



Title	頭頸部悪性腫瘍の局所浸潤に関する画像診断への深層学習の応用の検討
Author(s)	中川, 純一
Citation	北海道大学. 博士(医学) 甲第16058号
Issue Date	2024-06-28
DOI	10.14943/doctoral.k16058
Doc URL	http://hdl.handle.net/2115/92786
Type	theses (doctoral)
File Information	NAKAGAWA_Junichi.pdf (この論文の全文)



[Instructions for use](#)

学位論文

頭頸部悪性腫瘍の局所浸潤に関する
画像診断への深層学習の応用の検討

(Investigations on the application of deep learning
to the diagnostic imaging for local invasion
in head and neck malignant tumor)

2024 年 6 月

北海道大学
中川 純一

学位論文

頭頸部悪性腫瘍の局所浸潤に関する
画像診断への深層学習の応用の検討

(Investigations on the application of deep learning
to the diagnostic imaging for local invasion
in head and neck malignant tumor)

2024年6月

北海道大学
中川 純一

目次

発表論文目録および学会発表目録	1
要旨	2
略語表	5
全体の緒言	6
第一章	
緒言	10
方法	11
結果	18
考察	26
第二章	
緒言	28
方法	30
結果	37
考察	46
全体の結論	48
謝辞	50
利益相反	51
引用文献	52

発表論文目録および学会発表目録

本研究の一部は、以下の論文に発表した

1. Nakagawa J, Fujima N, Hirata K, Tang M, Tsuneta S, Suzuki J, Harada Y, Ikebe Y, Homma A, Kano S, Minowa K, Kudo K.
Utility of the deep learning technique for the diagnosis of orbital invasion on CT in patients with a nasal or sinonasal tumor.
Cancer Imaging 2022 Sep 22;22(1):52. doi: 10.1186/s40644-022-00492-0.
2. Nakagawa J, Fujima N, Hirata K, Harada Y, Wakabayashi N, Takano Y, Homma A, Kano S, Minowa K, Kudo K.
Diagnosis of skull-base invasion by nasopharyngeal tumors on CT with a deep-learning approach
Japanese Journal of Radiology 2024 Jan 27. doi: 10.1007/s11604-023-01527-7.

本研究の一部は、以下の学会に発表した

- Nakagawa J, Fujima N, Hirata K, Tang M, Tsuneta S, Suzuki J, Homma A, Minowa K, Kudo K.
Deep learning assistance for CT diagnosis of orbital invasion by nasal or sinonasal tumors - Like a specialist giving you the answers
EUROPEAN CONGRESS OF RADIOLOGY ECR 2023, Vienna, 1-5/January/2023

要旨

【背景と目的】頭頸部腫瘍は直視的・内視鏡的な観察・生検が他の領域より容易である一方、眼窩内や頭蓋底等の深部方向の視覚評価や病理診断が困難であるため、画像診断は腫瘍進展の評価に重要である。機械学習は人工知能技術の一つで、明示的にプログラムすることなくデータから規則性を導く手法である。その中でも深層学習は規則性を反映する情報である特徴量の自動抽出が可能で、特に畳み込みニューラルネットワーク（Convolutional Neural Network：CNN）は画像認識に長けている。CNNを悪性腫瘍の局所浸潤の診断に応用し、放射線科医と同等以上の診断性能を示した既報は散見されるが、頭頸部領域の報告は非常に少ない。そこで頭頸部腫瘍による様々な部位への局所浸潤の画像診断について、CNNモデルは頭頸部非専門の放射線科医より高い診断性能を示すという仮説を立てた。第一章では鼻副鼻腔悪性腫瘍による眼窩骨膜を越えた浸潤、第二章では上咽頭悪性腫瘍による頭蓋底浸潤のCT診断についてCNNモデルを作成し、診断性能を放射線科医の読影実験と比較し検証した。

○第一章

【対象と方法】2009年1月～2021年3月に当院を受診した鼻副鼻腔悪性腫瘍の168例を頭頸部専門の放射線診断専門医2名（画像診断の経験年数6・15年）の合議で眼窩骨膜を超えた浸潤のCT所見の有無を評価した（陽性：陰性＝81：87）。対象を約7：3にランダムに分配し、7割をCNNモデルの訓練用（n=119。陽性：陰性＝56：63）に、残りを作成したモデルのテスト用（n=49。陽性：陰性＝25：24）とした。前処理として冠状断CTで眼窩に正方形ROIを設定し、眼窩の連続スライスを抽出した。訓練データはデータ拡張を行い、学習済みCNNモデルのVGG16（Visual Geometry Group 16）に入力した。作成したCNNモデルに前処理済みのテストデータの画像を入力し、各スライスの浸潤陽性・陰性を出力した。これを各症例の陽性・陰性の診断に変換するために、モデルが陽性と判定したスライスの連続枚数でROC曲線を作成し、Youden indexでカットオフ値を決定した。読影実験として頭頸部非専門の一般放射線科医2名（経験年数6・3年目）がテストデータの冠状断CTの元画像を読影し、各症例の陽性・陰性を診断した。約2ヶ月後、CNNモデルの各症例の陽性・陰性の結果を知った上でCT画像を再評価した。

【結果】作成した CNN モデルは ROC 曲線で AUC が 0.94、カットオフを陽性スライスが連続 3 枚以上とすると、精度は 0.92 であった。2 名の一般放射線科医は補助なしの 1 回目の読影（読影者 1：AUC 0.49・精度 0.49、読影者 2：AUC 0.45・精度 0.45）よりも、CNN モデルによる陽性・陰性の結果を知った上で読影した 2 回目で有意に成績が向上した（読影者 1：AUC 0.94・精度 0.94、読影者 2：AUC 1.00・精度 1.00）（ $p < 0.001$ ）。

【考察】鼻副鼻腔悪性腫瘍による眼窩骨膜を越えた浸潤の CT 診断において、CNN モデルは 2 名の一般放射線科医より有意に高い診断能を発揮した。本研究の限界として、単一施設のためサンプルサイズが小さい事、ROI の手動設定が必要である事、眼窩骨膜を超えた浸潤の正解ラベルの基準が CT 所見のみで病理や MRI を用いておらず頭頸部専門の放射線科医の意見に正解ラベルが依存してしまう事があげられる。第一章に続き、様々な浸潤部位における CNN の有用性を検証すべく第二章では眼窩浸潤と並び実臨床でよく遭遇する頭蓋底浸潤に CNN を適応した。かつ頭蓋底浸潤のゴールドスタンダードである MRI も参考にすることで正解ラベルの客観性を重視した。

○第二章

【対象と方法】2008 年 1 月～2022 年 3 月に当院を受診した上咽頭悪性腫瘍の 100 例を頭頸部専門の放射線診断専門医 2 名（画像診断の経験年数 7・16 年）の合議で MRI を参考にしながら頭蓋底浸潤の CT 所見の有無を評価した（陽性：陰性＝42：58）。対象を約 7：3 にランダムに分配し、7 割を CNN モデルの訓練用（ $n=70$ 。陽性：陰性＝29：41）に、残りを作成したモデルのテスト用（ $n=30$ 。陽性：陰性＝13：17）とした。前処理として水平断 CT で上咽頭と頭蓋底を含むように長方形 ROI を設定し、上咽頭と頭蓋底の連続スライスを抽出した。訓練データはデータ拡張を行い、学習済み CNN モデルの VGG16、ResNet50（Residual Networks 50）、Xception に入力し、訓練データでの検証精度が最も高いものを採用することとした。作成した CNN モデルに前処理済みのテストデータの画像を入力し、各スライスの浸潤陽性・陰性を出力した。これを各症例の陽性・陰性の診断に変換するために、モデルが陽性と判定したスライスの連続枚数で ROC 曲線を作成し、Youden index でカットオフ値を決定した。読影実験として頭頸部画像診断の経験が豊富な放射線診断専門医（経験年数 13 年。Senior reader：SR）、経験の少ない放射線科医（経験年数 1 年。Junior reader：JR）がテストデータの水平断 CT の元画像を読影し、各症例の陽性・陰性を診断した。また、ROI 設定者の違いによる CNN モデルの診断結果の変化を評価するために、別の放射線科医（経験年数 2 年）がテストデータ

の水平断 CT に長方形 ROI を設定し、抽出した画像を作成した CNN モデルに入力して診断性能を評価した。

【結果】 検証精度が最も高かった ResNet50 で作成した CNN モデルは ROC 曲線で AUC が 0.95、カットオフを陽性スライスが連続 2 枚以上とすると、精度は 0.97 であった。2 名の放射線科医の読影実験の結果は SR : AUC 0.83 ・精度 0.83、JR : AUC 0.62 ・精度 0.60 であった。AUC は CNN モデル、SR、JR の順に有意に高かった ($p < 0.05$)。別の ROI 設定者の抽出画像での CNN モデルの ROC-AUC は 0.97 で、オリジナルの ROI 設定時の AUC と有意差は無かった ($p = 0.730$)

【考察】 上咽頭悪性腫瘍の頭蓋底浸潤の CT 診断において、作成した CNN の診断モデルは放射線科医よりも有意に高い診断性能を示した。本研究の限界として、単一施設のためサンプルサイズが小さいこと、ROI の手動設定が必要である事、頭蓋底浸潤の病理学的検証が得られていない事があげられる。ROI については別の設定者でも有意差の無い診断性能が得られたことから、臨床使用において操作者間の変動はある程度許容できる可能性を示している。病理が得られていない点については、実臨床における判断基準も主に MRI での信号変化の有無に基づいており乖離は少ないと考えられた。

【全体の結論】 頭頸部腫瘍の局所浸潤の CT 診断について転移学習を用いた CNN モデルを作成し、第一章・第二章ともに放射線科医の診断性能と同等以上の結果が得られた。仮説が立証できたと考えられ、頭頸部画像診断の専門家がない施設での補助ツールになると期待される。また、今回の研究から様々な部位への腫瘍浸潤の評価に CNN モデルが応用可能と考えられる。

略語表

本文中および図中で使用した略語は以下の通りである。

adam	adaptive moment
AI	artificial intelligence
AUC	area under the curve
CNN	convolutional neural network
CT	X-ray computed tomography
DICOM	Digital Imaging and Communication in Medicine
DL	deep learning
Fs	fat-saturation
GdT1WI	gadolinium-enhanced T1 weighted image
Grad-CAM	gradient-weighted class activation map
HU	Hounsfield units
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
JPEG	Joint Photographic Experts Group
MRI	magnetic resonance imaging
NPV	negative predictive value
PPV	positive predictive value
ResNet50	Residual Networks 50
ROC	receiver operating characteristic
ROI	region of interest
sgdm	stochastic gradient descent with momentum
T1WI	T1 weighted image
T2WI	T2 weighted image
UICC	Union for International Cancer Control
VGG16	Visual Geometry Group 16
XAI	explainable artificial intelligence

全体の緒言

1. 背景

(1) 頭頸部領域の悪性腫瘍について

頭頸部とは頭蓋底より尾側、鎖骨より頭側を指し、主に耳鼻咽喉科・頭頸部外科、歯科・口腔外科、眼科が診療の対象とする領域である。頭頸部腫瘍は鼻副鼻腔、咽喉頭、口腔、唾液腺などから発生する腫瘍の総称であり、その中でも頭頸部癌は世界で5番目に頻度の高い癌とされている (Bray et al., 2018)。頭頸部腫瘍は直視的あるいは内視鏡的な観察・生検が脳、胸部、腹部など他の領域より施行しやすい一方で、眼窩内や頭蓋底、頭蓋内といった深部方向への広がりについては視覚評価および病理学的な検索が困難である。そのため、原発巣の評価において X-ray computed tomography (CT)、magnetic resonance imaging (MRI) を中心とした画像診断には腫瘍の存在診断や質的診断、鑑別診断だけでなく、進展評価が特に求められ、Union for International Cancer Control (UICC) の TNM 分類に代表される病期分類の決定および治療方針の選択にも大きく影響する。

(2) 頭頸部腫瘍の画像診断について

CT は臨床現場に広く普及しており、MRI よりも短い時間で撮像可能なため、通常は頭頸部腫瘍の最初の評価に用いられる。その短い撮像時間から嚥下や呼吸を含めた体動のアーチファクトの影響を最小限に抑えることができ、重篤な患者や長時間の安静臥床が困難な症例にも施行可能である。また、MRI が禁忌である金属や植え込み型デバイスを有する症例も利用できる。加えて、薄層の高分解能画像を取得することで細かな解剖構造の評価が可能であり、多断面再構成による多方向からの観察が可能である。CT は骨皮質・石灰化や空気の検出・評価に優れているが、軟部組織の評価は MRI より劣るため、腫瘍と正常構造とのコントラストが不十分になりえる。また、CT は歯科補綴物などの金属のアーチファクトが MRI より目立ち、診断の妨げとなりやすい (S.Y. Kim et al., 2021)。

一方、MRI は軟部組織のコントラストが高く、骨髄、軟部組織、隣接臓器浸潤の評価に優れている。特に傍咽頭間隙、頭蓋底、頭蓋内、蝶形骨洞進展において CT よりも優れるとされ、顎骨についても骨皮質の評価は CT が有用で

ある一方で骨髄の浸潤には MRI が有用とされている。以上のことから本邦の画像診断ガイドラインでは T 因子病期診断において上咽頭、中咽頭、口腔については MRI が推奨されている。一方で、MRI において隣接する空気による磁化率アーチファクトや嚥下等の体動アーチファクトの影響が強くなる下咽頭、喉頭は CT が推奨されている（日本医学放射線学会, 2021）。また、頭頸部癌診療ガイドラインにおいて、鼻・副鼻腔については CT や MRI 単独で判断するよりも両者を併せて判断するほうが頭蓋内進展の感度が向上するとされている（日本頭頸部癌学会, 2022）。

(3) 人工知能技術について

人工知能（artificial intelligence : AI）は人間の知能を必要とするタスクを実行するシステムの作成を目的としたコンピュータサイエンスの一分野で、医療分野にも応用されている。機械学習は AI 技術の一つであり、明示的にプログラムすることなく、コンピュータが与えられたデータから情報を整理して規則性を導く手法である（Chartrand et al., 2017）。古典的な機械学習では導きたい規則性をよく反映する情報である“特徴量”を人間が抽出・設定する必要があったが、近年発展している深層学習（deep learning : DL）では特徴量の自動抽出が可能となった（間下, 2018）。

DL は機械学習モデルの一つであるニューラルネットワークを基にしている。これは複数のニューロンの興奮がシナプスを介して次のニューロンへ伝わるといった階層構造を模しており、複数の入力値に重み付けをして組み合わせ、結果を出力するものである。DL はこのニューラルネットワークを多層に積み重ねたものといえる（Chartrand et al., 2017）。DL 技術の一つである畳み込みニューラルネットワーク（convolutional Neural Network : CNN）は一次視覚野を模したアルゴリズムで画像認識に長けている。CNN は畳み込み層とプーリング層が交互に構成されており、畳み込み層では入力データにフィルタやカーネルと呼ばれる小さな行列をかけあわせて入力画像のエッジや曲線などの特徴を抽出した“特徴マップ”を作成し、プーリング層では特徴マップの解像度を低くして画像内の対象物体の小さな移動や歪みに影響されにくくなる（Do et al., 2020）。

(4) 深層学習の手法

機械学習や DL における代表的な学習方法として教師あり学習と教師なし学習がある。教師あり学習は入力データが何を示しているかという正解を予め付

けておく必要があり、その正解のことを“正解ラベル”や“ground truth”と呼ぶ (Do et al., 2020)。このラベル付けされたデータから一般的な規則や傾向を学習し、未知のデータに対して予測や分類をするものである。一方で教師なし学習はラベルが付与されていないデータを与えて、その中から類似性や規則性を学習し、グループ分けするものである (森康, 2018)。

収集したデータは“訓練データ”と“テストデータ”に分けられる。訓練データは DL モデルを目的のタスクを遂行させるために最適化させるために使用され、テストデータは作成した DL モデルの性能を最終的に評価するために使用される。データを分割する際は、両者のデータセットは重複しないようにすることが重要である (Do et al., 2020)。DL モデルに入力された訓練データから導かれる予測値と、正解ラベルに相当する訓練データの実測値との差分を誤差や損失と呼び、この誤差・損失を最小化するように、パラメータを修正・更新することを“学習”と呼ぶ。

DL においてモデルの学習の前に人の手で予め設定する必要のあるパラメータをハイパーパラメータと呼ぶ。これには、訓練データ全体を使用する反復回数である“エポック数”、訓練データを数組に分割した際の分割数である“イテレーション数”、分割された 1 組の中にあるサンプル数である“ミニバッチサイズ”、学習によってモデル内のパラメータを更新する際の変化率である“学習率”などがある。また、パラメータを更新するために誤差・損失の最小値を求めるためのアルゴリズムのことを“最適化アルゴリズム (optimizer)”と呼ぶ。

医療画像に限らず、画像認識の DL モデルの多くは教師あり学習を用いているが、ラベル付けされたデータ数の不足が課題となる。少ないデータでも DL モデルの精度を向上するために用いられる手法として、“転移学習 (transfer learning)”と“データ拡張 (data augmentation)”がある。転移学習は事前に別のタスクでトレーニングされた DL モデルに、目的のために用意したデータを学習させて目的のタスクに転用することである。画像認識においては、約 1000 種類の植物や動物などの自然画像のカラー写真が 1400 万枚以上集められた大規模データベースである ImageNet (<http://image-net.org/index>) (Simonyan et al., 2015) で学習した CNN モデルを転用するのが一般的である (Chartrand et al., 2017)。データ拡張は訓練データとなる画像に反転、回転、平行移動、ズームなどの処理を加えて、人工的に訓練データをかさ増しする方法である。これにより本質的に同じ所見でもわずかに異なる外観を示すデータを学習することで、モデルを堅牢で汎用性のあるものにすることができる (Do et al., 2020)。

