Data-driven automated classification algorithms for acute health conditions: Applying PheNorm to Anaphylaxis

Joshua C. Smith, PhD[1], Daniel Park BS[1], Jill M. Whitaker, MSN[1], Michael F. McLemore, BSN[1], Robert Winter, BA[1], Arvind Ramaprasan, MS[2]
David Cronkite, MS[2], Saranrat Wittayanukorn, PhD[3], Danijela Stojanovic, PharmD, PhD[3], Yueqin Zhao, PhD[3], Sarah Dutcher, PhD[3]
Kevin B. Johnson, MD, MS[4], David S. Carrell, PhD[2], Brian D. Williamson, PhD[2]

[1]Vanderbilt University Medical Center, Nashville, TN; [2]Kaiser Permanente Washington Health Research Institute, Seattle, WA;
[3]US Food and Drug Administration, Silver Spring, MD; [4]University of Pennsylvania, Philadelphia, PA

ABSTRACT

Accurate identification of anaphylaxis using observational data is important for medical product safety surveillance, but difficult to diagnose clinically or recognize algorithmically. Traditional phenotyping methods rely on expensive gold standard training data and manual feature engineering. We have instead applied an automated approach, PheNorm, to create a computable phenotype for identifying patients with anaphylaxis using NLP, machine learning, and low-cost silver-standard training labels. Performance was comparable to a recently published, higher-cost manual phenotyping effort.

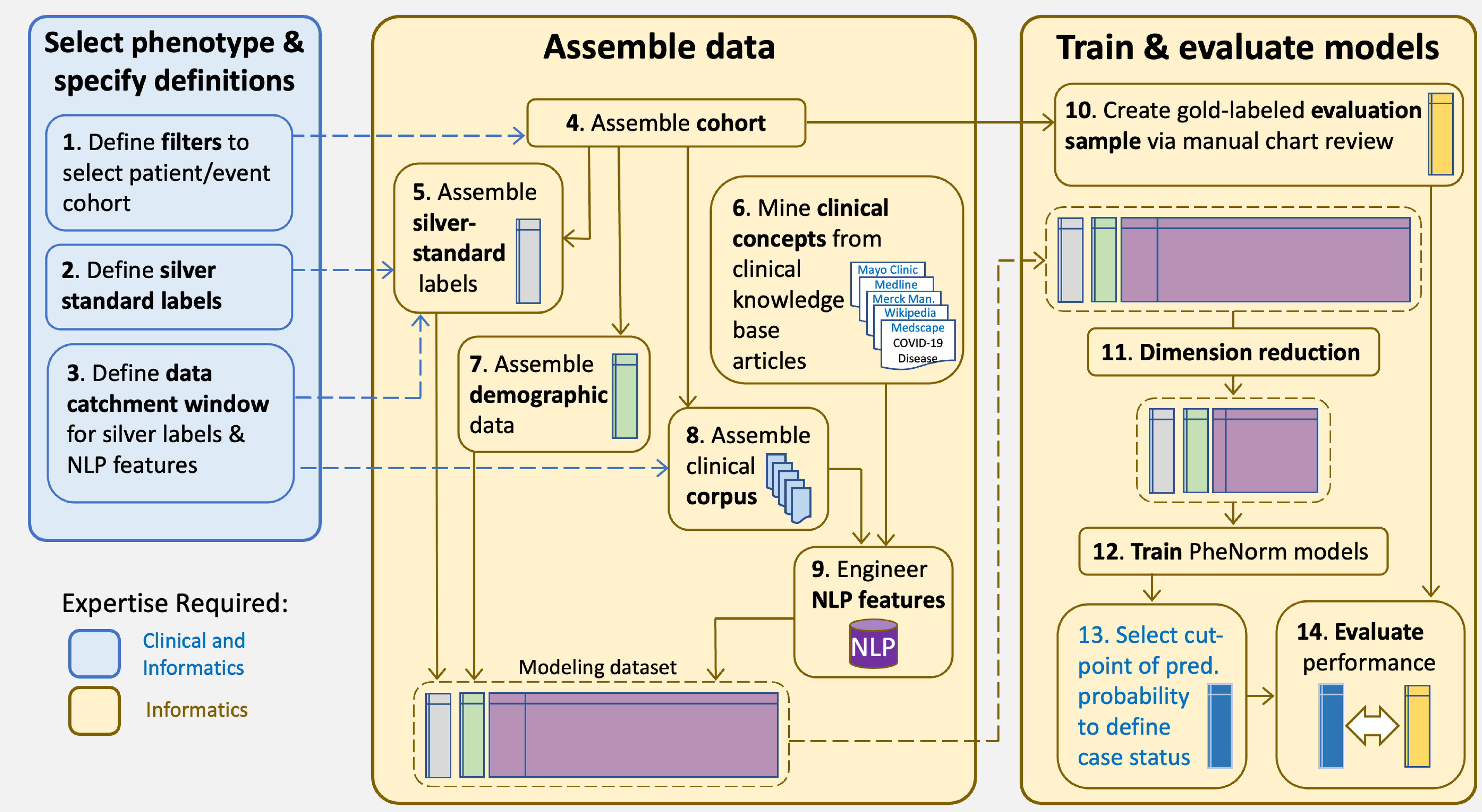
METHODS

We created four PheNorm models, one for each silver label, and a fifth that averaged predicted probabilities generated by the others. To evaluate performance, we used manual chart review to assign gold-standard labels based on the NIAID/FAAN Consensus Criteria for Anaphylaxis for a random sample of 255 patients set aside from our study cohort. Charts were initially reviewed by two reviewers (kappa 0.67). We evaluated model performance using predicted probabilities in the samples of patients with gold-standard labels. Gold labels were not used in model training.

