

# 光干渉断層計画像から緑内障の視野を推測する

## 3次元畳み込みニューラルネットワークモデルの構築

著者

古山 誠<sup>1</sup>

著者の所属機関

1: 南子安眼科

住所 〒299-1162 千葉県君津市南子安2-8-30、日本

責任著者

氏名 古山 誠

住所 〒299-1162 千葉県君津市南子安2-8-30、日本

電話番号 0439-27-1022

電子メールアドレス minamikoyasuganka@gmail.com

## 要約

**目的:** 光干渉断層計 (OCT) 画像から視野を推定し、その精度を向上させるため、症例数を増やす事が困難な従来の2次元モデルに代えて、セグメンテーション不要の3次元畳み込みニューラルネットワークモデル (3DCNN) により学習を行った。

**方法:** 単一施設の後向き全例調査である。OCT に加えハンフリー視野計(HFA) 24-2もしくは10-2検査を受けた、すべてのタイプの緑内障または緑内障の疑いの患者を含む連続3,416人 (6,356眼) を調査した。学習すべき視野及び視野の傾きの教師値は、視野閾値の測定点毎の回歸直線から算出した。学習対象と評価対象を分け、学習時に低 signal strength index (SSI) 及び眼内疾患合併の有無で4群に分けて3DCNN モデルによる5分割交差検証を施行し、同一条件の評価対象で最良の結果が得られた学習条件で10分割交差検証を施行した。

**結果:** 5分割交差検証の結果、全ての SSI 及び眼内疾患を含んで学習した群がいずれの対象に対しても最も推測性能が高く、10分割交差検証の結果、緑内障患者のペアデータあたりの二乗平均平方根誤差は、HFA24-2で $3.27 \pm 2.20$ dB [平均 ± 標準偏差]、HFA10-2で $3.02 \pm 2.28$ dB であった。

**結論:** 今回のモデルにより、OCT 画像からの高い視野推測精度を達成した。容易に症例数を増やす事ができるため、今後の多施設共同研究が期待される。

**臨床への橋渡し:** 本モデルは、OCT 画像から高い精度で視野を推定できるため、患者の負担を軽減することが可能である。

## 1 緒言

緑内障は非可逆的な網膜神経の障害を引き起こし、視野が欠損していく疾患<sup>1</sup>であり、世界の中途失明の原因の上位に位置する疾患である<sup>2</sup>。その進行により患者の生活の質は大きく低下する恐れ<sup>3</sup>があるが、緑内障に対して点眼や手術等の適切な治療を行うとその進行速度を遅くする事ができる<sup>1</sup>。緑内障の視野欠損の状況を過小評価すると治療の遅れに繋がり、逆に過大評価すると本来不要であるはずの治療を行う原因となりうるため、緑内障の状況を評価するために視野検査は必須である。視野検査とは、見える範囲がどの程度かを調べる検査の事であり、その測定は世界的に最も一般的な機種であるハンフリー視野計(HFA)によって行われる事が多い。HFA は Carl Zeiss Meditec の製品であり、視野を測定するための最も世界的に広く使用されている機種である。緑内障による視野欠損は、中心24度もしくは中心30度の鼻側周辺から出現する機会が多い<sup>4</sup>ため、一般的には HFA による中心24度(24-2)もしくは中心30度(30-2)の測定条件の視野検査が行われる事が多い。緑内障が進行すると、中心10度の視野欠損が出現する機会が多く<sup>4</sup>、また中心10度の視野欠損は生活の質の低下に直結する<sup>5</sup>ため、中心24度だけではなく中心10度の視野検査も大切であり、それは HFA による中心10度(10-2)条件での計測が可能である。視野欠損の評価を正確に行うために、HFA 24-2に加え10-2の視野検査を頻繁に行うと、視野検査は時間のかかる検査である<sup>6</sup>ため、患者側及び医療側の両者の負担が大きくなってしまう。その点、光干渉断層計(OCT)は視野検査よりも短い時間で測定でき、緑内障が進行するほど OCT で測定される網膜の内層側の厚みが薄くなり、両者には相関が得られる事が知られている<sup>7-10</sup>ため、OCT 画像から視野を推測する事は大切な事である。OCT の画像から直接視野を測定する事はできないが、多数の OCT 画像を入力値とし、その実際の視野を正解値として AI モデルに学習させる事で、OCT 画像からその学習済モデルを使用して視野を推測する事が可能であり、その報告は過去に多数<sup>7-16</sup>ある。近年、医療の分野では AI モデルの導入が目覚ましく、特に放射線科の画像診断の分野においては胸部レントゲンの異常所見を指摘する AI 等の臨床応用が既に実用化されている<sup>17</sup>。同様に、OCT 画像から視野を推測する優れた精度の AI モデルを作成して臨床応用する事ができれば、OCT 画像を元に学習済み AI モデルを使用して視野検査を比較的正確に推測する事で、視野検査を行う頻度を減らす事やより正確な視野の評価が可能になる可能性が考えられる。

網膜の構造変化は OCT 画像によって検出され、機能変化は HFA の値によって検出される。構造変化と機能変化は直線の関係ではなく、構造変化が生じてから機能変化が生じるまでには長い期間があり、また緑内障の末期になると機能変化に比べて構造変化はわずかであり<sup>18,19</sup>、それは floor effect と呼ばれる。さらにもともとの網膜厚は近視の度や眼球の形状による個人差が大きい<sup>20,21</sup>ため、AI によって OCT 画像から視野を推測する精度には限界があった。

もともとの OCT 画像は3次元であるが、3次元の OCT 画像をそのまま AI 学習するためには大量のコンピューターの演算能力及びメモリが必要であるため、従来は3次元の画像を網膜の層毎に

分離（セグメンテーション）して情報を2次元化し、2次元画像を元にして AI 学習を行う場合が多く、2次元 AI の中でも推測性能に優れた2次元 CNN を用いたモデルが多かった<sup>7-16</sup>。2次元 CNN モデル(2DCNN)の利点としては、コンピューターの演算能力をそれほど要求されない事及び、多くの既報の手法をそのまま利用できる事が挙げられ、逆にその一番の欠点としては、しばしばセグメンテーションエラーを起こす事が挙げられる。セグメンテーションエラーの有無を人間が全て確認するのはとても大変な作業になるため、黄斑上膜等の合併疾患を除外して、ノイズが少なく綺麗に撮影できた OCT のみで学習する必要がある、その結果2次元モデルではどうしても学習対象となる症例数が限られてしまう傾向にあった<sup>7-16</sup>。

機械学習の学習精度を上げるには、大量の症例数を準備する必要がある。もし3次元の OCT 画像をそのまま3次元の機械学習モデルで学習すれば、3次元画像をそのまま3次元モデルにて学習するため、次元を削減する工程であるセグメンテーションは不要である。黄斑上膜等の緑内障以外の合併疾患やノイズが乗った OCT 画像を2次元 AI モデルで学習する時にはセグメンテーションエラーが発生する恐れがあるが、3次元のモデルであればセグメンテーションが不要なため、緑内障以外の合併疾患も含んで学習する事も可能で、ノイズの多い OCT も全て学習対象とする事ができる。そのように考え、基礎疾患を含む緑内障症例も推測できるようにするため、3次元の機械学習モデルの中でも推測精度の優れた、3次元畳み込みニューラルネットワーク(3DCNN)を使用して、教師あり学習による機械学習モデルを作成した。

もしも、基礎疾患を含む症例やノイズの乗った OCT 画像も全て含んで学習した場合の3DCNN の推測性能が高ければ、基礎疾患の有無や OCT 画像のノイズの有無を人間がチェックする工程を減らす事ができるため、今後症例数を比較的容易に増やす事ができると考えられる。学習対象をどの程度まで広げるのが最も良い結果が出るのかを検証するため、緑内障以外の眼内疾患も全て含んで学習した群と、眼内疾患の合併症を除外して緑内障のみの症例で学習した群の2つに分け、結果を比較した。さらに、OCT 画像の信号強度指数 (SSI) が低い (=ノイズの多い) OCT 画像を含んで学習した群と、SSI が低い OCT を除外して学習した群の2つも分け、結果を比較した。

もし緑内障による視野欠損の進行速度を早期に推測できれば、進行速度が早い症例に対してより早期に手術や点眼等の適切な治療に介入する事により視野の進行速度を遅くする事ができ<sup>1</sup>、その結果将来の視野維持に役立つため、現在の視野だけではなく、同時に視野の進行速度の推測も大切である。視野の進行速度を推測するには、視野検査を複数回行い、Mean Deviation(MD 値)等の指標の回帰直線を作成して視野の指標の進行速度を推測する方法が臨床において一般的に用いられているが、その進行速度を判定するためには複数回の視野検査が必要となり、そのためには長い期間が必要である<sup>22</sup>。過去には、HFA の視野を元にして視野の進行速度を推測するベイズ法を利用した AI モデルの報告<sup>23</sup>はあるが、OCT 画像を元にして視野の進行速度を推測する AI モデルは我々が調べた限りでは見当たらない。もし単回の OCT 画像を元にして視野の進行速度が推

測できれば、従来よりも早期に進行速度を推測することができるため、進行速度が早い症例に対して早期に積極的な治療を行う事で視野を維持できる可能性が高くなる事が期待される。そのように考え、今回のモデルでは、OCT 測定時の視野だけではなく、視野の進行速度も計算し、それらを同時に教師値として3DCNN に与える事で、単回の OCT 画像を元にして視野と視野の進行速度を同時に学習した。

今回の論文の主な目的は2つである。1つ目は、3DCNN モデルを用い、合併疾患及び SSI の低い OCT 画像を含んで学習した群が、含まない条件で学習した群と比べ同等以上の HFA24-2及び10-2の視野の推測性能が得られるかどうかであり、2つ目は、最も高い推測性能が得られた学習条件で、3DCNN モデルを学習して解析する事である。

## 2 方法

### 2.1 研究方法

後ろ向きの単一クリニックの全例調査である。南子安眼科（当院）の院内に倫理委員会を設立し承認を行い(No.2022-002, 2022年3月19日)、研究情報を院内及びホームページに開示して拒否の権利を保証し、ヘルシンキ宣言に則り、人を対象とする医学系研究の倫理指針に準拠し、個人情報保護に関する法律についてのガイドライン Q&A2-15（利用目的による制限の例外）に準拠し実施した。当院は、緑内障及び緑内障以外の疾患の患者が来院する、入院施設及び手術設備のない、医師一人の普通の眼科診療所である。

### 2.2 症例選択

当院を開院してから現在まで（2011年3月8日～2022年8月4日）の単一施設の後ろ向き全例調査である。

OCTの黄斑マップに加え、HFA 24-2/30-2もしくは10-2を1回以上施行した症例を対象にした。オプトアウトを希望した症例はいなかった。OCT の黄斑マップの検査を受けた対象症例は、上記の対象期間の連続62,951テスト、7,641例、15,150眼であった。HFA24-2もしくは30-2の検査を受けた対象症例は、上記の対象期間の連続32,134テスト、3,884例、7,680眼であった。HFA10-2の検査を受けた対象症例は、上記の対象期間の連続11,884テスト、1,368例、2,603眼であった。視野検査を行った症例は、緑内障の有無に関わらず緑内障以外の疾患の患者も全て含めて研究対象とした。緑内障が進行すると視野の閾値が低下し、該当する部位の HFA の測定点の閾値が低下する。MD 値は視野全体を示す指標であり、緑内障が進行すると値が低下していく。今回の調査では、HFA の閾値及び MD 値を学習対象として用いた。

既報に合わせ<sup>8,14,16</sup>、偽陽性33%、偽陰性33%、固視不良33%をカットオフ値とし、カットオフ値以上の HFA は除外した。HFA30-2は周辺を削って HFA24-2のデータとして学習対象にした。

カットオフ値以上のHFAを除外した結果、対象はHFA24-2で25,268テスト、3,680例、6,864眼となり、HFA10-2で10,038テスト、1,320例、2,414眼となった。

## 2.3 診断

疾患の診断は全て1人の医師が行い、同一の医師がカルテを元に後ろ向きに症例を数えた。頭蓋内疾患例及びアーチファクトを含有した症例は3DCNNでの学習は困難と考え、それらを除外条件の疾患として表1に記載した。眼内疾患及び視神経疾患の症例（以下、眼内疾患含有と記載）は、それらの症例を学習対象に含めるかどうかで群を分けるために、緑内障を合併した場合も合併しなかった場合も全て含んで表2に記載した。何も疾患の無い正常症例に視野検査を行う事は無かったため、表1の症例を除外後の対象症例から表2の疾患を除外した症例は全て緑内障もしくは緑内障疑いの症例となった。眼底検査は全例Ocular社のOsher MaxField 78Dレンズを細隙灯顕微鏡で使用し、無散瞳での検査の症例も含まれる。緑内障の診断は、緑内障に特徴的な視神経乳頭所見（視神経乳頭陥凹拡大、視神経乳頭のノッチ、NFLD）及び、緑内障に特徴的なOCT所見及び、視野の所見から判定した。

いずれの群も学習時には緑内障・緑内障疑い・視神経乳頭陥凹拡大は区別せず、表1及び表2に記載の無い、白内障、後発白内障、強度近視、角膜混濁、点状表層角膜炎、虹彩炎、翼状片等の前眼部・中間透光体病変の合併例は除外せず全ての群で学習対象とした。いずれの群も、緑内障の病型の区別は行っておらず、隅角の所見による区分、原発性・続発性の区分、急性・慢性の区分、眼圧による区分はどれも行っていない。強度近視合併の緑内障及び後部ぶどう腫合併の緑内障の場合も除外せず、全ての群で学習対象及び評価対象とした。近視性視神経症による視野欠損は緑内障による視野欠損との鑑別が困難な場合が多いと考え、今回は両者の区別をせずに全て緑内障として学習対象及び評価対象とした。但し、強度近視による網脈絡膜萎縮もしくは網膜剥離等の合併疾患を認めた場合は、表2の疾患に区分した。

## 2.4 視野検査の結果の取得方法

当院では、HFAからの出力は全て株式会社ビーラインのBeeFiles for HFAへ保存しており、HFAデータの抽出を全例HFAから直接手作業で行うと極めて長い時間がかかるため、BeeFiles for HFAから間接的に行った。自作のマクロを作成して1症例ずつBeeFiles for HFAから視野のexportを繰り返し、textデータとして出力された全てのファイルを元に、Pythonを利用したプログラムを使用してデータを抽出し、HFA30-2・24-2ではパターン偏差・パターン標準偏差・緑内障半視野テストからAnderson-Patellaの基準に該当するかどうか自動的に計算し、データベースへ登録した。

## 2.5 OCT の 3 次元画像の取得方法

OCT は Nidek 社の RS-3000 もしくは RS-3000 Advance2 を使用した。2011 年 3 月 8 日～2022 年 8 月 4 日の対象期間の中で、2022 年 4 月 21 日～2022 年 5 月 12 日及び 2022 年 6 月 16 日～2022 年 8 月 4 日の期間は RS-3000 Advance2 を使用し、それ以外の期間は RS-3000 を使用した。撮影条件は、RS-3000 及び RS-3000 Advance2 の設定を統一し、黄斑マップの wide スキャン 9mm の X-Y 方向、幅 9mm、高さ 9mm、ScanPointA=512、ScanPointB=128、ScanCenterX=399、ScanCenterY=299、ScanWidth1=30.0、ScanWidth2=30.0 に設定し、他の撮影条件の画像は破棄した。3 次元黄斑マップ + 3 次元乳頭マップの両方を同時に学習対象とした方が良い結果が出る可能性も考えられたが、3 次元乳頭マップは当院での症例数が少ないため採用を断念し、3 次元黄斑マップのみを採用した。

普段の診療では 2 次元に圧縮された OCT 画像を利用する機会が多いが、実際の画像情報は 3 次元のまま内部に保存されている。推測性能向上のため、今回の 3DCNN モデルではその 3 次元 OCT 画像をそのまま学習に使用する必要があった。但しその保存場所は不明であり、OCT 内部のデータベースも非公開のため、患者 ID 及び撮影日時から 3 次元 OCT 画像の保存場所を特定するために以下の作業を行った。Tesseract-OCR を使用した Python プログラムを作成し、OCT から jpeg 形式でエクスポートされた画像を元に、撮影日時の部分を自動で切り取って OCR 解析してテキストデータへ変更し、同時に SSI や focus(近視の度に類似)の値も同様に読み取り、データベースに登録した。過去の jpeg エクスポート画像は OCT のソフトウェアのバージョンによってテキストの部位が異なるので、それぞれのソフトウェアのバージョンに対応した部位で自動的に切り取りが行えるようにプログラムを作成した。OCR 解析の際には、認識可能な文字を数字と目的とする記号のみに限定したために、誤った文字に変換される可能性は極めて低いと考えられた。但し、focus の小数点が抜けて OCR 解析される事があったため、focus に小数点が存在しない場合には小数点以下の有効桁数が 2 桁になる位置に自動的に小数点を追加した。解析された撮影日時の形式及び focus の形式及び SSI の形式を全て自動的に確認し、いずれも本来存在しない記号が無い事及び正しい桁数である事を自動的に確認した。撮影日時は日時の形式に正常に変換できる事を自動的に確認し、調査の期間以外の過去や未来の期間に誤変換されていない事を自動的に確認した。focus の値は 0.25D 刻みである事を自動的に確認し、全ての OCT 画像の focus の最小値及び最大値を取得し、得られた最小値及び最大値の画像が実際の OCT 画像の値と一致している事を確認した。その操作により、誤認識によるあり得ない大きすぎるもしくは小さすぎる focus の値は存在しない事が確認できた。SSI は 0 から 10 の中に収まっている事を自動的に確認した。3 次元の非圧縮 OCT 画像のディレクトリに存在する xml ファイルに記録されている撮影日時と、jpeg エクスポート画像の撮影日時が一致する物を自動的に探す事により、患者 ID 及び撮影日時データと 3 次元の非圧縮画像を過去に遡り全て紐付けしてデータベースに登録した。最後に、ランダム

に200枚の画像を選択し、それらが全て正しく OCR 認識されてデータベースに登録されている事を確認した。

## 2.6 OCT と HFA のペアデータの作成方法

今回の論文では、機械学習の中でも、教師あり学習を用いて学習を行った。教師あり学習の場合、ペアデータを作成し、入力値から教師値を推測するように学習する事が一般的である。今回は、機械学習モデルに与える値として OCT の画像、機械学習モデルから出力される値の教師値としての HFA をペアにして、ペアデータを作成した。

「単回の OCT の 3次元画像」を入力データとし、「HFA24-2・10-2のそれぞれの閾値と MD 値及びその傾き」を教師データとした。そのペアデータの作成方法は以下の通りである。

HFA24-2、10-2とも、緑内障の症例に対しては SITA-Standard 及び SITA-FAST が代表的な測定条件であり、SITA-FAST の方が SITA-Standard よりも測定時間が短い<sup>24</sup>ため、当院では患者側の負担を考え、緑内障患者の視野検査は HFA24-2, 10-2の SITA-Fast で行っていた。SITA-FAST と SITA-Standard は、既報では両者の値には強い正の相関性があり、また視野の進行判定能力にはほとんど差がないとの事<sup>25,26</sup>であるが、測定毎の測定点毎の閾値及び MD 値のばらつきが大きい<sup>25,26</sup>ため、そのままでは機械学習の教師値としては不適切と考え、眼毎に測定点毎の実測閾値の時間軸に対する通常の最小二乗法による回帰直線を作成し、OCT 撮影時の回帰直線からの値を視野閾値の教師値とした。但し、測定回数が少ない場合の回帰直線を延長するとばらつきが大きくなってしまったため、測定回数×6ヶ月間を有効期限とし、OCT 撮影時と直近の HFA 測定時の期間（絶対値）が有効期限を超えた場合には OCT データを破棄した。また、緑内障は非可逆的な疾患<sup>1</sup>のため、回帰直線の傾きが0以上になる事は無いはずであるため、回帰直線の傾きが0以上の場合には測定誤差と考えて傾きを0とした。視野閾値の教師値は0dBを最小値、35dBを最大値とした。MD 値も同様に回帰直線から計算し、MD 値に対しては最小値及び最大値の設定は行わなかった。

視野の進行速度に関して、MD 値の傾き（MD スロープ）は「スピアマンの順位相関係数から眼毎に計算した p 値<5%」の場合に最小二乗法による時間軸に対する回帰直線の傾きを採用した。視野の測定点毎の傾きは、単純に「測定点毎の p 値<5%」のみで採用すると偶然5%未満として採用された本来の傾きとは異なる大量の外れ値が出てしまう恐れがあるため、「MD 値の傾きの p 値<5%」且つ「測定点毎の p 値<5%」の場合に採用した。

尚、線維柱帯切除術・切開術を受けた症例や緑内障点眼治療を行っている症例も全て学習対象としたため、未治療での進行速度というのは不明で、未治療での自然経過よりも進行速度が遅く判定されるはずである。

MD 値の傾き及び視野の傾きの学習を SSI が低い OCT 画像も含めて行くと学習毎のばらつきが



大きくなる恐れがあるため、MD 値の傾き及び視野の傾きの学習は SSI 7 以上の OCT 画像に限定した。

閾値及び傾きには多くの欠損値があるため、その部位はマスク値を 1 とし、学習時の loss の計算の時に(1 - マスク値)を掛ける事により学習に影響を与えないようにし、検証時にも計算から除外した。