また DL の進歩により、人間の介入を最小限に抑えながら複雑な医用画像の

解析が容易になった。しかし、多くの DL モデルは、その非線形な基礎構造により、しばしば「ブラックボックス」とみなされる。このため、DL モデルの透明性を高めることを目的とした、説明可能な人工知能 (explainable artificial intelligence : XAI) の研究が登場している (Qian et al., 2023)。

2. 何がわかっていて、なにがわかっていないのか

AI 技術は医療画像分野においては画像分類、病変検出、セグメンテーションにおいてよく用いられている。開発されている診断支援システムの例としては胸部 X 線写真や CT による肺病変の診断、乳房、前立腺、結腸の病変の検出、X 線による骨折の診断、MRI による半月板や軟骨病変の診断、CT による脳出血の検出、血液流量の定量化などがある (Blum et al., 2017)。

前述のように頭頸部腫瘍の評価には最初に CT が用いられる。CT 上の悪性腫瘍の局所浸潤診断について CNN 技術を応用した研究としては、肺癌による胸膜浸潤 (Choi et al., 2021) や膀胱癌による筋層浸潤 (Zhang et al., 2021; Yang et al., 2021) で報告されている。しかしながら、検索した限りでは頭頸部悪性腫瘍の原発巣による局所浸潤の CT 診断への応用に関する報告は確認できず、関連するものとしては悪性腫瘍の浸潤に限らず骨折なども含めた甲状軟骨の異常 (Santin et al., 2019) や、頭頸部扁平上皮癌のリンパ節転移における節外浸潤の診断 (Kann et al., 2020) への応用が報告されているのみであった。

3. 本研究で何を明らかにしようとしているか

頭頸部は狭い範囲に重要な解剖構造が密集しており、画像診断医でも読影や解釈に苦手意識をもたれやすい領域でもある。そこで、AI 技術を応用することで画像診断を支援できるかを検討した。本研究に際して、頭頸部腫瘍による様々な解剖構造に対する局所浸潤について、転移学習を用いた CNN モデルは頭頸部を専門としない放射線科医より高い診断能を示すという仮説を立てた。第一章では鼻副鼻腔悪性腫瘍による眼窩骨膜を越えた眼窩浸潤の CT 診断について、第二章では上咽頭悪性腫瘍による頭蓋底浸潤の CT 診断について、CNN モデルを作成し、その診断性能を放射線科医の読影実験の結果と比較することで検証した。

第一章 鼻副鼻腔悪性腫瘍による眼窩浸潤に関する CT 診断への深層学習の応用

緒言

鼻腔または副鼻腔腫瘍の病期分類と術前評価には、切除可能性を判断するためにも画像診断が不可欠である (Meerwein et al., 2020; Salfrant et al., 2021)。鼻副鼻腔の悪性腫瘍はしばしば眼窩に浸潤し、眼窩骨壁への浸潤は上顎洞悪性腫瘍の 60%~80% (Carrau et al., 1999)、および、篩骨悪性腫瘍の 66~82%

(Ganly et al., 2005) に見られるとされている。鼻副鼻腔癌の眼窩への浸潤は UICC の臨床 TNM 病期分類基準第 8 版によると眼窩底・内側壁浸潤が T3、眼窩内容前部が T4a、眼窩先端が T4b に分類されている (Huang et al., 2017)。眼窩骨膜は内眼窩の骨膜であり、眼窩腔の前方開口部から視神経管および上眼窩裂のある眼窩尖部まで、4 つの眼窩壁を覆っている (Cornelius et al., 2014)。眼窩骨膜は、眼窩内容物への腫瘍進展に対する効果的なバリアであると考えられてきた (Eisen et al., 2000)。そのため腫瘍と眼窩骨膜との関係は手術選択の判断基準の 1 つとされ、眼窩骨膜の破綻を伴う眼窩浸潤では眼窩内容を摘出、眼窩骨膜の破綻がない場合は眼窩内容を温存するケースが多いとされる (Eisen et al., 2000; Suarez et al., 2008)。また、眼窩骨膜を越えた浸潤は生存率の有意な低下と関連するとされ、予後不良を示唆する所見である (Perry et al., 1988; Suarez et al., 2008)。CT および MRI において、外眼筋の巻き込みや眼窩内脂肪の混濁は、眼窩周囲を超える眼窩浸潤の高い陽性適中率を有することが報告されている (Eisen et al., 2000)。しかし、眼窩骨膜を超える眼窩浸潤の診断には経験豊富な放射線科医であっても難しく、この領域の専門家でない放射線科医にとっては更に困難である。

そこで、AI 技術を用いることで、CT 上で眼窩骨膜を超える眼窩浸潤の判定が可能になり、放射線科医の読影に役立つのではないかと考えた。本研究では、眼窩骨膜を越える眼窩浸潤の CT 所見の有無を診断するための CNN モデルを作成し、作成した CNN モデルの診断性能を放射線科医の読影実験と比較することで評価することを目的とした。

方法

1. 対象

本研究は、北海道大学病院自主臨床研究審査委員会での承認を得て行われ（研究番号：021-0001）、研究参加に際してはオプトアウトとした。

診療録に基づいて、2009年1月から2021年3月までに当院を受診した鼻副鼻腔腫瘍患者を233人・234病変（うち1人は異時性多発癌）が、以下の選択基準に基づいて組み入れられた：（1）病理学的に悪性鼻副鼻腔腫瘍と確定されている、（2）治療前のCTにおいて腫瘍を含む範囲の軟部条件の冠状断像が作成されている。一部の患者は以下の除外基準により除外された：（1）血液腫瘍のような手術不要な腫瘍の患者（20例）、（2）CTの撮像または再構成パラメータが得られない場合（スライス厚、マトリックスサイズなど）（12例）、（3）原発腫瘍が眼窩壁から明らかに離れた位置にある患者（32例）、（4）義眼など標的病変の画像所見に重大な影響を与える眼窩内構造物を有するもの（2例）。最終的に、167人の患者とその168病変が本研究の対象となった。

2. CT画像

168病変のCT画像が、4社の様々なCTスキャナーで撮像された（主な装置：Aquilion PRIME、東芝メディカルシステムズ、大田原。Sensation64、SIEMENS Medical Solutions、Malvern。Brilliance iCT 256、Philips Medical Systems、Best。LightSpeed VCT、GE Healthcare、Milwaukee）。168病変のうち、150病変は造影後画像であり、残りの18病変は非造影画像であった。評価には軟部条件の冠状断CT画像を用いた。スライス厚は1～3mm、主なマトリックスサイズは512×512であった。

3. 眼窩浸潤の正解ラベル付け

頭頸部放射線診断を専門とする2名の放射線診断専門医（放射線診断の経験年数は6年と15年）が、CT画像の画質が判読に適切かどうかを判断した。DICOM（Digital Imaging and Communication in Medicine）ビューア（XTREK、J-MAC SYSTEM、札幌）を用いて、冠状断CT画像における眼窩壁の骨破壊、

腫瘍と眼窩内容物との境界の不整、腫瘍による外眼筋の巻き込み、腫瘍周囲の眼窩内脂肪の混濁などの画像所見を注意深く評価した。評価する眼窩の範囲は鼻涙管開口部から中頭蓋窩先端までと定義した。前述の2名の放射線診断専門医の合議により、すべての画像所見を加味して各画像においてCT所見上の眼窩骨膜の浸潤陽性群と浸潤陰性群（以下、浸潤陽性と浸潤陰性）に分類した。この画像診断の結果、81病変が浸潤陽性、残りの87病変が浸潤陰性と診断された。

4. 正解ラベルの信頼性評価

浸潤陽性・陰性の判定の信頼性を評価するために、前述の頭頸部を専門とする2名の放射線診断専門医による正解ラベル付けのための合議による画像評価から約9ヶ月後、同じ2名の放射線診断専門医がすべての症例を個別に再評価し、浸潤陽性・陰性に分類して、正解ラベルと両者の再評価の間での一致度を判断した。

加えて、前述とは別の頭頸部放射線診断の経験を持つ2名の放射線診断専門医（経験年数は7年と12年）が、合議により全症例を浸潤陽性・陰性の判定を行い、正解ラベルとの一致度を評価した。

5. 画像処理と深層学習

(1) 診断モデル作成のための症例の分割

収集した168病変を、浸潤陽性群（81病変）、浸潤陰性群（87病変）のそれぞれについて約7:3の割合となるように無作為に分配した。うち約7割の119病変（浸潤陽性56:陰性63）を診断モデル作成のための訓練データセットとして、残りの49病変（浸潤陽性25:陰性24）を作成したモデルの性能を評価するためのテストデータセットとして使用した。

以下に、症例選択からデータセットの分配までの流れを図示する（図 1-1）。

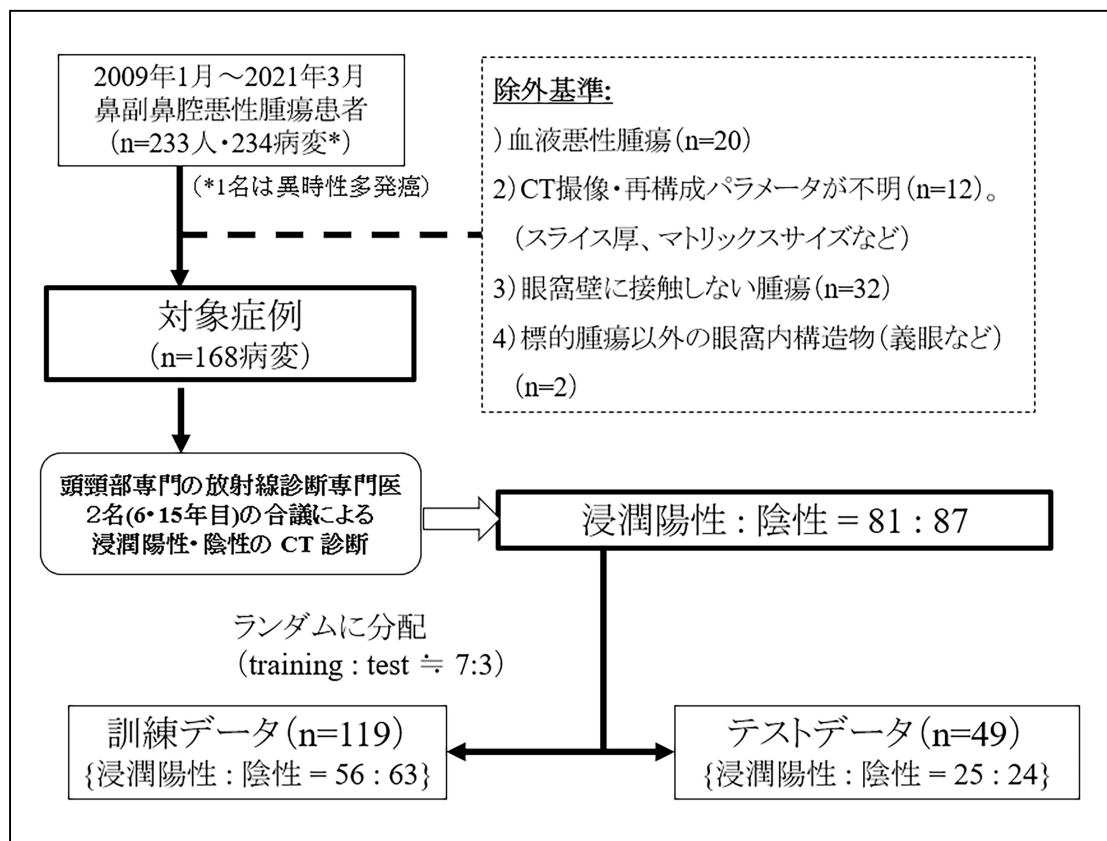


図 1-1 : 症例選択、正解ラベル付け、データセットの分配の流れ

(2) 画像の前処理

画像の前処理としてすべての冠状断 CT 画像について病変の以下のように抽出を行った。まず、眼窩骨壁を十分含むように正方形の関心領域 (regions of interest : ROI) を手動で設定した。腫瘍が正中線付近に存在して両側の眼窩に接触していた 3 病変 (訓練データセットの浸潤陽性 1 例、テストデータセットの浸潤陽性 1 例・陰性 1 例) は、左右別々に ROI を設定した。正解ラベル付けで用いた眼窩の評価範囲である鼻涙管開口部から中頭蓋窩先端について画像を抽出し、正方形の眼窩の連続画像を作成した。次に右眼窩の抽出画像を水平方向に反転し、すべてを左眼窩となるように統一した。病変と無関係な位置にある眼窩の上部と外側を三角形にマスクした。すべての画像の CT ウィンドウはウィンドウレベル 60 HU (Hounsfield unit)、ウィンドウ幅 300 HU に調整した。最後に、各々の画像を JPEG (Joint Photographic Experts Group) として出力した。

(3) データセットの画像選択とデータ拡張

訓練データセットの浸潤陽性群（56 病変）については、正解ラベル付けを行った 2 名の放射線診断専門医の合議によって CT 上の浸潤の所見があると判定したスライス（以下、浸潤陽性スライス）のみを用いることとし、408 枚が選択された。一方で浸潤陰性群（63 病変）については腫瘍が眼窩壁接触していないスライスは除いて、腫瘍が眼窩壁に接触ないし圧迫しているスライスを用いることとし、635 枚が選択された。また、モデルの汎化性能を向上させるために、各画像に対してランダムな回転と、垂直方向および水平方向の平行移動によるデータ拡張が行われ、各画像に対して 10 枚の追加画像が生成された。

テストデータセットについては、浸潤陽性と陰性の両群におけるすべての抽出画像を使用し、総計 640 枚のスライスが得られた。なお、テストデータセットは、データ拡張なしで用いられた。

(4) 深層学習によるモデル作成と性能評価

画像分類に特化した事前学習済み CNN モデルを用いた転移学習により、冠状断 CT 画像における浸潤陽性または陰性の分類に適応した CNN モデルの作成を試みた。本研究で使用したオリジナルモデルは、オックスフォード大学で開発された Visual Geometry Group 16 (VGG16) である。VGG は大規模な画像データベースである ImageNet を用いた画像認識の競技会である ILSVRC

(ImageNet Large Scale Visual Recognition Challenge) の 2014 年大会で第 2 位になったモデルである (Simonyan et al., 2015)。VGG16 は、5 つの畳み込みブロック (13 の畳み込み層) と 3 つの全結合層の組み合わせによる 16 層で構成され、ImageNet データセットで学習・評価されている。そのシンプルな構造のおかげで、VGG16 は小さなデータセットの転移学習によく適応している

(Santin et al., 2019)。CNN モデルの学習において、最後の全結合層の重みを今回作成した画像データによる学習によって変更し、他の先行層のパラメータは元の VGG16 の重みに固定した。これにより、VGG16 の汎用的な特徴を維持し、限られた数の訓練可能なパラメータによってモデルを CT 画像に適応させることができた。本研究では入力した画像について浸潤陽性または陰性の 2 値分類としてモデルが出力するようにした。

転移学習のハイパーパラメータは 15 エポック、ミニバッチサイズ 32、学習率 1.0×10^{-5} に設定し、最適化アルゴリズム (optimizer) には stochastic gradient descent with momentum (sgdm) を使用した。訓練データのうち 30% を検証データとして使用し、各ミニバッチにおける精度を評価した。

作成した CNN モデルの診断性能を、テストデータセットを用いて評価した。まず、前述の前処理を施したテストデータセットの冠状断 CT 画像を作成した CNN モデルに入力し、スライス毎に浸潤陽性または陰性に分類した。次に各病変において CNN モデルが浸潤陽性と判定した連続スライス枚数に基づいて、病変毎の浸潤陽性または陰性の診断に変換した。最後に、各病変における CNN モデルの診断結果と正解ラベルの浸潤陽性・陰性との整合性を確認した。

ここまでの訓練データセットを用いた転移学習による CNN モデル作成、並びにテストデータセットを用いた各病変性能評価の手順を図示すると以下のようになる (図 1-2)。

