A Multi-Aspect Technical Performance Evaluation of CovRNN

Laila Rasmy[1], Masayuki Nigo[1], Bijun Sai Kannadath[2], Ziqian Xie, Bingyu Mao, Angela Ross[1], Hua Xu[1], Degui Zhi[1]

[1]University of Texas Health Science Center in Houston, [2]College of Medicine, University of Arizona

UTHealth The University of Texas Health Science Center at Houston School of Biomedical Informatics

CovRNN

CovRNN[1] is a deep learning model to predict COVID-19 patients' outcomes on admission, including inhospital mortality (iMort), mechanical ventilation (mVent), and prolonged length of stay (pLOS)

CovRNN was trained on Cerner Real world data Covid-19 cohort v.20Q3, using our established pytorch_EHR framework

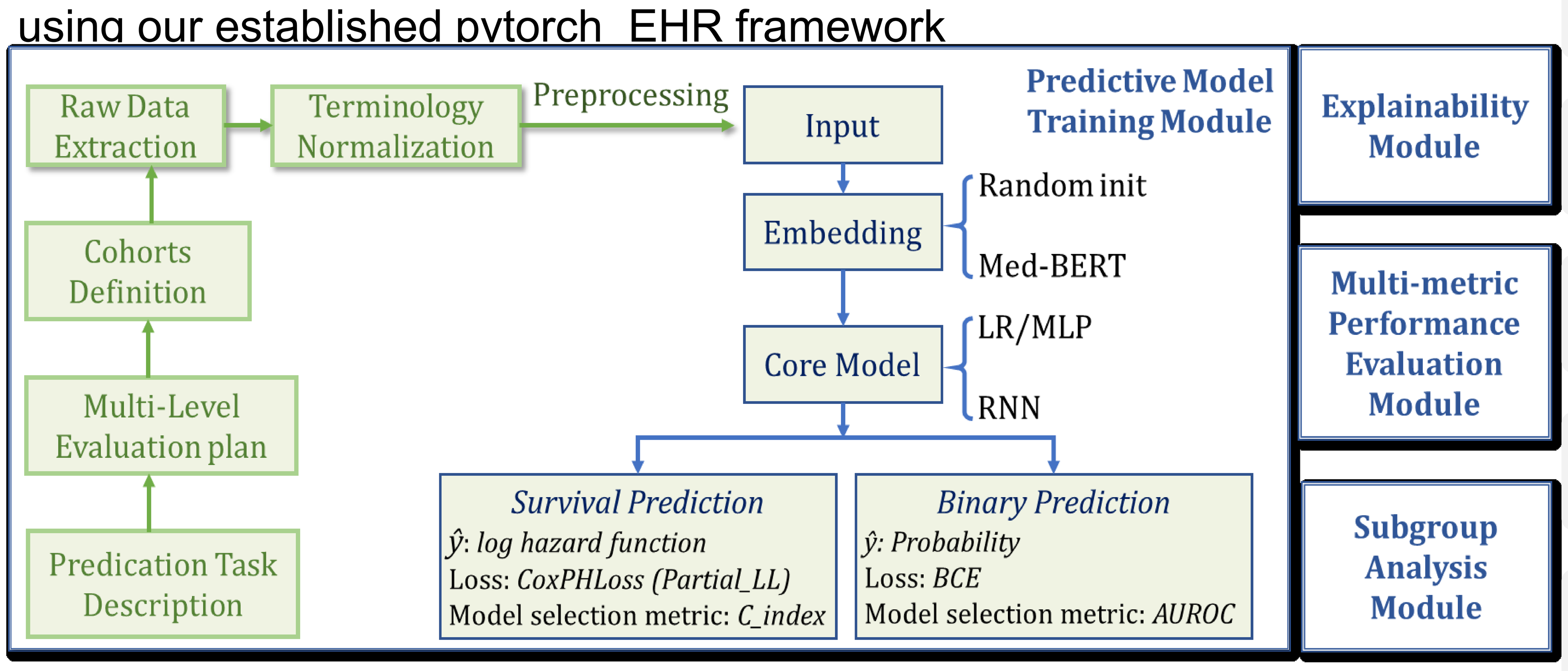


Figure 1. pytorch_EHR Framework

Implementability Evaluation Factors

Performance: How far predictions deviate from actual observation on a testing dataset