OCT と HFA のペアデータ作成の結果、HFA24-2と OCT のペアデータで41,544ペア、3,533例、6,585眼となり、HFA10-2と OCT のペアデータで22,410ペア、1,291例、2,365眼となった。

## 2.7 中枢疾患及びアーチファクトの除外

中枢の合併疾患及び、アーチファクトも学習対象とすると、本来の目的である緑内障の推測精度に悪影響が出てしまう恐れがあるため、表 1 の症例は全ての群の学習対象から除外した。除外の結果、対象は HFA24-2と OCT のペアデータで39,722ペア、3,324例、6,153眼となり、HFA10-2と OCT のペアデータで21,828ペア、1,241例、2,267眼となった。

## 2.8 HFA24-2と10-2の同一 OCT ペアデータへの混合

同一の OCT 画像に対して HFA24-2と10-2のデータがそれぞれ存在した場合、単一のペアデータとなるように混合した。即ち、HFA24-2の52点及び、HFA10-2の68点の閾値を、一つのペアデータとして統合した。HFA24-2もしくは10-2のいずれかが欠損していた場合には該当する値をマスク値とした。統合の結果、41,420ペア、3,416例、6,356眼となった。

## 2.9 学習時に眼内疾患を含めたかどうか

もしも、緑内障以外の眼内疾患も含んで学習した場合の推測性能が、緑内障以外の眼内疾患を除外して学習した場合の推測性能と同等以上であれば、眼内疾患を除外せずに学習する事ができるため、学習時に眼内疾患の有無のチェックの手間が省け、症例数を今後集めやすくなると期待される。緑内障以外の症例も全て含んで学習した方が推測性能の向上が認められるかどうかを確認するため、表 1 の疾患（主に中枢疾患か、もしくは視野が正しく計測できなかった症例）を除外した全ての症例を対象に、学習時の条件により以下の 2 群に分けた。眼内疾患含有群 は、学習時に表 2 の疾患（主に眼内・視神経疾患）を含め、緑内障の有無に関わらず緑内障以外の疾患も含めて全ての HFA と全ての OCT の 3 次元黄斑マップ画像を元に機械学習を行った。眼内疾患除外群 は、学習時に表 2 の疾患を除外し、緑内障・緑内障疑い・視神経乳頭陥凹拡大以外の疾患を合併している症例を全て除外して学習した。

眼内疾患除外群は、HFA24-2/10-2と OCT のペアで38,368ペア、3,274例、5,961眼（内、HFA24-2と OCT のペアで36,824ペア、3,185例、5,778眼、HFA10-2と OCT のペアで19,942

ペア、1,171例、2,079眼) となった。

## 2.10 学習時に低 SSI を含めたかどうか

ノイズ混じりの OCT から学習した方が良い推測性能が得られるかどうか確認するため、全ての SSI を含んで学習した群 (全 SSI 群) と、SSI が 7 未満の症例を除外して学習した群 (SSI 制限群) に前述の 2 群をさらにそれぞれ 2 つに分けてモデルを作成し、最後に評価対象を同一に揃え比較を行った。SSI が極めて低い症例の緑内障の診断は、白内障や後発白内障が進行して SSI が低くなった場合がほとんどであるため、その症例の眼底が見えた過去の時点の所見、もしくは手術後に眼底が見えるようになった後の所見を元に診断を行った。初診時から終診時まで一貫して白内障が強く眼底が見えない症例は、もともと視野検査を施行していないため、今回の研究対象には含まれていない。

眼内疾患含有・SSI 制限群は、HFA24-2/10-2 と OCT のペアで 36,066 ペア、3,286 例、6,011 眼 (内、HFA24-2 と OCT のペアで 34,661 ペア、3,201 例、5,818 眼、HFA10-2 と OCT のペアで 18,884 ペア、1,208 例、2,177 眼) となった。眼内疾患除外・SSI 制限群は、HFA24-2/10-2 と OCT のペアで 33,660 ペア、3,156 例、5,658 眼 (内、HFA24-2 と OCT のペアで 32,377 ペア、3,073 例、5,484 眼、HFA10-2 と OCT のペアで 17,403 ペア、1,142 例、2,001 眼) となった。

## 2.11 Anderson-Patella の基準外を HFA24-2 の評価対象から除外

当院では、検診で視神経乳頭陥凹拡大を指摘された方に OCT 及び HFA24-2 を行う事も多かったため、その場合の多くは正常な視野であり、モデルからの推測が比較的容易であると推測される。そのような症例も全て含んで評価を行うと、本来の推測性能よりも良い結果が出てしまう恐れがあるため、学習時にはいずれの群も Anderson-Patella の基準 (パターン偏差で  $p < 5\%$  の点が 3 つ以上隣接しそのうち 1 点以上が  $p < 1\%$ 、パターン標準偏差  $< 5\%$ 、緑内障半視野テストが正常範囲外のいずれか)<sup>27</sup> に合致しない症例も全て対象に含めたが、評価時には眼毎に最終 HFA24-2 が Anderson-Patella の基準に合致する HFA24-2 のみを評価した。緑内障の発症初期の場合、Anderson-Patella の基準に後から合致する症例もあるため、そのような症例は途中経過も含め評価対象に含めた方が良くと考え、途中の視野が例え Anderson-Patella の基準に該当していない場合でも、最終 HFA24-2 が Anderson-Patella の基準に該当している場合には眼毎に過去に遡って全て評価対象とした。

尚、当院での HFA10-2 は、OCT 所見から見て中心視野が欠損している可能性が考えられる症例に限り行っており、また HFA24-2 が Anderson-Patella の基準に該当していない場合でも HFA10-2 の中心視野欠損を来す症例も存在するため、HFA10-2 は HFA24-2 の Anderson-Patella の基準とは無関係に評価した。

Anderson-Patella の基準外を上述の方法で除外した結果、HFA24-2と OCT の評価時のペアデータは、眼内疾患含有・全 SSI 群で27,712ペア、2,407例、3,569眼、眼内疾患除外・全 SSI 群で25,090ペア、2,234例、3,220眼、眼内疾患含有・SSI<7除外群で23,678ペア、2,290例、3,340眼、眼内疾患除外・SSI<7除外群で21,633ペア、2,130例、3,030眼となった。

## 2.12 視野の進行速度の補正方法

以下の2種類の進行速度の補正操作を、予め学習前に全てのペアデータに関して行った。

### 2.12.1 正の傾きの進行速度の補正方法

1種類目の進行速度の補正操作は正の傾きの補正に関連するものである。

眼毎の真の視野の進行速度を測定する事は不可能であるが、視野検査を繰り返して回帰直線を作成する事で真の視野の進行速度に近い値を得る事ができる可能性が高くなるはずである。その計算上の回帰直線の傾きは、測定誤差により真の視野の進行速度よりも大きくなる事もあれば小さくなる事もあると考えられる。真の視野の進行速度がもともと遅い場合、回帰直線の傾きが正の方向にずれた場合には計算上の傾きが0より大きくなる事がしばしばある。本来、緑内障による視野欠損は非可逆的な変化<sup>1</sup>のため、傾きが正の値になる事は無いはずである。傾きが正の場合に単純に傾きを0に置き換えた場合、正の傾きの測定誤差のみを補正した事になり、負の傾きの測定誤差は判別不能なため、負の傾きの測定誤差によって実際の臨床よりも進行速度の傾きが負の方向に大きくなる恐れがある。その解決策として、HFA24-2及び HFA10-2のそれぞれに関し、「MD 値の傾き」及び「視野の測定点毎の傾き」の両者とも、時間軸に対する最小二乗法による回帰直線の傾きが正の場合に傾きを0と置き換えた上で、「傾きが正の場合も受け入れた時の傾きの平均」と、「傾きが正の場合に傾きを0と置き換えた時の傾きの平均」の比を全てのペアデータに対して掛ける事により、誤差の補正を行った。正の傾きの補正に関連する進行速度の最終的な補正值の比率は、HFA24-2の視野の傾きは0.883倍、HFA24-2の MD 値の傾きは0.822倍、HFA10-2の視野の傾きは0.933倍、HFA10-2の MD 値の傾きは0.881倍となった。

### 2.12.2 SSI 低下に関連する進行速度の補正方法

2種類目の進行速度の補正操作は SSI 低下に関連するものである。

例えば中には患者が手術を希望せずに白内障や後発白内障が進行する事により見かけ上の視野欠損が進行してしまう症例もあり、そのような症例を含んでそのまま学習すると、本来の緑内障による進行速度よりも早い進行速度で学習してしまう恐れがある。その問題を解決するため、以下の通りに HFA24-2及び HFA10-2の進行速度の補正を行った。一般的には、白内障や後発白内障が進行すると眼内へ光が透過し難くなるために OCT の SSI の数値が低下<sup>28,29</sup>し、また OCT の SSI

の数値が低下すると MD 値や視野閾値の低下が認められる (図3~図6)。年齢と共に SSI は低下する傾向にあるため (図7~図8)、SSI が有意に減少した場合との差分を計算し、その差分を元にして進行速度の補正を行った。その具体的な計算方法は以下の通りである。今回の研究での表2の疾患を除外した全てのペアデータに関し、「SSI に対する MD 値(dB/SSI)」の最小二乗法による回帰直線の傾き及び、「SSI に対する平均視野(dB/SSI)」の最小二乗法による回帰直線の傾きを算出する事により、SSI が1低下する度に平均どの程度 MD 値及び平均視野が悪化するのかを計算した。次に、加齢により SSI の数値は低下するため、表2の疾患を除外した全てのペアデータに関し眼毎に「年齢に対する SSI (SSI/year)」の最小二乗法による回帰直線の傾きを算出する事により、年齢が1歳増える度に平均で SSI がどの程度減少するのかを計算した。その上で、眼毎に「時間軸に対する SSI (SSI/year)」の最小二乗法による回帰直線の傾きの「時間軸に対する SSI のスピアマンの順位相関係数から計算した p 値<5%」を有意と判定し、有意な場合にはその傾き (SSI/year)と上記の「全てのペアデータの年齢に対する SSI (SSI/year)」を比較し、「全てのペアデータの年齢に対する SSI (SSI/year)」よりもその傾き (SSI/year)の絶対値が大きい場合には、「眼毎の時間軸に対する SSI の傾き (SSI/year)」から「全てのペアデータの年齢に対する SSI (SSI/year)」を引く事により平均的な SSI の低下速度との差分を算出し、その値に「SSI に対する MD 値(dB/SSI)」もしくは「SSI に対する平均視野(dB/SSI)」の値を掛ける事により、SSI の減少による分の「MD 値の傾き(dB/year)」もしくは SSI の減少による分の「視野の傾き(dB/year)」が得られ、次にその値を元の「MD 値の傾き(dB/year)」もしくは元の「測定点毎の視野の傾き(dB/year)」に対し減算する事により、SSI の減少による見かけ上の傾きへの影響を打ち消した。その結果、例えば白内障や後発白内障が進行した事による見かけ上の視野欠損進行をある程度打ち消した進行速度が算出されるため、補正前と比べてその分遅い進行速度となり、最も知りたい緑内障の進行による視野欠損の速度により近いものになるはずである。尚、傾きが0よりも大きくなる場合には0とし、「時間軸に対する SSI のスピアマンの順位相関係数から計算した p 値<5%」が得られない場合には、SSI が低下しているかどうか不明であるため上述の補正を行っていない。そのため、SSI 減少による進行速度への影響を完全に打ち消す事はできず、補正を行った眼と補正を行っていない眼の両方がペアデータに存在する事になる。SSI 低下に関連する進行速度の補正操作による全体への影響を計算すると、最終的な補正值の平均値の比率は、HFA24-2の視野の傾きは0.974倍、HFA24-2の MD 値の傾きは0.973倍、HFA10-2の視野の傾きは0.979倍、HFA10-2の MD 値の傾きは0.970倍となった。

尚、視野の測定点毎の傾きはばらつきが大きいため、上記の2種類の進行速度の補正とも、最終的な補正の比率は全ての測定点の補正の比率の平均値から算出した。表3, 表4に記載の値は上記の2種類の進行速度の補正後の値である。

### 2.13 データセットの振り分け方法

全てを10分割交差検証で学習するためには極めて長い時間が必要なため、今回は5分割交差検証及び10分割交差検証の両方を用いた。5分割交差検証は複数のモデルの推測性能を比較するためのみに使用し、複数の5分割交差検証の中で最も良い結果が得られた1つのモデルの推測性能を厳密に調べるため、最良の学習条件に対し10分割交差検証を用いて再度モデルを作成し学習を行った。患者毎に全症例をランダムに等分し、5分割交差検証では学習群：検証群兼テスト群を4：1、10分割交差検証では学習群：検証群：テスト群を8：1：1に振り分けた。モデル間の比較のため、ランダムに等分する際に使用する乱数のシード値は各モデルで共通の値を採用した。

患者によってペアデータ数が多い症例もあれば、逆に少ない症例もあり、単純に全てランダムに患者毎に振り分けた場合には分割交差検証の群間でのペアデータ数が偏り、学習の妨げとなる可能性があるため、以下のように振り分けた。まず、患者毎にHFA24-2、10-2のそれぞれの時系列に対するMD値の「スピアマンの順位相関係数から計算したp値<5%」が左右眼の少なくともどちらかで該当するか、それとも両眼とも該当しないかで患者を分けた。即ち、「少なくとも片眼のHFA24-2のMD値のp値<5%」を満たす群と満たさない群の2群に分割し、さらにそれぞれ「少なくとも片眼のHFA10-2のMD値のp値<5%」を満たす群と満たさない群に分割し、合計4群に分割した。それぞれの分割した4群に対し、ランダムに5分割もしくは10分割し、最終的に5分割もしくは10分割毎にその4群からの患者をそれぞれ混合して加える事により、均等に5分割交差検証用もしくは10分割交差検証用に患者を全て振り分けた。

表1及び表2の症例の除外は患者毎ではなく眼毎であるため、5分割交差検証用もしくは10分割交差検証用の患者の振り分けを行った後に、眼毎に眼内疾患及び眼外疾患の除外を行った。その際、両眼とも除外される場合もあるため、最終的な5分割交差検証用もしくは10分割交差検証用の患者数は各群で必ずしも一致せず、また患者によってペアデータ数が異なるために各群でペアデータ数も一致しないが、上述の方法により、完全にランダムに振り分けた場合よりも均等に近くなる可能性が高いはずである。

学習群ではOCT画像から教師データを推測するように学習し、テスト群ではOCT画像からの推測データと正解データから精度を計算し、5分割交差検証では5回群を順番にずらして、10分割交差検証では10回群を順番にずらして最終結果を算出した。その際、それぞれの交差検証にて保存していた検証群での最も良いモデルを読み込んでテスト群から推測し判定した。5分割交差検証では早期終了を使用しないで検証群無しで学習しても良いが、その場合には学習のばらつきが大きくなり、学習条件を変えた時の比較が困難になってしまうため、敢えて検証群とテスト群を兼用とした。その結果、実際よりも良い結果が出てしまう恐れがあるため、最終的なモデルの精度を算出する時には、最良の値が得られた同一学習条件にて10分割交差検証のモデルを再度作成

し直して検証群とテスト群を分けて検証した。

MD 値及び傾きを加えて同時に学習すると、学習毎のばらつきが大きくなってしまい、どの学習条件が最適かの推測が困難になってしまうため、5分割交差検証では HFA24-2・10-2 閾値のみを学習対象とし、その後の10分割交差検証では HFA24-2・10-2の MD 値及び傾きも学習対象とした。

尚、上記の5分割交差検証及び10分割交差検証の振り分けはいずれも患者毎であるため、それぞれの群に同時に同一の患者が存在する事は無い。

#### 2.14 視野の教師値の処理方法

視野の閾値の教師値は0dB から35dB と定義したため、35で割る事により正規化して0から1の範囲内に収める事が可能であるが、視野の進行速度の絶対値上限は母集団によって異なるため、0から1の範囲内への正規化ではなく、平均値0、標準偏差1となる標準化が望ましいと考えた。1つのモデルで同時に学習する際、同時に標準化と正規化の教師値が混在すると、スケールの違いにより誤差逆伝播法時の学習の妨げとなる恐れがあるため、教師値は全て標準化で統一した。即ち、HFA10-2・24-2のそれぞれの学習群に対し、HFA の閾値、MD 値、HFA の閾値の傾き、MD スロープの標準偏差及び平均値を計算して標準化を行った。その際、検証群及びテスト群の教師値はモデルからは不明であるため、10分割交差検証毎もしくは5分割交差検証毎に、学習群・検証群・テスト群とも、学習群から算出した平均値及び標準偏差を使用して標準化の計算を行った。そのため、検証群及びテスト群に関しては平均値0、標準偏差1とはならないが、それに近い値となるはずである。最終結果を算出する時にはいずれも逆の計算を行い、元の数値に戻した。

#### 2.15 OCTデータの処理方法、モデルの学習方法

今回のモデルには、PyTorch の3DCNN モデルの1つである EfficientNet3D-b0<sup>30</sup> 2021年7月20日版を使用した。EfficientNet<sup>31</sup>は2次元の CNN の中でも特に高い推測性能と高い学習効率が特徴のモデルで、今回使用した EfficientNet3D はその3次元版である。EfficientNet3D にも様々な大きさのモデルがあり、学習効率の面から今回はその中で最も小さい EfficientNet3D-b0を使用した。

元々512 × 512 × 128 pixel × 8bit グレースケールの3次元 OCT 画像を、224 × 224 × 128 pixel × 8bit グレースケール画像へ Lanczos 法で変換し、その値が0から1の範囲に収まるように256で割る事により正規化して EfficientNet3D-b0へ入力した。学習毎に画像を変換すると学習時間が極めて長くなるため、予め変換した Tensor イメージを local の高速なストレージに保存して読み出すようにした。過学習を抑えるため、EfficientNet3D-b0の head 部分を入れ替え、モデル内部の出力直前の1280次元からの元の dropout20%に加え、dropout をさらに30%追加した。

dropout の後は隠れ層無しの linear の全層結合で出力した。モデルからの出力は、10分割交差検証では244次元の「HFA24-2・10-2のそれぞれの、閾値及び傾き及び、MD 値及び傾き」であり、5分割交差検証では120次元の「HFA24-2・10-2のそれぞれの閾値」である。モデルからの出力と教師値の平均二乗誤差を最小にするようにモデルの学習を行った。ミニバッチサイズは4、学習率は $6e-4$ から $1e-3$ まで3epoch かけて漸増して、その後5epoch かけて $6e-4$ まで漸減した。Optimizer は Adam<sup>32</sup>を使用し、Weight Decay は非使用である。転移学習及びファインチューニングは EfficientNet3D-b0の学習済みデータが見当たらなかったため使用していない。それらのハイパーパラメータの決定は以下の方法で行った。

### 2.15.1 事前のハイパーパラメータの調整

機械学習の推測性能を向上させるためには、予め十分な時間をかけてハイパーパラメータを調整しておく事は大切である。ハイパーパラメータとは、学習プロセスを制御するために使用される値の事である。

中でも、適切な学習率の値の設定は学習を行う上で非常に大切であり、学習率が大きすぎても小さすぎても最適解から逸脱する危険性が高くなると考えられている<sup>33</sup>。学習率とは、勾配法によってどの程度大きくパラメータの更新を行うかを決定するハイパーパラメータ値の事である。予め、様々な学習率での学習を繰り返し、高い推測性能が得られる最適な学習率の値の範囲を調査した。最初から大きな学習率で学習してしまうと局所的な最適解に陥ってしまい推測性能が伸び悩む危険性があるため、小さい学習率から徐々に学習率を上げていく warmup という手法を用いる事により、局所的な最適解に陥る危険性が低くなる事が知られている<sup>34</sup>。

また、ミニバッチサイズを小さくするほど推測性能が向上する傾向がある反面、学習時間が長くなり、学習のばらつきが大きくなり、局所的な最適解に陥ってしまい推測性能が伸び悩む危険性もある<sup>35</sup>。ミニバッチサイズとは、データセットをサブセット毎に区切って学習する際のサブセットの大きさの事であり、ハイパーパラメータの1つである。

可能な限り推測性能を向上させるため、今回のモデルでは warmup を使用し、局所的な最適解に陥らない最小のミニバッチサイズを、学習率・warmup・ミニバッチサイズの条件を変えて学習して調査した。

また、warmup 後に一定の学習率ではなく学習率を徐々に減らしていく事により最終的な推測性能が向上する可能性が高くなると考えられている<sup>33</sup>。繰り返し学習する事で、最適と考えられる学習率及び学習率の減少率を調査した。

また、過学習を防ぐため、EfficientNet3D の head 部分を入れ替え、追加した Dropout の割合を変更して学習し、最も良い推測性能が得られた値を予め調査した。Dropout とは、学習時に、一定割合のノードを不活性化させながら学習を行う事で過学習を緩和し、推測性能を向上させる

