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