

# Lecture 10 - Contents

An overview of the parts in the multimodal medical AI lecture.

## Part 1

Vision-Language Models

## Part 2

Additional Modalities

## Part 3

Fusion Strategies

## Hands-on

Multimodal Hands-on

This outline is for guidance. Navigate the slides with the left/right arrow keys.



Lecture 10:

# **Multimodal Medical AI: Integrating Text, Images, and Signals**

**Ho-min Park**

[homin.park@ghent.ac.kr](mailto:homin.park@ghent.ac.kr)

[powersimmani@gmail.com](mailto:powersimmani@gmail.com)

## Multimodal Medical Learning Overview

멀티모달 학습은 텍스트, 영상, 신호 등 다양한 데이터 유형을 통합하여 더 정확하고 포괄적인 의료 AI 모델을 구축



### Text

임상 노트, 진단 보고서, EHR 데이터



### Images

X-ray, CT, MRI, 병리 슬라이드



### Signals

ECG, EEG, 생체 신호, 파형



### Genomics

DNA/RNA 시퀀스, 유전자 변이



### Video

수술 영상, 내시경, 동작 분석



### Audio

폐음, 심음, 청진 데이터

### 통합 진단

### 시너지 효과

단일 모달리티 대비 향상된 성능

### 임상 실용성

실제 의료 환경과 유사한 통합 접근

다양한 데이터 소스로부터 종합적인  
진단

Part 1:

# **Medical Vision-Language Models**

## Medical Image Encoders

### CNN (Convolutional Neural Networks)

전통적이지만 강력한 영상 특징 추출 방법

- 공간적 계층 구조 학습
- 로컬 패턴 인식에 강점
- ResNet, DenseNet, EfficientNet
- 의료 영상의 텍스처, 경계 검출

### Vision Transformer (ViT)

패치 기반 어텐션으로 글로벌 관계 학습

- Self-attention 메커니즘
- 긴 범위 의존성 포착
- 대규모 데이터셋에서 우수
- 병변 간 관계, 전체 맥락 이해

### 의료 특화 인코더

의료 도메인에 사전 학습된 모델

- MedCLIP, BiomedCLIP
- 대규모 의료 영상-텍스트 쌍 학습
- 도메인 지식 내재화
- 적은 데이터로 높은 성능

### Hybrid Architectures

CNN과 Transformer의 결합

- CNN: 로컬 특징 + ViT: 글로벌 관계
- CoAtNet, MaxViT
- 계산 효율성과 성능 균형
- 다양한 스케일의 병변 검출

정규화  
(Normalization)

윈도잉  
(Windowing)

리샘플링  
(Resampling)

증강  
(Augmentation)

