

ポスター | 画像情報・生体信号処理

ポスター6

画像情報・生体信号処理

2019年11月23日(土) 14:40 ~ 15:40 ポスター会場1 (国際展示場 展示ホール8)

[3-P1-3-01] 少数カラー眼底写真による緑内障推定精度の検証

○広田 雅和¹、井上 賢治²、三村 達哉³、林 孝雄¹、澤 智博⁴、溝田 淳³ (1. 帝京大学医療技術学部視能矯正学科, 2. 井上眼科病院, 3. 帝京大学医学部眼科学講座, 4. 帝京大学医療情報システム研究センター)

キーワード : Ophthalmology, Glaucoma, Deep learning, Fundus photography

【目的】眼科領域における畳み込みニューラルネットワーク(CNN)を用いた眼疾患の推定は、数千例以上の症例を同一機器で測定したデータを学習させていることが多く、少数のデータでCNNの性能評価を行った例や眼底カメラの機種別で予測精度検証した報告は少ない。本研究では少数のカラー眼底写真を用いたCNNによる緑内障眼の予測精度および別機種の眼底カメラで撮影した眼底写真における緑内障眼の予測精度を検証した。

【方法】実験1では、緑内障患者 130 眼、健常者 130 眼のカラー眼底写真 (Topcon) を訓練データ各90枚、検証データ各40枚に振り分けた。視神経乳頭が含まれる 512×512 px を抽出し、128×128 px に縮小した。CNN モデルは VGG16 を改造した。学習回数は 100 回とした。元画像のみを使用した場合を対照として、訓練用画像を各 3000枚に水増しした場合(DAC)の受信者動作特性曲線下面積(AUC)を比較した。水増ししたカラー画像をグレースケール画像に変換した場合(DAG)と DACの AUCを比較した。実験2では、別機種 (Nidek) で撮影した緑内障眼の眼底写真79枚を DAGと DACモデルに予測させた。

【結果】実験1 では、DAC(0.751)は対照(0.607)よりも AUCが有意に高かった($P=0.024$)。DAG(0.796)は DACよりも AUCが有意に高かった($P=0.031$)。実験2では、DAG(0.304)の正解率は DAC(0.797)よりも有意に低かった($P<0.001$)。

【結論】少数の緑内障眼のカラー眼底写真を深層学習に使用する際は、画像の水増しとグレースケール化を組み合わせると判定精度が向上することが示唆された。一方で、予測精度は眼底カメラの機種に影響される可能性があることが示唆された。

少数カラー眼底写真による緑内障推定精度の検証

広田 雅和^{*1}、井上 賢治^{*2}、
三村 達哉^{*3}、林 孝雄^{*1,2}
澤 智博^{*4}、溝田 淳^{*2}

*1 帝京大学 医療技術学部 視能矯正学科、*2 帝京大学 医学部 眼科学講座、
*3 井上眼科病院、*4 帝京大学 医療情報システム研究センター

Evaluation of Accuracy in Glaucoma Identification with Small Sample Size using a Convolutional Neural Network

Masakazu Hirota^{*1}, Kenji Inoue^{*2}, Tatsuya Mimura^{*3}
Takao Hayashi^{*1}, Tomohiro Sawa^{*4}, Atsushi Mizota^{*3}.

*1 Department of Orthoptics, Teikyo University Faculty of Medical Technology, *2 Inoue Eye Hospital, *3 Department of Ophthalmology, Teikyo University School of Medicine, *4 Teikyo University Medical Information Systems Research Center.

The purpose of the present study was to evaluate the accuracy of the glaucoma identification with data augmentation and gray conversion using fundus photographs with a small sample size using a convolutional neural network (CNN). The dataset was prepared using color fundus photographs captured with two fundus cameras (VX-10i with CCD of KD-211C and D200) which have resolution of 1604 × 1216 and 3872 × 2592 px. The training dataset consisted of 180 images, and the validation dataset contained 80 images. The color channels were converted into 8-bit grayscale. All images were resized to size of 128 × 128 px and were augmented to 3000 images. CNN model used fine-tuned with VGG16. The diagnostic performance was significantly higher in data augmentation with color images (DAC-model) than in the none of data augmentation ($P = 0.024$). Furthermore, the diagnostic performance was significantly higher in the data augmentation with grayscale images (DAG-model) than in the DAC-model ($P = 0.031$). However, the diagnostic performance was significantly lower in the DAG-model than in the DAC-model in D200 ($P < 0.001$). These findings suggest that the diagnostic performance depends on data augmentation and gray conversion in the color fundus photographs using a fundus camera.

