SUPPLEMENTAL MATERIAL

Supplemental Table I: Machine Learning Reporting Guideline Checklist (MI-CLAIM[16])

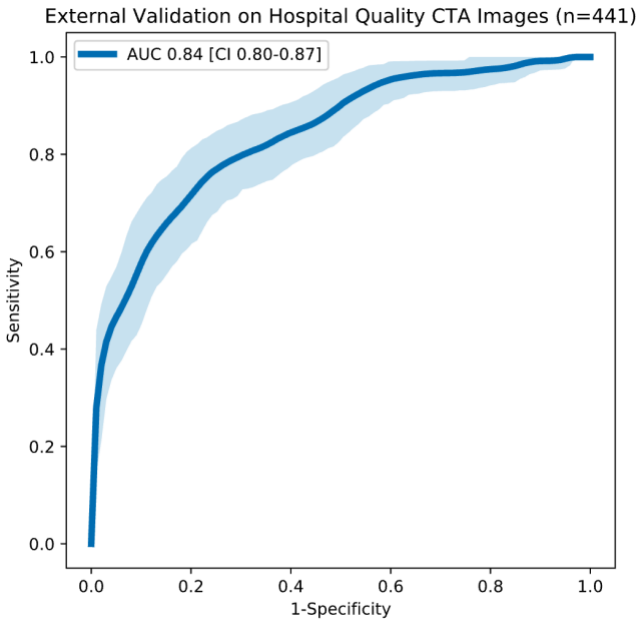
Study Design (Part 1)	**Completed (page number)**	**Notes if not completed**
The clinical problem in which the model will be employed is clearly detailed in the paper.	7	
The research question is clearly stated.	7	
The characteristics of the cohorts (training and test sets) are detailed in the text.	6	
The cohorts (training and test sets) are shown to be representative of real-world clinical settings.	6	
The state-of-the-art solution used as a baseline for comparison has been identified and detailed.	6	
Data and Optimization (Parts 2, 3)	**Completed (page number)**	**Notes if not completed**
The origin of the data is described and the original format is detailed in the paper.	5	
Transformations of the data before it is applied to the proposed model are described.	6	
The independence between training and test sets has been proven in the paper.	6	
Details on the models that were evaluated and the code developed to select the best model are provided.	6	
Is the input data type structured or unstructured?	Unstructured	
Model Performance (Part 4)	**Completed (page number)**	**Notes if not completed**
The primary metric selected to evaluate algorithm performance (e.g., AUC, F-score, etc.), including the justification for selection, has been clearly stated.	6	
The primary metric selected to evaluate the clinical utility of the model (e.g., PPV, NNT, etc.), including the justification for selection, has been clearly stated.	7	
The performance comparison between baseline and proposed model is presented with the appropriate statistical significance.	7	
Model Examination (Part 5)	**Completed (page number)**	**Notes if not completed**
Examination technique (sensitivity analysis)	9, Supplemental Figure 2	

A discussion of the relevance of the examination results with respect to model/algorithm performance is presented.	9-10	
A discussion of the feasibility and significance of model interpretability at the case level if examination methods are uninterpretable is presented.	10	
A discussion of the reliability and robustness of the model as the underlying data distribution shifts is included.	10	
Reproducibility (Part 6): choose appropriate tier of transparency	**Notes**	
Tier 1: complete sharing of the code		
Tier 2: allow a third party to evaluate the code for accuracy/fairness; share the results of this evaluation		
Tier 3: release of a virtual machine (binary) for running the code on new data without sharing its details	We are working to host our software for access for research purposes.	
Tier 4: no sharing		

Supplemental Table II: Sensitivity and Specificity at Varying DeepSymNet-v2 cutoffs

DeepSymNet-v2 Probability Cutoff	Sensitivity	Specificity
0.95	98%	31%
0.80	95%	50%
0.65	88%	69%
0.57	76%	73%

Supplemental Figure I. DeepSymNet-v2 performance for LVO detection on in-hospital testing dataset. ROC curve with AUC and 95% confidence intervals.



Supplemental Figure II. DeepSymNet-v2 performance for LVO detection in the two MSU cohorts. ROC curves with AUC curves and 95% confidence intervals. UTH represents the Houston cohort and UCLA represents the Los Angeles based cohort.