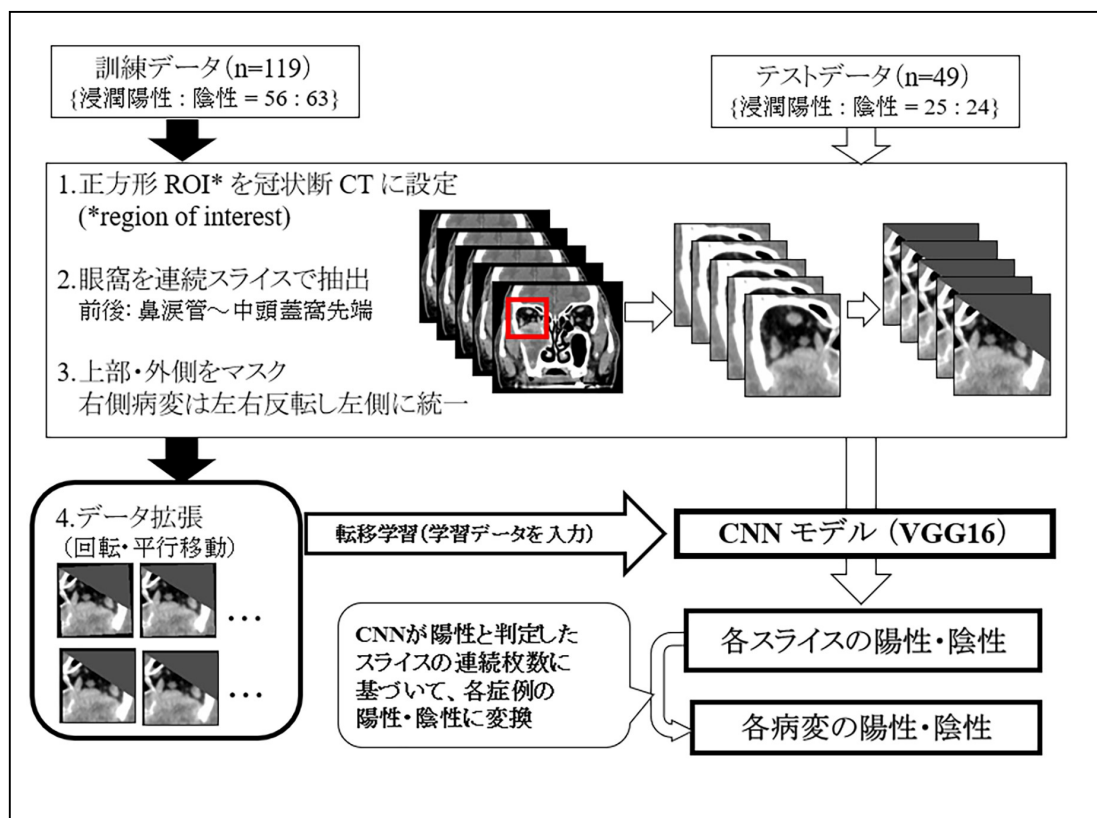


図 1-2 : 画像の前処理、転移学習、作成した CNN モデルの結果出力の流れ

画像の前処理は Metavol ソフトウェア(<https://www.metavol.org>) (Hirata et al., 2014) 、CNN モデルの作成および評価は MATLAB (R2021a, MathWorks, Natick, MA, USA)を用いて行った。転移学習による CNN モデルの作成とその性能評価に使用した PC のスペックを以下に記す : Ubuntu 18.04 long-term support (LTS)-based server with a Core i9 10980XE 18core/36thread 3.0-GHz central processing unit (CPU), four NVIDIA Quadro RTX8000 graphics processing unit (GPU) cards, and 128-GB (16GB×8) DDR4-2933 quad-channel memory。

6. 放射線科医による読影実験

頭頸部を専門としない 2 人の一般放射線科医（放射線診断の経験年数は 6 年と 3 年）が、テストデータセットに含まれる 49 病変の画像を独立して読影し、腫瘍による骨膜を越えた眼窩浸潤の有無を診断した。読影には正方形 ROI によって抽出されていない全視野の冠状断 CT 画像を参照した。

上記の読影実験から約 2 ヶ月後、同じ 2 人の一般放射線科医が CNN モデルによって提供された病変毎の診断結果を聞いた上でテストデータセットの全ての冠状断 CT 画像を再度見直し、浸潤陽性または陰性を診断した。その際、CNN モデルの診断性能は伝えていない。

7. Grad-CAM による視覚的評価

Grad-CAM (Gradient-weighted class activation mapping) は、CNN モデルが注目している部位の視覚的評価を強化するために、最も一般的に使用されている XAI 手法の 1 つである (Selvaraju et al., 2020) 。本研究では、畳み込み層から特徴マップを抽出して Grad-CAM を作成し、入力画像に応じてネットワークの注意が強調されている部位を示すヒートマップで視覚化を行った。テストデータセットの前処理済み CT 画像に Grad-CAM ヒートマップを用いることで、CNN モデルが腫瘍や眼窩内容物に注目できるかどうかを検証した。

8. 統計解析

病変の局在について、正解ラベルにおける浸潤陽性群と陰性群、および、訓練データセットとテストデータセットとの間での比較について χ 二乗検定を用いて偏りがあるかを検討した。 χ 二乗検定で有意差があったものについては、残渣分析によりどの局在に 2 群間の偏りがあるかを確認した。

浸潤陽性・陰性の正解ラベルの信頼性評価については、頭頸部を専門とする 2 名の放射線診断専門医の合議により決定された正解ラベル付けと同 2 名によ

る再読影の診断結果との一致度、および、頭頸部放射線診断の経験を持つ別の2名の合議による診断結果と正解ラベルとの一致度を評価するためにカッパ係数を用いた。カッパ係数が0.40未満は一致度が低い (poor)、0.41~0.57はまずまず (fair)、0.58~0.74は良い (good)、0.75以上は優れている (excellent) と解釈された (Pinto et al., 1988)。また、正解ラベル作成者と別の2名の合議による診断結果は正解ラベルをゴールドスタンダードとして ROC (receiver operating characteristic) 曲線における曲線下面積 (area under the curve : AUC) によっても評価された。

テストデータセットを用いて CNN モデルの診断性能を評価する際には、各病変において CNN モデルによって判定された浸潤陽性スライスの連続枚数をプロットして ROC 曲線を作成し、浸潤陽性スライスの連続枚数の最適なカットオフ値を Youden index から求め、病変毎の浸潤陽性・陰性の結果に変換した。この CNN モデルを用いて得られたテストデータセットの診断性能に加えて、2名の一般放射線科医によるテストデータセットの読影の診断性能、および、CNN モデルの診断結果を聞いた上での2名の一般放射線科医による再読影の診断性能を、以下の指標を算出してそれぞれ評価した：ROC-AUC、精度 (accuracy)、感度 (sensitivity)、特異度 (specificity)、陽性的中率 (positive predictive value)、陰性的中率 (negative predictive value)。AUCの比較には χ^2 検定を用いた。統計的な有意性は p 値 < 0.05 と定義した。すべての統計解析はエクセル統計 (BellCurve for Excel、株式会社社会情報サービス、東京) を用いて行った。

結果

1. 症例特性

対象となった鼻副鼻腔悪性腫瘍の症例 167 人（168 病変）の特徴は以下の通りであった。異時性多発癌を発症した 1 例については別の病変としてカウントした。年齢は悪性腫瘍が病理学的に診断された時点として決定された。

年齢の中央値（範囲）は 65 歳（29～91 歳）。男女比は 129 : 39。腫瘍の局在は上顎洞 117 例、鼻腔 30 例、篩骨洞 17 例、蝶形骨洞 2 例、上顎歯肉 2 例。組織型は扁平上皮癌 115 例、低・未分化癌 9 例、悪性黒色腫 9 例、腺様嚢胞癌 5 例、その他の癌 14 例（紡錘細胞癌、神経内分泌癌、腺癌、粘表皮癌など）、嗅神経芽細胞腫 6 例、非上皮性腫瘍 10 例（横紋筋肉腫など）。治療前の原発巣の臨床病期は、癌腫 142 例および悪性黒色腫 9 例については UICC の TNM 分類に基づき診断され、T1 が 2 例、T2 が 4 例、T3 が 41 例、T4a が 62 例、T4b が 42 例であった。嗅神経芽細胞腫 6 例については Kadish 分類（Kadish et al., 1976）に基づき診断され、グループ B が 1 例、グループ C が 5 例であった。初回治療選択については超選択的動注化学療法併用放射線療法が 96 例、手術が 21 例、化学放射線療法が 18 例、放射線療法単独が 18 例、全身化学療法が 10 例、無治療・best supportive care が 5 例であった。なお手術例については眼窩内容摘出術を併用された症例はなく、眼窩骨膜浸潤が病理学的に確認された症例はいなかった。

病変の局在について、正解ラベルにおける浸潤陽性群（81 例）と浸潤陰性群（87 例）との間での比較を表 1-1、訓練データセット（119 例）とテストデータセット（49 例）との間での比較を表 1-2 にまとめた。浸潤陽性群と陰性群との間には χ^2 二乗検定で有意差があった（ $p=0.007$ ）。残渣分析では上顎洞（ $p=0.004$ ）、鼻腔（ $p<0.001$ ）について有意差が認められ、篩骨洞（ $p=0.681$ ）、蝶形骨洞（ $p=0.170$ ）、上顎歯肉（ $p=0.959$ ）には有意差がなかった。訓練データセットとテストデータセットについては局在の偏りに有意差はなかった（ $p=0.863$ ）。

	総計 (n=168)	浸潤陽性 (n=81)	浸潤陰性 (n=87)
上顎洞	117	65	52
鼻腔	30	6	24
篩骨洞	17	9	8
蝶形骨洞	2	0	2
上顎歯肉	2	1	1

	総計 (n=168)	訓練データセット (n=119)	テストデータセット (n=49)
上顎洞	117	84	33
鼻腔	30	22	8
篩骨洞	17	11	6
蝶形骨洞	2	1	1
上顎歯肉	2	1	1

2. 正解ラベルの信頼性評価

正解ラベルを作成した頭頸部専門の放射線診断専門医 2 名が全病変の冠状断 CT 画像を再読影して診断した浸潤陽性・陰性の判定について、2 名の放射線科医間のカップパ係数は 0.769 (95%CI : 0.671-0.867)、正解ラベルと 1 人の放射線科医とのカップパ係数は 0.808 (95%CI : 0.720-0.897)、もう 1 人の放射線科医とのカップパ係数は 0.844 (95%CI : 0.764-0.925) であり、いずれも優れた一致度であった。

正解ラベル作成者と別の 2 名の放射線診断専門医の合議による診断結果と正解ラベルとのカップパ係数は 0.810 (95%CI : 0.722-0.898) であり、こちらも優れた一致度であった。

3. 転移学習による CNN モデルの作成

転移学習において、モデルの学習の反復回数と精度および損失の推移を示したグラフである学習曲線を以下に提示する（図 1-3）。訓練データセットを繰り返して学習していくに連れてモデルの精度が向上し、誤差が減少しているのが視覚的に確認できる。

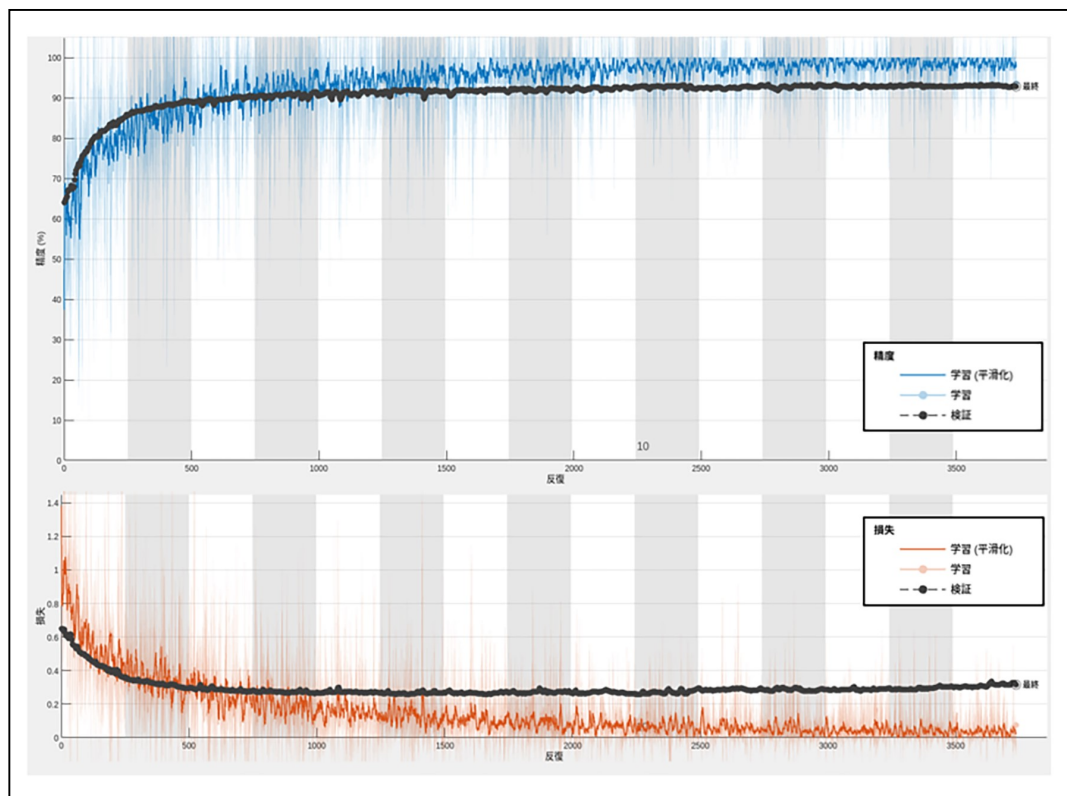


図 1-3：転移学習における学習曲線

横軸は学習の反復回数、上図の縦軸は精度、下図の縦軸は損失の進行状況が表示されている。上図の青線はミニバッチ毎の分類精度で黒点は検証データに対する分類精度、下図の赤線はミニバッチ毎の損失で黒点は検証データに対する損失を示す。

4. 診断性能の評価

(1) CNN モデルの診断性能

作成した CNN モデルに前処理済みのテストデータセットの画像を入力し、判定された浸潤陽性スライスの連続枚数を用いて、病変毎の浸潤陽性・陰性の診断に変換するための ROC 曲線解析を行った。ROC 曲線の AUC は 0.940

(95%CI : 0.873-1.000) となった。病変毎の浸潤陽性・陰性を区別する浸潤陽性スライスの連続枚数のカットオフ値は、ROC 曲線の Youden index に基づいて連続 3 枚以上となった。このカットオフ値を用いた場合、診断性能として以下の値が得られた：精度 0.922 (95%CI : 0.811-0.978)、感度 0.923 (95%CI : 0.749-0.991)、特異度 0.920 (95%CI : 0.740-0.990)、陽性的中率 0.923 (95%CI : 0.749-0.991)、陰性的中率 0.920 (95%CI : 0.740-0.990)。

(2) 一般放射線科医による読影実験の診断性能

CNN モデルによる補助のない最初の読影実験では、二人の頭頸部非専門の一般放射線科医の診断性能は以下のとおりであった。

読影医 1 : ROC-AUC 0.491 (95%CI : 0.351-0.631)、精度 0.490 (95%CI : 0.348-0.634)、感度 0.500 (95%CI : 0.291-0.709)、特異度 0.481 (95%CI : 0.287-0.681)、陽性的中率 0.462 (95%CI : 0.266-0.666)、陰性的中率 0.520 (95%CI : 0.313-0.722)。

読影医 2 : ROC-AUC 0.451 (95%CI : 0.311-0.590)、精度 0.451 (95%CI : 0.311-0.597)、感度 0.462 (95%CI : 0.266-0.666)、特異度 0.440 (95%CI : 0.244-0.651)、陽性的中率 0.462 (95%CI : 0.266-0.666)、陰性的中率 0.440 (95%CI : 0.244-0.651)。

次に、CNN モデルが出力した診断結果を知ったうえでの再読影では、両者はいずれも以下のように高い診断性能を達成した。

読影医 1 : ROC-AUC 0.941 (95%CI : 0.875-1.007)、精度 0.941 (95%CI : 0.838-0.988)、感度 0.962 (95%CI : 0.757-0.991)、特異度 0.920 (95%CI : 0.789-0.999)、陽性的中率 0.926 (95%CI : 0.804-0.999)、陰性的中率 0.958 (95%CI : 0.740-0.990)。

読影医 2 : ROC-AUC 1.000 (95%CI : 1.000-1.000)、精度 1.000 (95%CI : 1.000-1.000)、感度 1.000 (95%CI : 1.000-1.000)、特異度 1.000 (95%CI : 1.000-1.000)、陽性的中率 1.000 (95%CI : 1.000-1.000)、陰性的中率 1.000 (95%CI : 1.000-1.000)。

最適なカットオフ値（連続陽性スライス枚数3枚以上）を用いた場合の CNN モデルの診断性能、および、2人の頭頸部非専門の一般放射線科医の2回の読影実験における診断性能を表1-3にまとめた。

表 1-3. CNN モデルおよび一般放射線科医の診断性能					
	CNN	読影医 1 補助なし	読影医 1 補助あり	読影医 2 補助なし	読影医 2 補助あり
ROC-AUC	0.940	0.491	0.941	0.451	1.000
精度	0.922	0.490	0.941	0.451	1.000
感度	0.923	0.500	0.926	0.462	1.000
特異度	0.920	0.481	0.958	0.440	1.000
陽性的中率	0.923	0.462	0.962	0.462	1.000
陰性適中率	0.920	0.520	0.920	0.440	1.000

テストデータセットにおける CNN モデルから得られた ROC 曲線と、2人の頭頸部非専門の一般放射線科医による2回の読影実験（1回目：CNNモデルの補助なし、2回目：補助あり）から得られた感度と特異度の点を以下に図示する（図1-4）。CNNモデルのROC曲線はCNNモデルが判定した各病変の浸潤陽性スライスの連続枚数に基づいてプロットされている。CNNモデルのAUCは、2人の頭頸部非専門の一般放射線科医のAUCを有意に上回った（いずれも $p < 0.001$ ）。さらに、CNNモデルの補助を受けた場合のAUCは、補助なしの場合のAUCよりも有意に高かった（ $p < 0.001$ ）。

