



Title	インプラント治療の術式決定を支援する人工知能モデルの開発
Author(s)	堺, 貴彦
Citation	大阪大学, 2023, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/91846
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

学 位 論 文

インプラント治療の術式決定を
支援する人工知能モデルの開発

大阪大学大学院歯学研究科

口腔科学専攻

顎口腔機能再建学講座（歯科理工学教室）

堺 貴彦

目次

I. 緒言	1
II. (実験 1) AI モデルの選択と条件の最適化	4
1) 目的	
2) 実験方法	
①被験データ	
②AI モデルの実装	
③評価	
3) 実験結果	
①3 種の深層学習モデルを用いた AI モデルの性能	
②6 種の最小化アルゴリズムを用いた AI モデルの性能	
③3 種の学習回数での AI モデルの性能	
④異なる画像選択方法を用いた AI モデルの性能	
III. (実験 2) 抽出した画像の仕様が AI モデルの性能に与える影響の評価	13
1) 目的	
2) 実験方法	
①異なる画像サイズが AI モデルの性能に及ぼす影響の評価	
②画像の枚数が AI モデルの性能に及ぼす影響の評価	
③異なる関心領域が AI モデルの性能に及ぼす影響の評価	
3) 実験結果	
①異なる画像サイズで学習した AI モデルの性能	
②画像の枚数を変化させて学習した AI モデルの性能	
③異なる関心領域で学習した AI モデルの性能	
IV. (実験 3) AI モデルの信頼性と解釈性の評価	18
1) 目的	
2) 実験方法	
①画像間の類似度の評価	
②AI モデルが注目した領域の可視化	
3) 実験結果	
①画像の相互情報量	
②アテンションマップの重畳表示	
V. 考察	22
VI. 結論	30
VII. 謝辞	31

VIII. 参考文献	32
IX. 图表	38

I. 緒言

近年，歯科用インプラントは，欠損歯に対する補綴装置として部分的あるいは完全無歯顎患者にとって重要な選択肢となりつつある[1]．インプラントを口腔内で長期的に良好に機能させるためには，顎骨や歯周組織などの状態と，使用するインプラント体の長さやアバットメントの径などを総合的に考慮して治療計画を立案する必要がある[2]．とくに，インプラント治療の成否を左右するオッセオインテグレーション[3-5]を獲得するうえで，埋入インプラント体と骨との初期固定を確実に得ることが重要であり[6,7]，インプラント体埋入方法は，この点において治療成績に大きな影響を及ぼす因子である．

インプラント体埋入の術式や使用するインプラント体の種類は，一般的に，患者の骨質に基づいて決定される[8]．したがって，術者には，骨質を正確に診断し，それに応じて埋入窩を適切に形成することが求められる[3-5]．骨質の判断は，エックス線写真の読影と術中の術者の手指感覚によってなされるが，より客観的な情報という意味で，前者は重要な判断根拠である．以前より，Zarb ら[9]によって提唱された分類に基づいて埋入窩形成の術式を決定することで，インプラント治療の予後が向上することが知られている[10-12]．これは，エックス線写真で確認できる皮質骨と海綿骨の割合を基本にしたもので，ほとんどが皮

質骨の Type I, 厚い皮質骨と密な海綿骨からなる Type II, 薄い皮質骨と密な海綿骨からなる Type III, ならびに, 薄い皮質骨と疎な海綿骨からなる Type IV の 4 種に分類するというものである. そして, 各 Type に対応する埋入窩形成の方法として, Type I, II の骨については, 硬い骨に対して下穴を空けながらねじ切りを行うドリル (タッピングドリル) を使用しての形成, Type III の骨についてはインプラント体と同等の径までの下穴の形成, Type IV の骨に対してはインプラント体よりも小さな径までに下穴の形成を留めるという推奨が提案されている [13–15].

前述のように, インプラント治療における術前の骨質診断には, 長年, デンタルエックス線写真やパノラマエックス線写真などの 2 次元のエックス線画像が使われてきた. しかし, Cone-beam computed tomography (CBCT) の登場により, 顎骨構造を 3 次元で高解像度 (80-400 μ m) [16] で詳細に観察できるようになり, 骨の幅や高さ, 下歯槽神経の走行や隣接歯の歯根までの距離などの精密な診断が可能となった [17,19]. CBCT を応用した支援システムを用いれば, 術前に決定した位置にインプラント体を正確に埋入することができる [19] ことから, CBCT の使用はますます増加している. しかしながら, たとえ CBCT を応用したシステムで位置決定はできても, エックス線写真と手指感覚に基づいて骨質を正確に判断し, 最適な埋入窩形成の術式を決定することはそう容易ではない. と

くに手術経験の少ない歯科医師にとっては、確実な初期固定獲得のために最適な埋入窩形成の術式を術前に決定できることは、臨床的に大きな意義がある。

ところで人工知能 (Artificial intelligence; AI) は、ウェブでのキーワード検索、ソーシャルネットワーク上のコンテンツフィルタリングなど、現代社会のさまざまな場面で活用されている[20]. AI のうちの機械学習は、入力データと出力データの関係から規則性を見出すことで、未知の画像からその画像が何であるかを判別することを可能にする[20,21]. ただし、データから規則性を検出または分類するために、機械学習では高度な技術と専門知識が必要とされていた[20]. 一方、深層学習と呼ばれる技術では、予測に必要な特徴をコンピュータで自動的に抽出可能であることから、これまで人が特徴を抽出することが困難であった複雑な問題にも対応可能である[20,21]. したがって、深層学習を用いれば、術前の CBCT 画像から人が認識することが困難な特徴に基づいて骨質を客観的に判断し、埋入窩形成の術式を的確に導き出せるものと予想され、インプラントの術式決定を支援する強力なツールとなり得るものと考えられる。

そこで本研究では、実際の臨床におけるインプラント埋入前の CBCT 画像と埋入窩形成の術式に関する記録をもとに、深層学習をベースにして、さまざまな条件のもとで人工知能モデル (AI モデル) の作成を行い、確実な初期固定獲得のための埋入窩形成術式の決定を支援するツールとしての有用性を評価した。

II. (実験 1) AI モデルの選択と条件の最適化

1) 目的

深層学習の種類, 深層学習で用いる最小化アルゴリズムの種類, 学習回数, ならびに画像の選択方法をさまざまに変化させて埋入窩形成の術式を選択する AI モデルを作成し, その性能を比較検討した.

2) 実験方法

①被験データ

大阪大学歯学部附属病院口腔補綴科にてインプラント治療を受けた患者のうち, インプラント体埋入後 1 年以内でのインプラント体の喪失もしくは明らかな歯肉の炎症症状を認めた患者は除外し, 骨造成や抜歯即時埋入を受けておらず, 初期固定が得られた 27 症例 (表 1) を抽出した (大阪大学大学院歯学研究科・歯学部及び歯学部附属病院倫理審査委員会承認番号: R2-E11). なお, 初期固定獲得の基準は, インプラント体埋入時にトルクレンチで 30 N/cm 以上のトルクが得られた場合とした. これら 27 症例から, CBCT 撮影装置 (AlphardVEGA, ASAHIROENTOGEN, Kyoto) を用いて, ボクセル解像度 $200 \times 200 \times 200 \mu\text{m}$ で撮影した匿名化された CBCT 画像を得た. 1 名の歯

科医師により、インプラント埋入後の CBCT 画像から埋入部位の骨領域を矢状面断、冠状面断、水平面断の 3 方向から目視で確認し、術前の CBCT 画像における当該領域を画像解析ソフトウェア (Image J, NIH, Maryland, USA) を用いて 20×20 pixel の画像として 25 枚抽出し (図 1)、27 症例で計 675 枚の画像を得た。実際に行われた埋入窩形成の術式に基づき、27 症例を、A : タッピングドリルを使用して形成した症例 (図 2A)、B : インプラント体の径に最も近い径まで形成した症例 (図 2B)、ならびに、C : インプラント体の径よりも小さい径まで形成した症例 (図 2C) の 3 群 (9 症例ずつ) に分類した。各群 225 枚の画像 (1 症例につき 25 枚×9 症例) を、AI モデルの学習に用いる訓練用画像 200 枚 (8 症例) と、AI モデルの性能評価に用いるテスト用画像 25 枚 (1 症例) に分割した。

②AI モデルの実装

AI モデルは、Graphic processing unit を用いた深層学習モデルの実行が可能なソフトウェアである NVIDIA DIGITS (Version 6.0.0, NVIDIA, California, USA) を使用して、ThinkPad X280 (Core i7-8650 CPU, 16GB RAM, Windows 11 64 bit Pro, Lenovo) で実行した。

③評価

まず, LeNet[22], GoogleNet[23], AlexNet[24] の3種の深層学習構造で AI モデルを作成し, 訓練用画像に基づく正確率, テスト用画像に基づく正確率, 適合率, ならびに再現率を比較した. 学習回数は 100 回とした. LeNet [24] は 1998 年に LeCun によって考案された深層学習で, 人の脳を構成する神経細胞間のつながりに倣って設計されたニューラルネットワークに, 畳み込み層と呼ばれるフィルターを組み込んだ初期の深層学習である. 手書き文字の認識など, 判別される分類が少ない場合に高い性能を発揮する. GoogleNet [23]は 2014 年に考案され, フィルターサイズの異なる複数の畳み込み層を持ち, 多種の画像についての特徴抽出に優れるとされる. AlexNet [24]は 2012 年に考案され, 非常に深い畳み込み層をもつことで知られている.

AI モデルは学習時に, 正解ラベルに対応する値と予測結果に対応する値の差を誤差として, その誤差が可能なかぎり小さくなるようにパラメータを更新することで正解に近づけていく. この誤差を関数とみなし, 関数の傾きが負となる方向 (勾配降下), すなわち最小値へと向かう方向を探索する手法が勾配降下法である[25,26]. 勾配降下法においては勾配に沿って一度にどの程度降りるか (最小値を探索する範囲の大きさ) を示す学習率の設定が重要となる. 学習率が大きいと学習の効率は良くなるが, 最小値に近づいてか

らその探索が難しくなる場合がある。一方で、学習率が小さい場合は、学習の効率が下がってしまうが、最小値を正確に探索できる。勾配降下法では 6 種の最小化アルゴリズム (Adagrad[25], Adam[26], RMSprop[27], Adadelata[28], SGD[29], NAG[30]) が一般的に使用されるため、これらを用いて LeNet をベースとする AI モデルを作成し、訓練用画像に基づく正確率、テスト用画像に基づく正確率、適合率、ならびに再現率を比較した。学習回数は 100 回とした。Adagrad は学習が進むにつれ学習率を自動的に小さくすることができるため、効率的に最小値を探索できるが、最終的に学習率が 0 に収束する (学習が停止する) という欠点もある[25]。Adam は勾配の小さい方向でもパラメータの変更により、学習が停止することなく遂行されることから、適切な最小化アルゴリズムの選択が難しい場合、最初に用いられることが多い[26]。RMSprop, Adadelata は Adagrad の欠点であった学習率が 0 に収束するリスクをパラメータの調整方法で解決した手法である[27,28]。NAG は勾配降下法が最小値でない極値で学習が停滞してしまうことを回避することを目的に提案された手法である[30]。SGD は確率的勾配降下法と呼ばれ、すべてのデータの中からランダムに選択されたデータ 1 つのみで勾配を求め、パラメータの更新をデータの数だけ行う方法である[29]。

一般的に、学習の回数が多くなるにつれて AI モデルの性能は向上するこ

とが想定される．そこで，3 種の学習回数（30 回，50 回，100 回）で LeNet をベースとする AI モデルを作成し，訓練用画像に基づく正確率，テスト用画像に基づく正確率，適合率，ならびに再現率を比較した．

さらに，訓練用画像とは異なる症例からテスト用画像を選択する方法 (T1, 図 3A) と，各症例から均等配分でテスト用画像を選択する方法 (T2, 図 3B) で LeNet をベースとする AI モデルを作成し，正確率，適合率，ならびに再現率を比較した．正確率，適合率，ならびに再現率の計算は以下の式 (1) ～ (3) に従い求めた．

$$\text{正確率} = \frac{TA+TB+TC}{TA+TB+TC+FA+FB+FC} \quad (1)$$

$$\text{適合率} = \frac{TA}{TA+FB+FC}, \frac{TB}{FA+TB+FC}, \frac{TC}{FA+FB+TC} \quad (2)$$

$$\text{再現率} = \frac{TA}{TA+FA}, \frac{TB}{TB+FB}, \frac{TC}{TC+FC} \quad (3)$$

ここで，TA は術式を A と正しく分類できた回数，TB は術式を B と正しく分類できた回数，TC は術式を C と正しく分類できた回数，FA は術式を A と正しく判断出来なかった回数，FB は術式を B と正しく判断出来なかった回数，FC は術式を C と正しく判断出来なかった回数を示す．得られた結果を，表 2 に示すように術式ごとにまとめた．

3) 実験結果

① 3種の深層学習モデルを用いた AI モデルの性能

LeNet をベースとした AI モデルの訓練用画像に基づく学習曲線を図 4 に示す。学習の正確率は 88.8%であった。テスト用画像に基づく学習の正確率は 36.0%，適合率は A, B, C それぞれ 12.0%，28.0%，68.0%，再現率は A, B, C それぞれ 37.5%，50.0%，32.1%であった（表 3）。GoogleNet をベースとした AI モデルの訓練用画像に基づく学習曲線を図 5 に示す。学習の正確率は 74.4%であった。テスト用画像に基づく学習の正確率は 24.0%，適合率は A, B, C それぞれ 4.0%，12.0%，56.0%，再現率は A, B, C それぞれ 5.9%，100.0%，25.5%であった（表 4）。AlexNet をベースとした AI モデルの訓練用画像に基づく学習曲線を図 6 に示す。学習の正確率は 71.3%であった。テスト用画像に基づく学習の正確率は 36.0%，適合率は A, B, C それぞれ 28.0%，20.0%，60.0%，再現率は A, B, C それぞれ 46.7%，41.7%，31.3%であった（表 5）。3種の深層学習構造のうち、訓練用画像に基づく学習では LeNet をベースとした AI モデルの正確率が最も高く、テスト用画像に基づく学習では LeNet, AlexNet をベースとした AI モデルの正確率が最も高くなった。また、いずれの AI モデルでも術式 C の適合率が最も高くなった。

②6種の最小化アルゴリズムを用いた AI モデルの性能

最小化アルゴリズムとして Adagrad を用いた AI モデルの訓練用画像に基づく学習曲線を図 7 に示す。学習の正確率は 88.8%であった。テスト用画像に基づく学習の正確率は 36.0%, 適合率は A, B, C それぞれ 12.0%, 28.0%, 68.0%, 再現率は A, B, C それぞれ 37.5%, 50.0%, 32.1%であった (表 6)。

Adam を用いた AI モデルの訓練用画像に基づく学習曲線を図 8 に示す。学習の正確率は 93.1%であった。テスト用画像に基づく学習の正確率は 26.7%, 適合率は A, B, C それぞれ 8.0%, 24.0%, 48.0%, 再現率は A, B, C それぞれ 20.0%, 35.3%, 25.0%であった (表 7)。

RMSprop を用いた AI モデルの訓練用画像に基づく学習曲線を図 9 に示す。学習の正確率は 86.3%であった。テスト用画像に基づく学習の正確率は 34.7%, 適合率は A, B, C それぞれ 8.0%, 16.0%, 80.0%, 再現率は A, B, C それぞれ 18.1%, 57.1%, 36.8%であった (表 8)。

Adadelta を用いた AI モデルの訓練用画像に基づく学習曲線を図 10 に示す。学習の正確率は 34.4%であった。テスト用画像に基づく学習の正確率は 40.0%, 適合率は A, B, C それぞれ 0.0%, 24.0%, 96.0%, 再現率は A, B, C それぞれ 0.0%, 85.7%, 33.3%であった (表 9)。

SGD を用いた AI モデルの訓練用画像に基づく学習曲線を図 11 に示す。学習の正確率は 73.1%であった。テスト用画像に基づく学習の正確率は 34.7%, 適合率は

A, B, C それぞれ 4.0%, 4.0%, 96.0%, 再現率は A, B, C それぞれ 100.0%, 50.0%, 33.3%であった (表 10). NAG を用いた AI モデルの訓練用画像に基づく学習曲線を図 12 に示す. 学習の正確率は 71.9%であった. テスト用画像に基づく学習の正確率は 41.3%, 適合率は A, B, C それぞれ 8.0%, 16.0%, 100.0%, 再現率は A, B, C それぞれ 66.7%, 100.0%, 36.8%であった (表 11). 6 種の最小化アルゴリズムのうち, 訓練用画像に基づく学習では Adam を用いた場合に最も正確率が高く, テスト用画像に基づく学習の正確率では NAG を用いた場合に最も高くなった.