### 특징 추출

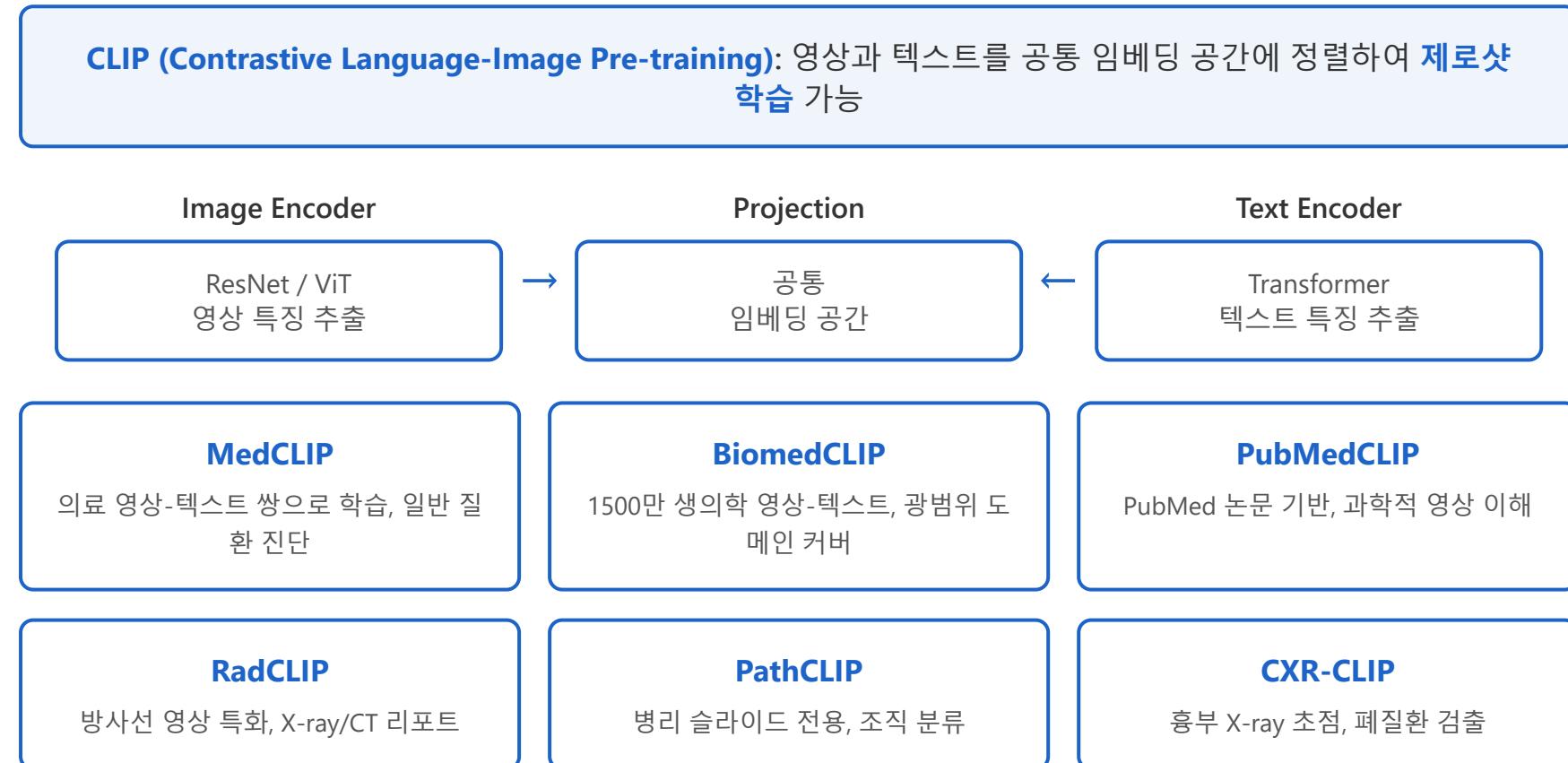
고차원 영상을 저차원 벡터 표현으로 변환하여 의미 있는  
임상 특징 포착

