

深層学習を用いた Digitally Reconstructed Radiographs に基づく変形性股関節症の多指標自動分類

梶田 修慎[†] 崇風 まあぜん[†] 大竹 義人[†] 上村 圭亮^{††} 高尾 正樹^{†††}

菅野 伸彦^{††} 佐藤 嘉伸[†]

[†] 奈良先端科学技術大学院大学 〒630-0192 奈良県生駒市高山町 8916-5

^{††} 大阪大学 〒565-0871 大阪府吹田市山田丘 2-2

^{†††} 愛媛大学 〒791-0295 愛媛県東温市志津川 454

E-mail: †{masuda.masachika.mp2,msoufi,otake,yoshi}@is.naist.jp

あらまし 変形性股関節症（HOA）は軟骨の変性により発生し、疼痛及び日常生活動作の低下に繋がる。HOA は通常、X 線画像で診断され、診断の際には軟骨変性の程度に加え、寛骨臼からの大腿骨頭の転位（亜脱臼・脱臼）の程度も同時に評価する事が望まれる。近年、畳み込みニューラルネットワーク（CNN）を用いた自動分類手法が報告されたが、HOA は二値分類問題として扱われ、疾患の進行に伴う臨床的に重要な形状や輝度値の変化を表現することができていない。そこで本研究は、Digitally Reconstructed Radiograph（DRR）における疾患進行情報を含む HOA の自動分類器を開発することを目的とした。新規性は、Crowe 分類（脱臼度）と Kellgren-Lawrence スコア（OA 重症度）の2つの分類指標を同時に分類することにより、臨床的に重要な変動を表現する点である。深層学習に基づいた3つの分類器を評価し、モデルの不確実性を推定するために、テスト時の dropout sampling の影響も評価した。その結果、VisionTransformer は DenseNet と VGG より高い精度が得られた。

キーワード 深層学習, 変形性股関節症, 不確実性

Automatic Multi-Measure Classification of Hip Osteoarthritis Based on Digitally Reconstructed Radiographs using Deep Learning

Masachika MASUDA[†], Mazen SOUFI[†], Yoshito OTAKE[†], Keisuke UEMURA^{††}, Masaki TAKAO^{†††},
Nobuhiko SUGANO^{††}, and Yoshinobu SATO[†]

[†] Nara Institute of Science and Technology 8916-5 Takayama-cho, Ikoma-shi, Nara, 601-0192 Japan

^{††} Osaka University 2-2 Yamadaoka, Suita-shi, 565-0871 Japan

^{†††} Ehime University 454 Shitsukawa, Toon-shi, 791-0295 Japan

E-mail: †{masuda.masachika.mp2,msoufi,otake,yoshi}@is.naist.jp

Abstract Hip Osteoarthritis (HOA) is usually diagnosed by radiographs. In addition to the degree of cartilage degeneration, the degree of dislocation of the femoral head from the acetabulum (subluxation or dislocation) should also be evaluated. Recently, an automatic classification method using Convolutional Neural Networks (CNNs) has been reported; however, HOA diagnosis was addressed as a binary classification problem, therefore it was not able to represent clinically important changes in shape and intensity values with the disease progression. Therefore, the purpose of this study was to develop an automated HOA classification approach that incorporates information on disease progression based on Digitally-Reconstructed Radiographs (DRRs). The novelty is that it simultaneously classifies each DRR into two diagnostic measures, i.e., the Crowe classification (degree of dislocation) and the Kellgren-Lawrence score (OA severity), to represent clinically important changes. Three deep learning-based classifiers, i.e. VGG, DenseNet and ViT, were evaluated. The impact of involving dropout sampling at test-time for model uncertainty estimation was also evaluated. The ViT performance was superior to the other classifiers in terms of the two-measure classification accuracy.

