The University of Osaka
Institutional Knowledge Archive

Title	視線データ解析及びXAIを用いた臨床医の口腔粘膜疾患に対する診断精度向上を目的とした研究
Author(s)	内田, 修爾
Citation	大阪大学, 2022, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/87944
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

視線データ解析及びXAIを用いた臨床医の口腔粘膜疾患に対する診断精度向上を目的とした研究

2022 年 3 月

大阪大学大学院歯学研究科口腔科学専攻
顎口腔病因病態制御学講座
口腔外科学第一教室

内田修爾

緒言

悪性腫瘍や前がん病変を含む口腔粘膜疾患は、消化器がんとは異なり内視鏡等のデバイスを用いず、視診にて発見可能である。定期的に口腔内の検診を行うことで病変の早期発見を行うことが可能となり[1]、早期治療を行うことで後の生存率の向上に大きく影響を及ぼす[2]。

成書では、口腔粘膜疾患に関する特徴を視覚的に表示したアトラスも出版されているが、それぞれ鑑別診断として挙げられる疾患が複数、多岐にわたっており、診断に苦慮する例も散見される。適切なリスク評価を行えず、高次医療機関への紹介が遅れることで、発見時には既に進展例に至っている症例も多く見受けられる[3]。口腔がん進展例では原発巣切除に加え、組織の再建手術が必要になり、術後は咀嚼・嚥下機能のみならず整容面を含めた QOL の低下を引き起こす。進展例に至る前に、適切なリスク評価を行うためにも、口腔粘膜疾患の検診技術は重要であり、視診、触診から口腔内を客観的に評価することは必須である。

触診に関しては、初期舌癌の硬さを、超音波検査を用いたエラストグラフィーで評価する報告[4]も存在するが、症例数や評価基準に制約があり、現時点では、臨床医の触診を正確に評価することは困難である。また、インドで約 20 万人を対象とした、口腔がんに対する検診の効果に関する大規模な RCT が報告[5]されているが、検診を行った臨床医に対する前がん病変、口腔がんの識別に関するト

レーニングに用いられたのは、2 冊の目視検査マニュアルのみで、視診上での適切な検診方法に対する詳細な規定はなかった。以上から、口腔粘膜疾患の診断能力向上のためには、視診に関する教育ツールの開発が重要であると考えられるが、現時点では、病態写真に対する視診の方法を評価した研究は存在しない。

昨今、本邦では COVID-19 の世界的なパンデミックの影響から、人工知能（AI:artificial intelligence）技術を利用した医用画像診断支援システムや、オンライン診療導入の機運[6, 7]が高まっている。当教室では、口腔粘膜疾患診断支援 AI を開発中であるが、薬事承認を取得した AI 医療機器によってなされた診断の最終的な責任は使用する医師にあるとされている[8]。AI 医療機器の診断精度は急速に向上しているが、従来、深層学習（DL:Deep Learning）による AI はブラックボックスと称され、得られた回答の根拠を説明することは困難であった。しかしながら近年、Explainable AI（XAI）の開発等により、根拠を可視化する取り組みがすすんでいる。高い診断精度を誇る AI の根拠を可視化し、臨床医の診断技術向上に応用する取り組みは現時点において存在しない。

またオンライン診療導入を見据え、モニタ上で、口腔粘膜疾患の診断を行う際、どのような検診方法が最適なのか、明らかにすることは重要である。モニタ上での口腔粘膜疾患検診時の視線データを取得し、診療経験に勝る熟練医の診断技術、傾向を抽出、臨床医の診断技術向上に応用する研究も現時点においては存在

しない。

上記背景を踏まえ、モニタ上での口腔粘膜疾患に対する最適な検診方法の確立を目的とし、①XAI を用い、AI の診断根拠の抽出を行うことが、臨床医の診断技術向上に有用であるか、②モニタ上での口腔粘膜疾患写真に対する視線データを取得、解析することが、臨床医の診断技術向上に有用であるかという研究を行った。

材料・方法

本研究は、大阪大学大学院歯学研究科・歯学部附属病院倫理審査委員会（承認番号：H29-E35-1）において承認を得て実施した。

研究環境

表 1 に研究で用いた計算機、表 2 に開発環境を示す。本研究では DL の中でも画像認識分野に適している Convolutional Neural Network（CNN）[9]を用いて、研究を行った。

Convolutional Neural Network（CNN）

CNN は、生物の神経細胞の情報伝達の仕組みを模倣[10]した人工ニューラルネットワーク（ANN）を改良した構造である。入力層の直後に畳み込み層、プーリング層が組み込まれ、それらを複数回繰り返すことで深い層を形成し、その後全結合層も同様に層を重ねる構成である。まず、画像を 1 ピクセル毎に RGB（ディスプレイ上で色を表現するための尺度）の値が割り振られた 3 次元行列（RGB 値、縦ピクセル数、横ピクセル数）として捉える。畳み込み層では入力画像を小サイズのフィルタで 2 次元的にスライド（畳み込み演算）させることで、画像の局所的な特徴を抽出し、特徴を強調したり、画像全体をぼかせることで特徴を顕著化した画像（特徴マップ）を生成する。さらにプーリング層にて、畳み込み層で生成された画像から一定領域の最大値のみを取り出した

ピクセルの集合を行うことで、低画素に圧縮された画像へ変換し、さらに特徴を抽出した小さいサイズのデータを生成する。CNN は、全結合層の重みだけでなく、畳み込み層のフィルタ自体も重みとして学習し、学習が進むことでフィルタ値が変化していく。この構造により、入力画像を効率よく学習するためのデータへと変換することが出来、画像認識が可能となる。CNN の構造（入力層、畳み込み層、プーリング層、全結合層、出力層）を様々な形で多層化、組み合わせることで、より詳細な画像認識を可能するための試行錯誤がなされている。

CNN の学習方法

CNN は、訓練データ（入力）とネットワークが表す関数（出力）の誤差の尺度である loss（誤差関数）を数学的に最小へと近づけることで学習を行う。その際に用いられる数学的手法に、勾配降下法、誤差逆伝播法がある。

勾配降下法とは、重みを **w** とし、loss である E（w）が最小となるように学習によって重みを更新する方法である。更新前の重みを **w**、更新後の重みを **w'** とし、学習係数を ε、勾配を∇E とした際、

$$\mathbf{w}' = \mathbf{w} - \varepsilon \nabla E$$

の様に表わされる。更新量の如何によっては、最小の誤差へ到達するまでに時

間を要するため、最適な更新量の設定が必要となり、そのための計算手法として Adam[11]などが用いられる。しかし、全ての訓練データを用いて誤差の計算を行うと、計算時間がかかり、毎回同じデータでの学習となることから、訓練データに対して最適化してしまい、ある一定の最小誤差から抜け出せなくなる。つまり、新たなデータに対する汎用性に欠ける学習となる可能性がある。この問題を解決するため、訓練データのうち一部のみ抜き出して重みを調整する、（ミニバッチ）確率的勾配降下法が用いられている。

上述のように、ANN は出力と入力の誤差を求めて学習するが、その際、個々の loss を最小になるように調整し、出力層から遡って、各ニューロンの重みを調整していく。出力とは逆方向に効率よく重みを伝播していくことで、誤りを訂正するように結合重みを更新する方法を誤差逆伝播法と呼ぶ。本研究では、CNN モデルである ResNet50,InceptionV3 を用いた。

ResNet50

Residual Network（ResNet）[12]は、2015 年に発表された CNN であり、ネットワークの模式図を図 1 に示す。50 は学習可能な層の数を表す。2015 年当時、画像認識においては CNN の層数を増やすことでより高次元の特徴を獲得し、認識精度が改善することが知られていた。しかし、誤差逆伝播法では層を遡る毎に、乗算回数が増えていくため、勾配が 0 に近づく勾配消失問題

（gradient vanishing problem）が生じ、深いネットワークの場合入力層近くまで勾配が伝播されなくなった。この問題を解決するために、複数の畳み込み層を跨ぐ shortcut connection と Residual ブロックが導入された。通常のネットワークの処理が入力 x に対する畳み込み層の出力を $F(x)$、最終的な出力を $H(x)$とした際、

$$H(x)=F(x)+x$$

と表現される。この処理を Residual モジュールと呼び、図 2 に示す。ResNet はこのモジュールを用いることで、以前よりも圧倒的に多くの層を重ね、学習を行うことが可能となった。後に、様々な改良モデルが発表されている。

InceptionV3

InceptionV3[13]は GoogleNet[14]を改良したモデルであり、模式図を図 3 に示す。GoogleNet は複数の畳み込み層から構成された Inception モジュールを積層した構造になっており、次元削減、計算コストの削減が行える。Inception モジュールは 1×1 の畳み込みを行った後、3×3 または 5×5 の畳み込みを行って得られた特徴マップと 1×1 の畳み込みによる特徴マップ、プーリング後に 1×1 の畳み込みを行った特徴マップを結合して特徴マップ群を出力する構成である。

InceptionV3 は、GoogleNet にバッチ正規化（Batch Normalization）[15]を導入、改良した CNN である。ミニバッチによる学習においてネットワークの構

造が深くなると、ある層の特徴の分布がバッチ毎に大きく変化し、学習が効率よく進まない。バッチ正規化（Batch Normalization）とは各層においてバッチ毎の分布を調整して、なるべく分布の偏りが生じないよう調整する方法である。

＜研究 1＞AI の診断根拠の抽出から臨床医の診断技術向上を目指した研究

使用したデータセット、検証方法

大阪大学歯学部附属病院口腔外科 1（制御系）で原発の舌扁平上皮癌と診断され、撮影された口腔内写真（学習用 525 枚、評価用 10 枚）と正常粘膜の口腔内写真（学習用 221 枚、評価用 10 枚）を準備した。準備した口腔内写真に対し、それぞれ Cancer/Normal のラベル付けを行い、事前学習用のデータセットを作成した。ResNet50/InceptionV3 を用いて転移学習（あるタスク向けの学習済みモデルを用い、類似したタスクを実行するモデルの開始点として使用する方法で、学習時間の短縮、精度の向上が見込まれる。）を行い、AI 学習モデルを作成した。InceptionV3 を用いた学習の際には、validation_split=0.2 に設定し、モデル作成を行った。モデルの accuracy（正しい予測を行ったデータ数/検証に用いた全ての評価用画像のデータ数）、loss をそれぞれ算出し、XAI を使用し、それぞれの画像の診断の根拠となる領域をそれぞれ可視化した。

研究に用いた XAI

Grad-CAM

Gradient-weighted Class Activation Mapping （Grad-CAM）[16]は、予測値に対する勾配を重み付けすることで、重要なピクセルを可視化する技術である。模式図を図 4 に示す。原理としては、CNN を通して、クラス分類を行う際に得られる畳み込み層の出力とクラス分類の出力を取り出す。出力された分類結果を用いて誤差を逆伝播し、畳み込み層の各要素に対するクラス分類出力の勾配を計算する。その後、各畳み込み層から得た勾配の平均値を取得（グローバルアベレージプーリング（GAP））し、順方向で得られた畳み込み層の値に重み付けを行う。すべての畳み込み層で加算することで、CNN が元画像のどの領域に注目して、画像のクラスを判断しているのかを可視的に出力、表示することが可能となる。

LIME

Local Interpretable Model-agnostic Explanations （LIME）[17]は、画像やテキストを含む多様なデータに対する AI モデルの予測結果について、予測に寄与したデータの特徴を線形近似によって算出することが可能である。LIME のコンセプトを図 5 に示す。複雑な識別境界や内部ロジックを持つ AI モデルであっても、説明対象データの近傍に限れば、単純な線形モデルで近似できると仮定し、その近似した線形モデルを予測根拠の説明のために使用する。通常、画像データを AI

モデルに入力する場合は、各画素に対応した RGB 値を使用するが、画素単位で説明を行おうとすると計算量が膨大になり、また得られる説明結果も画素単位で解釈が難しくなる。そのため、画像をセグメントと呼ばれる領域に分割し、セグメント毎に判定への寄与度を算出する。

＜研究 2＞視線データを用いたモニタ上での適切な検診方法取得に関する研究

研究 2-1　疾患毎の正答率及び診断に要した時間の比較

対象

大阪大学歯学部附属病院に在籍中の口腔外科、保存科、予防歯科、補綴科に所属する歯科医師および開業歯科医師計 78 名。被験者の内訳を表 3 に示す。被験者を①初期研修医群、②口腔外科若手群（臨床経験 7 年以下）、③口腔外科熟練者群（臨床経験 8 年以上）、④口腔外科以外を専門とする歯科医師を一般歯科医群とし、4 群に分類した。

視線データ取得、解析に用いたデバイス

Tobii pro nano®（Tobii technology 株式会社,東京,日本）

ラップトップ PC やタブレット上での眼球運動を用いた個人の視覚的注意を研究することを目的としており、性別や民族を問わずモニタ上での目の動きを非常に高い精度で追跡することが可能である。個人の視線データ取得の記録を開始する前に、キャリブレーションという被験者の目の幾何学的特徴を取得し、

正確に視点を算出するためのプロセスが存在する。被験者がモニタ上に表示される特定の点（キャリブレーションポイント）を注視している際に、内蔵された近赤外 LED を用いて、被験者の角膜に光の反射パターンを生成し、幾何学的特徴（中心窩の位置、形状、光の屈折や反射特性に関する情報）とともにアイトラッキングカメラで取得する。解剖学的な 3D アイモデルを作成して個々の視点を算出することが可能となる。

