A novel deep learning method for large-scale analysis of bone marrow adiposity using UK Biobank Dixon MRI data

Computational and Structural Biotechnology Journal 24 (2024) 89–104

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Research article

A novel deep learning method for large-scale analysis of bone marrow adiposity using UK Biobank Dixon MRI data

David M. Morris [a,b,1], Chengjia Wang [a,c,1], Giorgos Papanastasiou [b,d], Calum D. Gray [b], Wei Xu [e,2], Samuel Sjöström [a,3], Sammy Badr [f,g], Julien Paccou [f,h], Scott IK Semple [a,b], Tom MacGillivray [i], William P. Cawthorn [a,*]

[a] University/BHF Centre for Cardiovascular Science, University of Edinburgh, The Queen's Medical Research Institute, Edinburgh BioQuarter, 47 Little France Crescent, Edinburgh EH16 4TJ, UK
[b] Edinburgh Imaging, University of Edinburgh, The Queen's Medical Research Institute, Edinburgh BioQuarter, 47 Little France Crescent, Edinburgh EH16 4TJ, UK
[c] School of Mathematics and Computer Sciences, Heriot-Watt University, Edinburgh EH14 1AS, UK
[d] School of Computer Science and Electronic Engineering, Wivenhoe Park, The University of Essex, Colchester CO4 3SQ, UK
[e] Centre for Global Health, Usher Institute, University of Edinburgh, Edinburgh EH8 9AG, UK
[f] University of Lille, Marrow Adiposity and Bone Laboratory (MABlab) ULR 4490, F-59000 Lille, France
[g] CHU Lille, Department of Radiology and Musculoskeletal Imaging, F-59000 Lille, France
[h] CHU Lille, Department of Rheumatology, F-59000 Lille, France
[i] Centre for Clinical Brain Sciences, University of Edinburgh, The Queen's Medical Research Institute, Edinburgh BioQuarter, 47 Little France Crescent, Edinburgh EH16 4TJ, UK

ARTICLE INFO

Keywords:
Deep learning
Biomarkers
Predictive analytics
Magnetic resonance imaging
Bone marrow adipose tissue
Bone marrow adiposity
Bone marrow fat fraction
UK Biobank
Bone
Osteoporosis
Ageing
Sex differences

ABSTRACT

Background: Bone marrow adipose tissue (BMAT) represents > 10% fat mass in healthy humans and can be measured by magnetic resonance imaging (MRI) as the bone marrow fat fraction (BMFF). Human MRI studies have identified several diseases associated with BMFF but have been relatively small scale. Population-scale studies therefore have huge potential to reveal BMAT's true clinical relevance. The UK Biobank (UKBB) is undertaking MRI of 100,000 participants, providing the ideal opportunity for such advances.
Objective: To establish deep learning for high-throughput multi-site BMFF analysis from UKBB MRI data.
Materials and methods: We studied males and females aged 60–69. Bone marrow (BM) segmentation was automated using a new lightweight attention-based 3D U-Net convolutional neural network that improved segmentation of small structures from large volumetric data. Using manual segmentations from 61–64 subjects, the models were trained to segment four BM regions of interest: the spine (thoracic and lumbar vertebrae), femoral head, total hip and femoral diaphysis. Models were tested using a further 10–12 datasets per region and validated using datasets from 729 UKBB participants. BMFF was then quantified and pathophysiological characteristics assessed, including site- and sex-dependent differences and the relationships with age, BMI, bone mineral density, peripheral adiposity, and osteoporosis.
Results: Model accuracy matched or exceeded that for conventional U-Nets, yielding Dice scores of 91.2% (spine), 94.5% (femoral head), 91.2% (total hip) and 86.6% (femoral diaphysis). One case of severe scoliosis prevented segmentation of the spine, while one case of Non-Hodgkin Lymphoma prevented segmentation of the spine, femoral head and total hip because of T2 signal depletion; however, successful segmentation was not disrupted by any other pathophysiological variables. The resulting BMFF measurements confirmed expected relationships between BMFF and age, sex and bone density, and identified new site- and sex-specific characteristics.
Conclusions: We have established a new deep learning method for accurate segmentation of small structures from large volumetric data, allowing high-throughput multi-site BMFF measurement in the UKBB. Our findings reveal

* Correspondence to: University/BHF Centre for Cardiovascular Science, The Queen's Medical Research Institute, Edinburgh BioQuarter, 47 Little France Crescent, Edinburgh EH16 4TJ, UK.
E-mail address: W.Cawthorn@ed.ac.uk (W.P. Cawthorn).
[1] These authors contributed equally to this work.
[2] ORCID: 0009-0008-3338-4545
[3] ORCID: 0009-0001-3827-2218

https://doi.org/10.1016/j.csbj.2023.12.029
Received 20 September 2023; Received in revised form 20 December 2023; Accepted 23 December 2023
Available online 27 December 2023

new pathophysiological insights, highlighting the potential of BMFF as a novel clinical biomarker. Applying our method across the full UKBB cohort will help to reveal the impact of BMAT on human health and disease.

1. Introduction

Bone marrow adipose tissue (BMAT) accounts for up to 70% of total bone marrow (BM) volume and approximately 10% of total fat mass in lean, healthy humans [1]. BMAT further increases with ageing and in diverse clinical conditions, including osteoporosis, obesity, type 2 diabetes, oestrogen deficiency, chronic kidney disease, radiotherapy and glucocorticoid treatment [1]. In striking contrast to other adipose depots, BMAT also increases during caloric restriction in animals and in humans with anorexia nervosa [1–4]. Thus, BMAT is a major component of normal human anatomy; is distinct to other types of adipose tissue; and is altered in numerous clinical contexts.

These observations suggest roles for BMAT in normal physiological function and the pathogenesis of multi-morbidities, including major ageing-associated diseases. Indeed, clinical and preclinical studies suggest that BMAT can directly influence skeletal remodelling, haematopoiesis and energy homeostasis [1,5,6] and have revealed endocrine properties through which BMAT may exert systemic effects [3]. However, study of BMAT has been limited, especially in comparison to other major adipose depots [1]; hence, BMAT formation and function remains poorly understood.

Despite this relative ignorance, recent studies have revealed new fundamental knowledge of BMAT biology. One key finding is that BMAT's characteristics and functions differ according to its skeletal location. BMAT is proposed to exist in two broad subtypes, dubbed 'constitutive' and 'regulated' [7,8]: constitutive BMAT predominates in the appendicular skeleton, particularly at more-distal sites, whereas regulated BMAT develops in the axial skeleton and in proximal regions of the long bones, such as the femoral head and epiphysis. Adipocytes within regulated BMAT increase or decrease in size and/or number in response to altered environmental, physiological and pathological conditions, whereas those within constitutive BMAT are relatively resistant to expansion or breakdown in such contexts [7,8]. Thus, efforts to further elucidate BMAT formation and function must consider these fundamental site-specific differences.

Magnetic resonance imaging (MRI) and proton MR spectroscopy have emerged as key tools for non-invasively assessing BMAT properties in humans [9], including the extent of BM adiposity and the proportions of saturated and unsaturated lipids within the BM [10]. The former depends on analysis of BM fat fraction (BMFF) using chemical shift-encoding based water-fat separation methods. These approaches have been applied in various small- and mid-scale human cohort studies, revealing some insights into BMAT's association with human skeletal and metabolic health [11,12]. For example, multiple studies have shown that BMFF is increased in osteoporosis and is associated with lower bone mineral density (BMD) in non-osteoporotic subjects [11–13]. However, these cohort studies have never included more than 676 people [14], limiting the ability to detect other associations. Thus, analysis of BMFF on a larger scale has enormous potential to reveal fundamental new knowledge of BMAT formation and function, including the association with other physiological, pathological and genetic variables. This would provide new understanding about the factors that regulate BMAT development, as well as highlighting how altered BMFF impacts human health and disease.

The UK Biobank (UKBB) is undertaking the world's largest health imaging study [15], providing an ideal opportunity for such large-scale BMFF analysis. Of the 500,000 UKBB participants, 100,000 are undergoing MRI of the brain, heart and whole body, as well as dual-energy X-ray absorptiometry to measure BMD. As of Novmeber 2023, approximately 73,000 participants have been scanned. Efficient measurement of BMFF from these MRI datasets will require development of new automated analysis methods. Several groups have developed machine learning for automated segmentation of other anatomical regions from the UKBB MRI data [16–18]. One preprint also reports deep learning for segmentation of calvarial BM from UKBB MRI scans of the skull [19]. However, this study used only T1-weighted MR data and attempted to quantify BM adiposity based on raw MRI signal intensity, which has never been validated for this purpose [19]; the clinical significance of calvarial BM adiposity also remains uncertain. Machine learning has also recently been used to segment the knee or vertebral BM from Dixon images in smaller cohorts outwith the UKBB [20–22]; however, there are no peer-reviewed studies establishing machine learning for automated segmentation of the BM from other skeletal sites, and never using MR data from the UKBB. These were the goals of the present study.

Given the potential insights that could be gained from such large-scale BMFF analysis, herein our aims were to develop a deep learning pipeline for automated BM segmentation, at multiple skeletal sites, from UKBB MRI data; and to validate the resulting BMFF values by testing if they show pathophysiological relationships that are consistent with previous studies. Our findings establish the utility of deep learning for large-scale analysis of BMFF within the UKBB and the potential of this approach for revealing the impact of BMAT on human health and disease.

2. Materials and methods

2.1. UKBB Imaging study – participants

Full details of the UKBB imaging study have recently been reported by Littlejohns et al., who summarise the study as "*a population-based cohort of half a million participants aged 40–69 years recruited between 2006 and 2010. In 2014, UK Biobank started the world's largest multimodal imaging study, with the aim of re-inviting 100,000 participants to undergo brain, cardiac and abdominal magnetic resonance imaging, dual-energy X-ray absorptiometry and carotid ultrasound*" [15]. As of November 2023, approximately 73,000 participants have undergone the UKBB abdominal MRI protocol. In this study, we focussed on an initial cohort of 729 participants to train and validate our deep learning models; further details are provided below ("Training and validation cohort"), with participant characteristics reported in Table 1. The phenotypic and imaging data used in this study were obtained from UKBB and analysed under an approved project application (ID 48697). All work reported herein was done in accordance with UKBB ethical requirements.

2.2. UKBB – MRI acquisition

MRI data were acquired on a 1.5 T whole-body MR system (Magnetom Aera, Siemens Medical Solutions, Erlangen, Germany). Tridimensional two-point Dixon sequences were used to give coverage from neck to knees, For quantification of BM adiposity, the availability of two-point Dixon sequences only is one limitation of the UKBB imaging study, because these sequences do not allow accurate T2 * correction. This is a limitation because, within the BM, the presence of trabecular bone can cause T2 * decay effects that may differ in the water and fat components [9,10]. Consequently, two-point Dixon sequences do not allow quantification of the corrected proton-density fat fraction (PDFF), and therefore herein we calculated the dual-echo bone marrow fat fraction (BMFF); further details and considerations are reported in Section 2.8 ('Fat fraction mapping') and in the Limitations section of the Discussion.

The UKBB MRI sequences consist of six volumes (slabs), with the first slab starting at the neck and the sixth slab extending to the knees. In the

present study we analysed three of these slabs: the lower thorax and abdomen (slab 2), hips (slab 4), and upper leg (slab 5). For slabs 1–4, breath-hold sequences were acquired by using a 3D dual-echo spoiled gradient-echo (FLASH) T1-weighted acquisition using the following parameters: $TR/TE_{in\text{-}phase}/TE_{out\text{-}of\text{-}phase}$: 6.7/4.8/2.4 ms; field of view (FOV): 500 × 381 mm; slice thickness: 4.5 mm; isotropic in-plane spatial resolution of 2.2 mm; number of slices: 44. Parallel imaging factor 2 in both frequency/phase directions and a partial Fourier reconstruction of 71% were used to reduce acquisition time. For slab 5 (upper leg), slice thickness was reduced to 3.5 mm and 72 slices were acquired with the same resolution. Detailed technical parameters are available in previous papers reporting the UKBB imaging protocol [15,23].

2.3. UKBB – DXA scans for bone mineral density measurement and body composition

As part of the UKBB Imaging study, bone mineral density (BMD) was measured at the lumbar spine (L1–L4) and at the non-dominant hip for femoral neck and total hip by DXA scan (GE-Lunar iDXA). Machines were calibrated daily, and quality-assurance tests were carried out periodically. WHO criteria were used to define osteoporosis (BMD T-score ≤−2.5) and osteopaenia (BMD T-score between −1.0 and −2.5). All UKBB imaging participants also underwent total-body DXA scanning (GE-Lunar iDXA). Fat, lean, and bone masses for the total body and per region (arms, legs, and trunk) were measured and analyzed using the manufacturer's validated software, with visceral adipose tissue (VAT, kg) also measured. Daily quality-control and calibration procedures were performed using the manufacturer's standards.

