

主論文の要旨

**Age estimates from brain magnetic resonance images
of children younger than two years of age
using deep learning**

〔 深層学習を用いた2歳以下の小児頭部MRIの年齢推定 〕

名古屋大学大学院医学系研究科 総合医学専攻
発育・加齢医学講座 小児科学分野

(指導：高橋 義行 教授)

川口 将宏

【緒言】

2歳以下の小児において頭部MRIから年齢を推定することは、脳の発達を評価し中枢神経疾患を鑑別する上で重要である。実臨床においては、主に髄鞘化の進展に伴う画像変化をもとに年齢推定を行うが、その評価には熟練を要する。

近年、機械学習を用いて作成した人工知能(AI)モデルによる画像解析が普及し、医療画像にも応用が広がっている。機械学習の一種である深層学習は多層のニューラルネットワークを用いる手法であり、その一種である畳み込みニューラルネットワークは画像判別能力に優れ、多くの研究で用いられている。

AIモデルによる頭部MRIの年齢推定は機械学習の発展に伴い精度が向上しているが、多くの研究は成人を対象としており小児、特に乳児の報告は少ない。また、年齢推定のための特徴量として白質と灰白質のコントラストや各脳領域の体積など様々なデータが検討されているが、これらのデータを得るためには特殊なMRI撮像法や解析ソフトを用いた処理が必要になることが多い。

我々の研究の目的は、シンプルな深層学習を用いて作成したAIモデルで生後0~24か月の小児の頭部MRIの年齢を推定することである。加えて、作成したAIモデルを用いて医師との精度比較や髄鞘化遅延を認める画像の年齢推定を行なった。

【対象と方法】

dataset 1 (正常画像)

2016年4月から2019年3月に名古屋大学医学部附属病院小児科で撮像した生後0~24か月児の頭部MRI画像のうち、使用するスライスに異常所見のない画像を対象とした。在胎週数37週未満の早産児と、強いアーチファクトを認める症例は除外し、最終的に161症例、述べ187回の撮像の画像を用いた。

dataset 2 (髄鞘化遅延症例)

名古屋大学医学部附属病院小児科で撮像した生後0~24か月児の頭部MRI画像のうち、明らかな髄鞘化遅延を認めると判断された画像を対象とした。在胎週数37週未満の早産児と、髄鞘化遅延以外に粗大な異常所見を認める例や強いアーチファクトを認める症例は除外し、最終的に6症例、述べ7回の撮像の画像を用いた。

MRI画像の前処理

髄鞘化の目安となる脳梁膨大部レベルの体軸断T1およびT2強調画像を1枚ずつ選択した。画像形式をDICOMからJPEGへ変換し、次に画像の余白を正方形にトリミングし、最後に画像サイズを128×128pixelに統一した。dataset 1および2に含まれるすべての画像を同じ手順で処理した。

深層学習 (AIモデルの作成)

neural network console® (Sony Network communications Inc.)を用いて、4層の畳み込み層を持つ畳み込みニューラルネットワーク(図1)の学習を行った。学習ラベルとして撮像時の週齢を用いた。dataset 1の画像を、年齢を層別化した上でランダムに5等分し、trainingに60%、validationとtestに20%ずつを用いて4回のクロスバリデーショ

ンを行った。

評価と統計解析

AIモデルによる test データセットの年齢推定値に対し、平均絶対誤差 (MAE)、二乗平均平方根誤差 (RMSE)、実年齢と推定年齢のピアソン相関係数を算出した。小児神経医師 3 人による同じ画像の年齢推定値 (月齢) の平均値を医師の結果として比較対象に用いた。AI モデルの推定値を週齢から月齢に変換した後、AI モデルと医師の推定値に対し Wilcoxon の符号付き順位和検定および Bland-Altman 解析で統計学的検討を行った。

dataset 2 の年齢推定

作成した AI モデルを用いて、dataset 2 の画像の年齢推定を行った。AI モデルの推定値と実年齢に対して Wilcoxon の符号付き順位和検定を用いて統計学的検討を行った。

【結果】

dataset 1

dataset 1 に含まれた 187 回の撮像の画像のうち、性別は男児が 58%、年齢の中央値は 7 か月であった。新生児集中治療室での撮像件数が多かったため、生後 0 か月児の画像が 55 回と最多であった。

深層学習

学習過程で過学習は認めず、問題なく学習が実行された。

年齢推定

AI モデルと医師の推定値の散布図を図 2 に示す。推定値の指標は、AI モデルの MAE は 1.9 か月、RMSE は 2.9 か月、医師の MAE は 1.1 か月、RMSE は 1.5 か月であった。実年齢と推定年齢のピアソン相関係数は、AI モデルで 0.94、医師平均で 0.98 であり、両者ともに高い相関が見られた。統計検定では、AI モデルと医師の推定値の間で、Bland-Altman 解析での一致は認めなかったが、Wilcoxon の符号付順位和検定での有意差も認めなかった ($p=0.23$)。AI モデルの推定結果のうち、誤差の値が全体の ± 2 標準偏差を超えるものを外れ値と定義すると、2 例が外れ値と判定された。一例は実年齢 11 か月に対し推定年齢は 17.7 か月であり、別の一例は実年齢 24 か月に対し推定年齢は 13.8 か月であった。これら 2 画像について視察的な観察および撮像パラメータを他の画像と比較検討したが、誤差の理由は明らかではなかった。

dataset 2 の年齢推定

dataset 2 の全画像の診断や年齢推定値を表 1 に示す。AI モデルは dataset 2 の延べ 7 例中 6 例を実年齢より若く推定し、その推定値と実年齢の間に有意差を認めた ($p=0.028$)。

【考察】

本研究の AI モデルは、MAE 1.9 か月の誤差をもって 2 歳以下の小児の頭部 MRI の年齢を推定した。医師の推定結果と比較し、一致は認めなかったが有意差も認めず、AI モデルは医師と似た精度で年齢推定ができたと考えられた。また、AI モデルは髓

鞘化遅延症例を実年齢より有意に若く推定したこととから、本モデルは MRI 読影の際に髄鞘化遅延を見出す助けになりうると考えられた。

本研究の深層学習は一般的に撮像されている T1 および T2 強調画像を用いており、複雑な前処理も要しないためシンプルで簡便な手法である。学習に用いる特徴量が少ないと精度向上が難しい問題はあるが、特別な撮像法や処理を必要としないことは様々な環境で実装しやすい利点がある。本研究ではシンプルな学習方法を採用し、その範囲で十分な精度を得たと考えた。

外れ値は AI モデルの精度を下げる主な要因である。本研究では 2 例の外れ値を認めた。視察的には誤差が大きい理由は明らかではなかったが、学習データに含まれる症例が生後 0 か月に偏っており月齢の高い画像の学習が不十分であった可能性や、成長に伴い髄鞘化が完成に近づくとも月齢ごとの画像変化が小さくなり区別しづらくなるのが理由として考えられた。学習データの偏りや症例数の不足を改善することで外れ値を減らせる可能性があると考えられる。

【結論】

シンプルな方法で学習を行った本研究の AI モデルは、2 歳以下の頭部 MRI に対して、専門家に近い精度で年齢推定を行うことができた。また、本研究の AI モデルは髄鞘化遅延を見出すうえでの助けとなりうると思った。