### 다운스트림 태스크

추출된 특징을 분류, 검출, 세분화, 리포트 생성 등에 활용

## CLIP Medical Variants: Contrastive Learning

**CLIP (Contrastive Language-Image Pre-training):** 영상과 텍스트를 공통 임베딩 공간에 정렬하여 **제로샷 학습 가능**



### 대조 학습 (Contrastive Learning)

- 매칭 영상-텍스트 쌍: 높은 유사도
- 비매칭 쌍: 낮은 유사도

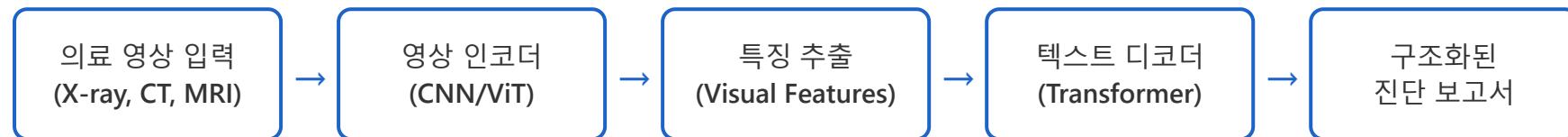
### 영상-텍스트 정렬 효과

- 제로샷 분류 가능
- 텍스트 쿼리로 영상 검색

- InfoNCE loss로 구분 학습

- 멀티모달 표현 학습

## Radiology Report Generation: 자동 의료 리포트 작성



### Chest X-ray

폐렴, 결핵, 흉수 등 흉부 질환 자동 기술

### CT Scan

장기별 병변 위치, 크기, 특성 상세 설명

### MRI

뇌, 척추, 관절 이상 소견 문장 생성

### Mammography

유방 석회화, 종괴 판독 보고서

### Ultrasound

초음파 소견 자동 문서화

### Multi-View

여러 각도 영상 통합 분석

### 구조화된 보고서 섹션

Findings  
(소견)

Impression  
(판단)

Comparison  
(비교)

Recommendation  
(권고)

## Pathology Slide Analysis: WSI 분석

**WSI (Whole Slide Imaging):** 기가픽셀급 병리 슬라이드를 **패치 단위로 분할**하여 조직 분류 및 병변 검출

### 1. 슬라이드 스캐닝

- 고해상도 디지털 스캔 (20x, 40x)
- 수십 GB 크기의 이미지
- Multi-resolution pyramid 구조

### 2. 패치 추출

- 256x256 또는 512x512 패치 분할
- 조직 영역 검출 (배경 제거)
- Stride 조절로 오버랩 설정

### 3. 패치별 분석

- CNN/ViT로 각 패치 특징 추출
- 정상/암/염증 등 조직 타입 분류
- Multiple Instance Learning (MIL)

