SPANDIDOS PUBLICATIONS

MOLECULAR AND CLINICAL ONCOLOGY 13: 46, 2020

Predictability of postoperative recurrence on hepatocellular carcinoma through data mining method

SHUICHI IWAHASHI[1], A. AMMAR GHAIBEH[2], MITSUO SHIMADA[1], YUJI MORINE[1], SATORU IMURA[1], TETSUYA IKEMOTO[1], YU SAITO[1] and JUN HIROSE[2]

Departments of [1]Surgery and [2]Medical Informatics, Institute of Health Biosciences, The University of Tokushima, Kuramoto-cho, Tokushima 770-8503, Japan

Received November 28, 2019; Accepted April 27, 2020

DOI: 10.3892/mco.2020.2116

Abstract. Hepatocellular carcinoma (HCC) is a highly lethal tumor and the majority of postoperative patients experience recurrence. In the present study, we focus on the predictability of postoperative recurrence on HCC through the data mining method. In total, 323 patients with HCC who underwent hepatic resection were included in the present study, 156 of whom suffered from cancer recurrence. Clinicopathological data including prognosis were analyzed using the data mining method for the predictability of postoperative recurrence on HCC. The resulting alternating decision tree (ADT) was described using data mining method. This tree was validated using a 10-fold cross validation process. The average and standard deviation of the accuracy, sensitivity, and specificity were 69.0±8.2, 59.7±14.5 and 77.7±10.2%, respectively. The identified postoperative recurrence factors were age, viral hepatitis, stage, GOT and T-cholesterol. Data mining method could identify the factors associated at different levels of significance with postoperative recurrence of HCC. These factors could help to predict the postoperative recurrence of HCC.

Introduction

The artificial intelligence (AI) has been widespread globally in various fields including the medical field. AI is used to identify key information from the large mass of academic reports and experimental data. The AI system then analyzes this database and displays appropriate information tailored to the personal preferences of medical professions. Subsequently, the AI system can indicate the most appropriate and useful information for the individual patient concerned (1).

Correspondence to: Dr Shuichi Iwahashi, Department of Surgery, Institute of Health Biosciences, The University of Tokushima, 3-18-15 Kuramoto-cho, Tokushima 770-8503, Japan
E-mail: shuichiiwahashi@yahoo.co.jp

Key words: data mining method, hepatocellular carcinoma, slternating decision tree, postoperative recurrence, artificial intelligence

The data mining method is a statistical method that employs the AI system. In conventional analysis, data are collected and organized for the creation of a summary or a graph. However, in the data mining system, the data are used to form a pattern. Many health workers have used this data mining method in various aspects of medicine due to their promising results (2,3). Data mining aims to extract useful information from available data by applying techniques including databases, statistics, and visualization (4). The data mining method is useful in the diagnosis, therapy and prognosis of patients. Regarding diagnosis, patterns in multivariate patient attributes are identified and classified allowing for selection from available treatments based on their effectiveness and suitability for individual patients. Concerning prognosis, future outcomes of patients can be predicted based on previous experience and current conditions.

In a recent report, we demonstrated that systemic inflammation was associated with increased prevalence of type 2 diabetes on an alternating decision tree (ADT) using the data mining method (5). In regard to the decision tree analysis, it was previously reported that this method was effective in predicting the future development of the aggressive behavior of dural arteriovenous fistulas (6,7).

Hepatocellular carcinoma (HCC) is a highly lethal tumor (8), with over 80% patients who received curative surgery having experienced recurrences within five years after surgery (9-11). It is important to identify the recurrence factor on HCC; however, to the best of our knowledge, no reports have focused on recurrence factors in HCC using the data mining method. Therefore, the aim of this study was to examine the utility of data mining on the predictability postoperative recurrence in HCC.

Materials and methods

Patients and methods. A total of 323 patients who received hepatic resections were included in this study. One hundred and fifty-six of these patients had suffered from cancer recurrence while 167 patients had no recurrence.

