

딥러닝을 이용한 당뇨성황반부종 등급 분류의 정확도 개선을 위한 검증 데이터 증강 기법

이태수

충북대학교 의과대학 의공학교실, 의학연구소

Validation Data Augmentation for Improving the Grading Accuracy of Diabetic Macular Edema using Deep Learning

Tae Soo Lee

Department of Biomedical Engineering, College of Medicine and Medical Research Institute, Chungbuk National University
(Manuscript received 30 October 2018 ; revised 10 March 2019 ; accepted 13 March 2019)

Abstract: This paper proposed a method of validation data augmentation for improving the grading accuracy of diabetic macular edema (DME) using deep learning. The data augmentation technique is basically applied in order to secure diversity of data by transforming one image to several images through random translation, rotation, scaling and reflection in preparation of input data of the deep neural network (DNN). In this paper, we apply this technique in the validation process of the trained DNN, and improve the grading accuracy by combining the classification results of the augmented images. To verify the effectiveness, 1,200 retinal images of Messidor dataset was divided into training and validation data at the ratio 7:3. By applying random augmentation to 359 validation data, $1.61 \pm 0.55\%$ accuracy improvement was achieved in the case of six times augmentation (N=6). This simple method has shown that the accuracy can be improved in the N range from 2 to 6 with the correlation coefficient of 0.5667. Therefore, it is expected to help improve the diagnostic accuracy of DME with the grading information provided by the proposed DNN.

Key words: deep learning, Validation data augmentation, Diagnostic accuracy, Diabetic macular edema

1. 서 론

당뇨성망막병증(DR: diabetic retinopathy)은 당뇨병성 만성합병증 중 하나로 전 세계적으로 중요 실명 원인 중 하나이다[1]. 특히 DR이 진행되어 황반부 망막이 붓게 되면 시력 저하가 발생할 수 있으며, 이를 당뇨성황반부종(DME: Diabetic Macular Edema)이라 하는데, 당뇨병 환자 중

4.6~10%가 시력을 저해하는 심각한 수준으로 이환되는 것으로 알려져 있으나[2], 질환 초기에는 시력에 큰 영향을 미치지 않으므로 환자가 인지하지 못하며, 비가역적인 시력 저하로 진행될 수도 있기 때문에 당뇨병 진단 후 DR에 대한 조기검진과 정기검진을 받는 것이 심각한 합병증을 예방할 수 있는 가장 좋은 방법이다.

DR의 진단에 가장 중요한 검사는 안저검사이며, 최근 인공지능기술의 발달로 안저검사 촬영 결과인 망막영상을 자동진단하는 기법들이 발달하여, iDx-DR과 같은 의료기기가 등장하게 되었다. 이는 인공지능기술 중에서 심층신경망(DNN: deep neural network)을 응용한 기기로서, 수십만 장의 망막영상을 사용하여 딥러닝(deep learning) 시켜서, 개발한 치료권고 당뇨망막병증(RDR: referable diabetic retinopathy) 유무를 진단하는 알고리즘을 내장하고 있다.

Corresponding Author : Tae Soo Lee
1 Chungdae-ro, Seowon-gu, Cheongju, Chungbuk, Korea, 28644
TEL: +82-43-269-6332, +82-10-5492-6332
E-mail: tslee@chungbuk.ac.kr
Messidor database was kindly provided by the Messidor program partners (see <http://www.adcis.net/en/DownloadThirdParty/Messidor.html>).

2016년 Abramoff 등은 IDx-DR 기기를 사용하여 RDR 진단의 민감도 96.8%, 특이도 87.0%, AUC 0.98을 얻었음을 보였으며[3], 2018년에는 900명의 피험자를 대상으로 하는 전향적 임상시험환경에서 이 기기를 사용하여, 민감도 87.2%, 특이도 85%를 얻었다[4]. 이러한 진단을 위해서는 망막영상의 특성에 따른 등급 분류기법이 필수적이며, 그 정확도가 진단의 신뢰도에 결정적인 요소이므로, 이를 개선할 수 있는 기법은 의료기기의 성능 향상에 직결된다고 할 수 있다. 한편, 2018년 Heijden 등은 IDx-DR 2.0 기기를 사용하여, RDR 진단에서 민감도 68%, 특이도 86%, AUC 0.87를 얻었고, 시력을 위협하는 DR에서는 민감도 62%, 특이도 95%, AUC 0.90을 얻었음을 보고하면서, 그 과정에서 전문의 간의 진단일치도를 검증해 보았는데 그 범위가 40%~61%임을 보였는데[5], 이는 정확도가 높은 객관적 등급 분류 기법의 필요성을 강조하는 사실로 판단된다.

