al., 2003). This is likely due to its inability to distinguish between bone types (cortical vs. trabecular), and as a result, DXA in the setting of renal disease has long been challenged (Leonard and Bachrach, 2001). Thus, it is necessary to identify early skeletal consequences of CKD, particularly those connected to the cortical bone, to allow for timely intervention that could reduce high fragility fracture rates in CKD.

High resolution peripheral quantitative computed tomography (HRpQCT) has emerged as a powerful imaging modality allowing detailed bone geometry and microarchitecture, including cortical porosity assessments using a nominal isotropic voxel size (60.7 μm) (Surowiec et al., 2022). Initial CKD studies have shown that HRpQCT can allow for detailed measurements of cortical bone properties, such as cortical bone mineral density (BMD), porosity, and thickness (Nickolas et al., 2013). However, in cases of notable cortical thinning, severe porosity, and endocortical trabecularization—common features in CKD (Davis et al., 2007; Nishiyama et al., 2010)—standard automated segmentation can fail at extracting the cortex (Sharma et al., 2022; Tsuji et al., 2022). This necessitates manual inspection of the 168-slice stacks and correction by an expert or trained user, introducing variability and resulting in additional time investment (Whittier et al., 2020a; Whittier et al., 2020b). While some have introduced novel automated segmentation routines to rescue fine microarchitectural features (Sadoughi et al., 2023), application in severe CKD images has not been evaluated.

While precise segmentation is a critical first step in image analysis, maximizing image information through techniques like textural feature extraction using machine learning could complement or surpass current cortical bone outcomes that rely on geometry and density. Textural analysis is a promising method for evaluating spatial patterns that indicate tissue heterogeneity (Ganesan and Miles, 2013; Lu et al., 2023). This approach has successfully detected early disease-related changes in non-bone tissues in medical images (Pantic et al., 2023; Ricardo et al., 2023; Baidya Kayal et al., 2021). Textural analysis can significantly benefit from machine learning algorithms' refined pattern recognition capabilities. Deep learning has greatly advanced medical imaging, especially in bone segmentation, by detecting intricate patterns in HRpQCT images (Sadoughi et al., 2023; Figueiredo et al., 2018; Li et al., 2015; Ohs et al., 2021; Valentinitsch et al., 2012). In our approach, we use deep learning to enhance segmentation accuracy. However, for the next step of analyzing the texture features in these segmented images, we opted for machine learning. Machine learning, particularly XGBoost, is better suited for smaller datasets and provides clearer insights into which features are important for classifying CKD-related bone abnormalities (Chen and Guestrin, 2016). By combining deep learning for segmentation with machine learning for classification (Seeja and Suresh, 2019), we are able to maximize the strengths of both techniques.

The primary objective of this proof-of-concept study is to develop an integrated AI framework that utilizes deep learning for segmentation and machine learning for classification, incorporating textural features from HRpQCT images of patients with advanced CKD and non-CKD individuals. This combined approach aims to accurately and reliably detect cortical bone abnormalities related to CKD compared to age- and sex-matched non-CKD control images. We hypothesize that machine learning classification will differentiate CKD from non-CKD bone in both the distal and diaphyseal radius and tibia. This approach is expected to surpass standard DXA and HRpQCT assessments, with potential benefits for patient management and personalized treatment planning, underscoring its clinical significance in CKD evaluation.

2. Methods

2.1. Study participants and high-resolution peripheral quantitative computed tomography

Fifteen CKD patients who underwent HRpQCT at the Musculoskeletal Function, Imaging, and Tissue Resource Core (FIT Core) of the Indiana Center for Musculoskeletal Health between 2018 and 2022 were initially included in the study. All participants provided written informed consent under Institutional Review Board approval from Indiana University. CKD patients were recruited either through ongoing studies or as part of the Core's 'all-comer' cohort and were flagged as 'CKD' in the FIT Core database. Inclusion criteria required participants to be 18 years or older with CKD stage 3 or higher. No exclusions were made based on race, ethnicity, or gender. Exclusions applied to pregnant or breastfeeding individuals, those with active implanted medical devices, or patients undergoing chemotherapy. Additionally, age- and sex-matched 29 non-CKD individuals were retrospectively identified from the FIT Core database.

As no public CKD HRpQCT datasets were available for testing our model performance, additional data from CKD and non-CKD participants meeting inclusion criteria were collected from Jan 1, 2023, to Mar 31, 2024. This 'independent validation cohort' included 42 individuals, with 24 CKD stage 5D patients and 18 non-CKD participants used to test model performance.

HRpQCT (XtremeCT II, Scanco Medical, Bruttisellen, Switzerland) was acquired on the participants' non-dominant arm and leg, as previously described (Warden et al., 2022a). Scan stacks were positioned at 4 % and 30 % of the bone length proximal to the radius reference line, and at 7.3 % and 30 % of the bone length proximal to the tibia reference line. Motion was assessed using a visual grading score (VGS) ranging from 1 (no artifacts) to 5 (significant artifacts) (Pialat et al., 2012). Images with a VGS score ≥ 3 were excluded. Consequently, 5 CKD and 9 non-CKD participants were excluded from the study cohort due to scores of 3 or higher in the distal radius and tibia, and diaphyseal radius and tibia. The final study dataset included scans from 10 CKD and 20 non-CKD participants (20,160 DICOM images), while the independent validation dataset comprised scans from 24 CKD and 18 non-CKD participants (28,224 DICOM images) (Fig. 1).

Participants' height (measured to the nearest 0.1 cm) and weight (measured to the nearest 0.1 kg) were obtained. Appendicular skeletal muscle mass relative to height ($\text{ASM}/\text{height}^2$; kg/m^2) and whole-body areal bone mineral density (aBMD) were evaluated through whole-body DXA (Norland Elite; Norland at Swissray, Fort Atkinson, WI). The same DXA scanner was used assess total hip and spine aBMD.

Standard HRpQCT outcomes were assessed by reconstructing raw scans and extracting cortical bone using manufacturer-supplied scripts with a low-pass Gaussian filter (sigma 0.8, support 1.0 voxel) and a fixed threshold of $450 \text{ mgHA}/\text{cm}^3$, as previously described (Warden et al., 2022b). The following conventional outcomes were recorded at the diaphyseal and distal site: volumetric BMD (Ct.vBMD, mgHA/cm^3), area (Ct.Ar, mm^2), thickness (Ct.Th, mm), and porosity (Ct.Po, %). μFE analysis (Scanco Medical FE software version 1.13) was used to estimate stiffness (kN/mm) using a pixel-wise assigned modulus of 10GPa and a Poisson's ratio of 0.3 under axial compression as previously described (Arias-Moreno et al., 2019).

2.2. Data preprocessing

Separate models for segmentation and classification of CKD and non-CKD were developed and tested using distinct datasets for each anatomical site (distal and diaphyseal radius and tibia). The data was divided into training (70 %, 4704 images for CKD, 9408 images for non-CKD), validation (10 %, 1344 images for non-CKD, 672 images for CKD), and testing (20 %, 2688 images for non-CKD, 1344 images for CKD) cohorts. Efforts were made to maintain the same average age and sex of

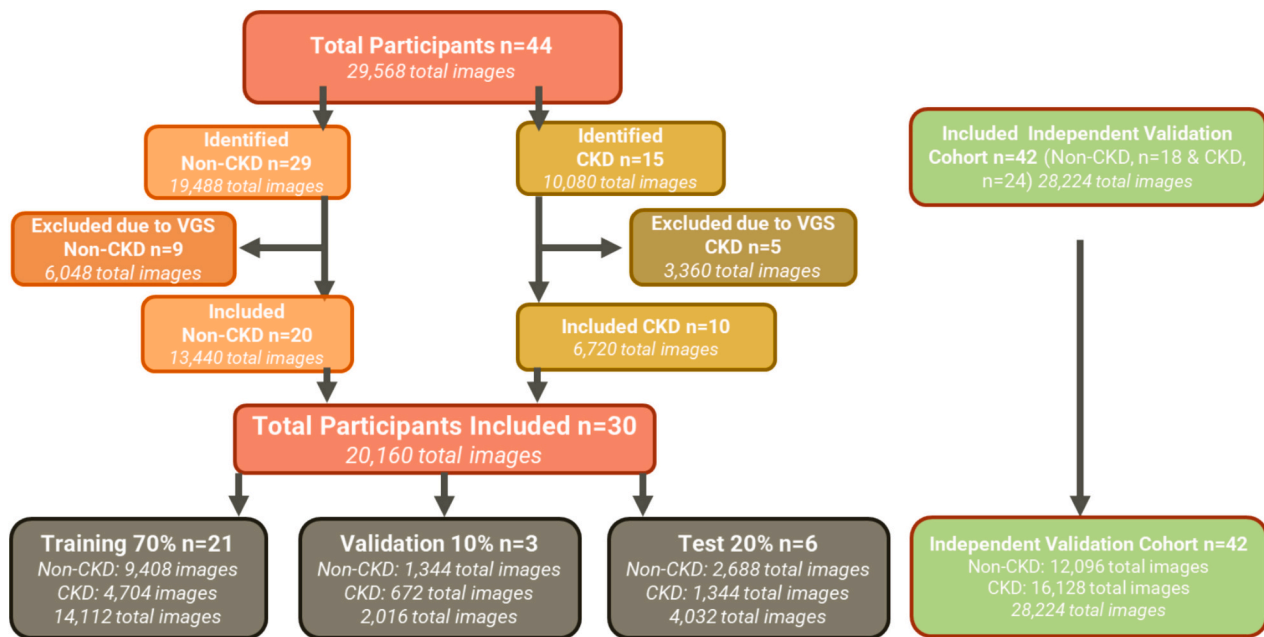


Fig. 1. Flow diagram for image inclusion. HRpQCT from 44 participants were retrospectively considered in the current study. Exclusions were made for nine non-chronic kidney disease (non-CKD) participants and five CKD participants who had a visual grading score (VGS) for motion artifacts of three or above in at least one location (distal or diaphyseal radius or tibia). The study ultimately included HRpQCT volumes from 30 participants, encompassing the distal and diaphyseal radius and tibia, totaling 20,160 individual images (non-CKD = 20 [13,440 images], CKD = 10 [6720 images]). The final dataset was randomly divided into a training cohort (70 %), a validation cohort (10 %), and a testing cohort (20 %), comprising a 7: 1:2 ratio for each stage of the project. An independent validation cohort of $n = 18$ and $n = 24$ (28,224 images) individuals whose images were not used during model development or in the initial validation stage were included to test model performance.

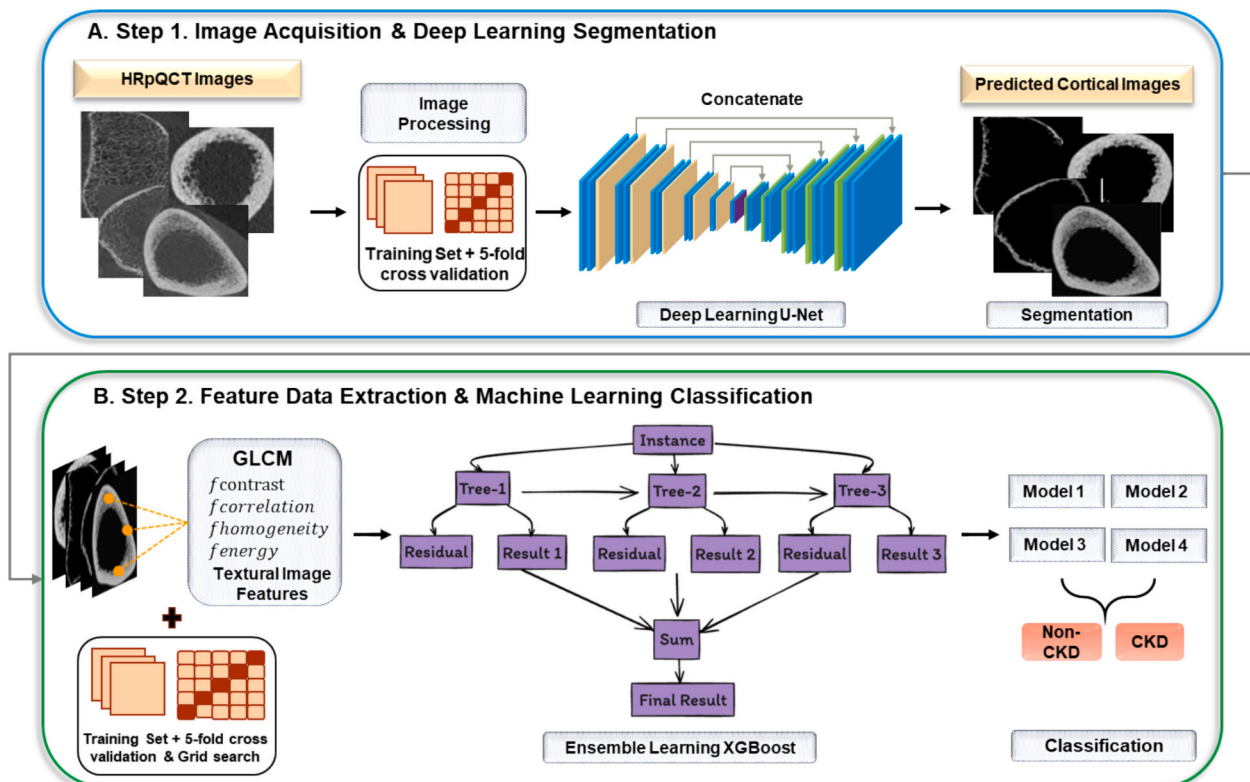


Fig. 2. Workflow of the two-stage cortical bone analysis using HRpQCT. A. The process initiates with the preprocessing of axial HRpQCT scans, followed by segmentation of non-CKD and CKD cortical bones using a U-Net model trained using 5-fold cross-validation to make the model robust. B. The isolated cortical bones, shown as grayscale images that maintain the original pixel values, are then used for GLCM feature extraction. These features are subsequently trained using 5-fold cross-validation and grid search for optimal performance. Subsequently, an XGBoost machine learning classification is employed to determine distinctive attributes between non-CKD and CKD cortical bones. Model 1 = Distal radius; Model 2 = Diaphyseal radius; Model 3 = Distal tibia; Model 4 = Diaphyseal tibia.

patients throughout all the datasets. Datasets were split by patient rather than by image to prevent the assignment of the same patients or the mixing of patients' images between the training, validation, and test datasets. The following four AI models were developed:

1. **Model 1:** Distal radius segmentation & classification model
2. **Model 2:** Diaphyseal radius segmentation & classification model
3. **Model 3:** Distal tibia segmentation & classification model
4. **Model 4:** Diaphyseal tibia segmentation & classification model

2.3. Manual annotations

Volumes were first cropped to contain only the tibia or radius, removing the fibula or ulna, respectively. The cortical bone was segmented from the axial images by applying a threshold followed by manual correction using 3D Slicer (<https://www.slicer.org>) to create manual annotations. Two independent experts reviewed the annotated masks, and any discrepancies were resolved through consensus discussions. Each image was then resized to 512×512 pixels without image augmentation and normalized to ensure consistent intensity values. Segmentation models generated binary masks for the cortical region, excluding trabecular bone and soft tissue. These masks were applied to the original images to isolate the cortical region, preserving the original CT intensity values.

2.4. Stage 1. Enhanced cortical bone segmentation via deep learning

We employed a deep learning approach to enhance the segmentation of the cortical bone, particularly in cases with severe cortical abnormalities (Fig. 2A). The U-net architecture, known for its effectiveness in image segmentation, serves as the foundation for the model due to its previous success in medical image segmentation tasks (Abedalla et al., 2020). To find the proper hyperparameters and ensure reliability and generalizability, a 5-fold cross-validation strategy was used (Wong and Yeh, 2019). Hyperparameters (Adam optimizer, learning rate set to 0.001, a batch size of 16, loss function of binary cross entropy) were chosen for both cohorts. After processing the cortical bone images, the two groups (CKD, non-CKD) for each of the four models (distal and diaphyseal radius and tibia) underwent a training phase. The training process was repeated, with the models being adjusted each time, until no further improvement was observed in their performance on a separate validation dataset across multiple training epochs, meaning that the entire training run was repeated until validation performance stopped improving. The model that achieved the lowest validation loss during this process was then selected for testing on the performance datasets. The segmentation model performances were evaluated using several well-defined metrics described in detail in the statistical analysis section before initiating the classification stage (stage 2). The model segmentation masks were combined with the original corresponding images to isolate the cortical region and guide further image analysis.

2.5. Stage 2. Utilizing machine learning for CKD vs. non-CKD image classification

2.5.1. Gray level co-occurrence matrix (GLCM) feature extraction

In the initial phase of stage 2, features were extracted from the segmented cortices of the HRpQCT images (Fig. 2B). Using the Gray Level Co-Occurrence Matrix (GLCM), we analyzed the texture of the cortical bone region by quantifying the occurrence of pixel value pairs at specified distances, following established methodologies (Zulpe and Pawar, 2012). The study focused on the following GLCM parameters: contrast, homogeneity, energy, and correlation.

2.5.2. Machine learning classification with XGBoost

After extracting texture features, we employed XGBoost (eXtreme Gradient Boosting) (Chen et al., 2015), a robust machine learning

algorithm. In simple terms, XGBoost is an advanced system that learns from data, such as texture patterns via GLCM, to improve its predictive accuracy over time. XGBoost excels at identifying patterns within large datasets. We optimized the model's settings using a grid search with variations in key parameters: number of estimator (50, 100, 200), maximum depth (3, 5, 7), and learning rate (0.01, 0.1, 0.2). This process identifies the best combination of hyperparameters to balance model accuracy with 5-fold cross-validation. The computations were performed on a system equipped with an Intel(R) Xeon(R) CPU @ 2.30GHz and a GPU A100.

2.6. Independent validation

The internal validation dataset was used to independently validate the model's ability to distinguish between CKD and non-CKD cortical bone images. Segmentations were automatically performed using the approach in stage 1. This was followed by classification described in stage 2.

2.7. Statistical analysis

Participant characteristics and conventional cortical HRpQCT outcomes are presented as mean \pm standard deviation. The Shapiro-Wilk test was employed to assess the normality of the data. For datasets exhibiting a normal distribution within each group, independent *t*-tests and chi-squared tests were applied as relevant for analysis. A significance threshold of $p < 0.05$ was applied.

We employed several metrics to assess segmentation model performance, including the Dice similarity coefficient (DSC), mean Intersection over Union (IoU), accuracy, and Hausdorff Distance (Huttenlocher et al., 1993). The DSC is a metric for segmentation accuracy and was used to assess the model's ability to identify bone regions within HRpQCT images. The DSC quantifies the level of overlap between the segmented regions predicted by the model and the ground truth annotations, with a higher DSC indicating superior segmentation performance. The mean IoU, another spatial overlap metric, was also used to gauge the degree of overlap between the segmented regions and ground truth annotations. Accuracy was used to assess overall segmentation model performance. Average HD measures the mean of the maximum distance between any point on the first set (endosteal surface of original data) and its nearest point on the second set (endosteal surface of predicted binary mask). The HD proves invaluable in validating CKD-related image-based texture variation by calculating the distance between each point in overlapped images of two sets, given its proficiency in quantifying the similarity or dissimilarity of intricate structures. In addition, we identified systemic and random errors in HRpQCT-derived automated contour segmentation.

Model performance for classification was evaluated using standard metrics. We first calculated the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) to measure the model's ability to distinguish CKD cortical bone from non-CKD cortical bone images. A higher AUC-ROC value indicates better classification performance and the model's effectiveness in distinguishing between the two groups (DeLong et al., 1988). Additional metrics derived from the confusion matrix, such as accuracy, precision, recall, and the F1 score, were used to evaluate the model's performance in classifying CKD and non-CKD images in both the original and independent validation datasets. To validate model consistency, we compared performance metrics, including True Positive Rate (TPR, sensitivity) and True Negative Rate (TNR, specificity), between the test set and an independent validation set. These metrics were assessed in terms of the model's accuracy in predicting CKD and non-CKD images.

