# CoRelation: Boosting Automatic ICD Coding Through Contextualized Code Relation Learning

**Junyu Luo**[1], **Xiaochen Wang**[1], **Jiaqi Wang**[1], **Aofei Chang**[1]
**Yaqing Wang**[2], **Fenglong Ma**[1]
[1]Pennsylvania State University, [2]Google
[1]{junyu, xcwang, jqwang, aofei, fenglong}@psu.edu, [2]yaqingwang@google.com

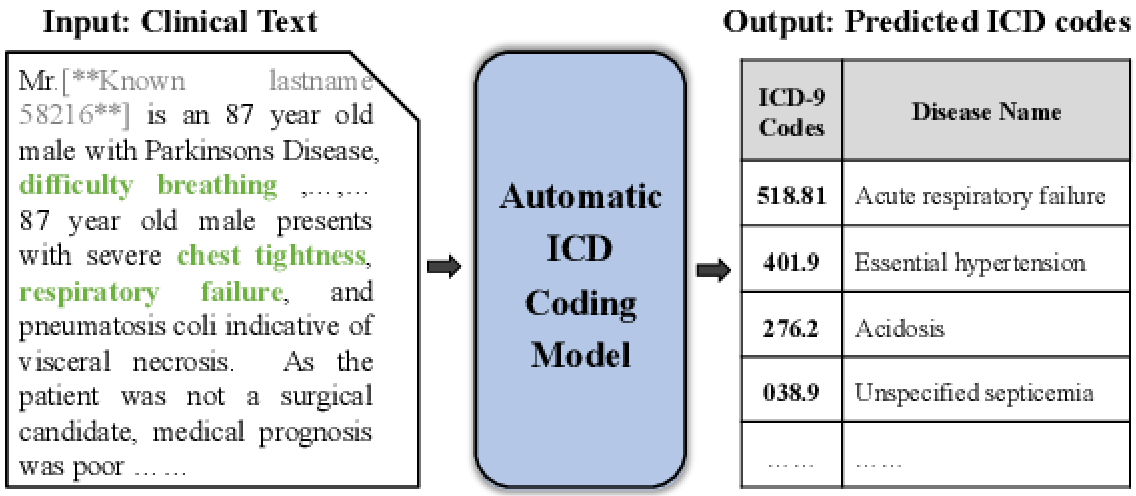
# Background

- ICD coding aims to automatically assign International Classification of Diseases (ICD) codes from unstructured clinical notes or discharge summaries

- However, manual code assignments are labor-intensive and prone to errors.

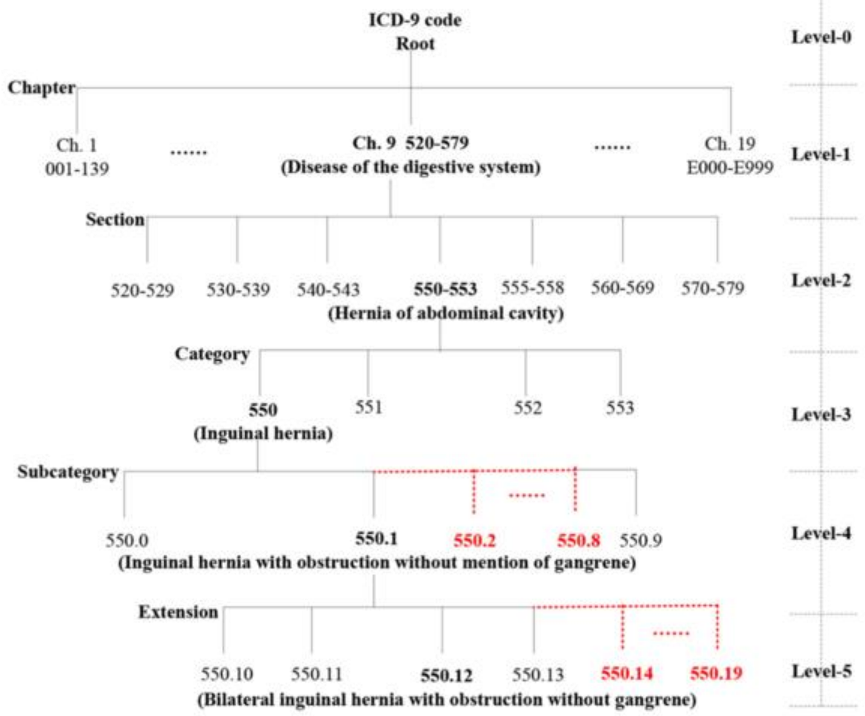
# Task

- Giving a clinical textual note, the models need to predict the correct codes as a multi-label classification task.



