A New Histogram-based Visualization Tool for Analyzing Anomaly Detection Algorithm Performance

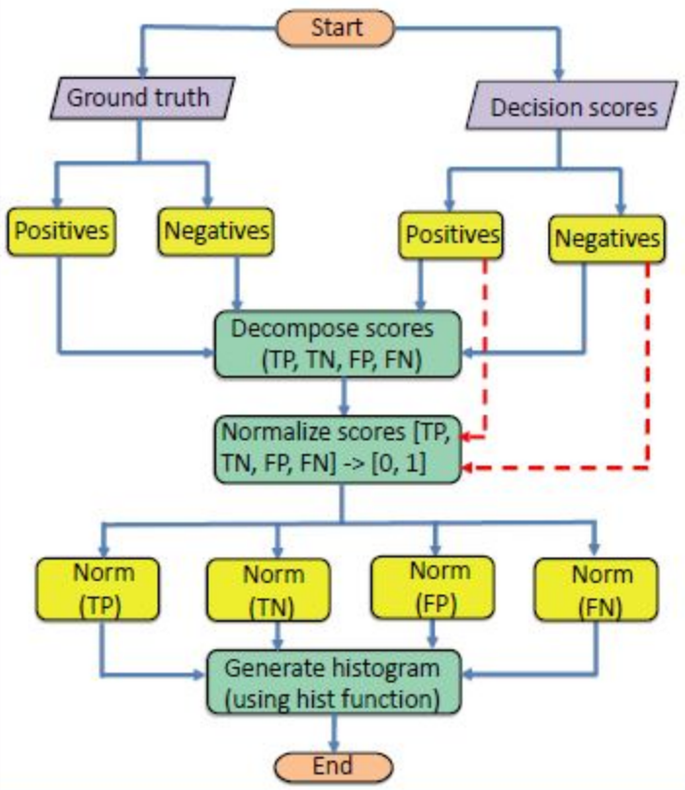
Emmanuel Aboah Boateng, J.W. Bruce
Department of Electrical and Computer Engineering

1. INTRODUCTION

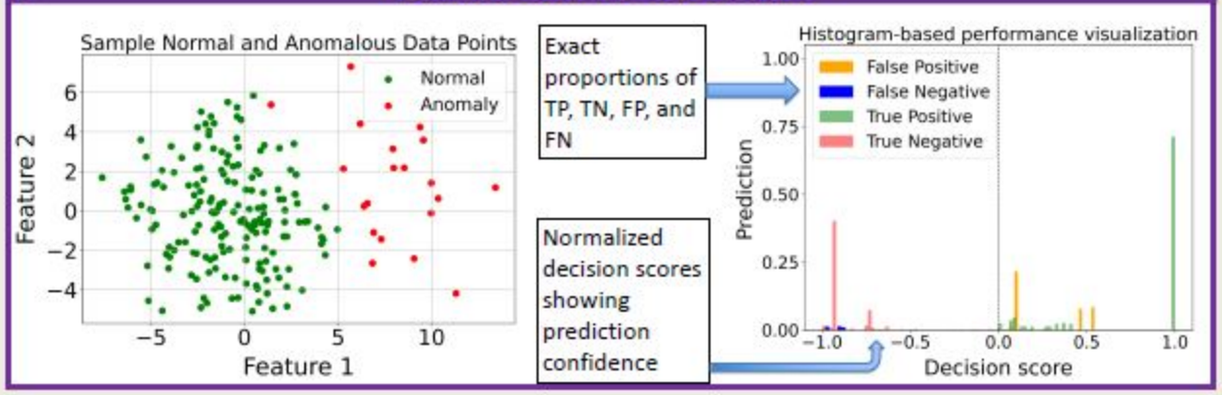
- Performance visualization of anomaly detection algorithms enables researchers to highlight trends and outliers in anomaly detection models results to gain intuitive understanding of detection models.
- Generally, anomaly detection algorithms produce negative and positive decision scores, representing normal and anomalous data points.
- Previous work relies on using histograms based on positive and negative scores for visualizing anomaly detection algorithms' performance [1].
- This work proposes a new histogram-based visualization approach that provides a better understanding of detection algorithms' performance by revealing the exact proportions of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) values.
- The proposed method also reveals the detection confidence of detection algorithms.

