PERFORMANCE OF OUTLIER DETECTION ON SMARTWATCH DATA IN SINGLE AND MULTIPLE PERSON ENVIRONMENTS

An analysis of the performance of different outlier detection methods on consumer-grade wearable data in environments with single and multiple subjects

AUTHOR

Luuk Wubben

SUPERVISORS

Arman Naseri Jahfari, Ramin Ghorbani

RESPONSIBLE PROFESSOR

Dr. David Tax

AFFILIATIONS

Delft University of Technology, Faculty of Electrical Engineering, Mathematics, and Computer Science (EEMCS)

INTRODUCTION

- Outlier detection is an essential part of modern systems. Examples include being able to predict battery failure from voltage data [1] and heart disease from heart rate data [2].
- Not a lot of research is dedicated to the pros and cons of outlier detection when used in an environment with a single subject versus multiple subjects.
- Research on this topic is valuable, as model parameters on one environment will not work on the other. Furthermore, subject data can be hard to combine (e.g. different dimensions) or separate (e.g. unlabelled).

RESEARCH QUESTION

"Do outlier detection methods perform better in a single person environment, compared to in a multiple person environment"

In the context of this research question, a single person environment is an environment or data set with one subject, while a multiple person environment has 2 or more subjects in its environment or data set.

METHODOLOGY

- Split heart rate and step count time series data into 6 hour windows and summarize with features.
- Define outliers based on one of two definitions.
  - Using another subject as outlier.
  - Defining outlier windows using the formula:

$$outlier\_indicator = \sqrt{\sum_{f \in F} f^2}$$

  Where:

  f = a feature in this window's feature space

  F = this window's feature space
- Implement, optimize, and test a Gaussian Mixture Model (GMM) and DBSCAN in the two environments using the two outlier definitions.
- Collect data on accuracy for both, Area Under the Curve (AUC) score for GMM, and Silhouette Coefficient score for DBSCAN.

ANALYSIS

- Both algorithms were tested in four scenarios (shown in Figure 1).
- DBSCAN generally outperformed the GMM.
- DBSCAN shows consistent accuracy with little deviation regardless of environment or outlier definition (Fig 2-5).
- GMM shows improvement with distance on between subjects outliers (Fig 4,5) and consistency with high standard deviation on within subject outliers (Fig 2,3).
- Exclusion of heart rate or step count shows step count is responsible for most of the standard deviation in results. Accuracy does not drop in most cases when either is excluded.
- Due to low deviation in heart rate, a GMM using between subjects outliers performs worse when excluding step count (~20% accuracy drop).