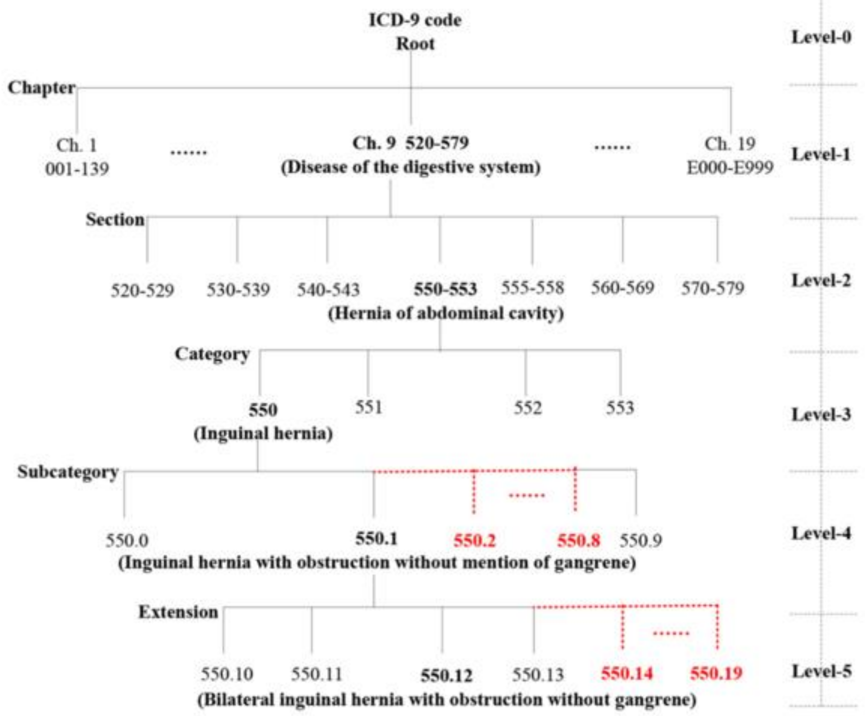
# Limitations of the Previous Methods

- In ICD code prediction one of the important aspects is that the ICD codes naturally have an ontology structure.

