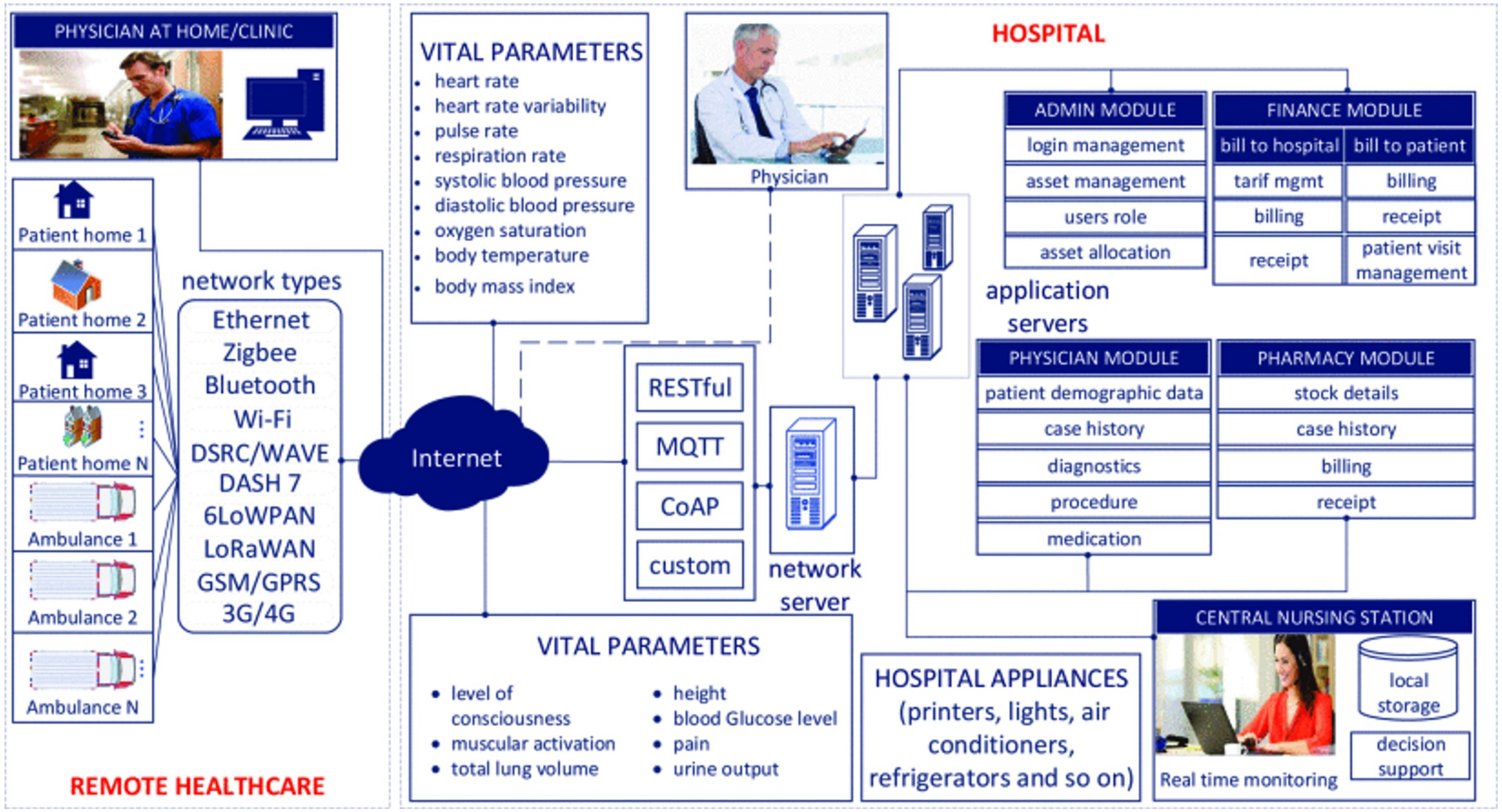
EFFICIENT INFORMATION DISTRIBUTION IN INTERNET OF MEDICAL THINGS (IOMT) SCENARIOS

TULLIA FONTANA, NICOLÁS ORTIZ DE ZARATE, NICOLE ZATTARIN

INTERNET OF MEDICAL THINGS (IOMT):