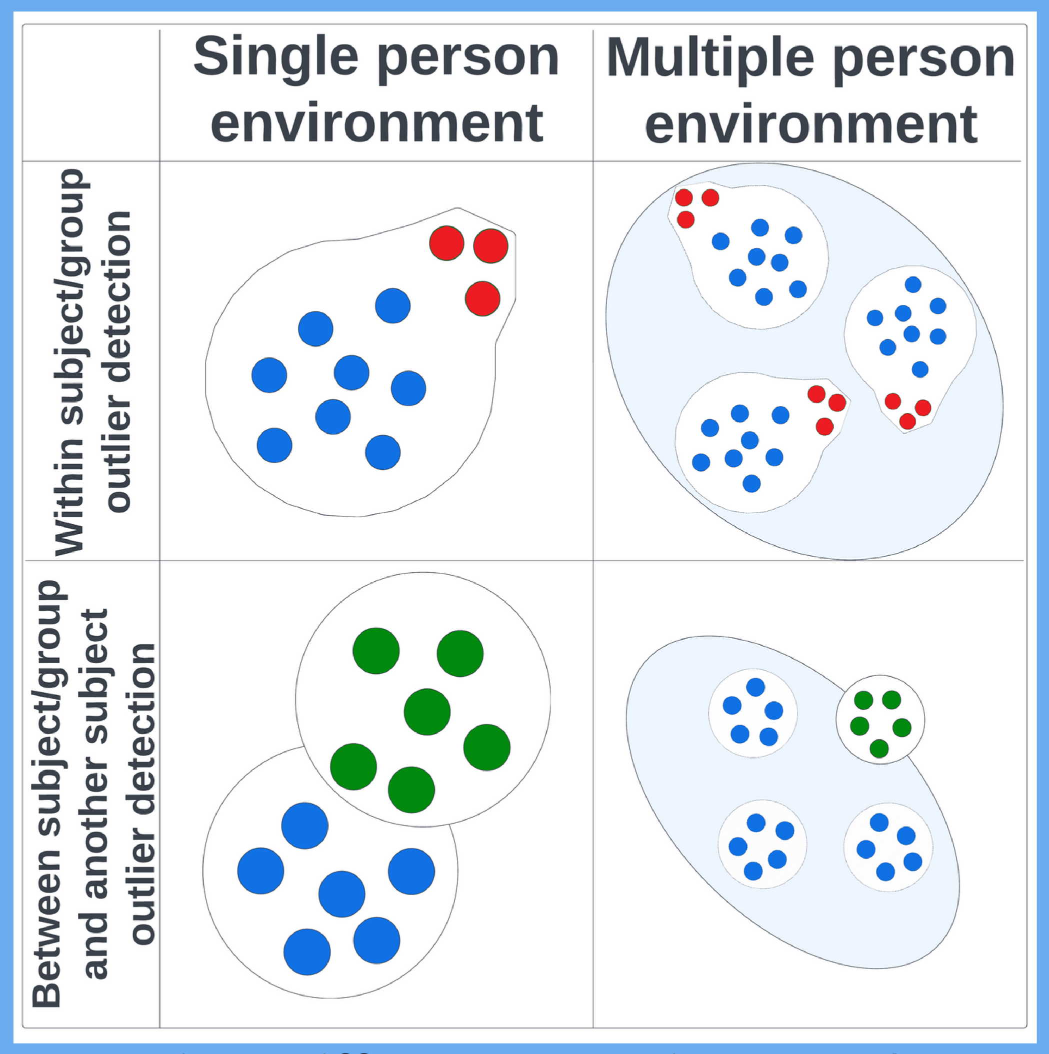

Fig 1: Different scenarios tested

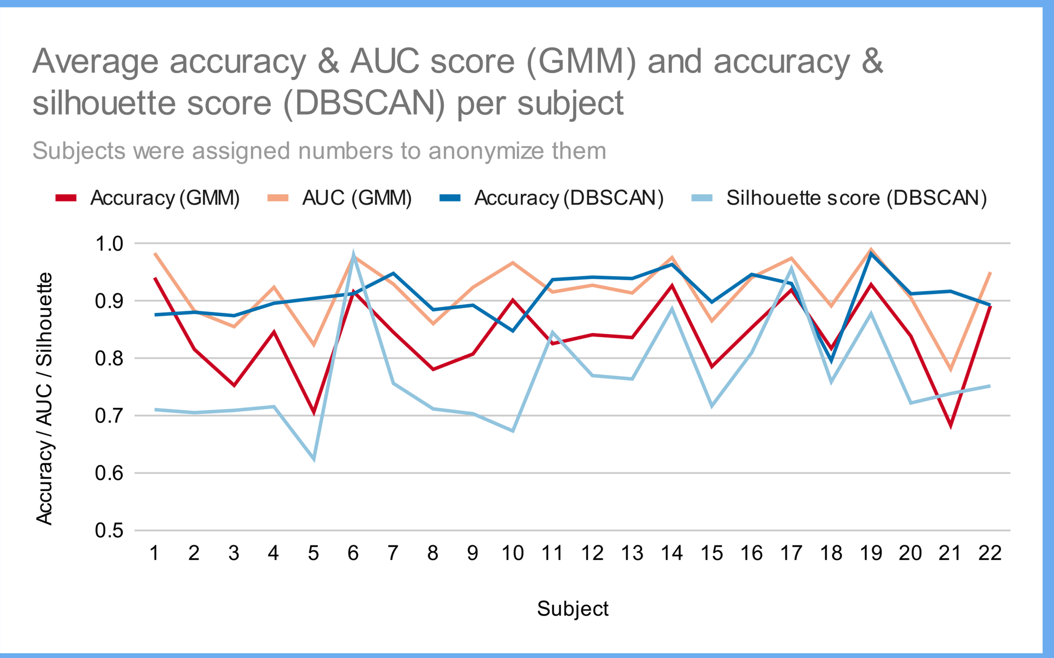

Fig 2: Performance of GMM and DBSCAN in a single person environment on within subject outliers