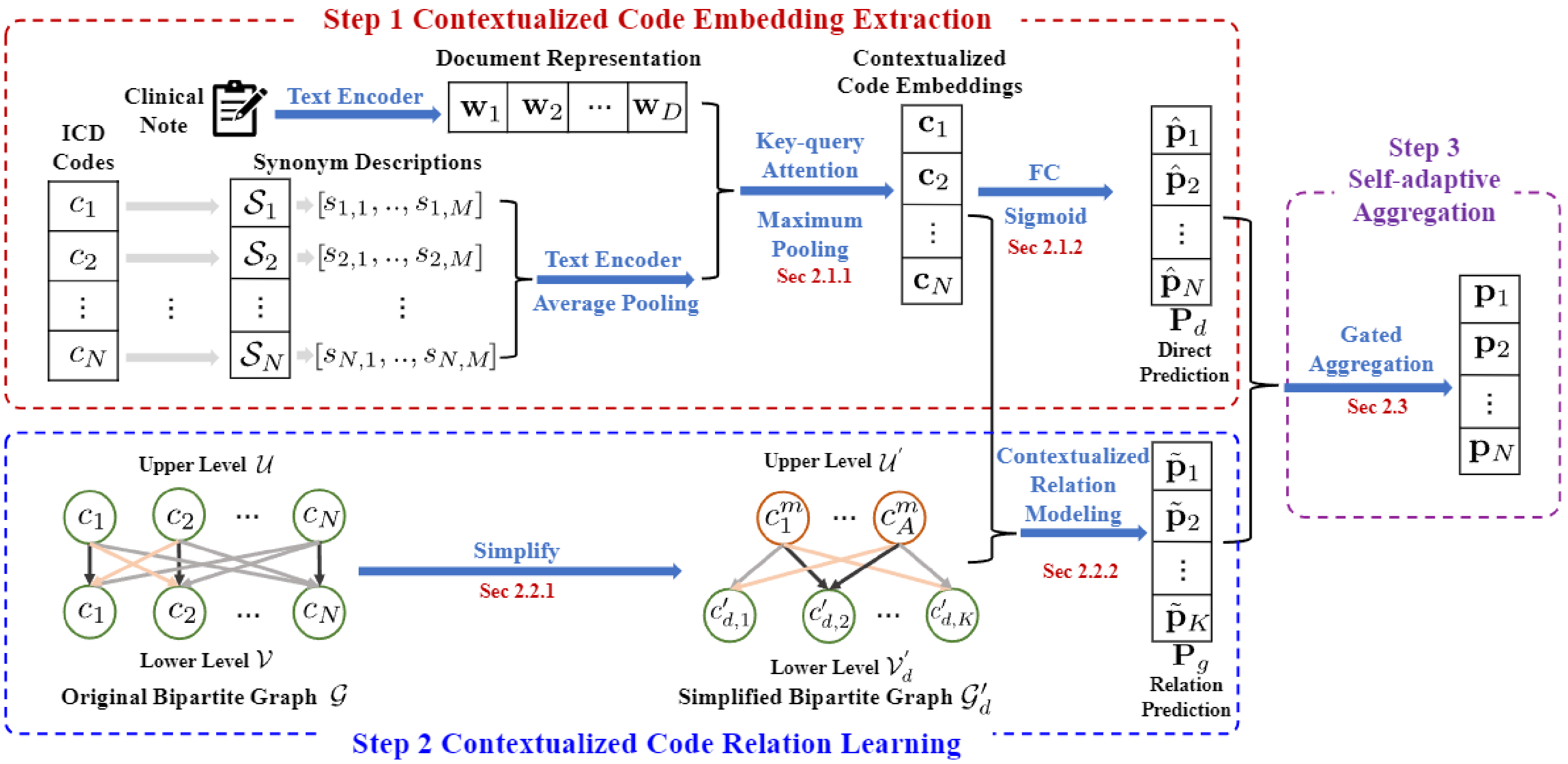


# Limitations of the Previous Methods

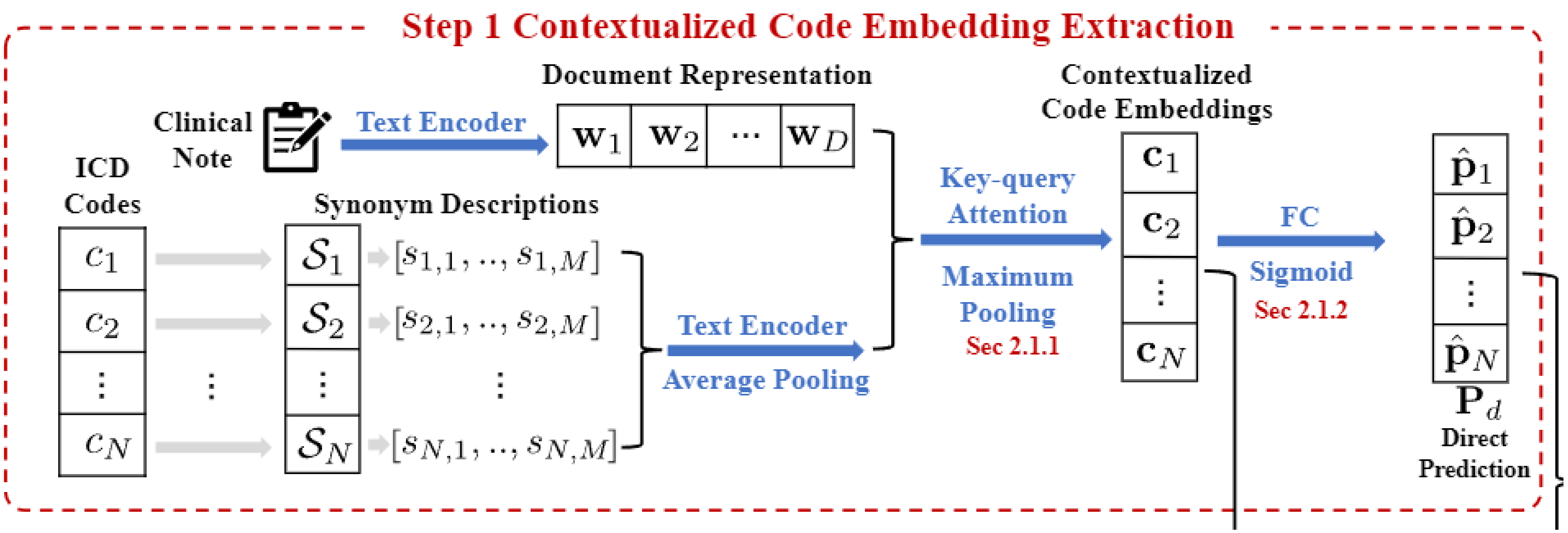
- Insufficiently Modeling Relations Among ICD Codes.
- Ignoring the Importance of Context.