XGBoost classifier performance was evaluated in Python 3.7 using Keras and TensorFlow. The best performing models were chosen for each task to quantify the entire integrated pipeline performance, and the resulting error in the cortices is presented. SPSS v28 was used for

statistical analysis.

3. Results

3.1. Participant characteristics and conventional HRpQCT outcomes

Participant characteristics are detailed in Table 1. Eight of the 10 CKD patients were classified as CKD stage 5D and receiving hemodialysis; the remainder were CKD stage 3 (Supplemental Table 1). The study featured participants from diverse racial backgrounds and included individuals self-reporting as non-Hispanic White ($n = 17$), non-Hispanic Black ($n = 11$), Hispanic White ($n = 1$) and Asian ($n = 1$). Shapiro-Wilks results indicated that data from all groups exhibited a normal distribution ($p > 0.05$), allowing us to proceed with further analyses.

Results comparing the non-CKD vs. CKD group's standard cortical HRpQCT outcomes can be found in Supplemental Table 2. No individual standard cortical HRpQCT outcome could effectively distinguish between cortical bone characteristics in non-CKD and CKD cases in our cohort across each site. Because our CKD cohort contained CKD types 3 and 5D, we conducted a sub-analysis to evaluate any differences between the HRpQCT outcomes when only CKD 5D vs. non-CKD were considered. We observed no changes in statistical outcomes when doing this evaluation. Thus, we moved forward with the 10 CKD patients containing both stage 3 and 5D for the remaining evaluation.

3.2. Stage 1: accuracy of the segmentation models

The segmentation and classification accuracy reported herein reflects an analysis of multiple cross-sectional images per subject, highlighting the texture-gradient boost's capability to discern CKD presence at the image level rather than subject diagnosis.

The DSC, Mean IoU, accuracy, and Average HD values from the segmentation models are presented in Table 2. All models of four sites employed the binary cross-entropy loss function. We also used early stopping to prevent overfitting and identify the optimal epoch, and each model was assessed on the 1344 test set images. Overall, the diaphyseal cortical CKD models (radius, tibia) outperformed the distal cortical CKD models (radius, tibia) with the following metrics: DSC = 0.97, 0.96 vs. 0.88, 0.91; mean IoU = 0.96, 0.95 vs. 0.84, 0.86; accuracy = 0.97, 0.97 vs. 0.96, 0.95; respectively.

Qualitatively, nearly all outcomes from the trained models corresponded visually with the ground truth masks and did not overestimate the manually segmented masks (examples shown in Supplemental Fig. 1). Each DSC training graph of the segmentation algorithm is included in Supplemental Fig. 2. The automated method generated by HRpQCT incorrectly identified regions, misclassifying large trabecular bone areas as cortical bone and large pores or low-density cortical bone as trabecular bone (see Supplemental Fig. 3).

Table 1
Subject demographics.

Characteristics	Mean (standard deviation), n (%)					
	Healthy Volunteers (n = 20)	CKD Patients (n = 10)	P-value	Independent Validation		
				Healthy Volunteers (n = 18)	CKD Patients (n = 24)	P-value
Age (years)	53.23 ± 11.9	54.06 ± 10.38	0.91	54.58 ± 8.37	51.63 ± 14.51	0.42
Height (cm)	169.10 ± 11.82	173.43 ± 10.18	0.32	172.88 ± 8.53	168.48 ± 11.19	0.16
Weight (kg)	80.39 ± 15.44	79.57 ± 14.91	0.76	77.48 ± 11.79	85.93 ± 18.46	0.08
BMI (kg/m ²)	26.82 ± 4.53	26.45 ± 4.22	0.37	25.87 ± 3.00	30.43 ± 7.07	0.01
Whole body Total T-score	0.09 ± 0.45	−0.37 ± 1.02	0.21	−0.24 ± 1.39	−0.88 ± 1.58	0.18
Whole body Total Z-score	0.56 ± 0.67	−0.13 ± 1.02	0.05	0.04 ± 1.12	−0.45 ± 1.22	0.19
Sex (Female)	10 (50 %)	5 (50 %)	0.80	9 (50 %)	10 (42 %)	0.71

Note: The p -values between non-CKD and CKD patients were determined utilizing t-tests or chi-squared tests. Values are shown as means \pm standard deviation or as counts (percentages). P -values are bolded when significance is reached. BMI = body mass index. T- and Z-scores were determined by DXA.

Table 2

Performance of cortical bone segmentation using the original test dataset.

Model	Location	Segmentation Performance Metrics				
		Group	DSC	Mean IoU	Accuracy	Average Hausdorff Distance
Model 1	Radius Distal	Non-CKD	0.88	0.86	0.95	24.51
		CKD	0.88	0.84	0.96	34.28
Model 2	Radius Diaphyseal	Non-CKD	0.98	0.96	0.97	12.70
		CKD	0.97	0.96	0.97	15.20
Model 3	Tibia Distal	Non-CKD	0.91	0.85	0.95	18.24
		CKD	0.91	0.86	0.95	23.83
Model 4	Tibia Diaphyseal	Non-CKD	0.95	0.92	0.96	29.12
		CKD	0.96	0.95	0.97	35.01

Note: A probability threshold of 0.5 was used to evaluate the DSC, IoU, accuracy, and Hausdorff Distance. The highest values in each column are bolded. CKD = chronic kidney disease; DSC = dice similarity coefficient; Mean IoU = mean intersection over union.

3.3. Stage 2: evaluation of the classification models

Extracted GLCM texture features from the segmented cortical HRpQCT images were used as inputs to the ensemble XGBoost machine learning model. We optimized the model's hyperparameters (number of estimators: 200, learning rate: 0.2, maximum depth: 5) via the grid search method and compared it with the XGBoost default parameters. Table 3 presents performance of XGBoost with optimized hyperparameters and Supplemental Table 3 presents the performance using default XGBoost parameters.

With optimization, the diaphyseal tibia model almost achieved peak performance with a score of 99 % across precision, recall, F1-score, accuracy, and AUC-ROC (Supplemental Fig. 4). The confusion matrix for the classification models was assessed based on a 5-fold validation technique with- and without optimized hyperparameters, illustrated in Fig. 3 and Supplemental Fig. 5, respectively. In short, the confusion matrix data shows better classification performance following hyperparameter adjustment, with classification at the diaphyseal radius outperforming all sites. Details on GLCM feature importance are found in Supplemental Fig. 6.

Some images in the CKD dataset were inaccurately classified as non-CKD, and conversely, some non-CKD images were misclassified as CKD. Representative cross-sectional images of these misclassifications are depicted in Supplemental Fig. 7. Upon close examination, it is evident that non-CKD cases misclassified as CKD exhibited increased cortical porosity and visually thinner cortices. Conversely, CKD cases misclassified as non-CKD were associated with lower pore numbers and

Table 3
Performance of XGBoost machine learning classification with hyperparameter optimization in distinguishing between non-CKD and CKD cortical bone from HRPQCT images using the test dataset.