Background factors were obtained preoperatively. Clinicopathological data and patient prognosis were analyzed using the data mining method for predictability of postoperative recurrence on HCC. The data mining method was performed using Rapid Miner Studio version 6.4 software.

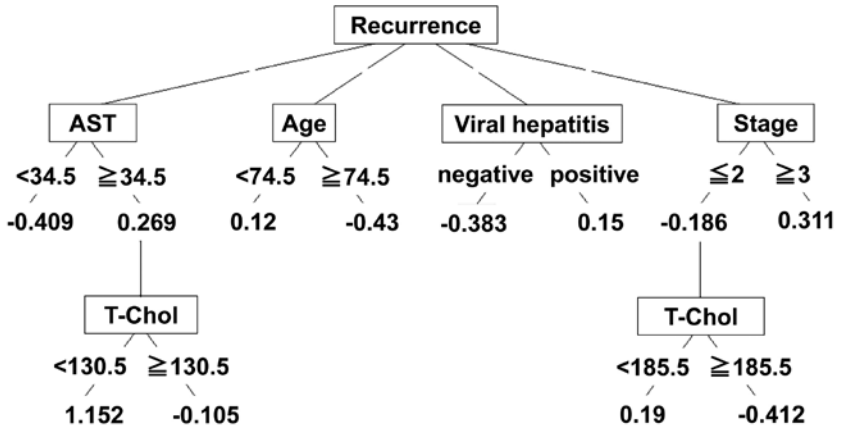

Figure 1. Alternating decision tree analysis of postoperative recurrence on HCC. HCC, hepatocellular carcinoma.

This study was authorized by the Institutional Review Board of the University of Tokushima Graduate School (approved ID no. 3215).

Data mining method. Decision trees are classification methods widely used in data mining because they are visually informative. ADT is generated by recursively dividing and partitioning the patients according to the values of a statistically important factor. The statistical importance of each factor is evaluated with the C4.5 algorithm using the entropy function applied in information theory (6). Factors with high entropy reduction are considered of statistical importance. First, the most important factor for all patients is selected, and the patients are divided into subgroups based on the selected factor. The procedure is then repeated, and each subgroup is divided again based on its most important factor until all the patients in a specific subgroup are in the same class or the subgroup is no longer subject to further splitting. In data mining, the group that contains all the patients is the root, each subgroup is a node, and the subgroups not subject to further division are the leaves. The leaves embody the final decision. Using the recorded factors of a new patient, risk can be predicted by following the decision tree path from the root to a node to one of the leaves. For decision tree accuracy, it is important that the tree provide accurate results, not only with respect to existing cases, but also to new cases. One method for estimating the performance of the decision tree with new cases is to test the tree by using a test case that is different from the same population. However, in this approach the calculated performance is highly dependent on the selected test case. To overcome the effect of this dependency, we used 10-fold cross-validation to estimate the decision tree accuracy, sensitivity, and specificity. We randomly divided the existing data into 10 equally sized subsets and used each of the 10 subsets once as a testing set. The remaining nine subsets were used for generating a decision tree. Finally, the overall accuracy, sensitivity, and specificity of the decision tree were calculated as the average of its performance with the 10 test sets (6).

Naive Bayes assume that all the features are independent with each other. It has often performed well on the real data. If the features are not redundant then this algorithm will work with best accuracy. Naive Bayes is simplified version of bayes theorem that is used to classify the unknown instances into relevant class. The probability of each class is calculated based on given attribute value associated with each tuple. We used 10-fold cross-validation to estimate the accuracy, sensitivity, and specificity.

Results

ADT analysis of postoperative recurrence on HCC is shown in Fig. 1. The resulting tree was validated using a 10-fold cross validation process. The average of the accuracy, sensitivity, and specificity were 69.0±8.2, 59.7±14.5 and 77.7±10.2%, respectively. The identified postoperative recurrence factors were age, viral hepatitis, stage, AST and Total-cholesterol.