# Methodology



# Step 1 Code Embedding Extraction



# Embedding Extraction

- We encode the document and codes into the embeddings.
- Follow standard key-query attention to extract the contextualized code embedding.

$$[\mathbf{w}_1, \cdots, \mathbf{w}_D] = \mathrm{TextEncoder}(d),$$

$$[\mathbf{ws}_1^{i,j}, \cdots, \mathbf{ws}_L^{i,j}] = \mathrm{TextEncoder}(s_{i,j}).$$

$$\mathbf{s}_{i,j} = \mathrm{Pool}([\mathbf{ws}_1^{i,j}, \cdots, \mathbf{ws}_L^{i,j}]).$$

$$\mathbf{c}_{i,j} = \mathrm{KeyQueryAttention}(\mathbf{s}_{i,j}, [\mathbf{w}_1, \cdots, \mathbf{w}_D]),$$

$$\mathbf{c}_i = \mathrm{Pool}([\mathbf{c}_{i,1}, \cdots, \mathbf{c}_{i,M}]).$$

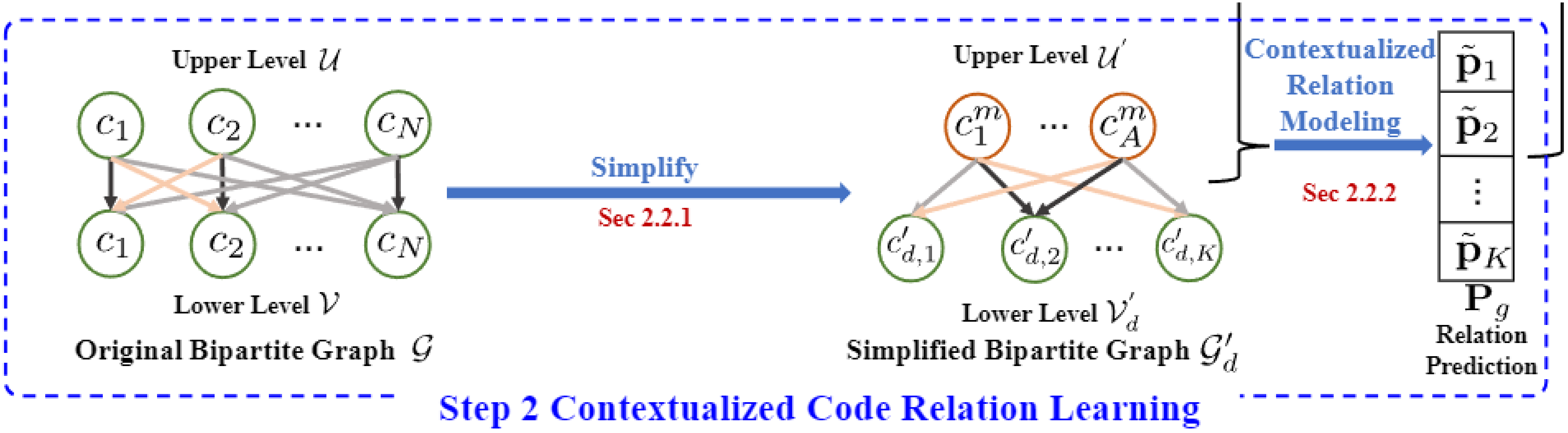
# Direct Code Prediction

- Initial results are predicted based on the extracted code embeddings.

