Deep Learning for Automated Medical Coding

Shaoxiong Ji

shaoxiong.ji@aalto.fi

July 1, 2022 at LMU Munich

Medical Coding

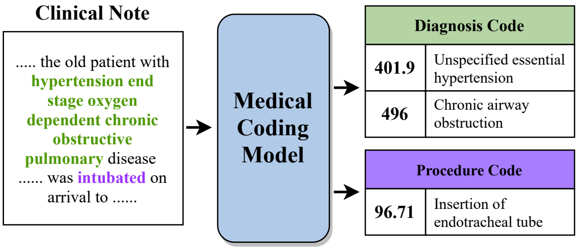


Figure: An example of medical coding with ICD codes

- Standard translation of written patient descriptions
- Standardized treatment alignment; insurance reimbursement
- Extreme multi-label multi-class classification

Effective CNN Encoding

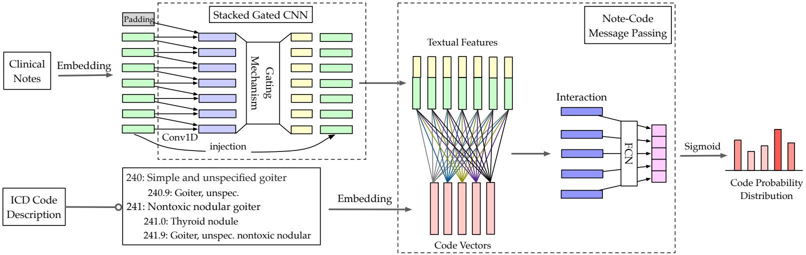
Challenges

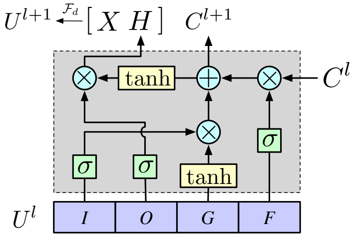
- Complex diagnosis information: professional medical vocabulary and noise e.g., non-standard synonyms and misspellings
- Lengthy documents: from hundreds to thousands of tokens.

Solutions:

- Effective feature representation learning
- Effective convolutional networks (this talk)
- Improved BERT-baed (hierarchical) models or efficient transformers, e.g., DLAC (Feucht et al., 2021)

Gated Convolutional Networks





GatedCNN-NCI, Findings of ACL 2021 (Ji et al., 2021)

Effective CNN Encoding: Results

Table: Results on MIMIC-III dataset with top-50 ICD codes. "-" indicates no results reported in the original paper.

Model	AUC-ROC Macro	AUC-ROC Micro	F1 Macro	F1 Micro	P@5
C-MemNN (Prakash et al., 2017)	83.3	-	-	-	42.0
Attentive LSTM (Shi et al., 2017)	-	90.0	-	53.2	-
CAML (Mullenbach et al., 2018)	87.5	90.9	53.2	61.4	60.9
MultiResCNN (Li and Yu, 2020)	89.9±0.4	92.8±0.2	60.6±1.1	67.0±0.3	64.1±0.1
HyperCore (Cao et al., 2020)	89.5±0.3	92.9±0.2	60.9±0.1	66.3±0.1	63.2±0.2
GatedCNN-NCI (ours)	**91.5**±0.3	**93.8**±0.1	**62.9**±0.5	**68.6**±0.1	**65.3**±0.1

Aalto-yliopisto
Aalto-universitetet
Aalto University

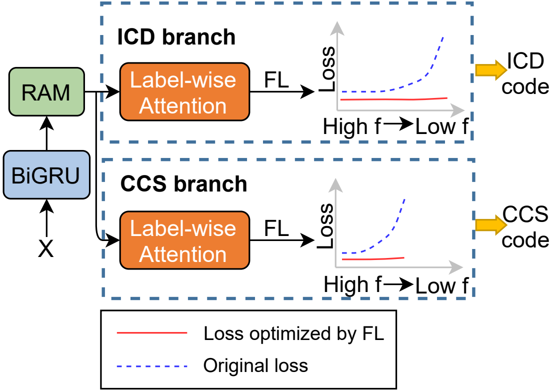
Effective CNN Encoding: Parameters

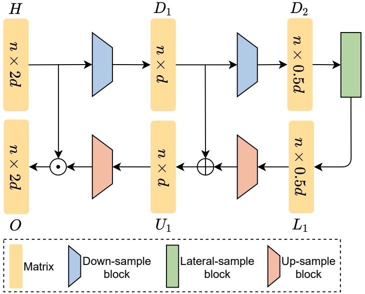
Table: Number of trainable parameters

Model	num. params.
CAML (Mullenbach et al., 2018)	6.2M
DCAN (Ji et al., 2020)	8.7M
MultiResCNN (Li and Yu, 2020)	11.9M
ClinicalBERT (Alsentzer et al., 2019)	113.8M
GatedCNN-NCI	7.6M

Multitask Medical Coding

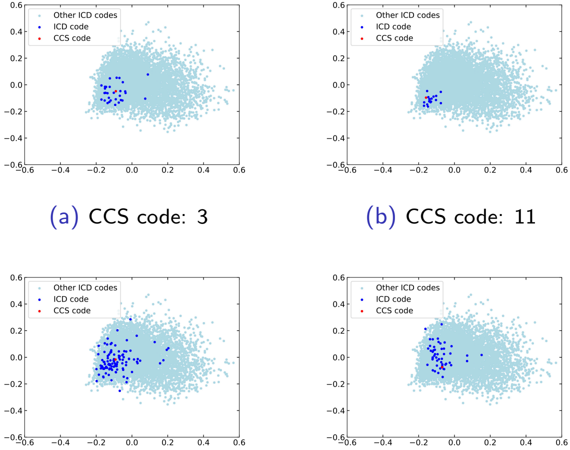
- High-dimensional label space
- Different disease classification systems
- Multitask learning with different granularities





MT-RAM, ECML-PKDD 2021 (Sun et al., 2021b)
MARN, Preprint (Sun et al., 2021a)

Can multitask learning connect different medical coding systems?



