

의료 데이터를 활용한 데이터 분석

01 의료 인공지능 개요



목차

- 01. 의료 서비스 패러다임의 변화
- 02. 의사에게 딥러닝이란...
- 03. 의료 데이터의 유형과 활용
- 04. 인공지능망이 의료 데이터를 학습하는 방법
- 05. 의료 인공지능 활용 사례
- 06. 의료 인공지능의 윤리적 문제



커리큘럼

➤ 의료 서비스 패러다임

2020년 1월 9일 데이터 3법 통과, 딥러닝 기술의 발전 및 다양한 웨어러블 디바이스의 등장으로 대량의 의료데이터의 생성속도가 급증함에 따라 헬스케어 패러다임이 데이터 기반 중심으로 어떻게 바뀌는지 알아보니다.

➤ 의사에게 딥러닝이란....

딥러닝의 발전으로 방대한 의료데이터를 활용하기 위해 의사에게 요구되는 역량이 무엇인지 알아보니다.



커리큘럼

➤ 의료 데이터의 유형과 활용

대표적인 의료데이터 유형을 알아보고 개인의 건강 및 생체 정보, 질병과 관련된 가족력 등의 의료 정보가 디지털 헬스케어에 어떻게 활용되는지 알아봅니다.

➤ 의료데이터와 인공지능

복잡한 EMR 데이터와 이미지로 나타낼 수 있는 의료 데이터를 딥러닝 기술이 어떻게 판독하고 분석하는지 알아봅니다.



커리큘럼

➤ 데이터기반 의료사례

빅데이터 분석을 통해 온전히 데이터에만 기반하여 질환의 발병에 어떤 위험 요인이 있는지 예측하는 방법을 알아봅니다.

➤ 의료 인공지능 윤리적 문제

의료 인공지능의 도입이 의사-환자의 관계를 어떻게 변화시킬 것이며 어떤 사회윤리적 차원의 문제점이 있는지 알아봅니다.

추천대상

의료 현업 실무자

의료 인공지능 분야에 연구 계획이 있거나 관심이 있는 자

보건의료 전공자 출신 인공지능 꿈나무

딥러닝을 적용시키고자 하는 보건의료 전공자 및 관련 면허·자격 소지자

인공지능 연구 도전자

인공지능을 활용하고자 하는 의료인과 생소한 의료 영역에 도전하고 싶은 자

01

의료 서비스 패러다임의 변화



✓ 과거의 헬스케어

근거중심의학 / 지식 기반 / 치료 중심

- 질병에 걸렸을 때, 질병을 치료하기 위한 목적으로만 의료 서비스를 이용
- 사전에 질병 이환 여부를 알아볼 수 있는 데이터 부족
- 전염병 예방 및 질병 치료법, 약제 개발을 위한 인프라 부족
- 의료서비스 공급 체계 부족

✓ 의료 환경의 변화



IOT, 웨어러블 기기, 클라우드



의료 빅데이터·인공지능 도입



데이터 3법 통과



데이터 기반/ 예방 중심

✓ The Rise of the Data-Driven Physician

스탠포드 의대에서 2020년 발표한 Health Trends Report

: Data-driven physician이 대두되고 있으며, 웨어러블 디바이스와 IoT, 환자 앱의 임상적 활용



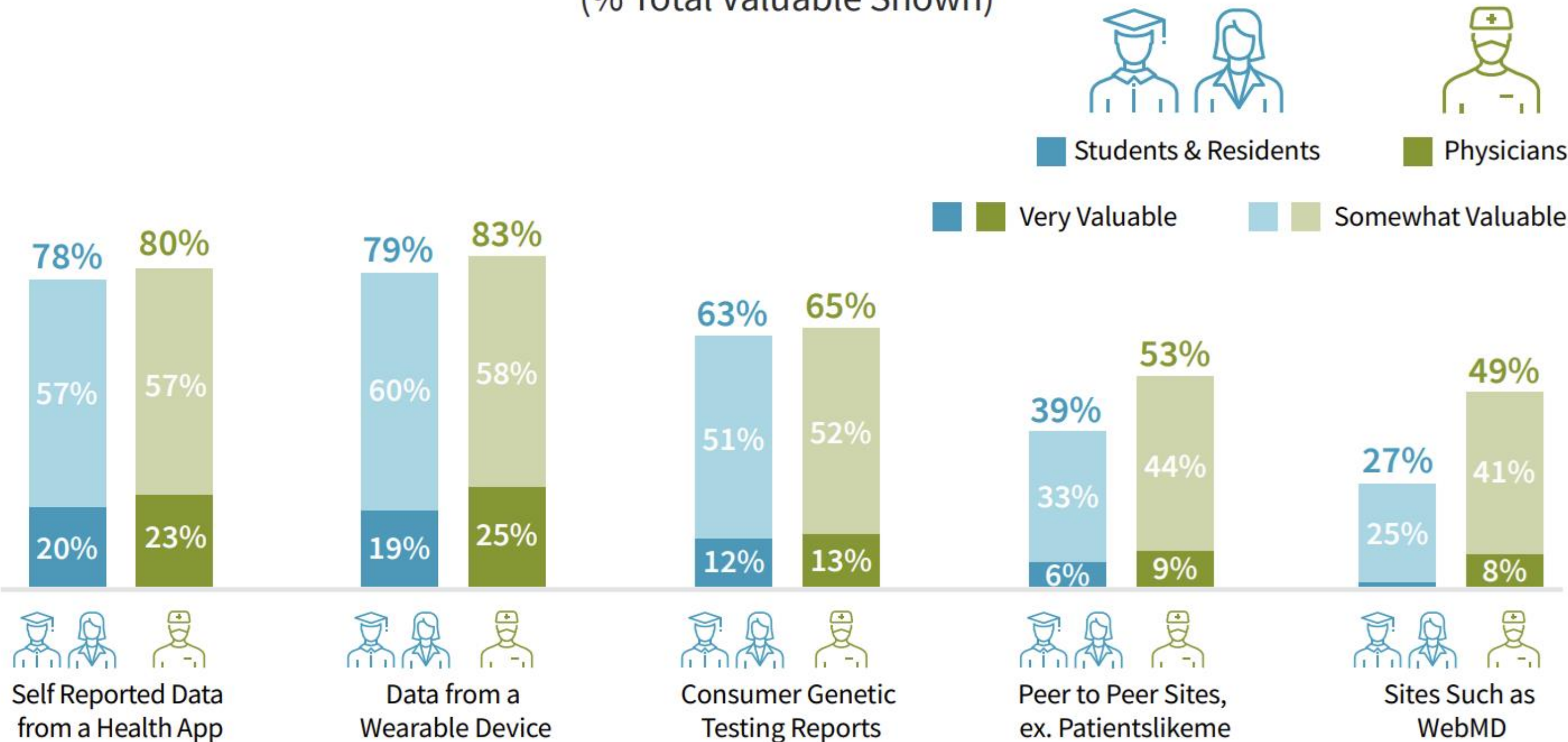
The 2020 Stanford Medicine Health Trends Report identifies the **Rise of the Data-Driven Physician** as one of the **industry's most consequential developments**—one that has **significant implications for patients**.

- [Stanford Medicine's 2020 Health Trends Report spotlights the rise of the data-driven physician | News Center | Stanford Medicine](#)
- [Stanford Medicine 2020 Health Trends Report | Stanford Medicine | Stanford Medicine](#)

✔ The Rise of the Data-Driven Physician

The next generation of Data-Driven Physicians understand that valuable data can come from many different sources, including apps and wearable devices.

If a patient provided you with the following sources of information, how much clinical value do you believe it would provide?
(% Total Valuable Shown)



✔ 의료 환경의 변화: 데이터 3법 통과

데이터 3법이란?

개인정보보호법, 정보통신망 이용촉진 및 정보보호 등에 관한 법률, 신용정보의 이용 및 보호에 관한 법률
개인정보를 가명화하여 과학적 연구 및 통계작성 목적으로 활용할 수 있도록 허용함 (2020년 1월 9일)

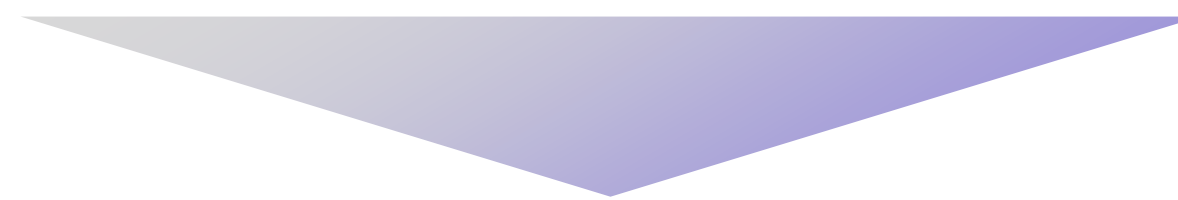
	개념	활용가능 범위
개인정보	특정 개인에 관한 정보, 개인을 알아볼 수 있게 하는 정보	사전적이고 구체적인 동의를 받은 범위 內 활용 가능
가명정보	추가정보의 사용없이 특정 개인을 알아볼 수 없게 조치한 정보	다음 목적에 동의 없이 활용 가능 (EU GDPR 반영) ① 통계작성 (상업적 목적 포함) ② 연구 (산업적 연구 포함) ③ 공익적 기록보존 목적 등
익명정보	더 이상 개인을 알아볼 수 없게 (복원 불가능할 정도로) 조치한 정보	개인정보가 아니기 때문에 제한없이 자유롭게 활용

✓ 의료 환경의 변화: 의사에게 데이터 3법이란?



데이터 3법: 정부가 아닌 병원과 같은 민간에서 개인정보를 가명화하여 과학적 연구 및 통계작성 목적으로 활용할 수 있도록 허용

- 분석할 수 있는 데이터 증가
- 데이터를 수집 분석 통계화하여 다양한 서비스와 융합의 가능성이 열림



- **환자의 인적 사항, 발병데이터, 처방데이터 신약개발에 박차를 가할수 있음**
- **병원 진료 데이터, 검진결과, 생체 정보 등을 클라우드 서비스에 저장해 원격으로 전문 관리를 받을 가능성**
- **건강+금융 데이터로 개인 맞춤형 서비스 제공 가능성**

✓ 의료 환경의 변화: 의사에게 데이터 3법이란?

[2021년] 보건의료 데이터 활용
가이드라인

2021. 1.

개인정보보호위원회
보건복지부

주요 내용

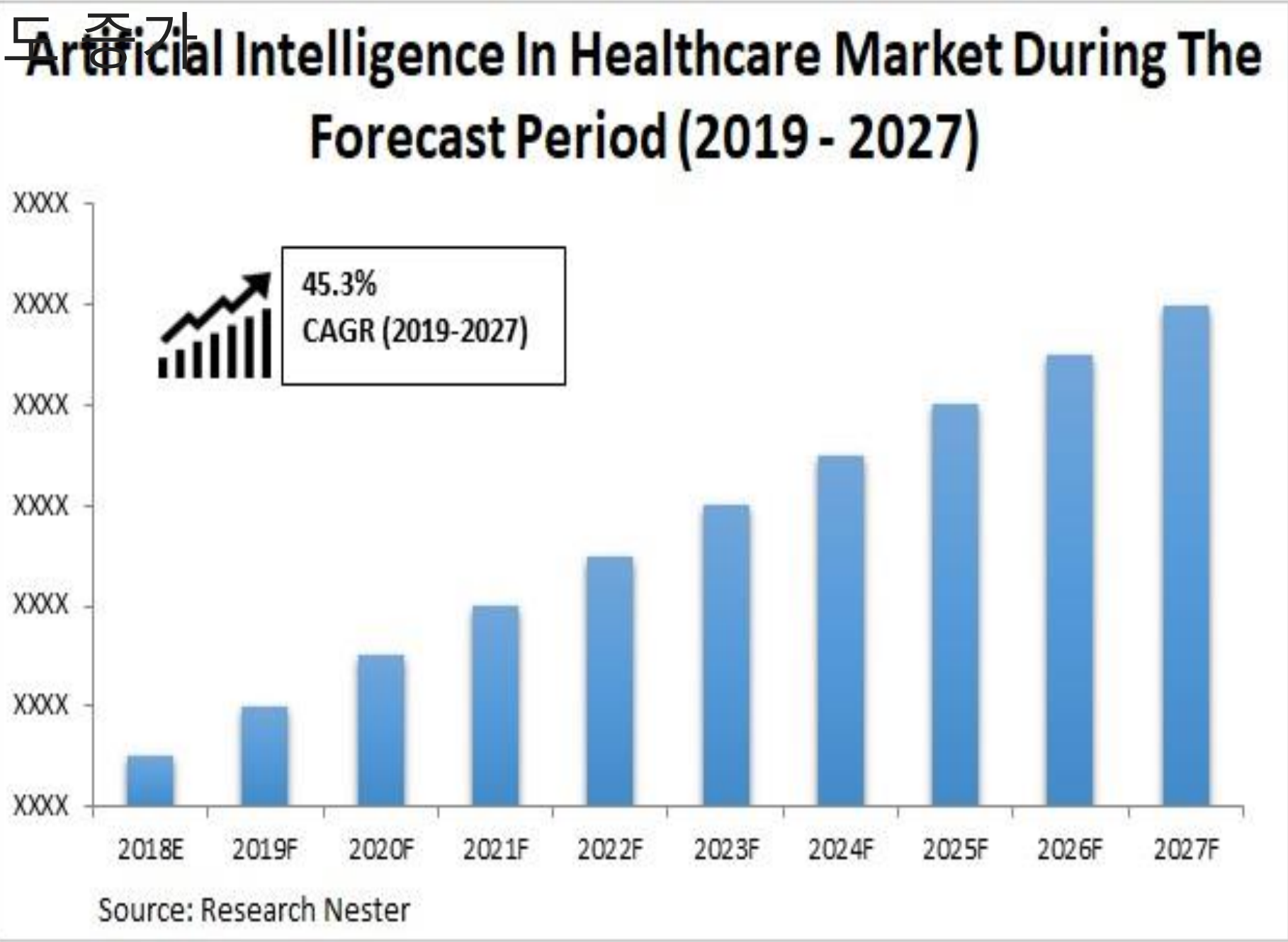
- 가명화 처리가 필요한 ‘건강’ 정보의 범위를 정의함
- 주요 의료 데이터에 대해 가명처리 방법을 정의함
 - 환자 식별자: 삭제 또는 일련번호로 대체하면 가명화 인정
 - 물리적, 화학적 측정 과정을 통해 만들어진 정형화 정보
(예. 체중, 키, 혈압, 혈당, 산소포화도, 심전도, 걸음 수 등)
: 별도 가명화 조치하지 않아도 됨
 - 영상 데이터, 유전체 데이터: 제한적인 형태로 사용 가능 (가이드라인 참조)
 - 음성 데이터: 현존 가명화 기술이 완벽하지 않으므로 본인 동의 기반으로만 사용, 등.
- 의료기관 내 데이터 심의위원회 운영

현재 1차 개정 완료하였으며, 앞으로 의료 현장의 니즈를 담아 지속적으로 개정 운영될 예정임

✓ 의료 인공지능 산업 트렌드

EMR, 보건의료 데이터, 딥러닝 활용도 증가

헬스케어 업계 또는 병원이 환자 질병을 예측하고 관리하는데 정밀 데이터 분석에 대한 의존



AI 헬스케어 시장규모

AI Total Earnings Calls Mentions by sectors, 2018-19
Source: Prattle, 2019.

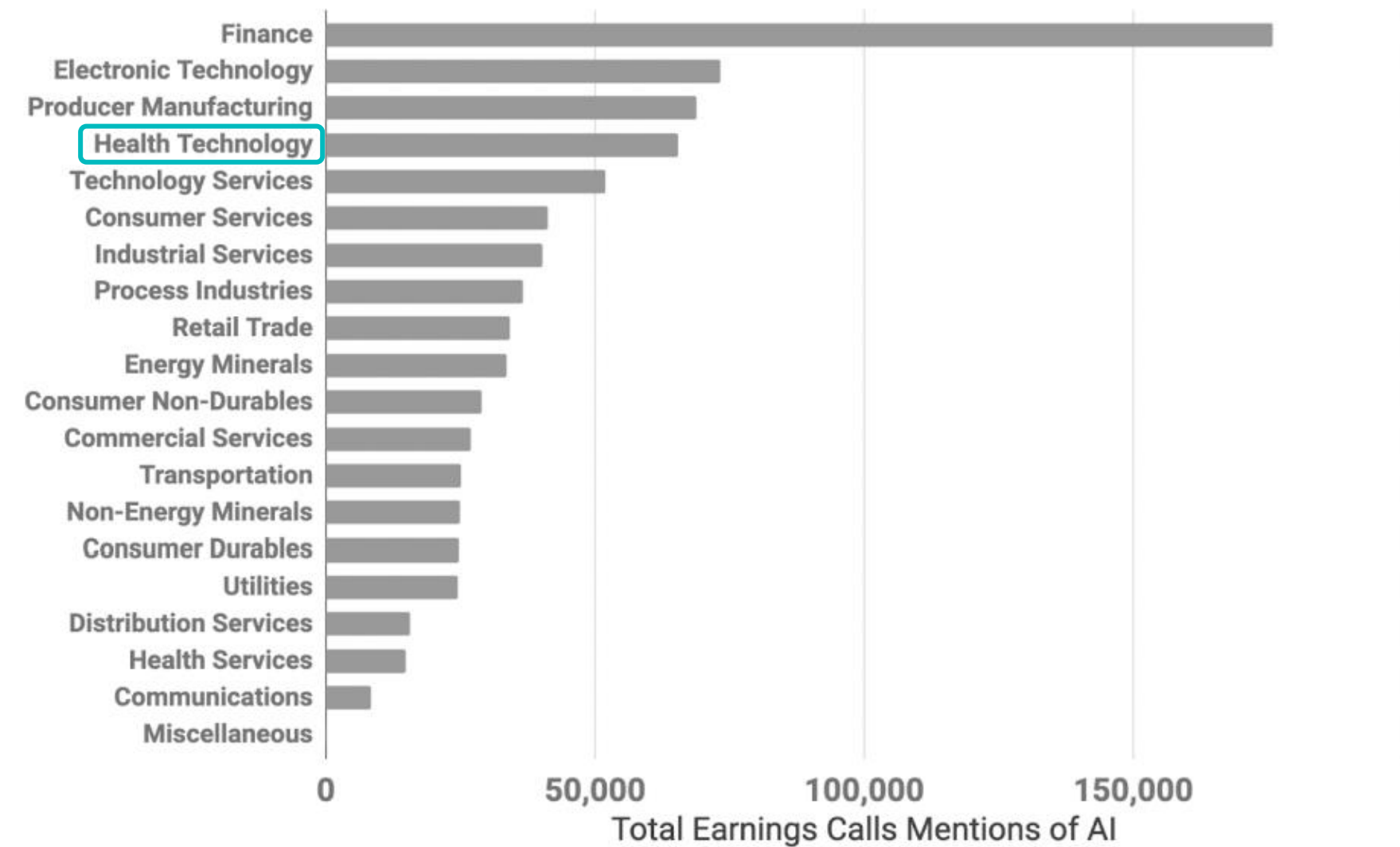


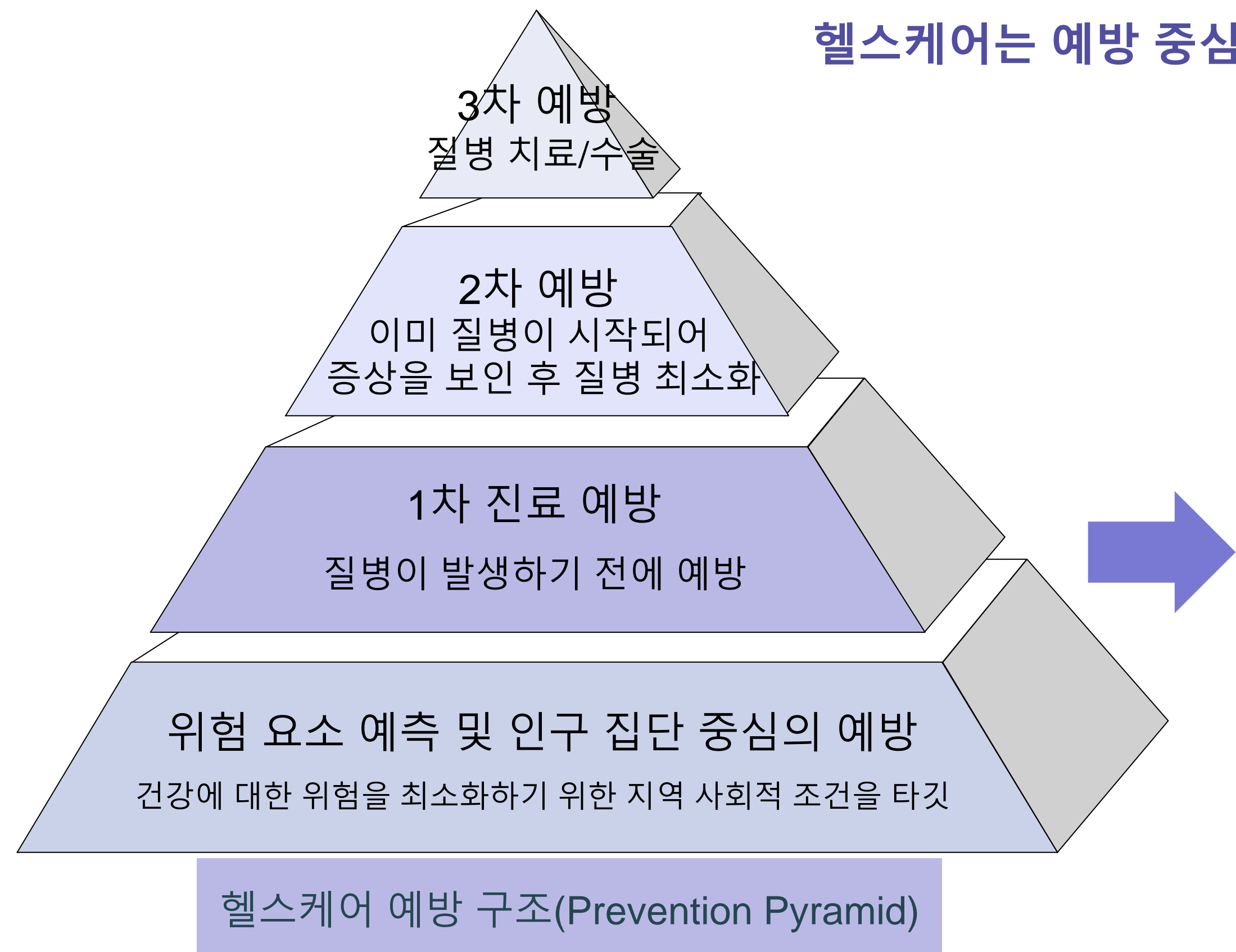
Fig. 7.4b.

산업별 AI 어닝콜 수

Earning Call(어닝콜): 주식시장에 상장된 기업이 분기별로 실적을 발표하고, 이후 기업 운영의 전망을 내놓는 행사

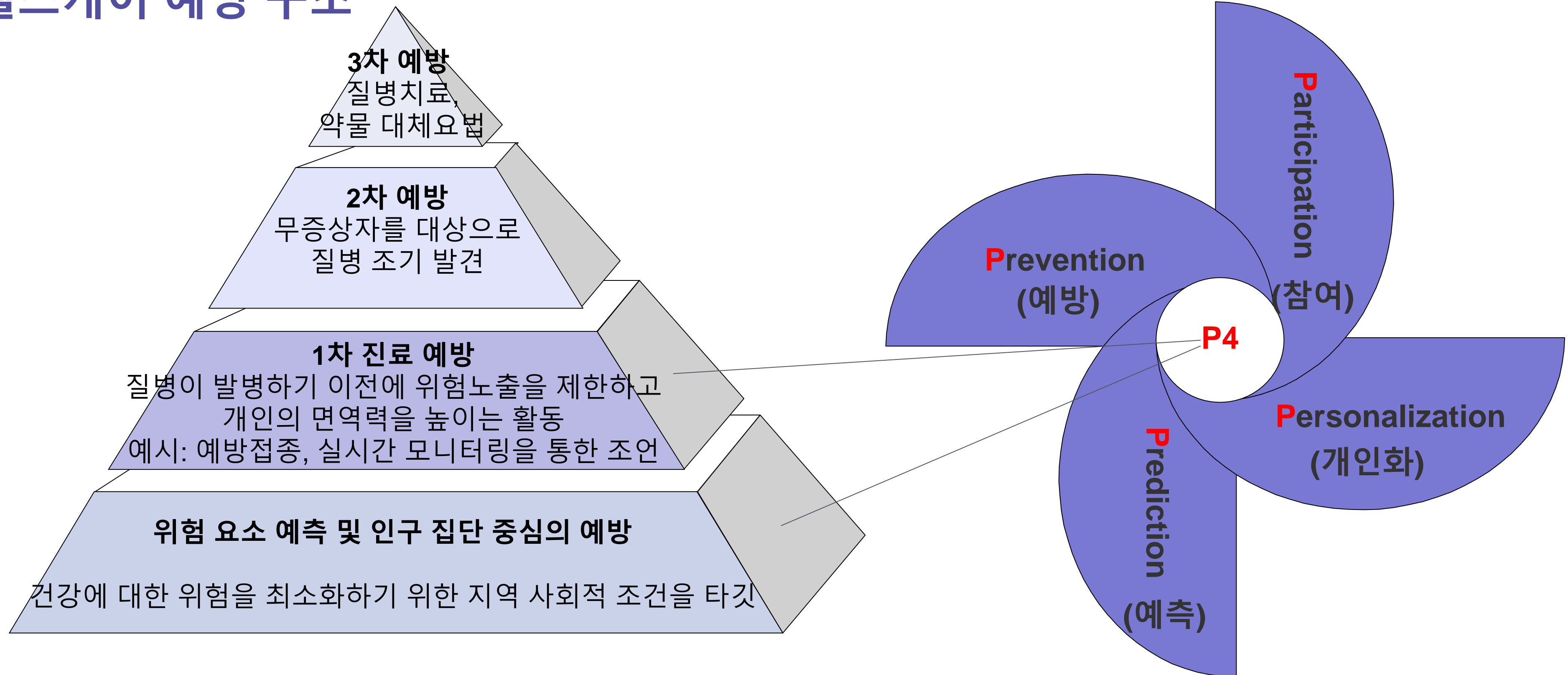
✓ 헬스케어 예방 구조

의료 빅데이터·인공지능 도입과 데이터 3법 통과 이후
헬스케어는 예방 중심으로 전환



개인 유전체 분석 정보와 과거병력, 치료전력, 생활 습관 등 환자 유래 데이터의 수집과 분석을 통해 가장 적합한 약물과 치료방법을 선택할 수 있게 된다.