Keywords: Ophthalmology, Glaucoma, Deep learning, Fundus photography.

緒論

緑内障は全世界における不可逆的な中途失明率原因第一位の疾患である。先行研究において、緑内障の有病率は3.5% (64.3 百万人)であることが報告されている¹⁾。緑内障の患者数は年々増加しており、2040 年には1 億人を超えることが予測されている。

緑内障の早期発見は中途失明の予防に繋がる。緑内障は視神経乳頭周囲が特徴的に変化するため、眼科医は直像鏡および倒像鏡を用いて視神経乳頭の形状および網膜神経繊維を観察し、緑内障を診断する²⁾。

他覚的検査として視神経乳頭周囲の構造変化はカラー眼底写真や光干渉断層計から得られた画像を評価している。カラー眼底写真は緑内障のスクリーニングに用いられている。しかしながら、カラー眼底写真からの緑内障診断は眼科医の自覚的なものであり、かつ読影には技術が必要である³⁾。

深層学習の技術である畳み込みニューラルネットワーク (CNN) は画像認識や画像分類の分野において主流となっている⁴⁾。CNN は人間の視覚野における神経のパターンを模しており、畳み込み層を含めた中間層では入力画像に対し、複数のフィルターを適応することで画像内の特徴量を抽出する。

近年、先行研究において CNN による画像分類によってカラー眼底写真から緑内障の自動判別ができたこと⁵⁾、眼科医よりも精度の高いモデルができたことが報告されている³⁾。こ

れらの先行研究は、CNN が眼疾患を他覚的に判定する有用なツールになり得ることを示唆している。しかしながら、先行研究では数千以上の症例を同一のカメラで測定したデータを学習させた CNN モデルの精度に主眼が置かれており、CNN の精度向上に繋がるデータセットの前処理や別機種のカメラで測定した眼底画像に転用可能か検証した報告は少ない。

目的

本研究では、少数のカラー眼底写真を用いた CNN による緑内障眼の予測精度、および別の撮像素子に組み替えた解像度の異なる眼底カメラで撮影した眼底写真における緑内障眼の予測精度を検証した。

方法と結果

本研究は帝京大学医学部附属病院における倫理委員会の承認を得て、すべての被験者からインフォームドコンセントを得た上で実施した (18-161)。

データセットは VX-10i (CCD, KD-211C; Kowa) で撮影したカラー眼底写真とした。すべてのカラー眼底写真は視野角 50 度、解像度 1604 × 1216 px の jpg 画像だった。緑内障例は 130 枚、健常例は 130 例だった。すべての緑内障例は単一の緑内障専門医 (K.I.) が眼科的検査結果から診断した。

訓練用 90 枚、検証用 40 枚にデータ振り分けた。各カラー眼底写真から視神経乳頭が含まれる 512×512 px を抽出し、 128×128 px にリサイズした。画像処理は OpenCV ライブラリ (Python 3.6.5 on Windows) を用いた。

CNN モデルは imagenet を学習済みの VGG16 (Oxford Univ.) を改造した^{6,7)}。学習中、VGG16 の block4 までを凍結し、block5 以降の重みを学習させた。

1. 実験 1

実験 1 では画像の水増しによる緑内障の予測精度を検証した。

各訓練用画像は、回転±5 度、水平移動 20%、垂直移動 20%、輝度倍率 0.2 – 2.0 倍の範囲内においてランダムで各パラメータが設定され jpg 画像として出力した。水増しは 180 (計 270 枚, DAC-270), 360 (計 450 枚, DAC-450), 810 (計 900 枚, DAC-900), 1440 (計 1530 枚, DAC-1530), 2910 枚 (計 3000 枚, DAC-3000) とした。水増しなし (水増しなしモデル) を対象としてモデル間の緑内障予測精度を比較した。

画像の水増し効果は、受信者動作特性 (ROC) 曲線を用いて、ROC 曲線下面積 (AUC) を計算することで評価した。統計解析には JMP 14 Pro (SAS Inc.) を用い、有意水準 5% 未満を統計学的有意差ありと定義した。

