

<研究課題>

AI（人工知能）を用いた、胸部単純X線写真から骨密度を計測するアルゴリズムの開発研究

代表研究者	名古屋大学大学院医学系研究科	整形外科	佐藤洋一
共同研究者	宮本整形外科病院	整形外科	山本乃利男
	東京慈恵医大附属柏病院	整形外科	稲垣直哉
	名古屋医療センター	整形外科	家崎雄介
	加古川医療センター	整形外科	高原俊介

【抄録】

高齢化に伴い、骨粗鬆症患者が世界的に増えている一方で、診断や治療が十分に行えていない。スクリーニングを効率的に行うためには、一般的かつ撮影頻度が高い撮影機器を用いる必要がある。本研究では、深層学習モデルを用いて胸部X線写真から骨密度の予測、骨粗鬆症や骨量減少の予測が可能か検討した。2010年から10年間に6つの病院で骨密度検査と胸部X線写真撮影の両方を受けた患者の10102人、17899枚を含むデータセットを収集した。骨密度検査結果をラベルとして、胸部X線写真・年齢・性別を深層学習モデルにアンサンブル学習させた。深層学習モデルは、骨密度を相関係数 $R=0.75$ の精度で推測した。また骨密度検査結果に基づく診断（正常、骨量減少、骨粗鬆症）の予測精度はArea Under the Curve (AUC)=0.89、0.70、0.84であった。これらの結果から、深層学習モデルが胸部X線写真を用いて骨密度の推測を行うことで、臨床環境における骨粗鬆症患者のスクリーニングができる可能性を示唆された。

1. 研究の目的

本邦では加齢に伴う骨粗鬆症患者数が1300万人に上るといわれている一方、骨粗鬆症検診受診率は約5%に留まる。検診受診率が低い理由について、検診実施施設側の要因としては特殊な検査機器が必要である点被検者側の要因としては受診の動機づけが難しい点、つまり行動変容が必要である点が挙げられる。ここで我々は一般的かつ撮影頻度が高い画像検査を利用し、その画像検査を用いた他の検診と一緒にすることで、骨粗鬆症検診受診率向上に寄与できると考えた。

先行研究において、深層学習モデルを用いて胸部X線写真から骨粗鬆症や骨量減少を診断するといった報告がある。これらのリミテーションとしては連続値である骨密度（Bone Mineral Density; BMD）の回帰問題ではない点、正常・骨量減少・骨粗鬆症といったマルチクラス分類ではない点、いずれも小さいデータセットで学習しており、過学習の可能性があ

る点が挙げられる。そこで本研究の目的は大規模データセットから得られた胸部X線画像・年齢・性別から・大腿骨近位部・腰椎のBMDの予測・T-scoreに基づく診断（正常、骨量減少、骨粗鬆症）の予測を行う深層学習モデルを開発することである。

