ADAM: automated digital phenotyping and morphological texture analysis of bone biopsy images using deep learning

Satvika Bharadwaj[1] , Florence Lima[2] , Tilak Bahadur Pathak[1], Rohan Dhamdhere[1], Pingfu Fu[3] , Hartmut Malluche[2], Madhumathi Rao[2] , Anant Madabhushi[1,4,*]

[1]Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA 30332, United States
[2]Division of Nephrology, Bone and Mineral Metabolism, University of Kentucky, Lexington, KY 40508, United States
[3]Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, United States
[4]Atlanta Veterans Affairs Medical Center, Atlanta, GA 30033, United States

*Corresponding author: Anant Madabhushi, Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, 1750 Haygood Drive, Atlanta, GA 30322, United States (anantm@emory.edu).

Abstract

Histomorphometric analysis of undecalcified bone biopsy images provides quantitative assessment of bone turnover, volume, and mineralization using static and dynamic parameters. Traditionally, quantification has relied on manual annotation and tracing of relevant tissue structures, a process that is time-intensive and subject to inter-operator variability. We developed ADAM, an automated pipeline for digital phenotyping, to quantify tissue and cellular components pertinent to static histomorphometric parameters such as bone and osteoid area, osteoclast and osteoblast count, and bone marrow adipose tissue (BMAT) area. The pipeline allowed rapid generation of delineated tissue and cell maps for up to 20 images in less than a minute. Comparing deep learning-generated annotation pixels with manual annotations, we observed Spearman correlation coefficients of $\rho = 0.99$ for both mineralized bone and osteoid, and $\rho = 0.94$ for BMAT. For osteoclast and osteoblast cell counts, which are subject to morphologic heterogeneity, using only brightfield microscopic images and without additional staining, we noted $\rho = 0.60$ and 0.69, respectively (inter-operator correlation was $\rho = 0.62$ for osteoclast and 0.84 for osteoblast count). The study also explored the application of morphological texture analysis (MTA), measuring relative pixel patterns that potentially vary with diverse tissue conditions. Notably, MTA from mineralized bone, osteoid, and BMAT showed differentiating potential to identify common pixel characteristics between images labeled as low or high bone turnover based upon the final diagnostic report of the bone biopsy. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) obtained for BMAT MTA features as a classifier for bone turnover, was 0.87, suggesting that computer-extracted features, not discernable to the human eye, hold potential in classifying tissue states. With additional evaluation, ADAM could be potentially integrated into existing clinical routines to improve pathology workflows and contribute to diagnostic insights into bone biopsy evaluation and reporting.

Keywords: bone histomorphometry, osteoclasts, osteoblasts, diseases and disorders of/related to bone, disorders of calcium/phosphate metabolism
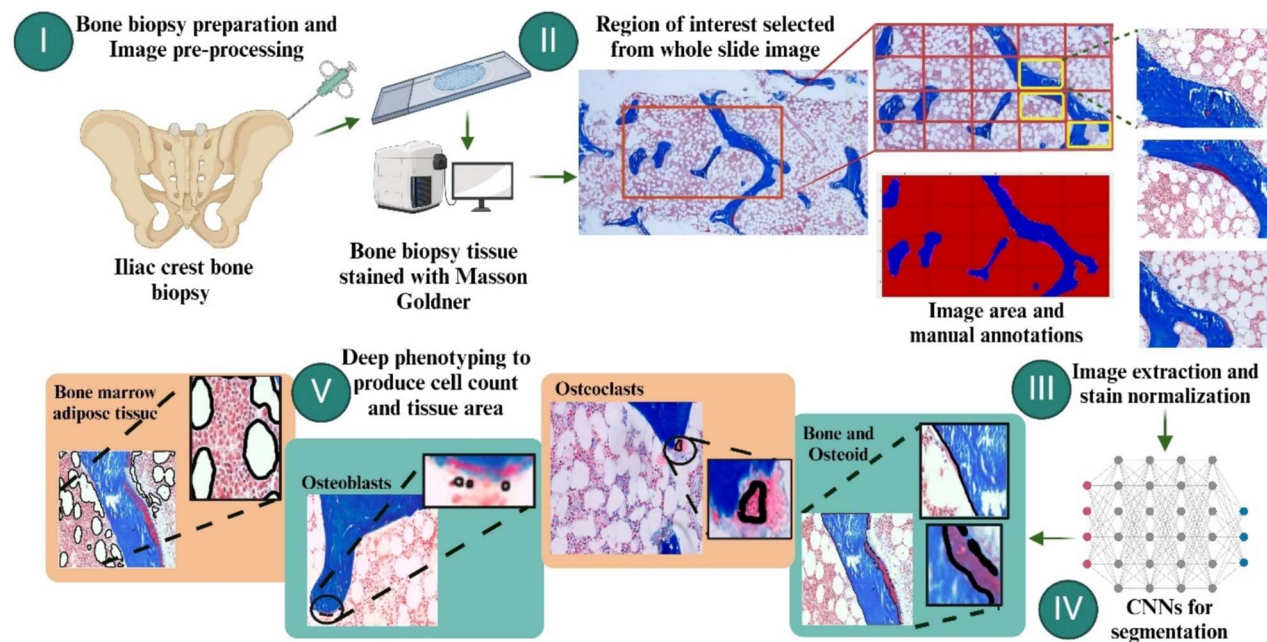
Lay Summary

Bone biopsies are used to study bone diseases at a microscopic level. For precise measurements, detailed tracings of tissue and cell structures are hand-drawn on images, a process that is time-consuming and prone to variability between observers. Advances in artificial intelligence, particularly deep learning, enable automatic identification of tissue and cell parts using their shape, color, texture, and location. We developed ADAM, the first automated deep learning-based pipeline to trace tissue and cellular structures, and extract texture patterns embedded in images of undecalcified bone samples. These automated tracings allow quicker and easier processing of bone biopsy images for diagnostic reporting.

Graphical Abstract

Introduction

Diagnostic bone biopsies play a pivotal role in evaluating metabolic bone disorders such as osteoporosis, osteomalacia, hyperparathyroid bone disease and renal osteodystrophy.[1–5] Although imaging modalities such as quantitative and high-resolution CT or MRI[6] offer non-invasive and 3D information on bone structure, it is through bone histopathology that we obtain reliable insights into bone formation and resorption, thereby explaining clinical phenotypes of bone loss and quality.[7] Undecalcified bone histology, used in conjunction with fluorochrome labeling[8] is the gold standard for evaluating bone turnover (coupling of bone formation and resorption[9]) and mineralization (incorporation of calcium phosphate nanocrystals into organic bone matrix, essential for the hardness and strength of bone[10]); these processes provide standard therapeutic targets for anti-remodeling, anabolic or other medications.[11]

Traditionally, histomorphometric reporting relies on visual identification of histologic primitives (structures) by experienced pathologists, taking approximately 60 min per slide.[12] Detailed feature maps[13] are manually created by point labeling cells like osteoblasts and osteoclasts, and contour tracing outlines of mineralized bone and osteoid for areal dimensions, erosion cavities for width and depth, and fluorochrome double labels for extent and inter-label width. In the early 1980s, Malluche et al.[14,15] developed a semi-automatic method based on stereological principles for calculation of 2D and 3D bone histomorphometric parameters. Initially marketed by Zeiss as Osteoplan,[16] these principles were also adapted by other specialized image analysis software such as Bioquant[17] and OsteoMeasure[18] that are now commonly used to quantify histomorphometric feature maps. The American Society of Bone and Mineral Research (ASBMR) Histomorphometry Nomenclature Committee ASBMR committee has standardized reporting and nomenclature[20] for both static and dynamic parameters. Static parameters include perimeters of mineralized bone and osteoid, osteoblast and osteoclast numbers and dimensions of bone-osteoblast, bone-osteoclast interfaces, at a single timepoint; dynamic histomorphometry assesses fluorochrome label uptake from two or more time points to indirectly estimate bone mineralization over time.[21] Measurements and their precision rely largely on the expertise of the pathologist, are time-consuming, and can introduce subjectivity and inter- and intra- operator variability.[22,23] As a result, histomorphometric reporting of bone biopsies is often limited to research settings, while clinical diagnostics rely more on qualitative interpretation.[19]

Efforts at automated and semi-automated quantification methods have thus far been limited[21,23] to mice models[21,24] and require a certain degree of manual intervention and therefore run-time. However, the advent of deep learning (DL) models, particularly nnU-Net,[25] introduced a powerful semantic-based network for biomedical image segmentation that focuses on spatial relationships between pixels[26] rather than just pixel color information. Furthermore, computational methods allow for the extraction of embedded image texture information, usually not discernable to the human eye.[26] These texture features, collectively termed as morphological texture analysis (MTA) include occurrences and patterns of relative pixel intensities and have found application in radiology[27] and pathology[28] studies to classify normal and abnormal tissue states.

