

## 人工知能を用いた後縦靭帯骨化症の進行予測とその精度検証

大阪市立大学 整形外科学教室

玉井 孝司

### 1. 諸言

後縦靭帯骨化症 (ossification of posterior longitudinal ligament : OPLL) は、脊椎椎体の後縁を連結し、脊柱のほぼ全長を縦走する後縦靭帯が骨化することにより、脊椎管狭窄を来し、脊髓又は神経根の圧迫障害を来す疾患である。頸椎に最も多く生じるとされており、厚生労働省難病情報センターより難病指定 (指定難病 69) を受けている疾患である。多くの関連原因が指摘されているが、現在のところ主たる骨化の原因は不明とされている。初発症状は頸部痛、上肢のしびれ、痛みで始まることが多い。進行すると下肢のしびれ、痛み、知覚鈍麻、筋力低下、上・下肢の腱反射異常、病的反射などが出現し、痙性麻痺を呈する。手術加療が必要となることが多い。一般的に、骨化が脊柱管前後径の 60% を超えると、ほぼ全例で脊髓障害が出現するため、手術加療が必要となるとされている。近年、OPLL の画像的評価には、診断精度の観点より、単純 X 線写真ではなく、CT の利用が推奨されている<sup>1)</sup>。一方で、CT 撮影による医療被爆も社会問題となっている<sup>2),3)</sup>。われわれは、人工知能による深層学習を用いて、頸椎単純 X 線写真側面像にて OPLL の有無を同定するシステムを開発した。そのため本研究の目的は、同システムによる OPLL 診断能を評価する事である。

### 2. 方法

頸椎単純 X 線写真及び頸椎 CT が同時期に撮影され、CT にて後縦靭帯骨化を認めた症例 243 例 (OPLL 群、平均年齢 63.5±10.1 歳、男性 157 人、女性 86 人) を用いた。また、CT にて OPLL が認められなかった患者のうち年齢・性別をマッチさせた 243 例 (コントロール群、平均年齢 64.9±11.2 歳、男性 153 人、女性 89 人) を抽出した。分類機は、Imagenet で学習済みの EfficientNetB2 のモデルに対し、平均プーリング層、全結合層、ドロップアウト層および出力層を接続して構成し、ファインチューニングによる脊椎周辺の部分画像 (224x224) に対する分類学習を行った。学習時は教師データに対し、ランダムなコントラスト変更、ガンマ変換、回転、mix-up 等手法によるデータ拡張を行った。全症例を用いてアルゴリズムの精度検証を行った。その後、無作為に抽出した 50 例 (OPLL 症例 25 例、

OPLL を認めない症例 25 例) をアルゴリズムに投入し、具体的なアルゴリズムの返答を提示した。最後に、上記 50 例を脊椎外科専門医 4 人にも独立して OPLL の有無を評価させ、OPLL の有無に関するコンセンサス回答を作成した。その上で、コンセンサス回答の正答数とディープラーニングアルゴリズムの正答数を、カイ 2 乗検定を用いて統計学的に比較した。

### 3.結果

#### 3.1 アルゴリズムの診断精度

学習用データ 436 例にて作成した分類器による診断精度は、真陽性 219 例、偽陽性 34 例、真陰性 209 例、偽陰性 24 例であり、正確度 0.88、感度 0.90、特異度 0.86 であった。CT にて評価した骨化巣の有無をアウトカムとした ROC 曲線を描くと、AUC 0.944 (95%信頼区間 [CI] 0.923-0.966) と高い診断能であった。(図 1)

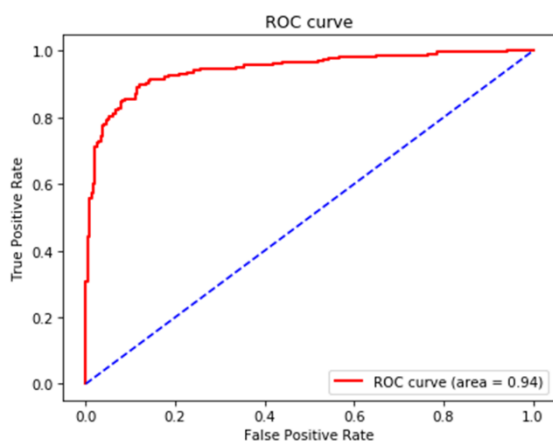


図 1 ROC 曲線

われわれが作成した OPLL を判別するディープラーニングアルゴリズムの診断精度は、AUC 0.944 (95% CI 0.923-0.966)、 $P < 0.001$  であった。

#### 3.2 アルゴリズムが OPLL あり診断した画像

単純 X 線写真データと CT データを基に作成されたアルゴリズムに対して、無作為に抽出した OPLL を有する単純 X 線写真を投入した結果画像を提示する (図 2)。