RESULTS

The study cohort consisted of 1213 distinct patients and 1283 potential anaphylaxis events. Of note, using the cutoff where PPV is 0.79, model sensitivity is 0.65; at a PPV of 0.80, sensitivity is 0.60. The automated lower-cost PheNorm approach produced results comparable to a recently published higher-cost effort trained using gold-standard training labels and manually engineered features (sensitivity of 0.66 at PPV=0.79) (Carrell, et al., Am J Epidemiol 2023).

PHENORM AUTOMATED PHENOTYPING APPROACH

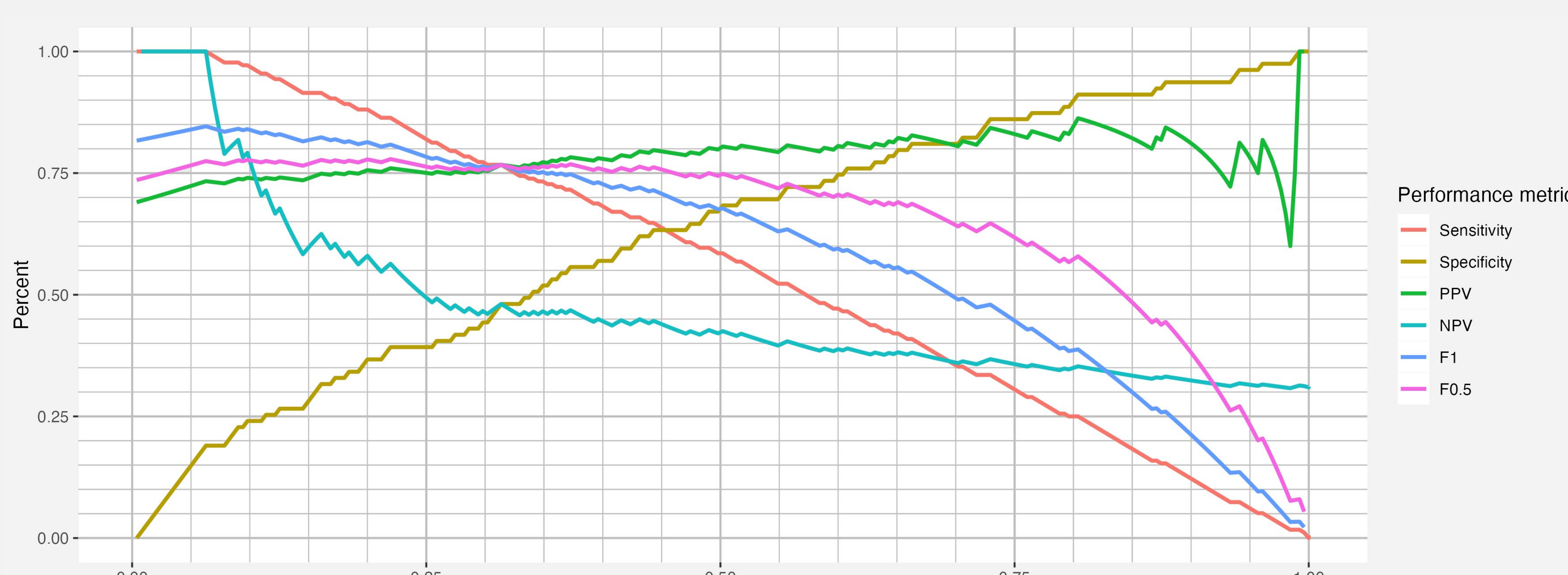


Smith, et al. Data-driven automated classification algorithms for acute health conditions: applying PheNorm to COVID-19 disease. JAMIA 2024.

RESULTS

Model performance for each silver label for identifying anaphylaxis at quantiles of model-predicted risk that maximize F1.

Silver Standard Labels	AUC (95% CI)	Max F1	Sensitivity	Specificity	PPV	NPV
Encounters with anaphylaxis diagnoses	0.679 [0.606, 0.752]	0.846	1.000	0.190	0.733	1.000
Anaphylaxis mentions in notes	0.613 [0.533, 0.693]	0.839	0.994	0.165	0.726	0.929
Anaphylaxis NLP concepts from notes	0.573 [0.492, 0.654]	0.824	0.955	0.190	0.724	0.652
Anaphylaxis/epinephrine mentions	0.615 [0.537, 0.694]	0.831	0.977	0.165	0.723	0.765
Aggregate	0.614 [0.534, 0.694]	0.842	1.000	0.165	0.727	1.000



Performance for the best-performing model trained on count of unique encounters coded with ICD anaphylaxis diagnoses.

CONCLUSION

Automated clinical phenotyping algorithms can be developed for acute health conditions with reasonable success and with far less effort and expertise than required by traditional approaches.

VANDERBILT UNIVERSITY MEDICAL CENTER

Department of Biomedical Informatics

IPHI

Center for Improving the Public's Health through Informatics

Sentinel

joshua.smith@vumc.org