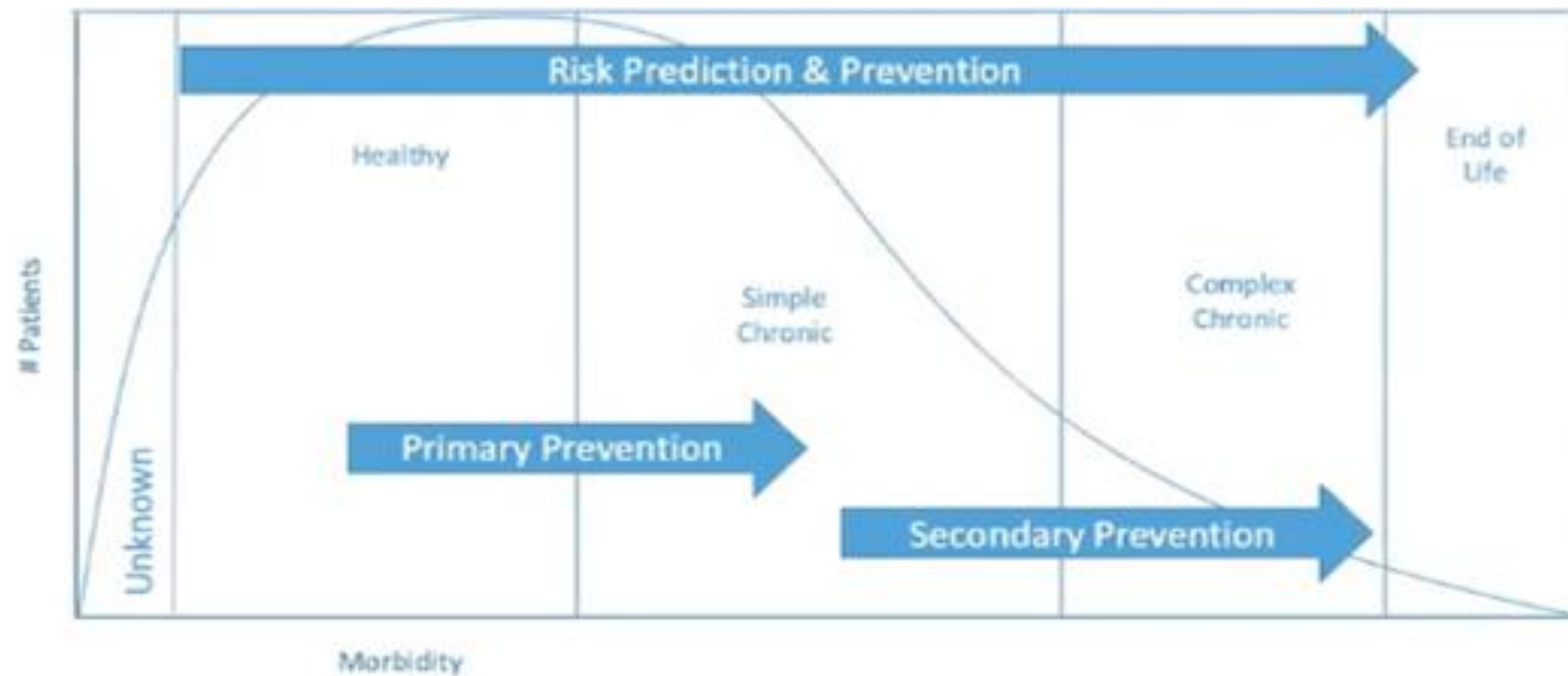
✓ 헬스케어 예방 구조



21세기 의료 서비스는 과거의 치료 중심에서 **예방(Prevention)**으로, 환자가 직접 의료 과정에 **참여(Participation)**하고, 개인의 데이터에 따라 분류하고 맞춤형 의학적인 결정을 내리거나 처방하는 **맞춤(Personalization)** 의료, 질병의 예후를 **예측(Prediction)** 하는 예측 중심의 의료로 변화하고 있다.

✓ 치료 중심에서 예방 중심으로

임상 헬스케어 모델의 중요성: 예방 >>> 치료



- Yesterday: Excellent post-hip fracture care
- Today: Proactive case management of fall risk factors
- Tomorrow: Management of elder frailty as early driver of fall risk

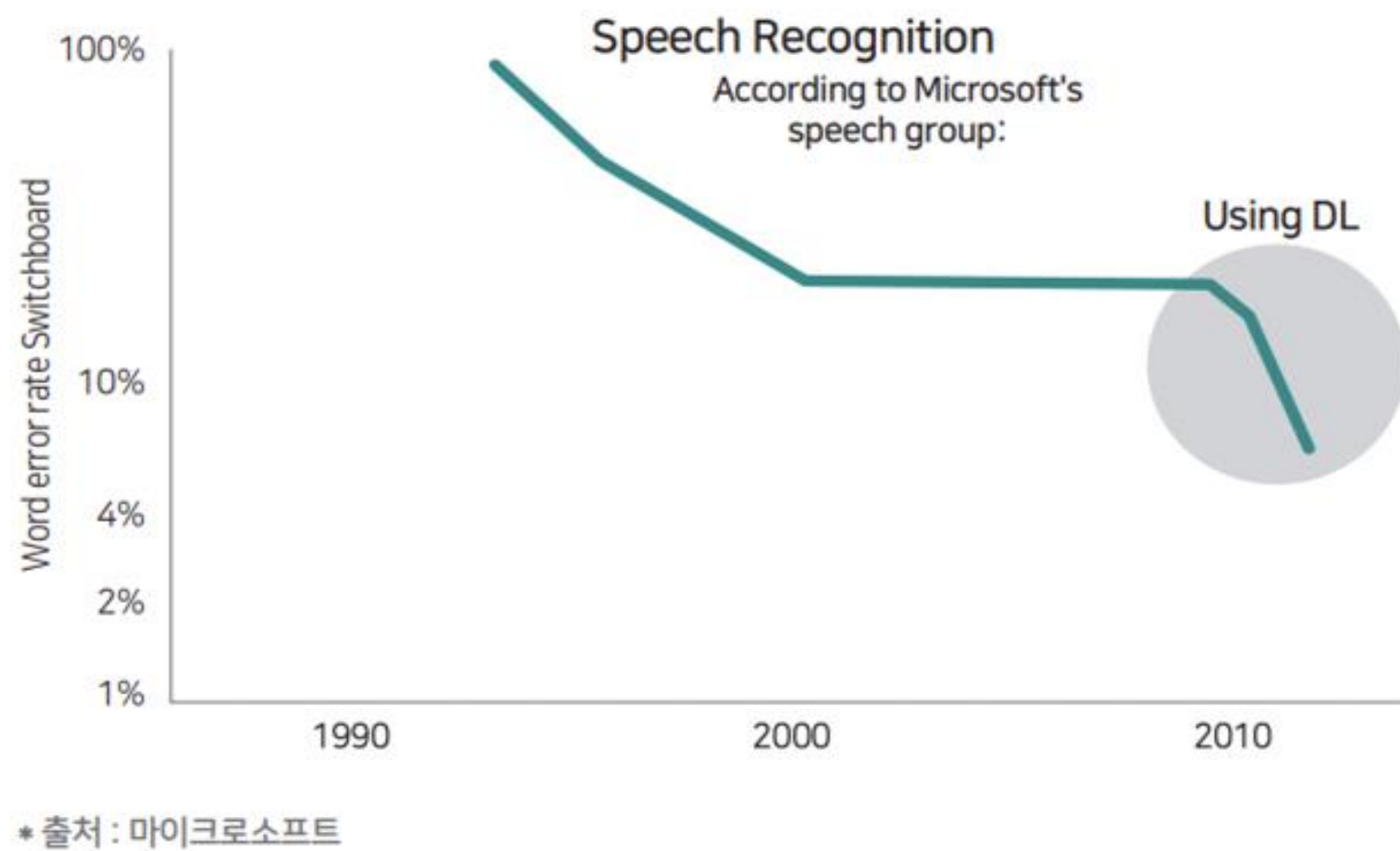
02

의사에게 딥러닝이란...

