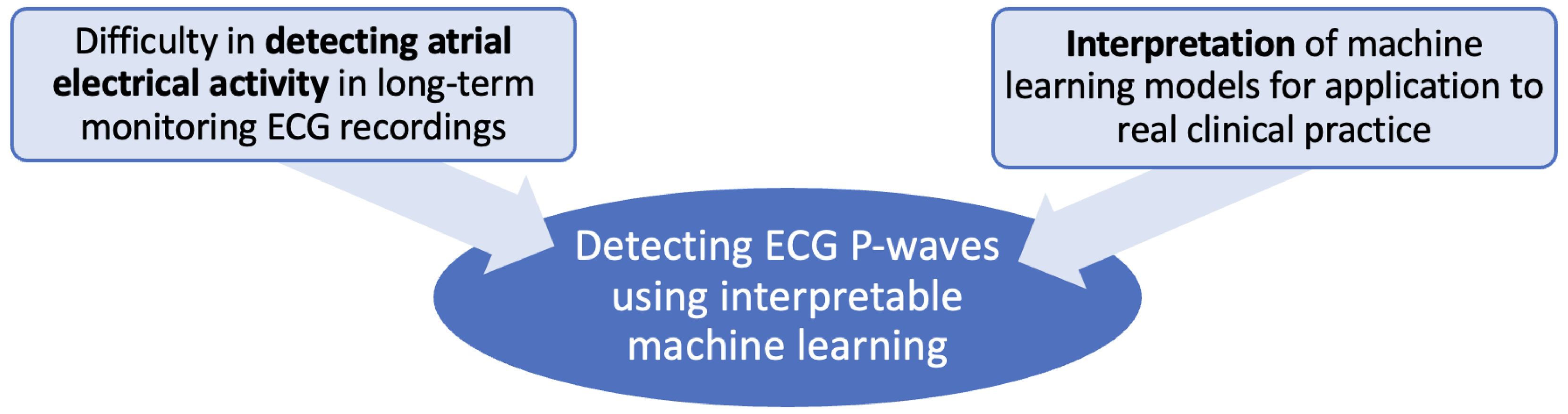
Interpretable ECG P-wave detection using autoencoders

Carmen Plaza-Seco, Kenneth E. Barner and Manuel Blanco-Velasco

Department of Electrical and Computer Engineering, University of Delaware
Department of Signal Theory and Communications, University of Alcalá (Spain)

Motivation

Electrocardiogram (ECG) P-waves, which represent atrial contraction of the heart, are of significant importance to cardiac health due to the high prevalence of atrial fibrillation. Two main **challenges** are associated with detecting the P-wave using machine learning tools:



Methods

Autoencoders (AE). Dimensionality reduction models that compress the input data to extract relevant patterns (*encoder*) in order to reconstruct the input data to the output (*decoder*).