본 논문에서는 딥러닝의 대표적인 기법인 DNN을 이용해서 망막영상의 DME 진단을 위한 등급 분류의 정확도 개선을 목적으로 검증 데이터 증강 기법을 제안하고자 한다. 데이터 증강기법은 DNN의 딥러닝 데이터 준비 과정에서 하나의 영상을 무작위적인 좌우상하 이동, 회전, 뒤집음, 크기 조절 등을 통해 여러 영상들로 변화시킴으로써 데이터의 다양성을 확보하기 위해 기본적으로 적용하는 기법이다. 이 기법을 본 논문에서는 딥러닝을 완료한 등급 분류기의 검증 과정에서 분류하고자 하는 주어진 영상을 데이터증강을 통해 여러 영상으로 변화시키고, 그 분류결과들을 종합하여 분류기의 정확도를 개선하고자 하는 것이 기본적인 아이디어이며, 그 유효성을 검증할 뿐 아니라 데이터 증강의 횟수에 따라 개선되는 정도가 어떻게 달라지는지를 확인해 보고자 한다.

II. 재료 및 방법

1. 컴퓨팅 환경과 데이터셋

DME 등급 분류를 위한 컴퓨팅 환경은 윈도우10 운영체제이며, DNN 구성을 위해 매트랩 2018b의 딥러닝 툴박스 및 병행컴퓨팅 툴박스를 설치하였고, 이를 지원하기 위해 GTX1060 GPU를 사용하였다. 데이터셋은 프랑스 정부의 연구부 및 국방부(French Ministry of Research and Defense)의 지원에 의해 2004 TECHNO-VISION의 일환으로 수행된 Messidor 연구프로그램에 의해서 컴퓨터 지원 DR 진단 연구용으로 제작된 망막영상데이터베이스이다[6]. 1,200장의 영상으로 구성되어 있으며, 3곳의 안과에서 Topcon TRC NW6로 촬영한 것으로서, 그 해상도는 1440*960(588장), 2240*1488(400장), 2304*1536(212장) 세가지이며, 24비트 RGB 컬러영상이다. 미세동맥류 수, 출

혈수 및 신생혈관 유무에 따라서 DR 등급을 4 가지로 나누었으며, 경성 삼출액 유무 및 황반과의 거리에 따라 DME의 임상적 중요도(CS: clinical significance) 등급을 3 가지로 분류하였다. 본 연구에서는 이중에서 3등급 DME 분류를 사용하였으며, 이는 class 0(정상) 974장(81%), class 1 (non-CSME) 75장(6%), class 2 (CSME) 151장(13%)으로 각각 구성되어 있다.

2. 영상 전처리

세가지 해상도로 촬영된 망막영상들을 512x512의 해상도를 가지는 영상으로 정규화 하였으며, 검사 과정에서의 조명 등 환경 변화에 영향을 덜 받는 데이터로 변환하기 위해 영상전처리를 하였다. 전처리는 4가지 영상처리를 단계적으로 수행하는데, 먼저 RGB컬러를 CIE 1976 L*a*b* 컬러로 변환하여, 세 채널 중에서 밝기 채널(L) 영상에 대조도 제한적응형히스토그램평활화(CLAHE: contrast limited adaptive histogram equalization)를 적용하고, 이를 RGB 컬러영상으로 재변환한 후, RGB 각 채널에 메디언 필터(median filter)를 적용하여 DNN입력영상을 구하였다. CLAHE의 대조도 제한은 0.05, 창 크기는 8x8로 하였고, 영상잡음을 제거하기 위한 메디언 필터의 커널 크기는 3x3으로 하였다.

3. 심층신경망

딥러닝 대상인 DNN은 Inception v3를 사용하였으며, 이는 299x299 입력 데이터를 받아들여므로 사전학습된 DNN 중에서는 해상도가 상대적으로 높으며, 2,300 백만개의 파라미터를 사용하고 컨볼루션 레이어의 깊이가 48층으로 구성되어 있다. 이는 Gulshan 등이 RDR 진단에 활용하여 우수한 결과를 보고한 연구에서 사용한 바 있으며[7], 현재 발표된 DNN 중에서 같은 해상도를 가지는 Inception-ResNet v2 에 비해 상대적으로 정확도가 열등하나, 파라미터 수가 작은 장점이 있어서 컴퓨팅 용량과 학습시간 등을 종합적으로 고려할 때 본 연구의 목적에 최적인 것으로 판단되었다.

4. 훈련 데이터 증강

훈련 데이터의 증강을 위한 영상의 좌우상하 평행이동 화소 영역은(-30, 30), 회전의 각도 영역은(-90, 90), 크기 조절 영역은(0.5, 1.5) 로 하고 이에 추가하여 뒤집음을 적용하였으며, 이는 각각 훈련 시에 무작위로 정해지게 하였다.

