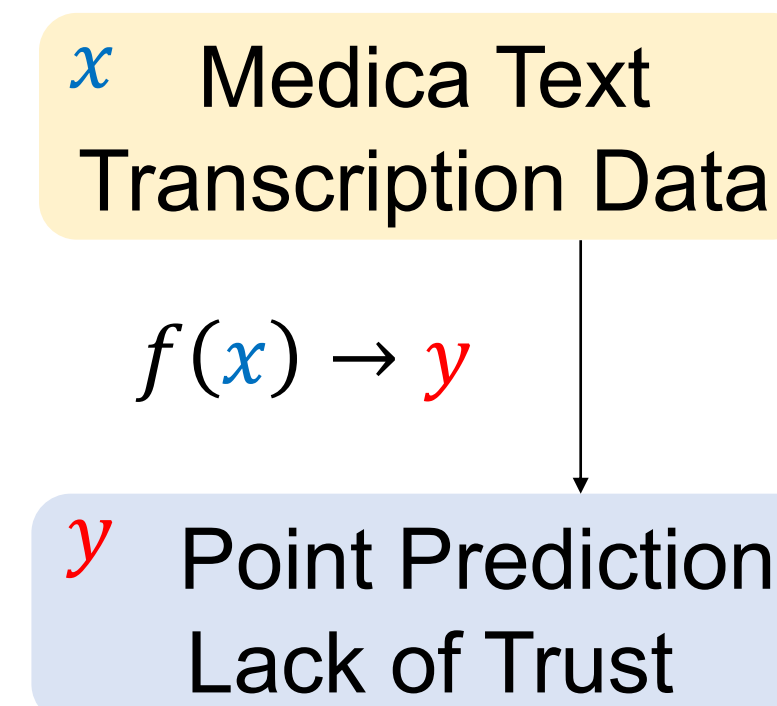


Enhancing Risk Aware Decision in Healthcare with Uncertainty Quantification

Rahul Vishwakarma, Jinha Hwang, Benyamin Ahmadnia
California State University Long Beach

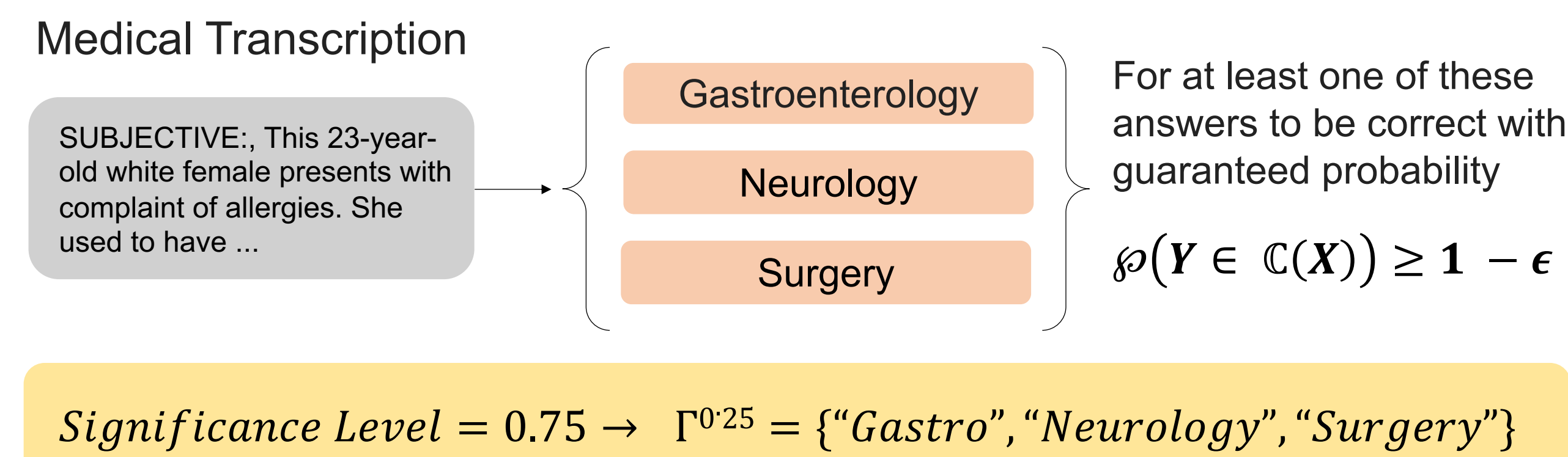


Motivation: Unreliable point prediction



- Most of the Machine Learning model lacks **calibration**.
- No guarantee** for uncertainty estimations for new data.

Conformal Prediction



Inference: p-values

p-values								y (sig = 0.20)
p-0	p-1	p-2	p-3	p-4	p-5	p-6	p-7	
0.125	0.122	0.131	0.373	0.122	0.122	0.130	0.124	3
False	False	False	True	False	False	False	False	