③3 種の学習回数での AI モデルの性能

学習回数を 30 回とした AI モデルの訓練用画像に基づく学習曲線を図 13 に示す. 学習の正確率は 90.6%であった. テスト用画像に基づく学習の正確率は 29.3%, 適合率は A, B, C それぞれ 8.0%, 20.0%, 60.0%, 再現率は A, B, C それぞれ 16.7%, 45.5%, 28.8%であった (表 12). 学習回数を 50 回とした AI モデルの訓練用画像に基づく学習曲線を図 14 に示す. 学習の正確率は 93.1%であった. テスト用画像に基づく学習の正確率は 34.7%, 適合率は A, B, C それぞれ 12.0%, 28.0%, 69.6%, 再現率は A, B, C それぞれ 33.3%, 45.5%, 31.3%であった (表 13). 学習回数を 100 回とした AI モデルの訓練

用画像に基づく学習曲線を図 15 に示す。学習の正確率は 93.1%であった。

テスト用画像に基づく学習の正確率は 36.0%，適合率は A，B，C それぞれ 12.0%，28.0%，68.0%，再現率は A，B，C それぞれ 37.5%，50.0%，32.1%であった（表 14）。

④異なる画像選択方法を用いた AI モデルの性能

訓練用画像とは異なる症例からテスト用画像を選択する方法（T1）で作成した AI モデルの訓練用画像に基づく学習曲線を図 16 に示す。学習の正確率は 93.1%であった。テスト用画像に基づく学習の正確率は 34.7%，適合率は A，B，C それぞれ 12.0%，28.0%，69.6%，再現率は A，B，C それぞれ 33.3%，45.5%，31.3%であった（表 15）。各症例から均等配分でテスト用画像に基づく方法（T2）で作成した AI モデルの訓練用画像に基づく学習曲線を図 17 に示す。学習の正確率は 92.5%であった。テスト用画像に基づく学習の正確率は 94.7%，適合率は A，B，C それぞれ 92.0%，96.0%，96.0%，再現率は A，B，C それぞれ 100.0%，96.0%，88.9%であった（表 16）。訓練用画像に基づく学習の正確率は T1 で作成した AI モデルが高くなった一方で、テスト用画像に基づく正確率では T2 が大幅に高くなった。

III. (実験 2) 抽出した画像の仕様が AI モデルの性能に与える影響の評価

1) 目的

実験 1 では、高性能な AI モデルを作成するためのベースとなる深層学習モデルと学習条件を決定した。しかし、AI モデルの性能をさらに向上させるためには、学習に用いる入力データの質を上げる必要がある。そのため、実験 2 では、入力として用いる画像の特徴（サイズ、枚数）や抽出する領域が AI モデルの性能に与える影響を評価した。

2) 実験方法

実験 1 と同様に、AI モデルは、Graphic processing unit を用いた深層学習モデルの実行が可能なソフトウェアである NVIDIA DIGITS (Version 6.0.0, NVIDIA) を使用して、ThinkPad X280 (Core i7-8650 CPU, 16GB RAM, Windows 11 64 bit Pro, Lenovo) で実行した。学習条件は、実験 1 で得られた結果に基づき、ベースとなる AI モデルとして LeNet を選択し、誤差を最小化するためのアルゴリズムとして Adam, 学習回数を 50 回、ならびに画像の選択方法を T2 とした。

① 異なる画像サイズが AI モデルの性能に及ぼす影響の評価

訓練用画像, テスト用画像として 3 種の画像サイズ (5×5 pixel, 10×10 pixel, 20×20 pixel) で学習した AI モデルを作成し, 正確率, 適合率, ならびに再現率を求め, 作成した AI モデルの性能を比較した.

② 画像の枚数が AI モデルの性能に及ぼす影響の評価

画像の枚数が AI モデルの性能に及ぼす影響を評価するために, これまでの実験で用いた症例数を 27 から 60 (表 17) に増やし, 画像枚数を 60 症例×20 枚=1,200 枚とした. 訓練用画像, テスト用画像の合計を 2 種の枚数 (675 枚, 1,200 枚) で学習した AI モデルを作成し, 正確率, 適合率, ならびに再現率を求め, 作成した AI モデルの性能を比較した.

③ 異なる関心領域が AI モデルの性能に及ぼす影響の評価

初期固定の獲得には皮質骨の存在が大きくかかわるとされている[31]ため, 骨縁の皮質骨を含む領域 (インプラント体のプラットフォームに相当する領域) と, 皮質骨を含まない海綿骨のみの領域 (図 18) を各々 20×20 pixel で 抽出した画像を学習した AI モデルを作成し, 正確率, 適合率, ならびに再現率を求め, 作成した AI モデルの性能を比較した.

3) 実験結果

① 異なる画像サイズで学習した AI モデルの性能

関心領域として抽出した画像のサイズを 5×5 pixel とした AI モデルの訓練用画像に基づく学習曲線を図 19 に示す. 学習の正確率は 50.0%であった. テスト用画像を用いた学習の正確率は 60.0%, 適合率は A, B, C それぞれ 36.7%, 75.0%, 42.9%, 再現率は A, B, C それぞれ 88.0%, 24.0%, 12.0%であった (表 18). 関心領域として抽出した画像のサイズを 10×10 pixel とした AI モデルの訓練用画像に基づく学習曲線を図 20 に示す. 学習の正確率は 61.3%であった. テスト用画像に基づく学習の正確率は 56.0%, 適合率は A, B, C それぞれ 41.7%, 72.2%, 61.9%, 再現率は A, B, C それぞれ 60.0%, 52.0%, 52.0%であった (表 19). 関心領域として抽出した画像のサイズを 20×20 pixel とした AI モデルの訓練用画像に基づく学習曲線を図 21 に示す. 学習の正確率は 95.2%であった. テスト用画像に基づく学習の正確率は 94.7%, 適合率は A, B, C それぞれ 92.0%, 96.0%, 96.0%, 再現率は A, B, C それぞれ 100.0%, 96.0%, 88.9%であった (表 20). 3 種の画像サイズのうち, 20×20 pixel の画像を入力データとして学習した AI モデルが, 訓練用画像, テスト用画像に基づいた学習ともに最も高い正確率を示した.

② 画像の枚数を変化させて学習した AI モデルの性能

画像枚数 675 枚で学習した AI モデルの訓練用画像に基づく学習曲線を図 22 に示す。学習の正確率は 95.2%であった。テスト用画像に基づく学習の正確率は 94.7%，適合率は A，B，C それぞれ 92.0%，96.0%，96.0%，再現率は A，B，C それぞれ 100.0%，96.0%，88.9%であった（表 21）。1,200 枚で学習した AI モデルの訓練用画像に基づく学習曲線を図 23 に示す。学習の正確率は 87.8%であった。テスト用画像に基づく学習の正確率は 78.2%，適合率は A，B，C それぞれ 77.6%，75.4%，80.0%，再現率は A，B，C それぞれ 81.8%，78.2%，76.2%であった（表 22）。1,200 枚の画像を入力データとして学習した AI モデルは、訓練用画像、テスト用画像に基づく学習の双方で、正確率が減少した。

③ 異なる関心領域で学習した AI モデルの性能

海綿骨のみを含む領域の画像を入力データとして学習した AI モデルの訓練用画像に基づく学習曲線を図 24 に示す。学習の正確率は 87.8%であった。テスト用画像に基づく学習の正確率は 78.2%，適合率は A，B，C それぞれ 77.6%，75.4%，80.0%，再現率は A，B，C それぞれ 81.8%，78.2%，72.7%であった（表 23）。皮質骨を含む領域の画像を入力データとして学習した AI モ

デルの訓練用画像に基づく学習曲線を図 25 に示す．学習の正確率は 100.0%であった．テスト用画像に基づく学習の正確率は 95.2%，適合率は A, B, C それぞれ 90.0%, 96.4%, 98.0%，再現率は A, B, C それぞれ 98.2%, 96.3%, 89.1%であった（表 24）．皮質骨を含む領域の画像を入力データとして学習した AI モデルの正確率は，訓練用画像，テスト用画像に基づく学習の双方で向上した．

IV. (実験 3) AI モデルの信頼性と解釈性の評価

1) 目的

実験 1 で AI モデルのベースとなる深層学習モデルと学習条件の最適化を行い、実験 2 では AI モデルの性能を向上する入力データの仕様を決定した。実験 3 では、作成した AI モデルが過学習していないことを検証するために、訓練用画像とテスト用画像の類似度を評価した。また、作成した AI モデルは高い性能を発揮したものの、深層学習では予測の過程がブラックボックスであるため、臨床応用を検討するためには判断の根拠を解明する必要がある。そこで、予測した結果を解釈するために、AI モデルが注目した領域の可視化を行った。

2) 実験方法

① 画像間の類似度の評価

術式 A, B, C の 3 種の分類ごとに、訓練用画像 (320 枚) とテスト用画像 (80 枚) の全ての組み合わせ ($320 \times 80 = 25,600$ 通り) で相互情報量[32]を求め、比較した。相互情報量は 2 つの情報間の依存度を表す指標であり、以下の関係が成立する。

1. X と Y が最も依存していない（独立している）ときに、相互情報量は最小となる.
2. X と Y が最も依存している（同一である）ときに、相互情報量は最大となる.

すなわち、画像においては、画像 X の画素値と画像 Y の画素値に対して以下の方法で相互情報量を計算することで類似度を評価することができる [32]. 画像 X, Y (階調 N) のすべての画素を網羅し、画像 X の画素値 (X_i), 画像 Y の画素値 (Y_j) の組み合わせ (X_i, Y_j) の出現頻度からヒストグラムを作成する. その後、画素 X_i の出現確率 $P(X_i)$, 画素 Y_j の出現確率 $P(Y_j)$, 画素 X_i, Y_j の組み合わせの出現確率 $P(X_i, Y_j)$ を算出し、以下の式 (4) に代入することで相互情報量 (MI) を求める.

$$MI = \sum_i^N \sum_j^N P(X_i, Y_j) \log_2 \frac{P(X_i, Y_j)}{P(X_i)P(Y_j)} \quad (4)$$

テスト用画像 80 枚について、 $80 \times 80 = 6,400$ 通りの相互情報量を求め、その中の最大値を、全く同じ画像が含まれていた場合の相互情報量とした.

② AI モデルが注目した領域の可視化

AI モデルの予測に対する局所的な説明（解釈性）を与えるために考案された Gradient-weighted Class Activation Mapping (Grad-CAM) [33,34]を、実験 1 と実験 2 を通して得られた AI モデル（図 26）の特徴抽出器の最後の層に追加し、術式の予測に影響を与えた特徴を抽出することで（図 27）得られたアテンションマップをテスト用画像上に重畳表示した。実験 1 と実験 2 で用いた NVIDIA DIGITS では、既存の深層学習モデルの編集ができないため、Keras ライブラリ (Version 2.2.4) と呼ばれる深層学習の実装に必要な関数群を用いて、LeNet をベースとした AI モデルをプログラミング言語のひとつである Python (Version 3.7.2) で実装し、特徴抽出器の最後の層に Grad-CAM を加えた。各種プログラムは Atom エディター (Version 1.60.0) で記述し、ThinkPad X280 (Core i7-8650 CPU, 16GB RAM, Windows 11 64 bit Pro, Lenovo) でコマンドプロンプトから実行した。

3) 実験結果

① 画像の相互情報量

術式 A, B, C それぞれの画像の相互情報量を図 28-30 に示す。テスト用画像の相互情報量の最大値は、術式 A, B, C それぞれ 7.390, 7.161, 7.295 であっ

た．訓練用画像とテスト用画像のすべての組み合わせで相互情報量を求めた結果，いずれの値もテスト用画像の相互情報量の最大値を超えなかったことから，AI モデルの学習に用いた訓練用画像とテスト画像の間に重複する画像の組み合わせが存在しないことが確認された．

②アテンションマップの重畳表示

Grad-CAMで生成したアテンションマップを図31に示す．赤く色づけされた領域は，AIモデルが予測の過程でとくに参考にした領域を示している．成功例と失敗例の典型的なアテンションマップを図32に示す．成功例ではA, Bの術式については暗い画素に注目している傾向が認められるのに対して，術式Cでは明るい画素に注目している画像が多くみられた．失敗例では前述の傾向が逆転していることもあり，画素の明暗が適切に判別されていない画像が多くみられた．

V. 考 察

科学，工学，医療分野において，AI は画像分類や物体認識などすでに多くの領域で応用されている[35]. AI の中でも機械学習は，人が特徴を定義したデータからパターンを学習することで，新たに与えられたデータの挙動を予測する[36]. 機械学習のうち，とくに人の脳を構成する神経細胞間のつながりに倣って提案されたニューラルネットワークに基づいた学習法が深層学習[20,37,38]であり，歯科領域においても，歯周病[39–42]やう蝕[43–48]病変の検知のために応用が試みられている．補綴領域では，インプラント体の判別[49–54]，オッセオインテグレーションやインプラント治療の予後の予測[55–58]，インプラント体デザインの最適化[59–61]を行う AI モデルなどが報告されているが，インプラント埋入手術における診断支援を行う AI モデルについてはいまだ作成されていない．インプラント埋入窩形成の術式は，術者がエックス線写真の読影と術中の手指感覚によって骨質を診断することで選択される．エックス線写真は，より客観的な情報として重要であり，これを基に術式を導き出す AI モデルの作成に成功すれば，これまで困難であった術前の術式決定が可能になり，インプラント治療の成功率を高めることに繋がるものと考えられる．そこで本研究では，術前の CBCT 画像と実際に行われたインプラント埋入窩形成の術式を深層学習により学習する

ことで、CBCT 画像からインプラント埋入窩形成の術式を決定できる AI モデルの作成を目指した。

深層学習は、特徴量の抽出が困難なデータセットに対して、事前に正解をラベル付けしたデータを必要とせずにコンピュータが特徴を自動的に抽出することを可能とする手法として開発された[62]。その後、LeCun ら[20]により、畳み込み層と呼ばれるフィルターをニューラルネットワークへ組み込む手法が確立され、各層で抽出される特徴量の数が削減されることで不変的な特徴量を効率的に抽出することが可能となった。LeNet は、LeCun らによって開発された古典的な畳み込み層を持つニューラルネットワークで構成されており、アルツハイマー病の発見（2 分類、画像枚数 36,120 枚、正確率：96.85%）[63]、交通状況の認識（6 分類、画像枚数 51,839 枚、正確率：99.75%）[64]、人の表情の認識（6 分類、画像枚数 981 枚、正確率：97.6%）[65]、幼児歩行者の認識（2 分類、画像枚数 64,468 枚、正確率：93.7%）[66]などに利用されている。本研究で作成した LeNet をベースとした AI モデル（3 分類、学習画像枚数 1,200 枚、正確率：95.8%）も、これらの先行研究と同様に高い性能を発揮することに成功した。LeNet が白黒の手書き文字を識別するために開発された深層学習モデルであることから、白黒の濃淡で表現された CBCT 画像でも高い性能を発揮したものと考えられた。GoogleNet と AlexNet は、カラー画像を入力として学習することを前提に開発さ

れており，入力画像のサイズも LeNet の 32×32 pixel と比べて 256×256 pixel と大きくなる[23,24]．本研究で用いたサイズの小さな濃淡画像を入力しても，元の画像が疑似的に引き延ばされてしまうことから，十分に特徴量の抽出が出来なかった可能性が考えられた．

AI モデルの学習で重要となるのが，誤差を最小にするために用いられるパラメータを最適化する方法である．AI モデルは，学習時に，正解ラベルに対応する確率と予測結果に対応する確率の差を誤差として，その誤差が可能なかぎり小さくなるように重み係数を更新することで真の解に近づける[25,26]．この誤差を表現する関数は様々であり，どの方法を使用するかで結果が変化する[25]．深層学習では勾配降下法[26,67]が一般的に用いられ，SGD (Stochastic Gradient Descent) [29,30]，NAG (Nestrov accelerated gradient descent) [30]，Adagrad (Adaptive Gradient Algorithm) [25]，Adadelata[28]，RMSprop[27]，ならびに Adam[26]の 6 種が代表的な最適化アルゴリズムとして提案されている．訓練用画像に基づいた学習においては，Adam，Adagrad，RMSprop，SGD，Adadelata，NAG の順に高い正確率を示した．これは，Adam や Adagrad が，SGD や NAG では誤差を小さくするためのパラメータ調整が困難な学習を扱うことができるという特長に一致する[26]．また RMSprop では，学習に不必要な過程を省くための補正を行わないため，パラメータの更新回数が多くなる傾向がある[27]．Adam は，Adagrad と

RMSprop が共に有する，誤差を小さくするためのパラメータ調整能力を組み合わせたアルゴリズムであり，様々なパラメータを含む誤差の最小化に適している[26]ことから，色情報や境界が明瞭な構造などが無い画像でも高い性能を示したものと考えられた．テスト用画像を用いた学習では Adam がもっとも低い正確率を示したが，これは各症例から均等配分でテスト用画像を選択する手法により解決した．

学習とは，訓練用画像を用いて予測した値と正解値との誤差が，一定の閾値を満たすまで何度も繰り返す処理のことを指すが，この繰り返し回数のことを学習回数と呼ぶ．一般的に，学習回数が多くなるにつれて，訓練用画像を用いた学習における誤差は小さくなり，AI モデルの性能は向上していくが，学習中のある時点で再び低下し始めることがある[68–70]．これは過学習と呼ばれ，訓練用画像に対しては高い性能を発揮するものの，学習に用いられていない，未知の画像に対する予測値と正解値の誤差が増加する状態である．深層学習における過学習を防ぐ手法として，層の総数を小さくする方法[69]と，誤差が大きくなった時点で学習を停止する方法[68–70]がある．誤差が大きくなった時点で学習を停止する方法の代表的なものが Early stopping 法[68–70]である．Early stopping 法は理解や実装が容易であり，他の手法よりも優れているとの報告が多い[68–70]ため，広く用いられている．Early stopping 法は，(1) 学習に用いるデータを訓練集

