Analysis of Photoplethysmographic Signals in Low-Dimensional Latent Spaces

Enrique Feito-Casares[1,2], Francisco M Melgarejo-Meseguer[1],
Alejandro Cobo[2], Luis Baumela[2], José Luis Rojo-Álvarez[1]

[1] Departamento de Teoría de la Señal y Comunicaciones. Universidad Rey Juan Carlos, Madrid, Spain
[2] Universidad Politécnica de Madrid, Madrid, Spain

Abstract

Physiological signals are inherently complex yet serve as a critical source of clinical knowledge. Integrating this data into artificial intelligence systems has shown promising capabilities in supplementing usual clinical practices. However, the applicability of these systems may be constrained by the lack of transparency in the algorithms. The input space of photoplethysmography (PPG) signals is sourced from representative datasets with diverse purposes and extraction techniques. These signals are projected into a latent space using unsupervised (Autoencoders, AE), supervised (Fully Connected Neural Networks, FCNNs), and semi-supervised methods (Uniform Manifold Approximation and Projection, UMAP). The optimal hyperparameters are found with a grid search using hold-out cross-validation: reconstruction error for AE, classification loss for FCNN, and the Davies-Bouldin index for UMAP. UMAP demonstrates detailed geometry of the latent space accounting for the periodicity of signals. AE produces a compact representation, enabling reconstruction from hidden variables, while FCNN provides insight into the classification framework. Reduced-dimensionality latent spaces retain amplitude, frequency, periodicity, and morphological insights of PPG signals.

1. Introduction

Clinicians have long relied on sophisticated heuristic rules, such as photoplethysmography (PPG), to interpret physiological signals. These heuristics are not merely a collection of guidelines but are supported by extensive scientific research, rendering them highly effective diagnostic tools. The success of these rules lies not only in their diagnostic capabilities but also in their ability to be systematically taught, learned, and discussed within the clinical community.

While promising, the integration of AI with physiological signals remains unplaced due to the intricate, often opaque nature of AI models, which stands in contrast to the transparent frameworks used by clinicians. AI systems, particularly those employing deep learning, operate on principles different from heuristic rules. These models process vast amounts of data, identifying patterns beyond human capability, but lack the transparency and interpretability essential in clinical practice [1].

Manifold learning techniques offer a compelling data-driven solution to process physiological signals. By enabling the visualization of high-dimensional data in lower-dimensional spaces, these techniques can uncover the underlying structure of signals, retaining only the most essential variables and identifying intrinsic features [2]. Nevertheless, manifold learning algorithms have not been utilized to obtain a low-dimensional PPG signal representation. This study aims to evaluate the feasibility of obtaining lower-dimensional representations of PPG signals across various application contexts and to inspect the most evident intrinsic features extracted.

2. Materials and Methods

Next, we present the experimental datasets used, the remote PPG signal acquisition procedure, the preprocessing steps performed on the input signals, and the models used to process the signals: Autoencoders (AE), Fully Connected Neural Networks (FCNNs), and Uniform Manifold Approximation and Projection (UMAP).

2.1. Experimental Dataset

The PPG signals required for this study were sourced from PhysioNet, an online repository encompassing diverse physiological signals from human subjects [3].

Wrist PPG dataset: Measurements acquired utilizing a PPG unit positioned on the left wrist of study participants. These individuals performed different physical activities: walking, light running, and indoor cycling under high and low-resistance conditions [4].

BIDMC PPG and Respiration Dataset: Contains sig-

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.338

nals obtained from critically ill patients in a hospital environment utilizing conventional PPG instrumentation [5].

2.2. rPPG Signals Acquisition

The principle of PPG is based on a light source and a photodetector to observe changes in the volume of subcutaneous blood vessels. This technique involves illuminating the skin and employing a photodetector to detect minor changes in the intensity of light reflection or transmission, which occur due to variations in blood circulation, thereby generating PPG signals [6]. Additionally, remote photoplethysmography has been developed as a non-contact alternative, utilizing camera technology to capture facial videos from which subtle changes in skin coloration, indicative of the rPPG signal, can be extracted [7].

PhysFormer++ is an end-to-end video transformer-based architecture engineered for extracting rPPG signals from facial recordings, which sets state-of-the-art performance across various benchmark datasets [8]. A pre-trained PhysFormer++ model is used to perform inference on the FaceForensics++ dataset, which consists of 977 YouTube videos featuring clear, frontally-oriented faces without hands, thus ensuring their suitability for rPPG analysis [9].