2. FLOWCHART OF PROPOSED METHOD

Input 1 ← {Decision scores} ∈ R
Input 2 ← {Ground truth} ∈ $[-1,1]$



3. BENEFITS OF PROPOSED METHOD

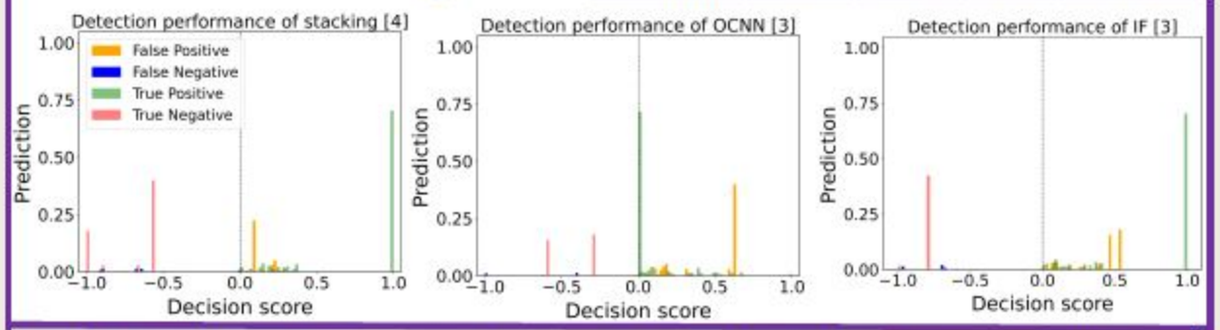


4. RESULTS

- Previous work such as [2] relies on table 1 for detection algorithms performance evaluation and comparison.
- How do we know the true performance of the models in table 1? **New Histogram-based approach**

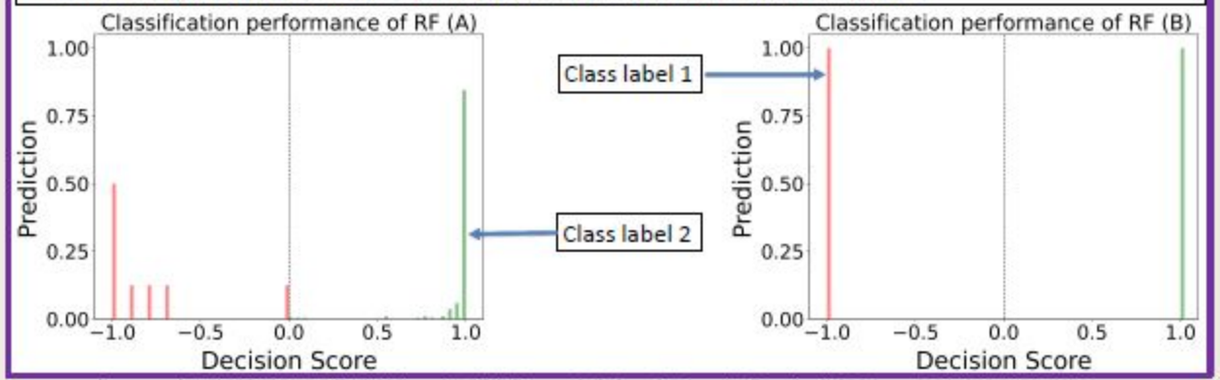
Table 1: Performance comparison of anomaly detection models using dataset in [3]

Method	Accuracy	Precision	Recall	F1-score
OCNN [3]	0.91	0.90	0.91	0.91
IF [3]	0.91	0.91	0.90	0.91
Stacking [4]	0.92	0.92	0.92	0.92



Extension of the visualization approach to supervised machine learning involving binary classification

- Random Forest (RF) classifier A is plotted using the histogram-based approach in this work, whereas RF (B) is plotted using the visualization approach in [1]. RF (A) and RF (B) have perfect accuracies.



5. INSTALLATION/USAGE

- The histogram-based visualization software is open-source and can be installed by:

```
$ pip install hist-plot
```

- Usage:

```
from hist_plot import AnomalyScoreHist
fig = AnomalyScoreHist(dec_score, g_truth)
fig.plot_hist(fig_name)
```

Where:
- dec_score: decision score output of anomaly detection model
- g_truth: ground truth label
- fig_name: optional name for the plot

6. CONCLUSION

- This work has introduced a better way of visualizing and analyzing anomaly detection algorithm performance using a histogram-based approach.
- Results show that the proposed method provides a better meaning of detection algorithm performance as compared to previous work.
- The proposed method can be applied to the performance visualization of supervised machine learning models involving binary classification.

7. REFERENCES

[1] R. Chalapathy, et. al, "Anomaly detection using one-class neural networks." *arXiv preprint arXiv*:1802.06360 (2018)
[2] E. Aboah Boateng, "Anomaly Detection for Industrial Control Systems Based on Neural Networks with One-Class Objective Function." *Proc. of Student Research and Creativity Inquiry Day* 5 (2021).
[3] E. Aboah Boateng, and J.W. Bruce. "Unsupervised Machine Learning Techniques for Detecting PLC Process Control Anomalies." *Journal of Cybersecurity and Privacy* 2.2 (2022): 220-244..
[4] E. Aboah Boateng, and J.W. Bruce. "Unsupervised Ensemble Methods for Anomaly Detection in PLC-based Process Control." IEE Transactions (under review).

8. ACKNOWLEDGEMENTS

Support from funds provided by CMR and the state of Tennessee to TN Tech in recognition of the University Carnegie Classification, R2, is appreciated.