In this work, we present ADAM, an end-to-end pipeline for digital phenotyping of brightfield bone biopsy images to generate feature maps for static histomorphometry by delineating histologic primitives. We used a set of 443 bone histology images annotated by two of the authors (MR, FL) with training and experience in bone histomorphometry. These images were selected expressly for feature identification

for training DL models. Algorithms were constructed for the following:

1. Feature mapping and cell segmentation: to delineate mineralized bone, osteoid, osteoblasts, osteoclasts, the intent being to reduce the time required[21] for manual annotations. We assessed the agreement between feature maps generated by our DL models against manually annotated feature maps.

2. Additionally, MTA was performed on histologic primitives segmented in (a). This additional and novel exploratory step generated a methodology for extracting and identifying unique structured or unstructured pixel patterns in mineralized bone, osteoid and bone marrow adipose tissue (BMAT), that could potentially associate with normal or anomalous tissue states.

Materials and methods
Bone biopsy preparation
The study utilized a selected set of 443 histology images from anterior iliac crest bone biopsies from 29 patients (female 77%, age range 22-71 yr) investigated for metabolic bone diseases. Bone core samples were approximately 0.3 cm in diameter, with adequate representation of cortical and trabecular bone compartments. Bone cores were fixed in ethanol at room temperature, subjected to dehydration, and then embedded in methyl methacrylate, following the procedures outlined by Malluche et al.[29] Sections were stained with the modified Masson–Goldner trichrome stain (Masson, 1929; Goldner, 1938), the Aurin tricarboxylic acid stain (Lillie, 1976) and solochrome azurine stain (Denton, 1984).

Slide digitization and image selection
Brightfield light microscopy images were acquired through an MTI color video camera attachment to a Zeiss Axioplan microscope, interfaced with OsteoMeasure (OsteoMetrics, Inc.), a widely used commercial software for histomorphometric analysis.[30] A total of 29 regions of interest (ROIs) or image areas were selected from 29 whole slide images (Figure 1, step II) based on representative tissue characteristics that provided the morphologic diversity needed for training of the machine learning algorithms (Figure S1). Image areas in biopsies from patients with osteoporosis provided variability in mineralized bone area whereas image areas from patients with osteomalacia provided variability for unmineralized osteoid regions. Diversity in cellularity was captured using image areas from low-turnover osteoporosis (low cellularity) and hyperparathyroid states (high cellularity) in patients with renal osteodystrophy. A total of 443 images at a magnification of 20x were in turn extracted from the 29 image areas, based on the grid generated by the OsteoMeasure software. These images containing the representative features were then used to train and test the DL models.

Annotation and image preprocessing steps
The images were processed to match the image color statistics of a reference image using the Reinhard color normalization technique from tissue image analytics stain normalization toolbox,[31] to address any variation in color maps (Figure S1). Furthermore, all images were cropped and resized from an initial size of $1171 \times 1580 \times 3$ to a standard size of $512 \times 512 \times 3$ to ensure consistency with the requirements of the convolutional neural network.

Two of the authors (M.R., F.L.) trained and experienced in histomorphometry, annotated the histologic features using OsteoMeasure to create feature maps. These manually annotated feature maps served as the reference data, and provided primary static histomorphometric indices, namely, area of mineralized bone and osteoid, osteoblast number and osteoblast surface, osteoclast number, osteoclast surface and outlines of BMAT areas.

Deep learning dataset
Distinct nnUnet DL models for tissue and cellular elements in biopsies were developed. Table S1 summarizes the data split for each histologic primitive based on the annotations available for model training. Leveraging the nnU-Net deep learning architecture, the models were rigorously trained to achieve precise segmentation, considering not just color, and shape characteristics but also the spatial information for each histologic primitive. The train and test datasets were independent and different depending on the annotations available for the histologic primitive being considered (Table S1).

Digital phenotyping of bone biopsy images: automated segmentation of tissue and cellular components and morphological texture analysis
Mineralized bone and osteoid
The mineralized bone and osteoid regions have distinct blue and red regions respectively in Masson-Goldner trichrome stained bone biopsies (Figure 2). Traditional segmentation techniques like global thresholding and connected component[32] analysis were initially considered. However, due to diverse color maps of image areas (Figure S1), these techniques would be inadequate and incapable of filtering out artifacts caused by the folding or tearing of tissue, especially in thin biopsy slices. A single nnUnet model was employed for the segmentation of bone and osteoid regions. The train dataset comprised of 107 annotated images of size $512 \times 512 \times 3$ and the independent test set consisted of 77 images.

Osteoclasts
Osteoclasts, responsible for bone resorption, are generally situated along the surface of mineralized bone.[33] Due to the limited number of osteoclast labels within images, additional augmentation techniques such as rotation and reflection were deployed to expand the training dataset. Dividing the image size from $512 \times 512 \times 3$ into 64 $64 \times 64 \times 3$ images provided finer local details of the cellular structures, improving the generalizability and adaptability of the deep learning model to heterogenous presentations. An nnUnet model was trained using 5696 subdivided images (from 89 images of size $512 \times 512 \times 3$) with osteoclastic activity and was tested in 1987 subdivided images (from 31 images of size $512 \times 512 \times 3$). The resultant images from the DL model were reconstructed to the original image size of $512 \times 512 \times 3$ images (Figure 3). This was followed by the implementation of a cell counting algorithm to determine the number of osteoclasts in each reconstructed image.