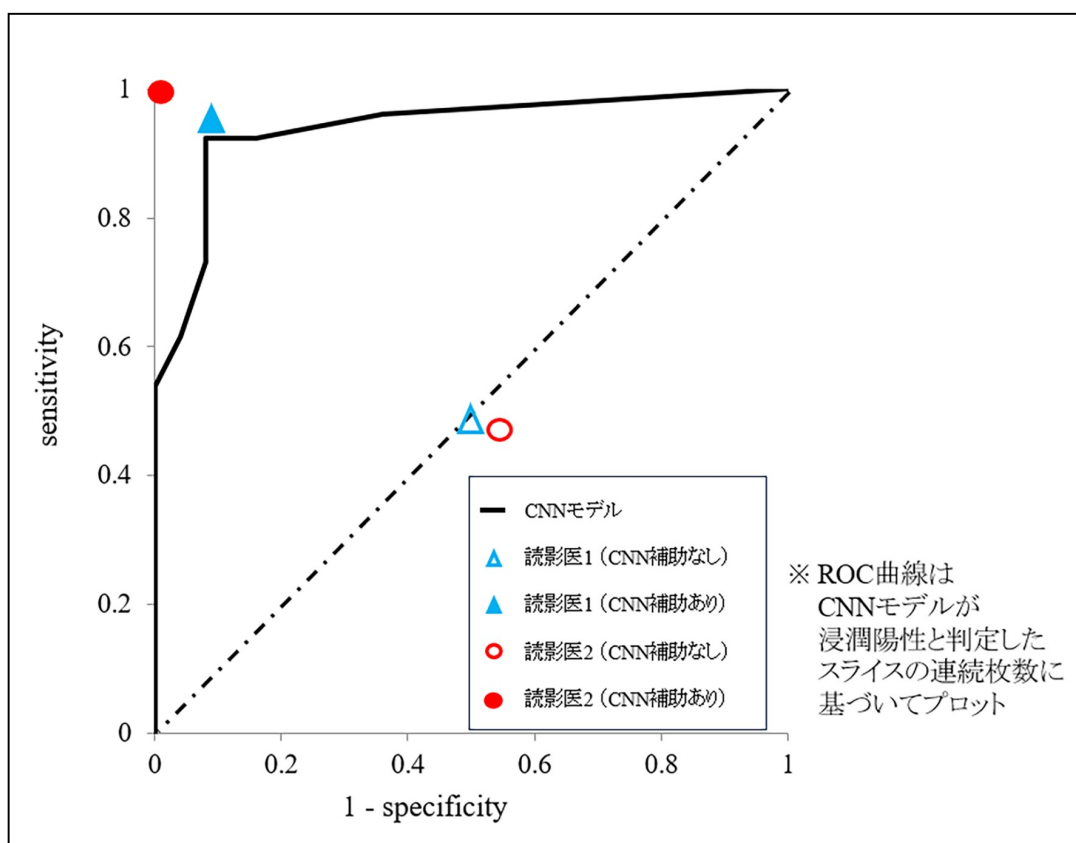


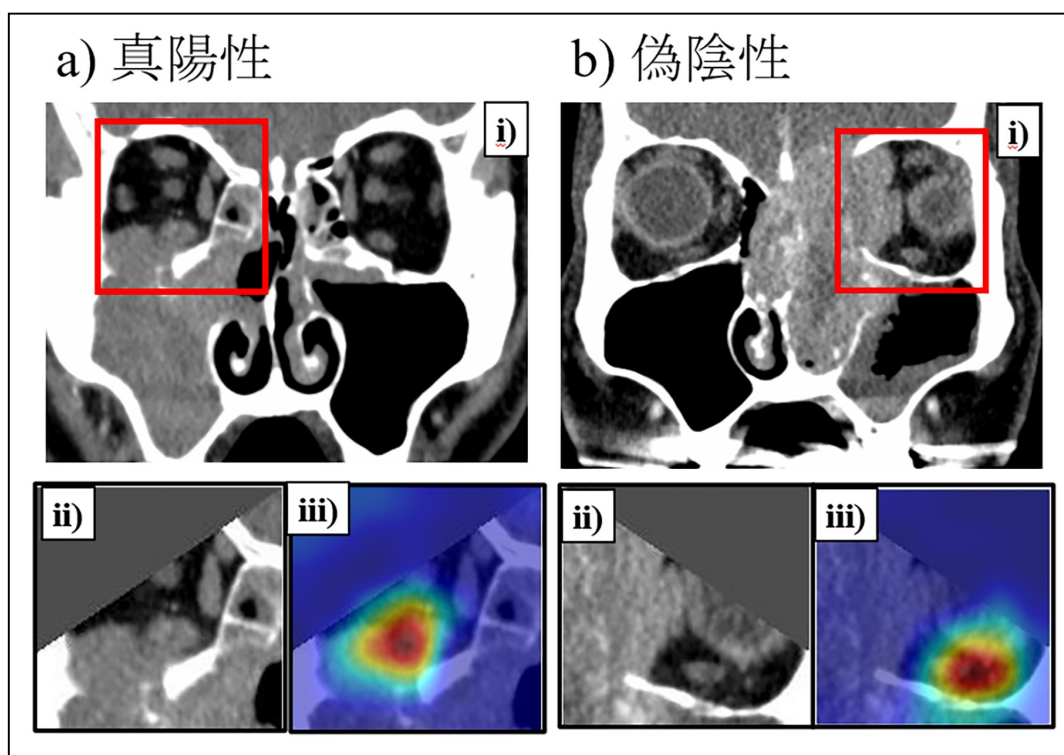
図 1-4 : CNN モデルの ROC 曲線と放射線科医の読影実験の結果

CNN モデルが眼窩骨膜を越えた浸潤陽性と判断した連続スライス数を、その病変における浸潤の有無の閾値として ROC 曲線をプロットした。CNN モデルによる補助のない 1 回目と補助のある 2 回目の読影実験における 2 人の一般放射線科医の感度と特異度を示す点も示した。

5. 代表症例と Grad-CAM による視覚的評価

最適なカットオフ値を用いた CNN モデルによるテストデータセット（合計 49 病変、浸潤陽性 25：陰性 24）の病変毎の浸潤陽性・陰性の診断において、真陽性が 23 例、偽陰性が 2 例、真陰性が 22 例、偽陽性が 2 例であった。

図 1-5 に CNN モデルによる a) 真陽性、b) 偽陰性、c) 真陰性、d) 偽陽性の代表的な症例を示した。真陽性症例のヒートマップは、眼窩へ突出した腫瘍と腫瘍周囲の眼窩内脂肪織の混濁に焦点を当てていた。偽陰性症例の腫瘍は眼窩内側壁に及んでいたが、ヒートマップ上のホットスポットは下壁周辺に観察された。真陰性症例の腫瘍は内側壁に及んでいたが、腫瘍の辺縁は明瞭であり、ヒートマップでは腫瘍辺縁と腫瘍周囲の眼窩脂肪に焦点が当てられていた。偽陽性症例では、腫瘍が眼窩に突出した部分にヒートマップ上で焦点が合っていたが、辺縁は平滑で、菲薄化した壁が確認でき、眼窩内脂肪織の混濁はなかった。



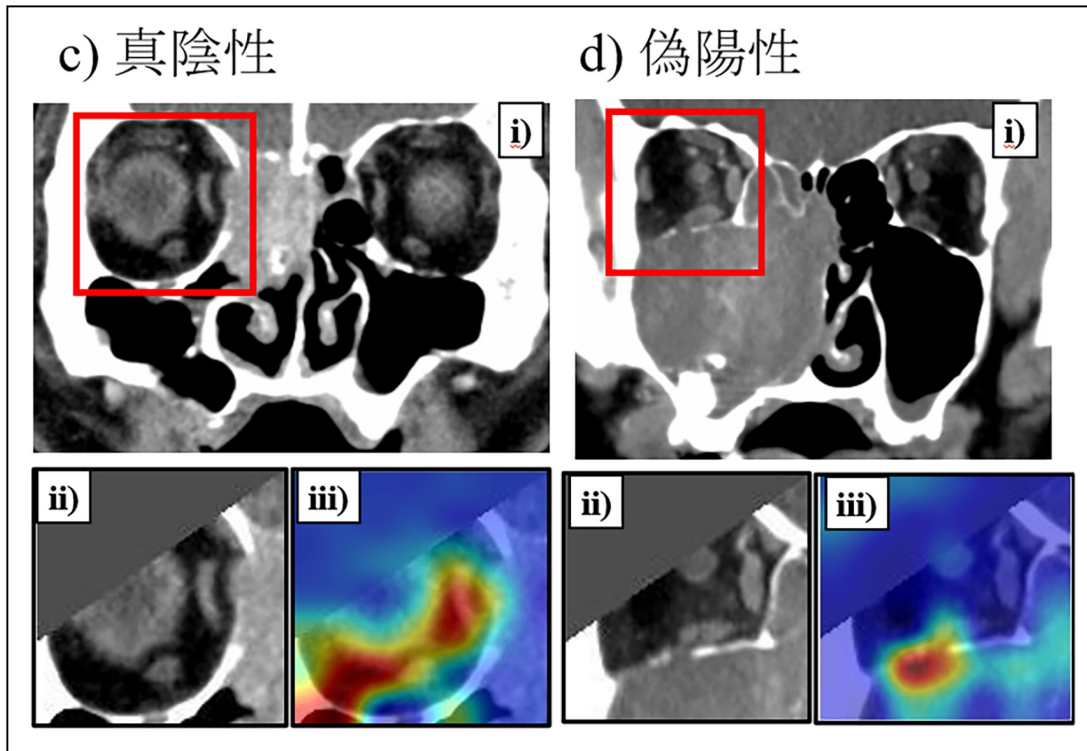


図 1-5：代表的な症例の画像

a)真陽性、b)偽陰性、c)真陰性、d)偽陽性の代表症例を提示する。

また各症例の画像は i) 元の冠状断 CT 画像（赤い四角は正方形 ROI の抽出範囲）、ii) 前処理済み CT 画像（抽出・マスク・左揃え）、iii) Grad-CAM ヒートマップをそれぞれ表示している。

(a) 真陽性例のヒートマップは眼窩へ突出した腫瘍と周囲の脂肪織混濁に注目していた。(b) 偽陰性例の腫瘍は眼窩内側壁に及んでいたが、ヒートマップで注目できていなかった。(c) 真陰性例の腫瘍は内側壁に及んでいたが辺縁は明瞭であり、ヒートマップは腫瘍辺縁と周囲の脂肪に注目していた。(d) 偽陽性例は腫瘍が眼窩に突出した部分にヒートマップが注目していたが、菲薄化した壁が確認でき脂肪織混濁はなかった。

考察

本研究では CT で観察された眼窩骨膜を越えた眼窩浸潤の放射線学的徴候を識別するために、DL 法の一つである CNN を応用した診断モデルを作成した。開発した CNN モデルによる診断性能は頭頸部を専門としない一般放射線科医の診断性能よりも高く、CNN モデルの補助により診断性能が向上した。このように、DL 技術は CT 画像上の眼窩骨膜を越えた眼窩浸潤の評価において、特に頭頸部画像診断の専門家ではない放射線科医にとって、診断支援ツールとなる可能性がある。この技術のさらなる発展により、鼻副鼻腔悪性腫瘍の正確な病期分類と適切な治療選択が可能となり、最終的に患者の予後改善につながる事が期待される。

検索した限り、頭頸部悪性腫瘍による眼窩浸潤の画像診断への DL 法の応用を評価した先行研究はない。CT 画像における悪性腫瘍の浸潤の有無を診断するための DL アプローチの有用性に関する種々の研究において (Choi et al., 2021; Zhang et al., 2021; Yang et al., 2021)、いずれも良好な診断性能を達成しており、経験豊富な放射線科医による診断と同等かそれ以上の結果が報告されているものもある。また、いくつかの研究では、DL に基づく診断モデルの支援により診断能が向上する支援ツールとしての DL 法の有用性が示されている (Kann et al., 2020; Duron et al., 2021; Li ZY et al., 2021)。本研究の結果は、これらの先行研究と同じ傾向の結果を示したと考えられる。

Grad-CAM は CNN モデルによる画像分類の根拠を視覚的、直感的に理解できることが実証されている。例えば、肺のすりガラス結節 (subsolid nodule) を 3 つのクラス (良性/前浸潤性病変、微小浸潤性腺癌、浸潤性腺癌) に分類した研究では、Grad-CAM ヒートマップは病変のエッジの形状、浸潤影の成分、気管支透亮像に焦点を当て、放射線科医が悪性結節を診断する際に評価すべき重要な特徴と一致した (Jiang et al., 2021)。今回の研究では Grad-CAM ヒートマップは眼窩へ突出した腫瘍と腫瘍周囲の眼窩内脂肪織に焦点を当てていた。これらの所見は日常臨床において眼窩骨膜を越えた眼窩内浸潤の有無を評価する際に放射線科医が注意すべきポイントと類似している。

本研究で作成した CNN モデルの診断において、2 例の偽陰性症例と 2 例の偽陽性症例が検出された。偽陰性症例は 2 例とも下方ではなく内側に病変があった。内側の病変は鼻腔、篩骨洞、蝶形骨洞の腫瘍であり、この局在の浸潤陽

性の病変数は全データセット 168 例中の 15 例 (9%) と、上顎洞・上顎歯肉の浸潤陽性の病変数 (66 例、39%) に比べて少なかった。かつ、正解ラベルの浸潤陽性群・陰性群における病変の局在分布の比較において、上顎洞病変は陽性例が多く、鼻腔病変は陰性例が多く、その偏りに有意差が確認された。そのため、CNN モデルの開発のための内側壁浸潤の画像所見によるトレーニングの量と質が十分でなかった可能性がある。一方、偽陽性症例は 2 例とも眼窩内にある程度突出していたが、病変の辺縁は明瞭であり、菲薄化した眼窩骨壁が確認でき、眼窩内脂肪織の不整はなかった。このような画像所見を有する症例は対象となった全データセット中の 13 例 (8%) と少なかったため、十分なトレーニングが行われなかった可能性がある。

本研究の限界として、単一施設による研究デザインのためサンプルサイズが小さい事、眼窩の連続スライスを抽出するための正方形 ROI は手動設定が必要である事、眼窩骨膜を超えた眼窩内浸潤の正解ラベルの判断基準には CT の所見のみを評価しており病理および MRI の所見を用いていない事があげられる。サンプルサイズについてはデータ拡張により限界を部分的に克服したと考える。正解ラベルの判断基準に関しては、本研究で対象となった症例のほとんどが放射線治療や化学療法、あるいはその組み合わせによって治療されており、手術例についても眼窩内容摘出術を併用した治療法は選択されておらず、病理学的な眼窩周囲浸潤の診断が確認できなかった。これは、当施設において進行した鼻副鼻腔悪性腫瘍に対する根治的治療法として化学放射線療法、特に上顎洞癌においては超選択的動注化学療法併用放射線療法が主に選択されるため、病理学的結果が得られにくいという特徴によるものである。また、眼窩骨膜浸潤の陽性・陰性の診断には MRI の所見が有用であることが報告されているが (Reyes et al., 2020)、本研究の対象患者においては MRI を受けなかったあるいは撮像条件や画質が不十分な症例が多く、正解ラベルの基準にできなかった。

第一章に続き、様々な進展部位における CNN の有用性を検証すべく、後述の第二章の研究では眼窩浸潤に並んで実臨床でよく遭遇する頭蓋底浸潤に対して CNN を応用した。その際には、上述のように第一章の研究では腫瘍による眼窩骨膜を超えた眼窩浸潤の CT 所見の有無という正解ラベルが頭頸部専門の放射線科医の意見に依存してしまうという問題があったため、頭蓋底浸潤の診断のゴールドスタンダードである MRI 所見も用いて正解ラベルを決定することで客観的な評価を重視することとした。

第二章 上咽頭悪性腫瘍による頭蓋底浸潤に関する CT 診断への深層学習の応用

緒言

1. 論文の背景

上咽頭腫瘍は内視鏡検査により比較的容易に検出および生検できるため病理学的な確定診断がしやすい。そのため、CT や MRI の最も重要な役割は、原発腫瘍の進展範囲および転移の有無を決定することである。上咽頭は頭蓋底の直下に位置するため、上咽頭腫瘍はしばしば頭蓋底に浸潤する。上咽頭癌の頭蓋底への浸潤は UICC の臨床 TNM 病期分類基準第 8 版によると T3 に分類されており (Huang et al., 2017)、局所再発のリスクが高く、生存率が低いことを示す予後因子と考えられている (Roh et al., 2004)。また、本邦の頭頸部癌診療ガイドライン 2022 年版において、いずれのステージングでも放射線治療と化学療法の組み合わせによって治療されるため、原発巣の局所評価については手術の可否ではなく放射線照射範囲の決定に重要である (日本頭頸部癌学会, 2022)。

CT では頭蓋底の浸透性またはびらん性の骨変化や骨孔に沿った広がりを含め、また骨髄の濃度上昇・硬化性変化や皮質骨の肥厚も検出されることがある (Razek et al., 2012)。MRI では腫瘍浸潤によって T1 強調像 (T1 weighted image : T1WI) において高信号を示す骨髄が低信号に置換され、ガドリニウム造影剤によって増強効果が認められ (Chen et al., 2011)、脂肪抑制 T2 強調像 (fat-saturation T2 weighted image : FsT2WI) で高信号を認めることも、骨髄浸潤の診断に有用である。上咽頭癌と比較して、上咽頭の悪性リンパ腫は、深部浸潤の頻度や範囲が少ない傾向があるが (King et al., 2003)、浸潤の所見自体は上咽頭癌と同様である。

