Fiterau M, Wang J, Dubrawski AW, Clermont G, Hravnak M, Pinsky MR. Using expert review to calibrate semi-automated adjudication of vital sign alerts in step-down units *Critical Care Medicine*. 2015; *in press.*

Objectives: Machine Learning (ML) has shown predictive utility in analyzing vital sign (VS) data collected from physiologically unstable monitored patients. Training ML classifiers requires labeled ground truth data obtained via laborious annotation or events or manual chart reviews by expert clinicians. We aim to reduce the annotation effort through semi-automated adjudication.

Methods: Noninvasive monitoring data including ECG-derived heart rate (HR), respiratory rate (RR), systolic and diastolic blood pressure (BP), and pulse oximetry (SpO_2) were sampled at 1/20Hz and alerts were issued whenever VS exceed any of preset stability thresholds ($40 \leq HR \leq 140$/min, $8 \leq RR \leq 36$/min, $SpO_2 > 85\%$). Statistical features were extracted from each raw VS stream independently during the alert window. ML models using Informative Projections (IP) were trained separately on 91 RR and 194 SpO_2 events. 80 of remaining events (40 RR and 40 SpO_2) were then automatically selected by the system and annotated by 2 expert clinicians. Then the experts adjudicated the same alerts using the available chart time series. We have determined annotation confidence categories based on the confidence and agreement of the experts.

Results: Out of the 80 events, 36 are labeled with the same confidence (and label) irrespective of the manner of annotation, 3 events that could not be annotated based on the VS trace were annotated by analyzing the IPs, 10 of the events were annotated with more confidence based on the VS, while the remaining 31 (39% of total) could not be annotated based on the IPs, but could be adjudicated using VS. Informative-projection-assisted annotation was useful in labeling of 35 events, 53.03% of the non-ambiguous cases, reducing the need for chart-based adjudication to 31 events, 46.97% of the non-ambiguous cases.

Conclusions: Effective training of ML-based automatic alert adjudication systems, calibrated on selective human annotation, can improve accuracy of automated adjudication while reducing human effort to prepare training data for ML.

FUNDING:NIH R01NR013912, NSF 1320347.