OPEN ACCESS

University of BRISTOL

Lu, S., Fuggle, N. R., Westbury, L. D., Ó Breasail, M., Bevilacqua, G., Ward, K. A., Dennison, E. M., Mahmoodi, S., Niranjan, M., & Cooper, C. (2023). Machine learning applied to HR-pQCT images improves fracture discrimination provided by DXA and clinical risk factors. *Bone*, *168*, [116653]. https://doi.org/10.1016/j.bone.2022.116653

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
10.1016/j.bone.2022.116653

Link to publication record in Explore Bristol Research
PDF-document

This is the final published version of the article (version of record). It first appeared online via Elsevier at https://doi.org/10.1016/j.bone.2022.116653.Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/

Bone 168 (2023) 116653

Contents lists available at ScienceDirect

Bone

journal homepage: www.elsevier.com/locate/bone

Full Length Article

Machine learning applied to HR-pQCT images improves fracture discrimination provided by DXA and clinical risk factors

Shengyu Lu [a,1], Nicholas R. Fuggle [b,c,1], Leo D. Westbury [b], Mícheál Ó Breasail [d], Gregorio Bevilacqua [b], Kate A. Ward [b,e], Elaine M. Dennison [b,f], Sasan Mahmoodi [a,2], Mahesan Niranjan [a,2], Cyrus Cooper [b,e,g,*,2]

[a] Faculty of Engineering and Physical Sciences, Electronics and Computer Science, University of Southampton, UK
[b] MRC Lifecourse Epidemiology Centre, University of Southampton, Southampton, UK
[c] The Alan Turing Institute, London, UK
[d] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[e] NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK
[f] Victoria University of Wellington, Wellington, New Zealand
[g] NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK

ARTICLE INFO

Keywords:
Osteoporosis (OP)
Fracture risk
High-resolution peripheral quantitative computed tomography (HR-pQCT)
Machine learning
Computer vision
Bone microarchitecture

ABSTRACT

Background: Traditional analysis of High Resolution peripheral Quantitative Computed Tomography (HR-pQCT) images results in a multitude of cortical and trabecular parameters which would be potentially cumbersome to interpret for clinicians compared to user-friendly tools utilising clinical parameters. A computer vision approach (by which the entire scan is 'read' by a computer algorithm) to ascertain fracture risk, would be far simpler. We therefore investigated whether a computer vision and machine learning technique could improve upon selected clinical parameters in assessing fracture risk.

Methods: Participants of the Hertfordshire Cohort Study (HCS) attended research visits at which height and weight were measured; fracture history was determined via self-report and vertebral fracture assessment. Bone microarchitecture was assessed via HR-pQCT scans of the non-dominant distal tibia (Scanco XtremeCT), and bone mineral density measurement and lateral vertebral assessment were performed using dual-energy X-ray absorptiometry (DXA) (Lunar Prodigy Advanced). Images were cropped, pre-processed and texture analysis was performed using a three-dimensional local binary pattern method. These image data, together with age, sex, height, weight, BMI, dietary calcium and femoral neck BMD, were used in a random-forest classification algorithm. Receiver operating characteristic (ROC) analysis was used to compare fracture risk identification methods.

Results: Overall, 180 males and 165 females were included in this study with a mean age of approximately 76 years and 97 (28 %) participants had sustained a previous fracture. Using clinical risk factors alone resulted in an area under the curve (AUC) of 0.70 (95 % CI: 0.56–0.84), which improved to 0.71 (0.57–0.85) with the addition of DXA-measured BMD. The addition of HR-pQCT image data to the machine learning classifier with clinical risk factors and DXA-measured BMD as inputs led to an improved AUC of 0.90 (0.83–0.96) with a sensitivity of 0.83 and specificity of 0.74.

Conclusion: These results suggest that using a three-dimensional computer vision method to HR-pQCT scanning may enhance the identification of those at risk of fracture beyond that afforded by clinical risk factors and DXA-measured BMD. This approach has the potential to make the information offered by HR-pQCT more accessible (and therefore) applicable to healthcare professionals in the clinic if the technology becomes more widely available.

* Corresponding author at: MRC Lifecourse Epidemiology Centre, University of Southampton, Southampton, UK.
E-mail addresses: Shengyu.Lu@soton.ac.uk (S. Lu), nrf@mrc.soton.ac.uk (N.R. Fuggle), lw@mrc.soton.ac.uk (L.D. Westbury), micheal.obreasail@bristol.ac.uk (M. Ó Breasail), gb@mrc.soton.ac.uk (G. Bevilacqua), kw@mrc.soton.ac.uk (K.A. Ward), emd@mrc.soton.ac.uk (E.M. Dennison), sm3@ecs.soton.ac.uk (S. Mahmoodi), mn@ecs.soton.ac.uk (M. Niranjan), cc@mrc.soton.ac.uk (C. Cooper).
[1] Joint first authors.
[2] Joint senior authors.

https://doi.org/10.1016/j.bone.2022.116653
Received 6 May 2022; Received in revised form 16 December 2022; Accepted 21 December 2022
Available online 27 December 2022
8756-3282/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Osteoporosis is characterised by a reduction in bone mineral density leading to a predisposition to fracture [1] and is associated with substantial morbidity and mortality [2,3]. The global prevalence of individuals at high risk of fragility fracture is >158 million and is set to double by the year 2040 [4]. This will see economic costs associated with osteoporotic fractures rising beyond a baseline level of €37 billion in the European Union in 2010 [5]. However, identifying those at high risk of fractures means that they can be treated with effective medications to reduce their fracture risk [6] and improve outcomes.