合と検証集合に分割し，(2) 訓練集合のみで学習を行った後に検証集合で適宜サンプルごとの誤差を評価する，(3) 検証集合における誤差が前回の評価よりも大きくなった時点で学習を停止する，という手法である[70]．本研究において，訓練用画像を用いた学習過程において，学習回数が 30 回の AI モデルの正確率は 90.6%となり，50 回，100 回では 93.1%であった．テスト用画像を用いた学習では，学習回数が 30 回の AI モデルの正確率は 29.3%となり，50 回では 34.7%，100 回では 36.0%となった．学習回数が増えるにつれて AI モデルの性能は向上したが，50 回での誤差が 0.57 となり，100 回での誤差が 0.63 となったことから，誤差が学習の経過にともない減少から増加に転じた時点で学習を停止するという Early stopping の手法[70]にしたがい，学習回数を 50 回とするのが最適であると判断された．

T2 を用いた場合，テスト用画像と重複しない画像を AI モデルの訓練に使うことが，高い性能の発揮に寄与していると考えられた．一般的な医療用画像の学習においては，同一の患者から得られた複数のスライスを学習に用いることは過学習を誘発するために避けるべきとされている[32]．しかし，本研究で用いた CBCT 画像の関心領域は，サイズが小さく (20×20 pixel)，同一患者の隣接するスライス間であっても，画像の類似度の指標である相互情報量[32]が小さくなったことから，テスト用画像に用いた画像が，訓練用画像と重複していない，す

なわち過学習を誘発しない条件で AI モデルの訓練が行われたことが示された。

初期固定の獲得には皮質骨の存在が大きくかかわるとされており[31]，本研究においても，皮質骨を多く含む領域に着目して抽出した 20×20 pixel の領域を入力画像として学習した AI モデルが最も高い性能を示した．初期固定の獲得に関わる皮質骨[33]を多く含む領域を可能な限り広く抽出することで，AI モデルが正確に埋入窩形成の術式を選択できることが分かった。

1,200 枚の画像で学習した場合，675 枚の場合と比較して性能が低下したことから，単に画像枚数を増やすことよりも，入力画像として用いる画像の質を向上することの方が，AI の性能が向上する可能性が示唆された．これは質の高い特徴量を学習することで AI モデルの性能が向上すること[71,72]と一致する．関連性の高い特徴を選択し，関連性のない特徴を排除することは機械学習において重要である[71,72]．また，特徴量の選択は元の特徴集合から部分集合を得る過程であり，冗長な特徴や無関係な特徴を除外し，データ処理の規模を縮小する役割を果たす[71,72]．優れた特徴量の選択は学習精度を向上するとともに，学習時間の短縮に有効で，AI モデルの社会実装にとって重要な項目と言える[73,74]．深層学習モデルは，様々なコンピュータビジョンのタスク（画像分類[75,76]，物体認識[77]，物体検出[78]など）において革新的発展をたどっているが[33]，優れた性能を発揮する一方で，個々の構成要素に分解することができないため，AI モ

デルの透明性（どのように機能しているか）や AI モデルの事後的な説明（結果からわかること）の解釈が困難である[79]. その結果，作成した AI モデルが誤った判断をした場合，警告や説明もないために，ユーザーは誤った出力に対してなぜ誤ったかの解釈をすることができない [33]. したがって，作成した AI モデルを社会実装するためには，説明可能なモデルとする必要がある[33]. 本研究で用いた CBCT 画像には明瞭な特徴は認められないことから，画像のみから人が埋入窩形成の術式を予測することは難しい．しかしながら，作成した AI モデルでは高い正確率で術式を予測することが可能であったため，その判断の根拠となった画像内の着目部位を確認する必要があった．実験 3 で追加された Grad-CAM により生成されたアテンションマップでは，AI モデルが，人の目視では識別することの難しい骨梁構造に注目していることが分かった．すなわち，AI モデルでは，抽出された特定の領域の特徴（骨梁構造）から，その分布比率に基づいて骨質を正確に判断し，よりの確に埋入窩形成の術式を選択できるものと考えられた．

本研究では，術前の CBCT 画像から埋入窩形成の術式を導き出すことに成功したが，用いた CBCT 画像は全て同一の装置で撮影されたものであるため，他の装置で撮影された CBCT 画像に対してどれほどの性能を示すかは現時点では不明である．また，インプラント体の形状が初期固定の獲得に影響を与えるとい

う報告[80]もあり，本研究では埋入されたインプラント体の形状は分類せずに学習しているため，臨床応用に向けて頑健性を高める必要がある．一方，埋入窩形成の術式に関しては，本研究で用いた 3 種以外のもの[81]を追加することで，比較的容易に，より臨床に即した AI モデルの作成が可能である．さらに，低解像度の入力画像を高解像度化して出力することができる超解像度化技術[82]を用いて，低解像度の CBCT 画像では抽出が困難な骨梁構造など，入力画像から多くの特徴を抽出することで，より高い性能の AI モデルの作成が可能であると考えられ，本研究において開発した AI モデルのさらなる改良に向けて取組みを進めたいと考えている．

VI. 結 論

本研究において，深層学習を用いて適切な条件で AI モデルを作成し，皮質骨を含む領域から抽出した画像を用いれば，高い性能で埋入窩形成の術式を選択できることが明らかとなった．以上のことから，本研究で開発した AI モデルは，インプラント治療の術式決定を支援するツールとして有用であることが示された．

VII. 謝 辞

稿を終えるにあたり，本研究の機会を与えていただき，御指導と御高配を賜りました大阪大学大学院歯学研究科歯科理工学教室の今里 聡教授に対し，深甚なる謝意を表します．

また，本研究の遂行にあたり，大阪大学大学院歯学研究科歯科理工学教室の山口 哲准教授に心より感謝申し上げます．

最後に，本研究を行うに際し，多大なる御協力と御助言を頂いた大阪大学大学院歯学研究科クラウンブリッジ補綴学分野の中野 環助教ならびに歯科理工学教室の教室員各位に厚く御礼申し上げます．

VIII. 参 考 文 献

1. Naser S, Amir M, Hamid S. Comparing the quality of life of patients requesting dental implants before and after implant. *Open Dent J* 11: 485–91, 2017.
2. Chuang SK, Wei LJ, Douglass CW, Dodson TB. Risk factors for dental implant failure: a strategy for the analysis of clustered failure-time observations. *J Dent Res* 81(8): 572–7, 2002.
3. Maria BG, Daniel GO, Romulo LC. Research on implants and osseointegration. *Periodontolo* 2000 79(1): 178–89, 2019.
4. Cochran DL, Morton D, Weber HP. Consensus statements and recommended clinical procedures regarding loading protocols for endosseous dental implants. *Int J Oral Maxillofac Implants* 19(Suppl): 109–13, 2004.
5. Aparicio C, Rangert B, Sennerby L. Immediate/early loading of dental implants: a report from the Sociedad Española de Implantes World Congress consensus meeting in Barcelona, Spain. *Clin Implant Dent Relat Res* 5(1): 57–60, 2003.
6. Nkenke E, Hahn M, Weinzierl K, Radespiel-Tröger M, Neukam FW, Engelke K. Implant stability and histomorphometry: a correlation study in human cadavers using stepped cylinder implants. *Clin Oral Implants Res* 14(5): 601–9, 2003.
7. Nasatzky E, Gultchin J, Schwartz Z. The role of surface roughness in promoting osteointegration. *Refuat Hapeh Vehashinayim* 20(3): 8–19, 2003.
8. Ekfeldt A, Christiansson U, Eriksson T, Lindén U, Lundqvist S, Rundcrantz T, Johansson LA, Nilner K, Billström C. A retrospective analysis of factors associated with multiple implant failures in maxillae. *Clin Oral Implants Res* 12(5): 462–7, 2001.
9. Zarb GA, Zarb FL. Tissue integrated dental prostheses. *Quintessence Int* 16(1): 39–42, 1985.
10. Miyamoto I, Tsuboi Y, Wada E, Suwa H, Iizuka T. Influence of cortical bone thickness and implant length on implant stability at the time of surgery—clinical, prospective, biomechanical, and imaging study. *Bone* 37(6): 776–80, 2005.
11. Herrmann I, Lekholm U, Holm S, Kultje C. Evaluation of patient and implant characteristics as potential prognostic factors for oral implant failures. *Int J Oral Maxillofac Implants* 20(2): 220–30, 2005.
12. Ikumi N, Tsutsumi S. Assessment of correlation between computerized tomography values of the bone and cutting torque values at implant placement: a clinical study. *Int J Oral Maxillofac Implants* 20(2): 253–60, 2005.
13. Pauwels R, Jacobs R, Singer SR, Mupparapu M. CBCT-based bone quality assessment: are Hounsfield units applicable? *Dentomaxillofac Radiol* 44(1): 279–85, 2015.
14. Beer A, Gahleitner A, Holm A, Tschabitscher M, Homolka P. Correlation of insertion torques with bone mineral density from dental quantitative CT in the mandible. *Clin Oral Implants Res*

- 14(5): 616–20, 2003.
15. Shalabi MM, Wolke JGC, de Ruijter AJE, Jansen JA. Histological evaluation of oral implants inserted with different surgical techniques into the trabecular bone of goats. *Clin Oral Implants Res* 18(4): 489–95, 2007.
16. Pauwels R, Jacobs R, Bogaerts R, Bosmans H, Panmekiate S. Determination of size-specific exposure settings in dental cone-beam CT. *Eur Radiol* 27(1): 279–85, 2017.
17. Braut V, Bornstein M, Kuchler U, Buser D. Bone dimensions in the posterior mandible: a retrospective radiographic study using cone beam computed tomography. Part 2— analysis of edentulous sites. *Int J Periodontics Restorative Dent* 34(5): 639–47, 2014.
18. Bornstein MM, Horner K, Jacobs R. Use of cone beam computed tomography in implant dentistry: current concepts, indications and limitations for clinical practice and research. *Periodontol 2000* 73(1): 51–72, 2017.
19. Kernen F, Kramer J, Wanner L, Wismeijer D, Nelson K, Flügge T. A review of virtual planning software for guided implant surgery - data import and visualization, drill guide design and manufacturing. *BMC Oral Health* 20(1): 251–6, 2020.
20. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 521(7553): 436–44, 2015.
21. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. *Adv Exp Med Biol* 1213: 3–21, 2020.
22. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 86(11): 2278–324, 1998.
23. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. *Proc IEEE Conf Comput Vis Pattern Recognit*: 1–12, 2014.
24. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 60(6): 84–90, 2017.
25. Duchi J, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Res* 12: 2121–59, 2011.
26. Kingma DP, Ba J. Adam: a method for stochastic optimization. *Proc Int Conf Learn Rep*: 1–13, 2015.
27. Tieleman T, Hinton GE. Lecture 6.5-rmsprop, coursera: neural networks for machine learning. *Tech Report 6*: 26–31, 2012.
28. Zeiler MD. ADADELTA: an adaptive learning rate method. *ArXiv*: 1212.5701, 2012.
29. Bottou L. Online learning and stochastic approximations. *Cambridge Uni Press*, pp.2-7, 1998.
30. Nesterov Y. A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$. *Mathe Doklady*: 543–7, 1983.
31. Chrcanovic BR, Albrektsson T, Wennerberg A. Bone quality and quantity and dental implant failure: a systematic review and meta-analysis. *Int J Prosthodont* 30(3): 219–237, 2012.

32. Thomas M, Thomas A. Elements of information theory 2nd edition. New York Acad Press, pp.13-24, 2006.
33. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis* 128(2): 336–59, 2020.
34. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang GZ. XAI-Explainable artificial intelligence. *Sci Robot* 4(37): 1347–52, 2019.
35. Revilla-León M, Gómez-Polo M, Vyas S, Barmak BA, Galluci GO, Att W, Krishnamurthy VR. Artificial intelligence applications in implant dentistry: a systematic review. *J Prosthet Dent* 3913(21): 272–9, 2021.
36. Wooldridge M, Jennings NR. Intelligent agents: theory and practice. *Knowl Eng Rev* 10(2): 115–52, 1995.
37. Park WJ, Park JB. History and application of artificial neural networks in dentistry. *Eur J Dent* 12(04): 594–601, 2018.
38. Legg S, Hutter M. Universal intelligence: a definition of machine intelligence. *Minds Mach* 17: 391–444, 2007.
39. Feres M, Louzoun Y, Haber S, Faveri M, Figueiredo LC, Levin L. Support vector machine-based differentiation between aggressive and chronic periodontitis using microbial profiles. *Int Dent J* 68(1): 39–46, 2018.
40. Li H, Zhou J, Zhou Y, Chen Q, She Y, Gao F, Xu Y, Chen J, Gao X. An interpretable computer-aided diagnosis method for periodontitis from panoramic radiographs. *Front Physiol* 12: 655556, 2021.
41. Chang HJ, Lee SJ, Yong TH, Shin NY, Jang BG, Kim JE, Huh KH, Lee SS, Heo MS, Choi SC, Kim TI, Yi WJ. Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. *Sci Rep* 10(1): 1–8, 2020.
42. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: a systematic review. *Dentomaxillofac Radiol* 49(1): 1–22, 2020.
43. Devito KL, de Souza Barbosa F, Filho WNF. An artificial multilayer perceptron neural network for diagnosis of proximal dental caries. *Oral Surg Oral Med Oral Pathol Oral Radiol* 106(6): 879–84, 2008.
44. Wenzel A, Hintze H, Kold LM, Kold S. Accuracy of computer-automated caries detection in digital radiographs compared with human observers. *Eur J Oral Sci* 110(3): 199–203, 2002.
45. Firestone AR, Sema D, Heaven TJ, Weems RA. The effect of a knowledge-based, image analysis and clinical decision support system on observer performance in the diagnosis of approximal caries from radiographic images. *Caries Res* 32(2): 127–34, 1998.

46. Ducan RC, Heaven T, Weems RA, Firestone AR, Greer DF, Patel JR. Using computers to diagnose and plan treatment of approval caries detected in radiographs. *J Am Dent Assoc* 126(7): 873–82, 1995.
47. Heaven TJ, Firestone AR, Feagin FF. Computer-based image analysis of natural approximal caries on radiographic films. *J Dent Res* 71(3): 846–9, 1992.
48. Pitts NB. Detection and measurement of approximal radiolucencies by computer-aided image analysis. *Oral Surg Oral Med Oral Pathol Oral Radiol* 58(3): 358–66, 1984.
49. Takahashi T, Nozaki K, Gonda T, Mameno T, Wada M, Ikebe K. Identification of dental implants using deep learning-pilot study. *Int J Implant Dent* 6(1): 53–9, 2020.
50. Lee JH, Kim YT, Lee JB, Jeong SN. A performance comparison between automated deep learning and dental professionals in classification of dental implant systems from dental imaging: a multi-center study. *Diagnostics* 10(11): 9–18, 2020.
51. Kim JE, Nam NE, Shim JS, Jung YH, Cho BH, Hwang JJ. Transfer learning via deep neural networks for implant fixture system classification using periapical radiographs. *J Clin Med* 9(4): 11–7, 2020.
52. Sukegawa S, Yoshii K, Hara T, Yamashita K, Nakano K, Yamamoto N, Nagatsuka H, Furuki Y. Deep neural networks for dental implant system classification. *Biomolecules* 10(7): 1–13, 2020.
53. Lee JH, Jeong SN. Efficacy of deep convolutional neural network algorithm for the identification and classification of dental implant systems, using panoramic and periapical radiographs. *Medicine* 99(26): 1–7, 2020.
54. Hadj Saïd M, le Roux MK, Catherine JH, Lan R. Development of an artificial intelligence model to identify a dental implant from a radiograph. *Int J Oral Maxillofac Implants* 35(6): 1077–82, 2020.
55. Zhang H, Shan J, Zhang P, Chen X, Jiang H. Trabeculae microstructure parameters serve as effective predictors for marginal bone loss of dental implant in the mandible. *Sci Rep* 10(1): 1–9, 2020.
56. Papantonopoulos G, Gogos C, Housos E, Bountis T, Loos BG. Prediction of individual implant bone levels and the existence of implant “phenotypes.” *Clin Oral Implants Res* 28(7): 823–32, 2017.
57. Liu CM, Diao ZH, Huo WY, Kong LJ, Du JJ. Simultaneous removal of Cu²⁺ and bisphenol A by a novel biochar-supported zero valent iron from aqueous solution: synthesis, reactivity and mechanism. *Environ Pollut* 239: 698–705, 2018.
58. Ha SR, Park HS, Kim EH, Kim HK, Yang JY, Heo J, Yeo ISL. A pilot study using machine learning methods about factors influencing prognosis of dental implants. *J Adv Prosthodont* 10(6): 1–6, 2018.