Transparency: How a given technology reached a certain decision

Generalizability : The ability of the model to digest new data and make accurate predictions regardless of the setting or the population

Data Mechanics: How data can flow between systems and computational infrastructures

Efficiency: The amount of time and computational resources required by the model to work properly

Data Privacy: models should not require any PHI data

Results

Performance

Table 1. Multi-metric Performance Evaluation Results

	AUROC	AUPRC	Specificity @95%Sensitivity	Sensitivity	Specificity
iMort	93·0%	52·8%	70.9%	84.9%	85.7%
mVent	92·9%	79.5%	63.5%	83.4%	85.1%
pLOS	86.5%	60·0%	55.6%	79.7%	76.3%

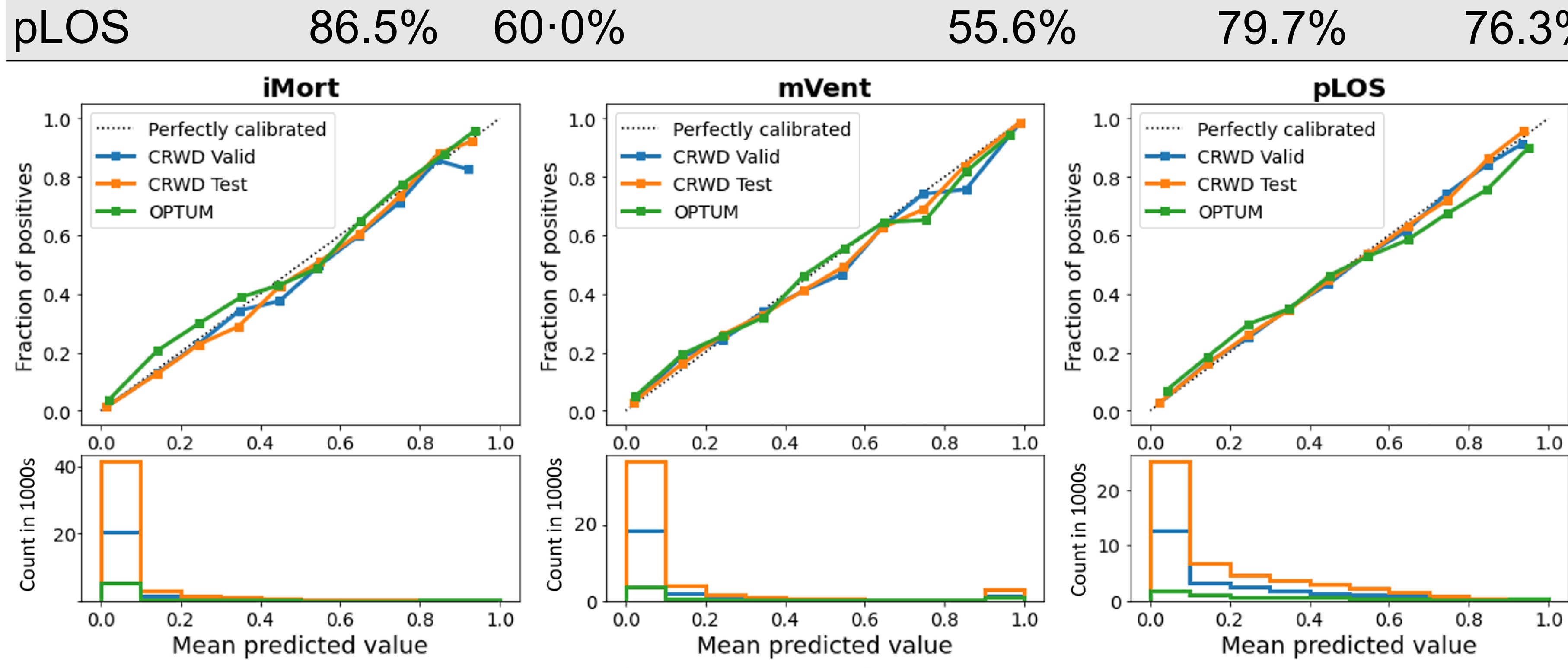


Figure 2. Calibration Curves

Generalizability

Table 2. AUROC across multiple external validation test sets.

