Oncology Treatment Discovery, 2026, Volume 4, Issue 1
http://ojs.bbwpublisher.com/index.php/OTD
Online ISSN: 2981-8079
Print ISSN: 3083-4996

Advances in MSCT Trabecular Bone Attenuation Combined with Natural Language Processing and Large Language Models for Opportunistic Osteoporosis Screening and Fracture Risk Assessment: A Review

Shiyu Zhong[1], Peng Tao[2], Hairong Gong[1], Xiangrong Deng[1], Jianghai Zhao[1], Fuxing Sun[1], Jinxiong Song[1], Dan Wu[1]*

[1]Department of Radiology, Chongzhou Traditional Chinese Medicine Hospital, Chongzhou, Chengdu 611200, Sichuan, China

[2]Department of Radiology, Affiliated Hospital of Chengdu University, Chengdu 610081, Sichuan, China

Corresponding author: Dan Wu, 2673302353@qq.com

Abstract: Osteoporosis and its consequent fragility fractures represent a major public health challenge in aging populations worldwide. Although dual-energy X-ray absorptiometry (DXA) remains the diagnostic gold standard, its utility for opportunistic screening is limited by equipment accessibility, low screening coverage, and inability to assess bone microarchitecture. Multislice computed tomography (MSCT), routinely performed for various clinical indications, offers a quantitative measure, trabecular bone attenuation in Hounsfield Units (HU), that can be opportunistically leveraged for osteoporosis evaluation. Recent advances in natural language processing (NLP) and large language models (LLMs) enable automated extraction of unstructured clinical data from electronic health records and integration of multimodal information. This review systematically summarizes the evidence base for MSCT-derived HU values in osteoporosis diagnosis, explores the application of NLP for extracting key variables from radiology reports and clinical notes, and discusses the potential of LLMs for multimodal data fusion and predictive modeling. We propose a novel framework combining HU measurements, NLP-extracted clinical features, and LLM-driven analysis to construct a Nomogram model for opportunistic osteoporosis screening and fracture risk prediction. This approach may expand the screened population beyond postmenopausal women to include men and individuals with chronic diseases, ultimately enabling opportunistic, individualized fracture risk assessment integrated into routine clinical workflows.

Keywords: Osteoporosis; Multislice computed tomography; Hounsfield units; Natural language processing; Large language models; Nomogram; Opportunistic screening; Fracture risk

Online publication: March 18, 2026

1. Introduction

Osteoporosis is a systemic skeletal disorder characterized by low bone mass and microarchitectural deterioration, leading to increased bone fragility and susceptibility to fractures [1]. With global population aging, the prevalence of osteoporosis continues to rise, particularly among postmenopausal women. However, fracture risk is not confined to this demographic; elderly men and patients with secondary osteoporosis (e.g., due to glucocorticoid use, diabetes mellitus) are also at substantial risk [2]. Current diagnosis and management rely primarily on DXA-measured bone mineral density (BMD). Yet DXA has well-recognized limitations: limited availability in primary care settings, low screening coverage, and inability to capture trabecular microarchitecture. Approximately 50% of individuals sustaining fragility fractures have DXA BMD values above the osteoporotic threshold [3], underscoring the need for complementary assessment tools.

Opportunistic screening—using imaging data acquired for other clinical indications to assess osteoporosis risk without additional radiation or cost—has emerged as a promising strategy [4]. MSCT, widely performed for oncologic, cardiovascular, or abdominal indications, provides high-resolution three-dimensional data. The trabecular bone attenuation measured in HU on MSCT has been shown to correlate with BMD and reflect microstructural changes [5–7]. Nevertheless, most studies have focused on isolated HU measurements and have not integrated the rich clinical information contained in electronic health records (EHRs).

In parallel with advances in CT-based bone assessment, recent breakthroughs in artificial intelligence (AI) offer new opportunities. NLP techniques can automatically extract structured information from unstructured clinical texts, such as fracture history, medication use, and comorbidities [8]. LLMs, with their capacity for multimodal understanding and generation, can fuse imaging features with textual clinical data to build more accurate predictive models [9]. A Nomogram, a graphical representation of a statistical prediction model, provides an intuitive tool for individualized risk estimation that can be readily adopted in clinical practice [10].