59. Zaw K, Liu GR, Deng B, Tan KBC. Rapid identification of elastic modulus of the interface tissue on dental implants surfaces using reduced-basis method and a neural network. *J Biomech* 42(5): 634–41, 2009.
60. Roy S, Dey S, Khutia N, Chowdhury AR, Datta S. Design of patient specific dental implant using FE analysis and computational intelligence techniques. *Appl Soft Comput* 65: 272–9, 2018.
61. Li H, Shi M, Liu X, Shi Y. Uncertainty optimization of dental implant based on finite element method, global sensitivity analysis and support vector regression. *Proc Inst Mech Eng H* 233(2): 232–43, 2019.
62. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Comput* 18(7): 1527–54, 2006.
63. Loddo A, Buttau S, di Ruberto C. Deep learning based pipelines for Alzheimer’s disease diagnosis: A comparative study and a novel deep-ensemble method. *Comput Biol Med* 141: 816–20, 2022.
64. Cao J, Song C, Peng S, Xiao F, Song S. Improved traffic sign detection and recognition algorithm for intelligent vehicles. *Sensors* 19(18): 1–21, 2019.
65. Wang G, Gong J. Facial expression recognition based on improved LeNet-5 CNN. *Proc 2019 CCDC*: 5655–60, 2019.
66. Zhang CW, Yang MY, Zeng HJ, Wen JP. Pedestrian detection based on improved LeNet-5 convolutional neural network. *J Algorithm Comput Technol* 13: 174–83, 2019.
67. Ruder S. An overview of gradient descent optimization algorithms. *ArXiv*: 1609.04747, 2016.
68. Prechelt L. Automatic early stopping using cross validation: quantifying the criteria. *Neural Net* 11(4): 761–7, 1998.
69. Bengio Y, Louradour J, Collobert R, Weston J. Curriculum learning. *Proc the 26th Ann Int Conf Mach Learn*: 1–8, 2009.
70. Prechelt L. Early Stopping — But when? *Neural Net Trick Trade* 1524: 55-69, 1996.
71. Langley P. Selection of relevant features in machine learning. *Proc the AAAI Fall symposium on relevance* 184: 245–71, 1994.
72. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: a new perspective. *Neurocomputing* 300: 70–9, 2018.
73. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res*: 1157–82, 2003.
74. Langley P. Selection of relevant features in machine learning. *Proc the AAAI Fall symposium on relevance* 184: 245–71, 1994.
75. Evgeny A, Smirnov DM, Timoshenko SN, Andrianov SN. Comparison of regularization methods for imagenet classification with deep convolutional neural networks. *Aasri Procedia*

- 6: 89–94, 2014.
76. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proc 2016 IEEE Conf Comp Vis Pattern Recognit*: 770-8, 2016.
 77. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc 2014 IEEE Conf Comp Vis Pattern Recognit*: 580–7, 2014.
 78. Long JS, Evan S, Darrell T. Fully convolutional networks for semantic segmentation. *Proc 2015 IEEE Conf Comp Vis Pattern Recognit*: 3431-40, 2015.
 79. Lipton Z. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. *Queue* 16: 31-57, 2016.
 80. Almutairi AS, Walid MA, Alkhodary MA. The effect of osseodensification and different thread designs on the dental implant primary stability. *F1000Res*: 17292.1, 2018.
 81. Inchingolo AD, Inchingolo AM, Bordea IR, Xhajanka E, Romeo DM, Romeo M, Zappone CMF, Malcangi G, Scarano A, Lorusso F, Isacco CG, Marinelli G, Contaldo M, Ballini A, Inchingolo F, Dipalma G. The effectiveness of osseodensification drilling protocol for implant site osteotomy: a systematic review of the literature and meta-analysis. *Materials* 14(5): 306–12, 2021.
 82. Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell* 38(2): 295-307, 2016.

IX. 図 表

表 1. 症例情報 (27 症例)

症 例 番 号	性 別	年 齢	埋入 位置	インプラント体 の製造社	インプラ ント体の 形状	インプラ ント体の 直径(mm)	インプラ ント体の 長さ(mm)	埋入 窩形 成術 式
1	M	63	36	Nobel Biocare	Straight	4.3	8.5	A
2	F	41	12	Straumann	Tapered	4.1	12.0	A
3	M	73	46	Nobel Biocare	Straight	4.3	8.5	A
4	F	80	36	Nobel Biocare	Tapered	5.0	8.0	A
5	M	59	36	Straumann	Tapered	4.8	10.0	A
6	F	79	15	Nobel Biocare	Straight	4.3	11.5	A
7	F	40	37	Straumann	Tapered	4.1	8.0	A
8	F	72	26	Nobel Biocare	Straight	5.0	10.0	A
9	F	69	35	Straumann	Tapered	4.1	8.0	A
10	F	60	15	Straumann	Straight	3.3	8.0	B
11	F	54	35	Nobel Biocare	Straight	4.3	8.5	B
12	F	70	17	Straumann	Tapered	4.8	12.0	B
13	M	67	46	Nobel Biocare	Straight	4.3	8.5	B
14	F	38	45	Nobel Biocare	Straight	4.3	10.0	B
15	M	63	23	Nobel Biocare	Straight	4.3	10.0	B
16	F	36	46	Nobel Biocare	Straight	4.3	10.0	B
17	M	48	16	Straumann	Tapered	4.8	8.0	B
18	F	56	25	Nobel Biocare	Straight	4.3	8.5	B
19	F	80	24	Straumann	Tapered	3.3	10.0	C
20	F	21	15	Nobel Biocare	Straight	3.5	10.0	C
21	F	62	17	Nobel Biocare	Straight	4.3	10.0	C
22	M	63	26	Nobel Biocare	Straight	4.3	10.0	C
23	F	65	46	Nobel Biocare	Straight	4.3	10.0	C
24	F	71	11	Nobel Biocare	Tapered	3.5	11.5	C
25	F	70	35	Nobel Biocare	Straight	3.5	8.0	C
26	F	67	16	Nobel Biocare	Tapered	4.3	10.0	C
27	F	71	26	Nobel Biocare	Straight	4.3	8.0	C

表 2. 予測された術式と実際の術式の関係

実際の術式	予測された術式			
		A	B	C
	A	正解 (TA)	不正解 (FB)	不正解 (FC)
	B	不正解 (FA)	正解 (TB)	不正解 (FC)
	C	不正解 (FA)	不正解 (FB)	正解 (TC)

表 3. LeNet をベースとした AI モデルのテスト用画像に基づく学習結果

真のラベル	推定ラベル			適合率(%)
	A	B	C	
A	3	0	22	12.0
B	4	7	14	28.0
C	1	7	17	68.0
再現率(%)	37.5	50.0	32.1	正確率=36.0%

表 4. GoogleNet をベースとした AI モデルのテスト用画像に基づく学習結果

真のラベル	推定ラベル			適合率(%)
	A	B	C	
A	1	0	24	4.0
B	5	3	17	12.0
C	11	0	14	56.0
再現率(%)	5.9	100.0	25.5	正確率=24.0%

表 5. AlexNet をベースとした AI モデルのテスト用画像に基づく学習結果

真のラベル	推定ラベル			適合率(%)
	A	B	C	
A	7	4	14	28.0
B	1	5	19	20.0
C	7	3	15	60.0
再現率(%)	46.7	41.7	31.3	正確率=36.0%

表 6. Adagrad を用いた AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	3	0	22	12.0
B	4	7	14	28.0
C	1	7	17	68.0
再現率(%)	37.5	50.0	32.1	正確率=36.0%

表 7. Adam を用いた AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	2	0	23	8.0
B	6	6	13	24.0
C	2	11	12	48.0
再現率(%)	20.0	35.3	25.0	正確率=26.7%

表 8. RMSprop を用いた AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	2	0	23	8.0
B	7	4	14	16.0
C	2	3	20	80.0
再現率(%)	18.1	57.1	36.8	正確率=34.7%

表 9. Adadelat を用いた AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	0	0	25	0.0
B	0	6	19	24.0
C	0	1	24	96.0
再現率(%)	0.0	85.7	33.3	正確率=40.0%

表 10. SGD を用いた AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	1	0	24	4.0
B	0	1	24	4.0
C	0	1	24	96.0
再現率(%)	100.0	50.0	33.3	正確率=34.7%

表 11. NAG を用いた AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	2	0	23	8.0
B	1	4	20	16.0
C	0	0	25	100.0
再現率(%)	66.7	100.0	36.8	正確率=41.3%

表 12. 学習回数を 30 回とした AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	2	0	23	8.0
B	6	5	14	20.0
C	4	6	15	60.0
再現率(%)	16.7	45.5	28.8	正確率=29.3%

表 13. 学習回数を 50 回とした AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	3	0	22	12.0
B	5	7	13	28.0
C	1	6	16	69.6
再現率(%)	33.3	45.5	31.3	正確率=34.7%

表 14. 学習回数を 100 回とした AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	3	0	22	12.0
B	4	7	14	28.0
C	1	7	17	68.0
再現率(%)	37.5	50.0	32.1	正確率=36.0%

表 15. 訓練用画像とは異なる症例からテスト用画像を選択する方法（T1）で作成した
AIモデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	3	0	22	12.0
B	5	7	13	28.0
C	1	6	16	69.6
再現率(%)	33.3	45.5	31.3	正確率=34.7%

表 16. 各症例から均等配分でテスト用画像を選択する方法（T2）で作成した
AIモデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	23	0	2	92.0
B	0	24	1	96.0
C	0	1	24	96.0
再現率(%)	100.0	96.0	88.9	正確率=94.7%

表 17. 症例情報 (60 症例)

症 例 番 号	性 別	年 齢	埋 入 位 置	インプラント体 の製造社	インプラ ント体の 形状	インプラ ント体の 直径(mm)	インプラ ント体の 長さ(mm)	埋入 窩形 成術 式
1	M	63	36	Nobel Biocare	Straight	4.3	8.5	A
2	F	41	12	Straumann	Tapered	4.1	12.0	A
3	M	73	46	Nobel Biocare	Straight	4.3	8.5	A
4	F	80	36	Nobel Biocare	Tapered	5.0	8.0	A
5	M	59	36	Straumann	Tapered	4.8	10.0	A
6	F	79	15	Nobel Biocare	Straight	4.3	11.5	A
7	F	40	37	Straumann	Tapered	4.1	8.0	A
8	F	72	26	Nobel Biocare	Straight	5.0	10.0	A
9	F	69	35	Straumann	Tapered	4.1	8.0	A
10	F	60	15	Straumann	Straight	3.3	8.0	B
11	F	54	35	Nobel Biocare	Straight	4.3	8.5	B
12	F	70	17	Straumann	Tapered	4.8	12.0	B
13	M	67	46	Nobel Biocare	Straight	4.3	8.5	B
14	F	38	45	Nobel Biocare	Straight	4.3	10.0	B
15	M	63	23	Nobel Biocare	Straight	4.3	10.0	B
16	F	36	46	Nobel Biocare	Straight	4.3	10.0	B
17	M	48	16	Straumann	Tapered	4.8	8.0	B
18	F	56	25	Nobel Biocare	Straight	4.3	8.5	B
19	F	80	24	Straumann	Tapered	3.3	10.0	C
20	F	21	15	Nobel Biocare	Straight	3.5	10.0	C
21	F	62	17	Nobel Biocare	Straight	4.3	10.0	C
22	M	63	26	Nobel Biocare	Straight	4.3	10.0	C
23	F	65	46	Nobel Biocare	Straight	4.3	10.0	C
24	F	71	11	Nobel Biocare	Tapered	3.5	11.5	C
25	F	70	35	Nobel Biocare	Straight	3.5	8.0	C
26	F	67	16	Nobel Biocare	Tapered	4.3	10.0	C
27	F	71	26	Nobel Biocare	Straight	4.3	8.0	C
28	F	54	36	Nobel Biocare	Straight	4.3	10.0	A
29	M	72	36	Straumann	Tapered	4.8	8.0	A
30	M	72	46	Straumann	Tapered	4.8	8.0	A

31	F	48	36	Nobel Biocare	Straight	4.3	8.5	A
32	M	70	26	Nobel Biocare	Tapered	5.0	10.0	A
33	M	70	46	Nobel Biocare	Straight	4.3	10.0	A
34	F	69	46	Straumann	Tapered	4.1	10.0	A
35	F	65	36	Nobel Biocare	Tapered	5.0	11.5	A
36	F	58	35	Straumann	Straight	3.3	8.0	A
37	F	60	37	Nobel Biocare	Straight	4.3	8.5	A
38	M	54	36	Nobel Biocare	Tapered	5.0	11.5	A
39	F	55	24	Nobel Biocare	Straight	3.5	11.5	B
40	M	63	16	Nobel Biocare	Tapered	4.3	11.5	B
41	M	63	26	Nobel Biocare	Tapered	4.3	11.5	B
42	F	66	26	Straumann	Tapered	4.1	10.0	B
43	F	53	36	Nobel Biocare	Straight	4.3	10.0	B
44	M	70	24	Straumann	Tapered	4.1	10.0	B
45	F	55	47	Straumann	Tapered	4.1	10.0	B
46	F	53	27	Nobel Biocare	Tapered	4.3	11.5	B
47	F	32	45	Straumann	Tapered	3.3	10.0	B
48	F	62	36	Nobel Biocare	Straight	5.0	8.5	B
49	F	59	46	Nobel Biocare	Tapered	4.3	10.0	B
50	M	52	16	Nobel Biocare	Straight	5.5	7.0	C
51	F	19	45	Nobel Biocare	Straight	3.5	10.0	C
52	F	50	16	Nobel Biocare	Straight	5.5	7.0	C
53	F	70	13	Nobel Biocare	Straight	4.3	11.5	C
54	F	70	21	Nobel Biocare	Straight	4.3	11.5	C
55	F	48	36	Nobel Biocare	Tapered	4.3	10.0	C
56	M	71	36	Nobel Biocare	Straight	4.3	10.0	C
57	F	57	27	Nobel Biocare	Tapered	4.3	10.0	C
58	F	62	47	Straumann	Tapered	4.8	8.0	C
59	F	46	47	Nobel Biocare	Straight	4.3	8.5	C
60	M	70	16	Straumann	Tapered	4.8	8.0	C

実験 1 で用いた 27 症例から 60 症例に増加

表 18. 関心領域として抽出した画像のサイズを 5×5 pixel とした AI モデルの
テスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	22	1	2	36.7
B	17	6	2	75.0
C	21	1	3	42.9
再現率(%)	88.0	24.0	12.0	正確率=60.0%

表 19. 関心領域として抽出した画像のサイズを 10×10 pixel とした AI モデルの
テスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	15	4	6	41.7
B	10	13	2	72.2
C	11	1	13	61.9
再現率(%)	60.0	52.0	52.0	正確率=56.0%

表 20. 関心領域として抽出した画像のサイズを 20×20 pixel とした AI モデルの
テスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	23	0	2	92.0
B	0	24	1	96.0
C	0	1	24	96.0
再現率(%)	100.0	96.0	88.9	正確率=94.7%

表 21. 画像枚数 675 枚で学習した AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	23	0	2	92.0
B	0	24	1	96.0
C	0	1	24	96.0
再現率(%)	100.0	96.0	88.9	正確率=94.7%

表 22. 画像枚数 1,200 枚で学習した AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	45	6	4	77.6
B	6	43	6	75.4
C	7	8	40	80.0
再現率(%)	81.8	78.2	76.2	正確率=78.2%

表 23. 海綿骨のみを含む領域の画像を入力データとして学習した AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	45	6	4	77.6
B	6	43	6	75.4
C	7	8	40	80.0
再現率(%)	81.8	78.2	72.7	正確率=78.2%

表 24. 皮質骨を含む領域の画像を入力データとして学習した AI モデルのテスト用画像に基づく学習結果

	推定ラベル			
真のラベル	A	B	C	適合率(%)
A	54	1	0	90.0
B	1	53	1	96.4
C	5	1	49	98.0
再現率(%)	98.2	96.3	89.1	正確率=95.2%

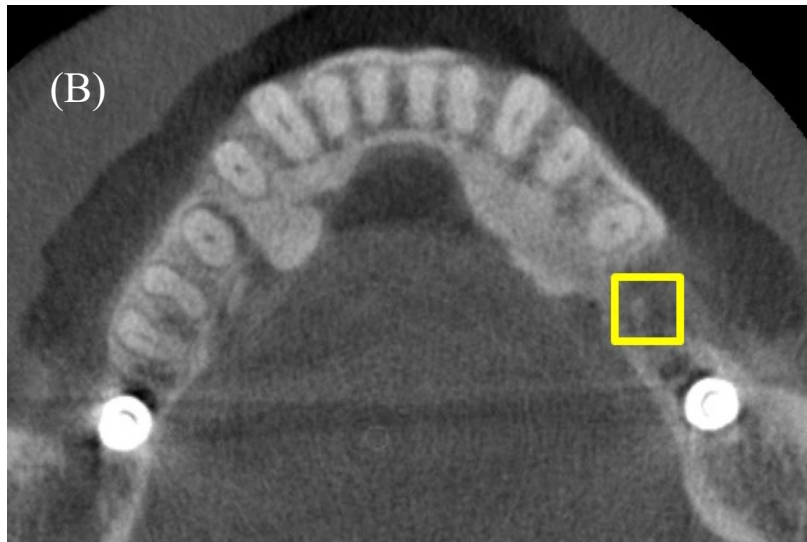
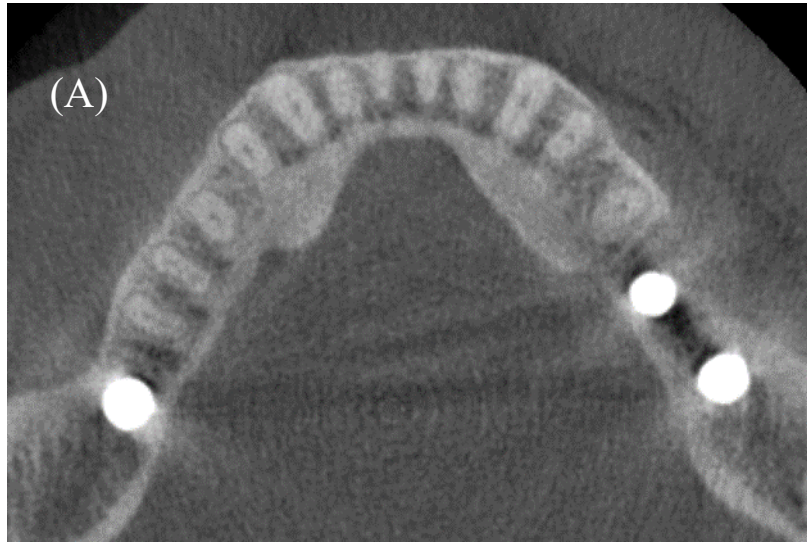


図 1. (A) 術後の CBCT 画像, (B) 術前の CBCT 画像.