手法の事であり<sup>36</sup>、dropout の割合はハイパーパラメータの1つである。

Optimizer は Adam<sup>32</sup>以外にも、AdamW<sup>37</sup>、AdaBelief<sup>38</sup>、SGD with momentum<sup>39</sup>、SAM<sup>40</sup> + SGD with momentum、ASAM<sup>41</sup> + SGD with momentum を試し、予めそれぞれ学習し、上述のハイパーパラメータの手作業での微調整をそれぞれの optimizer に関して繰り返した。Optimizer 毎に特有の複数のハイパーパラメータを含む事もあり、それらを含め、最適となる推測性能が得られるようにそれぞれのハイパーパラメータの組み合わせを手作業で試した。ハイパーパラメータの探索の際に使用した対象群は、眼内疾患を除外し、SSI を制限した群で、5分割交差検証と同じ方法で作成した1番目の学習群とテスト群を使用して学習を行った。最も良い結果が得られたハイパーパラメータ設定を、5分割交差検証の2番目の学習群とテスト群を使用して、再度繰り返し値をずらして学習し、手作業で微調整を行った。その結果、最終的に最も良い推測性能が得られたハイパーパラメータの値を決定し、前項に記載の通り採用した。検証群毎にハイパーパラメータを変化させる方法は極めて長い学習時間がかかる恐れがあるため採用せず、今回の論文では全てのモデルで統一したハイパーパラメータの値を使用した。

## 2.16 水増し・反転方法

効率的に学習するため、左眼の OCT データは水平反転し、擬似的に右眼のデータとして学習し、それに対応した教師値も水平反転を行った。学習データを2倍に水増しするため、OCT データは垂直反転を使用し、それに対応した教師値も垂直反転を行った。垂直反転は学習時のみではなく検証時にも使い、垂直反転した値と垂直反転していない値の平均的を取り最終出力とした。表3、表4は水増し前の値である。

## 2.17 評価対象

評価対象は、学習対象と同じ基準で症例選択を行ったが、HFA24-2の評価は眼毎の最終 HFA 24-2が Anderson-Patella の基準に該当した場合にのみ対象とした。

解析の母集団が異なるとモデルの精度比較が困難になるため、学習対象と評価方法は切り分けて考え、それぞれの学習対象で学習したモデルに対し、最後に同一条件のテストデータで評価した。即ち、眼内疾患含有群と眼内疾患除外群は学習対象が異なるが、精度比較のため評価対象を同一に揃え、例えば両群とも「緑内障以外の合併疾患症例を除外した同一群に対して評価」のように行った。評価対象を変えた時にも単回の学習で対応できるように、モデルからの出力時に全ての症例からの推測を保存し、後で評価対象の条件に応じて除外するという方法を取った。

## 2.18 計算方法

テスト群からの推測データと正解データから、ペアデータ毎に測定点毎の二乗平均平方根誤差



(Root Mean Square Error; RMSE) を算出した。既報では先に平方根を取りその後全ての症例の平均値を計算している報告<sup>8,11,14</sup>が多いため、今回は既報に合わせ、先に平方根を取り計算した。つまり、ペアデータ毎に

$$RMSE = \sqrt{\frac{1}{n} \sum_{d=1}^n (\hat{Y}_d - Y_d)^2} \quad \text{とした。}$$

上記の $\hat{Y}$ は推測データ、 $Y$ は正解データ、 $n$ は測定点の総数、 $d$ は測定点の番号である。HFA24-2の閾値及び閾値の傾きの場合には  $n=52$ 、HFA10-2の閾値及び閾値の傾きの場合には  $n=68$ 、HFA24-2・HFA10-2のMD値及びMD値の傾きの場合には  $n=1$ となる。

RMSE同様、テスト群からの推測データと正解データからペアデータ毎に平均絶対誤差 (Mean Absolute Error; MAE) を算出した。閾値及び閾値の傾きの MAE は2種類計算し、測定点毎の MAE と、視野全体の平均の MAE のそれぞれを計算した。つまり、ペアデータ毎に

$$\text{測定点毎の MAE} = \frac{1}{n} \sum_{d=1}^n |\hat{Y}_d - Y_d| \quad \text{とし、}$$

$$\text{視野全体の平均の MAE} = \frac{1}{n} \left| \sum_{d=1}^n \hat{Y}_d - \sum_{d=1}^n Y_d \right| \quad \text{とした。}$$

MD値及びMD値の傾きの場合、もともとMD値が視野全体の指標のため、上記のどちらで計算しても  $n=1$ から同一の MAE となる。

計算は全て Python 3.8.9で行った。スピアマンの順位相関係数、ピアソンの積率相関係数は Python の SciPy 1.7.3を使用し、RMSE 及び MAE は NumPy 1.22.0を使用した。SciPy 及び NumPy は世界的に最も広く使用されている数値解析及び数値計算のパッケージである。

## 3 結果

### 3.1 症例選択

HFA24-2、10-2とも、偽陽性33%、偽陰性33%、固視不良33%をカットオフ値とし、33%のカットオフ値以上のテストを棄却した。HFA24-2、10-2とも、OCTの黄斑マップ画像とのペアデータを作成し、表1の合併疾患の眼を棄却し、HFA24-2のペアデータとHFA10-2のペアデータを同一のOCT画像の条件で混合した。表2の合併疾患の有無及びSSI<7の有無によって学習対象を分け、最終のHFA24-2がAnderson-Patellaの基準に該当しない眼をそれぞれテスト対象のHFA24-2から除外した。症例選択のフローチャートを図2に示し、その結果得られたそれぞれの母集団の詳細を表3及び表4に示した。表3はHFA24-2/10-2とOCTのペアデータであり、HFA24-2は学習時、HFA10-2は学習時及び評価時の母集団である。表4は最終のHFA24-2がAnderson-Patellaの基準に該当しない眼をそれぞれテスト対象のHFA24-2から除外した後の表で

あり、HFA24-2と OCT のペアデータを示し、HFA24-2の評価時の母集団である。

眼内疾患含有・全 SSI 群の学習時の母集団は、HFA24-2/10-2 は 41,420 ペア、3,416 例、6,356 眼となり、その平均閾値は、HFA24-2 にて  $26.5 \pm 6.91$  dB、HFA10-2 にて  $27.8 \pm 8.30$  dB であった。

眼内疾患除外・SSI < 7 除外群の評価時の母集団は、HFA10-2 は 17,403 ペア、1,142 例、2,001 眼、教師値の平均閾値は  $28.4 \pm 8.02$  dB であり、最終の HFA24-2 が Anderson-Patella の基準に該当しない眼を除外した後の HFA24-2 は 21,633 ペア、2,130 例、3,030 眼、教師値の平均閾値は  $25.6 \pm 7.42$  dB であった。

### 3.2 OCT の SSI 低下に関連する要素

OCT の SSI 低下による見かけ上の視野の傾きの悪化を補正するための基礎データを取得するため、表 2 の疾患を除外し、全ての SSI を含んだ全てのペアデータに対し、以下の通り SSI に関連する項目を調べた。

#### 3.2.1 OCT の SSI と実際の視野の関係

白内障等により OCT 画像にノイズが多くなるほど、実際の視野が悪くなるかどうかを調べるため、表 2 の疾患を除外した全てのペアデータに対し、OCT の SSI と視野の関係を調べ、図 3 から図 6 に示した。図 3、図 5 は HFA24-2 を、図 4、図 6 は HFA10-2 を対象とし、図 3、図 4 は視野の閾値の平均値を対象とし、図 5、図 6 は MD 値を対象とした図である。図の箱はデータの第一四分位値 (Q1) から第三四分位値 (Q3) まで伸び、中央値に線が引かれている。ひげは、Q1 と Q3 のそれぞれ上下に箱から Q1 から Q3 の範囲の 1.5 倍だけ伸びている。点は、ひげの終点を超えた値を示している。以下の図でも同様の表記である。

OCT の SSI と実際の視野の間に有意な正の相関が得られた (HFA24-2 の平均閾値のスピアマンの順位相関係数は  $p < 0.001$ 、 $\rho = 0.220$ 、(図 3)、HFA10-2 の平均閾値のスピアマンの順位相関係数は  $p < 0.001$ 、 $\rho = 0.230$  (図 4)、HFA24-2 の MD 値のスピアマンの順位相関係数は  $p < 0.001$ 、 $\rho = 0.166$ 、(図 5)、HFA10-2 の MD 値スピアマンの順位相関係数は  $p < 0.001$ 、 $\rho = 0.188$  (図 6))。

従って、OCT 画像のノイズが多い、つまり SSI が低いほど、実際の視野が悪化する傾向が確認でき、得られた傾きを 2.12.3 章で述べた SSI 低下による視野の傾きの補正式に用いた。

#### 3.2.2 年齢と OCT の SSI の関係

年齢と共に OCT の SSI が低下するかどうかを調べるため、表 2 の疾患を除外した全てのペアデータに対し、年齢と OCT の SSI の関係を調べ、図 7、図 8 に示した。図 7 は HFA24-2 のペアデータを、図 8 は HFA10-2 のペアデータを対象とした図である。

年齢と OCT の SSI の間に有意な負の相関が得られた (HFA24-2 のペアデータのスピアマンの順位相関係数は  $p < 0.001$ 、 $\rho = -0.279$ 、(図7)、HFA10-2 のペアデータのスピアマンの順位相関係数は  $p < 0.001$ 、 $\rho = -0.293$  (図8))。

従って、年齢が高齢になるほど OCT 画像のノイズが多くなる、つまり SSI が低くなる傾向が確認でき、得られた傾きを 2.12.3 章で述べた SSI 低下による視野の傾きの補正式に用いた。

### 3.3 5 分割交差検証の結果

3次元 OCT 画像を用いて緑内障患者の視野を推測するために、3DCNN モデルの開発を行った。学習条件の違いによる OCT からの視野の推測性能の違いを調べるため、まず異なる条件で 5 分割交差検証により 3DCNN モデルを学習し、その後、統一の条件下でテストを行った。学習条件及び評価条件毎の 5 分割交差検証の結果一覧を表 5 に示す。

モデル間の精度を比較するための 5 分割交差検証の結果 (表 5)、SSI を 7 以上に制限して学習したモデル (モデル 3・4) よりも、全ての SSI で学習したモデル (モデル 1・2) の方が RMSE は低く、特に SSI 7 未満の評価対象に対しての RMSE の差が大きかった。即ち、ノイズ混じりの OCT を含んで学習した方が、ノイズ混じりの OCT からの推測性能が良い傾向が見られた。

モデル 1 はモデル 2 に比べ、HFA24-2・10-2 とともに、RMSE が低くピアソンの  $r$  も大きかったが、その差はわずかであった。モデル 3 とモデル 4 の差はほとんど認められなかった。

以上を総合して評価すると、モデル 1 が概ね最も安定した良い推測性能が得られ、モデル 2 よりもわずかに良い推測性能が得られた。つまり、今回の結果では学習時には全ての SSI を含んで学習した群の方が低い SSI を除外して学習した群よりも推測精度が高く、また眼内疾患を除外せずに学習した群は除外して学習した群よりもわずかに推測精度が高かった。但し、モデルの優劣を厳密に検定するためには、同一のモデルを複数回学習し直す必要があるため、少なくとも合計で数ヶ月間以上の長い学習時間がかかると予想されるため、今回は行っていない。

### 3.4 10 分割交差検証の結果

モデル間の精度を比較するための 5 分割交差検証の結果を受け、その推測性能を調べるため、今回の 5 分割交差検証にて比較的安定した結果が出たモデル 1 の条件、つまり学習時に眼内疾患及び全ての SSI を学習対象に含めた条件にて、10 分割交差検証を用いて検証群とテスト群を分離したモデルを再度作成し直した。その主な評価結果を表 6 に示す。

評価対象から表 2 の合併疾患を除外した図を図 9 から図 18 に示す。その中の図 9 から図 16 は、評価対象の SSI を 7 以上に限定した図である。図 17 及び図 18 は SSI の解析の図であるため、全ての SSI を評価対象にした。

### 3.4.1 推測閾値と実際の閾値の関係

推測閾値と実際の閾値に相関性があるかどうかを調べるため、10分割交差検証において、眼内疾患及び全ての SSI を学習対象に含めた条件で学習したモデルを使用し、眼内疾患及び SSI<7を除外したテスト群に対して推測した閾値と、実際の閾値の関係を図9から図12に示した。図9及び図10は測定点毎の推測閾値及び測定点毎の実際の閾値の関係を示し、図11及び図12は推測閾値の平均値及び実際の閾値の平均値の関係を示している。図9、図11は HFA24-2を、図10、図12は HFA10-2を対象とした図である。

各測定点における推測閾値と各測定点における実際の閾値の間に有意な高い相関が得られた (HFA24-2のピアソンの積率相関係数は  $p<0.001$ 、 $r=0.848$ 、(図9)、HFA10-2のピアソンの積率相関係数は  $p<0.001$ 、 $r=0.882$  (図10))。また、推測閾値の平均値と実際の閾値の平均値には有意な非常に高い相関が得られた (HFA24-2のピアソンの積率相関係数は  $p<0.001$ 、 $r=0.899$  (図11)、ピアソンの積率相関係数は  $p<0.001$ 、 $r=0.930$  (図12))。

今回の結果は、眼内疾患を含み、ノイズを含んだ OCT 画像で学習した3DCNN モデルによって、高い精度で OCT 画像から緑内障症例の視野を推測する事が可能であったという事を示している。

### 3.4.2 実際の閾値の平均値と RMSE の関係

視野が悪くなるほど視野の推測が困難になるかどうかを調べるため、10分割交差検証において、眼内疾患及び全ての SSI を学習対象に含めた条件で学習したモデルを使用し、RMSE と実際の閾値の平均値の関係を図13及び図14に示した。RMSE は、眼内疾患及び SSI<7を除外したテスト群に対して推測した閾値と実際の閾値から計算した。図13は HFA24-2を、図14は HFA10-2を対象とした図である。RMSE と実際の閾値の平均値の間に有意な負の相関が得られた (HFA24-2のスピアマンの順位相関係数は  $p<0.001$ 、 $\rho=-0.763$ 、(図13)、HFA10-2のスピアマンの順位相関係数は  $p<0.001$ 、 $\rho=-0.797$  (図14))。

今回の結果から、視野が悪化するほど視野の推測が困難になる傾向が確認できた。

### 3.4.3 OCT の focus と RMSE の関係

近視が強いほど視野の推測が困難になるかどうかを調べるため、10分割交差検証において、眼内疾患及び全ての SSI を学習対象に含めた条件で学習したモデルを使用し、OCT の focus(近視の度に類似)と RMSE の関係を図15及び図16に示した。RMSE は、眼内疾患及び SSI<7を除外したテスト群に対して推測した閾値と実際の閾値から計算した。図15は HFA24-2を、図16は HFA10-2を対象とした図である。

OCT の focus と RMSE 間に有意な正の相関が得られた (HFA24-2のスピアマンの順位相関係数は  $p<0.001$ 、 $\rho=0.0430$ 、(図15)、HFA10-2のスピアマンの順位相関係数は  $p<0.001$ 、 $\rho$

=0.102 (図16))。

このモデルでは、近視が強いほど視野の推測精度が高い傾向にあった。

### 3.4.4 OCT の SSI と RMSE の関係

ノイズの多い OCT 画像、つまり SSI の低い OCT 画像から視野を推測する事がより困難かどうかを調べるため、10分割交差検証において、眼内疾患及び全ての SSI を学習対象に含めた条件で学習したモデルを使用し、OCT の SSI と RMSE の関係を図17及び図18に示した。RMSE は、眼内疾患を除外し、全ての SSI を含むテスト群に対して推測した閾値と実際の閾値から計算した。図17は HFA24-2を、図18は HFA10-2を対象とした図である。

OCT の SSI と RMSE 間に有意な負の相関が得られた (HFA24-2のスピアマンの順位相関係数は  $p < 0.001$ 、 $\rho = -0.115$ 、(図17)、HFA10-2のスピアマンの順位相関係数は  $p < 0.001$ 、 $\rho = -0.128$  (図18))。

従って、OCT 画像のノイズが多いほど、視野の推測が困難となる傾向があった。

## 4 考察

OCT は視野検査よりも短い時間で測定できるため、OCT から視野を推測する事は临床上重要である。より多くのペアデータを確保し、高い推測精度を得るため、従来の2次元のモデルの代わりに3DCNN を用いた。教師値は、測定毎のばらつきを抑えるために HFA の測定点毎の回帰直線から計算して用いた。正の傾きを0とした事による影響及び、SSIの低下による見かけ上の視野の進行の影響を減らすため、傾きの教師値に予め補正を行った。学習対象を変更する事によって得られる推測性能の違いを比較するため、それぞれの学習条件での5分割交差検証モデルを作成し、共通した評価対象での推測性能を比較した結果、全ての眼内疾患を含有して全ての SSI を対象として学習した群が最も良い結果が得られた。今回の結果からは、必ずしも眼内疾患を除外して学習する必要性は無いという可能性が示唆され、また、低い SSI を敢えて除外して学習する必要性は無いという可能性も示唆された。眼内疾患を除外する必要が無ければ、後ろ向きに眼内疾患の有無を手作業で調べる労力が減り、新たな症例を集めやすくなる可能性が考えられる。機械学習の推測性能を向上させるためには多くの症例を集める必要があるが、従来の2次元モデルではセグメンテーションエラーの有無のチェック及び合併疾患の有無のチェックが必要であり、ノイズの乗った OCT 画像も除外する必要があったため、症例数を増やす事は困難であり、その結果、推測性能の向上には限界があった。それに対して、3DCNN はセグメンテーション不要のモデルであり、今回の結果からは、眼内疾患を含有してノイズが乗った OCT も含んで学習する事が可能という可能性が示唆されたため、その場合には中枢疾患及びアーチファクトを除いて、ほとんどの症例を除外せずにそのまま学習できるという事になる。既報では、眼科の分野では無いが、CNN モ

デルの1つである EfficientNet<sup>31</sup>を使用した画像の分類タスクにわざとノイズを付加して学習させる事で推測性能を向上させるという手法もあるため<sup>42</sup>、セグメンテーション不要という3DCNN の特徴を活かすためには、ノイズの乗った OCT 画像を除外せずに学習したほうが良い推測性能が得られる可能性が考えられる。もしそうであれば、今後多くの施設でさらに症例数を容易に増やす事ができ、さらなる推測性能の向上が期待できる。合併疾患やノイズが乗った OCT 画像を同時に学習する事の利点はもう1つある。実際の臨床では、緑内障に何らかの他の疾患が合併したり、OCT 画像にノイズが乗ったりする事をしばしば経験する。例えば黄斑上膜を合併した緑内障の場合、セグメンテーションがうまくできず、OCT 画像にノイズが乗ってしまうため、従来は OCT 画像からの視野の推測は困難であった。今回の3DCNN モデルを利用すれば、そのような合併疾患を丸ごと含んで学習する事により、従来では困難であった合併疾患を有する緑内障患者の視野も推測できる可能性が期待できる。同様に、ノイズが乗った OCT 画像で学習すれば、ノイズが乗った OCT 画像からの推測性能の向上も期待できる。

5分割交差検証で最も良い結果が得られた学習条件と同一の学習条件で10分割交差検証モデルを作成し学習し、最終的な推測精度を計算し、モデルの構成要素である、focus、SSI、平均閾値と RMSE との関係を解析した。今回のモデルによって、単回の OCT の黄斑マップから極めて精度の高い HFA24-2及び10-2の閾値推測が可能であった。OCT の黄斑マップ単独では範囲が狭いため、従来は HFA24-2の閾値推測は困難と考えられていた<sup>7,8,13</sup>が、今回のモデルでは HFA24-2も高い推測性能が得られた。OCT の黄斑マップの撮影範囲外の視野は、多数の同様の視野から学習してモデル内部で作り出して推測している可能性が考えられる。

今回のモデルでは従来一般的な考え<sup>20</sup>とは逆に、近視が強いほど高い推測精度が得られた。その主な理由として考えられるのは、3DCNN モデルを採用した事である。強度近視眼では網膜厚が薄い事が知られており<sup>21</sup>、眼球の形状によって部位別に生理的な網膜厚が異なる可能性が考えられる。2次元のモデルでは、3次元の情報を2次元に圧縮する際に形状の情報の多くが失われるが、3DCNN モデルであれば、3次元の情報をそのまま学習する事ができるため、強度近視の症例でより正確に推測できるようになったのではと考えた。

過去に OCT から HFA の閾値を推測する報告は多数あり<sup>7-16</sup>、そのほとんどは2次元のモデルであった。直近の既報の値は、母集団が異なり既報では大学病院の重症例が多いため、より視野の推測が困難であったと考えられるために単純な比較はできないが、HF10-2の RMSE:5.92dB (Xu ら)<sup>13</sup>に対して今回のモデルでは3.02dB、HFA10-2 測定点毎の MAE:5.0dB (Asaoka ら)<sup>8</sup>に対して今回のモデルでは2.24dB、HF24-2の RMSE:4.51dB (Shin ら)<sup>10</sup>に対して今回のモデルでは3.27dB であり、今回のモデルは既報の値をいずれも大きく上回る推測性能を達成した。その原因として主に考えられるのは、セグメンテーション不要の3DCNN モデルを用いて3次元画像のまま学習した事、低い SSI も含めて学習した事、眼内疾患を含めて学習した事、教師値に回帰直線を

用いて OCT 測定時の視野を計算した事、ペアデータ数をはるかに多い事、左眼のデータを反転として擬似的に右眼のデータとして学習した事、上下反転をテスト時にも用いて平均値を算出した事、3DCNN の中でも特に推測性能の優れた EfficientNet3D をモデルに採用した事、ハイパーパラメータの調整、検証群の採用など、複数の要因の積み重ねが可能性として考えられ、以下ではそれらに関して考察する。