Performance Metrics

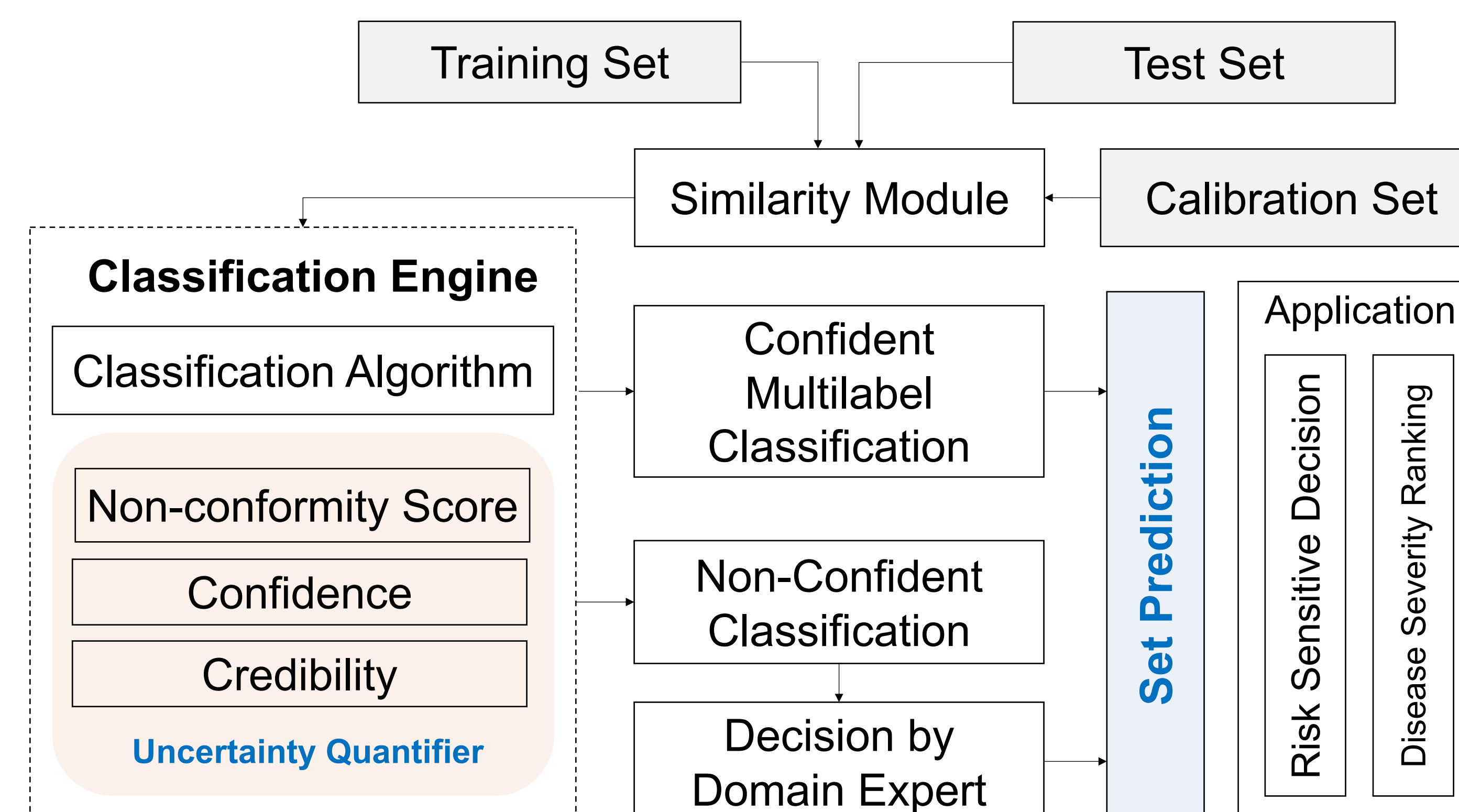
sig	mean_err	avg_c
0.5	0.506	0.604
0.7	0.701	0.371
0.8	0.804	0.251
0.9	0.894	0.115

- Coverage: proportion of true target values that fall within the pred intervals.
- Efficiency: how tight the prediction intervals are.

Prior Work

- Confidence intervals:** Not be well-suited for complex data and model.
- Bayesian Inference:** Require careful selection of prior distributions.
- Bootstrap:** Computationally expensive.

Proposed Solution



Reject Predictions: null set

sig	conf	prediction
0.5	0.61	{NULL}
0.15	0.86	{surgery}
0.10	0.87	{surgery, others}

Algorithm agnostic method for “**reject**” – model cannot provide a confident prediction for this input.

Method: Set valued prediction

- Idea:** Instead of calibrating the probabilities for each individual outcome (e.g., $Y = 1$), apply calibrated probabilities to a set of potential outcomes (e.g., $Y \in \{1, 2, 3\}$).
- Goal:** Forecast a limited collection of reasonable responses.
- Benefit:** Model Agnostic.

Experiment

Data: MTSamples Medical Transcription
Clinical Labels: {0} Cardiovascular/Pulmonary, {1} Consult History and Phy, {2} Gastroenterology, {3} General Medicine, {4} Neurology, {5} Obstetrics/ Gynecology, {6} Surgery, {7} Others

Multinomial Naive Bayes								
	0	1	2	3	4	5	6	7
0	57	0	1	1	0	0	1	3
1	0	28	2	0	0	0	0	2
2	32	7	404	22	18	17	30	15
3	0	0	0	34	0	0	2	0
4	0	0	2	0	32	0	0	1
5	1	1	5	0	0	29	2	0
6	0	0	0	1	0	0	21	0
7	0	1	3	0	0	0	0	12

Conformal Inference								
	0	1	2	3	4	5	6	7
0	8	2	34	5	4	1	4	5
1	3	2	20	0	1	5	0	1
2	68	34	290	39	27	33	17	37
3	4	4	17	3	1	3	1	3
4	2	5	19	4	1	1	1	2
5	5	0	23	3	0	3	0	4
6	2	2	10	2	0	0	3	3
7	2	0	11	1	1	0	0	1

Disease Severity Ranking

$Confidence(x) = \sup\{1 - \epsilon : |\Gamma_{\epsilon}(x)| \leq 1\}$ Derive the confidence of each predicted label.

$c(label_1) > \dots, c(label_n)$ Rank the instances of same label based on the confidence score.

Conclusion

- Reject option (null set) for a more trustworthy model.
- NCM based on Lesk Score for conformal prediction.
- Refine classification decisions based on ranking.