Key words Deep Learning, Hip Osteoarthritis, Uncertainty

1. ま え が き

変形性股関節症 (HOA) は現代の超高齢社会において、重要視されている疾患であり、多くの患者の疼痛や日常生活動作の低下を引き起こしている。HOA は通常、X 線画像で診断され、詳細な疾患進行の評価は整形外科医の臨床専門性を必要とする。そのため、疫学研究において同様の分類を大規模データに対して行う際には、長時間の労力を必要とし、大きな負担になる。近年、画像認識手法の一つである畳み込みニューラルネットワーク (CNN) を用いた自動分類手法が報告されたが[1], HOA は二値分類問題として扱われ、疾患の進行に伴う臨床的に重要な形状や輝度値の変化を評価することができていない。そこで、我々の研究では、脱臼度を示す指標である Crowe 分類と、OA の重要度を示す Kellgren-Lawrence スコア (KL)[2] の 2 つの分類指標を同時に分類することにより、臨床的に重要な変動の表現を試みた。Crowe 分類を組み合わせ、より細分化することは、疾患の進行具合をより正確に把握し、適切な手術計画の立案を可能にする。

2. 手 法

2.1 DRR データセット

本研究では、大阪大学の保有する、股関節周辺の CT 画像から作成された Digitally Reconstructed Radiograph(DRR) の 396 枚 (198 症例) を用いた。これらの画像には整形外科医 2 名による診断に基づいた Crowe と KL のラベルが 1 から 4 のグレードで付与されており、その組み合わせは 7 通り存在する。表 1 に組み合わせの種類とそのデータ数を示す。本来 Crowe 及び KL は 0 から 4 の 5 段階で評価されるが、0(正常)と 1(疑いあり)の境界が曖昧なことから、本研究では除外している。図 1 にラベルが付与された画像例を示す。

表 1 Crowe と KL の組み合わせ及びデータ数

2.2 深層学習モデル

本研究では三つの深層学習モデルを用いて自動分類を検証した。

2.2.1 Vision Transformer Base16

Vision Transformer(ViT)[3] は Dosovitskiy らによって 2020 年に発表された画像認識モデルで、自然言語処理で用いられる Transformer[4] を画像認識に適用させたものである。具体的には、画像を複数のパッチに分割し、それぞれを単語のように線形に変換して Attention 機構によって注目度を算出することで高度な画像解釈を可能にしている。また、大規模な事前学習が

図 1 7 つのクラスの画像例

認識精度を向上させることが分かっており、JFT-300M という 3 億枚のデータセットでの事前学習を行ったモデルは、当時の State-of-the-Art を達成した。

2.2.2 VGG16

VGG[5] は CNN の一種で、 3×3 という小さいカーネルを用いた、16 層及び 19 層の深いネットワークである。畳み込み層とプーリング層、線形層のシンプルなアーキテクチャであるにも関わらず、2014 年の ImageNet チャレンジのローカリゼーション・クラシフィケーションタスクで 1 位と 2 位を収めている。

2.2.3 DenseNet161

DenseNet[6] は複数の Dense Layer と Transition Layer からなるネットワークで、各層の出力がすべての後続の層の入力に使われる特徴がある。これにより、勾配消失問題の緩和、特徴量の伝播の強化、パラメータ数削減などの様々な利点を持つ。

2.3 モデルの学習

モデルの学習は、股関節の DRR(396 枚) による Patient-wise の 4fold 交差検証により行った。また、表 1 から 7 つのクラスの学習データ数には大きな偏りがあることがわかるため、各 fold に同じ比率でクラスを振り分けるために、Scikit-Learn の Stratified-Group-KFold を用いて fold を作成した。各 fold は 3 通り作成し、最後に集約することで総合的に評価した。また、データの分割は全データの 75% を訓練データ、25% をテストデータとし、検証データは、訓練データを Stratified-Group-KFold で 10 分割し、その中の 1fold を用いた。それぞれのモデルの重み初期化には PyTorch の提供するライブラリである torchvision の ImageNet による事前学習済みの重みを用い、ファインチューニングを行った。出力層の活性化関数は softmax とし、損失関数は Focal Loss($\gamma=2$)、パッチサイズは 32、学習率を 5×10^{-5} ～ 3×10^{-5} とし、スケジューラには Cosine Annealing を用いた。