Traditionally fracture risk prediction to target preventative measures has rested upon clinical risk factors and bone mineral density [7–9]. More recently HR-pQCT has been proposed as an alternative method to assess fracture risk and we have demonstrated bone microarchitecture phenotypes associated with high risk of fracture [10,11] which suggests that this imaging modality might help predict fracture occurrence. However, HR-pQCT results in a large number of variables and there is no practicable way in which these can be adequately integrated into a convenient fracture risk assessment tool. Novel statistical methods and computer science techniques have the potential to assist.

Machine learning has become increasingly popular because it can automatically learn the features from current instances and automatically provide prediction for new cases [12]. Machine learning methods have demonstrated success in many medical tasks [13,14]; some researchers have proposed the use of computer-assisted diagnostic algorithms to diagnose osteoporosis or predict fracture risk [15–18]. However, in the field of imaging, these studies have been limited to specific imaging features (for example Finite Element Analysis) [19] or conventional computed tomography images [20,21] and have not utilised the detailed, textural information on bone microarchitecture which can be gleaned from HR-pQCT via computer vision.

This study proposes a method that automatically extracts texture features from HR-pQCT images and combines this information with data on clinical variables and femoral neck BMD to identify previous fractures. Therefore, the aim of this study was to deploy a computer vision approach to HR-pQCT images in order to predict those at risk of fracture and to compare the diagnostic performance of this approach against the traditional methods of clinical risk factors and femoral neck DXA.

2. Methods

2.1. The Hertfordshire Cohort Study

The Hertfordshire Cohort Study (HCS) comprises 2997 men and women born in Hertfordshire from 1931 to 39 and who still lived there in 1998–2004. In 2004, 642 of the 966 participants from East Hertfordshire were recruited to a musculoskeletal follow-up study. In 2011-12, 376/642 participated in a further bone follow-up study. The HCS has been described in detail previously [22,23].

2.2. Ascertainment of participant characteristics

Dietary calcium intake was ascertained in 1998–2004 using a food-frequency questionnaire [24]; all characteristics stated below were ascertained in 2011–2012. Height was measured (wall mounted SECA stadiometer) along with weight (calibrated SECA 770 digital floor scales, SECA Ltd., Hamburg) and used to derive BMI. Fractures since aged 45 years were ascertained through self-report. Morphometric vertebral fractures were diagnosed from a lateral spine view imaged using a Lunar Prodigy Advance DXA scanner (GE Medical Systems) and graded based on the Genant semi-quantitative method [25]. Participants with a vertebral or self-reported fracture were regarded as having had a previous fracture. HR-pQCT scans (XtremeCTi Scanco Medical AG, Switzerland) of the non-dominant distal radius and tibia were performed; dominant limbs were scanned if the non-dominant limb had fractured. 110 parallel CT slices were obtained, representing a volume of bone 9 mm in axial length with a nominal resolution (voxel size) of 82 μm. The scan protocol was in accordance with manufacturer's guidelines and as described by Boutroy et al. [26]. Using the method of Pauchard and colleagues [27], scans with excessive motion artefact (grade 5) were excluded. Manufacturer standard evaluation and cortical porosity scripts were used for image analysis [28]. Extended cortical analysis was performed for all scans [29].

2.3. Additional image processing

The KHKs MicroCT tool [30] was used to crop the radius and tibia in the image and remove the fibula and surrounding soft tissues. There were 396 HR-pQCT images (205 radius scans and 191 tibia scans). Images were archived in HDF5 format for further analysis. A 3D local binary patterns method was then implemented to extract 3D texture features from images using textural descriptors [31]; further details are included in Supplementary Material. The texture features of each image were represented as a vector containing 352 entries [32].

2.4. Classification algorithm to predict fracture

A random forest classifier was implemented to group participants according to whether or not they had experienced a previous fracture; further details of this algorithm are included in Supplementary Material. A sample of 22 consecutive slices was selected for each HR-pQCT image. There were fewer participants with previous fractures compared to those without. Therefore, an oversampling strategy was used for individuals with previous fractures [33]. Samples for training the random forest classifier and for validation (testing) were selected randomly with a ratio of 4:1, and five-fold cross-validation was implemented to evaluate performance. All analyses to assess the diagnostic performance of the random forest classifier were based on the validation dataset.

Discrete values of variables used in the random forest classifier were first normalised to the [0,1] range. Information on the six clinical factors considered in the random forest classifier (age, sex, height, weight, BMI, and dietary calcium) was stored as a vector containing 6 entries; femoral neck BMD was represented as a vector containing one entry. Vectors of samples capturing image, clinical and BMD data were combined and then processed using the random forest classifier to generate fracture probabilities for samples from each participant, as shown in Fig. 1. The highest sample probability from each participant was selected as the participant's predicted fracture probability. Participants in the validation dataset were estimated to have had a previous fracture if their predicted fracture probability was >0.5.