This review aims to: (1) summarize the evidence for MSCT-derived HU values in osteoporosis assessment; (2) examine NLP applications for extracting osteoporosis-related data from EHRs; (3) discuss the potential of LLMs for multimodal fusion; and (4) propose a framework for constructing a Nomogram model integrating HU, NLP-extracted features, and LLM-driven analysis for opportunistic screening and fracture risk prediction across diverse populations.

2. MSCT trabecular bone attenuation: Foundations and extensions

2.1. Measurement principles and clinical evidence

MSCT measures X-ray attenuation through tissues, quantified in HU, which reflects linear attenuation coefficients. For bone, HU values are determined by mineral density and trabecular microarchitecture. Within vertebral bodies, the three-dimensional trabecular network, including trabecular thickness, number, and connectivity, directly influences biomechanical strength. In osteoporosis, reductions in both mineral content and trabecular connectivity contribute to decreased HU values [5–7].

Measurement typically targets the central trabecular region of vertebral bodies (commonly L1–L4) on bone window images, avoiding the cortical shell and basivertebral plexus [11]. Multiple studies have reported moderate to strong correlations between lumbar trabecular HU and DXA BMD ($r = 0.6–0.8$), with areas under the receiver operating characteristic curve (AUC) of 0.75–0.89 for diagnosing osteoporosis [12]. For postmenopausal women, proposed diagnostic thresholds range from < 110 HU to < 100 HU, with sensitivities of 80–85% and specificities of 78–82% [12,13].

2.2. From single vertebral HU to radiomics

Conventional HU measurement yields an average value for a selected vertebral region. Radiomics extends this by extracting high-dimensional quantitative features from CT images, including shape, first-order statistics, texture, and wavelet features, capturing trabecular heterogeneity [14]. Lin et al. [10] demonstrated that a radiomics model based on a single thoracic vertebral body (T11 or T12) from chest CT achieved an AUC of 0.896 for osteoporosis screening, significantly outperforming a clinical parameter-only model (AUC 0.744). This suggests that deeper mining of microstructural information from MSCT can enhance early detection sensitivity.

2.3. Expanding the target population

Most research has focused on postmenopausal women, but opportunistic screening can benefit all patients undergoing MSCT for any indication. Future studies should broaden the population to include:

(1) Elderly men, particularly those with hypogonadism, smoking, or alcohol abuse.
(2) Patients with chronic diseases predisposing to secondary osteoporosis, such as diabetes mellitus, chronic kidney disease, and inflammatory bowel disease.
(3) Individuals on medications affecting bone metabolism, including glucocorticoids, aromatase inhibitors, and proton pump inhibitors.
(4) Diagnostic thresholds and predictive performance of MSCT HU values in these populations require validation through large-scale multicenter studies.

3. Natural language processing: Unlocking clinical information from electronic health records

3.1. Value of unstructured clinical data

EHRs contain a wealth of information relevant to osteoporosis assessment, including radiology reports, discharge summaries, medication records, and laboratory results. Key variables such as prior fracture history, DXA results (T-scores), anti-osteoporotic medication use, fall history, and family history are often embedded in unstructured narrative text [15]. Manual extraction is time-consuming and impractical for large-scale research or real-time decision support.

3.2. NLP for bone density information extraction

Fodeh et al. [15] developed an NLP pipeline to extract BMD measurements from radiology reports and clinical notes, achieving 82.8% accuracy for standardized BMD indicators. Subsequently, a rule-based NLP algorithm named “BoneScore” was developed, using 20 regular expressions to extract femoral neck T-scores from DXA reports with 98% accuracy by capturing text windows around the keyword “femoral” [16]. These studies demonstrate the feasibility of automated BMD data extraction from EHRs.

3.3. From rule-based systems to semantic understanding

Rule-based NLP methods rely on predefined dictionaries and patterns, which may fail with linguistic variations. LLMs such as BERT and GPT enable deep semantic understanding of clinical text [17]. In osteoporosis research, LLMs can automatically identify fracture sites, distinguish acute from chronic fractures, and extract fracture mechanisms (low-energy vs. high-energy trauma), providing refined phenotypes for risk stratification.

