University of Pittsburgh

Auton Lab
Carnegie Mellon University

Active machine learning to increase annotation efficiency in classifying vital sign events as artifact or real alerts in continuous noninvasive monitoring

Hravnak M; Chen L, Fiterau M, Dubrawski A, Clermont G, Guillame-Bert M, Bose E, Pinsky MR

University of Pittsburgh

Carnegie Mellon University

Disclosures

- PI on NIH NINR RO1NR013912 “Predicting Patient Instability Noninvasively for Nursing Care (PPINNC)”
- No Conflict of Interest

Rationale

Bedside monitors issue vital sign (VS) alarms when individual parameters crosses threshold, but many are due to monitoring artifacts, causing alarm fatigue. Better approaches for artifact filtering are required, but the task of classifying events as artifacts vs. real alerts is complex, requiring sophisticated algorithms acting on real-time data.

Active machine learning is one approach to learning feature differences in real alerts vs. artifact signals for subsequent classification. Machine learning does, however, first require a bank of events to be annotated by experts as real or artifact from which to commence learning, with larger banks providing the most robust classification.

Unfortunately, creating the bank of correctly annotated events to serve as ground truth in developing these algorithms requires considerable expert effort.

Purpose

We therefore proposed to use active learning to first learn a model based on a small number of existing annotated samples, and then use the model to further iteratively select small numbers of VS events for expert annotations that might be most helpful in further model building.

The samples to be annotated are selected in a manner which decreases the uncertainty in classifying the unlabeled samples and leads, after only a limited number of annotations, to a test set performance comparable to using a large number of expert-annotated events upfront, thereby reducing expert effort.

Methods

We recruited 314 admissions to a 24-bed stepdown unit, recording 18,314 hrs of noninvasive VS monitoring data at 1/20Hz frequency for continuous peripheral pulse oximetry (SpO_2), heart rate, and respiratory rate, and noninvasive blood pressure (BP) measured every 2h.

There were 219 events of low SpO_2 events <85%, which were visually adjudicated and annotated by two experts (MRP, MH) as artifact or real. Artifact comprised 17% of these events

BP events (systolic BP<80 or >200 mmHg, diastolic BP>110 mmHg; n=96) were similarly annotated, yielding 37.5% artifact.

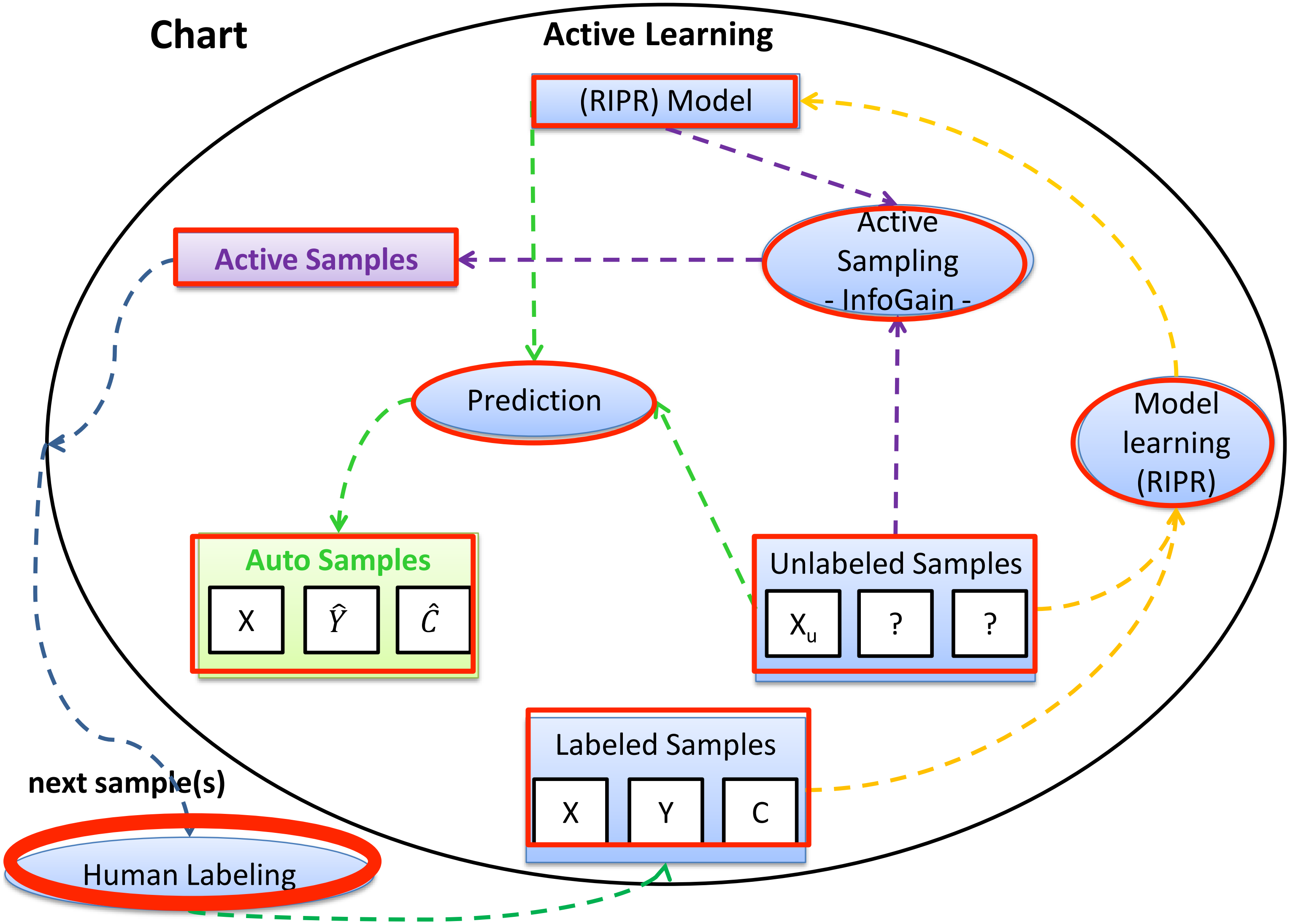
Methods (cont.)

We then simulated an active learning system using the expert-annotated SpO2 and BP data as ground truth. First, the system builds a logistic regression model with a fraction of expert-annotated data points (e.g. 10%). It then proposes a list of other events to be annotated, based on classification uncertainty.

Once feedback is received (ground truth label supplied) the model is updated, and, based on this new information, the system proposes the next batch of events. We tracked model performance at each stage on hold-out data.

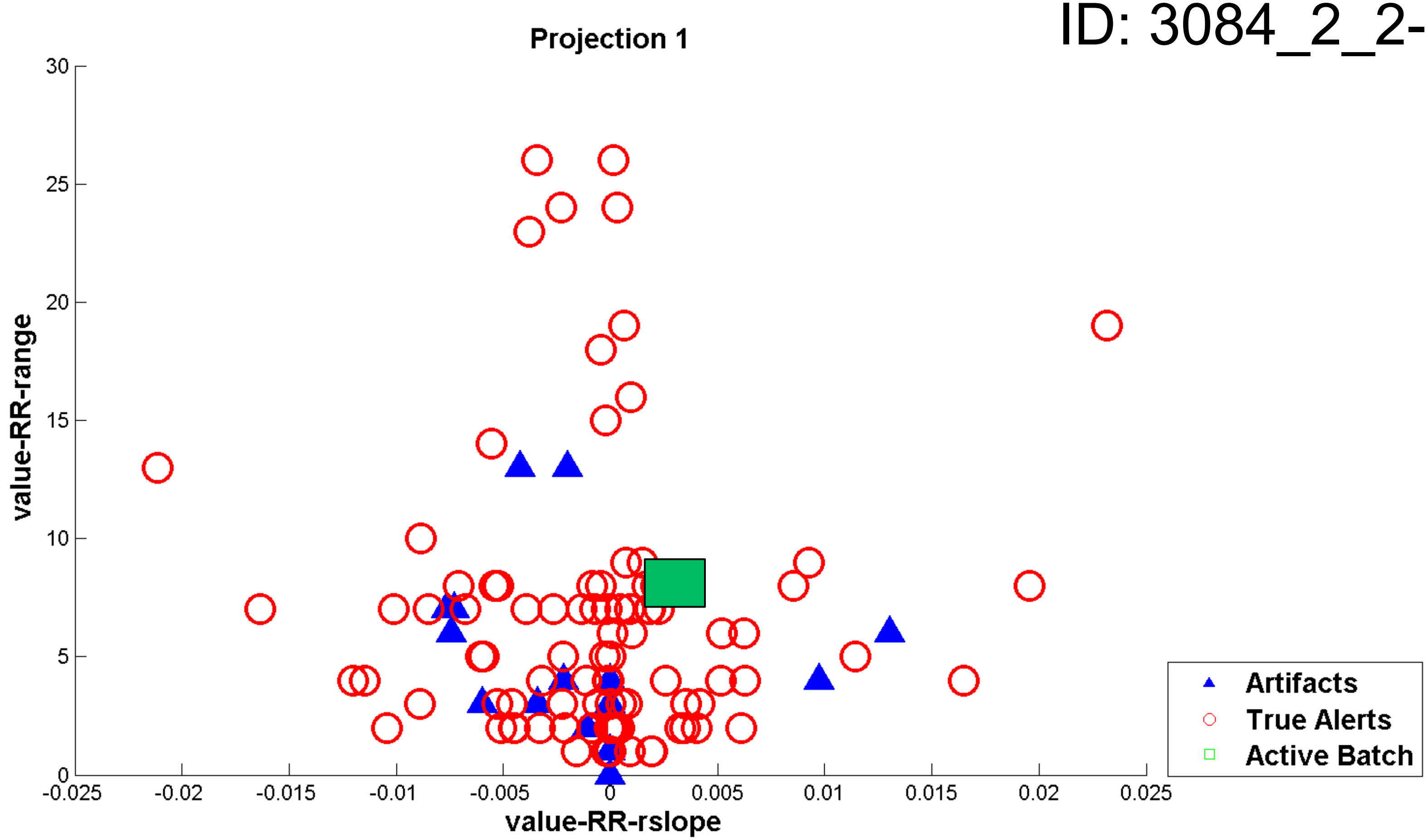
The machine learning approach utilized is RIPR (Regression for Informative Projection Recovery) which uses a regression type approach to explore multiple feature projections and models around a query point

