

GOLD 2, GOLD 3, GOLD 4) and different dyspnoea scores (MMRC 0, MMRC 1, MMRC 2, MMRC 3, MMRC 4) is performed through a 10-fold random-partitioning cross-validation process.

We firstly divided the data from the 132 subjects, comprising healthy subjects and matched COPD patients, into ten subsets. At each iteration, one of the subsets was used as the test set and the other nine subsets formed a training set. From each subject in the training set we randomly selected one night, represented by its characteristic vector of activation probabilities $\vartheta = [\vartheta_1, \dots, \vartheta_k]$ over the sleep modality $\beta_{1:k}$. We used these distributions to compute a square dissimilarity matrix A between pairs of nights according to *Kullback–Leibler* divergence as in [157], in which $A(i, j)$ denotes the dissimilarity between the i^{th} and j^{th} randomly selected nights $A(i, j) = \sum_k \theta_k^i \log \frac{\theta_k^i}{\theta_k^j}$ with θ_k^i and θ_k^j the activation probabilities of the sleep modality β_k for the nights represented by θ^i and θ^j . The choice of the dissimilarity measure is critical and must fit the nature of the features in question, which in this case are discrete probability functions. Secondly, we calculated the eigenvectors and eigenvalues of A so that $AV = DV$, where D is a diagonal matrix of eigenvalues and V a matrix whose columns are the corresponding right eigenvectors and there are as many eigenvectors and eigenvalues as there are rows in the initial matrix. Eigenvalues were ranked from the greatest to the least. Using the transformation $V_T D_T^{-1}$, with V_T the truncated eigenvector matrix of the first n eigenvectors of V and D_T the associated and truncated eigenvalue matrix, we summarized and attempted to represent inter-nights dissimilarities in a lower dimensional space.

We iteratively projected into $V_T D_T^{-1}$ all the nights of each subject in the training set and test set such that the between-object dissimilarities are preserved as well as possible. In particular, given a vector of sleep modality activation probabilities ϑ representing one night, we calculated the vector x of pairwise dissimilarities between ϑ and the nights used to compute $V_T D_T^{-1}$. Then we assigned to ϑ a location in a low-dimensional space projecting x into the learned space $V_T D_T^{-1}$ according to $v' = x V_T D_T^{-1}$. Iterating this operation for each night in the training set and test, the positions of points relative to each other did not change but the coordinate systems changed resulting in a rotation of the data. In a nutshell we created a transformed feature set which rows represent a night of a patient and columns the projection of the pairwise dissimilarities into the space of the first n eigenvectors and eigenvalues learned using one single night per subject. We performed class recognition by using a Random Forest (RF) [158] classifier with 50 trees. RFs are ensembles of weakly correlated decision trees that vote on the classification and have been shown to provide good generalization compared to individual decision trees. We used the transformed features representing the nights in the training set to train the ensemble of 50 classification trees. We report the results for the classification of the single nights and of the subjects as results of the vote of his assessed nights.

We applied the same procedure to evaluate the classification of the five disease severity grades (healthy, GOLD 1, GOLD 2, GOLD 3, and GOLD 4) and of the five dyspnoea scores (MMRC 0, MMRC 1, MMRC 2, MMRC 3, and MMRC 4) for a total of 1125 subjects and 690 subjects, respectively. We explored all possible combinations for different number of topics β_k (with k varying from 2 to 20) and selecting the first n eigenvectors (with n varying from 1 to 20).

For comparison we also evaluated the classification performances in the case standard features extracted during night-time such us total night sleeping time, number of nocturnal sleeping bouts and duration of sleeping bouts are used.