
CNN기반 대상포진 조기진단 AI모델 구축

3조 : ██████████ 최우연

목차

연구 개요

1. 연구 주제 및 동기
2. 연구 배경 및 문제 정의
3. 연구 목적 및 기대 효과
4. 연구 목표 및 차별점

관련 연구 및 데이터 전략

5. 관련 연구 탐색 및 분석
6. 데이터 세트 출처 및 수집
7. 데이터 전처리 및 증강 전략

모델 설계 및 학습

8. 모델 구조 및 비교 대상 설계
9. 모델 선택
10. KDE-CT 학습 방식 설계
11. 실험 구성 및 학습 파라미터

결과 분석 및 향후 계획

12. 평가 지표 설명 및 기준
13. 실험 결과 및 성능 비교
14. 모델 학습 파이프라인 및 서비스 적용 방안
15. 추후 학습 / 응용 확장 계획
16. 연구 한계 및 개선 방향
17. 참고문헌

1. 연구 주제 및 동기

대상포진(HerpesZoster)이란?



- 바리셀라 조스터 바이러스*(수두 바이러스)의 재활성화로 인해 발생
- 평생 발병 확률 10%~30%
- 피부 발진과 극심한 신경통 유발
- 대상포진 후 신경통(PHN)*은 수개월~수년 지속되어 삶의 질을 크게 저하

조기 진단의 중요성

- 조기 진단을 통해 치료를 신속히 시작함으로써 증상 악화를 효과적으로 예방
- PHN 발생률은 조기 치료 시 9% -> 치료 지연 시 30% 이상으로 증가 [1]
- 72시간 내 항바이러스 치료 미 제공 시 만성 질환으로 진행 가능 [2]
- 발병 초기 인지가 어려워 의료진의 정확한 분류가 중요
 - 수포 발생 전 비특이적 통증(근육통, 화끈거림 등)만 나타나는 경우가 많아 일반 통증과 혼동됨
- 신속한 예측 시스템이 있으면 만성 통증 예방 및 의료 비용 30~50% 절감 가능 [5]

사회적 문제 및 연구 필요성

- 고령화로 대상포진 환자가 지속 증가하면서 의료 시스템 부담 가중
 - 50세 이상 인구의 약 30%가 대상포진 경험 보유 [5]
 - 연간 약 100만 건의 대상포진 신규 발생, 이 중 약 20% 이상이 PHN으로 진행 [1]
- 조기 진단은 생산성 회복과 공공 보건 효율 개선에 기여
 - 근로 복귀 가속 및 불필요한 의료 비용 절감
- 피부 병변의 초기 변화를 객관적으로 분류하기 어려운 문제가 있음

딥러닝 도입 필요성

- CNN은 의료 영상/이미지에서 패턴 인식에 강점
 - 피부 병변* 진단에서 CNN 모델은 전통적 진단 대비 정확도 향상에 기여 [6]
- 피부 질환 이미지에서 색상, 경계, 텍스처 등 시각적 특징을 자동 추출 가능
 - 초기 병변 단계에서 다른 질환과 구별 가능성 증가 -> 오진 최소화
- 모바일 기반 진단 도구로 확장 가능 -> 접근성, 실시간 확보
 - 빠른 예측 -> 초기 치료 개입 가능성 증대 + 의료비 절감

"본 연구는 딥러닝(CNN)을 활용해 대상포진을 조기에 정확히 예측함으로써, 의료.사회적 부담을 줄이고 환자의 삶의 질을 향상시키고자 함"

대상포진 후 신경통(PHN)* : 대상포진 피부 병변이 치유된 후에도 90일 이상 지속되는 만성 신경병성 통증을 일컫는 용어
바리셀라 조스터 바이러스* : 수두와 대상포진을 유발하는 헤르페스 바이러스 계열의 신경침화성 바이러스
병변* : 질병으로 인해 피부나 조직에 나타나는 이상 변화 또는 손상 부위

2. 연구 배경 및 문제 정의

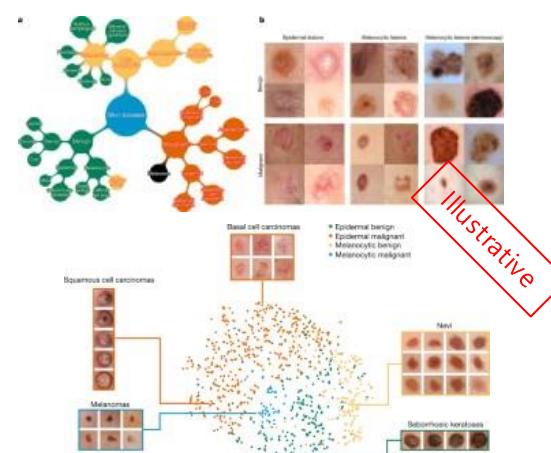
기존 분류의 한계

- ▶ 대상포진 병변은 진단 초기 단계에서 다른 피부 질환과 시각적으로 유사함
 - 발병 초기에 흉반, 수포 등이 나타남 - 접촉성 피부염, 단순포진, 곤충 물림 등과 유사한 외형
- ▶ 병변의 경계, 색상, 형태가 명확하지 않아 의료진 간 판별 일관성에 차이 발생
- ▶ 육안 진단은 경험의존성이 높아 객관적 분류 체계 부족
 - 피부과 전문의 간 진단 일치율 : 경계성 병변에서 평균 67% 수준 [7]



의료 영상 정량 분석 가능성

- ▶ 디지털 피부 영상은 색상 분포, 경계의 선명도, 병변의 텍스처* 변화 등을 수치화 및 정량 분석 할 수 있음
- ▶ 병변 주변 피부와의 명암 대비, 형태적 기하학적 패턴은 질환별 고유 시각 정보로 활용 가능
- ▶ 전통 진단에선 활용되지 않던 미세 시각 정보를 딥러닝 모델은 효과적으로 포착 가능 [6]

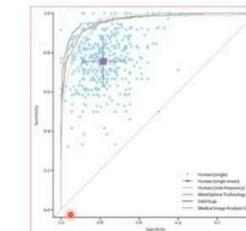


CNN 기반 모델의 적합성

- ▶ CNN(Convolutional Neural Network)은 이미지에서 가장자리, 질감, 색상 변화와 같은 국소 시각적 특징을 자동으로 학습하고 추출하는 데 특화된 딥러닝 알고리즘
- ▶ 이전 특징 선택 과정 없이, 시각적 차이를 계층적으로 학습할 수 있어 대상포진 병변 분류에 유리
- ▶ 특히 다양한 피부 타입, 조명 조건에서도 높은 일반화 성능을 기대할 수 있음
 - CNN 모델은 피부 병변의 경계, 색상 변화 등을 기반으로 전문의 수준 이상의 분류 정확도 달성 [9]

Dermoscopic images: melanoma vs nevus

- Dermoscopic image alone : AI > dermatologists



병변의 텍스처* : 표면의 거칠기, 융기, 평坦함 등의 측각적, 시각적 특성을 의미

3. 연구 목적 및 기대 효과

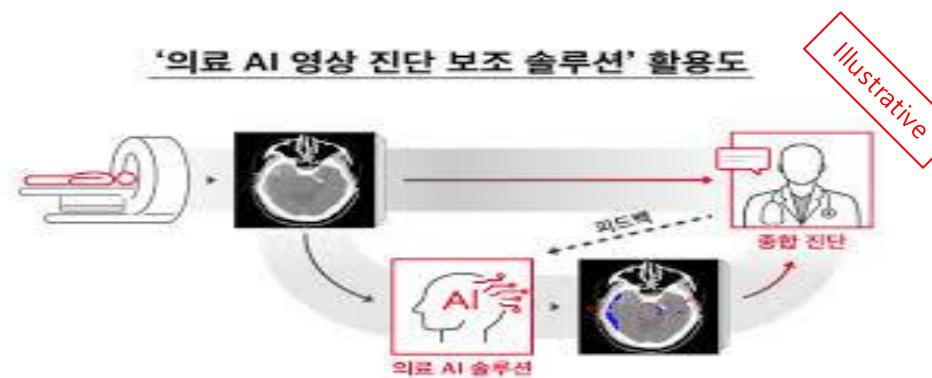
| 연구 목적 | 문제 및 해결전략 | 기대 효과 |
|--|---|---|
| <ul style="list-style-type: none">딥러닝 기반 영상 분석 기술(CNN)을 활용하여, 대상포진의 피부 병변을 조기에 정확히 진단 할 수 있는 인공지능 모델을 개발하고자 함기존 진단 방식의 시각적•임상적 오판 위험성을 보완비전문가 또는 1차 진료 기관에서도 정확한 조기 감별 진단 가능성 확보향후 모바일 기반진단 시스템으로 확장 가능성 고려  | <ul style="list-style-type: none">육안 진단의 주관성과 병변 간 유사성으로 인한 오진 가능성 -> CNN이 색상, 경계, 텍스처 등 시각적 특징을 정량 분석하여 진단의 일관성과 정확성 향상조기 진단 지연으로 인한 대상 포진 후 신경통(PHN) 발생 위험 증가 -> AI모델이 병변의 초기 시각적 패턴을 자동 분류하여 신속한 치료 개입 지원의료 접근성 부족(지방/고령층) -> 1차 진료 기관 활용이 가능한 경량 AI 시스템 진단 도구로 신속한 진단 가능 | <ul style="list-style-type: none">대상포진 후 신경통(PHN) 발생률 감소<ul style="list-style-type: none">조기 진단 시 PHN 발생률 약 30% -> 약 9%이하로 감소[4]의료비 절감 및 생산성 향상<ul style="list-style-type: none">대상포진/PHN 환자 1인당 연간 평균 추가 의료비(최대 \$3,000) [5] 절감AI 진단 정확도 향상<ul style="list-style-type: none">피부병변 분류 정확도 : CNN 약 86.6%, 피부과 전문의 평균 약 76% [9]모바일 기반 원격 진단 확장성<ul style="list-style-type: none">피부질환 AI 앱 진단 정확도 87%, 개발도상국 진단률 보다 높음 [12] |