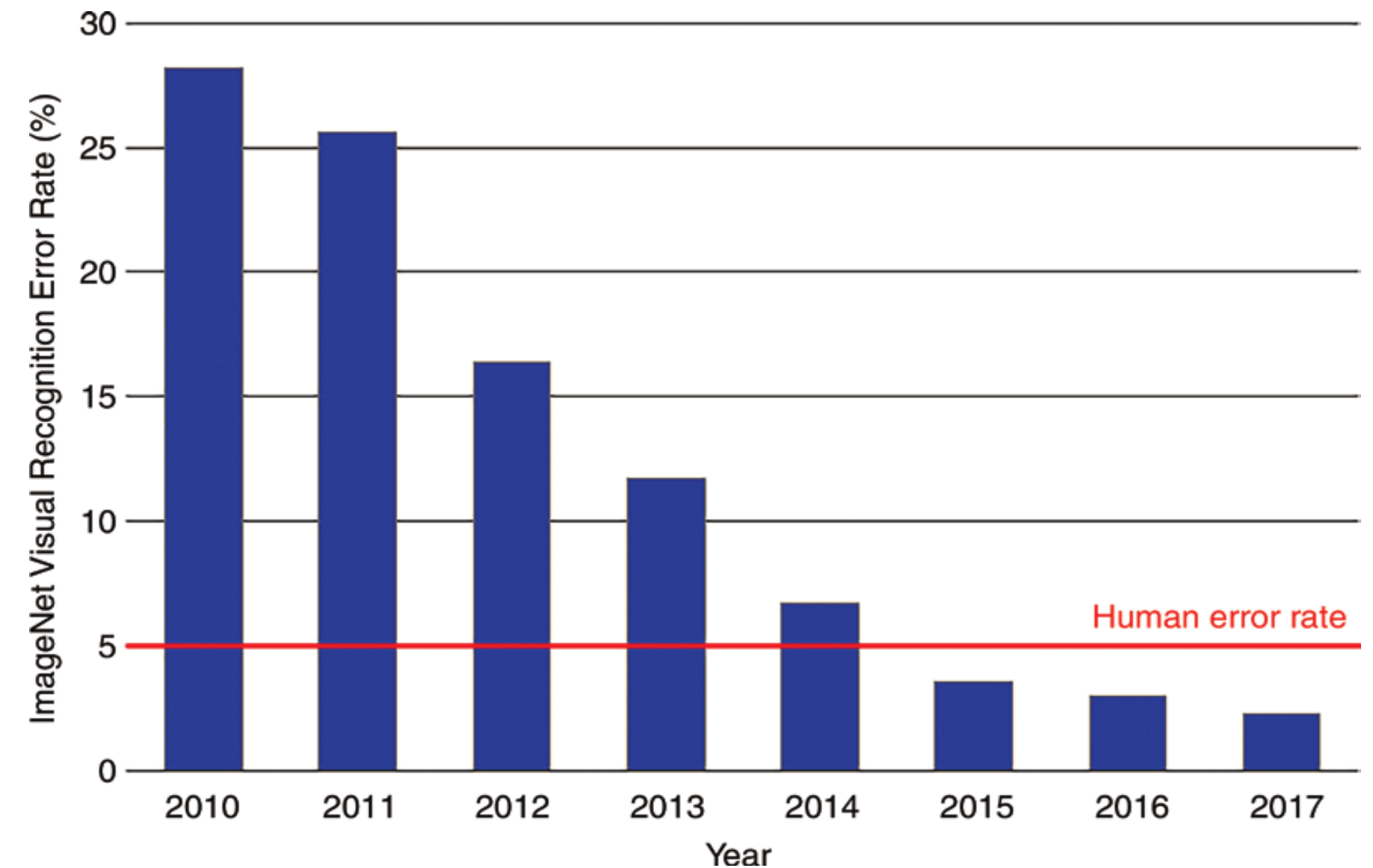


✓ 딥러닝 기술의 발전

- 1) 딥러닝 기술 + 의료 기기 발전으로 의료 데이터가 질적, 양적으로 크게 개선
- 2) 의료 분야에 딥러닝 및 머신러닝 방법론이 확장



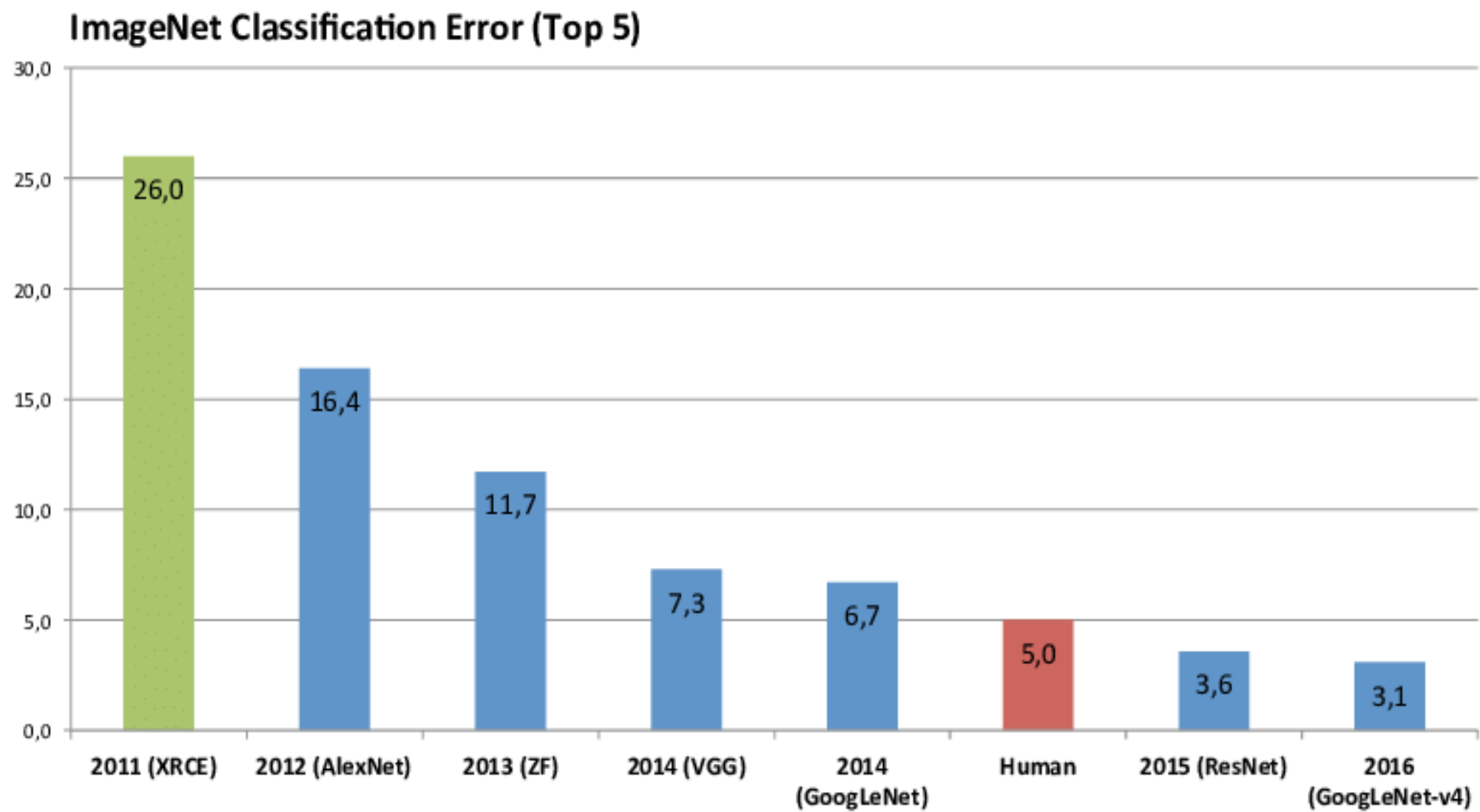
마이크로소프트의 딥러닝 음성 인식 에러율 감소



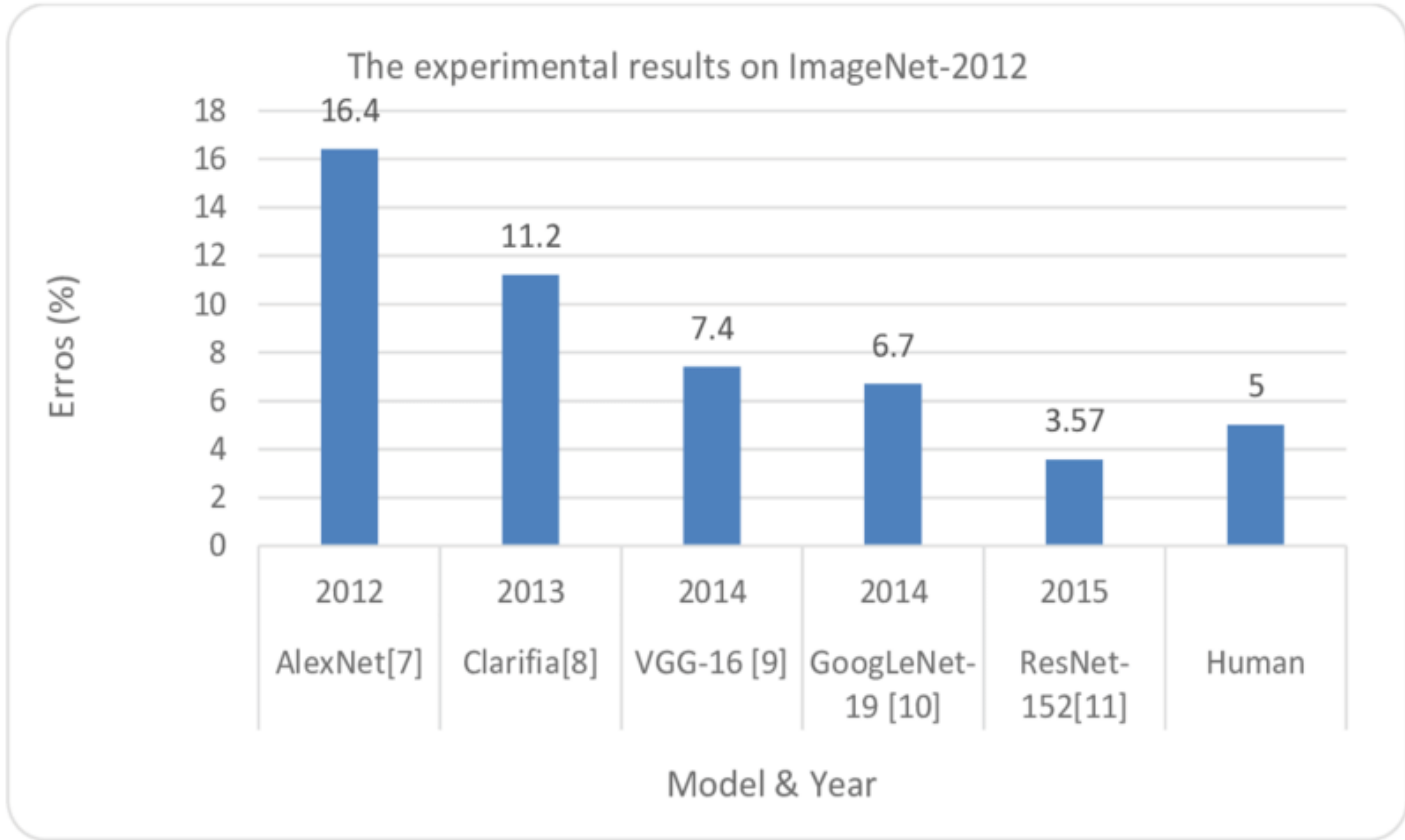
이미지넷의 이미지 인식 에러율 감소

✓ 딥러닝 이미지 판독 기술의 발전

의료 기기 및 디지털 기술의 발전이 딥러닝 이미지 판독 기술을 이끔



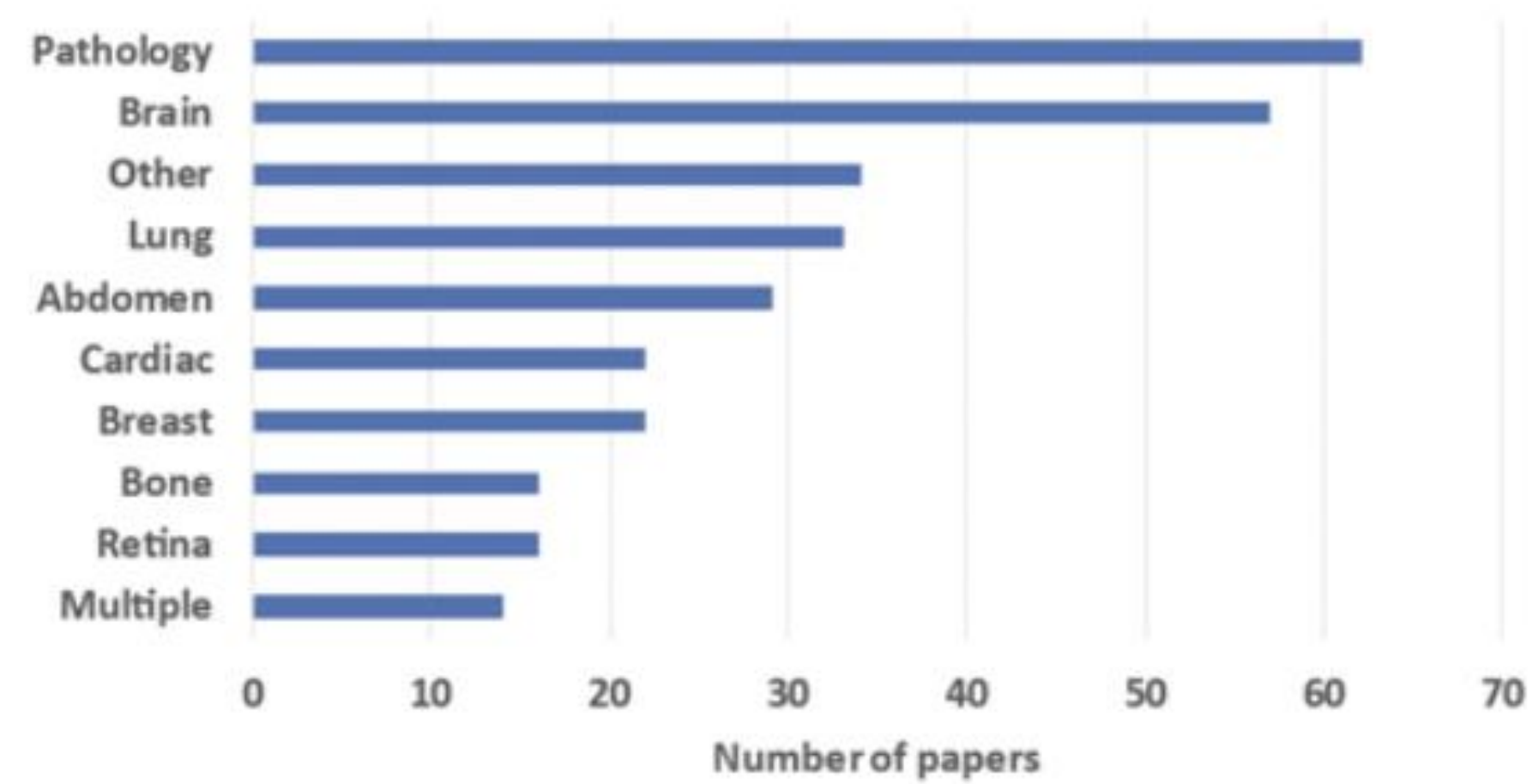
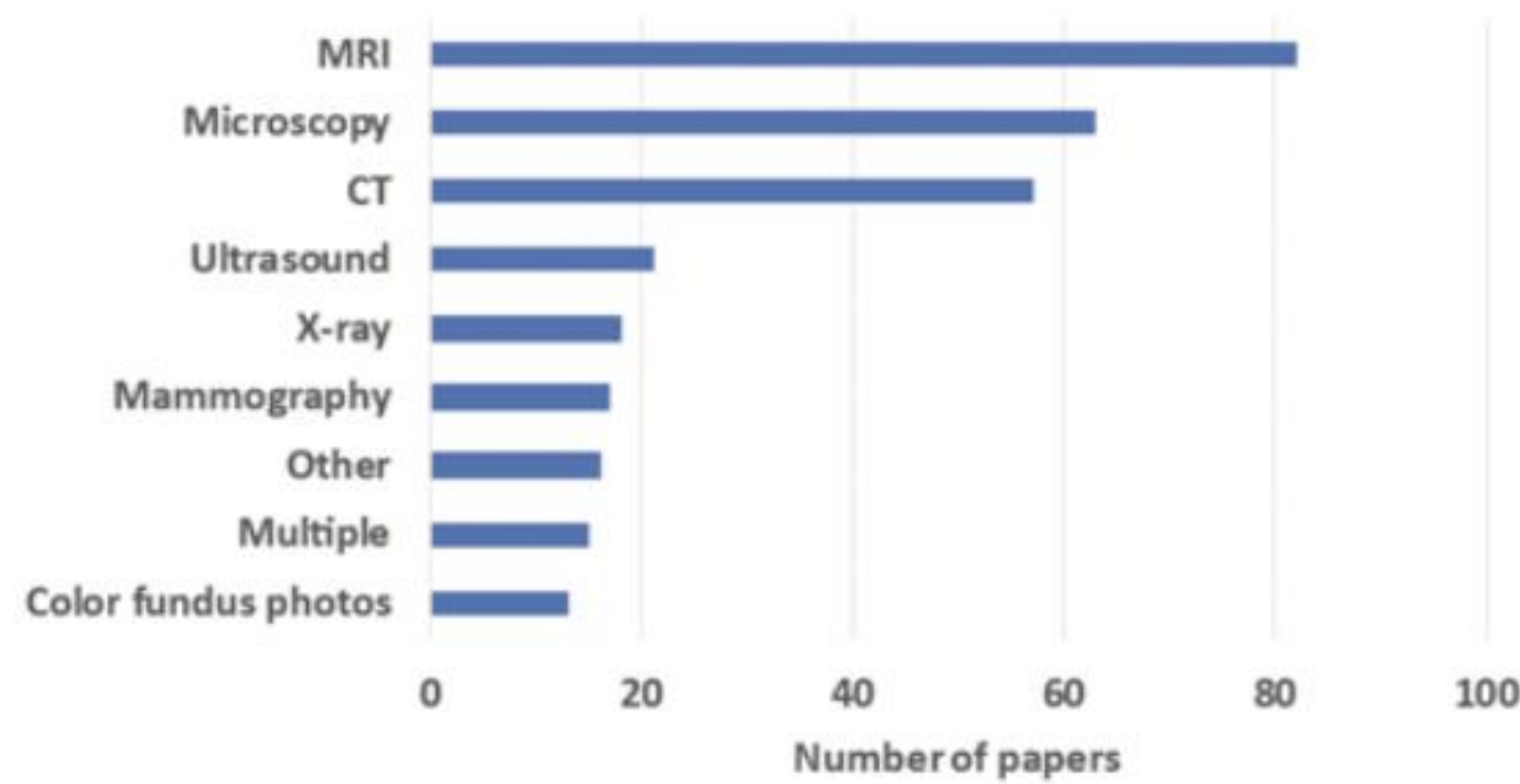
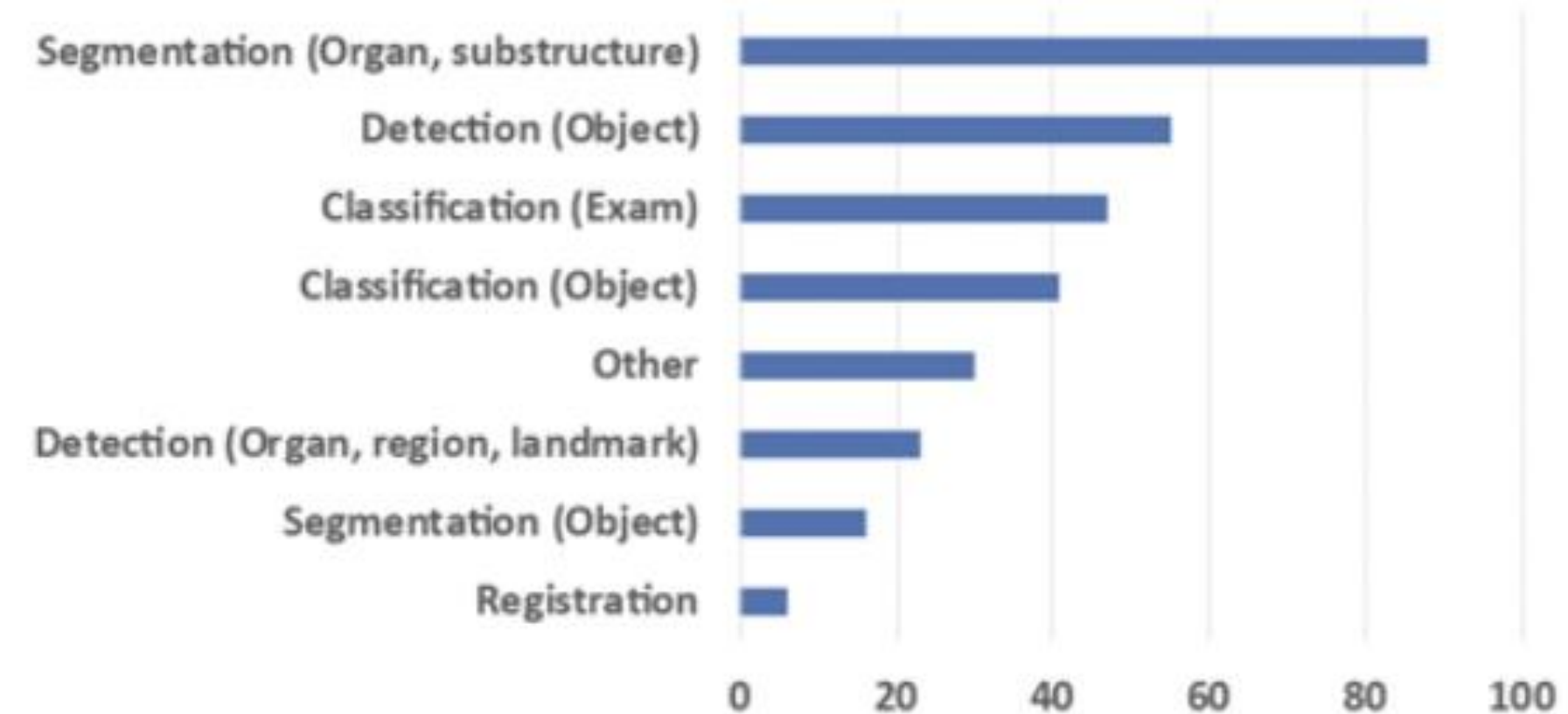
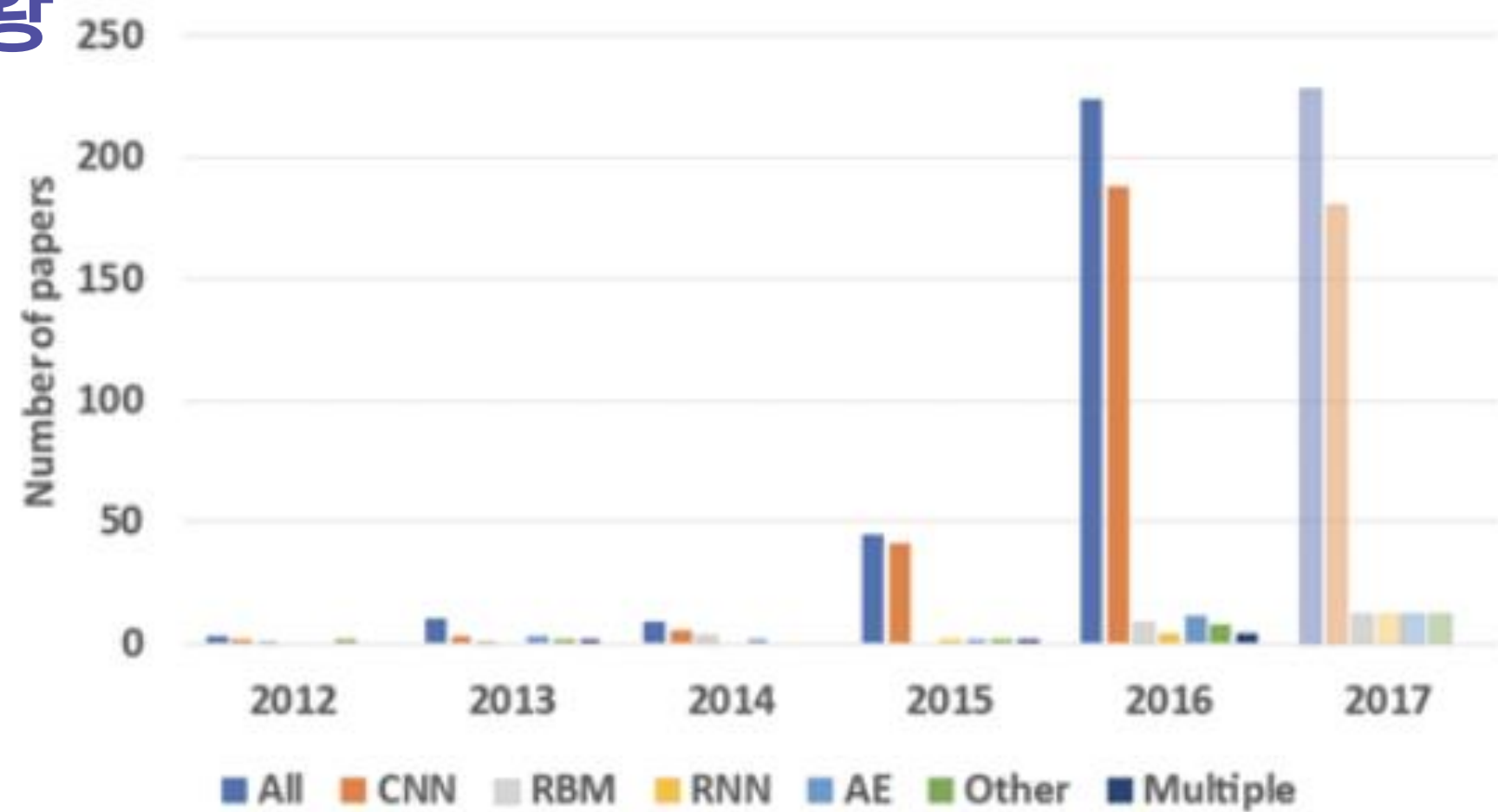
- 이미지넷 알고리즘의 분류 에러율 감소



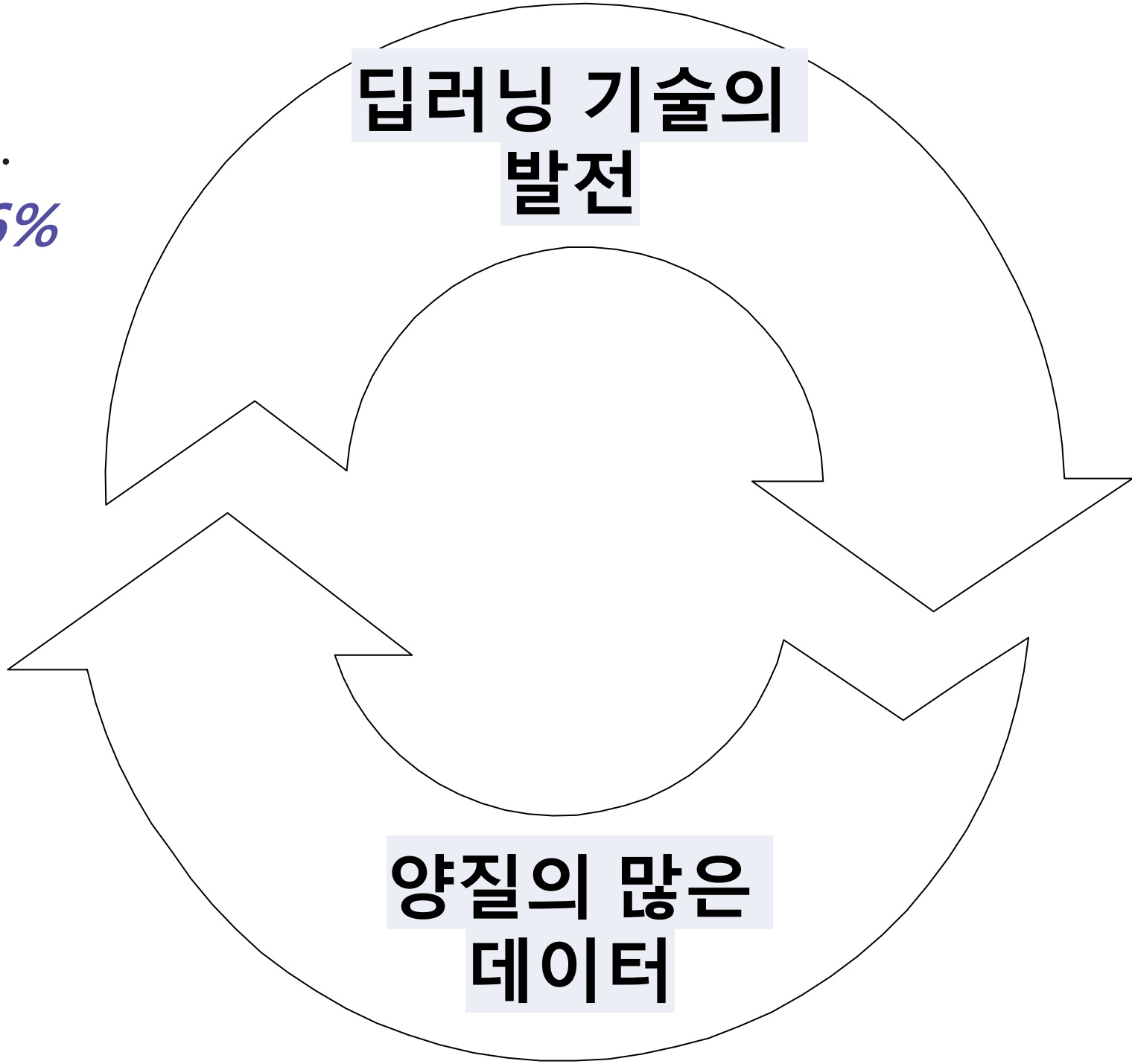
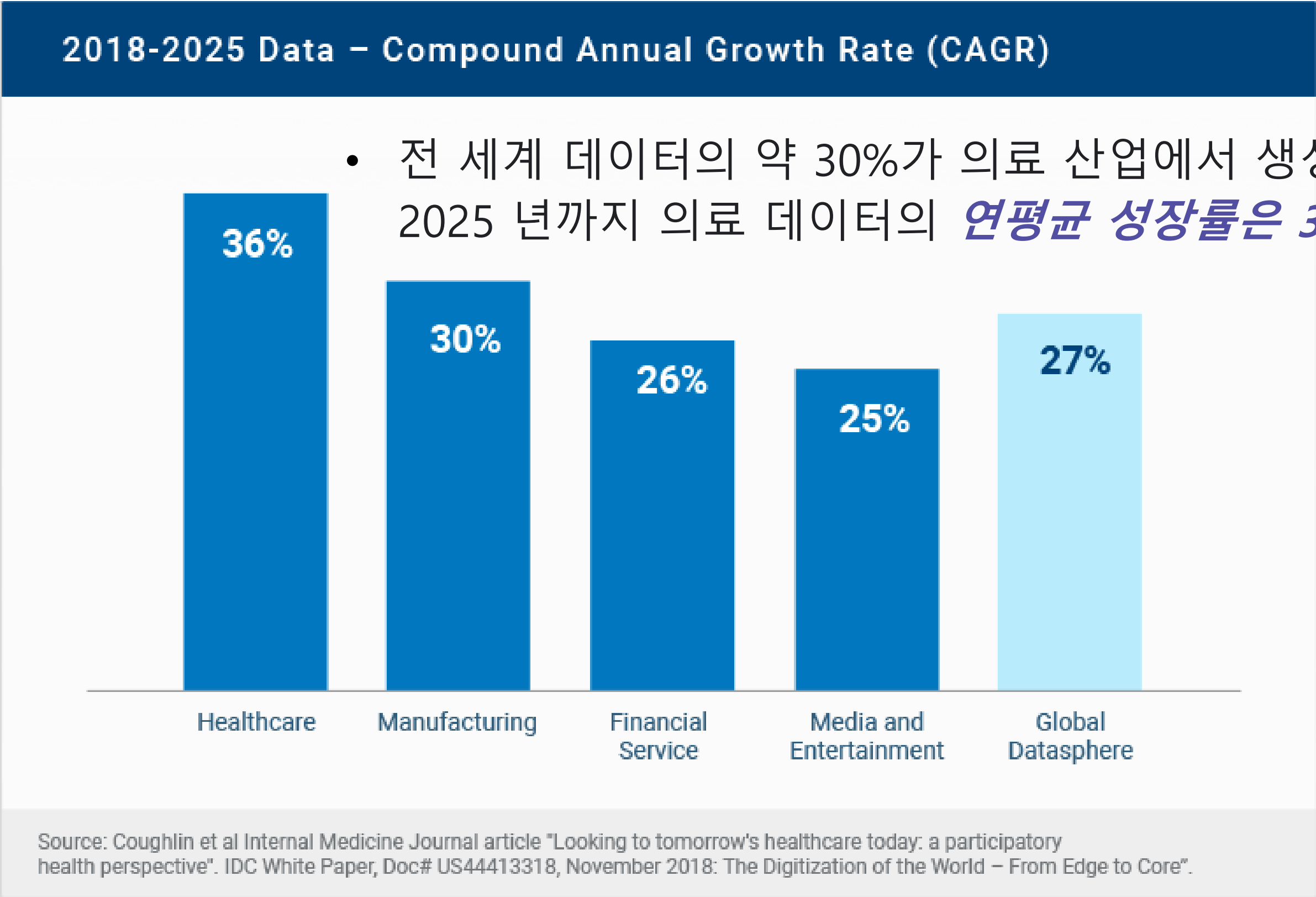
- 사람의 인식 능력과 비교

✓ 딥러닝을 이용한 영상 의료 데이터 분석 연구 현황

출처: <https://www.sciencedirect.com/science/article/pii/S1361841517301135>



✓ 의료데이터의 증가와 딥러닝의 상보성



산업별 데이터 수집 및 활용률

✓ 의사의 역할 VS 인공지능의 역할

인공지능이 인간을 초월하는 역량: 패턴 인식, 지식 분류, 암기, 판별 (인공지능의 역할)



앞으로 의사와 병원에게 더욱 요구되는 것(의사의 역할)

1

인공지능의 도움을 받아
환자에게 더 나은 의료치료를
제공할 방안을 제시

2

인공지능 분석 기반 치료의 효과는
높이면서 의료 비용도 가능하면
낮추는 방안 제시하는 의료 서비스
컨설팅

3

의료 데이터 리터러시

✓ 의사에게 요구되는 역량





인공지능을 언제 활용하며, 어떠한 기준과 원칙 아래에 활용되는지에 따라 환자에 미치는 역할, 의학적인 효용, 치료에 미치는 파급력이 달라질 수 있다.

- 데이터를 객관적으로 분석하고 활용하기 위한 **의료 데이터 리터러시**
- 인공지능 분석 결과를 바탕으로 효율적인 의학적 최종 의사결정을 내리기 위한 **딥러닝에 대한 이해**
- 진료 현장에서 **인공지능을 효과적으로 활용할 수 있는 능력**

✓ 의료 인공지능(머신러닝)의 정의

머신러닝: 데이터 학습을 통해 자동으로 성능을 개선하는 **generic** 알고리즘

	 일반 의학 통계 (Biostatistics)	 머신 러닝 (Machine Learning)
접근방식	가설 중심의 모델 개발	데이터에서 학습한 시스템 생성
목표	추론, 변수들의 상관관계	최적화, 예측 정확성 높이기
가설	주어진 데이터 관련 기본 지식	필수로 하지 않음
데이터 복잡성	대체로 범주/특징 등이 단순함	범주/특징이 매우 복잡함
성공의 기준	새로운 예측에 응용할 수 있는 모든 발견 들	기존 지식 또는 선입견을 배제한 오로지 데이터 기반의 정확한 예측

03

의료 데이터의 특성



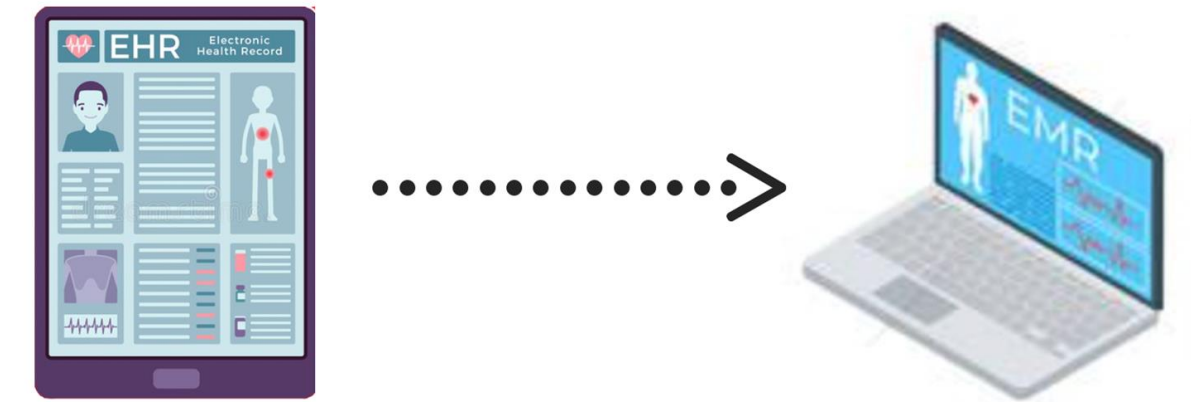
✓ 의료 데이터 종류

1. 복잡한 의료 데이터
2. 이미지 형식의 의료 데이터
3. 연속적 의료 데이터

✓ 의료 데이터 종류

- **복잡한 의료 데이터** - 전자의무기록(EMR)이나 차트에 저장된 진료기록(치료 계획 및 약물, 예방 접종 세부 사항 및 날짜) 및 진료비 청구 데이터, 유전체 데이터, 임상시험 데이터 등을 총칭
- **이미지 형식의 의료 데이터** - MRI, PET, X-ray, CT, 방사선 이미지, 초음파, 유방 촬영술, 안저 사진 등
- **연속적 의료 데이터** - 생체 신호(호흡수, 체온, 심전도, 심장박동, 혈압, 혈당, 체온, 산소포화도, 뇌파, 피부전도도 등) 의 다양한 수치들을 직간접적으로 나타낸 데이터

✓ 복잡한 의료 데이터



전자의무기록(EMR)이나 차트에 저장된 진료기록 데이터 등을 ○○

- 병력 발병 기록
- 치료 계획
- 약물 및 알레르기
- 예방 접종 세부 사항 및 날짜
- 진료비 청구 정보
- 활력 징후 데이터
- 연령 및 체중을 포함한 개인 통계
- 유전체 데이터
- 임상시험 결과

The screenshot shows a medical software interface titled "Handy patients enterprise edition". The interface includes a patient profile for David (8 months and 10 days old), a list of appointments, a "Digestive" section with inspection, auscultation, and palpation findings, and a diagram of the digestive system with red arrows indicating areas of concern. The interface also includes a "Diagnosis" section, a "Notes" section, and a "Documents" section.

https://en.wikipedia.org/wiki/Electronic_health_record

✓ 복잡한 의료 데이터(EMR) 특징

- 개인의 진단 기록, 신체 검사 및 약물 처방 기록, 임상 테스트 검사 기록 등 방대한 데이터를 처리, 분석, 저장, 안전한 접근 권한 부여 및 활용을 위해 서버, 통신망 저장공간 같은 IT 인프라가 중요

장점

- 환자를 추적하는 데 유용하며 개인 유전 정보와 결합할 시 효용성 증가
- 시간 경과에 따른 추세를 추적하거나 후속 치료가 필요하거나 예방 접종과 같은 필요한 치료가 뒤쳐진 환자를 빠르게 살펴 보는데 유용함
- 불필요한 건강진단 최소화, 헬스케어 비용 절감 (수술 횟수, 입원 시간 등)

