LREC-COLING 2024

# Auxiliary Knowledge-Induced Learning for Automatic Multi-Label Medical Document Classification


Xindi Wang[1,2], Robert E. Mercer[1], and Frank Rudzicz[2,3,4]

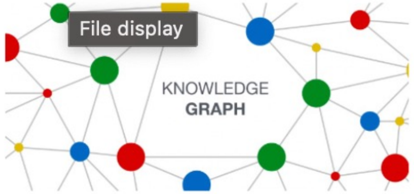


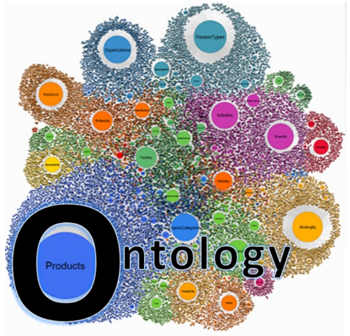
[1]University of Western Ontario, Canada
[2]Vector Institute for Artificial Intelligence, Canada
[3]Dalhousie University, Canada
[4]University of Toronto, Canada

# What is Knowledge in NLP?

Knowledge is any external information absent from the input but helpful for generating the output

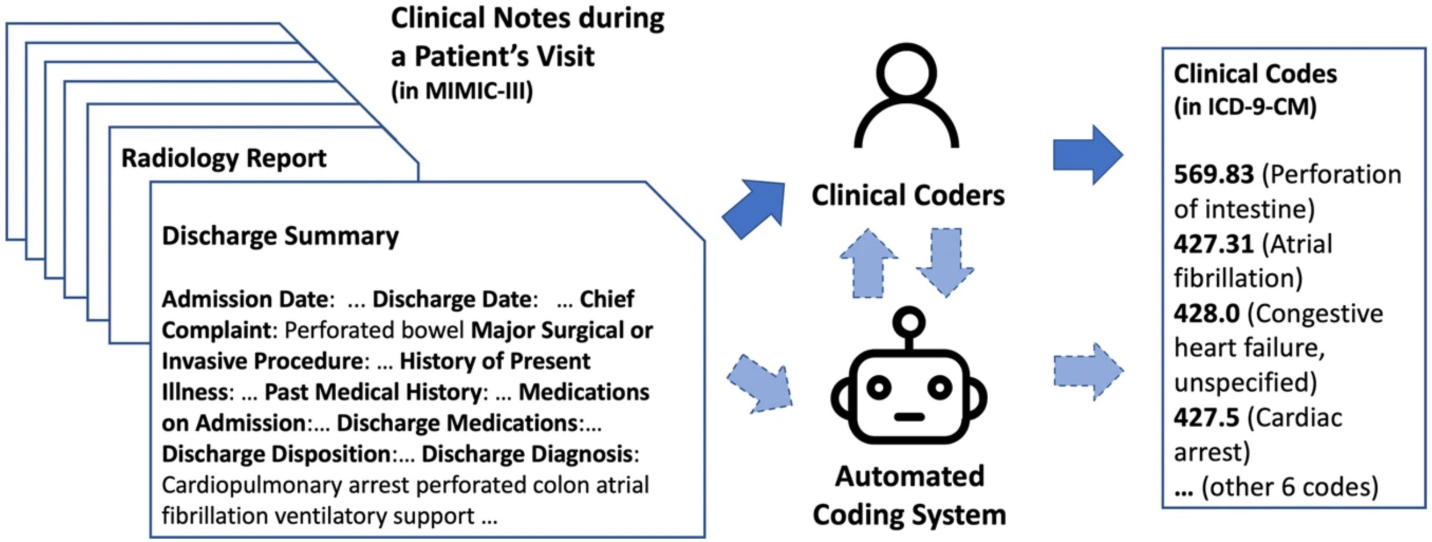
- Structured:
  - Knowledge graph: a meta-representation of knowledge, common sense, entities, relations
  - Ontology: a set of concepts, knowledge within a domain and the relationships that hold between them
- Unstructured:
  - Text data: knowledge from data without a predefined format, e.g., document, emails
  - Video / images