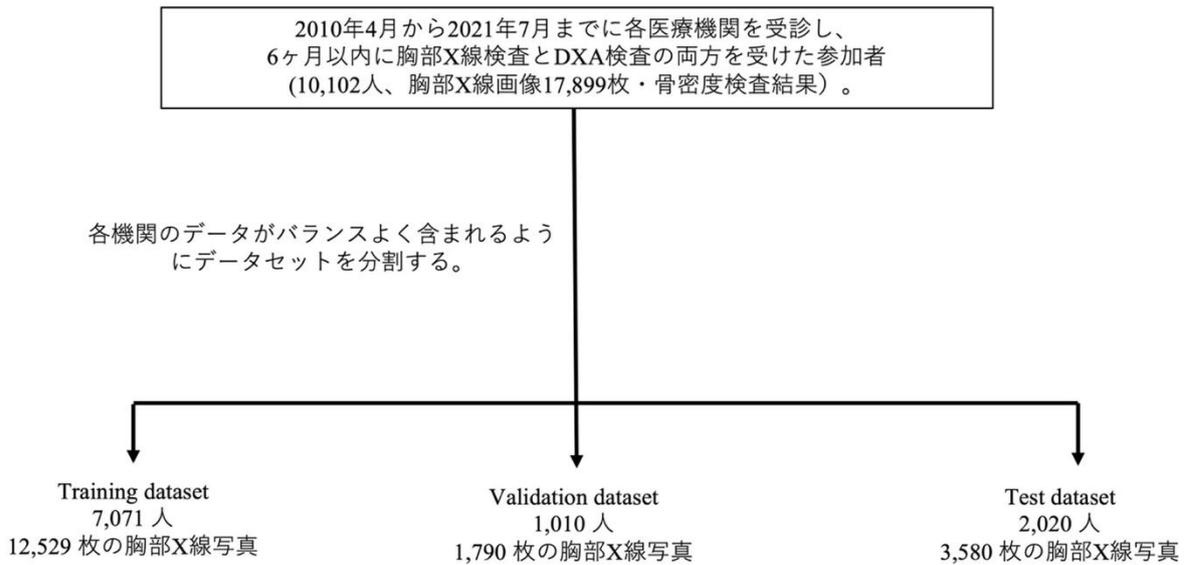
2. 研究方法と経過

2010年から2021年の間で各施設を受診した20歳以上の男女のうち、胸部X線写真を撮影し、骨密度検査をその6ヶ月以内に施行した10102名、17899枚を対象とした。このデータセットをトレーニングデータセット、バリデーションデータセット、テストデータセットに分割した。（図1）学習済み深層学習モデルに対して胸部X線写真、年齢、性別および大腿骨近位部・腰椎のBMD、T-scoreに基づく診断（正常、骨量減少、骨粗鬆症）を学習させた。BMDに関しては回帰問題T-scoreに関してはマルチクラス分類問題として学習させた。検討項目は

深層学習モデルのイメージ



図 1. データセット構成



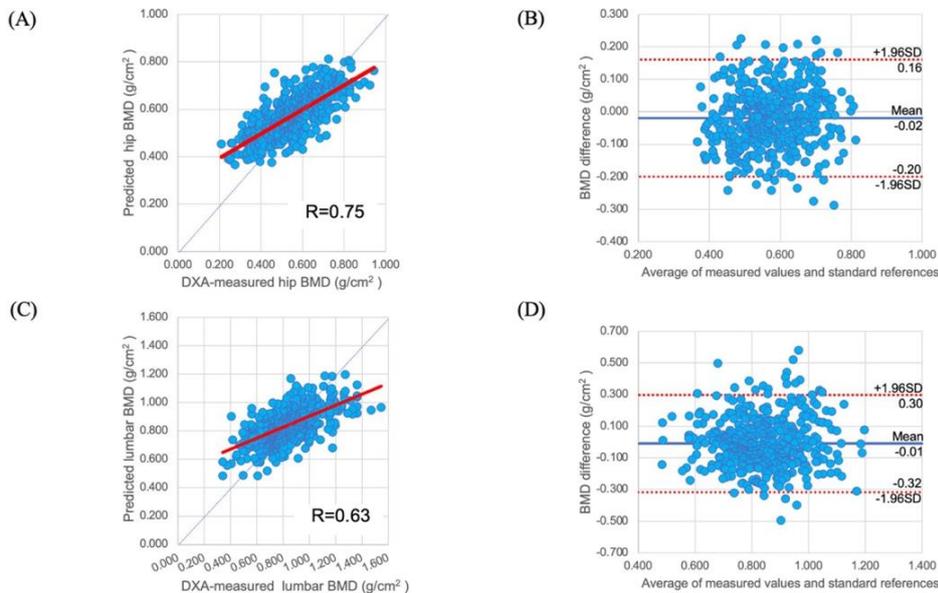
BMD に関して、予測した BMD と真の BMD との相関、T-score に関して、予測した診断と真の診断との正誤とした。

3. 研究の成果

大腿骨に関する、予測された BMD と真の BMD の相関図と Bland-Altman plot を示す (図 2)。相関係数は 0.75 となった。次に、腰

椎に関する、予測された BMD と真の BMD の相関図と Bland-Altman plot を示す。相関係数は 0.63 となった。T-score に関して、予測された診断と真の診断との正誤に関する予測精度を示す (図 3)。T-score -1.0、-2.5 をカットオフとしたマルチクラス分類の AUC はそれぞれ 0.89、0.70、0.84 となった。

図 2. BMD に関する回帰問題の予測精度。



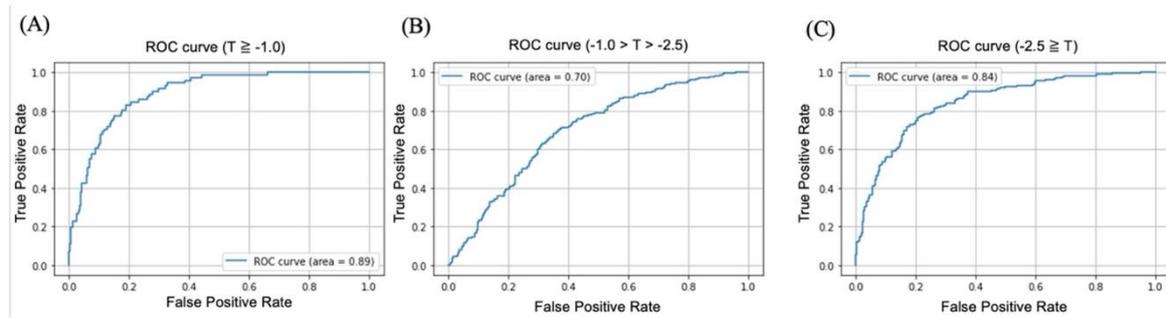
(A)大腿骨 BMD 予測用の学習済みモデルの Linear fitting curve と(B) Bland-Altman plot。

