

A. DATA COLLECTION TASKS

Collect urine quantitative data (numeric dataset).

Collect stool image dataset (real + GAN samples).

Gather disease labels (ground truth) for each sample.

B. DATA PREPROCESSING TASKS

Urine Numeric Data

Handle missing values (KNN imputation).

Remove outliers (IQR or Z-score).

Normalize/scale numeric features (MinMax or StandardScaler).

Encode categorical fields (if any).

Perform class balancing (SMOTE / undersampling) if needed.

Stool Image Data

Remove low-quality GAN samples (using FID, SSIM, or manual review).

Resize all images (e.g., 224×224).

Normalize images (ImageNet mean/std).

Apply augmentation:

- Rotation
- Flip
- Contrast/brightness jitter
- Random crop

Split dataset into train/val/test sets (stratified).

C. MODEL DESIGN TASKS

Urine Model (MLP)

Build an MLP network for numeric features.

Add dense layers with ReLU + dropout.

Generate a urine feature vector (e.g., 64-D).

Stool Image Model (CNN)

Select CNN backbone (EfficientNet, ResNet, MobileNet, ViT).

Load pretrained ImageNet weights.

Freeze early layers (transfer learning).

Add global average pooling.

Generate stool image feature vector (256–1024-D).

Fusion Model

Concatenate urine vector + stool vector.

Build fusion dense layers.

Add dropout + batch normalization.

Add final classification layer (sigmoid/softmax depending on task).

D. TRAINING TASKS

Define loss function:

- Binary cross entropy (binary disease)
- Categorical cross entropy (multi-class)
- Multi-label BCE (multiple diseases)

Select optimizer (Adam/AdamW).

Set learning rates:

- 1e-3 for initial training
- 1e-5 for fine-tuning

Train MLP + fusion head (CNN frozen).

Unfreeze last CNN layers for fine-tuning.

Train full fusion model end-to-end.

Monitor training using validation loss.

Implement early stopping.

Save best model weights.

E. EVALUATION TASKS

Predict on test dataset.

Compute:

- **Accuracy**
- **Precision**
- **Recall (Sensitivity)**
- **F1-score**
- **Specificity**
- **ROC-AUC**

Create:

- Confusion matrix
- ROC curve
- Precision–recall curve

Evaluate:

- Urine-only model
- Stool-only model
- Fusion model

Compare performance and identify best model.

F. RESULTS INTERPRETATION TASKS

Identify important urine features (SHAP, feature importance).

Identify important stool regions (Grad-CAM).

###Urine Model Tasks

Preprocess urine numeric dataset

Scale & clean features

Train urine MLP/XGBoost

Evaluate urine model metrics

###Stool Image Model Tasks

Preprocess stool images

Train CNN/ViT model

Evaluate stool model metrics

Fusion Tasks (No pairing needed)

###Create ensemble:

Weighted average

Soft voting

Stacking ML model (meta-learner)

Combine predictions

Compare fusion vs individual models