# International Classification of Disease (ICD) Classification

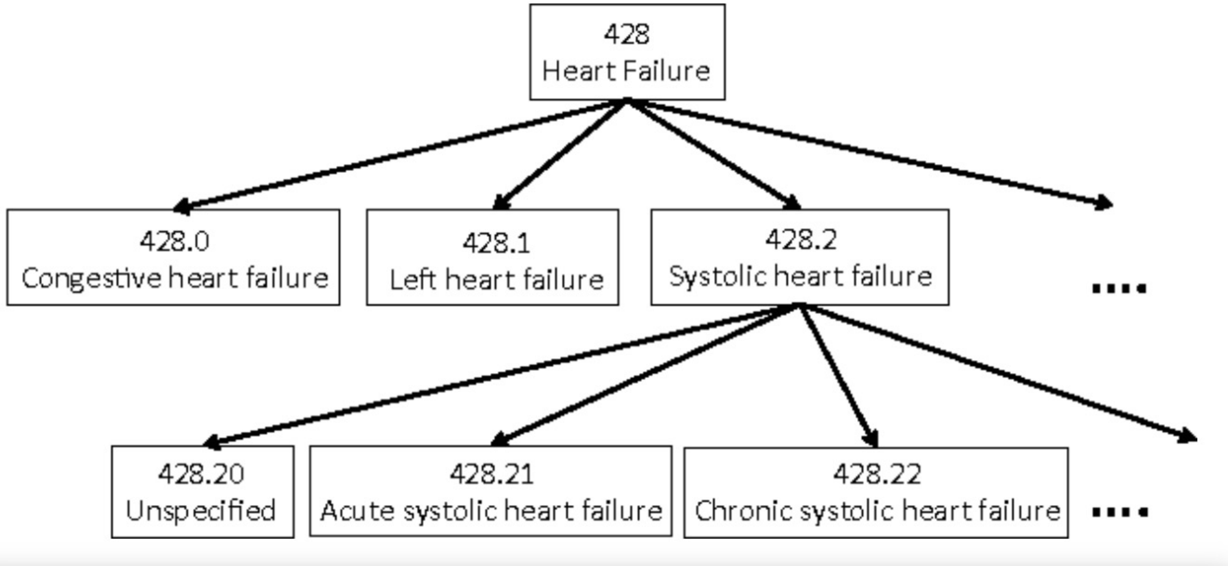
- Problem Definition – Extreme multi-label classification:
  - the automatic assignment of the most relevant subset of labels (ICD codes) to a clinical note
  - differently from the classic multi-label problem: the label set size is large, usually in the order of ten




Dong, H., Falis, M., Whiteley, W. et al. Automated clinical coding: what, why, and where we are?. npj Digit. Med. 5, 159 (2022).

# International Classification of Disease (ICD) Codes

- International Classification of Disease (ICD):
  - maintained by the World Health Organization (WHO)
  - organizes hierarchically (shallow, 3 levels)
  - Revised periodically, and is currently in its 11th revision
  - ICD-11 has 17,000 codes in total
  - Used worldwide for morbidity and mortality statistics, reimbursement systems and automated decision support in healthcare
  - annotated by the coders

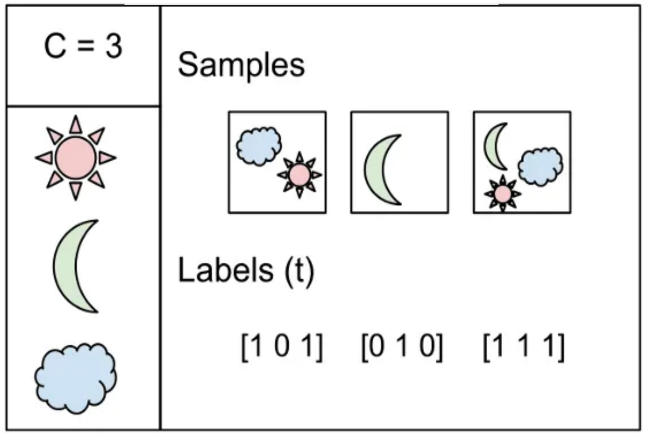


# International Classification of Disease (ICD) Classification

Problem Formulation:

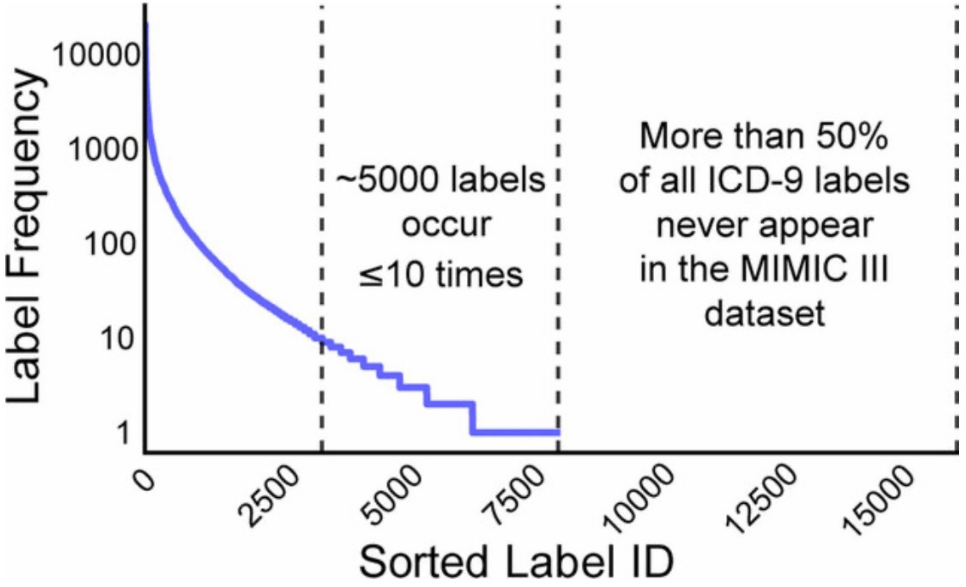
- Training data $\{(x_i, y_i)\}_1^n, x_i \in \chi \subseteq \mathbb{R}^D, y_i \in \{0,1\}^L$
- Learning a mapping $f\colon\ \chi \to \{0,1\}^L$
- Each document $x_i$ is associated with a set of relevant labels $y_i$

