

The Application of Decision Tree Model in Differential Diagnosis of Myelodysplastic Syndrome and Aplastic Anemia

J. SONG*, Y. WANG, MEIRONG YANG¹ AND RUI LIU

Chemoradiotherapy Department, North China University of Science and Technology Affiliated Hospital, Tangshan, Hebei, 063000, P.R. China, ¹Department of Hematology, North China University of Science and Technology Affiliated Hospital, Tangshan, Hebei, 063000, P.R. China

Song *et al.*: Decision Tree Model in Myelodysplastic Syndrome and Aplastic Anemia

The differential diagnosis models of C5.0, classification and regression trees and QUEST decision-making trees were established to provide a basis for the diagnosis of myelodysplastic syndrome and aplastic anemia. Patients with myelodysplastic syndrome and aplastic anemia hospitalized in hematology hospital of the Chinese Academy of Medical Sciences from January 2008 to December 2014 were selected as the study subjects. The general condition, clinical examination and laboratory examination data of the patients were collected using a self-designed questionnaire. The differential diagnosis models of C5.0, classification and regression trees and QUEST decision-making trees were established with 2 blood diseases as dependent variables and their differential diagnosis indices as independent variables. The performance of the 3 models was compared using the indices of accuracy, sensitivity and specificity. The prediction accuracy, sensitivity, specificity, F1 and Youden's index of C5.0 decision-making tree model were 78.12, 87.50, 66.67, 81.48 and 0.54 %, respectively. The prediction accuracy, sensitivity, specificity, F1 and Youden's index of classification and regression trees decision-making tree model were 73.75, 76.14, 70.83, 81.03 and 0.51 %, respectively. The prediction accuracy, sensitivity, specificity, F1 and Youden's index of QUEST decision-making tree model were 76.88, 89.77, 61.11, 76.14 and 0.47 %, respectively. There was no statistical difference in the accuracy, specificity, F1 and Youden's index of the 3 models. The specificity of C5.0 decision-making tree and QUEST decision-making tree models were significantly higher than classification and regression trees decision-making tree model ($p < 0.05$). The C5.0 decision-making tree model has higher prediction accuracy, sensitivity, specificity, F1, and Youden's index, which were superior to the other two models and can be used as auxiliary models for differential diagnosis of myelodysplastic syndrome and aplastic anemia.

Key words: Myelodysplastic syndrome, aplastic anemia, decision tree, differential diagnosis

Myelodysplastic syndrome (MDS) is a kind of acquired hematopoietic stem/progenitor cell clonal disease with the clinical features of myelocyte dyshaematopoiesis and transferring to acute myelogenous leukemia at high risk. Aplastic anemia (AA) refers to primary bone marrow hematopoiesis failure syndrome with etiology unknown, which mainly represents bone marrow hematopoiesis hypofunction and pancytopenia, with bleeding and infection. As treatment and prognosis of the 2 diseases are obviously different, differential diagnosis of MDS and AA has important clinical significance. At present, differential diagnosis of MDS and AA is mainly made from hematology, cell morphology and cytogenetics in clinic. At different stages of disease development, both MDS and AA

patients' peripheral blood (PB) could show reduction in one type blood cell, two types of blood cells or even 3 types of blood cells at the same time^[1,2]. The dyshaematopoiesis is a key index for clinical diagnosis of MDS, but there are the defects of poor repeatability, specificity and sensitivity. But the dyshaematopoiesis shows non-specificity because of some AA patients also exhibit dyshaematopoiesis or the cases of MDS without dyshaematopoiesis being found in some studies^[2,3]. In the past, cytogenetic abnormality is considered to be a reliable criterion to diagnose MDS, but the chromosomal abnormality detection rate of MDS patients was 40-60 %, which was even lower in hypo-MDS group and it is clear that MDS abnormal cytogenetics ratio is not high, which indicates that this index is of poor

*Address for correspondence

E-mail: wlqpf@126.com

specificity. In recent years, flow cytometry (FCM) has received more attention in the differential diagnosis of AA and MDS, but the sensitivity of using single immunophenotyping index to diagnose hypo-MDS and AA is too low and it is too difficult to use FCM, a milestone in morphologic diagnosis of MDS, to evaluate erythroid dyshaematopoiesis, which has limited the extensive use of FCM in MDS diagnosis^[4-9]. It can be concluded that pathological and clinical features of hypo-MDS and AA are very similar and there are many differential diagnosis indices, but specificity of both is not high. Thus, it is difficult to make differential diagnosis for these 2 diseases clinically.