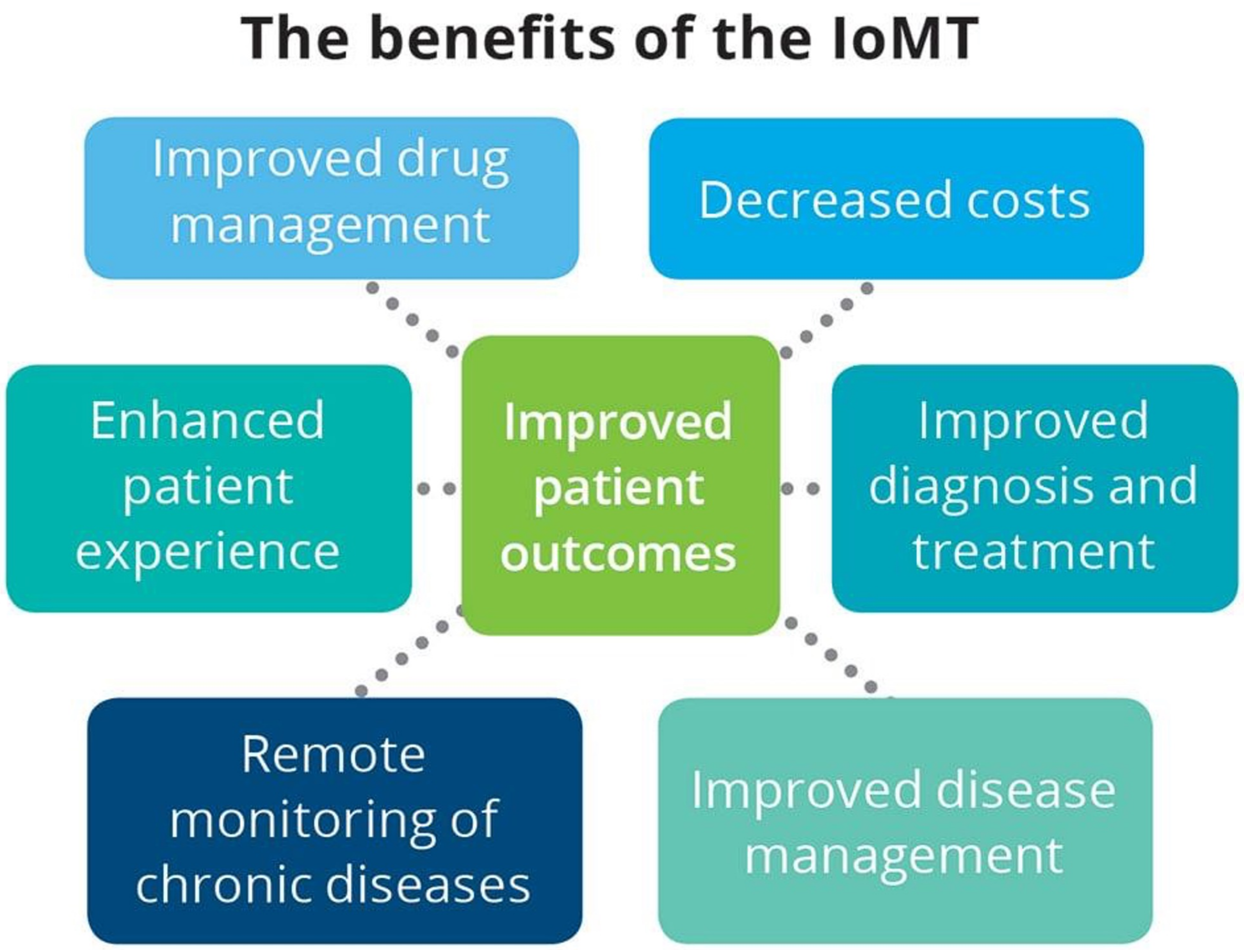
ARCHITECTURE



Architecture of Internet of Medical Things (IoMT), taken from [9]

INTERNET OF MEDICAL THINGS (IOMT):

ADVANTAGES



INTERNET OF MEDICAL THINGS (IOMT):

CHALLANGES

Problems:

- Large amount of data
- Fast processing
- Limited capacity of communication networks

Trade-off:

Transmitted information – Channel saturation

WORK OUTLINE

Correlations: Principal Component Analysis [2]

Dimensionality reduction: Kmeans Clustering [4]

Validation: binary classification task

Network design

Heterogeneous analysis: data of the same kind

Homogeneous analysis: data of different kind

DATASET

Dataset:

OPPORTUNITY Activity Recognition Dataset [1]

Termed activity of daily living (ADL) dataset:

- Start: lie on the deckchair, get up
- Groom: move in the room
- Relax: go for a walk
- Prepare/drink coffee
- Prepare/eat sandwich
- Cleanup
- Break: lie on the deckchair