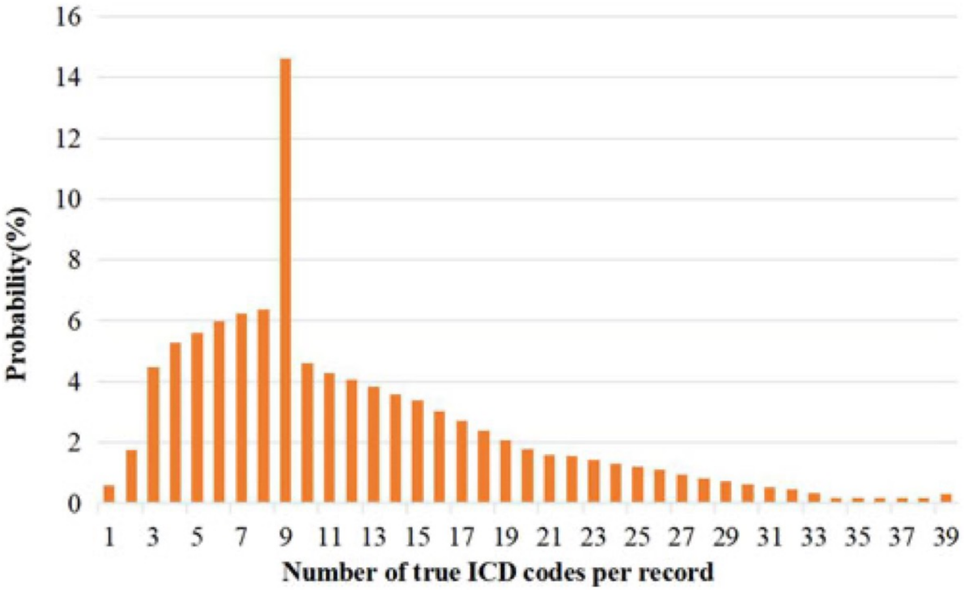
Multi-Label



# ICD Classification -- Challenges

- The number of labels is large, and they are varying occurrence frequencies




- The number of labels assigned to each document varies

# Motivations -- External Knowledge in ICD Classification

**Discharge Summary (text):**

**...Patient is a 83 year-old man with a history of hypertension, prostate ca (per son this has been stable, untreated for several months), and dementia who presented with an upper gastrointestinal bleed and was noted at his NSG home to have malaise, poor PO intake and low grade fevers (no note of fever in paperwork) for past 2d … For his upper GI bleeding, the patient received IV fluids and was transfused with [**Year/Month/Day **]. He received intravenous pantoprazole therapy. Patient underwent an EGD that showed edematous mucosa and thickened folds concerning for malignancy with no evidence of active. H. pylori testing was positive and he was started on lansoprazole amoxicillin, and clarithromycin. He had biopsies taken during endoscopy… He had dark maroon colored stool… Abnormal mucosa in the stomach (biopsy)…The patient did not have any active issues regarding his dementia. He was continued on his Namenda and Aricept during this admission.**

**ICD codes (label):**

- 401.9 Unspecified essential hypertension
- 151.9 Malignant neoplasm of stomach, unspecified site
- 285.1 Acute posthemorrhagic anemia
- 331.0 Alzheimer's disease
- 185 Malignant neoplasm of prostate
- 294.10 Dementia in conditions classified elsewhere without behavioral disturbance
- 45.16 Esophagogastroduodenoscopy [EGD] with closed biopsy
- 041.86 Helicobacter pylori [H. pylori]

...

**Auxiliary Knowledge:**

CPT Codes:

- 99231 Hospital inpatient services

DRG Codes:

- 2402 Digestive Malignancy

…

Prescriptions:

- Bicalutamide
- Pantoprazole
- Midazolam
- Namenda
- Donepezil
- Atorvastatin
- Potassium Chloride

…

Western University · Canada | Vector Institute | Dalhousie University | University of Toronto

# ICD Classification – Knowledge Representation

- Auxiliary knowledge:
  - Diagnosis related group (DRG) Codes: facilitate inpatient billing and reimbursement
  - Current Procedural Terminology (CPT) Codes: describe clinical procedures and services in healthcare
  - Drug prescriptions

- Knowledge representation:
  - Conditional probabilities

$$P(L_i \mid M_j) = \frac{C_{L_i \cap M_j}}{C_{M_j}},$$

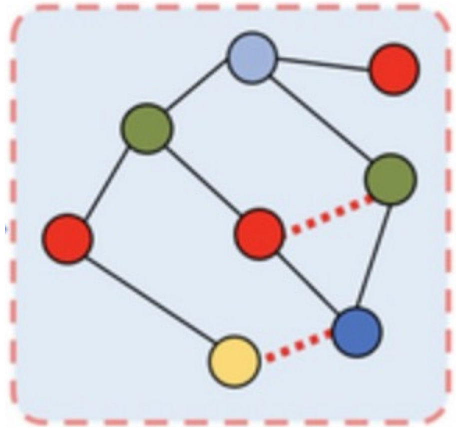
  - where $C_{L_i \cap M_j}$ is the number of co-occurrences of $L_i$ and $M_j$, and $C_{M_j}$ is the number of occurrence of $M_j$ in the training set.