5. 전이학습

사전학습된 inception v3를 사용하여 SGD (stochastic gradient descent)기법으로 전이 학습시켰으며, 학습 환경

파라미터는 최대 에폭 50, 미니배치 크기 32, L2조절치 0.0002, 모멘텀 0.9, 학습률 초기치 0.001로 했고, 10 에폭마다 학습률을 절반으로 감소시켰다. 학습용 데이터는 전체 망막영상 1,200장 중에서 70%인 841 장을 사용하였다. 여기에는 클래스 0, 1, 2의 영상이 각각 682, 53, 106장이 포함된다.

6. 검증 데이터 증강

검증 데이터는 학습데이터와는 완전히 분리된 전체 데이터의 30%인 359장의 망막영상을 사용하였으며, 여기에는 클래스 0, 1, 2의 영상이 각각 292, 22, 45장이 포함된다. 검증 데이터 증강과정에서 학습데이터가 재사용되지 않도록 특별히 유의하였다. 학습이 완료된 DNN의 정확도 개선을 위해 영상을 등급 분류할 때, 검증영상 원본을 N (2~6)가지로 무작위의 이동, 회전, 뒤집음을 적용하여 증강시켰다. N 가지의 증강 때마다 각각 다르게 변화된 영상이 분류기에 입력되면, 각 클래스에 포함될 확률 {p0, p1, p2}을 N번 출력하게 되고, 식(1)과 같이 각 클래스 단위로 평균하여 {mp0, mp1, mp2}를 계산하고, 식(2)와 같이 이들 중에서 mpj 가 최대치를 가지면 j-번째 클래스로 분류하게 하였다.

$$mp_i = (p_{i1} + p_{i2} + \dots + p_{iN}) / N, \text{ for } i=0, 1, 2 \quad \text{식 (1)}$$

식(1)에서 N은 증강회수이며, pik는 k번째 증강데이터가 i클래스에 포함될 확률이고, mpi는 그 평균을 나타낸다.

$$j = \text{maximum index of } \{mp_0, mp_1, mp_2\} \quad \text{식 (2)}$$

식(2)에서 j는 0, 1, 2 클래스의 평균 확률인 mp0, mp1, mp2 중에서 최대치를 보이는 인덱스를 나타낸다.

III. 실험 및 결과

1. 영상 전처리

그림1의 좌측 영상은 Messidor 데이터베이스의 2005 1020_64007_0100_PP 사진으로 class 2 (CSME 등급)이며, 2240×1488 화소를 가지고 있다. 우측 영상은 전처리 결과 사진으로 망막의 혈관 및 병변 부위가 보다 선명하게 처리되었음을 알 수 있다. 전처리 과정에서 밝기 부분만 변경되고, 잡음이 제거되었을 뿐 컬러 정보가 유지되는 것은 본 기법의 장점으로서, 의학영상은 컬러 정보가 임상적으로 의미가 있기 때문이다. 뿐만 아니라 이러한 전처리과정이 DNN 내에서 딥러닝 되기를 기대하기는 어려우므로 결과에 좋은 영향을 미칠 것으로 판단된다.

2. 훈련 데이터 증강

그림2는 훈련 데이터 중에서 처음 8장이 각각 데이터 증강을 통해 어떻게 변환되어 DNN에 입력되어 학습에 사용되는지를 나타내고 있다.

3. 전이학습

그림3에서 흑색점이 연결된 선은 각 에폭 수에 따른 전체 검증데이터의 정확도와 로스 값의 변화이며, 가는 실선은 iteration에 따른 미니배치 데이터 내에서 정확도와 로스 값의 변화를 나타내고, 굵은 실선은 가는 실선 값을 스무딩한 값의 변화를 나타낸다. 마지막 부분에서 미니배치의 정확도