Osteoblasts
Osteoblasts are responsible for bone formation and are typically found bordering osteoid seams.[33] Segmentation of osteoblasts was achieved through the integration of
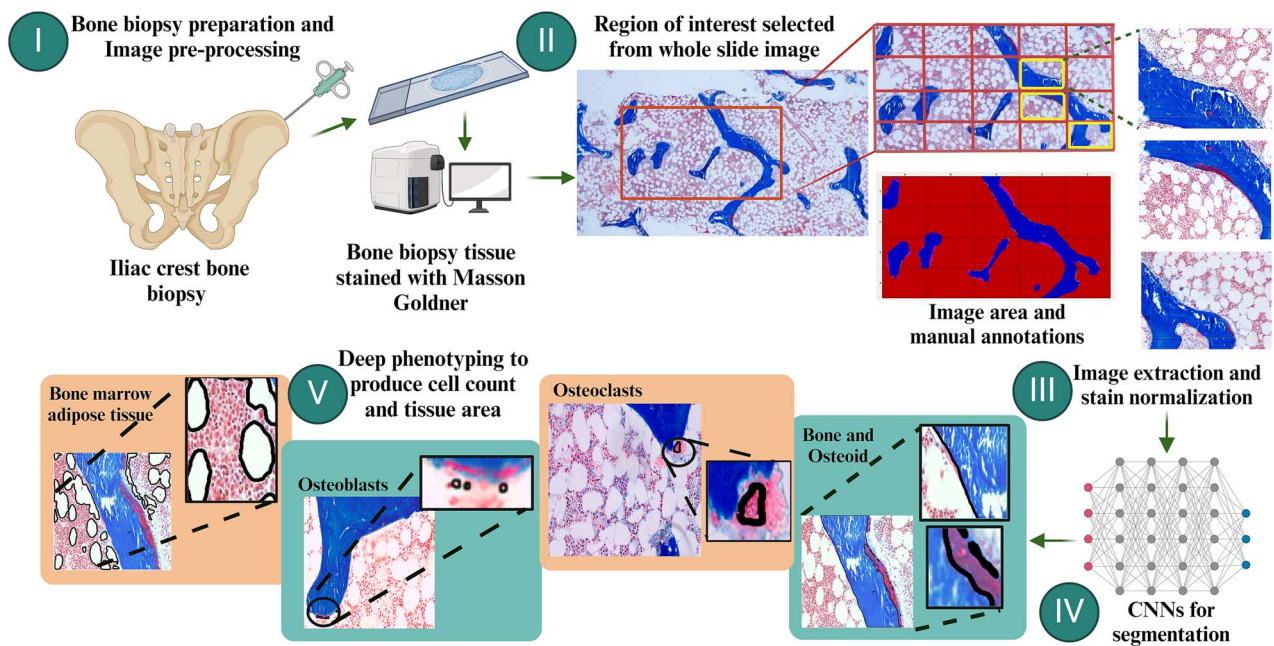
Figure 1. ADAM for digital phenotyping of bone biopsy images; step I bone biopsy preparation: Masson–Goldner trichrome stained bone biopsies are subjected to slide digitization; step II involves selection of image areas (for a one-time deep learning training process) which are further subjected to image extraction and stain normalization (step III); step IV comprises deep learning models for digital phenotyping, followed by quantification of cells and tissue area segmented (step V).

computer vision and deep learning techniques as their structure proved to be more complex than the relatively large osteoclasts. The final model was a fusion of computer vision and deep learning techniques, as shown in Figure S2. The DL model, similar to osteoclast segmentation, used $512 \times 512 \times 3$ images each subdivided into 64 $64 \times 64 \times 3$ images. The model was trained using 4416 subdivided images (from $69\ 512 \times 512 \times 3$ images) and 1920 subdivided images to test (from $30\ 512 \times 512 \times 3$ images). Subsequently, the DL model's output, which consisted of smaller $64 \times 64 \times 3$ images, was reassembled to recreate the original $512 \times 512 \times 3$ image dimensions (Figure 3).

The DL model's output images underwent post-processing with a computer vision algorithm, particularly focusing on regions near the osteoid where the model exhibited low confidence in its predictions (osteoid mask was obtained from bone-osteoid DL model described in Mineralized Bone and Osteoid). A contiguous region of interest was selected using the osteoid mask, and cells within this region having a radius greater than two pixels were detected and combined with the cells predicted from the DL model.

Bone marrow adipose tissue

BMAT consists of adipocytes in close contact with bone and hematopoietic cells.[34] During processing, alcohol fixation removes the lipid content of fat cells, leaving behind residual cell boundaries and the surrounding extracellular matrix. Though BMAT is generally left unmeasured in workflows,[35] we made their segmentation an essential step since their texture was analyzed in MTA. The ground truth annotations for training the model were obtained by manually tracing BMAT boundaries (Figure 4). An nn-UNet model was trained on 76 images and tested on 29 images (both data splits comprising images of size $512 \times 512 \times 3$).

Morphological texture analysis

MTA yielded a broad spectrum of quantitative features, such as entropy, energy[36] (derived from gray-level co-occurrence matrix, a 2D histogram containing the counts/probabilities of co-occurring intensity values[37]), and intensity (features from pixel intensities in an image[38]). For this study, an exploratory analysis was performed to relate pixel information obtained from MTA of histological primitives to the bone turnover status ("low" vs "high") specified in the final diagnostic report of the bone biopsy. Images with the diagnosis of osteomalacia were excluded for these analyses to focus on the singular dimension of bone turnover unaffected by any variability in mineralization. The definition of turnover status by the pathologist in the diagnostic report was based on (1) histologic findings (brightfield light microscopy) of cellular surfaces and assessment of bone formation and resorption (2) fluorescent light microscopy with tetracycline labels for extent of mineralizing surface. Bone turnover was categorized as "high" for histology showing high osteoblast surface with high osteoid surface indicative of bone formation, high osteoclast surface with eroded surface indicative of resorption, and fluorescent microscopy showing high mineralizing surface with mainly double tetracycline labels. Similarly, bone turnover was categorized as "low" for histology showing low osteoblast surface with low osteoid surface indicative of decreased bone formation, low osteoclast surface, and fluorescent microscopy showing low mineralizing surface with no, or few, mainly single tetracycline labels.

For MTA analysis, 379 images were subjected to segmentation using DL models to delineate bone, osteoid, and BMAT regions. All images contained BMAT regions, however, bone and osteoid regions were present in just 358 images. Classification models including support vector classifiers, RF, and linear regression, were then trained on top decorrelated features to classify low and high turnover image datasets.
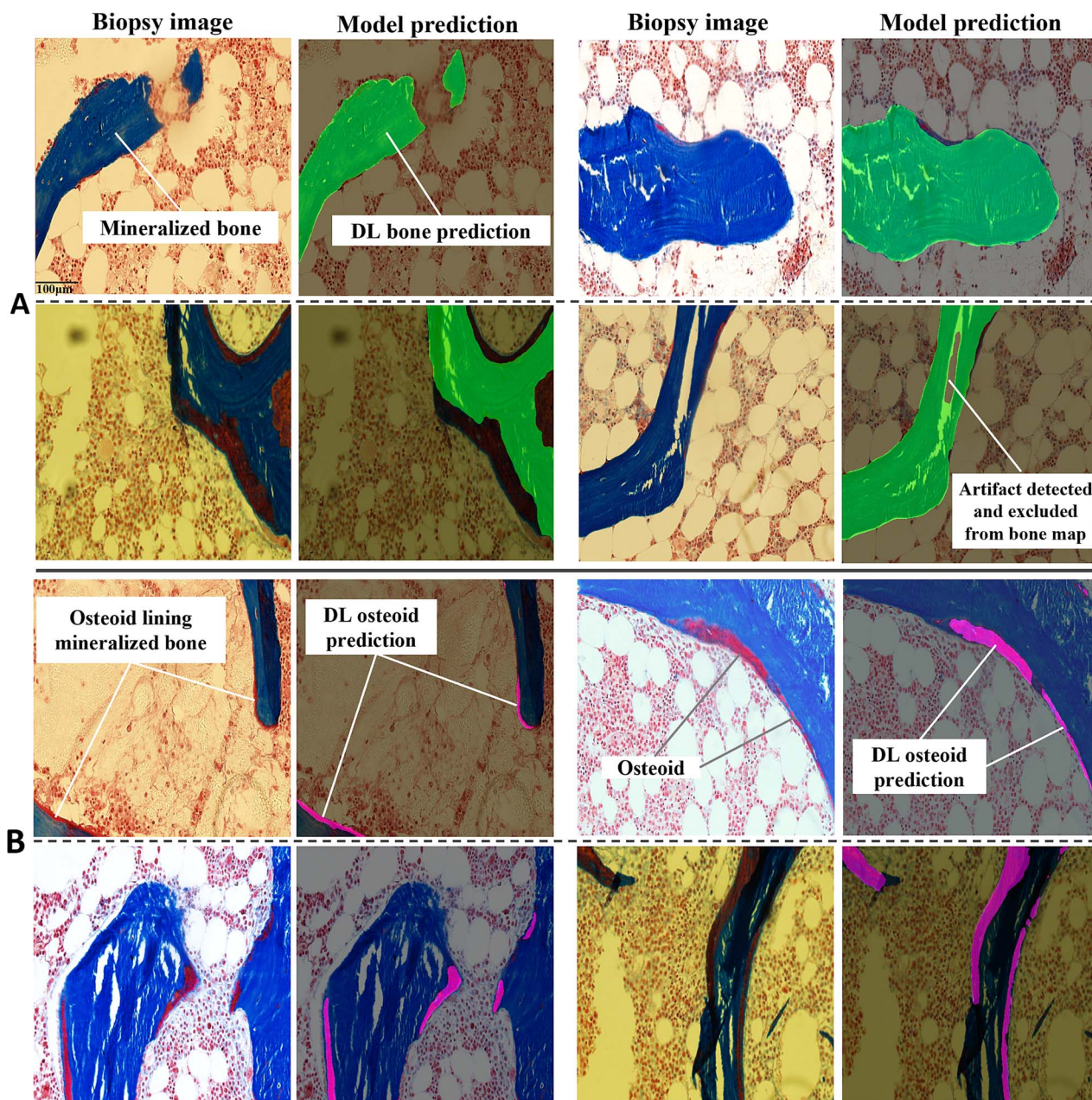
Figure 2. Deep learning-derived segmentation of bone and osteoid regions in bone biopsy images. (A) Illustrates bone regions in green, predicted by the model and overlaid on biopsy images; the osteoid regions predicted by the model are represented with a pink overlay in (B). The images display a range of color variations in biopsy images prior to color normalization.