一般に、骨びらんの描出では MRI よりも CT の方が優れているのに対し、軟部組織の異常の描出では MRI の方が優れていることが示されている (Chong et al., 1996)。また、頭蓋底浸潤の検出に関しては、CT よりも MRI の方が感度は高いという報告がいくつかある (Chong et al., 1996; Ng et al., 1997)。しかしながら、別の研究では、MRI のみで示唆される頭蓋底浸潤が

必ずしも局所再発の高リスクと関連するとは限らないと報告している (Nishioka et al., 2000)。したがって、上咽頭腫瘍による頭蓋底浸潤が疑われる症例では、CT と MRI の両方の所見を確認することが重要である。

MRI は撮影時間が長く、特に痛みや呼吸困難などの症状がある患者では、モーションアーチファクトによる画質の低下が起こる可能性がある。閉所恐怖症、金属インプラント、心臓ペースメーカーなどの電子デバイスのために MRI を受けられない人もいる。そのような場合には、CT を用いて頭蓋底浸潤の有無を評価する必要がある。

第一章の研究から鼻副鼻腔悪性腫瘍の眼窩骨膜を越えた浸潤の CT 診断について CNN モデルが有用であると考えられ、上咽頭悪性腫瘍による頭蓋底浸潤の CT 診断についても同様の手法を適応することで有用性を検証した。第一章では病理組織学的評価や MRI のような CT 以外のモダリティを用いた正解レベルの客観的な裏付けが得られていない点が研究の限界の一つであったため、第二章では実臨床で頭蓋底浸潤の診断のゴールドスタンダードとなっている MRI 所見も併用することで正解レベルにより客観性を持たせることとした。また、第一章で示したもう一つの限界である ROI の手動設定については、設定者の違いによる CNN モデルの診断性能の変化を評価した。

方法

1. 対象

本研究は、北海道大学病院自主臨床研究審査委員会での承認を得て行われた（研究番号：021-0001）。

診療録に基づいて、2008年1月から2022年3月までに当院を受診した上咽頭腫瘍患者115例が、以下の選択基準に基づいて組み入れられた：(1)病理学的に悪性上咽頭腫瘍が確認されている、(2)治療前のCTで腫瘍病変を含む軟部条件の水平断が作成されている。一部は以下の除外基準により除外された：(1)CTのスライス厚が4mm以上と厚い（11例）、または(2)MRIが実施されていない（4例）。最終的に100人の患者が本研究の対象となった。

2. CTおよびMRI画像

100症例のCT画像が、4社の様々なCTスキャナーで撮像された（主な装置：Aquilion PRIME、東芝メディカルシステムズ、大田原。Sensation64、SIEMENS Medical Solutions、Malvern。Brilliance iCT 256、Philips Medical Systems、Best。Optima CT580、GE Healthcare、Milwaukee）。100症例のうち、7例は非造影画像のみ、76例は造影画像のみ、残りの17例は非造影画像と造影画像の両方が撮像されていた。評価には軟部条件の水平断CT画像を用いた。スライス厚は1.25~4mm、主なマトリックスサイズは512×512であった。

MRIは5社の様々なスキャナーで撮像された（主な装置：TRILLIUM OVAL 3.0T、日立メディコ、東京。ACHIEVA 3.0T、Philips Medical Systems、Best。MAVNETOM Avanto 1.5T、Siemens Medical Solutions、Malvern。Discovery MR750w 3.0T、GE Healthcare、Milwaukee。Vantage 1.5T、東芝メディカルシステムズ、大田原）。MRIはCTの前後2か月以内に撮像され、T1WI、FsT2WI、ガドリニウム造影T1強調像（Gadolinium-enhanced T1WI：GdT1WI）または脂肪抑制ガドリニウム造影T1強調像（FsGdT1WI）のうち少なくとも2つのシーケンスが用いられた。100症例のうち、6例は非造影画像のみ、残りの94例は非造影画像と造影画像の両方が撮像されていた。撮影方向は水平断に加えて、冠状断または矢状断の少なくとも一方が撮像された。スライス厚は1~5mm、主なマトリックスサイズ512×512であった。

3. 頭蓋底浸潤の正解ラベル付け

頭頸部放射線診断を専門とする2名の放射線診断専門医（放射線診断の経験年数は7年と16年）がDICOMビューア（XTREK、J-MAC SYSTEM、札幌）を用いて、CTおよびMRI画像が読影に適切かどうかを判断した。頭蓋底浸潤の所見の有無を評価する範囲は、水平断における蝶形骨洞の底部から後頭骨の斜台下端までとした。頭蓋底浸潤の判定は、画像診断については斜台の変化を中心に読影され、MRI所見（骨髄がT1WIで低信号、FsT2WIで高信号、造影後に増強される）の有無、CT所見（浸透性またはびらん性の骨変化や骨破壊、および骨硬化）の有無、これらに加えて患者のすべての利用可能な医療記録に基づき、前述の2名の頭頸部専門の放射線診断専門医が合議により評価した。この画像評価の結果、CTとMRIの両方で浸潤の所見を確認できた42例が浸潤陽性、残りの53例が浸潤陰性と診断された。

4. 画像処理と深層学習

(1) 診断モデル作成のための症例の分割

収集した100症例を、浸潤陽性群（42例）、浸潤陰性群（53例）のそれぞれについて約7：3の割合となるように無作為に分配した。うち約7割の70病変（浸潤陽性29：陰性41）を診断モデル作成のための訓練データセットとして、残りの30病変（浸潤陽性13：陰性17）を作成したモデルの性能を評価するためのテストデータセットとして使用した。

以下に、症例選択からデータセットの分配までの流れを図示する（図 2-1）。

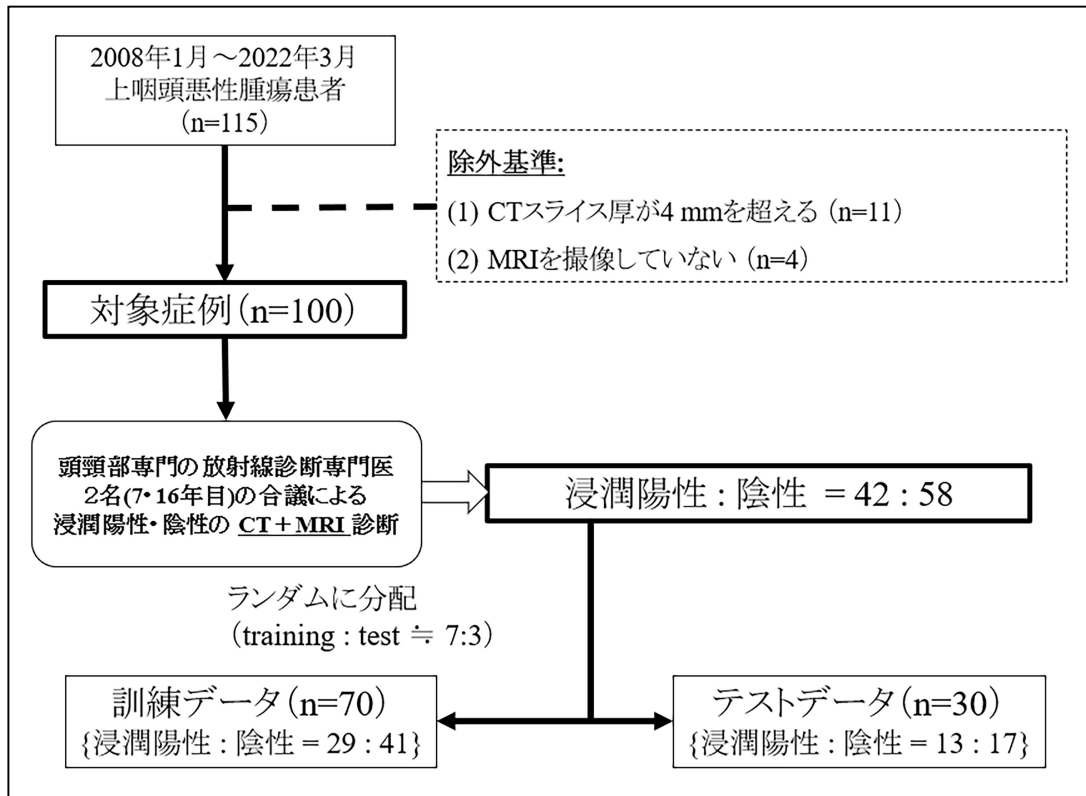


図 2-1：症例選択、正解ラベル付け、データセットの分配の流れ

(2) 画像の前処理

画像の前処理としてすべての水平断 CT 画像について病変の以下のように抽出を行った。まず、上咽頭と頭蓋底（特に斜台）を十分含むように長方形の ROI を手動で設定した。長方形の ROI の範囲は、前後方向が翼状突起から斜台後端、左右方向が両側内頸動脈を目印にして設定した。上下方向の範囲は前述のラベル付け時と同様に蝶形骨洞底部から斜台下端までを使用し、長方形の連続画像として抽出した。すべての画像の CT ウィンドウはウィンドウレベル 60 HU、ウィンドウ幅 600 HU に調整した。最後に、各々の前処理済み画像を JPEG 形式として出力した。なお、非造影 CT と造影 CT のどちらも撮像された症例についてはどちらも別々に前処理を行い、画像を抽出した。

(3) データセットの画像選択とデータ拡張

訓練データセットの浸潤陽性の 29 症例（うち、非造影 CT のみ：1 例、造影 CT のみ：24 例、非造影 CT と造影 CT の両方：4 例）については、前述の正解ラベル付けを行った 2 名の頭頸部専門の放射線診断専門医の合議によって、CT と MRI の両方で頭蓋底浸潤の所見があると判定したスライス（以下、浸潤陽性スライス）のみを用いることとし、146 枚の画像が選択された。一方で浸潤陰性の 41 症例（うち、非造影 CT のみ：4 例、造影 CT のみ：31 例、非造影 CT と造影 CT の両方：6 例）については、抽出されたすべてのスライスを用いることとし、226 枚の画像が選択された。また、モデルの汎化性能を向上させるために、各画像に対して水平方向の反転、ランダムな回転と、垂直方向および水平方向の平行移動によるデータ拡張が行われ、各画像に対して 21 枚の追加画像が生成された。最終的に訓練データセットは、浸潤陽性群の 3,212 画像と陰性群の 4,972 画像から構成された。

テストデータセットについては、浸潤陽性の 13 症例（うち非造影 CT のみ：1 例、造影 CT のみ：10 例、非造影 CT と造影 CT の両方：2 例）と、浸潤陰性の 17 症例（うち、非造影 CT のみ：1 例、造影 CT のみ：11 例、非造影 CT と造影 CT の両方：5 例）の両群におけるすべての抽出画像を使用し、総計 168 枚のスライスが得られた。なお、テストデータセットは、データ拡張なしで用いられた。

(4) 深層学習によるモデル作成と性能評価

第一章の研究と同様、画像分類に特化した事前学習済み CNN モデルを用いた転移学習により、水平断 CT 画像における頭蓋底浸潤陽性または陰性の分類に適応した CNN モデルの作成を試みた。本研究では、第一章で使用した VGG16 のほか、系統の異なる事前学習済み CNN モデルである ResNet50 (Residual Networks 50) や Xception も使用し、訓練データでの検証精度が高いものを選択することとした。

ResNet は 2015 年の ILSVRC で優勝したモデルで、ある層の入力を数個の畳み込み層を経た後の出力にショートカットして足しあわせるスキップ接続 (skip connection) を用いており、このスキップ接続を用いた層の集合を残差ブロック (residual block) と呼ぶ。この残差ブロックをつなげることで深い層を構築することが可能となったモデルである (He et al., 2015)。ResNet50 は、1 つの 3×3 畳み込み層、1 つのマックス・プーリング層、16 の残差ブロックを含み、各残差ブロックは、1 つの 1×1 畳み込み層、1 つの 3×3 畳み込み層、1

つの 1×1 畳み込み層を含んでいる (Li Y et al., 2021)。

Xception は 2014 年の ILSVRC で優勝した GoogleNet の派生モデルである。GoogleNet は Inception モジュールという、一つの入力に対して複数の畳み込み層 (1×1、3×3、5×5 など) を並列に適応し、結果を連結して出力する構造を用いている。Xception ではこれを応用し、入力に対して 1×1 畳み込み層でチャンネル数を削減した後で、チャンネル数と同じ数の 3×3 畳み込み層を並列に適応することで計算を軽くしたモデルである (Chollet F., 2016)。

CNN モデルの学習において、最後の全連結層の重みを今回作成した画像データによる学習によって変更し、他の先行層のパラメータは元の事前学習済みモデルの重みに固定した。これにより、各モデルの汎用的な特徴を維持し、限られた数の訓練可能なパラメータによってモデルを CT 画像に適応させることができた。本研究では入力した画像について浸潤陽性または陰性の 2 値分類としてモデルが出力するようにした。

転移学習のハイパーパラメータは 15 エポック、ミニバッチサイズ 64、学習率 1.0×10^{-4} に設定し、最適化アルゴリズムは adaptive moment (adam) を使用した。訓練データのうち 30% を検証データとして使用し、各ミニバッチにおける精度を評価した。

作成した CNN モデルの診断性能を、テストデータセットを用いて評価した。まず、前述の前処理を施したテストデータセットの水平断 CT 画像を作成した CNN モデルに入力し、スライス毎に浸潤陽性または陰性に分類した。次に各病変において CNN モデルが浸潤陽性と判定した連続スライス枚数に基づいて、病変毎の浸潤陽性または陰性の診断に変換した。最後に、各症例における CNN モデルの診断結果と正解ラベルの浸潤陽性・陰性との整合性を確認した。

ここまでの訓練データセットを用いた転移学習による CNN モデル作成、並びにテストデータセットを用いた性能評価の手順を図示すると以下のようになる (図 2-2)。

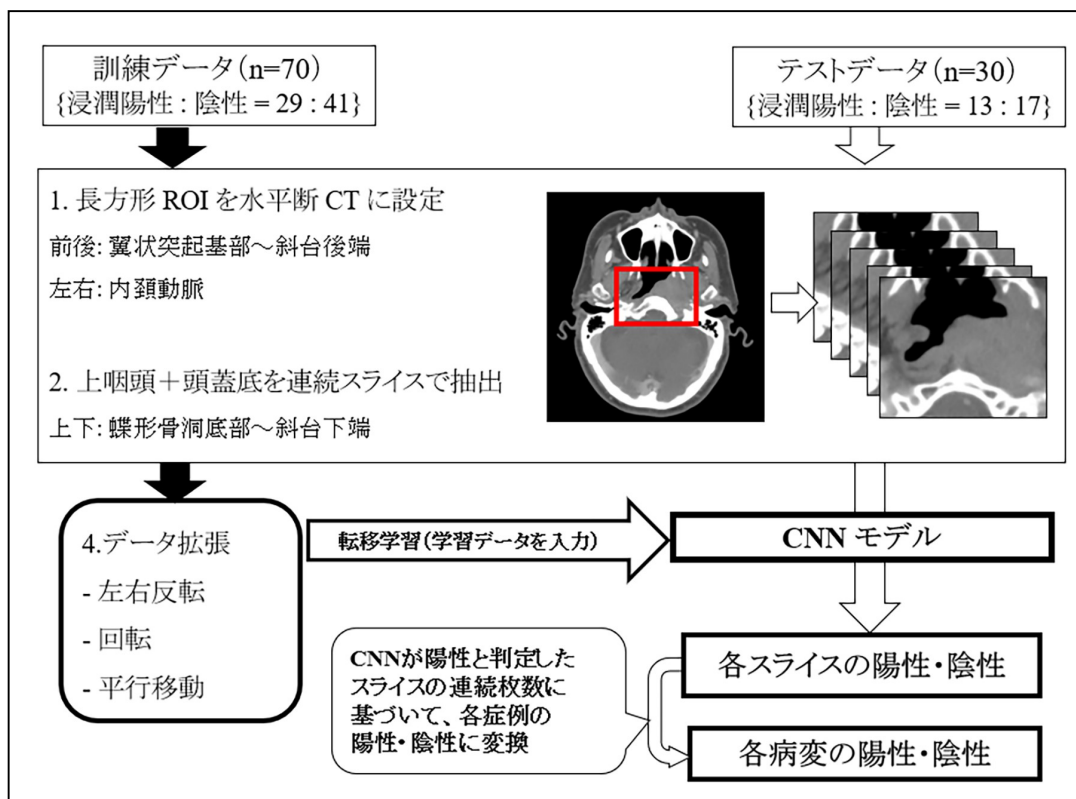


図 2-2 : 画像の前処理、転移学習、作成した CNN モデルの結果出力の流れ

画像の前処理は Metavol ソフトウェア(<https://www.metavol.org>) (Hirata et al., 2014)、CNN モデルの作成および評価は MATLAB (R2021a, MathWorks, Natick, MA, USA)を用いて行った。転移学習による CNN モデルの作成とその性能評価に使用した PC のスペックを以下に記す: Ubuntu 18.04 long-term support (LTS)-based server with a Core i9 10980XE 18core/36thread 3.0-GHz central processing unit (CPU), four NVIDIA Quadro RTX8000 graphics processing unit (GPU) cards, and 128-GB (16GB×8) DDR4-2933 quad-channel memory。