4. 연구 목표 및 차별점

| 연구 목표 |
|--|
| <ul style="list-style-type: none">▶ 정상 피부와 대상포진 병변을 명확히 구별할 수 있는 다중 클래스 AI 모델 개발▶ 실제 의료기관 영상 기반 지속 학습 가능한 경량 모델 구축▶ 모바일 진단 시스템 확장에 최적화된 진단 정확도 확보▶ 임상적 실용성과 신뢰도를 고려한 진단 리포트 기능 제공 |

| 연구의 차별성 | |
|-------------------|------------------------------|
| 기존 연구 한계 | 본 연구의 차별점 |
| 피부질환 간 다중 분류에 집중 | 정상 vs 질환을 구분하는 선별 진단 구조 설계 |
| 대규모 데이터 기반 고성능 모델 | 소량 데이터 기반 경량 모델 최적화 |
| 모델 성능에 대한 정량적인 보고 | 의료현장 활용 가능한 시스템화 및 리포트 기능 포함 |
| 고해상도·정제된 이미지 중심 | 1차 의료기관 수집한 이미지 데이터 기반 학습 반영 |
| 모델 학습 후 구조 고정 | 지속 학습 가능한 AI 구조 설계 |

| 기술적 구현 전략 |
|--|
| <ul style="list-style-type: none">▶ KDE-CT(Knowledge-Distilled Efficient CNN with Curriculum Training) 기법 적용<ul style="list-style-type: none">• 경량 모델의 진단 정확도 향상 및 학습 안정성 확보▶ 추후 MobileNet, EfficientNet-lite 등 경량화 모델에 범용 적용 가능한 구현 설계 |



5. 관련 연구 탐색 및 분석

모바일 진단을 위한 대상포진 딥러닝 분류 모델 [8]

데이터 세트 구성 전략

- 대상포진 진단용 피부 질환 이미지 데이터 세트 구성
 - 공개 데이터 세트 수집
 - SD-198(6,548장, 198개 질환 클래스) + SD-260(20,600장, 260개 질환 클래스) 수집 - 단, 대상 포진 이미지는 각각 24장, 12장으로 학습에 부족함
 - 커스텀 데이터 세트 수집
 - Google/Bing에서 'herpes zoster' 키워드로 웹 크롤링*
 - 중복 제거 및 전문가 검토를 통해 총 377장의 대상포진 이미지 확보
- 강건성* 평가를 위한 데이터 세트 생성
 - 대상포진 진단용 피부 질환 이미지 데이터 세트에 총 75가지 손상 적용하여 모델의 강건성 평가에 활용



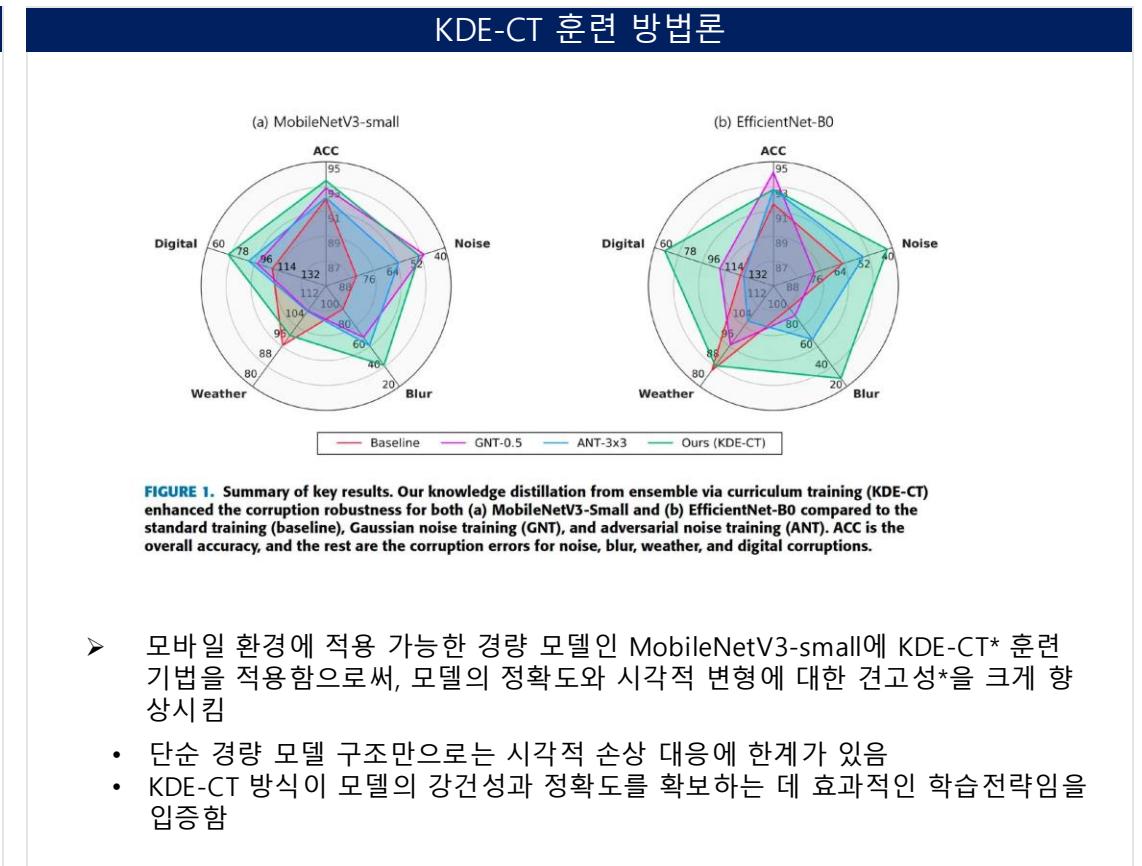
FIGURE 2. Examples images in SD-HZ dataset: (a) Acne (b) Herpes Zoster (c) Tinea (d) Other Disease.



FIGURE 3. Examples of corrupted images in SD-HZ-C dataset: (a)-(c) motion; (d)-(g) blur; (h)-(i) weather; and (j)-(k) digital corruptions with severity level of 3.

[대상포진 진단용 피부 질환 이미지 데이터 세트]

[강건성 평가를 위한 데이터 세트]



웹 크롤링* : 인터넷에서 자동으로 이미지를 수집하여 학습 데이터로 활용하는 과정

KDE-CT* : 지식 증류(teacher → student 모델)를 적용한 딥러닝 훈련 전략을 포함한 프레임워크
견고성 : 데이터 변화나 잡음에도 예측 성능이 크게 떨어지지 않는 일관되고 안정적인 학습 능력

5. 관련 연구 탐색 및 분석

딥러닝 단일 모델의 구조 및 성능 분석 [14]

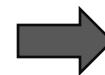
| 기본 CNN | |
|------------------|--|
| 1. 구성 요소: | <ul style="list-style-type: none">여러 개의 3x3 컨볼루션 레이어, ReLU 활성화 함수, 배치 정규화, 맥스풀링 레이어전결합층(Fully Connected Layer)과 Softmax 출력층으로 구성됨 |
| 2. 특징: | <ul style="list-style-type: none">단순하고 학습 속도가 빠르며 구현이 쉬움하지만 복잡한 피부 병변의 세부 패턴을 충분히 학습하기 어려움 |

| 지표 | CNN | ResNet-18 | EfficientNet-B0 |
|------------|------|-----------|-----------------|
| 정확도 | 80% | 87% | 97% |
| 민감도 | 64% | 82% | 99% |
| 특이도 | 95% | 93% | 93% |
| 정밀도 | 93% | 91% | 95% |
| F1-score | 76% | 86% | 97% |
| MCC | 0.75 | 0.82 | 0.94 |
| Type-I 오류 | 4% | 6% | 6% |
| Type-II 오류 | 3.6% | 1.8% | 1% |

| ResNet-18 | |
|------------------|---|
| 1. 구성 요소: | <ul style="list-style-type: none">총 18개의 레이어로 구성각 블록은 2개의 3x3 컨볼루션 레이어 + ReLU + 배치 정규화 포함잔차 연결(Residual Connection)을 통해 기울기 소실 문제를 해결 |
| 2. 특징: | <ul style="list-style-type: none">깊은 네트워크를 안정적으로 학습 가능CNN보다 복잡한 구조 학습에 유리함 |

흑생종 양성(Benign)과 악성(Malignant) 이진 분류

| EfficientNet-B0 | |
|------------------|---|
| 1. 구성 요소: | <ul style="list-style-type: none">MBConv 블록과 Depthwise Separable Convolution 기반Swish 활성화 함수와 Squeeze-and-Excitation 블록 포함입력: 224x224 이미지 → 점차 축소 → 7x7 특징 맵 → Global Average Pooling → Fully Connected → Softmax 출력 |
| 2. 특징: | <ul style="list-style-type: none">컴파운드 스케일링(compound scaling)으로 깊이, 너비, 해상도를 균형 있게 확장적은 연산량으로 높은 정확도를 달성하며, 모바일 환경에서도 효율적채널별 중요도를 재조정해주는 SE 블록 덕분에 민감도 향상 |



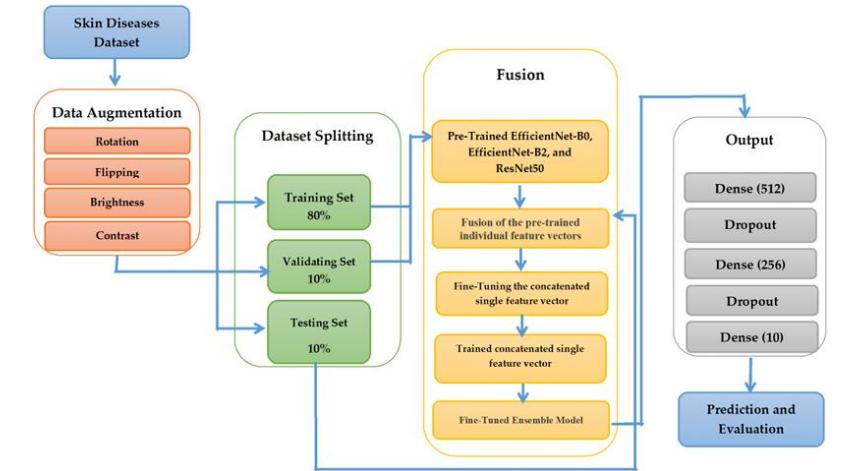
모든 지표에서 **EfficientNet-B0**가 가장 뛰어난 성능을 보여주며, 특히 민감도 99%는 흑색종과 같은 질환에 있어 매우 중요한 요소입니다.