✓ 이미지 형식의 의료 데이터

MRI, PET, X-ray, CT, 방사선 이미지, 초음파, 유방 촬영술, 안저 사진 등

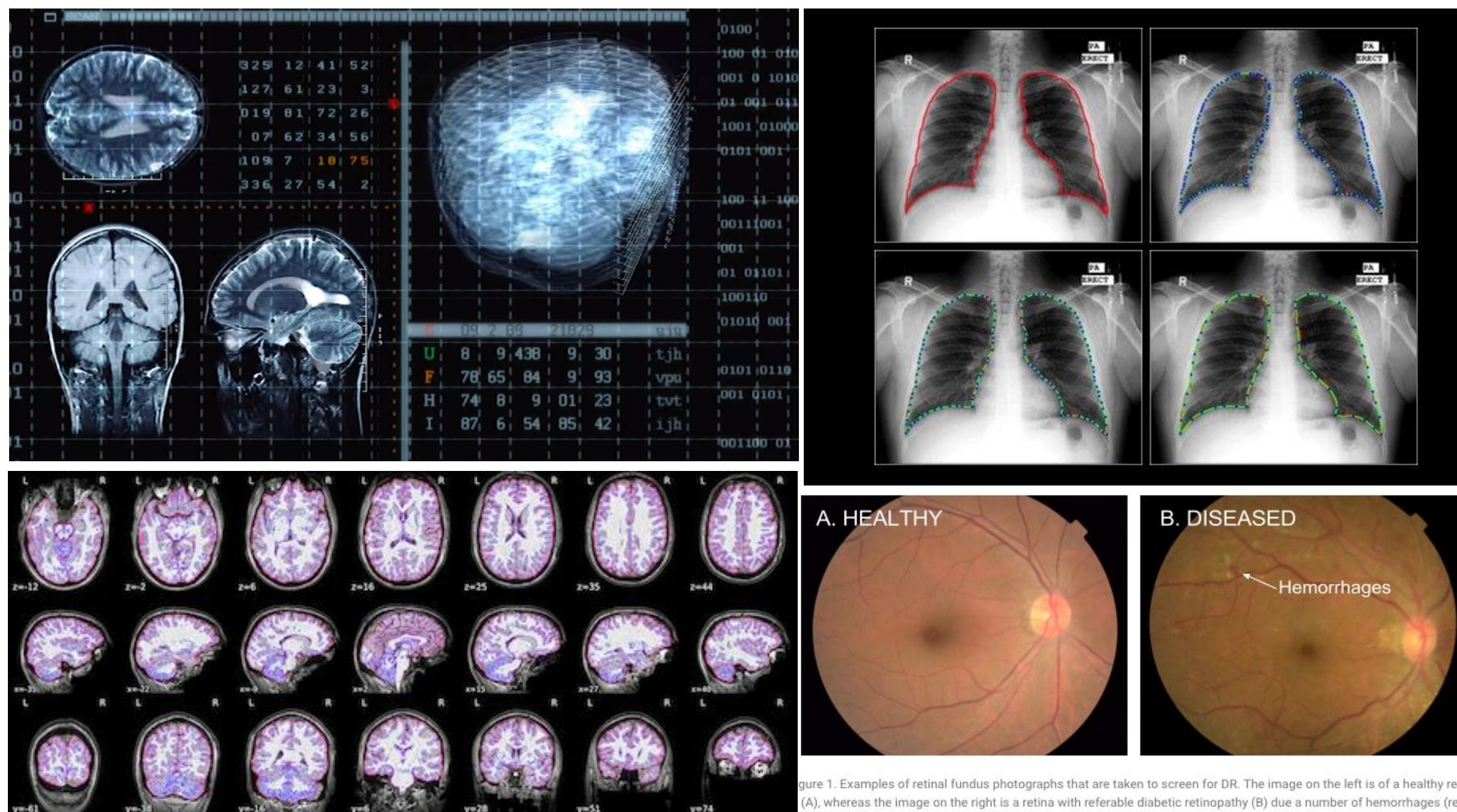
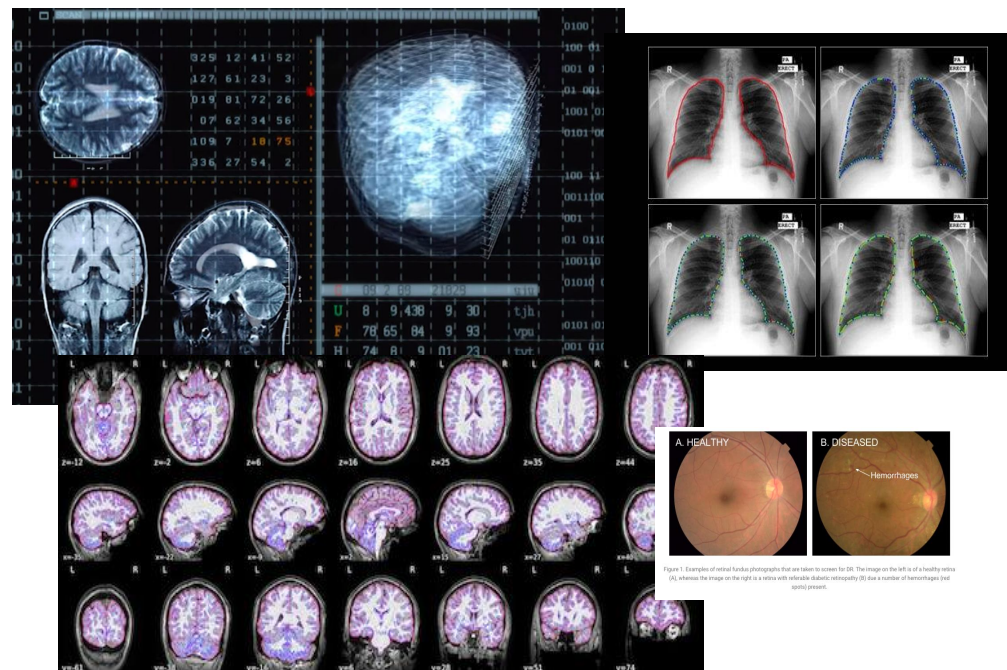


Figure 1. Examples of retinal fundus photographs that are taken to screen for DR. The image on the left is of a healthy retina (A), whereas the image on the right is a retina with referable diabetic retinopathy (B) due a number of hemorrhages (red spots) present.

✓ PACS (의료 영상 저장 전송 시스템)

영상의학 진단장치를 통하여 진단한 영상을 디지털 상태로 획득·저장하고, 그 판독과 진료기록을 함께 각 단말기로 전송·검색을 통합적으로 처리하는 디지털 의료 영상 저장 전송 시스템.



PACS 구성 요소

- 환자 정보 전송을 위한 보안 네트워크
- 이미지 해석 및 검토를 위한 워크스테이션
- 이미지 및 보고서의 저장 및 검색

PACS 가 처리하는 의료 영상 장비 및 이미지

컴퓨터단층촬영장치(CT)

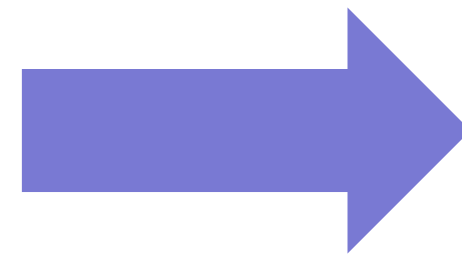
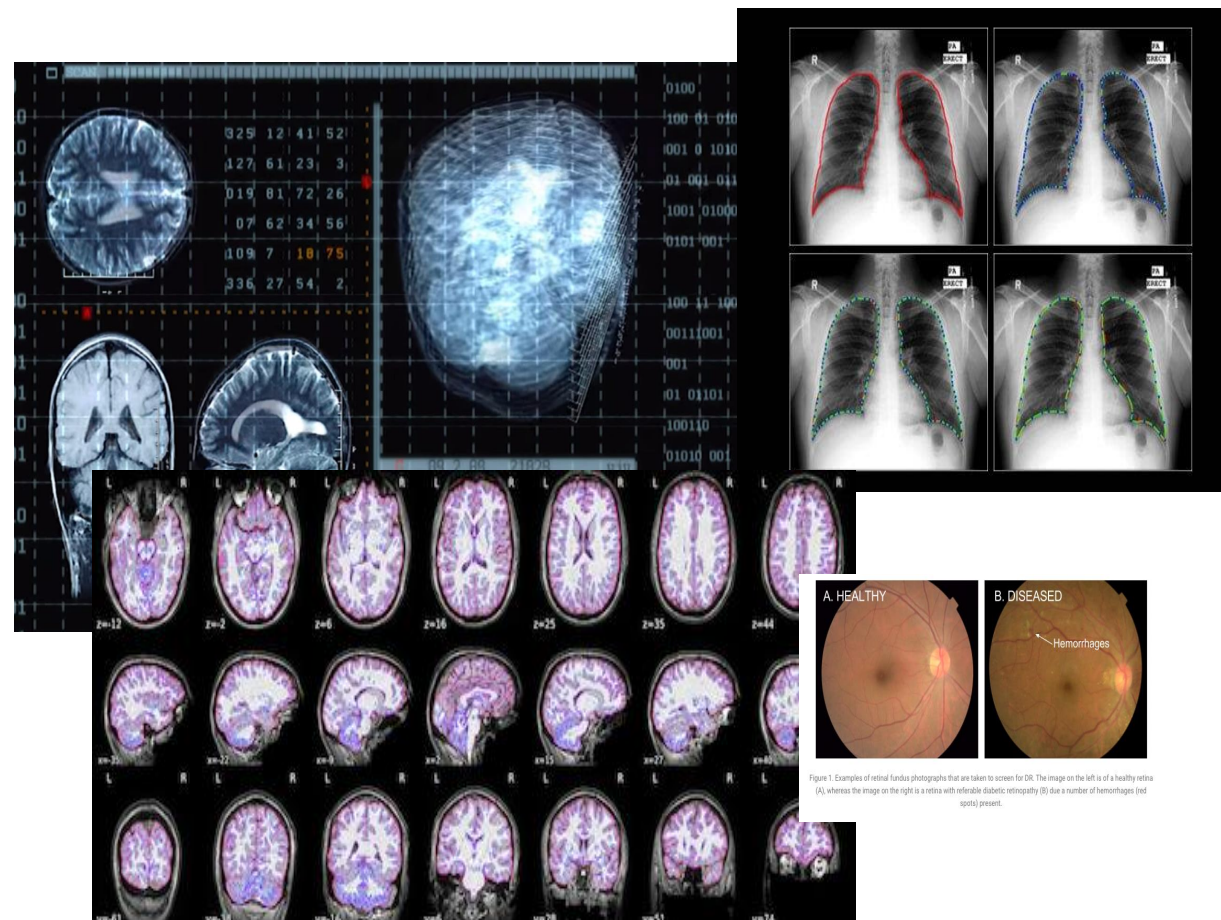
자기공명영상촬영장치(MRI)

초음파·내시경·현미경 등의 이미지

투시촬영장치·혈관조영장치·유방암검진기 등

✓ PACS (의료 영상 저장 전송 시스템)

- 필름 비용 절감, 미판독 이미지들의 감소
- 특수검사 판독 지연 감소 등의 효과
- 이미지의 동시 활용, 임상정보의 증대, 즉각적인 이미지 확보, 판독 의사와 임상 의사 간의 원활한 의사소통



영상의 디지털화 &
데이터베이스, 정보저장 관리

방사선의학뿐만 아니라 심장학, 종양학, 정형 외과 등 다양한 임상 영역에서 PACS에 통합가능한 의료 이미지 생성중

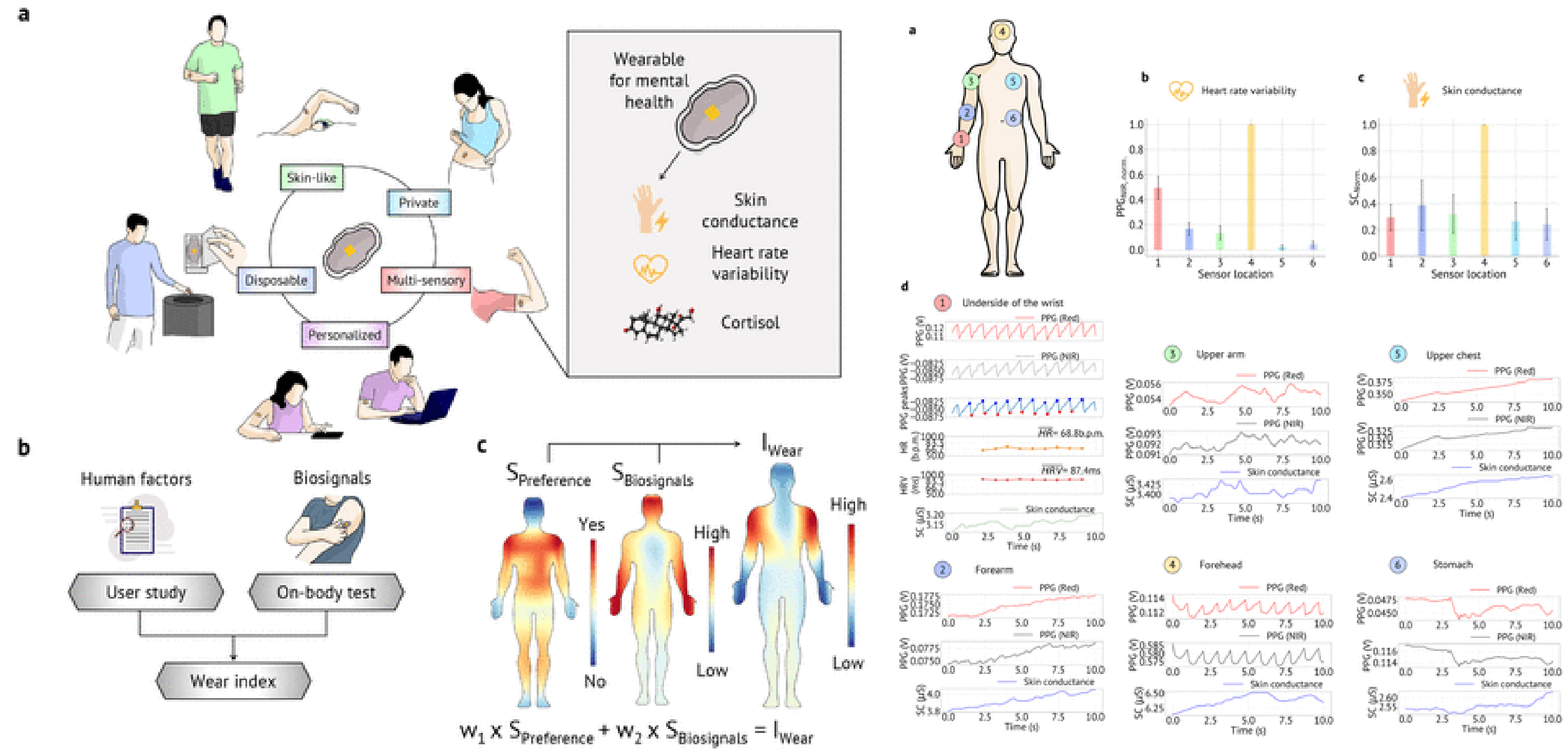
✓ 딥러닝 알고리즘에 사용되는 이미지 데이터의 조건

최적의 지도학습기반 딥러닝 알고리즘에 사용될 이미지 데이터의 조건

- 촬영술에서 보일 수 있는 모든 시각적인 증상 범위를 최대한 포함할 수 있는 대규모의 데이터
- 다양한 크기, 위치, 영상 소견들을 포함함 → 높은 수준의 정확도를 확보
- 다양한 장비, 검사 기관, 연령대, 인종 등을 폭넓게 포함한 데이터 → 딥러닝 일반화

연속적 의료 데이터

생체 신호(호흡수, 체온, 심전도, 심장박동, 혈압, 혈당, 체온, 산소포화도, 뇌파, 피부전도도, 생리적 모니터링 데이터 등)의 다양한 수치들을 직간접적으로 나타낸 데이터



✓ 연속적 의료 데이터의 특징

ECG 심전도 데이터



- 만성질환환자의 행동을 주기적으로 검시 및 향상 가능
- 대체로 다양한 병원/클리닉의 개개인 환자들에 의해 수집되어 여러군데 분포되어 있으며 민감한 개인정보를 담고 있음
- 스마트폰, 웨어러블 기기 등으로 추출됨

03

의료 데이터 관리의 중요성



✓ 데이터 리터러시란?

좁은 의미 : 개인이 통계적 수치를 비판적으로 읽는 역량

넓은 의미 : 정부 및 기업에서 데이터를 활용해 효과적으로 문제를 해결하는 역량



"빅데이터로 보다 예방적이고 예측 가능하며 개인화되고 정확한 의료 서비스를 추구할 수 있고 차세대 의사를 교육하는 전통적인 방법을 고수하는 것만으로는 충분하지 않다. 병원은 환자의 효율적인 치료를 위해 데이터 관리 기술을 발전시킬 책임이 있다."

Lloyd Minor, MD, 스탠포드 의과 대학 연구원, 교육자

✓ 의료 데이터 리터러시의 중요성

의료 데이터 소스 및 구성, 적용된 분석 방법 및 기술에 대한 이해와 사용된 케이스 및 결과 값을 설명하는 능력을 포함해서 데이터를 맥락을 이해하고, 활용하고, 통신하는 능력

- 헬스케어 업계의 데이터 의존도 증가
- 의사가 질병을 예측하고 관리하는 데 도움이 될 수 있는 정밀 데이터 분석에 대한 업계의 의존도 증가

✓ 데이터 품질의 중요성

수집된 방대한 데이터를 최대한 효율적으로 활용하기 위한 전략이 필요하다

좋은 데이터란?

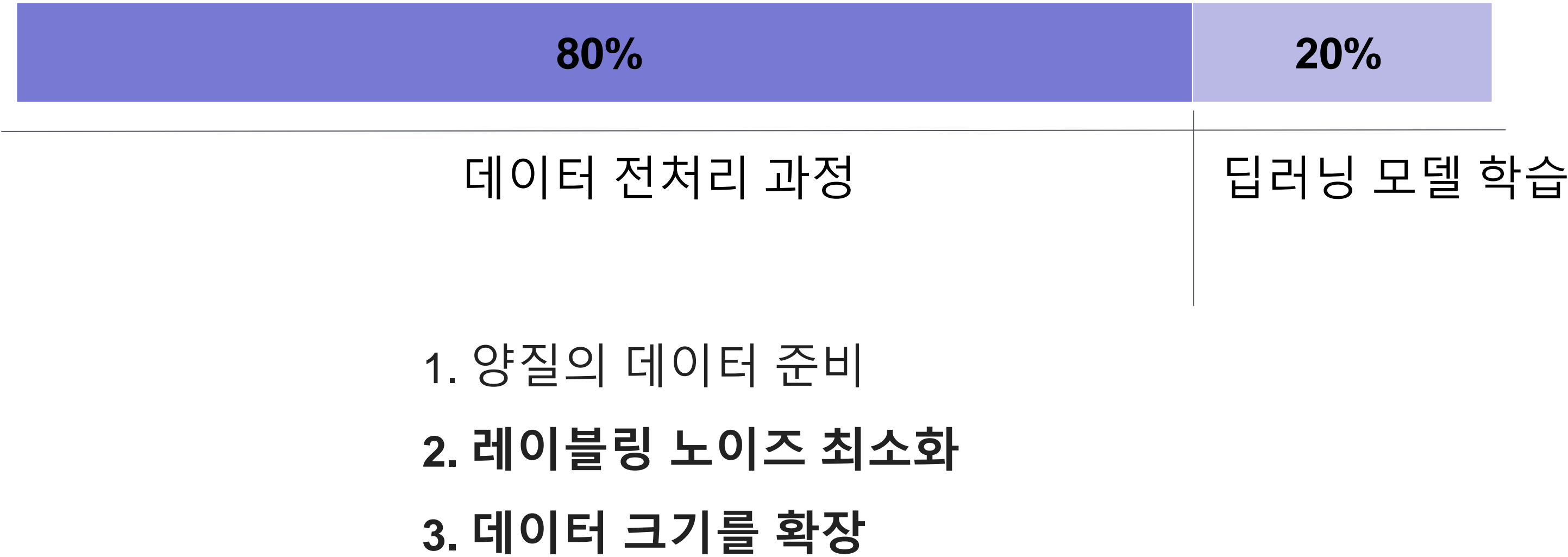
- 일관성있게 레이블 됨 (레이블의 정의가 애매하지 않고 명확함)
- 다양한 input x를 포함됨
- 일관성있는 형식의 많은 양의 데이터 (이미지라면 사이즈와 필터등이 적절하게 배분됨)
- 생산된 데이터에 대한 적시에 피드백을 받을 수 있음

✓ 코드 향상시키기 VS 데이터 퀄리티 향상시키기

딥러닝 알고리즘 개발을 위한 비용 대비 이미지 인식 효과 비교

	강철 결함 감지	태양 전지판	표면 검사
베이스라인	76.2%	75.68%	85.05%
모델 중심	+0% (76.2%)	+0.04% (75.72%)	+0.00% (85.05%)
데이터 중심	+16.9%	+3.06% (78.74%)	+0.4% (85.45%)

✔ 딥러닝 알고리즘 개발을 위한 비용 대비 효과 극대화: 데이터 전처리



✓ 데이터 전처리(1) - 레이블링 노이즈 최소화

- 레이블이 어떻게 설정되었는지에 대해 명확하게 파악하여 설정 기준을 고려해야 함
- 레이블링이 구체적이고 세밀할 수록 딥러닝 학습 효율이 높아짐
- 레이블을 추출할 때 특정 레이블에 대한 데이터의 총괄적인 검토를 반복해서 함

예: 심장 마비에 걸린 환자의 데이터 레이블을 추출할 때 퇴원 요약 노트를 검토해서 식별함

✓ 데이터 전처리(2) - 데이터 크기 확장

- 데이터 증강 기법(Data Augmentation)

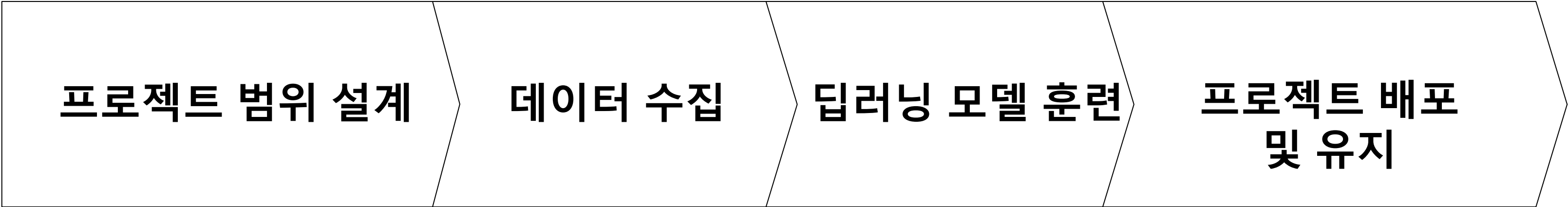
크기 조절, 반전, 자르기, 회전 이동 윈도우 분할 등으로 데이터 크기를 확장하여 과적합 방지

- **과적합(Over-fitting)이란?** : 딥러닝 모델이 너무 데이터에만 치중하여 데이터의 무작위 변동, 노이즈까지 학습하여 오차가 증가하고 일반화의 성능이 떨어지는 경우