2.3. Preprocessing

Data preprocessing transforms the raw data from collected datasets, as well as signals obtained through rPPG. The following workflow is followed:

Trend removal: Eliminate low frequencies caused by noise that disrupts the baseline. To achieve this, the process involves averaging the signal over segments, followed by interpolation, and ultimately, differentiating the original signal from the resultant signal.

Filtering: Applied to eliminate high-frequency noise extraneous to a typical PPG signal using a 10Hz low-pass Butterworth filter.

Resampling: Due to the diversity of the sources employed, each dataset exhibits a distinct sampling frequency: 256Hz for the Wrist PPG dataset, 125Hz for the BIDMC PPG dataset, and 30Hz for the signals acquired via rPPG. Consequently, all signals are resampled to match the lowest frequency observed across all sources.

Normalization: Min-Max normalization is employed on every signal $\boldsymbol{x}_i$ to scale all signals to fall within the interval $[0, 1]$ and facilitates comparison across the different datasets and measurement scales.

Buffering: Partitions input signal $\boldsymbol{x}_i$, where i in an index of the observations in the datasets, into n overlapping data segments of 20 samples with an overlap s of 5 samples. Given a signal vector $\boldsymbol{x}_i = [x_i^1, x_i^2, \ldots, x_i^L]$ with length L, then $K = \left\lfloor \frac{L-5}{15} \right\rfloor$ denotes the number of segments formed. The complete matrix of segments $\boldsymbol{X}$ is formed by vertically concatenating all segments.

2.4. Machine Learning Algorithms

Fully Connected Neural Network: An FCNN is a recurrent neural network architecture for classification tasks. This architecture minimizes a specific loss function by adjusting a set of weight matrices and biases. In a single hidden layer FCNN, this can be mathematically expressed as,

- For the hidden layer, the activation can be computed as:

$$\boldsymbol{H}^{[1]} = f^{[1]}(\boldsymbol{W}^{[1]}\boldsymbol{X} + \boldsymbol{b}^{[1]}) \tag{1}$$

- For the output layer, using the softmax function σ. the final output $\boldsymbol{y}$ is given by:

$$\boldsymbol{y} = \sigma(\boldsymbol{W}^{[2]}\boldsymbol{H}^{[1]} + \boldsymbol{b}^{[2]}) \tag{2}$$

where the activation function f are applied to the weighted sum of the input matrix $\boldsymbol{X}$ with the weight matrix $\boldsymbol{W}$ and the biases $\boldsymbol{b}$ The performance of the model can be evaluated using the cross-entropy loss function for multi-class classifications:

$$L(\boldsymbol{y}, \hat{\boldsymbol{y}}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(\hat{y}_{ij}) \tag{3}$$

where C is the number of classes N the number of observations, y_{ij} is a binary indicator of whether class j is the correct classification for observation i, and $\hat{y}_{ij}$ represents the predicted probability that observation i belongs to class j. In the present work, the multi-label classification task is utilized as a fake task to obtain a latent representation. The hidden layer constrains instances represented on a viewable space $\boldsymbol{H}$.

Autoencoder: An autoencoder is a specialized neural network architecture for unsupervised learning. It comprises two main components: an encoding stage and a decoding stage. The encoding process can be mathematically represented as follows:

$$\boldsymbol{H} = f(\boldsymbol{W}_e\boldsymbol{X} + \boldsymbol{b}_e) \tag{4}$$

where $\boldsymbol{H}$ denotes the latent representation. The decoding process, aiming to reconstruct the input from the latent representation, is expressed as:

$$\hat{\boldsymbol{X}} = f(\boldsymbol{W}_d\boldsymbol{H} + \boldsymbol{b}_d) \tag{5}$$

The parameters of the network are optimized to minimize a loss function that measures the discrepancy between the original input $\boldsymbol{X}$ and its reconstruction $\hat{\boldsymbol{X}}$ to capture the most informative features of the data.

