PRECISION AND RECALL FOR TIME SERIES

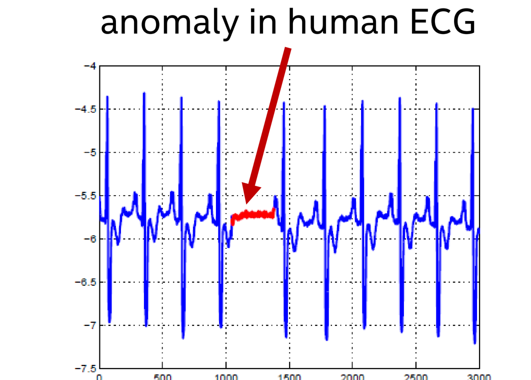
Nesime Tatbul (Intel Labs and MIT), Tae Jun Lee (Microsoft),
Stan Zdonik (Brown University), Mejbah Alam (Intel Labs), Justin Gottschlich (Intel Labs)

PROBLEM

Time Series Anomaly Detection

- **Anomalies** are patterns that do not conform to expected behavior.
- Time series anomalies are **range based**, i.e., they occur over a period of time.
- Detecting and mitigating anomalies can be safety critical.

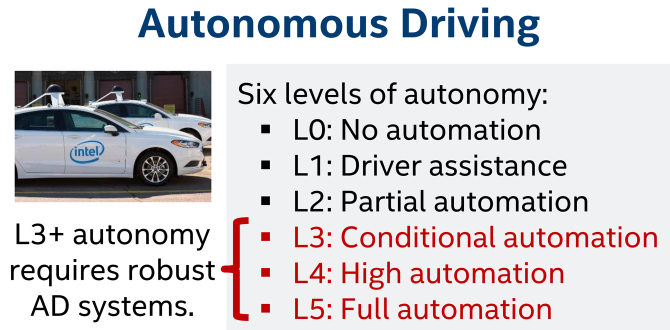
Atrial Premature Contraction anomaly in human ECG



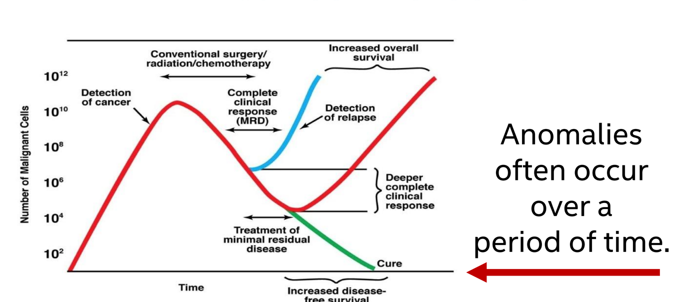
Source: Chandola et al., "Anomaly Detection: A Survey", ACM Computing Surveys, 41(3), 2009.

Application Diversity

- Applications of anomaly detection are numerous and diverse.



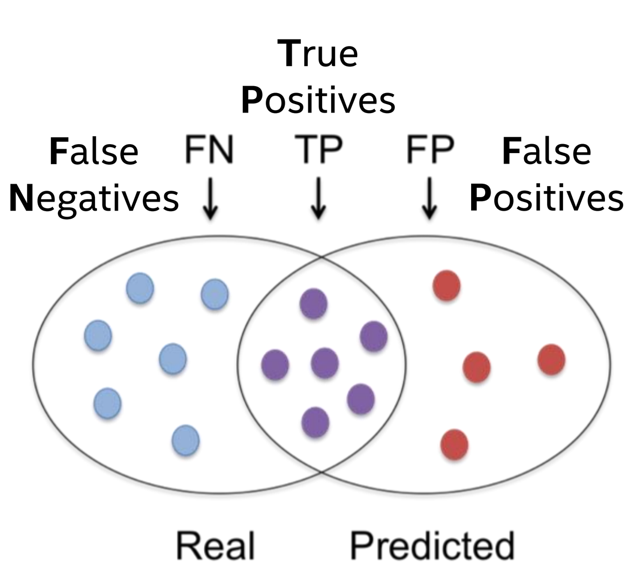
Source: Society of Automotive Engineers (SAE), National Highway and Traffic Safety Administration (NHTSA)