# ICD Classification – Knowledge Representation

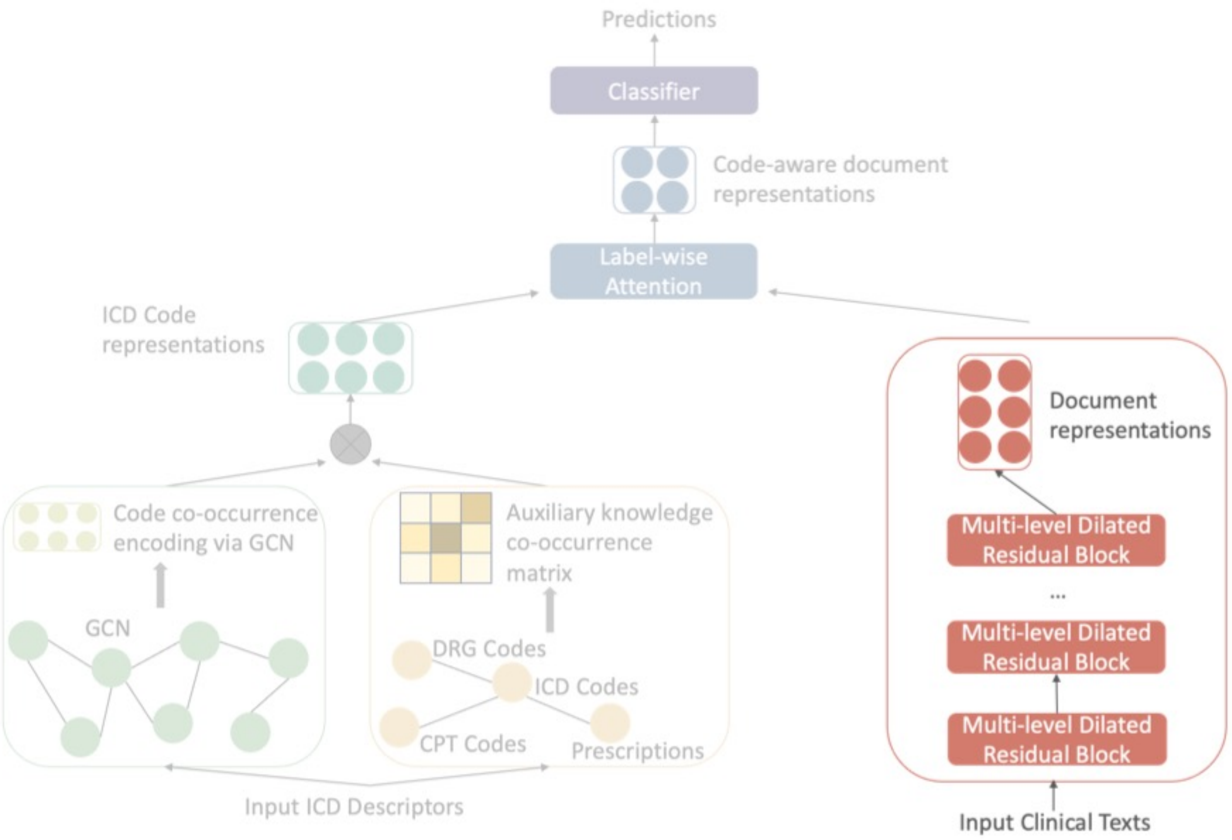
- Label co-occurrence information:
  - Use graph operation to encode the label co-occurrence information
  - Nodes: labels
    - Node feature: average of word embeddings in the label description

$$v_i = \frac{1}{Z}\sum_{j=1}^{Z} w_j, i = 1, 2, ..., L,$$

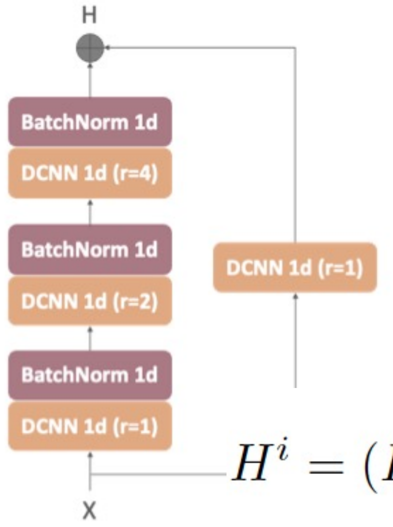
  - Edges: label co-occurrence information