Results of the naive Bayes analysis identified postoperative recurrence factors including AST, tumor number, HBeAb, Sf and surgical margin (Fig. 2). Probability (P) (Recurrence. negative) was calculated as: P_{GOT} x P_{HBeAb} x P_{Sf} x P_{sm} x $P_{Tumor\ number}$ and P (Recurrence. positive) was calculated as: P_{GOT} x P_{HBeAb} x P_{Sf} x P_{sm} x $P_{Tumor\ number}$. The average of the accuracy, sensitivity, and specificity were 69.6±7.7, 56.2±13.5 and 81.9±7.4%, respectively.

Discussion

In the present study, we mentioned the predictability of postoperative recurrence on HCC through the data mining method. To the best of our knowledge, this is the first report showing the predictability of postoperative recurrence on HCC through the data mining method.

In regard to HCC recurrence, there were some reports on the prediction of liver cancer recurrence with machine learning (12-14). However, there was no report of data mining method on HCC recurrence. To the best of our knowledge, this study was the first report of data mining method in HCC recurrence. Furthermore, compared with other statistical analyses, ADT analysis created the map between input and output factors and the map is learned automatically from available data and this is not possible with classical statistical methods. An important feature of ADT analysis exists in this surface in addition to yielding a highly accurate final prediction (6). There were the additional merits of the data mining method. The representation of data in the form of the tree is easily understood and this method can handle multidimensional