2.4. Training and validation cohort

To develop a deep learning method for automated BM segmentation we focussed on a subset of UKBB Imaging participants, consisting of 729 male and female subjects aged 60–69 years old (Table 1). This cohort was selected to include control subjects (with normal BMD) and subjects with osteopaenia or osteoporosis. Subjects with obesity and type 2 diabetes were excluded because these conditions can influence BMFF [1,6], leaving only non-diabetic subjects with a body mass index (BMI) within the normal range (18.5–25 kg/m^2). No other skeletal conditions were particularly prevalent among this cohort, as assessed by systematic analysis of PheCodes for these conditions [24,25] (see Supplemental Data file).

2.5. Data management and workflow

MRI data was downloaded from UKBB, consisting of multiple volumes acquired using the two-point Dixon technique, based on the parameters listed above. For each volume the in- and out-of-phase, fat and water images were available. The data were downloaded in flat format and sorted by sequence to expedite data access. The volumes required were identified by their sequence number assuming a standard acquisition protocol, which was determined from the data. As shown in Fig. 1, we began by downloading and analysing data from the 729-subject training and validation cohort.

2.6. Manual segmentation of MRI data

A training dataset of 75 subjects (Fig. 1 A) was extracted from the test dataset to be used for the training and validation of the deep learning algorithms. Each of these 75 datasets was segmented by a single observer (D.M.M.) for consistency, generating manual segmentations. For each subject, the fat images were used to define four distinct volumes of interest (VOIs) corresponding to BM regions of pathophysiological relevance: the spine, the femoral head, the total hip, and the femoral diaphysis. The spine consisted of all the vertebral marrow in the principle abdominal volume (slab 2), which contained 6–7 vertebrae ranging from T8 to L3. The reason for this range of vertebrae is that the multiple abdominal acquisitions have a fixed volume and are continuous across the patient's body; hence, the range of vertebrae within each abdominal volume depends on the patient's height. The femoral head and total hip regions were segmented from the hip volume (slab 4). Here, the total hip consisted of the femoral neck and the hip between the lesser and greater trochanter. The femoral diaphysis, located in the upper leg volume (slab 5), was segmented at the mid-shaft of the femur, which was identified by locating the point of the shaft with the narrowest cross section. Each femoral volume was segmented from the non-dominant left femur to allow more-direct comparison with DXA measurements, which are usually performed at the non-dominant hip. Femoral BMFF does not show significant contralateral differences [26], meaning that BMFF measurements from the left femur should be representative of both sides. Segmentation was performed on the native axial images on a slice-by-slice basis in Analyze 12.0 software

Table 1
Characteristics of subjects in training and validation cohort. Normally distributed data are reported as mean ± SEM while non-normally distributed data are reported as median [interquartile range]. BMI, body mass index; DXA, dual-energy X-ray absorptiometry; VAT, visceral adipose tissue. Within each sex, significant differences between control subjects and osteopaenic or osteoporotic subjects are indicated by * ($P < 0.05$), * * ($P < 0.01$) or * ** ($P < 0.001$). Within control subjects, significant differences between males and females are indicated by $^{\#\#}$ ($P < 0.01$) or $^{\#\#\#}$ ($P < 0.001$).

	Males (n = 277)			Females (n = 452)		
	Control (n = 138)	Osteopaenic (n = 146)	Osteoporotic (n = 17)	Control (n = 134)	Osteopaenic (n = 262)	Osteoporotic (n = 70)
Age (years)	65 [63,67]	65 [63,67]	64.47 ± 0.7	65 [62,67]	65 [62,67]	65 [63,67]
BMI (kg/m^2)	23.6 [22.8, 24.3]	23.3 [22.3, 24.1]	22.04 ± 0.20 ***	22.9 [21.7, 23.9]$^{\#\#}$	22.6 [21.3, 23.7]	21.67 ± 0.40 *
BMD T-score (L1-L4)	0.65 [− 0.2, 1.775]	-1 [− 1.575, − 0.1] ***	-3 [− 3.25, − 1.55] ***	0.15 [− 0.4, 0.9]	-1.5 [− 1.9, − 0.8] ***	-2.8 [− 3.1, − 2.6] ***
BMD T-score (total femur, left)	0.2 [− 0.3, 0.7]	-1.12 ± 0.05 ***	-2.2 ± 0.14 ***	0 [− 0.4, 0.475]	-1.4 [− 1.8, − 1] ***	-2.22 ± 0.09 ***
BMD T-score (femoral neck, left)	-0.3 [− 0.7, 0.275]	-1.5 [− 1.8, − 1.2] ***	-2.45 ± 0.13 ***	-0.15 [− 0.7, 0.4]	-1.45 [− 1.8, − 1.1] ***	-2.11 ± 0.07 *defined**
Android tissue fat% by DXA	30.6 [24, 34.6]	30.0 [22.8, 35.7]	24.4 ± 2.0	34.8 [27.8, 40.7]$^{\#\#\#}$	32.5 ± 0.6	31.0 ± 1.1
Gynoid tissue fat% by DXA	24.3 ± 0.4	24.4 ± 0.4	23.5 ± 1.0	37.6 ± 0.4$^{\#\#\#}$	38.5 ± 0.3	38.7 ± 0.6
Legs tissue fat% by DXA	20.9 ± 0.3	21.2 ± 0.3	21.3 ± 1.0	35.2 ± 0.5$^{\#\#\#}$	36.9 ± 0.3	37.1 ± 0.6
Trunk tissue fat% by DXA	29.1 [23.7, 32.0]	28.6 [23.0, 33.4]	24.3 ± 1.5	35.4 [29.9, 39.5]$^{\#\#\#}$	33.3 ± 0.4	32.3 ± 0.9
Total tissue fat% by DXA	24.6 ± 0.4	25.6 [21.6, 28.5]	22.9 ± 1.8	34.7 [30.9, 37.38$^{\#\#\#}$	34.3 ± 0.3	33.9 ± 0.6
VAT mass (g)	949.4 ± 35.25	783.5 [465.5, 1131]	586 ± 79.6 **	407 [225.5, 717]$^{\#\#\#}$	346.5 [217, 563.5]	296 [193.3, 526.5]

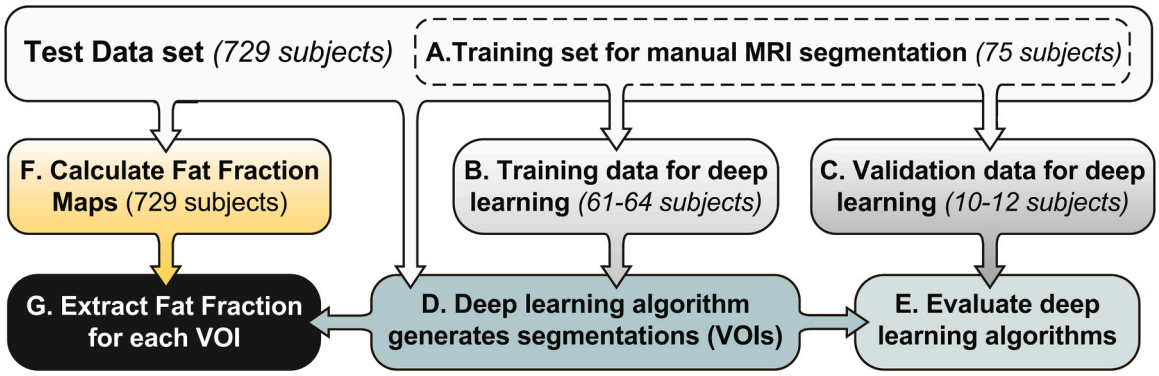


Fig. 1. Workflow for data management, manual segmentation and application and validation of deep learning. The test dataset comprised the validation cohort of 729 subjects (described in Table 1), from which datasets from 75 subjects were manually segmented (A) to generate four VOIs per subject (spine, femoral head, total hip, and femoral diaphysis). The manual segmentations from 61–64 of these subjects were used to train the deep learning models for each VOI (B), while those from 10–12 subjects were kept as 'unseen' segmentations that had not been used to train the models (C). The models were then used to segment all datasets from the 729-subject cohort (D), with deep learning segmentations from the 10–152validation datasets then compared to the corresponding manual segmentations to calculate dice coefficients for each model (E). Finally, FF maps were generated from each MRI dataset (F) and the deep learning segmentations applied to these to obtain the BMFF for each VOI (G).

(AnalyzeDirect, Overland Park, KS, USA) following an overall inspection of each volume to determine the extent of each region excluding partial volume, defined as a drop in signal intensity $>$ 50% compared to the centre of the region.

Of the 75 manually segmented datasets (Fig. 1 A), 64 were used to train the deep learning model for the spine; 61 were used for the femoral head and diaphysis; and 62 were used for the total hip (Fig. 1B). To do so, the fat images and their corresponding manual segmentations were used iteratively to build a separate model to segment each region individually and generate a deep learning segmentation (Fig. 1D). The remaining datasets (Fig. 1 C) were not used in training the models but instead were used as unseen validation data to test the models: 12 datasets were used for testing the spine, 11 for the femoral head, and 10 each for the total hip and diaphysis models. For these validation datasets, comparison of the deep learning segmentations with the manual segmentations (Fig. 1E) allowed dice coefficients to be calculated for the four different algorithms (Table 2).

All the deep learning segmentations for the training and validation datasets were manually checked. This identified several data issues and segmentation failures that required the development of specific error-checking rules. These rules were based on determining if the VOIs generated were physiologically appropriate: VOIs could not consist only of single voxels, nor were gaps allowed within the VOIs. Therefore, the initial error-checking steps automatically removed any single-voxel VOIs and joined together any discontinuous VOIs. Additional error checking was used to identify those segmentations that were outliers within the distribution of regions generated. This was based on the centre of mass being greater than 3 standard deviations from the mean of the training dataset. This is useful for identifying erroneous segmentations that have been caused by data quality issues or deviations from the standard MRI protocol.

2.7. U-Net design and rationale

Directly segmenting 3D data using a traditional U-Net [27] has several drawbacks: i), the size of input data and the depth of the model are limited by the available GPU memory; ii), due to the highly imbalanced distribution between classes, the traditional 3D U-Net [27] tends to label all voxels as background; and iii), the fixed size of the receptive field limited the ability of the model to effectively utilize the global correlations between local features.

To address these issues, we developed a novel light-weight attention-based U-Net model for simultaneous detection and segmentation of tiny structures in large 3D data. Fig. 2 shows the architecture of this new Attention ROI U-Net model. The encoding subnetwork output feature maps four resolution levels [28]. Each encoding block consists of a conventional U-Net convolutional layer (3D conv + Relu + Instance normalization), a convolutional layer equipped with a modified convolutional block attention module (CBAM) [29], and a down-sampling layer implemented as a stride 2 3X3X3 convolution operation. The last encoding block consists of two CBAM convolutional layers with a non-local spatial attention layer [30] inserted between them. Unlike the original CBAM, which generates two attention maps using average and max pooling, we used 1X1X1 convolution to generate one single fixed-size attention map from each CBAM layer. The 5 attention maps are all resized to 96X96X96 and then fused by a mini convolutional neural network (CNN) with a *Softmax* layer to generate a probability map $\boldsymbol{P}$. The centre, $(x,y,z)_{ROI}$, of a region of interest (ROI), which indicates the location of the segmented anatomical structure, is then given by:

$$(x,y,z)_{ROI} = \boldsymbol{P} \odot (\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}),$$

Here, $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}$ are grid of data coordinates normalized to $[-1, 1]$. With this centre, a cubic ROI is extracted from the encoder feature maps of all resolution levels with sizes 32, 16, 8 and 4. The U-Net decoder then generates the segmentation of this ROI. The final segmentation results are produced by recovering the ROI location within the original data volume.

Detection of the ROI location is realised by minimizing a ROI centre localization loss, L_{loc}, defined on the predicted and ground-truth ROI centres. We use the conventional Dice loss, L_{ROI}, to optimize the segmentation of the detected ROI. Because minimize bias in traduced by the class imbalance on the final segmentation results, we also compute a weighted Dice loss, L_{seg}, using the full image segmentation, where the weight of each class is defined as the reciprocal of the number of voxels. To sum up, the loss function for trains ing our new U-Net model is defined as:

$$L = L_{seg} + \lambda_1 L_{ROI} + \lambda_2 L_{loc},$$

Table 2
Segmentation Accuracy (dice scores) of the traditional U-Net and our CBAM-ROI-attention U-Net.