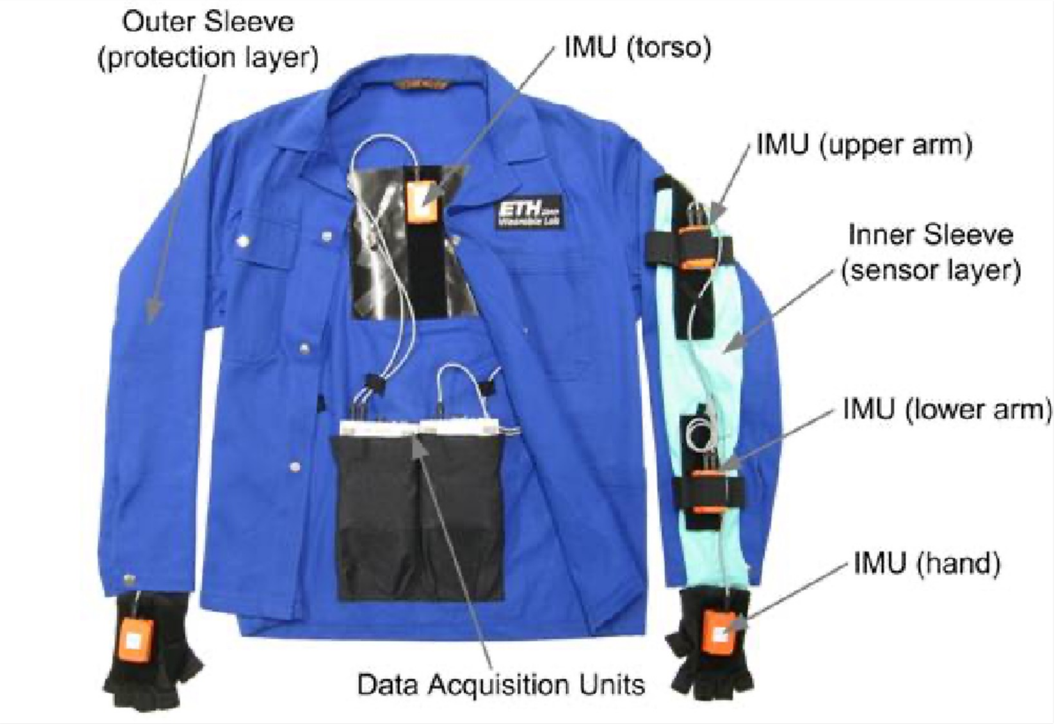


Figure 1: *Wearable motion jacket on which sensors are attached. Figure taken from [1].*

DATASET

Sensor types:

- Inertial Measurement Units (IMU) accelerators and gyroscopes
- Triaxial accelerometers

Pre-processing:

$$M = \sqrt{x^2 + y^2 + z^2}$$

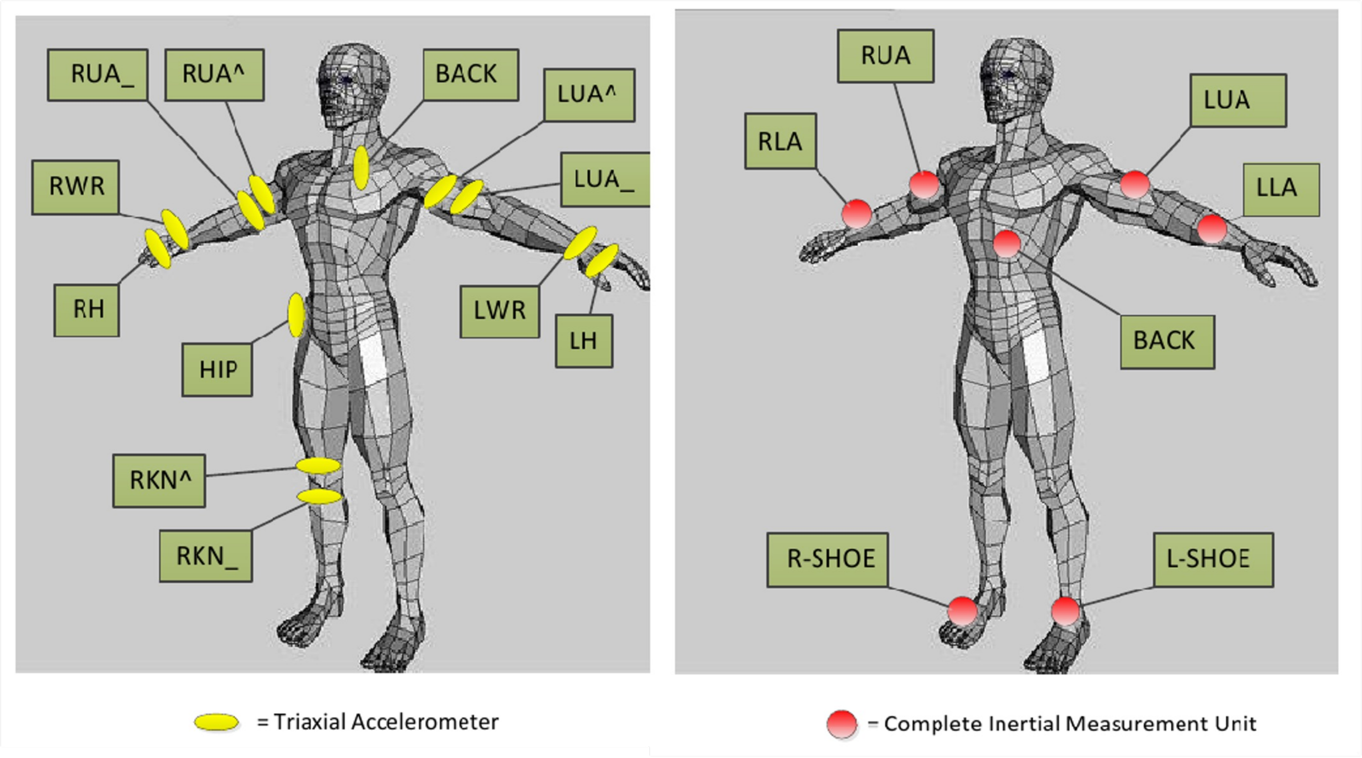


Figure 2: *Body sensor placement over the subject, for what concerns Inertial Measurement Unit on the right and Triaxial accelerometers on the left. Figure taken from [1].*