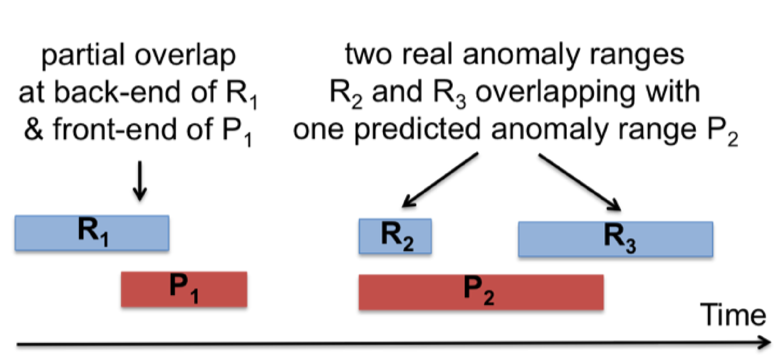


Source: http://www.vaccinogeninc.com/oncovax/science-due-diligence/overview-part-1

- There are **domain-specific preferences**.
  - Cancer detection, Real-time systems: Early response, Avoid FN!
  - Robotic defense systems: Delayed response, Avoid FP!
  - Emergency braking: Neither too early nor too late, Avoid FN!

Point-based vs. Range-based Anomalies





$Precision = TP \div (TP + FP)$
$Recall = TP \div (TP + FN)$

$Precision = ?$
$Recall = ?$

- Prior work: Classical model, Numenta model, Activity recognition metrics
- Lack of support for partial detection and flexible time bias

How to Measure Accuracy?

SOLUTION

Range-based Precision and Recall

Notation	Description
R, R_i	set of real anomaly ranges, the i^{th} real anomaly range
P, P_j	set of predicted anomaly ranges, the j^{th} predicted anomaly range
N, N_r, N_p	number of all points, number of real anomaly ranges, number of predicted anomaly ranges
α	relative weight of existence reward
$\gamma(), \omega(), \delta()$	overlap cardinality function, overlap size function, positional bias function

$$Recall_T(R,P) = \frac{\sum_{i=1}^{N_r} Recall_T(R_i,P)}{N_r}$$

$$Recall_T(R_i,P) = \alpha \times ExistenceReward(R_i,P) + (1-\alpha) \times OverlapReward(R_i,P)$$

$$ExistenceReward(R_i,P) = \begin{cases} 1, \text{if } \sum_{j=1}^{N_p} |R_i \cap P_j| \geq 1 \\ 0, \text{otherwise} \end{cases}$$

$$OverlapReward(R_i,P) = CardinalityFactor(R_i,P) \times \sum_{j=1}^{N_p} \omega(R_i, R_i \cap P_j, \delta)$$

$$CardinalityFactor(R_i,P) = \begin{cases} 1 & \text{, if } R_i \text{ overlaps with at most one } P_j \in P \\ \gamma(R_i,P), & \text{otherwise} \end{cases}$$

$$Precision_T(R,P) = \frac{\sum_{i=1}^{N_p} Precision_T(R,P_i)}{N_p}$$

$$Precision_T(R,P_i) = CardinalityFactor(P_i,R) * \sum_{j=1}^{N_r} \omega(P_i, P_i \cap R_j, \delta)$$

- Our model **subsumes** the classical point-based model, when:
  - all ranges are represented as unit-size ranges, and
  - $\alpha=0$, $\gamma()=1$, $\omega()$ is as below, and $\delta()$ = Flat.

