ECG-Based Machine Learning Model Identifies Patients At High Risk For Incident Pulmonary Hypertension

Poster/Abstract #

Lee G.[1], Hakim N.[1], Chen R.[1], Pfeifer J.[1], Vidmar D.[1], Thompson W.[1], Nemani A.[1], Miotto R.[1], Morland K.[2], Nelsen A.[2], Fornwalt B.[1]
[1]Tempus AI, Inc., Chicago, IL, United States, [2]United Therapeutics, Research Triangle Park, NC, United States

INTRODUCTION

- Patients with Pulmonary Hypertension (PH) suffer from diagnostic delays and remain underdiagnosed
- Machine learning (ML) models trained and deployed on electrocardiograms (ECG) may reduce these diagnostic gaps
- Existing models may lack real-world generalizability due to poor-quality labels and biased evaluation
- **Hypothesis:** ECG models can identify PH and demonstrate performance on a population with no pre-existing diagnosis

METHODS

- Links were created between ECGs and a natural-language-processing (NLP) phenotype which yielded 1.8M ECGs for model development, of which 55,023 ECGs were labeled positive (Figure 1)
- Positive Labels: ECG occurs 12 months prior to the diagnosis
- Negative Labels: ECG occurs up to 12 months prior to most-recent encounter or death, with no PH diagnosis
- A convolutional neural network (CNN) was trained to identify PH from ECG, age, and sex