	Spine	Femoral head	Total Hip	Femoral Diaphysis
U-Net	0.925	0.951	0.904	0.69
ROI-Attention-U-Net	0.912	0.945	0.912	0.866

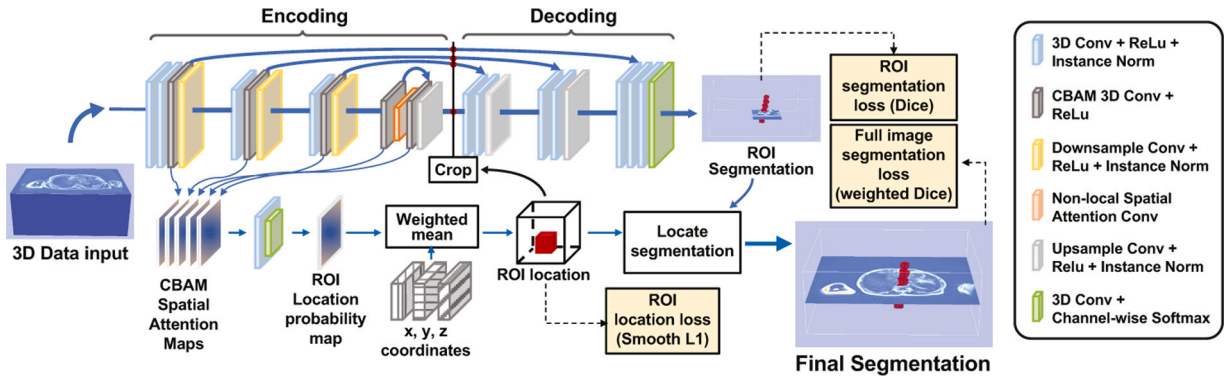


Fig. 2. Architecture of our CBAM Attention ROI U-Net for segmenting small structures from large 3D data. Each convolutional block in the U-Net encoding subnetwork (or contracting path) includes one or two CBAM (convolutional block attention module) layers. A fixed-size single channel spatial attention map is generated by each CBAM layer through 1X1X1 convolutions and trilinear interpolation. These attention maps are then combined to produce a probability map of object location with which a ROI is defined. The encoded features of all resolution-levels are then cropped to the ROI and input into the decoder which produces the segmentation results within the detected ROI. A non-local spatial attention layer is inserted in the final block to generate globally sensitive features. The final segmentation results are then generated by implanting the ROI back into the whole data volume.

where λ_1 and λ_2 control the weights between different losses. In this work, we set $\lambda_1 = \lambda_2 = 1$. The proposed algorithm was implemented in Pytorch [31] with an Adam optimizer [32].

2.8. Fat fraction mapping

Fat fraction (FF) measurements from MRI data allow for the determination of the relative quantities of water and fat present within tissue, based on the different resonant frequencies of hydrogen atoms bound to fat and water. Acquisition of in- and out-of-phase images allows fat and water images to be generated. Based on the intensities of these images the FF was calculated as a percent of the voxel volume. This was done for all volumes of interest. The specific VOIs, segmented using our novel U-Net model, were then applied to the FF maps to allow extraction of the FF for each VOI. This used the fat and water images for each volume of interest and nearest-neighbour smoothing was applied to the images before the maps were calculated to minimise the influence of any noise spikes in the data. In house code (Matlab 2019B, The Mathworks Inc, Natick, Massachusetts, USA) applied the deep learning segmentations to the FF maps after erosion of the spine, head and total hip regions by a single boundary voxel in plane to ensure measurements were from marrow and not bone. This erosion step was not applied to the diaphysis segmentations because of the small cross section of this region (for some patients the diaphyseal cross section is so small that it would be eliminated by the erosion step).

2.9. Data presentation and statistical analysis

Data were analysed for normal distribution using the Anderson-Darling test. For results tables of summary statistics, normally distributed data are reported as mean ± SEM and were compared using one-way or two-way ANOVA with Šídák's test for multiple comparisons. Non-normally distributed data are reported as median [interquartile range] and were compared using the Kruskal-Wallis test, with Dunn's test for multiple comparisons; the latter was also used when comparing normally distributed data with non-normally distributed data. Images of manual and deep learning segmentations were generated using 3DSlicer (v4.11) and colours adjusted using GIMP2. Graphs of summary data are presented as Violin plots overlaid with individual data points. Visualisation and statistical analysis of these summary data were done using Prism software (v10.1.1, GraphPad, USA). Univariable regression analyses were done in RStudio v2023.06.1 (Build 524), with multivariable regression performed using finalfit (R package v1.0.5) [33]. Subjects with any erroneous measurements (e.g. a BMD of 0 g/cm^2, or BMFF values derived from abnormal segmentations) were excluded from the regression analyses. A Bonferroni-adjusted P-value < 0.05 was considered statistically significant.

2.10. Data and code availability

All data for FF and segmentation volumes will be uploaded to the UKBB. Code for the deep learning models will be made available via GitHub. Code for regression analyses will be made available via DataShare (https://datashare.ed.ac.uk). Until these data are publicly available, the authors will agree to all reasonable requests for code and data sharing, in accordance with UKBB guidelines.

3. Results

3.1. U-Net development and training

We first used MRI data from 61–64 subjects for manual segmentation of four VOIs: the spine, consisting of lumbar and thoracic vertebrae; the femoral head; total hip; and femoral diaphysis. We then trained separate U-Net models for each VOI and tested their performance on 10–12 subjects in a validation dataset (Fig. 1). Fig. 2 shows the architecture of our new U-Net, while Table 2 shows the comparison Dice index results between the conventional U-Net and our new U-Net models for each site. Visual comparison of manual vs deep learning segmentations further confirmed the accuracy of the outputs from each of our deep learning models (Fig. 3). Notably, the conventional U-Net performed well for the spine, femoral head and total hip, but poorly for the diaphysis (accuracy of only 69%). In contrast, our CBAM-ROI-attention U-Net greatly improved segmentation accuracy for the diaphysis (to nearly 87%) while being comparable to the conventional U-Net for each of the other regions (Table 2).

3.2. Determining if technical or biological factors influence deep learning segmentation outputs

To further test if our CBAM-ROI-attention U-Net models yield robust segmentation outputs and reliable BMFF results, we next applied them to a cohort of 729 UKBB participants (Table 1). This cohort was chosen to include both males and females aged 60–69, comprising individuals with osteoporosis, osteopaenia, or normal BMD. The rationale for this is as follows: first, BMFF increases with age and, for humans aged 60–69, vertebral BMFF is expected to be greater in females than in males [34,35]; second, BMFF is increased in osteoporosis and negatively associated with BMD [1,6,12]; and finally, BMFF is greater in the femur than in the

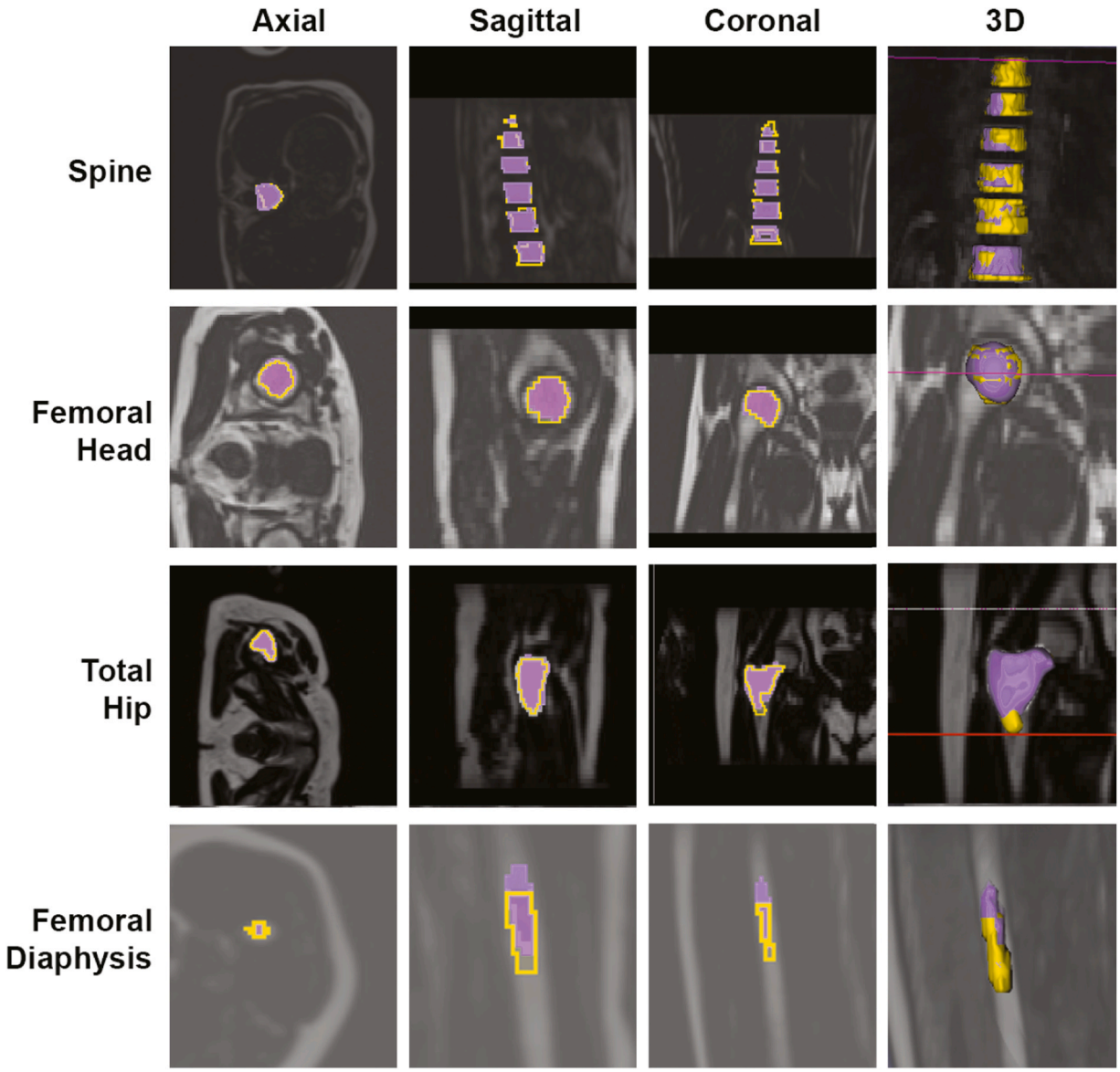


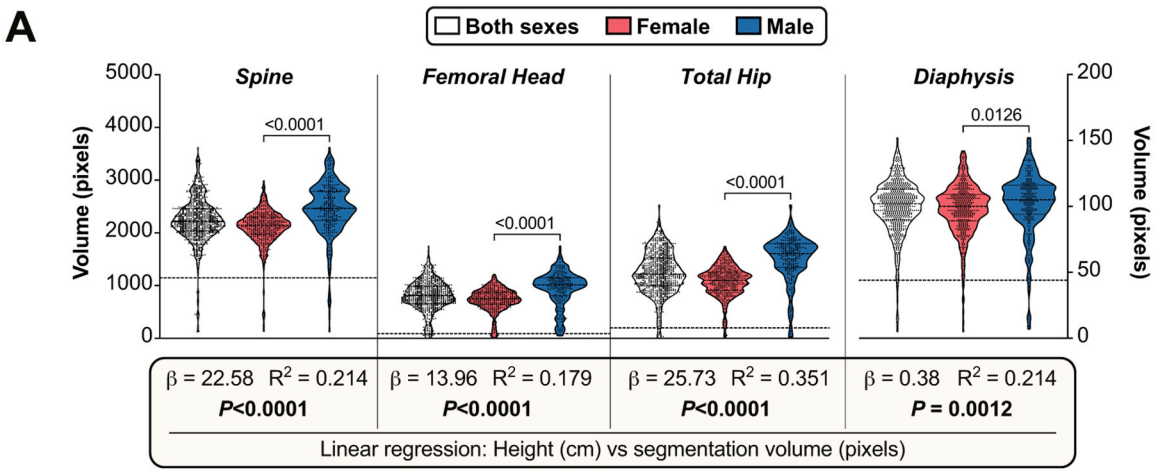
Fig. 3. Visual comparison of manual vs deep learning segmentations. Deep learning segmentation results (purple) are displayed on top of the ground-truth (manual) segmentations (yellow). Representative images from the axial, coronal and sagittal plane are shown, along with a 3D rendering. Note that the Total Hip includes the intertrochanteric region.

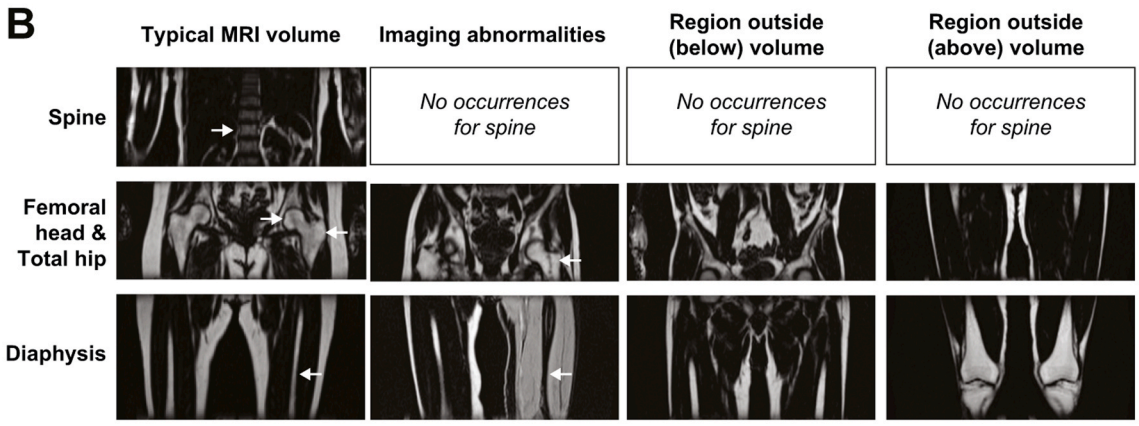
lumbar spine [1,36]. Thus, applying our U-Net models to analyse spinal and femoral BMFF in this cohort allowed us to test if the resulting deep learning segmentations yield BMFF values that show these expected associations with sex, age, BMD, and anatomical site. If so, this would validate the accuracy of our new models for high-throughput BM segmentation and BMFF analysis.

