

修士論文

対照学習による
顎骨骨髓炎発症範囲の自動推定

令和4年度修了

三重大学大学院工学研究科情報工学専攻
ヒューマンコンピュータインタラクション研究室

星野 秀晃

はじめに

顎骨骨髓炎は顎骨中心部である骨髓に起こる感染症であり、口腔外科において最も難治性の高い疾患の一つである。発症要因は骨吸収抑制剤の使用や頭頸部への放射線治療であり、他の症状の治療経過において発症する。骨露出による痛みと持続性、顎の重苦しさやしびれ、歯ぐきの腫れや膿が症状の大半であり、無痛の場合や患者が症状に無自覚の場合もある。感染が進むと、病的骨折や皮膚瘻孔を形成し、炎症が広範囲に及んだ場合には、開口障害により口からの食事ができなくなる。

近年では薬剤投与による顎骨の保存的治療が患者に多大な苦痛を与えるため、発症範囲を切除する外科的治療が検討されている。外科的治療においては術前に切除範囲を推定する必要があり、その際には骨髓炎発症範囲の不明瞭な術前 CT・MRI・PET 画像を参照する。この術前画像から明確な切除範囲を決めることは難しく、切除不十分や過切除が起こり、患者の予後悪化を招く。

本研究の目的は顎骨骨髓炎の治療において、術前画像と術後病理画像の関連性を明らかにし、術前における顎骨骨髓炎発症範囲の高精度推定をすることである。これにより手術計画と外科的治療に要する時間を短縮することができ、また再手術を抑制することで医師・患者の負担軽減をする。日常診療に応用することで早期発見も見込まれる。

関連研究はいくつかあるものの、顎骨骨髓炎発症範囲を推定する手法は検討・研究されていない。先行研究 [1] では、同時生起行列を用いて術前 CT 画像の特徴量を抽出し、三種類の機械学習器による学習を通して骨髓炎部位を推定した。しかし顎骨の正常部と骨髓炎部のデータ数が不均衡となり、推定領域は骨髓炎部位のみならず顎骨全体に広がってしまい、骨髓炎部位を正確に推定できない問題があった。

本研究では A Simple Framework for Contrastive Learning of Visual Representations (SimCLR) [2] による対照学習での特徴抽出と、Support Vector Machine(SVM) による特徴の学習、正常/骨髓炎の判定を行った。対照学習はラベルを必要としない自己教師あり学習のため、顎骨骨髓炎データ不均衡性の問題を解消した。被験者 9 名による術前頭部三次元 CT 画像を用いた実験により、骨髓炎発症有無を F 値 0.734 の精度で推定でき、先行研究の精度を上回った。また、発症有無を参照して顎骨における骨髓炎発症領域を推定するこ

とができた。本手法は被験者データ不足により、顎骨骨髄炎の普遍的な特徴を学ぶには至らなかったため、今後の課題として高精度なデータ拡張等によるデータ不足解消が挙げられる。

目次

はじめに	i
第 1 章 緒言	1
1.1 研究背景	1
1.2 関連研究	2
1.3 先行研究	2
1.4 研究目的	2
第 2 章 準備	4
2.1 データセット	4
2.2 SimCLR	7
2.3 SVM	8
第 3 章 提案手法	9
3.1 前処理	10
3.2 パッチ画像表現の学習	12
3.3 骨髄炎パッチ推定	14
第 4 章 実験	15
4.1 実験データ	15
4.2 精度評価	16
4.3 実験手順	17
4.4 結果と考察	20
4.5 追加実験	25
第 5 章 結言	26
付録 A 付録	27
A.1 プログラム	27

A.2	データ	27
謝辞		28

第 1 章

緒言

1.1 研究背景

顎骨骨髓炎は顎骨中心部である骨髓に起こる感染症であり、口腔外科において最も難治性の高い疾患の一つである。細菌感染により骨髓が炎症を起こし、顎骨の壊死を引き起こす。顎骨骨髓炎の発症原因には大きく分けて 2 つのタイプがある。薬剤関連顎骨壊死 (MRONJ) は、がん関連疾患や骨粗鬆症の治療に使われるビスフォスフォネート (BP) やデノスマブ (Dmab) 等の骨吸収抑制剤が原因で発症し、放射線性骨髄炎は、頭頸部への放射線照射が原因で発症する。顎骨骨髓炎の症状としては、痛みを伴う持続的な骨露出・顎が重い感じやしびれ感・歯肉の腫脹や排膿などがある。ただし、患者が痛みを感じなかったり、顎骨骨髓炎を自覚していない場合もある。感染が進むと、病的骨折や皮膚瘻孔を形成し、炎症が広範囲に及んだ場合には、開口障害により口からの食事ができなくなる。

顎骨骨髓炎の治療法は、抗菌薬による非外科的な保存的治療と発症範囲の切除による外科的治療である [3]。保存的治療はすべての症状段階で有効で、早期の症状を安定させたり治癒させたりすることが可能である。この治療法は現在も研究が進められており、新規治療薬であるテリパラチドは MRONJ の治療薬として期待されている [4]。ただし、保存的治療は患者に対して長期に渡る疼痛・排膿や病的骨折などで多大な苦痛を与えるため治療に難渋している。

進行期の患者には早期の外科的治療が有効であると考えられており、病気の進行を抑える治療法として外科的治療を検討・提示することが必要である。顎骨切除の外科的治療が保存的治療より優れているとする論文がある。[5–7]。顎骨骨髓炎発症患者に対しての顎骨切除の予後を検討した論文では、積極的な外科治療が 116 例中 109 例で成功している [8]。MRONJ に対する切除を想定して、上顎洞手術との併用切除が良好であることも示された [9]。

外科手術の問題点は、訓練を受けた放射線科医でさえ、CT, MRI, PET 画像において

曖昧な顎骨骨髓炎の領域を正確に決定することができないことである。手術計画に時間を要し、術前に推測した切除領域が実際の領域と異なることが多く、切除が不十分なために顎骨骨髓炎が再発したり、過剰な切除による患者の QOL の悪化などの不正確な手術計画による予後の悪化が問題となっている。

1.2 関連研究

近年、機械学習アルゴリズムによる医用画像処理・解析の応用が盛んである。医用画像処理における活用例は以下のようなものがある。アルツハイマー病の診断や軽度認知障害の進行を機械学習で予測することで、専門家である放射線科医よりも優れた診断ができ、医師のサポートが可能となった [10]。肺癌の検診において利用される肺 CT 検診による肺結節の早期自動検出に、CT 画像パッチから肺微小結節と非結節を学習した Convolutional Neural Network(CNN) が利用され、偽陽性率を低減させた [11]。

骨髓炎に関連する骨壊死の自動検出については、糖尿病患者の足部 MRI を用いて、足指の骨の感染を正確に判断した研究例もある [12]。顎骨骨髓炎患者について年齢や薬剤投与歴などの因子を機械学習器に分析させ、任意の患者における顎骨骨髓炎の発症を判断する研究では、発症の有無を予測するのみであり、明確な骨髓炎領域は推定されていない [15]。また、顎骨の骨壊死に関連する生体組織の変化については研究されているが [13, 14]、骨髓炎とその自動検出については全く検討されていない。CT, MRI, PET 画像による骨髓炎の自動検出は、これまで研究されていないことが現状である。

1.3 先行研究

先行研究として盛田らの研究 [1] がある。術前 CT 画像から同時生起行列 Gray Level Co-Occurrence Matrix (GLCM) を用いて特徴量を抽出し、Neural Network (NN) や Support Vector Machine (SVM), Random Forest (RF) の 3 種類の識別器を用いて 3 名分の顎骨骨髓炎領域の推定を行った。SVM は最も高い分類精度を示したが、SVM が推定した骨髓炎領域は健常な顎骨を含む顎骨全体に及んでいた。顎骨骨髓炎が顎骨全体の一部にのみ発症するため、顎骨の正常部と壊死部のデータ数が不均衡となり、学習が失敗してしまったことが示唆された。