50

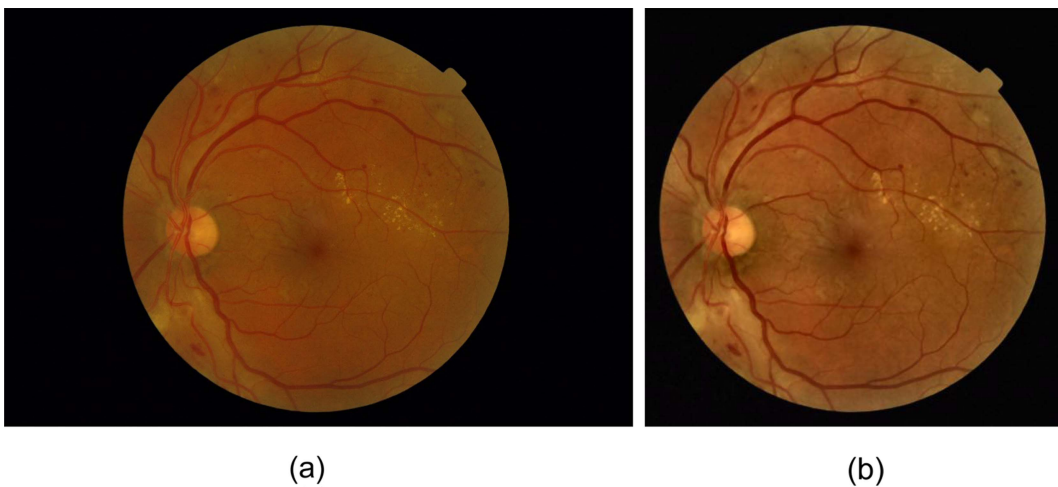


그림 1. 망막의 원 영상과 전처리 후 영상. (a) 원영상, (b) 전처리 후 영상.
Fig. 1. Typical example of original and preprocessed retinal image. (a) original image, (b) preprocessed image.

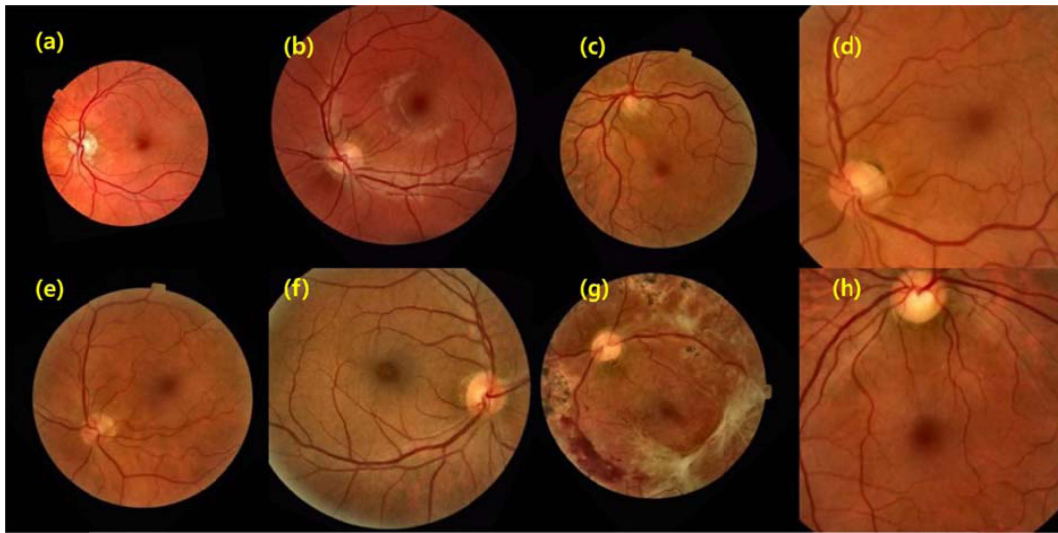


그림 2. 증강된 훈련데이터의 예. (a)-(h): 증강된 훈련데이터의 8장의 사진 예시(n) racy(%) of total and mini-batch validation data.
 Fig. 2. Typical examples of augmented training images. (a)-(h): 8 example images of augmented training data.

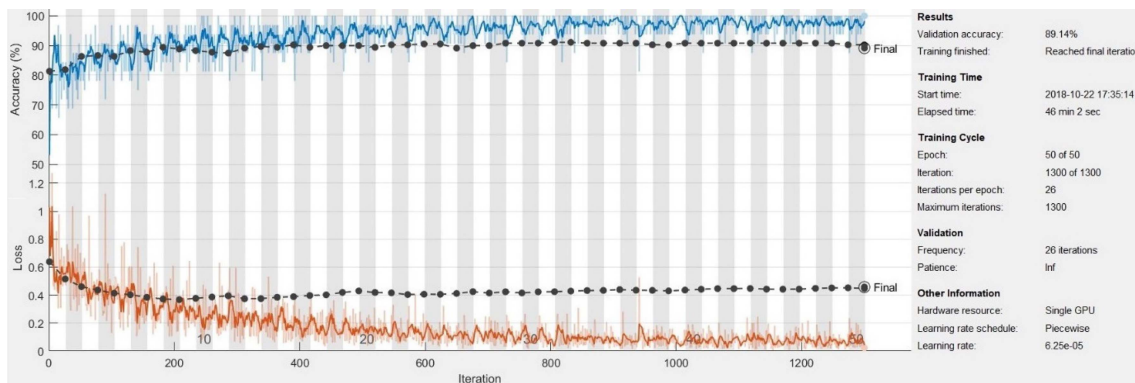


그림 3. 학습 중 정확도(좌상)와 크로스 엔트로피 로스 함수(좌하)의 변화 및 딥러닝 옵션변수들(우). x축: epoch 와 iteration, y축: 위 점선 및 실선: 전체 및 미니배치 검증데이터의 정확도(%)의 변화, 아래 점선 및 실선: 전체 및 미니배치 검증데이터의 로스 값의 변화.
 Fig. 3. Accuracy (upper left), Cross-entropy loss function (lower left), and option variables of deep learning (right). x-axis: number of epochs and iterations, y-axis: upper dotted and solid lines: accuracy (%) of total and mini-batch validation data, lower dotted and solid lines: loss values of total and mini-batch validation data.

