Hands-on: Multimodal Model Implementation Practice

Implement image+text fusion model with PyTorch, data loader, and training loop

Step 1

Data Preparation

Load image and text paired dataset

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

Step 2

Define Encoders

Build separate encoders for image/text

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

Step 3

Fusion Layer

Feature concatenation and projection

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

Step 4

Training Loop

Loss function and optimization

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

GitHub:
Multimodal-Medical

Colab Notebook:
Practice Examples

Dataset:
MIMIC-CXR