(a) Tokenization

MIMIC-III & MIMIC-IV

patients.csv
admissions.csv
diagnoses_icd.csv

Sort → Groupby → Tokenize → Mapping

Token Details

START_RECORD | Pneumonia | Fluid and electrolyte disorders | Fluid and electrolyte disorders | ICD9-40301: Hypertensive chronic kidney disease | ICD9-486: Pneumonia, organism unspecified | END_VISIT | ICD9-28521: Anemia in chronic kidney disease | ICD9-33829: Other chronic pain | END_RECORD | PADDING

Hypertension with complications | Chronic kidney disease | Complications of surgical/medical care | END_LABEL | ICD9-5855: Chronic kidney disease, Stage V | ICD9-4254: Other primary cardiomyopathies | ICD9-58381: Nephritis and nephropathy | ICD9-45829: Other iatrogenic hypotension | END_VISIT | PADDING

(b) Hierarchical & Semantic Embds

Hierarchical levels

Chapter: A00-B99 ··· N00-N99
Section: ··· N17-N19
Category: ··· N18
Subcategory_1: N18.3, N18.5
Subcategory_2: N18.32

ICD9-5855 to ICD10-N18.5

GNN → Hierarchical Embd

Text description lookup → "Chronic kidney disease, stage 5" → Pre-trained Clinical LMs → Semantic Embd

(c) Model Structure

Consistency Regularization: Hierarchical Head, Semantic Head
Next code prediction: Code Head

Hidden States
Transformer Decoder Blocks x 6
Positional Embd + Input Embd

Hierarchical Embd | Semantic Embd | Code Token