CORRELATIONS: PRINCIPAL COMPONENT ANALYSIS (PCA) [2]

Dimensionality reduction technique based on the correlation among features;

Principal Components (PCs): linear combination of original features;

Explained Variance: measurement of the percentage of variance which can be attributed to each of the PCs:

$$EV_i = \frac{\lambda_i}{\sum_{k=0}^{d-1} \lambda_k}$$

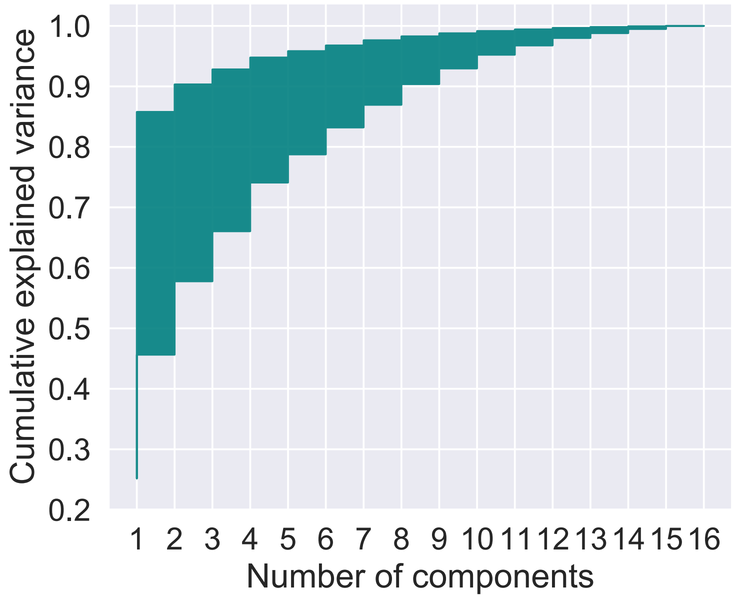


Figure 4: *Cumulative explained variance of each component, computation performed with all the different sensor types. The filled area represents the area between the minimum and the maximum for each component, among different combination of subject and run.*

CORRELATIONS: PRINCIPAL COMPONENT ANALYSIS (PCA)

