

인공지능을 이용한 안저영상 판독 기술 동향

조현성

한국전자통신연구원 책임연구원

I. 서론

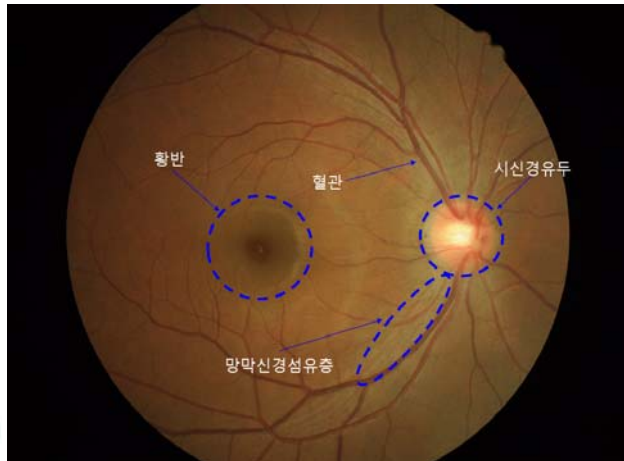
“몸이 10냥이면 눈은 9냥이다”라는 옛 속담이 있다. 인지과학이나 생명과학 분야에서는 사람이 감각기관을 통해 획득하는 정보량의 60~70%는 시각을 통해 이루어진다고 한다. 이는 사람의 신체 중에 눈이 그만큼 중요하다는 의미를 전하고 있다. 인체를 구성하는 조직 중에서 이렇게 중요한 부분을 차지하는 눈과 관련된 연구는 안과 질환을 중심으로 많은 발전이 있었고, 현재 또한 활발한 연구를 진행하고 있다.

2006년 진행한 메타 분석 연구에 따르면, 미국에서 각종 질병에 의해 발생된 실명으로 인한



(a) 안저카메라

〈자료〉 ㈜한국탐콘 홈페이지 제공



(b) 안저영상

〈자료〉 건양대학교병원 제공

[그림 1] 안저카메라와 안저영상

* 본 내용은 조현성 책임연구원(042-860-6734, hsc@etri.re.kr)에게 문의하시기 바랍니다.

경제적 손실은 연간 35억 달러에 이른다고 추정하고 있다[1],[2]. 아일랜드에서는 2010년 총 12,995명이 실명되었고, 이로 인한 직접 경제 손실이 2억 7,700만 유로(€)에 이르렀으며, 8억 900만 유로의 간접 경제 손실이 발생했다고 보고하고 있다[3]. 이처럼 눈과 관련된 질병으로 인한 사회 경제적 손실을 줄이고, 환자의 건강상태를 증진하기 위해서는 관련 질환의 예방과 조기진단이 매우 중요하다.

눈과 관련된 다양한 연구 중에 최근 가장 주목받고 있는 것이 바로 안저영상이다. 안저영상은 비침습적으로 간편하고 저렴하게 인체의 혈관, 신경, 망막 등의 상태를 측정하고 관찰할 수 있는 특징이 있다. 이러한 특징 때문에 안과에서는 다양한 안과 질환 진단을 위해 안저영상을 활용하고 있다.

[그림 1 (b)]와 같이 안저카메라로 촬영한 안저영상에서는 망막에 분포하고 있는 혈관, 망막신경 섬유층, 황반, 시신경유두 등을 관찰할 수 있으며, 촬영한 결과를 통해 해부학적 조직의 특징을 면밀히 분석하여 질환 여부를 판단하는 유용한 자료로 사용하고 있다.

안저영상을 판독하는 데 있어서 현실적인 문제는 미세한 병변을 정밀하게 판독하기 위해서는 고도의 숙련된 전문의가 필요한데, 검사 수요대비 안저영상 판독 전문의의 공급이 턱없이 부족하므로 검사 결과의 판독에 많은 어려움이 있다. 특히, 당뇨, 고혈압과 같은 만성질환 환자의 경우 내과 또는 가정의학과와 별개로 안과 분야를 전문적으로 검사하는 과정을 거쳐야 하는 번거로움이 있어 많은 불편함이 있다.

다른 한편으로는 충분한 전문성과 경험을 보유한 안과 전문의 사이에서도 안저영상의 판독 일치율에 많은 차이가 있다는 결과도 보고되고 있기에, 최근 판독의 정밀성과 일관성을 위해 인공지능 기술을 도입하고자 하는 시도가 많이 발표되고 있다. 특히, 미국과 같이 의료 접근성이 열악한 국가에서는 인공지능을 접목한 안저영상 판독은 원격의료를 통해 서비스할 수 있는 장점 때문에 인공지능 기술을 안저영상 판독에 활용하는 분야가 특별히 주목받고 있다.

본 고에서는 망막을 촬영하여 획득한 안저영상(Fundus Image, Fundus Photography)을 바탕으로 인공지능 기술을 적용한 연구 동향을 살펴본다. 특히, 주요 실명 유발 질환인 당뇨병망막병증과 녹내장에 관련된 주제는 심도 있게 관련 기술을 살펴본다. 이 두 가지 안과 질환 이외에도 앞으로 관심을 가지고 접근해야 할 대상인 황반변성, 심혈관계 질환, 치매 등을 안저영상과 인공지능을 이용해 진단하는 주제도 역시 살펴본다.