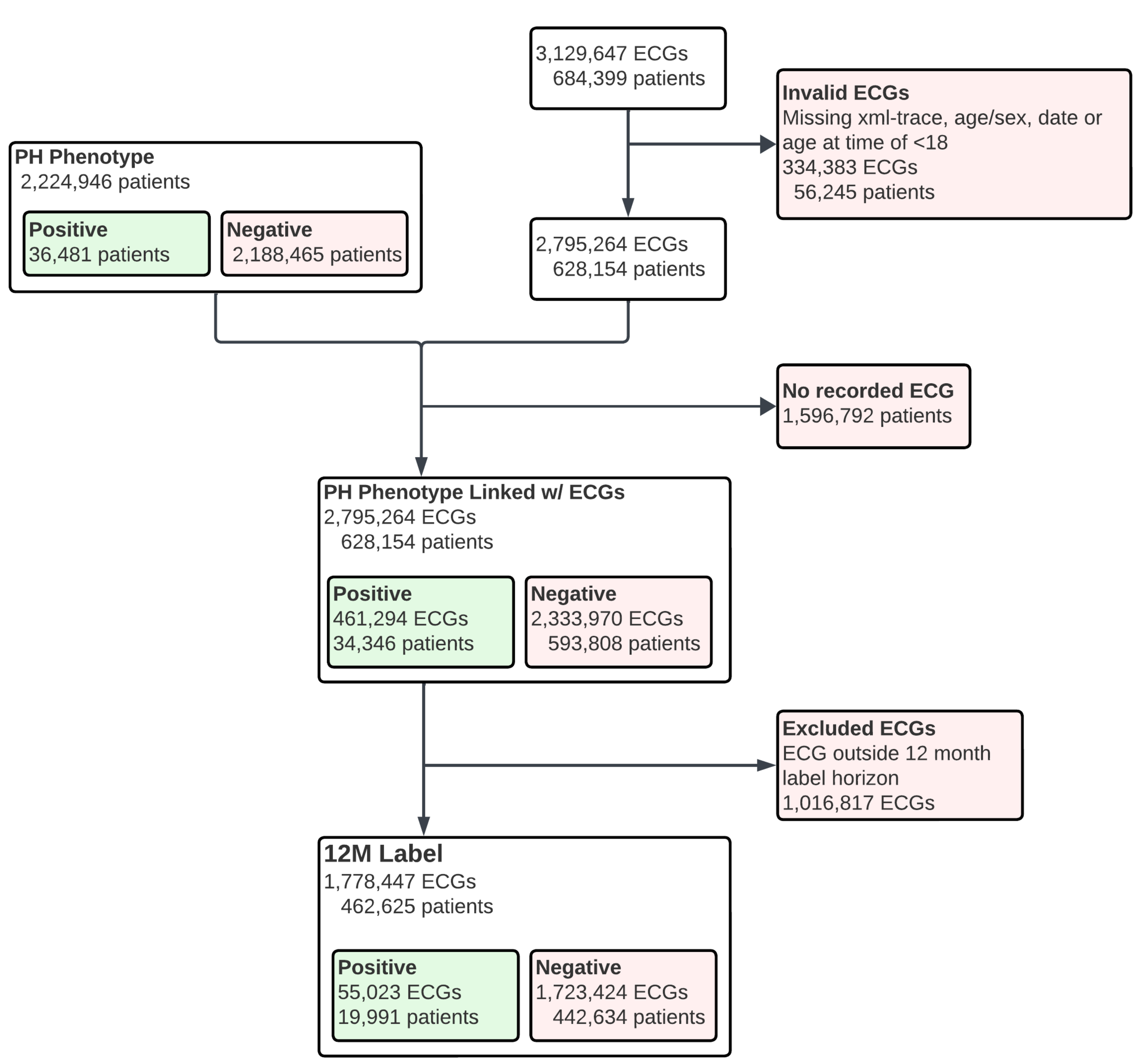

Figure 1. PH Patient Cohort. CONSORT diagram of the ECG extraction pipeline. Note that patients cannot have both positive and negative ECG due to label windowing.

SUMMARY

- Our ECG-based ML model showed good retrospective performance to identify risk of incident PH diagnosis
- We plan to implement this model in a prospective clinical trial to further investigate generalizability and real-world performance