2. 実験 1 の結果

画像を水増した DAC-900 (AUC, 0.748; $P = 0.029$), DAC-1530 (0.750; $P = 0.027$), DAC-3000 (0.751; $P = 0.024$) は水増しなしモデル (AUC, 0.607) よりも有意に緑内障予測精度が高かった (図 1)。DAC-3000 は DAC-270 (0.682) よりも有意に緑内障予測精度が高かった ($P = 0.035$; 図 1)。

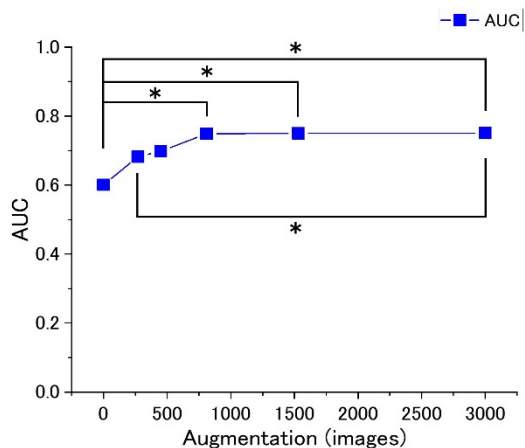


図 1 画像の水増しによる緑内障予測精度の変化
*: $P < 0.05$.

3. 実験 2

実験 2 ではカラー画像とグレースケール化した画像における緑内障予測精度を検証した。

実験 1 では画像を 3000 枚水増したとき最も良好な緑内障予測精度が得られた。DAC-3000 の学習時に使用した画像を 8 bit グレースケール化した。グレースケール画像 3000 枚で学習させたモデル (DAG-3000) とカラー画像 3000 枚で学習させたモデル (DAC-3000) に対してカラー

の検証用画像を用いて緑内障予測精度を比較した。

4. 実験 2 の結果

DAG-3000 (AUC, 0.796) は DAC-3000 (0.751) よりも有意に緑内障予測精度が高かった ($P = 0.031$; 図 2)。

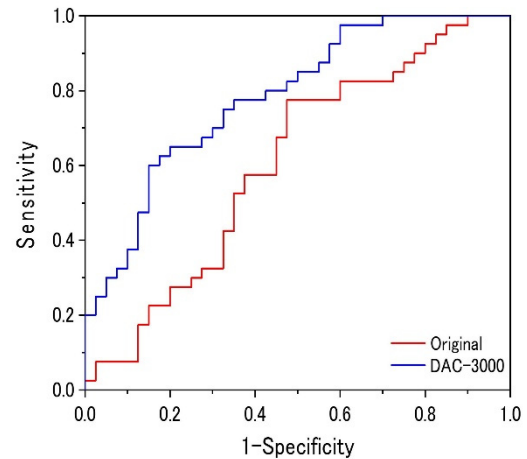


図 2 データ水増しによる緑内障予測精度の差
赤線は水増しなし、青線は 3000 枚に水増ししたときの ROC 曲線をそれぞれ示す。ROC: 受信者動作特性。

5. 実験 3

実験 3 では VX-10i の撮影用カメラを純正品の KD-211C から Nikon D200 (Nikon Corp.; 解像度, 3872×2592) に変更して撮影した緑内障例のカラー眼底写真 79 枚を用いた。各画像は実験 1 と同様、視神経乳頭周辺の 512×512 px を抽出し 128×128 px にリサイズした。DAG-3000 と DAC-3000 の緑内障予測精度を比較した。

6. 実験 3 の結果

DAG-3000 (0.304) の緑内障予測精度は DAC (0.797) よりも有意に低かった ($P < 0.001$; 図 3)。

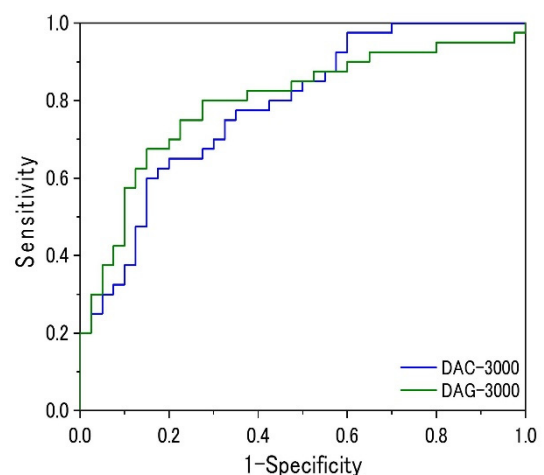


図 3 グレースケール化による緑内障予測精度の差
緑線と青線はグレースケール画像とカラー画像を 3000 枚に水増ししたときの ROC 曲線をそれぞれ示す。ROC: 受信者動作特性。

