

学位論文の全文に代えてその内容を要約したもの

愛知学院大学

乙 第 号	論文提出者 桑名良輔
論文題目 パノラマ X 線画像による上顎洞病変自動診断のための 深層学習システム	

I. 緒言

近年、畳み込みニューラルネットワークを用いた深層学習 (Deep learning:DL) アルゴリズムを活用したコンピュータ支援診断/検出 (CAD) システムが様々な医用画像に適用されている。CAD システムの重要な役割として、臨床で数多くの画像を読影する必要のある歯科放射線科医の負担を軽減することや、経験の浅い読影者が重要な病変を見落とさないようにサポートすることがあげられる。

本研究の目的は、パノラマエックス線画像における「片側顎裂」および「両側顎裂」の検出に有用な DL モデル (研究①) と、顎裂 (片側及び両側) を有する患者の「口蓋裂の有無」を診断するために有用な DL モデル (研究②) を作成し、それらの性能を観察者の診断能と比較することである。

II. 研究①：片側または両側顎裂の自動検出

本研究は、愛知学院大学歯学部倫理委員会の承認 (承認番号：496、577) を得て、ヘルシンキ宣言に則って行われた。

1. 対象

2004年8月から2020年7月の間に愛知学院大学歯学部附属病院においてパノラマエックス線撮影を行った顎裂を有する患者 446 例 (女性 192 例、男性 254 例、平均年齢 8.8 歳) で、そのうち片側顎裂患者が 353 例 (女性 153 例、男性 200 例)、両側顎裂患者が 93 例 (女性 39 例、男性 54 例) であった。顎裂の有無は診療録の記載を参考に CT または CBCT を基にして決定した。また、顎裂のない対照群は顎裂患者の平均年齢と性別の分布に一致する 210 名とした。

2. 方法

1) 深層学習システムの構成

深層学習システムは、Ubuntu16.04.2 の OS、11GB の GPU (NVIDIA GeForce GTX 1080 Ti; NVIDIA, Santa Clara, CA, USA) を使用し、トレーニングシステムとして Digits version 5.0 training system (NVIDIA) を用いて構築された。CNN は DIGITS Library に提供されている DetectNet をカスタマイズし Caffe のフレームワークで使用した。DetectNet は「物体検出」と、「検出した領域における分類 (診断)」が可能である。最適化の手法として ADAM (Adaptive Moment Estimation) solver を用い、学習率は 0.0001 とした。

2) モデルの作成

学習に用いる「データの種類」と「データの量」の違いが、作成したモ

デルの性能に与える影響を評価するため、4つのモデルを作成した。モデルUは「片側顎裂群」と「対照群」を、モデルBは「両側顎裂群」と「対照群」を用いて作成した。また、モデルC1とC2は、「片側顎裂群」、「両側顎裂群」、「対照群」を用いて作成した。モデルU、B、C1の学習には同じ数の顎裂 ($n=184$) を使用したが、モデルC2はより多くの顎裂のデータ ($n=479$) を用いて作成した。すなわち、「データの種類」の違いによる性能の評価はモデルU、B及びC1を用いて行い、「データの量」の違いによる性能の評価はモデルC1とC2を用いて行った。

学習過程には、訓練データと検証データが使用されるが、これらのデータそれぞれについてラベルが必要となる。ここでラベルとはパノラマエックス線画像における顎裂部の regions of interest (ROI) の座標をテキスト形式に変換したものである。900×900ピクセルにトリミングしたパノラマエックス線画像の左上を原点として、顎裂部に設定した長方形のROIの左上 (x_1, y_1) と右下 (x_2, y_2) の座標を記録し、テキスト形式に変換した。対照群では座標のないラベルを作成した。

3) 検出能の評価方法

対照群を含む75例(顎裂数=60)の同一テストデータを各モデルに適用した。学習モデルの検出能の評価として「再現率」、「適合率」、「F1値」を用いた。

さらに、モデルの性能を観察者と比較するため、1名の歯科放射線科医と1名の臨床研修歯科医がモデルの評価に用いたテストデータにおける顎裂の有無を評価した。作成した4つのモデルと観察者における「検出した顎裂」と「未検出の顎裂」の割合の比較は、 χ^2 検定を用いて行い、危険率0.05以下を有意差ありとした。

3. 結果

モデルU、B、C1、C2において、再現率はそれぞれ0.60、0.73、0.80、0.88であった。モデルUとモデルC1 ($p=0.01$)、モデルUとモデルC2 ($p<0.001$)、モデルBとモデルC2 ($p=0.036$)の結果において、顎裂の「検出数」と「未検出数」の割合に有意差が認められた。しかし、モデルBとモデルC1、モデルC1とモデルC2の間に有意差は見られなかった。観察者では、顎裂の再現率は歯科放射線科医が0.93、臨床研修歯科医が0.83であった。モデルC2の性能は4つのモデルの中で最も高く、歯科放射線科医と同等であった。

4. 考察

同一の学習データ量で作成したモデルU、モデルB及びモデルC1において、モデルC1(0.80)がモデルU(0.60)、モデルB(0.73)と比較して高

い再現率を示した。これらの結果から、パノラマエックス画像上の顎裂を検出するために有効な DL モデルを作成するには、「片側顎裂群」と「両側顎裂群」の両方のデータが必要であることがわかった。