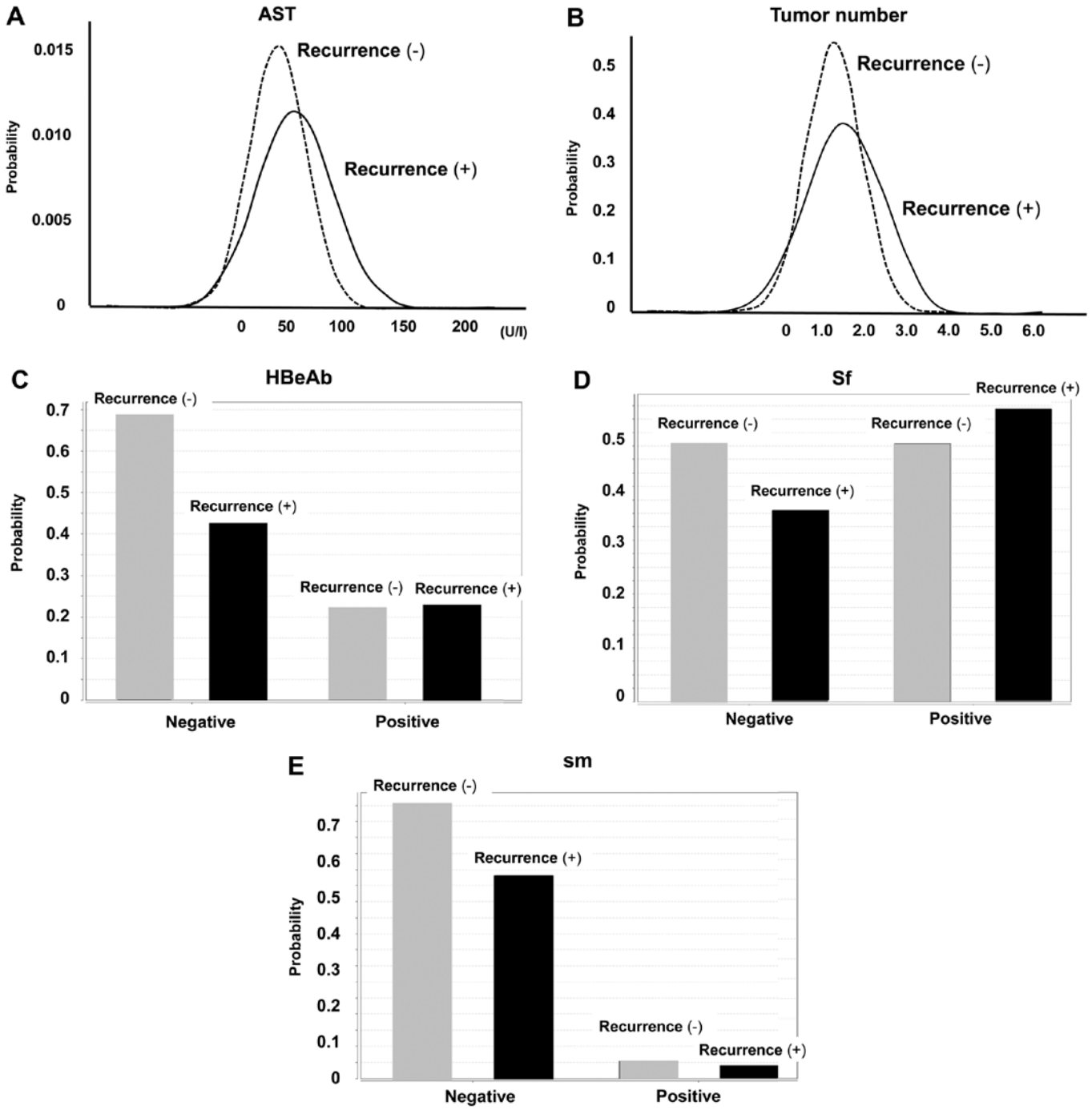

Figure 2. Naive Bayes analysis of postoperative recurrence on HCC. (A) AST, (B) tumor number, (C) HBeAb, (D) Sf and (E) surgical margin (sm). HCC, hepatocellular carcinoma.

data. Recently, the data mining method has been reported in the field of heart disease (15).

Regarding the effectiveness of the ADT analysis in solid tumor, the authors of that study mentioned the prediction of lymph node metastasis in breast cancer patients using a decision tree. The ADT model showed a high accuracy for predicting metastasis in patients with breast cancer (16). Previous findings showed that the expression of CD24 and CD44 would be useful predictive markers in breast cancer patients, via ADT analysis (17). ADT analysis has been expected to be the beneficial method in determining the predictability of tumor aggressiveness of solid tumors in the future.

In addition, results of the naive Bayes analysis revealed AST and the surgical margin as postoperative recurrence factors. Previous findings demonstrated that the AST/ALT ratio and surgical margin constituted risk factors for the recurrence of HCC patients (18). Those data are consistent with our findings in this study.

Nevertheless, the present study has a limitation. In this report, analysis was performed using only clinical factors. The roles of transcriptional factors about HCC recurrences have been elucidated gradually (19). In future, we plan to estimate the role of genetic or transcriptional factors on HCC recurrence using the data mining method.

In conclusion, the data mining method could identify factors with postoperative recurrence of HCC. The factors could be useful to predict the postoperative recurrence of HCC.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Authors' contributions

SI analyzed the data and drafted the manuscript. AAG performed the analysis of this study. MS conceived this study and reviewed the manuscript. YM, SI and TI contributed to the study design and helped conduct the analysis. YS provided clinical data. JH designed the experiments and supervised the study. All authors read and approved the final manuscript.

Ethics approval and consent to participate

This study was authorized by the Institutional Review Board of the University of Tokushima Graduate School (approved ID no. 3215).

