

研究テーマ

受傷時頸椎 X 線写真で外傷性頸部症候群「慢性化」を予測する人工知能開発

研究代表者

施設名 : 大阪公立大学医学部医学科
整形外科学教室

氏名 : 玉井 孝司

研究テーマ：受傷時頸椎X線写真で外傷性頸部症候群「慢性化」を予測する人工知能開発

所属先：大阪公立大学医学部医学科整形外科学教室

氏名：玉井 孝司

緒言

外傷性頸部症候群は交通外傷において最も高頻度に発生する病態である。警察庁の統計によると、2019年度の交通事故による軽傷者429,750人のうち、約60% (244,607人)が外傷性頸部症候群と診断されている[1]。さらに、外傷性頸部症候群を受傷した患者の7.7%が6ヶ月以上の加療を要する慢性症状へ移行すると報告されており[2]、現代社会における重要な医学的・社会的課題となっている。

本研究の目的は、交通事故受傷直後の頸椎X線写真側面像を用いて、外傷性頸部症候群受傷後の慢性化を予測するDeep learningアルゴリズムを確立することである。このアルゴリズムにより、早期から集学的治療を要するハイリスク症例を同定することが可能となる。

慢性疼痛化リスクを予測する人工知能(AI)アルゴリズムの開発により、外傷性頸部症候群患者の中から集学的治療(早期リハビリテーション介入、薬物療法、認知行動療法、適切な装具指導)を必要とする症例を効率的に抽出することが可能となる。これにより、限られた医療資源や医療費を効果的に活用し、交通外傷患者の疼痛緩和やQOL改善に寄与することが期待される。

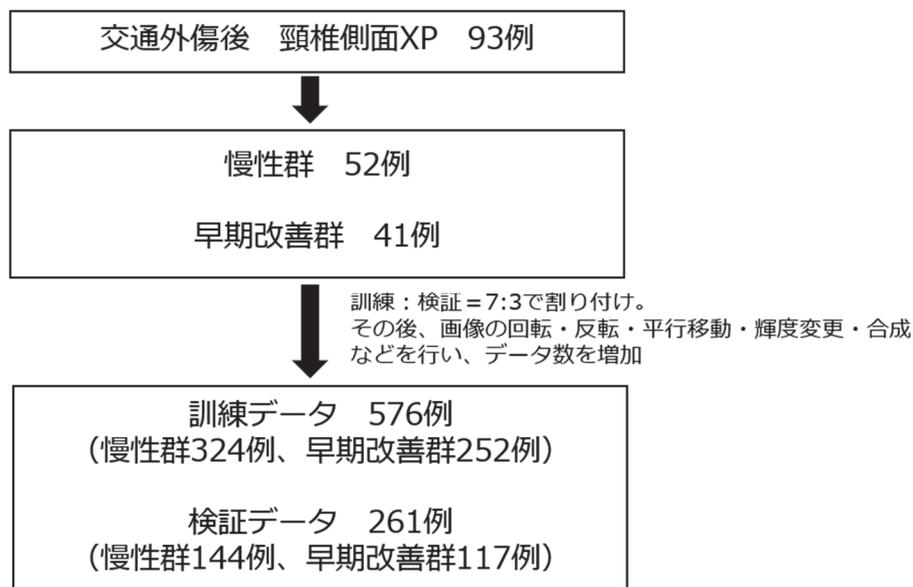
対象

本研究は、大阪公立大学医学部医学科整形外科学教室関連病院のうち、2次救急を受け入れる地域基幹病院にて2010年から2022年の間に外傷性頸部症候群と診断され、最終転帰が判明している93名(平均年齢42.5歳、女性33例)を対象とした。外傷性頸部症候群の重症度評価には、1995年にケベックむち打ち症関連障害特別調査団が報告したケベック分類[3]を採用した。本分類は以下の5段階で構成される: Grade 0: 臨床症状および理学的所見なし、Grade I: 頸部に疼痛、圧痛を認めるが、理学的所見なし、Grade II: 頸部に明確な愁訴があり、筋・骨格系の兆候(関節可動域制限、圧痛)を認める、Grade III: 頸部の愁訴に加え、神経学的兆候を伴う、Grade IV: 頸部の愁訴に加え、骨折または脱臼が認められる。本研究では、交通外傷受傷後4週時の臨床評価において、Grade II以上を慢性群、Grade I以下を早期改善群と定義した。慢性群は52例(平均年齢46歳、女性67%)、早期改善群は41例(平均年齢39歳、女性61%)であった。

