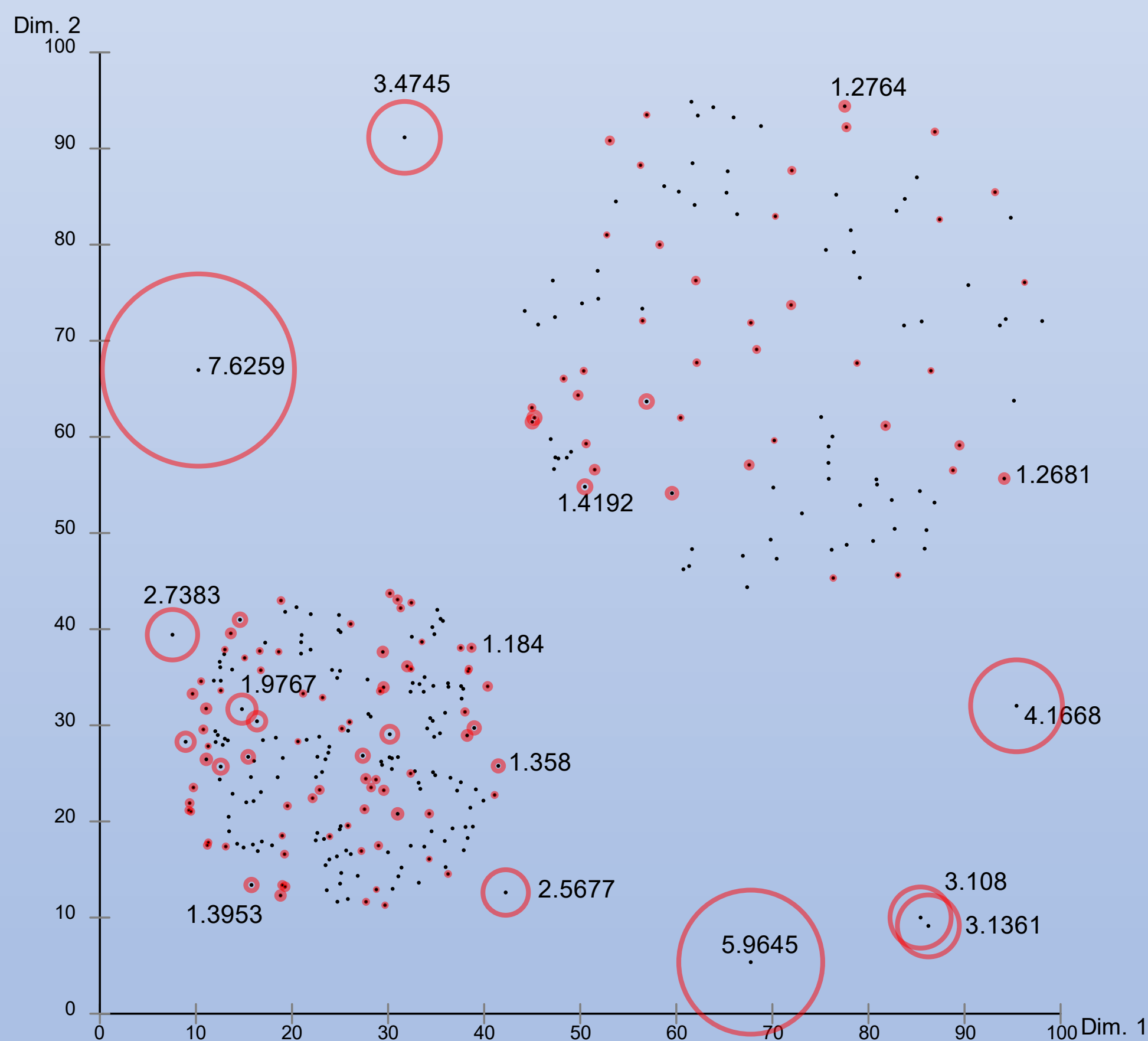


# Outlier/Anomaly Detection

Sayyed-Ahmad Naghavi-Nozad and Abdul Tauqeer  
University of Nevada, Las Vegas

## Introduction

In data analysis, outliers or anomalies (or even in some cases, novelties) are those rare items that deviate significantly from a well-known notion of normal behavior. Such examples may cause suspicions that these data samples might be deliberately created out of a discrepant mechanism or not be in harmony with the remainder of data. Outlier detection refers to methodologies that aim at discovering such rare and exceptional examples in data [1].