5. 放射線科医による読影実験

頭頸部画像診断の経験が豊富な放射線診断専門医 1 人 (以下、SR : Senior reader、放射線診断の経験年数 13 年) と経験の浅い放射線科医 1 人 (以下 JR : Junior reader、放射線診断の経験年数 1 年) が、テストデータセットに含まれる 30 病変の画像を独立して読影し、腫瘍による頭蓋底浸潤の有無を診断した。読影には長方形 ROI によって抽出されていない全視野の水平断 CT 画像を参照した。

6. Grad-CAMによる視覚的評価

第一章の研究と同様に、本研究でも畳み込み層から特徴マップを抽出して Grad-CAM を作成し、入力画像に応じてネットワークの注意が強調されている部位を示すヒートマップで視覚化を行った。テストデータセットの前処理済み CT 画像に Grad-CAM ヒートマップを用いることで、CNN モデルが腫瘍や頭蓋底に注目できるかどうかを検証した。

7. 別の放射線科医による ROI 設定実験

長方形 ROI を設定する人の違いによる CNN モデルの診断結果の変化を評価した。正解ラベル決定や読影実験を行った放射線科医とも異なる、放射線科医（経験年数 2 年）が病変抽出の範囲（”4-（2）画像の前処理”を参照）の指導を受けた後、テストデータセットの水平断 CT 画像に長方形 ROI を設定させた。この抽出した画像を、作成した CNN モデルに入力し、前述の手法に沿って診断性能を評価し、オリジナルの ROI 設定による抽出画像による診断性能と比較した。

8. 統計解析

テストデータセットを用いて CNN モデルの診断性能を評価する際には、各病変において CNN モデルによって判定された浸潤陽性スライスの連続枚数をプロットして ROC 曲線を作成し、浸潤陽性スライスの連続枚数の最適なカットオフ値を Youden index から求め、病変毎の浸潤陽性・陰性の結果に変換した。この CNN モデルを用いて得られたテストデータセットの診断性能に加えて、2 名の放射線科医（SR と JR）によるテストデータセットの読影の診断性能を、以下の指標を算出してそれぞれ評価した：ROC-AUC、精度、感度、特異度、陽性的中率、陰性的中率。AUC の比較には χ^2 検定を用いた。別の放射線科医による ROI 設定実験による診断性能についても ROC 曲線を用いて比較した。統計的な有意性は p 値 <0.05 と定義した。すべての統計解析はエクセル統計（BellCurve for Excel、株式会社社会情報サービス、東京）を用いて行った。

結果

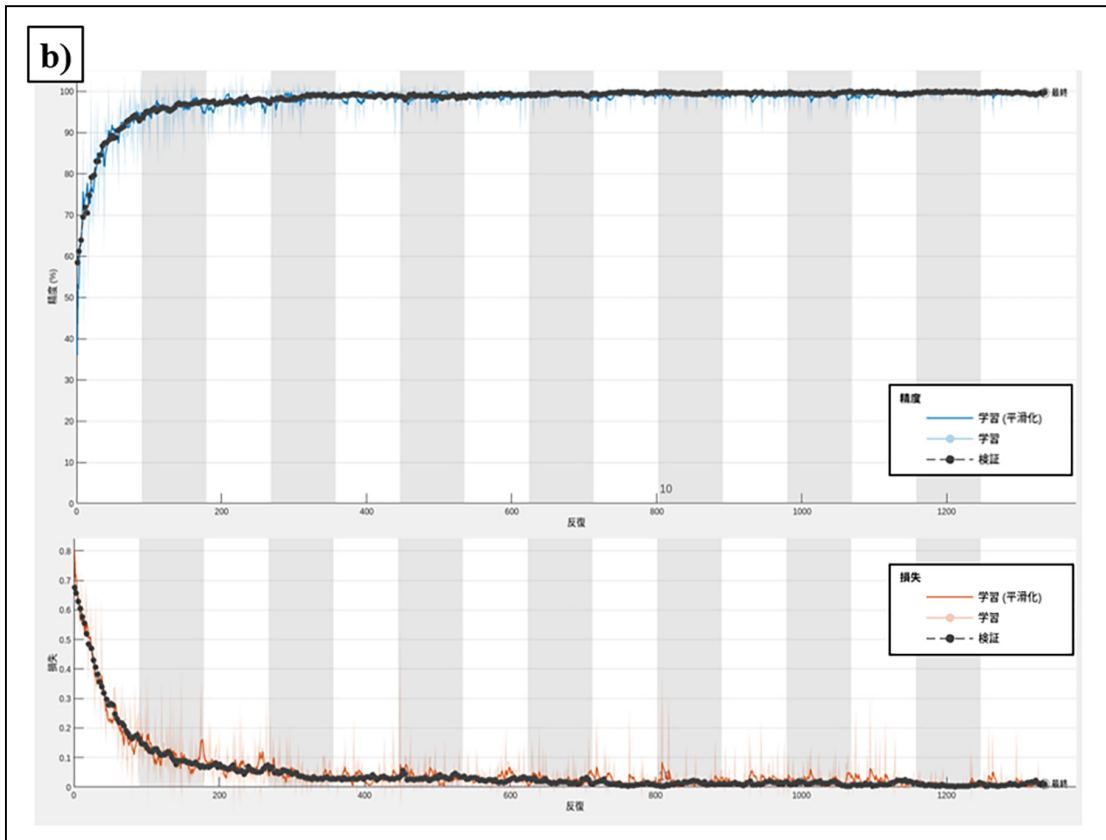
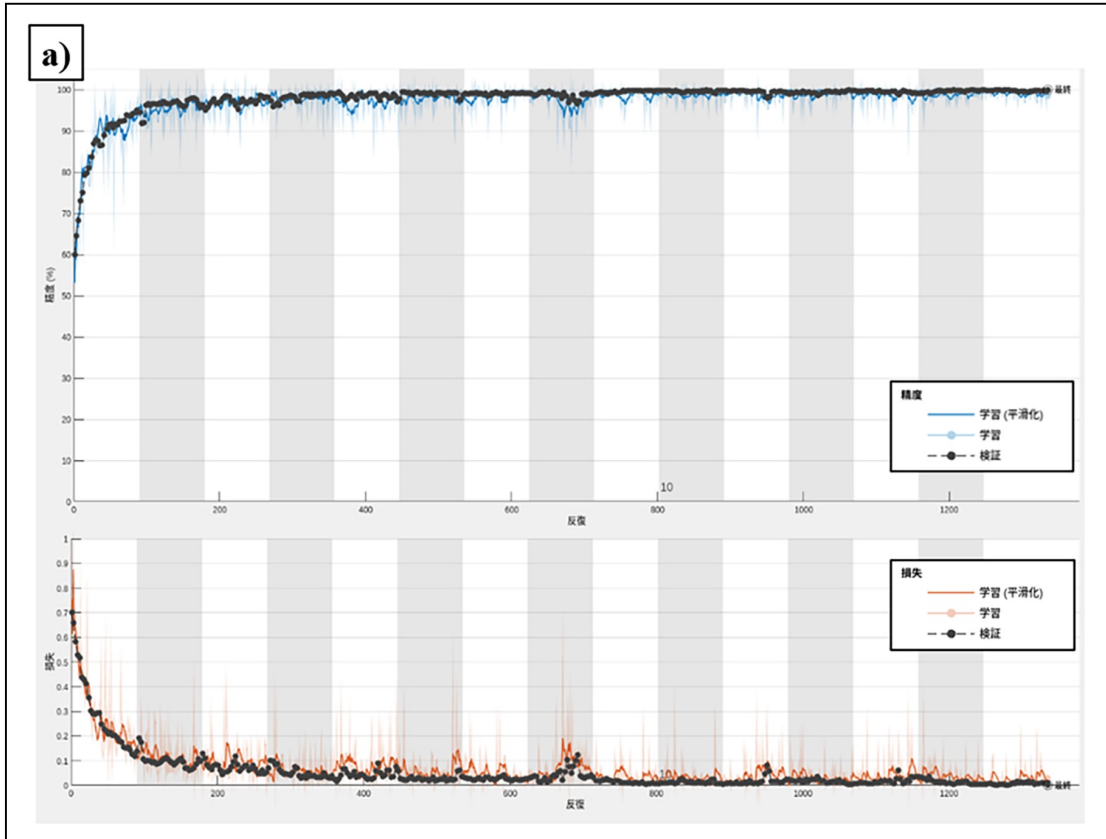
1. 症例特性

対象となった上咽頭悪性腫瘍の症例 100 人の特徴は以下の通りであった。年齢は悪性腫瘍が病理学的に診断された時点として決定された。

年齢の中央値（範囲）は 60 歳（18～85 歳）、男女比は 78：22。組織型は上咽頭癌が 94 例で、そのうち 89 例は扁平上皮癌（低～未分化型およびリンパ上皮癌を含む）に分類され、残りの 5 例は腺癌などの他の組織型に分類された。残る 6 例は悪性リンパ腫であった。治療前の原発巣の臨床病期は、上咽頭癌については UICC の TNM 分類に基づき診断され、T1 が 26 例、T2 が 18 例、T3 が 27 例、T4 が 22 例であった。

2. 転移学習による CNN モデルの作成と選択

転移学習において、モデルの学習の反復回数と精度および損失の推移を示したグラフである学習曲線を以下に提示する（図 2-3）。いずれのモデルも訓練データセットを繰り返して学習していくに連れてモデルの精度が向上し、誤差が減少しているのが視覚的に確認できる。訓練データのうち 3 割を検証データとして用いた際の最終検証精度は ResNet50 が 1.000、Xception が 0.997、VGG16 が 0.998 であり、より高い精度が得られた ResNet50 を用いた CNN モデルを後述の診断性能評価に使用することとした。



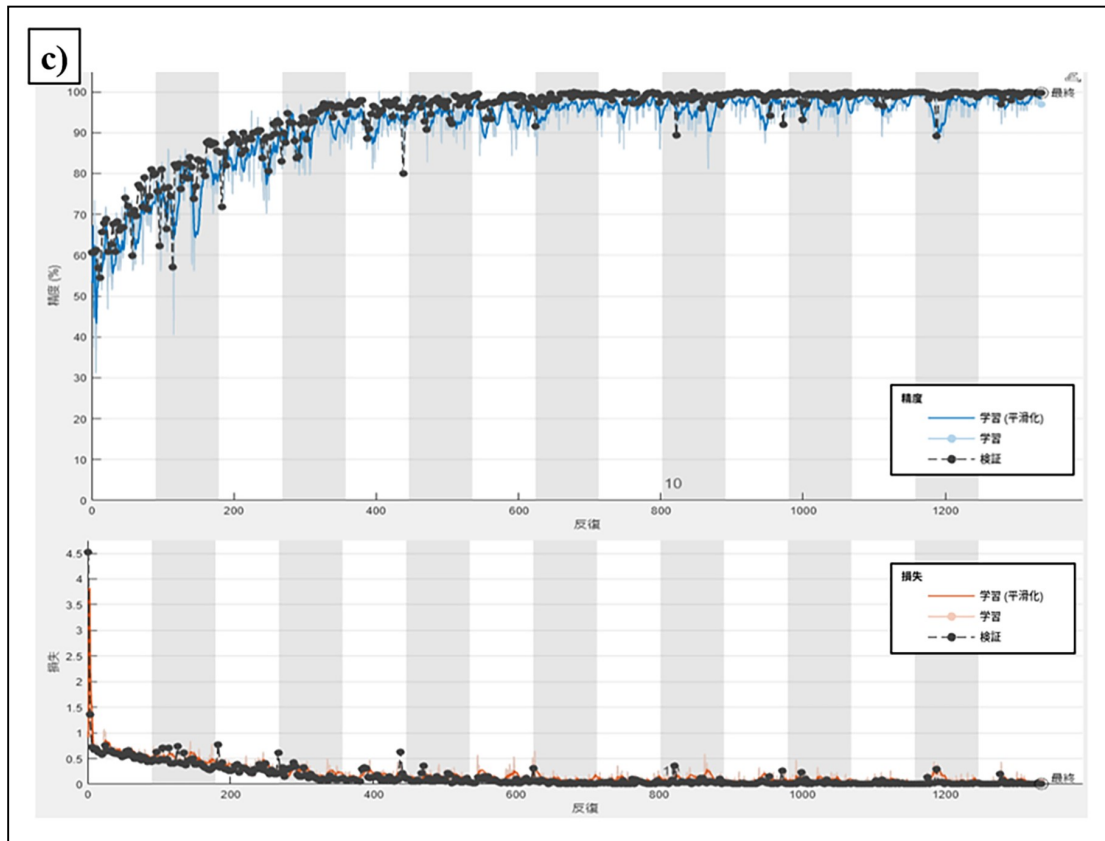


図 2-3：転移学習における学習曲線

a) ResNet50、b) Xception、c) VGG16

横軸は学習の反復回数、上図の縦軸は精度、下図の縦軸は損失の進行状況が表示されている。上図の青線はミニバッチ毎の分類精度で黒点は検証データに対する分類精度、下図の赤線はミニバッチ毎の損失で黒点は検証データに対する損失を示す。

3. 診断性能の評価

(1) CNN モデルの診断性能

作成した CNN モデルに前処理済みのテストデータセットの画像を入力し、判定された浸潤陽性スライスの連続枚数を用いて、病変毎の浸潤陽性・陰性の診断に変換するための ROC 曲線解析を行った。ROC 曲線の AUC は 0.953

(95%CI : 0.861-1.045) となった。病変毎の浸潤陽性・陰性を区別する浸潤陽性スライスの連続枚数のカットオフ値は、ROC 曲線の Youden index に基づいて連続 2 枚以上となった。このカットオフ値を用いた場合、診断性能として以下の値が得られた：精度 0.973 (95%CI : 0.858-0.999)、感度 0.933 (95%CI : 0.681-0.998)、特異度 1.000 (95%CI : 1.000-1.000)、陽性的中率 1.000 (95%CI : 1.000-1.000)、陰性的中率 0.957 (95%CI : 0.781-0.999)。

(2) 放射線科医による読影実験の診断性能

頭頸部画像診断の経験豊富な放射線診断専門医 (SR) と経験の乏しい放射線科医 (JR) の診断性能は以下のとおりであった。

SR : ROC-AUC 0.832 (95%CI : 0.704-0.960)、精度 0.833 (95%CI : 0.653-0.944)、感度 0.769 (95%CI : 0.462-0.950)、特異度 0.882 (95%CI : 0.636-0.985)、陽性的中率 0.833 (95%CI : 0.516-0.979)、陰性的中率 0.833 (95%CI : 0.586-0.964)。

JR : ROC-AUC 0.617 (95%CI : 0.459-0.774)、精度 0.600 (95%CI : 0.406-0.773)、感度 0.692 (95%CI : 0.386-0.909)、特異度 0.529 (95%CI : 0.278-0.770)、陽性的中率 0.529 (95%CI : 0.278-0.770)、陰性的中率 0.692 (95%CI : 0.386-0.909)。

最適なカットオフ値 (連続陽性スライス枚数 2 枚以上) を用いた場合の CNN モデルの診断性能、および、2 人の放射線科医の読影実験における診断性能を表 2-1 にまとめた。

表 2-1. CNN モデルおよび放射線科医の診断性能			
	CNN モデル	SR	JR
ROC-AUC	0.953	0.832	0.617
精度	0.973	0.833	0.600
感度	0.933	0.769	0.692
特異度	1.000	0.882	0.529
陽性的中率	1.000	0.833	0.529
陰性的中率	0.957	0.833	0.692

図 2-4 にテストデータセットにおける CNN モデルから得られた ROC 曲線と、2 人の放射線科医の読影実験から得られた感度と特異度の点を示す。CNN モデルの ROC 曲線は CNN モデルが判定した各病変の浸潤陽性スライスの連続枚数に基づいてプロットされている。AUC は CNN モデル、SR、JR の順で大きく、これらの値の間にも有意差があった（いずれも $p < 0.05$ ）。