$$\boldsymbol{\alpha}_i = \mathrm{FC}_\alpha(\mathrm{Pool}([\mathbf{s}_{i,1}, \cdots, \mathbf{s}_{i,M}])).$$

$$\hat{\mathbf{p}}_i = \sigma(\boldsymbol{\alpha}_i \cdot \mathbf{c}_i).$$

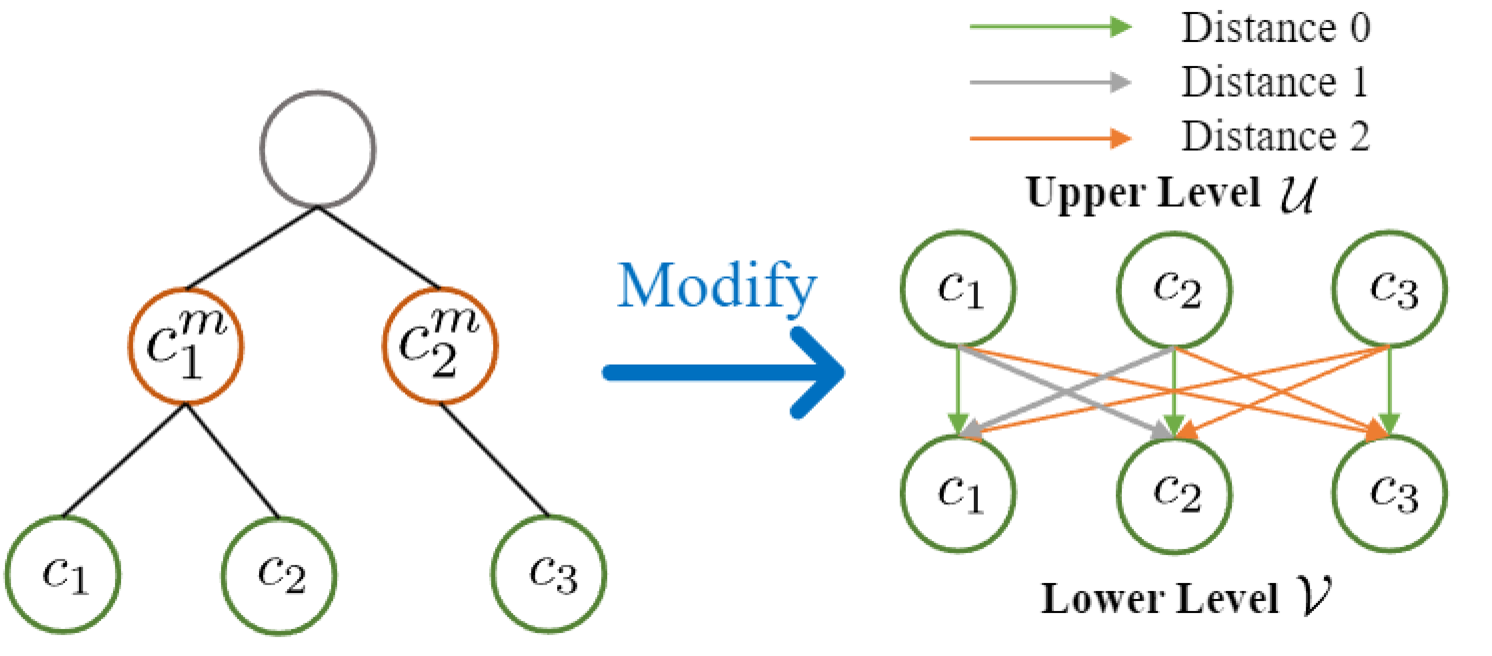
# Step 2 Code Relation Learning on Graph



# Building Flexible Bipartite Graph

250.03 (Diabetes mellitus without mention of complication, type I [juvenile type], uncontrolled)

250 (Diabetes mellitus)



