



# 多维特征融合骨质疏松风险评估模型研究\*

王朝亚<sup>1)\*\*</sup> 孟超<sup>2)</sup>

(<sup>1</sup>) 安徽卫生健康职业学院实验实训中心, 池州 247100; <sup>2</sup>) 安徽省第二人民医院骨科, 合肥 240041)

**摘要 目的** 本研究旨在构建并验证一种基于深度学习的多维特征融合风险评估算法模型, 以提升早期识别骨质疏松和预测骨折风险能力。**方法** 本研究纳入来自多组数据库的12 856名受试者数据, 通过多模态深度学习框架, 整合骨密度测量值、骨微结构参数、骨转换标志物、临床风险因素、基因标记物和体感数据等多维特征, 构建风险评估算法模型, 并在独立外部数据集上评估算法性能。**结果** 在测试集上, 该算法预测骨质疏松症的准确率为89.7%, 敏感度为87.5%, 特异度为91.2%, 曲线下面积(AUC)为0.936(95%置信区间(CI): 0.927~0.945), 简化模型AUC为0.917(95%CI: 0.905~0.931), 优于传统FRAX®模型(AUC=0.842, 95%CI: 0.829~0.855), 在独立外部验证集上AUC为0.918(95%CI: 0.905~0.931)。特征重要性分析显示, 骨密度、骨小梁分离度、I型胶原C端肽、平衡参数和特定基因多态性是重要的预测因素。亚组分析显示, 该算法在不同性别、年龄和种族人群中均表现良好。**结论** 基于多维特征融合算法模型可显著提高骨质疏松风险评估的准确性和个体化水平, 具有良好的泛化能力和临床应用前景, 有望为临床实践提供更精准的决策支持工具。

**关键词** 骨质疏松症, 机器学习, 多模态特征融合, 骨折风险, 预测模型

中图分类号 R318.08

DOI: 10.3724/j.pibb.2025.0374

CSTR: 32369.14.pibb.20250374

骨质疏松症(osteoporosis)是一种以骨量减少和骨微结构退化为特征的系统性骨骼疾病, 显著增加骨脆性和骨折风险<sup>[1]</sup>。据统计, 全球约有2亿人患有骨质疏松症, 每年导致890万例脆性骨折<sup>[2]</sup>。随着人口老龄化加剧, 预计骨质疏松性骨折患者数量将持续增加, 给社会经济带来沉重负担。因此, 早期识别骨质疏松和精确评估骨折风险对于制定预防和管理策略至关重要。目前, 骨质疏松症的诊断主要依靠双能X线吸收测定法(dual-energy X-ray absorptiometry, DXA)测量骨密度(bone mineral density, BMD)。按世界卫生组织(World Health Organization, WHO)标准<sup>[3]</sup>, 采用T值表示个体骨密度相对于同性别健康年轻成人峰值骨密度的标准差偏离, 用于评估骨量状态,  $T \leq -2.5 \text{ SD}$ 即可诊断为骨质疏松症。但单纯依靠BMD存在明显局限性, 约有50%的脆性骨折发生在 $T > -2.5 \text{ SD}$ 的患者<sup>[4]</sup>。为提高骨折风险评估的准确性, 研究人员开发了FRAX®、Garvan和QFracture等临床风险评分系统, 将年龄、性别、身体质量指数(body mass index, BMI)、既往骨折史等临床风险因素

纳入考量<sup>[5]</sup>, 但这些工具在不同人群中评估效果表现有差异, 且未能充分整合体感数据、新兴生物标志物和影像学参数等。

近年来, 骨质疏松症评估领域出现多项技术进展, 如高分辨率外周定量计算机断层扫描术(high-resolution peripheral quantitative computer tomography, HR-pQCT)能够评估骨微结构, 骨转换标志物可反映骨代谢状态, 基因组研究识别出与骨质疏松相关的基因位点, 穿戴设备可提供体感数据和步态参数等<sup>[6-7]</sup>, 这些进展为骨质疏松风险的全面评估提供了新的维度。机器学习和人工智能技术也为整合多维度特征提供了强大工具<sup>[8]</sup>。应用Vision Transformer(ViT)和Transformer架构处理医学复杂数据的研究表明<sup>[9-10]</sup>, 深度学习模型在

\* 安徽省教育厅自然科学重点研究项目(KJ2021A1564), 高校优秀青年骨干人才国内访学研修项目(gxgnfx2022211)和安徽省教育厅优秀青年教师培育重点项目(YQZD2024099)资助。

\*\* 通讯联系人。

Tel: 0566-3388850, E-mail: wcy\_2003@163.com

收稿日期: 2025-08-09, 接收日期: 2025-12-01

多模态医学数据融合中具有优势，能够捕捉特征间的复杂交互关系，已有研究尝试将机器学习应用于骨质疏松诊断和骨折风险预测<sup>[11]</sup>，但多数研究仅考虑单一或少数几种数据类型，缺乏全面的多维特征融合风险评估模型，且现有算法多未在外部独立数据集上进行验证，泛化能力存疑。

基于上述背景，本研究旨在开发一种基于深度学习的多维特征融合的综合性骨质疏松风险评估算法模型，整合传统骨密度、骨微结构参数、骨转换标志物、临床风险因素、基因标记物和体感数据等指标参数，并验证其在骨质疏松识别和骨折风险预测中的准确性和临床应用价值（图1）。