II. 인공지능을 이용한 안저영상 판독 사례

1. 당뇨병망막병증 판독 사례

가. 당뇨병망막병증 개요

당뇨성망막병증(diabetic retinopathy)은 만성질환인 당뇨병의 유병 기간이 길어진 환자의 경우 망막에 존재하는 혈관에서 문제가 발생하고, 이에 따른 출혈과 출혈에 동반되는 부작용으로 인해 망막 조직의 변성이 발생하여 생기는 질병이다. 이러한 변화의 결과를 관찰할 수 있는 병변으로는 소동맥류, 출혈, 황반부종 등이 있으며, 황반 부위의 침범 정도에 따라 시력에 중대한 영향을 주기도 하고, 실명에 이르기도 한다. 당뇨병망막병증에 대한 35개의 기존 연구를 메타 분석한 연구에 따르면, 당뇨병망막병증의 유병률은 34.6%, 증식성 당뇨병망막병증은 6.96%에 이른다고 보고하고 있다[4],[5].

당뇨성망막병증의 초기에는 비증식성 망막병증으로 시작하며, 국소적 출혈 또는 삼출물 등을 망막에서 관찰할 수 있고, 시력 저하 역시 심각한 수준이 아니다. 그러나, 비증식성 망막병증의 기간이 오래됨에 따라 증식성 망막병증으로 이행하고, 망막에 새로운 혈관이 생기면서 새롭게 생성된 혈관은 혈관 벽이 약하기 때문에 쉽게 파열되고, 이에 따른 추가적인 부작용들이 나타나 급격한 시력 저하나 실명에 이르게 된다.

당뇨성망막병증의 경우, 대부분 환자는 통증을 자각하지 못하며, 질환이 상당 수준 이상으로 진행이 되었을 시점에 시력 감퇴 등의 자각증상을 인지하여 병원에 찾아오는 경우가 많아서 초기 진단에 실패하는 경우가 종종 발생한다. 안저검사를 받더라도, 미세한 병변의 경우 판독 과정에서 놓치고 지나치기 쉬우므로 조기진단에 실패하는 사례도 있다.

나. Kaggle 사례

안저영상 판독에 최초의 인공지능 기술을 도입한 사례로, 세간의 관심을 불러 일으킨 배경에는 2015년 2월부터 7월까지 진행된 Kaggle의 “Diabetic Retinopathy Detection” 대회의 공헌이 크다고 할 수 있다[6]. 이 대회는 5개 등급으로 레이블링된 안저영상을 참여자에게 제공하여 판독 모델을 학습할 수 있게 하였으며, 팀별로 학습한 모델을 가지고 레이블을 부여하지 않은 테스트 데이터를 판독한 결과를 제출하여 quadratic weighted kappa 방식으로 예측 결과를 평가받아 순위를 정하였다. 이 대회에서는 레이블(기계학습 대상 클래스)을 정상, 경미한 당뇨병망막병증, 중증 당뇨병망막병증, 심각한 당뇨병망막병증, 증식성 당뇨병망막병증의 총 5개의 범주로 정의하였

다. 결과 평가 방법을 quadratic weighted kappa로 한 이유를 추정해 보면, 제시한 5개의 클래스 레이블이 순서의 의미를 내포하고 있으므로, 예측 모델이 정상을 초기 질병으로 오판한 결과보다는 정상을 중증 또는 심각한 상태의 병증으로 예측한 것에 대해서 더 많은 벌점을 적용하는 방식이 현장에서 사용하는데 더 합리적이라는 이유일 것이다. 661개 팀이 총상금 10만 달러를 두고 각축을 벌였으며, 이 대회에서 Min-Pooling 팀이 0.84957(quadratic weighted kappa score)의 점수로 1등을 차지하였다. Min-Pooling 팀이 사용한 모델 학습 방법은 여러 개의 심층신경망을 개별적으로 학습하고, 이 결과를 앙상블하는 방법을 사용하였다.

다. 구글 사례

Kaggle의 “Diabetic Retinopathy Detection” 대회가 끝난 이후, 세계적인 의학 저널인 JAMA, Ophthalmology, Nature 등에서 당뇨병성망막병증의 진단을 위해 딥러닝을 활용한 수많은 연구 결과들이 발표되었다[7]-[9],[13]. 당뇨병성망막병증을 판독하기 위해 딥러닝을 적용한 연구 결과를 의학 전문 저널에 최초로 발표한 사례는 2016년 구글이 JAMA에 게재한 연구 논문이다[7]. 논문에서 제시한 성능은 AUC 0.99 수준으로, 전문 안과 의사의 실력과 버금가는 수준이었다.



