

学位論文内容の要旨

愛知学院大学

論文提出者

西山 雅子

論文題目

異なる2施設のパノラマX線画像によって構築した関節突起骨折診断のための深層学習モデル

I. 緒言

近年、医療分野において畳込みニューラルネットワーク (CNN) を用いた深層学習 (DL : deep learning) の応用は高い関心を集めている。DL のなかでも画像分類と物体検出の技術は画像診断の分野で多く用いられており、これらの技術に基づく診断支援 (CAD) システムの有用性が報告されている。しかし、DL を関節突起骨折に適用した報告はこれまでにない。

パノラマ X 線画像を用いて行われた多くの研究では数百から数千枚程度の訓練用画像を用いて DL モデルが作成されているが、ほとんどの研究で訓練用画像およびテスト画像を同一施設から収集しており、高い診断能が報告されている。一方で、訓練用画像とは異なる施設のテスト画像を使用した外部検証に関する報告は未だ十分ではない。多くの場合、外部データを使用した検証では診断能は低下すると考えられている。DL モデルによる診断支援を実際の臨床において行うことを想定した場合、DL モデルの汎用性に関する外部検証の問題は重要な課題である。しかし、パノラマ X 線画像に関する外部検証の報告は我々の知る限りまだなされていない。

以上より、本研究の目的は 2 施設から得られたパノラマ X 線画像を用いて下顎骨関節突起骨折の診断のための DL システムを構築し、特に外部検証と内部検証の比較を中心に診断能を検証することとした。

II. 方法

本研究はヘルシンキ宣言に則り、歯学部倫理委員会の承認 (No. 586) および大垣市民病院倫理委員会の承認 (No. 20200423-13) を得て行った。

1. 対象

対象は愛知学院大学歯学部附属病院 (病院 A) と大垣市民病院 (病院 B) のデータベースから抽出した。CT 検査にて関節突起骨折を認めたものを骨折群とした。骨折なし群は CT 検査にて関節突起骨折を認めず、下顎頭に骨変形がないものとした。いずれの病院においても骨折群 100 顎関節、骨折なし群 100 顎関節のパノラマ X 線画像をそれぞれ抽出した。

2. 画像パッチの準備

パノラマ X 線画像から顎関節部を任意の大きさの矩形領域で切り出した。切り出し範囲の上端は関節窩の最深部に、下端は咬合平面に設定したが、無歯顎の場合は下顎枝の 1/2 の高さに設定した。近心端は下顎切痕、遠心端は関節窩の後端に一致させ、矩形領域の設定を行った。

3. DL システムの構築

今回使用した DL システムは Ubuntu OS ver. 16.04.2 を動作システムとし、11GB の GPU および 128GB のメモリを有するワークステーションに構築された。ネットワークは AlexNet および VGG-16 を使用した。

4. DL モデルの作成と評価

5 分割交差検証の手法を応用して DL モデルを作成し、評価した。DL モデルの作成にはそれぞれ 200 エポックの学習を実施した。

病院 A の画像を 5 グループに分割し、それぞれのグループが骨折群 20、骨折なし群 20 の計 400 画像パッチになるように無作為に割付けた。4 グループを訓練・検証用データとし、1 グループをテストデータとして診断能を検証した。この過程を、訓練データとテストデータを入れ替えながら 5 回繰り返した。作成されたモデルを ModelA1~5 とした。病院 B のデータで作成されたモデル (ModelB1~5) でも同様に行い、内部検証とした。

外部検証は、テストデータとして訓練用データとは異なる病院の画像で作成されたデータセットを使用した。

また、2 施設の訓練用画像を統合した DL モデルの作成を行った (ModelAB)。このモデルの検証にもテストデータとして病院 A、病院 B のそれぞれの画像パッチで作成されたデータセットを使用した。

診断能の評価には 200 のテストデータの結果から DL モデルごとの ROC 曲線を描記し、ROC 曲線下の面積 (AUC) を算出し結果を比較した。また ROC 曲線のグラフで座標の左上に最も近い点をカットオフ値として、感度、特異度、正診率を求めた。

5. 統計解析

2 病院間および骨折群と骨折なし群の間の年齢差の検定については t 検定を用いた。AUC、男女比および骨折の部位および骨折の程度の分布の差の検定は χ^2 検定を用いた。 $p < 0.05$ の場合に有意差ありとした。