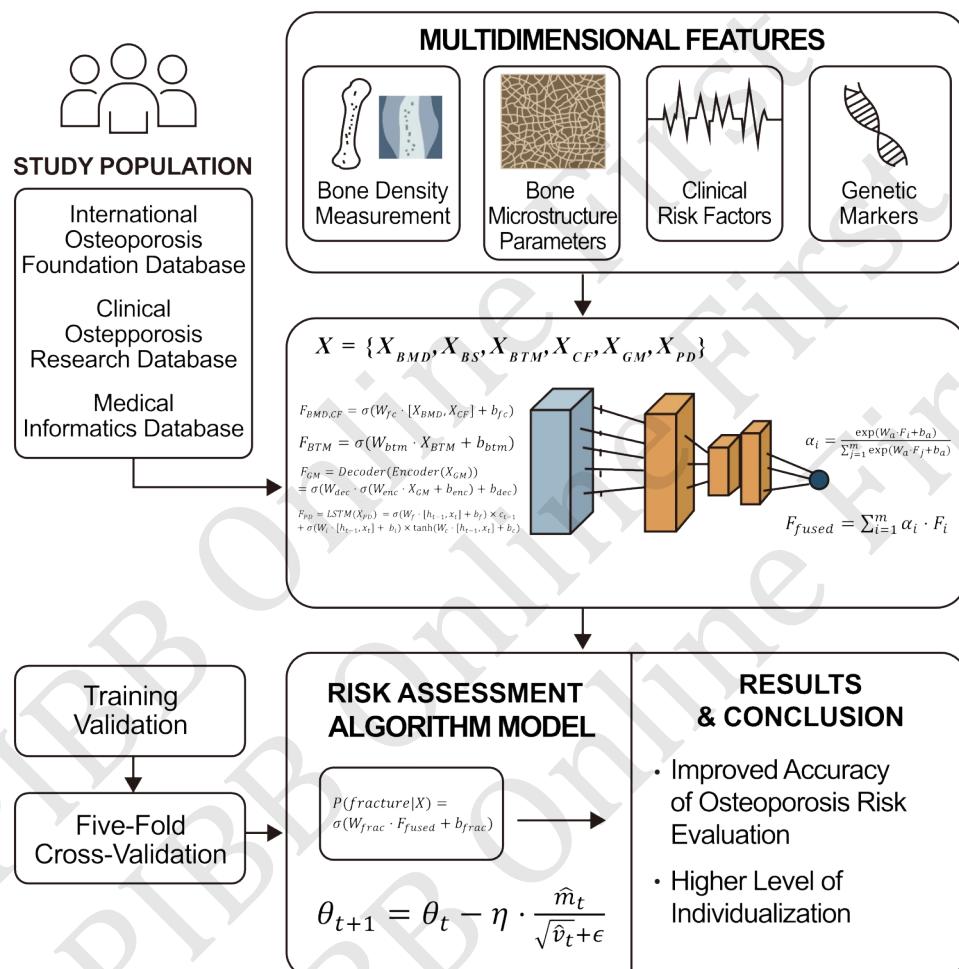


Fig. 1 Flowchart of data selection process

## 1 材料与方法

### 1.1 数据来源与研究设计

本研究数据集来源于3个公开数据库：a. NHANES 数据库 (<https://www.cdc.gov/nchs/nhanes/>)，包含全球26个国家的骨质疏松患者临床资料，主要提供标准化的骨密度测量、临床风险因素和体感数据等；b. UK Biobank 数据库 (<https://www.ukbiobank.ac.uk/>)，提供骨质疏松相关基因、骨代谢标志物、转录组及骨折事件长期追踪等研究

数据；c. OAI (Osteoarthritis Initiative) 数据库 (<https://ndb.nih.gov/oai/>)，提供综合医疗记录，包括影像学特征、结构参数、步态体征数据等。外部独立验证数据采集自安徽省第二人民医院，研究已获得医院伦理委员会批准（批准号：(R) 2025-124）。

数据采集过程采用以下步骤：a. 数据提取：使用统一数据提取协议从3个数据库中抽取标准化的骨质疏松相关数据，采用模糊匹配算法（基于性别、年龄±1岁、种族、首次就诊日期±30 d）识别

可能的重复记录。在数据库间进行去重, 消除样本重叠, 最终确认无重复患者; b. 身份去标识化: 移除所有可能识别个人身份的信息; c. 变量标准化: 建立统一的变量命名和编码系统, 确保跨数据库的一致性; d. 数据质量控制: 使用异常值检测和逻辑校验方法识别并处理数据异常; e. 缺失数据处理: 使用多重插补法 (multiple imputation by chained equations, MICE) 处理缺失数据。采用 Logistic 回归 (分类变量) 和线性回归 (连续变量) 作为插补模型, 通过 Rubin 规则汇总多重插补结果; f. 数据合并: 通过匹配关键变量将不同来源的数据整合到单一分析数据集内。数据集共计纳入 12 856 名受试者, 包括 6 234 名骨质疏松症患者 ( $T \leq -2.5$  SD)、3 987 名骨量减少者 ( $-2.5 \text{ SD} < T < -1.0 \text{ SD}$ ) 和 2 635 名骨量正常者 ( $T \geq -1.0 \text{ SD}$ )。采用分层随机划分方法将数据分为训练集 (80%,  $n=10\,284$ ) 和内部测试集 (20%,  $n=2\,572$ ), 按骨质疏松分类状态进行分层以保证各类别比例一致。从安徽省第二人民医院骨科采集独立外部验证数据集, 用于评估模型的外部泛化能力。纳入标准: 年龄  $\geq 40$  岁; 具备完整 DXA 骨密度测量数据; 至少包含 6 个月随访数据。排除标准: 继发性骨质疏松症; 恶性肿瘤; 接受影响骨代谢药物治疗超过 3 个月; 重要数据缺失超过 30%。依据 WHO 骨质疏松症诊断标准<sup>[12]</sup>, 将受试者分为 3 组: 骨量正常组 ( $T \geq -1.0$  SD)、骨量减少组 ( $-2.5 \text{ SD} < T < -1.0 \text{ SD}$ ) 和骨质疏松组 ( $T \leq -2.5 \text{ SD}$ )。T 值取腰椎、股骨颈和全髋中最低值。

## 1.2 数据特征提取及处理

采用多模态算法拟合以下 6 大类特征数据: a. 骨密度测量值: 通过 DXA 测量腰椎、股骨颈和全髋 BMD 及相应 T 值和 Z 值, 其中 Z 值表示个体骨密度相对于同年龄、同性别、同种族正常人群平均水平的标准差偏离, 用于评估该年龄段骨密度是否异常; b. 骨微结构参数: 包含部分受试者 (4 876 名), 通过 HR-pQCT 采集骨微结构参数, 包括骨体积分数、骨小梁数量、骨小梁厚度、骨小梁分离度、皮质骨厚度等; c. 骨转换标志物: 血清骨形成标志物 (骨特异性碱性磷酸酶、骨钙素、I型前胶原 N 端肽 (N-terminal propeptide of type I procollagen, P1NP)) 和骨吸收标志物 (I型胶原 C 端肽 (C-terminal telopeptide of type I collagen, CTX)、抗酒石酸酸性磷酸酶 5b); d. 临床风险因素: 年龄、性别、BMI、既往骨折史、父母髋部骨

折史、吸烟状态、饮酒习惯、糖皮质激素使用、类风湿关节炎、继发性骨质疏松症危险因素、钙摄入量、维生素 D 水平等; e. 基因标记物: 部分受试者 (3 245 名) 提供基因数据, 检测与骨质疏松相关的 63 个单核苷酸多态性 (single nucleotide polymorphisms, SNPs), 包括核因子 κB 受体活化因子配体 (receptor activator of nuclear factor-κB ligand, RANKL)、骨保护素 (osteoprotegerin, OPG)、维生素 D 受体 (vitamin D receptor, VDR)、雌激素受体 1 (estrogen receptor 1, ESR1)、I 型胶原 α1 链 (collagen type I alpha 1 chain, COL1A1) 以及低密度脂蛋白受体相关蛋白 5 (low-density lipoprotein receptor-related protein 5, LRP5) 等基因位点; f. 体感数据: 部分受试者 (2 768 名) 通过穿戴设备或平衡测试平台采集平衡和步态参数, 包括姿势摆动、单腿站立时间、步态速度、步长变异性等。