# Proposed Model – Document Representation

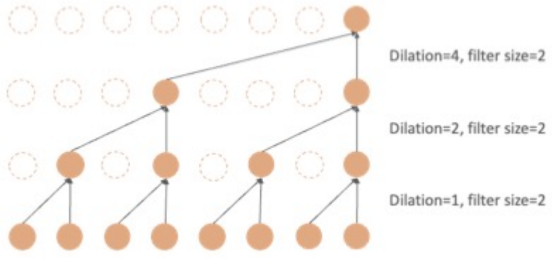



(a) The architecture of our model




$$H^i = (E *_r f)(s) = \sum_{j=0}^{K-1} f(i) \cdot E_{s-r \cdot j}$$

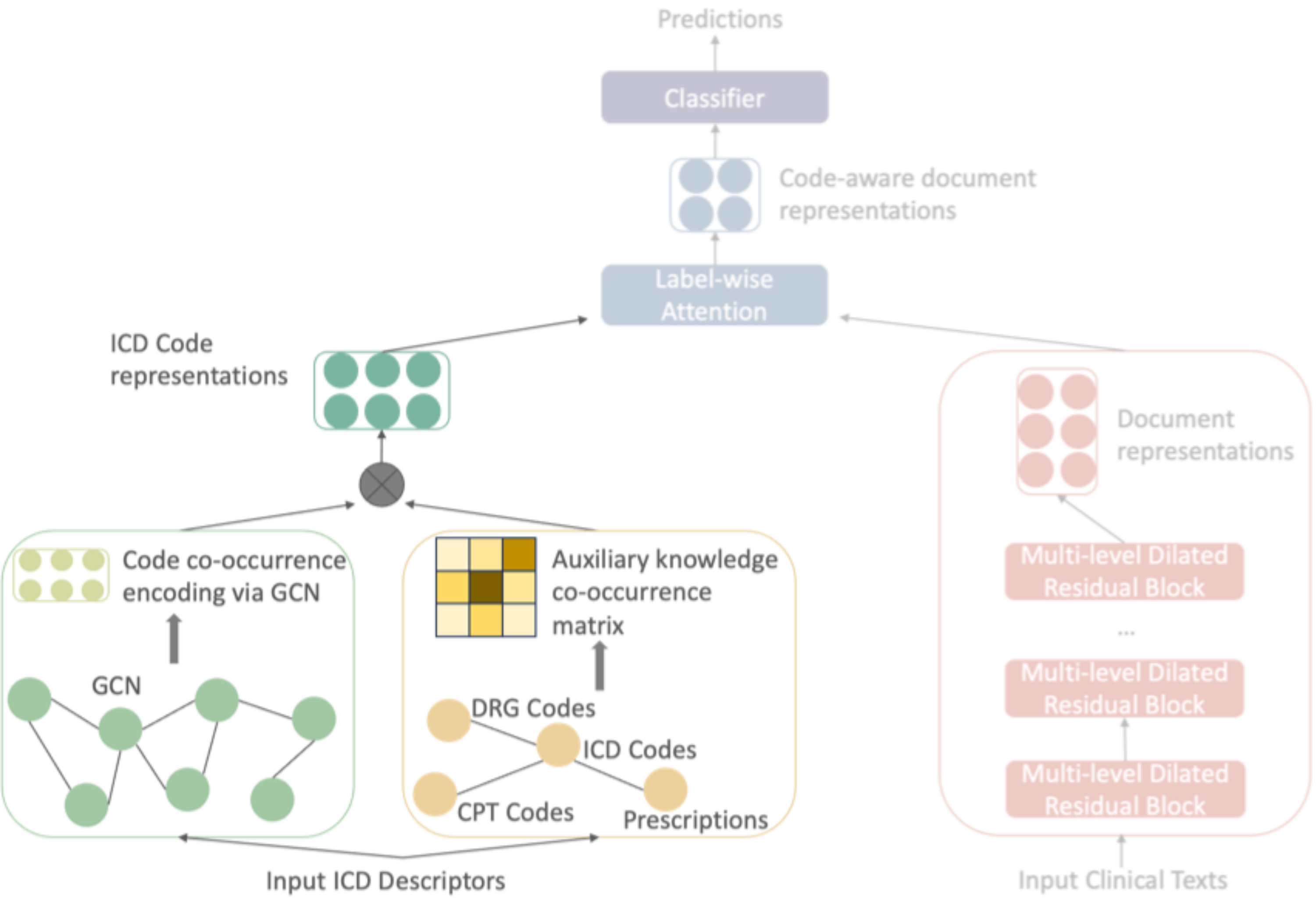