Data Mining, also known as Knowledge Discovery in Database (KDD), is a process to transfer much information in database into valuable knowledge. Classification is a very important task in data mining, and its most common methods include logistic regression, neural network, decision tree, Bayesian networks and support vector machines. Compared to neural networks and Bayesian, the classification rules of decision tree are easier to be explained and the rules formed are easy to be accepted and understood by clinicians, so as to better assist the clinicians to make differential diagnosis. At present, decision tree models have been widely used in the field of medical classification^[10-12].

As to the problem of misdiagnosis rate of MDS and AA is high in clinical practice, this research is expected to provide clinical evidence for clinicians to make differential diagnosis of these 2 diseases through collecting differential indices and building MDS and AA decision tree classification models.

MATERIALS AND METHODS

Patients and diagnostic criteria:

Seven hundred and eighty new primary MDS and AA patients, who were diagnosed at the Chinese Academy of Medical Sciences Hematology Hospital upon expert consultation from January 2008 to December 2014, were selected as study subjects. There were 413 MDS cases, which included 227 male (54.96 %) and 186 female (45.04 %) patients in the age range of 3-81 y with a median age of 38 y. The AA cases were 367, of which 203 (55.31 %) were male and the rest 164 (44.69 %) were females and the age range of the patients was 3-80 y with a median age of 28 y. All the selected MDS and AA cases conformed with the MDS classification criteria revised by WHO in 2008 and criteria of diagnostic and therapeutically effects of blood diseases (Edition III)^[13,14].

Observational indices:

Patient related information was collected, which included age, gender, the indices related to hematology, virology, serology, immunology, the analysis parameters of blood and marrow smears, inspection result of FCM analysis and indices of stem cell colony culture.

Statistical methods of data analysis:

EpiData 3.1 was used to build a database, then SPSS 22.0 software was adopted to analyze the data. The enumeration data were represented as ratio and percentage and chi-square test was used for comparing differences among groups, in which the differences were considered significant at $p < 0.05$. SPSS modeler 17.0 was applied to build the 3 decision tree models based on C5.0, Classification and regression trees (CART) and QUEST respectively, and then the merits of 3 models were determined by forecast accuracy, precision, sensitivity, specificity, F1 measure and Youden's index.

RESULTS AND DISCUSSION

Making use of build-in partitioning function of the C5.0 decision tree model, the data were divided into training set and test set randomly, with training set accounting for 80 % and test set accounting for 20 %. Then, the classification model for MDS and AA were built based on partitioned data and C5.0 decision tree model. This model had 3 layers of nodes and the model results were shown in fig. 1. According to the results of C5.0 decision tree model, the decision rules to differentiate the 2 diseases were extracted, as the decision tree rules should start from root node, i.e. mature lymphocyte ratio to the leaf nodes. For example, node 11 referred that when mature lymphocyte ratio was more than 45.45 % in the bone marrow flow cytometry test and myeloblast was not found on bone marrow smear and rubriblast ratio was not more than 3 %, then the probability for the patient to have AA was 92.8 %. Among the 620 cases in the training set, this model had accurately differentiated 520 cases, accuracy rate of which was 83.87 %. Among the 160 cases in the test set, this model had accurately differentiated 125 cases, the accuracy rate of which was 78.12 % as shown in Table 1.