采用模糊匹配算法 (基于性别、年龄  $\pm 1$  岁、种族、首次就诊日期  $\pm 30$  d) 在不同数据库间进行去重, 消除样本重叠, NHANES 数据库与 UK Biobank 数据库重叠率为 3.2%, NHANES 数据库与 OAI 数据库重叠率为 1.8%, 三库总体重叠率  $< 5\%$ 。

## 1.3 算法开发

本算法基于多模态深度学习框架开发, 主要包括以下步骤: a. 数据预处理: 对连续变量进行标准化处理 (Z-分数), 即将每个数值减去该变量的总体均值后再除以其标准差, 使其转化为均值为 0、标准差为 1 的标准化尺度。对分类变量进行编码, 缺失数据通过多重插补法处理; b. 特征提取: 针对不同数据类型设计特定的特征提取模块。骨微结构数据通过 3D 卷积神经网络提取特征, 临床数据通过全连接神经网络处理, 基因数据通过自编码器降维, 时序数据 (体感数据) 通过长短期记忆网络 (long short-term memory, LSTM) 处理; c. 特征融合: 采用多尺度注意力机制融合不同模态的特征表示, 动态调整各类特征的权重; d. 预测输出: 设计双输出结构, 同时预测骨质疏松状态 (正常/骨量减少/骨质疏松) 和 10 年骨折风险; e. 训练策略: 采用五折交叉验证方法训练策略, 使用 Adam 优化器, 学习率为 0.001, 批量大小为 64, 采取早停策略防止过拟合。考虑到临床实际情况, 算法设计为模块化结构, 能够适应不同组合的输入特征, 在数据不完整 ( $< 20\%$ ) 情况下仍能提供评估结果。

## 1.4 算法数学表示

本算法可通过以下数学公式表示：

对于输入特征

$X = \{X_{BMD}, X_{BS}, X_{BTM}, X_{CF}, X_{GM}, X_{PD}\}$ , 其中,  $X_{BMD}$ : 骨密度测量特征向量,  $X_{BS}$ : 骨微结构参数特征向量,  $X_{BTM}$ : 骨转换标志物特征向量,  $X_{CF}$ : 临床风险因素特征向量,  $X_{GM}$ : 基因标记物特征向量,  $X_{PD}$ : 体感数据特征向量。

### 1.4.1 特征提取

每类特征通过特定的特征提取网络处理：

骨密度和临床特征（全连接网络）：

$$F_{BMD,CF} = \sigma(W_{fc} \cdot [X_{BMD}, X_{CF}] + b_{fc}) \quad (1)$$

骨微结构特征（3D-CNN）：

$$F_{BS} = CNN_{3D}(X_{BS}) = \sigma(\sum_{i=1}^n W_i * X_{BS} + b_i) \quad (2)$$

骨转换标志物（全连接网络）：

$$F_{BTM} = \sigma(W_{btm} \cdot X_{BTM} + b_{btm}) \quad (3)$$

基因标记物（自编码器）：

$$F_{GM} = Decoder(Encoder(X_{GM})) = \sigma(W_{dec} \cdot \sigma(W_{enc} \cdot X_{GM} + b_{enc}) + b_{dec}) \quad (4)$$

体感数据（LSTM）：

$$\begin{aligned} F_{PD} = LSTM(X_{PD}) &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \times \\ &\quad c_{t-1} + \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \times \tanh(W_e \cdot \\ &\quad [h_{t-1}, x_t] + b_e) \end{aligned} \quad (5)$$

其中,  $\sigma$  表示激活函数 (ReLU),  $W$  和  $b$  分别表示权重矩阵和偏置向量,  $*$  表示卷积操作。

### 1.4.2 多模态特征融合

采用注意力机制进行特征融合：

首先计算每个特征模态的注意力权重：

$$\alpha_i = \frac{\exp(W_a \cdot F_i + b_a)}{\sum_{j=1}^m \exp(W_a \cdot F_j + b_a)} \quad (6)$$

其中,  $F_i$  表示第  $i$  个特征模态的表示,  $m$  是特征模态的总数。

然后进行加权融合：

$$F_{fused} = \sum_{i=1}^m \alpha_i \cdot F_i \quad (7)$$

为捕捉特征间的交互作用, 融合层后添加一个残差块 (residual block) 和交叉特征交互层 (cross-feature interaction layer), 通过双线性池化 (bilinear pooling) 计算特征对之间的交互项。

### 1.4.3 预测模型

对于骨质疏松状态分类（三分类问题）：

$$\begin{aligned} P(y_{oste} = cl|X) &= \text{softmax}(W_{oste} \cdot F_{fused} + b_{oste}) = \\ &\quad \frac{\exp(W_{oste,c} \cdot F_{fused} + b_{oste,c})}{\sum_{k=1}^3 \exp(W_{oste,k} \cdot F_{fused} + b_{oste,k})} \end{aligned} \quad (8)$$

其中,  $c \in \{\text{正常, 骨量减少, 骨质疏松}\}$

对于骨折风险预测（回归问题）：

$$P(fracture|X) = \sigma(W_{frac} \cdot F_{fused} + b_{frac}) \quad (9)$$

其中,  $\sigma$  为 Sigmoid 激活函数, 确保输出值在 [0, 1] 区间内。

### 1.4.4 损失函数

多任务学习损失函数：

$$L = \lambda_1 L_{CE} + \lambda_2 L_{BCE} + \lambda_3 L_{reg} \quad (10)$$

其中:  $L_{CE}$  是骨质疏松状态分类的交叉熵损失  $L_{CE} = -\sum_{i=1}^n \sum_{c=1}^3 y_{i,c} \log(P(y_i = cl|X_i))$ ;  $L_{BCE}$  是骨折风险预测的二元交叉熵损失  $L_{BCE} = -\sum_{i=1}^n [y_i \log(P(fracture|X_i)) + (1 - y_i) \log(1 - P(fracture|X_i))]$ ; 防止过拟合正则化项  $L_{reg} = \sum_j \|W_j\|_2^2$ ; 权衡不同损失贡献的超参数 ( $\lambda_1, \lambda_2, \lambda_3$ ) 设置为 (1.0, 1.0, 0.01)。

### 1.4.5 模型优化

采用 Adam 优化器更新参数：

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (11)$$

其中,  $\eta$  是学习率,  $\hat{m}_t$  和  $\hat{v}_t$  分别是一阶矩估计和二阶矩估计的校正值,  $\epsilon$  是小常数防止除零。

### 1.4.6 简化模型

对于临床环境中不可获取全部特征的情况, 设计了简化模型 (仅使用 BMD+临床特征+骨转换标志物)。通过迁移学习应用全模型的知识, 使用知识蒸馏 (knowledge distillation) 方法：

$$L_{simplified} = \arg \min_f \sum_{i=1}^n |f(X_{i,\text{subset}}) - f_{full}(X_i)|^2 + \mu \cdot \Omega(f) + \tau \cdot D_{KL}(P_{full} || P_{simplified}) \quad (12)$$