Machine Workflow Chart



Active Sample Annotation Example 1

ID: 3084_2_2--RR

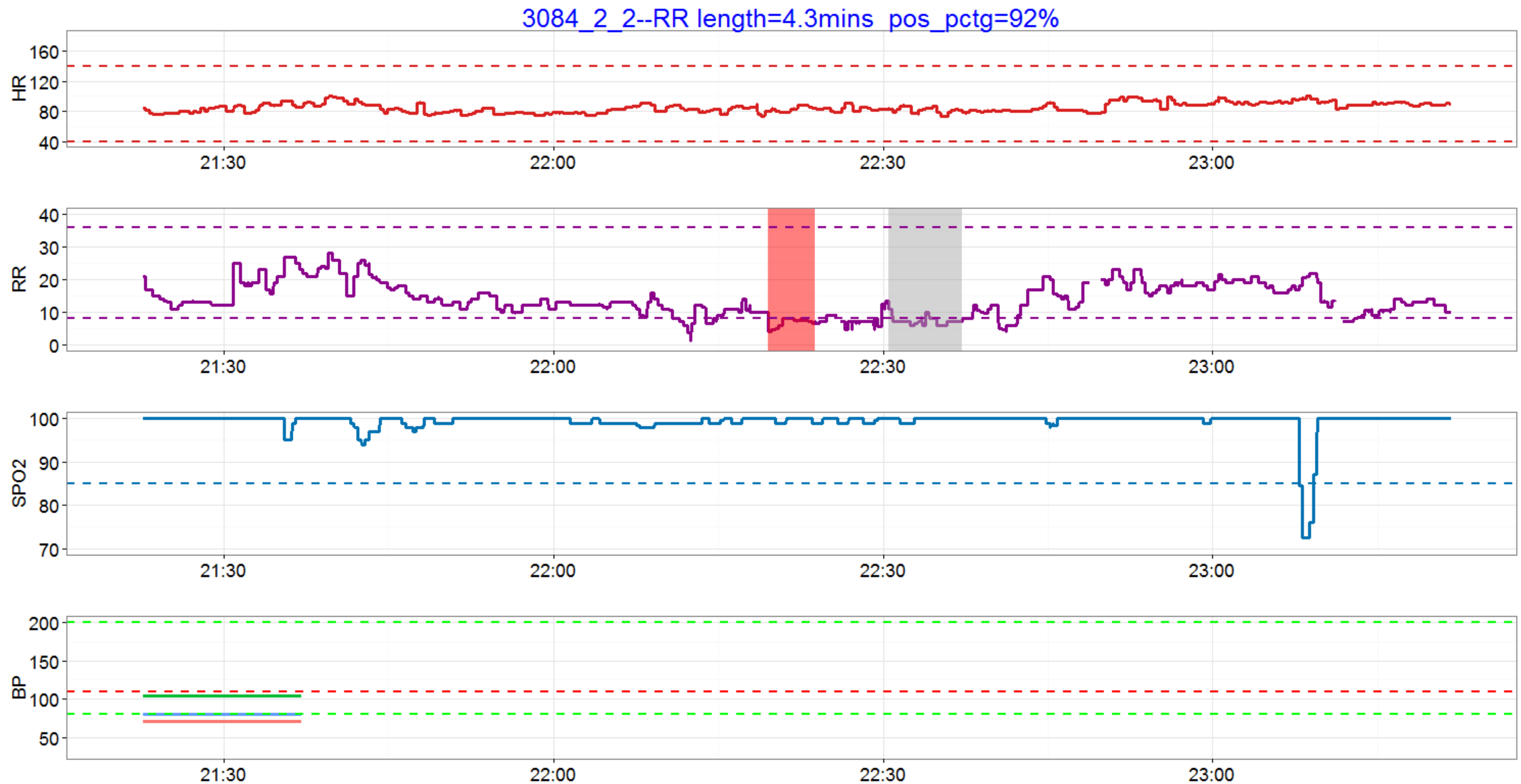


The sample ***can*** be confidently classified as a true alert.