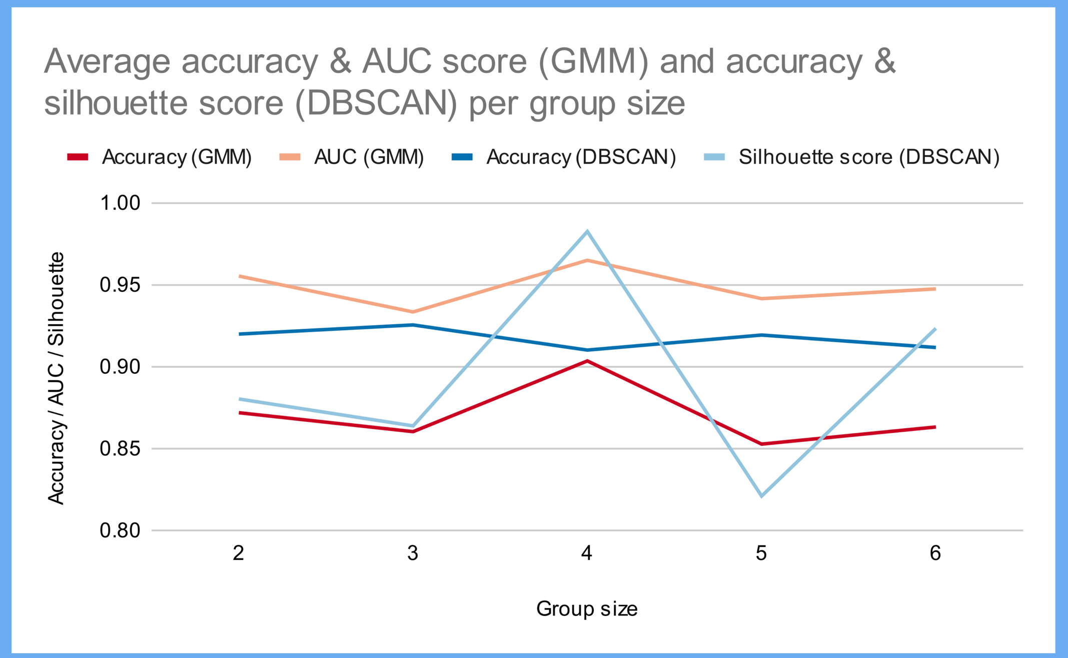

Fig 3: Performance of GMM and DBSCAN in a multiple person environment on within subject outliers

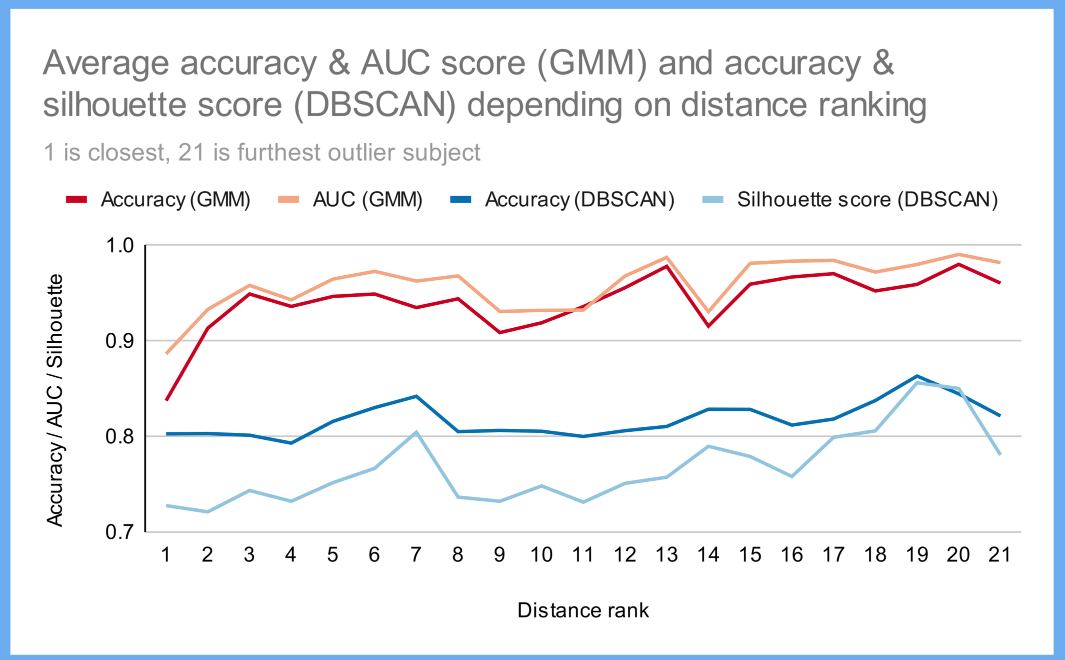

Fig 4: Performance of GMM and DBSCAN in a single person environment on between subjects outliers