其中,  $f_{full}$  是全特征模型,  $\Omega(f)$  是模型复杂度惩罚项,  $\mu = 10^{-4}$  是正则化参数,  $D_{KL}$  是 Kullback-Leibler 散度 (衡量两个分布的差异),  $\tau = 0.5$  是知识蒸馏的温度参数。

### 1.5 模型评估与比较

使用以下指标评估算法性能, 骨质疏松诊断为准确率、敏感度、特异度、阳性预测值、阴性预测值、曲线下面积 (area under curve, AUC); 骨折风险预测为 AUC、校准曲线、决策曲线分析 (decision curve analysis, DCA)、净重分类指数 (net reclassification improvement, NRI)、综合判别改善指数 (integrated discrimination improvement, IDI); 使用 DeLong 检验比较本模型与骨折风险评估工具 (fracture risk assessment

tool, FRAX) 之间受试者操作特征 (receiver operating characteristic, ROC) AUC 差异; 亚组分析在不同性别 (男性/女性)、年龄段 (<65岁/≥65岁)、骨密度水平 (正常/骨量减少/骨质疏松) 和种族 (亚洲/欧洲/非洲) 人群中评估算法性能。

Brier 评分 (Brier score) 评估模型校准性:  

$$\text{Brier} = \frac{1}{n} \sum_{i=1}^n (P_i - y_i)^2$$
, 值越小表示校准性越好。

## 1.6 统计分析

连续变量以均值±标准差或中位数 (四分位间距) 表示, 分类变量以频数 (百分比) 表示。正态分布检验使用 Shapiro-Wilk 检验。符合正态分布的连续变量组间比较采用独立样本 *t* 检验, 非正态分布连续变量采用 Mann-Whitney U 检验, 分类变量采用卡方检验。三组数据比较采用单因素方差分析 (ANOVA) 或 Kruskal-Wallis 检验, 事后分析采用

Bonferroni 校正。生存分析采用 Kaplan-Meier 方法和 Cox 比例风险模型, 并计算 95% 置信区间 (confidence interval, CI)。所有统计分析使用 R 4.1.0 和 Python 3.8.0 进行,  $P<0.05$  差异具有统计学意义。

## 2 结果

### 2.1 研究对象基线特征

算法开发数据集包含 12 856 名受试者, 平均年龄  $65.8\pm10.2$  岁, 女性占 76.2%。测试数据集包含 2 346 名受试者, 平均年龄  $66.2\pm9.8$  岁, 女性占 74.8%。外部独立验证数据集包含 380 名受试者, 平均年龄  $62.7\pm8.1$  岁, 女性占 68.4%。使用 Logistic 回归模型基于年龄、性别、BMI 等协变量计算倾向性评分匹配 (propensity score matching, PSM), 标准化差异  $<0.1$  (表 1)。

Table 1 Baseline characteristics of subjects in the algorithm development dataset

| Characteristics  | Normal bone mass<br>(n=2 635) | Osteopenia (n=3 987) | Osteoporosis (n=6 234) | P value |
|--|-------------------------------|----------------------|------------------------|---------|
| Age (years)  | 64.1±9.3                      | 64.8±9.5             | 65.2±9.7               | 0.157   |
| Female, n (%)  | 1, 970 (74.8)                 | 3, 015 (75.6)        | 4, 755 (76.2)          | 0.396   |
| BMI (kg/m <sup>2</sup> )                               | 25.8±4.1                      | 25.5±4.0             | 25.1±3.9               | 0.068   |
| Prior fracture, n (%)                                  | 186 (7.1)                     | 628 (15.8)           | 1, 847 (29.6)          | <0.001  |
| Parental history of hip fracture, n (%)                | 367 (13.9)                    | 521 (13.1)           | 1, 008 (16.2)          | <0.05   |
| Current smoking, n (%)                                 | 421 (16.0)                    | 712 (17.9)           | 1, 156 (18.5)          | 0.108   |
| High alcohol intake, n (%)                             | 251 (9.5)                     | 408 (10.2)           | 718 (11.5)             | 0.062   |
| Glucocorticoid use, n (%)                              | 178 (6.8)                     | 356 (8.9)            | 887 (14.2)             | <0.001  |
| Rheumatoid arthritis, n (%)                            | 148 (5.6)                     | 283 (7.1)            | 598 (9.6)              | <0.05   |
| Lumbar spine BMD (g/cm <sup>2</sup> )                  | 1.14±0.12                     | 0.95±0.09            | 0.79±0.11              | <0.05   |
| Femoral neck BMD (g/cm <sup>2</sup> )                  | 0.94±0.11                     | 0.79±0.08            | 0.65±0.09              | <0.001  |
| Total hip BMD (g/cm <sup>2</sup> )                     | 1.02±0.12                     | 0.868±0.08           | 0.715±0.10             | <0.001  |
| Osteocalcin (μg/L)                                     | 15.1±5.3                      | 19.8±6.9             | 25.6±9.4               | <0.05   |
| C-terminal telopeptide of type I collagen (CTX) (μg/L) | 0.33±0.15                     | 0.49±0.22            | 0.67±0.29              | <0.001  |
| [25(OH) D] (nmol/L)                                    | 76.2±25.8                     | 67.5±24.1            | 58.9±22.3              | <0.001  |

Notes: Serum 25(OH) D levels did not follow a normal distribution and are therefore presented as median (interquartile range). Group comparisons were performed using the Kruskal-Wallis test.