1.4 研究目的

本研究は、難治性顎骨骨髓炎の治療において、術前 CT と術後病理画像の関連性を明らかにすることで、術前における骨髓炎発生領域の高精度推定を可能とすることを目的

とする。提案システムは、手術計画時間の短縮や適切な骨髄炎切除を実現し、医師・患者の再手術等の負担軽減が見込まれる。また、術前画像と病理画像の関連性を明らかにし、応用することで、日常診療における骨髄炎の早期発見と予防医療の推進も期待できる。本研究では、A Simple Framework for Contrastive Learning of Visual Representations (SimCLR) [2] と SVM を組み合わせて、正常/骨髄炎判定の高精度化を行う。

第 2 章

準備

2.1 データセット

本研究では被験者 9 名分の頭部三次元 CT 画像を用いる (図 2.1). これは神戸大学附属病院より提供され, この研究は神戸大学附属病院倫理委員会より承認を得ている. CT 画像のサイズと空間分解能は表 2.1 の通りである. また, 提供された頭部三次元 CT 画像を参照し, 手動で顎骨領域マスク (図 2.2) と骨髄炎領域マスク (図 2.3) を作成した.

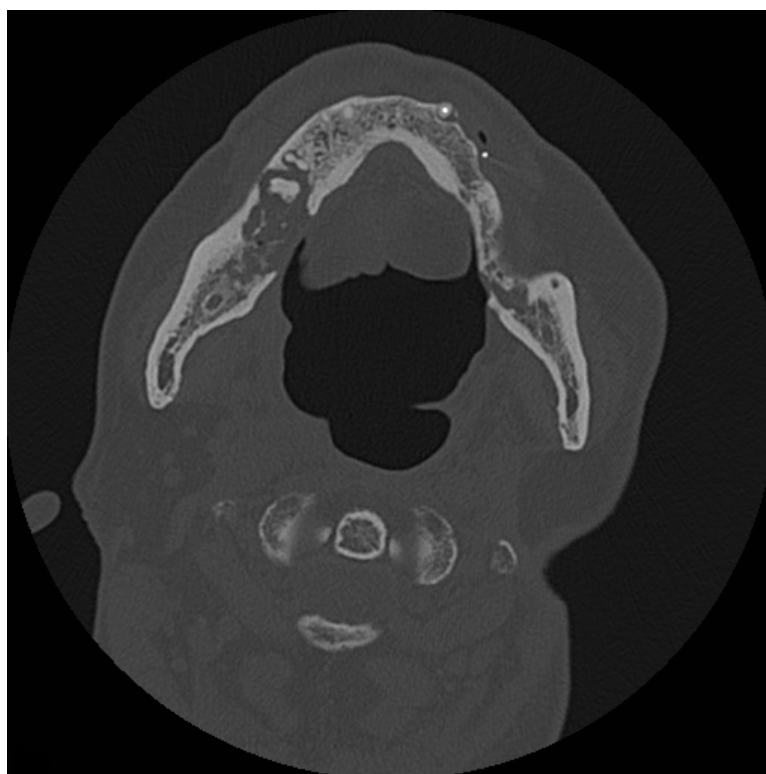


図 2.1: 頭部 CT スライス

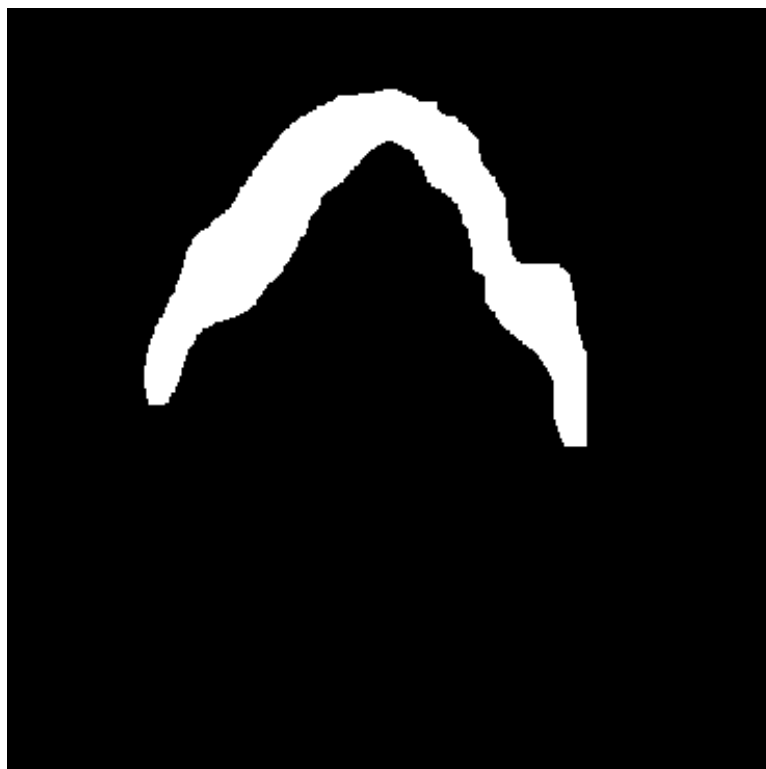


図 2.2: 顎骨領域マスク



図 2.3: 骨髄炎領域マスク

表 2.1: 頭部三次元 CT 画像詳細

被験者	サイズ (<i>voxel</i>)	空間分解能 (mm^3)
No.1	$512 \times 512 \times 401$	$0.351 \times 0.351 \times 0.3$
No.2	$512 \times 512 \times 246$	$0.390 \times 0.390 \times 1.0$
No.3	$512 \times 512 \times 282$	$0.391 \times 0.391 \times 1.0$
No.4	$512 \times 512 \times 346$	$0.352 \times 0.352 \times 0.4$
No.5	$512 \times 512 \times 286$	$0.352 \times 0.352 \times 0.4$
No.6	$512 \times 512 \times 296$	$0.352 \times 0.352 \times 0.4$
No.7	$512 \times 512 \times 601$	$0.351 \times 0.351 \times 0.3$
No.8	$512 \times 512 \times 401$	$0.352 \times 0.352 \times 0.3$
No.9	$512 \times 512 \times 299$	$0.352 \times 0.352 \times 0.4$

2.2 SimCLR

SimCLR [2] は深層学習モデルの一つである。ラベル付けを行うことなくデータを比較し学習する自己教師あり対照学習アルゴリズムを簡略化して表現したものである。特徴空間において類似したデータを近い位置に、異なるデータを遠い位置に配置するようにデータを変換する。対照学習は主に画像処理、自然言語処理の分野で使われている。SimCLR は 1 つのデータを変換フィルターを用いて異なる 2 つのデータに増幅する。これら二つのデータを正のペアとして扱う。また正のペア (i, j) 以外のペア $(i, k_{[k \neq j]})$ は負のペアとして扱う。SimCLR は ResNet と多層パーセプトロンで構成され、入力データを 128 次元の特徴量に変換する。正のペアは同一のデータより生成されたため、正のペアの 128 次元の特徴量同士が似たものになるよう、コサイン類似度を最大にするように学習する。負のペアは別のデータより生成されたため、特徴量同士が異なったものになるよう、コサイン類似度を最小にするように学習する。SimCLR で用いられる損失関数は、正規化温度スケールクロスエントロピー損失関数 (NT-Xent) といい、正のペア (i, j) の場合、次式で定義される。

$$l_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)} \quad (2.1)$$

ここで $\mathbf{1}_{[k \neq i]} \in \{0, 1\}$ は $k \neq i$ の時に 1 として計算される定義式であり、 τ は温度定数と呼ばれるハイパーパラメーター、 z_i は i 番目のデータのネットワークにより変換された特徴量である。 $\text{sim}(z_i, z_j)$ は L2 ノルムで正規化した z_i と z_j との内積を表す。データ数 N は変換フィルターにより 2 倍の $2N$ に増幅される。