黄色枠は, 抽出した関心領域 (20×20 pixel) を示す.

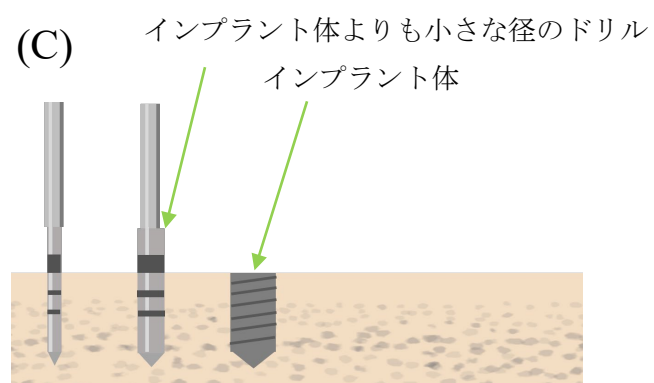
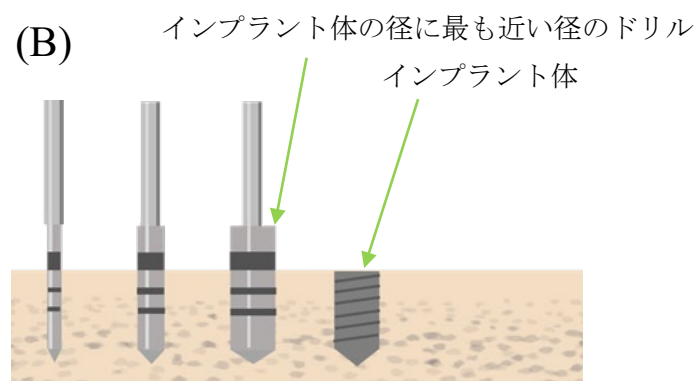
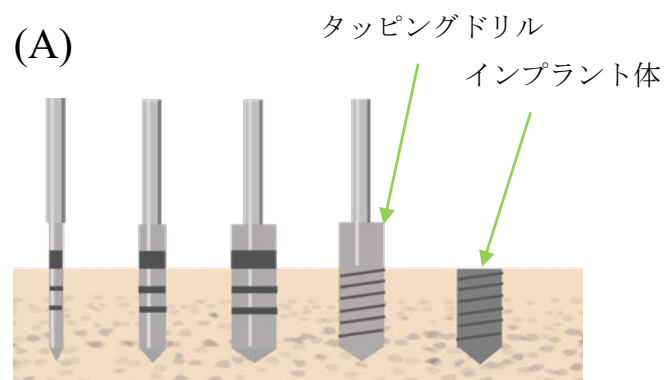
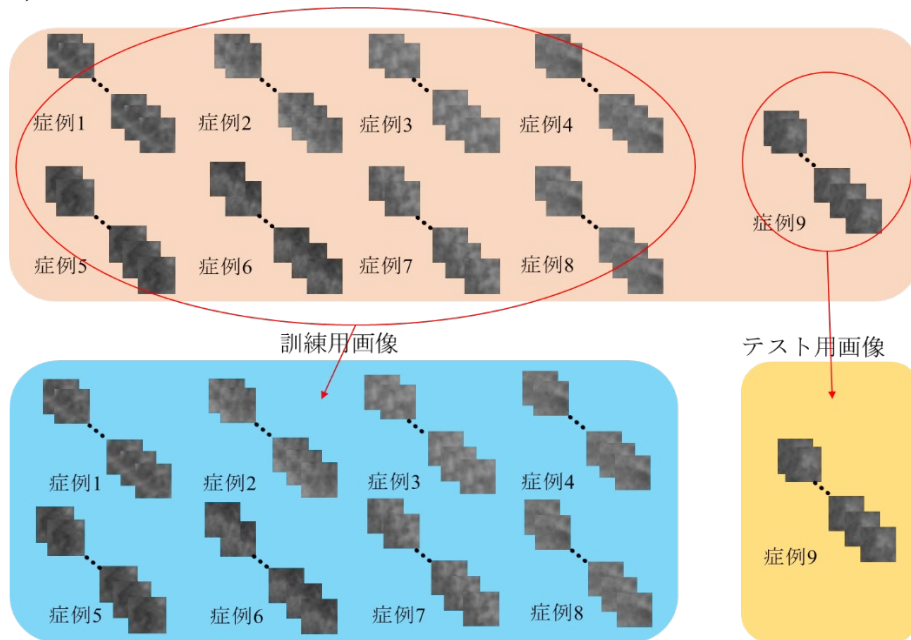


図 2. (A) タッピングドリルを使用して形成した症例, (B) インプラント体の径に最も近い径まで形成した症例, (C) インプラント体の径よりも小さい径まで形成した症例.

(A)



(B)

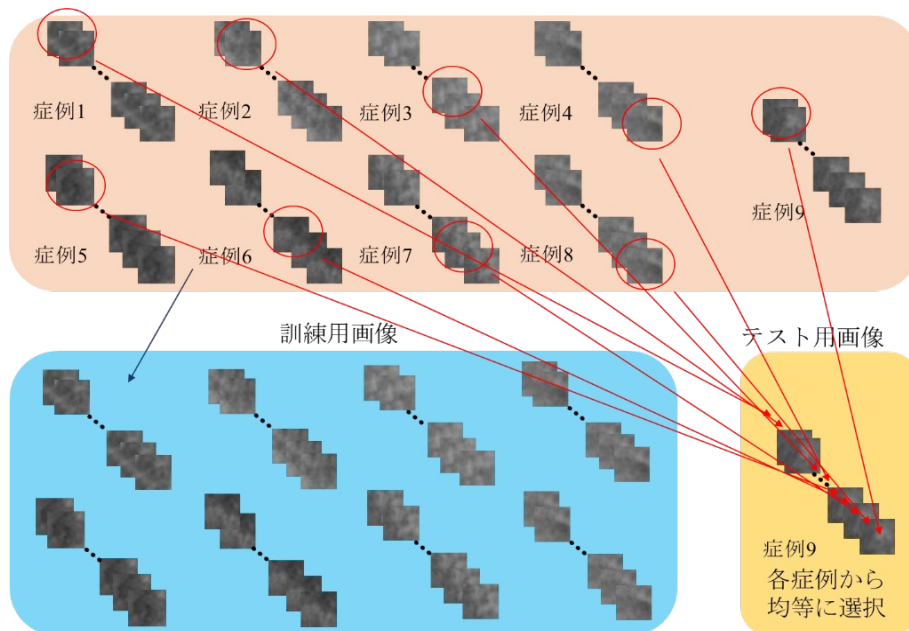


図3. (A) 訓練用画像とは異なる症例からテスト用画像を選択する方法 T1,

(B) 各症例から均等配分でテスト用画像を選択する方法 T2.