考察

本研究では少数の視神経乳頭周辺カラー眼底写真を用いた CNN による緑内障の予測精度、および別の撮像素子に組み替えた解像度の異なる眼底カメラで撮影した眼底写真における緑内障の予測精度を検証した。訓練用画像は 5 倍水増しまで緑内障予測精度が急激に増加したが、以降は水増し枚数に対する予測精度の向上が緩やかになった (図 1)。訓練用の画像をグレースケール化することで頭打ちに近かった予測精度が有意に増加した (図 2)。しかしながら、視野角が同じで解像度が異なるカメラで撮影したカラー眼底写真の予測精度はグレースケール画像で学習させたモデルのほうがカラー画像で学習させたモデルよりも有意に緑内障精度が低かった (図 3)。従って、視神経乳頭周辺画像を用いた深層学習ではグレースケール化した訓練用画像を水増しすることで予測精度を向上させることができるが、異なる解像度の画像に対する転用は困難であることが示された。

実験 1 において、訓練用画像は 5 倍水増しまで緑内障予測精度が急激に増加したが、以降は水増し枚数に対する予測精度の向上が緩やかになった (図 1a)。その理由として、オリジナルデータのバリエーションが挙げられる。緑内障は慢性進行性不可逆性の疾患である¹⁾。従って、オリジナルデータが緑内障初期から末期まで多様でなければ画像の水増しをしても、途中で予測精度が頭打ちになってしまうと考えられた。

実験 2 において、画像のグレースケール化は頭打ちになった緑内障予測精度を向上できることが示された (図 1b)。緑内障に典型的な網膜神経繊維欠損 (NFLD) はレッドフリー画像で強調できる⁸⁾。カラー画像をグレースケール化することで、NFLD の特徴を学習しやすくなったため、緑内障予測精度が向上したと考えられた。

実験 3 において、異なるカメラによるカラー眼底写真の緑内障予測精度は、グレースケール画像で学習させたモデルのほうがカラー画像で学習させたモデルよりも有意に低かった。視野角が同じで解像度のみが変更された場合、512 × 512 px の視神経乳頭探索アルゴリズムでは、視神経乳頭のみが抽出され、NFLD の情報量が減っている。グレースケール画像で学習させたモデルは視神経乳頭と NFLD の構造変化を緑内障と学習したため、視神経乳頭の情報に優先していたカラー画像で学習させたモデルよりも緑内障予測精度が低下したと考えられた。

本研究の研究限界は、視神経乳頭抽出アルゴリズムを画素単位でコーディングしたことにある。カメラの視野角と解像度、および眼軸長のデータが与えられれば、眼軸長補正を掛けた上で、抽出範囲を角度に換算できる。常に出力した画像の視野角が一定であれば異なるカメラで撮影した画像に対しても適応できると予想されるため、今後アルゴリズムを改良し、さらなる検討をしたい。

結論

同一眼底カメラで撮影した視神経乳頭周辺画像を用いた深層学習ではグレースケール化した訓練用画像を水増しすることで予測精度を向上させることができることが示された。

参考文献

1. Tham YC, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology* 2014;121(11):2081-90.

2. Chan WC, Morin JD, McCulloch C. Optic disc observations in glaucoma. *Can J Ophthalmol* 1976;11(2):134-9.
3. Shibata N, Tanito M, Mitsuhashi K, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Sci Rep* 2018;8(1):14665.
4. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. *Commun Acm* 2017;60(6):84-90.
5. Ting DSW, Cheung CYL, Lim G, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *Jama-J Am Med Assoc* 2017;318(22):2211-23.
6. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. *arXiv* 2015:arXiv:1512.03385.
7. Szegedy C, Vanhoucke V, Loffe S, et al. Rethinking the Inception Architecture for Computer Vision. *arXiv* 2015:arXiv:1512.00567.
8. Bennett TJ, Barry CJ. Ophthalmic imaging today: an ophthalmic photographer's viewpoint - a review. *Clin Exp Ophthalmol* 2009;37(1):2-13.