2.5. Statistical methods

Participant characteristics were described using summary statistics. The receiver operator characteristic (ROC) curve, sensitivity, specificity and the area under the curve (AUC) were used to assess the diagnostic capability of the random forest classifier regarding previous fracture [34]. The diagnostic capacity of the classifier with different combinations of input information (tibial HR-pQCT image data, clinical risk factors and BMD) was compared. The clinical risk factors considered included age, sex, height, weight, BMI, and dietary calcium. Of particular interest was whether the AUC for the classifier that contained HR-pQCT image data, clinical factors and BMD was substantially greater than the AUC for the classifier that was only based on clinical factors and BMD. Therefore, the statistical significance of the difference in these two AUCs was examined. Sensitivity analyses included stratifying analysis by sex, analysing distal radius scans instead of distal tibial scans, and including additional clinical factors (smoking history, alcohol consumption, physical activity, bisphosphonate usage, number of comorbidities, and occupational social class). The ascertainment of these additional clinical factors in HCS has been described previously [35].

Bone microarchitecture variables have been previously demonstrated to relate to fracture risk independently of DXA-measured areal BMD [36].

Odds ratios for previous fracture according to the clinical risk factors and BMD were estimated in the validation dataset using logistic regression. For comparison with the random forest classifier, the AUC for the mutually-adjusted logistic regression model with previous fracture as the outcome and these covariates included simultaneously as exposures was calculated. A significance level of 0.05 was used; $p < 0.05$ was considered statistically significant. The analysis sample comprised the 345 participants with data on previous fractures. Python 3.7 was used to extract multimodal features from participants and train the random forest classifier. All statistical analysis for predicted results was implemented in R, version 4.0. All analyses were performed on Intel (R) Core (TM) i5-6600 CPU 3.30GHz with HD Graphics 530.

3. Results

3.1. Participant characteristics

Participant characteristics in 2011–2012 are illustrated in Table 1. Overall, 180 males and 165 females were included in this study; all participants were over 72 years old. Participants with previous fractures (46 males, 51 females) had lower dietary calcium intakes and BMD values than those without fractures (134 males, 114 females). In addition, fractures occurred more frequently in women and among older participants, consistent with previous studies [37].

3.2. Accuracy of the random forest classifier regarding previous fracture

The sensitivity, specificity and AUC (95 % CI) values from the random forest classifier are presented in Table 2, according to the participant input information used; the corresponding ROC curves are shown in Fig. 2. When only femoral neck BMD was used as the input, the AUC (95 % CI) was 0.66 (0.59–0.73); this increased to 0.71 (0.57–0.85) with the addition of clinical data and to 0.90 (0.83–0.96) with the addition of clinical data and tibial HR-pQCT image data. Furthermore, there was a statistically significant difference between the AUCs obtained when tibial HR-pQCT image data was added as an input to the classifier based only on BMD and clinical data ($p < 0.02$). Compared with the use of BMD only (AUC: 0.66) and clinical data only (AUC: 0.70), tibial HR-pQCT image data in isolation demonstated a higher classification accuracy (AUC: 0.86). Although classifiers based on clinical data only and on a combination of clinical data and BMD had the highest specificity of 0.94, the sensitivity of these classifiers was low (<0.25).

Table 1
Participant characteristics in 2011–2012 according to fracture status.

Characteristics	Mean (standard deviation), n(%)		
	No previous fracture (n = 248)	Previous fracture (n = 97)	P-value
Age (years)	76.2 (2.6)	76.9 (2.6)	0.10
Height (cm)	167.0 (9.0)	167.1 (10.2)	0.31
Weight (kg)	77.0 (13.4)	76.9 (14.7)	0.12
BMI (kg/m^2)	27.6 (4.2)	27.5 (4.4)	0.26
Weekly dietary calcium intake (mg)[a]	8326.0 (2312.1)	8030.2 (2840.7)	0.04
Sex (females)	114 (46 %)	51 (53 %)	<0.001
Femoral neck BMD (g/cm^2)	0.90 (0.14)	0.84 (0.15)	<0.001

Participants with a vertebral fracture or a self-reported fracture since age 45 years were regarded as having had a previous fracture.
P-values for differences between participants with and without previous fractures were calculated using *t*-tests and chi-squared tests as appropriate.
[a] Ascertained in 1998–2004.

The AUC (95 % CI) for the mutually-adjusted logistic regression model with previous fracture as the outcome and age, sex, height, weight, BMI, dietary calcium and femoral neck BMD as exposures was similar to the AUC of the random forest classifier which was based on this same input data (0.75 (0.63, 0.86) vs 0.71 (0.57–0.85)) (Supplementary Table A). When included separately as individual exposures, only weight and BMI were significantly associated with previous fracture (Supplementary Table A).

3.3. Sensitivity analyses

Similar results were observed in the following scenarios: stratification of analyses by sex (Table 3); analysis of HR-pQCT scans of both the distal radius and tibia (data not shown); and analysis of HR-pQCT scans of the distal radius alone (data not shown). In all these scenarios, the addition of HR-pQCT image data resulted in a statistically significant ($p < 0.05$) improvement in fracture classification as measured using AUCs