視野検査は測定毎のばらつきがある検査であり<sup>25,26</sup>、単回の視野では本来の視野からの乖離が大きいが、時系列による視野の回帰直線を作成する事により、本来の視野に近づくはずである。既報では OCT の直近の視野の値をそのまま教師値として採用する報告がほとんどであり、その結果、本来の視野から乖離した値を教師値として採用する事になってしまったため、推測性能の向上の障害となっていた可能性が考えられる。今回のモデルでは、回帰直線から計算した視野を教師値として採用する事で、より本来の視野に近い値を教師値として採用できる事になり、その結果推測性能が向上した可能性が考えられる。同時に、例えば OCT 画像にノイズが確認されたためにすぐに OCT を撮影し直した場合等の、ノイズが乗った OCT 画像等も全て学習対象として含み、回帰直線から計算した視野をそれぞれの OCT 画像の教師値として全て採用する事により、より多くのペアデータを作成する事ができたため、推測性能が向上した可能性も考えられる。

眼球は左右対象な構造であり、右眼の OCT 画像と左眼の OCT 画像は水平方向に反転すれば同一の形状になるため、必ずしもそのままの向きで学習する必要は無いはずであるが、既報では右眼の OCT 画像と左眼の OCT 画像をそのままの向きで学習する報告がほとんどであった。本来、推測性能を向上させるためには、推測したい画像と似たような画像の数を増やすことが大切である。右眼の画像と左眼の画像はそのままの向きでは形状が異なるため、そのままの向きで学習する代わりに、左眼の画像を水平反転して擬似的に右眼の画像として学習した結果、同じ形状のペアデータ数が約 2 倍に増える事になったため、推測性能が向上した可能性も考えられる。また、垂直反転して水増しして学習したモデルを使用し、既報ではそのままの方向の画像からの出力のみを使用した報告がほとんどであったが、今回のモデルでは学習時だけではなく推測時にも垂直反転した画像からも推測し、垂直反転していない画像からの出力との平均値を出力する事により、推測性能が向上した可能性も考えられる。

既報では CNN モデルとして ResNet<sup>43,44</sup>を採用している報告が多い。ResNet は CNN モデルとしては最も有名なモデルであり、眼科以外でも世界中の多くの分野で使用されているモデルである。CNN モデルの推測性能を向上させるためには、CNN モデルのネットワークの深さ、広さ、入力画像の大きさの要素が大切であり、それらの関係性を調べる事で推測性能の向上を達成した CNN モデルが EfficientNet<sup>31</sup>である。EfficientNet は従来の ResNet に比べてパラメータ数が少ないためメモリ使用量が少なく、学習時間が短く、推測性能が高いという特徴がある。今回採用したモデルは、その 3 次元版である EfficientNet3D であり、そのために推測性能が向上した可能性

も考えられる。

既報の OCT から視野を推測するモデルでは、ハイパーパラメータについての詳細な記載はほとんど見当たらないが、機械学習の推測性能を向上させるためには、予め十分な時間をかけてハイパーパラメータを調整しておく事は大切である。今回のモデルではハイパーパラメータの調整を予め繰り返し行っており、そのため推測性能が向上したという可能性も考えられる。

既報の OCT から視野を推測するモデルの多くは、学習群、テスト群の2群に分けていたため、過学習を起こした場合にテスト群での推測性能が低下してしまう恐れが考えられた。今回の研究では、学習群、検証群、テスト群と3つの群に分け、検証群での最良値が得られたモデルを利用する事により、過学習に陥る前の最良の推測性能のモデルを利用してテスト群を推測する事ができたため、より高い推測性能が得られた可能性も考えられる。逆に今回の手法の欠点としては、全体に占める学習群の比率が同一の場合、従来手法よりも学習時間が長くなるという点が挙げられる。

既報では、分割交差検証を使用せずに単純に学習群とテスト群に分けている報告もあった。その利点としては、学習時間が短く済むという点が挙げられ、逆にその欠点としては、学習群とテスト群の分け方によって結果にばらつきが生じてしまうという点が挙げられる。今回は10分割交差検証を使用した事により、分割を10回ずらして学習し全てのテスト対象を網羅したため、単純に2群に分けた1回のみ学習よりもより正確に推測性能を評価できた可能性が高い。

今回のモデルにより現在の視野だけではなく、同時に視野の進行速度も単回の OCT 画像からある程度推測できた。既報では、既存の多数の視野の進行速度を教師値として視野からその進行速度を推測するという報告<sup>23</sup>はあるが、OCT 画像を元にして視野の進行速度を推測するという報告は調べた限りでは見当たらない。視野のデータよりも OCT 画像のデータの方が情報量は多く、機械学習によって3次元の眼底構造から視野が進行しやすいかどうかを学習できる可能性が考えられ、さらに OCT 画像から視野を推測してその推測視野から既報と同様に視野の進行速度を推測する事も可能である。今後多数の症例が集まれば、視野を元にするよりも OCT 画像を元にしたほうが、視野の進行速度の推測精度の向上が見込める可能性も考えられる。また、今回の研究では単回の OCT 画像を元に進行速度を推測しており、OCT 画像の時系列データによる情報は加味していない。将来的には、今回の研究モデルを基本モデルとし、時系列での OCT 画像データを今回のモデルに入力して、出力された時系列でのデータに対して追加解析を行う新たな機械学習モデルを作成する事により、進行速度の推測精度をさらに向上できる可能性も考えられる。今回は単一の診療所の調査のため有意な進行速度を教師値に採用できた症例数が少ないため、推測性能を向上させるためには症例数を増やす事が今後の課題である。

今回の報告の母集団は単一クリニックであるため、後ろ向き全例調査であっても症例数は限られている。また既報の母集団は当院よりも重症例の多い大学病院での報告が多い。そのため、従



来の2DCNN の手法よりも今回の3DCNN の手法の方が推測性能に優れるかどうかの検討をするためには、今後の他施設の追加研究が必要である。

今回の3DCNN モデルにより、単回の OCT 画像から精度の高い24-2、10-2の視野の推測が可能であった。当院では既に今回のモデルを臨床応用しており、OCT 撮影後の画像を出力した直後に自動的に24-2、10-2の視野の推測+視野の進行速度の推測を行っているが、他院での使用のためには、OCT の機種や撮影条件が少しでも異なると正しく推測できないため、多くのメーカーの OCT で様々な撮影条件での学習が必要である。そのためには多施設共同研究が必須であり、今回の3DCNN による後ろ向き全例調査は、そのような将来の多施設共同研究の基礎となる手法として有用な可能性が考えられる。

## 謝辞および金銭的開示

- a. 資金援助は無い。
- b. 金銭的な開示は無い。
- c. American Journal Experts (AJE)の英文校正に感謝する。

## 著者の記述

**古山 誠**：概念化、方法論、ソフトウェア、検証、形式分析、調査、リソース、データキュレーション、執筆、可視化、プロジェクト管理。

## データの利用可能性に関する声明

研究データは共有されない。

本研究の性質上、本研究の参加者はデータを公開することに同意していないため、支援データは利用できない。

## 利益相反

該当無し。

**キーワード**：緑内障、光干渉断層計、視野、深層学習

## 参考文献

1. Weinreb, R. N. *et al.* Primary open-angle glaucoma. *Nature Reviews Disease Primers* vol. 2 Preprint at <https://doi.org/10.1038/nrdp.2016.67> (2016).
2. Allison, K., Patel, D. & Alabi, O. Epidemiology of Glaucoma: The Past, Present, and Predictions for the Future. *Cureus* (2020) doi:10.7759/cureus.11686.
3. Altangerel, U., Spaeth, G. L. & Rhee, D. J. *Visual function, disability, and psychological impact of glaucoma.* *Curr Opin Ophthalmol* vol. 14 (2003).
4. Wang, M. *et al.* Characterization of Central Visual Field Loss in End-stage Glaucoma by Unsupervised Artificial Intelligence. *JAMA Ophthalmol* **138**, (2020).
5. Blumberg, D. M. *et al.* Association between undetected 10-2 visual field damage and vision-related quality of life in patients with glaucoma. *JAMA Ophthalmol* **135**, (2017).
6. Bansal, M., Arora, T. & Dada, T. Effect of a novel computer software simulating humphrey visual field (HVF) on patient performance of HVF. *Invest Ophthalmol Vis Sci* **55**, (2014).
7. Hashimoto, Y. *et al.* Predicting 10-2 Visual Field From Optical Coherence Tomography in Glaucoma Using Deep Learning Corrected With 24-2/30-2 Visual Field. *Transl Vis Sci Technol* **10**, (2021).
8. Asaoka, R. *et al.* A Joint Multitask Learning Model for Cross-sectional and Longitudinal Predictions of Visual Field Using OCT. *Ophthalmology Science* **1**, 100055 (2021).
9. Cirafici, P. *et al.* Point-wise correlations between 10-2 Humphrey visual field and OCT data in open angle glaucoma. *Eye (Basingstoke)* **35**, 868–876 (2021).
10. Shin, J., Kim, S., Kim, J. & Park, K. Visual field inference from optical coherence tomography using deep learning algorithms: A comparison between devices. *Transl Vis Sci Technol* **10**, (2021).
11. Sugiura, H. *et al.* Estimating glaucomatous visual sensitivity from retinal thickness with pattern-based regularization and visualization. in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 783–792 (Association for Computing Machinery, 2018). doi:10.1145/3219819.3219866.
12. Xu, L. *et al.* Improving Visual Field Trend Analysis with OCT and Deeply Regularized Latent-Space Linear Regression. *Ophthalmol Glaucoma* **4**, (2021).
13. Xu, L. *et al.* PAMI: A Computational Module for Joint Estimation and Progression Prediction of Glaucoma. in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 3826–3834 (Association for Computing Machinery, 2021). doi:10.1145/3447548.3467195.
14. Xu, L. *et al.* Predicting the Glaucomatous Central 10-Degree Visual Field From Optical Coherence Tomography Using Deep Learning and Tensor Regression. *Am J Ophthalmol* **218**, (2020).
15. Hashimoto, Y. *et al.* Deep learning model to predict visual field in central 10° from optical coherence tomography measurement in glaucoma. *British Journal of Ophthalmology* **105**, (2021).
16. Asano, S. *et al.* Predicting the central 10 degrees visual field in glaucoma by applying a deep learning algorithm to optical coherence tomography images. *Sci Rep* **11**, (2021).
17. Çallı, E., Sogancıoğlu, E., van Ginneken, B., van Leeuwen, K. G. & Murphy, K. Deep learning for chest X-ray analysis: A survey. *Medical Image Analysis* vol. 72 Preprint at <https://doi.org/10.1016/j.media.2021.102125> (2021).
18. Harwerth, R. S., Wheat, J. L., Fredette, M. J. & Anderson, D. R. Linking structure and function in glaucoma. *Progress in Retinal and Eye Research* vol. 29 Preprint at <https://doi.org/10.1016/j.preteyeres.2010.02.001> (2010).
19. Garcia-Medina, J. J., Rotolo, M., Rubio-Velazquez, E., Pinazo-Duran, M. D. & Del-Rio-vellosillo, M. Macular structure–function relationships of all retinal layers in primary open-angle glaucoma assessed by microperimetry and 8 × 8 posterior pole analysis of oct. *J Clin Med* **10**, (2021).
20. Akashi, A. *et al.* The ability of SD-OCT to differentiate early glaucoma with high myopia from

- highly myopic controls and nonhighly myopic controls. *Invest Ophthalmol Vis Sci* **56**, (2015).
21. Sezgin Akcay, B. I., Gunay, B. O., Kardes, E., Unlu, C. & Ergin, A. Evaluation of the Ganglion Cell Complex and Retinal Nerve Fiber Layer in Low, Moderate, and High Myopia: A Study by RTVue Spectral Domain Optical Coherence Tomography. *Semin Ophthalmol* **32**, (2017).
  22. Anderson, A. J. Significant glaucomatous visual field progression in the first two years: What does it mean? *Transl Vis Sci Technol* **5**, (2016).
  23. Murata, H., Araie, M. & Asaoka, R. A new approach to measure visual field progression in glaucoma patients using variational bayes linear regression. *Invest Ophthalmol Vis Sci* **55**, (2014).
  24. Yu, S. *et al.* Comparison of SITA Faster 24-2C test times to legacy SITA tests. *Invest Ophthalmol Vis Sci* **60**, (2019).
  25. L.J., S., R.A., R. & D.P., C. Standard or Fast?-Differences in precision between SITA Standard and SITA Fast testing algorithms and their utility for detecting visual field deterioration. *Invest Ophthalmol Vis Sci* **55**, (2014).
  26. Saunders, L. J., Russell, R. A. & Crabb, D. P. Measurement precision in a series of visual fields acquired by the standard and fast versions of the swedish interactive thresholding algorithm analysis of large-scale data from clinics. *JAMA Ophthalmol* **133**, (2015).
  27. Anderson DR, P. V. *Automated Static Perimetry, 2nd edition.* (Mosby, St. Louis, 1999).
  28. Nakatani, Y., Higashide, T., Ohkubo, S., Takeda, H. & Sugiyama, K. Effect of cataract and its removal on ganglion cell complex thickness and peripapillary retinal nerve fiber layer thickness measurements by fourier-domain optical coherence tomography. *J Glaucoma* **22**, 447–455 (2013).
  29. Kim, M., Eom, Y., Song, J. S. & Kim, H. M. Effect of Cataract Grade according to Wide-Field Fundus Images on Measurement of Macular Thickness in Cataract Patients. *Korean Journal of Ophthalmology* **32**, (2018).
  30. shijianjian. EfficientNet-PyTorch-3D. github.com. Updated July 20, 2021. Accessed May 10, 2022. <https://github.com/shijianjian/EfficientNet-PyTorch-3D> (2021).
  31. Tan, M. & Le, Q. v. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv preprint:1905.11946* (2019).
  32. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. (2014).
  33. Konar, J., Khandelwal, P. & Tripathi, R. Comparison of Various Learning Rate Scheduling Techniques on Convolutional Neural Network. in *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2020* (Institute of Electrical and Electronics Engineers Inc., 2020). doi:10.1109/SCEECS48394.2020.94.
  34. Liu, L. *et al.* On the Variance of the Adaptive Learning Rate and Beyond. (2019).
  35. Qian, X. & Klabjan, D. The Impact of the Mini-batch Size on the Variance of Gradients in Stochastic Gradient Descent. (2020).
  36. Wu, H. & Gu, X. Towards dropout training for convolutional neural networks. *Neural Networks* **71**, (2015).
  37. Loshchilov, I. & Hutter, F. Decoupled Weight Decay Regularization. (2017).
  38. Zhuang, J. *et al.* AdaBelief optimizer: adapting stepsizes by the belief in observed gradients. *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems* 18795–18806 (2020).
  39. Qian, N. *On the momentum term in gradient descent learning algorithms.* <https://www.sciencedirect.com/science/article/abs/pii/S0893608098001166> (1999) doi:[https://doi.org/10.1016/S0893-6080\(98\)00116-6](https://doi.org/10.1016/S0893-6080(98)00116-6).
  40. Foret, P., Kleiner, A., Mobahi, H. & Neyshabur, B. Sharpness-Aware Minimization for Efficiently Improving Generalization. (2020).
  41. Kwon, J., Kim, J., Park, H. & Choi, I. K. ASAM: Adaptive Sharpness-Aware Minimization for Scale-Invariant Learning of Deep Neural Networks. (2021).
  42. Xie, Q., Luong, M.-T., Hovy, E. & Le, Q. v. Self-training with Noisy Student improves ImageNet classification. (2019).

43. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. (2015).
44. He, K., Zhang, X., Ren, S. & Sun, J. Identity Mappings in Deep Residual Networks. (2016).

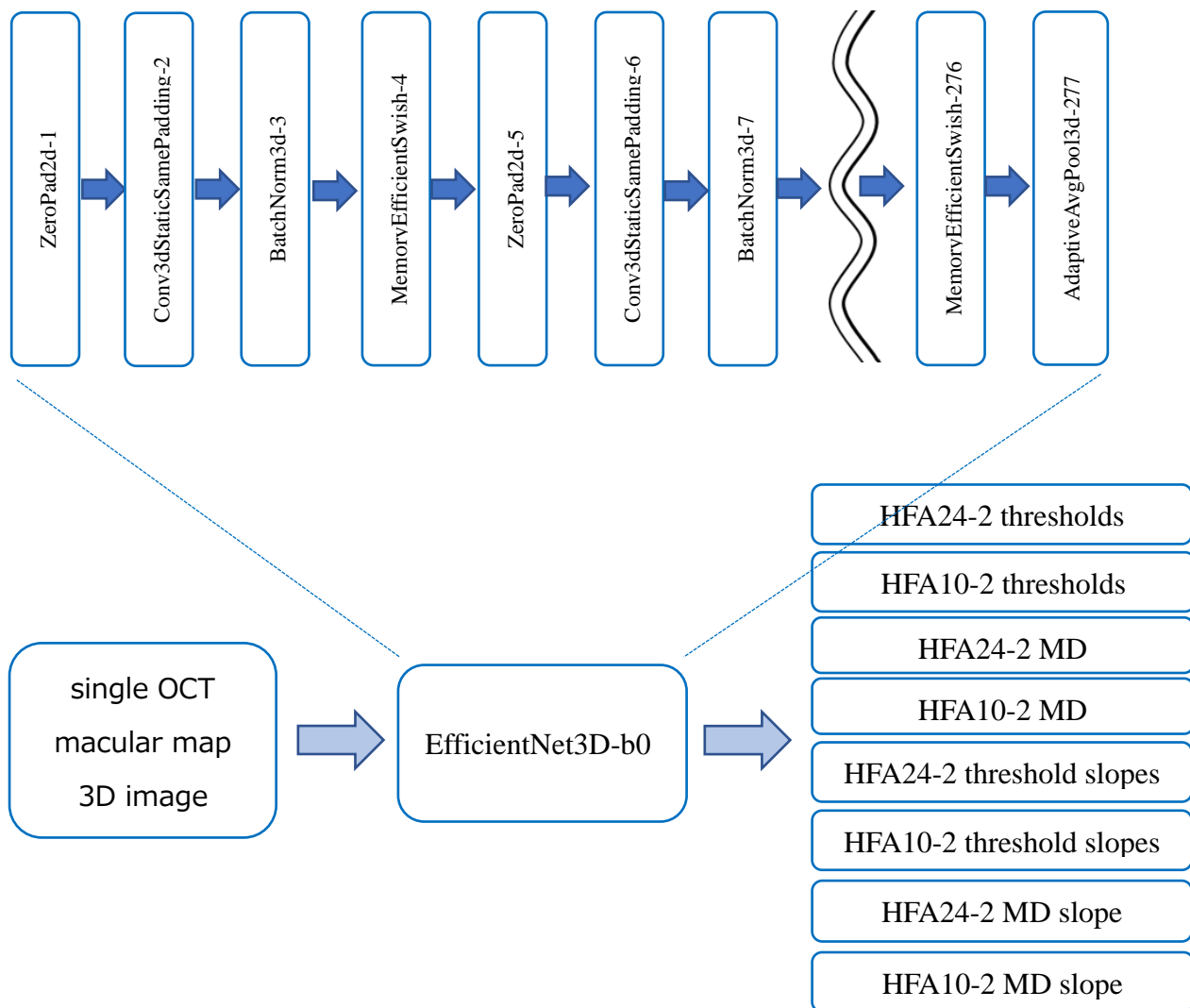


図1 今回のモデルのイメージ図。単回のOCTの黄斑マップ3次元画像の情報を3DCNNモデルのEfficientNet3Dへ与え、HFA24-2、10-2の閾値及びそのMD値とその傾きを同時に出力した。