Model	Location	Group	Classification Performance Metrics				
			Precision	Recall	F1 score	Accuracy	AUC-ROC
Model 1	Radius Distal	Non-CKD	0.97	0.96	0.97	0.96	0.98
		CKD	0.93	0.94	0.94	0.94	
Model 2	Radius Diaphyseal	Non-CKD	0.99	0.98	0.99	0.99	0.99
		CKD	0.98	0.98	0.98	0.98	
Model 3	Tibia Distal	Non-CKD	0.96	0.98	0.97	0.95	0.98
		CKD	0.96	0.92	0.94	0.92	
Model 4	Tibia Diaphyseal	Non-CKD	0.99	0.99	0.99	0.98	0.99
		CKD	0.99	0.99	0.99	0.99	

Note: XGBoost is a gradient-boosted machine-learning decision tree or classifier. A probability threshold 0.5 was used to evaluate the precision, recall, F1 score, accuracy, AUC-ROC. The highest values in each column are bolded. CKD = chronic kidney disease; DSC = dice coefficient similarity; AUC-ROC = area under the receiver operating characteristic curve; XGBoost = Extreme Gradient Boosting.

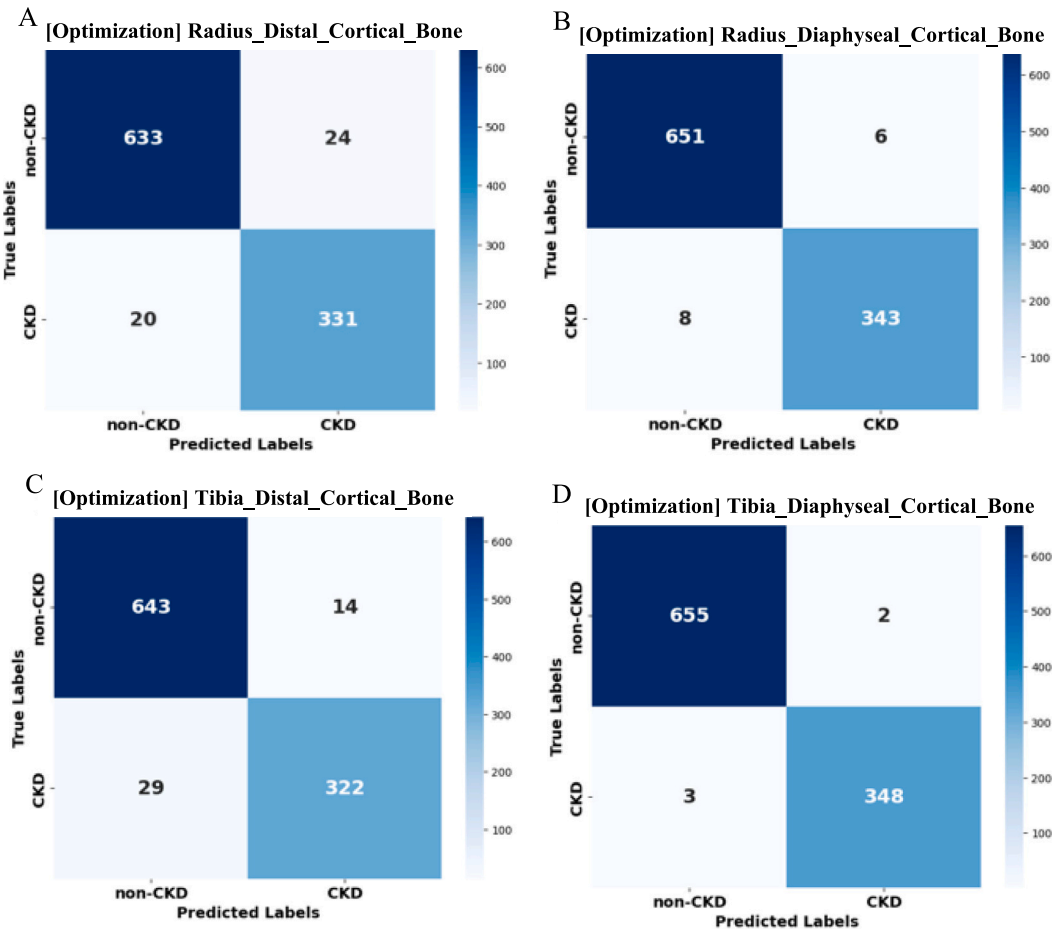


Fig. 3. Confusion matrix for binary (non-CKD, CKD) classification using the optimized XGBoost classification models in the study cohort. The classification model was based on 5-fold validation. Confusion matrices are shown for the (A) distal radius, (B) diaphyseal radius, (C) distal tibia, and (D) diaphyseal tibia sites. The number within each box represents the number of images classified.

thicker, more robust cortices. Notably, the misclassified images do not originate from a single individual but from several individuals within each group.

3.4. Evaluation of independent validation dataset

Supplemental Table 4 shows the results of predicting independent validation datasets using XGBoost based on hyperparameter optimization, achieving the highest evaluation scores at the diaphyseal tibia. Supplemental Fig. 8 and 9 display the confusion matrix and AUC-ROC

scores for binary (non-CKD, CKD) classification, respectively. Tibial locations performed better than radial locations. In particular, Model 4 accurately predicted nearly 4028 CKD images out of 4032 images (24 CKD patients) in the independent validation dataset. Compared to the test set, all performance metrics on the independent validation set dropped slightly, which could be due to variations in dataset characteristics or differences in texture image patterns of individuals. Finally, sensitivity and specificity were similar between the independent validation dataset and the study test dataset (Supplemental Table 5).

4. Discussion

Chronic kidney disease-related bone changes include changes in bone quality and can overlap with normal variations in non-CKD individuals, particularly when BMD is an outcome (Haarhaus et al., 2023), necessitating novel imaging techniques. Our two-stage integrated approach uses HRpQCT volumes from CKD patients and age/sex-matched non-CKD individuals. The first stage could overcome the challenges of automatically separating cortical from trabecular bone in cases of thin and trabeculated endosteal surfaces (CKD), resulting in a segmentation pipeline that required no manual correction. When analyzing the cortical bone of our pilot cohort, no singular standard HRpQCT cortical outcome showed significant differences between groups, even when limited to stage 5D patients. Implementing the second stage of our model, which leveraged textural features and thus maximized the image data from the HRpQCT, enabled differentiation of cortical bone in CKD vs. non-CKD individuals, surpassing conventional imaging outcomes. This proof-of-concept study provides important insights into the potential of machine learning to detect subtle differences in cortical bone, surpassing standard imaging outcomes, even in a clinically heterogeneous cohort.