最適化手法には Adam [7] を用い、学習時のデータ拡張は、画像の 224×224 へのリサイズ、ぼかし、ランダムコントラスト変化、ランダムドロップアウトを採用した。また学習の epoch 数は 200 で、学習及び推論時の dropout の確率は ViT が 0.1, VGG は 0.3, DenseNet は 0.2 とした。また、ViT の dropout 層はオリジナルの実装をそのまま使い、VGG は Classifier の dropout 層に加え、特徴マップのサイズが変化する層の ReLU の後ろにも追加した。DenseNet は DenseBlock の間にある Transition 層の後にそれぞれ追加した。そして、最終的な結果は、50 回の推論結果の平均の最大値を分類の推定ラベル、分散を不確実性として出力を得た。

2.4 評価

2.4.1 精度

4fold の交差検証によって分類精度の評価を行った、評価の指標として、Accuracy を用いた。この指標は正例と負例の関係から式 (1) によって計算される。

$$\text{ExactClassAccuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

また、本研究では隣り合うクラス、すなわち一近傍誤差を許容した Exact Class Accuracy も評価の対象とした。以降一近傍精度 (1-Neighbor Class Accuracy) とする。

2.4.2 不確実性

不確実性の評価には、推論時に dropout サンプリングを行うモンテカルロ dropout [8] を用いた。本研究では、不確実性の指標として以下の二つを比較した。(1) 推論時に dropout を行わなかった場合の softmax probability の出力、(2) 推論時の dropout により 50 回のサンプリングを行った出力の分散値。

3. 結果

3.1 分類精度

198 症例の変形性股関節症の分類精度を表 2 に示す。いずれのモデルも一近傍を含めた精度は 0.9 より大きく、特に ViT は、二つの条件において、3つのモデルの中で最も高い精度を得た。

表 2 自動分類結果

手法	Test-time dropout	Exact Class Accuracy \pm SD	1-Neighbor Class Accuracy \pm SD
ViT	×	0.657 \pm 0.136	0.956 \pm 0.141
VGG	×	0.644 \pm 0.167	0.941 \pm 0.172
DenseNet	×	0.629 \pm 0.160	0.945 \pm 0.167
ViT	✓	0.661 \pm 0.133	0.962 \pm 0.139
VGG	✓	0.641 \pm 0.154	0.945 \pm 0.165
DenseNet	✓	0.616 \pm 0.153	0.948 \pm 0.159

3.2 不確実性の評価

test-time dropout なしで推論回数が 1 回の時のボックスプロットを図 2 に、test-time dropout ありで推論を 50 回行った時のボックスプロットを図 3 に示す。なお、各予測クラスのボックスプロットの差が、統計的に意味を持つのかを確かめるために、

2 群の検定手法である、Mann-Whitney U Test を 3 つの組み合わせ、Exact と 1-Neighbor, 1 Neighbor と Others, Exact と Others に対して実施した。その結果、test-time dropout なしでの一回の推論で得られた結果は、ViT 以外は正解、不正解のクラスに有意差があることがわかった。一方、test-time dropout を有効にし、50 回の推論を行った結果では、ViT に有意差があり、最も正解と不正解を確信をもって区別できていることがわかる。

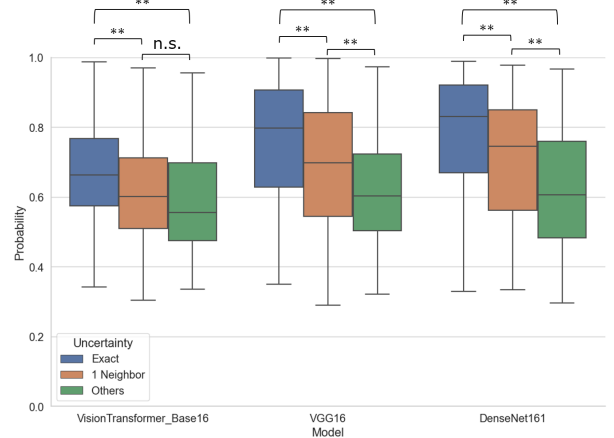


図 2 test-time dropout なしでの不確実性の評価結果 (softmax probability)

