A study of how outlier detectors can accurately authenticate multiple persons using the heart rate from consumer-grade wearables

Author
Matei Chirita - M.Chirita@student.tudelft.nl

Professor
David Tax - D.M.J.Tax@tudelft.nl

Supervisors
Arman Naseri Jahfari - A.NaseriJahfari@tudelft.nl
Ramin Ghorbani - R.Ghorbani@tudelft.nl

1. Introduction

Wearables

- Accessible to many people
- Already very good for health and fitness tracking

Person identification

- Many available methods (accounts, pins, biometrics, etc)
- Very useful and improves the quality of life
- Can help in life-saving situations

Person identification using wearables

- A fast and convenient way of authentication

2. Research Question

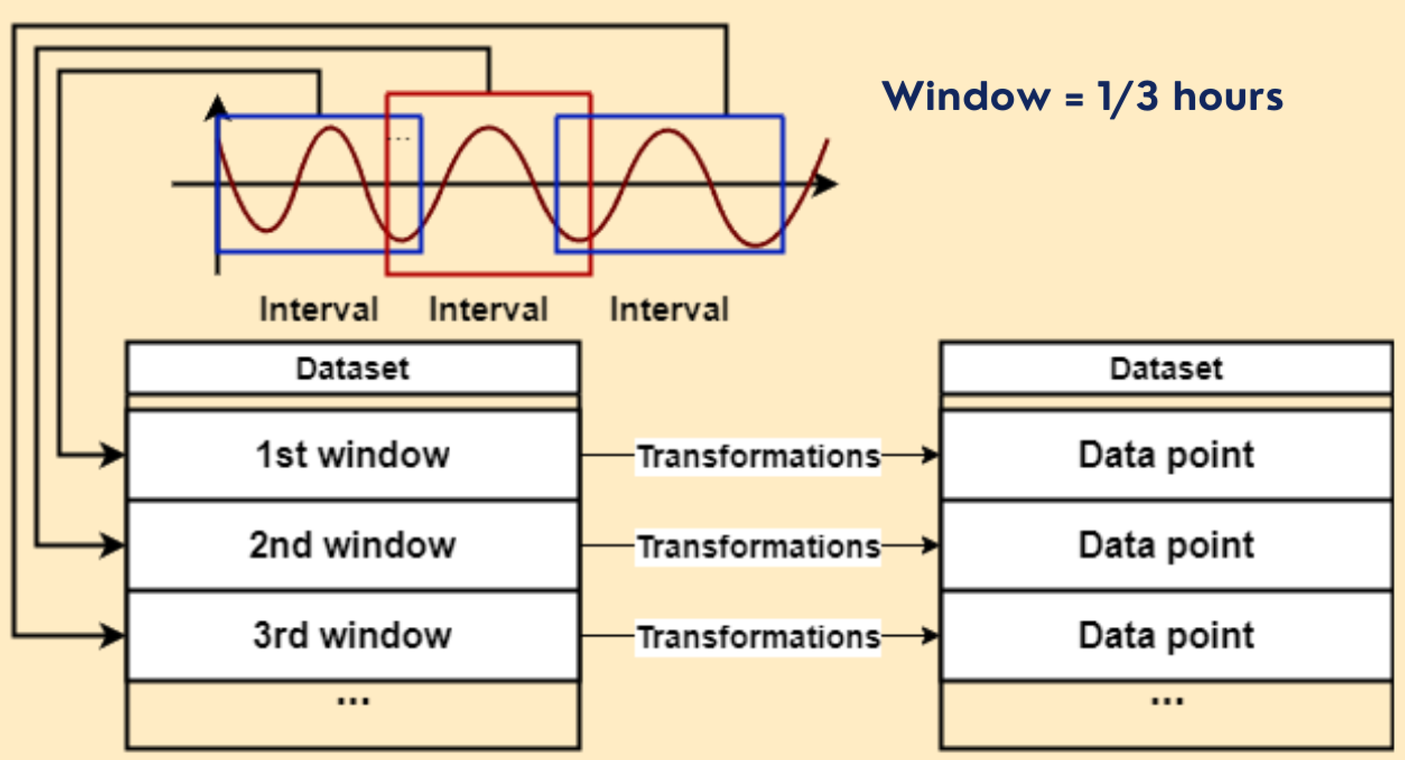
Gap in knowledge

- Most research is about identifying a specific user ([1], [2], [3])
- Data mostly gathered from complex devices
- When it comes to authentication ([4] and [5], there is not much research on outlier detectors

Using only heart rate data from a consumer-grade wearable, how well can Gaussian Mixture and One Class Support Vector Machine outlier detectors accurately distinguish multiple authorised persons from multiple unauthorised persons?