과적합이 발생하는 경우

- 새로운 데이터에 낮은 성능을 보일 때
- 모델이 복잡할 때 - (많은 공식을 사용하여 푸는 경우)
- 모델링에 영향을 주는 변수가 많을 때

✔ 머신러닝 프로젝트 생애주기



04

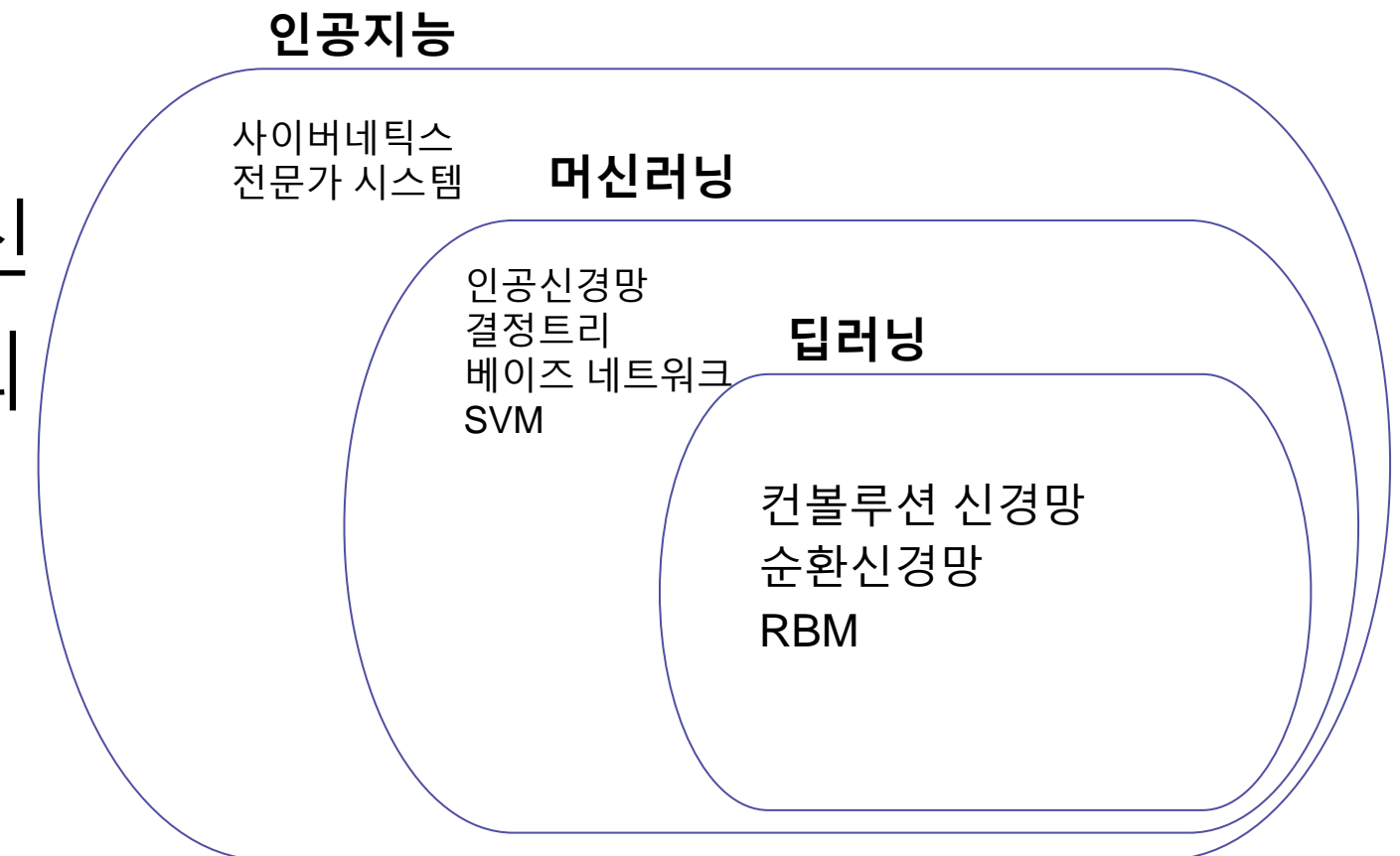
인공신경망이 의료 데이터를 학습하는 방법



✓ 딥러닝 vs 머신러닝

머신러닝: 의사 결정나무, 나이브 베이지, SVM(서포트벡터머신), 신경망 등 여러 모델을 포함하는 개념이며 입력 피쳐를 뽑는데 인간의 수작업이 필요함

딥러닝: 신경망이라는 하나의 개념을 의미하며 여러 단계의 계층적 학습 과정을 통해 적절한 피쳐(입력값)을 스스로 생성하여 학습함

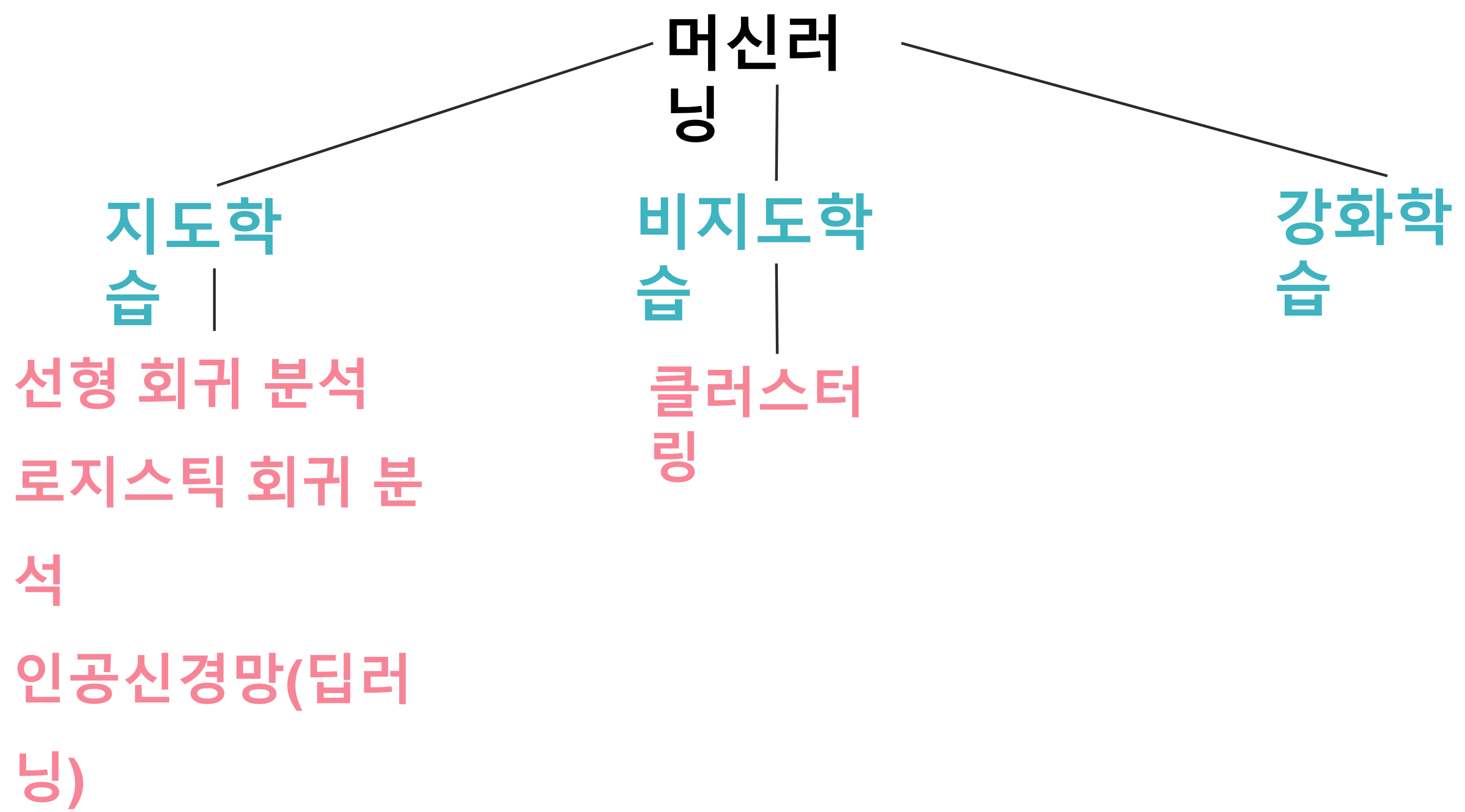


● 딥러닝과 CNN

딥러닝에는 컨볼루션 신경망, 순환신경망, 강화학습 등 다양한 방법론이 있음

그중에서 컨볼루션 신경망(CNN)이 딥러닝의 대표적인 알고리즘이며 이미지 인식에 놀라운 성능을 보임

✔ 지도학습 vs 비지도학습



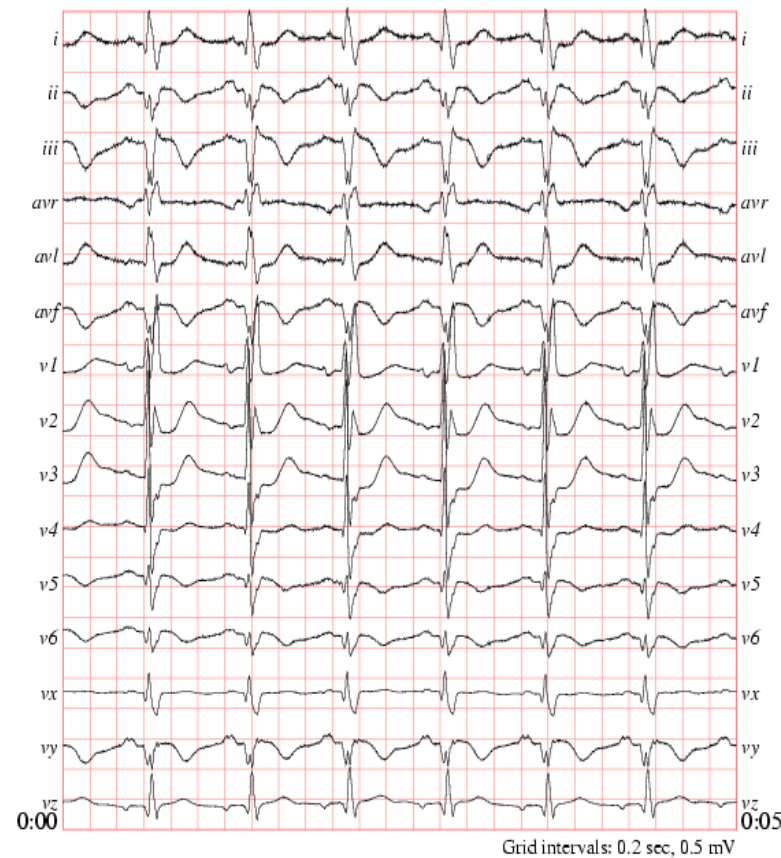
- 대부분의 의료데이터 학습의 접근방식은 지도학습

의사결정나무

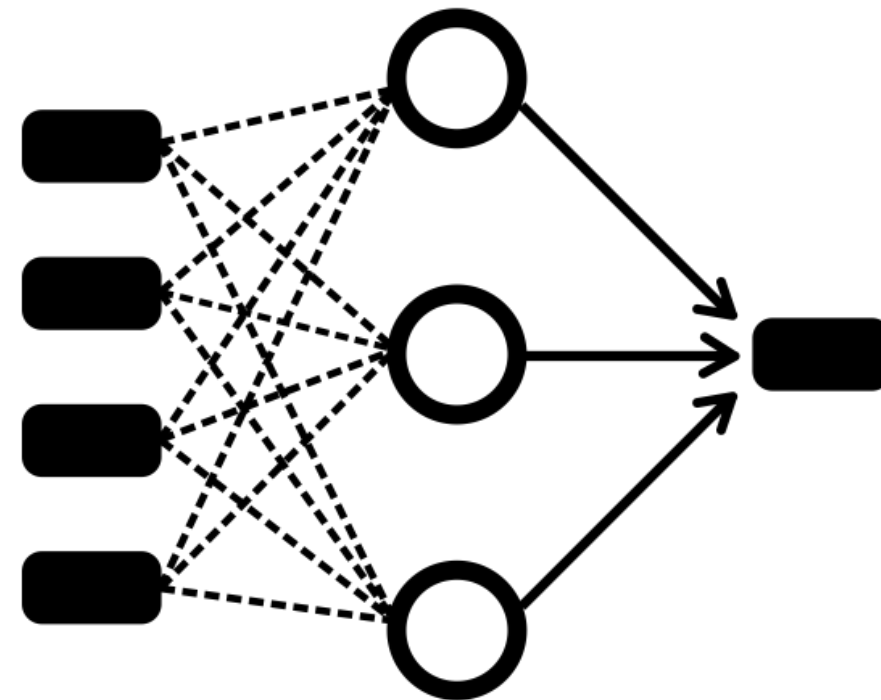
✓ 지도 학습

지도학습은 프로그램이 수많은 input 과 output 쌍을 학습하여 그 둘을 연결하는 함수를 학습함
즉 레이블이 없는 데이터를 수많은 함수를 발전시켜서 레이블을 예측함

Input
ECG (심전도 검사) 데이터



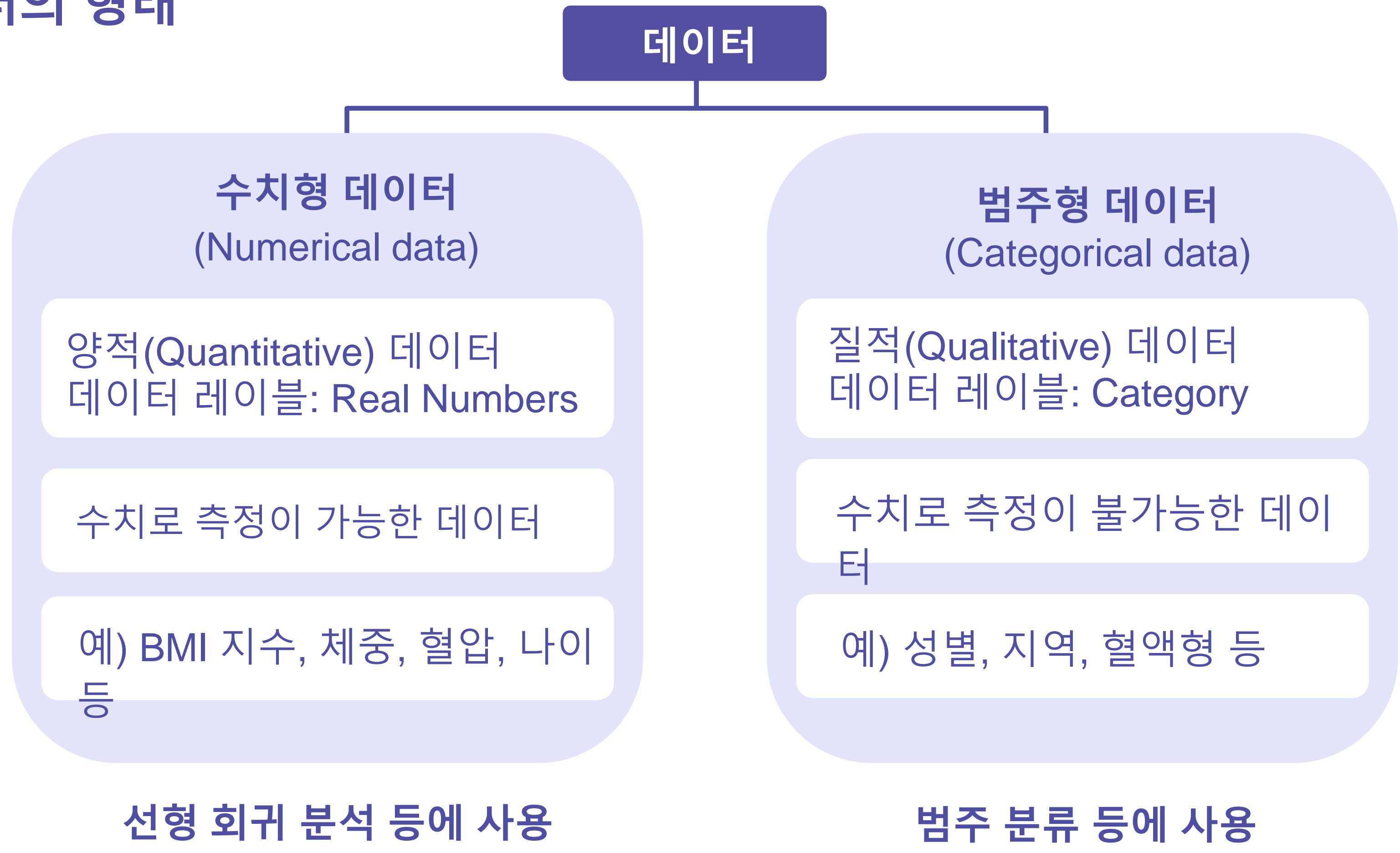
학습된 함수



심장마비 여부

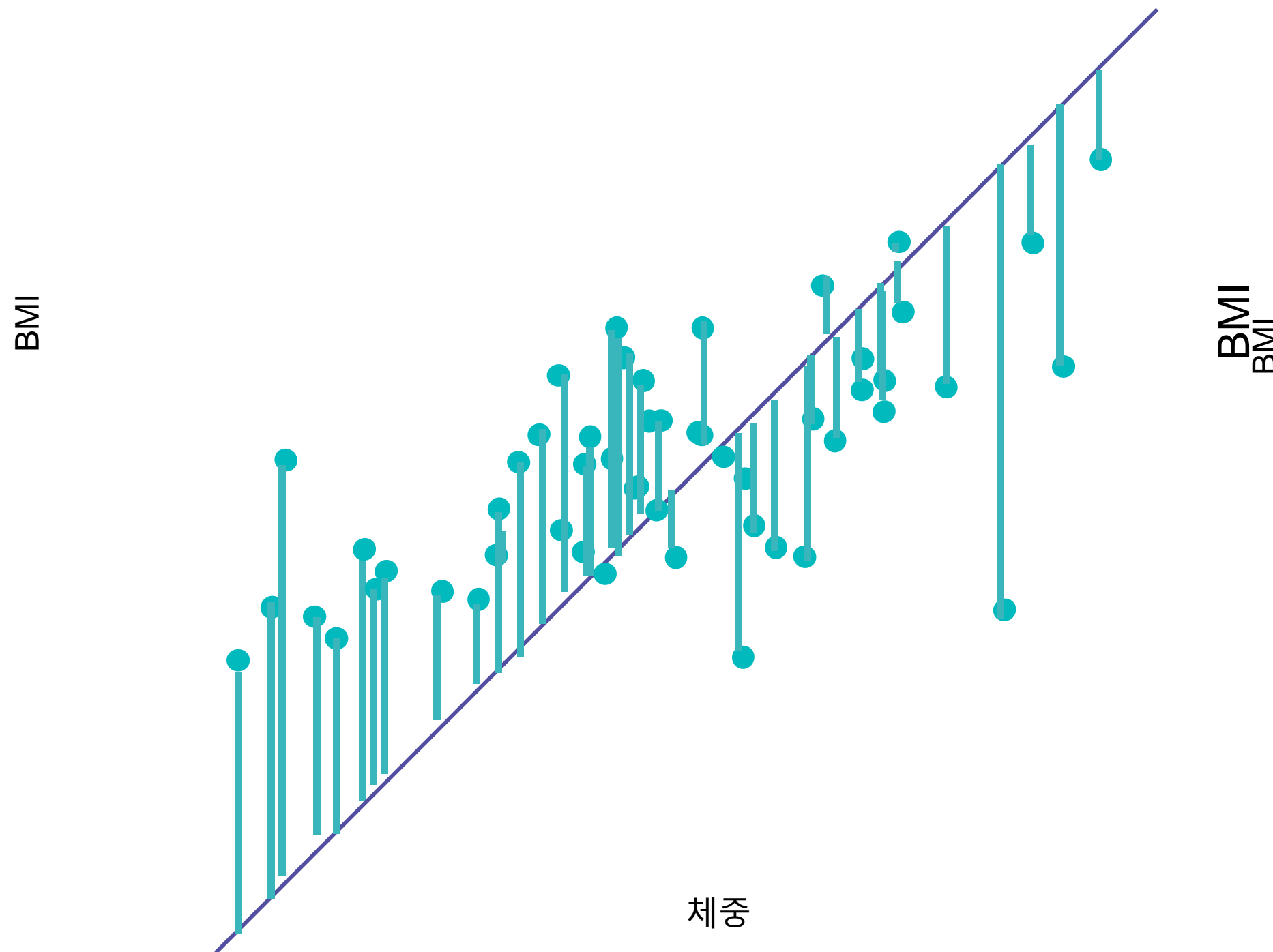


✓ 데이터의 형태

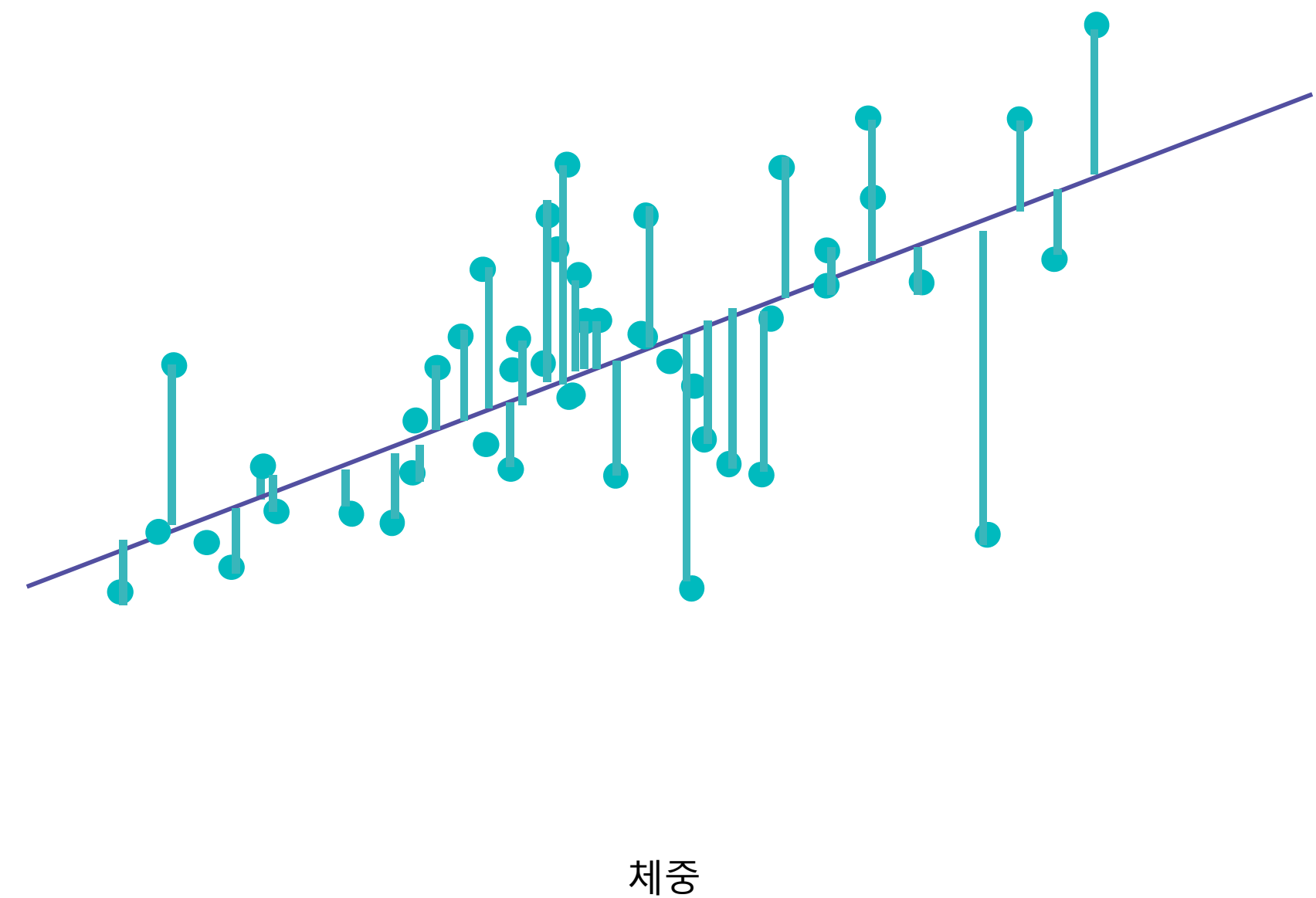


✓ 딥러닝의 Loss 함수(손실함수) 최소화

결과값 (y)와 함수 기울기의 차이가 적을수록 잘 훈련된 모델이다



label과 함수의 예측값의 차이가 평균적으로 큼
High cumulative loss



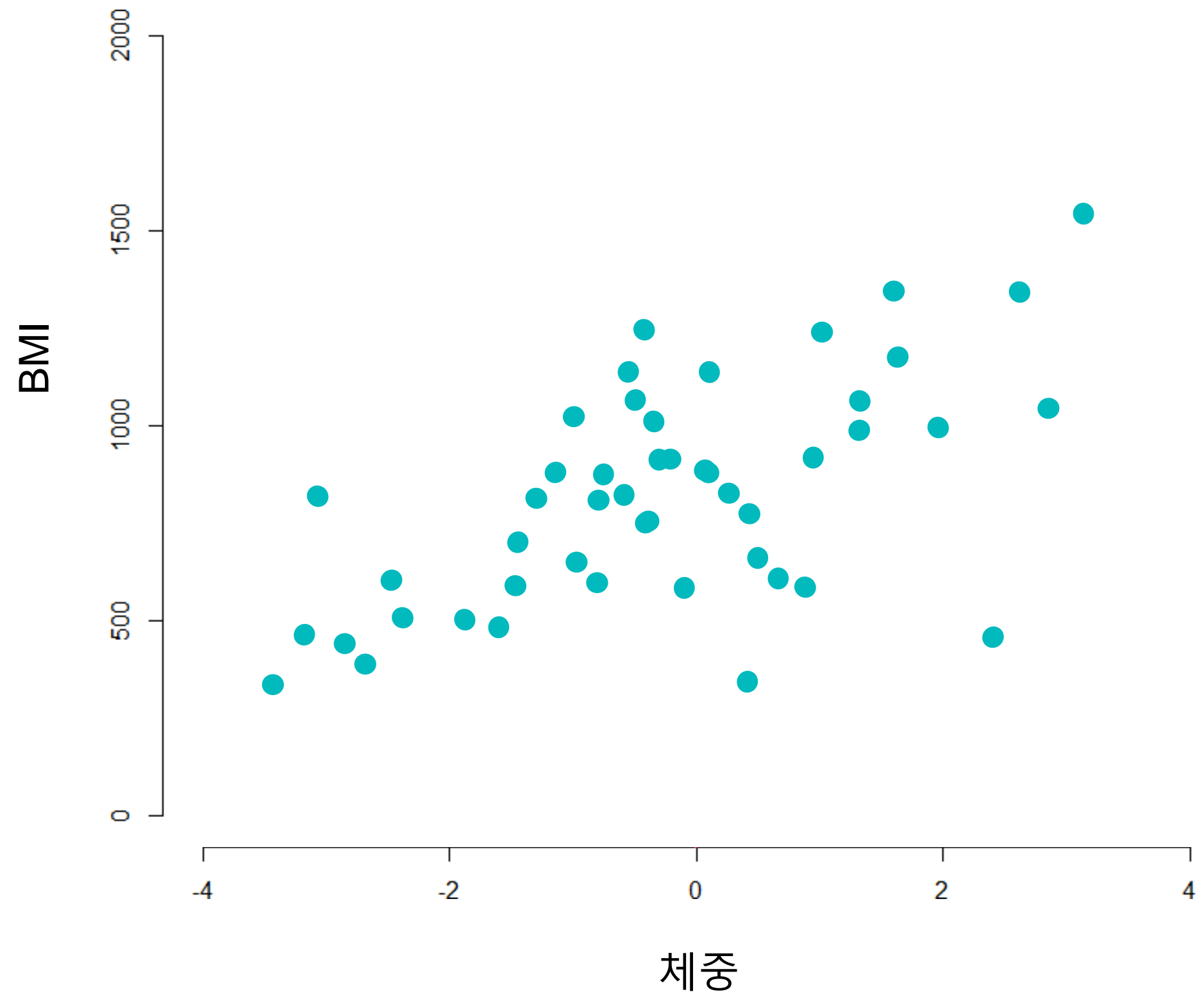
label과 함수의 예측값의 차이가 평균적으로 작음
Minimum cumulative loss