#### 3.2 アルゴリズムが OPLL なし診断した画像

単純 X 線写真データと CT データを基に作成されたアルゴリズムに対して、無作為に抽出した OPLL を有しない単純 X 線写真を投入した結果画像を提示する (図 3)。

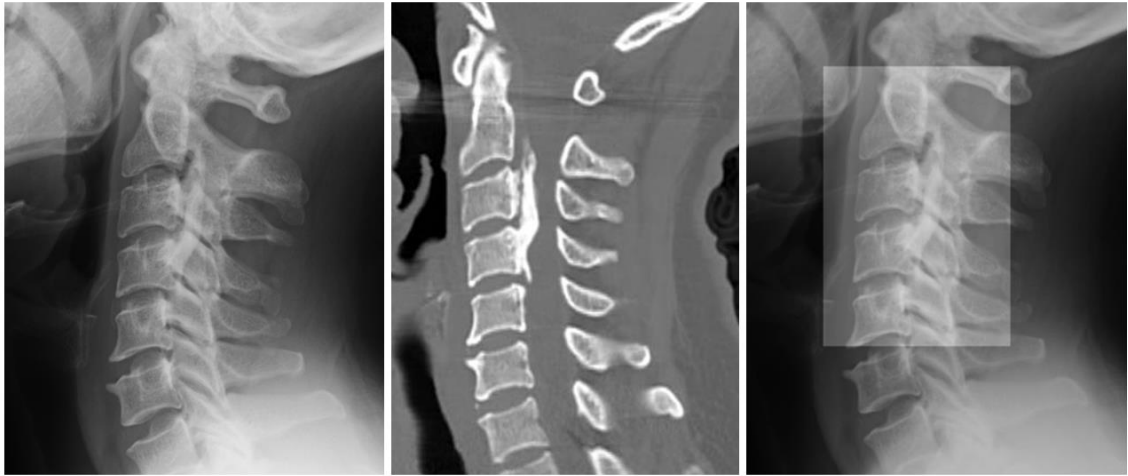


図2：アルゴリズムが OPLL 有り と判定した画像

左図：アルゴリズムに投入された単純 X 線写真

中央図：アルゴリズムには与えられていない CT データ。第 4 頸椎より上方へ延びる粗大な骨化巣を認めることが分かる。

右図：アルゴリズムからの返答。アルゴリズムは投入された単純 X 線写真には OPLL 有り と判定した。また、同アルゴリズムは OPLL がある可能性高い部位を明示するように設計されているが、左図のように CT 画像にて OPLL を認める部位と同部位を示す結果であった。

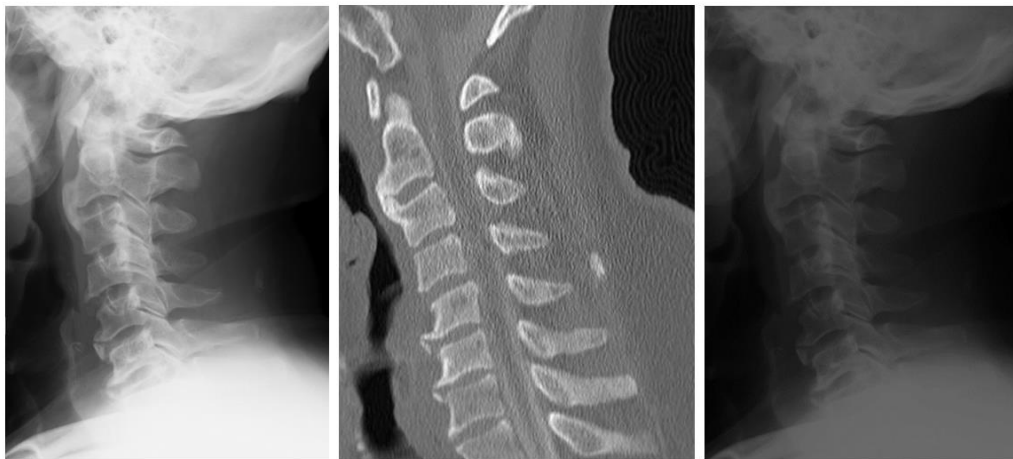


図3：アルゴリズムが OPLL なし と判定した画像

左図：アルゴリズムに投入された単純 X 線写真

中央図：アルゴリズムには与えられていない CT データ。第 2-3 頸椎間には前縦靱帯の骨化による骨性架橋を認めるが、脊柱管内の骨化巣は認めない事が分かる。

右図：アルゴリズムからの返答。アルゴリズムは投入された単純 X 線写真には OPLL なし

と判定した。また、同アルゴリズムは OPLL がある可能性高い部位を明示するように設計されているが、図 2 の左図と違い、どこも明示された領域がないことが分かる。