図 2.4 は SimCLR で用いられるコサイン類似度の考え方である。コサイン類似度ではベクトル同士の角度が小さいほど二つの特徴ベクトルが類似していることを示す。図の左側ではクラス A の二つのサンプルは別位置に存在するため、二つの特徴ベクトル間の角度が大きくなり、コサイン類似度が小さくなる。右側ではクラス A の二つのサンプルが近い位置に存在するため、二つのベクトル間の角度が小さくなり、コサイン類似度が大きくなる。

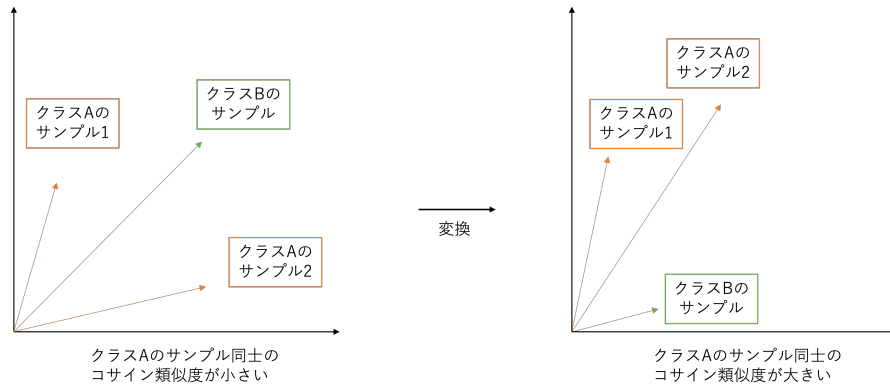


図 2.4: コサイン類似度の例

2.3 SVM

SVM は教師あり機械学習アルゴリズムの一種である。学習データ数が少ない場合のクラス分類問題や回帰問題で高い性能を示す。クラス分類問題において SVM はデータ群をクラスごとに分けるように線形の決定境界を求める。決定境界は各クラスのデータ群のうち一番近い点 (サポートベクター) との距離を最大にするよう求められる。線形な決定境界でデータ群を分けることができない場合、カーネルトリックを用いて、データを仮想的な高次元空間に射影した状態で決定境界を求める。このとき元の特徴空間において境界線が非線形となるため、これを非線形 SVM と呼ぶ。

第3章

提案手法

本論文では SimCLR と SVM を用いて術前 CT 画像の顎骨骨髓炎領域を推定する方法を提案する。提案手法の大きな流れを図 3.1 に示す。本手法では前処理として緩和処理、厳密処理の二種類、対照学習としてパッチ対照学習、ラベル対照学習の二種類を組み合わせた4通りの実験を行う。

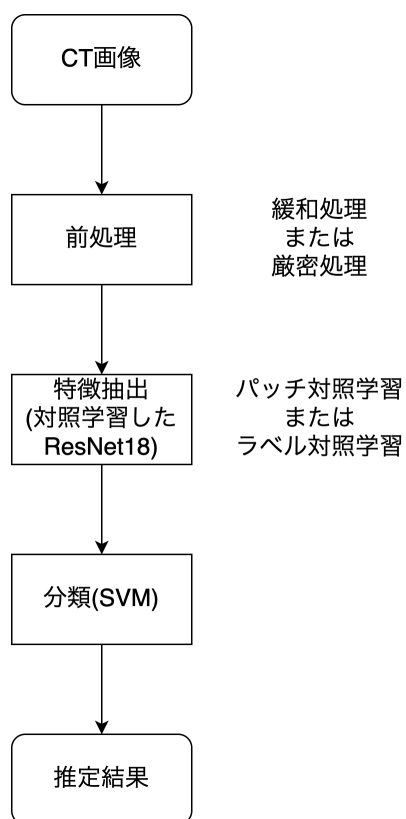


図 3.1: 提案手法の流れ

3.1 前処理

被験者の頭部三次元 CT 画像を z 方向にスライスし、体軸に垂直な面である横断面画像を得る。横断面をスライディングウィンドウ法 (ウィンドウサイズ 32×32 , スライド幅 16) で走査する。これにより得られた小領域パッチ画像に対し、顎骨マスクと骨髄炎マスクを参照し、ラベル付けを行う。ラベル付けに際しては二種類の前処理を用い、それぞれ別のデータセットとして扱う。

3.1.1 緩和処理

パッチ画像へ緩和処理のラベル付けを行う (図 3.2)。CT 画像に対応する顎骨マスクを参照してパッチ内の 65%以上が顎骨マスクと重なる場合、顎骨候補パッチとする。顎骨領域が 65%以下のパッチ画像は棄却する。顎骨候補パッチに対し、骨髄炎マスクを参照してパッチ内の 30%以上が骨髄炎マスクと重なる場合、パッチを骨髄炎とラベル付けする。パッチ内の 30%以下が骨髄炎マスクと重なる場合パッチを正常顎骨とラベル付けする。ここで、正常顎骨パッチは少量の骨髄炎領域を含んでいる点に注意されたい。

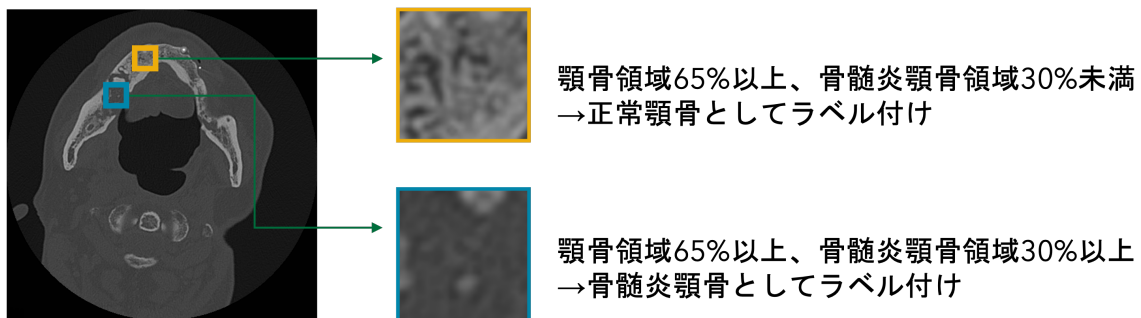


図 3.2: 緩和処理の顎骨データ抽出法

3.1.2 厳密処理

パッチ画像へ厳密処理のラベル付けを行う (図 3.3)。CT 画像に対応する顎骨マスクを参照してパッチ内の 50%以上が顎骨マスクと重なる場合、顎骨候補パッチとする。顎骨領域が 50%以下のパッチ画像は棄却する。顎骨候補パッチに対し、骨髄炎マスクを参照して 32×32 pixel のパッチ内に骨髄炎領域が 1pixel でも含まれている場合、パッチを骨髄炎とラベル付けする。パッチ内に骨髄炎領域が全く含まれていない場合、パッチを正常顎骨とラベル付けする。

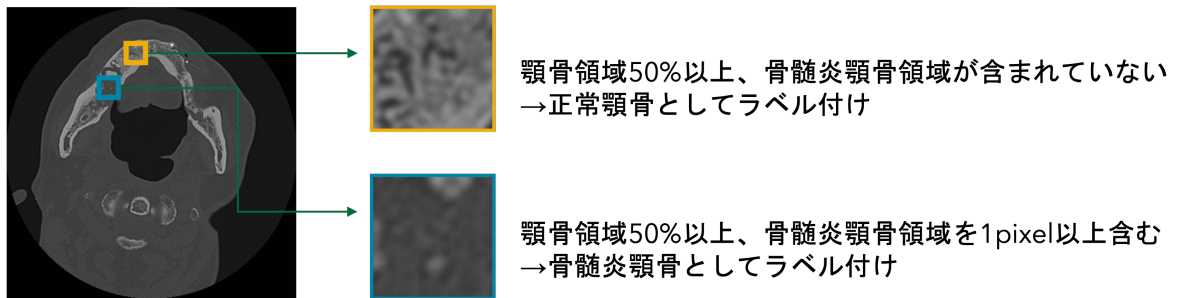


図 3.3: 厳密処理の顎骨データ抽出法

3.1.3 共通処理

ラベル付済みパッチデータセット二種類に対し画素値の正規化を行う。以下の式にてパッチ画像ごとに最大値 1, 最小値 0 にスケーリングする。

$$x'_i = \frac{x_i - \min(\mathbf{X})}{\max(\mathbf{X}) - \min(\mathbf{X})} \quad (3.1)$$