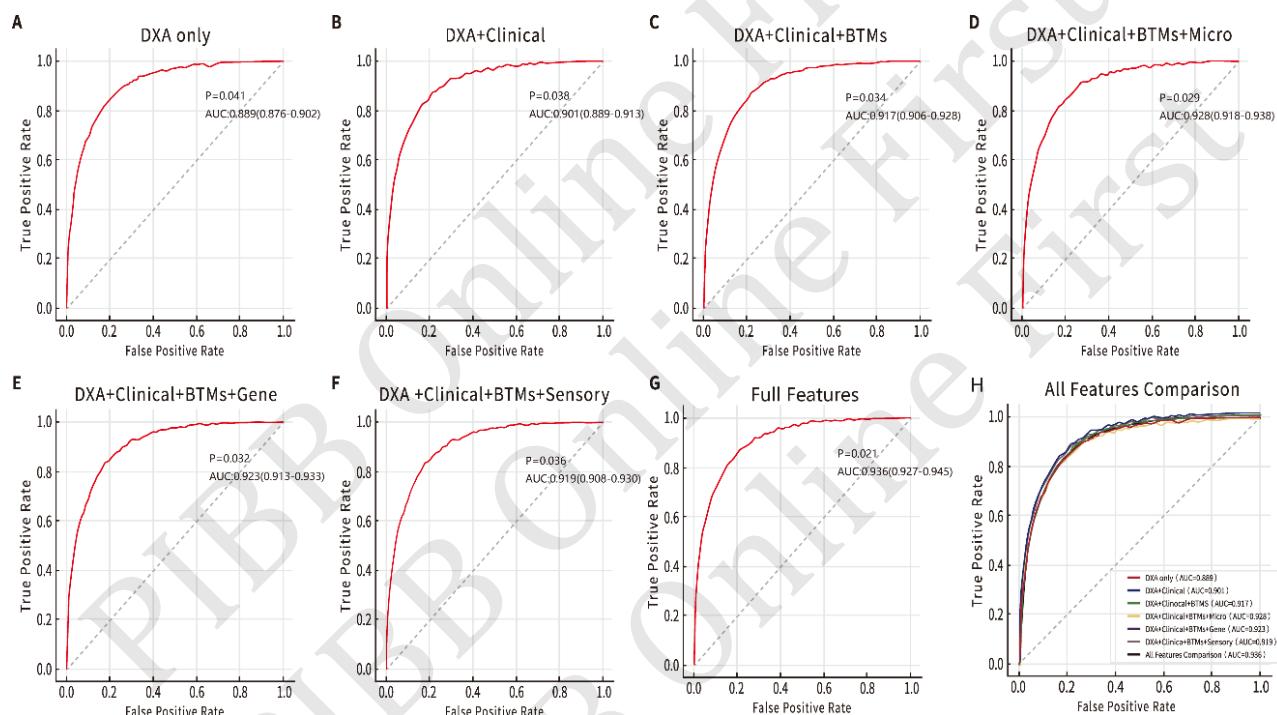
## 2.2 模型性能

在测试集上该算法预测骨质疏松症的准确率为 89.7%, 敏感度为 87.5%, 特异度为 91.2%, 阳性预测值为 90.8%, 阴性预测值为 88.3%, AUC 为 0.936 (95%CI: 0.927~0.945)。相比之下, 仅使用 DXA 的 AUC 为 0.889 (95%CI: 0.876~0.902), 传统 FRAX 模型的 AUC 为 0.842 (95%CI: 0.829~

0.855)。DeLong 检验显示, 该算法的 AUC 显著高于 DXA 和 FRAX ( $P<0.05$ )。在外部验证集上, 全特征模型的 AUC 为 0.918 (95% CI: 0.905~0.931), 敏感度为 85.3%, 特异度为 89.7%, Brier 评分为 0.087, 说明模型具有良好的校准性和外部泛化能力 (表 2, 图 2)。

**Table 2 Osteoporosis diagnostic performance of different feature combinations**

| Feature combination          | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC (95%CI)                | Brier score  |
|------------------------------|--------------|-----------------|-----------------|----------------------------|--------------|
| DXA                          | 81.2         | 83.6            | 79.5            | 0.889 (0.876–0.902)        | 0.124        |
| DXA+Clinical                 | 83.8         | 84.2            | 83.4            | 0.901 (0.889–0.913)        | 0.108        |
| DXA+Clinical+BTMs            | 86.5         | 85.3            | 87.2            | 0.917 (0.906–0.928)        | 0.093        |
| DXA+Clinical+BTMs+Bone Micro | 88.4         | 86.9            | 89.5            | 0.928 (0.918–0.938)        | 0.089        |
| DXA+Clinical+BTMs+Gene       | 87.9         | 86.3            | 89.1            | 0.923 (0.913–0.933)        | 0.094        |
| DXA+Clinical+BTMs+Sensory    | 87.3         | 85.8            | 88.5            | 0.919 (0.908–0.930)        | 0.098        |
| All features                 | <b>89.7</b>  | <b>87.5</b>     | <b>91.2</b>     | <b>0.936 (0.927–0.945)</b> | <b>0.083</b> |
| Simplified model             | 86.5         | 85.3            | 87.2            | 0.917 (0.906–0.928)        | 0.093        |
| Independent test set         | 81.2         | 85.3            | 89.7            | 0.918 (0.905–0.931)        | 0.087        |

**Fig. 2 ROC curves for osteoporosis diagnosis performance using different feature combinations**

### 2.3 特征重要性分析

通过 SHAP (SHapley Additive exPlanations) 值分析各特征的相对重要性，确定对算法预测结果贡献最大的特征。表3展示了预测骨质疏松症和骨折风险的前15个重要特征。

结果显示，骨微结构参数（骨小梁分离度、皮质骨厚度）、骨转换标志物（I型胶原C端肽、骨钙

素）、体感参数（单腿站立时间、步态速度）和特定基因多态性（RANKL、VDR基因）是预测骨质疏松和骨折风险的重要特征。

### 2.4 亚组分析

表4为亚组分析评估模型在不同人群中的表现，该模型在不同性别、年龄和种族人群中均表现良好。

**Table 3 SHAP value ranking of important features in this algorithm**

| Rank | Features for osteoporosis diagnosis             | SHAP | Features for fracture risk prediction | SHAP |
|------|---|------|---------------------------------------|------|
| 1    | Femoral neck BMD                                | 3.42 | Age                                   | 2.96 |
| 2    | Lumbar spine BMD                                | 3.18 | Prior fracture                        | 2.87 |
| 3    | Age   | 2.76 | Femoral neck BMD                      | 2.82 |
| 4    | Trabecular separation                           | 2.41 | Trabecular separation                 | 2.58 |
| 5    | C-terminal telopeptide of type I collagen (CTX) | 2.25 | Fall risk score                       | 2.43 |
| 6    | Osteocalcin                                     | 2.12 | Single-leg stance time                | 2.26 |
| 7    | Bone-specific alkaline phosphatase (BALP)       | 1.96 | CTX                                   | 2.15 |
| 8    | Cortical thickness                              | 1.85 | Gait speed                            | 1.98 |
| 9    | Serum 25(OH)D level                             | 1.74 | BMI                                   | 1.87 |
| 10   | BMI   | 1.68 | Cortical thickness                    | 1.75 |
| 11   | Sex   | 1.56 | Glucocorticoid use                    | 1.64 |
| 12   | Gait speed                                      | 1.45 | RANKL SNP rs2277438                   | 1.52 |
| 13   | RANKL SNP rs2277438                             | 1.38 | Trabecular number                     | 1.48 |
| 14   | VDR SNP rs1544410                               | 1.29 | Sex                                   | 1.37 |
| 15   | Serum calcium level                             | 1.23 | VDR SNP rs1544410                     | 1.28 |

SHAP values denote the mean absolute contribution of each feature to the model output and are unitless.