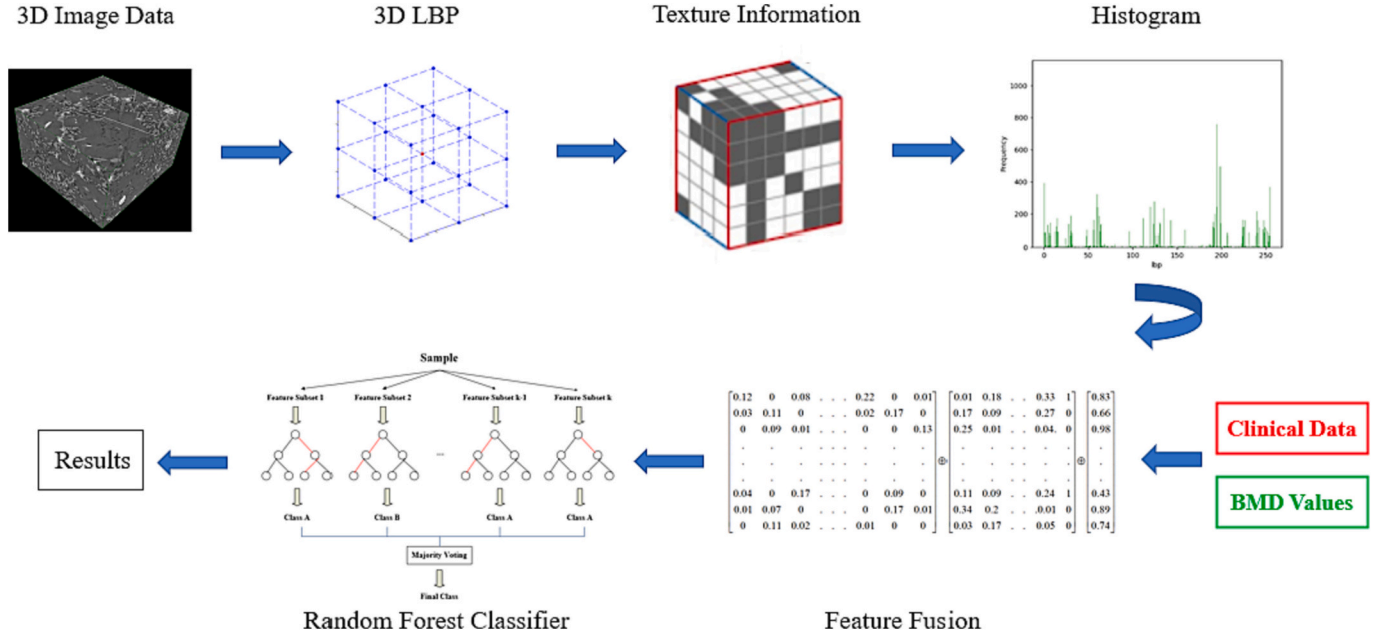

Fig. 1. A combination of HR-pQCT image data, clinical data and DXA BMD data was used in the machine learning algorithm.

Table 2
Diagnostic performance of the random forest classifier for previous fracture according to input data.

Input data	AUC (95 % CI)	Sensitivity	Specificity
Only BMD	0.66 (0.59–0.73)	0.30	0.86
Only clinical data	0.70 (0.56–0.84)	0.19	**0.94**
Only HR-pQCT image data	0.86 (0.78–0.94)	**0.87**	0.70
Clinical data and BMD	0.71 (0.57–0.85)	0.24	**0.94**
HR-pQCT image and clinical data	0.88 (0.81–0.96)	0.80	0.76
HR-pQCT image data, clinical data and BMD	**0.90 (0.83–0.96)**	0.83	0.74

The highest values in each column are highlighted in bold.
BMD: Femoral neck BMD.
HR-pQCT: high-resolution peripheral quantitative computed tomography.
Clinical data included age, sex, height, weight, BMI and dietary calcium intake.
HR-pQCT image data from tibia scans were used.
A probability threshold of 0.5 was used to estimate the sensitivity and specificity.

compared to classifiers that only used clinical data and BMD as inputs. No improvement in diagnostic performance was observed when additional clinical risk factors (smoking history, alcohol consumption, physical activity, bisphosphonate usage, number of comorbidities, and occupational social class) were included in the main sex-pooled analysis that utilised tibial scans.

4. Discussion

In this study, we proposed an automatic diagnostic algorithm which used various inputs (HR-pQCT tibia scan image data, clinical data and DXA femoral neck BMD) to discriminate between people with and without previous fractures. Although based on a small sample, the disgnostic performance of this algorithm was high; the highest AUC (0.90, 95 % CI: 0.83–0.96) was achieved using a combination of image data, clinical data and BMD; inclusion of image data significantly improved fracture discrimination compared to the use of clinical data and BMD only ($p < 0.02$). This may be because these clinical and BMD measures were similar between participants with and without a previous fracture in the analysis sample of this study. Furthermore, the use of tibial image data, clinical data and BMD significantly ($p < 0.05$) improved fracture discrimination compared to the use of any individual tibial HR-pQCT parameter, along with clinical data and BMD, as exposures in a logistic regression model (data not shown). This suggests that valuable information regarding fracture risk is utilised by image processing methods which is not captured by any individual HR-pQCT parameter.

HR-pQCT is an imaging modality which can substantially contribute to the assessment of a patient's bone health. Our previous work has used cluster analysis of HR-pQCT variables to identify two different bone microarchitecture phenotypes associated with fracture [10]. One phenotype demonstrated greater cortical deficiency with lower cortical thickness and cortical bone mineral density, and the other phenotype demonstrated greater trabecular deficiency with lower trabecular density and number than the sex-specific sample mean [10].