〈자료〉 Google, Google AI Blog, Deep Learning for Detection of Diabetic Eye Disease, 2016-11-29

[그림 2] 정상 안저영상과 당뇨병성망막병증 안저영상

구글은 심층신경망(Convolutional Neural Network: CNN)의 하나인 InceptionV3에 12만 여 장의 안저영상을 입력하여 기계학습을 하였고, 학습한 모델은 EyePACS-1[10] 데이터 8,788장과 Messidor-2[11] 데이터 1,745장을 가지고 성능을 테스트하였다. 구글이 발표한 논문에서는 영상의 품질과 레이블링 품질에 대해서 상세한 자료를 제공하고 있다.

그러나, EyePACS-1 데이터베이스에서는 학습과 테스트 모두에서 데이터 표본을 뽑은 것으로 나와 있는데, 이때 상호 중복이 없이 표본을 뽑았다는 언급이 없는 것이 아쉬운 점이다. 또한, 테스트 데이터의 구성에서 정상과 질병의 분포 편차가 크다는 점도 성능평가 결과의 해석에서 고려할 사항

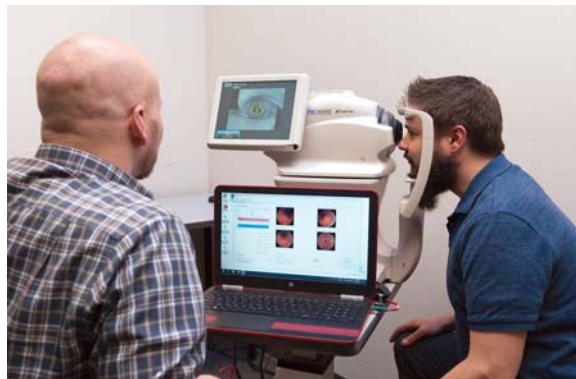
으로 보인다.

라. IDx-DR 사례

2018년 4월 미국의 IDx사는 자사가 개발하고 출시한 세계 최초 인공지능 의사가 FDA 승인을 받았다는 뉴스를 발표하였다[12]. IDx사에서 승인받은 제품은 IDx-DR이라는 인공지능 시스템인데, 안저영상을 가지고 당뇨병망막병증을 선별검사하는 인공지능 시스템이다.

이전에도 의료영상 판독을 위해 인공지능 기술을 사용해서 FDA 승인을 받은 몇몇 사례가 있었다. 이전의 사례는 임상 의사결정시스템(Clinical Decision Support System: CDSS)의 관점에서 접근한 허가사항이지만, IDx의 FDA 승인은 인공지능 시스템 독자적으로 최종 판독 결론을 도출한다는 점에서 차이가 있다. 이러한 차이를 IDx에서는 언론홍보뿐만 아니라 Nature와 같은 유력 학술저널에 기고하는 방식 등으로 홍보에 열을 올리고 있다[9],[12]-[14].

IDx-DR 시스템은 900명의 당뇨 환자를 대상으로 촬영한 망막 사진을 가지고 임상 시험을 진행했고, 성능은 민감도 87.4%, 특이도 89.5%로 보였으며, 촬영 장비는 Topcon NW400에 국한한 제한점이 존재한다[9].



〈자료〉 IDx, Press Release: IDx and Topcon sign exclusive AI platform agreement for the U.S. market, 2018. 10. 23.

[그림 3] IDx의 당뇨병망막병증 판독을 위한 AI 적용 제품

2. 녹내장 사례

가. 녹내장 개요

녹내장(Glaucoma)은 신경절세포(Retinal Ganglion Cell: RGC)와 그 축삭(Axon)이 손상됨으로써 실명에 이르는 질병이다[15],[16]. 녹내장은 만성적이고, 비가역적으로 진행되는 특성 때문에

조기에 발견한 녹내장은 치료제 또는 수술을 통해 진행을 늦출 수 있고, 치료 효과 역시 좋은 편이다. 하지만, 질환의 특성상 말기 단계까지 환자가 뚜렷하게 느낄 수 있는 시야 결손이나 시력 저하와 같은 주관적인 증상이 없는 경우가 많다. 또한, 진행 단계가 심한 녹내장의 경우 치료 예후가 나빠서, 검진을 통한 조기 발견이 중요하다.

최근 인구 조사에 근거한 50건의 글로벌 메타 분석에 따르면, 전 세계적으로 40세부터 80세까지의 녹내장 유병률은 약 6,430만 명에 해당하는 3.5%로 보고되고 있다. 인구 증가 및 고령화로 인해 이 수치는 2040년까지 1억 2,200만 명으로 증가할 것으로 예상되고 있다[15]. 특히, 싱가포르 중국인의 경우 약 85%, 미국 흑인 인구의 경우 같은 비율, 미국의 경우 전체 비율의 50%까지도 조기 진단이 어려운 상황이다[17].