Tobii pro Lab®（Tobii technology 株式会社,東京,日本）

アイトラッキングデータの取得、レコーディングの観察や解析、複数のデータを定量的に解析、HEATMAP による可視化が可能なソフトウェアである。手動で対象周囲の領域を設定し、同部位の注視回数、時間を算出することも可能である。

研究方法

17.3 型フル HD ディスプレイ（1920×1080）下に視線解析装置 Tobii pro nano®（Tobii technology 株式会社,東京,日本）を設置し、被験者を 65～70cm の距離に位置づけた。PC モニタに対して近遠心的な動きは、視線データ取得の妨げとなるため、被験者に出来るだけ同じ距離感、姿勢を保つよう指示した。実際の視線データ取得時の様子を図 6 に示す。ディスプレイ上 8 点でのキャリブレーションを行い、個々の視点の特徴を算出した後、診断の確定している口腔内写真（悪性腫瘍、口内炎、白板症、良性腫瘍、正常を各 5 枚）計 30 枚をモニタ上に

ランダムに表示し、口腔内写真毎に自由に検診を行い、臨床診断、病名をつけるよう指示した。それぞれの口腔内写真に対する視線座標データ、診断結果の正誤、診断に費やした時間を記録し、疾患、被験者群毎での正答率（各疾患,被験者群毎の正答数/各疾患,被験者群毎の全回答数）をそれぞれ算出、比較した。

研究 2-2　視線データを用いた臨床医の診断根拠の可視化に関する検討

使用した口腔内写真の選択基準

I. 舌に病変が存在する。

II. 病変の大きさの観点から、統計的な処理が可能である。（病変の領域面積がディスプレイ全体の 2%以上を占める。）

III. 正答率の観点から、偏りが少ない。（全体の正答率が 90%以下 10%以上,また口腔外科熟練者群の正答率が 50%以上）

視線データを取得した口腔内写真 30 枚のうち、上記の選択基準に該当した 10 枚を図 7 に示す。研究 2-2 以後ではこれらの口腔内写真を対象に研究を行った。

認知、理解における注視時間の閾値

視覚刺激の提示から情報が脳を介し、即座に行動に至るまでに経過した時間を単純反応時間と呼ぶが、現在に至るまで視覚刺激に対する単純反応時間に関する報告は多数なされている[18, 19]。立花ら[20]は、自然画像において、視覚から脳に情報が伝達され理解につながるには、250 ミリ秒以上の平均注視時間が必

要と報告している。そのため、研究 2-2 以後では 250 ミリ秒未満の注視点に関しては除外し評価した。

研究方法

研究 2-1 で取得した口腔外科熟練者群と一般歯科医群の視線座標データを用い、サポートベクトルマシン（SVM）[21],HEATMAP で各々の群の注視点の分布、注視領域の可視化を行った。

サポートベクトルマシン（SVM）

SVM は教師ありの機械学習アルゴリズムの一つで、入力データから基準を学習し、判断モデルを構築、それに基づいて未知の結果を予測する技術で、分類タスクに多く利用されている。本研究では 2 クラス（口腔外科熟練者群,一般歯科医群）の座標データ（写真 1 枚につき約 2500 個）から機械学習用ライブラリである scikit-learn を用いて、各クラスの非線形決定境界を作成した。研究に使用した口腔内写真と、口腔外科熟練者群と一般歯科医群の視線座標データを元に SVM で出力した画像を図 8 に示す。青プロットが口腔外科熟練者群、オレンジプロットが一般歯科医群の視点を表しており、青プロット周囲に同色の非線形領域が表示されている。この領域が、口腔外科熟練者群が注視している領域として表示される。病変中心部は両群とも注視点が集中しており、決定境界の作成が困難であることが分かる。

HEATMAP

視線が口腔内写真の一定箇所を注視している時間、または回数を算出し、そのデータを元に注目領域の注視の度合いを濃度勾配で表現しており、視覚情報を定量的に表示した可視化グラフの一種である。図 9 に示すカラーバーのように平均注視時間が長いほど、赤に近づき暖色に表示される。研究に使用した口腔内写真と視線の座標データを元に HEATMAP にて出力した画像を図 10 に示す。赤プロットが集積している領域は、画像内で注視時間が長い、もしくは注視回数が多い領域であるため、注目領域と意味づけられる。また、図 11 に示すように、多群間で注視点の散在を評価、比較することが可能となる。

研究 2-3　最適な検診方法及び診断の根拠となる領域に関する統計学的検討

研究方法

研究 2-1 で取得した口腔外科熟練者群と一般歯科医群のそれぞれの視線座標データを元に、疾患周囲の領域毎の注視回数、及び注視時間の比率（領域毎の注視時間/全注視時間）を比較、検討した。

病変の内外側領域の設定

写真内に存在する病変のどの部位をみているのか、具体的に評価するため、病変内外側を含め周囲に領域設定を行った。領域設定の方法を図 12 に示す。病変の外形を複数の臨床医の視点から基準として決定し、内部の面積が均等になる

様 4 個の領域に分割し、外側領域にも等面積に 2 個の領域を設定、計 6 領域を作成した。その領域を病変の中心部から順に領域 a～f と定義づけた。すなわち病変内部は領域 a～d、外部は領域 e、f と定義づけられる。

統計学的検証

EZR version 1.54[22]を使用し、多群間比較は one-way ANOVA にて分散分析を行い、交互作用が有意であった場合は Bonferoni 法で Post Hoc Test を行った。群内の有意差検定は Paired *t*- test、二群間の有意差検定は Student's *t*- test を用いた。

結果

結果 1　AI の診断根拠の抽出から臨床医の診断技術向上を目指した研究

事前学習用データセットを用いて作成したモデルの accuracy、loss を表 4 に示す。ResNet50,InceptionV3 から、作成したモデルは共に accuracy が 0.94 を超えており、高精度のモデルを作成することが出来た。上記のモデルを元に、XAI（Grad-CAM,LIME）を使用し、使用した画像の診断の根拠となる領域を可視化、出力した一例を図 13 に示す。LIME では、黒色に表示されている以外の領域、Grad-CAM では、赤色に表示されている領域が診断の根拠となる領域を示している。

XAI である LIME、Grad-CAM を用いることで、AI によって下された診断の根拠となる領域に関し、可視化は行えるが、細部に至る詳細な領域の抽出が難しく、人間の視覚情報に紐づけて臨床医の診断技術向上に役立てることは、現時点においては困難であった。

結果 2-1 疾患毎の正答率、診断に要した時間の比較

計測した疾患別の正答率、写真 1 枚当たりの平均診断時間について表 5，6 にそれぞれ示す。疾患別では、白板症が最も正答率が高く、口内炎は他の疾患に比べて、診断が困難であり正答率が 50%を下回る結果であった。口内炎の誤答例としては潰瘍、びらん、咬傷などの病態を回答している例、白板症、紅板症、舌がんというように誤答例が最も多岐にわたっていた。

また、口腔外科熟練者群は一般歯科医群と比較し正答率が高く、診断に要する時間が短い傾向を認め、全体を通して、正答率と診断時間との間に負の相関関係（相関係数（r）＝-0.721, $p<0.01$）があることが分かった。誤答者のほうが診断に要した時間が長く、診断に苦慮する傾向が得られた。

結果 2-2 視線データを用いた臨床医の診断根拠の可視化に関する検討

研究に使用した 10 枚の口腔内写真に対する視線座標データを元に、HEATMAP、SVM にて表示したところ、検診方法に 3 つの傾向が存在することが分かった。

初めに、HEATMAP にて共通した特徴を示した口腔内写真（図 14-A,D,G,J）と検診時の口腔外科熟練者群（図 14-B,E,H,K）、一般歯科医群（図 14-C,F,I,L）の視線座標データを元にそれぞれ HEATMAP 表示した結果（図 14）を示す。表面粗造な白色の悪性腫瘍（図 14-A）、辺縁均一な白板症（図 14-D）、均一な糜爛を認める口内炎（図 14-G）、均一な発赤を認める口内炎（図 14-J）の 4 写真に共通した傾向を認めた。二群間での HEATMAP を比較すると、口腔外科熟練者群が病変辺縁部も含め、注視点が広範囲に分布しているのに対し、一般歯科医群は病変中心部に注視が集中している傾向を認めた。

次に、SVM にて共通した特徴を示す口腔内写真（図 15-A,C,E,G）と、検診時の口腔外科熟練者群、一般歯科医群の視線座標データを元に、表示した決定境界

（図 15-B,D,F,H）を出力した結果（図 15）を示す。辺縁の形態がやや不均一な白板症（図 15-A,C）、色調不均一な血管腫（図 15-E）、発赤と白斑の色調の混在した潰瘍を伴う悪性腫瘍（図 15-G）の 4 写真に、共通した傾向を認めた。青色の決定境界で覆われた口腔外科熟練者群の注視領域が、病変辺縁部に一部一致しており、病変辺縁部から外側領域に位置している。

最後に、SVM にて共通した傾向を示した口腔内写真（図 16-A,C）において、表示した決定境界（図 16-B,D）を出力した結果を示す。内向型で表面性状に陥凹を認める悪性腫瘍（図 16-A）、外向型、表面に隆起を認める悪性腫瘍（図 16-C）のそれぞれの写真に共通した傾向を認めた。図 16-B は、口腔外科熟練者群の決定境界が図 15 に一致した傾向とは異なり、病変の外側辺縁部に一致して存在せず、病変の中心部に注視点が集中しており、領域決定が困難になっている。図 16-D は口腔外科熟練者群の決定境界が病変の内部の領域に至っており、いずれも病変の内側に口腔外科熟練者群の注視点が集中している傾向を認めた。

以上の結果から、HEATMAP と SVM を用いることで、注視点の離散具合、各群の注視領域の可視化が行え、口腔外科熟練者群と一般歯科医群の間で注視領域が異なることが分かった。

結果 2-3 最適な検診方法及び診断の根拠となる領域に関する統計学的検討

HEATMAP にて共通した傾向を示した 4 写真（図 14-A,D,G,J）について、口腔

外科熟練者群に関して、領域 a～f 毎の注視回数で比較した結果（図 17）、一般歯科医群に関して比較した結果（図 18）を示す。いずれの画像も口腔外科熟練者群の領域毎の注視回数は、領域 a と他の領域との間に有意差がなく、病変の内外側を問わず均等に視線データが存在している。一方、一般歯科医群の注視回数は領域 a と他の領域との間に、明確な差をつけた視診の傾向を認めた。

続いて SVM で共通した傾向を示した 4 写真（図 15-A,C,E,G）に関して、注視時間比率で比較した結果（図 19）を示す。一般的な統計的優位水準（p=0.05）に達しない写真も存在するが、いずれの写真も領域 d+e+f（病変の内側辺縁～外側領域）の注視時間比率で比較したところ、口腔外科熟練者群が一般歯科医群と比較して、同領域をよく診ている傾向を認めた。

最後に、SVM にて共通した傾向を認めた写真（図 16-A,C）を注視時間比率で比較した結果（図 20）を示す。図 17-A は口腔外科熟練者群が一般歯科医群と比較して、病変の中心部である領域 a、図 17-C は病変の内部全体である領域 a+b+c+d の注視時間比率が多く、口腔外科熟練者群が一般歯科医群と比較して、病変の内側領域を診ている傾向を認めた。これらは研究 2-2 で可視化した傾向と一致した。

以上の結果を踏まえ、肉眼的特徴と検診方法、診断の契機となる重要な領域の関係（図 21）についてまとめた。当研究内で、病変全体を均等に視診すること

が適切であるとされる写真群は、色調に関わらず、辺縁部が均一であるという肉眼的特徴が一致していた。病変の辺縁～外側が診断に重要であると予想される写真群は、色調が不均一であるという特徴が一致していた。病変の内側が診断に重要であると予想される写真群に関しては、表面性状に凹凸が存在するという共通した肉眼的特徴を有していた。このように検診方法が一致する病変は、肉眼的な特徴が類似していることが分かった。

考察

現時点において、医療分野における視線データを用いた解析例としては、CT、MRI、マンモグラフィー、パノラマエックス線写真といった、撮影方法が統一された画像上での、疾患に対する視線データを用いた研究[23-25]が報告されている。歯科の一次医療機関での診断能力向上や、将来的な AI 医療機器、オンライン診療導入を念頭に入れると、本研究で使用したような、撮影方法が統一されていない口腔内写真に対する視線データの解析が有用であると考えられる。

XAI を用いることで、AI の診断根拠の可視化は行えるが、詳細な領域抽出は困難であり、現時点においては臨床医の診断技術向上に役立てることは困難であった。Alejandro[26]らは、AI は作成したモデルの精度と、説明可能性との間に負の相関を認め、XAI によりその相関を解消すると説明している。本研究では高精度の AI 学習モデル（accuracy:0.94 以上）を用いたが、XAI が詳細な領域抽出を行えず、その役割を十分には果たせていないことから、今後臨床医の診断技術向上に役立てるためには、XAI の更なる技術革新や多様なデータセットを用いた研究、また他の CNN を用いて作成した AI 学習モデルにて、結果が改善するかについても研究が必要と思われる。