4. Large language models: A new engine for multimodal data fusion

4.1. Potential of LLMs in medical predictive modeling

LLMs excel not only in natural language understanding but also in integrating multimodal data [9,18]. Unlike traditional machine learning models requiring manual feature engineering, LLMs can directly process mixtures of image features, text embeddings, and structured variables, learning cross-modal interactions via attention mechanisms [9]. In fracture risk prediction, researchers have begun exploring models based on architectures like Gamma 7GB and LLaMA to combine imaging and clinical data [9].

4.2. Pathways for cross-modal fusion

An LLM-based osteoporosis risk prediction framework could follow these steps:

(1) Imaging modality: Extract HU values and radiomics features from MSCT images, transformed into feature vectors.

(2) Text modality: Use NLP to extract clinical entities (age, sex, fracture history, medications, comorbidities) and generate text embeddings.

(3) Structured data: Incorporate laboratory results (e.g., serum calcium, 25-hydroxyvitamin D, PTH) as structured inputs.

(4) Multimodal fusion: Feed the combined data into an LLM (e.g., fine-tuned ClinicalBERT or LLaMA), where attention mechanisms learn interactions across modalities and output a fracture risk probability.

4.3. Lessons from other fields: The oculomics example

LLM-based multimodal prediction has shown promise in other specialties, such as integrating ophthalmic data with demographic information for osteoporosis risk assessment [18]. This "oculomics" approach illustrates how combining imaging from different organ systems with clinical text can enhance prediction accuracy. Similarly, fusion of spinal CT with EHR text may yield synergistic gains.

5. Constructing a nomogram model: From complex algorithms to clinical tools

5.1. Current use of nomograms in osteoporosis

A Nomogram visualizes a regression model by assigning points to each predictor based on its coefficient, allowing calculation of an individual's risk probability [10]. In osteoporosis, Nomograms have been developed to integrate clinical risk factors with BMD. Lin et al. [10] constructed a radiomics Nomogram combining chest CT radiomics features with patient characteristics, achieving an AUC of 0.807 in an external validation set, outperforming a clinical-only model (AUC 0.741). Nomograms thus provide interpretable, clinically applicable tools for risk stratification.

5.2. A novel framework: Imaging–NLP–LLM fusion nomogram

We propose a three-stage framework for constructing a multimodal Nomogram:

(1) Stage 1: Multimodal Data Acquisition and Preprocessing

(a) Imaging data: Retrieve MSCT spine images (DICOM) from PACS. Automatically segment vertebral bodies (T11–T12 or L1–L4) using deep learning segmentation models [19]. Compute trabecular HU values and extract radiomics features (texture, shape, wavelet).

(b) Text data: Apply NLP tools (e.g., ClinicalBERT-based entity recognition) to extract key variables from

EHRs: demographics (age, sex, BMI), fracture history (site, timing, mechanism), medication use (glucocorticoids, osteoporosis drugs), comorbidities (diabetes, rheumatoid arthritis, chronic kidney disease), and laboratory results (if available).

(c) Structured data: Merge laboratory values and any other structured fields with NLP-extracted variables to form a unified clinical feature set.

(2) Stage 2: LLM-Driven Multimodal Fusion Modeling

(a) Encode imaging features and NLP-generated clinical embeddings as input tokens for a pre-trained LLM (e.g., LLaMA-2 fine-tuned on medical corpora).

(b) Employ attention mechanisms to learn cross-modal interactions, outputting a probability for osteoporosis diagnosis or fracture risk.

(c) Use interpretability techniques (e.g., attention weight visualization) to identify the most influential predictors.

(3) Stage 3: Nomogram Construction and Validation

(a) Extract regression coefficients for key predictors from the trained LLM (or from a simpler logistic model fitted using the selected features).

(b) Construct a Nomogram where each predictor is assigned a point scale, and total points correspond to predicted risk.

(c) Validate the Nomogram in independent external cohorts, assessing discrimination (AUC), calibration (Hosmer–Lemeshow test and calibration plots), and clinical net benefit (decision curve analysis).

6. Challenges and future directions

6.1. Technical standardization and reproducibility

HU measurements vary with CT scanner, acquisition protocol (kV, mA), and reconstruction kernel. Cross-institutional comparability requires phantom-based calibration, standardized protocols (e.g., 120 kVp, bone kernel), and deep learning-based HU harmonization.