ここで $x_i \in \mathbf{X}$ は正規化前, x'_i は正規化後のピクセルごとの値である。 $\max(\mathbf{X})$ と $\min(\mathbf{X})$ はそれぞれ, パッチ画像 \mathbf{X} における画素値の最大値と最小値である。

また, パッチ画像は $32 \times 32 \times 1$ のグレースケール画像であるため, 同じパッチ画像を三枚重ね合わせ, $32 \times 32 \times 3$ の擬似的な RGB 画像として扱う。これは SimCLR が RGB 画像を入力として動作するためである。

3.2 パッチ画像表現の学習

前処理により得られたパッチデータを用い、2.2章の手法を基に対照学習を行う。機械学習モデルはパッチ画像を128次元の特徴量に変換する表現を学習する。

3.2.1 パッチ対照学習

パッチ対照学習の手法は、SimCLRの構造を用いている(図3.4)。任意の顎骨画像をランダムな画像変換により二枚に増幅する。ここで行う画像変換はリサイズ、クロップ、反転、ガウスぼかし、カラージッター(明るさ、コントラスト、彩度、色相の調整)をランダムに複数組み合わせたものである。同じ画像から生成された画像二枚を正のペアとして扱い、異なる画像から生成された画像二枚を負のペアとして扱う。ResNetを通して表現された特徴量について、正のペアの類似度が高くなるように学習し、負のペアの類似度が低くなるように学習する。SimCLRは正常・骨髄炎のラベル情報なしで、正常・骨髄炎の特徴を学習する。この学習により、骨髄炎パッチの数が比較的少ないことで、正常と骨髄炎の2クラス分類モデルの学習が困難である問題を解決することができる。

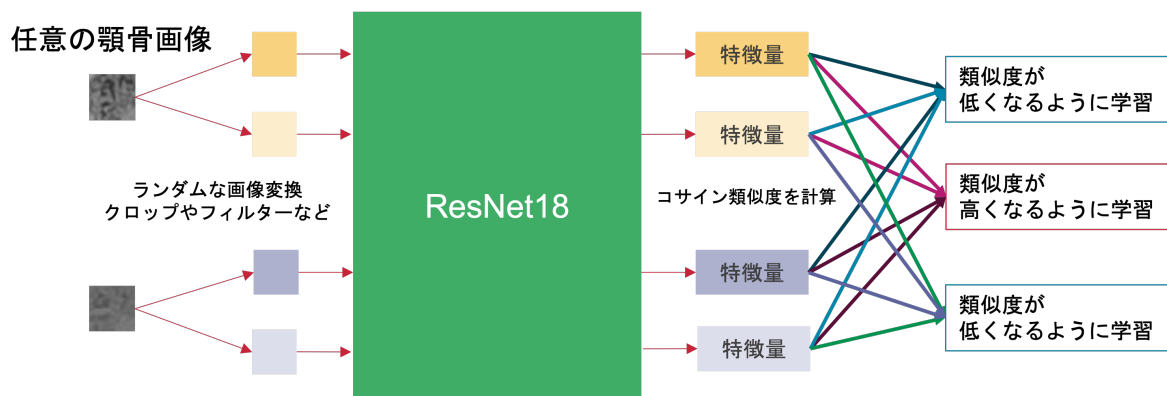


図 3.4: パッチ対照学習の手法

3.2.2 ラベル対照学習

ラベル対照学習の手法は、SimCLR の入力ペアを変化させたものである (図 3.5)。ラベル情報を持ったパッチ画像を ResNet に入力し、128 次元の特徴量に変換する。ここでラベル情報を元に、正常顎骨同士または骨髄炎顎骨同士の類似度を上げ、正常顎骨と骨髄炎顎骨のペアの類似度は下げるよう学習を進める。ラベル情報を用いるため、入力データ内の正常顎骨パッチペア数と骨髄炎顎骨パッチペア数のバランスを均等にする必要がある。

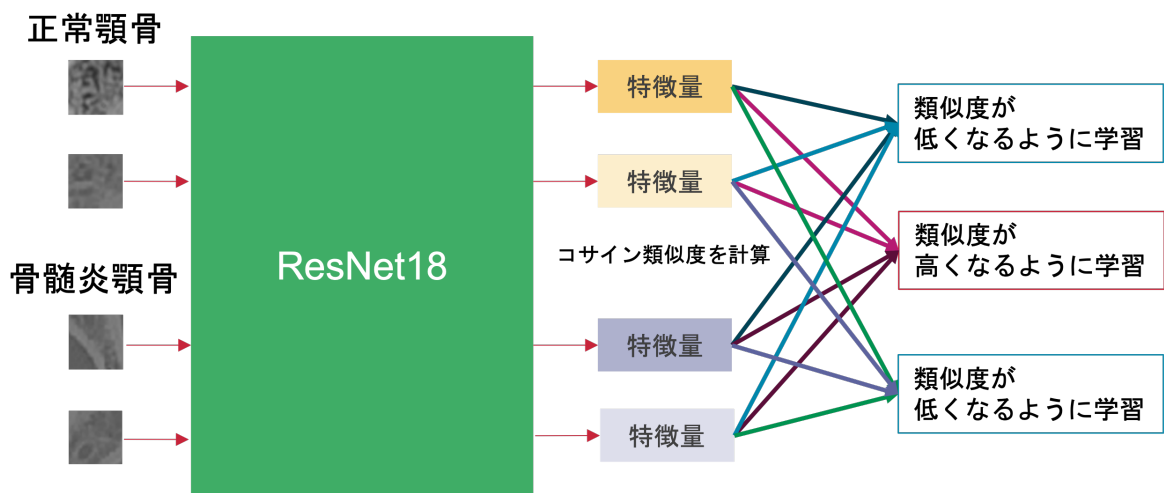


図 3.5: ラベル対照学習の手法

3.3 骨髄炎パッチ推定

対照学習モデルにより特徴量に変換されたデータを非線形 SVM(RBF カーネル) に学習させ、正常・骨髄炎の 2 クラス分類を行う (図 3.6, 図 3.7). ここでは、対象の患者一人の多量の正常顎骨と少量の骨髄炎顎骨を学習させ、その患者の他の骨髄炎を発見するという半自動的な手法を試行する.

