Reviewer’s report

Title: Using n-gram analysis to cluster heartbeat signals

Version: 1 Date: 30 March 2012

Reviewer: Shyamala Doraisamy

Reviewer’s report:

1. Minor Essential Revisions

2. In general, the paper presents a clear concept of using n-grams, symbolization and clustering to improve classification of heartbeat signal collections. However, further studies would be required to show robustness for heartbeat signals as opposed to being viable and realistic with other applications as discussed at the top of pg 13. This was not included in discussion and conclusion.

3. Abstract - Not necessary for sub-sections

4. CHF (abbreviation expanded in abstract but not in the first instance in main body of paper). Is there an expansion for RR?

5. Reference numbers in-text requires a thorough format checking as numbers are repeated outside the brackets. on pg 9, is it reference 16 or 17?

6. What is being referred to with 'Some early traditional linear methods ...' on top of pg 4

7. Reference Peng's team in-text as opposed to Peng et. al…. can be changed (to avoid personalization)

8. Pg 9 – Example 1 : Is the heart rate time series referring to the series of the data collection

9. Pg 10 – 2nd para – Cluster number 20, or is it number of clusters is 20/ k=20. Explanation of results does not include explanation of superiority in performance with CHF.

10. Pg 11 – since there is repetition of steps in generating the symbolic sequence, can these steps be summarized once and referenced to thereafter with changes in numbers (parameters).

11. Pg 13 – claim that the criterions of minimal total variance is not used in previous studies require reference to the previous studies being implied

12. Figures – unclear as to why short paragraphs of explanation is added under each figure caption

Level of interest: An article whose findings are important to those with closely
related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests