A Medical Decision Support System for Polyp Screening by Using Fuzzy Classification Trees
I-Jen Chiang, Ming-Jium Shieh, Jane Yung-jen Hsu, Jau-Min Wong

I. BACKGROUND
Colorectal cancer (CRC) has become the third causes (10.21% in 2001) of cancer death, which are the leading causes (27% in 2001) of death, in Taiwan. Despite advances in treatment, early detection can probably reduce CRC mortality more than any other approaches. Therefore, it is important to develop a cost-effective cancer screening policy in the hopes of reducing CRC mortality by detecting lesions at any early, curable stage.

II. INTRODUCTION
The prevalence of adenomatous polyp varies geographically in parallel with the incidence of colorectal cancer and an increasing risk of colorectal cancer [5], [6]. The concept is now widely accepted that adenomas are precancerous lesions and will degenerate into cancers ultimately. Nowadays, the majority of the pathogeneses of the colorectal cancer are attributed to the adenoma-adenocarcinoma sequence. Hence, the identification and removal of the precancerous lesion, an adenomatous polyp, has significant clinical implications and is now commonly recommended for the control of CRC. Endoscopy is considered the most sensitive diagnostic modality for detection of colorectal polyps. However, the effort and eventual cost involved based on this surveillance strategy are potentially enormous and not practical, except for high-risk groups. Owing to the shortage of medical resources at present, it is important to develop a most cost-effective and safe screening method to predict the existence of adenomatous polyps.

In order to determine the predictive value of the risk factors related to the existence of rectosigmoid colon polyps, physicians evaluate all putative risk factors obtained from checkup items. Bias inevitably occurs from this assumption, in that only factors that have been selected can be shown to have association. A collection of physical checkup data with the patients who underwent sigmoidoscopy enrolled for the polyp screening analysis.

Highly uncertain and noise among the patient records make it hard to be correctly predicted. Therefore, a vague classification method is needed to generate a predictive model that is able to classify an instance into all possible classes and each class is associated with a degree that shows how possible an instance is in that class. According to these degrees, we can discriminate the more possible classes from the less possible classes.

Fuzzy classification trees [1], [2] have been presented to generate a vague model for polyp screening on providing all possible classes associated with their degrees. Fuzzy classification trees which integrate decision tree techniques and fuzzy classifiers, provide the simple and efficient way to generate the classification model that can suffer from inadequately or improperly expressing and handling the vagueness and ambiguity associated with human thinking and perception [7]. In a fuzzy classification tree, an instance has a membership value at each leaf node. Instead of determining a single class for any given instance, fuzzy classification trees can predict the degree of possibility for every class. Using information-based measures, there is no need to generate multiple classification trees. Therefore, it requires less time and space than decision forests [3].

Comparing the classification results with C4.5 [4] by using three-fold cross validation testing on the patient records, we will see that the error rate on false negative of FCT is less than the error rates on false negative of C4.5. That is, FCT is more sensitive than C4.5. The decisions of C4.5 are always biased to the majority, if only a small proportion of population will get the disease. In medical applications, it is important that a classifier should give the estimate degrees of all potential classes.

REFERENCES