We first analysed segmentation results from across the 729-participant validation cohort to determine if any technical factors or participant characteristics compromised the deep learning outputs. Across this cohort, segmentation volume (pixels, mean ± SD) was greatest for the spine (2244 ± 438), followed by the total hip (1248 ± 404), femoral head (810 ± 286), and diaphysis (100 ± 21) (Fig. 4 A). The volumes for each site were greater in males than in females, likely because, on average, men are taller than women and therefore have larger bones. Consistent with this, for each site linear regression revealed a significant positive relationship between participant height and segmentation volume (Fig. 4 A); this relationship was the same in males and females ($P > 0.6$ for height*sex interaction at each site).

In some cases, the deep learning models generated an empty segmentation output. As shown in Table 3, this was most common for the femoral head (9.7% of all participants) and diaphysis (8%) but was less frequent for the total hip (3.6%) or spine (2.8%). For BMFF analysis we also excluded any small segmentation outputs, defined as having a volume > 2.5 SD below the mean for each region (Fig. 4 A); our rationale was that aberrantly small volumes will be more likely to yield inaccurate BMFF values. These 'small' outputs occurred for ~2% of all segmentations for each region, while both the empty and small outputs showed similar prevalence in males and females (Table 3).

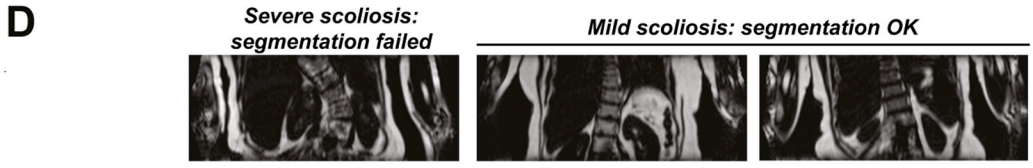
One concern is that empty or small segmentation volumes might result from biological factors, including skeletal abnormalities, that compromise the performance of our deep learning models. If so, this could limit the scope and generalisability of the resulting BMFF measurements. To address this, we manually inspected each MRI dataset to test if there were obvious causes of the abnormal segmentations; as shown in Table 3, we identified five broad categories of failure causes. The most-common category related to technical issues with the structure of the UKBB source MRI data ("Technical issue – Data structure"). Here, the MRI volumes required for segmenting the spine (slab 2), femoral





C

	Spine		Femoral head		Total hip		Diaphysis	
	OR (95% CIs)	P	OR (95% CIs)	P	OR (95% CIs)	P	OR (95% CIs)	P
Height vs empty segmentation	1.03 (0.98-1.09)	0.88	0.99 (0.96-1.01)	0.99	1.00 (0.96-1.04)	0.24	1.00 (0.97-1.03)	0.32
Height vs small or empty segmentation	1.03 (0.99-1.07)	0.43	1.00 (0.97-1.02)	0.14	0.97 (0.94-1.01)	0.14	0.99 (0.96-1.02)	0.87



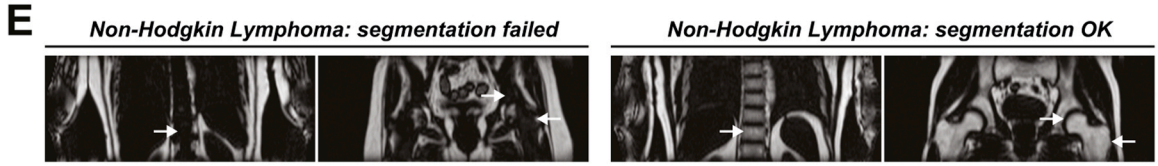


Fig. 4. Identification of technical and biological factors that influence segmentation outputs. (A) For each skeletal site, segmentation volumes (pixels) for both sexes (together or separately) are shown as violin plots overlaid with individual data points; the numbers for each group are shown in Table 3 ("OK" plus "Small" segmentations are included in the graph). For each region, horizontal dotted lines are drawn 2.5 SD below the mean to highlight the threshold used to exclude abnormally small volumes. Spine, femoral head and total hip are plotted on the same y-axis scale, whereas the much-smaller diaphysis is shown on a separate y-axis. The box beneath the graph shows the results of linear regression for height (cm) vs segmentation volume (pixels) at each site, for both sexes combined. The strong positive relationship did not differ between males and females. **(B)** Examples of coronal MRI volumes for each skeletal site, including typical volumes, those with imaging abnormalities, and those in which the target region fell partially or fully below or above the MRI volume. Arrows indicate the target BM regions (for clarity, only one arrow is shown for the spine, in which six vertebrae are segmented). **(C)** Results of logistic regression to investigate if participant height affects the odds of a segmentation being empty (top row) or small or empty (bottom row). **(D-E)** Coronal MRI volumes from participants with severe or mild scoliosis (D) or Non-Hodgkin Lymphoma (E). The differential effects on segmentation outcomes are indicated above the images.

Table 3
Characteristics of deep learning segmentation outputs from the 729-subject cohort. For each skeletal site, segmentation outputs were classified as 'OK' (volume not >2.5 SD below mean volume for that region), 'Small' (volume >2.5 SD below mean), or 'Empty' (no output generated from deep learning); Small or Empty volumes were excluded from the BMFF analyses. Columns 1–3 show the numbers of each type of output in both sexes (1), Females (2) or Males (3), and the % that these numbers represent for each sex. Columns 4–9 show the numbers of Small, Empty, or Small or Empty segmentation outputs, and the % these represent for each output type, for which there were technical issues with the source data structure (4) or imaging artefacts (5); the skeletal target site was partially (6) or fully (7) outside of the MRI slab volume; pathological skeletal abnormalities were apparent (8); or for which no obvious abnormalities were detectable (9). Column 10 shows, for each type of segmentation output, the number of participants having a PheCode for a skeletal disease and the % that this represents for each type of segmentation output; further details of these PheCodes and diseases are shown in the Supplemental Data file.

		Numbers (% of each group)			Numbers (% of faulty segmentation type)						
	Segmentation output	(1) Both sexes	(2) Females	(3) Males	(4) Technical issue - Data structure	(5) Technical issue -Imaging artefact	(6) Region partially outside slab	(7) Region fully outside slab	(8) Skeletal abnormality	(9) No obvious defect	(10)Prevalence of skeletal PheCode (% output type)
Spine	**OK**	696 (95.5%)	432 (96.3%)	264 (97.5%)	-	-	-	-	-	-	290 (41.7%)
	Small	13 (1.8%)	7 (1.6%)	6 (2.3%)	13 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (23.1%)
	Empty	20 (2.8%)	13 (2.9%)	7 (2.6%)	17 (85%)	0 (0%)	0 (0%)	0 (0%)	2 (10%)	0 (0%)	8 (40%)
	Small or Empty	33 (4.6%)	20 (4.5%)	13 (4.8%)	30 (91%)	0 (0%)	0 (0%)	0 (0%)	2 (6.1%)	0 (0%)	11 (33.4%)
Femoral Head	**OK**	646 (88.7%)	408 (90.9%)	238 (87.9%)	-	-	-	-	-	-	273 (42.3%)
	Small	13 (1.8%)	9 (2.1%)	4 (1.5%)	1 (7.7%)	1 (7.7%)	8 (61.6%)	0 (0%)	0 (0%)	3 (23.1%)	3 (23.1%)
	Empty	70 (9.7%)	35 (7.8%)	35 (13%)	17 (24.3%)	14 (20%)	10 (14.3%)	9 (12.9%)	1 (1.5%)	16 (22.9%)	25 (35.8%)
	Small or Empty	83 (11.4%)	44 (9.8%)	39 (14.4%)	18 (21.7%)	15 (18.1%)	18 (21.7%)	9 (10.9%)	1 (1.3%)	19 (22.9%)	28 (33.8%)
Total Hip	**OK**	693 (95.1%)	430 (95.8%)	263 (97.1%)	-	-	-	-	-	-	286 (41.3%)
	Small	10 (1.4%)	4 (0.9%)	6 (2.3%)	2 (20%)	0 (0%)	2 (20%)	6 (60%)	1 (10%)	0 (0%)	7 (70%)
	Empty	26 (3.6%)	18 (4.1%)	8 (3%)	19 (73.1%)	2 (7.7%)	1 (3.9%)	2 (7.7%)	0 (0%)	0 (0%)	8 (30.8%)
	Small or Empty	36 (5%)	22 (4.9%)	14 (5.2%)	21 (58.4%)	2 (5.6%)	3 (8.4%)	8 (22.3%)	1 (2.8%)	0 (0%)	15 (41.7%)
Diaphysis	**OK**	657 (90.2%)	411 (91.6%)	246 (90.8%)	-	-	-	-	-	-	272 (41.5%)
	Small	14 (2%)	7 (1.6%)	7 (2.6%)	1 (7.2%)	2 (14.3%)	3 (21.5%)	0 (0%)	0 (0%)	8 (57.2%)	4 (28.6%)
	Empty	58 (8%)	34 (7.6%)	24 (8.9%)	22 (38%)	12 (20.7%)	12 (20.7%)	1 (1.8%)	0 (0%)	10 (17.3%)	25 (43.2%)
	Small or Empty	72 (9.9%)	41 (9.2%)	31 (11.5%)	23 (32%)	14 (19.5%)	15 (20.9%)	1 (1.4%)	0 (0%)	18 (25%)	29 (40.3%)

head and total hip (slab 4), or femoral diaphysis (slab 5) (Fig. 4B) were located in an incorrect folder sorted from the source UKBB data. In some cases, the target folder contained the correct slab, but with the water image instead of the required fat image. In other cases, the participant's MRI data were distributed among a greater-than-expected number of folders; this was usually because the MRI acquisition began at the wrong landmark and so had to be repeated, resulting in an appended dataset in which the target MRI volume was no longer sorted into the correct folder number. Consequently, the models failed to generate a segmentation because they were presented with an incorrect MRI volume within the source data.

The second category of failed segmentations related to imaging artefacts or other abnormalities, which were most common for the diaphysis or femoral head (Table 3). For the diaphysis, all of the artefacts were fat-water swaps occurring contralaterally across slab 5, resulting in the targeted left leg containing a water image rather than a fat image (Fig. 4B). For the femoral head and total hip, the most common abnormalities were signal inhomogeneities within the proximal femur, often manifesting as distinct lines of hypointense T2 signal that resulted in an unclear segmentation target (Fig. 4B). In contrast, more-diffuse variation in T2 signal did not affect segmentation (i.e. 'Typical MRI volume' in Fig. 4B), and no cases of these artefacts or abnormalities were found among the faulty spine segmentations.

The third and fourth categories of failure causes related to the target region falling partially or entirely outside of the slab volume. This never occurred for the spine but was most common for the femoral head and diaphysis (Table 3; Fig. 4B). One concern is that these failures may be influenced by participant height: because the UKBB MRI protocol uses a fixed slice number for each slab volume, for shorter subjects the slabs will generally extend further down the body than for taller subjects. Thus, while the proximal femur and diaphysis midpoint typically fall within the middle of slabs 4 and 5, respectively (Fig. 4B), these regions may be more likely to fall partially or fully within slabs 3 and 4 for very short participants, or slabs 5 and 6 for very tall participants. However, we tested this using logistic regression and found no relationship between participant height and the odds of segmentation failure (Fig. 4 C). Instead, target regions typically fell partially or fully outside of the slab volume as a result of the MRI acquisition beginning slightly above or below the intended clavicular landmark *(not shown)*.

The fifth category of segmentation failure related to pathological abnormalities in skeletal morphology. This occurred only twice (Table 3): one case of severe scoliosis caused abnormal morphology that prevented spinal segmentation (Fig. 4D), while one participant with Non-Hodgkin Lymphoma had almost complete T2 signal depletion within the bone marrow, preventing segmentation of the spine, femoral head, and total hip (Fig. 4E). Notably, the validation cohort contained

several other cases of less-severe scoliosis and two other cases of Non-Hodgkin Lymphoma that did not impair segmentation (Fig. 4D-E); this variability in BM adiposity among Non-Hodgkin Lymphoma patients is consistent with previous reports [37].