### 4. 슬라이드 레벨 통합

- 패치 예측 집계 (평균, 최대, 어텐션)
- 전체 슬라이드 진단 (암 등급, 예후)
- 히트맵 시각화

암 검출 및 등급 분류

림프절 전이 판정

Ki-67 증식 지수 측정

## Cross-Modal Alignment: 모달 간 정렬

공통 임베딩 공간에서 서로 다른 모달리티를 의미적으로 가까운 위치에 배치하여 상호 이해 및 검색 가능

### Projection Layers

각 모달리티를 동일 차원의 공통 공간으로 투영하는 선형/비선형 변환

- 선형 레이어로 차원 매칭
- MLP로 비선형 변환
- L2 정규화로 단위 구상에 배치

### Contrastive Loss

매칭 쌍은 가깝게, 비매칭 쌍은 멀게 학습하여 정렬 품질 향상

- InfoNCE Loss 사용
- Temperature scaling 적용
- Hard negative mining

### Triplet Loss

Anchor, Positive, Negative 샘플로 상대적 거리 학습

- 동일 환자 데이터: Positive
- 다른 환자 데이터: Negative
- 마진 기반 거리 최적화

### Cross-Attention

Transformer 기반 어텐션으로 모달 간 상호작용 강화

- Query-Key-Value 메커니즘
- 모달 간 정보 교환
- 동적 가중치 학습

Zero-shot 분류

텍스트→영상 검색

영상→텍스트 검색

## 대조 학습 (Contrastive Learning)

양성 쌍(같은 환자의 다른 뷰, 동일 진단)과 음성 쌍을 구별 학습하여 의료 표현 품질 향상

### 양성 쌍 (Positive Pairs)

의미적으로 유사한 데이터 쌍

- 같은 환자의 다른 시점 영상
- 동일 질환의 다른 모달리티
- 영상-리포트 매칭 쌍
- 데이터 증강된 동일 샘플

### 음성 쌍 (Negative Pairs)

의미적으로 다른 데이터 쌍

- 다른 환자의 영상
- 다른 질환 카테고리
- 비매칭 영상-리포트
- 배치 내 다른 샘플들

### InfoNCE Loss

대조 학습의 핵심 손실 함수

- 양성 쌍: 높은 유사도 (코사인)
- 음성 쌍: 낮은 유사도
- Temperature 파라미터 조정
- Large batch size 필요

### 의료 도메인 적용

의료 특화 대조 학습 전략

- 해부학적 일관성 활용
- 시간적 연속성 보존
- 임상 메타데이터 통합
- 레이블 없이 사전학습

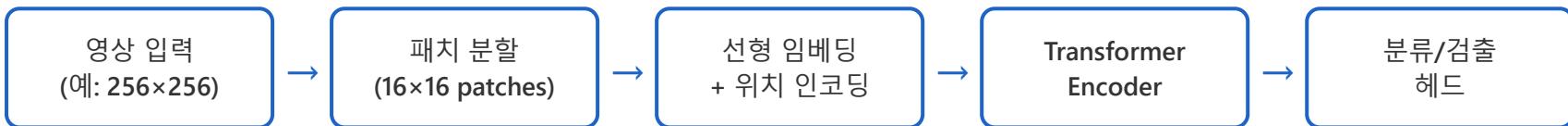
Self-supervised  
사전학습

Few-shot  
학습 가능

Transfer  
Learning 향상

## Vision Transformer (ViT) for Medical Imaging

패치 임베딩과 위치 인코딩으로 의료 영상을 처리, Self-attention으로 전역 관계 포착



### Self-Attention 메커니즘

전체 영상의 장거리 의존성 학습

- 모든 패치 간 관계 계산
- Query, Key, Value 변환
- 병변 간 공간적 관계 포착
- Multi-head attention

### 의료 영상 적용

ViT의 의료 도메인 특화

- 고해상도 영상 처리
- 가변 크기 입력 지원
- 3D 볼륨 확장 (3D-ViT)
- Medical-specific 사전학습

### 장점

ViT의 강점

- 전역 컨텍스트 이해
- 확장 가능한 구조
- 대규모 사전학습 효과
- 다양한 다운스트림 태스크

### 의료 ViT 모델

대표적인 의료 특화 ViT

- MedViT: 의료 영상 사전학습
- TransUNet: 세분화 특화
- Swin Transformer: 계층적
- CoTr: CT/MRI 복원

Part 2:

# **Beyond Images - Signals and Sequences**

## ECG Waveform Processing: 심전도 분석

12-lead ECG 신호로 부정맥 검출, QRS 복합체, ST 분절 변화 분석

### ECG 신호 특성

심전도 파형의 주요 구성 요소

- P wave: 심방 탈분극
- QRS complex: 심실 탈분극
- T wave: 심실 재분극
- 12-lead: 다각도 심장 활동

### 전처리 파이프라인

ECG 신호 정제 및 표준화

- 노이즈 제거 (필터링)
- 베이스라인 보정
- 정규화 (0-1 scaling)
- R-peak 검출

### 딥러닝 모델

ECG 분석을 위한 신경망

- 1D CNN: 시간축 컨볼루션
- LSTM/GRU: 시계열 패턴
- Residual Network: 깊은 구조
- Transformer: 장거리 의존성

### 멀티모달 통합

ECG와 다른 데이터 융합

- ECG + 흉부 X-ray
- ECG + 임상 노트
- ECG + 심초음파
- 시간적 정렬 필요

부정맥 분류  
(AF, VT, VF)

심근경색 검출  
(STEMI, NSTEMI)

급성 심장 이벤트  
예측

## Audio: Lung Sounds Analysis

청진음을 스펙트로그램으로 변환 후 CNN으로 폐렴, 천식, COPD 등 호흡기 질환 분류

### 폐음 유형

주요 청진음 패턴

- 정상 폐음: Vesicular sounds
- 천명음 (Wheezes): 기도 협착
- 수포음 (Crackles): 수분/염증
- 마찰음 (Rubs): 흉막 염증

### 오디오 전처리

음향 신호 변환 및 증강

- 스펙트로그램 생성 (STFT)
- Mel-frequency 변환
- 노이즈 제거 (심장음 등)
- 시간 분할 (윈도우)