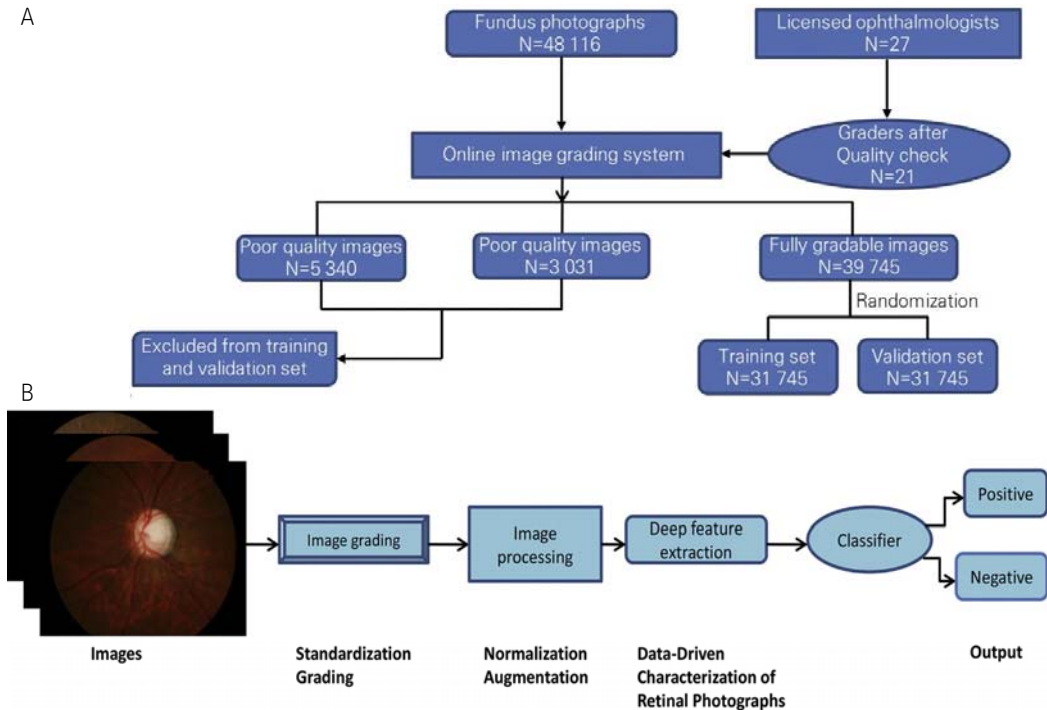
국내 역학조사 연구에 따르면 국내 녹내장 유병률은 전 국민의 2~4%(약 240만 명)에 이르며, 인구 고령화에 따른 유병률은 지속적으로 증가 추세에 있다고 한다[18],[19].

녹내장으로 인한 시력 손실의 대부분은 조기 발견과 이를 통한 적절한 치료를 통해 예방할 수 있으나, 대다수의 녹내장 환자의 조기진단에 실패하고 있다. 그 이유는 녹내장의 질환 특성상 질병의 진행 단계에서 중심 시력이 영향을 받는 만성 녹내장 단계에 이르기까지 대부분 환자가 자각증상을 느끼지 못하기 때문이다[20]. 녹내장 조기 진단의 또 다른 어려움으로는 간편하고 저렴한 검진 방법인 안저촬영 결과의 판독에 있어서 고도의 전문성과 경험이 필요하기 때문이다. 이러한 현상은 녹내장이 초기 단계에서 후기 단계로 진행됨에 따라 치료비가 4배 이상 증가하여 상당한 재정적 부담을 초래하고, 또한 녹내장 말기에 이루어지는 치료 역시 그 예후가 초기 단계부터 치료하는 것보다 상대적으로 좋지 않은 결과를 보인다[21].

나. 중국 Zhongshan 안과센터 사례

2018년 세계적 권위를 가지고 있는 미국 안과학회 저널인 Ophthalmology에 중국의 Zhongshan 안과센터가 딥러닝을 이용한 녹내장 판독 연구 결과를 발표하였다[22]. 중국 연구팀에서는 3만 1,000여 장의 안저영상으로 딥러닝 모델을 학습하였고, 8,000장의 안저영상을 가지고 성능을 시험하였다.

중국 연구팀에서 개발한 녹내장 판독 딥러닝 모델의 성능은 AUC 0.986 수준으로 매우 높았고, 민감도는 95.6%, 특이도는 92.0%를 보였다. 이 연구에서는 InceptionV3를 딥러닝 기본 모델로 사용하였다. 녹내장을 위한 데이터 레이블은 LabelMe라는 조직에서 보유한 7만 여장의 안저영상에 대해서 21명의 훈련받은 안과의를 동원하여 녹내장 의심과 정상으로 분류하고 영상마다 레이블을 정의하였다.



〈자료〉 <https://doi.org/10.1016/j.ophtha.2018.01.023>, Ophthalmology, Vol. 125, Issue 8, August 2018, Page 1,200.