データセット

交通事故受傷直後の頸椎X線写真側面像93枚をDICOM (Digital imaging and communications in medicine)からJPEG (Joint Photographic Experts Group)形式に変換し学習に使用した。訓練データと検証データは7:3の割合でランダムに分割した。データ数の不足とクラス不均衡に対処するため、画像の回転・反転・平行移動・輝度変更・合成などのデータ拡張を行い、最終的に訓練データを576例(慢性群324例、早期改善群252例)、検証データを261例(慢性群144例、早期改善群117例)とした(図1)。

図1 データフロー



CNNモデル構築

本研究では、過去の報告[4,5]を参考に2つのCNNモデルを構築した。具体的には、交通事故後X線写真を用いて早期改善群と慢性群の2値分類を行うために、以下のモデルを採用した：モデル1: EfficientNetB2、モデル2: VGG16、VGG19、DenseNet201、ResNet50を組み合わせたアンサンブル学習手法。損失関数には、クラスの不均衡を考慮し、クロスエントロピー損失関数にクラスの重みを適用した。学習にはAdamオプティマイザを用い、学習率スケジューラーを導入、過学習を防ぐためにearly stoppingの手法を用いた。アンサンブルモデルでは、各モデルの出力を平均化し、最終的な予測を行った。評価にはAUC (Area Under the Curve)、Accuracyに加えて適合率 (Precision)、再現率 (Recall)、F1-scoreを計算することで、モデルの分類能力を多角的に評価した。モデルの学習および推論にはGPU (NVIDIA GeForce GTX1080Ti)を使用し、コーディングはPythonで行った。CNNモデルの実装にはPyTorchを使用し、画像処理にはOpenCVライブラリーを用いた。判断根拠の可視化には、Gradient-weighted Class Activation Mapping (Grad-CAM)法を用いた。

結果

モデル1: EfficientNet B2

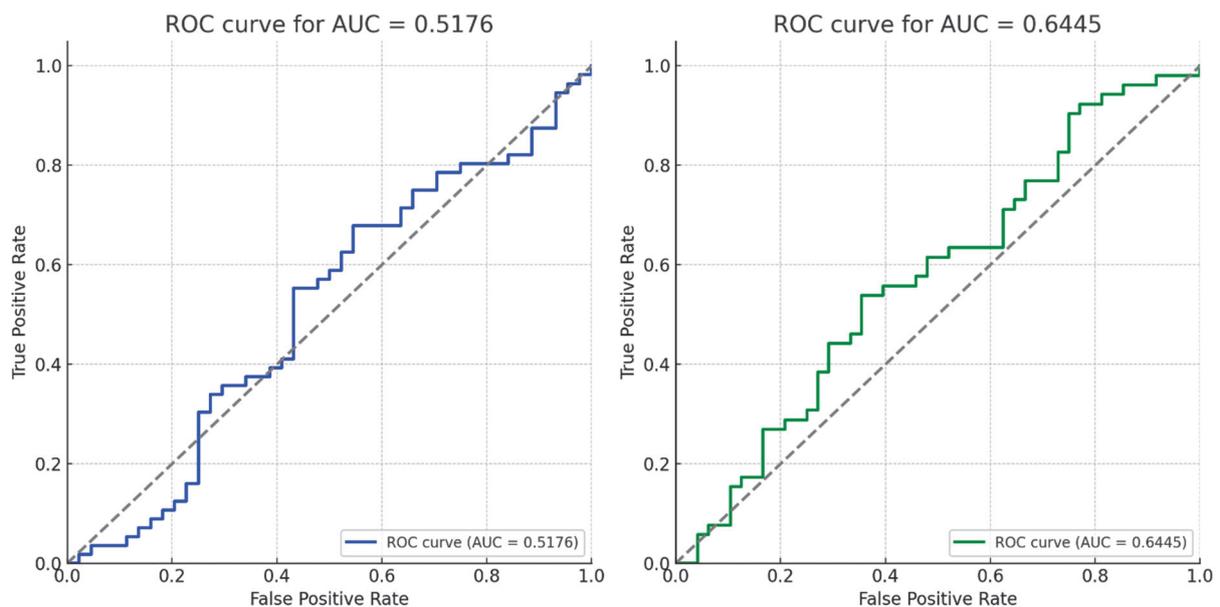
Final Test Accuracy: 57.09%, Test AUC: 0.517, Precision: 0.493, Recall: 0.472, F1-Score: 0.482