University of Pittsburgh

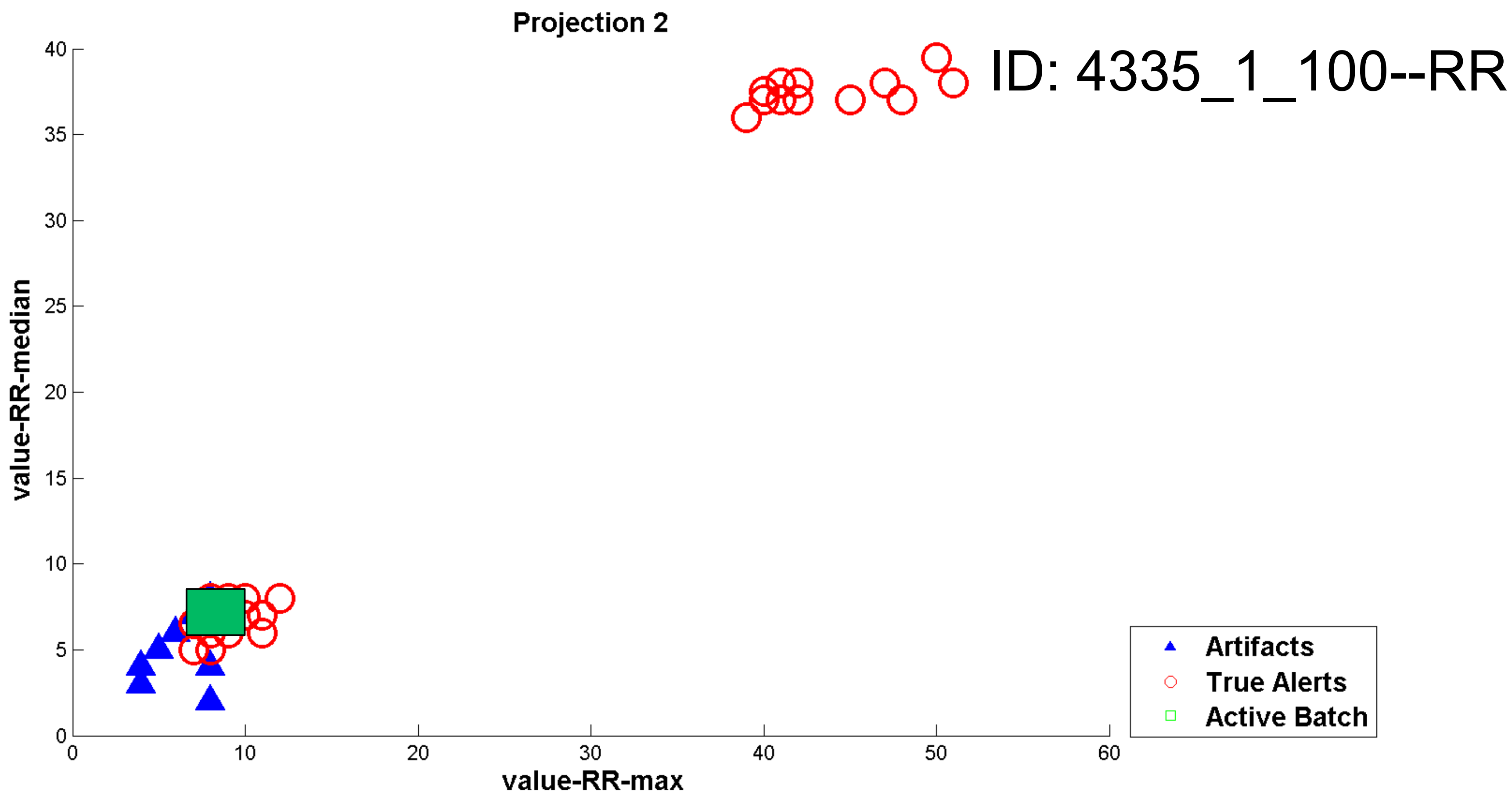
Auton Lab
Carnegie Mellon University

Active Sample Annotation Example 1



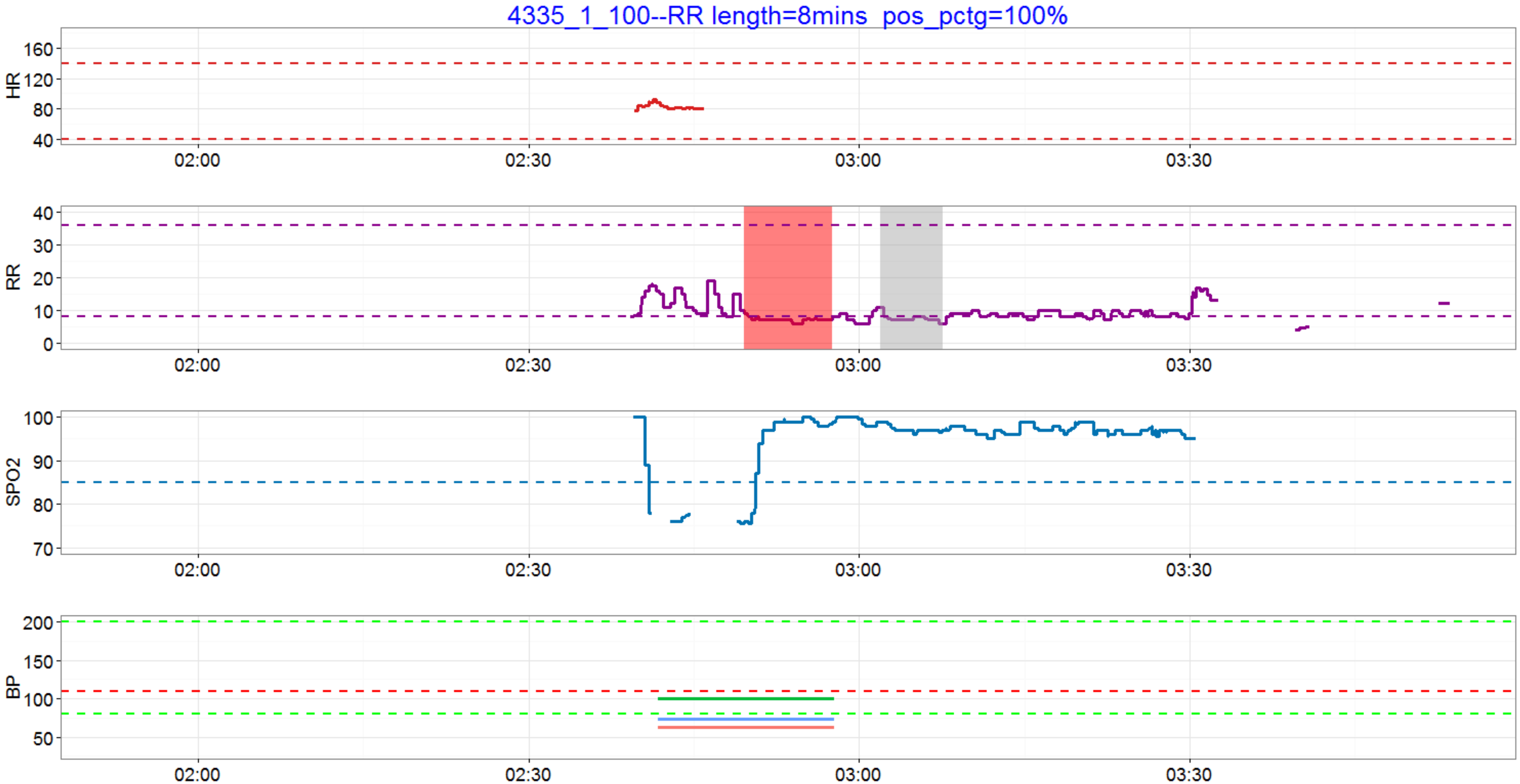
The sample can be auto-annotated and experts ***do not need to consult*** the vital sign trace.