3. Methodology

1) Data processing



Transformations = statistical features + Mel-frequency cepstral coefficients (MFCC) + Principal component analysis (PCA)

2) Outlier detectors training

Methods

- **one versus many** (authorised vs unauthorised)
- **many versus many**
  - one model that trains all the known persons
  - multiple models, one per known person, results aggregated

Models

- **Gaussian mixture model (GMM)**
- **One Class Support Vector Machine (One Class SVM)**

3) Conclusions

- interpret results in terms of performances
- Comparison of the data transformations
- Comparison between models' performances

4. Results

Gaussian mixture model (GMM)

The results are **a mean of 100 combinations** of 1 known and 11 unknown people

1) 1vN

	GMM distributions	MFCC components	PCA dimensionality	statistical features	mean AUC	std AUC
1.	10	5	10	[]	0.830	0.072
2.	10	5	20	[]	0.830	0.072
3.	10	5	450	[]	0.830	0.072
4.	10	5	500	[]	0.830	0.072

Table 1: Top 4 Gaussian Mixture models with 1h windows in one versus many case

	GMM distributions	MFCC components	PCA dimensionality	statistical features	mean AUC	std AUC
1.	4	5	0	[mean, median]	0.936	0.044
2.	4	5	4	[]	0.935	0.042
3.	4	5	100	[]	0.935	0.043
4.	4	5	500	[]	0.935	0.043

Table 2: Top 4 Gaussian Mixture models with 3h windows in one versus many case

1h windows:

- **No MFCC:** 0.634
- **No PCA:** 0.746

3h windows:

- **No MFCC:** 0.746
- **No statistical:** 0.89

2) NvN

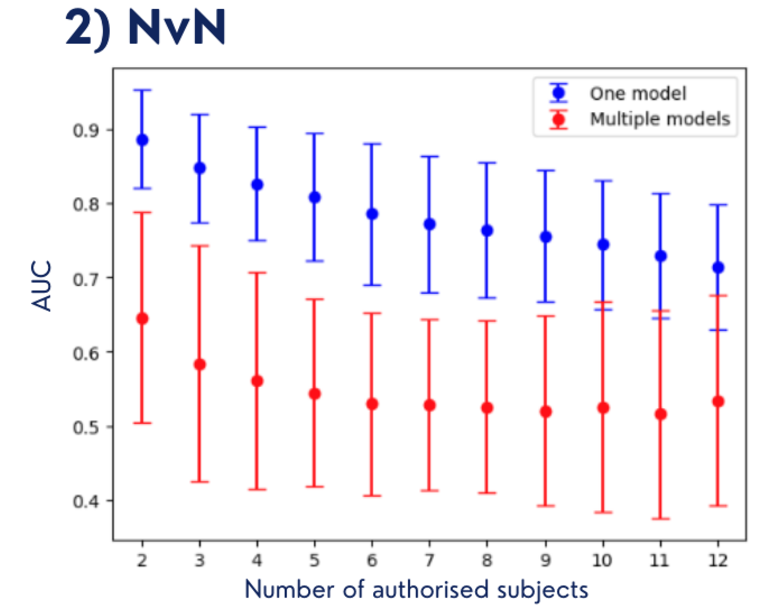


Figure 1: One model GMM performance versus multi-model GMM performance

The models in Figure 1 **are trained with the parameters of the best model** in 1vN

The results in Figure 1 are **a mean of 100 combinations** of 2-12 known and 6 unknown people

One Class SVM

The results are **a mean of 100 combinations** of 1 known and 11 unknown people

1) 1vN

	nu	MFCC components	PCA dimensionality	statistical features	mean AUC	std AUC
1.	0.8	20	100	[]	0.648	0.115
2.	0.8	20	250	[]	0.648	0.115
3.	0.8	20	4	[]	0.648	0.115
4.	0.8	20	50	[]	0.648	0.115

Table 3: Top 4 One Class SVM models with 1h windows in one versus many case

	nu	MFCC components	PCA dimensionality	statistical features	mean AUC	std AUC
1.	0.8	5	450	[]	0.785	0.112
2.	0.8	5	250	[]	0.785	0.112
3.	0.8	5	10	[]	0.785	0.112
4.	0.8	5	20	[]	0.785	0.112