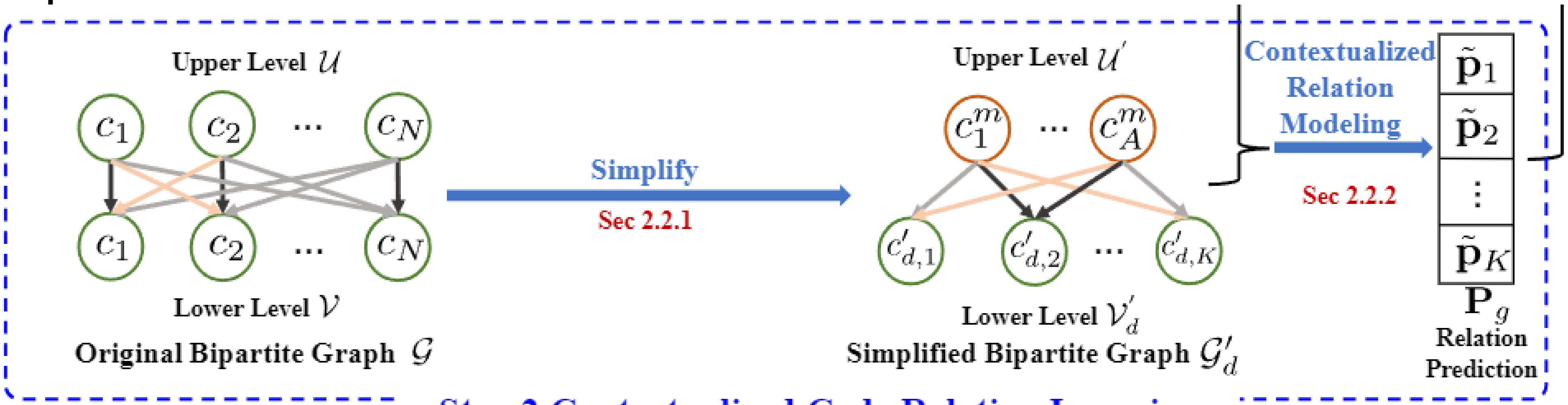

(a) Original ICD tree graph

(b) Flexible bipartite graph $\mathcal{G}$

**Major Code Substitution**

**Top-K Code Selection**

# Graph-Based Prediction




**Step 2 Contextualized Code Relation Learning**

$$\{\mathbf{U}^*, \mathbf{V}_d^*, \mathbf{E}_d^*\} = \mathrm{GraphTransformer}(\mathcal{G}'_d).$$

$$\boldsymbol{\beta}_i = \mathrm{FC}_\beta(\mathrm{Pool}([\mathbf{s}_{i,1}, \cdots, \mathbf{s}_{i,M}])).$$

$$\mathbf{V}_d^* = [\tilde{\mathbf{c}}_1, \cdots, \tilde{\mathbf{c}}_K]$$

$$\tilde{\mathbf{p}}_i = \sigma(\boldsymbol{\beta}_i \cdot \tilde{\mathbf{c}}_i),$$

# Step 3 Generating Final Prediction

- The two results are aggregated based on the wise product of code embedding and contextualized code embedding.

$$\gamma_i = \sigma(\mathrm{FC}_\gamma(\boldsymbol{\alpha}_i \odot \mathbf{c}_i)).$$

$$\mathbf{p}_i = (1 - \gamma_i)\hat{\mathbf{p}}_i + \gamma_i \tilde{\mathbf{p}}_i.$$

# Experiment

Table 5.3: Results on the MIMIC-III-50 test set.

| Category | Method | AUC | | F1 | | Pre | |
|---|---|---|---|---|---|---|---|
| | | Macro | Micro | Macro | Micro | P@5 | P@8 |
| PLM | HiLAT | 92.7 | 95.0 | 69.0 | <u>73.5</u> | 68.1 | 55.4 |
| | PLM-ICD | 90.2 | 92.7 | 64.8 | 69.6 | 65.0 | 53.0 |
| | KEPT | 92.6 | 94.8 | 68.9 | 72.9 | 67.3 | 54.8 |
| Non-PLM | CAML | 87.5 | 90.9 | 53.2 | 61.4 | 60.9 | - |
| | MultiResCNN | 89.9 | 92.8 | 60.6 | 67.0 | 64.1 | - |
| | HyperCore | 89.5 | 92.9 | 60.9 | 66.3 | 63.2 | - |
| | LAAT | 92.5 | 94.6 | 66.6 | 71.5 | 67.5 | 54.7 |
| | JointLAAT | 92.5 | 94.6 | 66.1 | 71.6 | 67.1 | 54.6 |
| | TwoStage | 92.6 | 94.5 | 68.9 | 71.8 | 66.7 | - |
| | MSMN | 92.8 | 94.7 | 68.3 | 72.5 | 68.0 | 54.8 |
| | CoRelation | **93.3** | **95.1** | **69.3** | **73.1** | **68.3** | **55.6** |

# Experiment

Table 5.4: Results on the MIMIC-IV-50 test sets.