Histomicstk,[37] a Python package for the analysis of digital pathology images was used to extract texture features and for MTA.

Statistical analysis

DL based feature mapping and cell segmentation

The tissue areas and the cellular structures were compared against the ground truth annotations using the Spearman correlation coefficient,[39] denoted as ρ (rho) applicable to non-normally distributed data. The histomorphometric parameters as outlined by Dempster et al.[20] were compared between ADAM-derived and OsteoMeasure-derived parameters (Figure S3). Correlation plots for tissue segmentation were generated by counting the pixels for each histological primitive, (bone, osteoid, BMAT), normalizing these against the total pixels in the image, and comparing the results with the pixels in manual annotations (Figure 5 and Figure S4). Furthermore, Dice similarity index[40] or Dice scores provided a spatial overlap index and a measure of the model's ability to accurately delineate target tissue regions. For cellular structures, ρ was reported for the correlation between the cell counts obtained from the DL model and manually from the ground-truth annotations. Inter-operator variability was computed as the Spearman correlation coefficient across the cell counts obtained by each of the manual annotations (Figure S5). Additionally, Bland–Altman[41] (BA) analyses were employed to depict the agreement between the two methods and the two observers (see Figures S6 and S7).
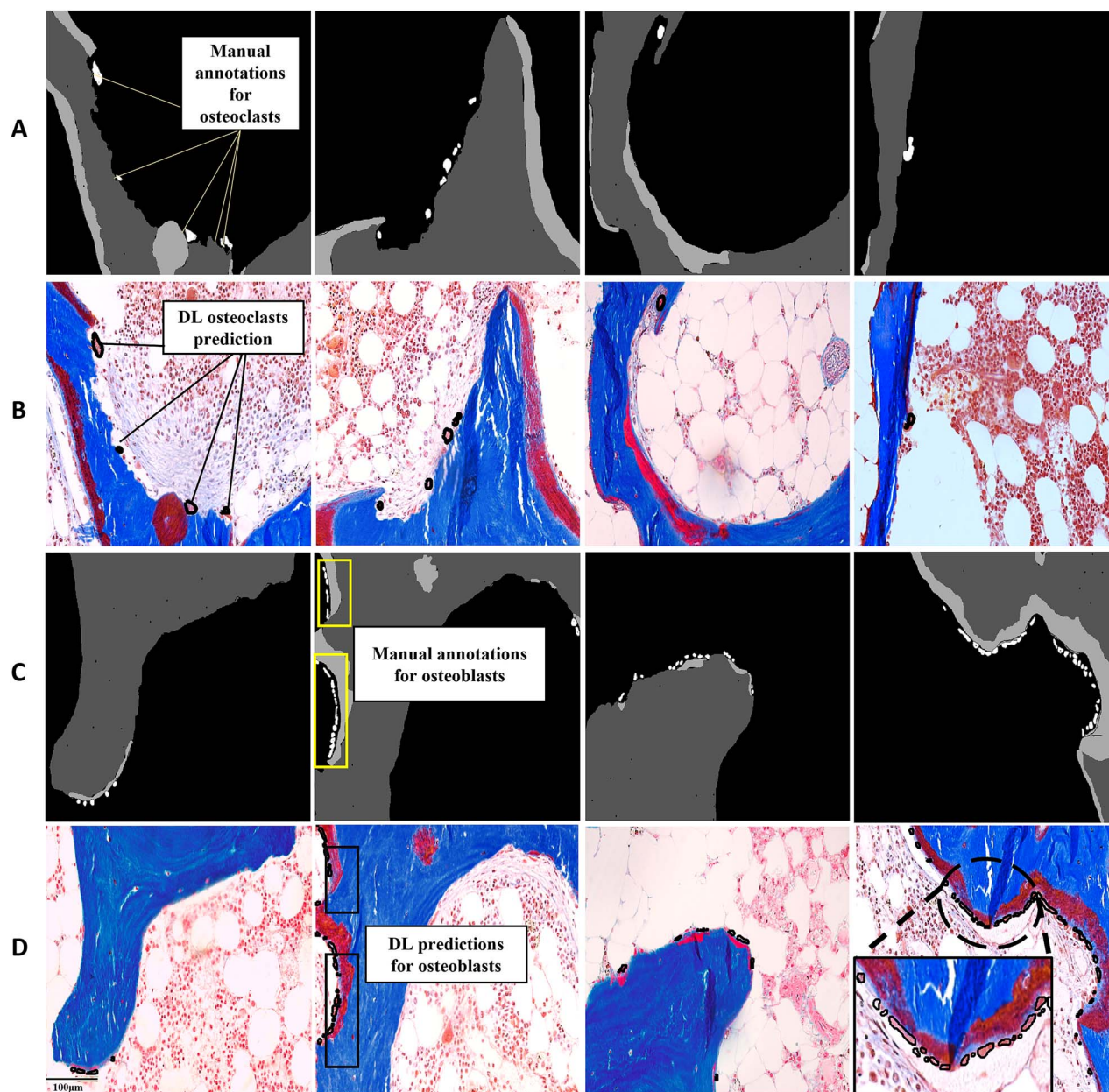
Figure 3. Osteoclasts and osteoblasts segmentation using ADAM; (A) and (C) show manual annotations for osteoclasts and osteoblasts for model validation respectively, a grayscale colormap is used for better visual representation; (B) and (D) show deep learning predictions of osteoclasts and osteoblasts respectively, overlaid on biopsy images with black cell outlines.