5. 관련 연구 탐색 및 분석

융합 딥러닝 모델의 구조 및 성능 분석 [13]

| 융합 모델 구조 요약 | |
|-----------------------------|--|
| 1. 세 가지 사전 학습된 CNN모델 | <ul style="list-style-type: none"> EfficientNet-B0, EfficientNet-B2, ResNet50 |
| 2. 특징 벡터 추출 | <ul style="list-style-type: none"> 세 모델은 병렬 구조로 구성되어 각기 다른 시각적 특징 벡터 추출 |
| 3. 통합 벡터 구성 | <ul style="list-style-type: none"> 추출된 특징 벡터는 병합되어 하나의 통합 벡터로 구성 |
| 4. 분류 수행 | <ul style="list-style-type: none"> 통합된 특징은 Dense 계층*을 통해 다중 질환 분류 수행 |

| 융합 딥러닝 모델*의 성능 분석 | |
|-------------------|---|
| 1. 모델 성능 | <ul style="list-style-type: none"> 정확도 (Accuracy) : 99.14% 정밀도 (Precision) : 0.9914 재현율 (Recall) : 0.9914 F1 Score : 0.9914 AUC (ROC) : 0.93 ~ 1.00 (클래스별 평균) 과적합 여부 : 없음 클래스 불균형 대응 : 효과적 |
| 2. 모델 평가 | <ul style="list-style-type: none"> 기존 단일 모델 대비 모든 지표에서 우수한 성능 소수 클래스에 인식 정확도에서 높은 민감도와 낮은 오탐률을 유지 |



| 모델 | Precision | Recall | F1-score | Accuracy | 실행 시간 (초) |
|--|---------------|---------------|---------------|---------------|-----------|
| EfficientNet-B2 | 0.84 | 0.84 | 0.84 | 0.84 | 140 |
| ResNet101V2 | 0.80 | 0.80 | 0.80 | 0.80 | 170 |
| MobileNet-V3 | 0.85 | 0.85 | 0.85 | 0.85 | 100 |
| InceptionNet-V3 | 0.82 | 0.81 | 0.81 | 0.81 | 150 |
| Proposed Model (EfficientNet-B0+B2 + ResNet50) | 0.9914 | 0.9914 | 0.9914 | 0.9914 | 200 |

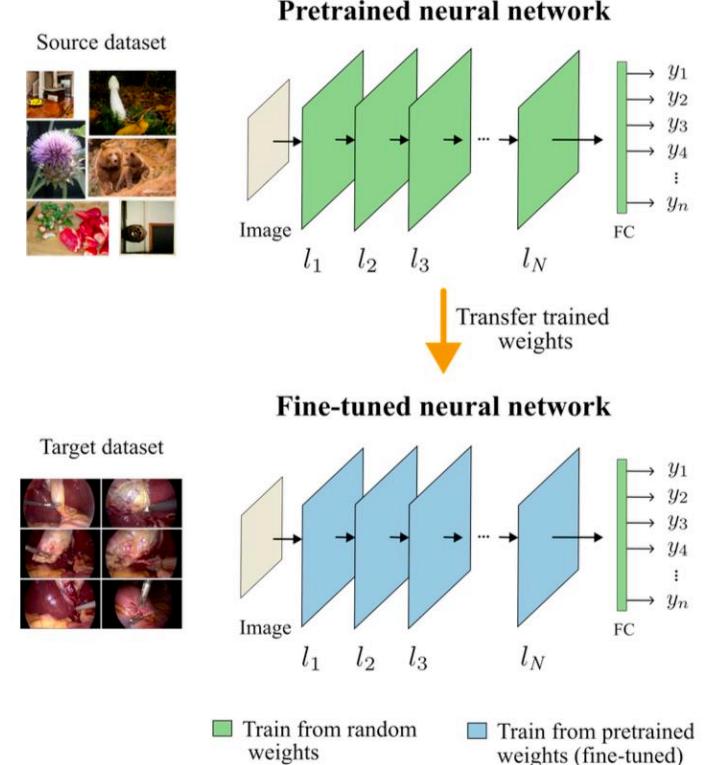
Dense 계층* : 다중 질환 분류를 위해 통합된 특징 벡터를 입력받아 최종 예측을 수행하는 완전 연결 신경망 계층
 융합 딥러닝 모델* : 서로 다른 CNN 기반 사전학습 모델들의 특성 벡터를 결합해 통합 학습한 양상을 기반 분류 모델

5. 관련 연구 탐색 및 분석

전이학습 및 Fine-Tuning에 의한 성능 향상 (1) [11]

| 문제 정의와 해결책 제시 |
|--|
| 1. 모델 적용의 한계 <ul style="list-style-type: none"> ▪ 사전 학습 모델의 의료 환경 직접 적용 부족 |
| 2. 데이터 특성 문제 <ul style="list-style-type: none"> ▪ 의료 데이터의 고유한 특성 ▪ 대규모 주석 데이터셋 부족 |
| 3. 전이 학습(Transfer Learning) 도입 <ul style="list-style-type: none"> ▪ 광범위한 데이터셋으로 훈련된 모델을 특정 의료 환경에 적용 ▪ 데이터 부족 문제 해결 |
| 4. 제약 사항 <ul style="list-style-type: none"> ▪ 원본-대상 도메인 일치도에 성능 의존 ▪ 분포 변화 시 성능 저하 위험 |

| 연구 범위 및 방법론 |
|--|
| 1. 연구 대상 도메인 <ul style="list-style-type: none"> ▪ 5개 의료 영상 분야 ▪ X-ray, MRI, 조직학, 피부과, 내시경 수술 |
| 2. Fine-tuning 전략 평가 <ul style="list-style-type: none"> ▪ 8가지 Fine-tuning 방법 ▪ 전체 계층 Fine-tuning, 분류기 계층만 Fine-tuning ▪ 점진적 Fine-tuning, 정규화 기반 Fine-tuning, 적응형 학습률 |
| 3. 실험 모델 <ul style="list-style-type: none"> ▪ 3개 CNN 아키텍처 ▪ ResNet-50, DenseNet-121, VGG-19 ▪ 다양한 학습 및 특징 추출 시나리오 포괄 |

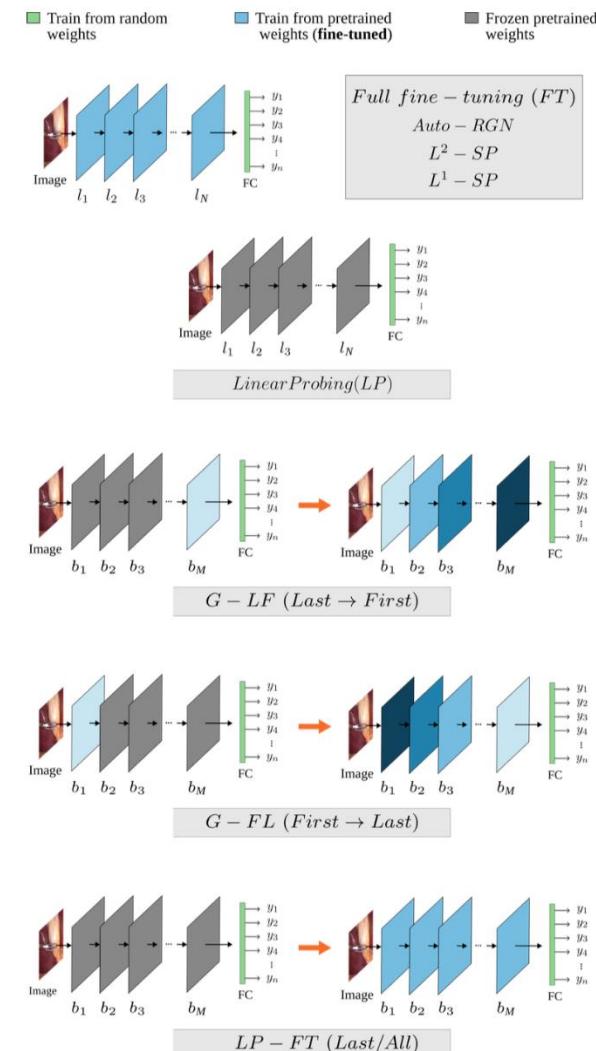


5. 관련 연구 탐색 및 분석

전이학습 및 Fine-Tuning에 의한 성능 향상 (2) [11]

| 평가된 Fine-tuning 전략 |
|--|
| 1. Full Fine-tuning (FT, 기본 전략) |
| ▪ 모든 계층 + 새 분류기 계층 동시 훈련 |
| 2. Linear Probing (LP, 기본 전략) |
| ▪ 사전 학습 계층 고정, 분류기 계층만 훈련 |
| 3. Gradual Unfreeze (Last First) (G-LF, 점진적 전략) |
| ▪ 마지막 블록부터 첫 번째 블록으로 순차적 해제 |
| 4. Gradual Unfreeze (First Last) (G-FL, 점진적 전략) |
| ▪ 첫 번째 블록부터 마지막 블록으로 순차적 해제 |
| 5. LP-FT (Hybrid)* |
| ▪ Linear Probing → Full Fine-tuning 단계적 전환 |
| 6. L ¹ -SP Regularization (정규화 기반 전략) |
| ▪ L ₁ -norm 기반 희소 정규화 적용 |
| 7. L ² -SP Regularization (정규화 기반 전략) |
| ▪ L ₂ -norm 기반 정규화 적용 |
| 8. Auto-RGN (적응형 학습률 전략) |
| ▪ 상대 기울기 노름(RGN)*에 따른 계층별 학습률을 자동 조정 |

| 주요 실험 결과 |
|---|
| 1. LP-FT (Linear Probing + Full Fine-tuning) 전략 |
| ▪ 평가 사례 50% 이상에서 개선 |
| ▪ 모든 아키텍처에서 mAP* 일관 개선 |
| 2. Auto-RGN 전략 |
| ▪ 특정 양식에서 최대 11% 개선 |
| ▪ X-ray, 조직학 (ResNet-50, DenseNet-121) |
| 3. DenseNet-121 모델 특성 |
| ▪ 대안적 Fine-tuning 접근법에서 뚜렷한 이점 |
| ▪ 24개 실험에서 Fine-tuning 전략 효과 확인 |
| ▪ 선호 전략: LP-FT, Auto-RGN |
| ▪ 기여도: 전체 개선 사례의 50% 차지 |
| 4. ResNet-50 & DenseNet-121 모델 특성 |
| ▪ 선호 전략: LP-FT, Auto-RGN |
| ▪ 기여도: 전체 개선 사례의 50% 차지 |
| 5. 결론 |
| ▪ 모든 데이터셋에서 일관된 우수 전략 없음 |
| ▪ 아키텍처와 의료 영상 유형별 전략 선택 중요 |