[그림 4] 중국 Zhongshan 안과센터의 녹내장 판독을 위한 딥러닝 적용 사례

중국 Zhongshan 안과센터의 연구 결과를 면밀히 살펴보면, 몇 가지 제약사항이 존재함을 알 수 있다. 먼저, 성능시험을 위한 테스트 데이터의 분포가 정상이 75%를 차지하고 녹내장 의심이 25%를 차지하고 있는 아쉬움이 있다. 기계학습을 위한 데이터의 클래스 분포는 연구자의 의도에 따라 어떻게 구성해도 큰 문제가 없어 보이나, 성능시험을 위해서는 정상과 비정상의 분포 균형이 맞추어진 구성으로 성능평가 결과를 보여주는 것이 좀 더 객관적인 결과 해석을 할 수 있다고 생각한다. 데이터의 레이블링 관점에서도 녹내장 의심 수준의 평가로 이루어진 아쉬움이 있다. 녹내장의 확진은 안저영상 하나만으로는 불가능하고, 시야검사와 기타 검사 데이터를 종합하여 최종 진단을 내리는 것이 학계의 정설이다. 이러한 점에서 중국 연구팀의 레이블링 결과는 말 그대로 녹내장이 의심되는 수준의 레이블링을 하였기 때문에 실전에서 효과성이 어느 정도 수준일지는 별개의 임상 시험을 통해 확인해 볼 필요가 있다.

다. 기타 녹내장 판독 사례

안저영상에 딥러닝을 적용하여 녹내장을 판독한 사례로는 일본의 Shibata 연구팀의 결과와[23]

호주의 Liu 연구팀에서 발표한 결과가[24] 있다. 일본 Shibata 연구팀의 경우 3,000여 장이라는 상대적으로 작은 규모의 학습 데이터를 사용했으며, 주요 특징으로는 테스트를 위해 OCT라는 별도의 검사 결과를 바탕으로 엄밀한 검증을 거친 데이터로 성능평가를 진행했다는 점이다. 일본 연구팀에서는 ResNet을 바탕으로 딥러닝을 진행했고, SVM과 Randomforest의 학습 결과와 비교한 결과를 제시하고 있다[23].

호주 연구팀 역시 상대적으로 작은 규모의 4,394장의 안저영상을 가지고 ResNet50 모델을 사용해서 녹내장을 판독하는 연구를 진행했다. 호주 연구팀에서 전체 연상의 80%를 학습에 사용하고 나머지 20%를 성능평가에 사용했다. 호주 연구팀에서 제시한 성능은 민감도 89.3%, 특이도 97.1%, AUC 0.97이다. 호주 연구팀의 연구에서 특기할 만한 사항은 안저영상에서 시신경유두 부분이 녹내장성인지 아닌지를 판독하는 목적으로 이루어졌다는 것이다. 녹내장에서 시신경유두 부분만으로 질환 여부를 판독하는 것에는 상당히 많은 한계점이 있다[24].

3. 기타 질환 사례

안저영상을 가지고 주요 안과질환인 당뇨병망막병증과 녹내장 이외에도 나이관련 황반변성, 심혈관계 질환, 치매까지도 진단하는 연구 결과가 속속 발표되고 있다. 앞서 언급한 바와 같이 안저영상으로 혈관과 신경조직까지 관찰할 수 있기 때문에 안과 이외의 질병으로 생성되는 병변의 특징을 딥러닝으로 검출하여 진단에 활용하고자 하는 시도는 충분히 의미 있는 접근이라 할 수 있다.

나이관련 황반변성은 백내장, 녹내장과 함께 WHO에서 정의한 3대 실명 질환 중의 하나이다. 안저영상과 딥러닝을 결합하여 나이관련 황반변성 위험도를 예측하는 연구 역시 당뇨병망막병증이나 녹내장과 같이 안과 분야에서는 주요 관심 질병으로 최근 괄목할 만한 연구 성과가 다수 발표되었다[25],[26]. Grassmann et al.(2017) 연구에서는 3,654명의 환자로부터 촬영한 12만여 장의 안저영상을 질병의 중증도에 따라 12개의 등급으로 레이블하고, 5,555장의 테스트 이미지로 성능평가를 하였다. 질병의 등급을 분류하는 실험 세팅으로 성능평가 결과 quadratic weighted kappa는 92.14, 정확도는 63.3%를 보였다. 딥러닝은 AlexNet, GoogLeNet, VGG, InceptionV3, Res Net101, InceptionResNetV2를 모두 사용하여 학습하고, 각각의 결과를 Randomforest로 앙상블하는 방법을 사용하였다[25].

Poplin et al.(2018)은 UK Biobank 데이터베이스에 있는 96,082장의 안저영상과 EyePACS 데이터베이스에 있는 1,682,938장의 안저영상을 가지고 InceptionV3를 사용해서 심혈관계 위험도를 예측하는 연구 결과를 발표하였다[27]. 연구 결과의 성능평가는 UK Biobank의 24,008장과

EyePACS-2K의 1,958장을 가지고 진행했으며, 딥러닝 알고리즘만으로 AUC 0.70 성능을 보였다. 실험 내용에서는 안저영상을 이용한 딥러닝 결과를 심혈관계 위험도와 상관관계가 높은 혈압, 연령, 성별 등의 지표와 연동하여 예측할 경우 AUC가 0.73까지 향상된다는 내용을 포함하고 있어서 다양한 형태의 데이터와 결합해서 성능 향상을 보이는 연구내용이 주목할 만하다.