Morphological texture features were extracted from bone, osteoid and BMAT, and visualized using the 3D uniform manifold approximation and projection[42] technique (Figure S8). A suite of 43 texture features were extracted from these tissue components and those exhibiting high correlation exceeding 95% were filtered out, resulting in 33 texture features for bone and BMAT, and 34 texture features for osteoid, and bone as detailed in Figure S9. Bone, osteoid and BMAT support vector classifiers[43] were trained using texture features to stratify the images based on turnover labels obtained from the diagnostic report of the bone biopsy (F1 score, Area under the receiver operating characteristic curve (AUC-ROC); Table S2). Subsequently, a deeper analysis of BMAT features was undertaken

based on the AUC-ROC[44] metric (Table S2). The dimensionality reduction of the feature vector using a stringent correlation threshold of 85% resulted in the selection of 24 BMAT texture features. Since BMAT texture features were found to have non-normal distributions (Shapiro–Wilk[45] test $p < .05$), the Mann–Whitney U test[46] was used to select significant texture features that tracked with the bone turnover labels. Applying a Bonferroni[47] corrected $p < .002$, three BMAT texture features satisfying the conservative p-value threshold were identified, and classifiers were trained (225 low turnover images and 78 high turnover images) and tested (57 low turnover and 19 high turnover). A flowchart shown in Figure S9 illustrates the steps involved in MTA. The performance of the binary classifiers for
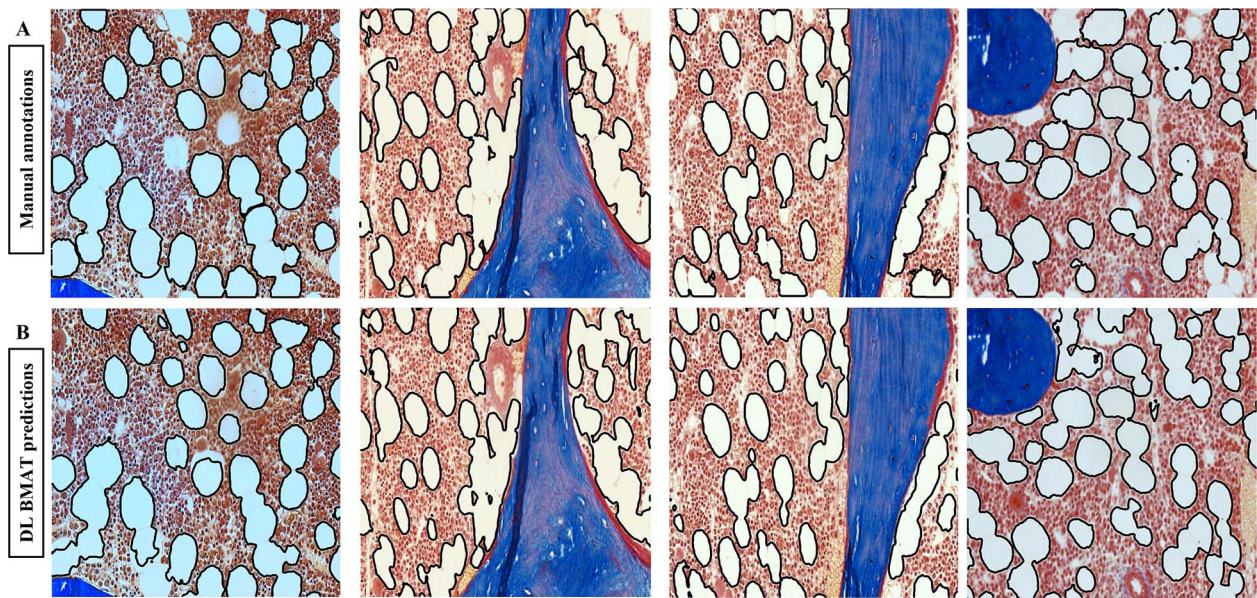
Figure 4. Representative manual annotations to segment the white oval BMAT regions (A) and deep learning predictions for BMAT using nnU-Net (B).
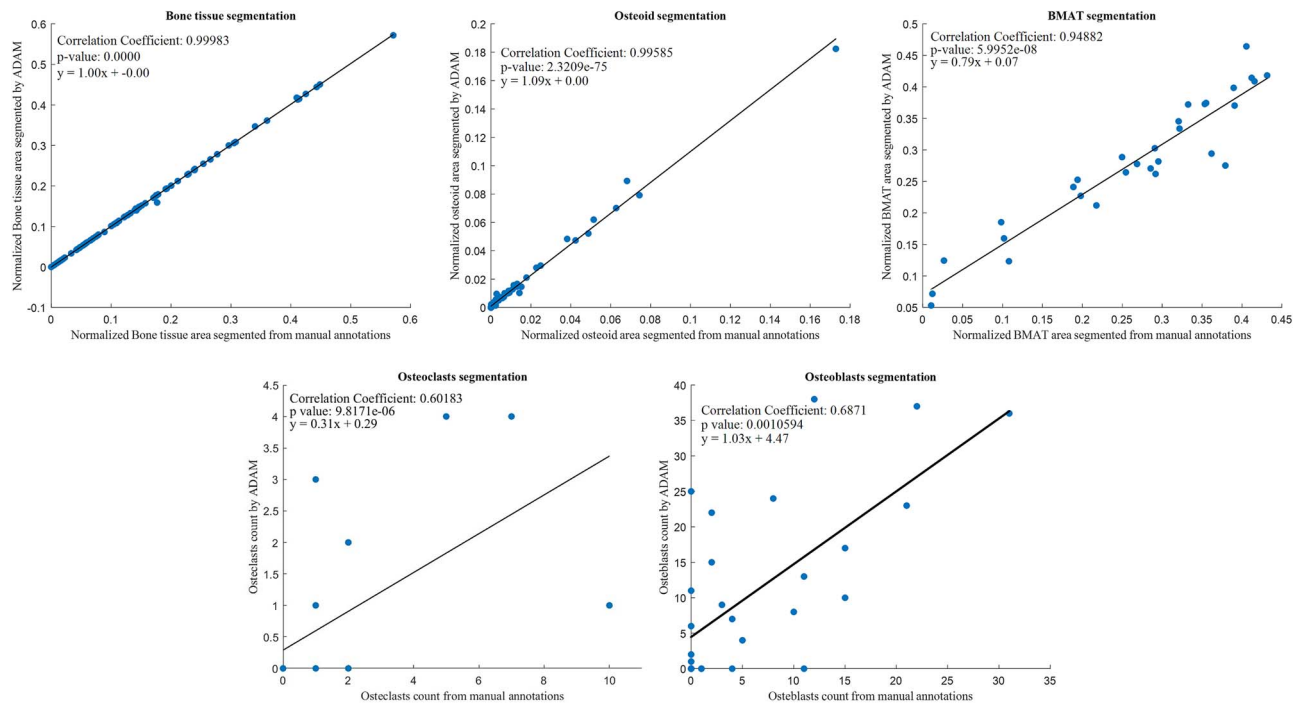

Figure 5. Correlation plots of phenotyped regions and cellular structures from manual annotations and automated segmentation; clockwise: Bone (0.99; 77 test images), osteoid (0.99; 77 test images), BMAT (0.95; 29 test images), osteoclasts (0.60; 31 test images) and osteoblasts (0.69; 30 test images).

turnover was evaluated using AUC-ROC, precision, recall, F1 score, and accuracy (Table S2).

Results
Results from deep learning models for feature extraction
Mineralized bone and osteoid
Spearman correlation coefficients (ρ) between model and manual annotations were reported to be 0.99 for bone (p = .0000) and osteoid segmentations (p = 2.89e^{-55}) (Figures 2 and 5). Additionally, Dice scores for bone and osteoid were found to be 0.96 and 0.73, respectively. The Spearman correlation between model and OsteoMeasure-derived areas was 0.92 for mineralized bone and 0.8 for osteoid (Figure S3).

Osteoclasts
The number of osteoclasts obtained from the DL model were compared against the manual annotations (Figure 3), yielding a correlation coefficient (ρ) of 0.60 (p = 9.82e^{-6}) (Figure 5). Similarly, the Spearman correlation between the DL model and OsteoMeasure clinical count was 0.66 (Figure S3). Bland–Altman analysis showed a mean difference or offset

of <2 between manual and automated counts for osteoclasts (Figure S6), with most counts lying within the 95% confidence interval. Additionally, the inter-operator correlation for osteoclasts count was $\rho = 0.62$ (Figure S5).