Active Sample Annotation Example 2



The sample ***cannot*** be confidently classified, .and requires expert review

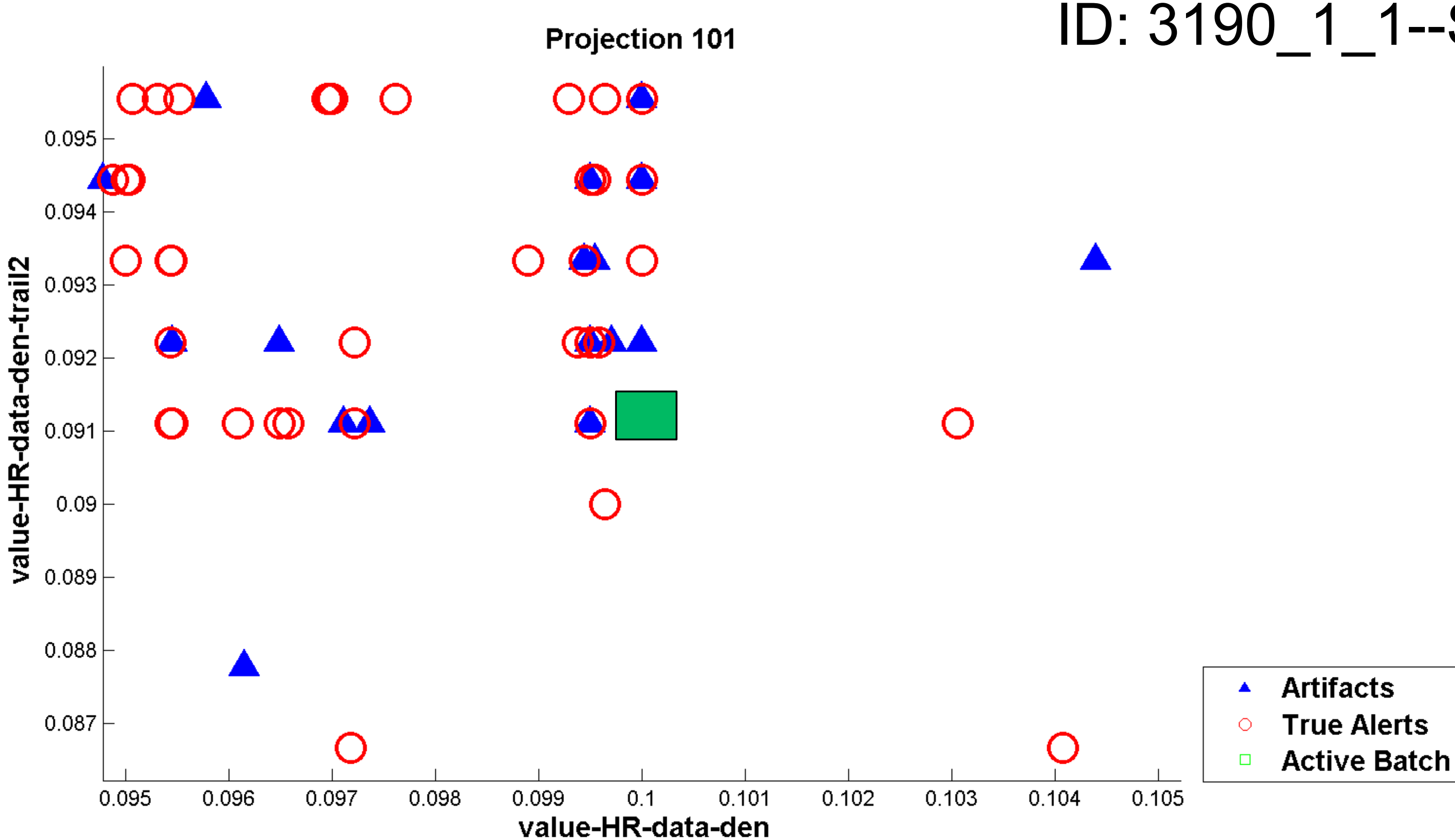
Active Sample Annotation Example 2



The experts ***need to be consulted to annotate*** the vital sign trace.