The cluster analysis findings were further developed to demonstrate that bone microarchitecture parameters (acquired via HR-pQCT) enhanced fracture prediction compared to areal BMD (of the proximal femur) alone [11]. When finite element analysis was incorporated, although this was found to add little to the fracture risk discrimination of the model, it did assist to identify clusters of features which predisposed to higher fracture risk including one cluster with lower Young modulus, cortical thickness, cortical volumetric bone mineral density and Von Mises stresses (compared to the wider sample) and another cluster in females with greater trabecular separation, lower trabecular volumetric bone mineral density and lower trabecular load (compared to the wider

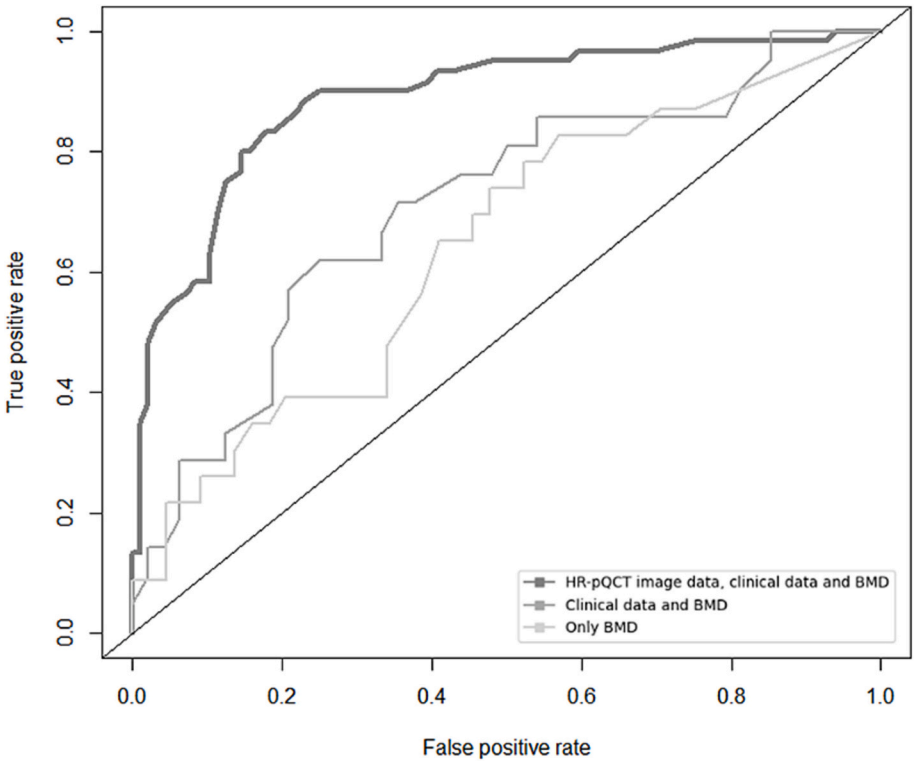

Fig. 2. Receiver operating characteristic curves for previous fracture depending on the input data used in the random forest classifier. HR-pQCT image data from tibia scans were used.

Table 3
Diagnostic performance of the random forest classifier for previous fracture according to input data (sex-specific analyses).

Input data	Males			Females		
	AUC (95 % CI)	Sensitivity	Specificity	AUC (95 % CI)	Sensitivity	Specificity
Only BMD	0.61 (0.52–0.71)	0.14	0.76	0.70 (0.62–0.79)	0.30	**0.82**
Only clinical data	0.70 (0.48–0.92)	0.27	**1.00**	0.70 (0.48–0.92)	0.50	0.70
Only HR-pQCT image data	0.91 (0.84–0.98)	**0.82**	0.88	0.85 (0.75–0.95)	0.81	0.62
Clinical data and BMD	0.71 (0.51–0.91)	0.18	**1.00**	0.71 (0.51–0.91)	0.50	0.74
HR-pQCT image and clinical data	0.91 (0.85–0.98)	**0.82**	0.86	**0.92 (0.84–0.99)**	0.87	0.73
HR-pQCT image data, clinical data and BMD	**0.92 (0.86–0.98)**	0.79	0.88	0.90 (0.82–0.98)	**0.94**	0.62

The highest values in each column are highlighted in bold.
BMD: Femoral neck BMD.
HR-pQCT: high-resolution peripheral quantitative computed tomography.
Clinical data included age, height, weight, BMI and dietary calcium intake.
HR-pQCT image data from tibia scans were used.
A probability threshold of 0.5 was used to estimate the sensitivity and specificity.

sample) [11].

There is growing support for the use of HR-pQCT in clinical fracture prediction with a random effects meta-analysis of individuals aged 10.9–84.7 years showing alterations in bone microarchitecture among individuals who had fractured compared to those who had not [38]. The key parameters identified included radial cortical and trabecular volumetric bone mineral density [38]. This study also lent support to the use of HR-pQCT for individuals in clinical practice as the differences observed in HR-pQCT parameters in relation to fracture status were greater than the differences expected due to the precision error of the instrument [38].

Imaging plays an important role in osteoporosis, with DXA BMD incorporated in the definition of the condition [39] and with the advent of computer vision techniques, a body of image-related machine learning research has started to develop. This has largely centred on using deep learning to assess for osteoporosis on routine CT scans via automated vertebral body segmentation and then training an algorithm to predict areal bone mineral density or a measure of volumetric bone mineral density [20,40,41]. A similar approach has also been used with hip radiographs to assess bone mineral density [42]. The only previous work using HR-pQCT utilised radial trabecular texture in 18 post-menopausal women with fragility fractures and 18 post-menopausal women without fragility fractures (a small number of participants compared to our study cohort) [43]. Our study also leveraged recent computer vision developments in texture analysis.

We employed a 3D local binary patterns (LBP) method to extract the texture density information from HR-pQCT images, making full use of the three-dimensionality of the HR-pQCT images and the bone tissue they depict. Two-dimensional textural analysis is used in clinical practice in the form of Trabecular Bone Scores (TBS) on lumbar spine DXA images [44]. However, these lose some of the information provided by the surrounding pixels [45,46]. To address this issue, we developed a method that constructs a 3D spatial cube for each voxel in the 3D images to calculate the feature descriptor, containing all the information in surrounding voxels to reflect the texture density information.