(C)腰椎 BMD 予測用学習済みモデルの Linear fitting curve と(D) Bland-Altman plot。

モデル予測は真の値と比較された。Linear fitting curve において、R はピアソン相関係数である。

Bland-Altman plot の各点は、DXA BMD と予測 BMD のペアを表し、横軸は平均を、縦軸は差を表している。

図 3. T-score に関する分類問題の予測精度



(A) T スコアが-1.0 以上、(B) T スコアが-1.0~-2.5、(C) T スコアが-2.5 以下。

注) ROC: the receiver operating characteristic curve

4. 考察・今後の課題

本研究において、胸部 X 線写真・年齢・性別をアンサンブル学習した深層学習モデルが、BMD の予測や T-score に基づく診断（正常、骨量減少、骨粗鬆症）の予測が行えることを示した。深層学習モデルによる画像認識技術は、胸部 X 線写真での骨粗鬆症スクリーニングに実用可能性があることが実証された。

予測された BMD と真の BMD との相関について考察する。先行研究では胸部 X 線写真での回帰モデルはない点で、本研究の新規性が挙げられる。一方、股関節や腰椎からの BMD 予測モデルはあり、その精度は $R=0.81-0.89$ と本研究はやや劣る結果となった。これは撮影部位、計測部位の位置の差異が関係していると思われる。また、本研究における大腿骨、腰椎の BMD 予測精度に関して腰椎は大腿骨に劣る結果となった。これは腰椎で発生する変形性脊椎症が予測精度に影響していると考えられる。次に、T-score を用いた予測された診断と真の診断との正誤について考察する。先行研究では胸部 X 線写真でのマルチクラス分類モデルはなく、この点で新規性があると考えられる。一方、T-score >2.5 をカットオフとした先行研究と比較してやや精度が劣る結果となった。先行研究では胸部 X 線写真全体ではなく特定の関心領域に絞ると精度が上がると報告しており本研究においても今後、解析方法の刷新が必要と考えられる。最後に実用性を考慮した場合の、検診を想定した精度に関して考察する。骨粗鬆症治療ガイドラインでは検診での判定基準は T-score >1.0 をカットオフとしている。骨粗鬆症スクリーニングに必要な精度は感度 90%以上、特異度 40-60%以上とされている。本研究結果からは感度 90.14%、特異度 72.24%と実用に耐えうると考えられる。一方、今後は外的妥当性の評価も必要である。

5. 研究成果の公表方法

上記に示した課題を含めた研究結果は第 95 回日本整形外科学会学術集会・第 40 回日本骨代謝学会学術集会にて口演を行った。また、海外雑誌へ投稿し、掲載された¹。また第 109 回 Radiological society of North America（北米放射線学会）にて発表予定である。

1. Sato, Y. *et al.* Deep Learning for Bone Mineral Density and T-Score Prediction from Chest X-rays: A Multicenter Study. *Biomedicines* 10, 1–11 (2022).

以上

Deep learning for bone mineral density and T-score prediction from chest X-rays: A multicenter study

Primary Researcher: Yoichi Sato
Nagoya University Graduate School of Medicine

Co-researchers: Norio Yamamoto
Miyamoto Orthopaedic Hospital
Naoya Inagaki
The Jikei University Kashiwa Hospital
Yusuke Iesaki
The National Hospital Organization Nagoya Medical Center
Shunsuke Takahara
Hyogo Prefectural Kakogawa Medical Center

Although the number of patients with osteoporosis is increasing worldwide, diagnosis and treatment are presently inadequate. In this study, we developed a deep learning model to predict the bone mineral density (BMD) and T-score from chest X-rays, which is one of the most common, easily accessible, and low-cost medical imaging examination methods. The dataset used in this study contained 17,899 images corresponding to 10,102 patients who underwent both dual-energy X-ray absorptiometry (DXA) and chest radiography at six hospitals between 2010 and 2021. For the learning labels, we used (1) BMD (g/cm²) of the hip and lumbar spine and (2) diagnosis based on the T-score of the hip or lumbar spine (normal, osteopenia, and osteoporosis). Then, we trained the deep learning model through ensemble learning of chest X-rays, age, and sex to predict the BMD using regression and T-score for multiclass classification. We assessed the following two metrics to evaluate the performance of the deep learning model: (1) correlation between the predicted and true BMDs and (2) consistency in the T-score between the predicted class and true class. The correlation coefficients for BMD prediction were hip = 0.75, lumbar spine = 0.63. The areas under the curves for the T-score predictions of normal, osteopenia, and osteoporosis diagnoses were 0.89, 0.70, and 0.84, respectively. These results suggest that the proposed deep learning model may be suitable for screening patients with osteoporosis by predicting the BMD and T-score from chest X-rays.