는 100%까지 상승하고, 로스는 2.27%까지 하강하는데 비해 검증데이터의 정확도와 로스는 각각 89.14%, 44.81%인 것으로 나타나고 있어서 훈련데이터에 의한 학습은 성공적으로 되고 있지만, 검증데이터에서는 오버피팅에 의한 로스 값의 증가가 있음을 알 수 있다.

4. 검증 데이터 증강

그림4는 검증 데이터 중에서 처음 8장이 각각 데이터 증강을 통해 어떻게 변환되어 DNN에 입력되어 분류에 사용되는지를 나타내고 있다. 이 변환은 증강 때 마다 무작위로 다르게 적용되며, 그림4는 한 예를 보이고 있다.

5. 검증 데이터 증강 횟수 N에 따른 정확도 개선 결과 분석

그림 5는 N에 따라서 등급 분류기의 정확도 개선 결과의 변화를 나타내고 있다. 통계적인 오차를 줄이기 위해 각 증강 횟수 N 마다 전체 검증데이터에 10회씩 적용하여 분류 결과를 집계하였고, 이를 평균한 결과이다. 회귀분석 결과 상관계수는 0.5667, 표준 오차는 0.177e-2로 나타났다. N=6에서의 정확도 개선은 $1.61 \pm 0.55\%$ 를 보였다.

6. 오차 행렬

그림 6의 오차행렬은 N=6 일 경우 한 예를 보인 것으로서, 특이도는 98.6%, CSME민감도 71.1%, NPV(음성판정

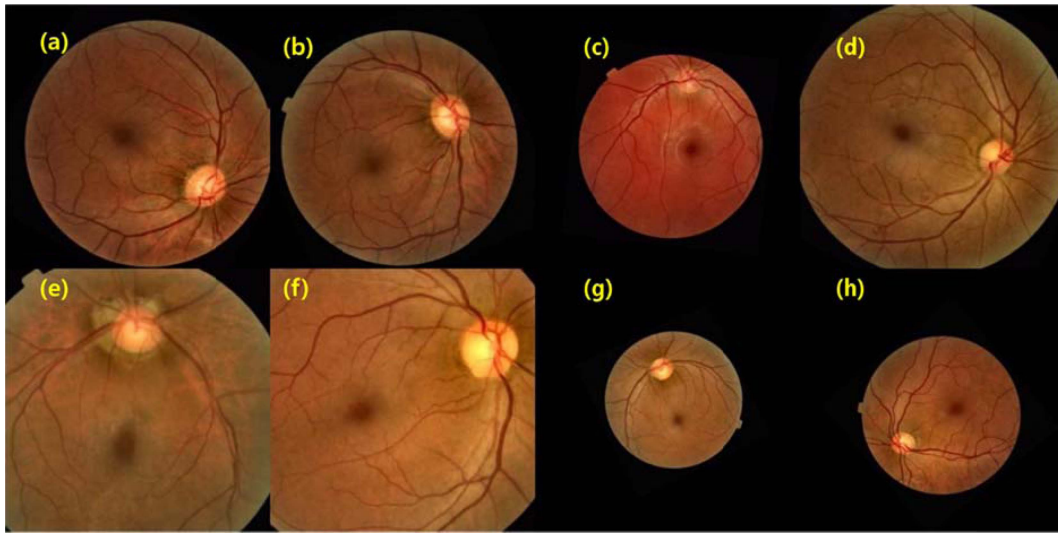


그림 4. 증강된 검증 데이터의 예. (a)-(h): 증강된 검증 데이터의 8장의 사진 예시.

Fig. 4. Typical examples of augmented validation images. (a)-(h): 8 example images of augmented validation data.

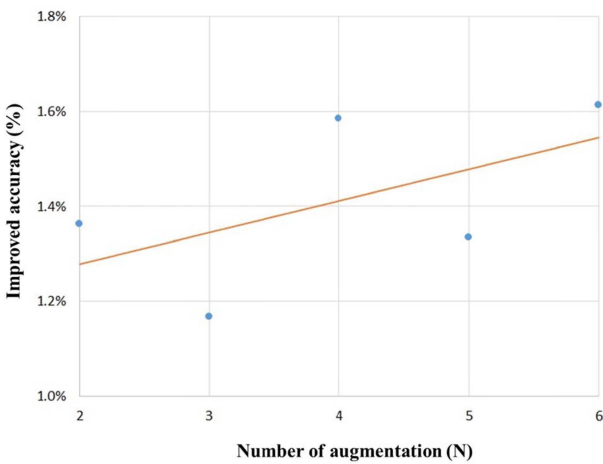


그림 5. 검증 데이터 증강 횟수와 개선된 정확도. X-축: N, Y-축: 개선된 정확도(%).