This study has some limitations. First, a healthy participant effect is, unsurprisingly, evident in HCS [22] and sample attrition across the various follow-up waves could have resulted in further selection effects. However, the cohort has been shown to be broadly comparable with participants in the nationally representative Health Survey for England [22]. Second, sample size calculations were not performed prior to analyses; analyses were based on all available data. Riley et al. suggest that suitable sample sizes for binary prediction models should be calculated a priori based on the number of participants and outcome events and also on the number of predictor variables considered [47]. However, a retrospective power calculation showed 90 % power to detect a statistically significant (at the 5 % significance level) improvement in AUC from 0.71 to 0.90; this change in AUC was observed in the validation dataset when HR-pQCT image data were added to the random forest classifier with clinical data and BMD as inputs. Third, although random forest classifiers have advantages for analysing small datasets [48] and similar sample sizes have been used in previous musculoskeletal research publications which have implemented this technique [49–52], a major limitation of this study is that the sample size was small (n = 345). Fourth, the data-driven nature of the classification method may result in limited generalisability and reproducibility of these findings. Fifth, previous (rather than incident) fracture history was used as the outcome. Sixth, further information on self-reported fractures, such as their location and type, was not available; at the 2011–2012 follow-up, participants were only asked whether they had broken any bones since age 45 years. Finally, the data were unbalanced which may have affected the performance of the classifier algorithm. For example, resampling techniques for the group of participants with previous fractures was required, otherwise the random forest classifier would be biased.

In conclusion, computer vision and machine learning may provide an exciting opportunity to operationalize a large number of diverse HR-pQCT parameters into a single score which can be readily used in clinical practice to identify individuals at high risk of fracture. This will potentially allow timely treatment and improved clinical outcomes; however, prior to deployment, this work requires replication in other cohorts.

Funding

The Hertfordshire Cohort Study was supported by the following organisations: Medical Research Council; British Heart Foundation; Versus Arthritis UK; International Osteoporosis Foundation; NIHR Southampton Biomedical Research Centre; NIHR Oxford Biomedical Research Centre; University of Southampton. The funders had no involvement in the following: the study design; the collection, analysis and interpretation of data; the writing of the report; and the decision to submit the article for publication.

Ethical approval

This research had ethical approval from the Hertfordshire and Bedfordshire Local Research Ethics Committee and the East and North Hertfordshire Ethical Committees.

Human and animal rights

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Written informed consent was provided by all participants.

CRediT authorship contribution statement

SL Methodology, Formal analysis, Writing - Original Draft; **NRF** Investigation, Writing - Original Draft; **LDW** Methodology, Writing - Review & Editing; **MOB** Writing - Review & Editing; **GB** Investigation, Resources, Writing - Review & Editing; **KAW** Conceptualization, Writing - Review & Editing, Project administration; **EMD** Writing - Review & Editing, Project administration; **SM** Conceptualization, Writing - Review & Editing, Supervision, Project administration; **MN** Conceptualization, Writing - Review & Editing, Supervision, Project administration; **CC** Conceptualization, Writing - Review & Editing, Project administration. All authors made substantial contributions to the manuscript and approved the final version.

Declaration of competing interest

CC reports personal fees (outside the submitted work) from Amgen, Danone, Eli Lilly, GSK, Kyowa Kirin, Medtronic, Merck, Nestle, Novartis, Pfizer, Roche, Servier, Shire, Takeda and UCB. EMD has received consultancy and speaker fees (outside the submitted work) from Pfizer, UCB, Viatris and Lilly. NRF has received travel bursaries from Pfizer and Eli Lilly (outside the submitted work). LDW, SL, MOB, GB, KAW, SM and MN declare that they have no conflicts of interest.

Data availability

Hertfordshire Cohort Study data are accessible via collaboration. Initial enquires should be made to CC (Principal Investigator). Potential collaborators will be sent a collaborators' pack and asked to submit a detailed study proposal to the HCS Steering Group.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.bone.2022.116653.

References

[1] C. Christodoulou, C. Cooper, What is osteoporosis? Postgrad. Med. J. 79 (2003) 133–138.

[2] M. Katsoulis, V. Benetou, T. Karapetyan, et al., Excess mortality after hip fracture in elderly persons from Europe and the USA: the CHANCES project, J. Intern. Med. 281 (2017) 300–310.

[3] P. Haentjens, J. Magaziner, C.S. Colón-Emeric, D. Vanderschueren, K. Milisen, B. Velkeniers, S. Boonen, Meta-analysis: excess mortality after hip fracture among older women and men, Ann. Intern. Med. 152 (2010) 380–390.

[4] A. Oden, E.V. McCloskey, J.A. Kanis, N.C. Harvey, H. Johansson, Burden of high fracture probability worldwide: secular increases 2010–2040, Osteoporos. Int. 26 (2015) 2243–2248.

[5] E. Hernlund, A. Svedbom, M. Ivergard, J. Compston, C. Cooper, J. Stenmark, E. V. McCloskey, B. Jonsson, J.A. Kanis, Osteoporosis in the European Union: medical management, epidemiology and economic burden. A report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA), Arch. Osteoporos. 8 (2013) 136.