### 딥러닝 아키텍처

음향 분류 모델

- 2D CNN: 스펙트로그램 처리
- ResNet: 깊은 특징 추출
- RNN: 시간적 패턴
- Audio Transformer: 어텐션

### 멀티모달 융합

청진음과 다른 데이터 통합

- 폐음 + 흉부 X-ray
- 폐음 + 폐기능 검사
- 폐음 + 임상 증상
- 공간-음향 정렬

폐렴 조기 발견

COPD 모니터링

천식 악화 예측

## Genomic Sequence Integration

DNA/RNA 시퀀스를 Transformer 또는 CNN으로 처리, 변이 검출 및 표현형 예측과 통합

### 게놈 데이터 유형

의료 AI에 활용되는 유전 정보

- DNA 시퀀스 (A, T, G, C)
- RNA 발현 프로파일
- 변이 (SNP, CNV, Indel)
- 후성유전학 (메틸화)

### 시퀀스 인코딩

염기서열을 수치 표현으로 변환

- One-hot encoding
- K-mer embedding
- Positional encoding
- DNA-BERT, DNA2Vec

### 딥러닝 모델

게놈 분석을 위한 신경망

- 1D CNN: 모티프 검출
- Transformer: 장거리 패턴
- Graph NN: 유전자 네트워크
- VAE: 잠재 표현 학습

### 멀티오믹스 통합

게놈 데이터와 다른 모달리티

- 게놈 + 영상 (Radiophenomics)
- 게놈 + 병리 (Pathogenomics)
- 게놈 + 임상 기록
- Multi-view learning

암 유전자 변이  
검출

개인 맞춤 치료  
예측

약물 반응성  
평가

## Time Series: Vital Signs Monitoring

ICU 생체 신호(HR, BP, SpO2)를 RNN/LSTM/Transformer로 시계열 모델링하여 악화 예측

### 주요 생체 신호

모니터링되는 바이탈 사인

- 심박수 (HR): 분당 박동수
- 혈압 (BP): 수축기/이완기
- 산소포화도 (SpO2): 혈중 산소
- 호흡수 (RR): 분당 호흡

### 시계열 특성

생체 신호의 시간적 패턴

- 불규칙 샘플링 간격
- 결측값 (센서 오류)
- 트렌드와 주기성
- 급격한 변화 (이벤트)

### 딥러닝 모델

시계열 예측 아키텍처

- LSTM: 장기 의존성
- GRU: 경량 순환 네트워크
- Temporal CNN: 확장 컨볼루션
- Transformer: 어텐션 기반

### 멀티모달 통합

생체 신호와 다른 데이터

- 바이탈 + 실험실 검사
- 바이탈 + 약물 투여 기록
- 바이탈 + 임상 노트
- 시간 정렬 필수

패혈증 조기 경보

심정지 위험 예측

ICU 재원기간 추정

## 3D Medical Imaging

3D CNN으로 CT/MRI 볼륨 데이터 처리, 공간적 맥락 활용하여 병변 검출 및 세분화

### 3D 영상 모달리티

볼륨 데이터 소스

- CT: 연속 슬라이스 (axial)
- MRI: 다중 시퀀스 (T1, T2, FLAIR)
- PET/SPECT: 기능 영상
- 초음파: 3D 재구성

### 3D CNN 아키텍처

볼륨 처리를 위한 네트워크

- 3D 컨볼루션 커널 ( $3 \times 3 \times 3$ )
- 3D U-Net: 세분화
- V-Net: 의료 특화
- nnU-Net: 자동 구성

### 공간적 특징

3D의 장점

- Z축 맥락 정보 활용
- 볼륨 측정 정확도
- 3D 구조 보존
- 인접 슬라이스 연관성

### 멀티모달 3D 통합

다중 볼륨 데이터 융합

- CT + PET: 해부학+기능
- MRI multi-sequence 융합
- 3D 영상 + 임상 데이터
- 공간적 정합 (Registration)