III. 結果

病院 A と病院 B の骨折群と骨折なし群の男女比に有意差はなかった。

骨折の部位 ($p=0.026$) および骨折の程度の分布 ($p=0.041$) では病院 A と病院 B の間で有意差を認めた。病院 A では上頸部骨折が多く、病院 B では下頸部骨折が多くみられた。また、病院 A では偏位や転位といった脱臼を伴わない症例が多かったが、病院 B では過半数が脱臼を伴う骨折であった。

AlexNet では、テストデータと訓練・検証用データとが同施設の場合 (内部検証)、AUC は Model A で 0.85、Model B では 0.86 であり、高い診断能を示した。一方で異なる施設の画像をテストデータとした場合 (外部検証) では AUC は Model A、Model B ともに 0.58 であり、有意に低い値を示した。Model AB では内部検証の結果と比べ、有意差はないもののわずかに高い値を示した。

VGG-16 では、内部検証では、Model A の AUC は 0.88、Model B では 0.98 とさらに高い診断能を示した。一方、外部検証では AUC が 0.51、0.60 と内部検証の場合に比較して有意に低い値を示した。Model AB の AUC は 0.91、0.98 であった。Model A および Model B の内部検証と比較して同等もしくはわずかに高い値を示したが有意差は認められなかった。

AlexNet と VGG-16 を比較すると、いずれのネットワークにおいても内部検証では診断能は高く、外部検証では低くなるという同様の傾向がみられ

た。また、VGG-16 では Model B および Model AB にテストデータとして病院 B の画像を使用すると、特に高い診断能を示した。両モデルとも AUC は 0.98、正診率は 94.5% であった。AUC は、AlexNet による Model B および Model AB と比較して有意差を認めた。

外部データで作成された DL モデルで関節突起骨折を検出できなかった症例は AlexNet を使用した場合、病院 A の症例で 24 例、病院 B の症例で 25 例であった。VGG-16 では病院 A の症例で 34 例、病院 B の症例では 31 例で骨折を検出できなかった。

IV. 考察

訓練・検証用データとテストデータが同施設から収集された場合（内部検証）の結果では、AlexNet、VGG-16 のいずれのネットワークでも AUC は 0.85 を超えていた。この結果は過去の報告と比較して、遜色ない結果であった。このことから、内部検証に関しては関節突起骨折の診断において DL システムは十分な診断能を有していると考えられた。

今回の研究において Model AB の結果は単独の病院のデータで作成したモデルの内部検証と比較して、同等もしくはわずかに高い値を示したが有意差は見られなかった。また訓練・検証用データに他施設の画像を含めても診断能の低下は見られなかった。他施設で作成された DL モデルを利用する際に自施設のデータセットを用いて再学習させるなどの調整を行うことで

診断能を向上させる可能性があることを示唆している。

本研究では、外部の画像をテストデータとした場合（外部検証）の診断能は内部検証と比較して有意に低いという結果を示した。これはドメインシフトと呼ばれる現象に起因していると考えられる。外部モデルを使用する場合にはこの問題を考慮する必要がある。

次に使用した AlexNet、VGG-16 の2つのネットワークの比較をすると、いずれのネットワークにおいても内部検証と比較して外部検証では診断能が有意に低下し、類似した傾向がみられた。しかし、VGG-16 では AlexNet を使用した同様の DL モデルと比較して有意に高い AUC を示した。今回使用した VGG-16 は 100 万枚を超える自然画像で事前に学習を行った状態で公開されているものを使用した。事前学習済みモデルでは訓練用データが少ない場合でも効率的に学習を行うことができる利点があるとされている。一方で、ModelA の外部検証では AlexNet と VGG-16 で同等の低い結果を示しており、必ずしも事前学習済みモデルの結果が高いとは言えない。事前学習済みモデルの使用に関してはさらなる検証が必要であると考えられる。

2 施設の診断能を比較するにあたって、施設間のデータの特徴に違いがあることも考慮する必要がある。骨折の部位および骨折の程度について、2 病院間に有意差を認めた。DL システムではどのような特徴を学習して診断を行なっているか一般的には明らかにされることはない、しかし、訓練データの症例分布に違いがあることから診断能に差があることが予想され

た。関節突起骨折には様々なバリエーションがあり、少ない症例数で学習を行うことは注意が必要である。また、ネットワークによって診断の傾向は類似しているものの、診断に使用した特徴的な所見は異なっている可能性も考えられる。

V. 結論

パノラマ X 線画像を用いた関節突起骨折診断のための DL システムの有用性が示唆された。しかし、訓練用画像を提供した施設とは異なる施設で DL モデルを使用する場合には診断能が低下することが明らかとなった。外部の施設で作成された DL システムを用いて診断支援を行う場合には注意が必要であることが示唆された。