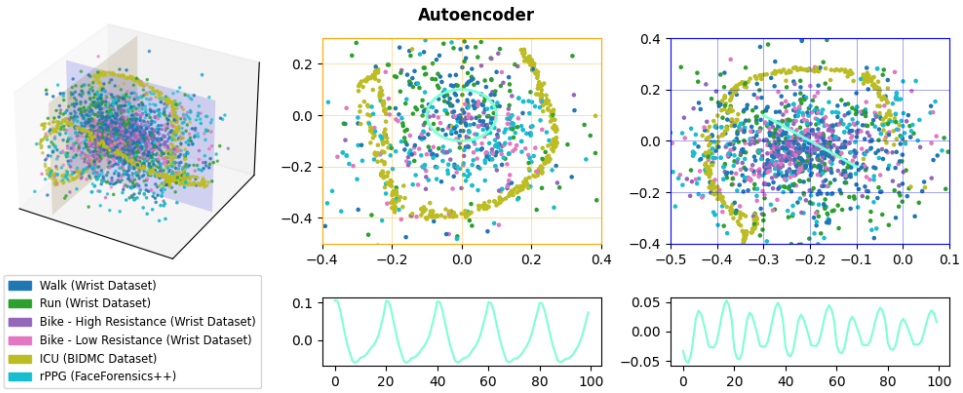
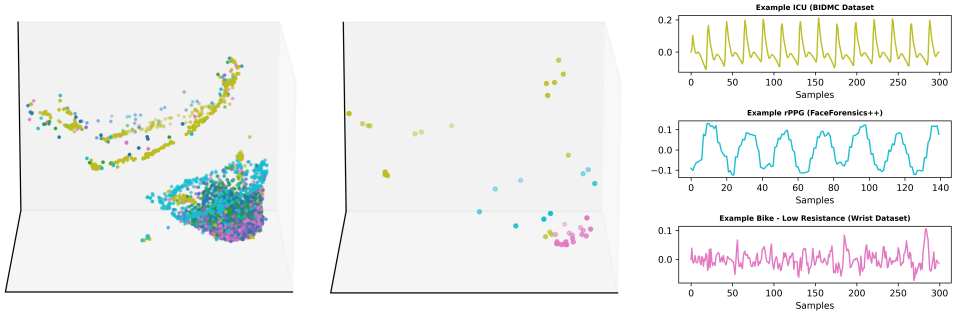


Figure 1: Latent space representation from AE for the complete test dataset with two perpendicular plane slices. Below each slice, a reconstruction of the PPG signal from the highlighted points is provided.



(a) (b) (c)

Figure 2: Latent space representation of UMAP for the complete test dataset (a) and for some example PPG signals (b) corresponding to the waveforms (c) .

Universal Manifold Approximation and Projection: UMAP is characterized by its adoption of a nonlinear dimensional reduction strategy, which is formulated upon three assumptions: first, that the data exhibits a uniform distribution across a Riemannian manifold; second, that the Riemannian metric remains locally constant or can be approximated as such; and third, that there exists local connectivity within the manifold [10].

3. Experiments and Results

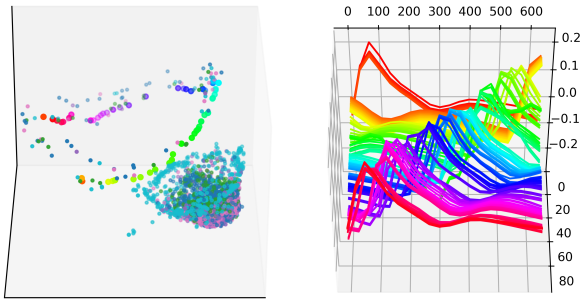
This section presents the performance of the proposed methods with the described data. The preprocessed dataset comprises 8,265 segments of the WRIST PPG dataset, 7,656 segments of the BIMDC dataset, and 2,912 segments from rPPG acquisition. Employing a standard hold-out cross-validation methodology, 60% of the segments are allocated for training, while the remaining 40% are evenly divided into validation and testing sets. For all models described, a grid search of hyperparameters is conducted.

The reconstruction error of AEs is assessed across various architectures and activation functions. The optimal configuration identified entails a 20-3-20 architecture utilizing Rectified Linear Unit activation, yielding a reconstruction error 0.038 (MAE). A depiction in Figure 1 exhibits the latent representation of AEs, elucidating their efficacy in reconstructing periodic geometries while introducing artifacts in the reconstruction of linear segments.

Figure 3a illustrates the projection of all test dataset PPG signals onto a conic shape, and 3b presents examples of PPG signals depicted at disparate altitudes within the projection space. Furthermore, Figure 2 shows the equivalence between periodic geometries in latent space and signal displacement in the input signal space.