視線データを用いたモニタ上での適切な検診方法取得に関する研究において、被験者を診療経験に乏しい初期研修医群、口腔外科を専門とする口腔外科医群、

口腔外科以外を専門とする一般歯科医群に分類した。日本口腔外科学会が定める、口腔外科専門医取得に必要な臨床経験年数が最低 8 年であるため、口腔外科医群を 7 年以下、8 年以上に分類し、それぞれを口腔外科若手群、熟練者群と定義づけた。

視線データ取得時の疾患別の正答率は白板症が最も高く、口内炎は 50%を下回る結果であった。白板症は全体を通して高正答率が得られたが、他疾患での誤答例（特に視診上、白色の病変を認める写真）として挙げられることも多く、理由として視診上で白色病変を認めた場合に、その性状に関わらず、白板症と答える被験者が多いことが考えられる。悪性腫瘍は、誤診が予後に直結する疾患であるが、誤答例として口内炎、潰瘍、咬傷が挙げられており、臨床現場と同じく被験者のなかで、疾患のリスク評価の見誤りが多数見受けられた。口腔外科熟練者群は 6 疾患全てで最も高い正答率を算出しており、4 群の中では最も口腔粘膜疾患に対して、診断が優れている群と意味づけられる。

研究 2-2,2-3 で使用した口腔内写真については、上述の 3 条件を選択基準として設定した。選択基準の採用理由としては、①我が国で口腔がんの部位別発生頻度が最多である舌に病変が存在する[27]写真は、より注意深く検診することが求められるという点、②病変の大きさが小さすぎると、領域毎の注視回数、時間が病変中央に集中し、統計的な比較検討が困難であるという点、③使用した口腔

内写真の診断が著しく難しい、もしくは簡単であった場合、研究の意義が乏しくなるという点、④口腔外科熟練者群の正答率が低い写真は、比較対象としての優位性が保てない点の 4 点が挙げられる。

また、口腔外科熟練者群に対して、一般歯科医群を比較研究の対象群とした理由は、実際に、歯科の一次医療機関で、口腔粘膜疾患を持つ患者のリスク評価を行い、高次医療機関に紹介する機会が多いと考えたからである。初期研修医群は、臨床現場での診療経験が乏しいという点から検討には用いなかった。

HEATMAP,SVM を用いることで、各群の注視している領域、注視点の離散具合の可視化を行え、口腔外科熟練者群と一般歯科医群の間で注視する領域が異なることが分かった。口腔外科熟練者群は各疾患において高い正答率を示しており、口腔外科熟練者群の注視領域が正しい診断の根拠、重要な領域である可能性が示唆された。しかし SVM を用いた決定境界の領域表示は、病変中心部においては座標が集中するため、詳細な表示が困難であり、病変内側に重要な領域が存在する場合は、異なったアプローチが必要であると考える。今後、研究を継続し、被験者数、注視点の座標数が増加することで、SVM における決定境界の可視化は、病変中心部のような座標が集中した領域において、より困難となると考える。一方で、病変の外側領域に関しては、現在より詳細な決定境界の作成を行うことが可能となり、精密な特徴抽出が実現出来ると予想される。

研究 2-3 において、注視回数は各領域の注視点の分布を評価するため、注視時間比率は、口腔内写真の診断を決定する際に、どの領域を重要視しているかを評価するために、それぞれパラメータとして採用した。

HEATMAP,SVM を用いることで、可視化した診断に重要な領域と、統計学的な比較から得た結果との間に一致した傾向を認めた。また、確定診断した病名での分類ではなく、肉眼的特徴から口腔内写真を分類することで、カテゴリ毎の最適な検診方法、及び診断に重要となる領域の存在が示唆された。この結果から、口腔粘膜疾患の検診を行う際は、写真内の病変の位置を確認し、肉眼的な特徴を把握したうえで、その特徴に応じた検診方法を実施することが、臨床医の診断技術向上への一助となると考える。本研究では、肉眼的特徴のカテゴリとして①色調に関わらず辺縁が均一、②辺縁部の色調が不均一、③粘膜の表面性状に凹凸が存在するの 3 つが挙げられた。いずれのカテゴリも病変の境界部の適切な認識が重要であり、認識が不十分であれば、病変の内外側領域の設定が曖昧となることから、正しい診断、検診を行うためには、まず病変の外形、境界部の適切に把握することが重要であると考える。

次に、各カテゴリの最適な検診方法について検討すると、色調に関わらず、辺縁が均一な 4 病変に関しては、病変の内外側を問わず均等に視診することが適切な検診方法であると示唆されている。その理由としては辺縁が均一であるが

故に、病変周囲も含め不整な領域がないか確認するためだと考えられる。辺縁部の色調が不均一な 4 病変に関しては、病変の境界部～外側領域が診断の根拠となる領域であると示唆されるが、同部位には白斑の不整部、発赤と白斑の混在した部位、色調の濃淡が異なった部位がそれぞれ存在し、それらの特徴を捉えるため注意深く検診していると考えられる。表面性状に凹凸が存在する 2 病変に関しては、病変の内側が診断の根拠となる領域であると示唆されるが、同部位には病変表面に存在する隆起、陥凹といった所見が存在し、それらの特徴を捉えるため注視時間比率が増えたと考えられる。3 カテゴリともそれぞれ特徴のある部位を確実に捉えることが、正しい診断に繋がると考えられる。

また、口腔癌の肉眼分類としては外向性発育をみせる隆起性腫瘍と内向性発育をみせる浸潤性腫瘍の 2 分類[28-31]、日本口腔腫瘍学会は、外向型、内向型に表在型を加え、3 つに分類することを提唱している[27]。本研究で用いた 10 枚の口腔内写真のうち、悪性腫瘍の写真は 4 枚含まれており、上述の分類に当てはめると、表在型が 2 枚（図 14-A,15-G）、内向型が 1 枚（図 16-A）、外向型が 1 枚（図 16-C）と分類される。従来の肉眼分類と、本研究の肉眼的特徴を元に定義したカテゴリとは、全く異なっている。しかし本研究で定義されたカテゴリには、悪性腫瘍だけでなく、他の口腔粘膜疾患も含んでおり、正しい検診傾向から得た分類として新たに提案することが出来る。

更に、初期舌癌（cT1N0,cT2N0）においては肉眼分類の内向型は表在型や外向型に比べ、頸部リンパ節転移率が高く、治療に際しては注意を要すると報告されている[32-35]。また、内向型は表在型、外向型に比べて原発巣再発率が高いとされている[35-39]。このような報告例から、本研究で定義したカテゴリにおいては、①粘膜の表面性状に凹凸が存在する病変＞②辺縁部の色調が不均一な病変＞③色調に関わらず辺縁が均一な病変の順に予後が悪い可能性が高く、注意深い経過観察が必要であると予想される。予後が悪いと予想される疾患の肉眼的特徴、対応した適切な検診方法を理解することが、悪性腫瘍の早期発見、高次医療機関への早期紹介に繋がると思われる。

現状、口腔がんの肉眼分類[27]が提唱されてはいるが、客観的な評価が難しく、判定に迷う症例も多く、前がん病変や早期がんを肉眼所見から正確に評価、分類する方法は存在しない[40, 41]。今後、本研究で示唆された肉眼的特徴を元にしたカテゴリが妥当であるか、口腔内写真の種類及び被験者を増やして検討する必要がある。

本研究に関しては二点の制約がある事を認める。まず、口腔内写真に選択基準を設けたことで、選択バイアスが否定できない点である。今後は、病変の部位や大きさ、診断への難易度に関わらず、統計処理を行える方法についても模索する必要がある。もう一点は、被験者の大多数が、大阪大学歯学部附属病院に在籍す

る歯科医師であり、ほぼ単施設でのデータ取得となったため、検診の傾向にバイアスが生じる点である。この点に関しては、今後、開業歯科医師の被験者数を増やすことで解消できればと考えている。

今後、視線データを蓄積することで得られた知見を利用し、将来的な教育ツールとして運用することを想定すると、本研究で用いた様な、疾患が含まれる口腔内写真だけでなく、定期的な歯科検診を想定した正常と似通った粘膜写真に対して、どのような手順で検診を行えば、疾患の見落としが防げるのかについても研究を行う必要があると考える。

結語

視診上での注目領域、注視点の離散具合を HEATMAP,SVM を使用し、可視化することで、口腔外科熟練者群と一般歯科医群の間で注視する領域が異なることが明らかになった。さらに、画像毎に口腔外科熟練者群の注視領域に異なる特徴を認め、同領域が正しい診断に寄与する根拠である可能性が示唆された。

また、肉眼的特徴から口腔内写真を分類することで、カテゴリ毎の最適な検診方法及び診断の根拠となる領域の存在が示唆された。

以上の結果から、口腔粘膜疾患写真に対する臨床医の視線データを取得、解析することで、診断技術向上に寄与する視診上での特徴抽出が可能であり、本研究がモニタ上における口腔粘膜疾患に対する最適な検診、診断方法確立のための基盤的データとなり得ることが示された。

謝辞

本論文を作成するにあたり、本研究を行う機会を与えていただき、的確なご助言を賜りました大阪大学大学院歯学研究科顎口腔病因病態制御学講座（第一口腔外科） 田中晋教授に深謝いたします。また本研究を行うにあたり、統計学的な側面から懇切なるご指導をいただきました大阪大学 感染症総合教育研究拠点 特任准教授 井出和希先生に心より御礼申し上げます。また本研究を進めるにあたり、PC 環境の構築及び Deep Learning に関する基本的な知識及び操作方法をご助言いただきました大阪大学大学院歯学研究科顎口腔病因病態制御学講座（第一口腔外科） 川村晃平先生に深謝いたします。また本研究に携わる契機となり、研究環境を整えていただき、ご多忙の中常時ご助言を頂きました大阪大学大学院歯学研究科顎口腔病因病態制御学講座（第一口腔外科） 平岡慎一郎先生に深く感謝いたします。

引用文献

1. Khalili J: Oral cancer and role of general dental practitioners in its early detection. Lik Sprava 10, 2007

2. Kujan O, Glenny AM, Oliver RJ, Thakker N, Sloan P: Screening programmes for the early detection and prevention of oral cancer. Cochrane Database Syst Rev Cd004150, 2006

3. 松浦 一登: 進行舌がんに対する治療戦略. 頭頸部癌 41:13, 2015

4. Shingaki M, Nikkuni Y, Katsura K, Ikeda N, Maruyama S, Takagi R, Hayashi T: Clinical significance of intraoral strain elastography for diagnosing early stage tongue carcinoma: a preliminary study. oral radiol. 33:204, 2017

5. Sankaranarayanan R, Ramadas K, Thomas G, Muwonge R, Thara S, Mathew B, Rajan B: Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial. Lancet 365:1927, 2005

6. 厚生労働省: 新型コロナウイルス感染症の拡大に際しての電話や情報通信機器を用いた診療等の時限的・特例的な取扱いについて. 2020 https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryou/rinsyo/index_00015.html

7. 厚生労働省: オンライン診療の適切な実施に関する指針. 2018 https://www.mhlw.go.jp/file/05-Shingikai-10801000-Iseikyoku-Soumuka/0000201789.pdf

8. 厚生労働省医薬・生活衛生局医療機器審査管理課: 人工知能技術を利用した医用画像診断支援システムに関する評価指標. 2019 https://www.mhlw.go.jp/content/10601000/000515843.pdf

9. LeCun Y, Haffner P, Bottou L, Bengio Y: Object recognition with gradient-based learning. Shape, Contour and Grouping in Computer Vision 1681:319, 1999

10. McCulloch WS, Pitts W: A logical calculus of the ideas immanent in nervous activity. 1943. Bull Math Biol 52:99, 1990

11. D. P. Kingma aJB: Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, May 2015

12. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, (ed., 2016, p 770

13. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z: Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (ed., 2016, p 2818

14. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A: Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, (ed., 2015, p 1

15. Ioffe S, Szegedy C: Batch normalization: Accelerating deep network training by reducing internal covariate shift. International conference on machine learning, (ed., PMLR, 2015, p 448

16. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D: Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision, (ed., 2017, p 618

17. Ribeiro MT, Singh S, Guestrin C: "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (ed. San Francisco, California, USA, Association for Computing Machinery, 2016, p 1135

18. 永井 大介, 長谷川 賢: 聴覚・視覚刺激反応時間に関する研究. 昭和医会誌 46:27, 1986

19. Hamburger H: Donald R. J. Laming. Information, theory of choice-reaction times. New York: Academic Press, 1968. Behav Sci 14:330, 1969

20. 立花良: 自然画像内における人物の知覚認知：時間情報処理プロセスと視点効果に関する実験心理学的検討. 東北大学大学院文学研究科博士学位論文(未公刊), 2018

21. Cortes C, Vapnik V: Support-Vector Networks. Mach. Learn. 20:273, 1995

22. Kanda Y: [Statistical analysis using freely-available "EZR (Easy R)" software]. Jpn J Clin Hematol 56:2258, 2015

23. Bahaziq A, Jadu FM, Jan AM, Baghdady M, Feteih RM: A Comparative Study of the Examination Pattern of Panoramic Radiographs Using Eye-tracking Software. J Contemp Dent Pract 20:1436, 2019

24. Fujii C, Onuma T, Nakamura F, Tome W, Sakai N, Sako N, Kitai N: Influence of dental education on eye gaze distribution when observing facial profiles with varying degrees of lip protrusion. J Dent Educ 85:476, 2021