RESULTS

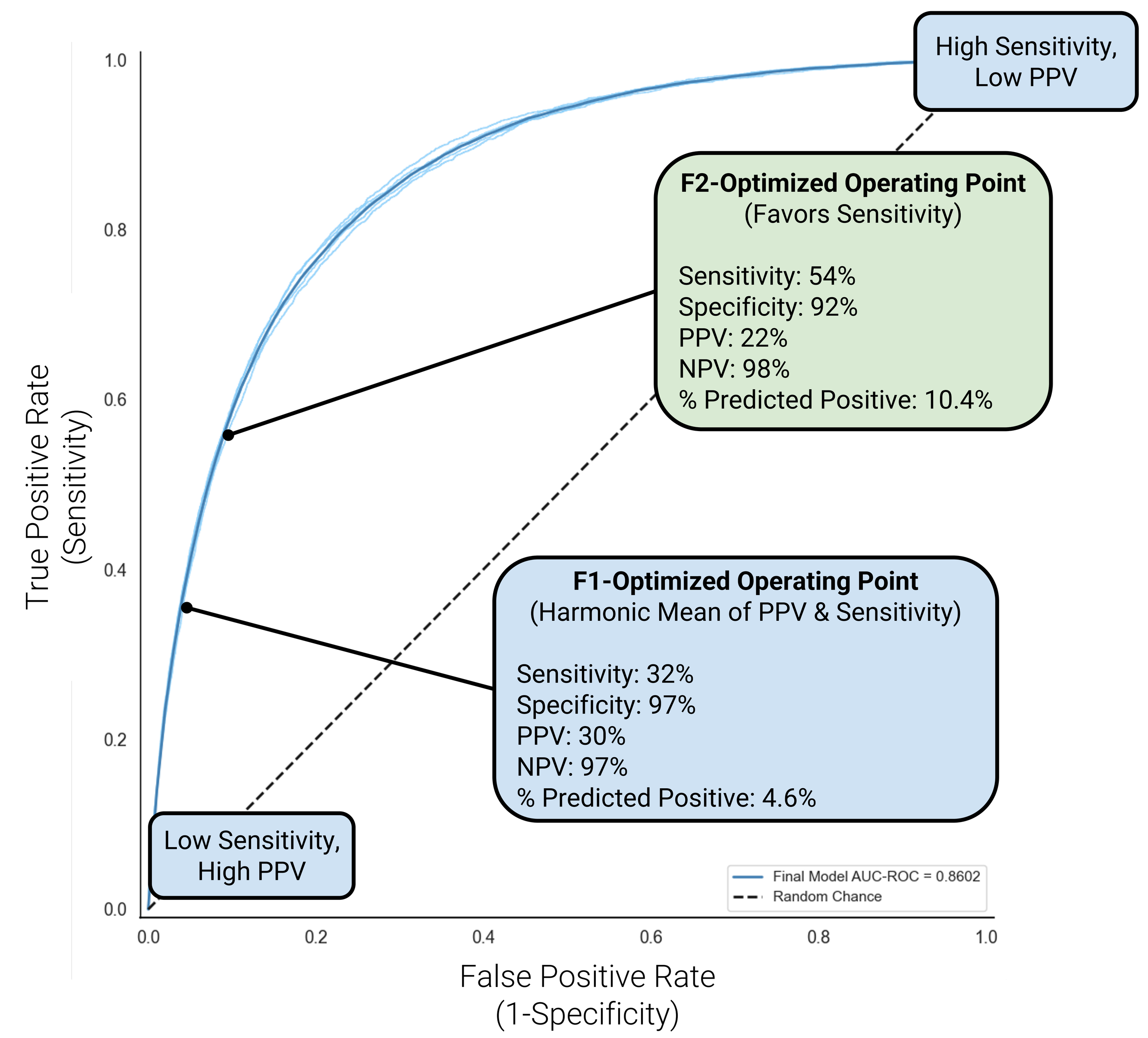

Figure 2. Operating Point Selection Impacts Clinical Utility. Selection of different operating points can steer a model towards different health system opportunities.

- A CNN performed well in identifying patients at high risk for PH within 12 months (Table 1)
- An F2-optimized operating point showed promise for real-world clinical deployment with >1/5 true positives out of 10% predicted high risk, while capturing 54% of the diseased population
- Performance was similar across all subgroups (Table 1) of the time windows analyzed (6,12, and 18 months prior to diagnosis).
- For a medium-sized health system of 1 million patients, these models may conservatively identify 112-431 new cases of PH annually

F2-optimized operating point				
Group		AUROC	AUPRC	F1-Score
Overall		0.86 [0.86,0.86]	0.25 [0.23,0.26]	0.31 [0.28,0.34]
Age	(18.0 40.0]	0.81 [0.78,0.83]	0.10 [0.06,0.15]	0.17 [0.09,0.30]
	(40.0, 60.0]	0.84 [0.83,0.84]	0.18 [0.16,0.20]	0.25 [0.20,0.30
	(60.0, 80.0]	0.80 [0.79,0.80]	0.24 [0.23,0.25]	0.30 [0.27,0.33]
	80+	0.73 [0.72,0.73]	0.30 [0.27,0.33]	0.36 [0.34,0.39]
Sex	Female	0.86 [0.86,0.86]	0.24 [0.23,0.25]	0.30 [0.27,0.33]
	Male	0.86 [0.85,0.87]	0.26 [0.24,0.28]	0.32 [0.29,0.35]
Race	Asian	0.92 [0.89,0.94]	0.29 [0.21,0.39]	0.32 [0.19,0.48]
	Black	0.88 [0.86,0.89]	0.24 [0.21,0.26]	0.30 [0.26,0.34]
	White	0.86 [0.85,0.86]	0.25 [0.23,0.26]	0.31 [0.28,0.34]
Ethnicity	Hispanic / Latino	0.89 [0.86,0.91]	0.23 [0.17,0.30]	0.29 [0.22,0.37]
	Not Hispanic / Latino	0.86 [0.86,0.87]	0.25 [0.24,0.26]	0.31 [0.28,0.34]
	Unknown	0.79 [0.78,0.81]	0.24 [0.22,0.26]	0.32 [0.29,0.34]

Table 1. ECG Model Performance. Values represent mean [95% CI] across cross-validation folds. All metrics are reported by selecting a random ECG per patient. Subgroups with proportion < 0.01 are not reported.

Disclosures: Tempus has received funding from the United Therapeutics Corporation to support development of predictive modeling technology.

Correspondence