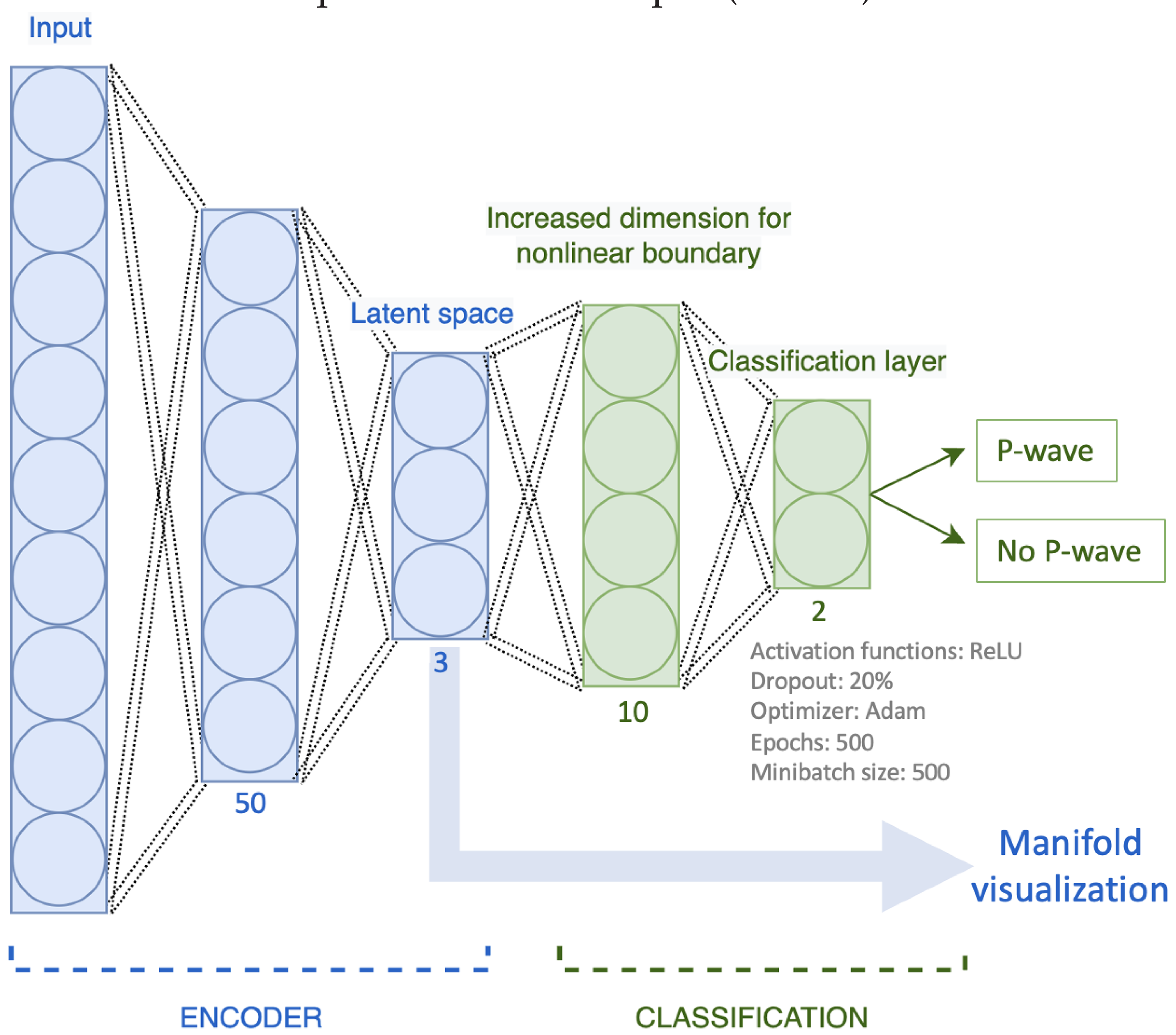


Figure 1: Neural network based on AE arquitecture for P-wave detection.

Based on AE arquitecture, a fully connected network has been created (Fig. 1), which compresses ECG data up to the *latent space* layer where the significant features are located.

Not only is it possible to perform classification on the latent space **manifolds**, but it **can** also **provide interpretability** to the functioning and decision making of the network.

Databases

Patient recordings have been segmented and labeled using a sliding window of 1-second (s), advancing in steps of 0.02-s, and a *proximity area* of 0.40-s, so that the **system can learn to work in a real situations** without prior beat delineation to detect P-waves (Fig. 2: top). Two ECG delineation databases have been used:

- **LUDB** [1]: 10-s long recordings of 200 patients have been used to train and validate the network due the availability of manual annotations. Randomly 85% of the segments to train and the remaining 15% to validate.
- **QTDB** [2]: 25 patients have been used to test the model. Only the manually labeled fragments of these 15-minutes (min) recordings have been used and segmented, approximately 1-min per patient.

Proximity area is defined as a means to obtain **balanced classes sets** for training and testing (Fig. 2: bottom).