Chan et al.(2017)은 딥러닝과 같은 기계학습 방법을 적용하지는 않았으나, 안저영상과 OCT에서 신경계의 특징을 추출하여 치매를 진단하는 검사 프로토콜을 제시하고 있다[28]. Chan et al. (2017)의 연구에서 얻을 수 있는 교훈은 안저영상과 딥러닝으로 치매를 진단하고 위험도를 예측하는 연구를 진행하는 것도 고려해볼 사항이라는 것이다. 고령화 사회와 밀접하게 연결된 사회문제가 바로 치매이며, 아직 효과적인 해결책이 보이지 않기에 저렴하고 비침습적인 안저검사와 인공지능을 결합한 치매 선별검사 방법에 관한 연구는 그 중요성이 매우 높다고 판단된다.

III. 안저영상의 기계학습 사례에 대한 고찰

1. 기술 발전의 공로자와 후원자

2012년 ImageNet 경진대회를 계기로 딥러닝에 관련된 연구가 폭발적으로 증가하기 시작하였다. 국내에서는 알파고와 이세돌의 대국으로 인해 인공지능에 대한 정부와 학계의 관심이 급격히 증가하였고, 이제는 고등학생도 딥러닝 관련 논문을 발표하는 시대가 도래하였다.

이러한 시대적 상황을 가능하게 한 배경을 나름 추정해보자면, 크게 2가지 핵심 요소가 그 원동력으로 작용했다고 본다. 첫 번째는 수많은 역경과 질곡을 이겨내며 평생을 걸쳐 기계학습 분야 연구에 매진하고, 수많은 제자를 키워낸 캐나다 토론토대학의 Geoffrey Everest Hinton 교수가 있었기 때문이다. 두 번째 요인은 이미지 분야에서 기계학습 실험이 가능하도록 ImageNet이라는 데이터를 만들어 공개한 미국 스탠포드대학의 Fei-Fei Li 교수 연구팀이 있었기 때문이다.

안저영상 분야에서도 EyePACS나 Messidor와 같은 데이터베이스가 있었기 때문이며, Kaggle에서 경진대회를 진행할 수 있도록 후원한 California Healthcare Foundation이[29] 있었기 때문에 가능한 것이었다.

이처럼 인공지능 분야에서 걸출한 결과물이 나올 수 있었던 배경에는 비단 알고리즘 분야의 전문가뿐만 아니라 유용한 학습 데이터를 구축하고 제공했던 숨은 공로자가 있었고, 사회발전을 위해 아낌없이 거금을 후원한 공익적 재단의 지원, 기술과 S/W를 타인과 공유할 수 있는 문화와 기반

환경, 그리고 여기에 적극적으로 동참한 참여자들의 노력이 함께한 결과임이 분명하다.

2. 학습 데이터 구축과 모델 성능평가

최근 의료 분야에 인공지능을 적용하여 괄목할 만한 연구 결과를 도출한 사례는 비단 안과 영역뿐만 아니라 의료계 전반에 걸쳐서 이루어지고 있는 하나의 거대한 흐름이다. 이러한 시점에서 의료 분야에 대한 인공지능 활용에 있어서 몇 가지 주의 깊게 접근해야 할 요소가 있다.

발표된 연구 결과의 해석에 있어서 실험의 구성이 어떠한지를 왜곡되지 않은 관점에서 살펴봐야 한다. 실험 구성이란 학습 데이터의 구성과 성능평가 지표를 어떻게 했느냐가 관건이다. 의료 분야에서 질병 유무를 판단하는 모델일 경우, 통상적으로 민감도, 특이도, AUC 지표를 가지고 성능을 평가한다. 질병 유무를 판독하는 것이 아니고 질병의 중증도를 판독하는 경우라면, 분류 결과가 순위의 의미를 내포하고 있기 때문에 더욱더 신중한 성능 지표를 사용해서 해석할 필요가 있다. 예를 들어, Kaggle에서 당뇨병망막병증의 중증도를 예측하는 경진대회에서는 quadratic weighted kappa라는 성능 지표로 성능을 평가했는데, 그 이유는 단순한 분류 문제를 해결하는 것이 아니라 분류 결과에 상, 중, 하와 같은 순서의 의미가 내포되어 있기 때문이다.

의료 분야의 기계학습 모델 성능평가를 위한 데이터 구성에 있어서 다른 분야와는 대비되는 또 다른 특성이 있다. 어떤 분야든지 학습에 사용한 데이터는 성능평가에 포함하지 않는 것이 기본이고, 모델의 성능은 한 번의 성능 평가결과가 아니라 10 폴드 교차검증과 같이 통계적 판단 척도를 같이 제공하는 것 역시 기본이다.

의료 분야에서, 특히 영상 데이터로 기계학습을 시도할 때 고려할 사항은 의미상 왜곡을 유발할 수 있는 상황을 배제해야 한다. 예를 들어, 어떤 사람의 안저영상을 어제 측정한 버전과 오늘 측정한 버전은 서로 다른 이미지임이 확실하다. 그러나 이 2장의 사진은 촬영자의 환경에 특이한 변화가 없는 한, 의미적으로는 동일한 사진으로 간주해야 한다. 실제 임상현장에 존재하는 의료영상에는 하루에도 여러 장을 반복 촬영한 경우가 많다. 이러한 상황에서 모든 데이터를 가감 없이 학습과 성능평가에 사용한다면, 결과에 심각한 왜곡을 가져올 수 있다.