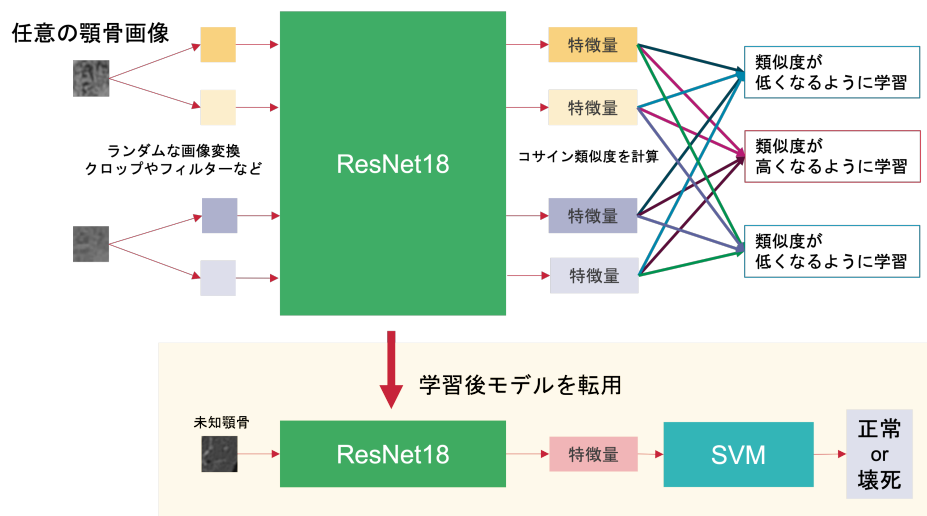


図 3.6: パッチ対照学習と SVM による分類

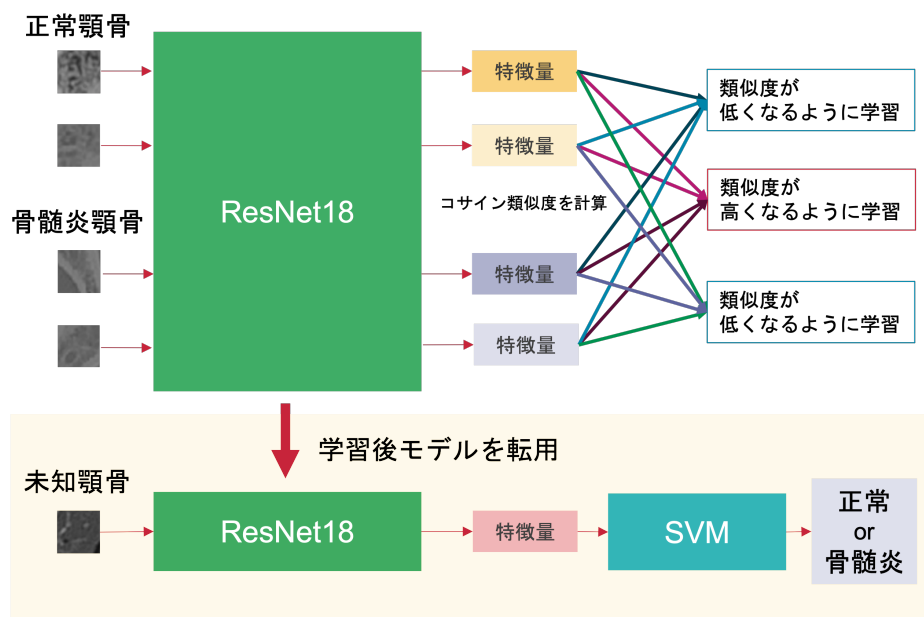


図 3.7: ラベル対照学習と SVM による分類

第4章

実験

4.1 実験データ

本研究では、9名分の頭部三次元CT画像を元に緩和処理, 厳密処理にて切り取った正常・骨髄炎ラベル付きパッチ画像を扱う(表4.1).

表4.1: 緩和処理, 厳密手法によるパッチ画像データセット

被験者番号	緩和処理データセット		厳密処理データセット	
	正常顎骨	骨髄炎顎骨	正常顎骨	骨髄炎顎骨
No.1	3886	174	4938	384
No.2	153	2	180	75
No.3	1794	228	2118	477
No.4	3165	77	4203	305
No.5	3347	732	4090	2090
No.6	4190	360	4871	1560
No.7	2453	897	3393	1462
No.8	3110	434	4103	899
No.9	1281	24	2139	152

4.2 精度評価

適合率 (*Precision*, 式 4.1) と再現率 (*Recall*, 式 4.2) の調和平均を表す F 値 (*F1score*, 式 4.3) を用いて分類精度の評価を行う (表 4.2).

表 4.2: 混同行列

		推測値	
		骨髄炎	正常
真値	骨髄炎	TP	FN
	正常	FP	TN

$$Precision = \frac{TP}{TP + FP} \quad (4.1)$$

$$Recall = \frac{TP}{TP + FN} \quad (4.2)$$

$$F1score = \frac{2Precision \times Recall}{Precision + Recall} \quad (4.3)$$

この時, TP は骨髄炎パッチを骨髄炎と識別, TN は正常パッチを正常と識別, FP は正常パッチを骨髄炎と誤認識, FN は骨髄炎パッチを正常と誤認識した数を示している. $Precision$ は骨髄炎と予測したものうちどれだけ本当に骨髄炎だったかを表し, $Recall$ はどれだけ骨髄炎を取りこぼしなく予測できたかを表す.

4.3 実験手順

緩和処理，厳密処理によるデータセットを用いて，パッチ対照学習，ラベル対照学習を行う．組み合わせにより4通りを実験する．

パッチ対照学習を行う場合，データセットの偏りは無視できる．被験者9名分のうち，6名を全て対照学習の学習用，残りの3名のうち正常顎骨の半分ずつをSVMの学習と推測に使い，骨髄炎顎骨の3割をSVMの学習，7割をSVMの推測に割り当てる．被験者No.7, 8, 9についてテストする場合には表4.3の通りとなる．SVMの学習時には，正常パッチ数と同等となるように骨髄炎パッチ数に重みを与える．

表 4.3: 緩和処理データセットを用いたパッチ対照学習にて被験者 No.7, 8, 9 についてテストする場合のデータ割り当て

被験者番号	正常顎骨				骨髄炎顎骨			
	学習		テスト	合計	学習		テスト	合計
	対照学習	SVM			対照学習	SVM		
No.1	3886	0	0	3886	174	0	0	174
No.2	153	0	0	153	2	0	0	2
No.3	1794	0	0	1794	228	0	0	228
No.4	3165	0	0	3165	77	0	0	77
No.5	3347	0	0	3347	732	0	0	732
No.6	4190	0	0	4190	360	0	0	360
No.7	0	1227	1226	2453	0	269	628	897
No.8	0	1555	1555	3110	0	130	304	434
No.9	0	641	640	1281	0	7	17	24
合計	16535	3423	3421		1573	406	949	

ラベル対照学習を行う場合、データセットが偏らないように調整する必要がある。被験者9名分のうち、6名を対照学習の学習用とし、正常顎骨データ数全体を $\frac{1}{30}$ 程度になるよう調整する。SVMについては先ほどと同様に、残りの3名のうち正常顎骨の半分ずつをSVMの学習と推測に用い、骨髄炎顎骨の3割をSVMの学習、7割をSVMの推測に割り当てる。被験者 No.7, 8, 9 についてテストする場合には表 4.4 の通りとなる。SVMの学習時には、正常パッチ数と同等となるように骨髄炎パッチ数に重みを与える。

表 4.4: 緩和処理データセットを用いたラベル対照学習にて被験者 No.7, 8, 9 についてテストする場合のデータ割り当て

被験者番号	正常顎骨				骨髄炎顎骨			
	学習		テスト	合計	学習		テスト	合計
	対照学習	SVM			対照学習	SVM		
No.1	130	0	0	130	174	0	0	174
No.2	5	0	0	5	2	0	0	2
No.3	608	0	0	60	228	0	0	228
No.4	106	0	0	106	77	0	0	77
No.5	112	0	0	112	732	0	0	732
No.6	140	0	0	140	360	0	0	360
No.7	0	1227	1226	2453	0	269	628	897
No.8	0	1555	1555	3110	0	130	304	434
No.9	0	641	640	1281	0	7	17	24
合計	1101	3423	3421		1573	406	949	

対照学習にて用いる SimCLR の学習パラメータを定める。学習率を 0.0001, 最適化関数を Adam, バッチサイズを 256, エポック数を 200, 損失関数に前述の NT-Xent, 温度定数を 0.07 とする。SVM は放射基底 (RBF) カーネルを使用し, ハイパーパラメータの C と γ は評価指標である F 値が最大となるよう調整を行う。本実験においては, プログラミング言語に Python version 3.8.10, ライブラリとして Pytorch version 1.8.1+cu111 と Scikit-learn version 0.23.1. を用いている。