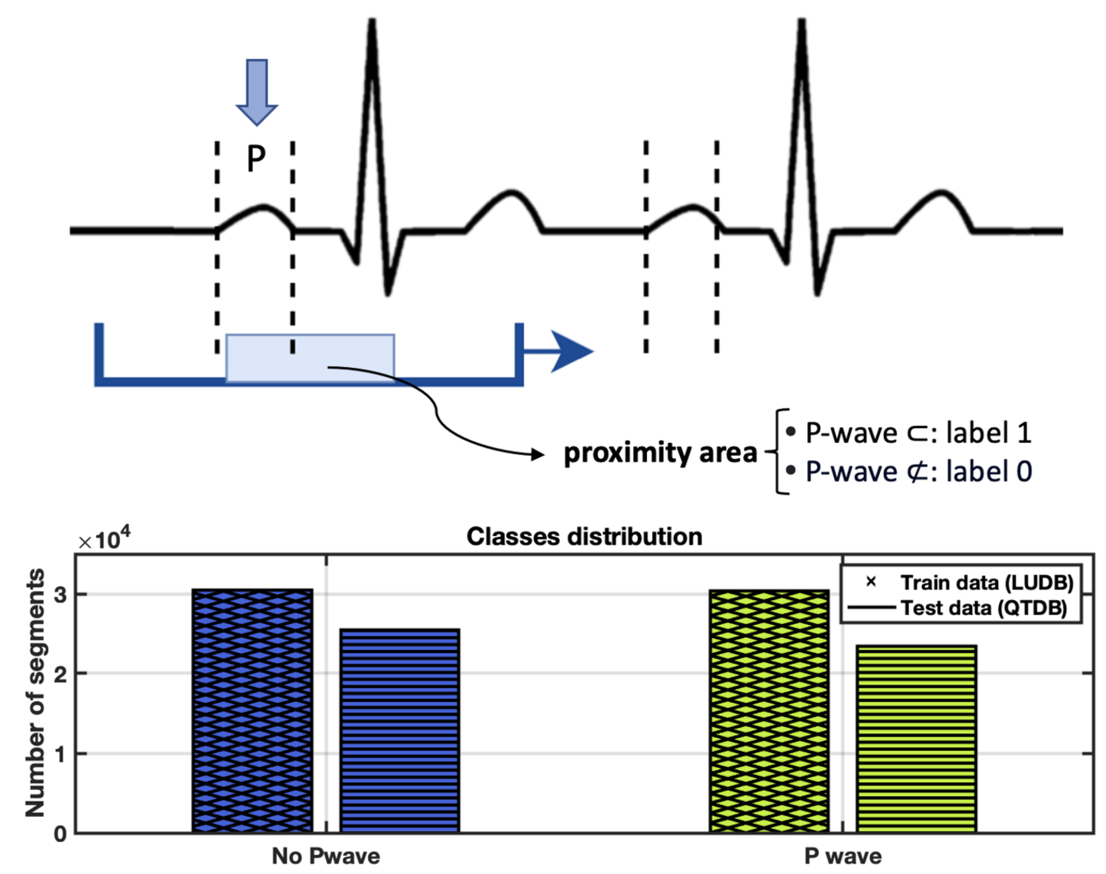


Figure 2: Recording segmentation and labeling process (top); Train-test classes distribution after segmentation (bottom).