**Table 4 Performance of the deep learning model in different subgroups**

| Category                   | Sample size                         | AUC (95%CI)         | Sensitivity (%) | Specificity (%) | P value* |
|----------------------------|-------------------------------------|---------------------|-----------------|-----------------|----------|
| Sex                        | Male (n=3 053)                      | 0.942 (0.932–0.952) | 88.6            | 92.1            | 0.013    |
|                            | Female (n=9 803)                    | 0.918 (0.901–0.935) | 85.2            | 89.7            |          |
| Age                        | <65 years (n=5 954)                 | 0.923 (0.909–0.937) | 84.8            | 90.3            | 0.036    |
|                            | ≥65 years (n=6 902)                 | 0.940 (0.929–0.951) | 89.5            | 92.0            |          |
| Bone Mineral Density Level | Normal bone mass (n=2 635)          | 0.896 (0.878–0.914) | 82.3            | 88.5            |          |
|                            | Osteopenia (n=3 987)                | 0.921 (0.907–0.935) | 85.7            | 90.4            | 0.007**  |
|                            | Osteoporosis (n=6 234)              | 0.945 (0.935–0.955) | 90.2            | 92.8            |          |
| Ethnicity                  | Asian (n=2 998)                     | 0.933 (0.919–0.947) | 87.1            | 91.4            |          |
|                            | European (n=8 693)                  | 0.938 (0.926–0.950) | 88.3            | 91.6            | 0.108**  |
|                            | African (n=1 165)                   | 0.904 (0.872–0.936) | 83.6            | 88.9            |          |
| Feature Completeness       | Complete (all 6 feature categories) | 0.936 (0.924–0.948) | 87.5            | 91.2            |          |
|                            | Partial (4–5 feature categories)    | 0.924 (0.911–0.937) | 85.8            | 90.1            | 0.039    |
|                            | Deficient (<3 feature categories)   | 0.891 (0.885–0.907) | 81.6            | 85.9            |          |

\*AUC differences across subgroups were assessed using the DeLong test; \*\*comparisons among the three groups were conducted using one-way ANOVA.

### 3 讨论

本研究开发并验证了一种基于多维特征融合的综合性骨质疏松风险评估算法模型。该算法可拟合骨密度、骨微结构、骨转换标志物、临床风险因素、基因标记物和体感信息等6大类多维度数据，显著提高骨质疏松诊断和骨折风险预测的准确性。传统骨质疏松评估主要依赖DXA测量的骨密度，而BMD单一指标存在明显局限性。Naylor等<sup>[13]</sup>的研究表明，约50%的骨折发生在BMD未达到骨

质疏松诊断标准的患者中，这与观察到骨量正常和骨量减少人群中骨折事件一致。FRAX等风险评分系统虽然纳入了临床风险因素，但这些工具在不同人群中预测准确性存在较大差异，且未能整合骨微结构参数、生物标志物、体感数据等关键因素<sup>[14]</sup>。本研究表明，整合多维特征的算法模型在骨质疏松诊断(AUC=0.936)和骨折风险预测(AUC=0.912)方面均优于传统方法。这一结果与近期几项机器学习研究一致，如Badgeley等<sup>[15]</sup>基于深度学习的骨折预测模型和Baik等<sup>[16]</sup>开发的多参数骨

质疏松评估系统。近期发表的 Hong 等<sup>[17]</sup>的研究使用卷积神经网络分析骨密度图像，结合传统风险因素，进一步支持了人工智能算法模型在骨质疏松诊断和预测中的潜力。

Samelson 等<sup>[18]</sup>纵向研究发现，骨微结构参数与骨折风险独立相关，即使在调整 BMD 后仍然显著。SHAP 分析进一步量化了这一贡献，骨小梁分离度在预测骨质疏松 ( $SHAP=2.41$ ) 和骨折风险 ( $SHAP=2.58$ ) 方面均排在前五位。这与 Boutroy 等<sup>[19]</sup> 和 Nishiyama 等<sup>[20]</sup> 关于 HR-pQCT 评估骨微结构对骨强度独立贡献的研究结果一致。Lu 等<sup>[21]</sup> 最新研究表明，结合骨微结构评估可将骨折风险预测 AUC 值显著提高，而本研究通过更全面的多维特征融合将其进一步提高至 0.912。采用多模态深度学习框架整合多维特征数据，相比单一特征或简单组合具有显著优势。Langdahl<sup>[22]</sup> 指出，骨质疏松是一种多因素疾病，不同患者的病理机制和风险因素存在显著异质性，这要求评估工具能够涵盖多维度特征。通过特征融合分析，随着特征维度增加，模型性能逐步提高，全特征模型 ( $AUC=0.936$ ) 优于任何单一或部分特征组合。这与 Stahlschmidt 等<sup>[23]</sup> 提出的多模态医学数据融合框架和 Shaik 等<sup>[24]</sup> 关于多源异构医疗数据整合的系统综述结论基本一致。

在特征重要性分析中，骨转换标志物 (I型胶原 C 端肽、骨钙素、骨特异性碱性磷酸酶) 显示出与传统骨密度互补的价值。这与 Eastell 等<sup>[25]</sup> 研究结果一致，即骨转换标志物反映骨代谢状态，可提前预测骨密度变化和治疗反应。特别是 I型胶原 C 端肽在本模型中排名第 5 ( $SHAP=2.25$ ) 和第 7 ( $SHAP=2.15$ )，分别用于骨质疏松诊断和骨折预测。Johansson 等<sup>[26]</sup> 的荟萃分析显示，I型胶原 C 端肽增加一个标准差与骨折风险增加 23% 相关。而在本模型中，通过与其他特征的协同作用，其预测价值得到进一步增强。基因标记物的整合显示，RANKL (rs2277438) 和 VDR (rs1544410) 多态性在模型中表现突出，VDR (rs1544410) 是经典的 VDR BsmI 位点，在维生素 D 受体基因 3' 端区域，是骨代谢研究里最常见的多态之一，这与 Styrkarsdottir 等<sup>[27]</sup> 的全基因组关联研究结果相符。但与 Estrada 等<sup>[28]</sup> 的研究不同，本研究发现，单个 SNP 的预测价值有限，只有在与临床和生化指标结合时才显著提高模型性能。这一发现支持了 Zheng 等<sup>[29]</sup> 提出的多基因-环境交互作用框架，即骨质疏

松的遗传风险通过与环境因素的复杂交互发挥作用。本研究首次将体感数据（平衡和步态参数）纳入综合风险评估模型，单腿站立时间和步态速度在骨折预测中排名第 6 和第 8，Harvey 等<sup>[30]</sup> 对 MrOS 队列的荟萃分析表明，既往跌倒史在调整 FRAX 概率后仍可独立预测骨折风险，提示将跌倒风险因素纳入综合评估模型具有重要意义。Montero-Odasso 等<sup>[31]</sup> 指出，跌倒风险评估是骨折预防策略的关键组成部分，但传统骨质疏松评估工具往往忽视了这一维度，多维特征融合提供了更全面的风险评估。

多维特征融合模型算法的主要优势在于：a. 提供更全面的骨骼健康评估，不仅包括骨量，还包括骨质量和跌倒风险；b. 适应个体差异，针对不同原因导致的骨质疏松提供个体化风险评估；c. 克服单一指标的局限性，如 BMD 正常但骨微结构异常的情况。