폐 결절 검출

뇌종양 세분화

장기 용적 측정

## Video: Surgical Analysis

수술 비디오를 프레임 단위 분석, 단계 인식, 도구 추적, 합병증 예측에 활용

### 수술 비디오 특성

비디오 데이터의 고유 속성

- 고해상도 (HD, 4K)
- 긴 지속 시간 (1-6시간)
- 복잡한 장면 변화
- 다양한 수술 도구

### 비디오 처리 기법

시공간 정보 추출

- 프레임 샘플링 (FPS 조정)
- 2D CNN + RNN/LSTM
- 3D CNN: 시공간 컨볼루션
- Optical flow: 움직임

### 수술 단계 인식

수술 워크플로우 분석

- Phase recognition (7-10 단계)
- Action segmentation
- Temporal CNN (TCN)
- 실시간 피드백

### 멀티모달 통합

비디오와 다른 데이터

- 비디오 + 키네틱 데이터
- 비디오 + 수술 기록
- 비디오 + 환자 정보
- 시간 동기화 중요

도구 추적 및  
사용 분석

수술 기술  
객관 평가

합병증 위험  
조기 감지

Part 3:

# Multimodal Fusion Techniques

## Early vs Late Fusion: 융합 시점 전략

Early Fusion: 특징 추출 전 결합 / Late Fusion: 각 모달 독립 처리 후 통합

### Early Fusion

입력 레벨 결합

- 원시 데이터 또는 초기 특징 결합
- 단일 모델로 공동 학습
- 모달 간 저수준 상호작용
- 계산 효율적

장점: 강한 통합, 빠른 추론

단점: 차원의 저주, 모달 불균형

### Intermediate Fusion

중간 레벨 결합

- 각 모달 부분 처리 후 결합
- 중간 표현 수준 통합
- 계층적 특징 결합
- 유연한 아키텍처

장점: 균형잡힌 접근

단점: 설계 복잡도

### Late Fusion

결정 레벨 결합

- 각 모달 독립 완전 처리
- 고수준 예측 결합
- 양상별 방식
- 모달별 전문화

장점: 모달 독립성, 확장성

단점: 약한 통합

### 의료 적용 예시

**Early:** CT + PET 픽셀 레벨 융합

**Intermediate:** X-ray + 텍스트 특징 결합

**Late:** 영상/유전자/임상 예측 통합

### 선택 기준

**Early:** 모달 크기/형식 유사, 밀접 관련

**Late:** 이질적 모달, 누락 가능성

**Hybrid:** 복잡한 멀티모달 시스템

## Attention-Based Fusion

Cross-attention과 Self-attention으로 모달 간 중요도를 동적으로 조절하여 융합

### Self-Attention

단일 모달 내부 관계

- Query, Key, Value 변환
- 모든 위치 간 유사도 계산
- 가중합으로 표현 업데이트
- 장거리 의존성 포착

### Cross-Attention

모달 간 상호작용

- 한 모달의 Query
- 다른 모달의 Key, Value
- 모달 간 정보 교환
- 선택적 정보 추출

### Multi-Head Attention

다양한 관점에서 융합

- 여러 어텐션 헤드 병렬
- 각 헤드는 다른 특징
- 결과 연결(Concatenate)
- 표현력 향상

### 의료 적용

Attention 기반 의료 융합

- 영상→텍스트 어텐션
- 다중 시점 영상 융합
- 시계열 + 정적 데이터
- 해석 가능한 가중치

영상-리포트  
정렬

다중 시퀀스  
MRI 융합

임상 데이터  
통합

## Graph Neural Networks (GNN) for Multimodal Fusion

의료 지식 그래프와 GNN으로 모달 간 관계를 모델링하여 융합 품질 향상

### 의료 지식 그래프

구조화된 의료 지식 표현

- 노드: 질환, 증상, 검사, 치료
- 엣지: 관계 (유발, 진단, 치료)
- 속성: 확률, 빈도, 심각도
- 온톨로지 통합 (SNOMED, ICD)

### GNN 아키텍처

그래프 기반 학습

- Graph Convolution (GCN)
- Graph Attention (GAT)
- Message Passing
- Graph Pooling