モデル2: VGG16、VGG19、DenseNet201、ResNet50を組み合わせたアンサンブル手法

Final Test Accuracy: 62.45%, Test AUC: 0.645, Precision: 0.652, Recall: 0.625, F1-Score: 0.638

図2に各モデルのROC曲線を示した。

図2 Receiver Operating Characteristic (ROC)曲線



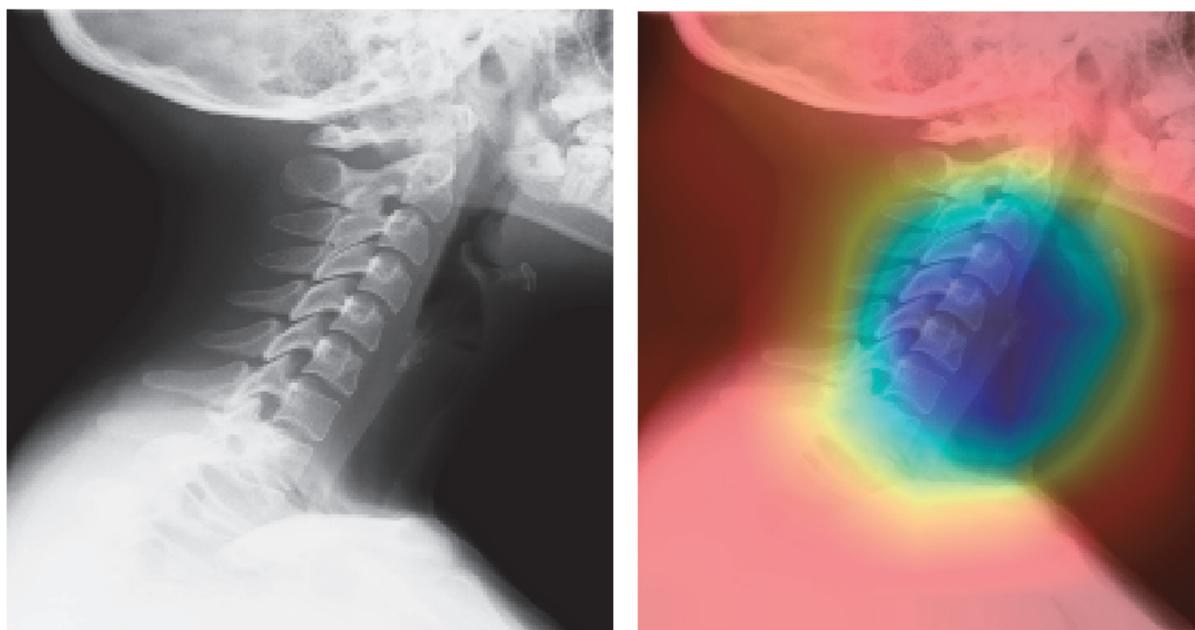
左図: モデル1: EfficientNet B2

右図: モデル2: VGG16、VGG19、DenseNet201、ResNet50を組み合わせたアンサンブル手法

判断根拠の視認化

Grad-CAMによる判断根拠の可視化では、CNNが椎体そのものよりも頸椎後方軟部組織に注目している傾向が観察された(図3)。

図3 CNNが着目した箇所を示すヒートマップ



左図: 頸椎単純X線側面

右図: CNNが着目し判断した箇所を示すヒートマップ

考察

本研究では症状慢性化の定義を交通外傷後4週時の症状のケベック分類にて判断した。むち打ち症関連障害に対するケベック診療ガイドラインでは交通外傷後に改善が認めない場合、3週間時に医師の再評価、6週間時には疼痛チームや、リハビリテーション分野の専門家による評価が必要とされている[3]。我々のAIアルゴリズム構築の一番の目的は集学的治療（早期リハビリテーション介入、薬剤介入、認知行動療法、適切な装具指導）を必要とする症例を絞り、より適切な治療を提供できるようにすることであるため、4週時の症状残存が残る症例を特定することは重要であると考えられる。

本研究ではAUC、Accuracyに加えて、モデル識別能力をより理解するためにPrecision, Recall, F1-scoreという指標を用いた。Precision(適合率)は、モデルが陽性と予測したサンプルのうち、実際に陽性であった割合を示す。これは誤検出(偽陽性)を最小化することが重要な場合に特に有用な指標である。Recall(再現率)は、実際に陽性であるサンプルのうち、モデルが正しく陽性と予測した割合を示す。この指標は、陽性サンプルの見落とし(偽陰性)を最小化することが重要な場合、例えば疾病検出や安全性に関わる問題にお