| Category | Method | MIMIC-IV-**ICD9**-50 | | | | | MIMIC-IV-**ICD10**-50 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AUC | | F1 | | Pre | AUC | | F1 | | Pre |
| | | Macro | Micro | Macro | Micro | P@5 | Macro | Micro | Macro | Micro | P@5 |
| PLM | PLM-ICD | 95.0 | 96.4 | 71.4 | 75.5 | 62.4 | 93.4 | 95.6 | 69.0 | 73.3 | 64.6 |
| Non-PLM | CAML | 93.1 | 94.1 | 65.3 | 69.2 | 58.6 | 91.1 | 93.2 | 64.3 | 67.6 | 59.6 |
| | LAAT | 94.9 | 96.3 | 70.0 | 74.5 | 62.0 | 93.2 | 95.5 | 68.2 | 72.6 | 64.4 |
| | JointLAAT | 94.9 | 96.3 | 69.9 | 74.3 | 62.0 | 93.4 | 95.6 | 68.4 | 72.9 | 64.5 |
| | MSMN | 95.1 | 95.5 | 71.9 | 75.8 | 62.6 | 93.6 | 95.7 | 70.3 | 74.2 | 65.2 |
| | CoRelation | **95.4** | **96.7** | **72.5** | **76.0** | **62.9** | **93.8** | **96.0** | **70.6** | **74.4** | **65.4** |

# Experiment

Table 5.5: Results on the MIMIC-III-Full test set.

| Method | AUC | | F1 | | Pre | | |
|---|---|---|---|---|---|---|---|
| | Macro | Micro | Macro | Micro | P@5 | P@8 | P@15 |
| PLM-ICD | 92.5 | 98.9 | 8.4 | 58.0 | 83.9 | 76.7 | 61.1 |
| CAML | 89.5 | 98.6 | 8.8 | 53.9 | - | 70.9 | 56.1 |
| MultiResCNN | 91.0 | 98.6 | 8.5 | 55.2 | - | 73.4 | 58.4 |
| HyperCore | 93.0 | 98.9 | 9.0 | 55.1 | - | 72.2 | 57.9 |
| LAAT | 91.9 | 98.8 | 9.9 | 57.5 | 81.3 | 73.8 | 59.1 |
| JointLAAT | 92.1 | 98.8 | **10.7** | 57.5 | 80.6 | 73.5 | 59.0 |
| TwoStage | 94.6 | 99.0 | 10.5 | 58.4 | - | 74.4 | - |
| MSMN | 95.0 | **99.2** | 10.3 | 58.4 | 82.5 | 75.2 | 59.9 |
| CoRelation | **95.2** | **99.2** | 10.2 | **59.1** | **83.4** | **76.2** | **60.7** |

# Experiment

Table 5.6: Results on the MIMIC-IV-Full test sets.