Osteoblasts

Osteoblast count from the DL model and corresponding manual annotations were compared (Figure 3) and a correlation coefficient of 0.69 ($p = 1.06e^{-3}$) was obtained (Figure 5). Additionally, the Spearman correlation between the DL model and OsteoMeasure clinical count was 0.53 (Figure S3). The Bland–Altman analysis (Figure S6) indicated a mean difference or offset of -5 between manual and automated cell counts, with most points lying within the 95% confidence interval. Furthermore, the inter-operator correlation gave a ρ of 0.84 for osteoblast count (Figure S5).

Bone marrow adipocyte tissue

The BMAT segmentation (Figure 4) achieved a Dice score of 0.65 and a correlation coefficient of 0.95 ($p = 5.99e^{-8}$) as determined between the DL model output and manual annotations was obtained (Figure 5).

Morphological texture grading and analysis

Computationally derived image texture and shape features demonstrated a performance of [AUC-ROC, F1 score] of [0.72, 0.61] for bone, [0.78, 0.67] for osteoid and [0.86, 0.74] for BMAT to classify images labelled as low and high turnover using diagnostic reports. Furthermore, three classifiers yielded performance metrics [AUC-ROC, F1 score] of [0.71, 0.70] for a SVM classifier, [0.67, 0.61] for linear regression and [0.71, 0.71] for a random forest classifier (Figure S11) when trained on the top significant BMAT texture features ($p < .002$).

Discussion

Undecalcified bone histology is a valuable diagnostic tool to understand pathologic processes such as mineralization defects and altered turnover states, which in turn can guide treatment.[7] Quantitative bone histology largely relies on manual annotation of histologic primitives that is time-consuming[21] and marked by inter-observer variability.[7] Though there are semi-automated software and workflows available,[21] there is a need for an automated pipeline that processes digitized bone biopsies to feature maps. The implementation of algorithms, however, has faced challenges due to the heterogeneity in cellular morphology[48] and limited availability of datasets that include annotated ground truths for bone histology.[21] In this study, we present ADAM, a static histomorphometry workflow, which segments the bone biopsy into five compartments (bone, osteoid, osteoblasts, osteoclasts, and BMAT) and provides a metric for the area and count for each histologic primitive. ADAM employs models and algorithms that allow for automated tissue segmentation and cell counts for human iliac crest biopsy images. By leveraging the potential of deep-learning models, ADAM generates approximately 20 feature maps in under a minute, providing support for clinical triage and decision-making.[26] Automated quantification could be a valuable addition to clinical workflows that currently provide qualitative and quantitative reporting and interpretation of bone biopsies.

Studies undertaken in automating segmentation of histologic primitives in bone biopsies have relied on manual or semi-automated methods using open-source or commercial software.[21] Van't Hof[23] proposed three applications for histomorphometric analyses, requiring manual selection of the region of interest. They demonstrated high reproducibility of the measurements while relying on the use of specific stains to study the tissue and cellular components. ADAM, on the other hand, was developed without a requirement for manual annotation or additional staining such as TRAcP to identify osteoclasts. Zhang et al.[12] implemented a Visiopharm algorithm for the histomorphometric analysis that showed ICC >0.97 for bone, cartilage and fibrous tissue between OsteoMeasure and Visiopharm. However, their approach relied on color training and required all elements in the slide to stain differently. We used a sematic segmentation-based DL approach, using spatial information for training as opposed to naive pixel-level DL classifiers.[26] For instance, incorporation of bone and osteoid masks into the osteoblast segmentation training data provided essential spatial cues for their detection. Malhan et al.[21] demonstrated a semi-automated algorithm to measure osteoclast activity, trabecular thickness, and trabecular separation using rat and sheep iliac crest biopsies. They again used a multi-stained approach which required manual intervention to preprocess the images by drawing contours around the bone to exclude the muscle regions.

Our results showed excellent correlation between predictions from ADAM and the ground truth annotations ($\rho > 0.9$) for tissue region segmentation such as mineralized bone, osteoid and BMAT. ADAM's predictions accurately recognized folding artifacts in bone areas, marking them as non-bone regions (Figure 2). In a subset of images ($n = 22$) with available OsteoMeasure histomorphometry values, we also show alignment between ADAM-derived and OsteoMeasure-derived measurements. Spearman correlation coefficients of $\rho > 0.8$ were obtained for bone and osteoid areas, and 0.66 and 0.53 for osteoclast count and osteoblast count respectively (Figure S3). The ability to automate precise measurements of mineralized bone and osteoid would also facilitate easier computation of indices such as bone volume/tissue volume and osteoid volume/bone volume, using a 3D correction factor.[20] Finally, while BMAT is generally left unmeasured in the clinical workflow, we show that ADAM enables BMAT segmentation in undecalcified bone biopsy images. Sarkis et al.[49] previously demonstrated BMAT segmentation in H&E-stained bone marrow trephine biopsies.

In addition to digital phenotyping, the study explored MTA for texture features extracted from bone, osteoid, and BMAT. Findings from BMAT MTA showed that images from biopsies labeled as high bone turnover had higher median values for pixel intensity uniformity compared to low turnover images (Figure 6). Additionally, low turnover images were noted to express higher median pixel intensity skewness and minimum pixel intensity levels compared to high turnover images (Figure S10). Our findings are interesting in the context of the paracrine role of BMAT in osteoblast and osteoclast functions.[35] As a novel exploratory step, this study introduces a methodology to extract and analyze MTA features and examine their relationship to a selected phenotype (such as bone turnover).

We acknowledge several limitations to our study. Firstly, the lower correlations observed between DL and manual annotations for osteoblast and osteoclast cell counts compared to tissue areas, could be due to morphologic heterogeneity and clustering or clumping of cells with
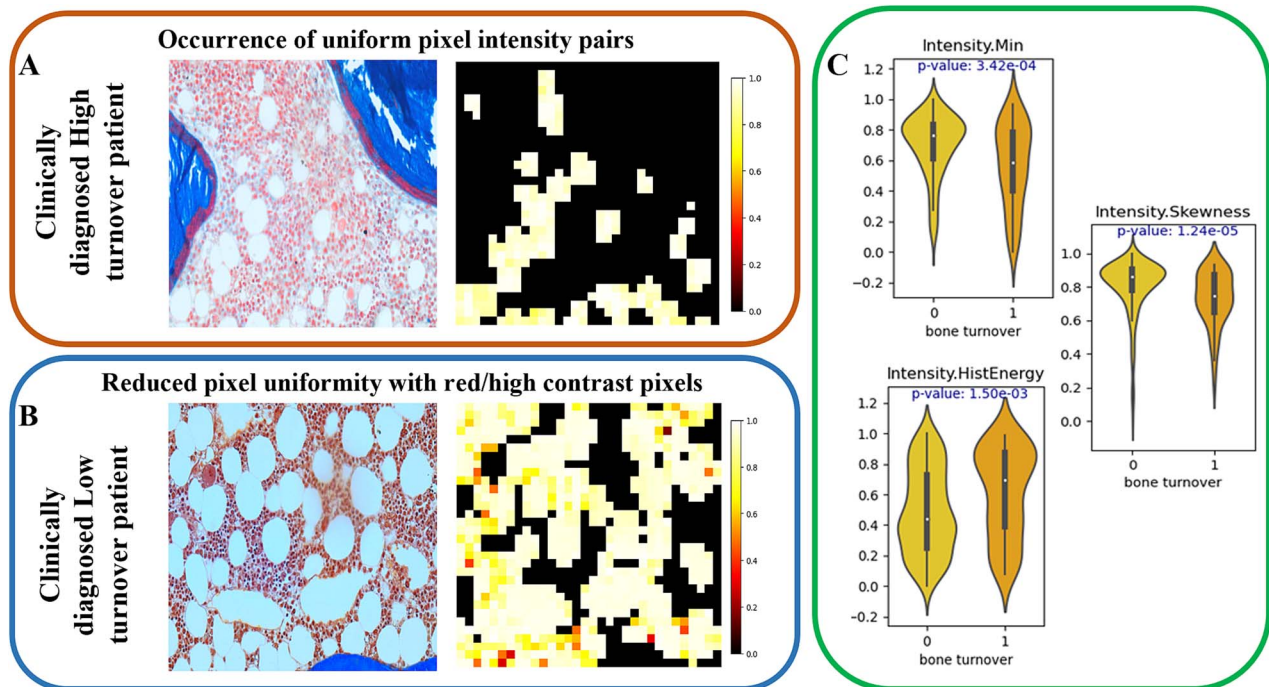
Figure 6. Representative images of BMAT regions with heatmaps of pixel intensities from biopsies labeled as "high" bone turnover (A) and "low" bone turnover (B); (C) violin plots for MTA features showing values for "low" (*x*-axis label "0") and "high" bone turnover images (*x*-axis label "1"). High turnover images exhibited a high median value for the energy feature (Intensity.Hist.Energy), while low turnover images had high median values for intensity-based features (Intensity.Min and Intensity.Skewness). Refer to Table S3 for a detailed description of MTA features.