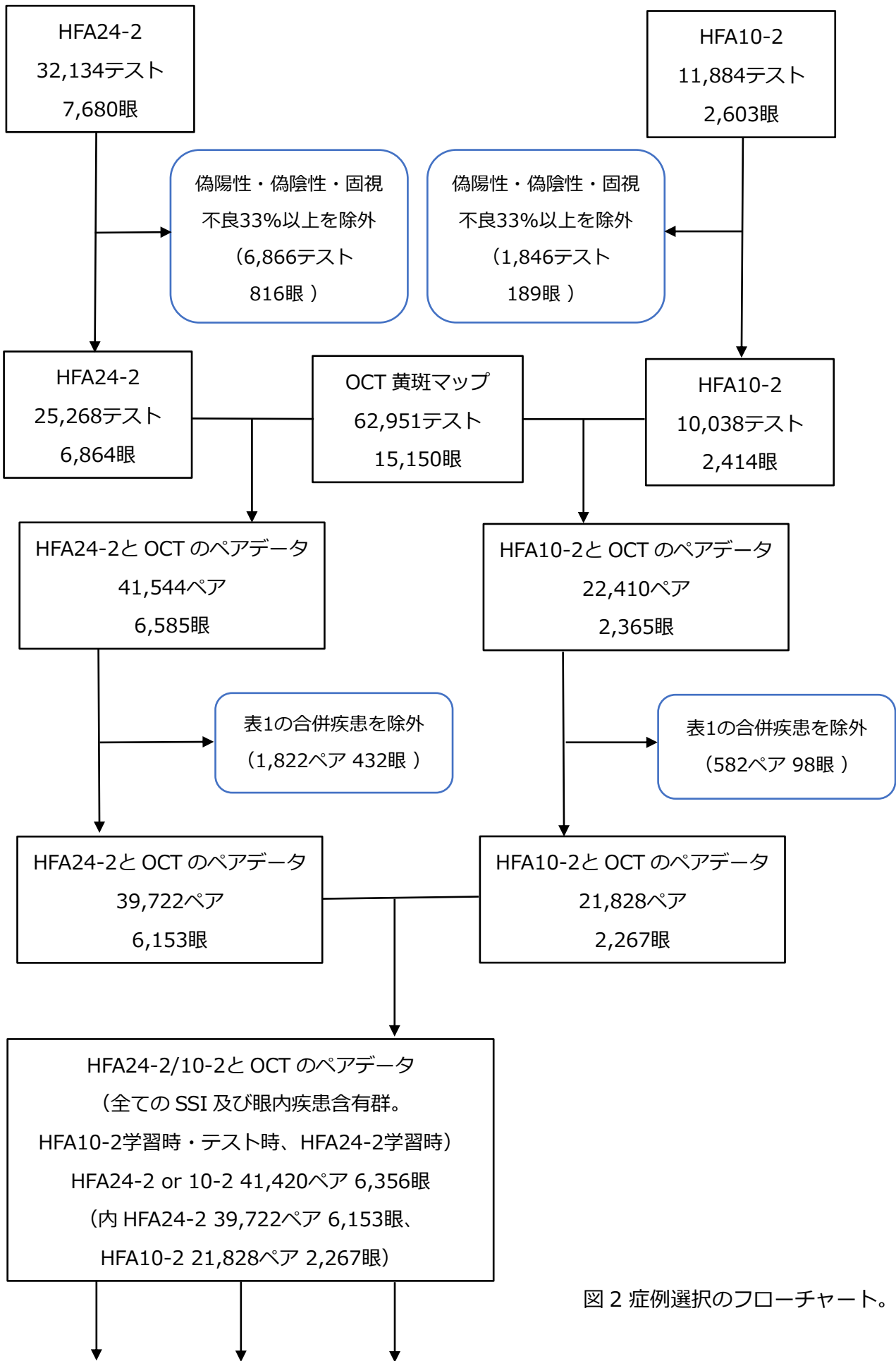


図2 症例選択のフローチャート。

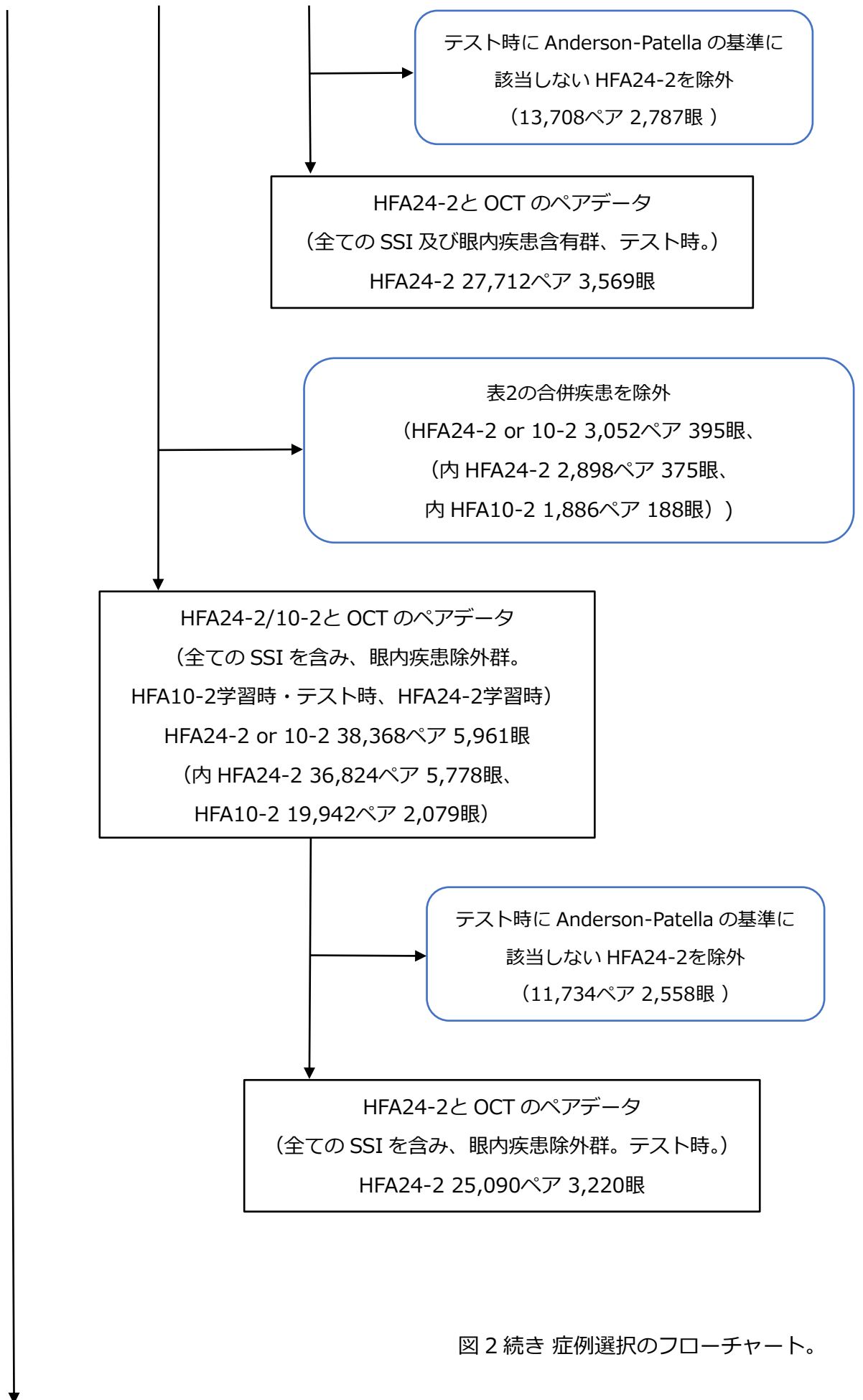


図 2 続き 症例選択のフローチャート。



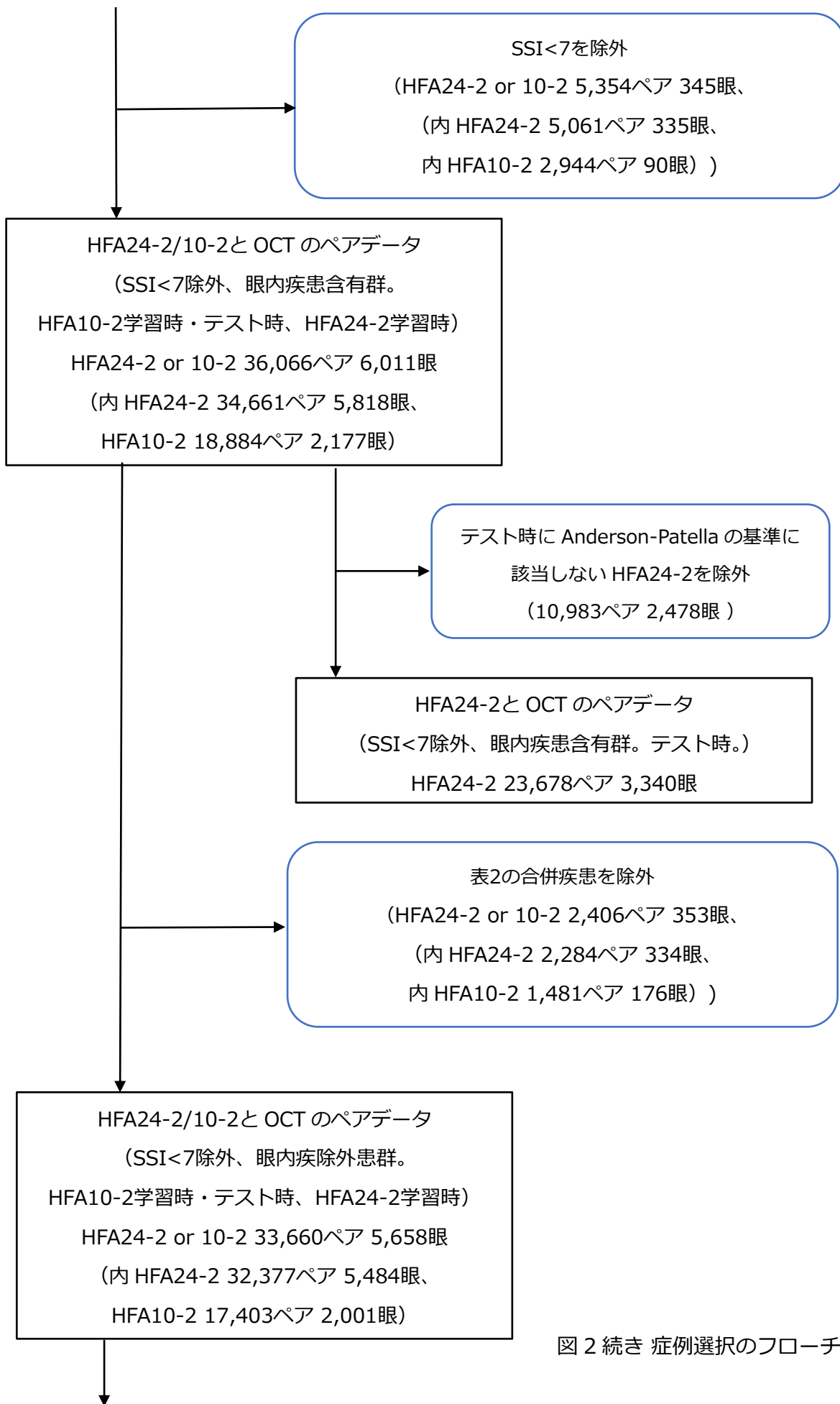


図 2 続き 症例選択のフローチャート。

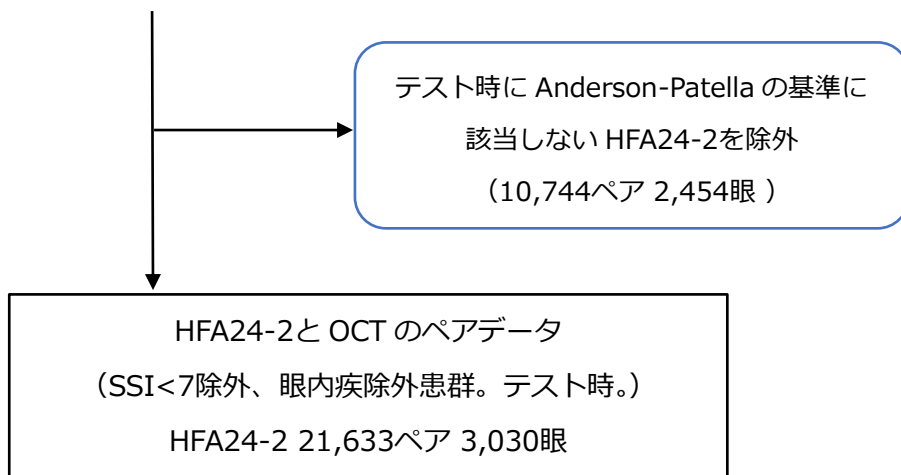


図 2 続き 症例選択のフローチャート。

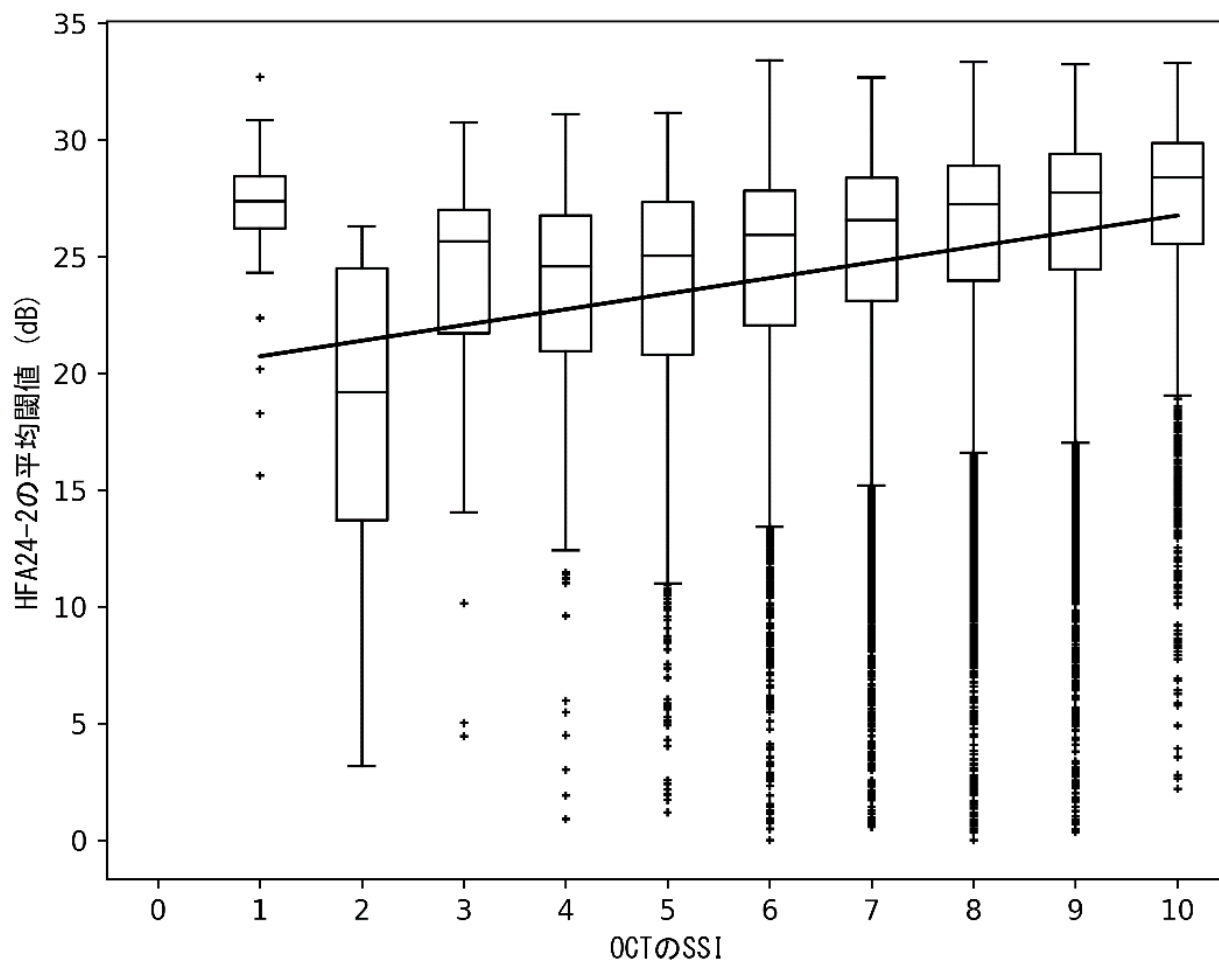


図3 OCTの全てのSSIとHFA24-2の平均閾値のグラフ。

SSIが低いほど平均閾値が低値であった。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = 0.220$ )

図の線は回帰直線。回帰係数 = 0.671 (dB/SSI)。

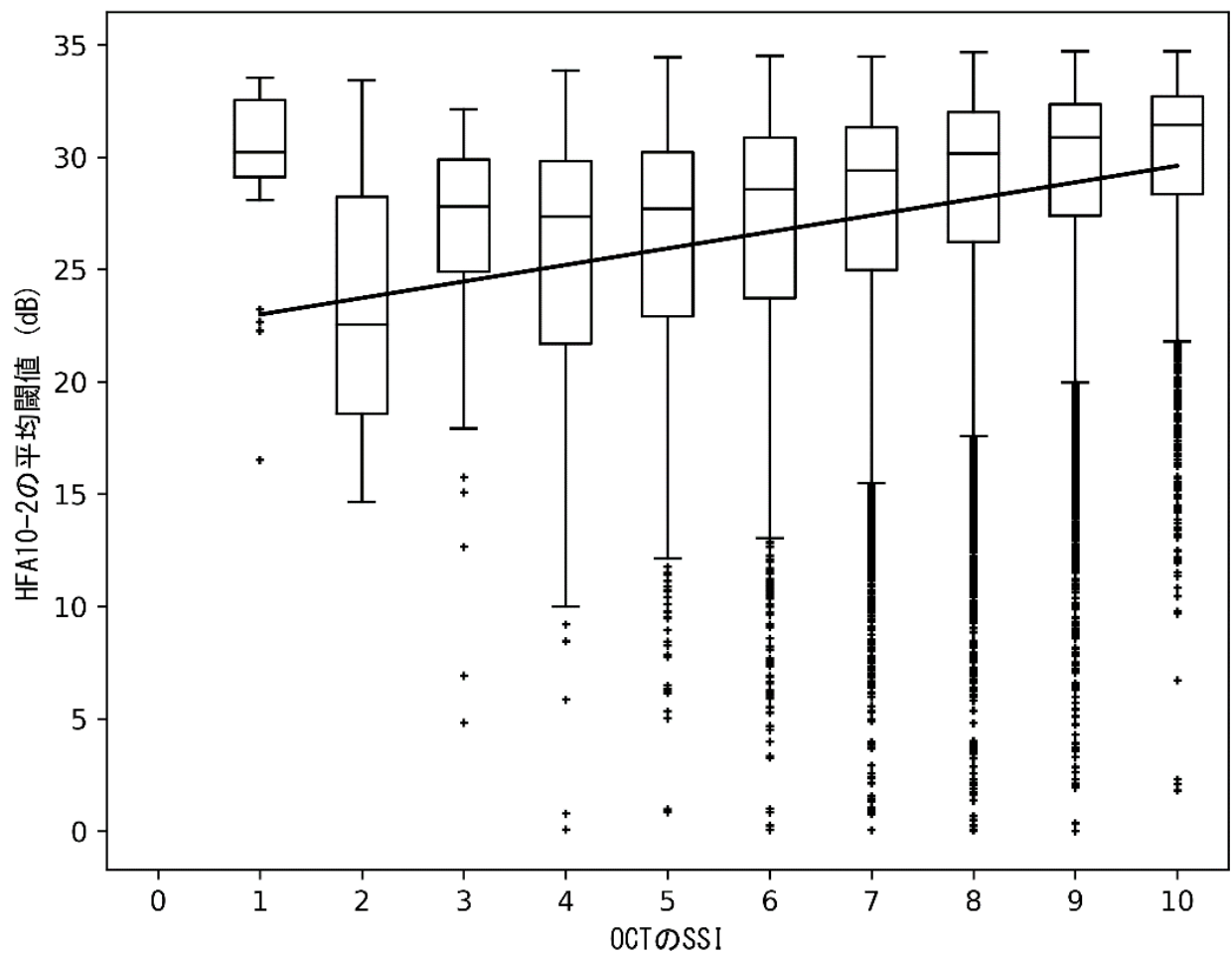


図4 OCTの全てのSSIとHFA10-2の平均閾値のグラフ。

HFA24-2同様、SSIが低いほど平均閾値が低値であった。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = 0.230$ )

図の線は回帰直線。回帰係数 =  $0.736$  (dB/SSI)。

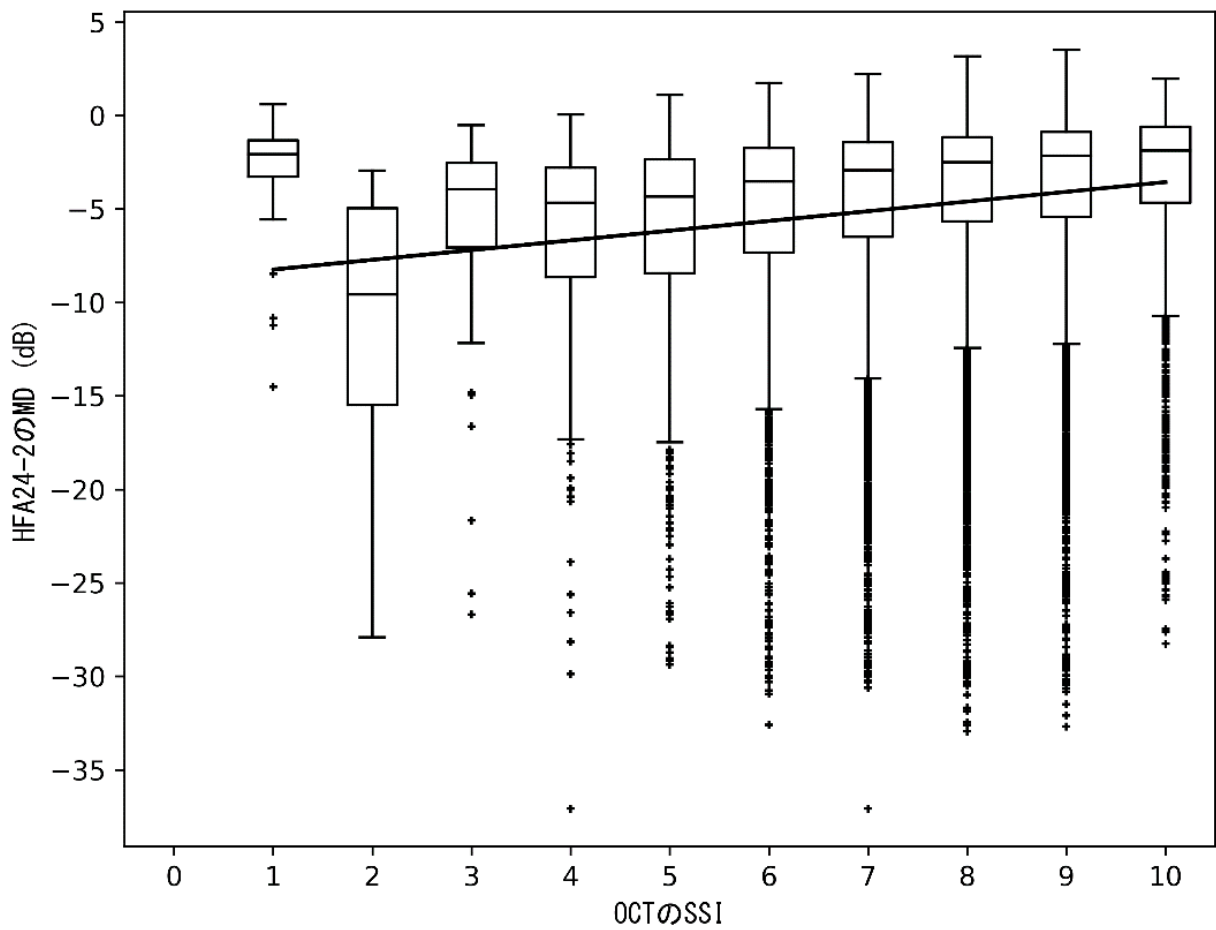


図5 OCTの全てのSSIとHFA24-2のMD値のグラフ。

SSIが低いほどMD値が低値であった。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = 0.166$ )