6.2. Portability of NLP models

NLP models trained on EHRs from one institution may not generalize to others due to differences in documentation style and terminology [15]. Domain adaptation techniques and shared medical language model repositories are needed.

6.3. “Black box” problem and clinical trust

The complexity of LLMs may hinder clinical adoption. Future work should focus on developing inherently interpretable LLM architectures and providing natural language explanations for predictions [17].

6.4. Multicenter prospective validation

Most current studies are retrospective and prone to selection bias. Large-scale, prospective, multicenter cohort studies are essential to validate the predictive performance and cost-effectiveness of the proposed multimodal Nomogram in real-world populations.

6.5. Ethical and privacy considerations

Combining imaging, clinical text, and personal data raises privacy concerns. Techniques such as federated learning can enable multicenter model training without sharing raw data, complying with data protection regulations [19].

7. Conclusion

MSCT-derived trabecular bone attenuation provides an accessible imaging biomarker for opportunistic osteoporosis screening. NLP enables automated extraction of rich clinical information from EHRs, overcoming the limitations of structured data alone. LLMs serve as powerful engines for fusing multimodal imaging and text data, potentially achieving predictive accuracy beyond conventional models. By translating these complex algorithms into a visual Nomogram, individualized risk assessment can be seamlessly integrated into routine clinical workflows. This paradigm not only expands osteoporosis screening beyond postmenopausal women to the broader population undergoing CT examinations but also offers a methodological template for opportunistic screening of other chronic diseases. Future efforts should focus on technical standardization, model interpretability, and rigorous prospective validation to translate this innovation into clinical practice.

Table 1. Comparison of technical approaches for opportunistic osteoporosis screening

Method	Data Sources	Key Features	Advantages	Limitations	Representative Study (AUC)
Conventional DXA	Dedicated DXA equipment	2D areal BMD	International standard, highly standardized	Limited availability, affected by artifacts, no microstructural info	–
Single MSCT HU	Routine CT	Mean HU of single vertebra	Opportunistic, no extra radiation, avoids cortical interference	Lacks microstructural detail, CT parameter dependent	Zou et al. [5](0.84)
CT Radiomics	Routine CT	High-dimensional features (texture, shape)	Captures microstructural heterogeneity, improved diagnostic performance	Features require standardization, less interpretable	Lin et al. [10](0.896)
HU + NLP clinical features	CT + EHR text	Imaging + clinical variables	Integrates multidimensional data	NLP model portability, need annotated training data	–
HU + NLP + LLM multimodal fusion	CT + EHR text + structured data	End-to-end multimodal fusion, semantic understanding	Fully leverages unstructured data, high predictive potential	LLM "black box," high computational demand	Huang et al. [17] (multimodal fusion)
Radiomics Nomogram	CT + clinical parameters	Visual scoring model	Intuitive, clinically applicable	Relies on manual feature selection, limited nonlinear capacity	Lin et al. [10](0.807)

Note: The oculomics example (ChatGPT-4, AUC 0.866) is not included due to the absence of a corresponding reference in the current list; interested readers may refer to recent literature on LLM applications in ophthalmology.

8. Analytical Pathway

The step-by-step analytical pathway is outlined in Section 5.2 and can be summarized as:

Data Acquisition: Collect MSCT images, EHR text, and structured laboratory data.

Preprocessing: Segment vertebrae, extract HU and radiomics features; apply NLP to extract clinical entities; merge all data.

Feature Fusion & Modeling: Use a fine-tuned LLM to learn cross-modal representations and predict osteoporosis/fracture risk.

Nomogram Construction: Select key predictors, fit a regression model, and create a visual Nomogram.

Validation: Assess discrimination, calibration, and clinical utility in external cohorts.

Funding

Health Commission of Chengdu (Project No. 2023488); Institutional Research Grant of Affiliated Hospital of Chengdu University, China (Project No. Y2024017); Health Commission of Chengdu (Project No. 2024015)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Gregson CL, Armstrong DJ, Bowden J, et al., 2022, UK Clinical Guideline for the Prevention and Treatment of Osteoporosis. Archives of Osteoporosis, 17(1): 58.