### 멀티모달 그래프 구성

이종 데이터를 그래프로 통합

- 각 모달리티 = 노드 유형
- 모달 간 관계 = 엣지
- 환자별 서브그래프
- 동적 그래프 업데이트

### 의료 적용

GNN 기반 융합 사례

- 질환 예측: 증상 그래프
- 약물 상호작용 모델링
- 환자 유사도 그래프
- 지식 기반 추론

다중 질환  
동반이환 예측

개인화  
치료 추천

임상 경로  
최적화

## Hierarchical Integration: 계층적 통합

다단계 융합으로 저수준 특징부터 고수준 의미까지 계층적으로 통합

### Level 1 저수준 특징 융합 (Low-level)

원시 데이터 또는 초기 특징: 텍스처, 엣지, 픽셀 강도, 주파수 성분

### Level 2 중간 수준 특징 융합 (Mid-level)

부분적 의미 표현: 해부학적 구조, 객체 부분, 패턴 조합

### Level 3 고수준 의미 융합 (High-level)

추상적 개념: 질환 카테고리, 진단 라벨, 임상적 해석

### Level 4 의사결정 레벨 (Decision-level)

최종 예측: 진단, 치료 계획, 예후, 위험도 점수

다중 스케일  
정보 활용

점진적  
추상화

해석 가능성  
향상

## Missing Modality Handling: 누락 모달리티 처리

일부 모달리티 부재 시 대체 전략 (zero-padding, imputation, robust fusion) 적용

### Zero-padding / Masking

간단한 기본 전략

- 누락 모달: 0으로 채우기
- 마스크 토큰 사용
- 구현 간단, 빠른 추론
- 성능 저하 가능

### Imputation (대치)

누락 데이터 예측하여 채우기

- 평균/중앙값 대치
- K-NN 기반 대치
- 생성 모델 (VAE, GAN)
- Cross-modal 예측

### Robust Fusion

누락에 강건한 융합 설계

- Dropout 기반 학습
- Late fusion 활용
- Ensemble 방식
- Gating mechanism

### Knowledge Distillation

완전 모델에서 지식 전이

- Teacher: 모든 모달 사용
- Student: 부분 모달 처리
- Soft label 학습
- 누락 시나리오 대비

훈련 시  
무작위 드롭

모달별  
신뢰도 가중치

동적  
아키텍처 조정

## Clinical Decision Support: 임상 의사결정 지원

멀티모달 CDSS로 통합 진단, 치료 추천, 예후 예측 제공

### 진단 지원

다중 데이터 기반 진단

- 영상 + 실험실 검사
- 증상 + 과거력
- 바이탈 + 신체검사
- 확률적 진단 제시

### 치료 추천

개인화 치료 계획

- 약물 선택 및 용량
- 수술 vs 보존적 치료
- 부작용 위험 평가
- 가이드라인 준수

### 예후 예측

장기 결과 추정

- 생존률 예측
- 합병증 위험도
- 재발 가능성
- 기능 회복 예상

### 실시간 모니터링

지속적 환자 관찰

- 악화 조기 경보
- 자동 알림 시스템
- 트렌드 분석
- 중재 시점 제안

응급실  
트리아지

ICU  
관리

만성질환  
관리

## Performance Benchmarks: 성능 평가

멀티모달 모델을 단일 모달 대비 벤치마크, 평가 메트릭(accuracy, AUC, F1) 비교

### Accuracy

전체 정확도  
올바른 예측 비율

### AUC-ROC

분류 성능  
민감도-특이도 균형

### F1-Score

정밀도-재현율  
조화 평균

### Sensitivity

민감도  
True Positive Rate

### Specificity

특이도  
True Negative Rate

### Dice/IoU

세분화 정확도  
영역 겹침 측정

## 폐렴 진단 성능 비교 예시

Model	Accuracy	AUC	F1
X-ray Only	84.2%	0.87	0.82
Clinical Only	79.5%	0.83	0.78
X-ray + Clinical (Multimodal)	<b>91.3%</b>	<b>0.94</b>	<b>0.90</b>