LP-FT* : 사전학습된 가중치를 고정한 후 분류기 계층만 학습(LP)한 뒤, 전체 계층을 점진적으로 학습(FT)하는 하이브리드 파인튜닝 전략

RGN* : 상대 기울기 크기를 기준으로 계층별 학습률을 자동 조정하는 적응형 파인튜닝 기법

MAP : 모델의 전체 예측 정확도를 평가하는 대표적인 성능 지표로, 클래스별 정밀도의 평균

6. 데이터 세트 출처 및 수집

다국적 피부 질환 공개 데이터 세트

- SD-198(6,548장, 198개 질환 클래스) [3][12]
 - 출처 : <https://huggingface.co/datasets/resyhgerwshshgdfghsdfgh/SD-198>
 - 수집 방식 : API 제공
 - 이미지(JPEG 형식)와 라벨(질병 라벨링) 매핑 파일(CSV 형식) 제공
- SD-260(20,600장, 260개 질환 클래스) [3][12]
 - 출처 : https://github.com/xpwu95/SPBL_Pytorch
 - 수집 방식 : 연구자에게 이메일로 요청
 - 이미지(JPEG 형식)와 라벨(단일 질병 라벨링) 매핑 파일(CSV 형식) 제공
 - SD-198을 확장·보완한 데이터 세트
- 피부과 전문의 등 의료 전문가에 의해 라벨링 된 데이터
- Herpes Zoster(대상포진) 라벨링 데이터 존재
 - 단, 대상 포진 이미지는 각각 24장, 12장으로 학습에 부족 함



AI-HUB 공개 데이터 세트

- 한국인 피부상태 측정 데이터(125,424장)
 - 출처 : <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=data&dataSetSn=71645>
 - 수집 방식 : 다운로드
 - 이미지(JPEG 형식)와 전문가에 의한 피부상태 항목(색소침착, 여드름 등) 라벨 매핑 파일(JSON 형식) 제공
 - Sensitive(민감도), Annotations(질환 정보 포함)을 참고하여 정상 클래스 이미지만 추출하여 활용할 예정
 - 연령·성별 다양 / 한국인 정상 피부 클래스 이미지 확보에 적합
- 피부 질환 진단 의료 이미지(67000장, 32개 질환 클래스)
 - 출처 : <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=data&dataSetSn=230>
 - 수집 방식 : 오프라인 안심존*에서 열람 가능한 제한사항이 있음
 - 대상포진 병변만 잘라낸 비 식별 이미지 데이터로 변환 -> 반출 신청 후 다운로드하여 활용할 예정
 - 국내 10개 대학 전문의에 의해 라벨링 된 데이터
 - Herpes 감염 클래스 이미지 포함
 - Herpes Zoster(대상포진) 케이스에 대해서는 SD-198/260 데이터 세트를 활용(전이 학습* 기반)하여 병변하여 활용할 예정

커스텀 수집

- 웹 크롤링 (Google 이미지) 통한 대상포진 이미지 수집

오프라인 안심존* : 외부 인터넷과 차단된 보안 환경에서 민감한 데이터를 안전하게 처리·학습할 수 있도록 구축된 물리적 보안 공간
전이 학습* : 이미 학습된 모델의 지식을 활용해 새로운 과제를 더 빠르고 적은 데이터로 학습하는 딥러닝 기법

7. 데이터 전처리 및 증강 전략

소수 클래스 대응 전략

| 데이터 증강 | | |
|--|--------------------|----------------------|
| <ul style="list-style-type: none">증강 라이브러리 : Alumentations<ul style="list-style-type: none">매우 빠르고 유연한 이미지 증강* 라이브러리의료 영상 처리에 적합한 연산들 포함PyTorch, TensorFlow 모두 호환 | | |
| <ul style="list-style-type: none">증강 전략 | | |
| 전략 | 설명 | 기대 효과 |
| Horizontal/ Vertical Flip | 이미지를 좌우 또는 상하 반전 | 좌우 대칭 질환 구조 대응 |
| Rotation | +15도 회전 | 다양한 촬영 각도에 대한 견고함 확보 |
| Zoom In/Out | 랜덤 비율로 중심 확대 또는 축소 | 병변의 크기 다양성 학습 |
| Translation | 병변 위치를 일정 픽셀 이동 | 위치 변화에 대한 불변성 강화 |
| Brightness/Contrast | 밝기/명암 대비 랜덤 조절 | 촬영 환경 변화 대응 |
| Color Jitter | Hue, Saturation 조절 | 조명이나 피부색 다양성 반영 |
| Gaussian Noise 추가 | 약한 랜덤 노이즈 삽입 | 촬영 장치의 노이즈 대응 |
| Cutout/GridMask | 이미지의 일부를 마스킹 | 특징 의존성 줄이고 학습 범용성 증가 |
| MixUp/CutMix | 두 이미지를 혼합 | 클래스 간 경계 일반화 학습 가능 |
| <ul style="list-style-type: none">증강 적용 범위<ul style="list-style-type: none">SD-198, SD-260 데이터 세트의 Herpes Zoster 클래스 이미지에 대해서만 적용(증강배수 : x10) | | |
| 합성 데이터 생성 | | |
| <ul style="list-style-type: none">원본 데이터 전처리<ul style="list-style-type: none">크기 표준화(224x224)색공간 정규화(RGB, HSV) - 밝기나 색상값을 일정 범위로 정규화Annotation 제거/ 병변중심 Crop - 사진의 가장 중요한 부분만 확대 | | |
| <ul style="list-style-type: none">GAN 기법* 활용 전략<ul style="list-style-type: none">사용 라이브러리: StyleGAN2, CycleGANSD-198-SD-260 데이터셋 활용: 공개된 피부질환 이미지 데이터세트 스스로 사용고해상도 합성: StyleGAN2 기반 전통적 GAN 구조로 학습해 고품질 합성 이미지 생성스타일 매핑: CycleGAN을 이용해 AI-HUB 한국인 피부상태 측정 데이터에 SD-198-SD-260의 대상포진 이미지를 매핑, 한국인 피부 특성이 반영된 스타일로 변환노이즈 제거 및 일반화: AI-HUB 피부질환 진단용 이미지를 SD-198-SD-260 데이터세트로 학습하여 환자별 고유 노이즈 제거, 평균적인 대상포진 특성만 남긴 합성 이미지로 변환하여 오프라인 안심존에서 반출하여 활용 | | |
| 전이 학습 | | |
| <ul style="list-style-type: none">사전 학습 (Pre-training)*<ul style="list-style-type: none">데이터 : SD-198 또는 SD-260목적 : 피부 병변 분류의 기본 시각 특징 추출 능력 확보모델 : RestNet50, DenseNet121, EfficientNet 등 CNN 계열 사용 | | |
| <ul style="list-style-type: none">전이 학습 (Transfer Learning)<ul style="list-style-type: none">대상 데이터 : AI-HUB + 웹 수집 이미지전이방식 : SD-198 또는 SD-260으로 학습된 모델에서 출력층만 교체 후 재학습목표 : Herpes Zoster 병변 분류 | | |
| <ul style="list-style-type: none">Optimizer* 설정<ul style="list-style-type: none">알고리즘 : Adam (빠르고 안정적인 최적화 방식)학습속도(η) : 1e-4 (학습 속도 조절. 너무 크면 불안정, 너무 작으면 느림)정규화(weight_decay) : 1e-5(모델 복잡도 제한, 과적합을 막기 위한 L2 정규화의 계수) | | |
| <ul style="list-style-type: none">Loss Function* (손실 함수)<ul style="list-style-type: none">함수 종류 : CrossEntropyLoss (다중 클래스 분류에 적합, 예측 정확도 향상)클래스 불균형 대응 : weight 인자 사용으로 소수 클래스에 가중치 부여 가능 | | |