25. Gandomkar Z, Mello-Thoms C: Visual search in breast imaging. Br J Radiol 92:20190057, 2019

26. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F: Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82, 2020

27. 口腔がん診療ガイドライン改訂合同委員会編: 口腔がん診療ガイドライン 2019 年版. 金原出版,

28. Harasymczuk M, Gooding W, Kruk-Zagajewska A, Wojtowicz J, Dworacki G, Tomczak H, Szyfter W, Whiteside TL: Head and neck squamous carcinomas with exophytic and endophytic type of growth have the same prognosis after surgery and adjuvant radiotherapy. Eur Arch Otorhinolaryngol 270:1105, 2013

29. Jakobsson PA, Eneroth CM, Killander D, Moberger G, Mårtensson B: Histologic classification and grading of malignancy in carcinoma of the larynx. Acta Radiol Ther Phys Biol 12:1, 1973

30. Sotiriou C, Lothaire P, Dequanter D, Cardoso F, Awada A: Molecular profiling of head and neck tumors. Curr Opin Oncol 16:211, 2004

31. Yilmaz T, Hoşal AS, Gedikoğlu G, Kaya S: Prognostic significance of histopathological parameters in cancer of the larynx. Eur Arch Otorhinolaryngol 256:139, 1999

32. 日本口腔腫瘍学会学術委員会: 舌癌取扱い指針
ワーキング・グループ案（第 1 版). 日本口腔腫瘍学会誌 17:13, 2005

33. 小浜 源郁, 野口 誠, 木戸 幸恵, 久保田 裕美, 金城 尚典, 宮崎 晃亘: 口腔扁平上皮癌の臨床所見と病理組織所見に基づいた外科療法
—481 例の分析—. 日口腔腫瘍会誌 13:33, 2001

34. Asakage T, Yokose T, Mukai K, Tsugane S, Tsubono Y, Asai M, Ebihara S: Tumor thickness predicts cervical metastasis in patients with stage I/II carcinoma of the tongue. Cancer 82:1443, 1998

35. 姉川 絵美子, 古賀 真, 岩本 修, 楠川 仁悟: 臨床発育様式からみた T1, T2 舌扁平上皮癌の悪性度評価に関する臨床病理学的検討. 日口腔腫瘍会誌 18:83, 2006

36. Capote-Moreno A, Naval L, Muñoz-Guerra MF, Sastre J, Rodríguez-Campo FJ: Prognostic factors influencing contralateral neck lymph node metastases in oral and oropharyngeal carcinoma. J Oral Maxillofac Surg 68:268, 2010

37. Kirita T, Okabe S, Izumo T, Sugimura M: Risk factors for the postoperative local recurrence of tongue carcinoma. J Oral Maxillofac Surg 52:149, 1994

38. Po Wing Yuen A, Lam KY, Lam LK, Ho CM, Wong A, Chow TL, Yuen WF, Wei WI: Prognostic factors of clinically stage I and II oral tongue carcinoma-A comparative study of stage, thickness, shape, growth pattern, invasive front malignancy grading, Martinez-Gimeno score, and pathologic features. Head Neck 24:513, 2002

39. Spiro RH, Guillamondegui O, Jr., Paulino AF, Huvos AG: Pattern of invasion and margin assessment in patients with oral tongue cancer. Head Neck 21:408, 1999

40. Fedele S, Lo Russo L, Mignogna C, Staibano S, Porter SR, Mignogna MD: Macroscopic classification of superficial neoplastic lesions of the oral mucosa: a preliminary study. Eur J Surg Oncol 34:100, 2008

41. Mignogna MD, Fedele S, Lo Russo L: Dysplasia/neoplasia surveillance in oral lichen planus patients: a description of clinical criteria adopted at a single centre and their impact on prognosis. Oral Oncol 42:819, 2006

表 1:計算機の環境

環境名 構成	HPCDIY-UPDL2	Nitro AN517-51
Operating System	Ubuntu18.04LTS	Windows10 Home
Central Processing Unit	Intel Xeon W-2245	Intel Core i7-9750H プロセッサー/4.50GHz
Memory	11GB　GDDR6	16GB
Graphical Processing Unit	NVIDIA® GeForce® GTX2080Ti	NVIDIA® GeForce® GTX1660Ti(6GB)

表 2:主な開発環境

ソフトウェア	検証 1	検証 2
Python	3.6.2	3.7.9
Keras	2.6.0	
Pillow	8.3.2	8.1.0
TensorFlow	2.6.0	
scikit-learn		0.24.1
matplotlib	3.2.2	3.3.4
Numpy	1.19.5	1.20.1
pandas	0.20.3	1.2.1

表 3:検証 2 に参加した被験者の内訳

		男性	女性	計
歯科医師歴	初期研修医	11	5	16
	2～7 年	26	16	42
	8～15 年	7	4	11
	16 年以上	5	4	9
専門領域	①初期研修医群	11	5	16
	②口腔外科若手群 (臨床経験 7 年以下)	18	8	26
	③口腔外科熟練者群 (臨床経験 8 年以上)	8	5	13
	④一般歯科医群 (口腔外科以外専門)	12	11	23
	計	49	29	78

表 4:作成した AI 学習モデルの評価パラメータ

CNN / パラメータ	ResNet50	InceptionV3
Accuracy	0.9441	0.9412
Loss	0.2834	0.1540

表 5:疾患別の正答率(%)

疾患別 / 専門/経験年数	悪性腫瘍	口内炎	正常	白板症	良性腫瘍
初期研修医群	52.08	42.70	51.04	82.29	50.00
口腔外科若手群	82.05	43.58	71.61	93.58	81.41
口腔外科熟練者群	91.02	60.25	85.71	94.87	97.43
一般歯科医群	64.49	34.05	49.27	55.79	44.92
全体	72.22	43.37	63.09	80.34	71.30

表 6:1 枚当たりの平均診断時間(秒)

疾患別 / 専門/経験年数	悪性腫瘍	口内炎	正常	白板症	良性腫瘍
初期研修医群	13.01	12.30	14.92	9.69	9.19
口腔外科若手群	10.89	12.62	13.63	7.23	10.22
口腔外科熟練者群	9.90	13.47	11.13	9.23	9.00
一般歯科医群	10.49	12.82	13.69	12.41	12.30
全体	11.04	12.76	13.49	9.59	10.42

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

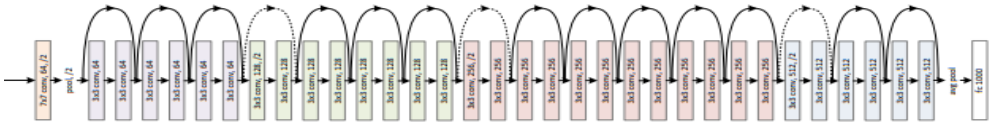


図 1:ResNet のネットワーク構造と模式図(13)

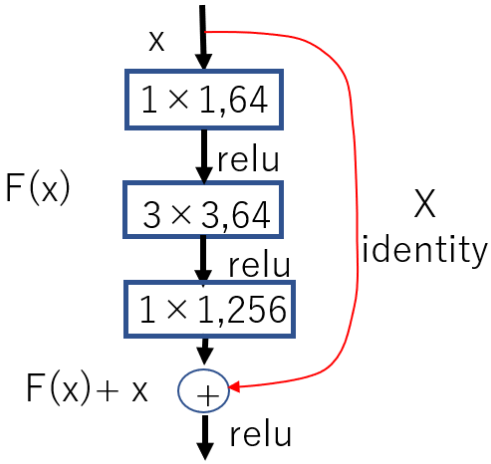


図 2:Residual モジュールの模式図

ResNet に多数導入されている。F(x)と x の残差算出を各畳み込み層が担っており、畳み込み層への入力を出力に加算している。

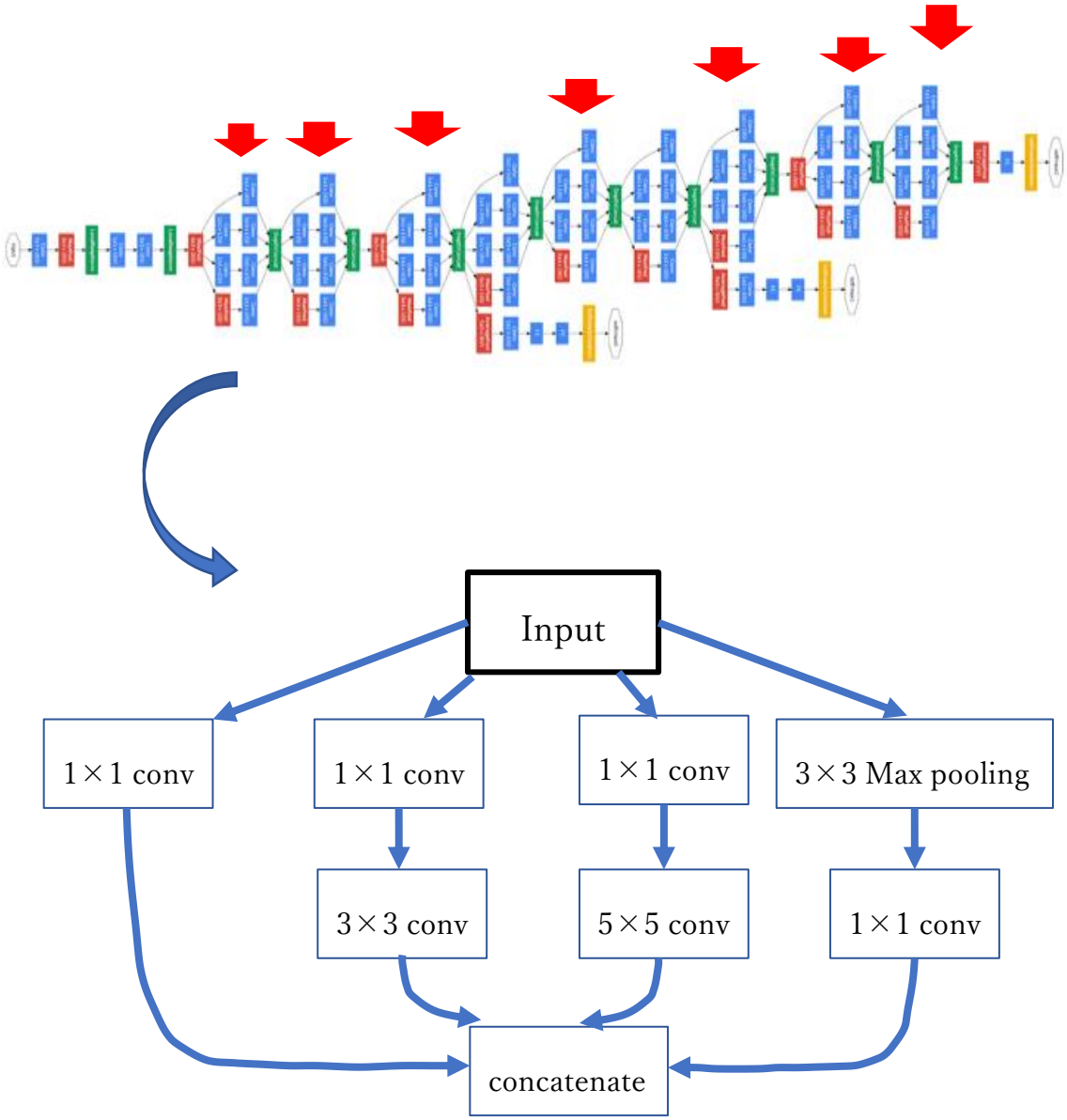


図 3:GoogleNet のネットワーク構造の模式図(14)

Inception モジュール(赤矢印)が多数導入されており、模式図を下図に示す。

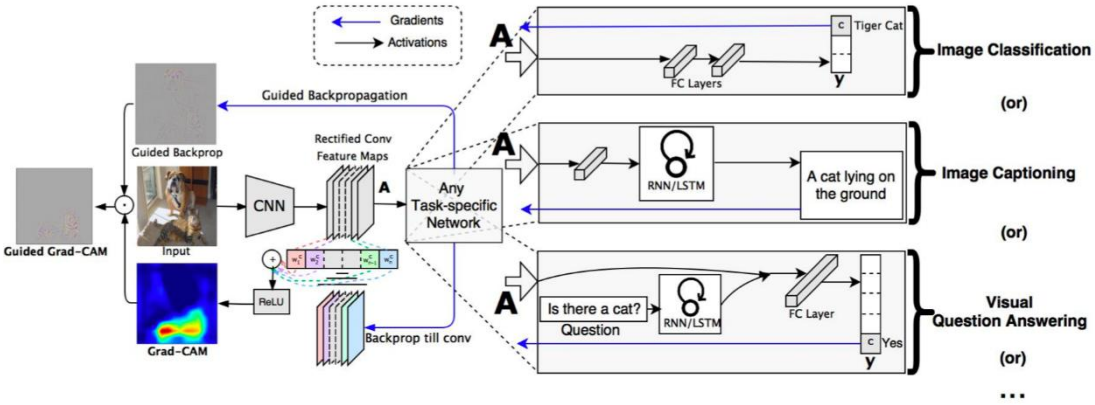


図 4:Grad-CAM の構造の模式図(16)