✓ 딥러닝의 Loss 함수로 Loss 최소화

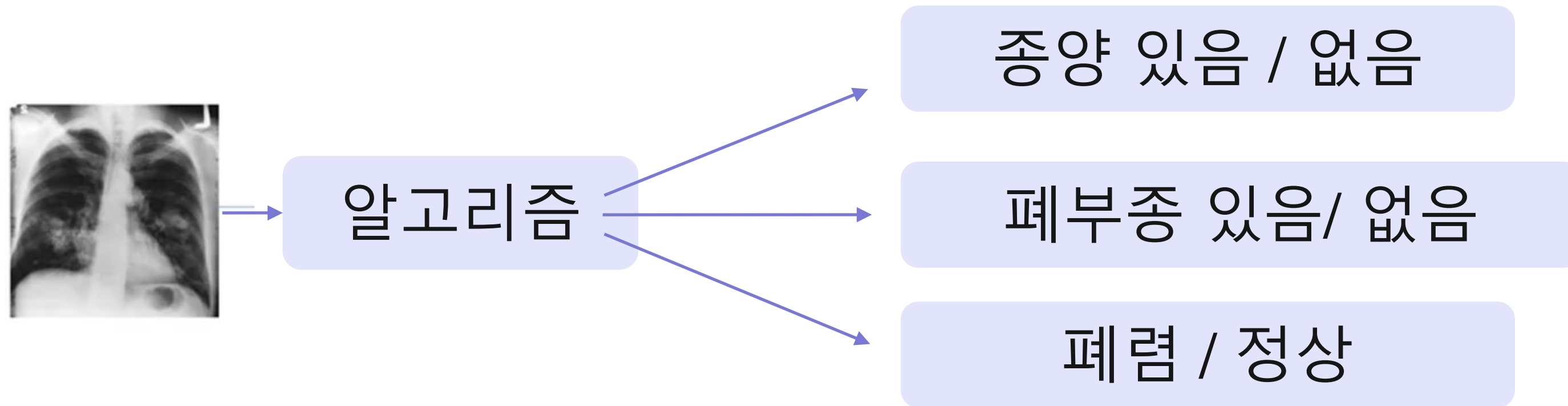
Loss 함수: 실제값과 예측값의 차이

(loss, cost)를 수치화해주는 함수

Loss 함수의 값을 최소화 하는 W , b 를
찾아가는것이 학습 목표이다.

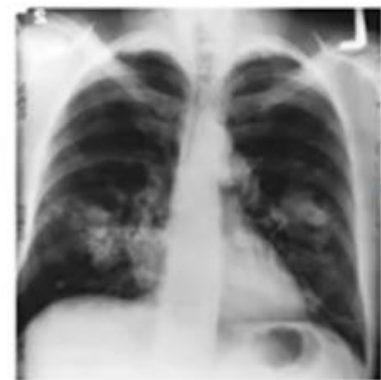


✓ CNN을 이용하여 흉부 X-Ray 이미지로 다양한 질병 분류하기



- 대부분의 이미지 인식 알고리즘은 지도학습
- 흉부 X-레이를 해석 할 수 있는 딥 러닝 모델을 구축하여 다양한 질병 원인을 분류하기 위해 이미지 분할(Image Segmentation) 을 사용하여 위치와 경계를 식별

✓ CNN을 이용하여 흉부 X-Ray 이미지로 질병 분류하기



알고리즘

종양 있음 / 없음

폐부종 있음 / 없음

폐렴 / 정상

1. X-Ray 이미지 전처리
2. Weight balancing 과 Sampling으로 데이터 **클래스 불균형(class imbalance)** 다루기
3. AUC-ROC 커브로 이미지 분류 모델 성능을 평가
4. **Grad-CAM**으로 딥러닝 분석 결과 시각화

✓ 딥러닝 모델의 성능 평가

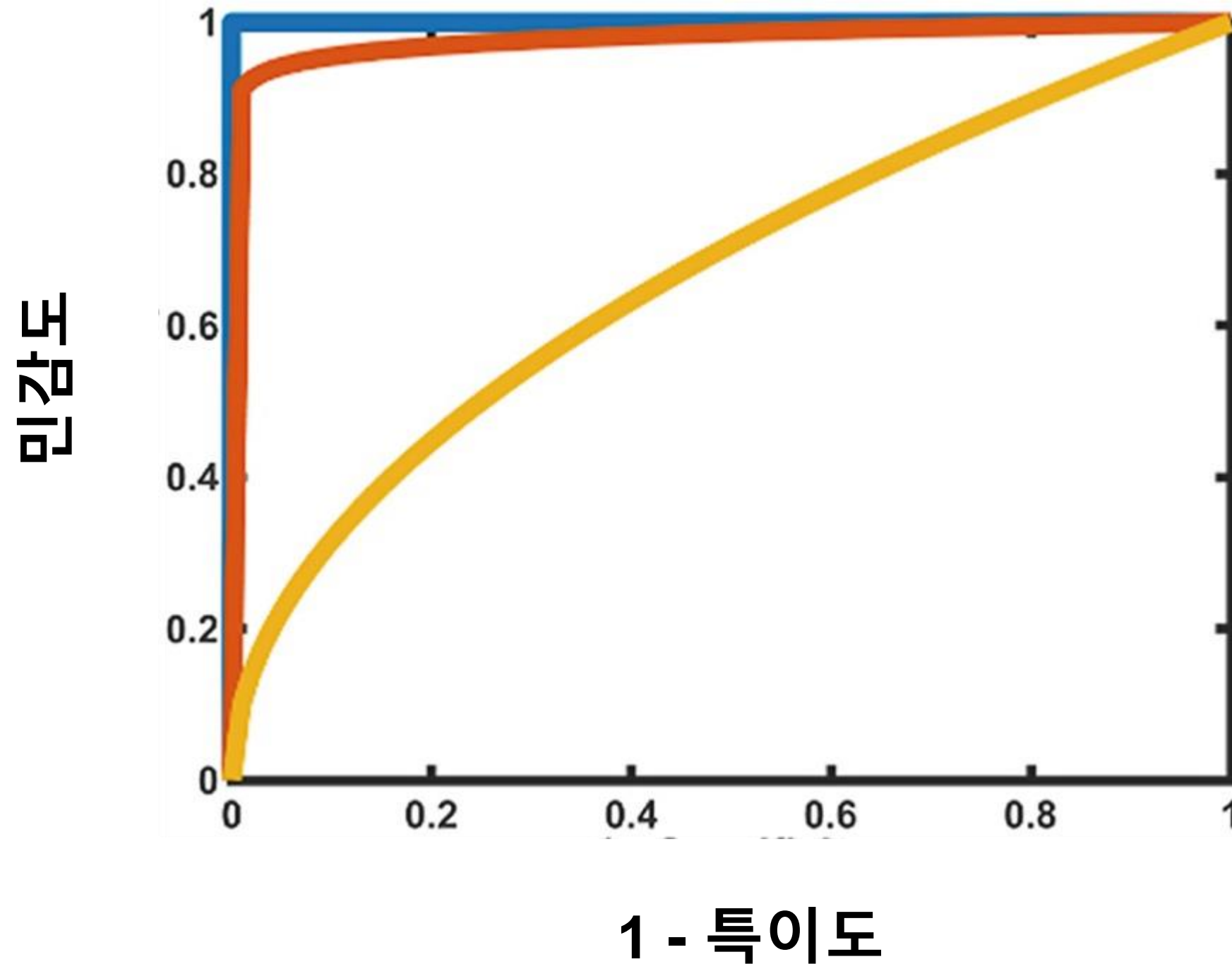
민감도와 특이도

- 민감도(sensitivity): 병이 있는 사람이 검사상 양성인 비율
- 특이도(specificity): 병이 없는 사람이 검사상 음성인 비율

민감도와 특이도가 모두 (실험 대상 모두 동시에) 100% 나올 경우에 가장 이상적

✓ ROC 와 AUC

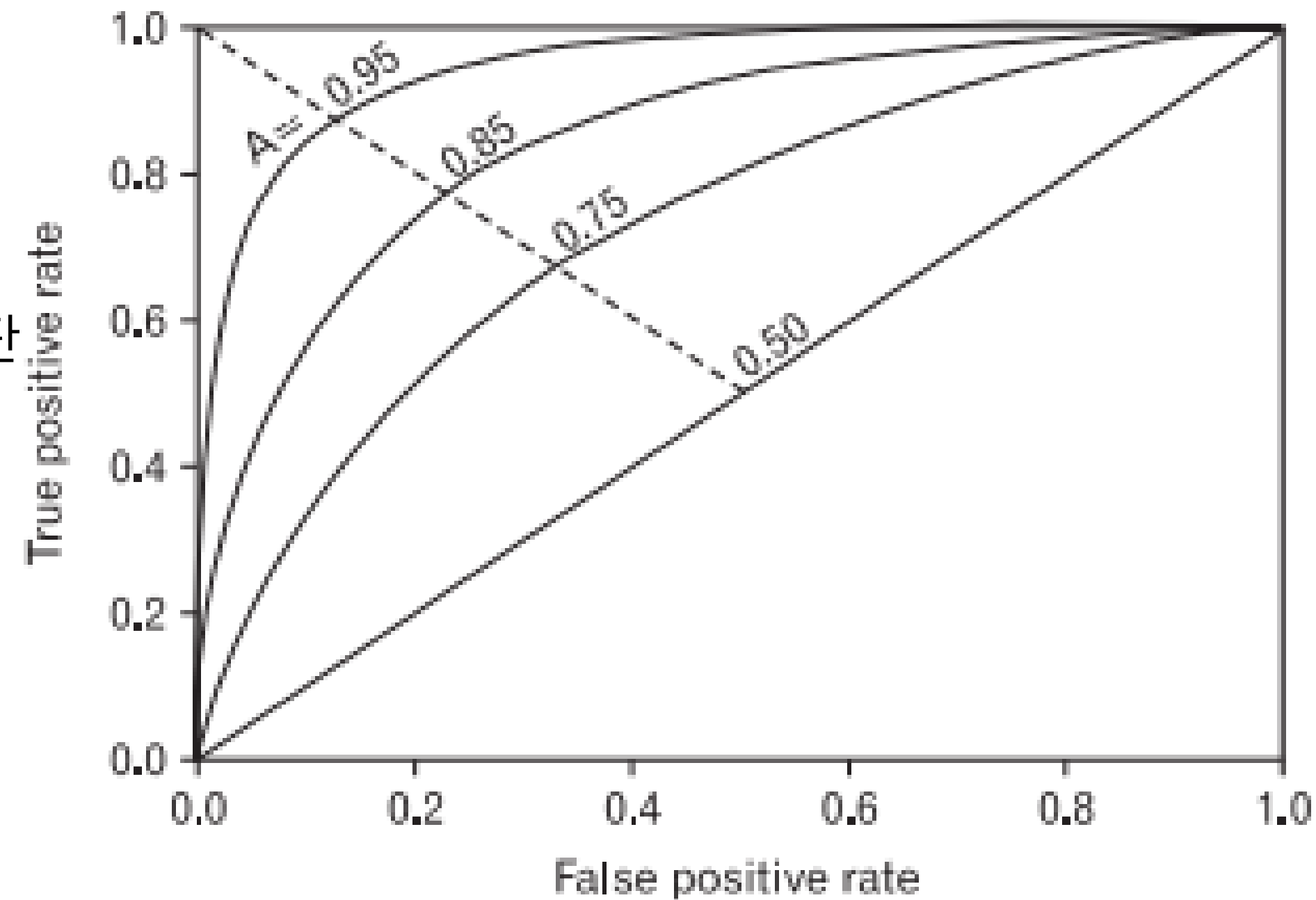
- 민감도가 크고 특이도가 큰 ROC 커브일 수록 , 즉 좋은 검사일 수록 곡선의 아래쪽 면적이 커짐



✓ ROC 와 AUC

- ROC 커브는 그 면적이 1에 가까울수록 (즉 왼쪽위 꼭지점에 다가갈수록) 좋은 성능
- $AUC = \text{ROC 커브 아래의 면적을 구한 값}$
이 면적은 항상 0.5~1의 범위를 갖는다. (0.5이면 랜덤에 가까운 성능, 1이면 최고의 성능)

병에 걸린 사람을 양성으로 판단
한 비율
(판단을 올바르게 한 경우)



병에 걸리지 않은 사람을 양성으로 판단한 비율
(판단을 틀리게 한 경우)

05

의료 인공지능 활용 사례



✓ 이미지 데이터 분석으로 얻는 의학적인 통찰

CNN을 사용하여 안저사진을 분석하여 당뇨병 망막병증을 진단하는 구글

- > 평균적인 안과 전문의보다 안저 판독에 대해서는 딥러닝이 더 정확함
- > 인간 의사와 함께 판독에 참여하여 의사가 놓친 당뇨병 망막병증이 없는지 파악하는데 민감도를 극대화하여 사용 가능

안과 전문의 F-스코어 = 0.91

구글 인공지능의 F-스코어 = 0.95

- F-스코어 : 모델의 종합적인 정확도 수치,
precision(정밀도)와 재현율(recall)
= 실제 질병이 있는 환자/ 질병이 있다고 모
델이 예측한 환자 총 수

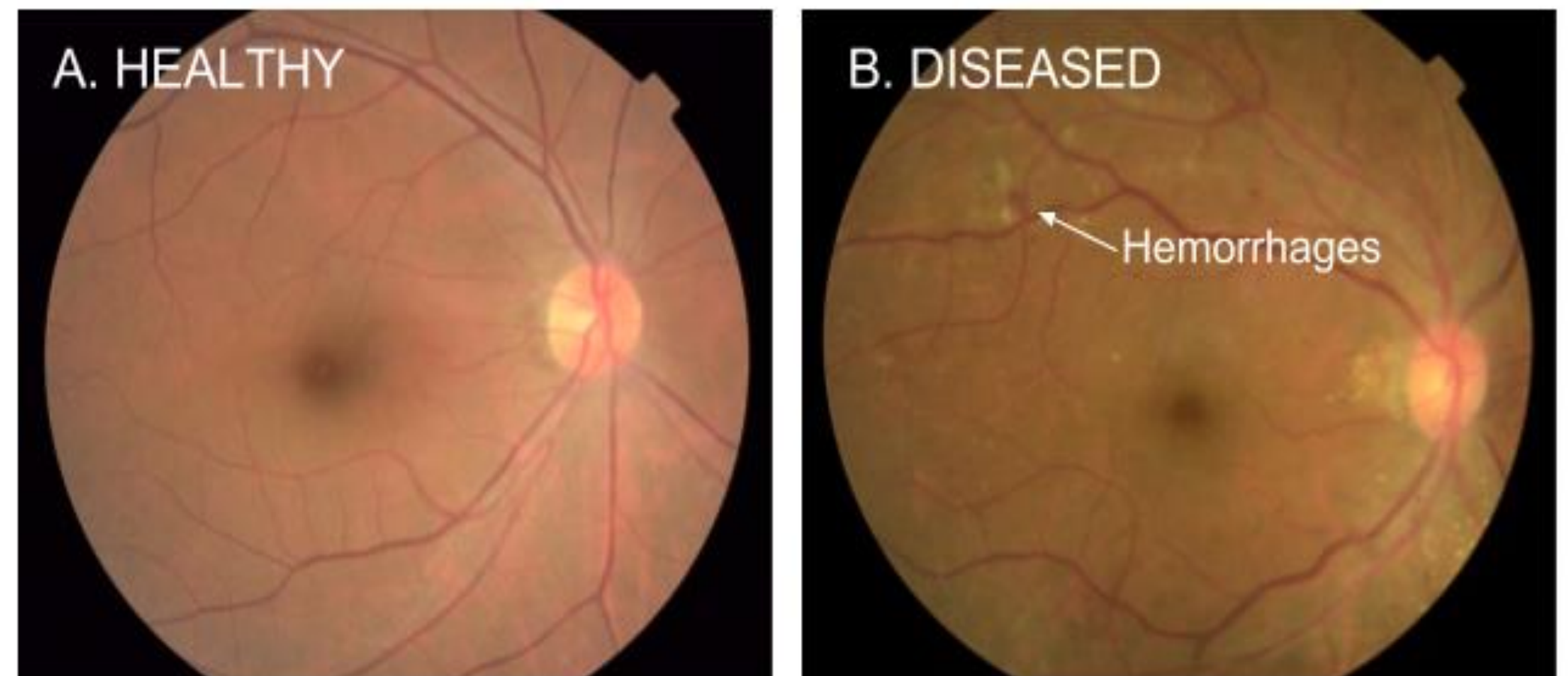
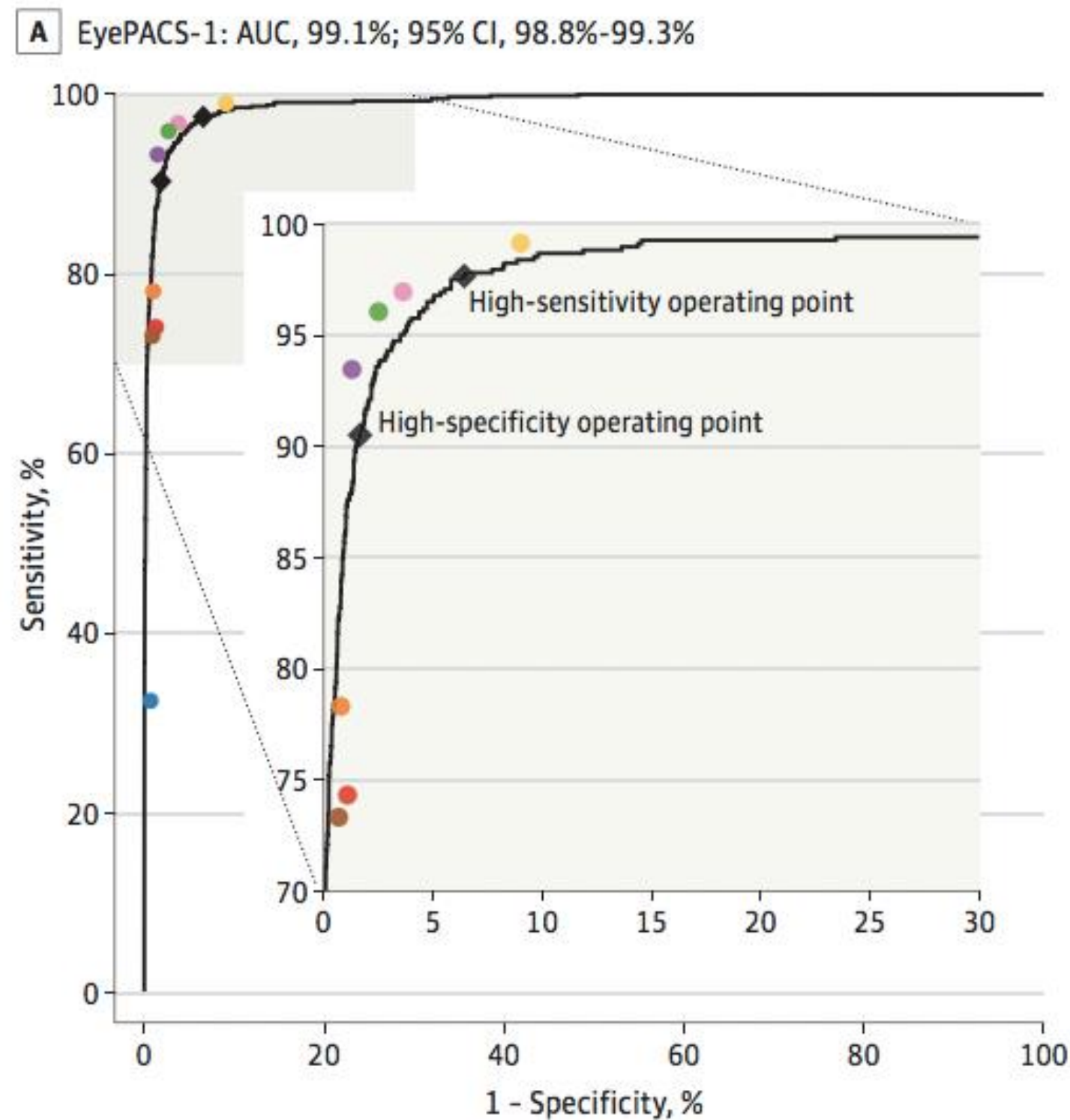


Figure 1. Examples of retinal fundus photographs that are taken to screen for DR. The image on the left is of a healthy retina (A), whereas the image on the right is a retina with referable diabetic retinopathy (B) due a number of hemorrhages (red spots) present.