図の線は回帰直線。回帰係数 = 0.520 (dB/SSI)。

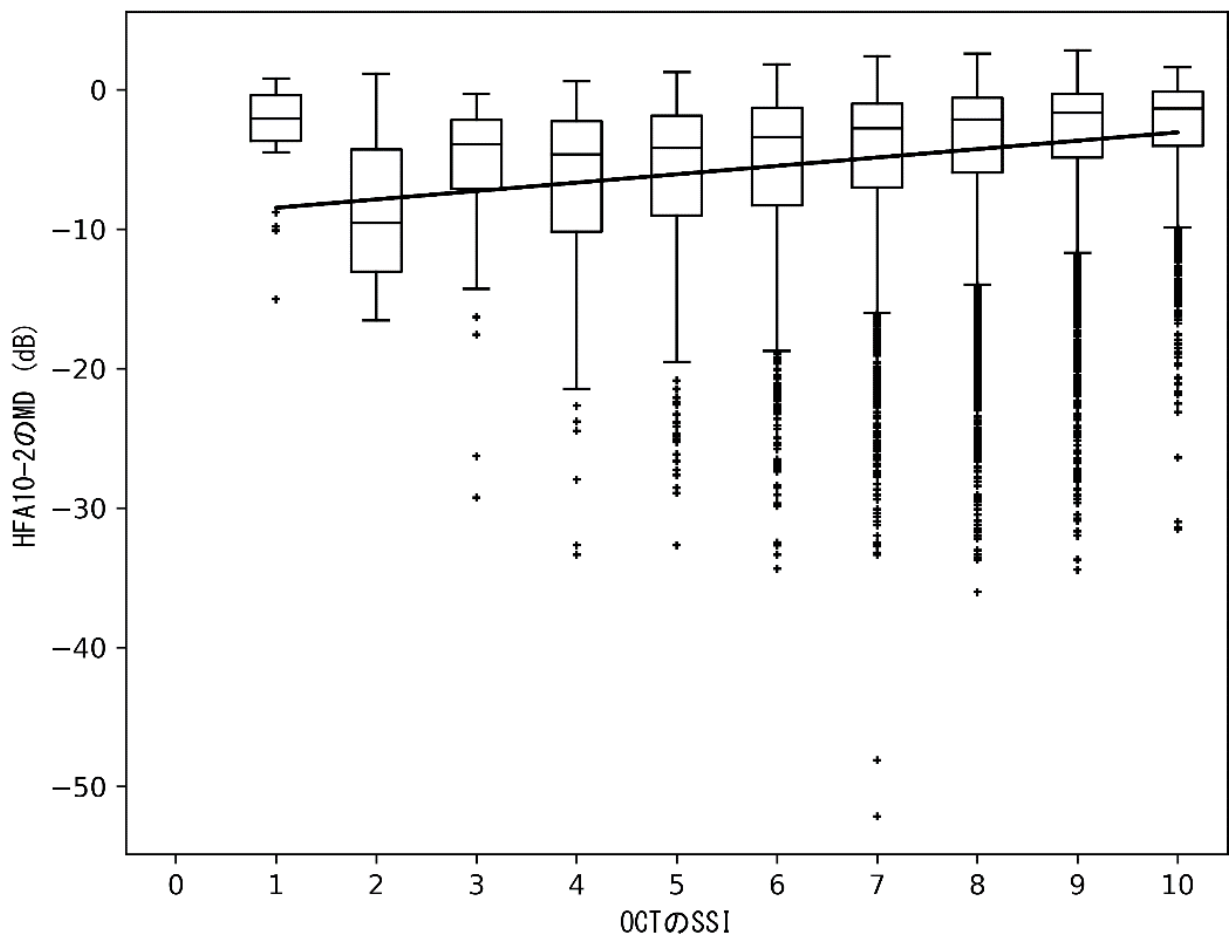


図6 OCTの全てのSSIとHFA10-2のMD値のグラフ。

HFA24-2同様、SSIが低いほどMD値が低値であった。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = 0.188$ )

図の線は回帰直線。回帰係数 = 0.601 (dB/SSI)。

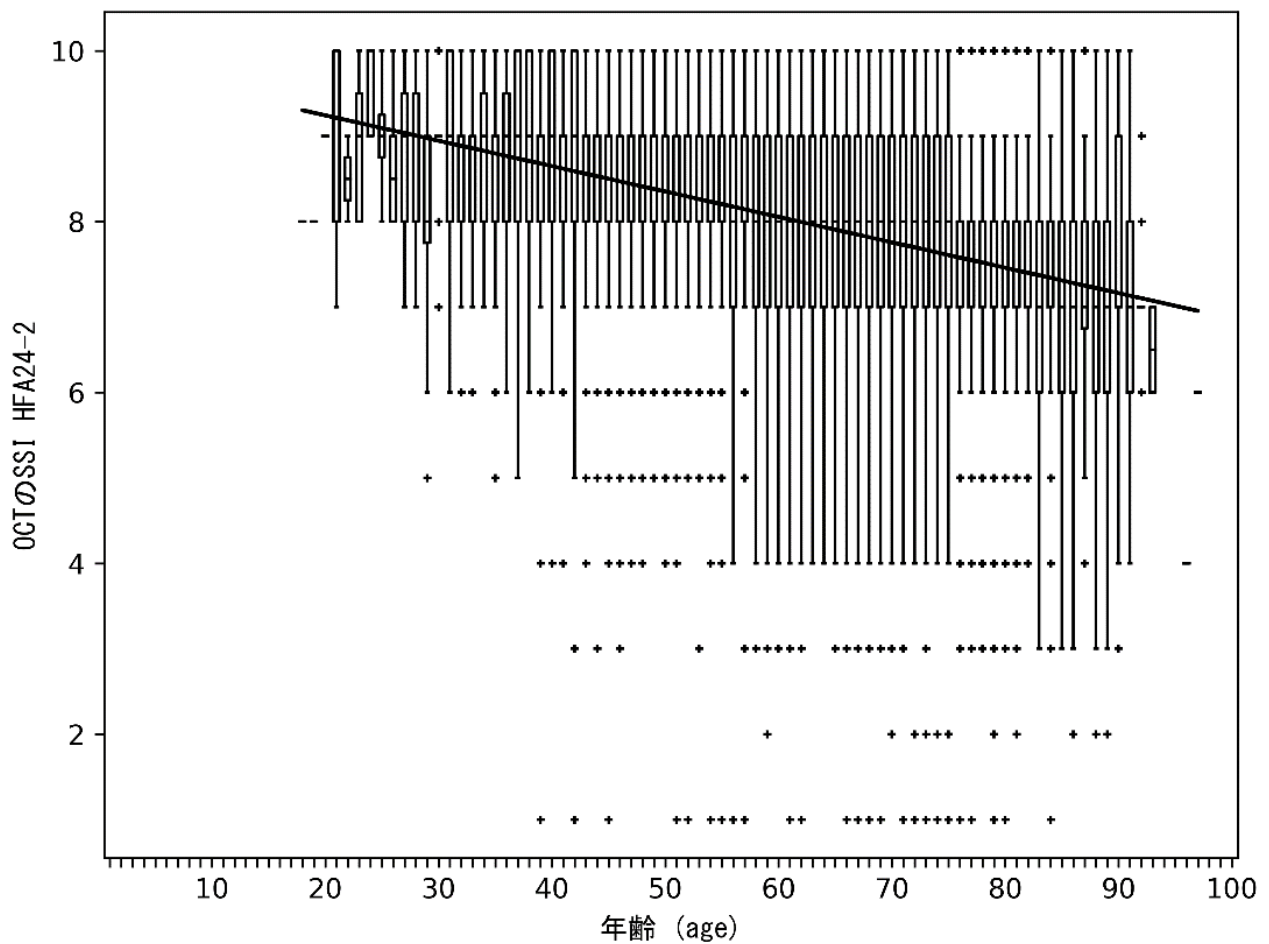


図7 HFA24-2のペアデータからの、年齢と全ての OCT の SSI の平均値のグラフ。

年齢が高齢になるほど SSI は低下していく傾向が認められた。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = -0.279$ )

図の線は回帰直線。回帰係数 =  $-0.0297$  (SSI/age)。

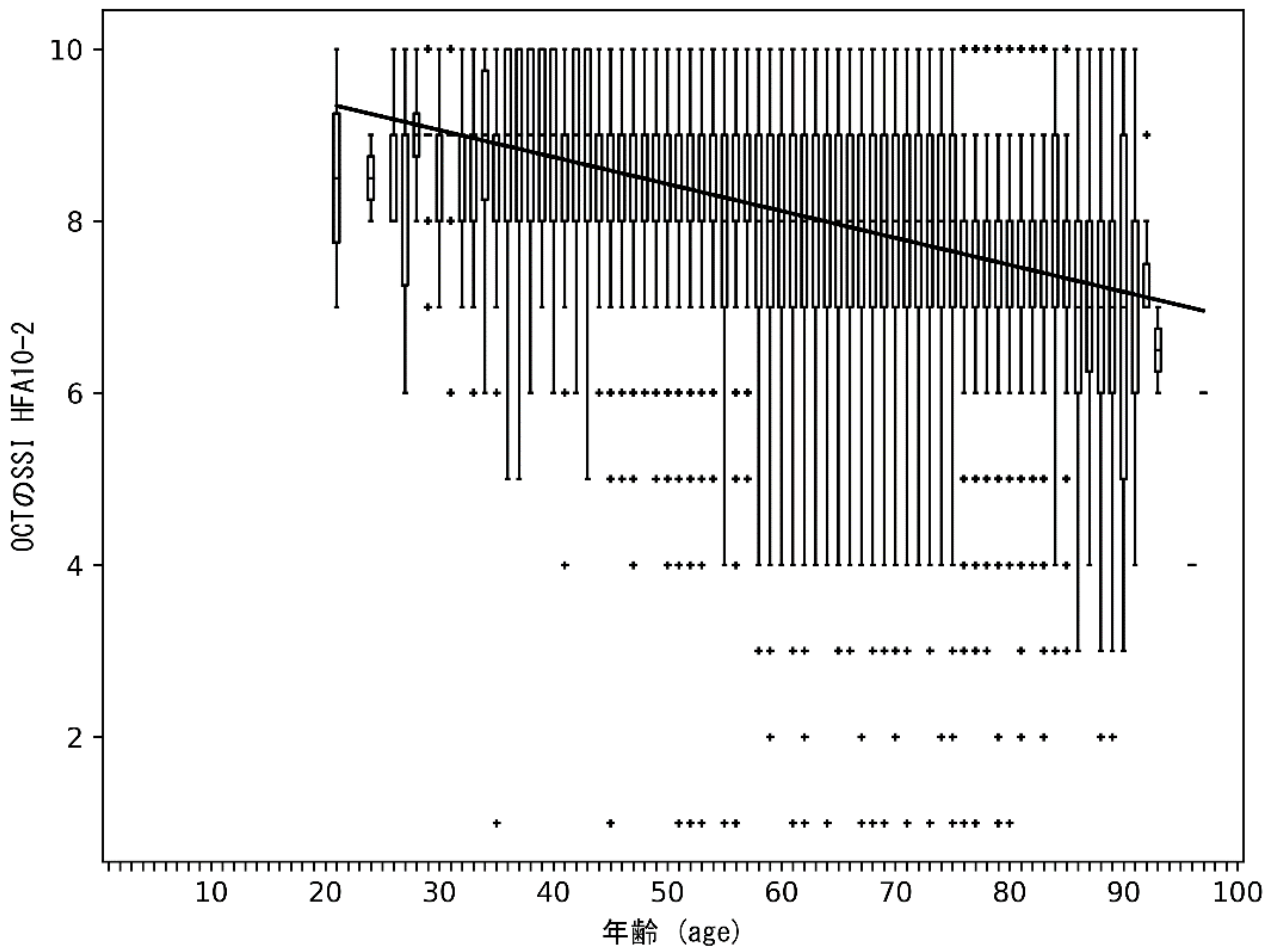


図8 HFA10-2のペアデータからの、年齢と全ての OCT の SSI の平均値のグラフ。

HFA24-2同様、年齢が高齢になるほど SSI は低下していく傾向が認められた。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = -0.293$ )

図の線は回帰直線。回帰係数 =  $-0.0314$  (SSI/age)。



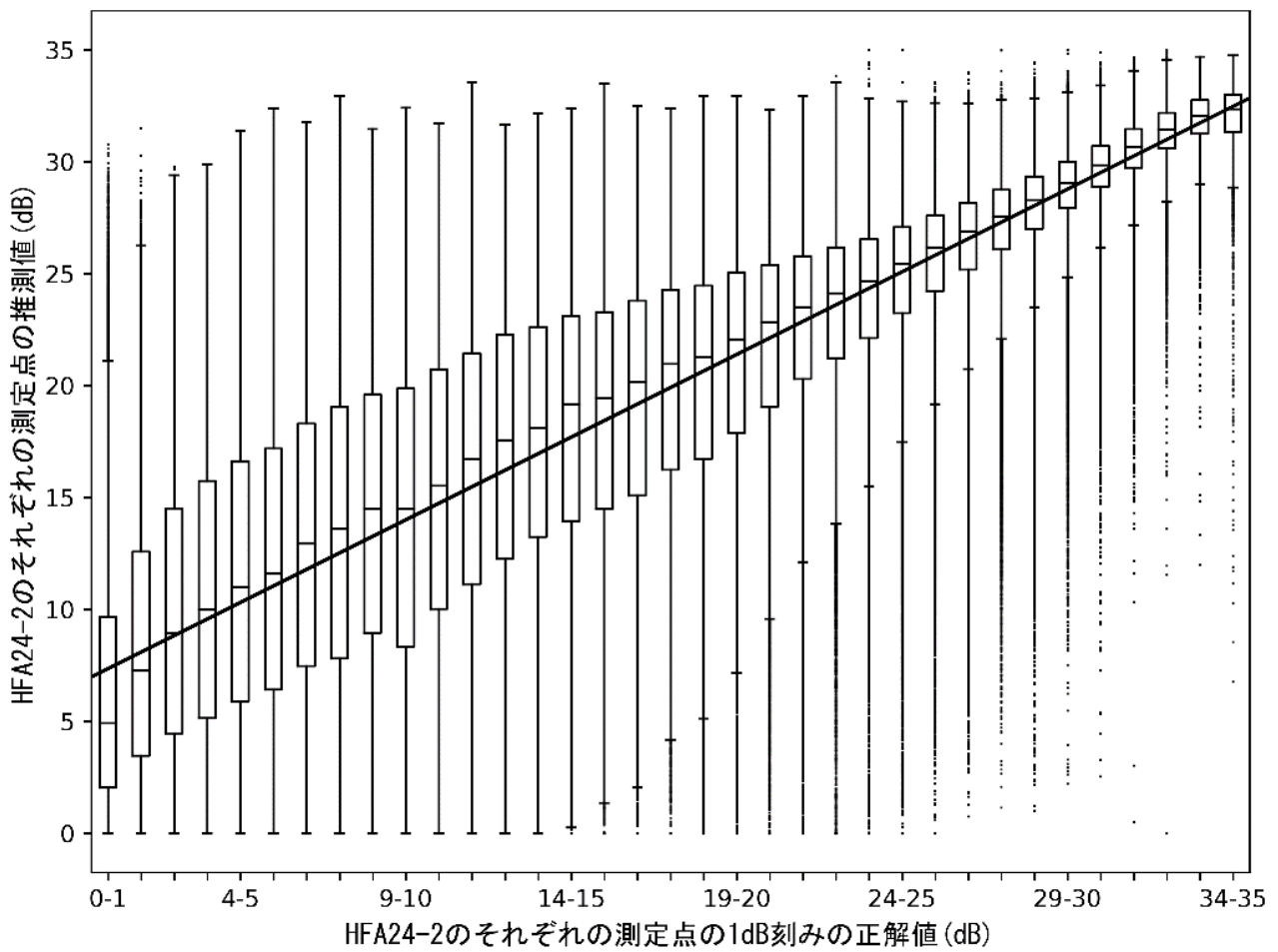


図9 OCT からの HFA24-2視野の測定点毎の推測値と正解値のグラフ。

高い相関が得られた。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = 0.814$ 、

ピアソンの積率相関係数  $p < 0.001$ ,  $r = 0.848$  )

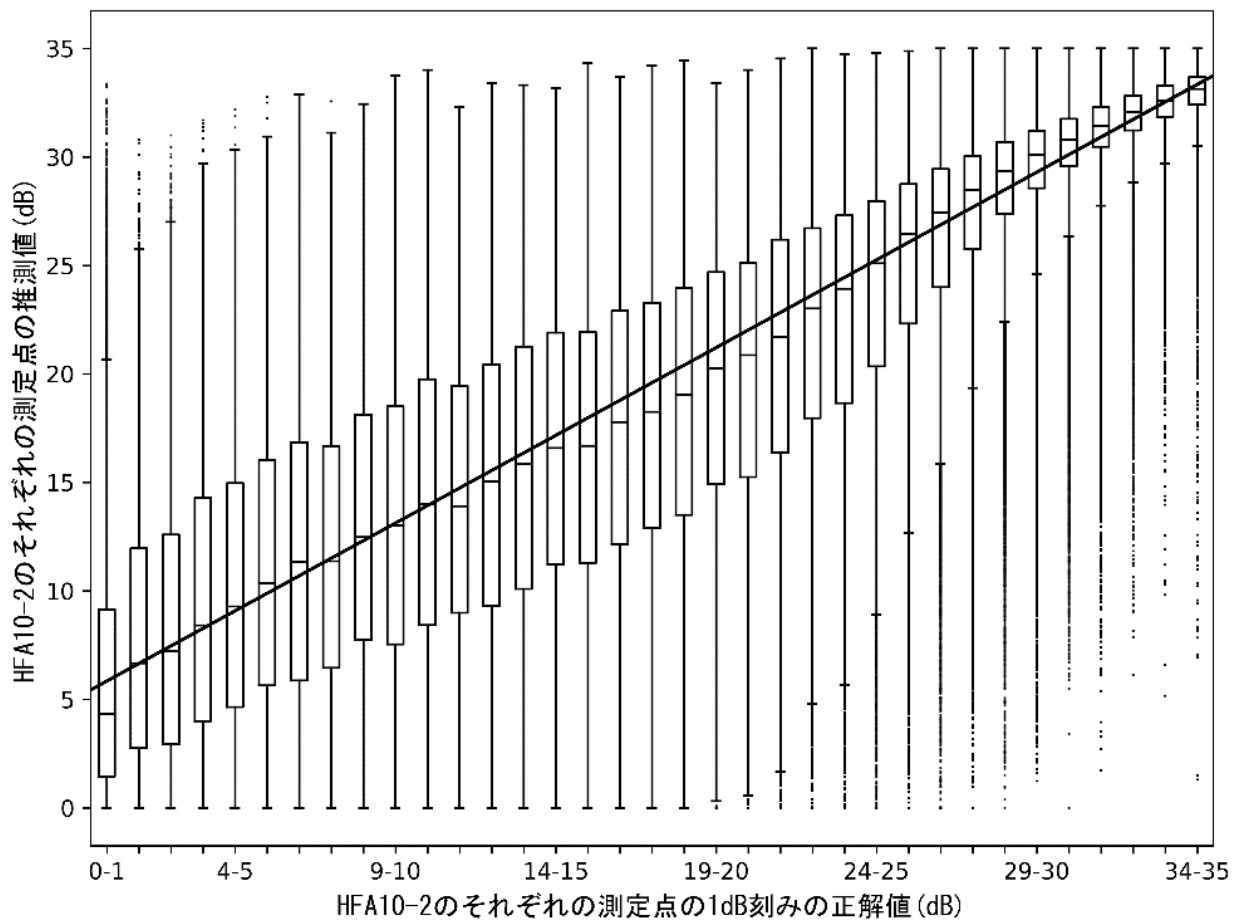


図10 OCT からの HFA10-2視野の測定点毎の推測値と正解値のグラフ。

HFA24-2同様、高い相関が得られた。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = 0.781$ 、

ピアソンの積率相関係数  $p < 0.001$ ,  $r = 0.882$  )