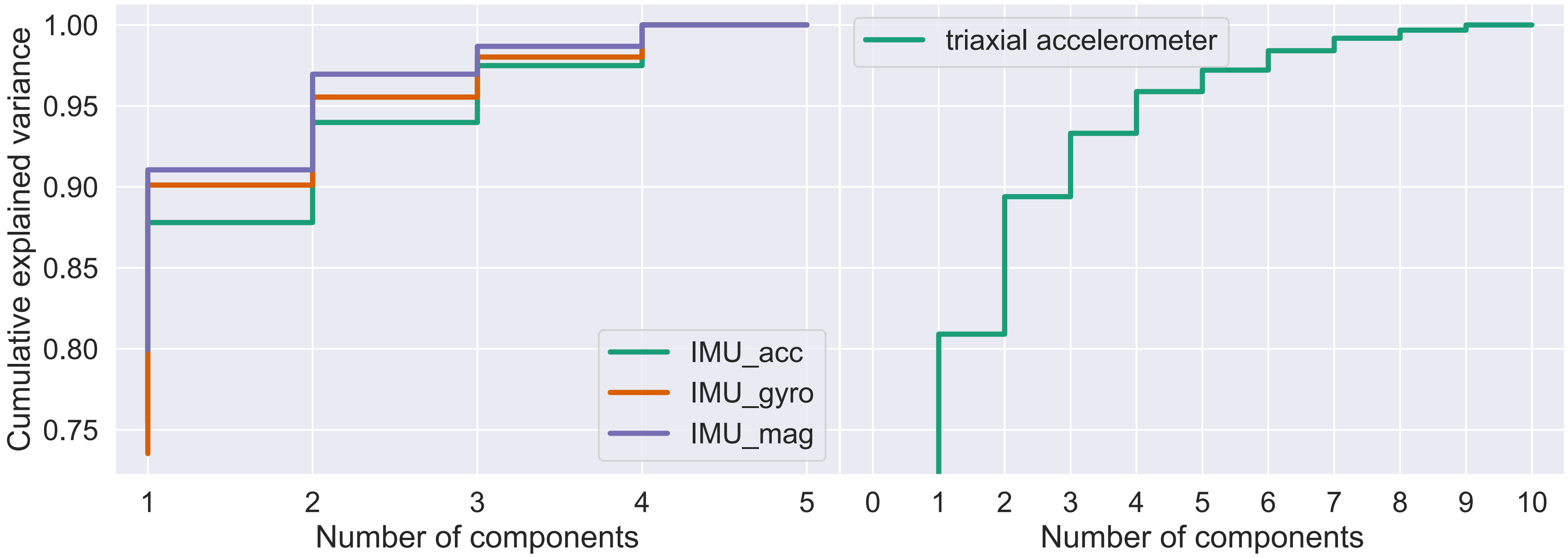


Figure 3: *Cumulative explained variance of each component for subject 1, run 1. Left panel shows the explained variance referring to different IMU sensors, while right panel refers to triaxial accelerometers.*

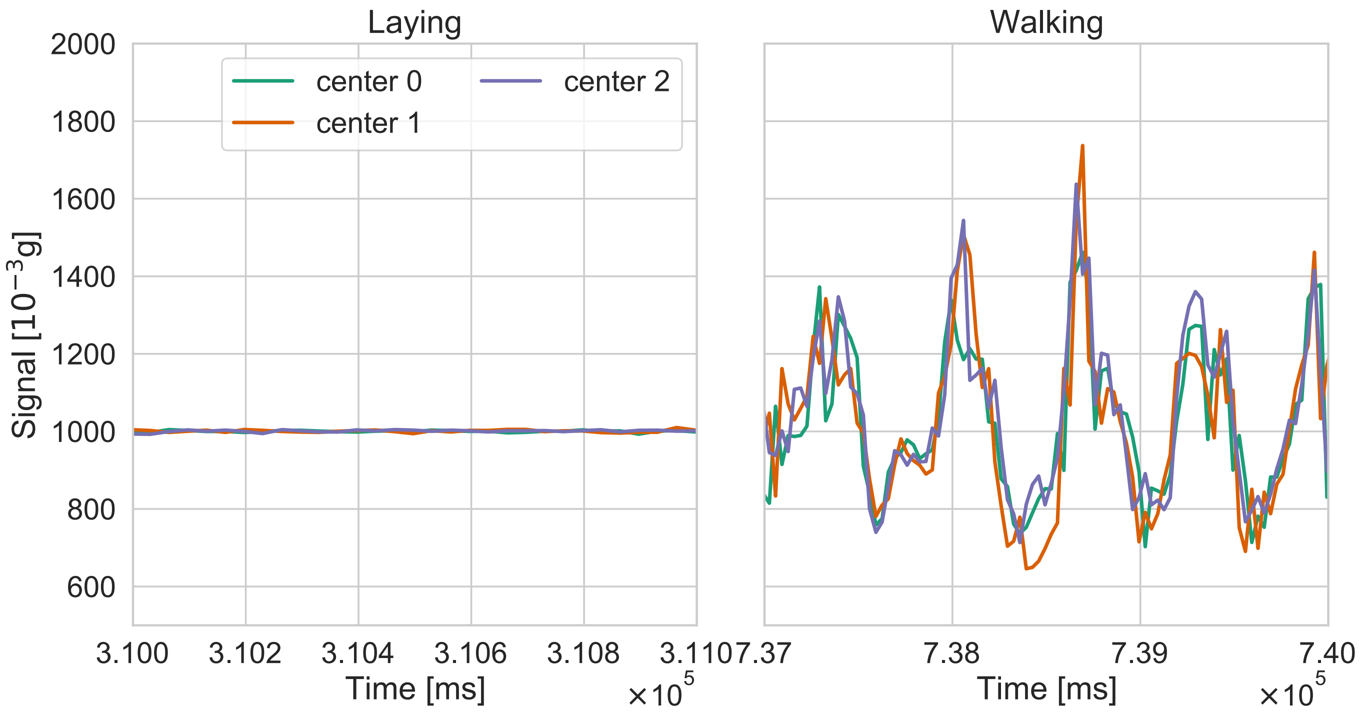
DIMENSIONALITY REDUCTION: CLUSTERING

Data: signals from different sensors while the subject is walking/laying;

Kmeans clustering [4] [5]: computes centroids with a fixed number of clusters;

Metrics: Euclidean, dynamic time wrapping distance [6];

