

## Symbolic Tree for Prognosis of Hepato Cellular Carcinoma Patients

Taerim Lee<sup>(1)\*</sup>, Hyosuk Lee<sup>(2)</sup>, and Edwin Diday<sup>(3)</sup>

<sup>(1)</sup> Department of Information & Statistics, Korea National Open University, Korea  
trlee@knou.ac.kr

<sup>(2)</sup> Department of Internal Medicine, Medical College, Seoul National University, Korea  
hsleemd@snu.ac.kr

<sup>(3)</sup> Department of Computer Science & Statistics University 9 at Dauphine Paris  
diday@ceremade.dauphine.fr

### Abstract

For the purpose of defining the analytic prognosis of Hepato Cellular Carcinoma, to fit the symbolic tree based survival model and to find the significant SNP covariates to detect the progression duration and survival time from liver cirrhosis to malignant liver cancer.

Using the symbolic data analysis which extends the data mining and exploratory data analysis, we can suggest the tree structured survival model on continuous response variable with new knowledge mining SDA approach. With the SDA classification, the new SDA application for huge Genomic SNP data, the availability could be proved statistically.

Visualization with SDA pyramidal clustering for hidden relationship of SNP for HCC, we can show the representativeness of data structure and to produce the quantitative statistics for evaluation of validity and stability of clustering.

The objectives of this paper is as follows:

1. Confirm of validity of application of SDA to the tree structured progression model.
2. Quantify the clinical lab data and SNP data for early diagnosis of HCC.
3. Construct the symbolic tree structured HCC progression and survival model.
4. Setting the Nomogram and provide the genomic information for HCC prevention and control for lifelong health of HCC patient.

This paper can get the following expected effect.

- To construct the representative model for HCC survival time and causal association with their SNP genomic data which contribute to promote the more healthy condition for old HCC patients and to make life long survival of HCC.
- To fit the simple and easy interpretation tree structured survival model which reduced from huge clinical and genomic data under the new statistical theory of knowledge mining SDA.
- To save the cost and time of HCC patients under the treatment with the reduced prognostic covariates and critical values for early diagnosis and prolong the life.
- To define advanced statistical theory with the statistical index of validity and stability of optimal number of cluster in SDA tree structured progression model of HCC for the future new patient's life long health.