indistinct borders. Secondly, we note comparable inter-operator variability and the correlation between DL predictions and manual annotations for osteoclasts was ρ of 0.62 and 0.6 respectively, but regarding osteoblast count, the inter-operator correlation appeared to be better than the correlation for manual and DL-based osteoblast segmentation (ρ of 0.84 and 0.69 respectively). The model seemed to overcount osteoblasts in sparsely populated images and undercount them in densely populated areas (Figure S6). The problem could likely be caused by the difficulty in accurately distinguishing closely packed or overlapping cells and the misclassification of flattened osteoblasts as active due to the state of the bone formation cycle. Potential improvements in model performance might include using larger datasets to capture phenotypic variability during training or opting for a higher (40x) image magnification that highlights local features of cells, especially when they are densely clustered. Regardless, we emphasize that, while DL could assist in cell counting and measuring tissue areas, manual verification of results by the pathologist, should remain an integral component of reporting workflows. The intended application of ADAM is to assist pathologists in analyzing bone biopsy images and is not validated for diagnostic purposes.

Secondly, the pipeline was developed using a set of highly selected images diagnostic for specific histologic primitives. The need for further training and validation with large independent datasets is required to overcome selection bias and to improve cellular segmentations. Although ADAM was tested on ROIs with available ground truth, its functionality is not limited by ROI selection and can be applied to whole slide analysis; this is a direction we will explore in future studies with larger multisite datasets. Furthermore, the pipeline must be extended to quantify dynamic histomorphometry

features using fluorescence microscopy images.[15] Despite these limitations, this preliminary investigation is valuable as it explores various computer vision and machine learning techniques in pursuit of digital phenotyping bone biopsies, reducing time and the dependency on operators and additional staining.

Conclusion

To the best of our knowledge, ADAM is the first automated machine learning driven image analysis pipeline that compartmentalizes five tissue components in undecalcified bone histologic images, generates accurate feature maps and provides quantitative parameters for static histomorphometry. It also illustrates the ability to conduct texture analysis for each histologic primitive and explore the relationship of non-traditional features, like BMAT to normal and pathological tissue states. These findings lay the foundation for further investigations to advance both diagnostic practices and research methodologies in bone histomorphometry, paving the way for more precise and reproducible reporting in metabolic bone disease.

Author contributions

Satvika Bharadwaj (Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing), Florence Lima (Conceptualization,

Supplementary material

Supplementary material is available at *JBMR Plus* online.

Funding

Conflicts of interest

The authors declare the following financial interests/personal relationships which may be considered potential competing interests: A.M. reports funding grants from AstraZeneca, Bristol Myers-Squibb, Boehringer-Ingelheim, and Eli Lilly; consulting or advisory roles with Picture Health Inc., Aiforia, and Simbiosys; and equity or stock ownership in Picture Health Inc., Inspirata Inc., and Elucid Bioimaging.

Data availability

Data will be shared on request to the corresponding author with permission of the involved institutions.