Customization Examples

Overlap Size Function

```
function ω(AnomalyRange, OverlapSet, δ)
    MyValue ← 0
    MaxValue ← 0
    AnomalyLength ← length(AnomalyRange)
    for i ← 1, AnomalyLength do
        Bias ← δ(i, AnomalyLength)
        MaxValue ← MaxValue + Bias
        if AnomalyRange[i] in OverlapSet then
            MyValue ← MyValue + Bias
    return MyValue/MaxValue
```

Positional Bias Function

```
function δ(i, AnomalyLength)                ▷ Flat bias
    return 1
function δ(i, AnomalyLength)                ▷ Front-end bias
    return AnomalyLength - i + 1
function δ(i, AnomalyLength)                ▷ Back-end bias
    return i
function δ(i, AnomalyLength)                ▷ Middle bias
    if i ≤ AnomalyLength/2 then
        return i
    else
        return AnomalyLength - i + 1
```

Cancer Detection:
- Set $\delta()$ = Front-end, $\beta = 2$

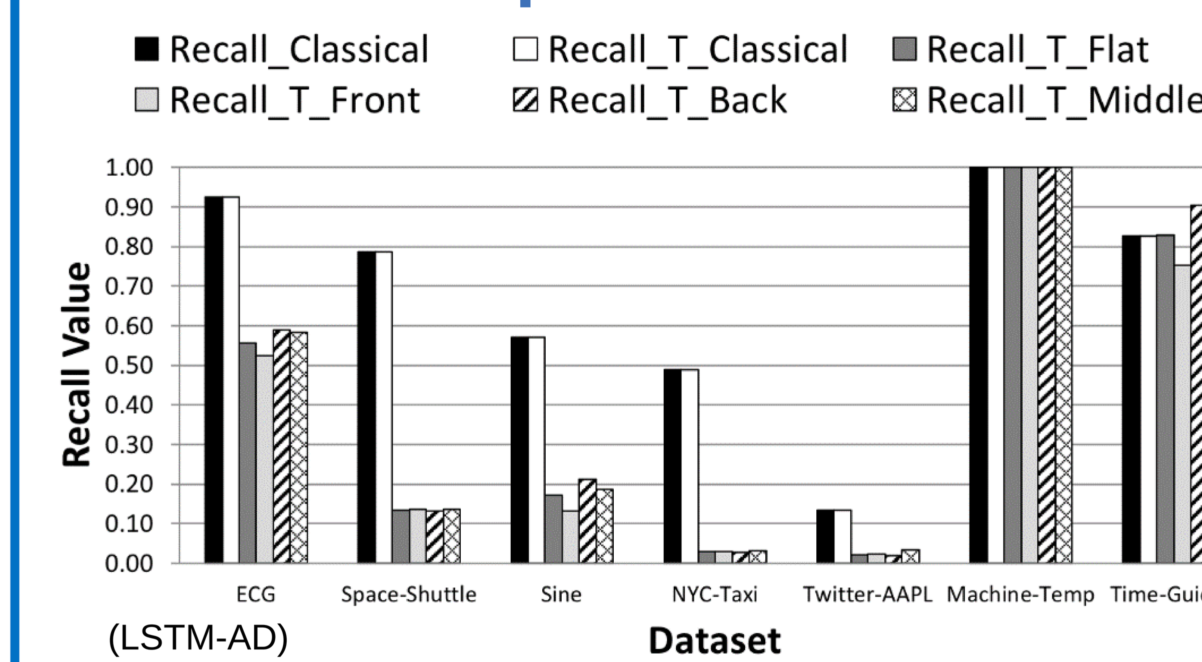
Robotic Defense:
- Set $\delta()$ = Back-end, $\beta = 0.5$