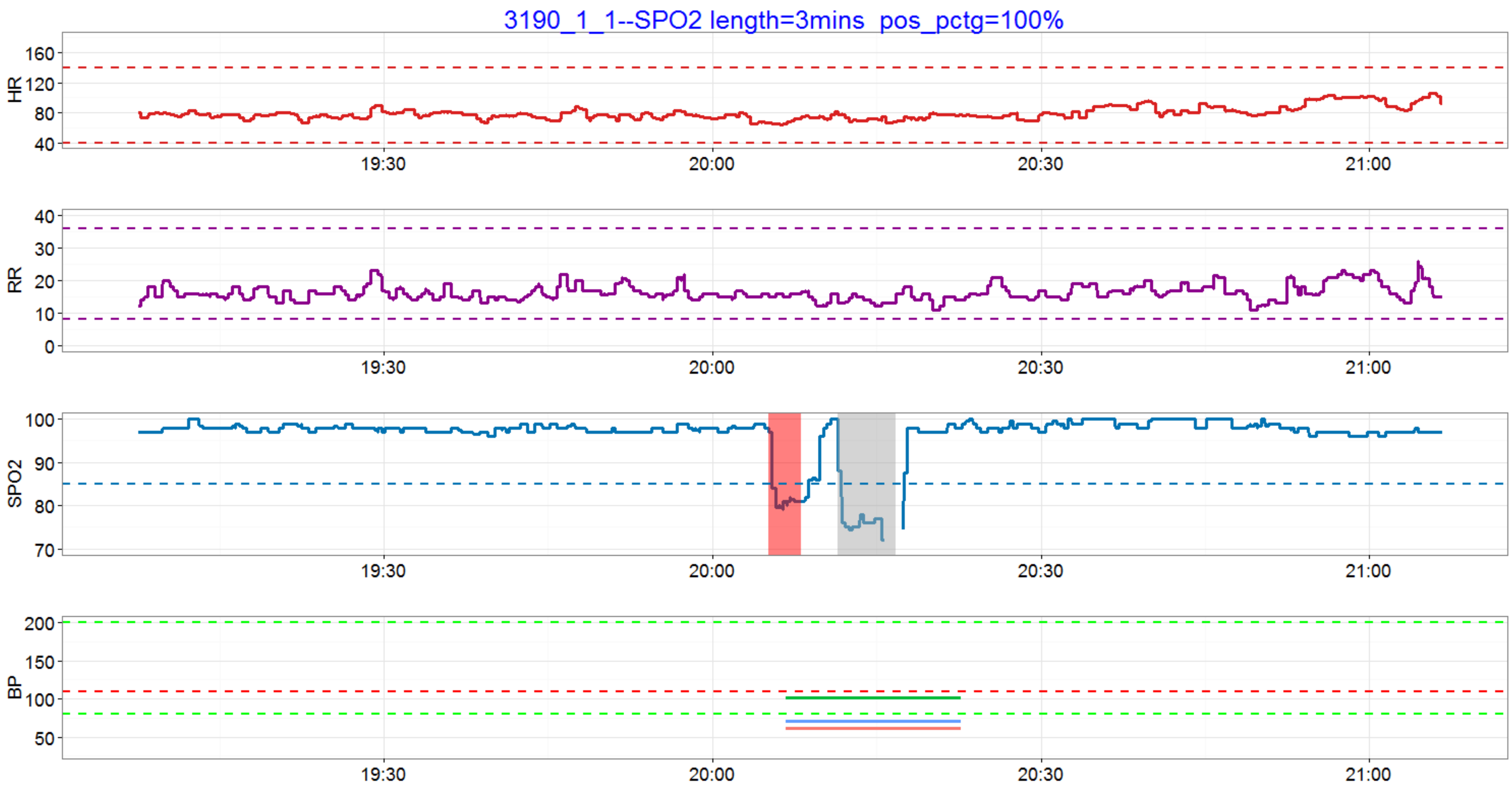
Active Sample Annotation Example 3

ID: 3190_1_1--SPO2



The sample ***can*** be somewhat confidently classified as an artifact.