References

1. Malluche HH, Mawad H, Monier-Faugere MC. Bone biopsy in patients with osteoporosis. *Curr Osteoporos Rep*. 2007;5(4):146-152. https://doi.org/10.1007/s11914-007-0009-x
2. Malluche HH, Langub MC, Monier-Faugere MC. Pathogenesis and histology of renal osteodystrophy. *Osteoporos Int*. 1997;7(Suppl 3):184-187. https://doi.org/10.1007/BF03194369
3. Malluche H, Faugere MC. Renal bone disease 1990: an unmet challenge for the nephrologist. *Kidney Int*. 1990;38(2):193-211. https://doi.org/10.1038/ki.1990.187
4. Malluche HH, Langub MC, Monier-Faugere MC. The role of bone biopsy in clinical practice and research. *Kidney Int*. 1999;56(Suppl 73):S20-S25. https://doi.org/10.1046/j.1523-1755.1999.07313.x
5. Malluche HH, Mawad HW, Monier-Faugere MC. Renal osteodystrophy in the first decade of the new millennium: analysis of 630 bone biopsies in black and white patients. *J Bone Miner Res*. 2011;26(6):1368-1376. https://doi.org/10.1002/jbmr.309
6. Florkow MC, Willemsen K, Mascarenhas VV, Oei EHG, van Stralen M, Seevinck PR. Magnetic resonance imaging versus computed tomography for three-dimensional bone imaging of musculoskeletal pathologies: a review. *J Magn Reson Imaging*. 2022;56(1):11-34. https://doi.org/10.1002/jmri.28067
7. Chavassieux P, Chapurlat R. Interest of bone histomorphometry in bone pathophysiology investigation: foundation, present, and future. *Front Endocrinol*. 2022;13:13Accessed October 15, 2023. https://www.frontiersin.org/articles/10.3389/fendo.2022.907914
8. Lindsay R, Zhou H, Cosman F, Nieves J, Dempster D. Double and quadruple tetracycline labeling of bone: impact of the label itself. *J Bone Miner Res*. 2013;28(1):222-223. https://doi.org/10.1002/jbmr.1818
9. Hlaing TT, Compston JE. Biochemical markers of bone turnover - uses and limitations. *Ann Clin Biochem*. 2014;51(2):189-202. https://doi.org/10.1177/0004563213515190
10. Marcus R, Dempster DW, Bouxsein ML. Chapter 2 – The Nature of Osteoporosis. In: Marcus R, Feldman D, Dempster DW, Luckey M, Cauley JA, eds. *Osteoporosis*. 4th ed. Academic Press, San Diego; 2013, 21-30. https://doi.org/10.1016/B978-0-12-415853-5.00002-9
11. Ramchand SK, Seeman E. Reduced bone modeling and unbalanced bone remodeling: Targets for antiresorptive and anabolic therapy. In: Stern PH, ed. *Bone Regulators and Osteoporosis Therapy. Handbook of Experimental Pharmacology*. Cham: Springer International Publishing; 2020:423-450. https://doi.org/10.1007/164_2020_354.
12. Zhang L, Chang M, Beck CA, Schwarz EM, Boyce BF. Analysis of new bone, cartilage, and fibrosis tissue in healing murine allografts using whole slide imaging and a new automated histomorphometric algorithm. *Bone Res*. 2016;4(1):15037-15039. https://doi.org/10.1038/boneres.2015.37
13. Novel-Catin E, Pelletier S, Fouque D, et al. Quantitative histomorphometric analysis of halved iliac crest bone biopsies yield comparable ROD diagnosis as full 7.5mm wide samples. *Bone*. 2020;138:115460. https://doi.org/10.1016/j.bone.2020.115460
14. Malluche HH, Sherman D, Meyer W, Massry SG. A new semiautomatic method for quantitative static and dynamic bone histology. *Calcif Tissue Int*. 1982;34(1):439-448. https://doi.org/10.1007/BF02411282
15. Manaka RC, Malluche HH. A program package for quantitative analysis of histologic structure and remodeling dynamics of bone. *Comput Programs Biomed*. 1981;13(3-4):191-201. https://doi.org/10.1016/0010-468x(81)90098-2
16. Scheller EL, McDonald M, Andersen TL, et al. Celebrating 50-years: the history and future of the International Society of Bone Morphometry. *JBMR Plus*. 2024;8(7). https://doi.org/10.1093/jbmrpl/ziae070
17. Human Histomorpometry. BIOQUANT. Accessed October 15, 2023. https://www.bioquant.com/bone/human-histomorphometry
18. OsteoMetrics. Accessed February 27, 2024. https://www.osteometrics.com/
19. Kulak CAM, Dempster DW. Bone histomorphometry: a concise review for endocrinologists and clinicians. *Arq Bras Endocrinol Metab*. 2010;54(2):87-98. https://doi.org/10.1590/S0004-27302010000200002
20. Dempster DW, Compston JE, Drezner MK, et al. Standardized nomenclature, symbols, and units for bone histomorphometry: a 2012 update of the report of the ASBMR Histomorphometry Nomenclature Committee. *J Bone Miner Res*. 2013;28(1):2-17. https://doi.org/10.1002/jbmr.1805
21. Malhan D, Muelke M, Rosch S, et al. An optimized approach to perform bone histomorphometry. *Front Endocrinol (Lausanne)*. 2018;9:666. https://doi.org/10.3389/fendo.2018.00666
22. Goldschlager T, Abdelkader A, Kerr J, Boundy I, Jenkin G. Undecalcified bone preparation for histology, histomorphometry and fluorochrome analysis. *J Vis Exp*. 2010;35(35):1707. https://doi.org/10.3791/1707
23. van, 't Hof RJ, Rose L, Bassonga E, Daroszewska A. Open source software for semi-automated histomorphometry of bone resorption and formation parameters. *Bone*. 2017;99:69-79. https://doi.org/10.1016/j.bone.2017.03.051
24. Hong SH, Jiang X, Chen L, Josh P, Shin DG, Rowe D. Computer-automated static, dynamic and cellular bone histomorphometry. *J Tissue Sci Eng*. 2012;Suppl 1:004. https://doi.org/10.4172/2157-7552.S1-004
25. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203-211. https://doi.org/10.1038/s41592-020-01008-z
26. Jayapandian CP, Chen Y, Janowczyk AR, et al. Development and evaluation of deep learning-based segmentation of histologic structures in the kidney cortex with multiple histologic stains. *Kidney Int*. 2021;99(1):86-101. https://doi.org/10.1016/j.kint.2020.07.044

27. Vaidya P, Bera K, Linden PA, et al. Combined radiomic and visual assessment for improved detection of lung adenocarcinoma invasiveness on computed tomography scans: a multi-institutional study. *Front Oncol*. 2022;12:902056. https://doi.org/10.3389/fonc.2022.902056

28. Chen P, Rojas FR, Hu X, et al. Pathomic features reveal immune and molecular evolution from lung preneoplasia to invasive adenocarcinoma. *Mod Pathol*. 2023;36(12):100326. https://doi.org/10.1016/j.modpat.2023.100326

29. Malluche HH, Faugere MC. *Atlas of Mineralized Bone Histology: 34 Tables*. Karger; 1986. https://doi.org/10.1159/isbn.978-3-318-04339-6

30. Savi FM, Lawrence F, Hutmacher DW, Woodruff MA, Bray LJ, Wille ML. Histomorphometric evaluation of critical-sized bone defects using osteomeasure and aperio image analysis systems. *Tissue Eng Part C Methods*. 2019;25(12):732-741. https://doi.org/10.1089/ten.TEC.2019.0179

31. Pocock J, Graham S, Vu QD, et al. TIAToolbox as an end-to-end library for advanced tissue image analytics. *Commun Med*. 2022;2(1):120-114. https://doi.org/10.1038/s43856-022-00186-5

32. Majanga V, Viriri S. Dental images' segmentation using threshold connected component analysis. *Comput Intell Neurosci*. 2021;2021(1):2921508. https://doi.org/10.1155/2021/2921508

33. Šromová V, Sobola D, Kaspar P. A brief review of bone cell function and importance. *Cells*. 2023;12(21):2576. https://doi.org/10.3390/cells12212576

34. Tratwal J, Labella R, Bravenboer N, 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. Accessed July 15, 2023. https://doi.org/10.3389/fendo.2020.00065

35. Hardouin P, Rharass T, Lucas S. Bone marrow adipose tissue: to be or not to be a typical adipose tissue? *Front Endocrinol*. 2016;7. Accessed October 18, 2023. https://doi.org/10.3389/fendo.2016.00085

36. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern B*. 1973;SMC-3(6):610-621. https://doi.org/10.1109/TSMC.1973.4309314

37. Histomics TK. Published online July 1, 2023. Accessed July 17, 2023. https://github.com/DigitalSlideArchive/HistomicsTK

38. Zwillinger D, Kokoska S. *CRC Standard Probability and Statistics Tables and Formulae*. Chapman & Hall/CRC; 2000. https://doi.org/10.1201/b16923

39. Mukaka M. A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J*. 2012;24(3):69-71.

40. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index: scientific reports. *Acad Radiol*. 2004;11(2):178-189. https://doi.org/10.1016/S1076-6332(03)00671-8

41. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-310.

42. Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2019;37(1):38-44. https://doi.org/10.1038/nbt.4314

43. Zhang Y. Support vector machine classification algorithm and its application. In: Liu C, Wang L, Yang A, eds. *Information Computing and Applications*. Berlin, Heidelberg: Springer; 2012:179-186. https://doi.org/10.1007/978-3-642-34041-3_27

44. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn*. 1997;30(7):1145-1159. https://doi.org/10.1016/S0031-3203(96)00142-2

45. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab*. 2012;10(2):486-489. https://doi.org/10.5812/ijem.3505

46. Nachar N. The Mann-Whitney U: a test for assessing whether two independent samples come from the same distribution. *Tutorial Quantitat Methods Psychol*. 2008;4(1):13-20. https://doi.org/10.20982/tqmp.04.1.p013

47. VanderWeele TJ, Mathur MB. Some desirable properties of the BONFERRONI correction: is the Bonferroni correction really so bad? *Am J Epidemiol*. 2019;188(3):617-618. https://doi.org/10.1093/aje/kwy250

48. Everts V, de Vries TJ, Helfrich MH. Osteoclast heterogeneity:: lessons from osteopetrosis and inflammatory conditions. *Biochim Biophys Acta (BBA) - Mol Basis Dis*. 2009;1792(8):757-765. https://doi.org/10.1016/j.bbadis.2009.05.004

49. Sarkis R, Burri O, Royer-Chardon C, et al. MarrowQuant 2.0: a digital pathology workflow assisting bone marrow evaluation in experimental and clinical hematology. *Mod Pathol*. 2023;36(4):100088. https://doi.org/10.1016/j.modpat.2022.100088