また本論文では, 従来の CNN による 2 クラス分類と本手法との比較も行う。CNN の学習には十分な数のデータが必要なため, 比較的データ数の多い厳密処理にて得たデータを用いる。被験者 No.1, 2, 3 での学習とテスト, No.4, 5, 6 での学習とテスト, No.7, 8, 9 での学習とテスト, の三種類の実験を行う。対照学習後の SVM での解析に合わせ, 学習に用いるデータ数は全体の 3 割, テストに用いるデータ数は全体の 7 割とする。CNN のモデルには EfficientNet [17] を採用する。学習率を 0.001, 最適化関数を Adam, バッチサイズを 16, エポック数を 100 とし, 損失関数として交差エントロピーを用いる。

4.4 結果と考察

先行研究の評価値を表 4.5 に、本手法と CNN の評価値を表 4.6 に示す。テスト対象被験者の異なる 3 回の平均値も求めている。

4.4.1 緩和処理データセットでの評価値考察

F 値に注目して比較すると、緩和処理データセットを用いた実験で最も良い精度を示したのはパッチ対照学習であった。先行研究の SVM を使った手法は本提案手法とよく似ているが、特徴量抽出方法の違いにより本提案手法が上回った。先行研究では GLCM を用いてパッチ画像の特徴量を計算しているが、顎骨特有の特徴を捉えきれていない。本手法では対照学習により顎骨を学習しているため、顎骨パッチ画像をより適した特徴量へと変換できた。ラベル対照学習と CNN はパッチ対照学習に劣った。緩和処理はパッチに顎骨が多く含まれるよう切り取りを行うため、骨髄炎データ数が比較的少ない。それぞれのクラスにふくまれるデータ数が均等でなければならぬラベル対照学習と CNN はこのデータには不適であった。

4.4.2 厳密処理データセットでの評価値考察

同様に F 値に注目して比較すると、緩和処理データセットを用いた実験で最も良い精度を示したのは CNN であった。厳密処理データセットでの骨髄炎データ数は緩和処理データセットより多い。そのため、CNN とラベル対照学習においても F 値が高くなった。また CNN は *Recall* が非常に高く、これは骨髄炎を見逃しにくいことを示している。骨髄炎の抽出漏れが少なくなるため、再発防止策として有用である。

表 4.5: 先行研究の評価値

	Precision	Recall	F1score
SVM	0.922	0.603	0.729
NN	0.827	0.540	0.674
RF	0.925	0.333	0.490

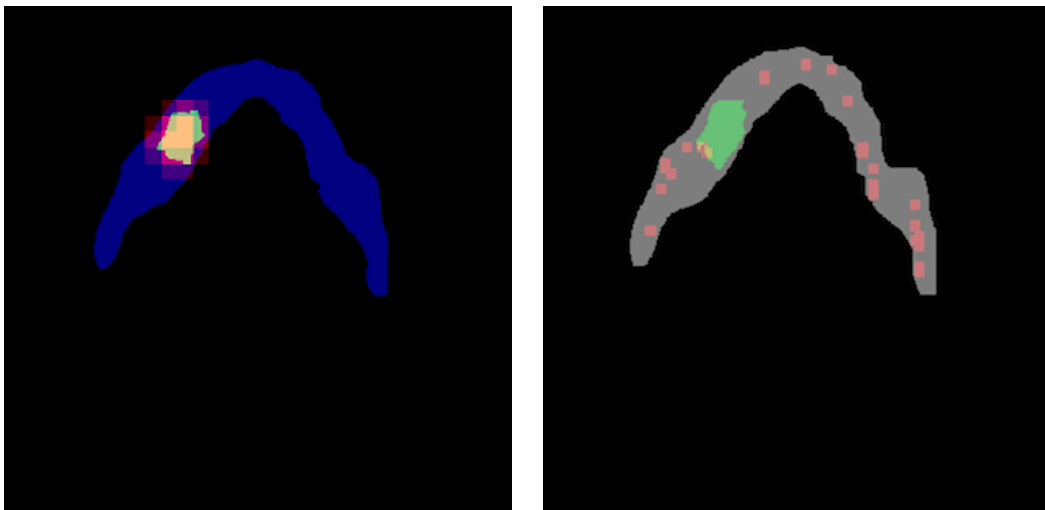
表 4.6: 対照学習と CNN の評価値

		緩和処理データセット			厳密処理データセット		
	テスト対象被験者	Precision	Recall	F1score	Precision	Recall	F1score
パッチ対照学習	No.1,2,3	0.643	0.602	0.622	0.373	0.820	0.513
	No.4,5,6	0.821	0.763	0.791	0.607	0.761	0.675
	No.7,8,9	0.885	0.709	0.788	0.523	0.776	0.626
	平均値	0.783	0.691	0.734	0.501	0.786	0.605
ラベル対照学習	No.1,2,3	0.429	0.532	0.471	0.466	0.621	0.533
	No.4,5,6	0.655	0.659	0.657	0.811	0.621	0.703
	No.7,8,9	0.710	0.584	0.641	0.547	0.656	0.597
	平均値	0.598	0.592	0.590	0.608	0.633	0.611
CNN	No.1,2,3	0.040	0.387	0.073	0.338	0.912	0.493
	No.4,5,6	0.022	0.131	0.038	0.654	0.947	0.773
	No.7,8,9	0.028	0.084	0.042	0.608	0.911	0.729
	平均値	0.030	0.201	0.051	0.533	0.923	0.665

目視評価のため、被験者 No.1 の類似 CT 画像スライスに対する提案手法の緩和処理とパッチ対照学習の組み合わせでの推定画像と先行研究の推定画像を図 4.1 に示す。黄色部が正答領域、赤色が過検出領域、緑色が未検出領域、提案手法においては青色が、先行研究においては灰色が顎骨全体である。先行研究では顎骨骨髓炎領域を全く捉えられていないのに対し、提案手法においては顎骨骨髓炎領域周辺のパッチを捉えている。

また他の被験者スライスに対する、推定成功例を図 4.2 に、推定失敗例を図 4.3 に示す。推定成功例においては、骨髓炎領域全体を覆うように推定ができた。しかし、推定失敗例では骨髓炎領域の一部を捉えるだけで、全体を把握できていない。また、骨髓炎領域が細かく散らばっている場合にはそれらをまったく検知できていない。

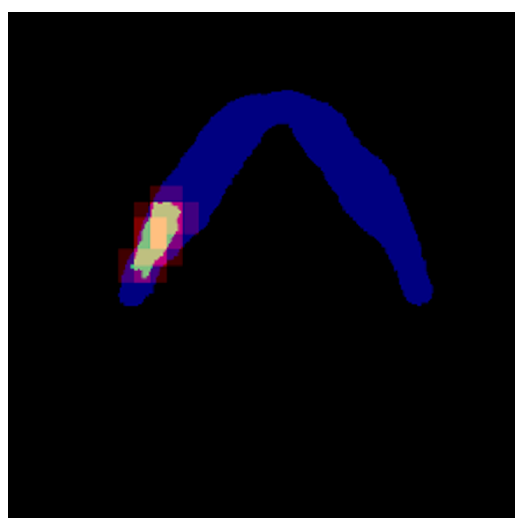
本手法はパッチ単位での推定のため、顎骨骨髓炎領域を画素単位で推定することはできない。よってどの推定スライスにおいても赤色の過検出が出現する。本手法は顎骨骨髓炎の大まかな位置を示すことは可能だが、詳細な領域は示すことができない。