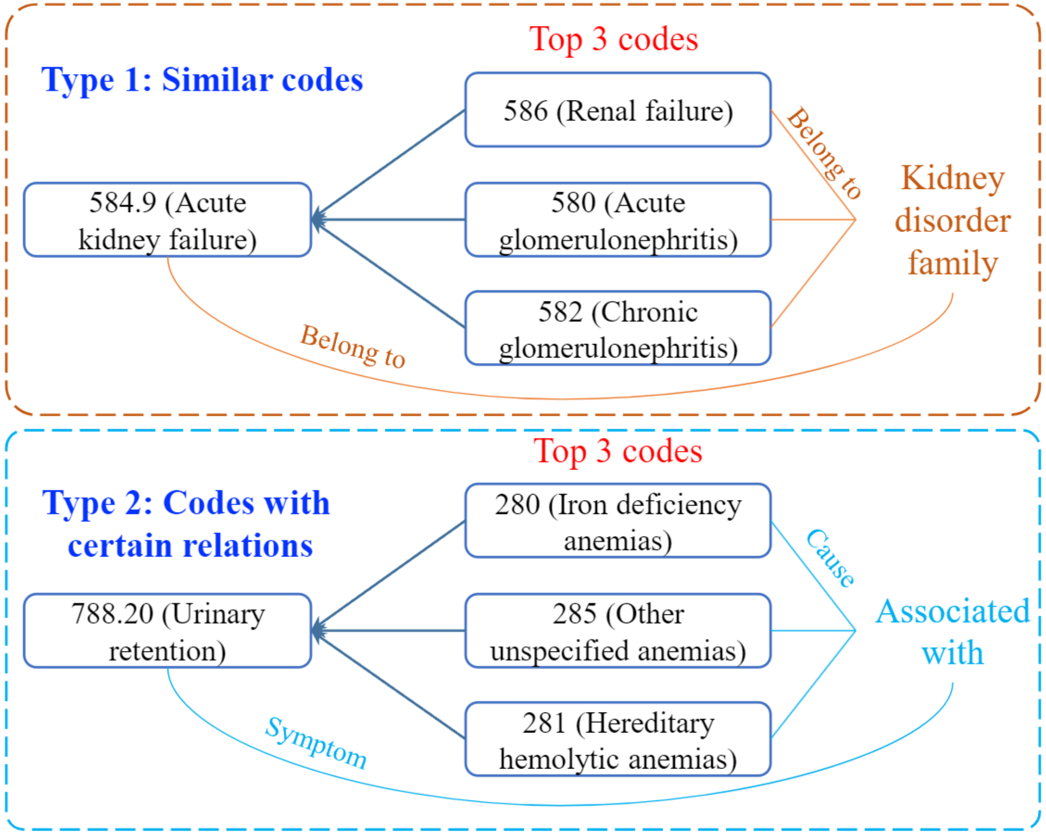
| Category | Method | MIMIC-IV-**ICD9**-Full | | | | | MIMIC-IV-**ICD10**-Full | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AUC | | F1 | | Pre | AUC | | F1 | | Pre |
| | | Macro | Micro | Macro | Micro | P@8 | Macro | Micro | Macro | Micro | P@8 |
| PLM | PLM-ICD | 96.6 | 99.5 | 14.4 | <u>62.5</u> | <u>70.3</u> | 91.9 | 99.0 | 4.9 | 57.0 | 69.5 |
| Non-PLM | CAML | 93.5 | 99.3 | 11.1 | 57.3 | 64.9 | 89.9 | 98.8 | 4.1 | 52.7 | 64.4 |
| | LAAT | 95.2 | 99.5 | 13.1 | 60.3 | 67.5 | 93.0 | 99.1 | 4.5 | 55.4 | 67.0 |
| | JointLAAT | 95.6 | 99.5 | 14.2 | 60.4 | 67.5 | 93.6 | 99.3 | 5.7 | 55.9 | 66.9 |
| | MSMN | **96.8** | **99.6** | 13.9 | 61.2 | 68.9 | 97.1 | **99.6** | 5.4 | 55.9 | 67.7 |
| | CoRelation | **96.8** | 99.5 | **15.0** | **62.4** | **70.1** | **97.2** | **99.6** | **6.3** | **57.8** | **70.0** |

# Ablation Study

Table 5.7: Results of ablation experiments on the MIMIC-III datasets.

| Dataset | MIMIC-III-50 | | | | | | MIMIC-III-Full | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | AUC | | F1 | | Pre | | AUC | | F1 | | Pre | | |
| | Macro | Micro | Macro | Micro | P@5 | P@8 | Macro | Micro | Macro | Micro | P@5 | P@8 | P@15 |
| CoRelation | **93.3** | **95.1** | **69.3** | **73.1** | **68.3** | **55.6** | **95.2** | **99.2** | 10.2 | **59.1** | **83.4** | **76.2** | **60.7** |
| W/O Relation | 93.1 | 95.0 | 69.0 | 72.6 | 68.1 | 55.2 | 95.2 | 99.1 | 9.3 | 58.9 | 82.8 | 75.7 | 60.5 |
| W/O FRG | 93.2 | 95.1 | 69.0 | 72.9 | 68.2 | 55.5 | 95.1 | 99.2 | 10.0 | 58.8 | 83.3 | 76.0 | 60.5 |
| W/O Context | 92.0 | 93.7 | 66.4 | 70.0 | 66.2 | 53.8 | 95.0 | 99.1 | **10.7** | 57.9 | 81.4 | 74.3 | 59.4 |
| W/O SAA | 92.5 | 94.7 | 68.6 | 72.2 | 67.9 | 55.0 | 95.0 | 99.1 | 9.7 | 58.8 | 82.9 | 75.9 | 60.1 |

# Case Study




- The model is able to learn meaningful results on the graph modality.

# Conclusion

- In addressing the ICD code prediction problem, providing better modeling on other than text modalities like the graph modality can help the model better capture the complex code relation and resulting better prediction performance.
- However, it is important to organically combine the two modalities.