[2] Kanis JA, Cooper C, Rizzoli R, et al., 2020, Correction to: European Guidance for the Diagnosis and Management of Osteoporosis in Postmenopausal Women. Osteoporosis International, 31(4): 801.

[3] Willers C, Norton N, Harvey NC, et al., 2022, Osteoporosis in Europe: A Compendium of Country-Specific Reports. Archives of Osteoporosis, 17(1): 23.

[4] Pickhardt PJ, Pooler BD, Lauder T, et al., 2013, Opportunistic Screening for Osteoporosis Using Abdominal Computed Tomography Scans Obtained for Other Indications. Annals of Internal Medicine, 158(8): 588–595.

[5] Zou D, Li W, Deng C, et al., 2018, The Use of CT Hounsfield Unit Values to Identify the Undiagnosed Spinal Osteoporosis in Patients with Lumbar Degenerative Diseases. European Spine Journal, 28(8): 1758–1766.

[6] Hiyama A, Sakai D, Katoh H, et al., 2024, Hounsfield Unit Values as an Adjunct Diagnostic Tool: Investigating Its Relationship with Bone Mineral Density and Vertebral Bone Quality in Lumbar Degenerative Disease Patients. World Neurosurgery, 183: e722–e729.

[7] Burke CJ, Didolkar MM, Barnhart HX, et al., 2016, The Use of Routine Non Density Calibrated Clinical Computed Tomography Data as a Potentially Useful Screening Tool for Identifying Patients with Osteoporosis. Clinical Cases in Mineral and Bone Metabolism, 13(2): 135–140.

[8] Wang Y, Wang L, Rastegar-Mojarad M, et al., 2017, Clinical Information Extraction Applications: A Literature Review. Journal of Biomedical Informatics, 77: 34–49.

[9] Moor M, Banerjee O, Abad ZSH, et al., 2023, Foundation Models for Generalist Medical Artificial Intelligence. Nature, 616(7956): 259–265.

[10] Lin X, Shen R, Zheng X, et al., 2024, Utilizing Radiomics Techniques to Isolate a Single Vertebral Body from Chest

CT for Opportunistic Osteoporosis Screening. BMC Musculoskeletal Disorders, 25(1): 785.

[11] Praveen AD, Sollmann N, Baum T, et al., 2024, CT Image-Based Biomarkers for Opportunistic Screening of Osteoporotic Fractures: A Systematic Review and Meta-Analysis. Osteoporosis International, 35(6): 971–996.

[12] Nam KH, Seo I, Kim DH, et al., 2019, Machine Learning Model to Predict Osteoporotic Spine with Hounsfield Units on Lumbar Computed Tomography. Journal of Korean Neurosurgical Society, 62(4): 442–449.

[13] Lambin P, Rios-Velazquez E, Leijenaar R, et al., 2012, Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis. European Journal of Cancer, 48(4): 441–446.

[14] Ford E, Carroll JA, Smith HE, et al., 2016, Extracting Information from the Text of Electronic Medical Records to Improve Case Detection: A Systematic Review. Journal of the American Medical Informatics Association, 23(5): 1007–1015.

[15] Fodeh S, Wang R, Murphy TE, et al., 2024, BoneScore: A Natural Language Processing Algorithm to Extract Bone Mineral Density Data from DXA Scans. Health Informatics Journal, 30(4): 14604582241295930.

[16] Shah NH, Entwistle D, Pfeffer MA, 2023, Creation and Adoption of Large Language Models in Medicine. JAMA, 330(9): 866–869.

[17] Huang SC, Pareek A, Seyyedi S, et al., 2020, Fusion of Medical Imaging and Electronic Health Records Using Deep Learning: A Systematic Review and Implementation Guidelines. NPJ Digital Medicine, 3: 136.

[18] Lessmann N, van Ginneken B, de Jong PA, et al., 2019, Iterative Fully Convolutional Neural Networks for Automatic Vertebra Segmentation and Identification. Medical Image Analysis, 53: 142–155.

[19] Lee SB, Hong Y, Cho YJ, et al., 2024, Enhancing Radiomics Reproducibility: Deep Learning-Based Harmonization of Abdominal Computed Tomography (CT) Images. Bioengineering, 11(12): 1212.