Evaluation Dataset	iMort	mVent	pLOS
Multi-hospital (Cerner)	93·0%	92·9%	86·5%
Hospital 1	91·8%	91·5%	87·2%
Hospital 2	97·0%	96·0%	88·3%
Multi-hospital (Optum)	91.3%	91.5%	81.0%

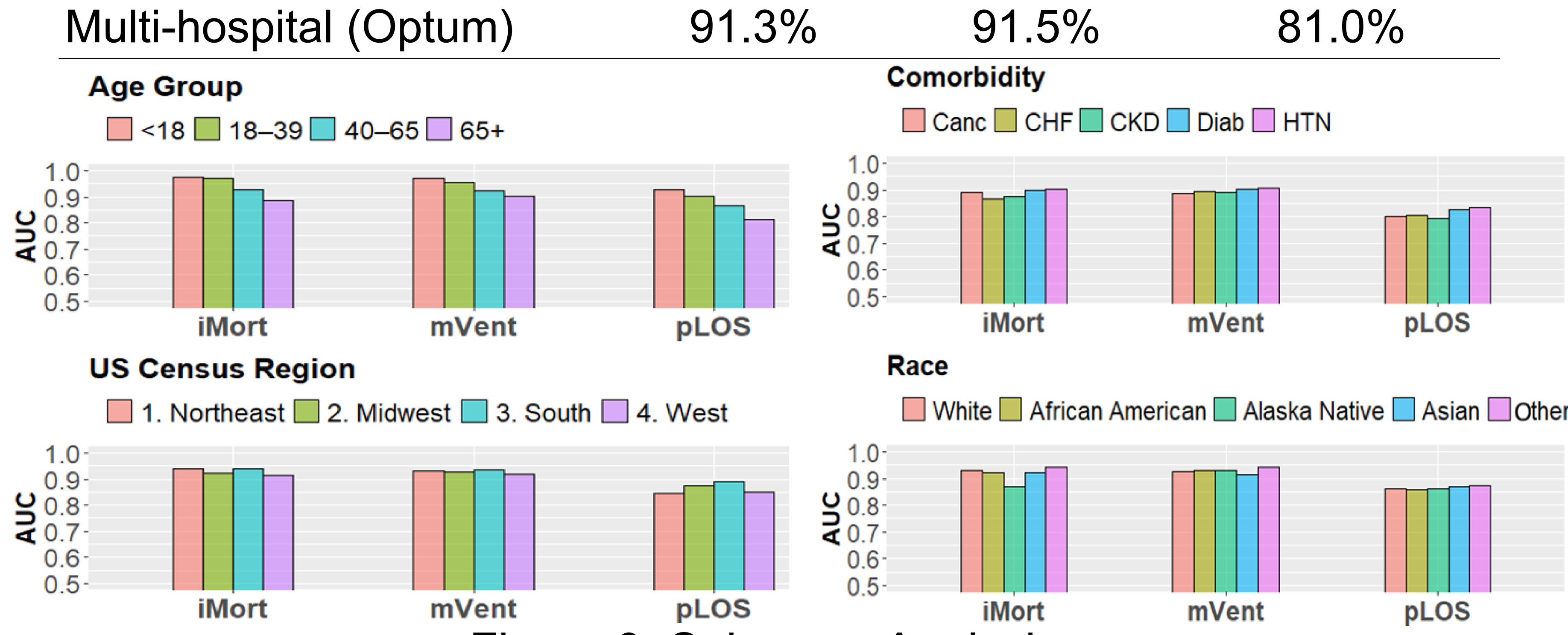
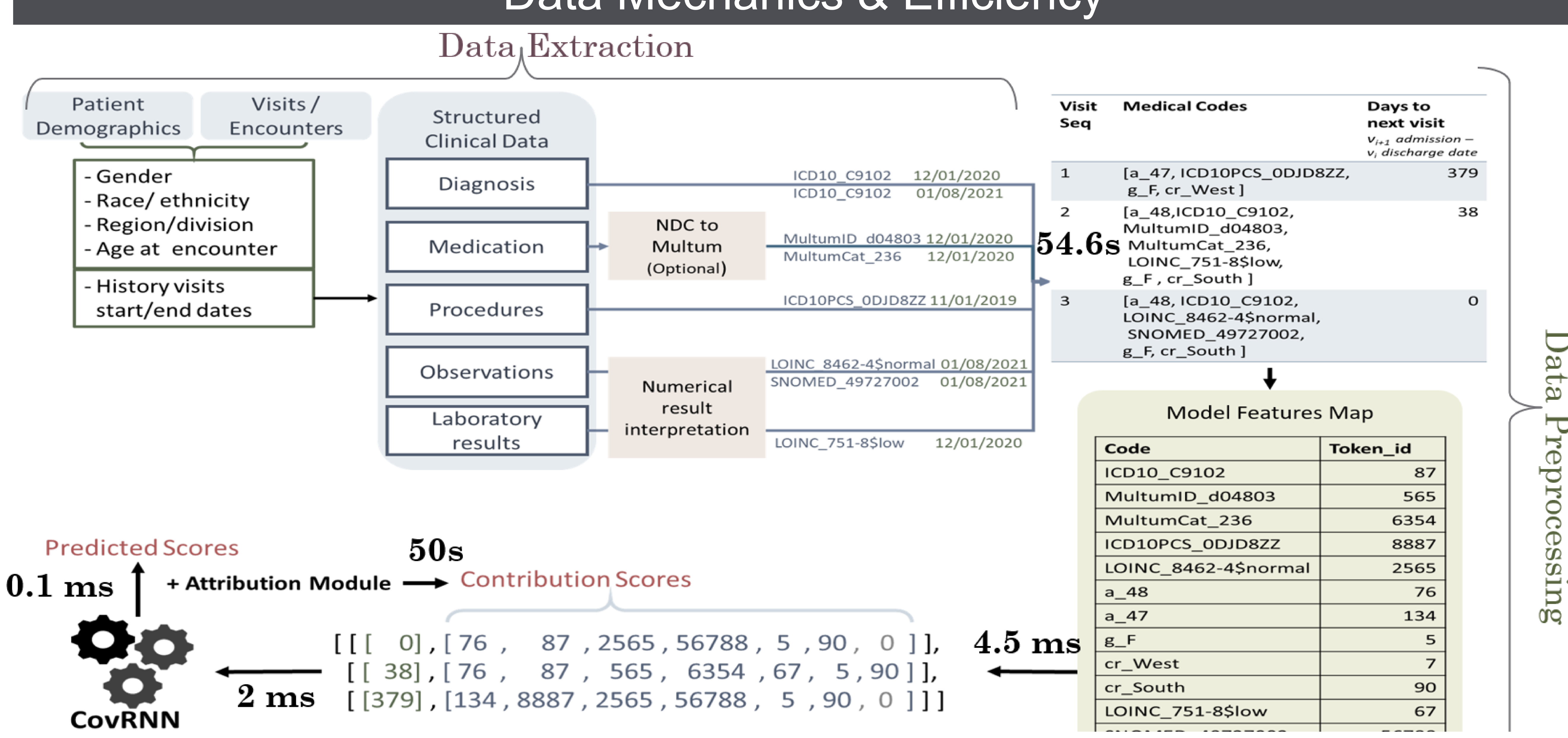