이미지 증강* : 학습 데이터를 다양하게 변형하여 모델의 일반화 성능과 견고성을 높이는 데이터 확장 기법

GAN 기법* : 두 개의 신경망이 경쟁하며 학습해 실제와 유사한 이미지를 생성하는 딥러닝 생성 모델

사전학습* : 대규모 데이터로 먼저 학습한 모델의 지식을 다른 과제에 효과적으로 활용하기 위한 초기 학습 단계

Optimizer* : 모델의 예측 오류를 줄이기 위해 가중치를 자동으로 조정하는 학습 알고리즘

Loss Function* : 모델의 예측이 정답과 얼마나 다른지를 수치로 계산해 학습 방향을 알려주는 기준 함수

8. 모델 구조 및 비교 대상 설계

| CNN Ensemble | Base Models | Mobile Models |
|---|--|--|
| <ul style="list-style-type: none"> ➤ 특징 <ul style="list-style-type: none"> • 서로 다른 학습기를 결합하여 최종 예측을 수행하는 기법 • 여러 개의 개별 CNN 모델을 조합하여 성능을 향상시키는 방법 • 단일 모델의 한계를 극복하고 일관된 예측과 높은 정확도를 얻기 위해 사용됨 ➤ 장점 <ul style="list-style-type: none"> • 정확도, F1*, AUROC* 등 거의 모든 지표에서 최고 성능 ➤ 단점 <ul style="list-style-type: none"> • GPU* 자원 소모가 큼 - 입출력까지 걸리는 시간(Latency)이 오래 걸리며 모델 파라미터 수(Params)가 많음 [12] ➤ 적합 용도 <ul style="list-style-type: none"> • 대형 병원 서버에서 정밀 진단용, 연구 목적의 고정확도 시스템 | <ul style="list-style-type: none"> ➤ 특징 <ul style="list-style-type: none"> • 전통적인 CNN 구조로 정확도는 높은 편이지만, Mobile 모델에 비해 파라미터 수와 연산량이 많고, Ensemble 모델 보다 추론 속도가 다소 느림 • 학습이 안정적이며 다양한 데이터셋에 적합 • 모델 구조가 표준화되어 있어 다양한 커스터마이징이 쉬움 ➤ 장점 <ul style="list-style-type: none"> • 상대적으로 높은 정확도, 중간 정도의 연산 비용 • 경량화 모델에 비해 결과의 일관성이 높고 예측의 변동성도 적음 ➤ 단점 <ul style="list-style-type: none"> • Latency*가 Mobile 모델보다 높음 [12] - GPU 환경이 요구되지만 Ensemble 모델 보다는 저사양의 서버에서 활용 가능 ➤ 적합 용도 <ul style="list-style-type: none"> • GPU 환경이 제한되지 않은 병원 시스템, 클라우드 기반 서비스 | <ul style="list-style-type: none"> ➤ 특징 <ul style="list-style-type: none"> • 경량화된 구조로 모바일/임베디드 디바이스에 적합 • 낮은 연산량과 짧은 Latency를 가짐 [12] • 실시간 응답성이 우수하며, 저전력 디바이스에서도 실행 가능 ➤ 장점 <ul style="list-style-type: none"> • 모델 크기가 작고 연산량이 적어, CPU나 모바일 AP* 환경에서도 고속 추론 가능 • 경량 모델임에도 불구하고 비교적 우수한 성능을 보임 ➤ 단점 <ul style="list-style-type: none"> • 구조 단순화로 복잡한 의료 영상 특징을 충분히 반영하지 못함 • 모델 해석력 및 학습 유연성도 상대적으로 부족 ➤ 적합 용도 <ul style="list-style-type: none"> • IoT 의료기기, 모바일 헬스케어 앱 등 |

TABLE 2. Performances of 15 different models on the SD-HZ dataset (Latency: millisecond, Params.: million, MACs: million).

| Category | Model | ACC | F1 | PR | RE | F1-HZ | AUROC | Kappa | Latency | Params. | MACs |
|----------------------|-------------------|-------------|-------------|-------------|-------------|-------------|--------------|--------------|-------------|------------|----------|
| Basic models | AlexNet | 0.93 | 0.84 | 0.87 | 0.82 | 0.80 | 0.946 | 0.794 | 21.0 | 57.0 | 710 |
| | Vgg-16 | 0.93 | 0.85 | 0.87 | 0.86 | 0.79 | 0.967 | 0.779 | 140.3 | 134.3 | 1,550 |
| | InceptionV3 | 0.92 | 0.83 | 0.85 | 0.82 | 0.78 | 0.960 | 0.766 | 73.2 | 21.8 | 285 |
| | ResNet-50 | 0.94 | 0.87 | 0.87 | 0.88 | 0.81 | 0.979 | 0.797 | 93.3 | 23.5 | 412 |
| | ResNext-101 | 0.93 | 0.87 | 0.87 | 0.89 | 0.82 | 0.975 | 0.809 | 320.7 | 86.8 | 1,551 |
| | SEResNext-101 | 0.94 | 0.88 | 0.89 | 0.87 | 0.85 | 0.981 | 0.846 | 226.8 | 46.9 | 805 |
| | DenseNet-121 | 0.93 | 0.85 | 0.87 | 0.84 | 0.80 | 0.969 | 0.791 | 328.6 | 7.0 | 1,552 |
| Mobile models | EfficientNet-b0 | 0.92 | 0.81 | 0.82 | 0.83 | 0.68 | 0.954 | 0.658 | 55.5 | 4.0 | 40 |
| | MNasNet | 0.92 | 0.84 | 0.85 | 0.84 | 0.76 | 0.947 | 0.749 | 38.6 | 3.1 | 33 |
| | MobileNetV2 | 0.90 | 0.79 | 0.80 | 0.83 | 0.61 | 0.955 | 0.586 | 39.5 | 2.2 | 32 |
| | MobileNetV3-Small | 0.92 | 0.84 | 0.85 | 0.83 | 0.82 | 0.927 | 0.811 | 20.6 | 1.5 | 5 |
| DNN Ensemble | Ensemble-M | 0.95 | 0.89 | 0.92 | 0.88 | 0.82 | 0.979 | 0.880 | 666.7 | 142.2 | 2,746 |
| | Ensemble-E | 0.97 | 0.93 | 0.94 | 0.92 | 0.78 | 0.992 | 0.910 | 699.5 | 144.6 | 2,781 |

Illustrative

AUROC* : 모델이 양성과 음성을 얼마나 잘 구분하는지를 나타내는 지표로, 1에 가까울수록 분류 성능이 뛰어남

GPU* : 대규모 연산을 병렬로 처리해 딥러닝과 영상 처리에 최적화된 고속 연산 장치

Latency* : 입력을 받아 출력이 나오기까지 걸리는 응답 시간

모바일 AP 환경* : 스마트폰 등 모바일 기기에서 연산, 그래픽, AI 기능을 통합 처리하는 저전력 시스템

9. 모델 선택

딥러닝 단일 모델과 융합모델의 구조 및 성능 비교

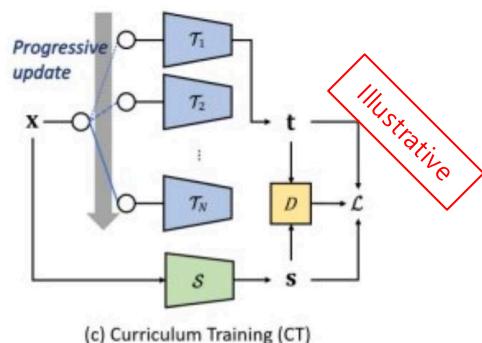
| EfficientNet-B0 | 융합 모델 |
|--|--|
| 1.구성 요소 <ul style="list-style-type: none">▪ MBConv 블록과 Depthwise Separable Convolution 기반▪ Swish 활성화 함수와 Squeeze-and-Excitation 블록 포함▪ 입력: 224x224 이미지 → 점차 축소 → 7x7 특징 맵 → Global Average Pooling → Fully Connected → Softmax 출력 | 1.세 가지 사전 학습된 CNN모델 <ul style="list-style-type: none">▪ EfficientNet-B0, EfficientNet-B2, ResNet5 |
| 2.특징 <ul style="list-style-type: none">▪ 컴파운드 스케일링(compound scaling)으로 깊이, 너비, 해상도를 균형 있게 확장▪ 적은 연산량으로 높은 정확도를 달성하며, 모바일 환경에서도 효율적▪ 채널별 중요도를 재조정해주는 SE 블록 덕분에 민감도 향상 | 2.특징 벡터 추출 <ul style="list-style-type: none">▪ 세 모델은 병렬 구조로 구성되어 각기 다른 시각적 특징 벡터 추출 |
| | 3.통합 벡터 구성 <ul style="list-style-type: none">▪ 추출된 특징 벡터는 병합되어 하나의 통합 벡터로 구성 |
| | 4.분류 수행 <ul style="list-style-type: none">▪ 통합된 특징은 Dense 계층을 통해 다중 질환 분류 수행 |

단일 모델과 융합 모델은 서로 다른 데이터 셋으로 성과를 도출하였으므로, 같은 데이터 셋으로 테스트가 필요.

10. KDE-CT 학습 방식 설계

KDE-CT 학습방식

- 여러 Teacher 모델*들의 지식을 집약해 하나의 Student 모델*에게 전달하는 방식
- 의료 진단처럼 다양한 특성을 포괄해야 하는 문제에서 일관성 + 일반화 성능을 극대화 할 수 있는 효과적인 전략
- 활용 목표
 - 여러 가지 서로 다른 특성을 가진 Base 모델*들을 Teacher로 사용하여, 각 모델의 예측 방식과 학습된 특징들을 종합
 - 이를 통해 하나의 가벼운 Base 모델(Student)이 여러 Teacher 모델로부터 지식을 배워, 더 풍부한 표현력과 높은 일반화 성능을 가질 수 있도록 학습하는 방식
 - 각 Teacher 모델이 학습한 중요한 특징(feature representation)을 통합하여, Student 모델이 더 넓은 시각과 판단 기준을 가질 수 있게 함