本研究尚存在一定局限性。如研究采用横断面设计，难以确立静态特征与骨折风险的因果关系，模型预测的是相关性而非因果性。研究样本也缺乏长期纵向随访数据。深度学习模型也存在“黑箱”特性，无法像传统统计模型一样提供直观的系数解释。多模态特征的融合虽提高了预测准确性，但在实际临床工作中各类特征的获取成本和难度差异显著。此外，基因标记物的贡献相对有限，未能充分捕捉复杂环境-遗传交互作用。多中心、跨地区的前瞻性验证研究以及与现有临床信息系统的集成策略仍有待进一步探索。

## 4 结论

本研究开发并验证了一种基于多维特征融合的综合性骨质疏松风险评估算法模型，通过整合骨密度、骨微结构、骨转换标志物、临床风险因素、基因标记物和体感等数据，显著提高了骨质疏松诊断和骨折风险预测的准确性。该算法在不同性别、年龄和种族人群中均表现良好，具有良好临床应用价值。SHAP 分析显示，骨微结构参数、骨转换标志物、体感数据和特定基因多态性在骨质疏松风险评估中的重要性，为骨质疏松个体化风险评估和精准治疗提供了新的视角。

## 参考文献

- [1] Compston J E, McClung M R, Leslie W D. Osteoporosis. Lancet, 2019, **393**(10169): 364-376
- [2] Pisani P, Renna M D, Conversano F, et al. Major osteoporotic

- fragility fractures: risk factor updates and societal impact. *World J Orthop.* 2016, **7**(3): 171-181
- [3] Kanis JA, Melton III L J, Christiansen C, et al. The diagnosis of osteoporosis. *J Bone Miner Res.* 1994, **9**(8): 1137-1141
- [4] Siris E S, Chen Y T, Abbott T A, et al. Bone mineral density thresholds for pharmacological intervention to prevent fractures. *Arch Intern Med.* 2004, **164**(10): 1108-1112
- [5] Kanis JA, Johnell O, Oden A, et al. FRAX™ and the assessment of fracture probability in men and women from the UK. *Osteoporos Int.* 2008, **19**(4): 385-397
- [6] Doshi K B, Moon S H, Whitaker M D, et al. Assessment of gait and posture characteristics using a smartphone wearable system for persons with osteoporosis with and without falls. *Sci Rep.* 2023, **13**(1): 538
- [7] Subramaniam S, Faisal A I, Deen M J. Wearable sensor systems for fall risk assessment: a review. *Front Digit Health.* 2022, **4**: 921506
- [8] Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med.* 2019, **380**(14): 1347-1358
- [9] Shamshad F, Khan S, Zamir S W, et al. Transformers in medical imaging: a survey. *Med Image Anal.* 2023, **88**: 102802
- [10] Takahashi S, Sakaguchi Y, Kouno N, et al. Comparison of vision transformers and convolutional neural networks in medical image analysis: a systematic review. *J Med Syst.* 2024, **48**(1): 84
- [11] Kruse C, Eiken P, Vestergaard P. Machine learning principles can improve hip fracture prediction. *Calcif Tissue Int.* 2017, **100**(4): 348-360
- [12] World Health Organization. Prevention and management of osteoporosis. Geneva: World Health Organization, 2003: 1 - 164
- [13] Naylor K, Eastell R. Bone turnover markers: use in osteoporosis. *Nat Rev Rheumatol.* 2012, **8**(7): 379-389
- [14] Kanis J A, Harvey N C, Johansson H, et al. A decade of FRAX: how has it changed the management of osteoporosis?. *Aging Clin Exp Res.* 2020, **32**(2): 187-196
- [15] Badgeley M A, Zech J R, Oakden-Rayner L, et al. Deep learning predicts hip fracture using confounding patient and healthcare variables. *NPJ Digit Med.* 2019, **2**: 31
- [16] Baik S M, Kwon H J, Kim Y, et al. Machine learning model for osteoporosis diagnosis based on bone turnover markers. *Health Informatics J.* 2024, **30**(3): 14604582241270778
- [17] Hong N, Cho S W, Shin S, et al. Deep-learning-based detection of vertebral fracture and osteoporosis using lateral spine X-ray radiography. *J Bone Miner Res.* 2023, **38**(6): 887-895
- [18] Samelson E J, Broe K E, Xu H, et al. Cortical and trabecular bone microarchitecture as an independent predictor of incident fracture risk in older women and men in the Bone Microarchitecture International Consortium (BoMIC): a prospective study. *Lancet Diabetes Endocrinol.* 2019, **7**(1): 34-43
- [19] Boutroy S, Van Rietbergen B, Sornay-Rendu E, et al. Finite element analysis based on *in vivo* HR-pQCT images of the distal radius is associated with wrist fracture in postmenopausal women. *J Bone Miner Res.* 2008, **23**(3): 392-399
- [20] Nishiyama K K, Shane E. Clinical imaging of bone microarchitecture with HR-pQCT. *Curr Osteoporos Rep.* 2013, **11**(2): 147-155
- [21] Lu S, Fuggle N R, Westbury L D, et al. Machine learning applied to HR-pQCT images improves fracture discrimination provided by DXA and clinical risk factors. *Bone.* 2023, **168**: 116653
- [22] Langdahl B L. Overview of treatment approaches to osteoporosis. *Br J Pharmacol.* 2021, **178**(9): 1891-1906
- [23] Stahlschmidt S R, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. *Brief Bioinform.* 2022, **23**(2): bbab569
- [24] Shaik T, Tao X, Li L, et al. A survey of multimodal information fusion for smart healthcare: mapping the journey from data to wisdom. *InfFusion.* 2024, **102**: 102040
- [25] Eastell R, Pigott T, Gossiel F, et al. DIAGNOSIS OF ENDOCRINE DISEASE: bone turnover markers: are they clinically useful?. *Eur J Endocrinol.* 2018, **178**(1): R19-R31
- [26] Johansson H, Odén A, Kanis JA, et al. A meta-analysis of reference markers of bone turnover for prediction of fracture. *Calcif Tissue Int.* 2014, **94**(5): 560-567
- [27] Styrkarsdottir U, Halldorsson B V, Gretarsdottir S, et al. Multiple genetic loci for bone mineral density and fractures. *N Engl J Med.* 2008, **358**(22): 2355-2365
- [28] Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet.* 2012, **44**(5): 491-501
- [29] Zheng H F, Forgetta V, Hsu Y H, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature.* 2015, **526**(7571): 112-117
- [30] Harvey N C, Odén A, Orwoll E, et al. Falls predict fractures independently of FRAX probability: a meta-analysis of the osteoporotic fractures in men (MrOS) study. *J Bone Miner Res.* 2018, **33**(3): 510-516
- [31] Montero-Odasso M, van der Velde N, Martin F C, et al. World guidelines for falls prevention and management for older adults: a global initiative. *Age Ageing.* 2022, **51**(9): afac205