Fig. 5. Number of augmentation of validation data and the improved accuracy. X-axis: N, Y-axis: improved accuracy (%).

예상적중율) 92.9%, CSME-PPV(양성판정예상적중율) 91.4%, 전체 정확도 91.1%로 나타났다. 여기서 class 1인 non-CSME 영상은 class 0인 정상과의 구분이 어렵고 혼동되고 있는 것을 알 수 있다. 이는 DR 등급 분류가 안과 전문의들 간 또는 동일인도 일치율이 60% 정도로 낮은 것으로 보고되고 있는 점을 감안하고[5], 인공지능이 사용하고 있는 데이터의 정답 신뢰도까지를 고려해 보면 이해할 수 있는 결과인 것으로 판단된다. DME 진단 측면에서 그림 6의 데이터로 계산된 진음성(TN), 위음성(FN), 진양성(TP),

위양성(FP)은 각각 288, 4, 45, 22이며, 이로부터 추정되는 DME 진단 정확도는 92.76%이다. 이 결과는 2016년 Liverpool대학의 Al-Bander 등이 Messidor 데이터셋을 사용하고 자체적으로 개발한 DNN을 적용해서 얻은 88.8%의 DME 진단 정확도 보다 3.96% 우수한 결과이다[8].

7. 수신자판단특성곡선

분류기에서는 각 클래스에 포함될 확률이 출력되므로 그림 7과 같이 각 클래스마다 수신자판단특성(ROC: receiver operating characteristic) 곡선을 나타낼 수 있으며, 분류기의 성능 지표인 AUC (area under curve) 값을 계산할 수 있다. 각각은 $AUC_0 = 0.9355$, $AUC_1 = 0.8840$, $AUC_2 = 0.9241$ 이었고, class 1의 AUC가 최소값을 나타내어 분류 성능이 다른 클래스에 비해 열등하다는 것을 알 수 있다.

8. Cohen's Kappa

분류기의 정량적 성능 지표 중의 하나인 Cohen's Kappa 값을 계산해 보았다. 그림 6의 오차행렬에 나타난 수치로 계산된 Quadratic weighted Cohen's Kappa 값이 0.7958 이었다. DME등급 분류기의 Kappa값을 보이는 보고는 아직 없고, 망막영상진단분야에서 Kaggle 경연대회 DR 등급 분류 결과가 공개되어 있으므로 이와 비교해 보았다. Kaggle 경연대회 우승 팀의 DR등급 분류기의 Kappa 값은 0.8495에 비해 열등하나, 등록된 목록 중에서 2% 내에 포함되는 우수한 분류기인 것으로 나타났다[9].



그림 6. N=6 에서 개선된 DME 등급 분류기의 오차 행렬.
Fig. 6. Confusion matrix of DME grading classifier for N=6.

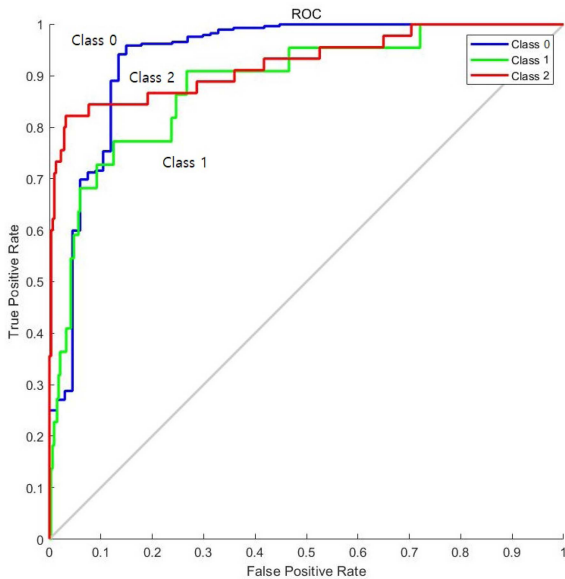


그림 7. N=6 에서 개선된 분류기의 수신자판단특성곡선.
Fig. 7. ROC curves of DME grading classifier for class 0, 1, 2, and N=6.