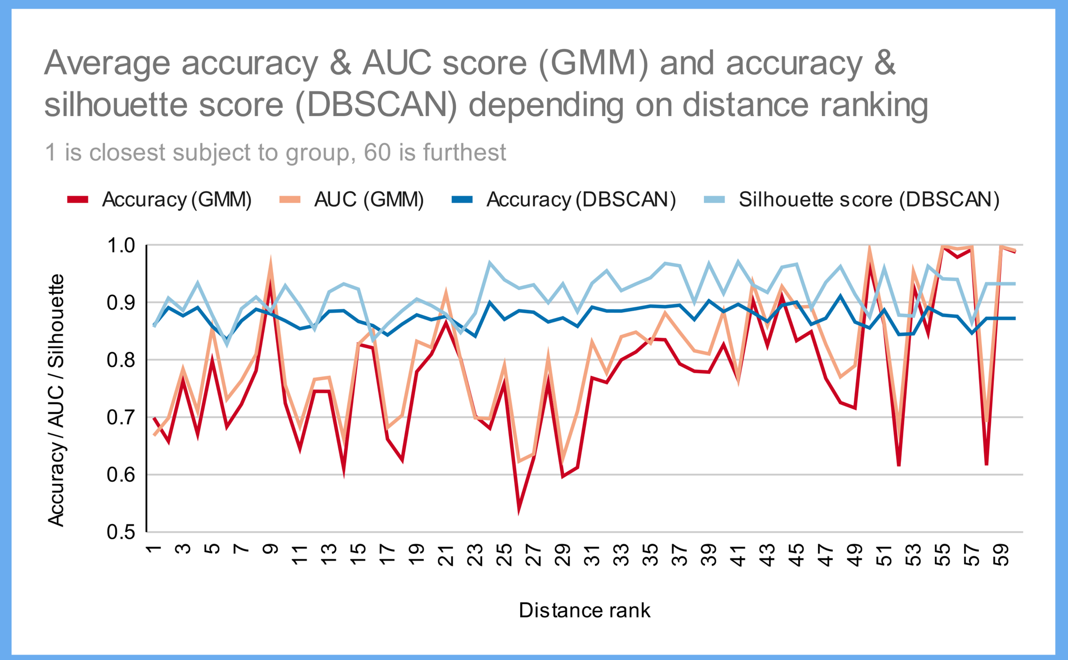

Fig 5: Performance of GMM and DBSCAN in a multiple person environment on between subjects outliers

RESULTS/FINDINGS

- For the chosen algorithms, there is a difference in performance between the two environments.
- This difference does not appear to be caused by the amount of subjects in the environment, but the outlier definition used.
- Thus, outlier detection methods can perform better in a single person environment, but it is not entirely dependent on the environment.
- A GMM struggles to differentiate subjects which are similar. DBSCAN does not.
- DBSCAN has much more consistent performance across the scenarios, with little to no deviation from its accuracy.
- Step count causes most deviation in accuracy. Removal could improve average performance unless using a GMM and between subjects outliers.

FUTURE WORK

- More algorithms should be tested to determine whether the conclusions apply to the bigger picture or are specific to this setup.
- Algorithms should be tested on data without windowing, to test efficacy in detecting small scale outliers.
- Test on data with labelled outliers, as the chosen outlier definitions might not be reflective of real world outliers.

RELATED LITERATURE

[1] Zonghai Chen, Ke Xu, Jingwen Wei, and Guangzhong Dong. Voltage fault detection for lithium-ion battery pack using local outlier factor. MEASUREMENT, 146:544—556, NOV 2019.

[2] Nan Yue and Stephan Claes. Wearable sensors for smart abnormal heart rate detection during skiing. Internet Technology Letters, 4(3):e230, 2021.

TU Delft