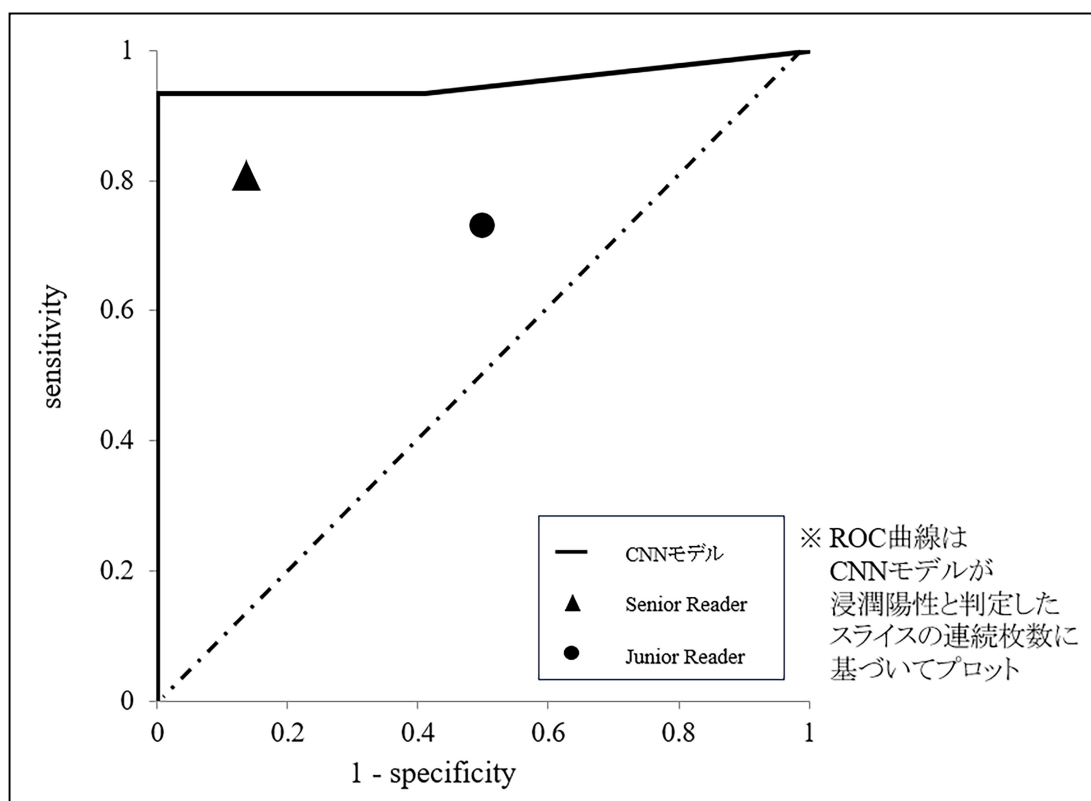


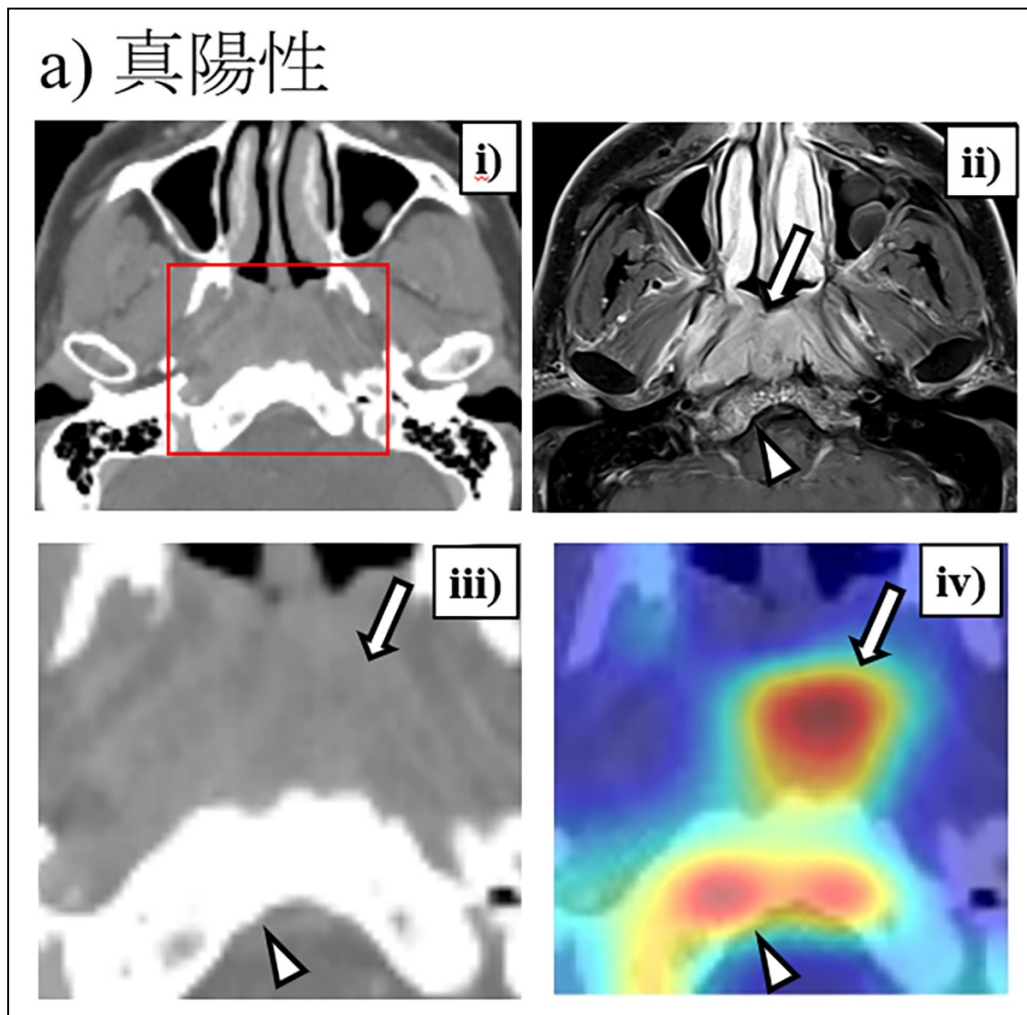
図 2-4 : CNN モデルの ROC 曲線と放射線科医の読影実験の結果

CNN モデルが頭蓋底浸潤陽性と判断した連続スライス数を、その病変における浸潤の有無の閾値として ROC 曲線をプロットした。頭頸部画像診断の経験が豊富な放射線診断専門医（Senior reader）と経験の少ない放射線科医（Junior Reader）の読影実験における感度と特異度を示す点も示した。

4. 代表症例と Grad-CAM による視覚的評価

最適なカットオフ値を用いた CNN モデルによるテストデータセット（合計 30 病変、浸潤陽性 13：陰性 17）の病変毎の浸潤陽性・陰性の診断において、真陽性が 12 例、偽陰性が 1 例、真陰性が 17 例で、偽陽性はいなかった。偽陰性の 1 例は MRI で斜台の造影効果に相当する所見が CT 上は僅かな骨硬化しか認めない症例だった。

図 2-5 に CNN モデルによる真陽性、真陰性の代表的な症例の CT と MRI 画像と Grad-CAM ヒートマップの組み合わせを示した。浸潤陽性例のうち正しく陽性と診断できた真陽性の 12 例は、ヒートマップは主に上咽頭の原因腫瘍や斜台の骨硬化ないし骨破壊に注目していた。浸潤陰性例は 17 例の全てを正しく陰性と診断できており、程度の差はあるもののヒートマップは主に骨硬化や軟部組織濃度の乏しい頭蓋底の骨髄に注目していた。



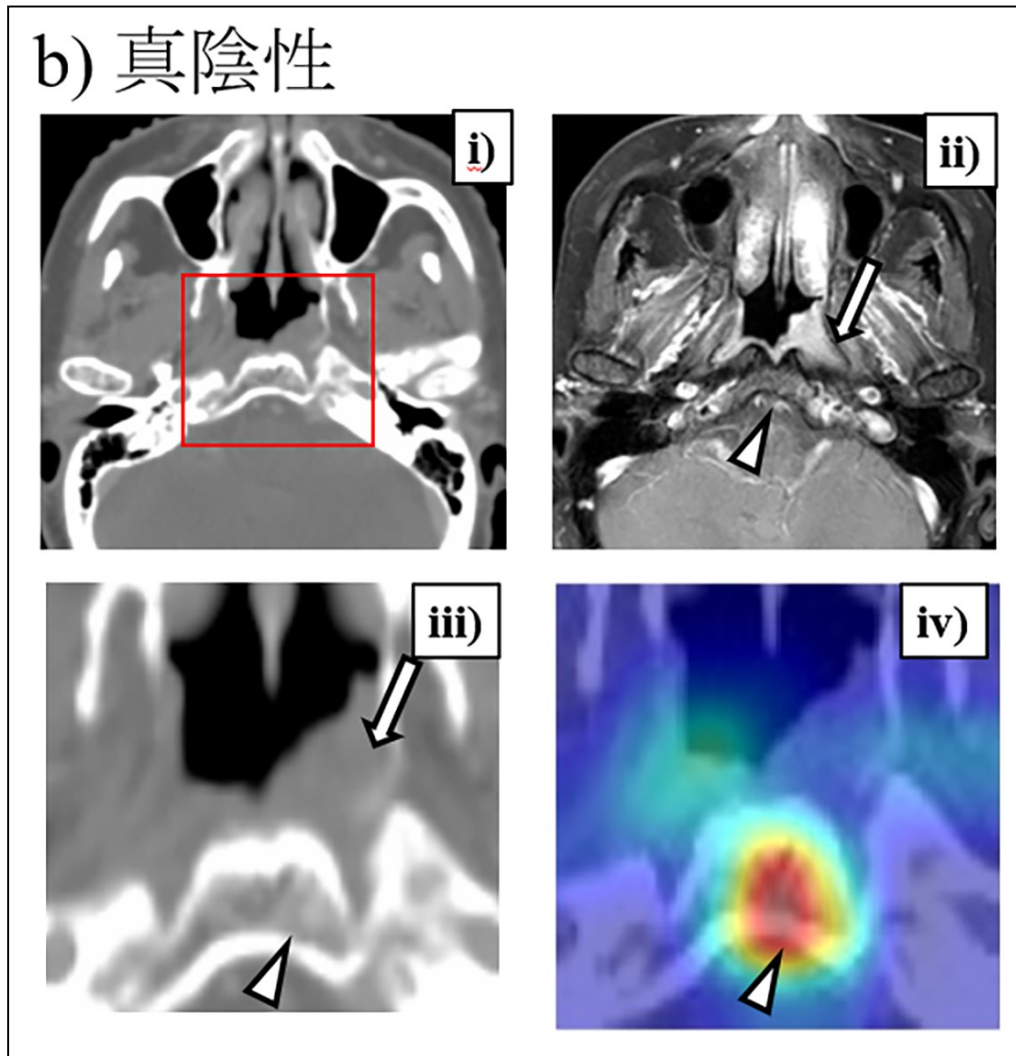


図 2-5：代表的な症例の画像

a)真陽性、b)真陰性の代表症例を提示する。

また各症例の画像は i) 元の水平断 CT 画像（赤い四角は長方形 ROI の抽出範囲）、ii) MRI の FsGdT1WI、iii) 前処理済み CT 画像、iv) Grad-CAM ヒートマップをそれぞれ表示している。

(a) 真陽性例では上咽頭に充満する軟部腫瘍があり（矢印）、CT で斜台に明瞭な骨硬化像を認め（i:矢頭）、MRI でも造影効果を認め（ii:矢印）、頭蓋底浸潤の所見を示す。ヒートマップは腫瘍と斜台に注目している。(b) 真陰性例では上咽頭左側壁に軟部腫瘍があり（矢印）、CT では斜台左側の骨髄に若干の濃度上昇があるように見えるが（i:矢頭）、(a)の真陽性例よりは弱い。MRI では腫瘍と斜台が離れているのが明瞭であり、斜台にも造影効果は認めない。ヒートマップは斜台に注目している。

5. 別の放射線科医による ROI 設定実験

正解ラベル決定や読影実験を行った放射線科医とも異なる放射線科医がテストデータセットの CT 画像に新たに設定した ROI で抽出した画像を CNN モデルに入力し、浸潤陽性スライスの連続枚数を用いて各患者における浸潤陽性・陰性分類の ROC 曲線解析を行った。この AUC は 0.971 (95%CI : 0.861-1.045) であり、オリジナルの ROI 設定時の ROC-AUC と有意差はなかった ($p=0.730$)。二つの ROC 曲線を重ねたものを以下に図示する (図 2-6)。

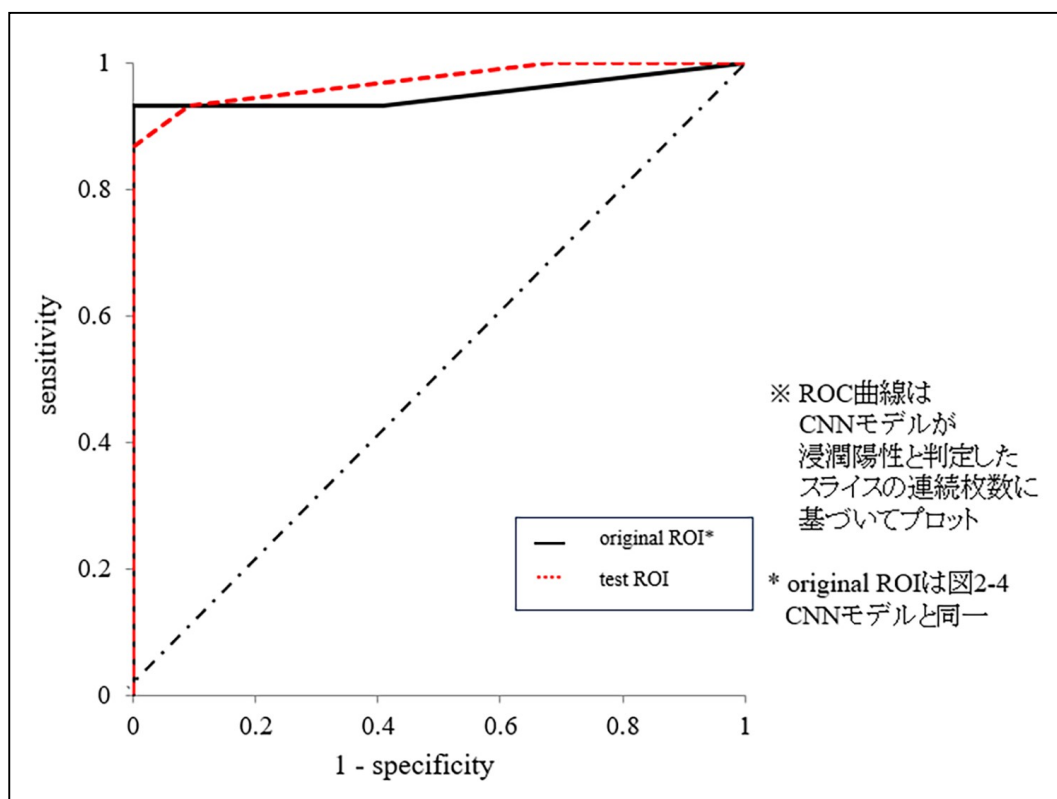


図 2-6 ROI 設定実験に基づいた ROC 曲線

テストデータセットの CT 画像に設定したオリジナルの ROI (黒実線 ; 図 2-4 と同じグラフ) と、別の放射線科医が新たに設定した ROI (赤破線) によって抽出した画像を入力した CNN モデルの診断性能を示す ROC 曲線をそれぞれ示した。

考察

本研究では CT で観察された上咽頭悪性腫瘍による頭蓋底浸潤の放射線学的徴候を識別するために、DL 法の一つである CNN を応用した診断モデルを作成した。頭蓋底浸潤の診断において、頭頸部画像診断の経験が豊富な放射線科医は、経験の少ない放射線科医よりも高い診断性能を示し、作成した CNN の診断モデルはいずれの放射線科医よりも高い診断性能を示した。

CT は骨、特に皮質骨浸潤を評価するのに有効なモダリティであるが、CT はビームハードニングアーチファクトの影響を受けやすく、骨髄の浸潤の評価が困難となることがある (Hiyama et al., 2019)。上咽頭癌における頭蓋底浸潤に対するさまざまなモダリティの診断能が報告されており、そのうちの 1 つの研究では、従来の CT における感度は 78.6%、特異度は 86.1%であったが、面検出器 CT (Area Detector CT) を用いた Bone subtraction iodine 画像を併用すると、感度は 92.9%、特異度は 95.6%となった (Hiyama et al., 2019)。もう 1 つの研究では、Dual energy CT を使用すると感度 75.0%、特異度 93.2%となり、ヨード濃度と実行原子番号 (Zeff) 値を組み合わせると、感度 90.7%、特異度 95.3%となったと報告されている (Zhan et al., 2023)。しかし、これらの画像診断装置や方法は新しい技術を必要とし、そのような画像診断が可能な施設は限られている。今回開発した CNN モデルの診断性能 (感度 93.3%、特異度 100%) は上記の研究結果と同等であると考えられ、かつ上記の研究とは異なりどの施設でも入手可能な従来型の CT 画像へ適応可能な技術である。

上咽頭腫瘍 (主に上咽頭癌) に対して CNN 技術を応用した研究は、画像セグメンテーション (Lin et al., 2019; Yang G et al., 2022)、疾患の分類 (Wong et al., 2021)、予後予測 (Li et al., 2023)、治療効果予測 (Yang Y et al., 2022) など、近年徐々に増加している。しかし、これらの研究のほとんどは、MRI 画像をトレーニングデータとして使用しており、CT 画像をトレーニングデータとして使用しているものはほとんどない。加えて、我々の知る限りでは、上咽頭腫瘍による頭蓋底浸潤を診断するための CNN モデルの報告はなく、本研究には新規性があると考えられる。

今回の研究では、Grad-CAM のヒートマップは浸潤陰性の症例では骨硬化や軟部組織濃度の乏しい斜台の骨髄に焦点を当てた。一方、浸潤陽性症例の多くでは、ヒートマップは上咽頭腫瘍そのものや斜台の骨硬化や骨破壊に焦点を合