IV. 토의 및 결론

본 연구에서는 DME 등급 분류 정확도의 개선을 위해 DNN에 입력되는 검증 데이터의 무작위적 증강 기법을 적용함으로써, 학습 최종 결과인 정확도 89.14%에 비해, N=6

일 경우 $1.61 \pm 0.55\%$ 개선된다는 것을 보였다. 뿐만 아니라 Al-Bander 등이 자체적으로 개발한 DNN은 입력영상의 해상도가 512×512 였음에도 DME 진단 정확도가 88.8% 이었으나, 229×229 입력영상을 사용한 본 연구에서는 92.76%로 더 우수하였다. 이는 전처리 과정의 우수성 및 검증데이터 증강기법의 우수성을 의미한다. 제안된 기법은 분류기의 개선에 일반적으로 적용되는 앙상블학습 (ensemble learning) 에 비해 간단하면서도 신뢰할만한 개선을 달성한다는 장점이 있으며, N 에 따라 2에서 6까지는 상관계수 0.5667, 표준 오차 $0.177e-2$ 의 선형 회귀가 가능한 수준에서 개선된다는 점을 보인 것이 본 연구의 특징이라 할 수 있다.

다만, 가용한 컴퓨팅 용량의 제한으로 인해 정확도 개선에 도움이 되는 다양한 기술들을 적용해 보지 않은 점들이 있다.

첫째로 입력영상의 해상도 향상과 이에 맞는 전용 DNN의 제작이다. DNN을 사용한 DR 등급분류 정확도 개선에 해상도가 결정적인 요소로 작용함을 2017년 Wang 등이 밝힌 바 있다[10]. 2018년 Lam 등은 Kaggle 데이터를 사용하여 GoogLeNet 과 AlexNet 을 학습시켜 ICDR {클래스 0, 1}과 {클래스 2, 3, 4} 의 2가지 그룹을 분류하는데 있어서 정확도가 74.5%가 됨을 보였으며[11], 이는 본 연구의 결과에 비해 정확도가 떨어지는데, 이는 GoogLeNet 과 AlexNet 의 입력영상 해상도가 각각 224×224 , 227×227 로 제한된 영향도 있는 것으로 판단된다. 2016년 Pratt 등은 Kaggle 데이터를 512×512 화소의 영상으로 변환해서 3x3 컨벌루션으로 구성된 13층의 VGG-net 과 유사한 DNN을 학습시킴으로써 95% 특이도와 75% 정확도를 얻을 수 있음을 보였으며 [12], 이 또한 본 연구 결과보다 정확도가 떨어지는데 이는 적용된 DNN의 제한에 의한 영향인 것으로 판단된다. Wang 등은 Kaggle 데이터를 사용하여, 18층의 Net-4와 20층의 Net-5를 학습시킬 때, 입력 영상해상도가 128×128 , 256×256 , 512×512 에서 Kappa score가 각각 0.70, 0.80, 0.81 이 됨을 보였으며 이는 본 연구 결과에 비해 우수하다[10]. 따라서 해상도를 현재 inception v3 의 299×299 에서 2배 정도 올리면 망막영상의 미세한 혈관 변화가 보다 명확히 분석될 수 있으며, 이를 위해서는 망막영상 진단 전용의 DNN을 제작해야 하는데, 이는 사전학습이 되어 있지 않으므로 학습과정에서 많은 시간이 소요되는 문제가 있겠지만, 시간과 컴퓨팅 용량이 지원된다면 우선적으로 시도해 볼 수 있을 것이다.

둘째로 학습 데이터의 질과 양 문제이다. 타 연구에서는 자체로 확보한 데이터 또는 공개된 데이터를 사용하고 있는데, 2016년 Gulshan 등은 EyePACS 데이터와 자체 확보 데이터로 구성된 망막 영상 $128,175$ 장으로 inception v3

를 딥러닝 시키고, Messidor-2 를 테스트데이터로 사용하여 RDR을 민감도 87.0%, 특이도 98.5% 수준으로 진단할 수 있음을 보였으며 [7], 2017년 Ting 등은 Singapore National Diabetic Retinopathy Screening Program (SIDRP 2010-2013)에 의해 확보된 494,661 장의 망막 영상으로 딥러닝 시킨 DNN의 AUC 값이 RDR 과 노인성황반변성(AMD: age-related macular degeneration) 에서 각각 0.936, 0.931 이 됨을 보였다. [13] 이들은 데이터의 양 문제를 해결하고 수십만장의 영상을 대상으로 하고 있으며, 우수한 결과를 내고 있다. 본 연구에서 사용한 Messidor 데이터셋은 공개된 타 데이터셋에 비해 영상 품질이 우수한 것으로 정평이 나 있음에도 불구하고, 등급 분류과정에서의 오류가 지적되고 있으며, 같은 영상의 중복 등록 문제도 있는 것으로 밝혀지고 있어서 이에 대한 보다 철저한 검증이 요구되며, 데이터의 양도 보다 확충할 필요가 있다. 뿐만 아니라 클래스 간 데이터 양의 불균형이 존재하여 전체 정확도에 영향을 미치고 있는 것으로 판단되며, 이는 안과 전문의와의 협력을 통한 데이터셋의 개선 및 균형 잡힌 데이터셋 마련을 통해 획기적인 개선이 이루어 질 것으로 기대되고 이는 향후 연구과제이다.