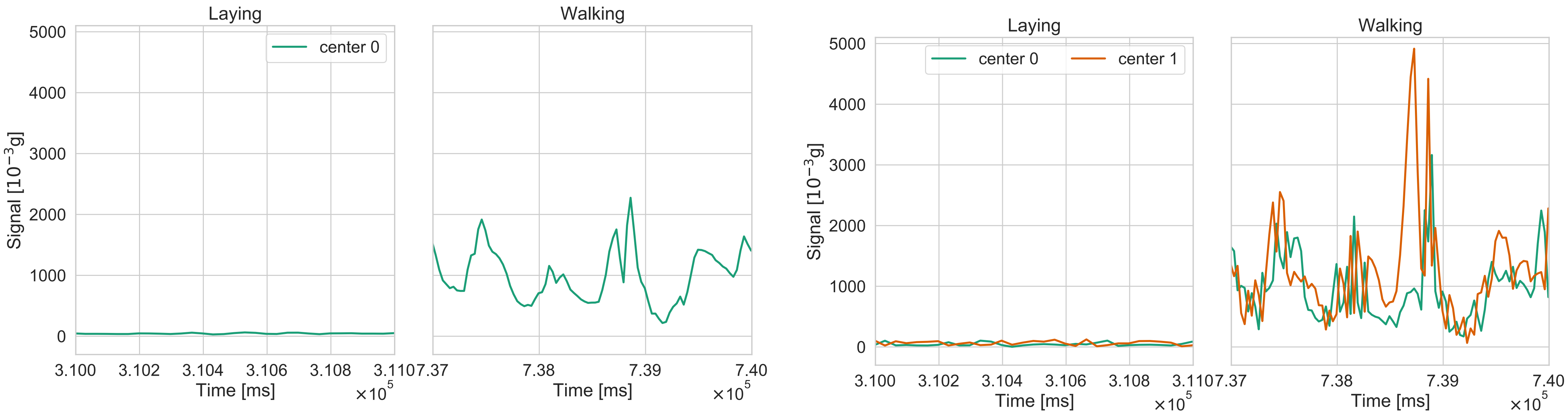
Approach: apply Kmeans at each millisecond.



(c) *Signal obtained with KMeans fixing 3 centers.*

Figure 5: *Signals of the centers obtained applying KMeans on IMU accelerometers for different numbers of clusters. Plots obtained for subject 1, run 1*

DIMENSIONALITY REDUCTION: CLUSTERING



(a) *Signal obtained with KMeans fixing 1 center.*

(b) *Signal obtained with KMeans fixing 2 centers.*

Figure 6: *Signals of the centers obtained applying KMeans on IMU gyroscopes for different numbers of clusters. Plots obtained for subject 1, run 1*

VALIDATION: BINARY CLASSIFICATION

How well can we still distinguish high level activity after clustering?

Train a binary classifier (walking/laying) on original features

Test on data obtained from the signals of KMeans centroids

Accuracy answers the question

ORIGINAL FEATURES

80% train set

20% test set

CLUSTERING DATA

CLASSIFICATION OF HOMOGENEOUS DATA

Linear model: logistic regressor

Dataset: amplitude of the signals in a window of 80 ms;

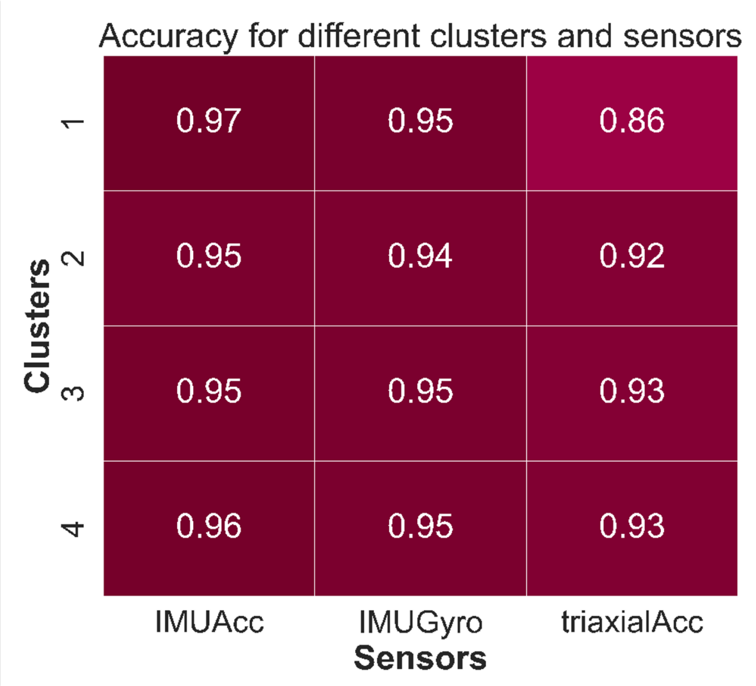


Figure 9: *Accuracy of the logistic model for each sensor type and number of centers considered for clustering.*