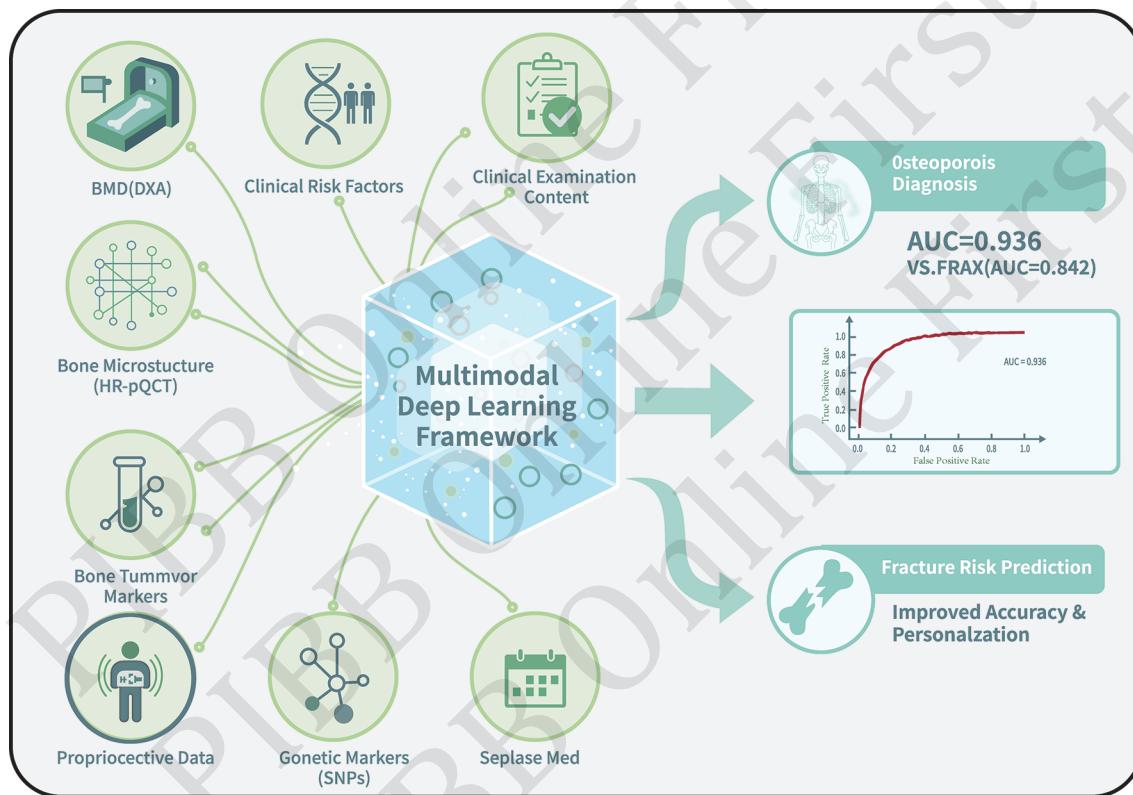
## Research on Multi-dimensional Feature Fusion Model for Osteoporosis Risk Assessment Based on Deep Learning\*

WANG Chao-Ya<sup>1)\*\*</sup>, MENG Chao<sup>22)</sup>

(<sup>1</sup>)Experimental Training Center, Anhui Health College, Chizhou 247100, China;

(<sup>2</sup>)Department of Orthopedics, Anhui Second People's Hospital, Hefei 240041, China)

### Graphical abstract



**Abstract Objective** Osteoporosis is a progressive metabolic bone disorder characterized by reduced bone mass and microarchitectural deterioration, leading to increased skeletal fragility and susceptibility to fracture. Conventional diagnostic and risk-assessment approaches, such as dual-energy X-ray absorptiometry (DXA) and the FRAX® algorithm, remain limited because they rely primarily on bone mineral density (BMD) and a restricted set of clinical factors, failing to capture the multidimensional determinants of bone strength. This study aimed to develop and validate a deep learning - based multi-dimensional feature fusion model that integrates heterogeneous biological, structural, functional, and genetic information to improve the early identification of osteoporosis and enhance fracture risk prediction. **Methods** A total of 12 856 participants were aggregated from three major data repositories: the International Osteoporosis Foundation database, a clinical research database on osteoporosis, and a large-scale medical informatics dataset. A unified data-extraction protocol was applied to ensure cross-database harmonization, followed by quality control, variable standardization, and missing-data handling using multiple imputation by chained equations (MICE). A multimodal deep learning framework was

constructed to integrate six categories of features: BMD measurements, quantitative bone microarchitecture parameters, bone turnover biomarkers, established clinical risk factors, osteoporosis-related genetic polymorphisms, and sensor-derived balance and gait metrics. A multi-task learning strategy was adopted to simultaneously predict osteoporosis status and 10-year fracture probability. Model training used five-fold cross-validation, and external validation was conducted in an independent clinical cohort. Model performance was benchmarked against DXA alone and the FRAX tool. **Results** In the internal test cohort, the proposed model achieved an *AUC* of 0.936 (95% *CI*: 0.927 – 0.945), with a sensitivity of 87.5% and a specificity of 91.2%, significantly outperforming DXA alone (*AUC*=0.889) and FRAX (*AUC*=0.842) (both *P*<0.05). External validation yielded an *AUC* of 0.918 (95% *CI*: 0.905 – 0.931) and demonstrated strong calibration (Brier score=0.087). SHAP analyses revealed that, beyond BMD, key predictors included trabecular separation, serum C-terminal telopeptide of type I collagen, balance-related metrics, gait speed, and specific SNPs within the RANKL and VDR loci. A simplified model incorporating only BMD, clinical features, and bone turnover markers preserved high accuracy (*AUC*=0.917), underscoring its feasibility for resource-limited clinical environments. **Conclusion** The deep learning – based multi-dimensional feature fusion model markedly enhances the precision and individualization of osteoporosis assessment compared with traditional tools. By integrating biological, structural, metabolic, genetic, and functional dimensions of bone health, the model provides a comprehensive representation of skeletal integrity and robustly improves both diagnostic accuracy and fracture risk prediction. Its strong generalizability across demographic subgroups highlights its clinical applicability. This work offers a promising direction for developing next-generation intelligent decision-support systems that may meaningfully improve osteoporosis screening, risk stratification, and preventive care.

**Key words** osteoporosis,machine learning,multimodal feature fusion,fracture risk,prediction model

**DOI:** 10.3724/j.pibb.2025.0374

**CSTR:** 32369.14.pibb.20250374

\* This work was supported by grants from the Key Research Project of Natural Science of the Anhui Provincial Department of Education (KJ2021A1564), the Domestic Visiting Scholar Program for Outstanding Young Faculty Members in Universities (gxgnfx2022211) and the Key Training Program for Outstanding Young Teachers of the Anhui Provincial Department of Education (YQZD2024099).

\*\* Corresponding author.

Tel: 86-566-3388850, E-mail: wcy\_2003@163.com

Received: August 9, 2025 Accepted: December 1, 2025