Optimizing HRpQCT segmentation has been the focus of several groups, including the introduction of a Laplace–Hamming binarization approach (Sadoughi et al., 2023) trialed in cadaveric bones and a subset of volunteers. The approach improved accuracy and reproducibility by preserving nuanced details lost in bone segmentation, resulting in reduced proportional bias and significantly lower errors in trabecular bone volume fraction (0.06 vs. 0.09, $p < 0.0005$) and trabecular thickness (0.03 vs. 0.07 mm, $p < 0.0001$) compared to the standard manufacturer method. In another work, Neeteson et al. (Neeteson et al., 2023) proposed a 2D U-Net architecture for segmenting five image stacks simultaneously, producing a complete stack of segmented images at distal and tibial sites. The model was trained on a cohort of male and female volunteers, and it showed strong agreement in morphological parameters for >93 % of images when compared to the manufacturer-supplied algorithm. However, neither approach has been tested on extreme cases like CKD, where severely compromised cortical microstructure present unique challenges in bone segmentation.

To the authors' knowledge, this is the first work to present a fully automated algorithm for segmentation of cortical bone at all four commonly used HRpQCT sites which was trained using normative and diseased (CKD) images. In the initial stage of our approach, the incorporation of deep learning enabled our model to discern patterns and variations in bone structure characteristic of CKD, resulting in a robust, fully automated segmentation scheme that required no manual intervention. Specifically, the DSC and mean IoU for the CKD group's diaphyseal sites (radius: DSC = 0.97, Mean IoU = 0.96; tibia: DSC = 0.96, Mean IoU = 0.95) underscore the model precision. Moreover, the consistent accuracy across all groups and the quantified Average HD of 12.70 and 15.20 particularly in the diaphyseal radius further substantiates the reliability of our automated segmentation with the distance of ground truth masks. Despite being based on a limited sample size, the automated segmentation performed exceptionally well, achieving the highest DSC in the diaphyseal radius cortical bone (non-CKD: 0.98, CKD: 0.97) and the lowest DSC in the distal radius cortical bone (non-CKD: 0.88, CKD: 0.88) compared to expert annotations. The disparity in performance metrics between diaphyseal and distal sites can be explained by the dimensions of the trainable area. The diaphyseal sections present a larger training region due to thicker cortices, common to both CKD and non-CKD subjects, as opposed to the more limited cortical areas of the distal bone. The findings on average HD in the study suggest that in patients with CKD, the endosteal and periosteal edges of cortical bone tend to be less smooth due to increased trabecularization and porosity than those without CKD. These changes result in larger HD measurements in CKD patients, reflecting greater irregularity.

While studies have been limited in the context of CKD, the benefit of

HRpQCT, including its ability to distinguish cortical porosity, has been highlighted. Nickolas et al. (Nickolas et al., 2010) showed variations in various standard HRpQCT outcomes at radial and tibial sites among CKD patients with and without prior fractures. However, no HRpQCT outcome differed consistently across all anatomical sites or effectively discriminated fractures based on ROC analysis. In our study, no standard HRpQCT cortical parameter consistently distinguished CKD from non-CKD across all sites, reflecting CKD's heterogeneous impact on the skeleton. To address this, we integrated GLCM texture analysis within our classification framework, complementing standard measurements. Originally introduced by Haralick (Haralick et al., 1973), GLCM effectively extracts statistical data on pixel distribution by analyzing pixel pairs in a set direction and distance. A key finding in our study is that GLCM-based analysis detected differences across all four anatomical sites, unlike standard HRpQCT outcomes.

CKD can adversely affect the cortical matrix, altering the properties of mineral, collagen, and water content in the bone's cortex (Kazama et al., 2015). The GLCM approach potentially captures these local and subtle disruptions to the matrix which are not fully appreciated in standard measurements and go beyond microstructural porosity. For example, despite insignificant differences in porosity as shown by *t*-tests, GLCM effectively identified textural anomalies allowing the model to distinguish between CKD and non-CKD images. Previous applications (Heilbronner et al., 2023; Maciel et al., 2020) of GLCM have predominantly centered on the trabecular region, leaving the cortical area less explored. To our knowledge, this is the first attempt to extract textural features from HRpQCT images of cortical bone. However, this approach has limitations, as it primarily captures second-order features. Future studies may address this by incorporating complementary texture analysis methods, such as CNN-based feature extraction, Local Binary Patterns (LBP), or Histogram of Oriented Gradients (HOG), to capture more complex, high-level structural features, potentially enriching our understanding of CKD-related bone morphology.

The classification performance of our XGBoost models, optimized through hyperparameter adjustments, shows strong results across both the testing set and an independent validation set. When misclassified cases were visually inspected, slices misclassified as 'non-CKD' showed no severe deterioration. In contrast, cases misclassified as 'CKD' showed patterns of trabecularization and cortical bone loss. Despite a slight decrease in performance metrics in the independent validation set as shown in Table 3 and supplemental Table 4, the models maintained high effectiveness. This indicates robustness and suggests that the superior initial performance in the testing set helped preserve high performance levels even in the validation set. For example, while Model 1 showed a decrease in precision from 0.97 to 0.87 for the non-CKD group from the testing to the validation set, the performance was still substantial. Similarly, Model 4 maintained nearly perfect metrics in both sets, especially in terms of TPR (as sensitivity) of 0.99. The slight drop in performance for the independent validation set may be due to differences in the dataset characteristics or variation in the disease severity's texture image patterns between datasets. Nonetheless, the consistency in high performance across different datasets highlights the models' reliability.

HRpQCT, while not yet FDA-approved for clinical diagnostics at the time of this study, remains a powerful research tool and is actively used in several clinical trials. This is particularly valuable for CKD patients, as standard imaging methods like DXA often fail to capture early bone changes. Our machine learning approach leverages HRpQCT to enhance the analysis of bone images by extracting textural features, improving the detection of CKD-related bone fragility. Integrating AI tools could streamline complex analyses as HRpQCT becomes more widely adopted in clinical settings. Such advancements could support earlier interventions and more targeted treatment plans, ultimately improving clinical outcomes. Future research with larger datasets and additional model refinements may further validate this approach, enhancing its potential to become a mainstay in bone health related assessments.

particularly in cases such as CKD where bone changes encompass both quantity and quality changes in the cortical bone.

5. Limitations

In this proof-of-concept study, we included both stage 3 and stage 5D CKD patients who underwent HRpQCT imaging at our institution. Conventional HRpQCT outcomes were not statistically different between non-CKD patients when excluding the two stage 3 cases or when including all ten CKD cases (stage 3–5D). The classification models were independently validated on a new dataset of non-CKD and CKD patients, distinct from the deep learning model's training set. After approximately 15 months of data collection, we obtained 18 non-CKD and 24 CKD cases for proper model validation. While model precision and internal validation was high, external validation with data from another institution was lacking. Future studies will expand the CKD cohort, incorporating data from multiple research centers to enhance the generalizability and applicability of our segmentation and classification algorithms.