Emergency Braking:
- Set $\delta()$ = Middle, $\beta = 1.5$

Customizable Precision and Recall

RESULTS

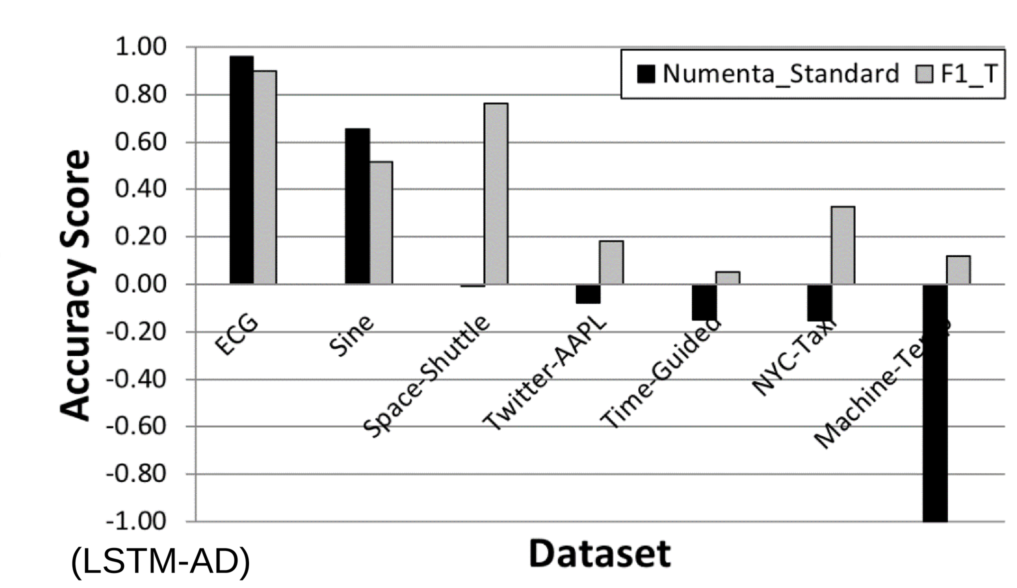
Comparison to Classical Model



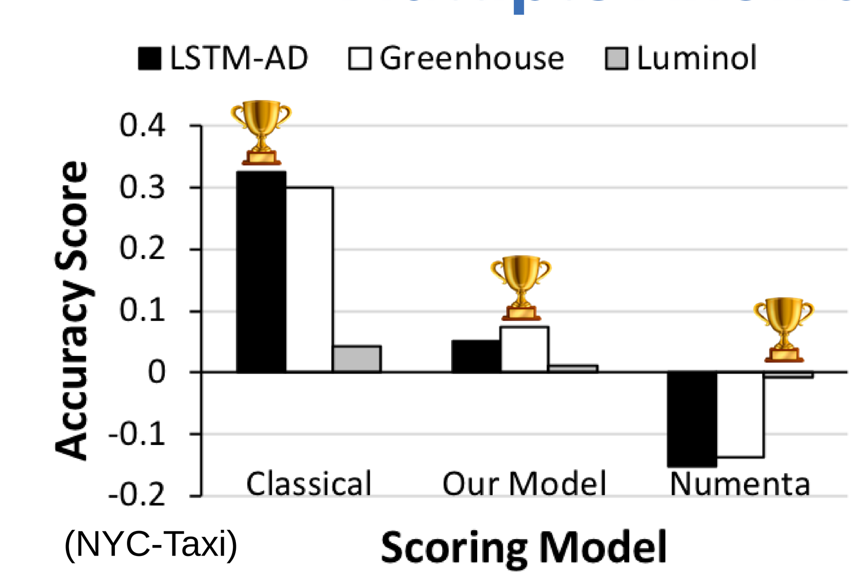
- Our model:
  - subsumes the classical model.
  - is sensitive to positional bias.
- Results are similar for Precision and F-Score.

Comparison to Numenta Model

- Our model can:
  - mimic Numenta by setting $\delta()$ = Front-end.
  - catch additional intricacies.
- Results are similar for all Numenta app profiles.



Multiple Anomaly Detectors



- Our model is more effective in:
  - evaluating multiple anomaly detectors.
  - capturing subtleties in data.
- Results are similar for other datasets.

Future Directions

- New training strategies for range-based anomaly detection
- Exploring use in other time series classification tasks and applications

More Information

Watch: https://www.youtube.com/watch?v=K5f-dUBiQP4
Read: https://arxiv.org/abs/1803.03639/
Use: https://github.com/IntelLabs/TSAD-Evaluator/

Expressive, Flexible, Extensible