의료 분야 데이터의 또 다른 특성을 살펴보면, 같은 목적과 결과를 도출하는 검사 기기라도 의료 기기 제작 회사에 따라 데이터 특성이 미세하게 달라지고, 이에 따른 왜곡 현상이 발생할 수 있다. 또한, 고가의 의료기기는 대부분 10년을 넘게 사용하는 경우도 있어서 기기의 센서가 열화되기 때문에, 데이터 품질의 일관성이 부족한 경우도 고려해야 한다.

3. 상용화를 위한 고려

의료 및 헬스케어 분야에서 인공지능 기술과 관련된 최근 산업 전망 보고서에 따르면, 관련 시장은 2023년에 2억 2,800만 달러에 달하며, 연평균 성장률은 48%에 이를 것으로 예측하고 있다 [30]. 관련 시장의 규모가 큰 만큼, 다국적 기업부터 새로 시작하는 스타트업에 이르기까지 다양한 제품을 출시하려는 경쟁이 이미 시작되었다. 이러한 상용화에 있어서 가장 중요한 부분은 역시 비용 대비 효과의 척도일 것이다. 인공지능이 아무리 좋다 한들 현장의 요구사항과 시장의 특성을 반영하지 못한다면, 그저 하나의 연구 결과이거나 이름 없이 사라질 제품에 불과하다.

연구 차원에서는 위에서 살펴본 다양한 질병들에 대해서 연구 주체에 따라 모두 독립적으로 1개의 목적, 즉, 특정 질병 하나만을 대상으로 하고 있다. 그러나, 이러한 제품은 개원의 관점에서 바라볼 때 반쪽짜리 제품에 불과할 것이다. 관련 제품을 도입하는 의료진 관점에서는 한 개의 제품으로 다양한 질환을 다룰 수 있어야 제품의 가치가 높아지고, 경제적 부담이 줄어들 수 있기 때문이다.

상용화에 있어서 또 다른 고려사항으로는 원본 데이터 획득환경을 반영한 제품이어야 한다는 점이다. 기계학습 과정에서는 최대의 성능을 위해 고품질의 데이터를 선별하여 학습하고 성능을 극대화한다. 또한, 의료 분야에서 데이터를 가지고 접근한 대부분 논문에서는 분석 대상 데이터의 포함 범주(inclusion criteria)와 제외 범주(exclusion criteria)를 엄밀하게 기술하고 있다.

현장에서 제품으로 사용할 모델은 모든 상황의 데이터가 입력으로 들어올 수 있다는 점을 간과해서는 안 된다. 예를 들어, 원본 데이터의 측정 품질이 매우 열악한 경우에도 상용제품에서는 대응할 수 있어야 하며, 제외 범주에 속한 데이터를 입력하더라도 적절한 응답을 할 수 있어야 한다. 이에 대한 간단한 해결책으로 질병 유무를 판단하기 전에 영상 품질을 확인하여 인공지능 시스템이 판독하기에 적합한지를 판단하는 기능을 추가하는 것을 생각해볼 수 있고, 하나의 영상에 여러 개의 질병에 관련된 병변이 존재하는 경우를 대비해서 여러 개의 모델을 사용하는 방법도 생각할 수 있다.

IV. 맺음말

인체에서 비침습적이고 경제적이며 쉽게 혈관, 신경, 망막 등을 관찰할 수 있는 유일한 해부학적 조직이 망막이다. 망막을 촬영한 안저영상은 다양한 안과 질환뿐만 아니라 심혈관계 질환과 치매까지도 진단할 수 있는 매우 유용한 도구로 활용할 수 있음을 본 고를 통해 살펴보았다.

안과 질환뿐만 아니라 다양한 만성질환의 합병증으로 극단의 경우에 실명에 이를 수 있어서 국민의 건강과 삶의 질 향상을 위해서는 안저영상을 통한 각종 질환의 조기진단이 필수적이고, 이 과정에

서 진단의 정확성과 효율성을 극대화하는 방법으로 인공지능 기술을 접목하는 것은 선택이 아닌 필수로 여겨진다.

인공지능을 이용한 안저영상의 자동 판독을 실현한다면, 안과 질환뿐만 아니라 다양한 질병을 조기 진단할 수 있고, 이에 따른 치료 효과 향상을 통해 대국민 의료비 지출을 절감할 수 있으며, 환자의 치료 만족도 역시 향상시킬 수 있다. 또한, 오진율 감소와 더불어 임상현장에서 의료진의 수고를 덜어줄 수 있으므로 의료서비스 품질 향상에도 공헌할 수 있다. 경제적 관점에서 관련 기술의 발전에 따른 신산업 창출이 가능하고 관련 산업 활성화에도 이바지할 수 있다.