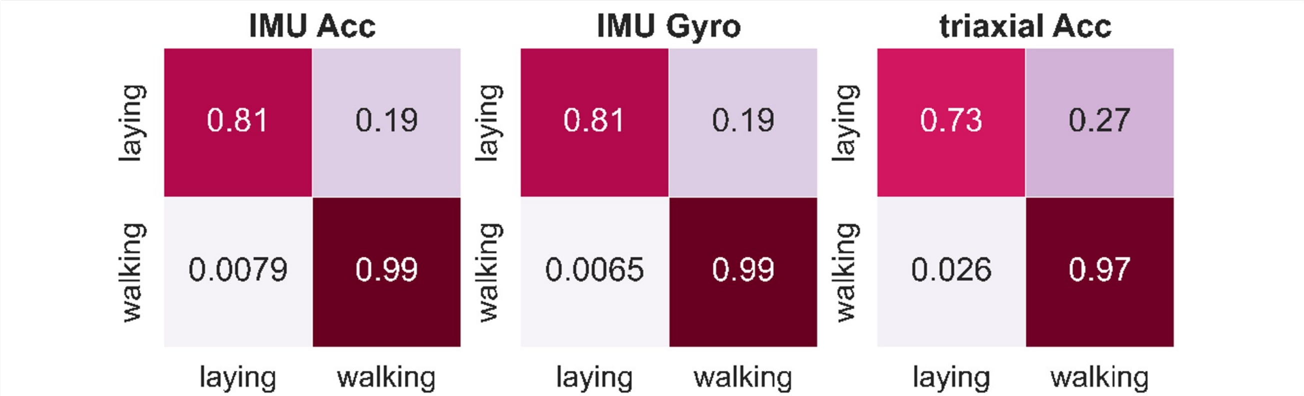


Figure 8: *Confusion matrices on test set for each sensor type logistic classifier.*

CLASSIFICATION OF HOMOGENEOUS DATA

Linear model drawback:

False positive rate: non negligible ~20%

Neural model:

- **Dataset**: entire time-series;
- **Accuracy:** high accuracy, diagonal confusion matrix;

Drawbacks:

- **Training:** slower and more delicate;
- **Tuning**: requires specific hyperparameters according to the specific kind of data;

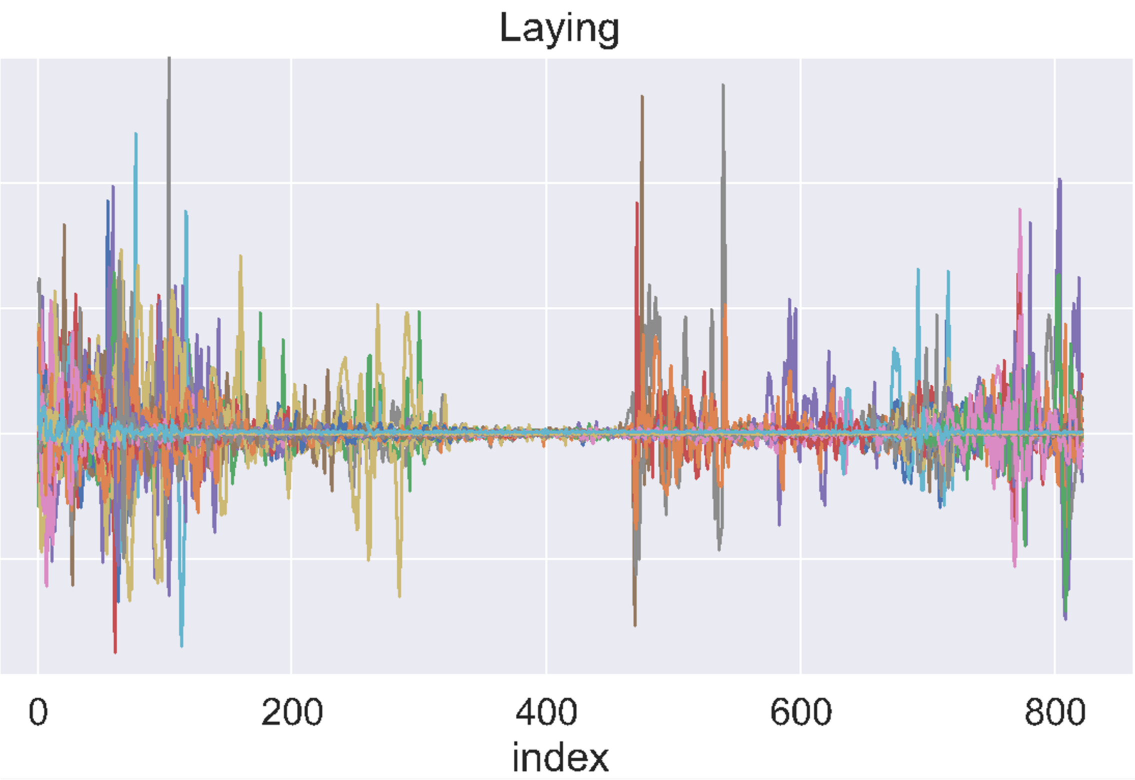


Figure: IMU accelerometer signals in laying conditions.