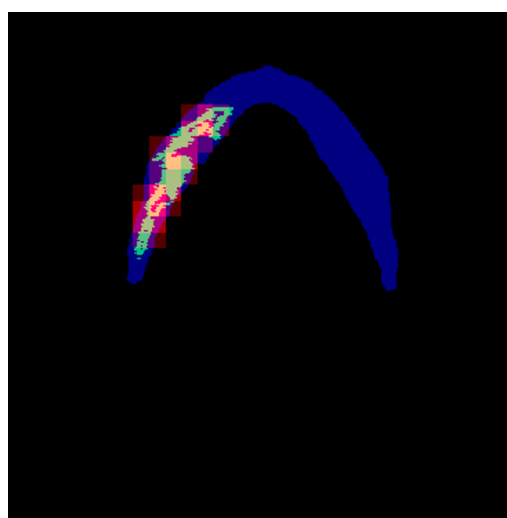
[a] 提案手法：緩和処理データセットを用いたパッチ対照学習

[b] 先行研究：SVM

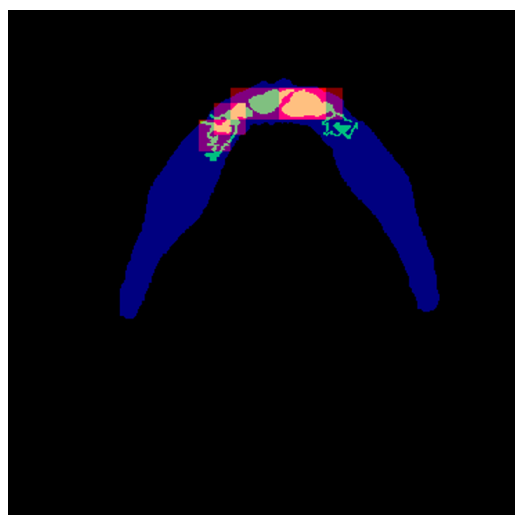
図 4.1: 被験者 No.1 に対する顎骨骨髓炎推定領域の目視評価



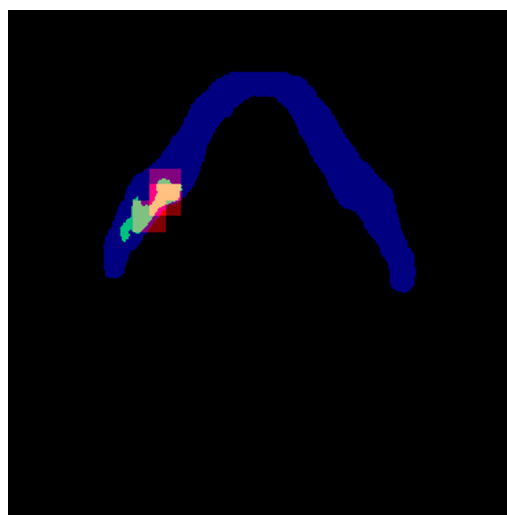
[a] 被験者 No.3



[b] 被験者 No.4

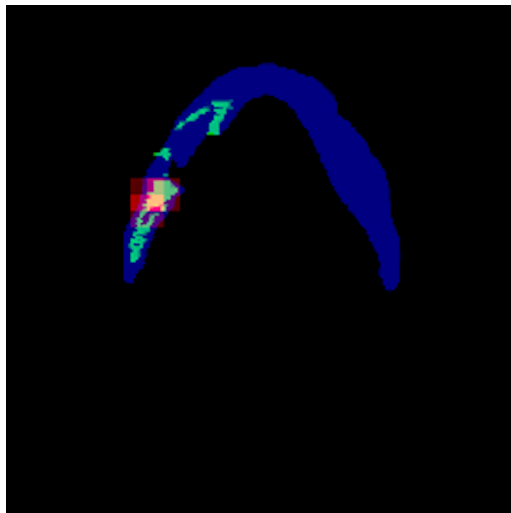


[c] 被験者 No.5

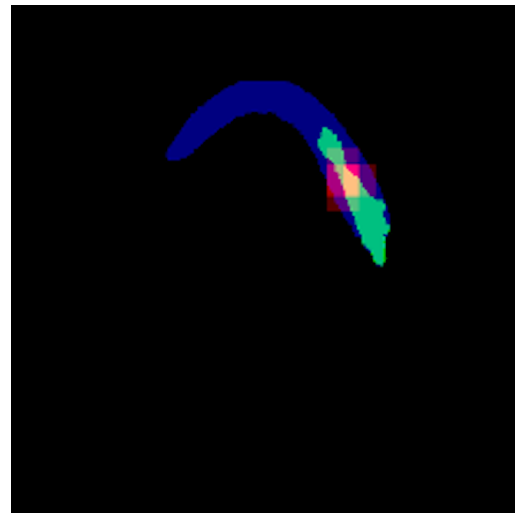


[d] 被験者 No.6

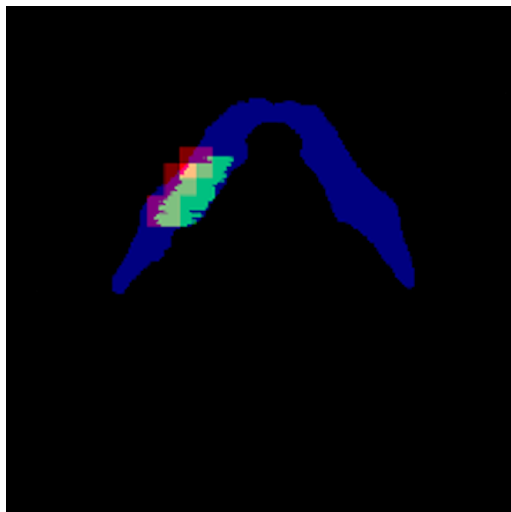
図 4.2: 緩和処理データセットを用いたパッチ対照学習の顎骨骨髓炎領域推定の成功例



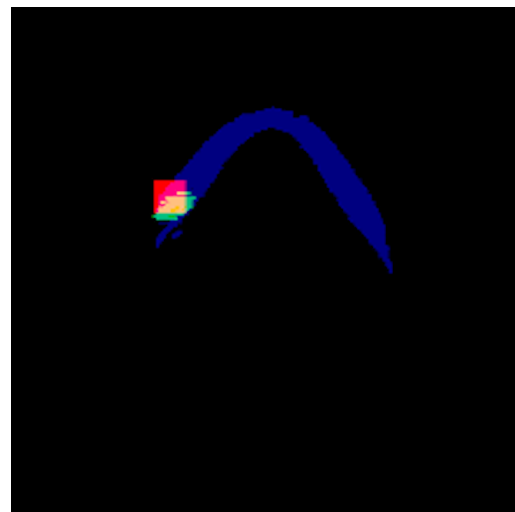
[a] 被験者 No.5



[b] 被験者 No.7



[c] 被験者 No.8



[d] 被験者 No.9

図 4.3: 緩和処理データセットを用いたパッチ対照学習の顎骨骨髓炎領域推定の失敗例

4.5 追加実験

前述の6種類の実験により、緩和処理とパッチ対照学習の推定精度が最も高いとわかった。この組み合わせを用いて追加実験を行う。被験者9名分のうち、6名分を全て対照学習の学習用とする。残りの3名のうち正常顎骨の半分ずつをSVMの学習と推測に用い、骨髄炎顎骨のうち、パッチ画像内に60%以上骨髄炎を含むものをSVMの学習に、30%以上60%以下骨髄炎を含むものをSVMの推測に用いる。ここで、顎骨骨髄炎が30%以下含まれているパッチ画像については正常顎骨として扱っている点に注意されたい。

この追加実験は、半教師あり学習手法である本手法において、医師が患者の顎骨骨髄炎パッチ画像をSVMに学習用画像として与える必要があるため、骨髄炎領域の広い、すなわち医師にとっても顎骨骨髄炎か否かを判別しやすいパッチ画像を学習用とする実験である。

実験結果を表4.7に示す。F値の平均は0.392となり、高い精度とは言えない。これは医者にとって判断の難しい顎骨骨髄炎と正常顎骨との間の曖昧な領域を、機械学習の手法でも捉えられていないことを示している。この実験ではパッチ画像内に骨髄炎顎骨を30%以下含む場合を正常顎骨としてラベル付けしている。骨髄炎顎骨と正常顎骨との間の曖昧な領域をどのようにラベル付けを行うかにより、実験結果が大きく異なってしまう。これは2クラス分類問題の限界であると考えられる。そのため、ラベルを用いた2クラス分類ではなく、パッチ画像内の顎骨骨髄炎のピクセル数、またはそのパーセンテージを用いた回帰分析手法が適すると推察される。

表 4.7: 追加実験：判別の難しい顎骨骨髄炎領域のテスト

テスト対象被験者	Precision	Recall	F1score
No.1,2,3	0.625	0.224	0.329
No.4,5,6	0.446	0.585	0.506
No.7,8,9	0.271	0.454	0.340
平均値	0.447	0.421	0.392

第 5 章

結言

本論文では発見の難しい顎骨骨髓炎領域を二種類の前処理手法により切り取り, SimCLR による対照学習と非線形 SVM を用いて推定した. その結果, F 値 0.734 を達成し, 先行研究の F 値 0.729 を上回ることができた. 推定画像から, 本手法は顎骨骨髓炎領域を覆うように捉えることができているが, 画素単位の詳細な領域を捉えきれてはいないことが確認できた.