셋째는 DNN 고유의 블랙박스 문제로서, 내부 변수의 가시화 등을 통해 설명 가능한 인공지능으로 개발해야 임상적 적용에 신뢰도를 더할 수 있게 될 것이며, 이 또한 연구 과제로 생각된다.

본 연구의 발전방향으로는 정확도의 개선과 같은 인공지능의 기술적인 측면뿐 아니라 환자의 조기진단 측면에서 휴대용정보단말을 사용해서 영상을 획득하고 분석하는 기술이 중요할 것으로 판단된다. 2018년 Rajalakshmi 등은 스마트폰 기반의 FOP (Fundus on phone) 기기와 인공지능 소프트웨어 EyeArt AI를 사용하여 296 명의 환자를 대상으로 한 DR 진단에서 민감도 95.8%, 특이도 80.2%, Kappa score 0.78을 얻었음을 보고하고 있다[14]. 따라서 이러한 모바일 헬스케어를 통해 임상적인 효과의 최대화를 목표로 환자 질환 관리의 성과 측면을 동시에 추구하는 것도 하나의 발전방향일 것이다.

결론적으로 본 연구는 상술한 여러가지 제한점에도 불구하고 검증 데이터의 증강을 통해 최종 DME 등급 분류기의 성능을 개선할 수 있다는 것을 정량적으로 검증한데 의의가 있으며, 이는 DME 진단의 정확도 향상에 도움이 될 것으로 기대한다.

참고문헌

[1] Gye HJ and Song SJ, "New Modalities for the Diagnosis and Treatment of Diabetic Retinopathy," *Korean J Med*, vol. 89, no. 3, pp. 271-276, 2015.

[2] Jee D, Lee WK and Kang S, "Prevalence and risk factors for diabetic retinopathy: the Korea National Health and Nutrition Examination Survey 2008-2011," *Invest Ophthalmol Vis Sci*, vol. 54, no. 10, pp. 6827-33, 2013.

[3] Abramoff MD, Lou Y, Erginay A, Clarida W, Amelon R, "Folk JC and Niemeijer M. Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning," *Invest Ophthalmol Vis Sci*, vol. 57, no. 13, pp. 5200-5206, 2016.

[4] Abramoff MD, Lavin PT, Birch M, Shah N and Folk JC, "Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices," *npj Digital Medicine*, vol. 1, no. 39, pp. 1-8, 2018.

[5] van der Heijden AA, Abramoff MD, Verbraak F, van Hecke MV, Liem A and Nijpels G, "Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System," *Acta Ophthalmol*, vol. 96, no. 1, pp. 63-68, 2018.

[6] Messidor. Available at: <http://www.adcis.net/en/DownloadThirdParty/Messidor.html>. Accessed on Sep. 1, 2017.

[7] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL and Webster DR, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, pp. 2402-2410, 2016.

[8] Al-Bander B, Al-Nuaimy W, Al-Tae MA, Williams BM and Zheng Y, "Diabetic Macular Edema Grading Based on Deep Neural Networks," *Proceedings of the Ophthalmic Medical Image Analysis International Workshop OMI*, vol. 2016, no. 1, pp. 121-128, 2016.

[9] Kaggle, Inc. Diabetic Retinopathy Detection. Available at: <https://www.kaggle.com/c/diabetic-retinopathy-detection>. Accessed on Sep. 1, 2016.

[10] Wang Z and Yang J, "Diabetic Retinopathy Detection via Deep Convolutional Networks for Discriminative Localization and Visual Explanation," *arXiv*, vol. 1703, no. 10757, pp. 1-9, 2017.

[11] Lam C, Darvin Yi D, Guo M and Lindsey T, "Automated Detection of Diabetic Retinopathy using Deep Learning," *AMIA Jt Summits Transl Sci Proc*, vol. 2017, pp. 147-155, 2018.

[12] Pratt H, Coenen F, Broadbent DM, Harding SP and Zheng Y, "Convolutional Neural Networks for Diabetic Retinopathy," *Procedia Computer Science*, vol. 90, pp. 200-205, 2016.

[13] Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, Hamzah H, Garcia-Franco R, San Yeo IY, Lee SY, Wong EYM, Sabanayagam C, Baskaran M, Ibrahim F, Tan NC, Finkelstein EA, Lamoureux EL, Wong IY, Bressler NM, Sivaprasad S, Varma R, Jonas JB, He MG, Cheng CY, Cheung GCM, Aung T, Hsu W, Lee ML and Wong TY, "Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes," *JAMA*, vol. 318, no. 22, pp. 2211-2223, 2017.

[14] Rajalakshmi R, Subashini R, Anjana RM and Mohan V, "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence," *Eye*, vol. 32, no. 6, pp. 1138-1144, 2018.