# Interpretability Challenges in Multimodal AI

멀티모달 모델의 복잡성으로 인한 해석 어려움, 어텐션 시각화 및 SHAP 활용

## 해석 가능성 필요성

의료 AI의 신뢰성 확보

- 임상의 의사결정 지원
- 규제 요구사항 (FDA, CE)
- 환자 설명 책임
- 모델 디버깅 및 개선

## 멀티모달 해석 어려움

복잡성의 원인

- 다층 융합 구조
- 모달 간 상호작용
- 비선형 변환
- 고차원 표현 공간

## Attention 시각화

어텐션 가중치 분석

- Cross-attention 맵
- 모달별 기여도
- 히트맵 오버레이
- 시간적 어텐션 패턴

## XAI 기법

설명 가능 AI 방법론

- SHAP: Shapley values
- LIME: 로컬 근사
- Grad-CAM: 그래디언트 기반
- Integrated Gradients

모달 중요도  
점수

특징 기여도  
분석

대조 사례  
제시

## Case Study: Emergency Room Multimodal Triage

응급실에서 텍스트(주소), 영상(X-ray), 생체신호(ECG, BP)를 통합하여 트리아지 우선순위 결정

환자 도착 & 초기 평가 → 다중 모달 데이터 수집 → 멀티모달 AI 분석 → 위험도 점수 산출 → 트리아지 우선순위



텍스트  
주 호소: "가슴 통증"  
증상 기간: 2시간  
과거력: 고혈압



영상  
흉부 X-ray  
심비대 소견  
폐부종 의심



생체신호  
HR: 110 bpm  
BP: 160/95  
ECG: ST 상승

### 통합 분석 결과

예측 진단: 급성 심근경색 가능성 높음 (87% 신뢰도)

위험도 점수: 9/10 (긴급)

권장 조치: 즉시 심장내과 협진, 응급 심도자술 준비

처리 시간: 3초 (실시간 지원)

## Hands-on: Multimodal Model 구현 실습

PyTorch로 영상+텍스트 융합 모델 구현, 데이터 로더, 학습 루프 작성

### Step 1

#### 데이터 준비

영상과 텍스트 쌍 데이터셋 로드

```
dataset = MultimodalDataset(  
    image_dir='./xrays',  
    text_file='reports.csv'  
)
```

### Step 2

#### 인코더 정의

영상/텍스트 각각 인코더 구축

```
img_encoder = ResNet50()  
text_encoder = BERT()
```

### Step 3

#### 융합 레이어

특징 결합 및 프로젝션

```
fusion = nn.Linear(  
    img_dim + text_dim,  
    hidden_dim  
)
```

### Step 4

#### 학습 루프

손실 함수 및 최적화

```
loss = contrastive_loss(  
    img_emb, text_emb  
)  
optimizer.step()
```

GitHub:  
Multimodal-Medical

Colab Notebook:  
실습 예제

Dataset:  
MIMIC-CXR

## Future Directions: 멀티모달 의료 AI의 미래

신규 모달리티 통합, 대규모 사전학습, 실시간 임상 적용, 설명가능성 향상

### Foundation Models

대규모 멀티모달 사전학습

- 의료 전용 GPT-4 수준 모델
- 다양한 모달리티 동시 학습
- Few-shot 적응 능력
- 범용 의료 AI 플랫폼

### 신규 모달리티

확장되는 데이터 유형

- 웨어러블 센서 통합
- 마이크로바이옴 데이터
- 소셜미디어 건강 정보
- 환경 요인 데이터

### 실시간 시스템

즉각적인 임상 지원

- Edge AI 배포
- 지연 시간 최소화
- 연속 학습 (Continual)
- 모바일 헬스케어

### 윤리와 규제

책임 있는 AI 개발

- 편향 완화 기술
- 프라이버시 보호 학습
- 국제 표준화
- 임상 검증 프로토콜

개인 맞춤  
정밀 의료

예방적  
건강 관리

의료 접근성  
향상

Thank You!

Multimodal Medical AI: 다양한 데이터 통합으로  
더 정확하고 포괄적인 의료 인공지능 구현

Ho-min Park

[homin.park@ghent.ac.kr](mailto:homin.park@ghent.ac.kr)

[powersimmani@gmail.com](mailto:powersimmani@gmail.com)