미국, 중국, 일본에서는 안저영상 판독에 인공지능을 적용하는 시도가 이미 활발하게 진행 중이며, 관련 시장 선점을 위한 다양한 활동들이 포착되고 있다. 또한, 이 분야는 아직 연구하고 검증해야 할 수많은 질병이 남아 있고, 제품으로 출시된 사례가 거의 없는 상황이다. 따라서, 안저영상 판독을 위한 인공지능 기술 개발에 관련된 기초연구와 더불어 상용화에 국내 학계와 관련 업계의 더 많은 관심과 노력이 필요한 시점이다.

[참고문헌]

- [1] Frick KDGower EWKempen JHWolff JL. Economic impact of visual impairment and blindness in the United States. Arch Ophthalmol 2007;125 (4) 544-550
- [2] Rein DBZhang PWirth KE et al. The economic burden of major adult visual disorders in the United States. Arch Ophthalmol 2006;124 (12) 1754-1760
- [3] Economics, Deloitte Access. The economic impact of vision impairment and blindness in the Republic of Ireland. NCBI Dublin, 2011.
- [4] Kobrin Klein, Barbara Eden. Overview of epidemiologic studies of diabetic retinopathy. Ophthalmic epidemiology, 2007, 14.4: 179-183.
- [5] Yau JW, Rogers SL, Kawasaki R, et al; Meta-Analysis for Eye Disease (META-EYE) Study Group. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012; 35(3):556-564.
- [6] Kaggle, Diabetic Retinopathy Detection, <https://www.kaggle.com/c/diabetic-retinopathy-detection>, access date: 2018. 10. 21.
- [7] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016.
- [8] Krause, Jonathan, et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. Ophthalmology, 2018.
- [9] Abramoff, Michael D., et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. npj Digital Medicine, 2018, 1.1: 39.
- [10] EyePACS, <http://www.eyepacs.com/>, access date: 2018. 10. 21.
- [11] Messidor-2 Decenciere et al.. Feedback on a publicly distributed database: the Messidor database. Image Analysis & Stereology, V.33, N.3, pp.231-234, aug. 2014. ISSN 1854-5165.

- [12] IDx, Press Release: FDA permits marketing of IDx-DR for automated detection of diabetic retinopathy in primary care, 2018. 4. 12.
- [13] Van Der Heijden, Amber A., et al. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta ophthalmologica*, 2018, 96.1: 63-68.
- [14] Keane, Pearse A.; Topol, Eric J. With an eye to AI and autonomous diagnosis. *npj Digital Medicine*, 2018.
- [15] Tham, Y. C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., & Cheng, C. Y. (2014). Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. *Ophthalmology*, 121(11), 2081-2090.
- [16] 박기호, 대한민국 최고의 명목이 들려주는 녹내장, 2015. 1. 30.
- [17] Tatham AJ, Weinreb RN, Medeiros FA. Strategies for improving early detection of glaucoma: the combined structure-function index. *Clin Ophthalmol*. 2014;8:611e621.
- [18] Kim YY, Lee JH, Ahn MD, Kim CY; Namil Study Group, Korean Glaucoma Society. Angle closure in the Namil study in central South Korea. *Arch Ophthalmol* 2012;130:1177-1183.
- [19] Yoon KC, Mun GH, Kim SD, et al. Prevalence of eye diseases in South Korea: data from the Korea National Health and Nutrition Examination Survey 2008-2009. *Korean J Ophthalmol* 2011;25: 421-33.
- [20] Gupta P, Zhao D, Guallar E, et al. Prevalence of glaucoma in the United States: the 2005e2008 National Health and Nutrition Examination Survey. *Invest Ophthalmol Vis Sci*. 2016;57:2577e2585.
- [21] Lee PP, Walt JG, Doyle JJ, et al. A multicenter, retrospective pilot study of resource use and costs associated with severity of disease in glaucoma. *Arch Ophthalmol*. 2006;124:12e19.
- [22] Li, Zhixi, et al. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *Ophthalmology*, 2018.
- [23] Shibata, Naoto, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Scientific reports*, 2018, 8. 1: 14665.
- [24] Liu, Sidong, et al. A Deep Learning-Based Algorithm Identifies Glaucomatous Discs Using Monoscopic Fundus Photographs. *Ophthalmology Glaucoma*, 2018, 1.1: 15-22.
- [25] Grassmann, Felix, et al. A Deep Learning Algorithm for Prediction of Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration from Color Fundus Photography. *Ophthalmology*, 2018.
- [26] Burlina, Philippe M., et al. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA ophthalmology*, 2017, 135.11: 1170-1176.
- [27] Poplin, Ryan, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2018, 2. 3: 158.
- [28] Chan, Victor TT, et al. Using retinal imaging to study dementia. *JoVE (Journal of Visualized Experiments)*, 2017, 129: e56137-e56137.
- [29] California Health Care Foundation, <https://www.chcf.org/>, access date: 2018. 10. 21.
- [30] Proquest, Report on the Artificial Intelligence in Healthcare Market, 2018.