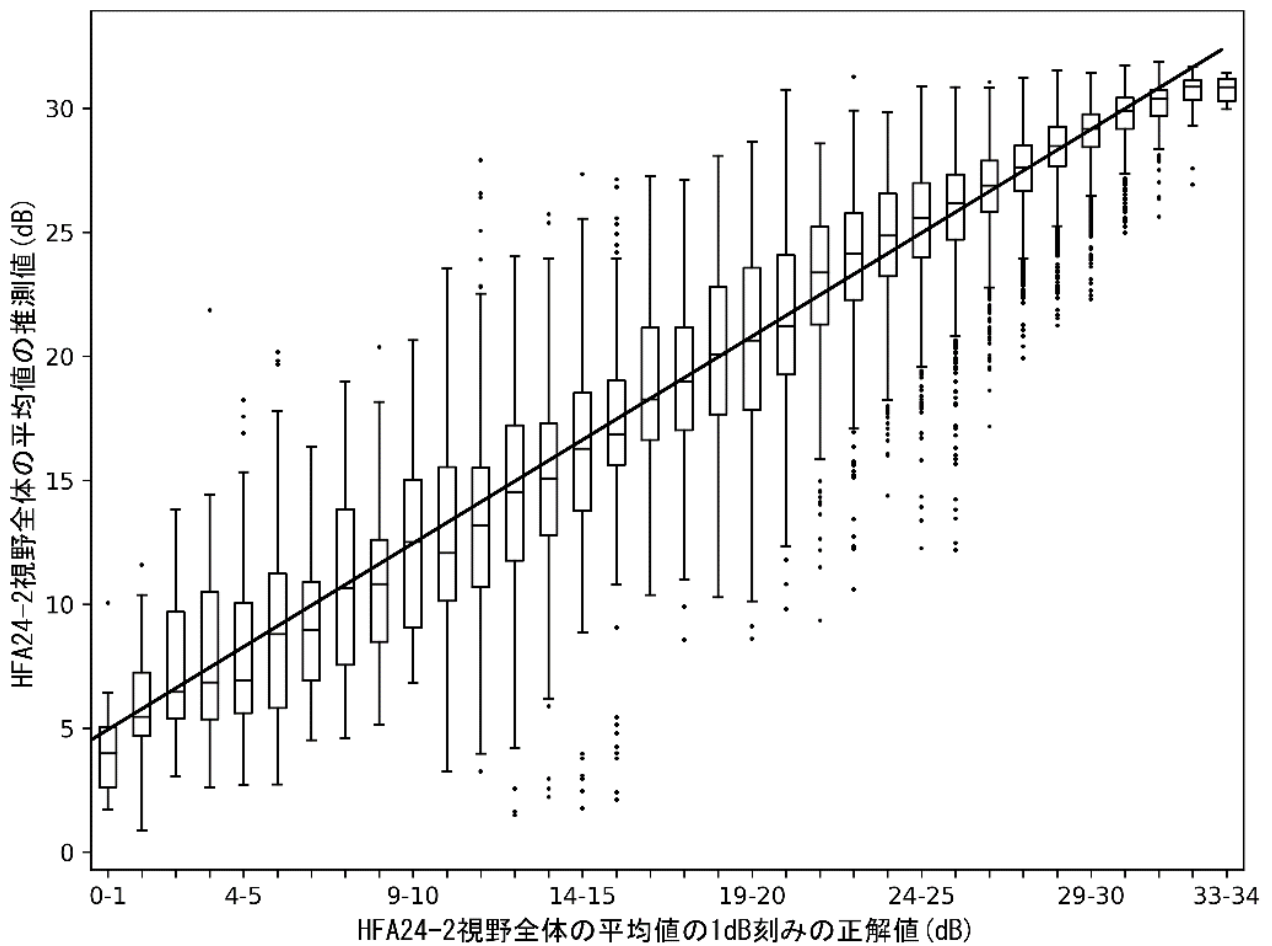


図11 OCTからのHFA24-2平均視野の推測値と正解値のグラフ。

極めて高い相関が得られた。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = 0.844$ 、

ピアソンの積率相関係数  $p < 0.001$ ,  $r = 0.899$ )

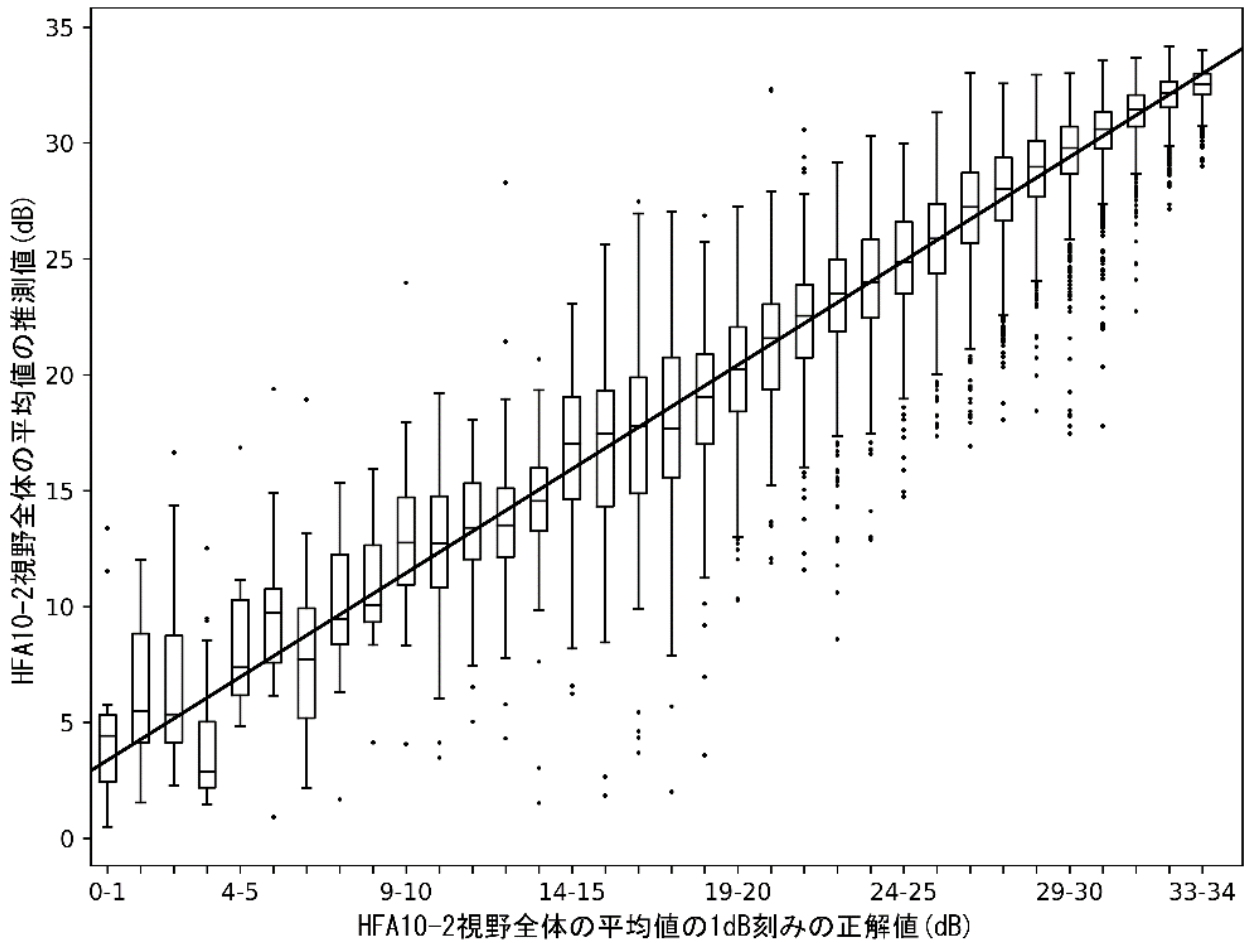


図12 OCTからのHFA10-2平均視野の推測値と正解値のグラフ。

HFA24-2同様、極めて高い相関が得られた。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = 0.890$ 、  
ピアソンの積率相関係数  $p < 0.001$ ,  $r = 0.930$ )

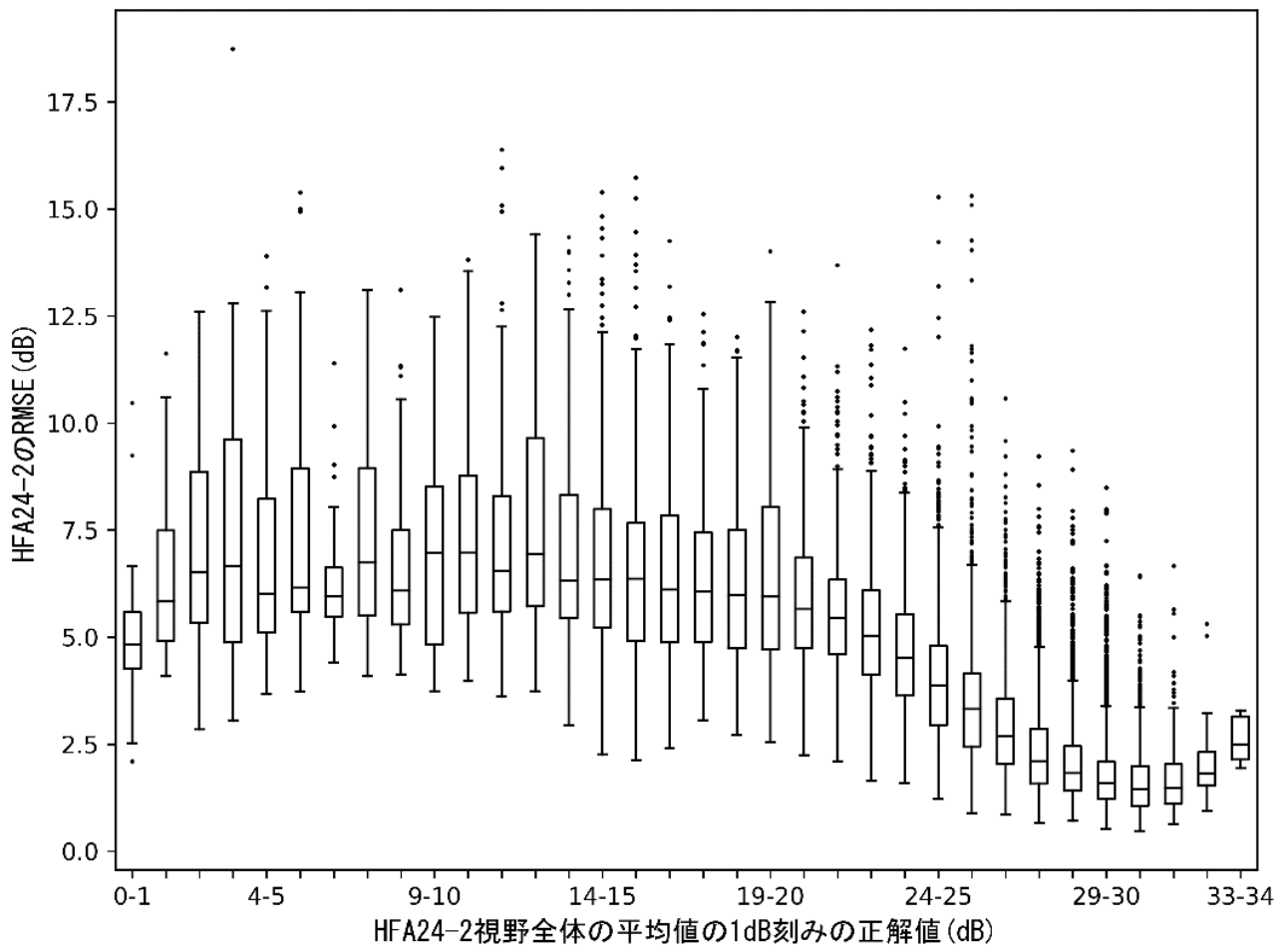


図13 HFA24-2視野全体の正解値と RMSE のグラフ。

平均視野が悪いほど推測が困難であった。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = -0.763$ )

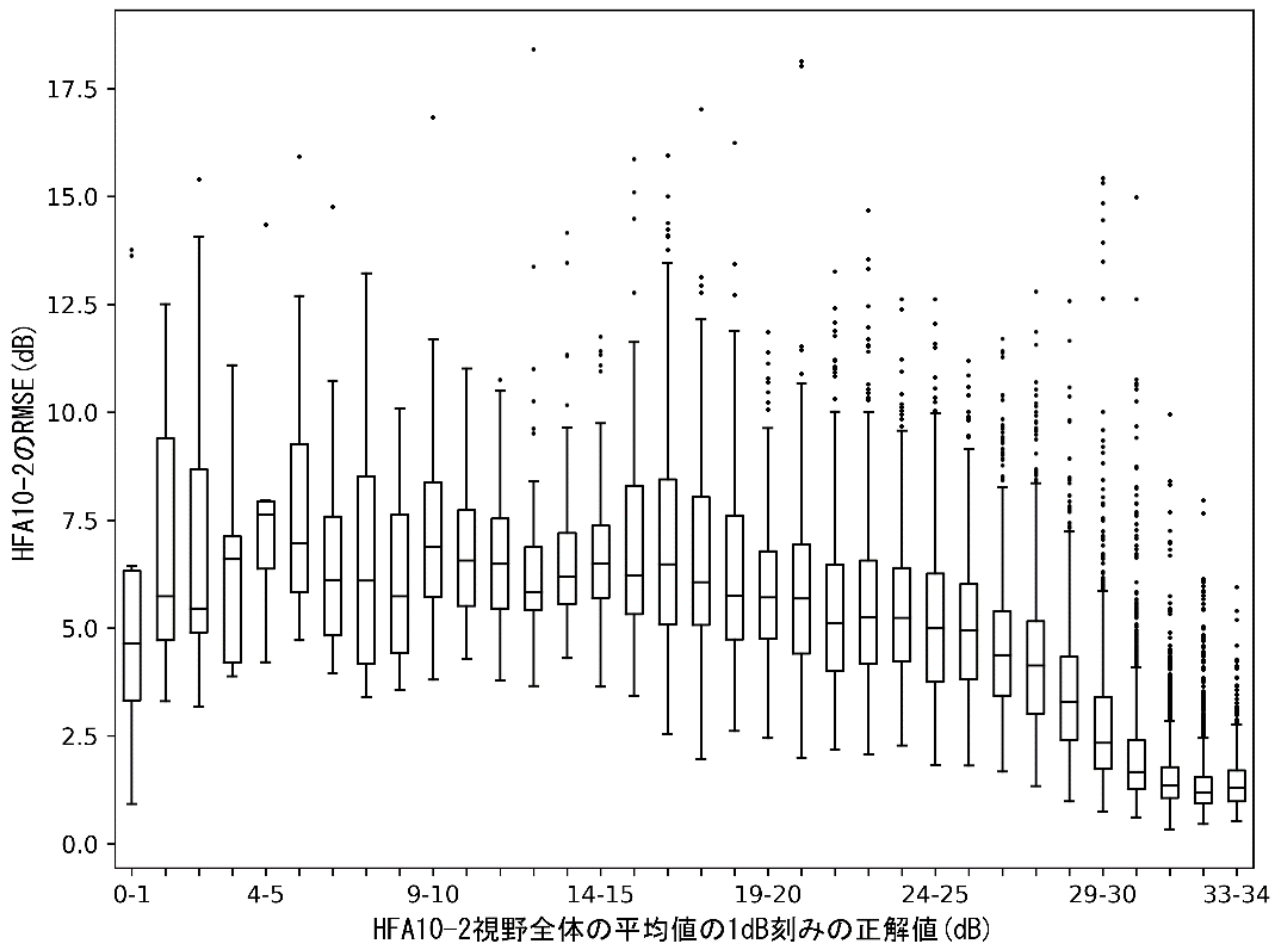


図14 HFA10-2視野全体の正解値と RMSE のグラフ。

HFA24-2同様、平均視野が悪いほど推測が困難であった。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = -0.797$ )

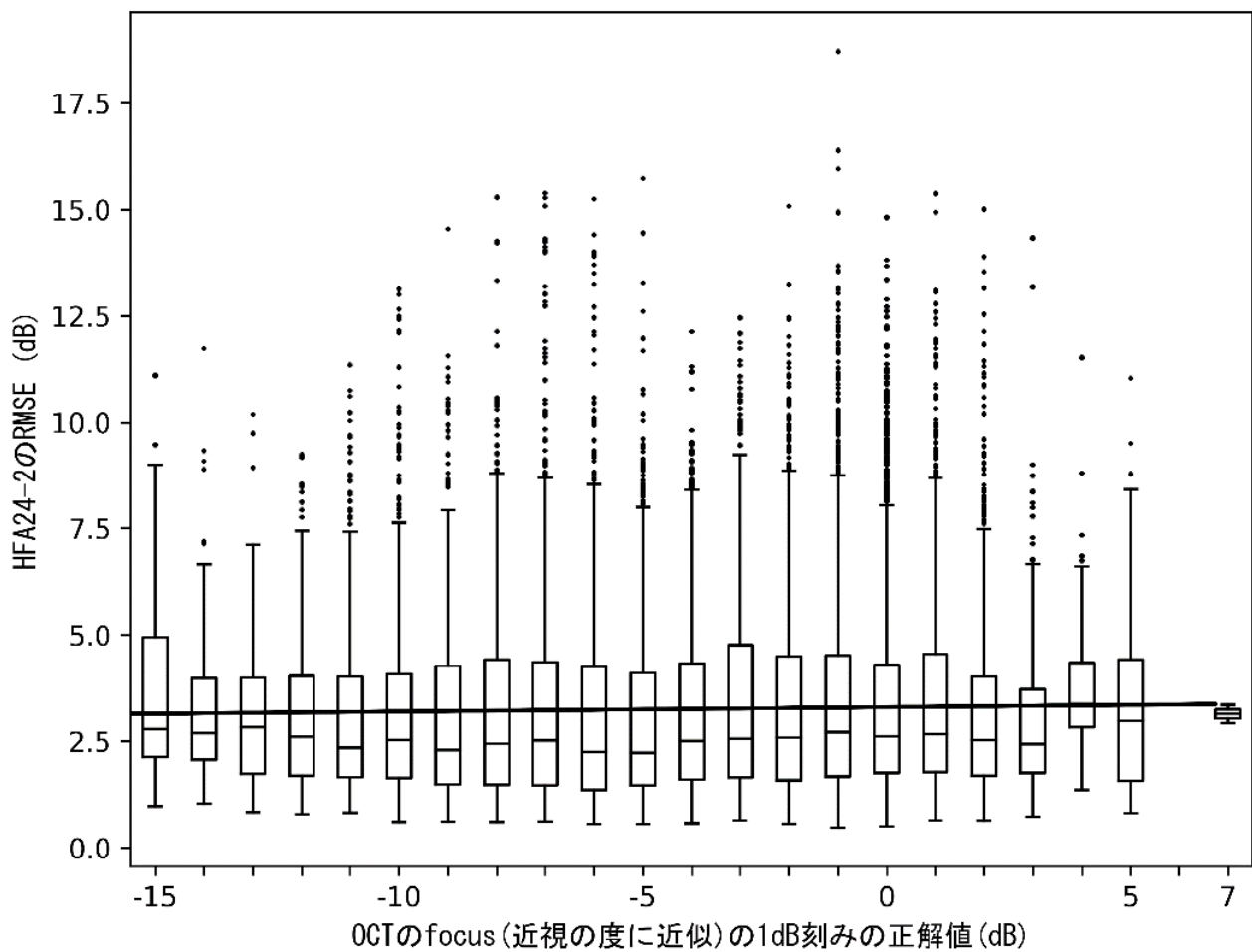


図15 OCTのfocus (近視の度に近似)とHFA24-2のRMSEのグラフ。

近視が強いほど高い推測精度が得られた。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = 0.0430$ )