To systematically test if any skeletal or haematological pathologies compromise segmentation, we next identified ICD codes for these diseases; mapped these to PheCodes [24,25]; and then assessed if any PheCodes were enriched among the faulty segmentation outputs (see Supplemental Data). No such enrichment was observed for any individual disease. Moreover, the prevalence of participants having one or more relevant PheCode was similar between those giving faulty vs successful segmentation outputs (Table 3, column 10).

Finally, for the femoral head and diaphysis there were some small or empty segmentations for which no obvious defects were apparent (Table 3, column 9). These accounted for ~20–25% of faulty segmentation outputs for each site, corresponding to < 3% of participants across the validation cohort.

Together, these observations show that, among the minority of faulty segmentation outputs, most errors result from technical issues relating to UKBB MRI acquisition or data outputs. In contrast, only two errors related to obvious pathological abnormalities in skeletal morphology or BM composition. Thus, our models provided robust segmentation volumes for the majority of participants analysed.

3.3. Fat Fraction mapping of training and validation cohort

We next applied the deep learning segmentations to FF maps from the 729-participant cohort, thereby measuring BMFF at each of the four sites. As shown in Fig. 5 A, we found that BMFF in healthy control subjects significantly differed across the four regions analysed. This was most obvious for the spine, where BMFF was lower than in each femoral region ($P < 0.0001$). However, BMFF also differed between each femoral region, being highest in the femoral head and then decreasing progressively in the total hip and diaphysis ($P < 0.0001$ for each pairwise comparison). There were also significant, region-dependent sex differences: spinal BMFF was greater in females than in males, whereas males had greater BMFF at each femoral site (Fig. 5 A).

To further understand the regional and sex differences in BMFF, we investigated if BMFF at one site is associated with BMFF at the other sites. As shown in Table 4, there were strong positive associations between BMFF at each femoral site, with the relationship between total hip BMFF and diaphyseal BMFF being stronger in males than in females. Spinal BMFF was not associated with diaphyseal BMFF; however, it was positively associated with femoral head BMFF in females, and with total hip BMFF in males and females; the latter relationship was also stronger in females than in males (Table 4). Thus, BMFF at one site is generally positively associated with BMFF at other sites, and this relationship differs between the sexes.

3.4. Effect of osteopaenia or osteoporosis on BMFF at each site

We next investigated the effect of osteopaenia or osteoporosis on BMFF at each site. As shown in Fig. 5B-D, osteopaenic or osteoporotic females had higher BMFF than control females at each site analysed. In males, osteopaenia was associated with significantly increased BMFF at the total hip and femoral diaphysis, and BMFF at the femoral head and total hip was also greater in osteoporotic vs control males (Fig. 5B-D). However, unlike in females, BMFF at the spine did not differ between normal, osteopaenic and osteoporotic males, while diaphyseal BMFF also did not differ between osteoporotic and normal males (Fig. 5B-D).

3.5. Univariable associations between BMD, BMFF and other traits

The lack of increased BMFF in the spine, total hip, and diaphysis of osteoporotic males was unexpected and may result from the low numbers in this group (Table 1). Thus, we next used univariable regression to determine if BMFF shows the expected inverse association with BMD at each site, regardless of osteoporotic status. We also investigated which other variables are associated with BMD at each site. As shown in Supplemental Table 1, BMD and BMFF were inversely associated at the spine and this relationship did not differ between the sexes. A similar relationship existed between spine BMD and legs fat%. In contrast, spine BMD was positively associated with visceral adipose tissue (VAT) mass, android fat%, trunk fat% and BMI, with the latter relationship being stronger in males than in females. There was no significant relationship between spine BMD and age, total fat% or gynoid fat%; however, females showed a trend for lower spine BMD with increasing age.

Univariable regression analyses for BMD at the femoral neck, total hip and femoral shaft are presented in Supplemental Tables 2, 3 and 4, respectively. For femoral neck BMD we detected robust inverse associations with BMFF at the femoral head, total hip and spine; the latter relationship was assessed to determine if spinal BMFF is a useful predictor of BMD at the femoral neck, given the clinical significance of fractures at this site. Femoral neck BMD also showed an inverse relationship with legs fat% and a positive association with BMI; however, no significant associations occurred for the other explanatory variables tested (Supplemental Table 2).

Similar relationships occurred for total hip BMD, including an inverse association with legs fat% and a positive association with BMI (Supplemental Table 3). Unlike for femoral neck BMD, total hip BMD also showed a positive association with VAT mass.

As for these other sites, femoral shaft BMD was inversely associated with BMFF at the femoral diaphysis while being positively associated with BMI. Weaker negative and positive associations were noted for legs fat% and VAT mass, respectively, and none of these relationships differed between the sexes (Supplemental Table 4).

3.6. Univariable associations between BMFF and age, BMI or adiposity traits

In addition to BMD, factors including age, BMI and peripheral adiposity have been associated with altered BMFF [1]. Thus, an important question is whether such other factors confound the relationships between BMFF and BMD. To address this, we first used univariable linear regression to identify other variables significantly associated with BMFF at each site, thereby identifying factors associated with BMFF and/or BMD. The results are presented in Supplemental Table 5.

We found that spinal BMFF was positively associated with age, VAT mass, total fat%, android fat%, gynoid fat% and trunk fat% in males and females, with no sex differences in these relationships. In contrast, spinal BMFF showed a positive association with legs fat% in males only (Supplemental Table 5).

Fewer variables were associated with BMFF at the femoral head or total hip. For the former, the strongest association was a positive relationship with gynoid fat% in females only. There were weaker positive associations between femoral head BMFF and legs fat% across both sexes, and with age and total fat% in females only; however, these were no longer significant after adjusting for multiple comparisons. Total hip BMFF was negatively associated only with BMI across both sexes, but no other variables were associated with BMFF at these two sites (Supplemental Table 5). In contrast, diaphyseal BMFF was associated with several of the variables assessed, often in a sexually dimorphic manner. Thus, across both sexes, diaphyseal BMFF was inversely associated with VAT mass, while inverse associations with total fat%, android fat% and trunk fat% showed significant sex differences, occurring in females but not in males. In contrast, in males, but not females, diaphyseal BMFF was positively associated with legs fat% (Supplemental Table 5).

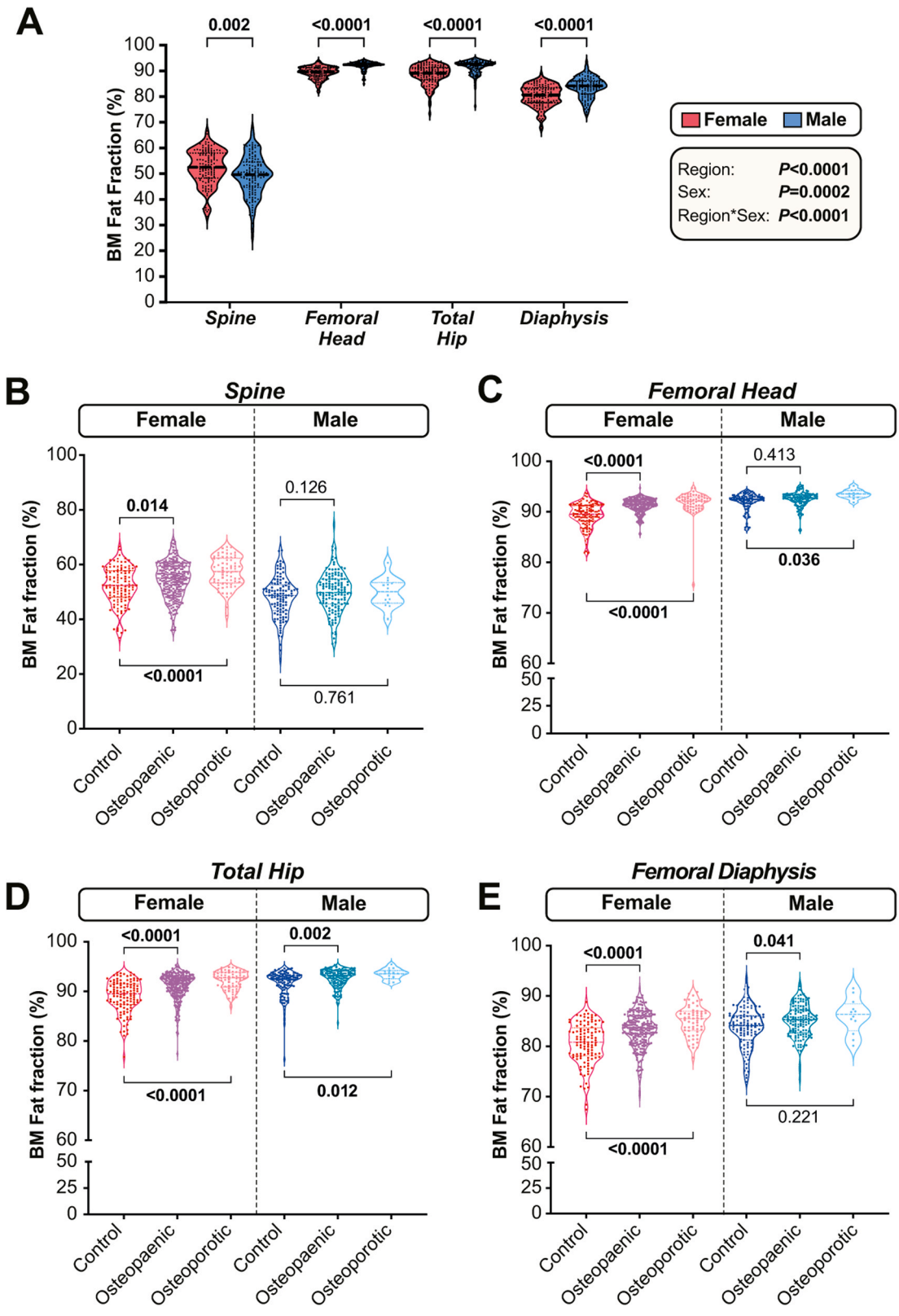


Fig. 5. Biological sex, osteopaenia and osteoporosis influence BMFF in a region-specific manner. BMFF for normal subjects (A) or control, osteopaenic and osteoporotic subjects (B-D) was assessed at each skeletal region. Data are shown as violin plots overlaid with individual data points; the numbers for each group are shown in Table 1. For (A), significant effects of region, sex, and region*sex interaction were assessed using a mixed-effects model with Šídák's multiple comparisons test. Overall *P* values for each variable, and their interaction, are shown in the box beside the graph, while significant sex differences within each region are indicated above the violins. For (B-D), significant differences between control and osteopaenic or osteoporotic subjects within each sex were assessed by one-way ANOVA (for normally distributed data: A) or the Kruskal-Wallis test (for non-normally distributed data: B-D). *P* values for each comparison are shown on each graph.

Table 4
Univariable and sex-stratified associations between BMFF for each region. To test if the explanatory-dependent relationship differs between males and F, a linear model was first analysed across both sexes, with sex included as an interacting variable. Beta coefficients are shown (with lower and upper 95% Cis in brackets), followed by the adjusted R2 (Adj. R2) and unadjusted P value for each explanatory variable (P Exp). P values were also calculated for the Explanatory*Sex interaction (P Exp*Sex); if significant, additional linear models were analysed in females and males separately. Because 12 correlations were assessed, the Bonferroni-adjusted alpha level for P (Exp) is 0.05/12 = 0.0042. Significant explanatory-dependent relationships are highlighted in bold.

Explanatory variable	Dependent variable	Sex	β (CIs)	Adj. R^2	*P* (Exp)	*P* (Exp*Sex)
BMFF Spine	BMFF Femoral Head	Both	**0.037 (0.015, 0.059)**	**0.015**	**1.25E-03**	**2.6E-04**
		Female	**0.109 (0.08, 0.138)**	**0.118**	**8.69E-13**	-
		Male	0.028 (0, 0.057)	0.013	0.049	-
	BMFF Total Hip	Both	**0.091 (0.063, 0.12)**	**0.055**	**4.48E-10**	**0.026**
		Female	**0.171 (0.132, 0.21)**	**0.145**	**2.16E-16**	-
		Male	**0.106 (0.069, 0.144)**	**0.107**	**7.22E-08**	-
	BMFF Diaphysis	Both	0.054 (0.01, 0.099)	0.007	0.017	0.801
BMFF Femoral Head	BMFF Total Hip	Both	**1.011 (0.939, 1.082)**	**0.552**	**1.18E-111**	0.474
	BMFF Diaphysis	Both	**0.805 (0.668, 0.943)**	**0.169**	**1.69E-26**	0.534
BMFF Total Hip	BMFF Diaphysis	Both	**0.766 (0.672, 0.86)**	**0.281**	**2.48E-48**	**0.001**
		Female	**0.65 (0.534, 0.766)**	**0.228**	**8.76E-25**	-
		Male	**1.05 (0.857, 1.238)**	**0.331**	**1.03E-22**	-

3.7. The inverse association between BMFF and BMD at each site persists after controlling for relevant covariables

Based on the univariable associations identified in Supplemental Tables 1–5, we next constructed multivariable models to estimate the true relationship between BMFF and BMD at each site. Table 5 shows the results for BMD spine as the dependent variable. Here, the best predictive model was obtained when including BMFF Spine, sex, BMI, Legs fat %, VAT mass and Android fat% as covariables (Model 4.6). Notably, the inverse association between spinal BMFF and spinal BMD persisted even when accounting for these other covariables. Moreover, inclusion of leg fat, VAT mass and android fat weakened the size of the sex effect, suggesting that increased spinal BMD in males is explained, at least in part, by their lower amount of leg fat and greater VAT mass and android fat.