KDE-CT 구성 전략

- Teacher Models 구성
 - 서로 다른 방식으로 영상 특징을 학습하여 Student 모델에게 다양한 관점의 지식을 전달하도록 구성
 - ResNet-50 : 깊은 계층 구조와 잔차 연결을 갖춘 안정적인 모델, 패턴을 일 반화하는 능력이 우수
 - SE-ResNext-101 : 병렬 경로 기반 구조에 SE라는 주의력 매커니즘*이 추가된 모델, 정확하고 정교한 학습이 가능
 - InceptionV3 : 다양한 커널크기를 활용한 구조, 다중 스케일의 정보를 효과적으로 학습할 수 있는 모델
- Student Model 구성
 - DenseNet-121
 - 이전 층의 특징을 다음 층에서 계속 활용할 수 있는 구조로 설계됨 - 여러 Teacher 모델들로부터 받은 지식을 효율적으로 통합
 - 연산량 및 파라미터 수가 상대적으로 적어 실시간 진단 시스템에 적합
 - Base 모델 계열 중 하나로, 추론 결과가 안정적이고 다양한 응용 환경에도 확장이 용이
- 기대 효과
 - 정확도 유지 : 여러 Teacher 모델들의 예측을 평균내어 Student 모델이 학습함으로 Ensemble 효과와 유사함
 - 경량화 : 학습 단계에서는 여러 Teacher 모델을 활용하지만 실제 추론 시에는 Student 모델 하나만 사용함으로 추론 속도가 빠르고 자원 부담이 적음
 - 소수 클래스 대응력 강화 : KDE-CT 기법이 경계 사례* 및 이상치에 강건성 부여

Teacher 모델* : Student 모델에게 예측 지식이나 표현 학습을 전달하는 학습 지도자 역할의 사전 학습된 모델
Student 모델* : Teacher 모델의 지식을 전달받아 경량화된 형태로 학습하며, 최종 추론을 담당하는 모델

Base 모델* : 다양한 구조의 CNN 중 성능 비교와 전이학습의 기준이 되는 기본 모델

주의력 매커니즘* : 입력 중 중요한 정보에 가중치를 부여해 모델이 핵심 특징에 집중하도록 돋는 기법

11. 실험 구성 및 학습 파라미터

| 데이터 분할 전략 | | |
|---|----------|-----------|
| [데이터 분할 구성] | | |
| Train(70%) | Val(15%) | Test(15%) |
| <ul style="list-style-type: none">Train : 모델 학습에 사용되는 데이터Validation : 학습 중 과적합 확인 및 하이퍼파라미터 조정Test : 최종 모델 성능 평가(훈련에 사용되지 않음) | | |
| [데이터 분할 기법] | | |
| <ul style="list-style-type: none">Stratified Split 적용<ul style="list-style-type: none">데이터를 나눌 때 각 클래스의 비율이 모든 그룹(Train, Val, Test)에 비슷하게 유지되도록 나누는 방법클래스 간 균형 유지 -> 학습 및 평가에 공정성 확보불균형 데이터 문제 완화모델이 소수 클래스도 학습할 수 있는 기회 제공 | | |
| <ul style="list-style-type: none">적절한 데이터 분할과 하이퍼파라미터 설정은 모델 일반화의 핵심모델 fine-tuning*은 학습 후반부에서 성능 상승에 매우 중요 | | |

| 학습 하이퍼파라미터 설정 | | |
|--|--|----------------------------|
| 항목 | 설명 | 적용 값 |
| Batch Size | 모델이 한번에 학습하는 입력 데이터 개수, CPU 메모리 고려하여 설정 | 16~32 |
| Epochs | 전체 학습 데이터를 모델이 한 번 모두 학습하는 횟수, Early Stopping* 기준 포함 | 50~100 |
| Learning Rate | 모델이 오차를 줄이기 위해 가중치를 얼마나 크게 조정할지 결정하는 학습 속도 | 1e-3 |
| Optimizer | 오차를 최소화할 수 있도록 가중치를 효과적으로 조정해주는 알고리즘 | Adam + Weight Decay (1e-4) |
| Loss Function | 모델의 예측값과 정답 사이의 오차를 수치로 계산해주는 함수 | CrossEntropy |
| Fine-tuning 전략 | | |
| <ul style="list-style-type: none">LP-FT* (Linear Probing + Full Fine-tuning) 적용.1단계 (Linear Probing): 모델의 Convolutional 계층*을 고정하고, 새로 추가된 분류기 계층만 대상포진 데이터로 훈련.2단계 (Full Fine-tuning): 이후 모든 계층을 고정 해제하여 전체 모델을 재훈련.학습률 설정: 단계별로 학습률을 조정하며 최적 값 탐색 (예: 초기 Linear Probing 시 높게, Full Fine-tuning 시 낮게).최적화 기법: Adam 옵티마이저 사용.조기 종료: 검증 성능 기반으로 5 에포크의 patience를 가진 조기 종료 적용.데이터 전처리: 이미지 크기 224x224로 조정 및 ImageNet 평균/표준편차로 정규화.데이터 증강: 모델 일반화 및 성능 향상을 위해 다양한 증강 기법 적용 (예: 회전, 반전 등).평가 지표: 정확도, F1-Score, AUROC를 주요 성능 지표로 활용.플랫폼: Hugging Face transformers* 라이브러리를 활용하여 구현 및 테스트 진행. | | |

fine-tuning* : 사전 학습된 모델을 특정 작업이나 데이터셋에 맞춰 추가 학습시켜 성능을 향상시키는 과정

Early Stopping* : 모델이 과적합되기 전에 검증 성능이 더 이상 향상되지 않으면 학습을 조기 종료하는 기법

LP-FT* : 사전 학습된 가중치를 고정한 후 분류기 계층만 학습(LP)한 뒤, 전체 계층을 점진적으로 학습(FT)하는 하이브리드 파인튜닝 전략

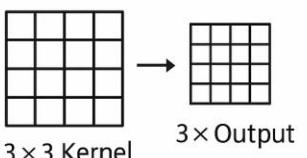
Hugging Face transformers : 사전 학습된 트랜스포머 모델을 쉽게 적용·파인튜닝할 수 있는 라이브러리

모델의 Convolutional 계층 : 이미지에서 중요한 특징(모양, 경계, 질감 등)을 찾아내는 필터 역할을 하는 계층

12. 평가 지표 설명 및 기준

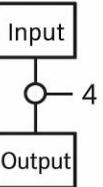
| 성능 지표 | | | 계산 효율성 지표 | | |
|---------------|--|-------------|-----------|---|---------------|
| 지표 | 설명 | 권장수준 [4] | 지표 | 설명 | 권장수준 [12] |
| F1-Score | 정밀도와 재현율의 조화 평균. 불균형 데이터셋에서 핵심 지표. | ≥ 0.85 | MACs | Multiply-Accumulate 연산 수(계산량) 연산량 기반 효율성 지표 | $\leq 2,000M$ |
| Recall | 재현율, 민감도. 질병 분류에서 놓치는 사례 방지를 위해 중요. | ≥ 0.85 | Latency | 실시간 응용을 위한 추론 지연 시간 기준 이미지 1장을 처리하는 시간(테스트 지연시간) | $\leq 100ms$ |
| Precision | 정밀도, 과진단 최소화를 위한 지표 | ≥ 0.85 | Params | 파라미터 개수(모델 저장 크기) 추론 및 배포 환경에 적합한 경량성 기준 | $\leq 10M$ |
| AUROC | ROC-CURVE* 하단 면적, 클래스 전체에 대한 예측 능력. 비교적 클래스 불균형에 덜 민감. | ≥ 0.90 | | | |
| F1-HZ | 대상포진(HZ) 클래스의 F1 점수 | ≥ 0.85 | | | |
| ACC-HZ | 대상포진(HZ) 클래스의 정확도 | ≥ 0.90 | | | |
| Cohen's Kappa | 분류기 예측과 실제 값의 일치도를 평가. | ≥ 0.80 | | | |

MACs 계산



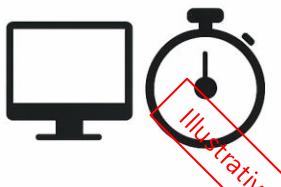
3x3 Kernel → 3x Output

Params 계산



Input → 4 → Output

Latency 측정



Monitor and Stopwatch

$F1-HZ = 2 \times \frac{Precision_{HZ} \times Recall_{HZ}}{Precision_{HZ} + Recall_{HZ}}$

대상포진 클래스에 대해 계산한 정밀도 (Precision) 와 재현율 (Recall) 의 조화 평균.

불균형 데이터에 적합한 F1-Score 및 Recall 중심의 성능 최적화 전략

$ACC-HZ = \frac{TP_{HZ} + TN_{HZ}}{TP_{HZ} + TN_{HZ} + FP_{HZ} + FN_{HZ}}$

대상포진 클래스에 대해 올바르게 예측한 비율

Illustrative

CNN 모델의 연산 복잡도와 추론 효율성을 정량적으로 평가하는 핵심 지표 비교

Base 모델의 일반화 성능 수준을 기준으로, KDE-CT 적용 후 향상된 성능을 비교하여 모델을 선택하는 전략

ROC-CURVE* : 분류 모델의 민감도(Recall)와 특이도(False Positive Rate) 간의 trade-off를 시각적으로 보여주는 그래프

13. 실험 결과 및 성능비교

Grad-CAM* 시각화

- 모델이 이미지 분류를 할 때 어떤 영역에 주목했는지를 시각적으로 보여주는 XAI*(설명 가능한 인공지능) 기법
- 모델이 예측을 위해 집중한 병변 영역이 붉게 표시됨

Illustrative

모델 성능 비교표

- 모델 성능 비교표를 통해 다양한 지표를 한눈에 정리하여 모델 간 차이를 직관적으로 파악

| Category | Model | ACC | F1 | PR | RE | F1-HZ | AUROC | Kappa | Latency | Params. | MACs |
|---------------------|---------------|------|------|------|------|-------------|-------|-------|---------|---------|-------|
| Basic models | AlxNet | 0.93 | 0.84 | 0.87 | 0.82 | 0.80 | 0.946 | 0.794 | 21.0 | 57.0 | 1550 |
| | Vgg-16 | 0.93 | 0.85 | 0.87 | 0.86 | 0.79 | 0.967 | 0.779 | 140.3 | 134.3 | 285 |
| | InceptionV3 | 0.92 | 0.83 | 0.85 | 0.82 | 0.78 | 0.960 | 0.766 | 73.2 | 21.8 | 285 |
| | ResNet-50 | 0.94 | 0.87 | 0.87 | 0.88 | 0.81 | 0.979 | 0.797 | 93.3 | 23.5 | 412 |
| | ResNext-101 | 0.93 | 0.87 | 0.87 | 0.89 | 0.82 | 0.975 | 0.809 | 320.7 | 86.8 | 1,651 |
| | SEResNext-101 | 0.94 | 0.88 | 0.89 | 0.87 | 0.85 | 0.981 | 0.846 | 226.8 | 46.9 | 805 |
| | DenseNet-121 | 0.93 | 0.85 | 0.87 | 0.84 | 0.80 | 0.969 | 0.791 | 328.6 | 7.0 | 285 |