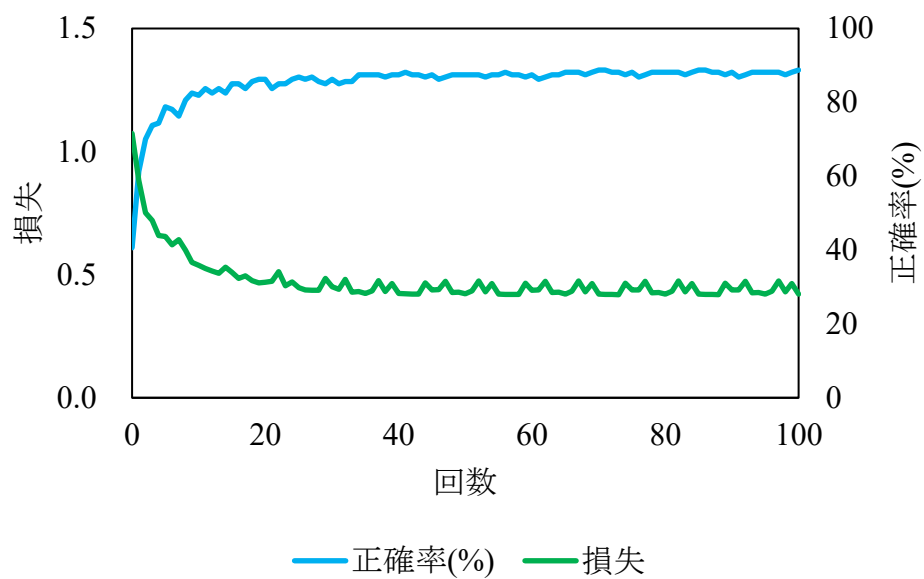


図 4. LeNet をベースとした AI モデルの訓練用画像に基づく学習曲線

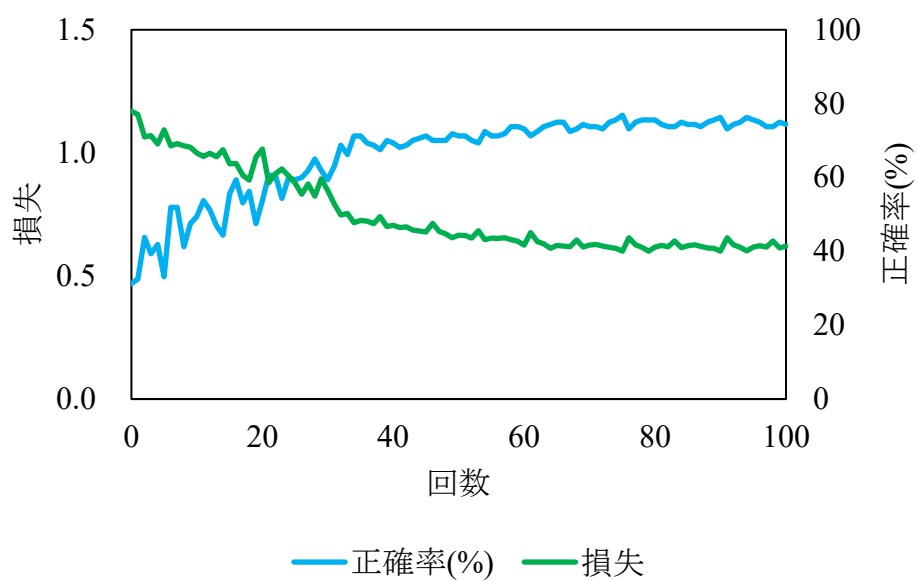


図 5. GoogleNet をベースとした AI モデルの訓練用画像に基づく学習曲線

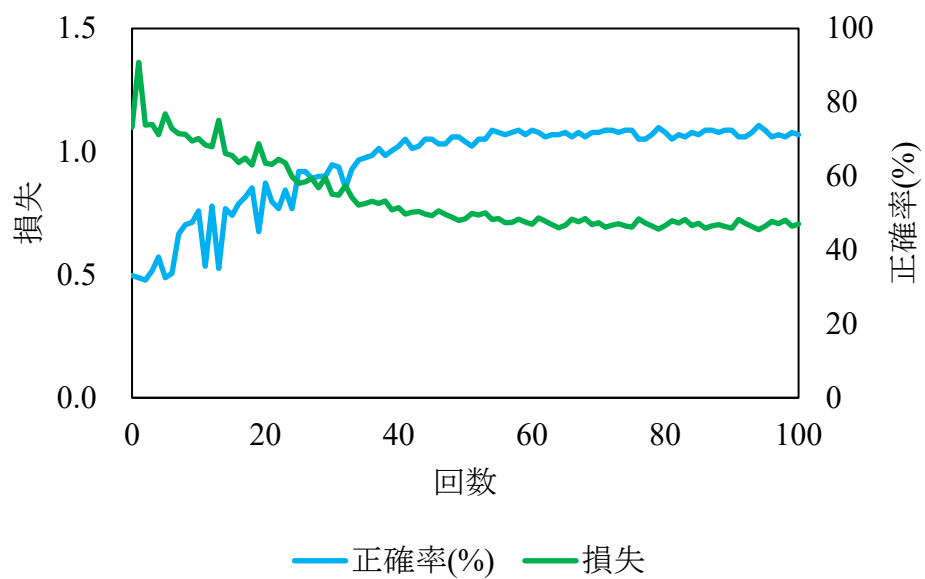


図 6. AlexNet をベースとした AI モデルの訓練用画像に基づく学習曲線

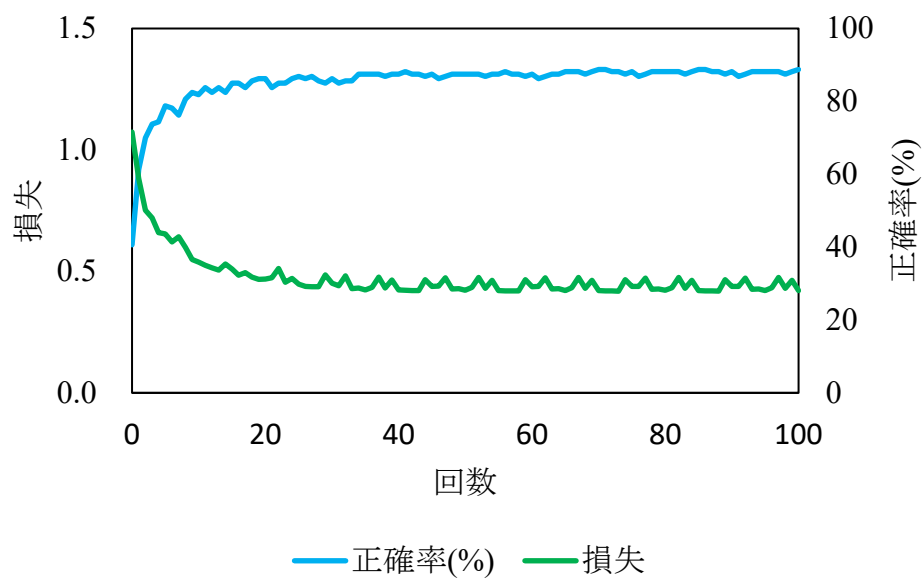


図 7. Adagrad を用いた AI モデルの訓練用画像に基づく学習曲線

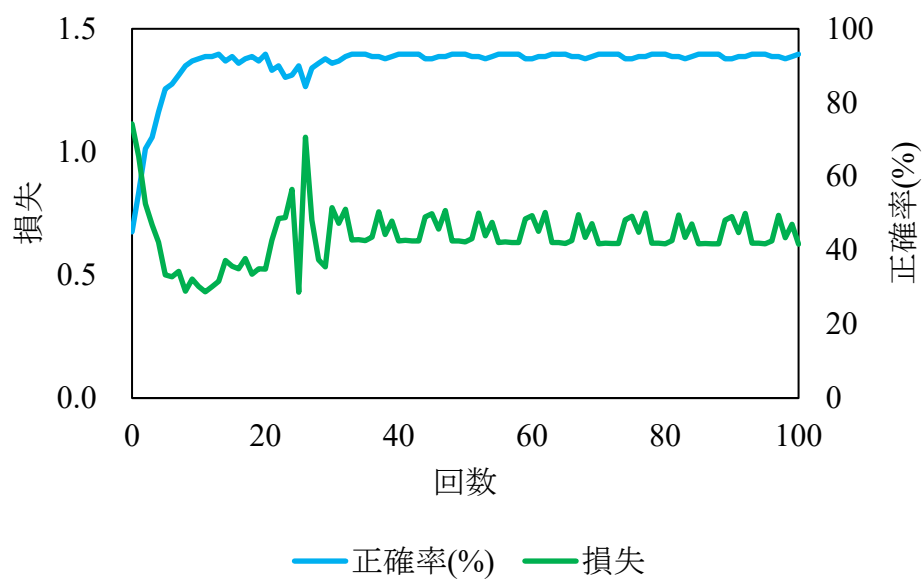


図 8. Adam を用いた AI モデルの訓練用画像に基づく学習曲線

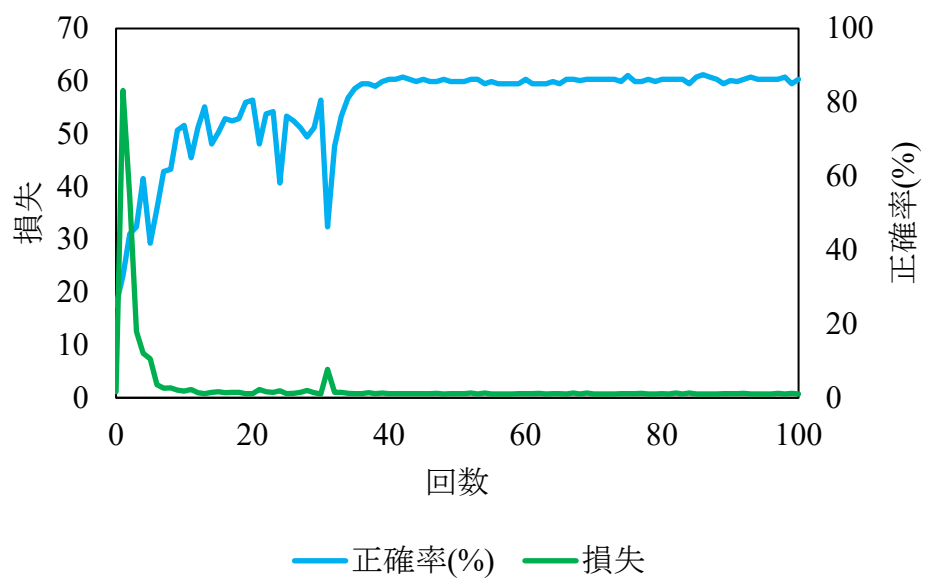


図 9. RMSprop を用いた AI モデルの訓練用画像に基づく学習曲線

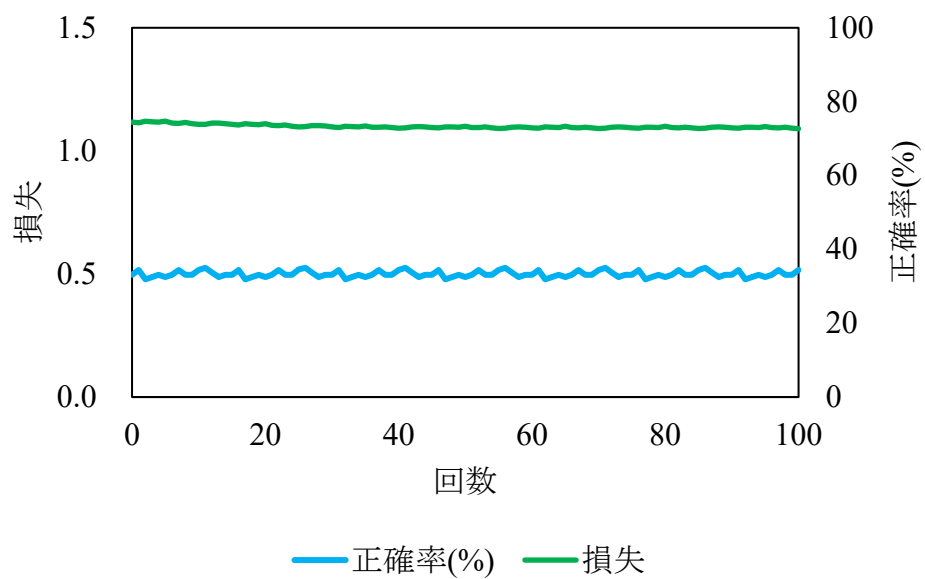


図 10. Adadelata を用いた AI モデルの訓練用画像に基づく学習曲線

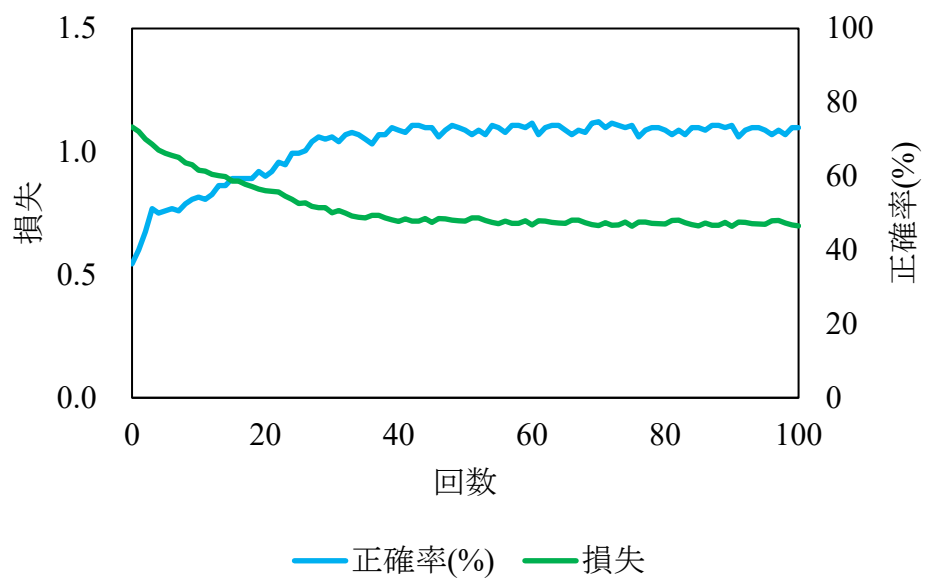


図 11. SGD を用いた AI モデルの訓練用画像に基づく学習曲線

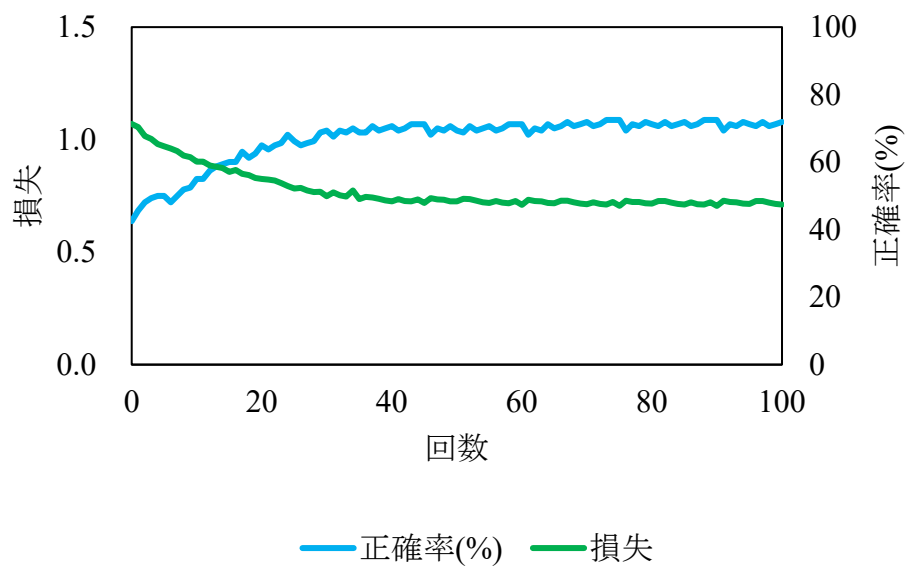


図 12. NAG を用いた AI モデルの訓練用画像に基づく学習曲線

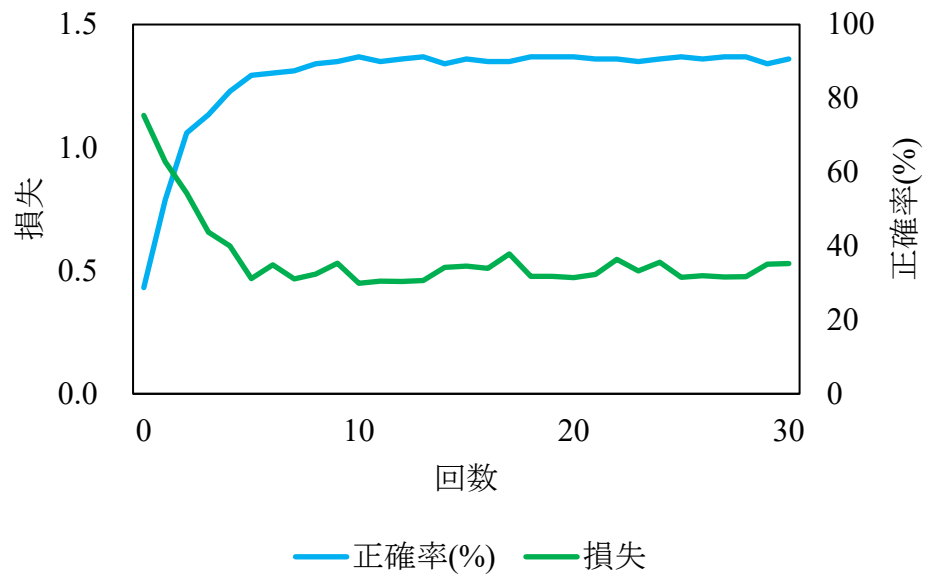


図 13. 学習回数を 30 回とした AI モデルの訓練用画像に基づく学習曲線

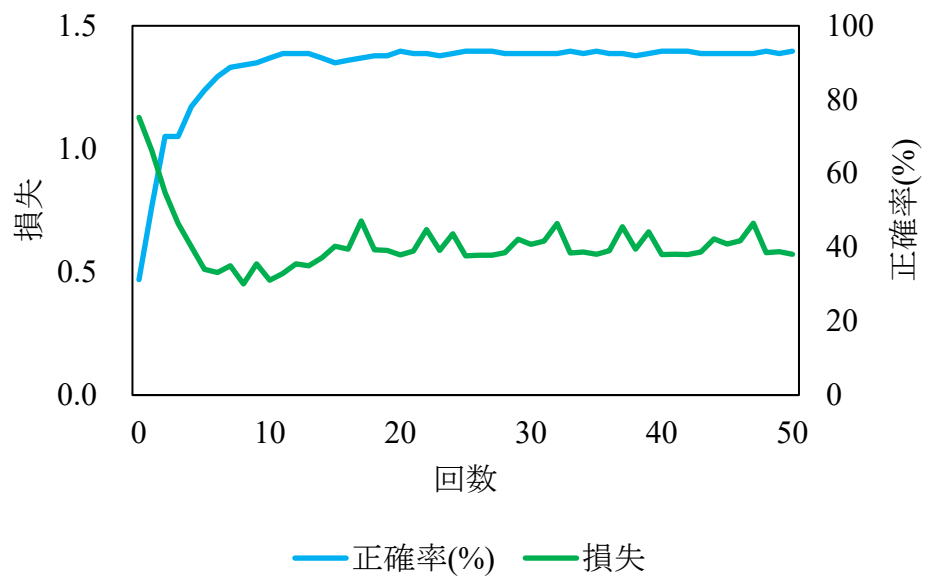


図 14. 学習回数を 50 回とした AI モデルの訓練用画像に基づく学習曲線

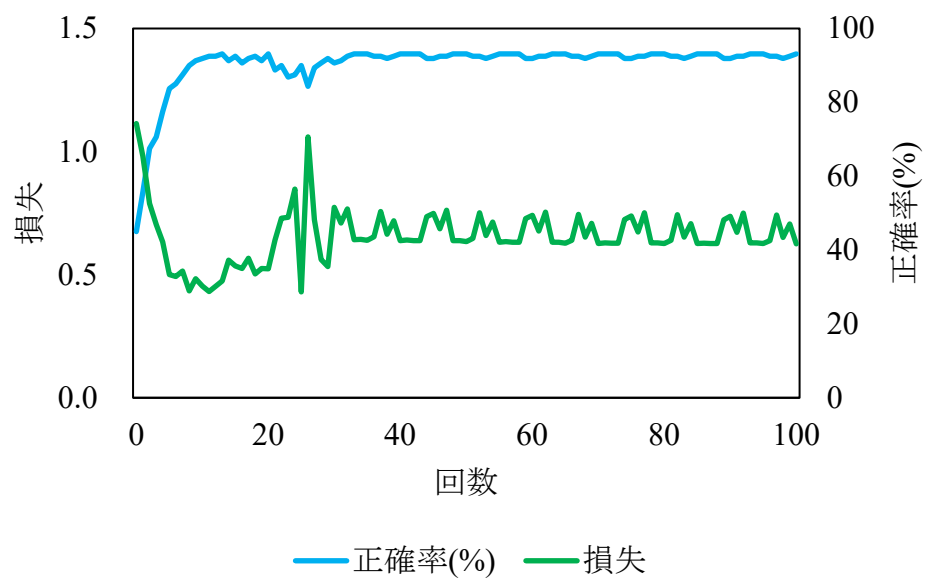


図 15. 学習回数を 100 回とした AI モデルの訓練用画像に基づく学習曲線

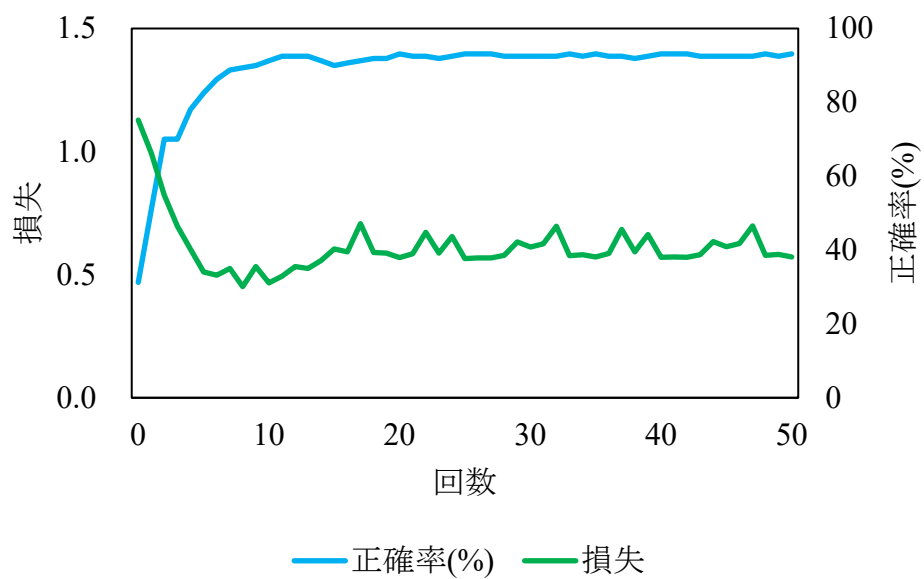


図 16. 訓練用画像とは異なる症例からテスト用画像を選択する方法で作成した AI モデルの訓練用画像に基づく学習曲線

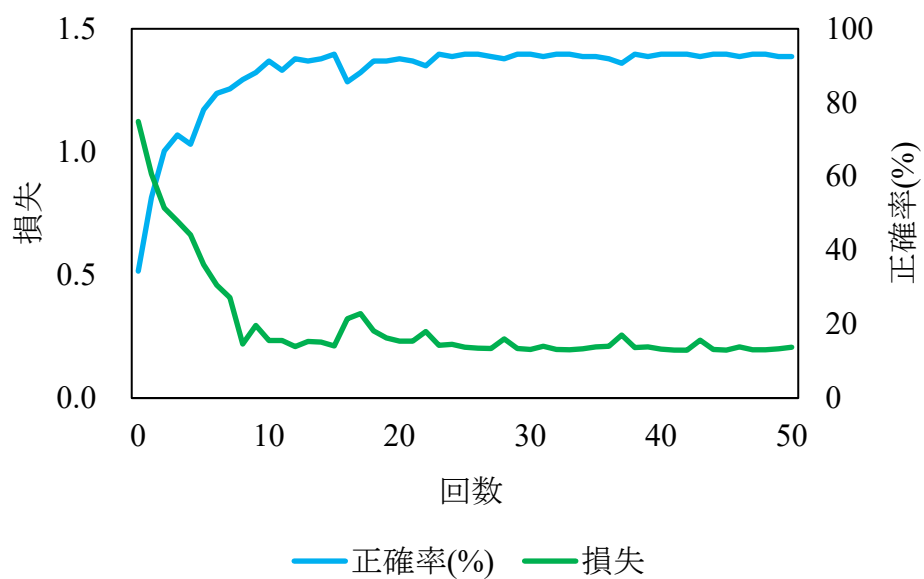
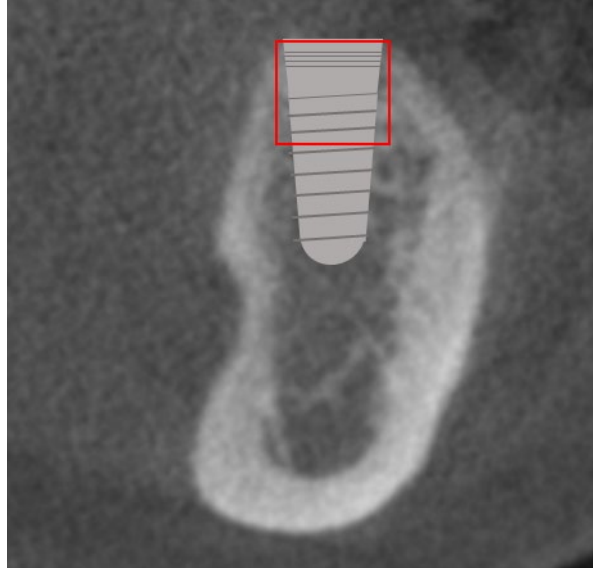