(b) Multi-level Dilated Residual Block




(c) Example of multi-level dilated convolution

# Proposed Model – Label Representation

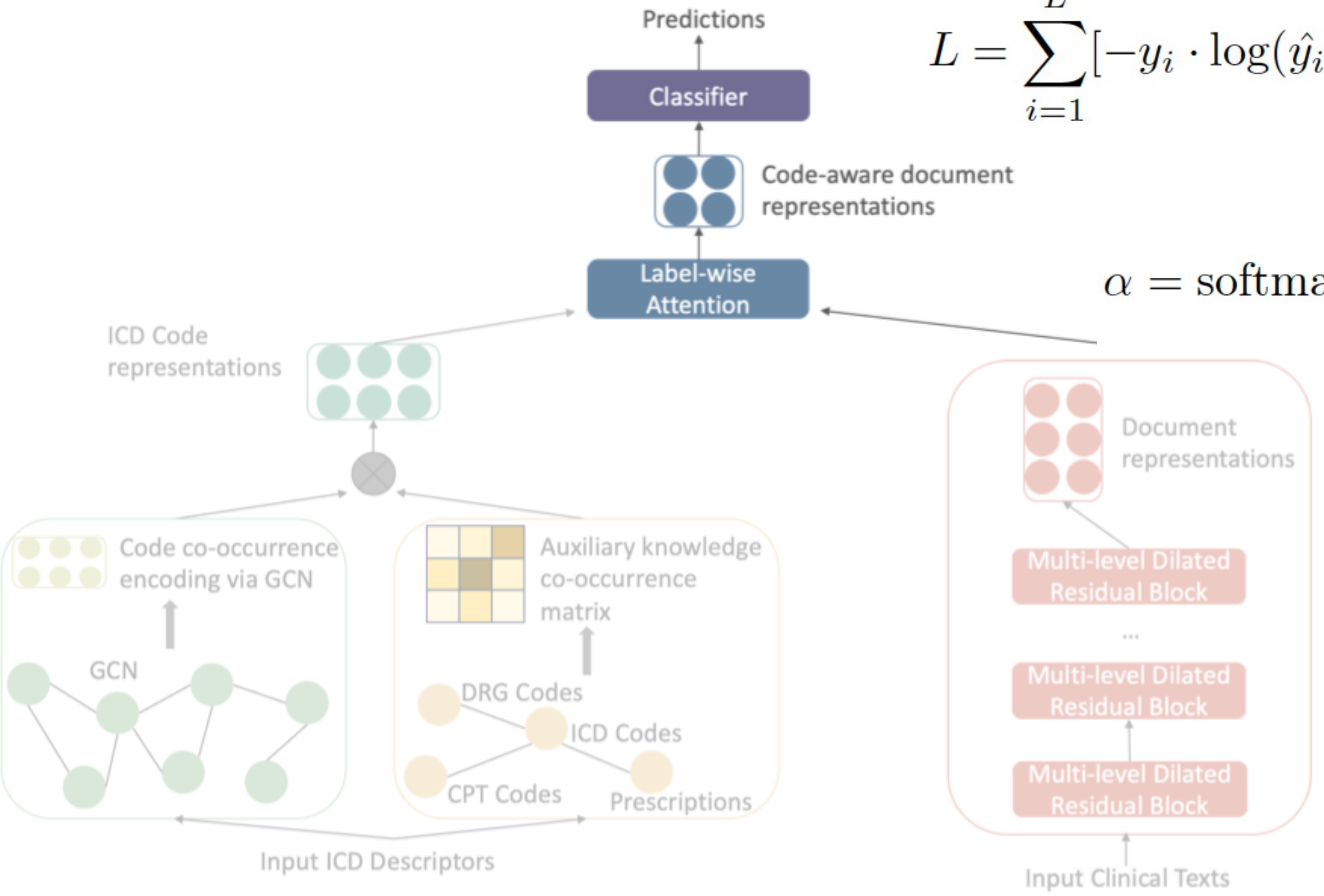
$$h^{l+1} = \sigma(A \cdot h^l \cdot W^l),$$