✓ 이미지 데이터 분석으로 얻는 의학적인 통찰

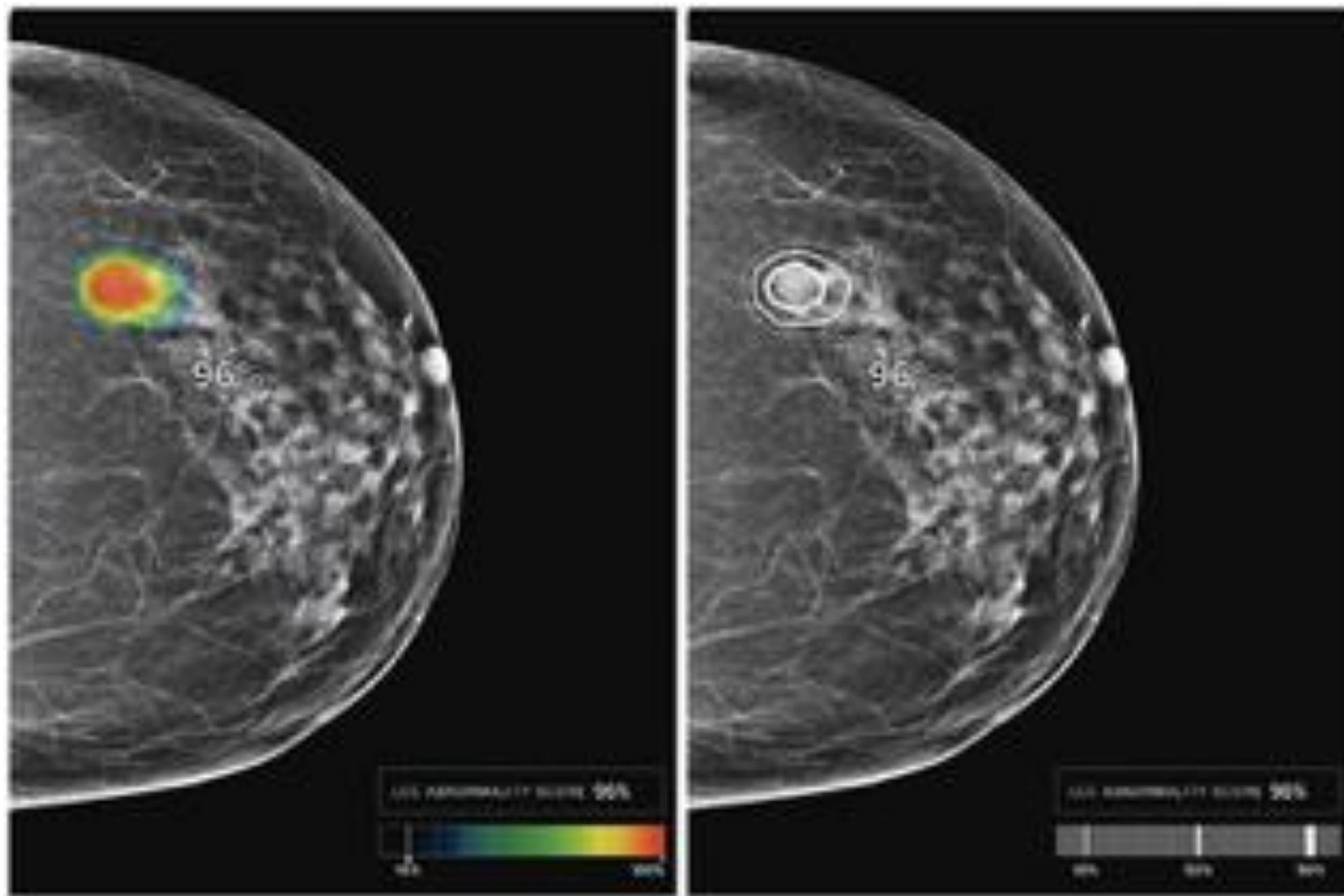
CNN을 사용하여 안저사진을 분석하여 당뇨병 망막병증을 진단하는 구글



- 각 점들: 구글의 인공지능 학습에 사용된 안과 전문의의 성적
- 안과 전문의 (7-8명)의 판독 성적이 ROC곡선 보다 위쪽에 위치하면 인공지능보다 나은 성적, 곡선보다 낮은 곳에 있으면 인공지능보다 판독이 성적이 좋지 않은 성적

✓ 이미지 데이터 분석으로 유방암 판독

유방촬영술에 CNN을 적용한 루닛

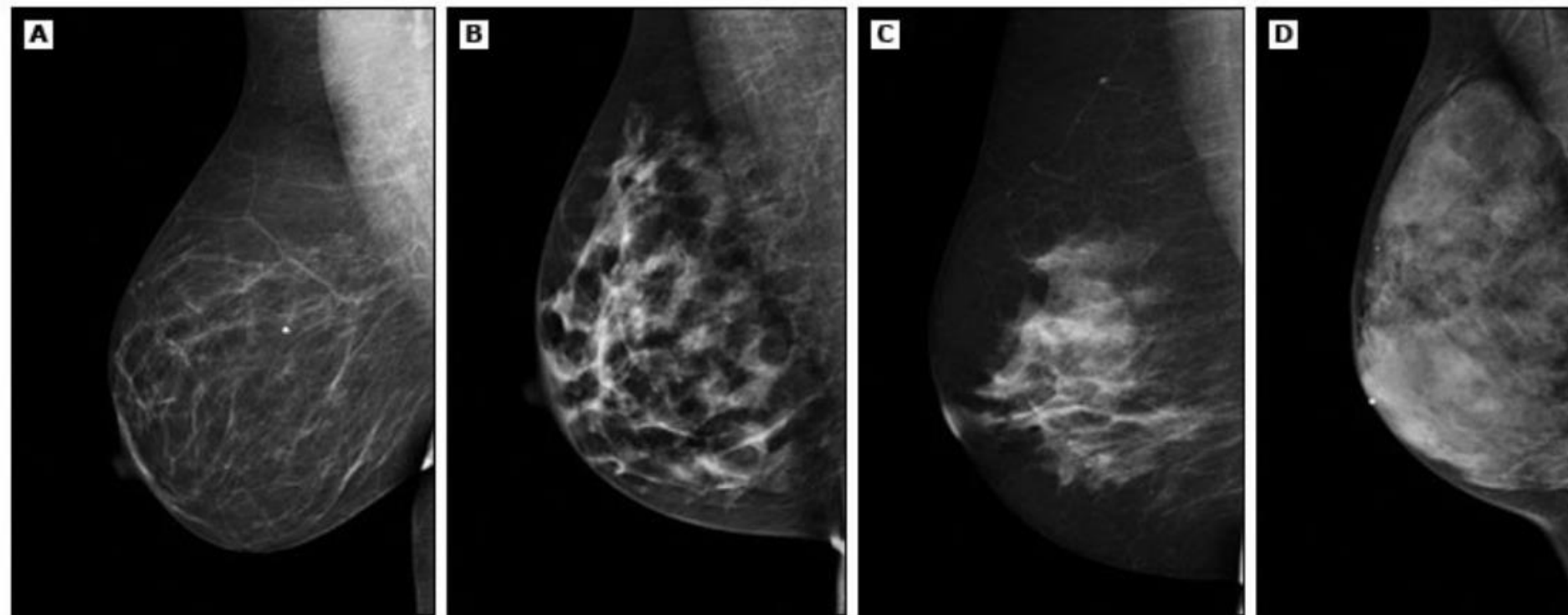


병변의 위치 정보에 대한 label(정답지)화

:

- a. 픽셀 값, 주변 픽셀과 해당 픽셀의 값 차이, 텍스처, 그리고 모양 등의 특징들을 분석하기
- b. 영상에서 정상조직이 아닌 것으로 추정되는 영역 찾기
- c. line(선)을 이용하여 유방암의 병변의 주변부를 그려주거나 병변 부위에 다양한 색으로 표기

✓ CNN 적용시 데이터 활용 효율을 높이기 위한 방법



(A) Almost entirely fatty. (B) Scattered areas of fibroglandular tissue. (C) Heterogeneously dense. (D) Extremely dense. - BIRAD 5th Classification (image from Courtesy of Phoebe Freer, MD and Priscilla Slanetz, MD, MPH, FACR.)

다른 영상 검사 데이터와 함께 활용

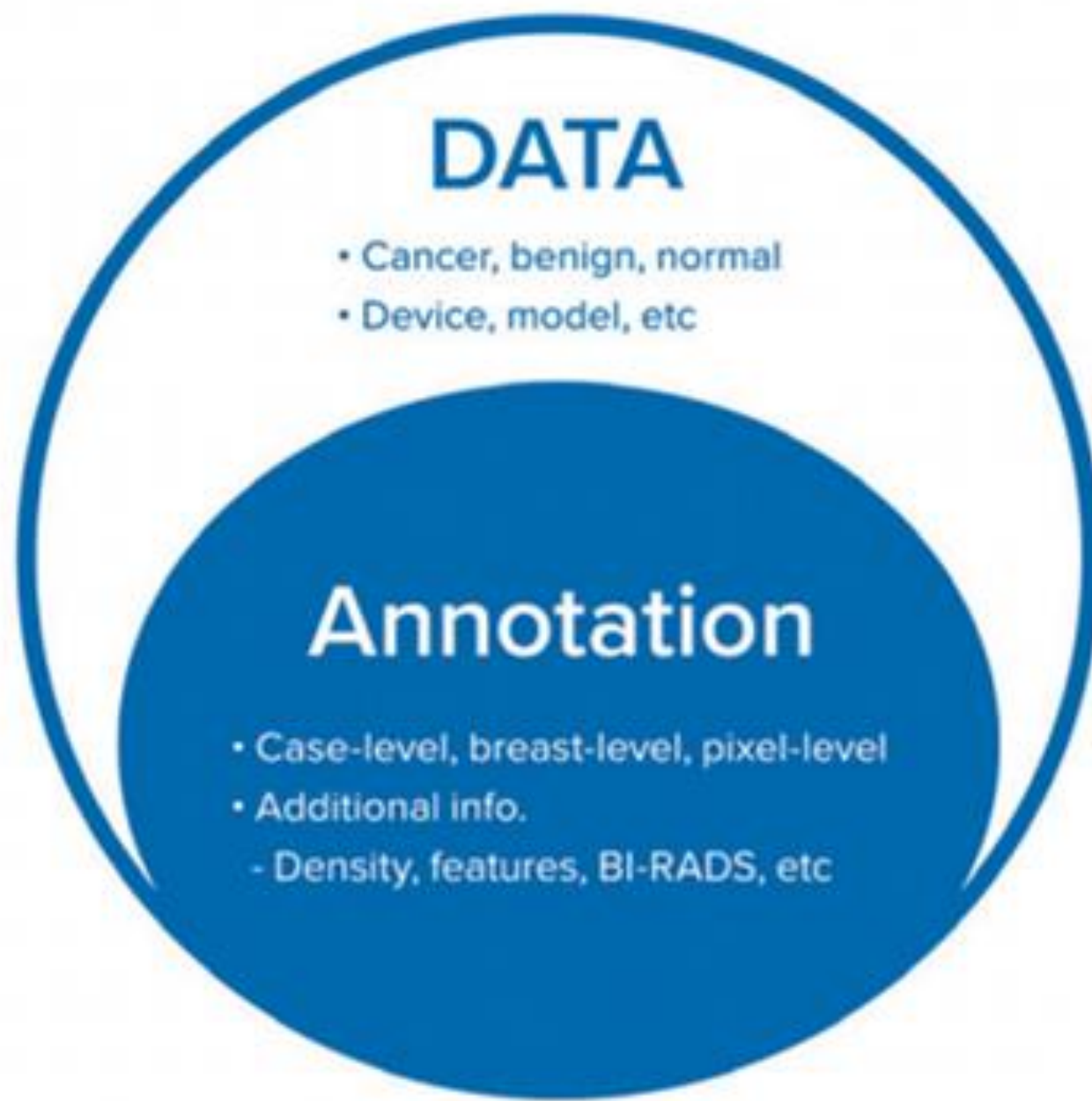
- 다른 영상 검사 결과들 또는 조직검사 결과들을 참고하여 유방암의 실제 위치 정보를 학습에 이용
- 유방 치밀도 또는 의사의 BI-RADS 판정 결과 등의 정보들을 함께 활용

✓ 이미지 의료데이터로 유방암 예측

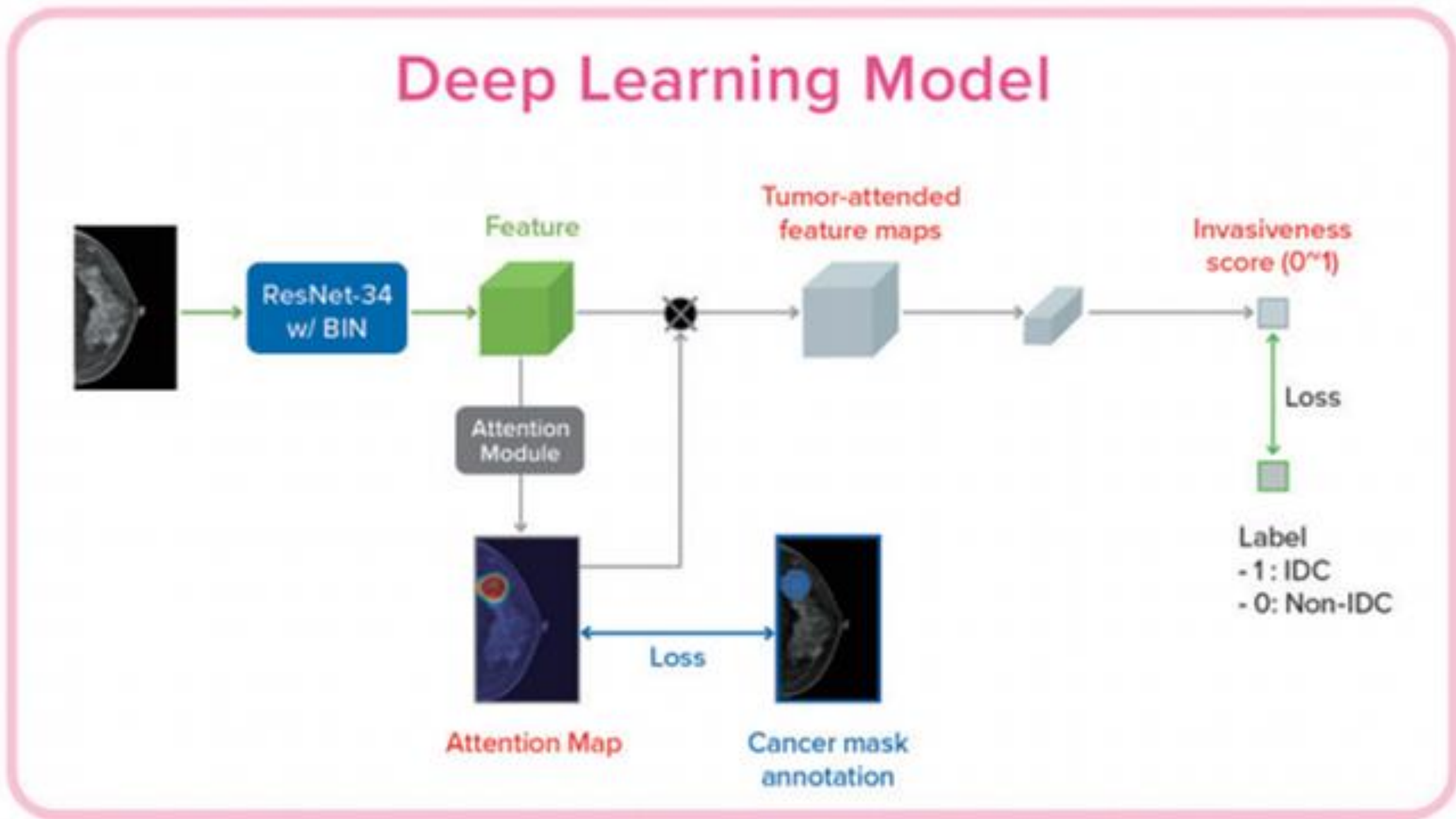
유방촬영술에 CNN 적용을 적용한 루닛

유방촬영술을 위한 최적의 딥러닝 알고리즘의 필수 요소

- 구체적인 label(정답지)과 추가 비용(cost)을 적절하게 타협
- 유방 치밀도의 평가에서 정확성이 높은 알고리즘



+

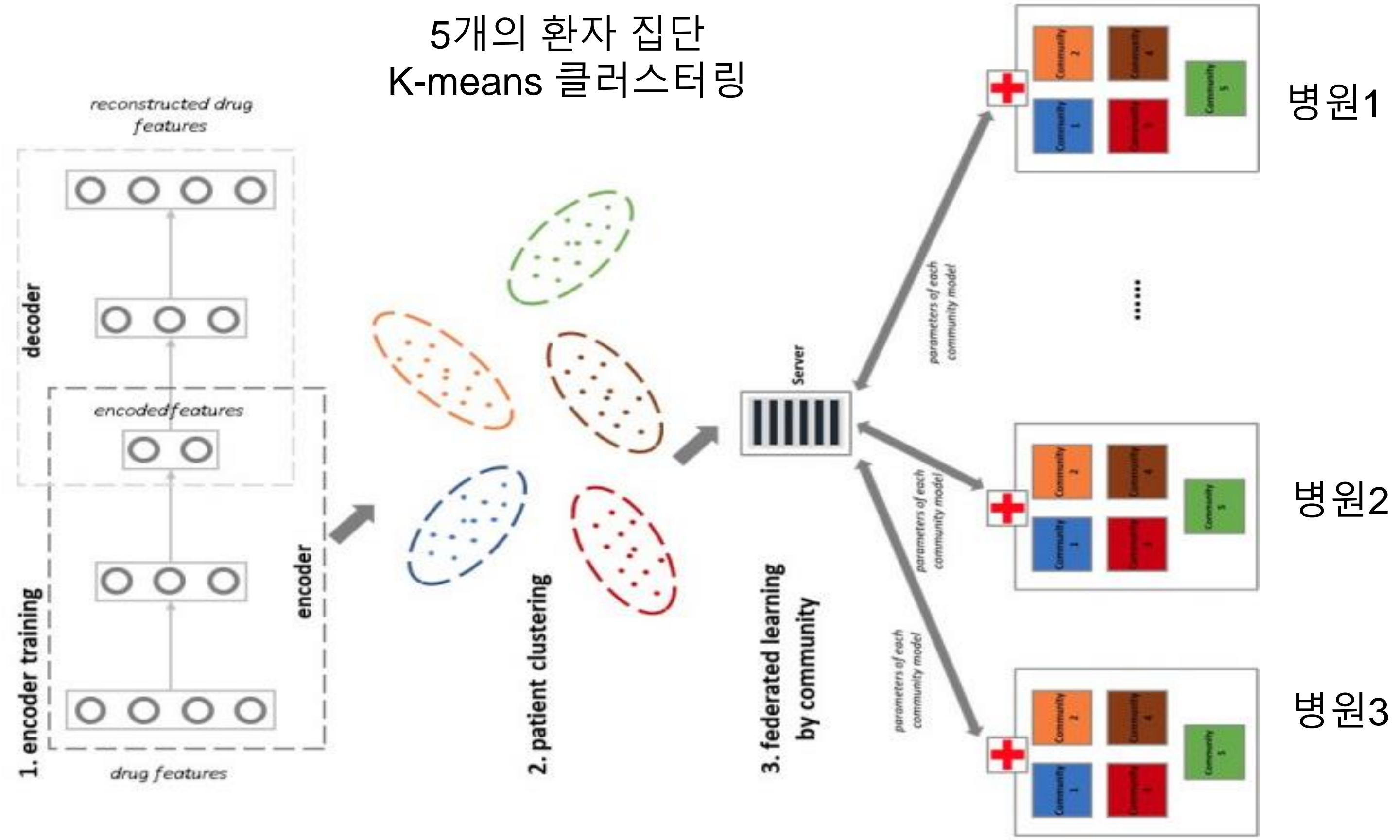


✓ 복잡한 의료 데이터 분석으로 의학적인 통찰을 얻는 연구

- 진자의무기록(EHR)의 분석 결과를 바탕으로 심혈관 질환의 **발병 원인을 정확하게 예측**하여 환자의 질병을 진료 또는 **사망률이나 재입원률을 낮추고 의료비를 절감**하는데 전략을 세우는데 보조
- 유전체 데이터를 바탕으로 개별 환자에게 **맞춤 처방** 권고
- 대규모 환자군의 과거 진료기록을 바탕으로 질병의 발병을 예측하고 입원중 **사망 가능성 예측**

환자 데이터(EMR) 클러스터링을 이용한 사망률 예측 사례

5개의 환자 집단
K-means 클러스터링



논문 출처: Li Huang, Andrew L. Shea, Huining Qian, Aditya Masurkar, Hao Deng, Dianbo Liu, Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records, Journal of Biomedical Informatics,

✔ 환자 데이터(EMR) 클러스터링을 이용한 사망률 예측 연구

접근방식: FL (Federated Learning)

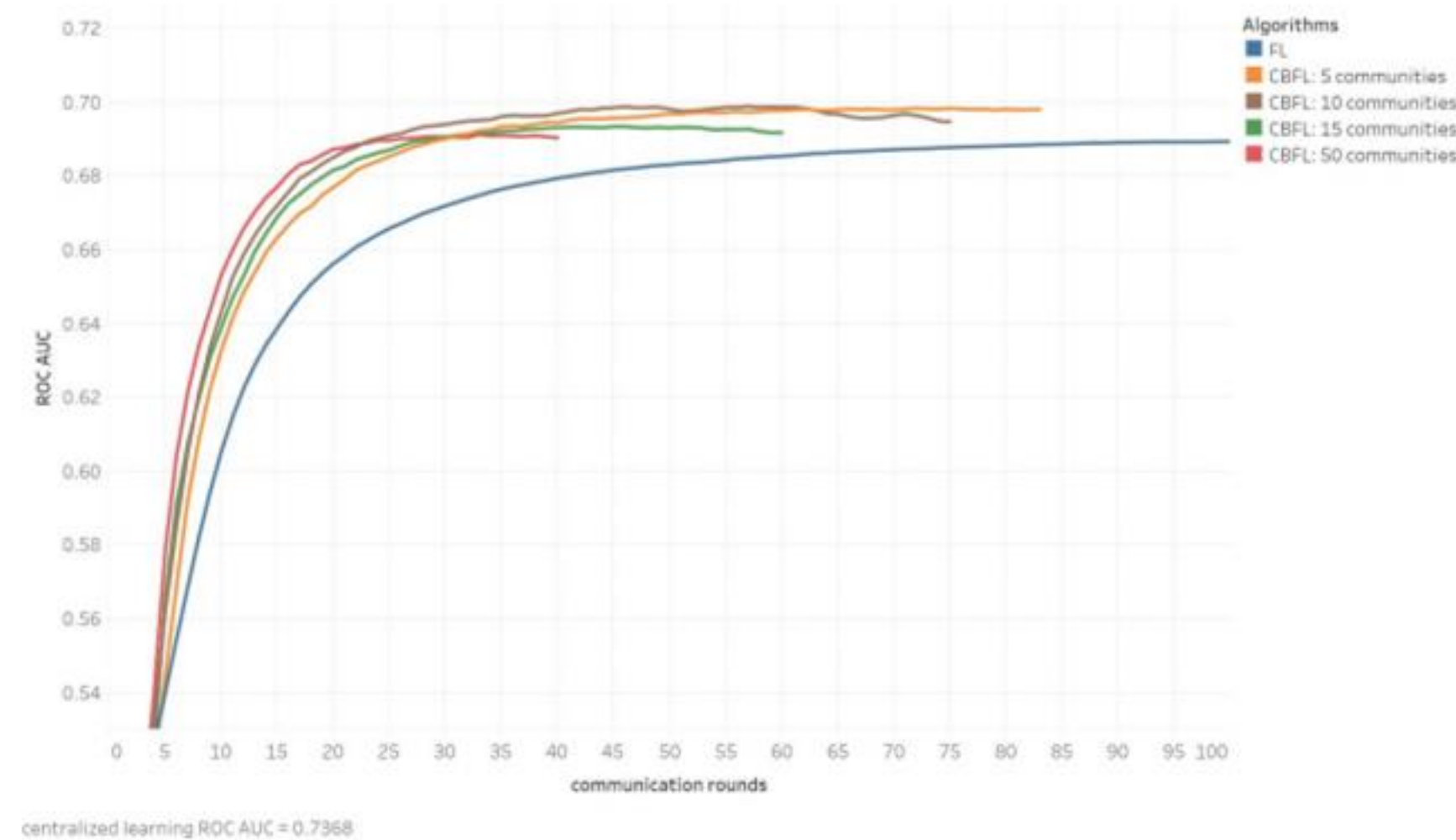
연합학습(FL: Federated Learning)이란? : 다수의 사물인터넷(웨어러블) 또는 스마트폰 기기와 하나의 중앙서버가 협력하여 데이터가 탈중앙화된 상황에서 데이터 모델을 학습 (데이터 프라이버시 향상과 커뮤니케이션 효율성 증가)

✔ 환자 데이터(EMR) 클러스터링을 이용한 사망률 예측 연구

연구 방법

1. 미국 전역의 208개의 병원에서 약 20만명의 환자들을 대상으로 ICU 데이터베이스를 추출
2. 28000명의 환자 중 비슷한 특징이 있는 환자들끼리 5 그룹으로 나누어 진단
3. 병원 레벨에서도 환자를 Clustering 해서 분류, 환자 집단들의 특성을 드러냄
4. FL(연합 학습)과 CBFL(집단 기반 연합학습)모델로 예측 정확도(ROC AUC / PR AUC 기반)를 분석

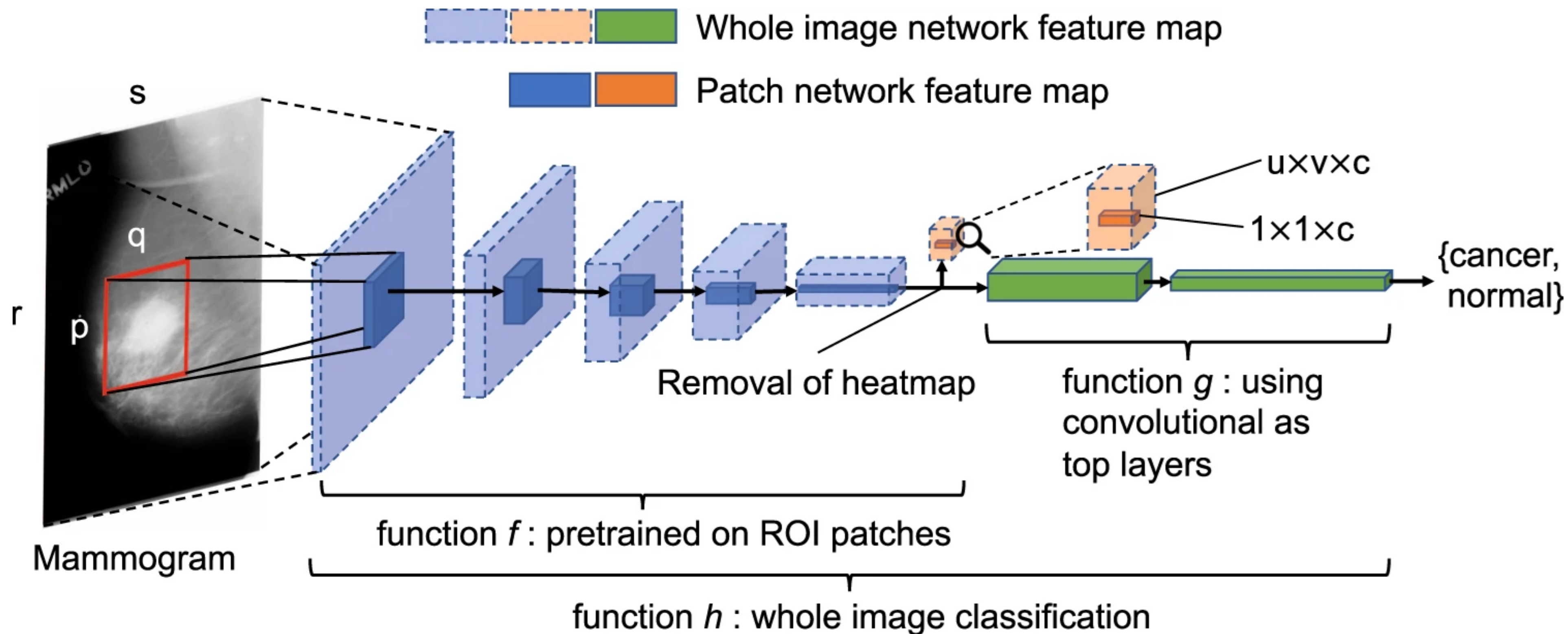
✓ 환자 데이터(EMR) 클러스터링을 이용한 사망률 예측 사례



연구 결과:

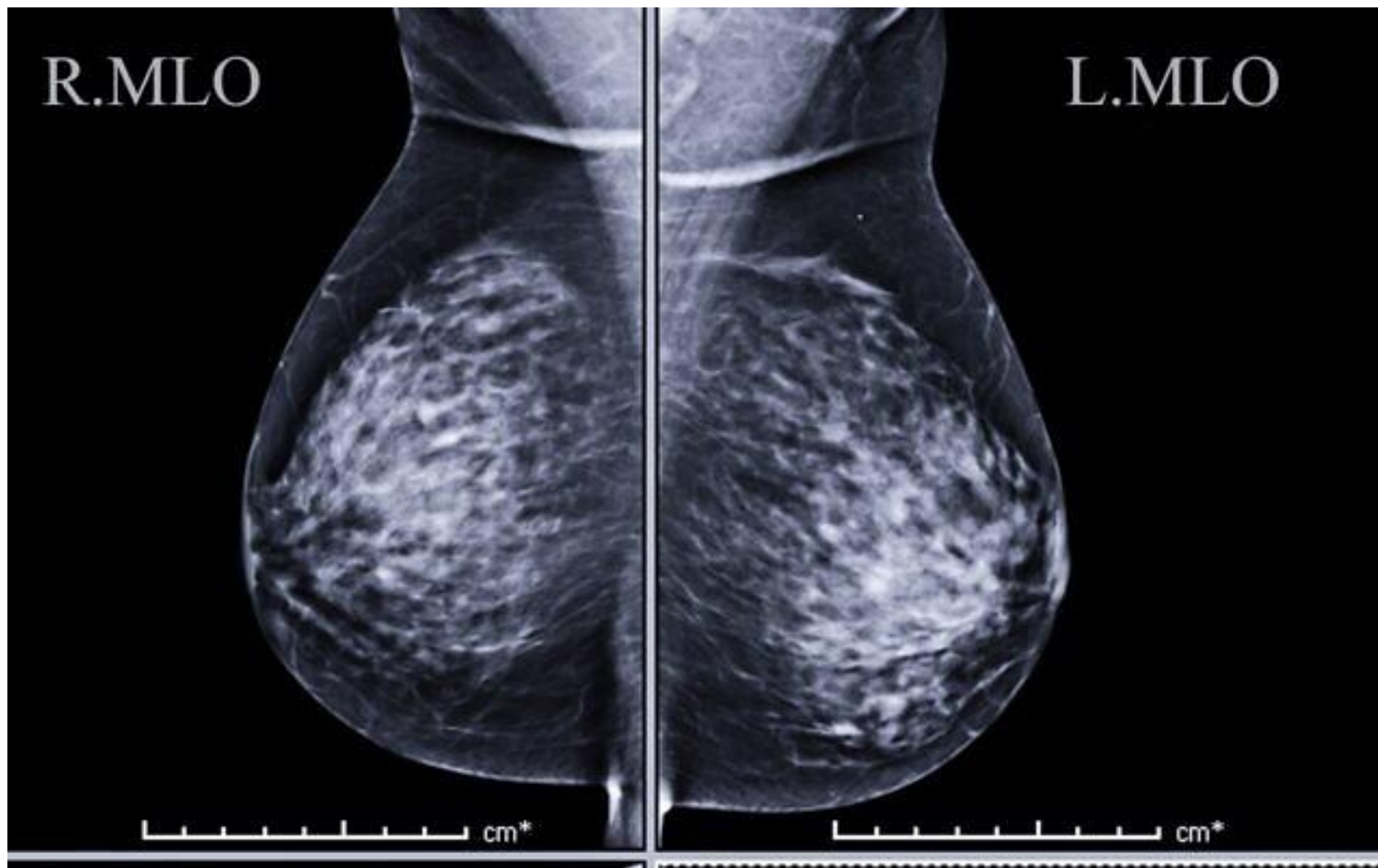
- 만성질환을 지닌 환자 집단을 대상으로 진단하여 사망률 및 입원시간 예측
- 불필요한 의료 진단 횟수 및 수술 비용 감소
- 집단 중심의 딥러닝 연합학습의 성능 및 평가 결과를 분석하였는데 AUROC 정확성과 PR AUC의 예측 결과가 기존 연합학습(FL)을 현저하게 넘어섰다.

✓ Lunit의 CNN을 적용한 유방촬영술



✓ Zebra Medical Vision 유방암 엑스레이 판독

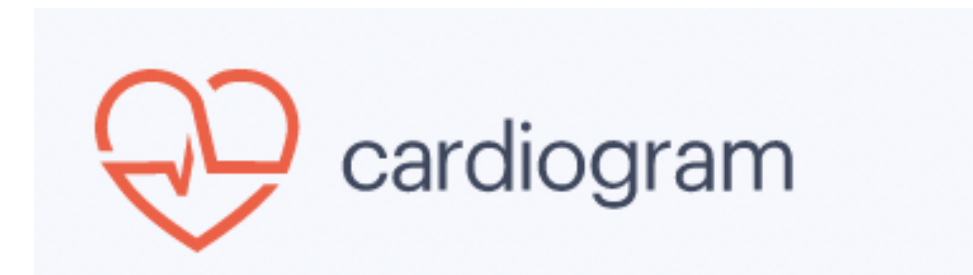
유방 촬영술을 통해서 유방암 의심 병변을 발견



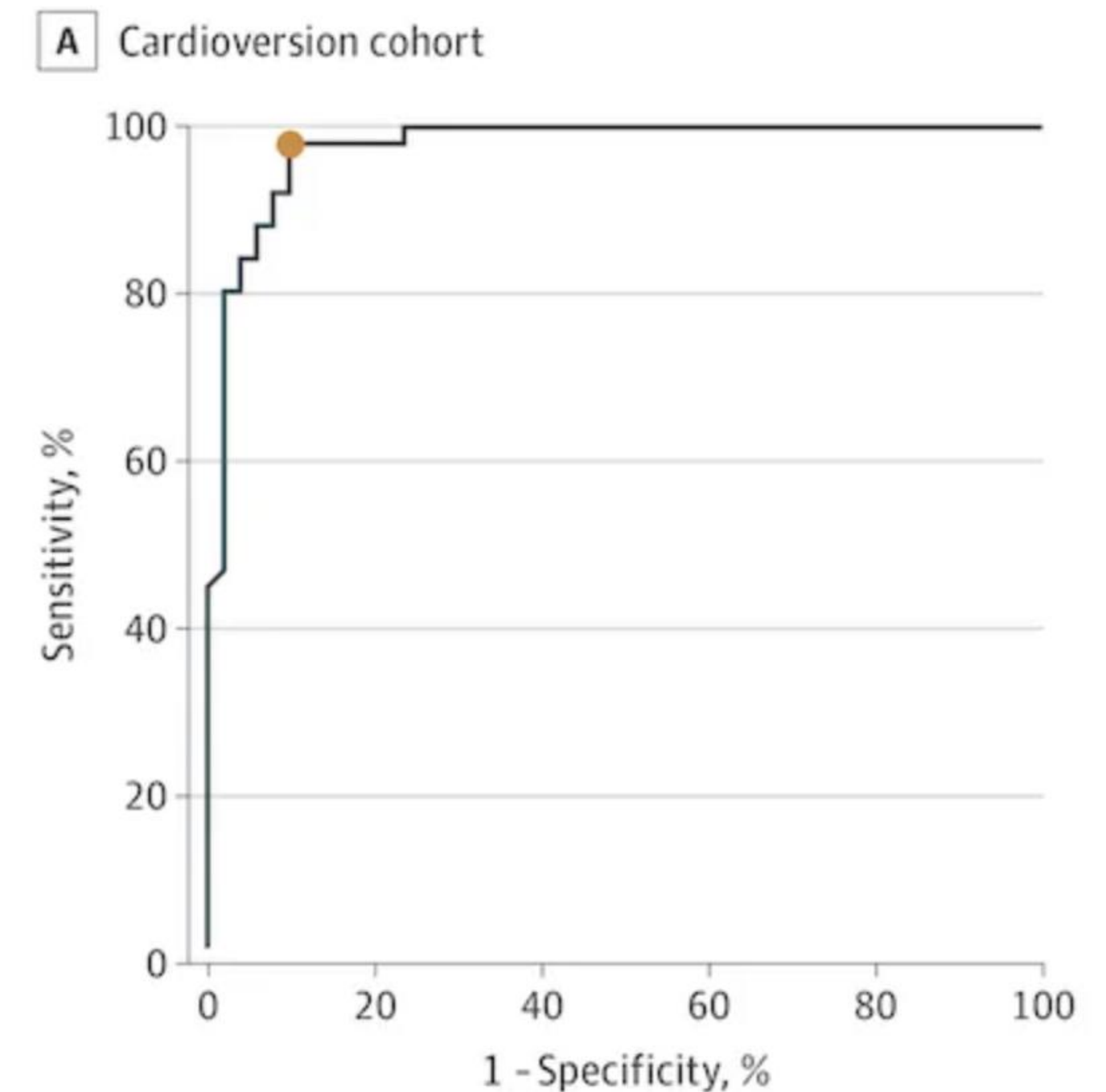
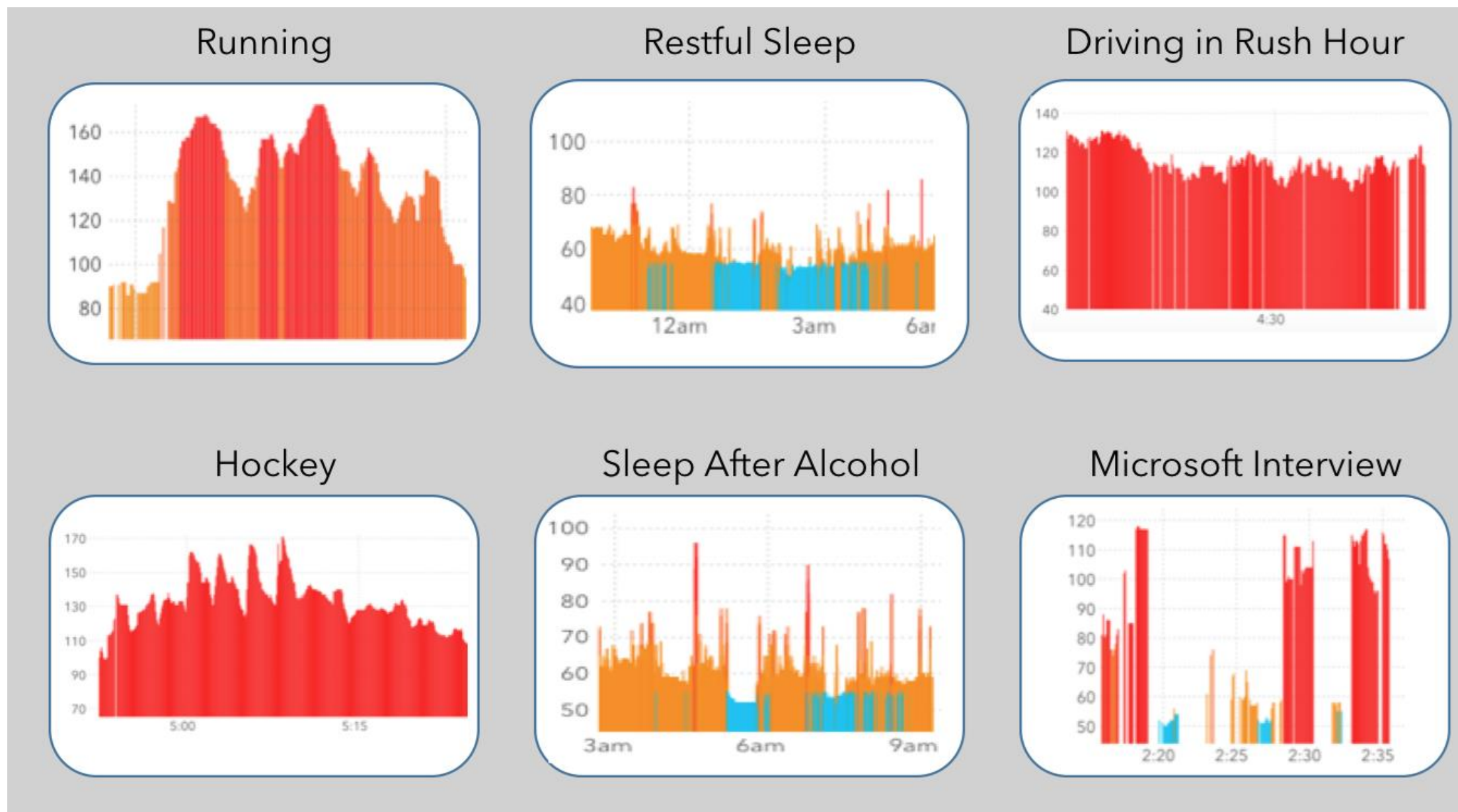
정기적인 유방 촬영술을 통한 검진은 지난 30년간 유방암 환자의 사망률을 30% 정도 감소시킴

유방촬영 영상을 판독하기위한 알고리즘 사용에 대해 미국식품의약국(FDA) 허가 받음

✓ 연속적 의료데이터로 질병 예측 사례

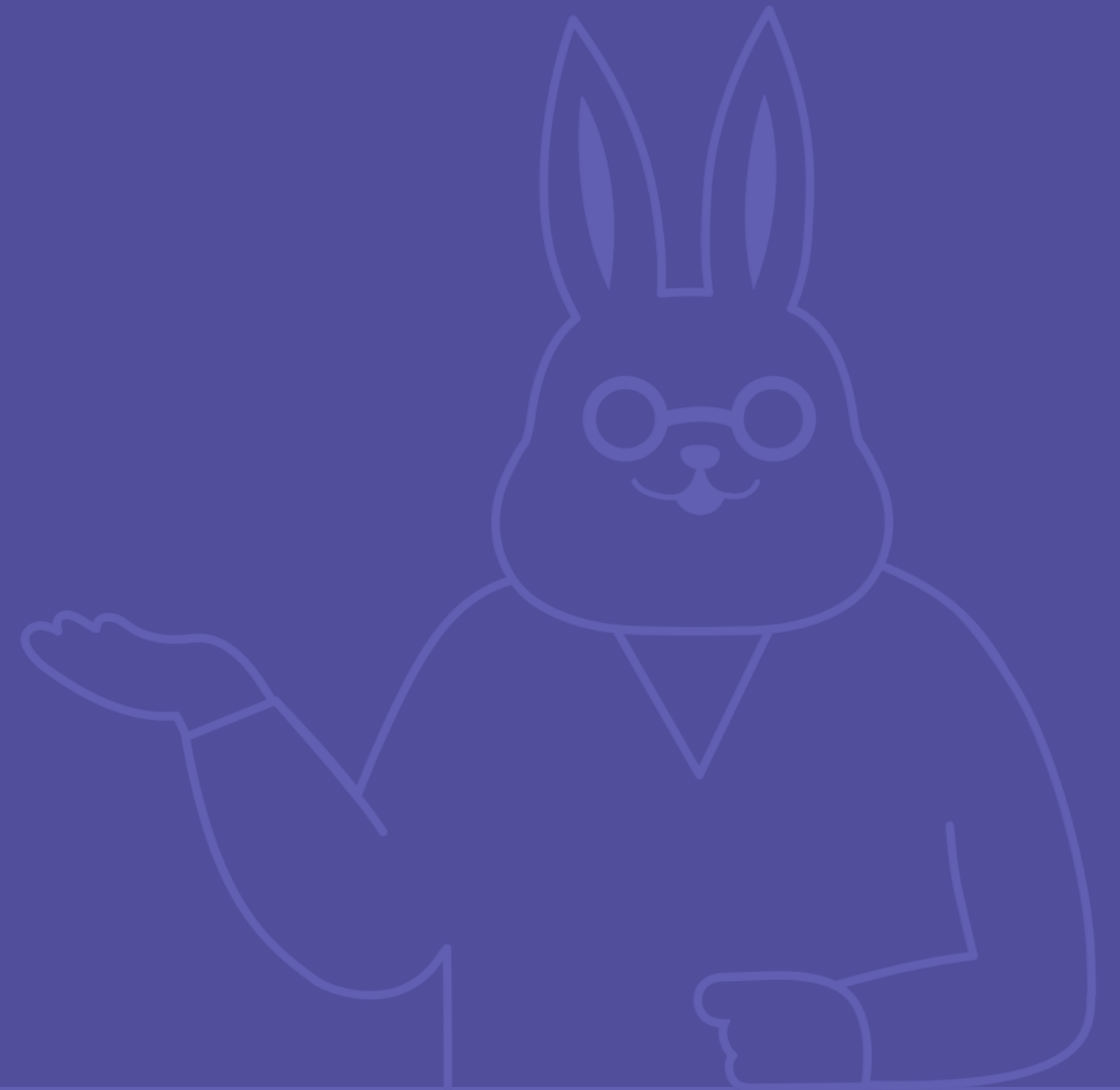


카디오그램은 심방세동을 AUC=0.97의 높은 정확도로 측정-> 심방세동 환자와 정상인을 정확하게 구분



06

의료 인공지능의 장점과 한계



✓ 딥러닝을 의료에 활용할 때의 장점

1. 매일 쏟아져 나오는 엄청난 분량의 의료연구논문들과 임상시험 결과들을 환자의 치료에 빠르게 반영 가능
2. 인공지능 기반으로 예측한 권고안을 제공하여 의사의 치료법 결정에 도움을 줌
3. 전자 의무기록을 분석 모델의 확장성
(데이터의 특징을 스스로 파악해서 데이터 전처리과정이 약 80% 감소함으로 새로운 종류의 전자 의무기록 데이터를 추가하거나 다른 병원에서 적용하기 쉬워짐)
4. EHR 데이터를 효과적으로 분석해서 환자의 치료와 병원의 운영에 유용한 인사이트를 얻을 수 있음

✔ 딥러닝의 단점

1. 데이터의 특징을 스스로 파악해서 학습하기 때문에 정확성과 임상적 효용성 등을 증명하기 어려움
2. 수많은 시간을 사용을 어떤 것을 학습하고 않을지의 결정하는데 의료전문가들이 관여해야 함
3. 특정 환자에 대한 판단을 받아보기 위해 환자의 동의를 받아 그 데이터를 전송하는 과정 및 절차가 까다로움

✔ 디지털 헬스의 제도적 불확실성

- 의료기기의 분류 기준이 명확하게 없음
- 개발의 기존 규제가 힘들거나 회색지대에 있는 경우가 많음

해결 방법

- 의료용-건강관리용 구분과 규제가 더 명확해져야 하며 일관성있게 적용되어야 함
- 헬스케어 애플리케이션과 의료기기 기획 초기 부터 식약처(FDA)와 미리 상의하면서 연구 및 개발을 진행해야 함

✓ 의료 인공지능의 활용에 따른 윤리적인 문제

- 의료 인공지능이 활용된 경우 의료사고가 발생할 경우: 환자의 존엄성 손상
- 의료의 기계화 및 객체화로 나갈수 있는 우려
- 데이터 유출 – 환자 프라이버시 침해 우려
- 인공지능이 도입된 의료서비스가 모든 사회 구성원에게 제공되지 않을 시 의료 격차의 심화 가능성

✔ 의료인공지능이 활용되면 의사와 환자와의 관계는 어떻게 변모할까?

- 의료 전문지식의 정보화가 이루어지면서 양자간 정보의 비대칭 성이 많이 완화
- 의료 영역의 지능정보화로 전문가와 비전문가의 간극이 좁아짐
- 환자의 의료주권이 강화
- 의사의 단 반복적인 업무 부담 경감으로 환자와의 직접적인 대면시간 및 진료/ 치료시간 확보

✓ 생각해 볼 것들

- 의료 데이터 이용 활성화와 개인정보 보호 필요성이 충돌할 경우 이를 어떻게 조화시킬 것인가?
- 의료 인공지능의 자율적 의사결정 수준이 발전할 경우, 의사의 의료과실 유무 및 주의 의무 위반 여부를 판단하는 기준이 무엇일까? (의료 인공지능의 역할을 어디까지 허용할 것인가)
- 의료 인공지능의 활용이 의사와 환자간의 정보 비대칭을 경감시킬 것인가 강화시킬 것인가?
- 의료 인공지능이 활용하면서 어떻게 치료적 대화와 상호작용을 늘릴 것인가?

연락처

TEL

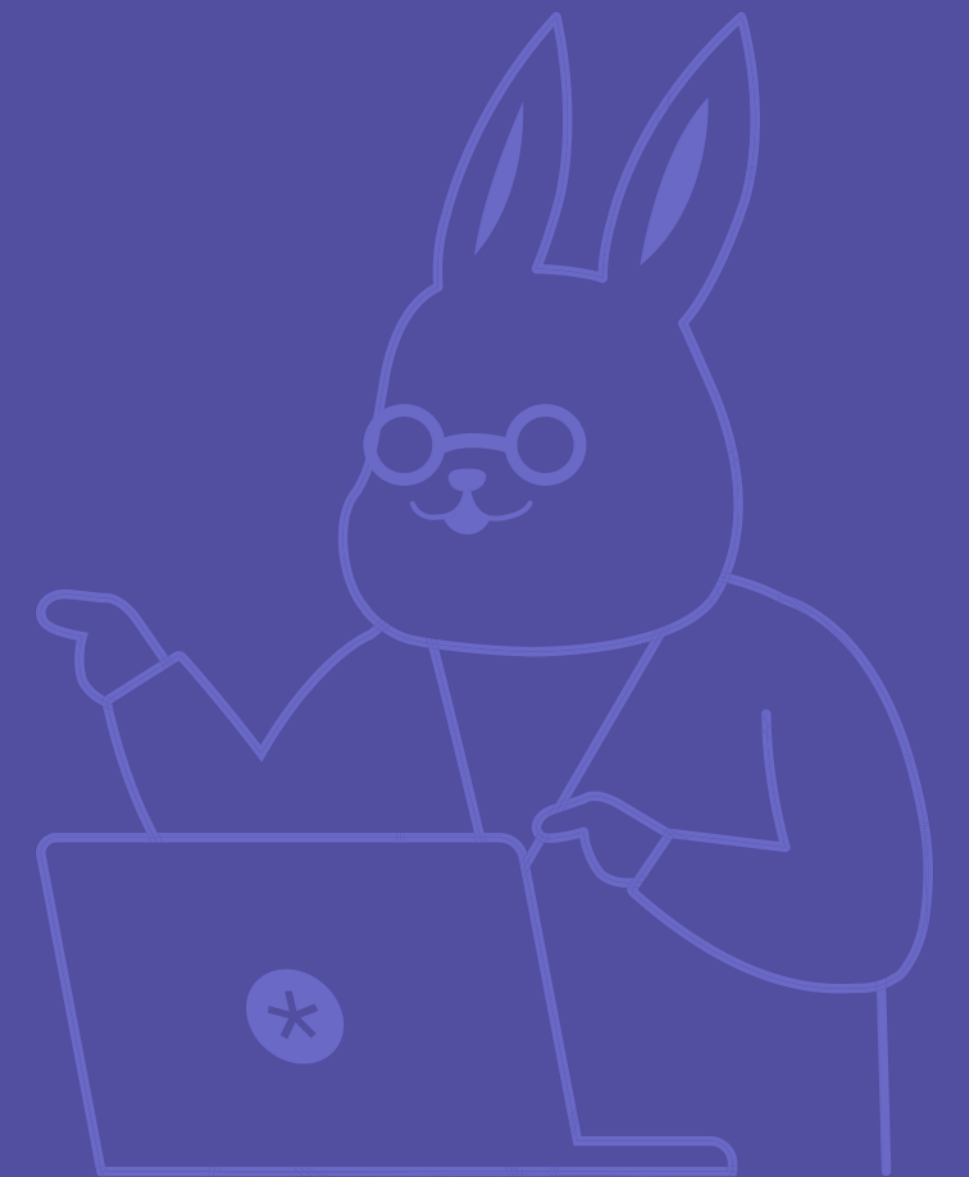
010-7254-3735

WEB

<https://elice.io>

E-MAIL

contact@elice.io



크레딧

코스 매니저
안조현

콘텐츠 제작자
안조현, 이상엽

강사
이상엽

디자이너
강혜정