Table 6 shows the results for femoral neck BMD as the dependent variable. Here, separate models were tested for BMFF at the femoral head, total hip or spine as the main explanatory variables. We found that the significant inverse association between BMFF femoral head and femoral neck BMD persisted when accounting for BMI and legs fat% (Model 5.3). Similarly, across both sexes, total hip or spine BMFF retained their inverse relationships with femoral neck BMD even after accounting for sex, BMI and legs fat% (Models 5.6 and 5.11). The best model for BMFF total hip also included Android fat% and Trunk fat% (Model 5.8). Notably, male sex was no longer associated with increased femoral neck BMD when controlling for BMFF spine, BMI and legs fat% (Model 5.11), suggesting that males have greater BMD at the femoral neck because they tend to have lower spinal BMFF, lower % leg fat and higher BMI than females.

Given that spine BMFF is positively associated with total hip BMFF (Table 4), we postulated that the inverse relationship between spine BMFF and femoral neck BMD may occur because spine BMFF is a surrogate for total hip BMFF. However, the inverse relationship between spine BMFF and femoral neck BMD persisted even when accounting for BMFF at the total hip (Model 5.12), demonstrating that these explanatory variables are acting at least partly independently of each other.

Multivariable regression for total hip BMD is presented in Table 7. The best predictive model included BMFF total hip, sex, BMI and legs fat % as the covariables (Model 6.3); inclusion of VAT mass (Model 6.4) did not further improve the model, despite VAT mass showing a significant univariable association with total hip BMD (Supplemental Table 3). Notably, the inverse relationship between total hip BMD and BMFF persisted even when accounting for sex, BMI and legs fat%, confirming total hip BMFF as an independent predictor of BMD at this site.

Finally, Table 8 shows the results of multivariable regression for femoral shaft BMD. Here, the best predictive model included diaphyseal BMFF, sex, BMI, legs fat% and android fat% (Model 7.5), although a similarly accurate model was obtained when VAT mass and trunk fat% were also included (Model 7.7). As for the other BMFF-BMD relationships, BMFF at the diaphysis retained its significant inverse association with femoral shaft BMD even when these other covariables were accounted for. Moreover, males no longer had significant increases in femoral shaft BMD when controlling for BMFF diaphysis, BMI and legs

Table 5
Multivariable regression analyses for spine BMD. Multivariable regression was done using BMD spine as the dependent variable; explanatory variables were selected based on those showing significant univariable association with BMD spine and/or BMFF at the relevant sites, as shown in Supplemental Tables 1–5. For each model the adjusted R2 (Adj. R2) and Akaike Information Criterion (AIC) are shown, along with multivariable beta coefficients (with lower and upper 95% Cis) for each variable. P values are indicated by * ($P < 0.05$), * * ($P < 0.01$) or * ** ($P < 0.001$), with significant associations highlighted in bold.

			Covariable					
	Adj. R^2	AIC	BMFF Spine	Sex (M)	BMI	Legs fat%	VAT mass (kg)	Android fat%
Model 4.1	0.39	-893.1	**-0.004 (−0.006 to −0.003)* ****	**0.177 (0.156 to 0.198)* ****	-	-	-	-
Model 4.2	0.43	-941.9	**-0.005 (−0.006, −0.003)* ****	**0.158 (0.137, 0.179)* ****	**0.023 (0.017, 0.030)* ****	-	-	-
Model 4.3	0.46	-975.3	**-0.004 (−0.006, −0.003)* ****	**0.061 (0.023, 0.098)* ***	**0.029 (0.023, 0.036)* ****	**-0.006 (−0.008, −0.004)* ****	-	-
Model 4.4	0.47	-990.7	**-0.005 (−0.006, −0.004)* ****	0.037 (−0.002, 0.075)	**0.022 (0.015, 0.029)* ****	**-0.006 (−0.008, −0.004)* ****	**0.064 (0.033, 0.095)* ****	-
Model 4.5	0.47	-990.1	**-0.005 (−0.006, −0.004)* ****	**0.058 (0.020, 0.095)* ***	**0.021 (0.014, 0.029)* ****	**-0.007 (−0.009, −0.005)* ****	-	**0.003 (0.001, 0.004)* ****
Model 4.6	0.47	-990.9	**-0.005 (−0.007, −0.004)* ****	**0.043 (0.004, 0.083)***	**0.021 (0.013, 0.028)* ****	**-0.007 (−0.009, −0.004)* ****	0.041 (−0.003, 0.085)	0.001 (−0.000, 0.003)

Table 6
Multivariable regression analyses for femoral neck BMD. Multivariable regression was done using femoral neck BMD as the dependent variable, with BMFF at the femoral head, total hip and spine chosen as the primary explanatory variables. Other explanatory covariables were selected, models constructed, and data presented as described for Table 5.

			Covariable							
	Adj. R^2	AIC	BMFF Femoral Head	BMFF Total Hip	BMFF Spine	Sex (M)	BMI	Legs fat%	Android fat%	Trunk fat%
Model 5.1	0.26	-1028.4	**-0.022 (−0.026, −0.018)* ****	-	-	-	-	-	-	-
Model 5.2	0.27	-1040.0	**-0.022 (−0.026 to −0.017)* ****	-	-	-	**0.011 (0.005 to 0.017)* ****	-	-	-
Model 5.3	0.29	-1054.9	**-0.021 (−0.025 to −0.016)* ****	-	-	-	**0.015 (0.009 to 0.021)* ****	**-0.004 (−0.006, −0.002)* ****	-	-
Model 5.4	0.24	-1079.4	**-**	**-0.015 (−0.018 to −0.012)* ****	-	**0.122 (0.104 to 0.139)* ****	-	-	-	-
Model 5.5	0.26	-1089.6	**-**	**-0.014 (−0.017, −0.011)* ****	-	**0.113 (0.095, 0.132)* ****	**0.010 (0.004, 0.016)* ****	-	-	-
Model 5.6	0.27	-1101.7	**-**	**-0.013 (−0.016, −0.010)* ****	-	**0.056 (0.021, 0.091)* ***	**0.014 (0.008, 0.020)* ****	**-0.005 (−0.005, −0.002)* ****	-	-
Model 5.7	0.27	-1103.4	**-**	**-0.013 (−0.016, −0.010)* ****	-	**0.056 (0.021, 0.091)* ***	**0.017 (0.010, 0.024)* ****	**-0.003 (−0.005, −0.001)* ***	**-0.001 (−0.002, −0.000)***	-
Model 5.8	0.28	-1104.5	**-**	**-0.014 (−0.017, −0.011)* ****	-	**0.065 (0.028, 0.101)* ****	**0.017 (0.010, 0.024)* ****	**-0.004 (−0.006, −0.002)* ****	**-0.007 (−0.014, −0.000)***	0.008 (−0.001, 0.017)
Model 5.9	0.20	-1039.1	-	**-**	**-0.004 (−0.006, −0.003)* ****	**0.077 (0.059, 0.096)* ****	-	-	-	-
Model 5.10	0.22	-1058.1	-	**-**	**-0.005 (−0.006, −0.003)* ****	**0.066 (0.047, 0.085)* ****	**0.014 (0.008, 0.019)* ****	-	-	-
Model 5.11	0.24	-1071.5	-	**-**	**-0.004 (−0.006, −0.003)* ****	0.007 (−0.028, 0.042)	**0.017 (0.011, 0.023)* ****	**-0.004 (−0.006, −0.002)* ****	-	-
Model 5.12	0.29	-1089.6	-	**-0.011 (−0.014, −0.008)* ****	**-0.003 (−0.004, −0.001)* ****	**0.053 (0.016, 0.090)* ***	**0.015 (0.008, 0.022)* ****	**-0.004 (−0.006, −0.002)* ****	**-0.007 (−0.014, −0.001)***	**0.009 (0.000, 0.018)***

Table 7
Multivariable regression analyses for total hip BMD. Multivariable regression was done using total hip BMD as the dependent variable, with BMFF at the total hip as the primary explanatory variable. Other explanatory covariables were selected, models constructed, and data presented as described for Table 5.

			Covariable				
	Adj. R^2	AIC	BMFF Total Hip	Sex (M)	BMI	Legs fat%	VAT mass (kg)
Model 6.1	0.34	-995.0	**-0.017 (−0.020 to −0.014) * ****	**0.170 (0.151 to 0.189) * ****	-	-	-
Model 6.2	0.37	-1026.8	**-0.016 (−0.019 to −0.013) * ****	**0.156 (0.137 to 0.175) * ****	**0.018 (0.012 to 0.024) * ****	**-**	**-**
Model 6.3	0.40	-1055.2	**-0.015 (−0.018 to −0.012) * ****	**0.068 (0.032 to 0.104) * ****	**0.024 (0.017 to 0.030) * ****	**-0.005 (−0.007 to −0.003) * ****	**-**
Model 6.4	0.40	-1049.3	**-0.015 (−0.018 to −0.012) * ****	**0.072 (0.034 to 0.109) * ****	**0.025 (0.018 to 0.032) * ****	**-0.005 (−0.007 to −0.004) * ****	-0.013 (−0.041 to 0.015)

fat% (Model 7.3–7.7). This suggests that males may have greater femoral shaft BMD because they have a higher BMI and lower % leg fat than females.

4. Discussion

Herein, we have developed a new deep learning method for analysis of BM adiposity using Dixon MRI data from the UKBB. This is the first study to establish deep learning for BM segmentation at multiple sites, and the first peer-reviewed study to do so, for any skeletal site, in the UKBB imaging study. Our models yield BMFF measurements that are consistent with previous observations, including sex differences in spinal BMFF and inverse associations with BMD. Moreover, empty or small segmentation outputs occur only in a minority of cases, mostly because of technical issues with UKBB source data rather than because of pathophysiological variation, and are readily excluded before BMFF analysis. This demonstrates the ability of our models to generate accurate, reliable BMFF measurements from the UKBB MRI data. We further reveal new site- and sex-specific associations that have not been reported previously, highlighting the potential of our methods to uncover new pathophysiological functions and implications of BMAT.

Table 8
Multivariable regression analyses for femoral shaft BMD. Multivariable regression was done using femoral shaft BMD as the dependent variable; explanatory covariables were selected, models constructed, and data presented as described for Table 5.

			Covariable						
	Adj. R^2	AIC	BMFF Diaphysis	Sex (M)	BMI	Legs fat%	VAT mass (kg)	Android fat%	Trunk fat%
Model 7.1	0.28	-693.2	**-0.015 (−0.018 to −0.012)*** **	**0.166 (0.143 to 0.188)*** **	-	-	-	-	-
Model 7.2	0.30	-711.7	**-0.015 (−0.018 to −0.012)*** **	**0.152 (0.128 to 0.175)*** **	**0.018 (0.011 to 0.025)*** **	-	-	-	-
Model 7.3	0.33	-743.1	**-0.015 (−0.017 to −0.012)*** **	0.038 (−0.007 to 0.083)	**0.025 (0.017 to 0.033)*** **	**-0.007 (−0.009 to −0.005)*** **	-	-	-
Model 7.4	0.33	-739.5	**-0.015 (−0.018 to −0.012)*** **	**0.049 (0.003 to 0.096)***	**0.029 (0.020 to 0.037)*** **	**-0.007 (−0.010 to −0.005)*** **	-0.032 (−0.068 to 0.004)	-	-
Model 7.5	0.34	-747.1	**-0.015 (−0.018 to −0.012)*** **	0.042 (−0.003 to 0.086)	**0.031 (0.022 to 0.040)*** **	**-0.006 (−0.009 to −0.004)*** **	-	**-0.002 (−0.004 to −0.000)***	-
Model 7.6	0.34	-745.3	**-0.015 (−0.018 to −0.012)*** **	0.039 (−0.006 to 0.084)	**0.030 (0.021 to 0.039)*** **	**-0.006 (−0.009 to −0.003)*** **	-	-	**-0.002 (−0.004 to −0.000)***
Model 7.7	0.34	-743.4	**-0.015 (−0.018 to −0.012)*** **	**0.051 (0.002 to 0.100)***	**0.030 (0.021 to 0.039)*** **	**-0.008 (−0.011 to −0.005)*** **	0.003 (−0.048 to 0.055)	**-0.011 (−0.020 to −0.002)***	0.011 (−0.000 to 0.023)

4.1. Potential of multi-site BMFF analyses across the UK Biobank

The development and validation of our models using UKBB MRI data is hugely significant because, unlike most other MRI datasets, the UKBB also provides extensive genetic and phenotypic data for each subject, including whole-genome sequencing and health records. This linked data allows comprehensive association studies to identify the genetic and pathophysiological factors associated with FF and other MRI-derived measurements. Indeed, Liu et al. recently demonstrated the power of this approach using deep learning for segmentation of abdominal organs from UKBB MRI data [17]. They identified genetic variants and clinical conditions associated with FF and other imaging-derived characteristics for each organ, as well as combinations of characteristics across multiple organs. Thus, by allowing multi-site BMFF measurements across the UKBB cohort, our models promise, for the first time, to reveal the genetic, physiological and clinical variables associated with BMFF.