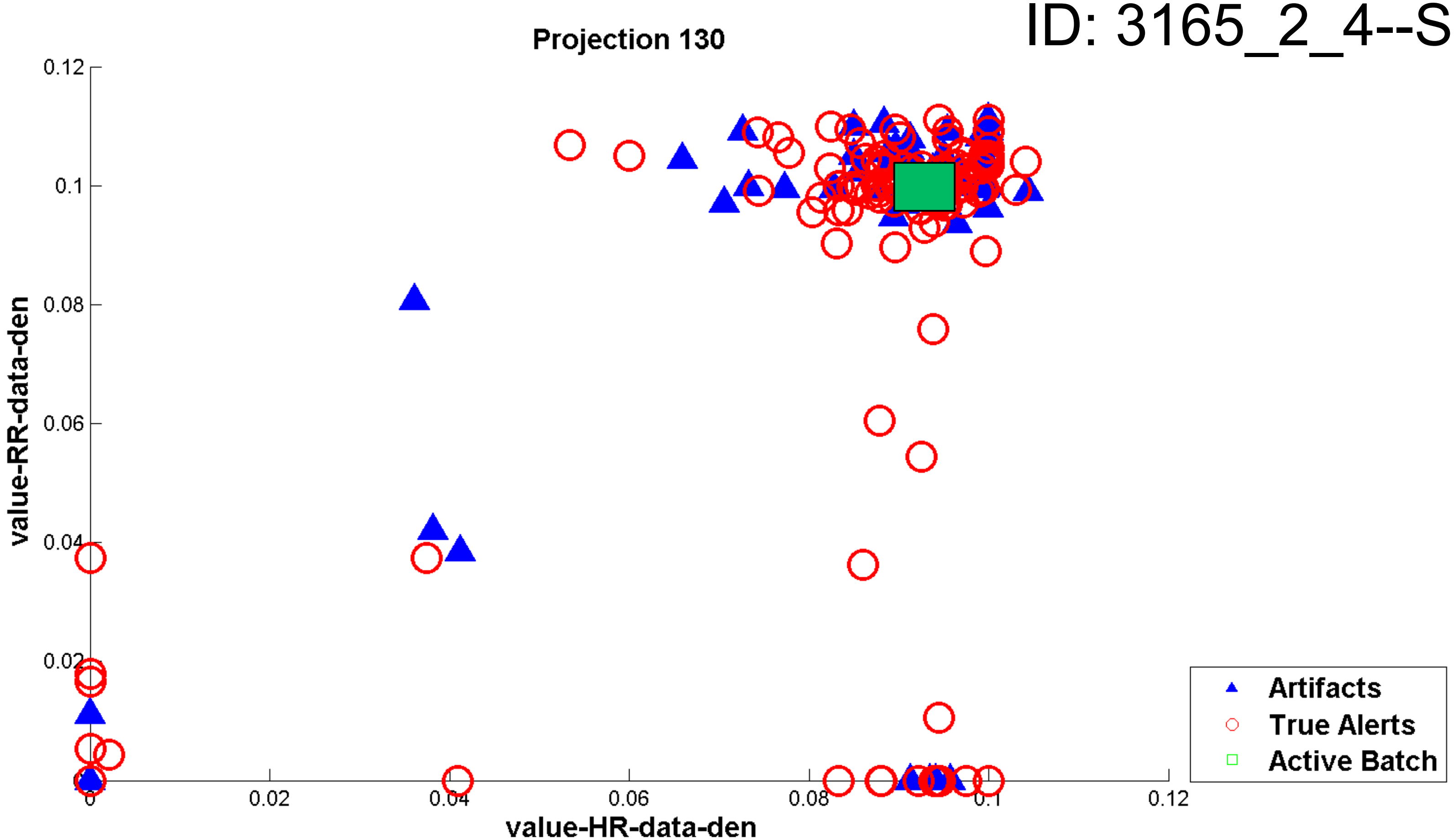
Active Sample Annotation Example 3



The experts ***might be needed to consult*** the vital sign trace.

Active Sample Annotation Example 4

ID: 3165_2_4--SPO2



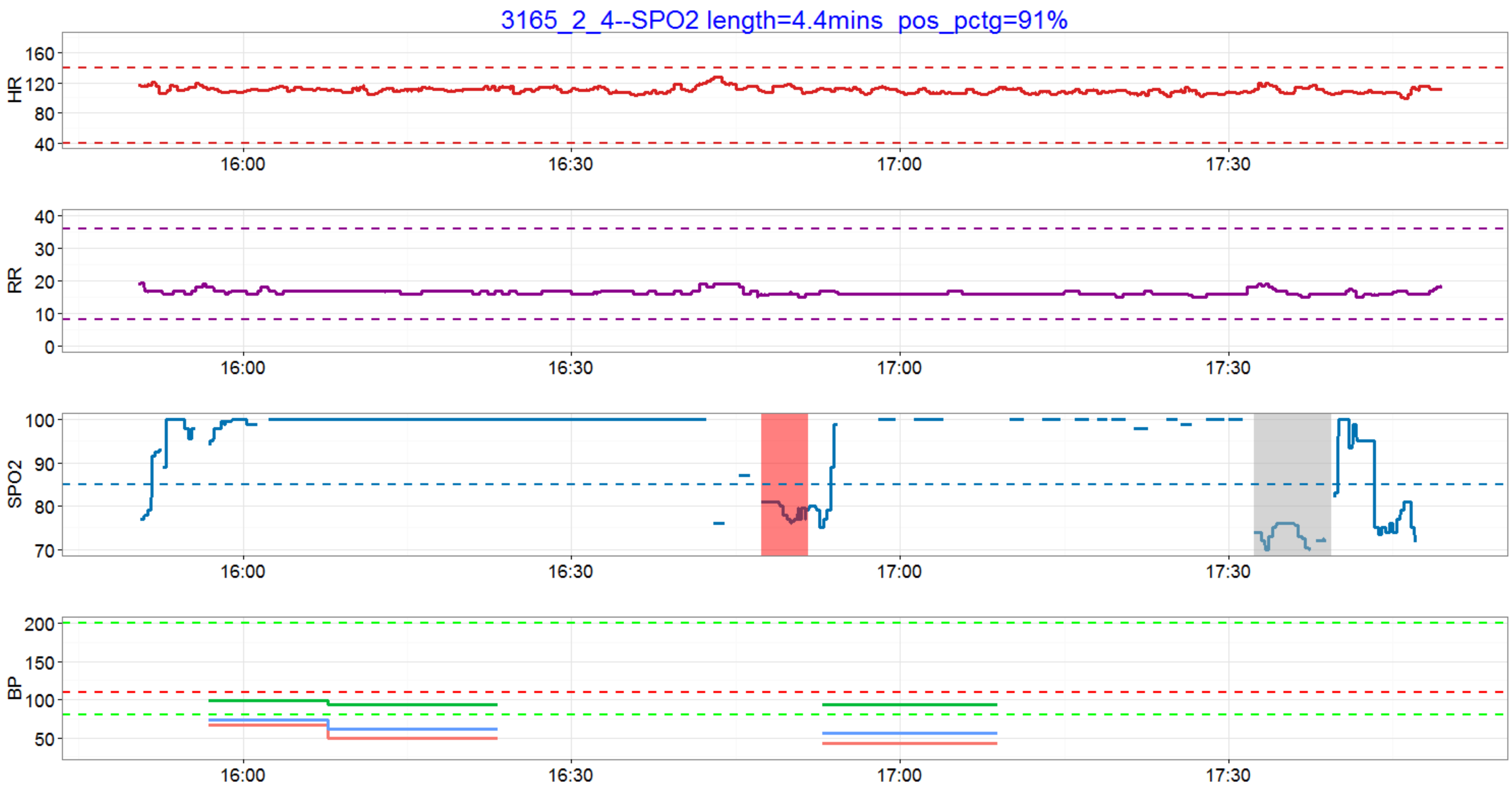
The sample ***cannot*** be confidently classified.

University of Pittsburgh

Auton Lab

Carnegie Mellon University

Active Sample Annotation Example 4



The experts ***need to consult*** the vital sign trace.

Results

Our active learning approach performance is reported as the ability to correctly adjudicate events as real or artifact by and ROC curve in a 10-fold cross validation setup vs. the percentage of expert-annotated data used in the process.