[6] M. Hoff, E. Skovlund, H.E. Meyer, A. Langhammer, A.J. Søgaard, U. Syversen, K. Holvik, B. Abrahamsen, B. Schei, Does treatment with bisphosphonates protect against fractures in real life? The HUNT study, Norway, Osteoporos. Int. 32 (2021) 1395–1404.

[7] J.A. Kanis, O. Johnell, A. Oden, H. Johansson, E. McCloskey, FRAX and the assessment of fracture probability in men and women from the UK, Osteoporos. Int. 19 (2008) 385–397.

[8] N.D. Nguyen, S.A. Frost, J.R. Center, J.A. Eisman, T.V. Nguyen, Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks, Osteoporos. Int. 19 (2008) 1431–1444.

[9] J. Hippisley-Cox, C. Coupland, Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores, BMJ 339 (2009), b4229.

[10] M.H. Edwards, D.E. Robinson, K.A. Ward, M.K. Javaid, K. Walker-Bone, C. Cooper, E.M. Dennison, Cluster analysis of bone microarchitecture from high resolution peripheral quantitative computed tomography demonstrates two separate phenotypes associated with high fracture risk in men and women, Bone 88 (2016) 131–137.

[11] L.D. Westbury, C. Shere, M.H. Edwards, C. Cooper, E.M. Dennison, K.A. Ward, Cluster analysis of finite element analysis and bone microarchitectural parameters identifies phenotypes with high fracture risk, Calcif. Tissue Int. 105 (2019) 252–262.

[12] G.A. Kaissis, M.R. Makowski, D. Rückert, R.F. Braren, Secure, privacy-preserving and federated machine learning in medical imaging, Nat.Mach.Intell. 2 (2020) 305–311.

[13] E. Tjoa, C. Guan, A survey on explainable artificial intelligence (XAI): toward medical XAI, IEEE Trans.Neural Netw.Learn.Syst. 32 (2021) 4793–4813.

[14] M. de Bruijne, Machine learning approaches in medical image analysis: from detection to diagnosis, Med. Image Anal. 33 (2016) 94–97.

[15] I.M. Wani, S. Arora, Computer-aided diagnosis systems for osteoporosis detection: a comprehensive survey, Med. Biol. Eng. Comput. 58 (2020) 1873–1917.

[16] A.S. Cruz, H.C. Lins, R.V.A. Medeiros, J.M.F. Filho, S.G. da Silva, Artificial intelligence on the identification of risk groups for osteoporosis, a general review, Biomed. Eng. Online 17 (2018), 12-12.

[17] C. Kruse, P. Eiken, P. Vestergaard, Machine learning principles can improve hip fracture prediction, Calcif. Tissue Int. 100 (2017) 348–360.

[18] N. Kilic, E. Hosgormez, Automatic estimation of osteoporotic fracture cases by using ensemble learning approaches, J. Med. Syst. 40 (2016) 61.

[19] K.K. Nishiyama, M. Ito, A. Harada, S.K. Boyd, Classification of women with and without hip fracture based on quantitative computed tomography and finite element analysis, Osteoporos. Int. 25 (2014) 619–626.

[20] A. Valentinitsch, S. Trebeschi, J. Kaesmacher, C. Lorenz, M.T. Löffler, C. Zimmer, T. Baum, J.S. Kirschke, Opportunistic osteoporosis screening in multi-detector CT images via local classification of textures, Osteoporos. Int. 30 (2019) 1275–1285.

[21] U.J. Muehlematter, M. Mannil, A.S. Becker, K.N. Vokinger, T. Finkenstaedt, G. Osterhoff, M.A. Fischer, R. Guggenberger, Vertebral body insufficiency fractures: detection of vertebrae at risk on standard CT images using texture analysis and machine learning, Eur. Radiol. 29 (2019) 2207–2217.

[22] H.E. Syddall, A. Aihie Sayer, E.M. Dennison, H.J. Martin, D.J. Barker, C. Cooper, Cohort profile: the Hertfordshire cohort study, Int. J. Epidemiol. 34 (2005) 1234–1242.

[23] H.E. Syddall, S.J. Simmonds, S.A. Carter, S.M. Robinson, E.M. Dennison, C. Cooper, The Hertfordshire Cohort Study: an overview, F1000 Res. 8 (2019) 82.

[24] S. Robinson, H. Syddall, K. Jameson, S. Batelaan, H. Martin, E.M. Dennison, C. Cooper, A.A. Sayer, Current patterns of diet in community-dwelling older men and women: results from the Hertfordshire Cohort Study, Age Ageing 38 (2009) 594–599.

[25] H.K. Genant, C.Y. Wu, C. van Kuijk, M.C. Nevitt, Vertebral fracture assessment using a semiquantitative technique, J. Bone Miner. Res. 8 (1993) 1137–1148.

[26] S. Boutroy, M.L. Bouxsein, F. Munoz, P.D. Delmas, In vivo assessment of trabecular bone microarchitecture by high-resolution peripheral quantitative computed tomography, J. Clin. Endocrinol. Metab. 90 (2005) 6508–6515.

[27] Y. Pauchard, A.M. Liphardt, H.M. Macdonald, D.A. Hanley, S.K. Boyd, Quality control for bone quality parameters affected by subject motion in high-resolution peripheral quantitative computed tomography, Bone 50 (2012) 1304–1310.

[28] E. Biver, C. Durosier-Izart, T. Chevalley, B. van Rietbergen, R. Rizzoli, S. Ferrari, Evaluation of radius microstructure and areal bone mineral density improves fracture prediction in postmenopausal women, J. Bone Miner. Res. 33 (2018) 328–337.