図 17. 各症例から均等配分でテスト用画像を選択する方法で作成した AI モデルの訓練用画像に基づく学習曲線

(A)



(B)

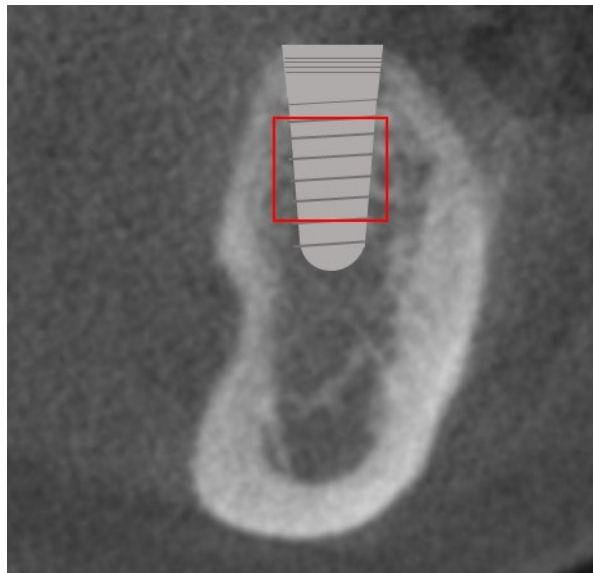


図 18. (A) 骨縁の皮質骨を含む領域, (B) 海綿骨のみを含む領域

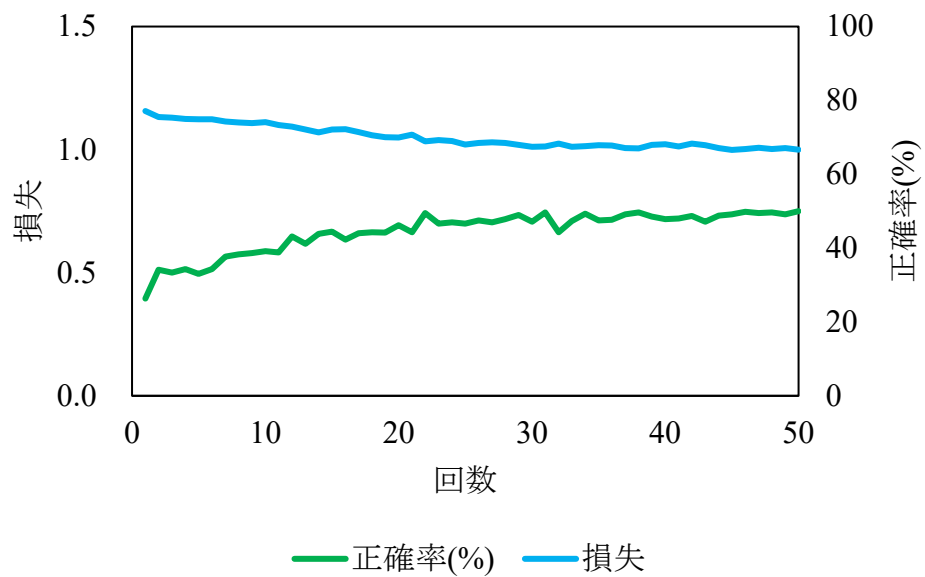


図 19. 関心領域として抽出した画像のサイズを 5×5 pixel とした AI モデルの
訓練用画像に基づく学習曲線

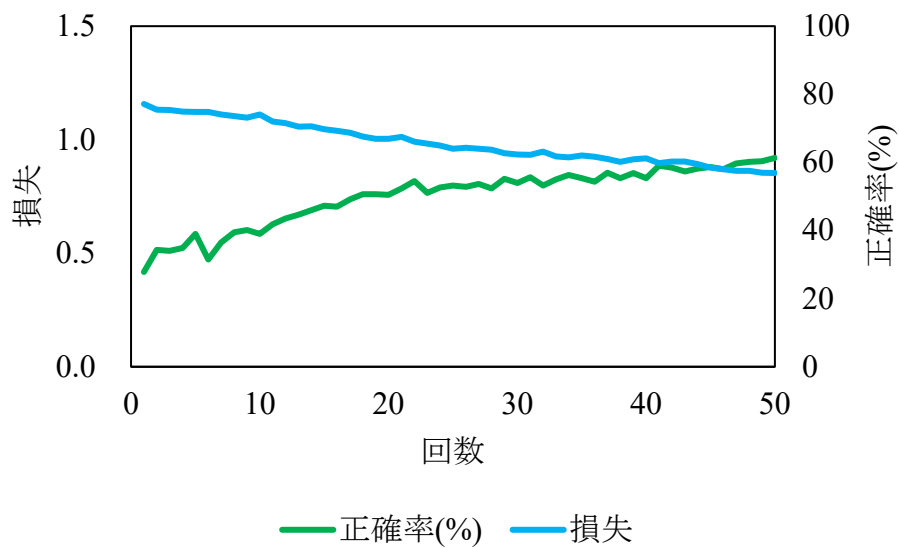


図 20. 関心領域として抽出した画像のサイズを 10×10 pixel とした AI モデルの
訓練用画像に基づく学習曲線

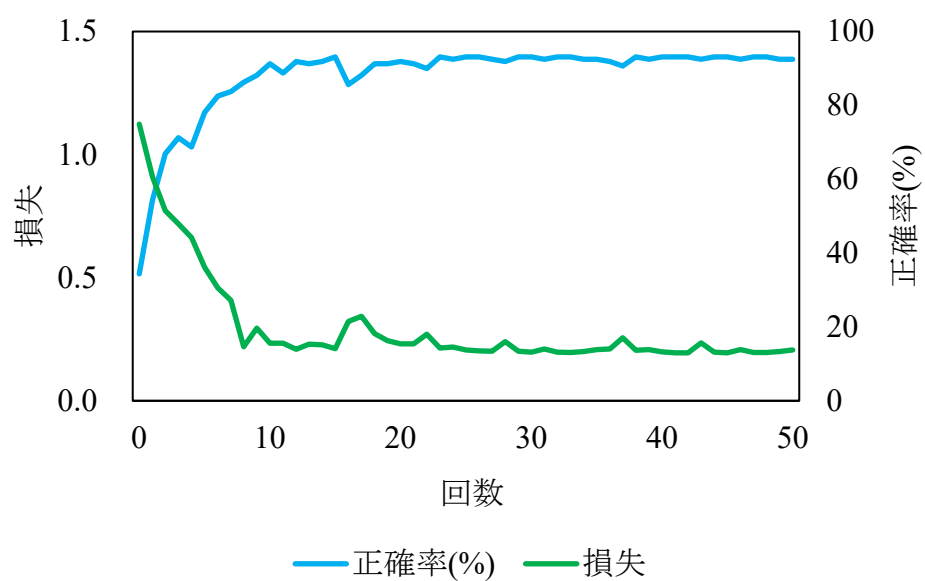


図 21. 関心領域として抽出した画像のサイズを 20×20 pixel とした AI モデルの
訓練用画像に基づく学習曲線

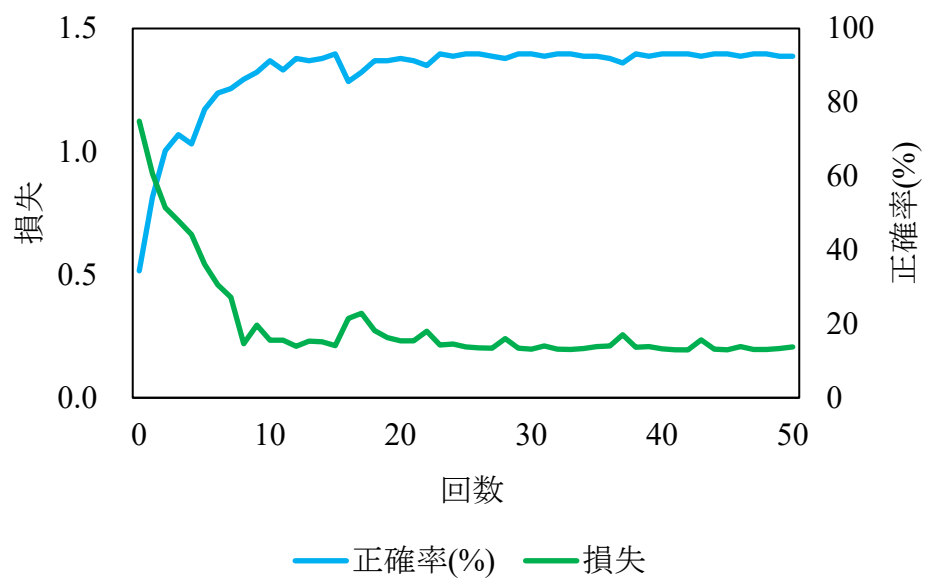


図 22. 画像枚数 675 枚で学習した AI モデルの訓練用画像に基づく学習曲線

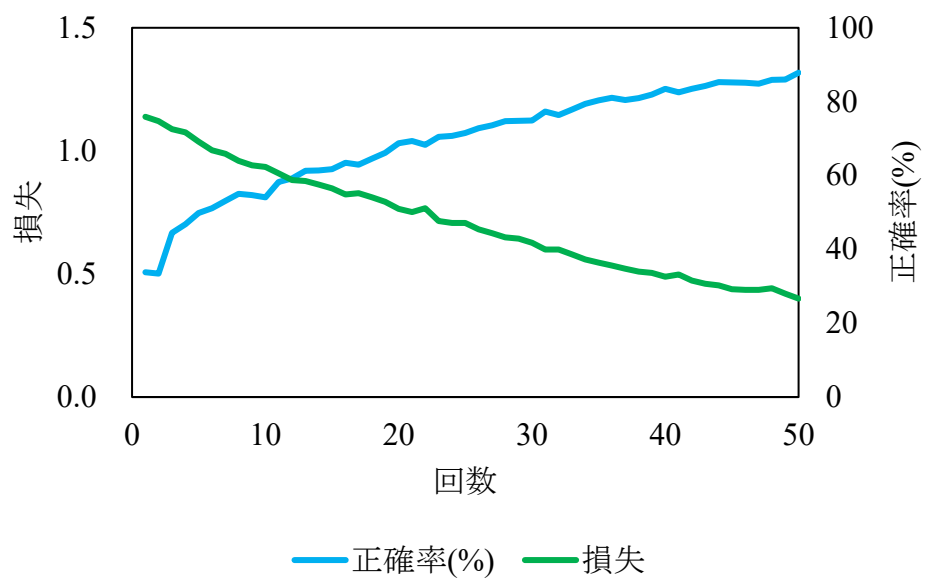


図 23. 画像枚数 1,200 枚で学習した AI モデルの訓練用画像に基づく学習曲線

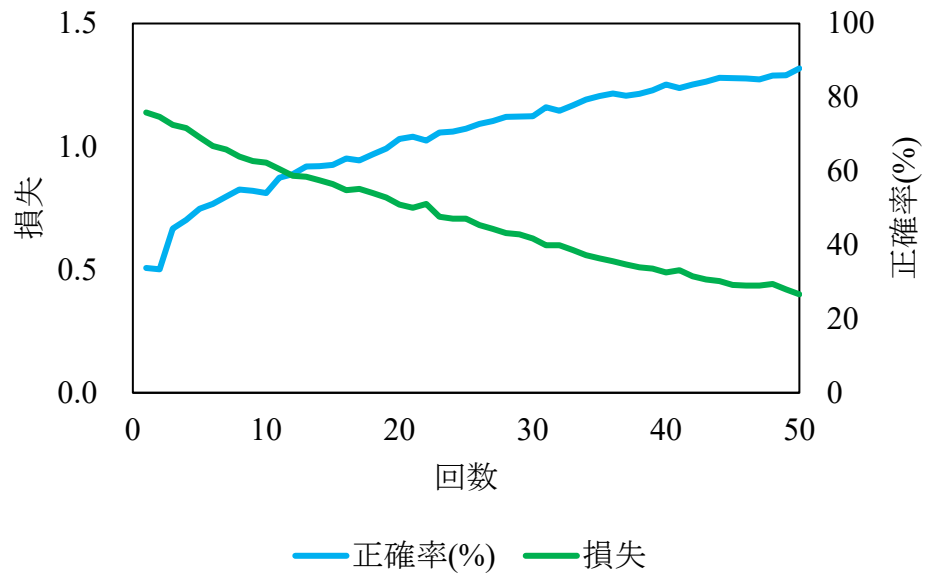


図 24. 海綿骨のみを含む領域の画像を入力データとして学習した AI モデルの訓練用画像に基づく学習曲線

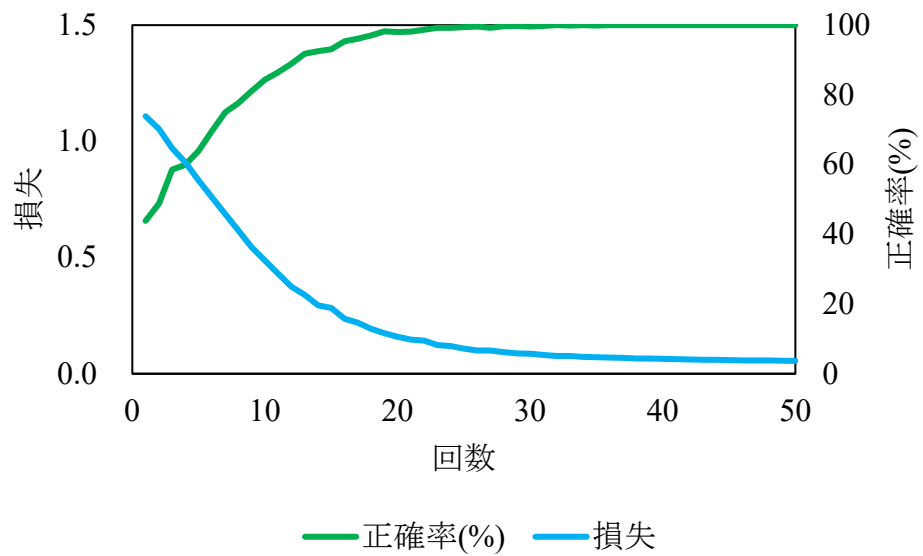


図 25. 皮質骨を含む画像を入力データとして学習した AI モデルの訓練用画像に基づく学習曲線

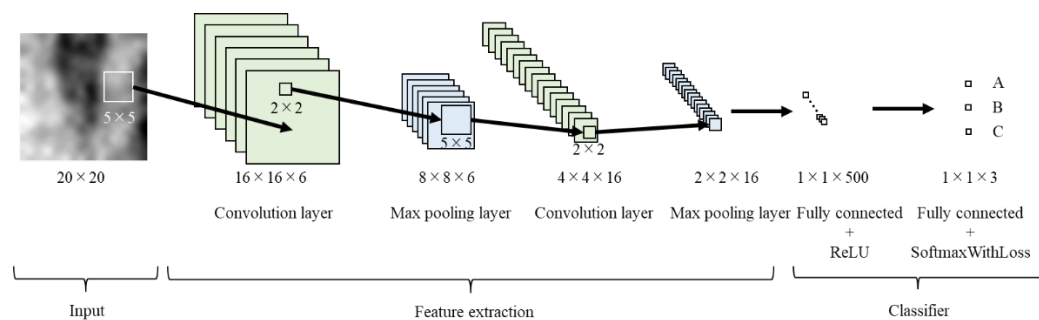


図 26. 実験 2 で作成した AI モデル

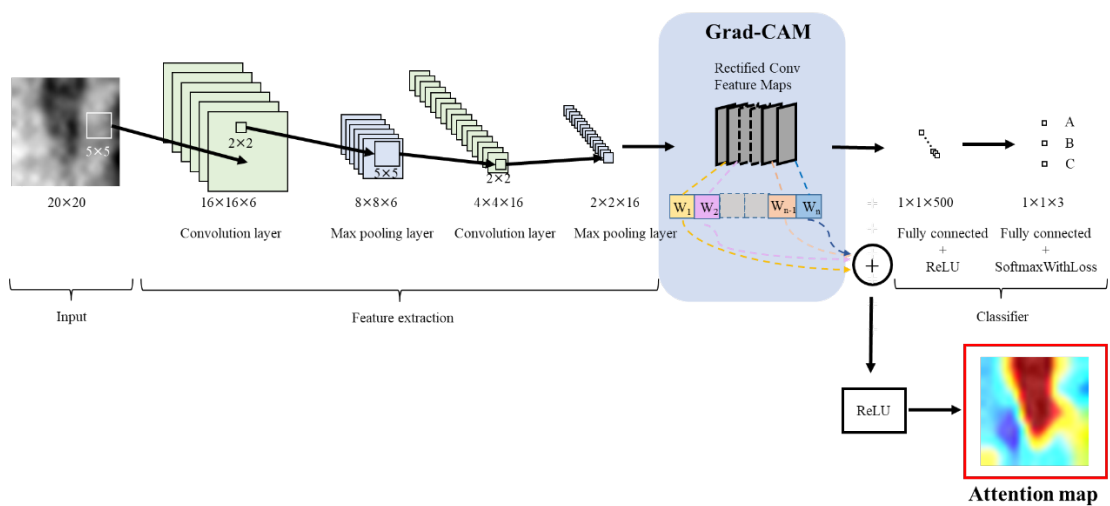


図 27. 実験 2 で作成した AI モデルに Grad-CAM を実装

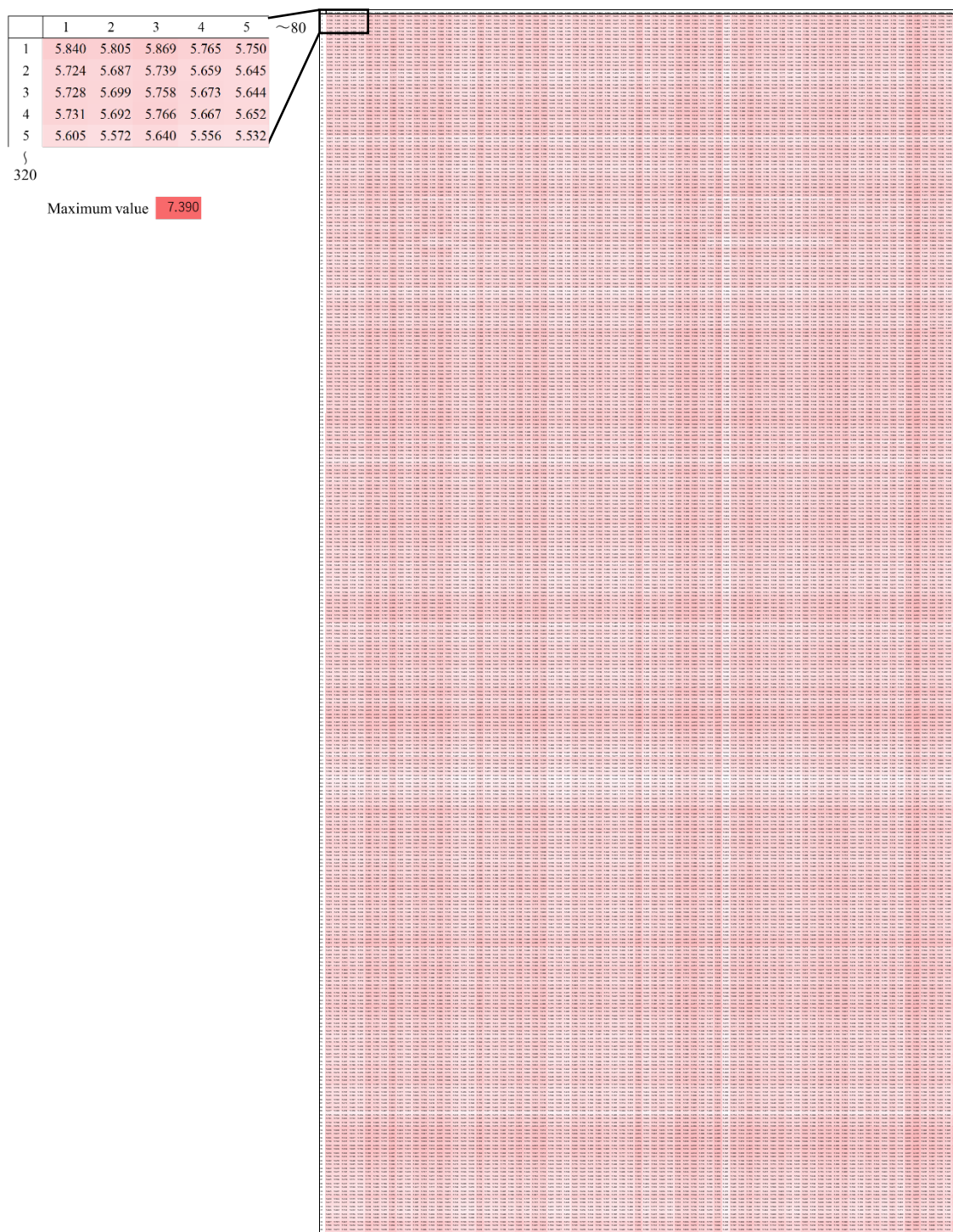


図 28. 術式 A の訓練用画像 320 枚とテスト用画像 80 枚のすべての組み合わせ（25,600 通り）の相互情報量. 術式 A における相互情報量の最大値は 7.390 であった. 各セルは相互情報量の値に応じて白（最小値）から赤（最大値）に色付けされており, 重複する画像が存在する場合に赤色となる. 左上に全セルの一部を拡大して表示する.