4.2. Deep learning for large-scale BM analysis

Several other recent studies have developed deep learning for automated BM segmentation from MRI data. For example, von Brandis et al. assessed the feasibility of deep learning for segmenting BM from T2-weighted Dixon water-only images, focusing on the knee region [22]; however, the best median dice score of their model was only 0.68, far below that obtained by our models (Table 2). Better accuracy was achieved by Zhou et al., who established a deep learning model for segmenting lumbar vertebrae from Dixon MRI data [20]. They trained their model using manual segmentations of 165 vertebrae from 31 subjects, with the model then tested on a validation set of 24 subjects. They achieved an average dice score of 0.849, below the accuracy of our vertebral ROI-Attention-U-Net (Table 2). More recently, Zhao et al. used deep learning for segmenting lumbar vertebrae from modified Dixon MRI data, using a training set of 142 subjects and a validation set of 64 subjects [21]. Their model achieved a mean dice score of 0.912, the same as that obtained by our vertebral ROI-Attention-U-Net (Table 2). Thus, among deep learning models for segmenting vertebral BM, our model achieves an accuracy that is similar or greater than that obtained by others.

Notably, our study is the first to develop deep learning for BM segmentation at the femoral head, total hip and femoral diaphysis. This is important because the properties of BMAT vary according to skeletal location [1,7,8]. Thus, to fully understand the health implications of BMAT and its potential utility as a clinical biomarker, it is critical to assess BMFF at other sites. Indeed, as discussed below, we found that the associations between BMFF, age, BMD, BMI and peripheral adiposity differ according to the BM region assessed, underscoring the importance of assessing BMFF across multiple sites. Finally, our model includes dedicated error-checking steps to remove inaccurate segmentation outputs, which is essential for reliable analysis of large-scale MRI data.

4.3. New ROI attention U-Net model

Another advance of the present study is our development of a new lightweight ROI attention U-Net model that allows accurate segmentation of small VOIs from large volumetric data. The traditional 3D U-Net has a fixed receptive field that is dependent on the size of convolutional kernels and network depth. To achieve state-of-the-art performance, the network architecture needs to be carefully designed to fit the sizes of the segmented objects and image resolution. As a result, in this study the traditional 3D U-Net generates highly accurate results for vertebrae and femoral head (Table 2), regions in which the segmented objects are relatively large. However, this traditional U-Net shows limited discriminative power when dealing with smaller structures such as the femoral diaphysis, where only a few pixels on each axial slice are annotated as foreground. On the contrary, our new ROI attention U-Net model can adaptively encode the local and global contextual information with its adjustive-attention mechanism. As shown in Table 2, it increases segmentation accuracy of the femoral diaphysis by over 25% and also slightly improves accuracy for the total hip region. Alongside these improvements, for the femoral head and vertebrae the ROI attention U-Net performs similarly to the carefully designed traditional 3D U-Net (Table 2).

Similar lightweight attention-based U-Net models have recently been developed for other imaging applications. Zhao et al. proposed such a model for segmentation of COVID-19 pneumonia lesions from 3D CT volumes [38], while Liu et al. developed an attention-based 3D model for brain tumour segmentation from MRI data [39]. A major difference between these studies and our segmentation task is that the femoral diaphysis is a particularly small anatomical target that commonly corresponds to < 5–10 pixels per slice; this is much smaller than COVID-19 pneumonia lesions and brain tumour segmentations. In addition, the pneumonia model from Zhao et al. increased the Dice score by 20.4% and produced an average score of only 78.7%; this relatively low accuracy may reflect this model's focus on reducing the network parameters to achieve a lighter weight, rather than for robust segmentation of small structures. In contrast, the brain tumour model from Liu et al. increased the Dice score by only 0.5 to 2%, yielding average scores of 79–90% [39]. Our model also differs to that of Liu et al. in two other ways. Firstly, they used a decoupled dilated convolutional operation and cascaded attention mechanism to extract multi-resolution features for a single receptive field. Secondly, they randomly cropped their training

images for training purposes. In contrast, our model shows a lighter weight theoretically at inference time because, rather than keeping more multi-resolution features, the decoder processes the feature maps of only the ROI, rather than the whole input image. As a result, we didn't require random cropping at the pre-processing stage, but instead removed the large empty backgrounds. Consequently, for our model the whole body is covered in the input data.

Attention-based models have also been developed for applications beyond biomedical imaging. For example, Zhu et al. developed an attention-based 3D model for human motion recognition, including extraction of both spatial and temporal features. Their model increased Dice scores by 5–10% over traditional models, yielding average scores of 84.8–91.6% [40]. Thus, our model's 25% improvement in diaphysis segmentation accuracy compares favourably to the accuracy gains produced by this and other recent attention-based U-Net models. Moreover, our attention-based model is the first to be developed for BM segmentation.

Our findings for scoliosis and Non-Hodgkin Lymphoma show that abnormal skeletal morphology or BM composition can impair segmentation. This did not occur for all cases of these diseases and appears to be limited to more-severe cases, confirming that this is not a universal limitation of our models. We also found that signal inhomogeneities in the proximal femur can disrupt femoral head and total hip segmentation (Fig. 4B); it is possible that these also have a biological basis, for example resulting from distinct foci of red marrow within the proximal femur. However, our comprehensive PheCode analysis shows that, generally, segmentation is not compromised by skeletal diseases. This type of segmentation failure is therefore likely to be relatively rare across the full UKBB cohort. If necessary, we will re-train our models to ensure that any common pathological abnormalities do not compromise segmentation.

Taken together, our new ROI attention model is the first accurate deep learning method designed for BM segmentation across multiple skeletal sites and varied anatomical sizes.

4.4. Association between BMFF and pathophysiological characteristics – confirmation of previous studies and new findings

The key aim of our study was to develop and validate deep learning models for automated BM segmentation of UKBB Dixon MRI data. Our group of 729 subjects is the largest cohort yet to undergo measurement of spinal BMFF, and by far the largest to include assessment of BMFF at any femoral site [12]. Consistent with previous reports, we find that spinal BMFF is lower than femoral BMFF (Fig. 4) [1,12,36]; is greater in females than in males (Fig. 4) [34,35]; increases with age (Supplemental Table 5) [12,34,35,41]; is elevated in osteopaenia or osteoporosis, at least in females (Fig. 5) [1,6,12]; exhibits a robust, inverse association with spinal BMD (Table 5) [1,6,12]; and is positively associated with visceral adiposity (Supplemental Table 5) [41,42].

Our results for femoral BMFF are also consistent with previous studies. For example, in a cohort of aged females, Griffith et al. found that BMFF in the femoral head, neck and diaphysis is increased in osteoporosis and inversely associated with BMD at each site [43]. We confirm these findings (Fig. 5, Tables 6–8) and further reveal that diaphyseal BMFF is typically inversely associated with peripheral adiposity in females but not in males, while BMFF at the femoral head or total hip is generally not associated with these peripheral adiposity traits (Supplemental Table 5); these observations confirm and extend those of a previous smaller-scale study [44]. The reasons for these variable site- and sex-dependent relationships between BMFF and peripheral adiposity remain to be determined; however, one possibility is that they reflect preferences for the partitioning of lipid storage between different adipose depots.

Many of our new findings relate to the fact that most previous MR-based studies of BM adiposity have focussed on vertebrae, with femoral sites being relatively overlooked [12]. For example, we show that, across both sexes, BMFF is highest in the femoral head and decreases progressively in the total hip and diaphysis, while BMFF at each femoral site is greater in males than in females (Fig. 5A). Unlike in the spine, age shows no relationship with BMFF at each femoral site (Supplemental Table 5). This could reflect the fact that, compared to the spine, these femoral sites contain a greater proportion of constitutive BMAT, which is less age responsive than the regulated BMAT that predominates in the axial skeleton [7,8]. However, it may be that age-related increases in femoral BMAT occur over a longer timeframe that would only be apparent when BMFF is assessed over a greater age range. If so, this should become apparent through BMFF analysis across the full UKBB imaging cohort.

Regarding constitutive vs regulated subtypes, we also find robust positive associations between BMFF at the four different sites analysed (Table 4), similar to the findings of Slade et al. [36]. However, we further reveal that these relationships exhibit sex differences and are strongest between the three femoral regions, with spinal BMFF showing no association with diaphyseal BMFF (Table 4). This may reflect differences in the development and function of regulated vs constitutive BMAT [7,8].

Together, our present findings confirm those of previous studies while also revealing new knowledge about BMAT's site- and sex-dependent characteristics. This underscores the ability of our deep learning models to yield reliable BMFF measurements and to identify new insights into the pathophysiology of BMAT.

4.5. Limitations

Despite these advances, there are several limitations to highlight. Firstly, our models were trained and tested using manual segmentations from only a single reader. In contrast, two previous BM segmentation models were trained and tested using manual segmentations produced by two independent human readers [20,22]; this multi-reader approach can help to ensure consistency in the ground truths. However, single-reader ground truths have also been used to successfully develop other recent deep learning models for bone or BM segmentation [21,45], and our ground truths were produced by a reader with extensive experience. Moreover, our deep learning segmentations yield BMFF values consistent with many established findings, as discussed above. Therefore, we can be confident that our single-reader segmentations provided reliable ground truths for robust model development.

A second limitation is that our models did not produce segmentations for all participants. As discussed above, this was generally not a result of pathological skeletal abnormalities; instead, in most cases it resulted from deviations in the structure of source data or image quality provided by UKBB, something that cannot be readily overcome by UKBB users. The next most-common cause of faulty segmentations was positioning issues during MRI acquisition, resulting in the target site (femoral head, total hip, or diaphyseal midpoint) falling partially or fully outside the expected MRI slab. If this issue persists across the full imaging cohort, then we will update our method by re-training our femoral head, total hip, and diaphysis models to segment slab volumes adjacent to the current target slabs and testing if this generates reliable segmentations for any affected participants. However, these positioning issues affected only 54 outputs, corresponding to $< 2\%$ of all segmentation outputs from our validation cohort. Another 37 of the faulty segmentations had no obvious cause of failure. While this represents only $\sim$1% of all segmentation outputs, it suggests that other unidentified factors can impair segmentation. We will further investigate this after applying our models across the full UKBB imaging cohort, which should allow the causes of any impaired segmentations to be more-comprehensively assessed. Importantly, the low failure rate means that the above issues should not substantially compromise BMFF analysis across the full UKBB cohort.

A third specific limitation is that our cohort included relatively few osteoporotic males. This restricted our ability, in males, to detect significant effects of osteoporosis on BMFF at each site. Our univariable and

multivariable regression analyses were still able to detect significant inverse associations between BMFF and BMD at each site; however, once we have measured BMFF across the full available UKBB cohort it will be informative to reassess the relationship between BMFF and osteoporosis. This analysis will also allow us to better account for other potential confounding factors, such as physical activity, dietary habits, and other pathophysiological parameters that may influence the relationship between BM adiposity and health outcomes.

There are two more-general limitations. Firstly, the UKBB imaging study is cross-sectional and so provides data for only one timepoint. Therefore, it is not designed capture longitudinal changes in BMFF and how these relate to health outcomes. A second general limitation relates to the UKBB MRI protocol, and in particular the use of two-point Dixon sequences. Participants in the UKBB imaging study visited several different imaging centres for acquisition of the MRI scans. Therefore, across these different imaging centres the MRI protocol parameters had to be standardised and harmonised, resulting in both advantages and drawbacks. For example, to simplify the procedure the Dixon sequences were based on only two echo times; however, with only dual-echo sequences, no accurate T2 * -correction could be applied and the complexity of the fat spectrum could not be considered in the BMFF mapping [10,15]. As a result, reported BMFF measurements can be affected by T2 * decay effects caused by the presence of trabecular bone, which in turn may differ in the water and fat components [9,10]. However, the moderately low flip angle (10°) is acceptable to limit T1-bias, and protocol standardisation compelled all examinations to be performed in similar conditions, with the exact same parameters [9,46]. Consequently, even if the more-accurate PDFF could not be quantified, a comparable estimate could be obtained through the reported BMFF, which permits group comparison and method cross-validation. Indeed, considering the sensitivity to detect BMFF changes between groups, the very large number of subjects in the UKBB imaging study helps to reduce any bias resulting from T2 * effects and thereby limits the improved sensitivity that is typically gained from multi-echo PDFF measurements. Furthermore, dual-echo Dixon-derived BMFF allows the derivation of consistent 3D BMFF measurements across all UKBB MR imaging centres. This is very important for our BMFF validation study, as it allowed us to assess and automate extraction of BMFF maps from multiple skeletal sites, on a 3D mode.