# Proposed Model – Label-wise Attention and Classifier

$$L = \sum_{i=1}^{L} [-y_i \cdot \log(\hat{y}_i) - (1 - y_i) \cdot \log(1 - \hat{y}_i)],$$

$$\alpha = \mathrm{softmax}(D \cdot H_{masked})$$




(a) The architecture of our model

# Results

| Models | MIMIC-III-full | | | | | | MIMIC-III-top 50 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | AUC | | F1 | | P@K | | AUC | | F1 | | P@5 |
| | Macro | Micro | Macro | Micro | P@8 | P@15 | Macro | Micro | Macro | Micro | |
| CAML (Mullenbach et al., 2018) | 0.895 | 0.986 | 0.088 | 0.539 | 0.709 | 0.561 | 0.875 | 0.909 | 0.532 | 0.614 | 0.609 |
| DR-CAML (Mullenbach et al., 2018) | 0.897 | 0.985 | 0.086 | 0.529 | 0.690 | 0.548 | 0.884 | 0.916 | 0.576 | 0.633 | 0.618 |
| MultiResCNN (Li and Yu, 2020) | 0.910 | 0.986 | 0.085 | 0.552 | 0.734 | 0.584 | 0.899 | 0.928 | 0.606 | 0.670 | 0.641 |
| LAAT (Vu et al., 2020) | 0.919 | 0.988 | 0.099 | 0.575 | 0.738 | 0.591 | 0.925 | 0.946 | 0.666 | 0.715 | 0.675 |
| Joint-LAAT (Vu et al., 2020) | 0.921 | 0.988 | 0.107 | 0.575 | 0.735 | 0.590 | 0.925 | 0.946 | 0.661 | 0.716 | 0.671 |
| EffectiveCAN (Liu et al., 2021) | 0.915 | 0.988 | 0.106 | 0.589 | 0.758 | 0.606 | 0.915 | 0.938 | 0.644 | 0.702 | 0.656 |
| MSMN (Yuan et al., 2022) | **0.950** | 0.992 | 0.103 | 0.584 | 0.752 | 0.599 | 0.928 | 0.947 | 0.683 | 0.725 | 0.680 |
| KEPTLongformer (Yang et al., 2022) | - | - | **0.118** | 0.599 | 0.771 | 0.615 | 0.926 | 0.947 | 0.689 | 0.728 | 0.672 |
| Ours | 0.948 ± 0.022 | **0.994** ± 0.013 | 0.112 ± 0.027 | **0.605** ± 0.021 | **0.784** ± 0.022 | **0.637** ± 0.011 | **0.928** ± 0.014 | **0.950** ± 0.018 | **0.692** ± 0.016 | **0.734** ± 0.012 | **0.683** ± 0.023 |

Table 2: Comparison to previous methods across three main evaluation metrics MIMIC-III dataset. We report the mean ± standard deviation of each result. Bold: best scores in each column.

# Takeaways

- We propose a novel end-to-end model integrating document features and label co-occurrence features for ICD classification.
- We use a novel auxiliary knowledge-induced learning to handle the large universe of candidate ICD codes and employ GCN in extracting label correlations.
- Experimental results demonstrate that our proposed model outperforms the baseline models on most of the metrics.

# Thank you!