### 3.3 アルゴリズムと脊椎外科専門医 4 人の正答率

無作為に抽出された 50 枚の画像をディープラーニングアルゴリズムと脊椎外科専門医に読影させた。その結果、アルゴリズムの診断精度は 0.92 と最も高く、脊椎外科医の診断精度は 0.80、0.78、0.76、0.74 であった (表 1)。

|                  | 真陽性 | 偽陽性 | 偽陰性 | 真陰性 | 精度   |
|------------------|-----|-----|-----|-----|------|
| AI アルゴリズム        | 24  | 1   | 2   | 23  | 0.92 |
| 医師 1 (25 年以上の経験) | 22  | 3   | 7   | 18  | 0.80 |
| 医師 2 (20 年以上の経験) | 20  | 5   | 8   | 17  | 0.74 |
| 医師 3 (10 年以上の経験) | 21  | 4   | 8   | 17  | 0.76 |
| 医師 4 (5 年以上の経験)  | 23  | 2   | 9   | 16  | 0.78 |

表 1 : アルゴリズムと脊椎外科専門医の診断精度

また、脊椎外科医 4 人の話し合いにて形成されたコンセンサス回答では正答 50 例中 39 例、誤答 11 例であったのに対して、AI アルゴリズムでは正答 50 例中 47 例、誤答 3 例となりカイ 2 乗検定で  $P=0.041$  と有意にアルゴリズムの正答率が高いことが判明した (表 2)。

|           | 正答 | 誤答 | P-value | 精度   |
|-----------|----|----|---------|------|
| AI アルゴリズム | 47 | 3  | 0.041*  | 0.92 |
| 医師のコンセンサス | 39 | 11 |         | 0.78 |

表 2 : アルゴリズムと脊椎外科専門医コンセンサスの正答・誤答数

## 4. 考察

当研究にて、われわれが作成したアルゴリズムは、経験のある脊椎外科医師が単純 X 線写真を用いた際の診断率よりも、有意に高い診断精度で OPLL を診断が可能なが判明した。

一般的に、ディープラーニングをふくむ人工知能は、人間の機能を超越することは出来ないとされている<sup>4)</sup>。つまり、AI は状況やデータと回答を多量に入力することで作成される。優れた AI システムとは与えられた状況やデータに対して正しい回答を高い正答率で返答するシステムを指す。しかし、AI 作成時に入力される回答は人間が作成したものであるため、回答を作成した人間の能力を超越することは出来ないとされているためである。しかし、

当アルゴリズムは、経験のある脊椎外科医よりも有意に高い診断率であった。これは、アルゴリズム作成時に投入された回答が単純 X 線写真のみではなく CT 画像のデータを用いたためと考えられる。つまり、単純 X 線写真データと CT 画像データを用いて作成されたアルゴリズムの単純 X 線写真のみでの OPLL 診断率は、純 X 線写真のみを用いた脊椎外科医の診断率よりも高い結果であったと考えられた。

AI による診断補助を用いながら医療用画像の診断精度を上げていく方法を総じて「Radiomics」という<sup>5)</sup>。当研究で作成したアルゴリズムも、実際の脊椎外科医とはまったく違う手順で画像を評価し OPLL の有無を判定している。そのため、脊椎外科医が当アルゴリズムを補助として利用することでより高い画像診断精度が得られると考えられる。

## 5.結語

人工知能を応用し作成した画像診断システムによる、頸椎単純 X 線写真を用いた OPLL の診断能は非常に高かった。当システムを併用することで、OPLL 見落としの減少や、CT 撮影機会減少による医療被曝の低減が期待される。

## 6.文献

- 1) Kawaguchi Y, Urushisaki A, Seki S, Hori T, Asanuma Y, Kimura T. Evaluation of ossification of the posterior longitudinal ligament by three-dimensional computed tomography and magnetic resonance imaging. *Spine J* 2011; **11**(10): 927-32.
- 2) Meulepas JM, Ronckers CM, Smets A, et al. Radiation Exposure From Pediatric CT Scans and Subsequent Cancer Risk in the Netherlands. *J Natl Cancer Inst* 2019; **111**(3): 256-63.
- 3) Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *Lancet* 2012; **380**(9840): 499-505.
- 4) Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 2019; **1**(6): e271-e97.
- 5) Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016; **278**(2): 563-77.

## 7.成果発表

### 学会発表

- 玉井孝司、星野雅俊、藪晋人、佐々木遼、田淵仁志、今西 勁峰、ほか. 人工知能技術を用いた画像診断システムによる単純 X 線写真での頸椎後縦靭帯骨化症患者診断能力の検証. 第 50 回日本脊椎脊髄病学会, 京都, 2020