Illustrative

Confusion Matrix

- 다중분류에서 특히 소수 클래스의 오분류 경향을 명확히 파악
- 불균형 클래스의 성능 저하를 진단하고 개선 방향을 제시하는데 필수적

| | | Predicted Class | | Sensitivity $\frac{TP}{(TP + FN)}$ |
|--------------|----------|--|---|---|
| | | Positive | Negative | |
| Actual Class | Positive | True Positive (TP) | False Negative (FN) Type II Error | |
| | Negative | False Positive (FP) Type I Error | True Negative (TN) | |
| | | Precision $\frac{TP}{(TP + FP)}$ | Negative Predictive Value $\frac{TN}{(TN + FN)}$ | Specificity $\frac{TN}{(TN + FP)}$ |
| | | | | Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$ |

Illustrative

Learning Curves*

- 모델의 학습 과정과 성능 변화를 직관적으로 파악 - 과적합(overfitting), 과소적합(underfitting), 학습 정체 등의 문제를 조기에 발견 가능

Illustrative

XAI* : AI 모델의 예측 결과를 사람도 이해할 수 있도록 설명하고 해석할 수 있게 만드는 기술

과적합* : 모델이 학습 데이터에 너무 치우쳐서 새로운 데이터에 일반화되지 못하는 현상

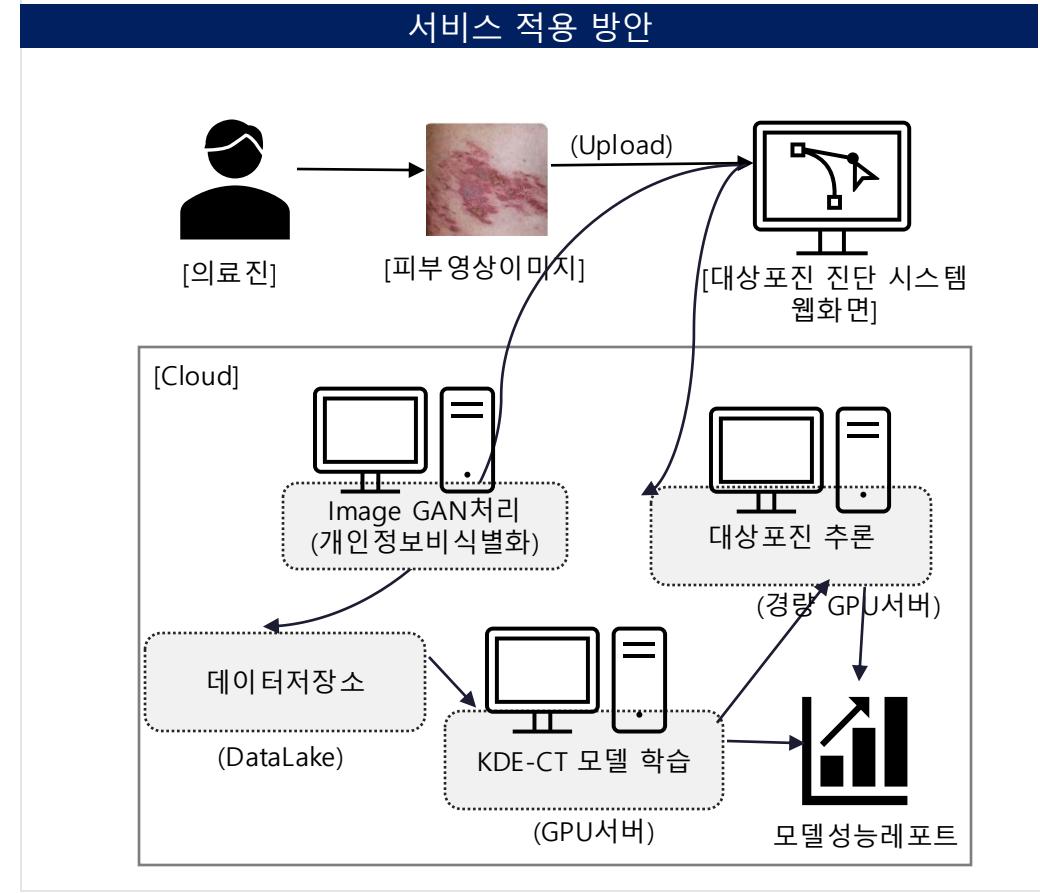
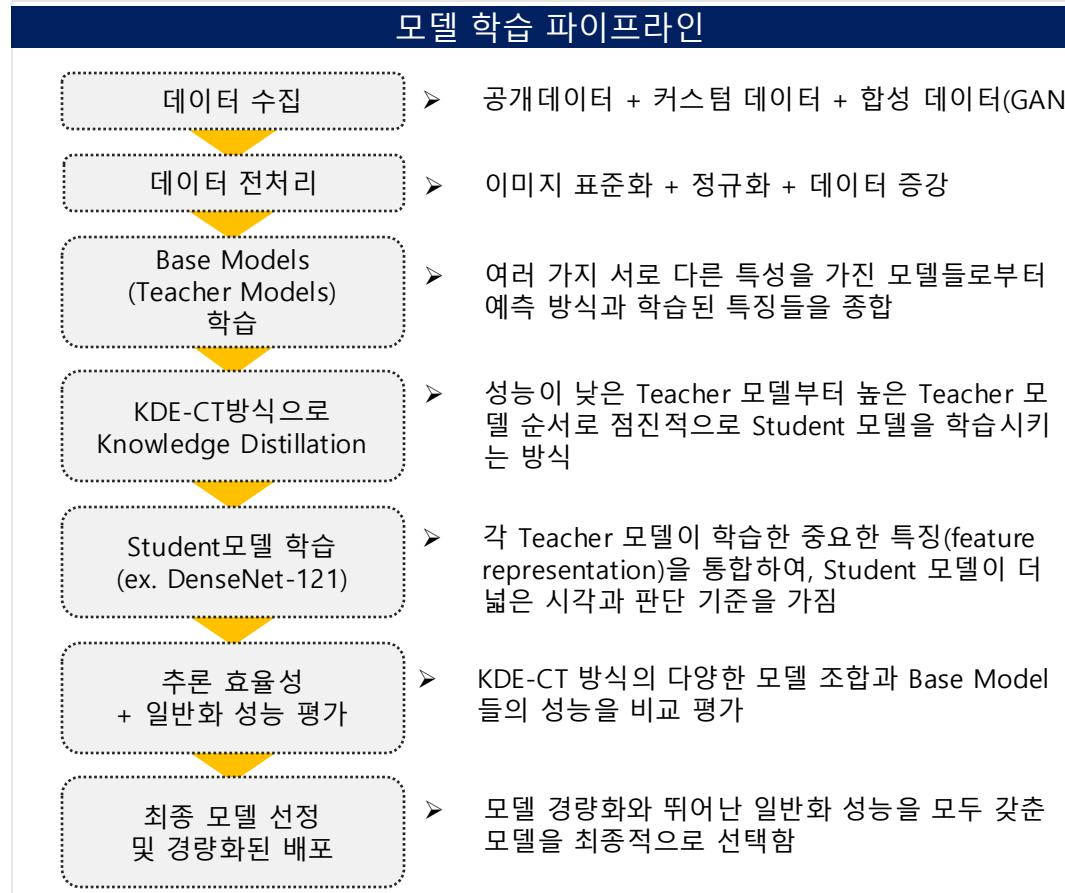
과소적합* : 모델이 학습 데이터의 패턴조차 제대로 학습하지 못해 성능이 낮은 상태

Grad-GAM* : CNN의 특정 클래스 예측에 기여한 이미지 영역을 기반으로 강조해 시각화하는 기법

Learning Curves* : 에폭(epoch) 또는 데이터 양에 따른 훈련-검증 성능 변화를 그래프로 나타내어 모델의 과대적합-과소적합 상태를 진단하는 도구

14. 모델 학습 파이프라인 및 서비스 적용 방안

1차 의료기관용 조기진단 시스템 구축을 위한 적용 전략



15. 추후 학습 / 응용 확장 계획

| 대상포진 병기별 자동 분류 모델 | 모바일 조기진단 시스템 | AI 맞춤형 뷰티 케어 추천 플랫폼 |
|--|---|--|
| <ul style="list-style-type: none"> ➤ 연구 목적 <ul style="list-style-type: none"> ○ 조기 단계부터 중증 단계까지 대상포진의 진행 정도를 자동 판단 ➤ 활용 방안 <ul style="list-style-type: none"> ○ 임상 치료 시기 결정 지원, 병기별 맞춤 진료 가능 ➤ 기술적 방안 <ul style="list-style-type: none"> ○ 병기별 라벨링된 데이터 수집 및 증강 ○ Grad-CAM 등 XAI 기법으로 신뢰성 확보 | <ul style="list-style-type: none"> ➤ 연구 목적 <ul style="list-style-type: none"> ○ 현장 접근성이 높은 AI 진단 솔루션 제공 ➤ 활용 방안 <ul style="list-style-type: none"> ○ 환자 자가진단, 1차 병원 또는 원격의료 보조 도구로 활용 ➤ 기술적 방안 <ul style="list-style-type: none"> ○ 경량화 모델(MobileNet, EfficientNet 등) 적용 ○ 모바일 앱/웹 플랫폼과 연동 ○ on-device 추론* 또는 클라우드 연산* 선택 가능 | <ul style="list-style-type: none"> ➤ 연구 목적 <ul style="list-style-type: none"> ○ 전문가가 없는 상황에서도 정확하고 신뢰할 수 있는 뷰티 컨설팅 경험 제공 ➤ 활용 방안 <ul style="list-style-type: none"> ○ 피부과 진료 전 AI 기반 사전 진단 & 시술 제안 ➤ 기술적 방안 <ul style="list-style-type: none"> ○ CNN 기반 피부 질환/피부 타입 분류 (여드름, 색소침착, 흥조 등) ○ Grad-CAM, ScoreCAM 등으로 투명성 확보 |