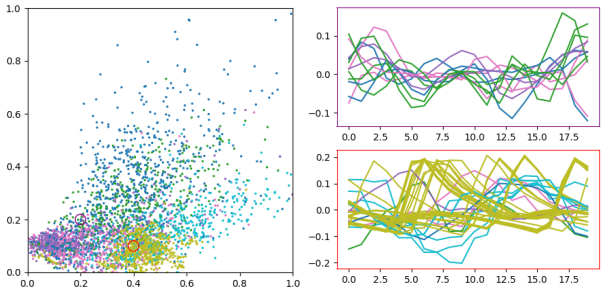
Regarding UMAP, hyperparameter tuning involves selecting the number of neighbors, with the David-Bouldin Index utilized to evaluate the latent space considering validation labels. A David-Bouldin Index of 8.2218 is achieved for 50 neighbors. Figure 3a illustrates the projection of all test dataset PPG signals onto a conic shape, and 3b presents examples of PPG signals depicted at disparate altitudes within the projection space. Furthermore, Figure 2 shows the equivalence between periodic geometries in latent space and signal displacement in the input signal space.

Finally, concerning FCNNs, a similar grid search procedure is implemented, employing categorical cross-entropy as the optimization criterion. The model achieves optimal



(a) (b)

Figure 3: (a) Latent space representation of UMAP for the complete test dataset with highlighted points from BIMDC Dataset (ICU). (b) Corresponding signal segments to highlighted points in (a)



(a) (b)

Figure 4: (a) Latent space representation extracted from FCNN for the complete test dataset. (b) Signal segments contained in highlighted red and purple circles in (a).

performance with a 40-2-40-20-6 architecture and ReLU activation function. Figure 4a provides visual evidence illustrating the impact of supervised learning on enhancing discrimination between distinct datasets. Furthermore, Figure 4(b) reveals consistent amplitude and morphology among signals localized within specific regions of the latent space.

4. Conclusions

The presented study proves that PPG signals can be represented in a reduced-dimension space while retaining a significant part of their structure. The methods have demonstrated promising outcomes in extracting salient features, including amplitude, frequency, periodicity, and insight into morphological characteristics. Future research aims to tackle the problem in more complex scenarios, focusing on identifying and interpreting pathophysiological features from low-dimensional latent spaces and integration into classification problems.

Acknowledgments

This work was supported by the Research Grants HERMES, LATENTIA, PCardioTrials, and ETHICFACE (PID2023-152331OA-I00, PID2022-140786NB-C31, and PID2022-140553OA-C42, and PID2022-137581OB-I00), funded by MICIU/AEI/10.13039/501100011033 and ERDF / EU. Also supported by Rey Juan Carlos University, project HERMES 2024/00004/00, by the Programa Propio of UPM and a grant from Comunidad de Madrid to the Madrid ELLIS Unit.

References

[1] Janiesch C, Zschech P, Heinrich K. Machine Learning and Deep Learning. Electronic Markets 2021;31(3):685–695.

[2] Pinaya WHL, Vieira S, Garcia-Dias R, Mechelli A. Chapter 11 - Autoencoders. Machine Learning. Academic Press, 2020; 193–208.

[3] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000;101(23):E215–E220.

[4] Jarchi D, Casson A. Description of a Database Containing Wrist PPG Signals Recorded during Physical Exercise with Both Accelerometer and Gyroscope Measures of Motion. Data Basel 2017;2(1):1.

[5] Pimentel MAF, Johnson AEW, Charlton PH, Birrenkott D, Watkinson PJ, Tarassenko L, Clifton DA. Toward a Robust Estimation of Respiratory Rate From Pulse Oximeters. IEEE transactions on biomedical engineering 2017; 64(8):1914–1923.

[6] Challoner AVJ, Ramsay CA. A Photoelectric Plethysmograph for the Measurement of Cutaneous blood flow. Physics in medicine biology 1974;19(3):317–328.

[7] McDuff DJ, Estepp JR, Piasecki AM, Blackford EB. A survey of remote optical photoplethysmographic imaging methods. In Annual International Conference of the IEEE Engineering in Medicine and Biology Society, volume 2015. IEEE, 2015; 6398–6404.

[8] Yu Z, Shen Y, Shi J, Zhao H, Cui Y, Zhang J, Torr P, Zhao G. PhysFormer++: Facial Video-Based Physiological Measurement with SlowFast Temporal Difference Transformer. International Journal of Computer Vision 2023; 131(6):1307–1330.

[9] Rossler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Niessner M. FaceForensics++: Learning to Detect Manipulated Facial Images. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019; 1–11.

[10] McInnes L, Healy J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. CoRR 2018;abs/1802.03426.

Address for correspondence:

Enrique Feito-Casares. University Rey Juan Carlos, Fuenlabrada (Madrid), Spain. Mail to: enrique.feito@urjc.es