Table 1. Mean Area Under the Curve (AUC) score for 10-fold cross validation at various stages of active learning to correctly annotate events as artifact or real alerts SpO2

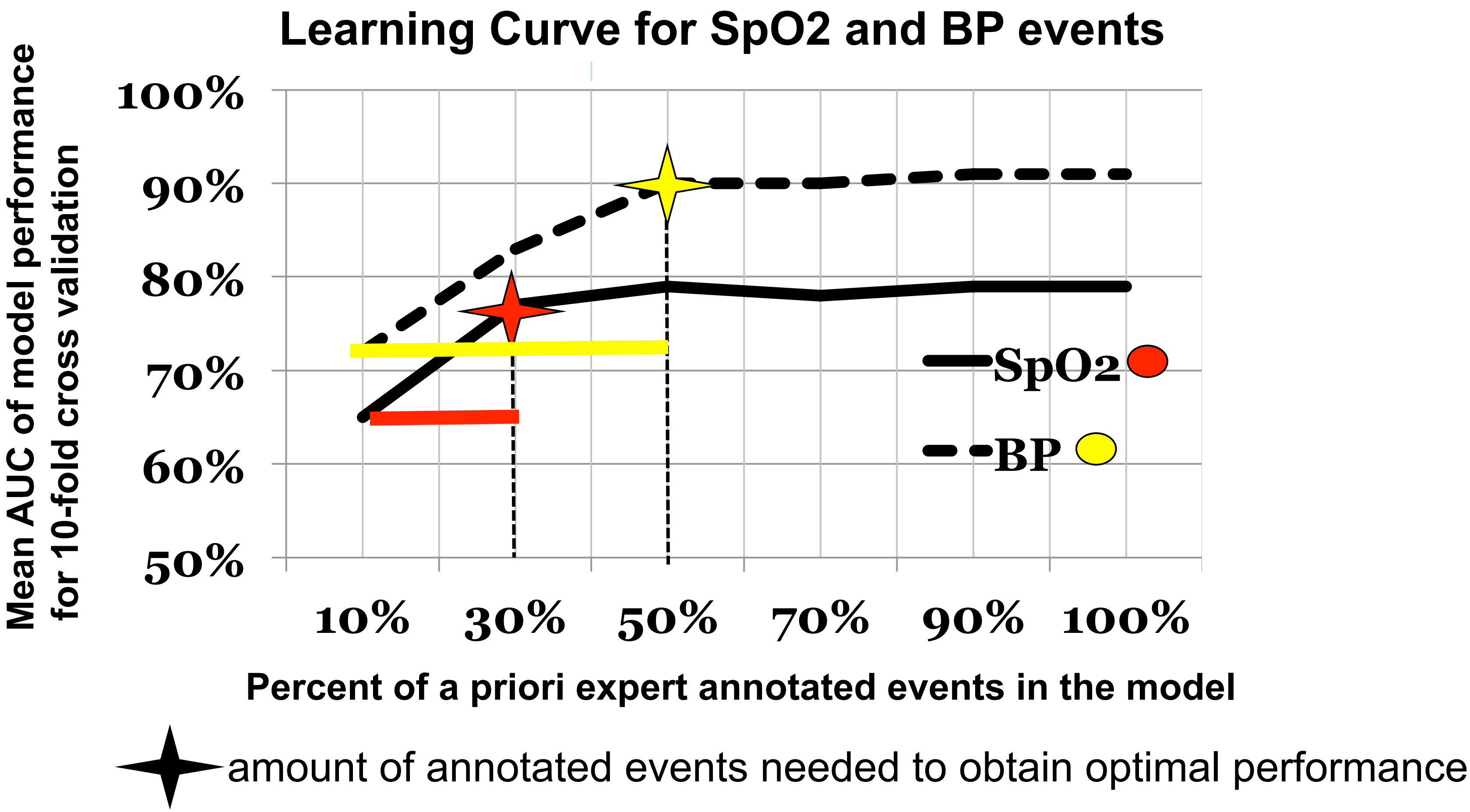
Percentage of a priori expertly annotated events in the model	SpO_2 mean AUC score on test set ± SD	Blood Pressure mean AUC score on test set ± SD
10%	65% ± 15% (84% of optimum)	72% ± 18% (80% of optimum)
30%	77% ± 11% (*optimum)	83% ± 15%
50%	79% ± 11%	90% ± 10% (*optimum)
70%	78% ± 10%	90% ± 8%
90%	79% ± 10%	91% ± 8%
100%	79% ± 10%	91% ± 8%

*denotes the stage when model performance does not substantially differ from optimal

University of Pittsburgh

Auton Lab
Carnegie Mellon University

Optimal model performance for SpO_2 was achieved at an AUC of 77%, and was reached using 30% of the expert-annotated data. Optimal BP performance was at an AUC of 90%, and was achieved using 50% of the data. When only 10% of the expert-annotated data were used for SpO_2 or BP events, each model achieved 80% of the eventual optimal performance.



Conclusions

An active learning method can reduce the amount of data needing human expert annotation when classifying monitoring events as artifact vs. real, although classification is more difficult for SpO_2. Nevertheless, further refinements of such algorithms hold promise for compiling robust datasets which can in turn to be used to build models which classify incoming monitoring data and inform clinical actions.

Next steps:

After artifact and real events are successfully classified, further machine learning can be applied to develop models which differentiate between features of stability and instability, and predict a future unstable state.

University of Pittsburgh

Marilyn Hravnak RN, PhD, ACNP-BC, FCCM FAAN [1]
Lujie Chen MS [3]
Madeline Fiterau MS,
Artur Dubrawski [3]
Gilles Clermont MD [2]
Mathieu Guillame-Bert PhD [3]
Eliezer Bose BS [1]
Michael R Pinsky MD, Dr hc [2]

University of Pittsburgh
School of Nursing [1] and Medicine [2]
Carnegie Mellon University
Robotics Institute; Auton Lab [3]