Figure 3. Subgroup Analysis

Transparency

1. We followed the TRIPOD standards to report our methods and results. 2. We reported the results of three ablation experiments to highlight the factors behind our model's good performance 3. We used the *Integrated Gradient* method to explain the individual predictions. The results and sample visualization are accessible through the QR code below which will direct you to our paper supplementary material

Data Mechanics & Efficiency



Discussion & Conclusion

- We considered the implementability factors while developing and evaluating the technical performance of CovRNN.
- Trained on a large heterogeneous dataset from 85 health systems, CovRNN showed one of the highest reported prediction accuracies on multiple outcomes as well as transferability and generalizability.
- We never use PHI data as features to our models.
- We found the explainability module is the most time-consuming step and we are working on mitagting as part of our interface design and usability evaluation stage.
- For reproducibility, we shared our code and trained models on https://github.com/ZhiGroup/CovRNN

References 1. Rasmy L, Nigo M, Kannadath BS, Xie Z, Mao B, Patel K, Zhou Y, Zhang W, Ross A, Xu H, Zhi D. CovRNN—A recurrent neural network model for predicting outcomes of COVID-19 patients: model development and validation using EHR data. *Lancet Digital Health*.

Please contact us:
Degui.Zhi@uth.tmc.edu
Laila.rasmy.gindybekhet@uth.tmc.edu