The data was divided into the training set and the test set randomly, training set accounting for 80 % and test set accounting for 20 %, according to the built-in partitioning function of the model. CART algorithm needed to divide the training set into 2 parts, including one part used to prevent the model from over-fitting. In

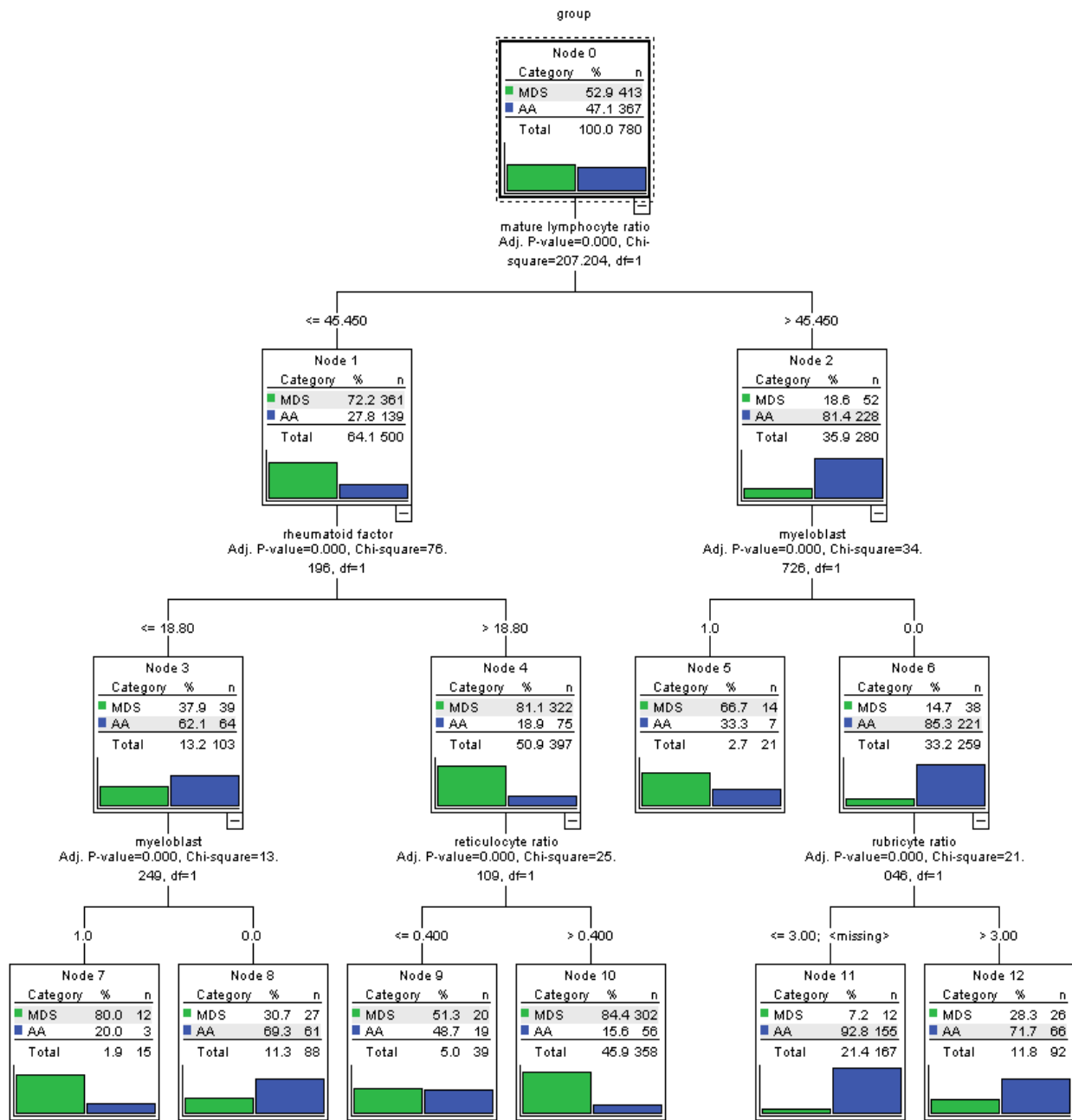


Fig. 1: The C5.0 Decision Tree of (■) MDS and (■) AA

TABLE 1: THE RESULT OF TRAINING SET AND TEST SET IN C5.0 MODEL

Partition	n	Correct (%)	Error (%)
Training set	620	520 (83.87)	100 (16.13)
Test set	160	125 (78.12)	35 (21.88)

this study, 30 % of the training set was randomly selected to prevent the model from over-fitting and minimum impurity change of the model was set as 0.0001. Then, a classification model for MDS and AA was built on the basis of CART method, as shown in fig. 2. This model had 3 layers of nodes. For example, node 2 indicated that the probability for the patient to have AA was 79.3 %

