**Recurrence Prediction in Colorectal Cancer Using Vision Transformers and Clinical Data**

**1. Cohort Selection: TCGA-COAD (Colon Adenocarcinoma)**

* **Source**: [TCGA via GDC](https://portal.gdc.cancer.gov/) and [TCIA](https://www.cancerimagingarchive.net/)
* **Data Available**:
  + Whole-slide images (H&E stained)
  + Clinical metadata (age, stage, treatment, recurrence status)
  + Optional: gene expression (open-access processed data)

**2. Base Model: Vision Transformer (ViT)**

**Option A: Pretrained ViT**

* Model: google/vit-base-patch16-224
* Source: Hugging Face
* Use: [CLS] token or pooled patch embeddings

**Option B: Self-Supervised ViT**

* Train on TCGA-COAD patches using:
  + **DINO**
  + **MAE**
  + **SimCLR** or **BYOL**

**3. Pipeline Overview**

**Step 1: WSI Preprocessing**

* Tile WSIs into 224×224 patches
* Filter background
* Normalize stain (e.g., Macenko method)

**Step 2: Feature Extraction**

* Feed patches into ViT
* Aggregate features per slide (mean pooling, attention pooling, MIL)

**Step 3: Clinical Fusion**

* Combine image features with clinical variables
* Normalize and encode categorical data

**Step 4: Recurrence Modeling**

* Use DeepSurv, Cox PH, or transformer-based survival model
* Predict disease-free interval or recurrence status

**4. Evaluation Metrics**

* **C-index**: Concordance between predicted and actual recurrence order
* **Hazard Ratio (HR)**: Risk stratification
* **Kaplan-Meier curves**: Survival stratification

**Comparison:**

* Baseline: Clinical-only model
* Your model: ViT + clinical fusion
* Optional: Compare with gene expression models

**5. Stretch Goals**

* Add multi-modal fusion with omics data
* Train ViT with self-supervised learning on COAD slides
* Apply to other TCGA cohorts (e.g., LUAD, PRAD)

**Tools & Libraries**

* **OpenSlide**, **PathML**, **Slideflow**: WSI handling
* **PyTorch**, **timm**, **transformers**: ViT models
* **lifelines**, **scikit-survival**: Survival analysis
* **MONAI**: Medical imaging workflows

Once you've pasted this into your editor, you can export it as a PDF with clean formatting. If you want help designing a title page or adding diagrams, I can help with that too.