Table 4: Top 4 One Class SVM models with 3h windows in one versus many case

1h windows:

- **No MFCC:** 0.648
- **No PCA:** 0.635

3h windows:

- **No MFCC:** 0.676
- **No PCA:** 0.758

2) NvN

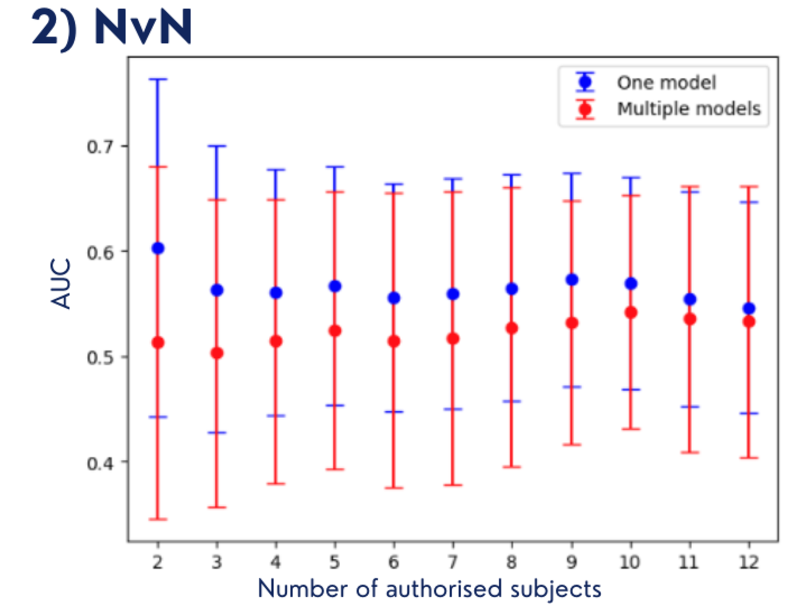


Figure 2: One model One Class SVM performance versus multi-model One Class SVM performance

The models in Figure 2 **are trained with the parameters of the best model** in 1vN

The results in Figure 2 are a **mean of 100 combinations** of 2-12 known and 6 unknown people

5. Conclusion

1) **in 1vN**

- **3h windows** give more significant features - the AUC scores are higher for models that use this
- **MFCC** has the most influence on the models
- the **Gaussian mixture model** performs the separation better than the **One Class SVM** with the best mean AUC of 0.936, against the best score of 0.785
- the **Gaussian mixture model** achieves more stable results (lower AUC standard deviation)

2) **in NvN**

- the AUC score **decreases with the more subjects are trained** as known
- the gap between the two methods decreases with each new authorised subject - **the trend indicates that at some point they will intersect**
- the mean **AUC score seems to converge to around 0.5 AUC**, but it's inconclusive because only 12 subjects were taken

6. Limitations

- Using the step counter could have improved the results, but it was hard to find subjects with enough step counter data
- The results improved from 1-hour to 3-hour windows, the next step would have been 6-hour windows
- The sliding window approach from 3-hour windows might have resulted in a temporal loss of the data
- In many versus many the convergence is not clear, experiments should be conducted with more than 12 subjects

References

[1] Issam Hammad and Kamal El-Sankary. Using machine learning for person identification through physical activities. In 2020 IEEE International Symposium on Circuits and Systems (ISCAS), pages 1–5, 2020.
[2] Christoph Lipps, Lea Bergkemper, Jan Herbst, and Hans Dieter Schotten. I know you by heart: Biometric authentication based on electrocardiogram (ecg) signals. In International Conference on Cyber Warfare and Security, volume 17, pages 135–144, 2022.
[3] Sudip Vhaduri and Christian Poellabauer. Wearable device user authentication using physiological and behavioral metrics. In 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pages 1–6, 2017.
[4] Eros Pasero, Eugenio Balzanelli, and Federico Caffarelli. Intruder recognition using ecg signal. In 2015 International Joint Conference on Neural Networks (IJCNN), pages 1–8, 2015.
[5] Alexa Muratyan, William Cheung, Sayanton V Dibbo, and Sudip Vhaduri. Opportunistic multi-modal user authentication for health-tracking iot wearables. In The Fifth International Conference on Safety and Security with IoT: SaSeIoT 2021, pages 1–18. Springer, 2022.

TU Delft Delft University of Technology