[29] A.J. Burghardt, H.R. Buie, A. Laib, S. Majumdar, S.K. Boyd, Reproducibility of direct quantitative measures of cortical bone microarchitecture of the distal radius and tibia by HR-pQCT, Bone 47 (2010) 519–528.

Karl-Heinz Kunzelmann. http://www.kunzelmann.de/4_software-imagej-import-export-utilities.html, 2012. (Accessed 19 December 2022).

[31] Y. Almakady, S. Mahmoodi, J. Conway, M. Bennett, Rotation invariant features based on three dimensional Gaussian Markov random fields for volumetric texture classification, Comput. Vis. Image Underst. 194 (2020), 102931.

[32] P. Liu, J.-M. Guo, K. Chamnongthai, H. Prasetyo, Fusion of color histogram and LBP-based features for texture image retrieval and classification, Inf. Sci. 390 (2017) 95–111.

[33] S. Maldonado, J. López, C. Vairetti, An alternative SMOTE oversampling strategy for high-dimensional datasets, Appl. Soft Comput. 76 (2019) 380–389.

[34] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, M. Müller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinforma. 12 (2011) 77.

[35] N.R. Fuggle, L.D. Westbury, G. Bevilacqua, P. Titcombe, M. Ó Breasail, N. C. Harvey, E.M. Dennison, C. Cooper, K.A. Ward, Level and change in bone microarchitectural parameters and their relationship with previous fracture and established bone mineral density loci, Bone 147 (2021), 115937.

[36] E.J. Samelson, K.E. Broe, H. Xu, et al., Cortical and trabecular bone microarchitecture as an independent predictor of incident fracture risk in older women and men in the Bone Microarchitecture International Consortium (BoMIC): a prospective study, Lancet Diabetes Endocrinol. 7 (2019) 34–43.

[37] A.S. Areeckal, M. Kocher, S.D. S, Current and emerging diagnostic imaging-based techniques for assessment of osteoporosis and fracture risk, IEEE Rev. Biomed. Eng. 12 (2019) 254–268.

[38] N. Mikolajewicz, N. Bishop, A.J. Burghardt, et al., HR-pQCT measures of bone microarchitecture predict fracture: systematic review and meta-analysis, J. Bone Miner. Res. 35 (2020) 446–459.

[39] Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group, in: World Health Organ Tech Rep Ser 843, 1994, pp. 1–129.
[40] M.T. Löffler, A. Jacob, A. Scharr, et al., Automatic opportunistic osteoporosis screening in routine CT: improved prediction of patients with prevalent vertebral fractures compared to DXA, Eur. Radiol. 31 (2021) 6069–6077.
[41] Y. Fang, W. Li, X. Chen, K. Chen, H. Kang, P. Yu, R. Zhang, J. Liao, G. Hong, S. Li, Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks, Eur. Radiol. 31 (2020) 1831–1842.
[42] N. Yamamoto, S. Sukegawa, A. Kitamura, et al., Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates, Biomolecules 10 (2020).
[43] A. Valentinitsch, J.M. Patsch, A.J. Burghardt, T.M. Link, S. Majumdar, L. Fischer, C. Schueller-Weidekamm, H. Resch, F. Kainberger, G. Langs, Computational identification and quantification of trabecular microarchitecture classes by 3-D texture analysis-based clustering, Bone 54 (2013) 133–140.
[44] B.C. Silva, W.D. Leslie, H. Resch, O. Lamy, O. Lesnyak, N. Binkley, E.V. McCloskey, J.A. Kanis, J.P. Bilezikian, Trabecular bone score: a noninvasive analytical method based upon the DXA image, J. Bone Miner. Res. 29 (2014) 518–530.
[45] S. Murala, Q.M. Jonathan Wu, Spherical symmetric 3D local ternary patterns for natural, texture and biomedical image indexing and retrieval, Neurocomputing 149 (2015) 1502–1514.
[46] S. Abbasi, F. Tajeripour, Detection of brain tumor in 3D MRI images using local binary patterns and histogram orientation gradient, Neurocomputing 219 (2017) 526–535.
[47] R.D. Riley, K.I. Snell, J. Ensor, D.L. Burke, F.E. Harrell Jr., K.G. Moons, G.S. Collins, Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes, Stat. Med. 38 (2019) 1276–1296.
[48] Y. Qi, Random forest for bioinformatics, in: C. Zhang, Y. Ma (Eds.), Ensemble Machine Learning: Methods And Applications, Springer US, Boston, MA, 2012, pp. 307–323.
[49] D. Hussain, S.M. Han, Computer-aided osteoporosis detection from DXA imaging, Comput. Methods Prog. Biomed. 173 (2019) 87–107.
[50] G. Cuaya-Simbro, A.I. Perez-Sanpablo, E.F. Morales, I. Quiñones Uriostegui, L. Nuñez-Carrera, Comparing machine learning methods to improve fall risk detection in elderly with osteoporosis from balance data, J. Healthc. Eng. 2021 (2021), 8697805.
[51] S.D. Mehta, R. Sebro, Random forest classifiers aid in the detection of incidental osteoblastic osseous metastases in DEXA studies, Int. J. Comput. Assist. Radiol. Surg. 14 (2019) 903–909.
[52] S.S.P. Gornale, R.R. P.U. Manza, Detection of osteoarthritis using knee X-ray image analyses: a machine vision based approach, Int.J.Comput.Applic. 145 (2016) 975–8887.