CLASSIFICATION OF HETEROGENEOUS DATA

Dataset: RUA, RLA, and BACK sensors for the IMU measurements and hip, back, RUA^, RUA_, RWR, RKN_ for the triaxial accelerators;

Preprocessing: standardization among data coming form different sensors;

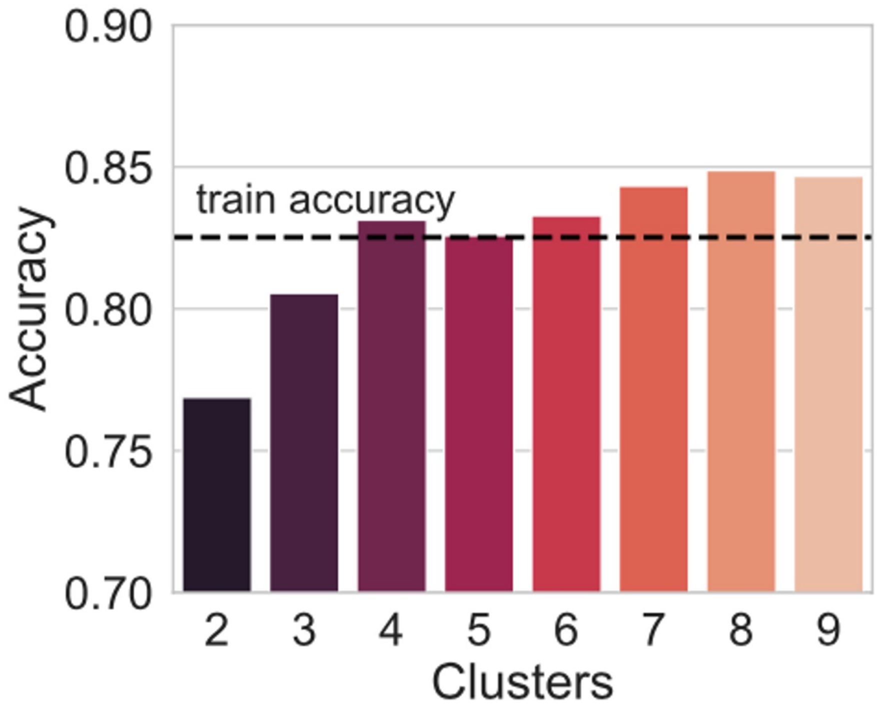


Figure 11: *Linear regression accuracy in an heterogenous scenario. We consider RUA, RLA, and BACK sensors for the IMU measurements and hip, back, RUA, RUA, RWR, RKN for the triaxial accelerators.*

CHECKPOINT

Homogeneous analysis:

IMU sensors (4 each): a single PC accounts for ~90% of variance → one cluster is enough to correctly classify with more than 95% accuracy;

Triaxial accelerometers (9 each): 2/3 PCs account for ~90% of variance → 2/3 clusters are needed to reduce the leak of information (~93% accuracy).

Heterogeneous analysis:

PCA: 8 PCs worst case - 2/3 best case to account for the 90% variance for all runs/subjects;

Clustering and classification: more than 4 centers to reach train accuracy;

REFERENCES

[1] Mirco Rossi Thomas Holleczek Gerhard Tröster Paul Lukowicz Gerald Pirkl David Bannach Alois Ferscha Jakob Doppler Clemens Holzmann Marc Kurz Gerald Holl Ricardo Chavarriaga Hesam Sagha Hamidreza Bayati Daniel Roggen, Alberto Calatroni and José del R. Millàn. Collecting complex activity data sets in highly rich networked sensor environments. *Seventh International Conference on Networked Sensing Systems (INSS'10), Kassel, Germany*, 2010.

[2] Ian Jolliffe. *Principal Component Analysis*, pages 1094–1096. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-04898-2. doi: 10.1007/978-3-642-04898-2_455. URL https://doi.org/10.1007/978-3-642-04898-2_455.

[3] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.

REFERENCES

[4] S. P. Lloyd. Least squares quantization in pcm. In *Technical Report RR-5497, Bell Lab, September.*, 1957.

[5] Romain Tavenard, Johann Faouzi, Gilles Vandewiele, Felix Divo, Guillaume Androz, Chester Holtz, Marie Payne, Roman Yurchak, Marc Rußwurm, Kushal Kolar, and Eli Woods. Tslearn, a machine learning toolkit for time series data. *Journal of Machine Learning Research*, 21(118):1–6, 2020. URL `http://jmlr.org/papers/v21/20-091.html`.

[6] S. Chiba H. Sakoe. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26:43–49, 1978.

[7] Ignacio Oguiza. tsai - a state-of-the-art deep learning library for time series and sequential data. Github, 2022. URL `https://github.com/timeseriesAI/tsai`.

[8] Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F. Schmidt, Jonathan Weber, Geoffrey I. Webb, Lhassane Idoumghar, Pierre-Alain Muller, and François Petitjean. InceptionTime: Finding AlexNet for time series classification. *Data Mining and Knowledge Discovery*, 34(6):1936–1962, sep 2020. doi: 10.1007/s10618-020-00710-y. URL `https://doi.org/10.1007%2Fs10618-020-00710-y`.

REFERENCES

[9] S. Vishnu, S. R. J. Ramson and R. Jegan, "Internet of Medical Things (IoMT) - An overview," 2020 5th International Conference on Devices, Circuits and Systems (ICDCS), 2020, pp. 101-104, doi: 10.1109/ICDCS48716.2020.243558.