16개 질문에 의한
7가지 피부 유형

- ① 보습 부족 건조 피부
- ② 노화 건조 피부
- ③ 피지 트러블 피부
- ④ inner dry 피부
- ⑤ 기미 주의 피부
- ⑥ 마스크성 민감 피부
- ⑦ 건조성 민감 피부

AI가 측정하는
6개 항목

- ① 기미
- ② 주름
- ③ 살결
- ④ 윤기
- ⑤ 오일 성분
- ⑥ 눈 밑 처짐

Illustrative

on-device 추론* : AI 모델이 클라우드가 아닌 사용자의 디바이스(예: 스마트폰)에서 직접 예측을 수행하는 방식

클라우드 연산* : 데이터 처리와 AI 추론을 인터넷 기반 서버에서 수행하여 디바이스의 부담을 줄이는 방식

Grad-CAM* : CNN의 특정 클래스 예측에 기여한 이미지 영역을 기울기 기반으로 강조해 시각화하는 기법

Score-CAM* : gradient 없이 출력 점수 기반으로 CNN이 주목한 영역을 시각화하는 CAM 기법으로, 안정성과 해석력이 향상된 방식

16. 연구 한계 및 개선 방향

| 연구 한계 | 개선 방향 |
|---|---|
| <ul style="list-style-type: none">➤ 의료 데이터 활용의 법적 제약 [10]<ul style="list-style-type: none">○ 의료법과 개인정보보호법이 의료 데이터 사용에 엄격한 규제를 두고 있어, 연구 목적으로 자유로운 활용이 어려움○ 고해상도 피부 영상은 환자 식별에 활용될 수 있는 정보가 남아 있어 민감정보로 분류될 위험이 있음○ 익명화(비식별 처리) 이후에도 법적 해석이 명확하지 않아, 데이터 활용 가능 범위를 확실히 판별하기 어려움➤ 법령 간 충돌과 데이터 활용 불확실성 [10]<ul style="list-style-type: none">○ 2020년 데이터3법* 개정으로 가명정보* 활용이 법적으로 허용되었지만, 의료법은 여전히 환자 진료 데이터 활용을 엄격히 제한하고 있음○ 정부·업계 가이드라인은 권고사항에 불과해 법적 강제력이 없으므로, 실제 준수 범위가 명확하지 않음○ 연구 현장에서는 어느 법령을 우선 적용해야 할지, 어떤 내부 절차를 수립·준수해야 할지, 위반 시 처벌이나 소송 가능성은 어떻게 되는지 등에 대한 혼란과 부담이 지속되고 있음➤ 모델 일반화의 어려움<ul style="list-style-type: none">○ 병원 간 장비, 조명, 인종, 피부톤 등 편차가 큼○ 도메인 편향(domain shift) 문제로 CNN 모델 일반화 성능 저하○ 공공 이미지 데이터셋 부족 → 학습 다양성 확보 어려움 | <ul style="list-style-type: none">➤ 보건의료데이터 전용 법 제정 및 가이드라인 강화<ul style="list-style-type: none">○ 「보건의료데이터법」 제정을 통해 → 의료영상 AI 연구의 법적 불확실성 제거, 활용 기준 명확화○ 표준 비식별 처리 프로토콜과 법적 보호 절차 마련 필요 → AI 학습용 의료 이미지 안전 활용 기반 구축➤ 데이터 거버넌스* 및 공공 인프라 구축<ul style="list-style-type: none">○ 의료기관 간 안전한 데이터 연계를 위한 데이터 거버넌스 체계 수립 필요○ 중앙기관 주도의 대상포진 이미지 공공 데이터셋 구축 → 모델 일반화 성능 향상 및 공정성 확보➤ 연합학습* 및 프라이버시 보호 학습* 적용<ul style="list-style-type: none">○ 연합학습(Federated Learning) → 데이터 이동 없이 각 병원 내에서 분산 학습, 개인정보 보호 강화○ 프라이버시 보호 기술(PETs) 도입 고려 → 예: Differential Privacy, Homomorphic Encryption 등 → 민감정보 유출 방지 + AI 학습 가능 |

의료영상 기반 대상포진 AI 진단 모델의 법적·기술적 한계를 극복하기 위해,
보건의료데이터법 제정, 공공 데이터 거버넌스 구축, 프라이버시 보호 기술 적용이 필수적

데이터 3법* : 개인정보 보호법, 정보통신망법, 신용정보법을 개정해 가명정보 활용을 허용한 데이터 기반 산업 활성화 법

가명정보* : 개인을 식별할 수 없도록 일부 정보를 삭제하거나 대체하여 처리한 개인정보로, 통계·연구 등 특정 목적 하에 제한적으로 활용될 수 있는 데이터

데이터 거버넌스* : 데이터의 수집, 관리, 활용, 보안을 체계적으로 규정하고 통제하는 조직의 정책과 절차 체계

연합학습* : 데이터를 중앙 서버에 모으지 않고 각 디바이스에서 분산 학습을 진행한 후 모델만 공유하는 개인정보 보호형 AI 학습 방식

프라이버시 보호 학습* : AI 모델이 개인 정보를 노출하지 않도록 암호화, 익명화 등 기술을 적용해 안전하게 학습하는 방법

17. 참고문헌

| No | 참고 문헌 | 저자 | 내용 |
|----|--|---|---|
| 1 | Herpes Zoster and Postherpetic Neuralgia | Robert H. Dworkin, ... | 대상포진 및 PHN의 병태생리, 임상 경과, 치료, 예방 전략 |
| 2 | Acyclovir Therapy of Herpes Zoster | Richard J. Whitley | 대상포진 치료 전략과 약물 비교, 조기 개입의 효과 |
| 3 | Self-Paced Balance Learning for Clinical Skin Disease Recognition | Jufeng Yang, Xiaoping Wu,... | 클래스 불균형 문제를 해결하기 위한 알고리즘을 제안 |
| 4 | The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets | Takaya Saito와 Marc Rehmsmeier | 불균형 데이터셋에서의 효율적인 성능지표 검증 |
| 5 | Healthcare costs related to herpes zoster and postherpetic neuralgia in adults | Barbara P. Yawn, Lin Y., ... | 대상 포진 조기 진단과 예방의 경제적 필요성을 강조 |
| 6 | Dermatologist-level classification of skin cancer with deep neural networks | Andre Esteva, Brett Kuprel, ... | 딥러닝(CNN)을 활용하여 인공지능 진단 체계 가능성 입증 |
| 7 | Reliability of skin lesion diagnosis and management decisions by dermatology trainees: a teledermatology study | Edward Tan et al. | 피부 병변의 영상 기반 진단에서 의료진 간 진단 일관성을 분석 |
| 8 | Robust Skin Disease Classification by Distilling Deep Neural Network Ensemble for the Mobile Diagnosis of Herpes Zoster | 백승혁, 이성주, 신성호, 유영옥, 유태경, 종재포미, 류승준, 이규빈 | 대상포진 모바일 진단을 위한 딥러닝 양상을 지식을 활용한 피부 질환 분류 기법 |
| 9 | Convolutional neural networks are superior to dermatologists in melanoma image classification | Tobias J. Brinker, Titus J. Hekler, Jochen S. Enk, et al. | 딥러닝(CNN) 모델이 피부암(특히 흑색종) 이미지 분류에서 피부과 전문의보다 높은 정확도를 보이는지 검증 |
| 10 | 보건의료 데이터 활용 가이드라인의 내용과 한계에 관한 연구 | 이기호, 김계현 | 의료 데이터 활용의 법적 문제점을 분석하고, 이를 해결하기 위한 법제도 개선 방안을 제시함 |
| 11 | Comparison of fine-tuning strategies for transfer learning in medical image classification | Ana Davila*, Jacinto Colanb, Yasuhisa Hasegawaa | 사전 학습된 CNN 모델의 Fine-tuning 전략 비교를 통해, LP-FT 방식이 다양한 의료 영상 분류 과제에서 가장 효과적이었음을 입증한 연구 |
| 12 | Augmented intelligence dermatology: Deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders | Seong Soo Han, Dae Hyun Ban, Sunyoung Park, et al. | AI 기반 피부질환 진단 앱(DNN model)의 정확도 평가 |
| 13 | An Integrated Deep Learning Model with EfficientNet and ResNet for Accurate Multi-Class Skin Disease Classification | Madallah Alruwaili, Mahmood Mohamed | EfficientNet과 ResNet 기반의 융합 딥러닝 모델을 활용해 피부 질환을 높은 정확도로 다중 클래스 분류하는 방법을 제안 |
| 14 | Melanoma Skin Cancer Classification Using Neural Networks | Rukhsar Sabir, Tahir Mehmood | 딥러닝 모델인 CNN, ResNet-18, EfficientNet-B0를 활용하여 피부 병변 이미지를 양성 및 악성으로 분류하는 연구 |

Thank you

"대상포진은 초기 증상이 애매하여 진단이 어렵고 조기 치료가 늦어지면 심각한 신경통과 높은 의료비 부담을 유발하므로 빠르고 정확한 진단이 매우 중요합니다. 본 연구는 육안 진단의 한계를 극복하고 피부 영상의 미세한 특징을 자동으로 학습하는 딥러닝 기반 CNN 모델을 개발하여 대상포진을 조기에 예측하고자 합니다. 이 모델은 환자의 고통을 줄이고 의료 시스템 부담을 덜어주며, 모바일 기반의 조기 진단 시스템으로 확장될 것으로 기대됩니다"