わせていた。これらの所見は、日常臨床において頭蓋底浸潤の有無を評価する際に、放射線科医が注意を払うべき重要なポイントに類似していると考えられた。

本研究にはいくつかの限界がある。単一施設での研究デザインのためサンプルサイズが小さいこと、上咽頭と頭蓋底を含む連続スライスを抽出するための長方形 ROI は手動設定が必要である事、上咽頭癌の標準治療が外科手術ではなく化学療法や放射線治療あるいはその併用であるため頭蓋底浸潤の病理組織学的検証が得られていない事があげられる。サンプルサイズについては今回もデータ拡張を用いることによって限界を部分的に克服した。長方形 ROI の手動設定が必要であるため完全に自動化されたパイプラインは実現できていないが、別の若手放射線科医が ROI 設定を行った場合でも、ほぼ同じ診断性能が得られたことから、臨床使用において操作者間の変動はある程度許容できる可能性を示している。病理組織学的な検証が得られていない点について、腫瘍に隣接する部位では反応性の炎症、浮腫、線維化によっても骨髄の信号変化が起こりうるとされており (Becker et al., 2008) 、今回の患者の MRI 所見についても病理学的には偽陽性の可能性がある。しかし、実臨床における判断基準も主に MRI での信号変化の有無に基づいており、診療録やカンサーボードによる臨床診断と本研究の正解ラベルとの間に大きな不一致はなかったことから、実臨床との乖離は少ないと考えられた。

全体の結論

1.本研究から得られた新知見

○第一章

- 鼻副鼻腔悪性腫瘍による眼窩骨膜浸潤を超えた眼窩浸潤の CT 診断に対して、事前学習済み画像分類モデルである VGG16 を元にした転移学習によって CNN モデルを作成し、診断性能を評価した。
- 作成した CNN モデルは頭頸部画像診断を専門としない一般放射線科医よりも有意に高い診断性能を達成し、CNN モデルの診断結果を参考にした場合には一般放射線科医の診断性能が向上した。
- CNN モデルにおける偽陰性例は腫瘍が眼窩内側にある症例、偽陽性例は腫瘍による眼窩下壁の変形はあるものの菲薄化した骨壁が確認できた症例で、訓練データ数の不足が原因と推察された。

○第二章

- 上咽頭悪性腫瘍による頭蓋底浸潤の CT 診断に対して、事前学習済み画像分類モデルである ResNet50 を元にした転移学習によって CNN モデルを作成し、診断性能を評価した。
- 作成した CNN モデルは頭頸部画像診断の経験が豊富な放射線科医、および、経験が乏しい放射線科医のいずれよりも有意に高い診断性能を達成した。

2.新知見の意義

CT 上における頭頸部悪性腫瘍の局所浸潤の診断について DL を応用した既報は極めて少なかった。本研究では第一章、第二章ともに作成した CNN モデルは放射線科医よりも有意に診断性能が高く、「転移学習を用いた CNN モデルは頭頸部を専門としない放射線科医よりも高い診断性能を示す」という仮説が立証できたと考えられ、頭頸部の専門家がない施設での補助ツールになると期待される。

3.本研究で得られた新知見から今後展開されうる研究

また、本研究の結果から様々な解剖構造への腫瘍浸潤の評価に CNN モデルが応用可能と考えられ、原発巣の病期分類の診断基準となる各種の解剖構造（例えば、翼状突起、脳神経など）についても、それぞれの診断モデルを作成

する研究が展開されうる。また、今回の CNN モデルは種々のメーカーの CT によって学習・テストされているものの、単一施設での検証に過ぎないため、他施設の CT でも同様の診断性能を発揮できるかどうかは検証する必要があると考えられる。

4.今後の課題

本研究では第一章、第二章ともに病変部への ROI の手動設定が限界の一つとなっている。DL 技術は病変検出やセグメンテーションなどにも応用されており、本研究のような画像分類モデルの前段階として病変の局在や範囲を正確に抽出するプログラムが導入されれば、ROI 設定が簡便化～自動化できる可能性があり、ROI 設定者の熟練度や経験に左右されず、診断結果のばらつきが低減されることが期待される。また、このような診断ソフトウェアが読影ビューアに実装されれば、CT 画像を入力するだけで腫瘍の浸潤範囲や浸潤された解剖構造およびそれに付随した病期分類に関する情報を自動生成し、レポート作成の有用な支援ツールになると期待される。

謝辞

本稿を結ぶに当たり、終始懇切なる指導を賜りました指導教官の北海道大学大学院医学研究院放射線科学分野画像診断学教室 工藤與亮 教授に深い感謝の意を表します。

併せて、本研究全般に渡って直接的なご指導を賜りました北海道大学病院 放射線診断科 藤間憲幸 講師に深い感謝の意を表します。

加えて、画像診断及び解析にご協力を頂いた、北海道大学大学院医学研究院放射線科学分野画像診断学教室 平田健司 准教授、原田太以佑 助教、唐明輝 特任助教、北海道大学死因究明教育研究センター オートプシーイメージング部門 池辺洋平 特任助教、北海道大学大学院歯学研究院口腔病態学分野放射線学教室 常田慧徳 助教、手稲溪仁会病院 放射線診断科 鈴木純 医師、北海道大学病院 放射線診断科 若林直人 医師や高野雄大 医師に深く感謝申し上げます。

また、臨床的な観点から様々な助言をいただきました、北海道大学大学院歯学研究院口腔病態学分野 放射線学教室 箕輪和行 教授、北海道大学大学院医学研究院耳鼻咽喉科・頭頸部外科学教室 本間明宏 教授や加納里志 講師に深く感謝いたします。

最後に、本論文は臨床データを使用させていただいた患者の皆様、当該分野内外の先生方、その他多くの方々のご指導とご協力に支えられて完成したものであり、研究に携わったすべての方々への心からの感謝の意を表します。

利益相反

本研究において開示すべき利益相反はありません。

引用文献

Abdel Khalek Abdel Razek A, King A. (2012) MRI and CT of nasopharyngeal carcinoma. *AJR Am J Roentgenol.* 198:11-8.

Becker M, Zbaren P, Casselman JW, Kohler R, Dulguerov P, Becker CD. (2008) Neoplastic invasion of laryngeal cartilage: reassessment of criteria for diagnosis at MR imaging. *Radiology.* 249:551-9.

Blum A, Zins M. *Radiology: Is its future bright? Diagn Interv Imag* 2017, 98(5):369-371. doi: 10.1016/j.diii.2017.04.002.

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A.(2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer J Clinicians*; 68: 394–424. doi: 10.3322/caac.21492.

Carrau RL, Segas J, Nuss DW, Snyderman CH, Janecka IP, Myers EN, D'Amico F, Johnson JT. (1999) Squamous cell carcinoma of the sinonasal tract invading the orbit. *Laryngoscope.* 109(2):230-235.

Chen L, Liu LZ, Mao YP, Tang LL, Sun Y, Chen Y, et al. (2011) Grading of MRI-detected skull-base invasion in nasopharyngeal carcinoma and its prognostic value. *Head Neck.* 33:1309-14.

Chong VF, Fan YF. (1996) Skull base erosion in nasopharyngeal carcinoma: detection by CT and MRI. *Clin Radiol.* 51:625-31.

Chong VFH, Fan YF, Khoo YBK. (1996) Nasopharyngeal carcinoma with intracranial spread: CT and MR characteristics. *J Comput Assist Tomo.* 20:563-9.

Choi H, Kim H, Hong W, Park J, Hwang EJ, Park CM, Kim YT, Goo JM. (2021) Prediction of visceral pleural invasion in lung cancer on CT: deep learning model

achieves a radiologist-level performance with adaptive sensitivity and specificity to clinical needs. *Eur Radiol.* 31(5):2866-2876.

Cornelius CP, Mayer P, Ehrenfeld M, Metzger MC. (2014) The orbits – Anatomical features in view of innovative surgical methods. *Facial Plast Surg.* 30(5):487-508.

Duron L, Ducarouge A, Gillibert A, Laine J, Allouche C, Cherel N, Zhang ZK, Nitche N, Lacave E, Pourchot A et al. (2021) Assessment of an AI aid in detection of adult appendicular skeletal fractures by emergency physicians and radiologists: A multicenter cross-sectional diagnostic study. *Radiology.* 300(1):120-129.

Eisen MD, Yousem DM, Loevner LA, Thaler ER, Bilker WB, Goldberg AN. (2000) Preoperative imaging to predict orbital invasion by tumor. *Head Neck-J Sci Spec.* 22(5):456-462.

Chollet, F. (2016) Xception: Deep Learning with Depthwise Separable Convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 1251-1258, arXiv2016, arXiv:1610.02357.

Gabriel Chartrand, Phillip M Cheng, Eugene Vorontsov, Michal Drozdal, Simon Turcotte, Christopher J Pal, Samuel Kadoury, An Tang. (2017) Deep Learning: A Primer for Radiologists. *Radiographics.* 37(7): 2113-2131. doi: 10.1148/rg.2017170077.

Ganly I, Patel SG, Singh B, Kraus DH, Bridger PG, Cantu G, Cheesman A, De Sa G, Donald P, Fliss DM et al. (2005) Craniofacial resection for malignant paranasal sinus tumors: Report of an international collaborative study. *Head Neck-J Sci Spec.*, 27(7):575-584. doi: 10.1002/hed.20165.

Hirata K, Kobayashi K, Wong KP, Manabe O, Surmak A, Tamaki N, Huang SC. (2014) A semi-automated technique determining the liver standardized uptake value reference for tumor delineation in FDG PET-CT. *PLOS ONE.* 9(8).

Hiyama T, Kuno H, Sekiya K, Tsushima S, Sakai O, Kusumoto M, et al. (2019)

Bone Subtraction Iodine Imaging Using Area Detector CT for Evaluation of Skull Base Invasion by Nasopharyngeal Carcinoma. *AJNR Am J Neuroradiol.* 40:135-41.
Huang SH, O'Sullivan B. Overview of the 8th Edition TNM Classification for Head and Neck Cancer. (2017) *Curr Treat Option On.* 18.

Jiang B, Zhang Y, Zhang L, G HdB, Vliegenthart R, Xie X. (2021) Human-recognizable CT image features of subsolid lung nodules associated with diagnosis and classification by convolutional neural networks. *Eur Radiol.* 31:7303-15.

Kadish S, Goodman M, Wang CC. (1976) Olfactory neuroblastoma. A clinical analysis of 17 cases. *Cancer.* 37: 15711576.

He K, Zhang X, Ren S, Sun J. (2015) Deep Residual Learning for Image Recognition. *Computer Vision and Pattern Recognition.* doi.org/10.48550/arXiv.1512.03385

Kann BH, Hicks DF, Payabvash S, Mahajan A, Du J, Gupta V, Park HS, Yu JB, Yarbrough WG, Burtness BA, Zain AH, Sanjay A. (2020) Multi-institutional validation of deep learning for pretreatment identification of extranodal extension in head and neck squamous cell carcinoma. *J Clin Oncol* , 38(12):1304-1311. doi: 10.1200/JCO.19.02031.

King AD, Lei KI, Richards PS, Ahuja AT. (2003) Non-Hodgkin's lymphoma of the nasopharynx: CT and MR imaging. *Clin Radiol.* 58:621-5.

Li S, Wan X, Deng YQ, Hua HL, Li SL, Chen XX, et al. (2023) Predicting prognosis of nasopharyngeal carcinoma based on deep learning: peritumoral region should be valued. *Cancer Imaging.* 23:14.

Li Y, Zhang Y, Zhang E, Chen Y, Wang Q, Liu K, et al. (2021) Differential diagnosis of benign and malignant vertebral fracture on CT using deep learning. *Eur Radiol.* 31:9612-9.

Li ZY, Kitajima K, Hirata K, Togo R, Takenaka J, Miyoshi Y, Kudo K, Ogawa T, Haseyama M. (2021) Preliminary study of AI-assisted diagnosis using FDG-PET/CT

for axillary lymph node metastasis in patients with breast cancer. *Ejnm Res.* 11(1).

Lin L, Dou Q, Jin Y-M, Zhou G-Q, Tang Y-Q, Chen W-L, et al. (2019) Deep Learning for Automated Contouring of Primary Tumor Volumes by MRI for Nasopharyngeal Carcinoma. *Radiology.*291:677-86.

Meerwein CM, Pazahr S, Soyka MB, Hullner MW, Holzmann D. (2020) Diagnostic accuracy of computed tomography and magnetic resonance imaging compared to surgical exploration for anterior skull base and medial orbital wall infiltration in advanced sinonasal tumors. *Head Neck-J Sci Spec.*, 42(8):2002-2012.

Ng SH, Chang TC, Ko SF, Yen PS, Wan YL, Tang LM, et al. (1997) Nasopharyngeal carcinoma: MRI and CT assessment. *Neuroradiology.* 39:741-6.

Nishioka T, Shirato H, Kagei K, Abe S, Hashimoto S, Ohmori K, et al. (2000) Skull-base invasion of nasopharyngeal carcinoma: magnetic resonance imaging findings and therapeutic implications. *Int J Radiat Oncol Biol Phys.* 47:395-400.

Perry C, Levine PA, Williamson BR, Cantrell RW. (1988) Preservation of the eye in paranasal sinus cancer surgery. *Arch Otolaryngol Head Neck Surg.* 114(6):632-634.

Pinto J, Paneth N, Kazam E, Kairam R, Wallenstein S, Rose W, Rosenfeld D, Schonfeld S, Stein I, Witomski T. (1988) Interobserver variability in neonatal cranial ultrasonography. *Paediatr Perinat Epidemiol.* 2(1):43-58.

Qian J, Li H, Wang J, He L. (2023) Recent Advances in Explainable Artificial Intelligence for Magnetic Resonance Imaging. *Diagnostics.*13:1571.

Reyes C, Patel M, Solares CA. Sinonasal malignancy and orbital exenteration Sparing cancer surgery. (2020) *J Neurol Surg Part B.* 81(04):369-375.

Roh JL, Sung MW, Kim KH, Choi BY, Oh SH, Rhee CS, et al. (2004) Nasopharyngeal carcinoma with skull base invasion: a necessity of staging subdivision. *Am J Otolaryngol.* 25:26-32.

Salfrant M, Garcia GCTE, Guichard JP, Bidault F, Reizine D, Auperin A, Bresson D, Verillaud B, Herman P, Moya-Plana A.(2021) Imaging of skull base and orbital invasion in sinonasal cancer: Correlation with histopathology. *Cancers*. 13(19).

Santin M, Brama C, Thero H, Ketheeswaran E, El-Karoui I, Bidault F, Gillet R, Teixeira PG, Blum A. (2019) Detecting abnormal thyroid cartilages on CT using deep learning. *Diagn Interv Imag*. 100(4):251-257.

Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. (2020) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*. 128:336-59.

Simonyan K, and Zisserman A, (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. *The 3rd International Conference on Learning Representations (ICLR2015)*. doi: 10.48550/arXiv.1409.1556

Soung Yung Kim, Meinrad Beer, Dechen W. Tshering Vogel. (2021) Imaging in head and neck cancers: Update for non-radiologist. *Oral Oncol*. Sep:120:105434. doi: 10.1016/j.oraloncology.2021.105434.

Suarez C, Ferlito A, Lund VJ, Silver CE, Fagan JJ, Rodrigo JP, Llorente JL, Cantu G, Politi M, Wei WI et al. (2008) Management of the orbit in malignant sinonasal tumors. *Head Neck-J Sci Spec*. 30(2):242-250.

Synho Do, Kyoung Doo Song, Joo Won Chung. (2020) Basics of Deep Learning: A Radiologist's Guide to Understanding Published Radiology Articles on Deep Learning. *Korean J Radiol*. 21(1): 33–41. doi: 10.3348/kjr.2019.0312

Wong LM, King AD, Ai QYH, Lam WKJ, Poon DMC, Ma BBY, et al. (2021) Convolutional neural network for discriminating nasopharyngeal carcinoma and benign hyperplasia on MRI. *Eur Radiol*. 31:3856-63.

Yang G, Dai Z, Zhang Y, Zhu L, Tan J, Chen Z, et al. (2022) Multiscale Local

Enhancement Deep Convolutional Networks for the Automated 3D Segmentation of Gross Tumor Volumes in Nasopharyngeal Carcinoma: A Multi-Institutional Dataset Study. *Front Oncol.*12:827991.

Yang Y, Wang M, Qiu K, Wang Y, Ma X. (2022) Computed tomography-based deep-learning prediction of induction chemotherapy treatment response in locally advanced nasopharyngeal carcinoma. *Strahlenther Onkol.* 198:183-93.

Yuhan Yang, Xiuhe Zou, Yixi Wang, Xuelei Ma. (2021) Application of deep learning as a noninvasive tool to differentiate muscle-invasive bladder cancer and non-muscle-invasive bladder cancer with CT. *Eur J Radiol.* 139:109666. doi: 10.1016/j.ejrad.2021.109666.

Zhan Y, Wang P, Wang Y, Wang Y, Tang Z. (2023) Dual-energy CT for the detection of skull base invasion in nasopharyngeal carcinoma: comparison of simulated single-energy CT and MRI. *Insights Imaging.*14:95.

Zhang GMY, Wu Z, Xu LL, Zhang XX, Zhang DM, Mao L, Li XL, Xiao Y, Guo J, Ji ZG et al: Deep Learning on Enhanced CT Images Can Predict the Muscular Invasiveness of Bladder Cancer. *Front Oncol* 2021, 11. doi: 10.3389/fonc.2021.654685.

日本医学放射線学会 編：画像診断ガイドライン 2021 年版.

日本頭頸部癌学会 編：頭頸部癌診療ガイドライン 2022 年版.

間下 以大. (2018) 人工知能と機械学習, 深層学習. 映像情報メディア学会誌. Vol.72, No.2, 236-240.

森 康久仁. (2018) 機械学習の基礎と医用画像への応用. 医用画像情報学会雑誌.35 卷 3 号, 42-47.