(a) CCS code: 3

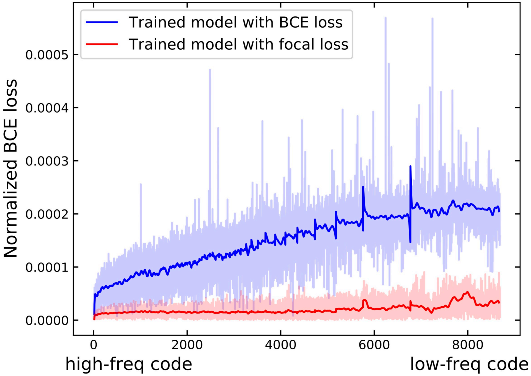
(b) CCS code: 11

(c) CCS code: 195

(d) CCS code: 223

- The embeddings of representative significant CCS codes and their corresponding ICD codes.
- The relevant ICD codes are clustered around the respective significant CCS code.
- MARN learns representations that capture informative relationships between the codes.

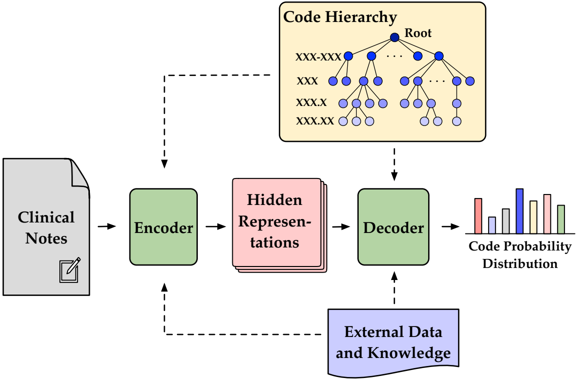
Does the model optimized with focal loss balance the learning between low- and high-frequency codes?



Normalized binary cross entropy loss loss of each ICD code, with x-axis sorted by code frequency. The high-frequency codes are on the left, the low-frequency codes on the right.

MARN optimized with focal loss can balance the learning of high- and low-frequency codes.

Unified Encoder-Decoder Framework



Categories	Functions	Representative Methods
Encoders	Extract text features	CNN, RNN, graph neural networks, attention, Transformers, capsule networks
Deep Connections	Build deep architecture	Stacking, residual networks, embedding injection
Decoders	Improve code prediction	Linear layer, attention, hierarchical decoders, multitask decoders, few-shot/zero-shot decoders
Auxiliary Data	Enhance feature learning	Code descriptions, code hierarchy, Wikipedia articles, chart data, entities and concepts

Table: Categorization of building blocks under the unified framework

Open Questions

- Long-term Dependency and Scalability
- Class Imbalance and Hierarchical Decoding
- Updated Guidelines and Data Shift
- Interpretability (post-hoc vs inherent interpretability)
- Novel Encoder-decoder Architectures

References I

E. Alsentzer, J. Murphy, W. Boag, W.-H. Weng, D. Jindi, T. Naumann, and M. McDermott. Publicly Available Clinical BERT Embeddings. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pages 72–78, 2019.

P. Cao, Y. Chen, K. Liu, J. Zhao, S. Liu, and W. Chong. HyperCore: Hyperbolic and Co-graph Representation for Automatic ICD Coding. In *Proceedings of ACL*, pages 3105–3114, 2020.

M. Feucht, Z. Wu, S. Althammer, and V. Tresp. Description-based label attention classifier for explainable icd-9 classification. In *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)*, pages 62–66, 2021.

S. Ji, E. Cambria, and P. Marttinen. Dilated convolutional attention network for medical code assignment from clinical text. In *Proceedings of the 3rd Clinical Natural Language Processing Workshop at EMNLP*, pages 73–78, 2020.

S. Ji, S. Pan, and P. Marttinen. Medical code assignment with gated convolution and note-code interaction. In *Findings of ACL-IJCNLP*, 2021.

F. Li and H. Yu. Icd coding from clinical text using multi-filter residual convolutional neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8180–8187, 2020.

J. Mullenbach, S. Wiegreffe, J. Duke, J. Sun, and J. Eisenstein. Explainable Prediction of Medical Codes from Clinical Text. In *Proceedings of NAACL-HLT*, pages 1101–1111, 2018.

A. Prakash, S. Zhao, S. A. Hasan, V. Datla, K. Lee, A. Qadir, J. Liu, and O. Farri. Condensed Memory Networks for Clinical Diagnostic Inferencing. In *Proceedings of AAAI*, 2017.

H. Shi, P. Xie, Z. Hu, M. Zhang, and E. P. Xing. Towards Automated ICD Coding Using Deep Learning. *arXiv preprint arXiv:1711.04075*, 2017.

W. Sun, S. Ji, E. Cambria, and P. Marttinen. Multitask balanced and recalibrated network for medical code prediction. *arXiv preprint arXiv:2109.02418*, 2021a.

W. Sun, S. Ji, E. Cambria, and P. Marttinen. Multitask Recalibrated Aggregation Network for Medical Code Prediction. In *Proceedings of ECML-PKDD*, 2021b.