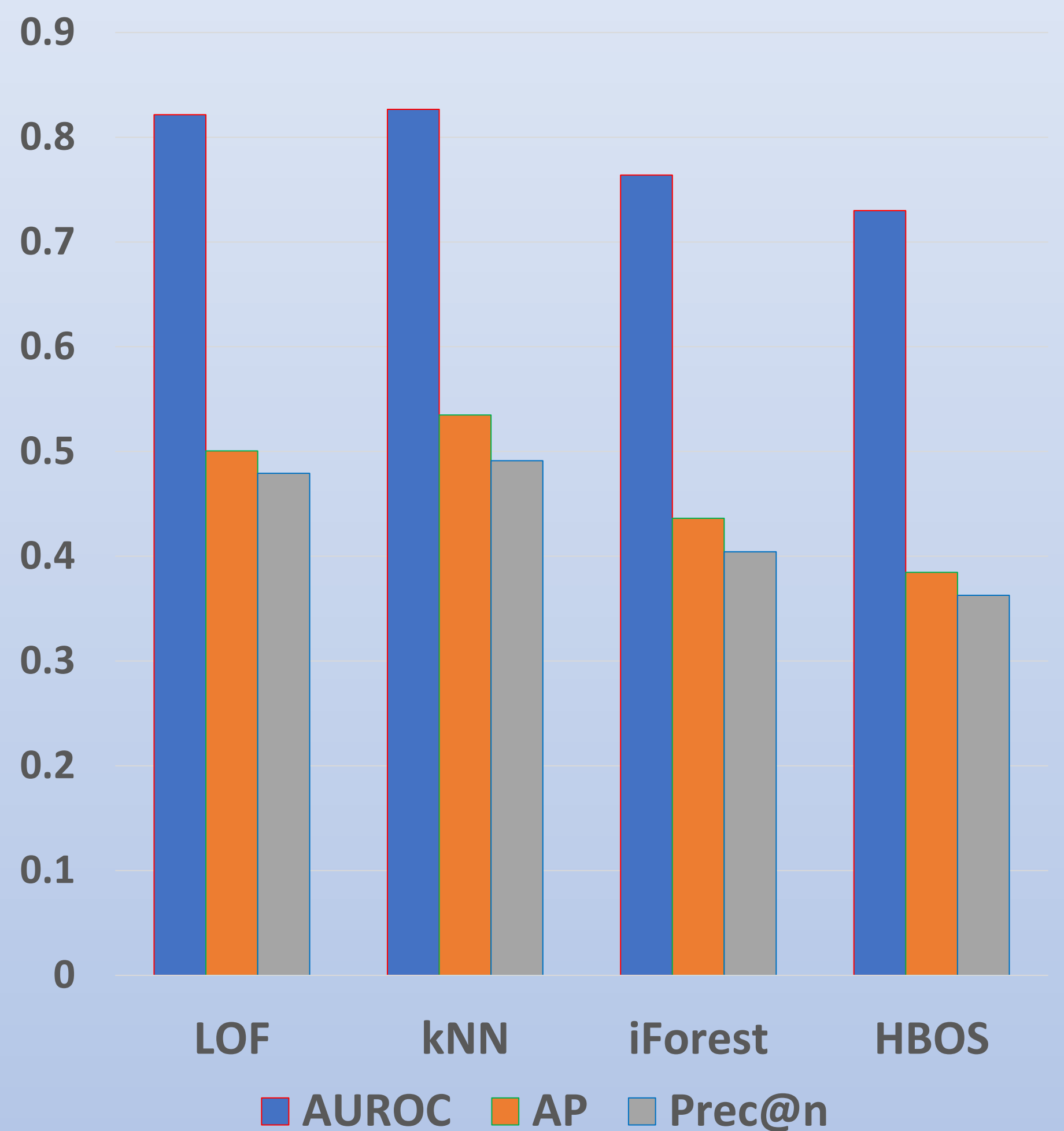


## Methods

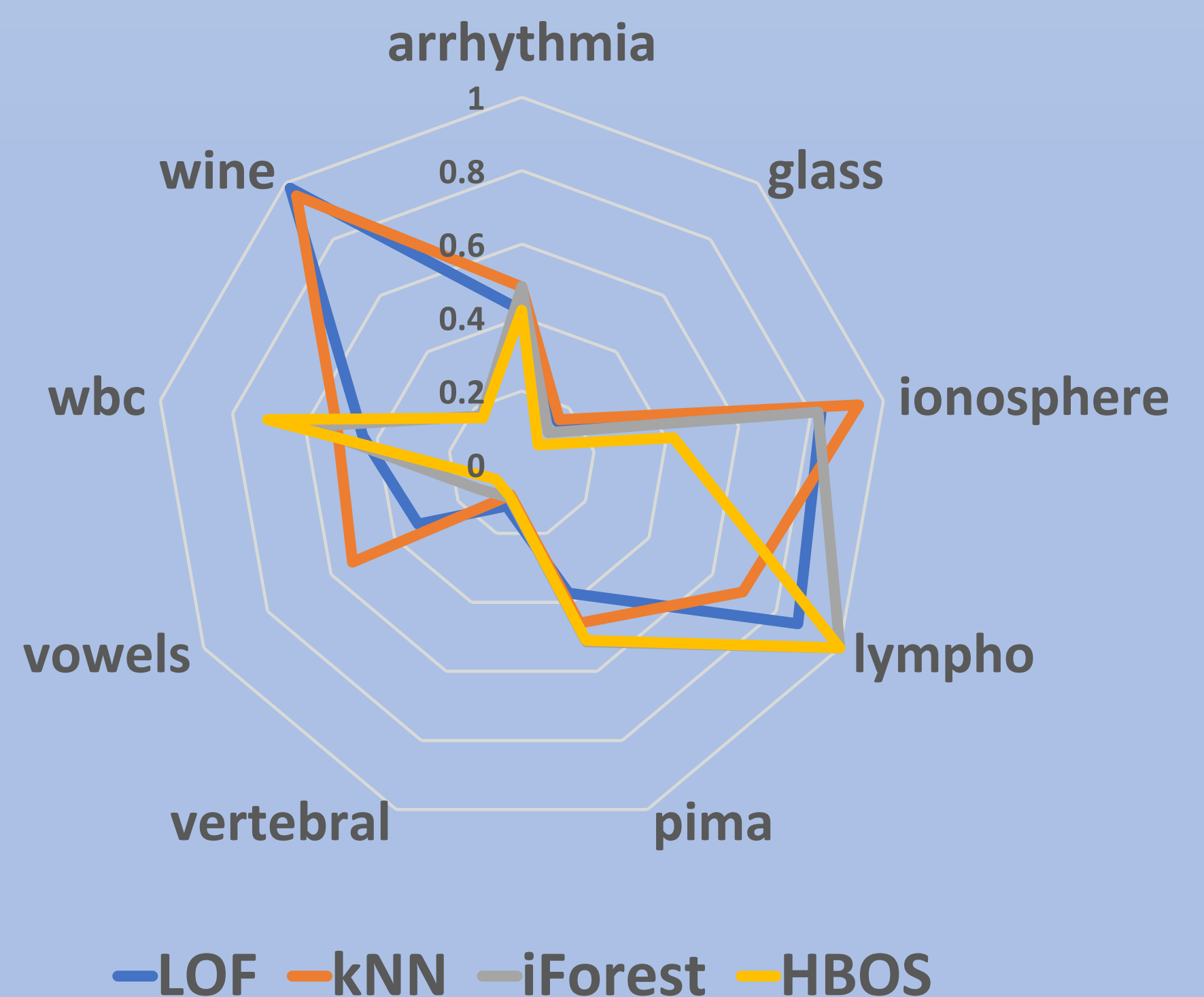
- **LOF:** Each sample gets an anomaly score entitled Local Outlier Factor. LOF for a point is higher, then it is more probable for it to be an anomaly. It computes the density fluctuation of every point w.r.t. its surrounding points, and after a normalization, the outlier scores are acquired [2].
- **kNN:** For every data element, it considers the distances to its neighboring points and derives the anomaly score out of them. Three different strategies including the largest distance, average distance, and median distance could be adopted [3].
- **iForest:** It strives to isolate every observation by choosing attributes at random and then finding a random split point; then, it starts the dividing process and building the isolation trees. The path length to reach a point in isolation trees will indicate the anomaly score [4].
- **HBOS:** Histogram-based outlier detection is a simple and fast method that presumes the data attributes being independent, and builds the outlier scores upon the respective histograms [5].

## Results

Average accuracies for different methods over 9 datasets



AP results over different datasets



## References

- [1] Hodge, Victoria, and Jim Austin. "A survey of outlier detection methodologies." *Artificial intelligence review* 22.2 (2004): 85-126.
- [2] Breunig, Markus M., et al. "LOF: identifying density-based local outliers." *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 2000.
- [3] Ramaswamy, Sridhar, Rajeev Rastogi, and Kyuseok Shim. "Efficient algorithms for mining outliers from large data sets." *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 2000.
- [4] Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation-based anomaly detection." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6.1 (2012): 1-39.
- [5] Goldstein, Markus, and Andreas Dengel. "Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm." *KI-2012: poster and demo track 9* (2012).