n.s. not - significant, ** $p < .01$

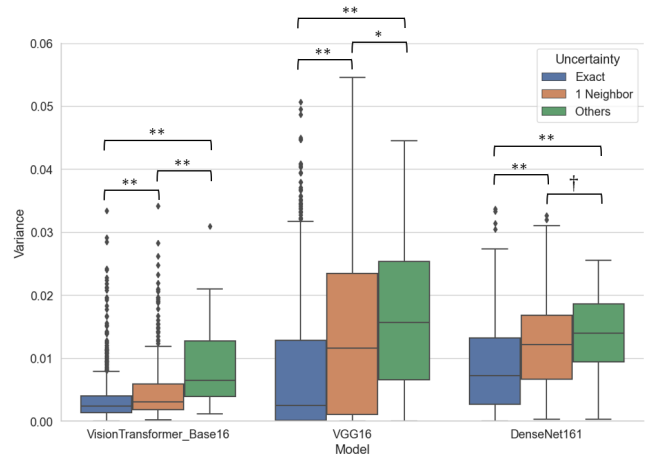


図 3 test-time dropout ありでの不確実性の評価結果 (softmax probability の分散)

† $p < .1$, * $p < .05$, ** $p < .01$

4. 考察

本実験では、従来の研究の二値分類とは違い、Crowe 指標を組み合わせて、さらにグレードを細分化した 7 クラスでの分類を行った。その結果、最も精度が高かったモデルである Vision Transformer_Base16 は、test-time dropout を有効にした 50 回のバイズ推論の結果、Exact Class Accuracy が 0.661 ± 0.133 , 1-Neighbor Class Accuracy は 0.962 ± 0.139 であった。そして、不確実性の評価から、不正解を最も分離できており、毎回の推

論の予測のばらつきが小さいのも ViT である。また、50 回の推論を行い、Softmax の平均の最大値を推定ラベルとした場合の外れ値として検出された、図 4 に示す二枚の画像は、Crowe や KL の値が大きく、我々の目視では比較的容易に判別できるものの、モデルは正確に予測できていない。特に、脱臼の程度が酷い場合は、臼蓋から離れた位置にある大腿骨頭の位置を適切に認識する必要があるが、予測に失敗していることから、Multi-Head-Attention により大域的な情報を参照することが可能 [9] な VisionTransformer の性能が十分に発揮できていないと考えられる。そしてこの問題は、サンプル数の少ないクラスの十分な学習データや、その他の症例の大量のデータが集まれば解消すると考えられる。本実験よりも大きなデータベースでの検証を進め、汎化性能の向上を目指したい。

図 4 予測結果が実際の分類と 2 クラス以上異なった画像の例

5. 結 論

本研究では、変形性股関節症のより詳細な進行具合を把握することを目的に、Crowe と KL の多クラス分類を行った。その結果、三つの深層学習モデルで Exact Accuracy が 60% 以上、1-Neighbor Class Accuracy が 90% 以上の分類精度が得られた。本研究で、198 症例 (396 枚) の画像から、疾患を高精度で予測し、バイズ推論を用いた不確実性の評価に成功したことは、今後、アノテーションが施されていない大規模な CT のデータベースに対するアクティブラーニングが可能なことを示している。予測に自信を持っていない症例を外れ値として、専門医にアノテーションを付与してもらい、学習データを増やして、再度学習する。このサイクルを繰り返すことで、アノテーションが施されていない大規模なデータベースのアノテーションに携わる医者の負担を減らすことができ、効率よく信頼性の高いデータベースを構築することが可能になる [10] と考えられる。

文 献

- [1] Robel K. Gebre, Jukka Hirvasniemi, Rianne A van der Heijden, Iikka Lantto, Simo S Saarakkala, J. Leppilahti, and Timo Jämsä. Detecting hip osteoarthritis on clinical ct: a deep learning application based on 2-d summation images derived from ct. *Osteoporosis International*, Vol. 33, pp. 355 – 365, 2021.
- [2] Jonas H Kellgren and JS1006995 Lawrence. Radiological assessment of osteo-arthritis. *Annals of the rheumatic diseases*, Vol. 16, No. 4,

p. 494, 1957.

- [3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, Vol. 30, , 2017.
- [5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [6] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
- [7] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [8] Yarín Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, 2015.
- [9] Jean-Baptiste Cordonnier, Andreas Loukas, and Martin Jaggi. On the relationship between self-attention and convolutional layers. *arXiv preprint arXiv:1911.03584*, 2019.
- [10] Yarín Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML'17*, p. 1183–1192. JMLR.org, 2017.