本手法は学習とテストに同被験者のデータを用いる半自動的手法である. 自動的推測を行うためにはモデルに一般的な顎骨の特徴を学習させる必要がある. また, 顎骨骨髓炎が顎骨の一部にのみ発症する特徴を利用し, 異常検知手法による領域推定も考えられる. その場合には正常な顎骨の CT 画像が大量に必要である. 画素単位の推測を行うためにセマンティックセグメンテーションの手法を用いることが考えられるが, そのためには被験者 9 名分の頭部三次元 CT 画像だけでは十分でない. 以上を踏まえ, 本手法の改善, 異常検知手法, セマンティックセグメンテーション手法, いずれにおいても, データ不足を解消する必要がある. 複数の病院と連携し顎骨骨髓炎のデータベースを準備する, または高精度なデータ拡張の利用検討が求められる.

付録 A

付録

A.1 プログラム

本研究に関するプログラムはすべて以下のディレクトリ

- /net/nfs2/export/home/hoshino/workspace/project/kobe-oral/necrosis/hoshino

に存在する.

A.2 データ

本研究に関するデータはすべて以下のディレクトリ

- /net/nfs2/export/home/hoshino/data

に存在する.

謝辞

若林哲史教授，盛田健人助教，白井伸宙助教には、的確なご指摘，助言，論文執筆をご教授いただきました。吉永みゆき事務員には日々の研究活動においてお世話になりました。心から感謝申し上げます。また，日頃からアドバイスをしていただいた研究室の先輩方，日々の活力をいただいた同期の皆様に感謝いたします。そして最後に，私の大学生活を支えてくれた家族に今一度感謝の意を示し，本論文の結びといたします。

参考文献

- [1] K. Morita, D. Takeda, T. Hasegawa, and T. Wakabayashi, “Osteoradionecrosis area Estimation Using Machine Learning,” The 6th International Symposium on Affective Science and Engineering, No.1–A–5, 2020.
- [2] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” International conference on machine learning, PMLR, 2020.
- [3] Salvatore L. Ruggiero, Thomas B. Dodson, T. Aghaloo, Eric R. Carlson, Brent B. Ward, and D. Kademani, “American Association of Oral and Maxillofacial Surgeons’ Position Paper on Medication-Related Osteonecrosis of the Jaws—2022 Update,” Journal of Oral and Maxillofacial Surgery, Vol.80, Issue 5, pp.920–943, 2022.
- [4] O. Sung-Woon, C. Seoung-Won, B. Soo-Hwan, and Y. Byoung-Eun. “Various Therapeutic Methods for the Treatment of Medication-Related Osteonecrosis of the Jaw (MRONJ) and Their Limitations: A Narrative Review on New Molecular and Cellular Therapeutic Approaches,” Antioxidants (Basel), 2021.
- [5] S. Hayashida, S. Soutome, S. Yanamoto, S. Fujita, T. Hasegawa, T. Komori, Y. Kojima, H. Miyamoto, Y. Shibuya, N. Ueda, T. Kirita, H. Nakahara, M. Shinohara, and M. Umeda, “Evaluation of the Treatment Strategies for Medication-Related Osteonecrosis of the Jaws (MRONJ) and the Factors Affecting Treatment Outcome: A Multicenter Retrospective Study with Propensity Score Matching Analysis,” Journal of Bone and Mineral Research, Vol.30, No.10, pp.2022–2029, 2017.
- [6] K. Okuyama, S. Hayashida, S. Rokutanda, A. Kawakita, S. Soutome, S. Sawada, S. Yanamoto, Y. Kojima, and M. Umeda, “Surgical strategy for medication-related osteonecrosis of the jaw (MRONJ) on maxilla: A multicenter retrospective study,” Journal of Dental Sciences, Vol.16, Issue 3, pp.885–890, 2021.
- [7] K. Rupel, G. Ottaviani, M. Gobbo, L. Contardo, G. Tirelli, P. Vescovi, R. Di Lenarda, M. Biasotto, “A systematic review of therapeutical approaches in bisphosphonates-related

- osteonecrosis of the jaw (BRONJ),” *Oral Oncology*, 2014 Vol.50 Issue 11, pp.1049–1057, 2014.
- [8] N. R. Choi, J. H. Lee, J. Y. Park, and D. S. Hwang, “Surgical Treatment of Medication-Related Osteonecrosis of the Jaw: A Retrospective Study,” *Int J Environ Res Public Health*, Vol.17, Issue 23, 2020.
- [9] S. Sawada, Yuka Kojima, H. Yasui, M. Kirihigashi, Y. Yun, S. Hayashida, S. Rokutanda, S. Soutome, S. Yanamoto, M. Umeda, and H. Iwai, “Treatment and outcome of maxillary sinusitis associated with maxillary medication-related osteonecrosis,” *Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology*, Vol.33, Issue 4, pp.408–415, 2021.
- [10] A. H. Syaifullah, A. Shiino¹, H. Kitahara, R. Ito, M. Ishida and K. Tanigaki, “Machine Learning for Diagnosis of AD and Prediction of MCI Progression From Brain MRI Using Brain Anatomical Analysis Using Diffeomorphic Deformation,” *Frontiers in Neurology*, Vol.11, 2021.
- [11] P. Monkam, S. Qi, M. Xu, F. Han, X. Zhao, and W. Qian, “CNN models discriminating between pulmonary micro-nodules and non-nodules from CT images,” *Biomedical Engineering Online*, Vol.17, No.1, pp.96, 2018.
- [12] M. Maftouni, I.B. Turksen, M.H. Fazel Zarandi and N.A. Yazdi, “Systematic bone infection detection in axial diabetic foot MRI,” *2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*, pp.1–5, 2015.
- [13] B. Liu, W. Wang and Z. Zheng, “Fuzzy recognition of necrosis area in pathological image of virus hepatitis,” *Proceedings of the 3rd World Congress on Intelligent Control and Automation*, Vol.3, pp.1742–1746, 2000.
- [14] R. Tanaka and T. Hayashi, “Computed tomography findings of chronic osteomyelitis involving the mandible: correlation to histopathological findings,” *Dentomaxillofacial Radiology*, Vol.37, No.2, pp.94–103, 2008.
- [15] SG. Choi, EY. Lee, OJ. Lee, et al, “Prediction models for early diagnosis of actinomycotic osteomyelitis of the jaw using machine learning techniques: a preliminary study,” *BMC Oral Health*, Vol.22, No.164, 2022.
- [16] V. Vapnik and A. Lerner, “Pattern recognition using generalized portrait method,” *Automation and Remote Control*, Vol.24, pp.774–780 1963.
- [17] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” *Proceedings of the 36th International Conference on Machine Learning*, Vol. 97, pp.6105–6114, 2019.