

Comparative study on the generalization ability of machine learning algorithms for PPG quality assessment

S. Mula, R. Zangroniz, O. Ayo-Martin, J.J. Rieta, R. Alcaraz

Research Group in Electronic, Biomedical and Telecommunication Engineering, University of Castilla-La Mancha, Cuenca, Spain

Background and Aim. Photoplethysmogram (PPG) signals are often too noisy to accurately measure heart rate. Methods for PPG quality assessment based on machine learning (ML) concepts have gained considerable attention, but they have been commonly developed on datasets recorded under specific conditions and not tested on others. Hence, the aim of the present work is to compare their performance and generalization ability on datasets representative of different conditions.

Methodology. Only datasets containing both PPG and ECG were used to label as clean or noisy on the difference in the inter-beat interval between both synchronous signals, using an error of 3 bpm as the threshold. Then, global and spectral features were extracted from the PPG segments and Decision Tree, SVM, and KNN classifiers were trained on WESAD, a dataset of wrist-captured PPG recordings in daily activities. The models were validated on the similar wrist-captured PPG DaLiA dataset and the finger-captured BIDMC dataset, which consists of recordings from ICU patients.

Results. The DT-based model reached 100% accuracy (Acc) on training, with a 30% drop in sensitivity (Se) and specificity (Sp) in DaLiA and an unbalanced Se-Sp on BIDMC of 31-84%. The SVM-based model reached an Acc of 82% on training, with a 3% performance drop on Se-Sp in DaLiA and 78-21% Se-Sp in BIDMC. Finally, the KNN-based model yielded an Acc of 85% on training and DaLiA, with an Se-Sp unbalance on BIDMC of 27-96%.

Significance. Whereas the DT-based model showed significant overfitting on the two validation datasets, the SVM-based and KNN-based models reported a successful performance on the database with similar PPG morphology as the training one. Therefore, the selection of ML classifier and datasets for training and validation are aspects that require special attention in PPG quality assessment.

	DT			SVM			KNN		
	Train	DaLiA	BIDMC	Train	DaLiA	BIDMC	Train	DaLiA	BIDMC
Accuracy	1.0000	1.0000	0.7081	0.8233	0.8233	0.7830	0.8498	0.8498	0.7781
Sensitivity	1.0000	0.7115	0.3126	0.8104	0.7888	0.2096	0.8581	0.7878	0.2629
Specificity	1.0000	0.7056	0.8466	0.8362	0.7765	0.9839	0.8415	0.7400	0.9586