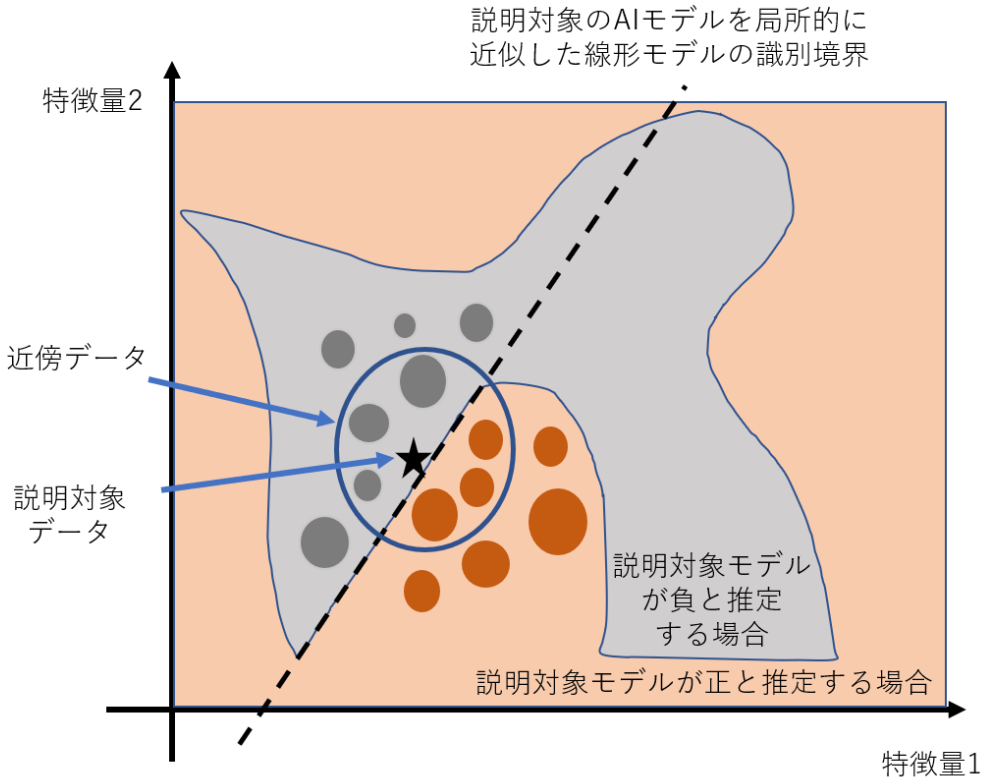
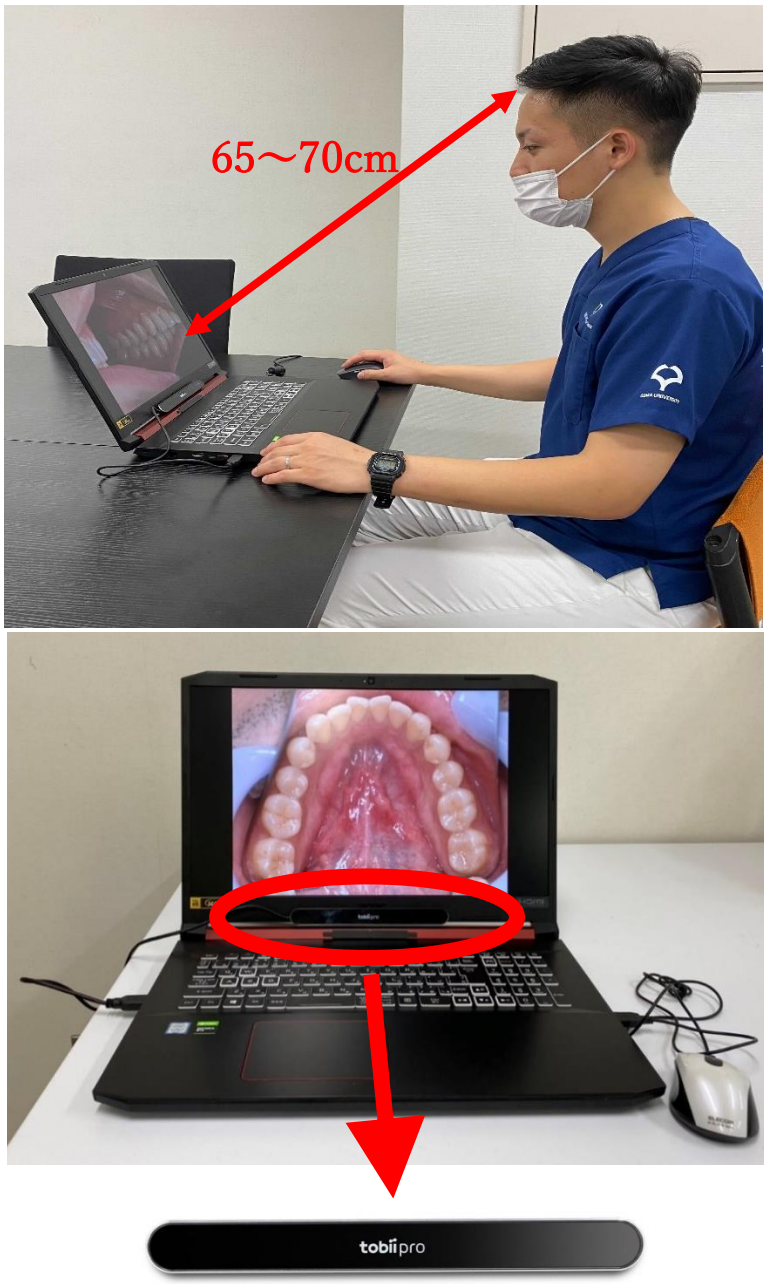


図 5:LIME のコンセプトの模式図(17)

説明対象データの近傍をセグメントと呼ばれる線形モデルで分割し、その近似した線形モデルを可視化、予測根拠の説明に利用する。



Tobii pro nano®(Tobii technology)

図 7:実際の視線データ取得時の様子

17.3 型フル HD ディスプレイ(1920×1080)下に視線解析装置 Tobii pro nano®(Tobii technology)を設置し、被験者の目との距離を 65～70cm に設定し、データ取得を行った

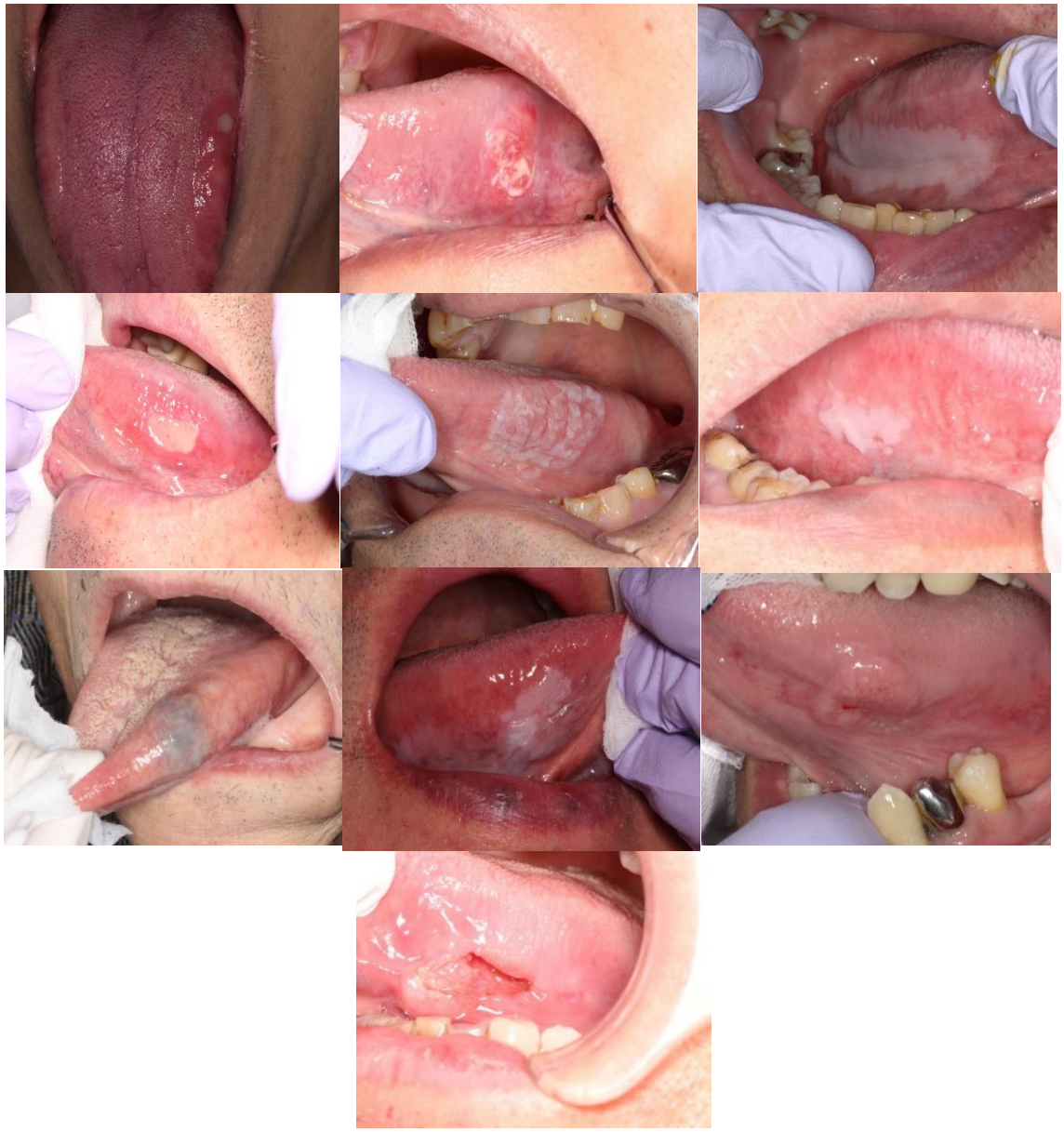


図 7:検証 2-2,2-3 に使用した口腔内写真

選択基準を満たした上記 10 枚の口腔内写真で検証 2-2,2-3 を行い、口腔外科熟練者群、一般歯科医群の検診時の視線データを取得、解析行った。

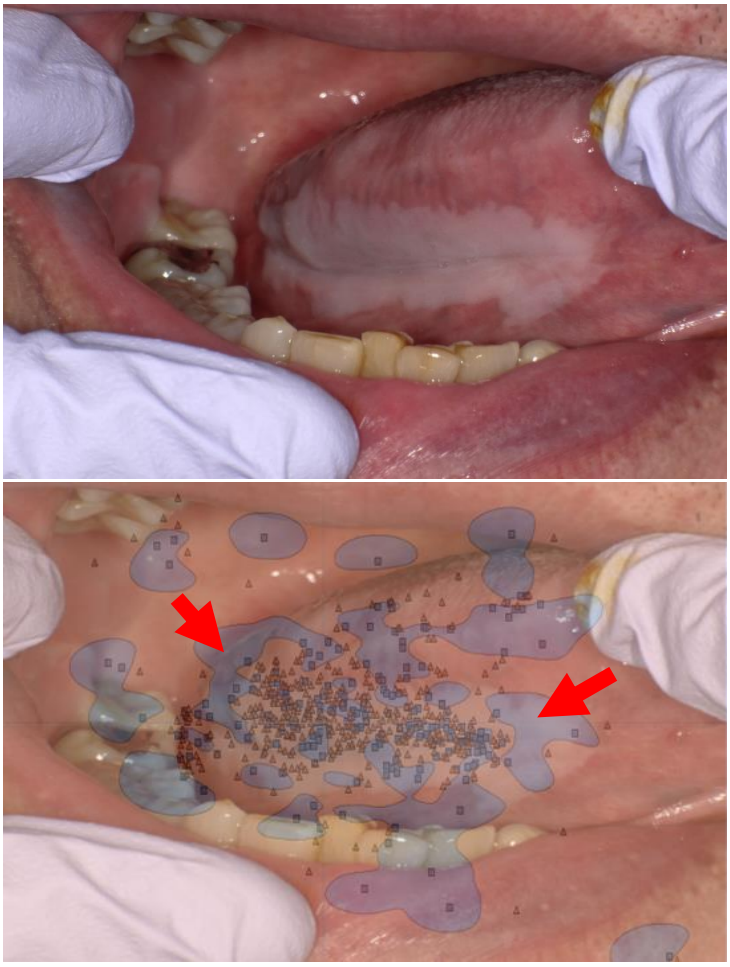


図 8:元画像(上)と視線座標データを基に

SVM にて出力(下)した画像例

青プロット:口腔外科熟練者群、オレンジプロット:一般歯科医群の注視点を表している。青プロット周囲青色の非線形領域(矢印)が表示される。この領域が、口腔外科熟練者群が注視している領域として表示される。