いて特に重要となる。F1-scoreは、PrecisionとRecallの調和平均であり、両指標のバランスを考慮した総合的な評価指標である。これらの指標を用いることで、モデルの性能をより多角的に評価することが可能となる。モデル2ではAccuracy, AUCに加えて、Precision, Recall, F1-scoreのスコアいずれも相対的に良好なスコアを示したことから、本研究においてはモデル2の方がより良いモデルと解釈できた。

興味深いことに、CNNが着目した個所を示すヒートマップでは、(詳細な判断根拠は不明であるものの)椎体そのものへ注目はされておらず、頸椎後方軟部組織に注目されていた。ケベックむち打ち症関連障害特別調査団による診療ガイドラインでは、最終治療として軟部組織修復の観点からその修復が完了する12週の時点での各分野の専門評価により最終治療と定めている[3]。さらに、むち打ち患者に関する長期MRI試験、最短10年間の追跡調査による報告によると、むち打ち患者の頸椎のMRIでの変性的変化の進展は98%の患者に認めたのに対して、頸椎椎体におけるMRI所見での臨床上的関連はなかったと報告されている[6]。我々が構築したAIモデルが椎体そのものではなく、頸部後方の軟部組織に注目していたことも上記研究などと矛盾しない結果であった。

本研究でのモデル構築ではアンサンブル手法を用いたモデルが精度62.4%であり、医療現場での実用化には不十分であり、更なるモデル向上が必要である。今後の改善点としては以下の2点が挙げられる。1つ目は、症例数の増加である。技術の進歩により、Big dataの解析が可能になってきており、画像の回転・反転・平行移動・輝度変更・合成などの技術を用いてデータ数を増加させる工夫を行ったが、実際の症例数は93例にとどまった。いずれのモデルにおいても学習を進めていくと、訓練データに対する予測能が高くなる一方で、テストデータに対するスコアが伸び悩んでしまう、いわゆる過学習の状態に陥ってしまう。Deep learningモデルの構築においては、良いデータを多く集めることが必須であり、より多くの画像からより多くの学習をすることができれば、モデルの汎化性能を向上させることができる。2つ目は画像情報以外からのデータの統合である。Deansらは、交通事故の状況(スピード、方向、安全装備)や被害者の受傷時の頸椎の状態など多くの因子により傷害の程度、発生部位は変化すると報告している[7]。本研究は頸椎側面XP画像一枚から症状遷延についてモデル構築を試みたが、その他の情報も交通事故には非常に大きな要因であるため、画像情報だけではなく、多因子を含めたモデル構築が今後期待される。

結論

外傷性頸部症候群と診断された頸椎XP画像を用いて、Deep learning algorithmを確立した。単一モデル(EfficientNet B2)を用いたモデルより、アンサンブル手法(VGG16、VGG19、DenseNet201、ResNet50)を用いたモデルの方が、外傷性頸部症候群受傷後の慢性化を予測精度は高い結果であったが、実用化にはさらなる改善が必要である。本研究のアプローチを発展させることで、早期から集学的治療を要するハイリスク症例を同定し、交通外傷患者の疼痛緩和やQOL改善に貢献することが期待される。今後は、症例数の増加と多因子を考慮したモデル構築を通じて、より高精度なシステムの開発を目指す。

参考文献

1. 交通事故統計情報のオープンデータ 警察庁 2019
2. 田中信弘 外傷性頸部症候群の 文献レビュー 臨床整形外科 2023
3. Spitzer WO, Skovron ML, Salmi LR, et al. Scientific monograph of the Quebec task force on whiplash-associated disorders : redefining “whiplash” and its management. *Spine (Phila Pa 1976)* 1995 ; 20(8 Suppl) : 1S-73S.
4. Tamai K, Terai H, Hoshino M, et al. Deep Learning Algorithm for Identifying Cervical Cord Compression Due to Degenerative Canal Stenosis on Radiography. *Spine (Phila Pa 1976)*. 2023;48(8):519-525.
5. Yabu A, Hoshino M, Tabuchi H, et al. Using artificial intelligence to diagnose fresh osteoporotic vertebral fractures on magnetic resonance images. *Spine J.* 2021;21(10):1652-1658.
6. Ichihara D, Okada E, Chiba K, et al. Longitudinal magnetic resonance imaging study on whiplash injury patients: minimum 10-year follow-up. *J Orthop Sci.* 2009;14(5):602-610.
7. Deans GT, Magalliard JN, Kerr M, et al. Neck sprain—a major cause of disability following car accidents. *Injury.* 1987;18(1):10-12.