5. Conclusions

Our new deep learning models allow accurate segmentation and BMFF measurements for the spine, femoral head, total hip, and femoral diaphysis from UKBB MRI data. While we have used these models to analyse BM, they are generally applicable for improved segmentation of small VOIs from any large volumetric MRI data. Thus, they could also be applied for precise, automated, large-scale analysis of other small anatomical structures of interest. We will next use our deep learning models to measure BMFF across the full UKBB imaging cohort, which will eventually include 100,000 subjects. This will allow us to identify the genetic, physiological and clinical conditions associated with altered BMFF at each site. Such knowledge will help to elucidate the mechanisms that influence BM adiposity and reveal, to an unprecedented extent, how BMAT impacts human health and disease.

Funding sources

This work was supported by a grant from the Medical Research Council (MR/S010505/1 to W.P.C., including support for W.X.). W.P.C. was further supported by a Chancellor's Fellowship from the University of Edinburgh. The British Heart Foundation supported C.W. (RG/16/10/32375) and S.S. (4-year BHF PhD studentship). C.D.G. and T.M. were supported by the Edinburgh Clinical Research Facility and NHS Lothian R&D.

CRediT authorship contribution statement

The authors confirm that they have each made the following contributions:

Conceptualisation, W.P.C.; Data curation, D.M.M., C.W., G.P., C.D.G., W.X., S.S. and W.P.C.; Formal Analysis, D.M.M., C.W., G.P., C.D.G., W.X., S.S. and W.P.C.; Funding Acquisition, S.I.K.S., T.M. and W.P.C.; Investigation, D.M.M., C.W., G.P., W.X., S.S. and W.P.C.; Methodology, D.M.M., C.W., G.P., C.D.G., W.X., S.S., S.B., J.P., S.I.K.S., T.M. and W.P.C.; Project administration, S.I.K.S., T.M. and W.P.C.; Resources, S.I.K.S., T.M. and W.P.C.; Software, D.M.M., C.W., G.P.; Supervision, S.I.K.S., T.M. and W.P.C.; Visualisation, D.M.M., C.W., C.D.G. and W.P.C.; Writing – Original Draft, D.M.M., C.W., S.B., J.P. and W.P.C.; Writing – Review & Editing, D.M.M., C.W., G.P., C.D.G., W.X., S.S., S.B., J.P., S.I.K.S., T.M. and W.P.C.

Declaration of Competing Interest

G.P. is currently an employee of Pfizer; however, Pfizer had no role in the design or interpretation of this research. All other authors declare no competing interest.

Acknowledgements

This work was supported by a grant from the Medical Research Council (MR/S010505/1 to W.P.C., including support for W.X.). W.P.C. was further supported by a Chancellor's Fellowship from the University of Edinburgh. The British Heart Foundation provided support to C.W. (RG/16/10/32375) and S.S. (4-year PhD Studentship). C.G. and T.M. were supported by the Edinburgh Clinical Research Facility and NHS Lothian R&D.

We are grateful to Dominic Job (Edinburgh Imaging, University of Edinburgh) for support with IT infrastructure, and Jimmy Bell, Louise Thomas and Brandon Whitcher (University of Westminster) for helpful discussions and advice regarding working with UKBB MRI data.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.12.029.

References

[1] Cawthorn, W.P. (2020) Bone Marrow Adipose Tissue. in Encyclopedia of Bone Biology (Zaidi, M. ed.), Oxford: Academic Press, Oxford, UK. pp 156–177. doi: 10.1016/B978-0-12-801238-3.11207-3.
[2] Devlin MJ, Cloutier AM, Thomas NA, Panus DA, Lotinun S, et al. Caloric restriction leads to high marrow adiposity and low bone mass in growing mice. J Bone Miner Res 2010;25:2078–88.
[3] Cawthorn WP, Scheller EL, Learman BS, Parlee SD, Simon BR, et al. Bone marrow adipose tissue is an endocrine organ that contributes to increased circulating adiponectin during caloric restriction. Cell Metab 2014;20:368–75.
[4] Cawthorn WP, Scheller EL, Parlee SD, Pham HA, Learman BS, et al. Expansion of bone marrow adipose tissue during caloric restriction is associated with increased circulating glucocorticoids and not with hypoleptinemia. Endocrinology 2016;157:508–21.
[5] Suchacki KJ, Tavares AAS, Mattiucci D, Scheller EL, Papanastasiou G, et al. Bone marrow adipose tissue is a unique adipose subtype with distinct roles in glucose homeostasis. Nat Commun 2020;11:3097.
[6] Veldhuis-Vlug AG, Rosen CJ. Clinical implications of bone marrow adiposity. J Intern Med 2018;283:121–39.
[7] Craft CS, Li Z, MacDougald OA, Scheller EL. Molecular differences between subtypes of bone marrow adipocytes. Curr Mol Biol Rep 2018;4:16–23.

[8] Scheller EL, Doucette CR, Learman BS, Cawthorn WP, Khandaker S, et al. Region-specific variation in the properties of skeletal adipocytes reveals regulated and constitutive marrow adipose tissues. Nat Commun 2015;6:7808.
[9] Tratwal J, Labella R, Bravenboer N, Kerckhofs G, Douni E, et al. Reporting guidelines, review of methodological standards, and challenges toward harmonization in bone marrow adiposity research. report of the methodologies working group of the international bone marrow adiposity society. Front Endocrinol 2020;11.
[10] Karampinos DC, Ruschke S, Dieckmeyer M, Diefenbach M, Franz D, et al. Quantitative MRI and spectroscopy of bone marrow. J Magn Reson Imaging 2018; 47:332–53.
[11] Cordes C, Baum T, Dieckmeyer M, Ruschke S, Diefenbach MN, et al. MR-based assessment of bone marrow fat in osteoporosis, diabetes, and obesity. Front Endocrinol 2016;7:74.
[12] Sollmann N, Löffler MT, Kronthaler S, Böhm C, Dieckmeyer M, et al. MRI-based quantitative osteoporosis imaging at the spine and femur. J Magn Reson Imaging 2021;54:12–35.
[13] Shen W, Chen J, Gantz M, Punyanitya M, Heymsfield SB, et al. MRI-measured pelvic bone marrow adipose tissue is inversely related to DXA-measured bone mineral in younger and older adults. Eur J Clin Nutr 2012;66:983–8.
[14] Shen W, Velasquez G, Chen J, Jin Y, Heymsfield SB, et al. Comparison of the Relationship Between Bone Marrow Adipose Tissue and Volumetric Bone Mineral Density in Children and Adults. J Clin Densitom 2014.
[15] Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. Nat Commun 2020;11:2624.
[16] Kart T, Fischer M, Küstner T, Hepp T, Bamberg F, et al. Deep Learning-Based Automated Abdominal Organ Segmentation in the UK Biobank and German National Cohort Magnetic Resonance Imaging Studies. Invest Radiol 2021;56: 401–8.
[17] Liu Y, Basty N, Whitcher B, Bell JD, Sorokin EP, et al. Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning. eLife 2021;10: e65554.
[18] Suinesiaputra A, Sanghvi MM, Aung N, Paiva JM, Zemrak F, et al. Fully-automated left ventricular mass and volume MRI analysis in the UK Biobank population cohort: evaluation of initial results. Int J Cardiovasc Imaging 2018;34:281–91.
[19] Kaufmann, T., Bjørnstad, P.M., Falck, M., O'Connell, K., Frei, O., *et al.* (2022) Quantifying bone marrow adiposity and its genetic architecture from head MRI scans. medRxiv, 2022.2008.2019.22278950.
[20] Zhou J, Damasceno PF, Chachad R, Cheung JR, Ballatori A, et al. Automatic Vertebral Body Segmentation Based on Deep Learning of Dixon Images for Bone Marrow Fat Fraction Quantification. Front Endocrinol (Lausanne) 2020;11.
[21] Zhao Y, Zhao T, Chen S, Zhang X, Sosa MS, et al. Fully automated radiomic screening pipeline for osteoporosis and abnormal bone density with a deep learning-based segmentation using a short lumbar mDixon sequence. Quant Imaging Med Surg 2022;12:1198–213.
[22] von Brandis E, Jenssen HB, Avenarius DFM, Bjørnerud A, Flatø B, et al. Automated segmentation of magnetic resonance bone marrow signal: a feasibility study. Pediatr Radiol 2022;52:1104–14.
[23] West J, Dahlqvist Leinhard O, Romu T, Collins R, Garratt S, et al. Feasibility of MR-Based Body Composition Analysis in Large Scale Population Studies. PLoS One 2016;11:e0163332.
[24] Wu P, Gifford A, Meng X, Li X, Campbell H, et al. Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. JMIR Med Inf 2019;7:e14325.
[25] Bastarache L. Using Phecodes for Research with the Electronic Health Record: From PheWAS to PheRS. Annu Rev Biomed Data Sci 2021;4:1–19.
[26] Gondim Teixeira, P.A., Cherubin, T., Badr, S., Bedri, A., Gillet, R., *et al.* (2019) Proximal femur fat fraction variation in healthy subjects using chemical shift-encoding based MRI. Scientific reports 9, 20212.
[27] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. Cham: Springer International Publishing; 2016.
[28] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Cham: Springer International Publishing; 2015.
[29] Woo S, Park J, Lee J-Y, Kweon IS. CBAM: Convolutional Block Attention Module. Cham: Springer International Publishing,; 2018.
[30] Wang, X., Girshick, R.B., Gupta, A., and He, K. (2018) Non-local Neural Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7794–7803.
[31] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., et al. (2017) Automatic differentiation in PyTorch. 31st Conference on Neural Information Processing Systems (NIPS 2017). Long Beach, CA, USA.
[32] Kingma, D.P., and Ba, J. Adam: A Method for Stochastic Optimization. arXiv 2015; 1412.6980v9. doi: 10.48550/arXiv.1412.6980.
[33] Harrison, E., Drake, T., and Ots, R. (2022) finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling. R package version 1.0.5 Ed.
[34] Griffith JF, Yeung DK, Ma HT, Leung JC, Kwok TC, et al. Bone marrow fat content in the elderly: a reversal of sex difference seen in younger subjects. J Magn Reson Imaging 2012;36:225–30.
[35] Baum T, Rohrmeier A, Syväri J, Diefenbach MN, Franz D, et al. Anatomical variation of age-related changes in vertebral bone marrow composition using chemical shift encoding-based water–fat magnetic resonance imaging. Front Endocrinol 2018;9:141.
[36] Slade JM, Coe LM, Meyer RA, McCabe LR. Human bone marrow adiposity is linked with serum lipid levels not T1-diabetes. J Diabetes Complicat 2012;26:1–9.
[37] Soliman AH. Diagnostic and prognostic relevance of bone marrow microenvironment components in non hodgkin's lymphoma cases before and after therapy. Asian Pac J Cancer Prev 2016;17:5273–80.
[38] Zhao, Q., Wang, H., and Wang, G. (2021) LCOV-NET: A Lightweight Neural Network For COVID-19 Pneumonia Lesion Segmentation From 3D CT Images. in 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). doi: 10.1109/ISBI48211.2021.9434023.
[39] Liu H, Huo G, Li Q, Guan X, Tseng M-L. Multiscale lightweight 3D segmentation algorithm with attention mechanism: brain tumor image segmentation. Expert Syst Appl 2023;214:119166.
[40] Zhu M, Bin S, Sun G. Lite-3DCNN combined with attention mechanism for complex human movement recognition. Comput Intell Neurosci 2022;2022:4816549.
[41] Hasic D, Lorbeer R, Bertheau RC, Machann J, Rospleszcz S, et al. Vertebral Bone Marrow Fat Is independently Associated to VAT but Not to SAT: KORA FF4—Whole-Body MR Imaging in a Population-Based Cohort. Nutrients 2020;12: 1527.
[42] Bredella MA, Torriani M, Ghomi RH, Thomas BJ, Brick DJ, et al. Vertebral Bone Marrow Fat Is Positively Associated With Visceral Fat and Inversely Associated With IGF-1 in Obese Women. Obesity 2011;19:49–53.
[43] Griffith JF, Yeung DK, Tsang PH, Choi KC, Kwok TC, et al. Compromised bone marrow perfusion in osteoporosis. J Bone Miner Res 2008;23:1068–75.
[44] Bredella MA, Fazeli PK, Miller KK, Misra M, Torriani M, et al. Increased bone marrow fat in anorexia nervosa. *J Clin Endocrinol Metab* 2009;94:2129–36.
[45] Klein A, Warszawski J, Hillengaß J, Maier-Hein KH. Automatic bone segmentation in whole-body CT images. Int J Comput Assist Radiol Surg 2019;14:21–9.
[46] Liu CY, McKenzie CA, Yu H, Brittain JH, Reeder SB. Fat quantification with IDEAL gradient echo imaging: correction of bias from T(1) and noise. Magn Reson Med 2007;58:354–64.