Experiments

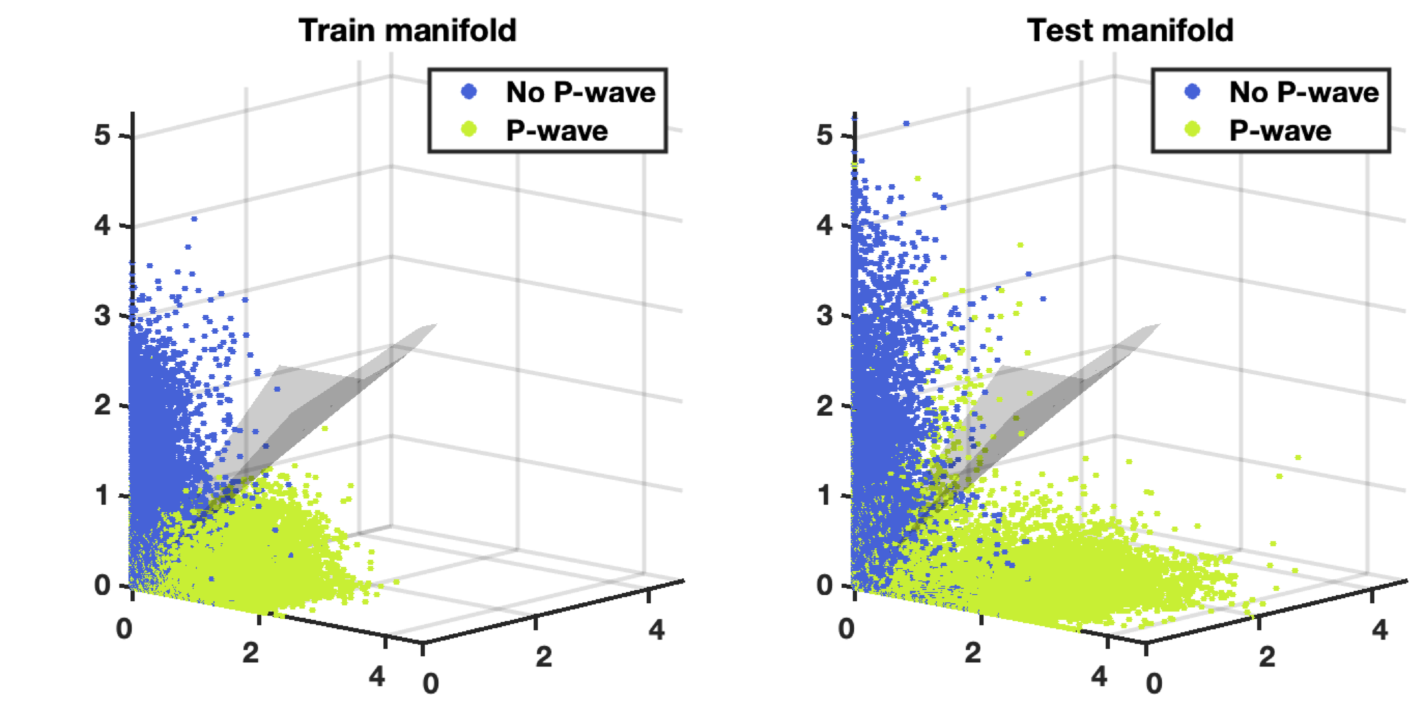


Figure 3: Train and test latent spaces manifolds with network classification boundary.

Different regions of the latent space can be **associated with each class**. Classification boundary and data projections (Fig. 3) provide a qualitative explanation of the network predictions.

The model is tested on new data consisting of real ECG fragments. The system yields **predictions close to the annotations**, although it still elicits some false positives (Fig. 4).

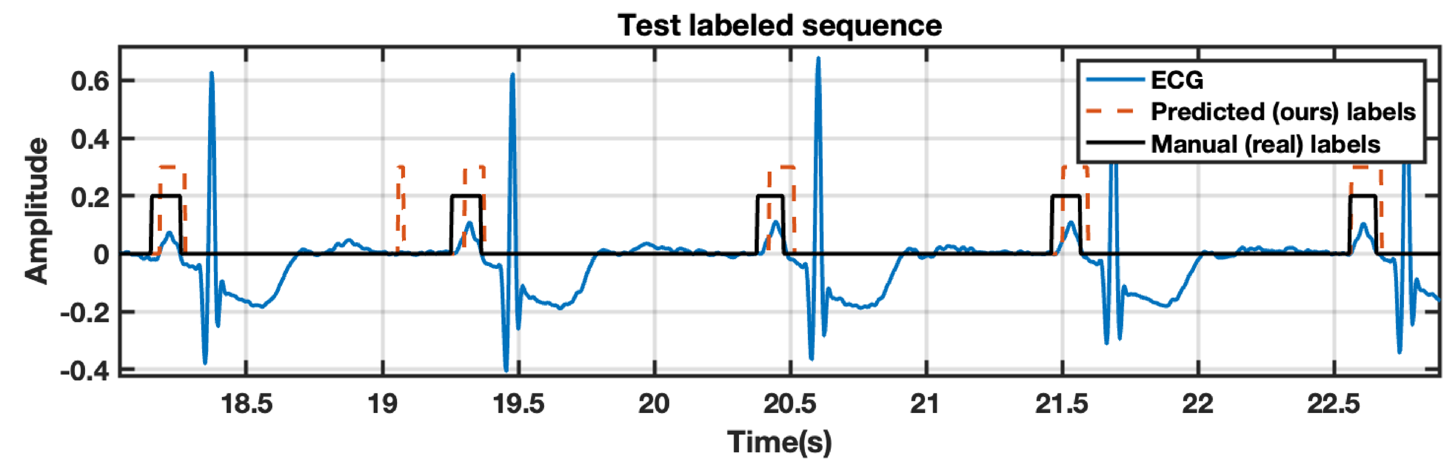


Figure 4: Test signal fragment with expert annotations and our predictions.

Table 1: Average of F1 scores of model performance after 10 simulations.

Train	Validation	Test
0.93 ±0.004	0.92 ±0.003	**0.90** ±0.007

Conclusion

- Joint visual inspection of data projections in latent spaces and classification boundaries provides the opportunity to **interpret model decision making**.
- This model provides tools to increase the **confidence of medical staff** in these systems.
- Confidence improvement will enable the **application** of these systems **to real clinical practice**.

Acknowledgement

Grant EPU-INV/2020/002 of the Comunity of Madrid and the University of Alcalá. Thank University of Delaware for the opportunity.

Future work

- **Different network layouts** with convolutional or LSTM layers.
- **Extrapolate to other waves**, such as T-wave related to sudden death pathology.
- Exploring **graph learning** models for ECG waveform detection.

References

[1] Kalyakulina, A., Yusipov, I., Moskalenko, V., *et al.*, "Lobachevsky university electrocardiography database," *PhysioNet*, 2021.

[2] P. Laguna, R.G. Mark, A. Goldberger, and G.B. Moody, "A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG," *Computers in Cardiology*, vol. 24, 673–676, 1997.