when mature lymphocyte ratio was more than 43.425 % in flow cytometry test result of bone marrow. Then, node 6 represented that when mature lymphocyte ratio was equal or lower than 43.425 % in flow cytometry test result of bone marrow and the rheumatoid factor was not more than 18.8 IU/ml in immunological tests and myeloblasts appeared on bone marrow smear, then the patient was determined to be MDS. Among the 620 cases in the training set, 478 cases have been correctly classified with an accuracy rate of 77.10 %; among the 160 cases in the test set, 118 cases have been accurately differentiated and the accuracy rate was 73.75 % (Table 2).

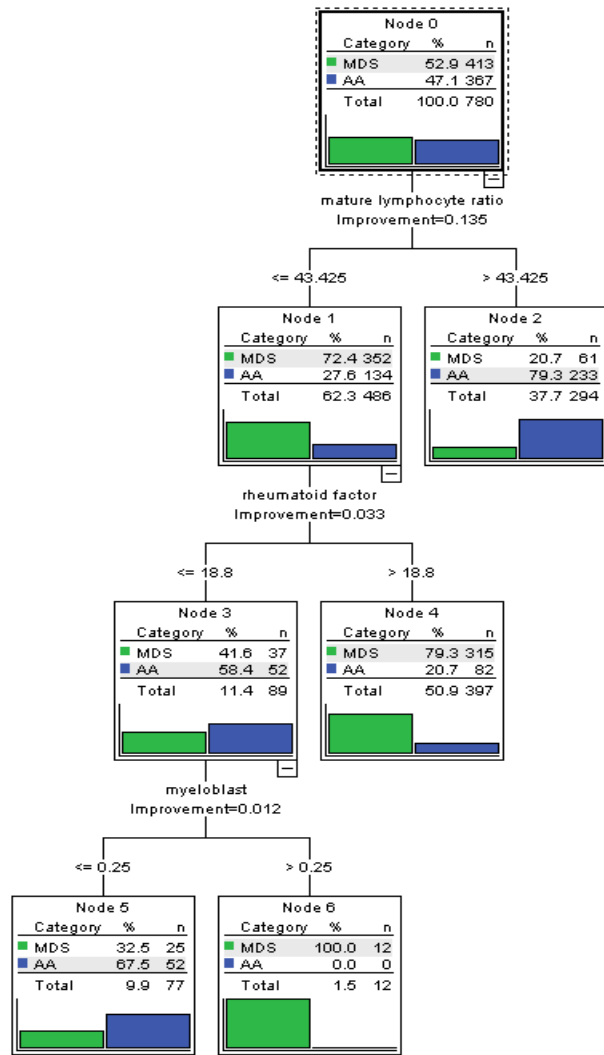


Fig. 2: The CART Decision Tree of (■) MDS and (■) AA

TABLE 2: THE RESULT OF TRAINING AND TEST SET IN CART MODEL.

Partition	n	Correct (9%)	Error (%)
Training set	620	478 (77.10)	142 (22.90)
Test set	160	118 (73.75)	42 (26.25)

The data was divided into the training set and the test set randomly that 80 % of data was from the training set and 20 % of that was from the test set. QUEST algorithm also required that the training set data be divided into 2 parts to avoiding over-fitting. In this study, 30 % of the training set was selected to prevent the model from over-fitting and model partition significance level was 0.05. Then the QUEST method was used to build the classification model for MDS and AA and the model was shown as fig. 3. It showed that this model was a binary tree with 3 layers of nodes and its classification rules were relatively simple. For example, node 2 indicated that when mature lymphocyte ratio was equal or lower 44.788 % in bone marrow flow cytometry test,

then probability for the patient to have MDS was 80.3 %. Node 5 suggested that when mature lymphocyte ratio was more than 44.788 % in bone marrow flow cytometry test, immature granulocyte ratio was equal or lower 22.74 % and immature red blood cell ratio was equal or lower 31.776 %, then probability for the patient to have AA was 95.7 % and when immature red blood cell ratio was more than 31.776 %, then probability for the patient to have MDS was 89.5 % (node 6). In the training set of QUEST model, 74.68 % of cases (463/620) were correctly differentiated and in the test set, 76.88 % of cases (123/160) were accurately differentiated as shown in Table 3.