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. El Naqa I, Haider MA, Giger ML and Ten Haken RK: Artificial intelligence: Reshaping the practice of radiological sciences in the 21st century. Br J Radiol 93: 1106, 2020.
2. Bellazzi R and Zupan B: Predictive data mining in clinical medicine: Current issues and guidelines. Int J Med Inform 77: 81-97, 2008.
3. Setoguchi Y, Ghaibeh AA, Mitani K, Abe Y, Hashimoto I and Moriguchi H: Predictability of pressure ulcers based on operation duration, transfer activity and body mass index through the use of an alternating decision tree. J Med Invest 63: 248-255, 2016.
4. Moreira LB and Namen AA: A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. Comput Methods Programs Biomed 165: 139-149, 2018.
5. Uemura H, Ghaibeh AA, Katsuura-Kamano S, Yamaguchi M, Bahari T, Ishizu M, Moriguchi H and Arisawa K: Systemic inflammation and family history in relation to the prevalence of type 2 diabetes based on an alternating decision tree. Sci Rep 7: 45502, 2017.
6. Satomi J, Ghaibeh AA, Moriguchi H and Nagahiro S: Predictability of the future development of aggressive behavior of cranial dural arteriovenous fistulas based on decision tree analysis. J Neurosurg 123: 86-90, 2015.
7. Podgorelec V, Kokol P, Stiglic B and Rozman I: Decision trees: An overview and their use in medicine. J Med Syst 26: 445-463, 2002.
8. Yang JD and Roberts LR: Hepatocellular carcinoma: A global view. Nat Rev Gastroenterol Hepatol 7: 448-458, 2010.
9. Imamura H, Matsuyama Y, Tanaka E, Ohkubo T, Hasegawa K, Miyagawa S, Sugawara Y, Minagawa M, Takayama T, Kawasaki S and Makuuchi M: Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy. J Hepatol 38: 200-207, 2003.
10. Poon RT, Fan ST, Ng IO, Lo CM, Liu CL and Wong J: Different risk factors and prognosis for early and late intrahepatic recurrence after resection of hepatocellular carcinoma. Cancer 89: 500-507, 2000.
11. de Lope CR, Tremosini S, Forner A, Reig M and Bruix J: Management of HCC. J Hepatol 56 (Suppl 1): S75-S87, 2012.
12. Liang JD, Ping XO, Tseng YJ, Huang GT, Lai F and Yang PM: Recurrence predictive emodels for patients with hepatocellular carcinoma after radio frequency ablation using support vector machines with feature selection methods. Comput Methods Programs Biomed 117: 425-434, 2014.
13. Divya R and Radha P: An Optimized HCC recurrence prediction using APO algorithm multiple time series clinical liver cancer dataset. J Med Syst 22: 193, 2019.
14. Xu D, Sheng JQ, Hu PJ, Huang TS and Lee WC: Predicting hepatocellular carcinoma recurrences: A data-driven multiclass classification method incorporating latent variables. J Biomed Inform 96: 103237, 2019.
15. Maheswari S and Pitchai R: Heart disease prediction system using decision tree and naive bayes algorithm. Curr Med Imaging Rev 15: 712-717, 2019.
16. Takada M, Sugimoto M, Naito Y, Moon HG, Han W, Noh DY, Kondo M, Kuroi K, Sasano H, Inamoto T, *et al*: Prediction of axillary lymph node metastasis in primary breast cancer patients using a decision tree-based model. BMC Med Inform Decis Mak 12: 54, 2012.
17. Horiguchi K, Toi M, Horiguchi S, Sugimoto M, Naito Y, Hayashi Y, Ueno T, Ohno S, Funata N, Kuroi K, *et al*: Predictive value of CD24 and CD44 for neoadjuvant chemotherapy response and prognosis in primary breast cancer patients. J Med Dent Sci 57: 165-175, 2010.
18. Wang ZX, Jiang CP, Cao Y, Zhang G, Chen WB and Ding YT: Preoperative serum liver enzyme markers for predicting early recurrence after curative resection of hepatocellular carcinoma. Hepatobiliary Pancreat Dis Int 14: 178-185, 2015.
19. Tovuu LO, Imura S, Utsunomiya T, Morine Y, Ikemoto T, Arakawa Y, Mori H, Hanaoka J, Kanamoto M, Sugimoto K, *et al*: Role of CD44 expression in non-tumor tissue on intrahepatic recurrence of hepatocellular carcinoma. Int J Clin Oncol 18: 651-656, 2013.