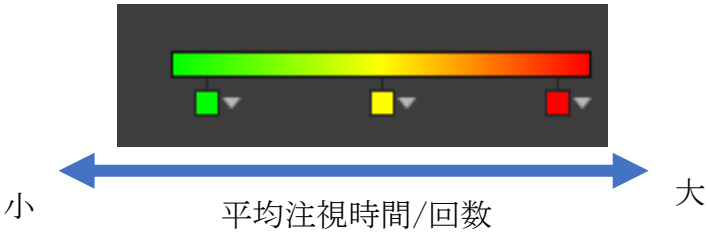


図 9:注視時間、回数の程度を視覚的に表すカラーバー

平均注視時間もしくは回数が少なければ緑に、多ければ赤に近づき暖色に表示される。

図 10:元画像(左)と視線座標データを元に

HEATMAP にて出力(右)した画像例

赤プロットされている領域(黄矢印)は画像内で注視時間が長いもしくは注視回数が多い領域であるので、注目領域として表示される。

図 11:異なった 2 群間の視線データを元に出力した HEATMAP 例

左画像は注視点が多く、離散しているのに対し、右画像は注視点が一か所に集中している。HEATMAP を用い、多群間で注視点の散在の度合いを可視化し比較することが可能である。

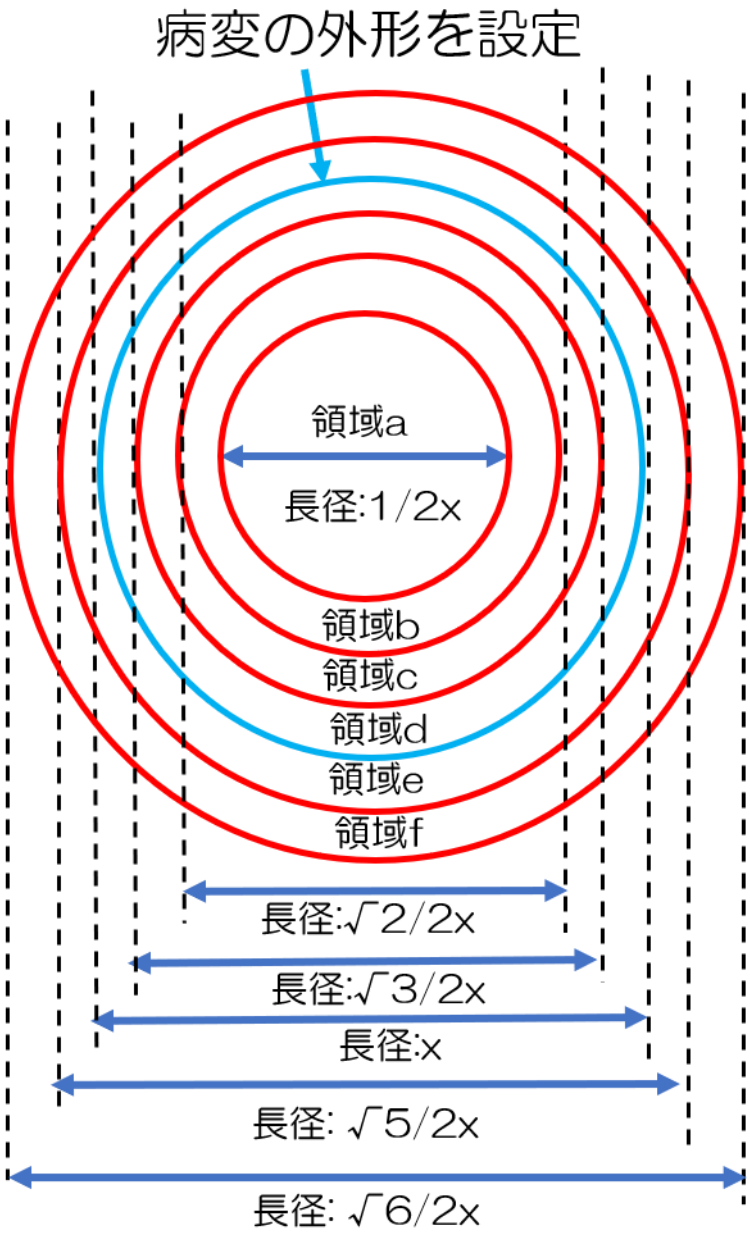


図 12:病変の内外側領域の設定方法

病変の外形を複数の臨床医の視点から基準として設定、内部の面積が均等になる様 4 個の領域に分割し、外側領域にも等面積に 2 個の領域を設定し、計 6 領域を作成した。その領域を病変の中心部から順に領域 a～f と定義づけた。

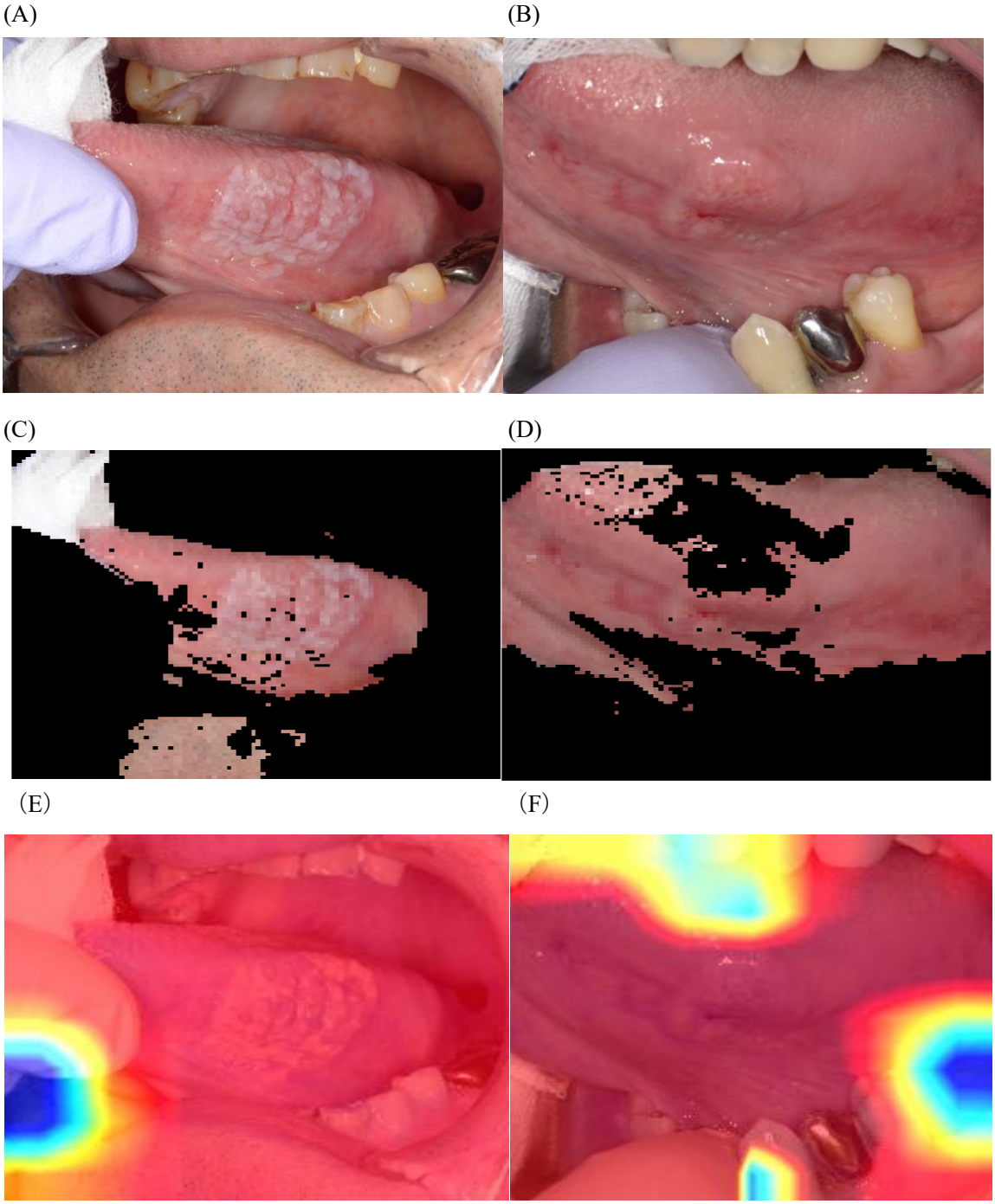


図 13:（結果 1）

A,B：元画像、C,D:LIME、E,F:Grad-CAM でそれぞれ出力された画像を示す。LIME は黒く表示されている以外の部位、Grad-CAM は赤く表示されている部位が AI の診断の根拠となった領域を意味する。

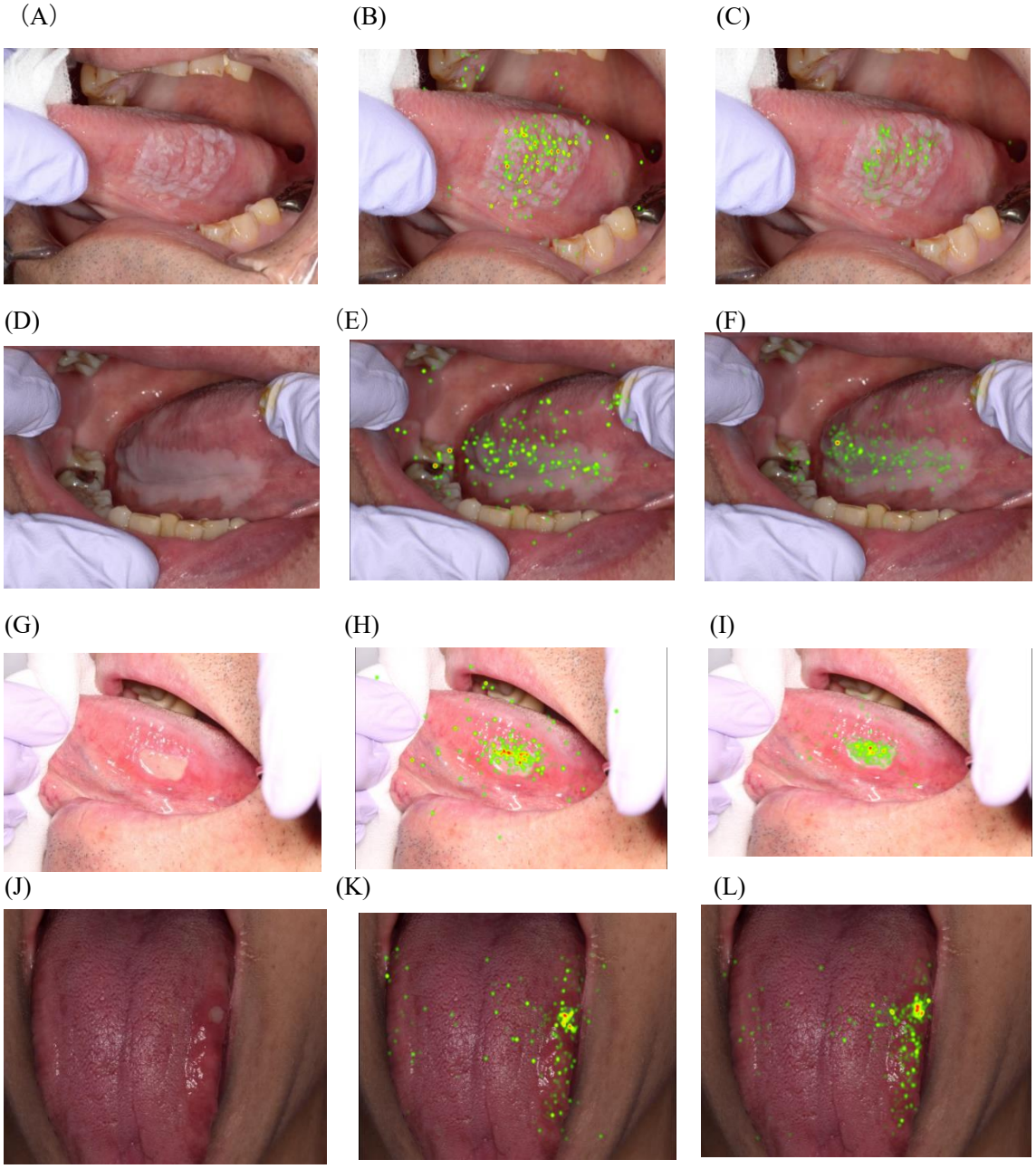


図 14：検診時の視線データの HEATMAP による比較

A,D,G,J：元画像（A）：表面粗造な白色の悪性腫瘍, （D）：辺縁均一な白板症
（G）：均一なびらんを認めるロ内炎, （J）：均一な発赤を認めるロ内炎
B,E,H,K：口腔外科熟練者群、C,F,I,L：一般歯科医群の視線データを元にそれぞれ HEATMAP 表示した画像を示す。口腔外科熟練者群が、病変辺縁部も含め注視点が広範囲に分布しているのに対し、一般歯科医群は病変中心部に注視点が集中している傾向を認めた。

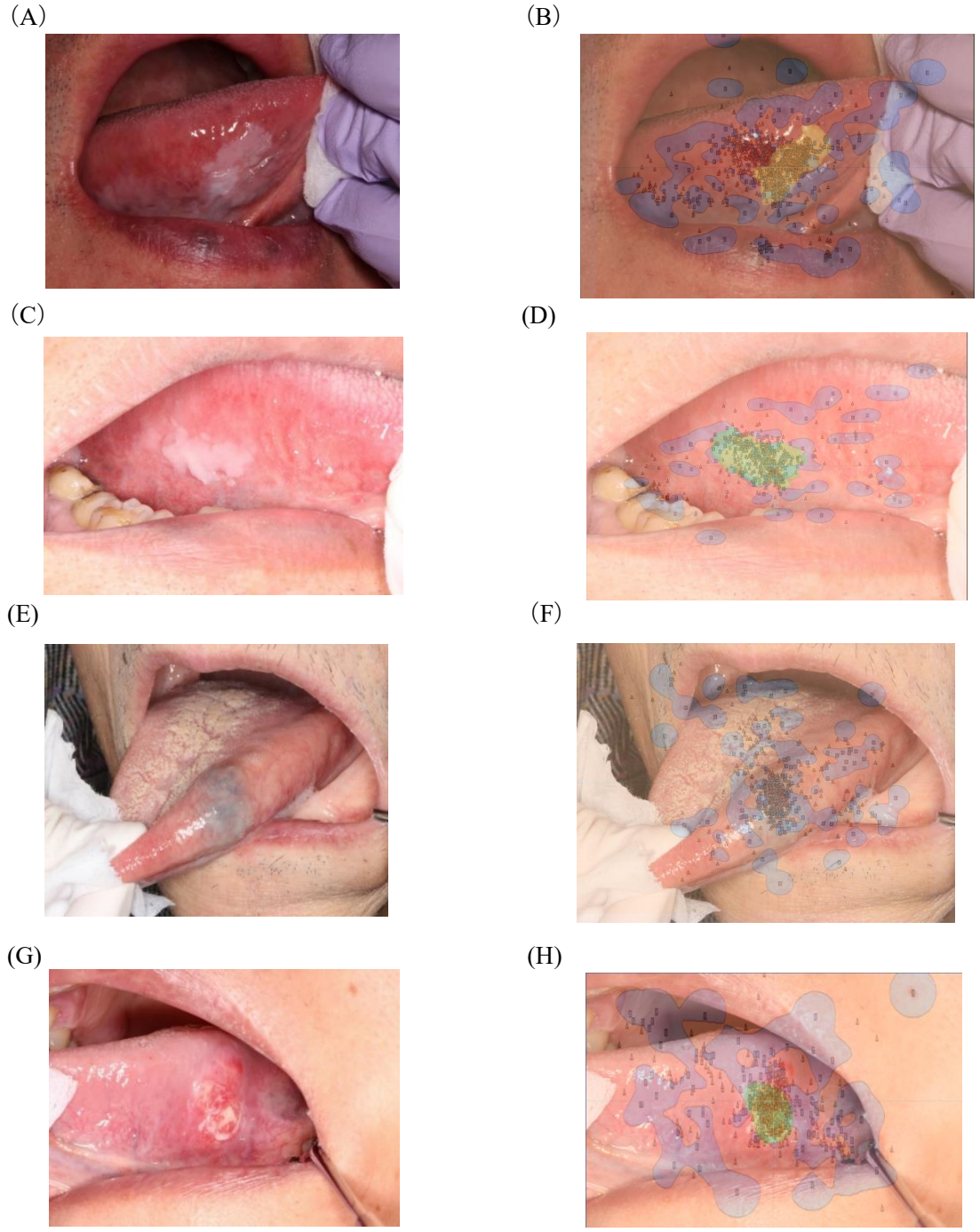


図 15：検診時の視線座標データの SVM による比較

A,C,E,G：元画像　（A,C）辺縁の形態がやや不均一な白板症、(E) 色調不均一な血管腫 (G) 発赤と白斑の色調の混在した潰瘍を伴う悪性腫瘍
B,D,F,H：口腔外科熟練者群, 一般歯科医群の視線データを元に SVM にて表示した画像を示す。口腔外科熟練者群の注視領域である青で示された決定境界が、緑で示された病変の外形の辺縁～外側にかけて存在した。

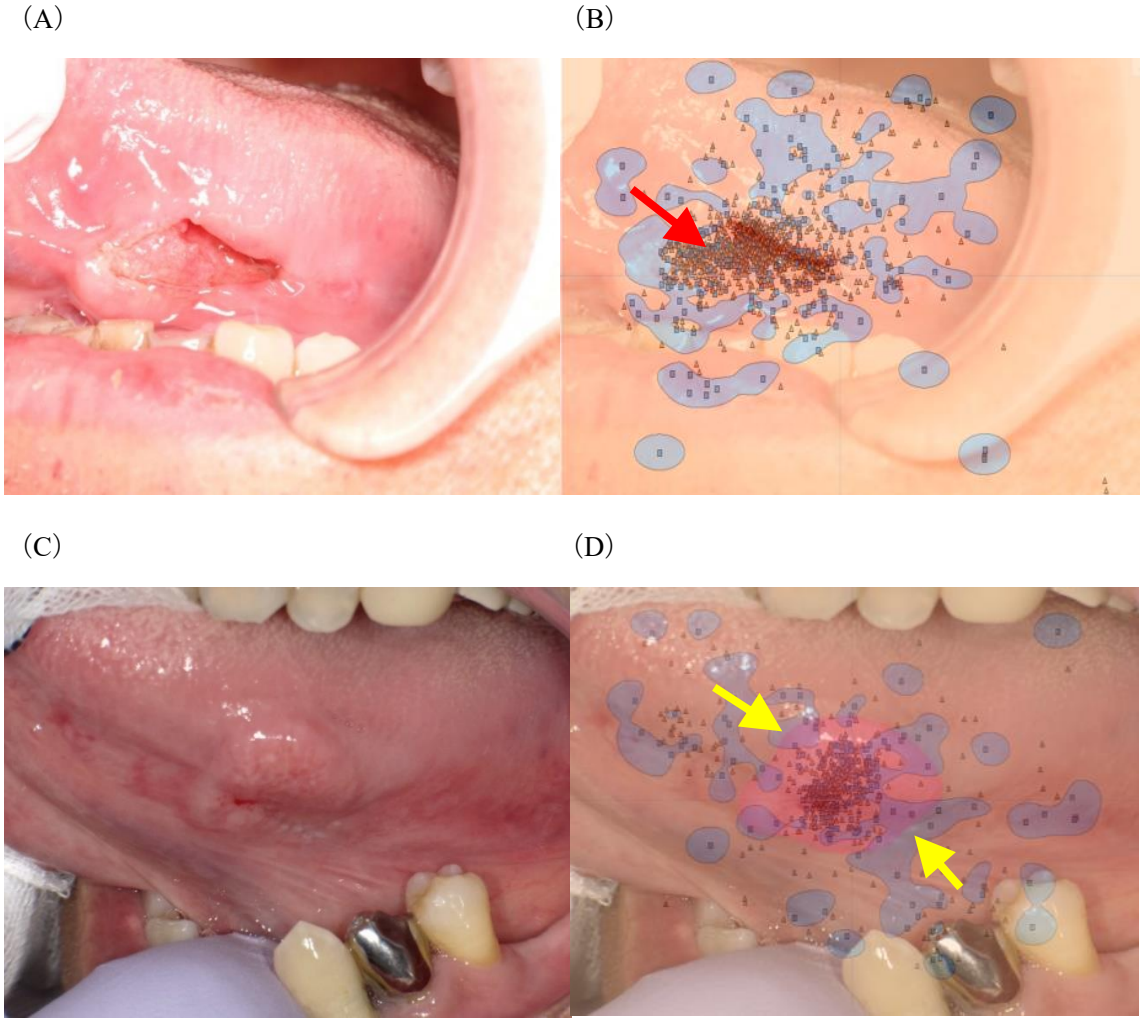


図 16:検診時の視線座標データの SVM による比較

A,C:元画像（A)内向型、陥凹を認める悪性腫瘍(C)外向型、隆起を認める悪性腫瘍　B,D:口腔外科熟練者群, 一般歯科医群の視線データを元に SVM にて表示した画像を示す。(B) 口腔外科熟練者群の注視領域である青で示された決定境界が病変内側に注視点が集中している(赤矢印)。 (D)赤で示す病変内側に口腔外科熟練者群の注視領域である決定境界が及んでいる(黄矢印)。

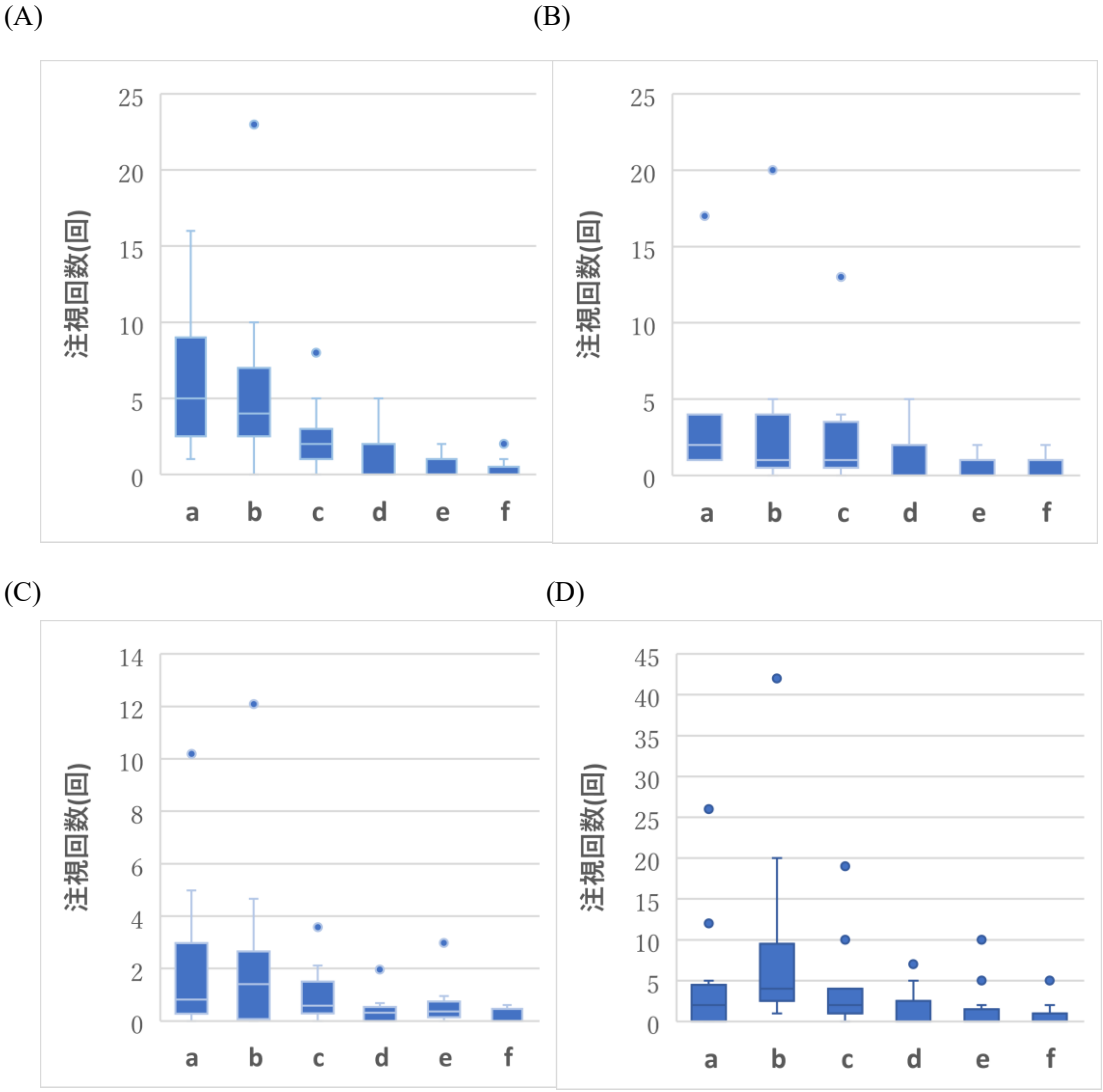
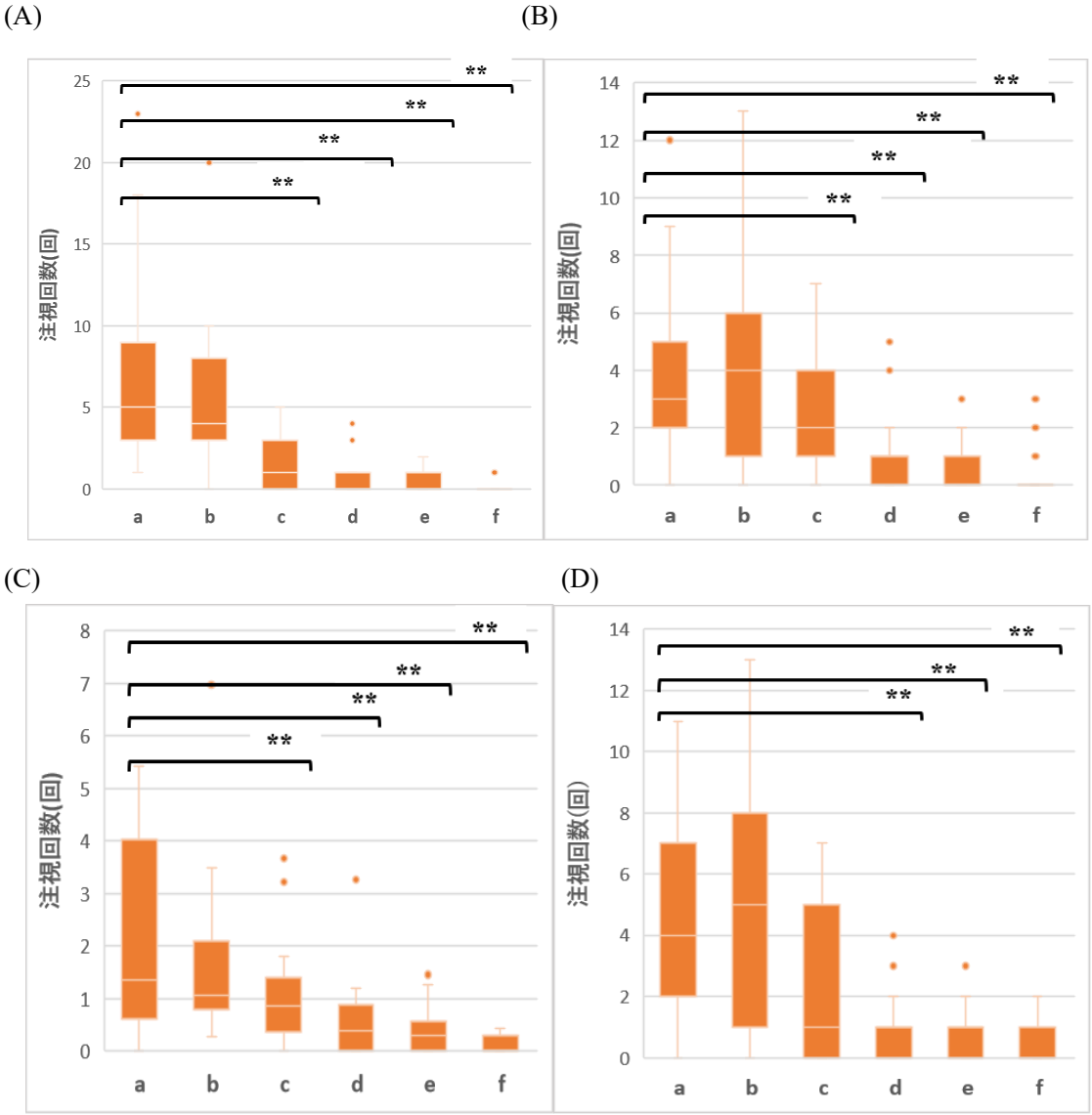


図 17: 口腔外科熟練者群の領域ごとの注視回数の比較

A:図 15-A、B:図 15-D、C:図 15-G、D:図 15-J をそれぞれ検診した際の口腔外科熟練者群の領域ごとの注視回数を比較した結果を示す。口腔外科熟練者群の領域ごとの注視回数を比較すると、領域 a と他領域との間に有意差がなく、病変を内外側問わず均等に診ている傾向を認めた。



**p<0.01,mean±SD, Bonferoni

図 18:一般歯科医群の領域ごとの注視回数の比較

A:図 15-A、B:図 15-D、C:図 15-G、D:図 15-J をそれぞれ検診した際の一般歯科医群の領域ごとの注視回数を比較した結果を示す。一般歯科医群の注視回数は領域 a と他領域の間に明確な差をつけ、診ている傾向を認めた。

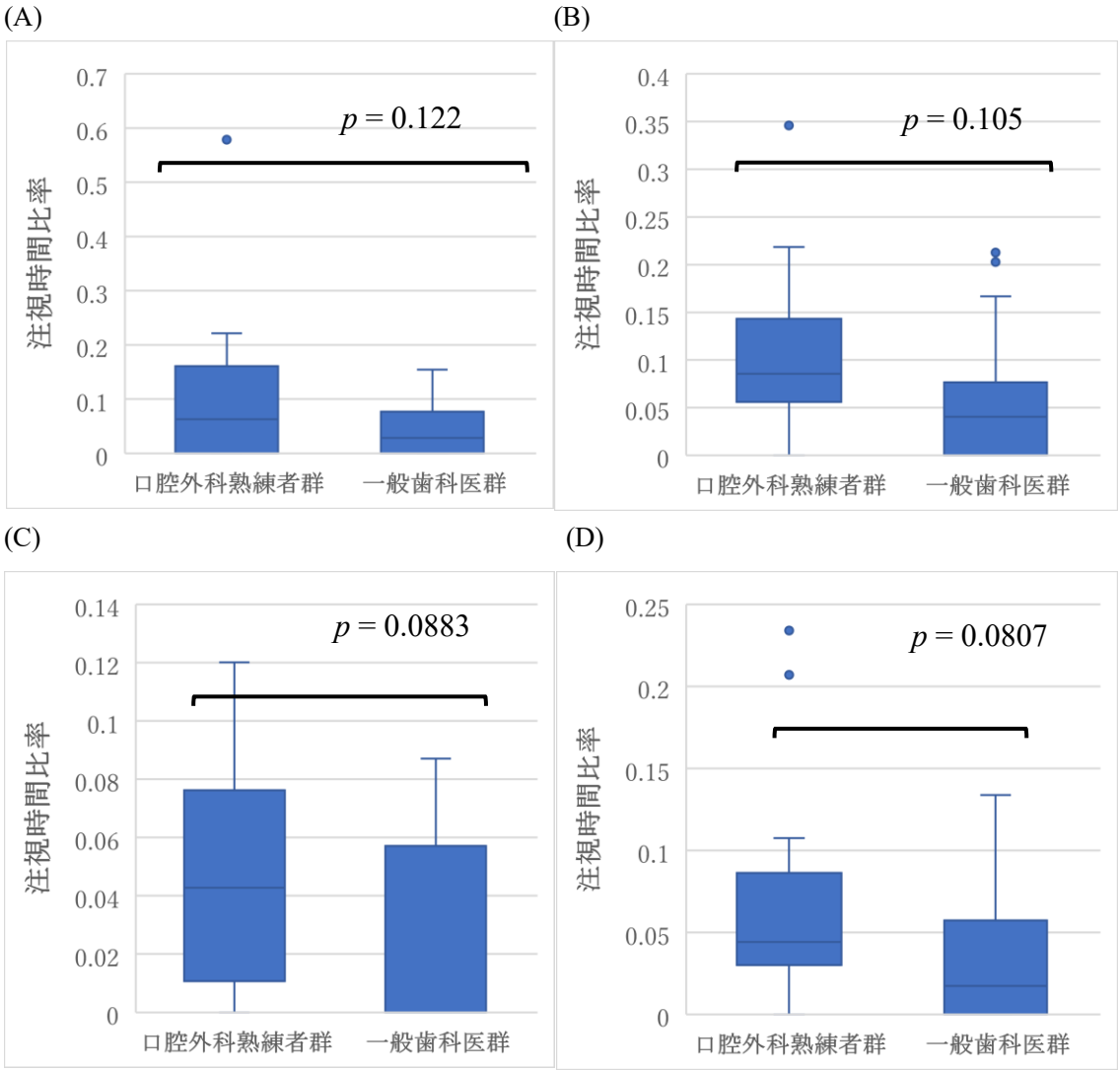
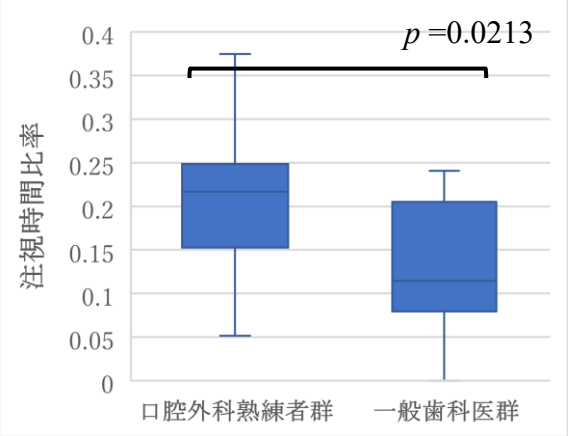


図 19：領域 d+e+f の注視時間比率の比較

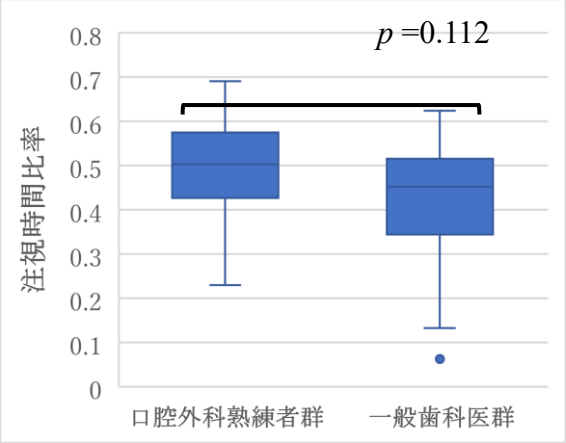
A:図 16-A、B:図 16-C、C:図 16-E、D:図 16-G の領域 d+e+f をそれぞれ検診した口腔外科熟練者群、一般歯科医群の注視時間比率（領域ごとの注視時間/全注視時間）を比較した結果を示す。口腔外科熟練者群が病変の内側辺縁～外側領域を診ている傾向を認めた。

(A)



領域 a 注視時間比率比較

(B)



領域 a+b+c+d 領域注視時間比率

図 20：病変内側領域の注視時間比率の比較

A:図 17-A の領域 a を検診した口腔外科熟練者群、一般歯科医群の注視時間比率（領域ごとの注視時間/全注視時間）を比較した結果を示す。B:図 17-C の領域 a+b+c+d を検診した口腔外科熟練者群、一般歯科医群の注視時間比率（領域ごとの注視時間/全注視時間）を比較した結果を示す。どちらも病変の内側領域の注視時間比率において、口腔外科熟練者群が病変の内側領域を診ている傾向を認めた。

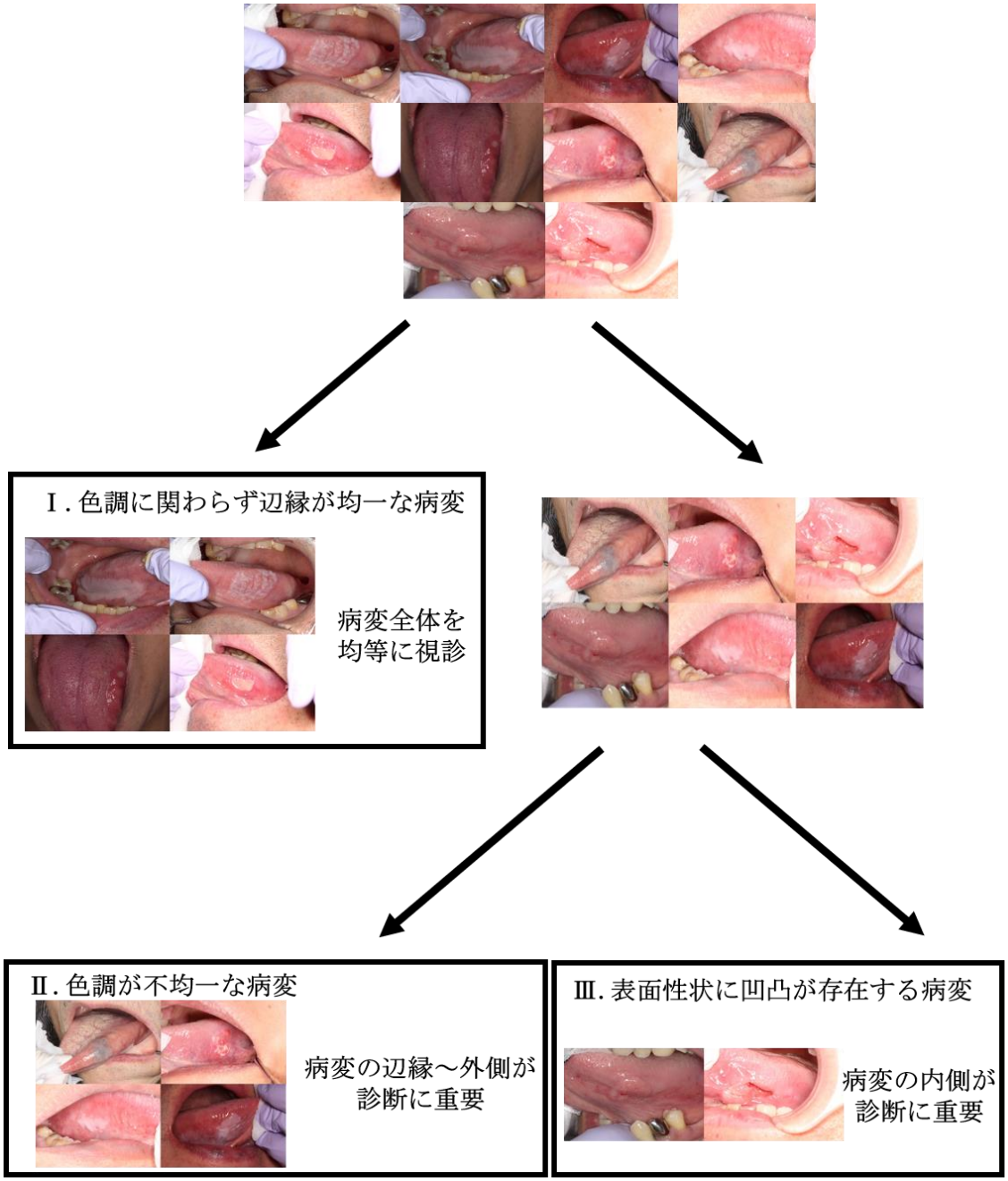


図 21：肉眼的特徴と検診方法、診断の契機となる領域の関係

検診方法が一致する病変は、肉眼的特徴が類似する傾向を認めた。