For stage 1 (segmentation), the distal radius showed a lower DSC of 0.88 in both non-CKD and CKD groups, compared to better results at diaphyseal sites. Future model iterations will incorporate more diverse training data to improve segmentation accuracy at the distal radius, emphasizing the structural differences in CKD patients. We utilized expert raters for validation, ensuring a direct assessment of our model's performance. This choice was made because manual adjustments for automatic endosteal contour using HRpQCT are laborious and susceptible to biases (Supplemental Fig. 9), especially when operators lack adequate training or experience. In addition, future work should consider the impact of different architectures such as DeepLabv3+, Mask R-CNN, and Attention U-Net to determine if they differentially capture finer structural details in trabeculated and porous cortical bone regions.

For stage 2 (classification), despite achieving a notably high ROC-AUC value, several cases were still misclassified. To address these issues, future versions will implement advanced feature selection techniques and explore the integration of additional bone biomarkers such as PTH and eGFR to improve classification precision. While our full pipeline included a segmentation and classification stage, future studies should evaluate classification performance using the existing segmentation pipeline from the HRpQCT manufacturer.

6. Conclusion

By implementing an automated system that integrates deep learning and machine learning, we describe a method that accurately segments diseased and normative cortical bone without manual correction and can effectively distinguish diseased cortical bone changes from non-CKD normative bone. The broader implications of this study could be significant, as our integrated network demonstrates precision in classifying the distinctive cortical bone changes associated with CKD (stages 3–5D) and distinguishing them from normal patterns in age and sex-matched otherwise healthy individuals. Ongoing work will evaluate the technique in earlier CKD stages, refine the classification algorithm, and assess its ability to differentiate between CKD stages.

CRedit authorship contribution statement

Youngjun Lee: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Wikum R. Bandara:** Writing – review & editing, Methodology, Formal analysis. **Sangjun Park:** Writing – review & editing, Methodology. **Miran Lee:** Writing – review & editing, Methodology. **Choongboem Seo:** Writing – review & editing, Methodology. **Sunwoo Yang:** Writing – review & editing, Methodology. **Kenneth J. Lim:** Writing – review & editing, Resources, Funding acquisition. **Sharon M. Moe:** Writing – review & editing, Supervision, Resources,

Funding acquisition. **Stuart J. Warden:** Writing – review & editing, Investigation, Funding acquisition, Formal analysis, Data curation. **Rachel K. Surowiec:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation.

Declaration of competing interest

All the authors state that they have no conflicts of interest.

Acknowledgement

We want to thank all the subjects who participated in this study and the Musculoskeletal Function, Imaging, and Tissue (MSK-FIT) Resource Core of the Indiana Center for Musculoskeletal Health for the use of their HRpQCT and their personnel for acquiring the images. We would also like to thank Amir Ali Dehghanpour, Farhan Sadik, Mohseu Rashid Subah and Olayooye Peter A for their help with valuable feedback on this study. My Korean peers (Seokkyoon Hong, Jinheon Jeong, and Doohyeong Jang) and my family always encourage me to excel in my research. Finally, I am always grateful to my PI, Dr. Rachel. This contribution was made possible by support from the National Institutes of Health (NIH/NIAMS P30 AR072581), (LRP 1L30DK130133-0).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bonr.2024.101821>.

Data availability

Data will be made available on request.

References

- Abedalla, A., Abdullah, M., Al-Ayyoub, M., Benkhelifa, E., 2020. The 2ST-UNet for pneumothorax segmentation in chest X-Rays using ResNet34 as a backbone for U-Net. *arXiv preprint arXiv:200902805*.
- Arias-Moreno, A.J., Hosseini, H.S., Bevers, M., Ito, K., Zysset, P., van Rietbergen, B., 2019. Validation of distal radius failure load predictions by homogenized-and micro-finite element analyses based on second-generation high-resolution peripheral quantitative CT images. *Osteoporos. Int.* 30, 1433–1443.
- Baidya Kayal, E., Kandasamy, D., Khare, K., Bakhshi, S., Sharma, R., Mehndiratta, A., 2021. Texture analysis for chemotherapy response evaluation in osteosarcoma using MR imaging. *NMR Biomed.* 34 (2), e4426. Feb. Epub 20201020.
- Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al., 2015. Xgboost: extreme gradient boosting. *R package version 04-2*, 1(4), pp. 1–4.
- Davis, K.A., Burghardt, A.J., Link, T.M., Majumdar, S., 2007. The effects of geometric and threshold definitions on cortical bone metrics assessed by in vivo high-resolution peripheral quantitative computed tomography. *Calcif. Tissue Int.* 81, 364–371.
- DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 837–45.
- Figueiredo, C.P., Kleyer, A., Simon, D., Stemmler, F., d'Oliveira, I., Weissenfels, A., et al., 2018. Methods for segmentation of rheumatoid arthritis bone erosions in high-resolution peripheral quantitative computed tomography (HR-pQCT). In: *Seminars in arthritis and rheumatism*. Elsevier, pp. 611–618.
- Ganeshan, B., Miles, K.A., 2013. Quantifying tumour heterogeneity with CT. *Cancer Imaging* 13 (1), 140–149. Mar 26. Epub 20130326.
- Haarhaus, M., Aaltonen, L., Cejka, D., Cozzolino, M., de Jong, R.T., d'Haese, P., et al., 2023. Management of fracture risk in CKD—traditional and novel approaches. *Clin. Kidney J.* 16 (3), 456–472.
- Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* 6, 610–621.
- Heilbronner, A.K., Koff, M.F., Breighner, R., Kim, H.J., Cunningham, M., Lebl, D.R., et al., 2023. Opportunistic evaluation of trabecular bone texture by MRI reflects bone mineral density and microarchitecture. *J. Clin. Endocrinol. Metabol.* 108 (8) e557–e66.
- Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J., 1993. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* 15 (9), 850–863.
- Kazama, J.J., Matsuo, K., Iwasaki, Y., Fukagawa, M., 2015. Chronic kidney disease and bone metabolism. *J. Bone Miner. Metab.* 33 (3), 245–252, 2015/05/01.