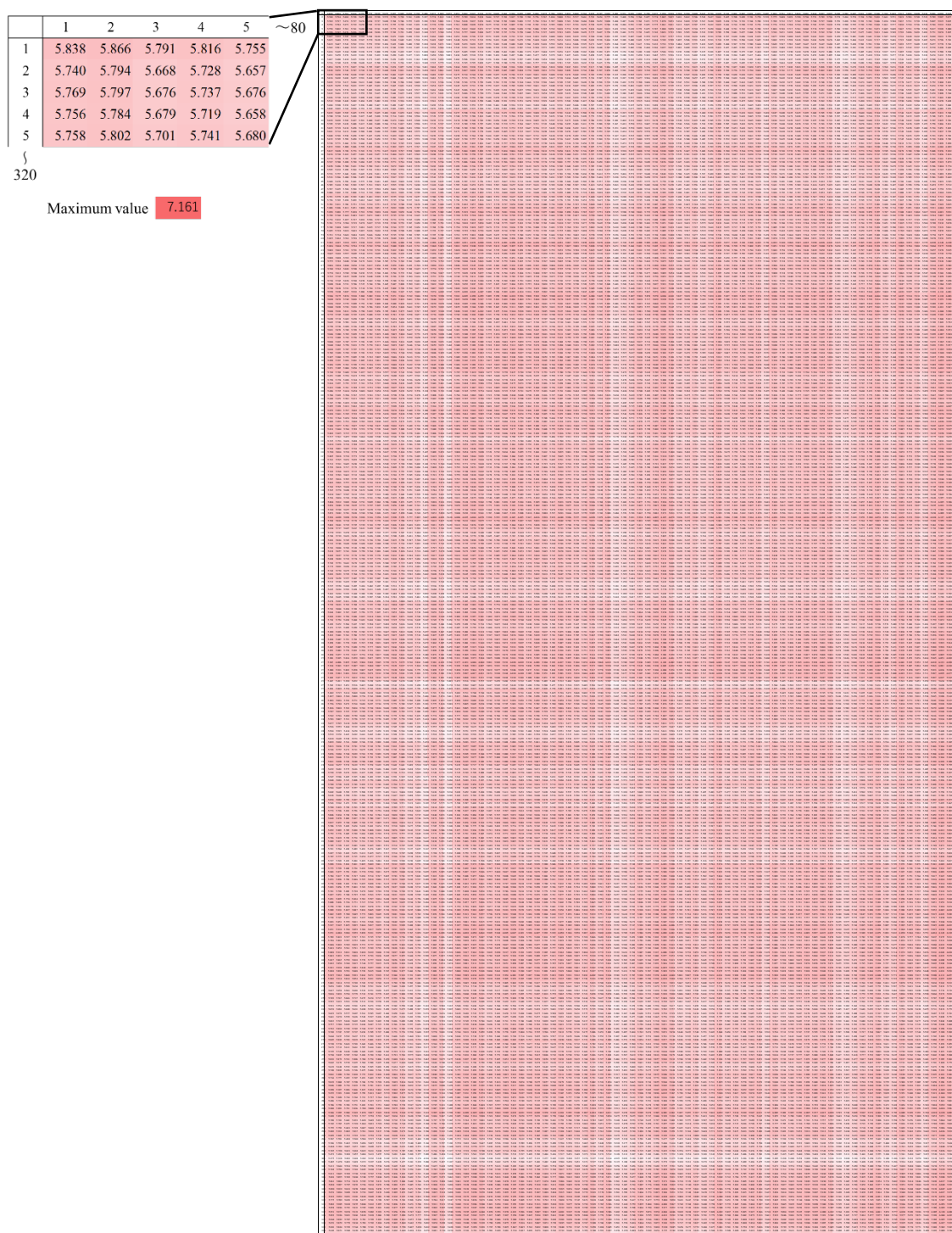


図 29. 術式 B の訓練用画像 320 枚とテスト用画像 80 枚のすべての組み合わせ（25,600 通り）の相互情報量。術式 B における相互情報量の最大値は 7.161 であった。各セルは相互情報量の値に応じて白（最小値）から赤（最大値）に色付けされており、重複する画像が存在する場合に赤色となる。左上に全セルの一部を拡大して表示する。

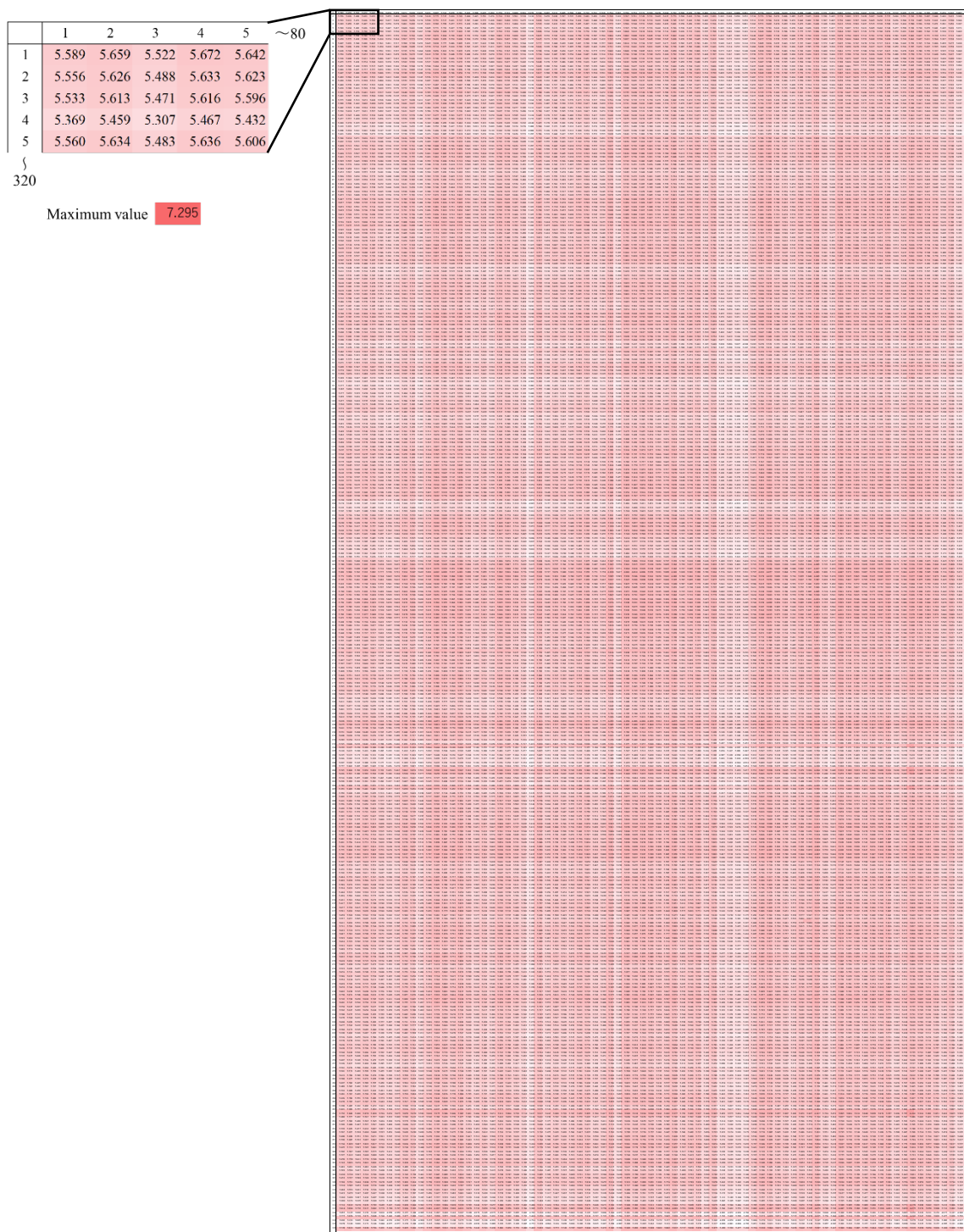
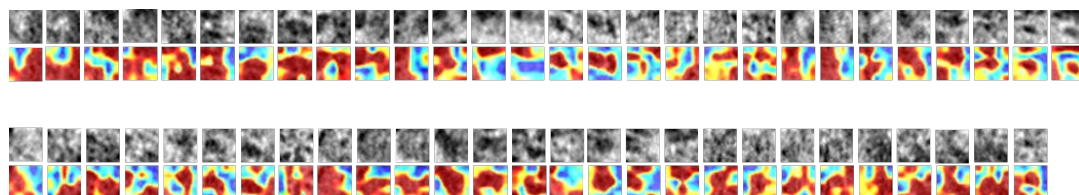
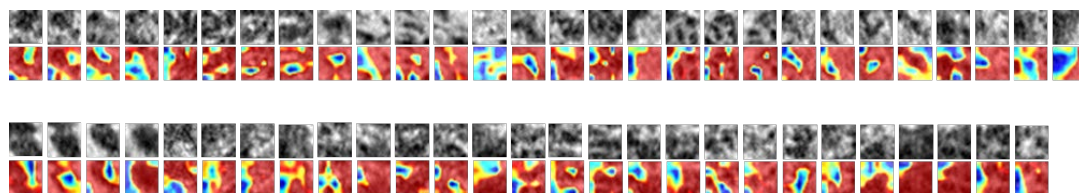


図 30. 術式 C の訓練用画像 320 枚とテスト用画像 80 枚のすべての組み合わせ（25,600 通り）の相互情報量．術式 C における相互情報量の最大値は 7.295 であった．各セルは相互情報量の値に応じて白（最小値）から赤（最大値）に色付けされており，重複する画像が存在する場合に赤色となる．左上に全セルの一部を拡大して表示する．

(A)



(B)



(C)

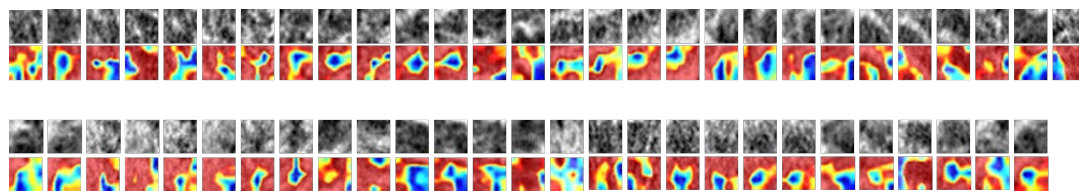
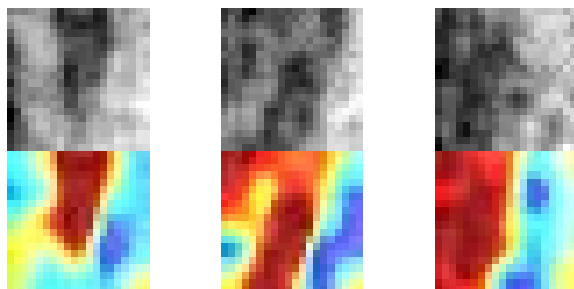
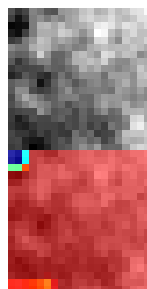


図 31. 各術式のテスト用画像に重畳表示したアテンションマップ。上段がテスト用画像，下段が対応するアテンションマップを示す。AI モデルが最も注目した領域がアテンションマップで赤色に色付けされている。(A) 術式 A，(B) 術式 B，(C) 術式 C.

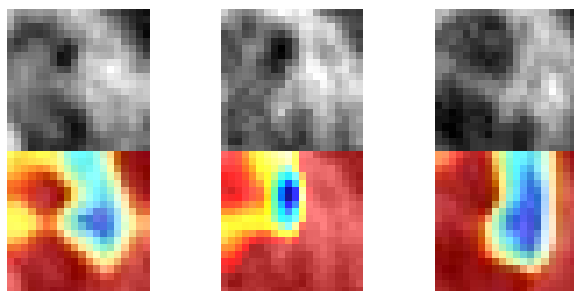
A の成功例



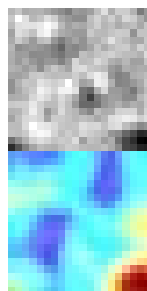
A の失敗例



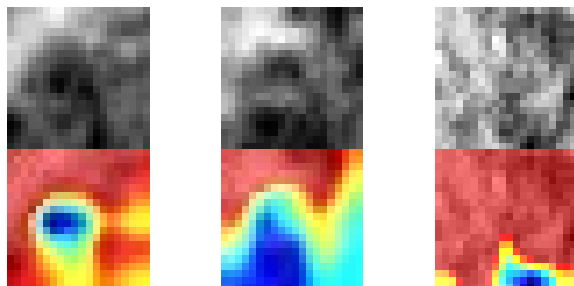
B の成功例



B の失敗例



C の成功例



C の失敗例

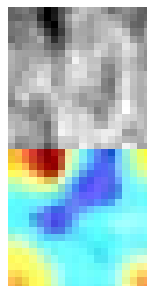


図 32. 成功例と失敗例の典型的なアテンションマップ