Reviewer's report

Title: Mining Casual Relationships among Clinical Variables for Cancer Diagnosis based on Bayesian Analysis

Version: Date: 17 October 2014

Reviewer: Dokyoon Kim

Reviewer's report:

The authors proposed a novel Bayesian learning strategy, Flexible K-dependence Bayesian Network, to identify causal relationships among clinical variables for cancer diagnosis. This paper addresses an important issue, which is an identification of causal-relationship between input variables associated with outcomes. Here are some comments for improving this manuscript.

Major Compulsory Revisions

1. An author should conduct the statistical tests on comparison between proposed method and other methods.

2. There are many previous studies that used colorectal cancer SEER dataset. The author should introduced and summarized them, especially in terms of causal-relationship between clinical variables associated survival outcome.

3. It would be nice if the author discuss about the application of the proposed method to the high dimensional genomic data. For this study, the author only used 19 variables so that the computation time to identify the causal-relationship between clinical variables would be reasonable. However, many papers published in Biodata Mining used (ultra) high-dimensional genomic data. For example, the input matrix consists of 1,000 samples and 1,000,000 SNPs. I'm really wondering if the proposed method will work for the high-dimensional genomic data in terms of the reasonable computational complexity.

4. There are lack of clinical implications based on the results. The author should add more useful clinical implications of causal-relationship between clinical variables and why those relationship could be associated with the survival outcome in colorectal cancer. I'm curious whether the causal-relationship can be detected only in colorectal cancer or in any other cancer types as well. It would be a great future work if the author uses other SEER dataset in several types of cancer and compare them.

5. Data description should be improved. The author should explain how to choose 19 variables among 200 variables. In addition, the number of total samples should be mentioned as well.

Minor Essential Revisions

1. The author should add more description for the figure legends in order to
improve the readability for readers.

**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:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.