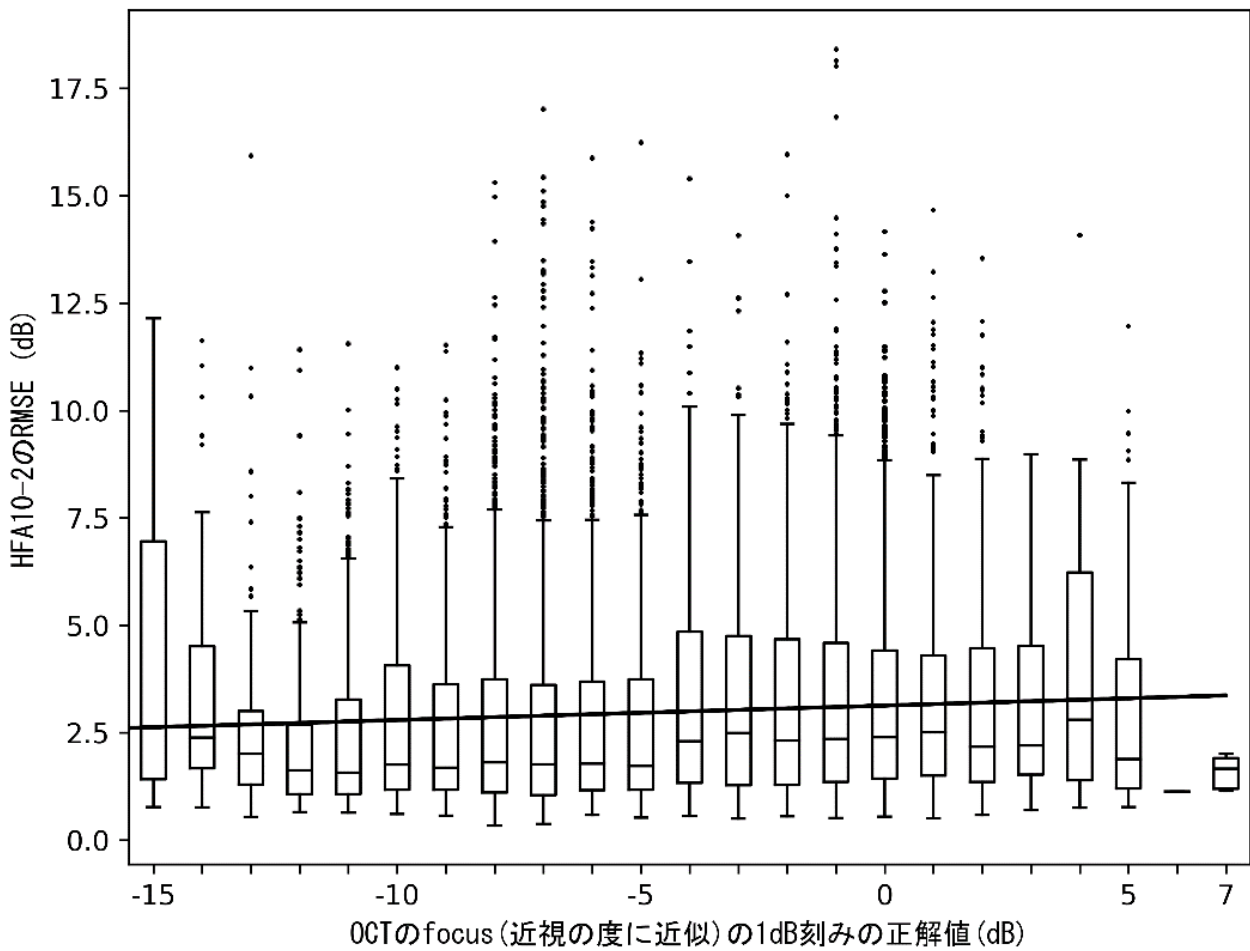


図16 OCTのfocusとHFA10-2のRMSEのグラフ。

HFA24-2同様、近視が強いほど高い推測精度が得られた。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = 0.102$ )



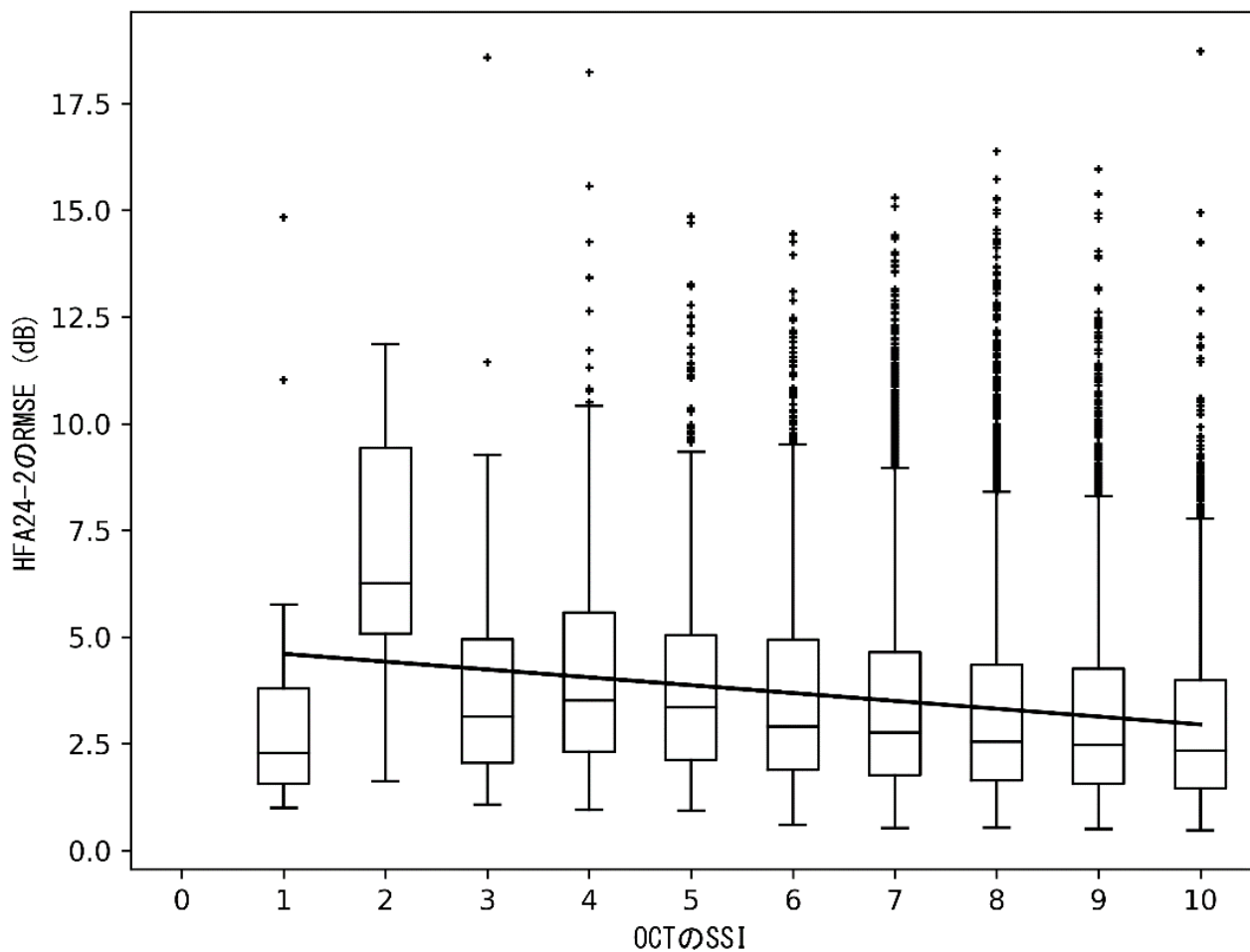


図17 OCTの全てのSSI（信号強度指数）とHFA24-2のRMSEのグラフ。

SSIが低い（=ノイズが乗ったOCT）ほど推測が困難であった。

（スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = -0.115$ ）

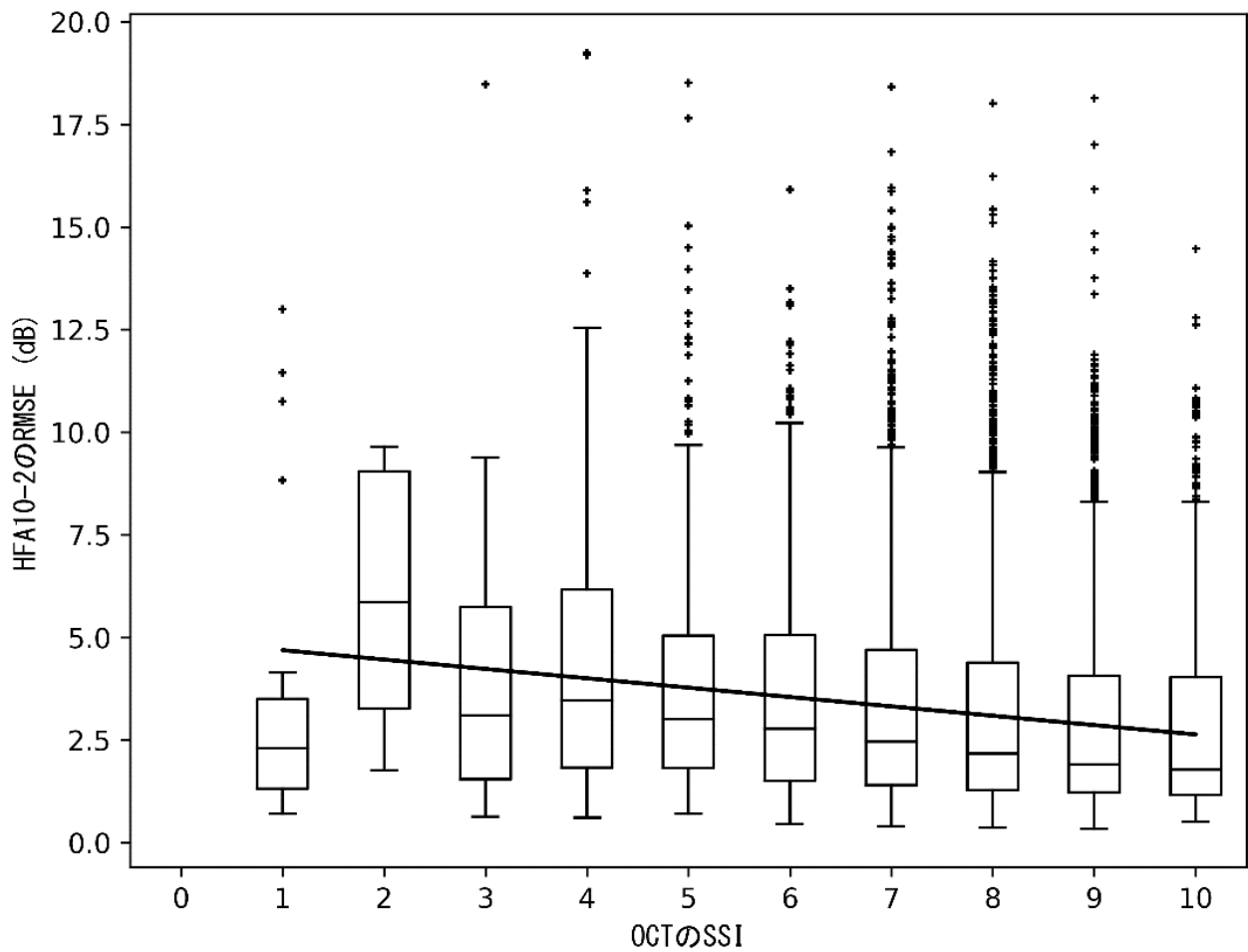


図18 OCTの全てのSSIとHFA10-2のRMSEのグラフ。

HFA24-2同様、SSIが低いほど推測が困難であった。

(スピアマンの順位相関係数  $p < 0.001$ ,  $\rho = -0.128$ )

<p>全ての群で学習対象から除外した疾患一覧  (主に中枢疾患か、もしくは視野が正しく計測できなかった症例) 合計445眼</p> <ul style="list-style-type: none"> <li>・両耳側半盲 89眼</li> <li>・同名1 / 4 盲 29眼</li> <li>・パーキンソン病 2眼</li> <li>・水頭症 2眼</li> <li>・頭蓋内出血 1眼</li> <li>・認知症 1眼</li> <li>・上眼瞼の視野欠損のある症例 139眼</li> <li>・レンズの枠の暗点のある症例 72眼</li> <li>・心因性視野障害、もしくは眼底・OCT 所見と矛盾する視野の症例 110眼</li> </ul>
--

表1 全ての群で学習対象から除外した疾患・症例一覧。

学習対象に含めるかどうかを、2群に分けて検討した疾患一覧（主に眼内・視神経疾患）

合計 395眼

- ・黄斑上膜 96眼
- ・糖尿病黄斑症(severe NPDR 以上) 65眼
- ・加齢黄斑変性 or 近視性黄斑症 53眼
- ・視神経（部分）低形成 43眼
- ・網膜静脈（分枝）閉塞症 37眼
- ・網脈絡膜萎縮 25眼
- ・非緑内障性視神経萎縮 21眼
- ・網膜動脈（分枝）閉塞症 20眼
- ・網膜色素変性症 13眼
- ・網膜剥離術後 6眼
- ・黄斑円孔 5眼
- ・脈絡膜炎 2眼
- ・Stargardt 病 2眼
- ・腎性網膜症 2眼
- ・多発消失性白点症候群(MEWDS) 1眼
- ・視神経炎 1眼
- ・急性帯状潜在性網膜外層症(AZOOR) 1眼
- ・中心性漿液性網脈絡膜症の光凝固術術後 1眼
- ・エタンブトール性視神経障害 1眼

表2 学習対象に含めるかどうかの疾患一覧。単一眼に複数の疾患を合併した場合、その中の主な疾患1つのみを記載した。

学習対象疾患	眼内疾患含有	眼内疾患除外	眼内疾患含有	眼内疾患除外
SSI の条件	全 SSI		SSI を7以上に制限	
ペアデータ総数	41,420	38,368	36,066	33,660
全患者数	3,416	3,274	3,286	3,156
内 HFA24-2の患者数	3,324	3,185	3,201	3,073
内 HFA10-2の患者数	1,241	1,171	1,208	1,142
全眼数	6,356	5,961	6,011	5,658
内 HFA24-2の眼数	6,153	5,778	5,818	5,484
内 HFA10-2の眼数	2,267	2,079	2,177	2,001
HFA24-2の平均閾値(dB)	26.5±6.91	26.7±6.68	26.8±6.74	27.0±6.53
HFA10-2の平均閾値(dB)	27.8±8.30	28.1±8.13	28.1±8.16	28.4±8.02
HFA24-2の閾値スロープ 平均(dB/年)	-0.816±1.03	-0.783±1.00	-0.816±1.03	-0.783±1.00
HFA10-2の閾値スロープ 平均(dB/年)	-1.22±1.34	-1.18±1.33	-1.22±1.34	-1.18±1.33
HFA24-2の 平均 MD 値(dB)	-3.59±5.23	-3.34±5.04	-3.34±5.05	-3.12±4.89
HFA10-2の 平均 MD 値(dB)	-4.54±5.69	-4.31±5.54	-4.27±5.54	-4.08±5.42
HFA24-2の 平均 MD スロープ (dB/年)	-0.312±0.418	-0.294±0.411	-0.312±0.418	-0.294±0.411
HFA10-2の 平均 MD スロープ(dB/年)	-0.482±0.508	-0.454±0.460	-0.482±0.508	-0.454±0.46
平均 SSI	7.93±1.34	7.97±1.33	8.29±0.97	8.31±0.972
focus(≒屈折の度)(D)	-2.55±3.79	-2.62±3.82	-2.56±3.75	-2.63±3.77
年齢(歳)	63.7±13.1	63.2±13.1	62.8±13.1	62.3±13.1
HFA24-2の閾値スロープの 教師値の採用率 (%)	5.52	4.91	5.52	4.91
HFA10-2の閾値スロープの 教師値の採用率 (%)	4.00	3.56	4.00	3.56
HFA24-2の MD スロープの 教師値の採用率 (%)	20.4	18.4	20.4	18.4
HFA24-2の MD スロープの 教師値の採用率 (%)	13.5	12.3	13.5	12.3

表3 HFA24-2, HFA10-2の学習対象の母集団及び HFA10-2の評価対象の母集団。±は標準偏差。  
HFA24-2 の場合、最終 HFA24-2 が Anderson-Patella の基準に該当する眼のみを評価対象としたため、学習時(表3)と評価時(表4)を分けて記載した。教師値の採用率の計算は、HFA24-2 全体もしくは 10-2 全体のそれぞれを分母として行った。

評価対象疾患	眼内疾患含有	眼内疾患除外	眼内疾患含有	眼内疾患除外
SSI 条件	全 SSI		SSI を7以上に制限	
HFA24-2ペアデータ数	27,712	25,090	23,678	21,633
HFA24-2の患者数	2,407	2,234	2,290	2,130
HFA24-2の眼数	3,569	3,220	3,340	3,030
HFA24-2の平均閾値 (dB)	25.1±7.70	25.3±7.52	25.4±7.59	25.6±7.42
HFA24-2の閾値スロープ 平均(dB/年)	-0.922±1.06	-0.891±1.04	-0.922±1.06	-0.891±1.04
HFA24-2の 平均 MD 値(dB)	-4.93±5.72	-4.67±5.58	-4.67±5.58	-4.45±5.46
HFA24-2の 平均 MD スロープ(dB/年)	-0.393±0.441	-0.375±0.437	-0.393±0.441	-0.375±0.437
平均 SSI	7.83±1.35	7.87±1.34	8.23±0.962	8.25±0.965
focus(≒屈折の度)(D)	-2.53±3.9	-2.62±3.94	-2.56±3.86	-2.63±3.89
年齢(歳)	65.6±12.3	65.1±12.3	64.7±12.5	64.2±12.4
HFA24-2の閾値スロープの 教師値の採用率 (%)	4.78	4.20	4.78	4.20
HFA24-2の MD スロープの 教師値の採用率 (%)	15.8	14.0	15.8	14.0

表4 HFA24-2の評価対象の母集団。教師値の採用率の計算は、HFA24-2全体を分母として行った。最終 HFA24-2が Anderson-Patella の基準に該当する眼のみを評価対象とした。

モデル名				モデル1	モデル2	モデル3	モデル4
学習時 条件	学習時に SSI を 7 以上に制限したか			全 SSI		SSI7以上	
	学習時に眼内疾患を対象に含めたか			含有	除外	含有	除外
HFA24-2 評価時 条件	眼内疾患 除外	SSI 7以上	測定点毎の ピアソンの 積率相関係数	<b>r=0.853</b> p<0.001	r=0.851 p<0.001	r=0.848 p<0.001	r=0.849 p<0.001
			RMSE(dB)	<b>3.23</b> ±2.17	3.26 ±2.14	3.28 ±2.18	3.28 ±2.18
	SSI 7未満	RMSE(dB)	<b>3.73</b> ±2.31	3.80 ±2.39	3.90 ±2.42	3.90 ±2.43	
	眼内疾患 含有	SSI 7以上	RMSE(dB)	<b>3.35</b> ±2.24	3.39 ±2.23	3.40 ±2.25	3.42 ±2.27
HFA10-2 評価時 条件	眼内疾患 除外	SSI 7以上	測定点毎の ピアソンの 積率相関係数	<b>r=0.881</b> p<0.001	r=0.880 p<0.001	<b>r=0.881</b> p<0.001	r=0.879 p<0.001
			RMSE(dB)	<b>3.01</b> ±2.33	3.05 ±2.32	3.04 ±2.29	3.05 ±2.33
	SSI 7未満	RMSE(dB)	<b>3.73</b> ±2.75	3.76 ±2.82	3.87 ±2.85	3.83 ±2.81	
	眼内疾患 含有	SSI 7以上	RMSE(dB)	<b>3.11</b> ±2.41	3.15 ±2.40	3.13 ±2.37	3.15 ±2.41

表 5 5分割交差検証の学習結果。学習時及び評価時の条件を変え、眼内疾患の除外の有無及び信号強度指数(SSI)の制限の有無により比較した。二乗平均平方根誤差 (RMSE) はペアデータ毎に計算した。RMSE は小さいほど良く、ピアソンの r は大きいほど良い。太字は最良値。HFA24-2は最終視野が Anderson-Patella の基準に該当する眼のみを評価対象とした。

条件	眼内疾患	SSI	評価項目	HFA24-2	HFA10-2
視野の閾値	眼内疾患 除外	SSI7 以上	測定点毎のピアソンの 積率相関係数	r=0.848 p<0.001	r=0.882 p<0.001
			測定点毎のスピアマンの順位相 関係数	$\rho$ =0.814 p<0.001	$\rho$ =0.781 p<0.001
			RMSE(dB)	3.27 ±2.20	3.02 ±2.28
			測定点毎の MAE(dB)	2.46 ±1.73	2.24 ±1.66
			視野全体の平均の MAE(dB)	1.62 ±1.70	1.39 ±1.46
			RMSE(dB)	3.81 ±2.43	3.72 ±2.72
	SSI7 未 満	測定点毎の MAE(dB)	2.96 ±2.03	2.85 ±2.13	
		視野全体の平均の MAE(dB)	2.01 ±2.02	1.86 ±1.94	
		RMSE(dB)	3.40 ±2.28	3.13 ±2.37	
	眼内疾患 含有	SSI7 以上	測定点毎の MAE(dB)	2.55 ±1.79	2.32 ±1.74
			視野全体の平均の MAE(dB)	1.68 ±1.76	1.46 ±1.56
			ピアソンの 積率相関係数	r=0.903 p<0.001	r=0.923 p<0.001
MD 値	眼内疾患 除外	SSI7 以上	スピアマンの 順位相関係数	$\rho$ =0.804 p<0.001	$\rho$ =0.844 p<0.001
			MAE(dB)	1.62 ±1.74	1.44 ±1.55
			測定点毎のピアソンの 積率相関係数	r=0.431 p<0.001	r=0.438 p<0.001
視野の 閾値の 進行速度	眼内疾患 除外	SSI7 以上	測定点毎のスピアマンの順位相 関係数	$\rho$ =0.521 p<0.001	$\rho$ =0.525 p<0.001
			測定点毎の MAE(dB)	0.594 ±0.738	0.819 ±0.900
			ピアソンの 積率相関係数	r=0.429 p<0.001	r=0.383 p<0.001
MD 値の進 行速度	眼内疾患 除外	SSI7 以上	スピアマンの 順位相関係数	$\rho$ =0.496 p<0.001	$\rho$ =0.481 p<0.001
			MAE(dB)	0.257 ±0.302	0.314 ±0.311
			測定点毎のピアソンの 積率相関係数	r=0.429 p<0.001	r=0.383 p<0.001

表6 10分割交差検証の評価結果。視野の閾値及び MD 値に関し、極めて高い推測性能が得られた。HFA24-2は最終視野が Anderson-Patella の基準に該当する眼のみを評価対象とした。