- Leonard, M.B., Bachrach, L.K., 2001. Assessment of bone mineralization following renal transplantation in children: limitations of DXA and the confounding effects of delayed growth and development. *Am. J. Transplant.* 1 (3), 193–196.
- Li, C., Jin, D., Chen, C., Letuchy, E.M., Janz, K.F., Burns, T.L., et al., 2015. Automated cortical bone segmentation for multirow-detector CT imaging with validation and application to human studies. *Med. Phys.* 42 (8), 4553–4565.
- Lu, S., Fuggle, N.R., Westbury, L.D., Breasail, M.O., Bevilacqua, G., Ward, K.A., et al., 2023. Machine learning applied to HR-pQCT images improves fracture discrimination provided by DXA and clinical risk factors. *Bone* 168, 116653.
- Maciel, J.G., Araújo, Imd, Trazzi, L.C., Azevedo-Marques, Pmd, Salmon, C.E.G., Paula, FjAd, et al., 2020. Association of bone mineral density with bone texture attributes extracted using routine magnetic resonance imaging. *Clinics* 75.
- Neeteson, N.J., Besler, B.A., Whittier, D.E., Boyd, S.K., 2023. Automatic segmentation of trabecular and cortical compartments in HR-pQCT images using an embedding-predicting U-Net and morphological post-processing. *Sci. Rep.* 13 (1), 252.
- Newman, C.L., Moe, S.M., Chen, N.X., Hammond, M.A., Wallace, J.M., Nyman, J.S., et al., 2014. Cortical bone mechanical properties are altered in an animal model of progressive chronic kidney disease. *PLoS One* 9 (6), e99262.
- Nickolas, T.L., Leonard, M.B., Shane, E., 2008. Chronic kidney disease and bone fracture: a growing concern. *Kidney Int.* 74 (6), 721–731.
- Nickolas, T.L., Stein, E., Cohen, A., Thomas, V., Staron, R.B., McMahon, D.J., et al., 2010. Bone mass and microarchitecture in CKD patients with fracture. *J. Am. Soc. Nephrol.* 21 (8), 1371–1380.
- Nickolas, T.L., Stein, E.M., Dworakowski, E., Nishiyama, K.K., Komandah-Kosseh, M., Zhang, C.A., et al., 2013. Rapid cortical bone loss in patients with chronic kidney disease. *J. Bone Miner. Res.* 28 (8), 1811–1820.
- Nishiyama, K.K., Macdonald, H.M., Buie, H.R., Hanley, D.A., Boyd, S.K., 2010. Postmenopausal women with osteopenia have higher cortical porosity and thinner cortices at the distal radius and tibia than women with normal aBMD: an in vivo HR-pQCT study. *J. Bone Miner. Res.* 25 (4), 882–890.
- Ohs, N., Collins, C.J., Touroille, D.C., Atkins, P.R., Schroeder, B.J., Blauth, M., et al., 2021. Automated segmentation of fractured distal radii by 3D geodesic active contouring of in vivo HR-pQCT images. *Bone* 147, 115930.
- Pantic, I., Cunic, J., Dugalic, S., Petroianu, G.A., Corridon, P.R., 2023. Gray level co-occurrence matrix and wavelet analyses reveal discrete changes in proximal tubule cell nuclei after mild acute kidney injury. *Sci. Rep.* 13 (1), 4025, 2023/03/10.
- Pialat, J., Burghardt, A., Sode, M., Link, T., Majumdar, S., 2012. Visual grading of motion induced image degradation in high resolution peripheral computed tomography: impact of image quality on measures of bone density and micro-architecture. *Bone* 50 (1), 111–118.
- Piraino, B., Chen, T., Cooperstein, L., Segre, G., Puschett, J., 1988. Fractures and vertebral bone mineral density in patients with renal osteodystrophy. *Clin. Nephrol.* 30 (2), 57–62.
- Ricardo, A.L.F., da Silva, G.A., Ogawa, C.M., Nussi, A.D., De Rosa, C.S., Martins, J.S., et al., 2023. Magnetic resonance imaging texture analysis for quantitative evaluation of the mandibular condyle in juvenile idiopathic arthritis. *Oral Radiol.* 39 (2), 329–340. Apr. (Epub 20220810).
- Sadoughi, S., Subramanian, A., Ramil, G., Burghardt, A.J., Kazakia, G.J., 2023. A laplace-hamming binarization approach for second-generation HR-pQCT rescues fine feature segmentation. *J. Bone Miner. Res.* 38 (7), 1006–1014.
- Seeja, R., Suresh, A., 2019. Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM). *Asian Pacific journal of cancer prevention: APJCP* 20 (5), 1555.
- Sharma, S., Mehta, P., Patil, A., Gupta, S., Rajender, S., Chattopadhyay, N., 2022. Meta-analyses of the quantitative computed tomography data in dialysis patients show differential impacts of renal failure on the trabecular and cortical bones. *Osteoporos. Int.* 33 (7), 1521–1533.
- Surowiec, R.K., Swallow, E.A., Warden, S.J., Allen, M.R., 2022. Tracking changes of individual cortical pores over 1 year via HR-pQCT in a small cohort of 60-year-old females. *Bone Reports* 17, 101633.
- Tsuji, K., Kitamura, M., Chiba, K., Muta, K., Yokota, K., Okazaki, N., et al., 2022. Comparison of bone microstructures via high-resolution peripheral quantitative computed tomography in patients with different stages of chronic kidney disease before and after starting hemodialysis. *Ren. Fail.* 44 (1), 381–391.
- Urena, P., Bernard-Poenaru, O., Ostertag, A., Baudoin, C., Cohen-Solal, M., Cantor, T., et al., 2003. Bone mineral density, biochemical markers and skeletal fractures in haemodialysis patients. *Nephrology Dialysis Transplantation* 18 (11), 2325–2331.
- Valentinitsch, A., Patsch, J.M., Deutschmann, J., Schueller-Weidekamm, C., Resch, H., Kainberger, F., et al., 2012. Automated threshold-independent cortex segmentation by 3D-texture analysis of HR-pQCT scans. *Bone* 51 (3), 480–487.
- Warden, S.J., Liu, Z., Fuchs, R.K., van Rietbergen, B., Moe, S.M., 2022a. Reference data and calculators for second-generation HR-pQCT measures of the radius and tibia at anatomically standardized regions in White adults. *Osteoporos. Int.* 1–16.
- Warden, S.J., Liu, Z., Fuchs, R.K., van Rietbergen, B., Moe, S.M., 2022b. Reference data and calculators for second-generation HR-pQCT measures of the radius and tibia at anatomically standardized regions in White adults. *Osteoporos. Int.* 33 (4), 791–806. Apr. (Epub 20210929).
- Whittier, D., Mudryk, A., Vandergaag, I., Burt, L., Boyd, S., 2020a. Optimizing HR-pQCT workflow: a comparison of bias and precision error for quantitative bone analysis. *Osteoporos. Int.* 31, 567–576.
- Whittier, D.E., Boyd, S.K., Burghardt, A.J., Paccou, J., Ghasem-Zadeh, A., Chapurlat, R., et al., 2020b. Guidelines for the assessment of bone density and microarchitecture in vivo using high-resolution peripheral quantitative computed tomography. *Osteoporos. Int.* 31, 1607–1627.
- Wong, T.-T., Yeh, P.-Y., 2019. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans. Knowl. Data Eng.* 32 (8), 1586–1594.
- Zulpe, N., Pawar, V., 2012. GLCM textural features for brain tumor classification. *International Journal of Computer Science Issues (IJCSI)* 9 (3), 354.