6.5 Results

Since PA measures during the weekdays and the weekend days are known to be different [44] results were computed separately. Kruskal–Wallis ANOVA test and Dunn's multiple comparisons test were used to determine significance of the results. A value of $P < 0.05$ was considered significant. Pearson's coefficient was used to investigate correlations in the entire dataset of 977 COPD patients. The findings are organized as follows. First, an interpretation of the discovered routines using the dataset of matched COPD and healthy subjects is given to highlight differences in intensity composition of routines and associated physiological responses. Second, results from the inference on window segments of 30 min are presented to qualitatively show two examples of daily routines patterns from a healthy subject and a COPD patient. Discovered routines are successively inferred on the first 6 h of all the assessed days of 66 healthy subjects and 977 COPD patients to quantitatively assess differences between selected groups of subjects and trends across different stages of COPD. Finally, the dataset of matched pairs is used to evaluate the routines in discriminative tasks, such as dividing healthy from patients and recognizing to which population each of the assessed days belongs.

6.5.1 Routine interpretation

Figure 39 illustrates the distribution of the discovered routines $\beta 1,..,15$ over the terms of the vocabulary. Three routines related to low intensity levels (R2, R8, R9), nine routines related to moderate-high intensity levels (R1, R3, R4, R5, R6, R10, R12, R14, R15) and one routine composed by a combination of VL, L, and MV descriptors (R7) were discovered in data from 66 COPD patients and 66 healthy subjects. Two separate routines characterizing, respectively, the sleeping behaviour (R11) and high intensity levels (R13) were also found. Each of these routines is characterized by the probability over the terms, where each term is defined by a certain combination of feature levels. For each routine, the three most important terms (i.e., terms of highest probabilities in this routine) can be found in Table XIII. The words composing a particular routine are listed together with their occurrence probability (e.g., the first word of R1: [MV 1 2 2] refers to a simplified feature vector V given by [MV, Ld1 , LST 2 , LV M 2] and has occurrence probability equal to 40%). Routines, that might be considered similar by only looking at their intensity composition, are different at the level of the descriptors. As an example, Figure 39 shows that R3 and R5 are formed for the 20% by MV descriptors and for the 50% by L descriptors. Analysing the letters that compose the descriptors shows that both the MV descriptor and the L descriptors of R5 have higher values in the feature level corresponding to the ST (second letter, see Table XIII). This illustrates that some routines differ depending on physiological responses. In absence of appropriate labels for these topics, such insights into routines are relevant for interpreting the inferences of topics per subject.