また、モデルの作成に使用するデータ数を変えて作成したモデル C1 と C2 を比較すると、モデル C2 では未検出の顎裂が少なく、再現率は 0.88 であった。この結果から、訓練データと検証データの数を増やすことで、モデルの性能が向上することが確認された。

Ⅲ. 研究②：顎裂患者における口蓋裂の自動診断

1. 対象

2004年8月から2020年7月の間に愛知学院大学歯学部附属病院においてパノラマエックス線撮影を行った顎裂を有する491例（女性214例、男性277例、平均年齢8.8歳）で、そのうち口蓋裂を有する患者が299例（片側顎裂209例、両側顎裂90例）、口蓋裂のない患者が192例（片側顎裂174例、両側顎裂18例）であった。診療録の記載を参考にして、CTまたはCBCTによって「口蓋裂あり群」と「口蓋裂なし群」に分類した。パノラマエックス線画像は900×900ピクセルにトリミングした。

2. 方法

1) 深層学習システムの構成

研究①で用いた同一の深層学習システムを用い、ネットワークは DetectNet と VGG-16 を用いた。VGG-16 は自動分類（診断）が可能なネットワークである。

2) 学習モデルの作成

研究②では、「DetectNet を用いて作成したモデル D」と「VGG-16 を用いて作成したモデル V」の2つの学習モデルを作成した。モデル D はパノラマエックス線画像における上顎前歯部領域を自動検出し、その領域における口蓋裂の有無を自動分類（診断）するモデルであり、モデル V はパノラマエックス線画像から直接口蓋裂の有無を自動分類（診断）するモデルである。モデル D とモデル V の学習で用いたデータセットは同一であり、訓練データ、検証データ、テストデータはそれぞれ口蓋裂あり群で 215 例、54 例、30 例、口蓋裂なし群で 129 例、33 例、30 例であった。

① モデル D

学習過程に用いるラベルの作成では、上顎前歯部領域に長方形の ROI を設定した。研究①で作成したモデルは、DetectNet の「自動検出」の機能のみを用いて作成したが、モデル D は「自動検出」に加え「自動分類」の機

能を用いた。したがって、ラベルには ROI の座標に加えて、「口蓋裂あり」または「口蓋裂なし」のクラス分類を含むこととなった。推論過程によって検出された領域はパノラマエックス線画像上に長方形のボックスとして描出されるが、「口蓋裂あり」と推論した場合には青色のボックス、「口蓋裂なし」と推論した場合には赤色のボックスが表示された。

② モデル V

学習に用いた 344 枚の訓練画像の鮮鋭度、明るさ、コントラストを調整し、約 2600 枚の画像にデータ拡張した。学習過程は 150 エポックとした。作成された学習モデルにテストデータを適用すると、1 枚の画像ごとの ground truth に対する予測値が提示され、自動分類が行われた。

3) 学習モデルの評価方法

作成したモデル D とモデル V に同一のテストデータを適用し性能を評価した。モデル D の検出能は研究①と同様に「再現率」、「適合率」、「F1 値」を用いて評価した。また、モデル D とモデル V の分類能は「口蓋裂あり」を陽性とし、それぞれ感度、特異度、正診率および ROC 曲線の面積 (AUC) を算出し評価した。

さらに、モデルの性能を観察者と比較するため、2 人の歯科放射線科医がモデルの評価に用いたテストデータにおける口蓋裂の有無を評価した。2 つのモデルと、2 人の歯科放射線科医の AUC を χ^2 検定で評価した。危険率 0.05 以下を有意差ありとした。

3. 結果

モデル D は 60 枚のテスト画像全てにおいて上顎前歯部を正しく検出したため、2 つのモデルと観察者において同一の画像を用いて分類能を評価し、その AUC を統計的に比較することができた。AUC はモデル D とモデル V でそれぞれ 0.95 と 0.93 であり、両モデル間に有意差は見られなかった。観察者の診断能は比較的 low (AUC=0.70、0.63)、2 人の歯科放射線科医間に有意差は認められなかったが、歯科放射線科医と両モデルの間には有意差が認められた。

4. 考察

モデル D とモデル V の分類能 (AUC) は歯科放射線科医と比較して有意に高い値を示した。モデル D の検出能においては、上顎前歯部のような解剖学的部位は、パノラマエックス線画像で ROI の座標がすべての症例において類似しているため、比較的容易に学習することができたと考えられる。

また、自動分類においては、一般的に学習に用いる画像を小さくすることでモデルの精度が向上するとされているが、本研究では臨床的有用性を

考慮し、パノラマエックス線画像を小さくトリミングせず適用し口蓋裂有無の自動分類が可能なモデルVを作成した。その結果、モデルV (AUC:0.93) はモデルD (AUC:0.95) と同等の十分高い性能を示した。この結果の理由として、口蓋裂の有無はパノラマエックス線画像で比較的広い範囲に影響を及ぼす可能性があり、口蓋裂の有無による所見の違いが広範囲に及ぶ可能性があることを示唆していると考えられる。

V. 結 論

片側または両側顎裂の自動検出 (研究①)

片側、両側顎裂の両方を検出するのに有効なモデルを作成するためには、「片側顎裂」と「両側顎裂」の両データが必要であることが明らかとなった。さらに、学習に使用するデータ量を増やすことで、モデルの性能が向上することがわかった。

顎裂患者における口蓋裂の自動診断 (研究②)

本研究で構築したモデルDとモデルVは、パノラマエックス線画像における口蓋裂の有無の診断に有用であることがわかった。