The forecasting accuracy was the most visual index used for comprehensive comparison of these models. Through analysing forecast conditions of the test

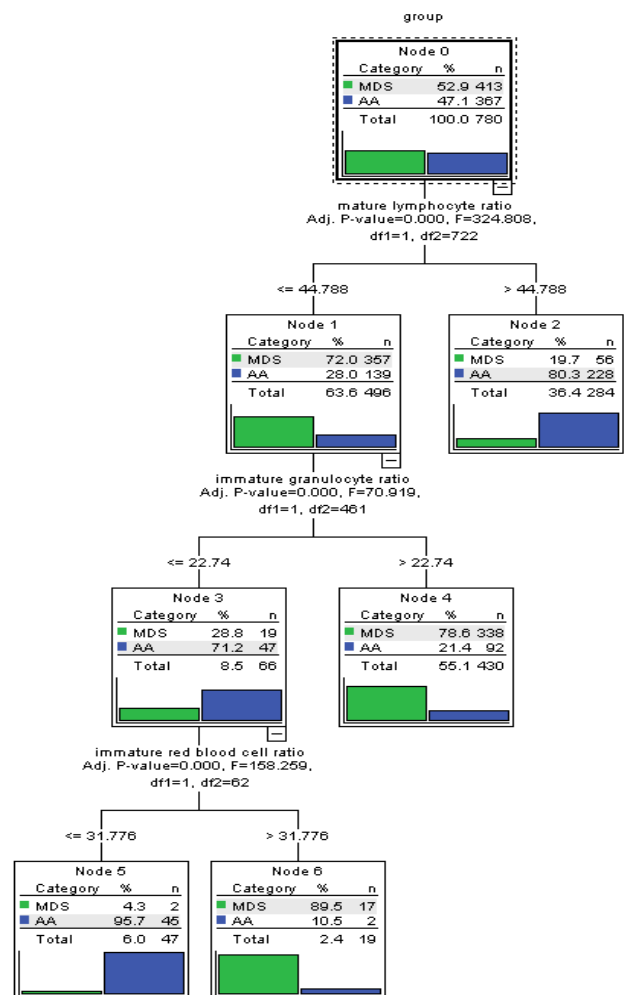


Fig. 3: The QUEST Decision Tree of (■) MDS and (■) AA

TABLE 3: THE RESULT OF TRAINING AND TEST SET IN QUEST MODEL

Partition	n	Correct (%)	Error (%)
Training set	620	463 (74.68)	157 (25.32)
Test set	160	123 (76.88)	37 (23.12)

set in all the 3 models, it was found that C5.0 model demonstrated the highest forecast accuracy, which was 78.12 %, and QUEST model's forecast accuracy was 76.88 %. Compared to these models, CART model's forecast accuracy was lower, which was 73.75 %. However, the differences of accuracy among models was not statistically significant (Table 4).

The effects of MDS and AA on the human body are different with MDS has a higher probability of transforming to malignant leukemia^[15]. It is essential to differentiate MDS from AA during early treatment. Thus, in order to facilitate analysis, MDS cases were set to be positive and AA cases were set to be negative in this study.

The two indices of precision and sensitivity could reflect the classification of MDS cases. Precision represented the percent of positive cases correctly classified among the total cases considered positive; sensitivity represented the probability of the positive cases correctly classified to the total real positive cases. As to C5.0, CART and QUEST models, the precision was 76.24, 76.14 and 73.83 % respectively, which was not significantly different between models, the sensitivity were 87.50 %, 76.14 % and 89.77 % separately. There was differences among the 3 models ($p < 0.05$), and QUEST model had the highest sensitivity as showed in Tables 5 and 6.

For a model to be successful both the precision and sensitivity should be very high. However, these often contradict with each other. In this study, F1 measure was introduced to make a comprehensive evaluation. F1 measure is the harmonic mean of precision and sensitivity and it can be calculated as, $F1 = 2 \times (\text{precision} \times \text{sensitivity}) / (\text{precision} + \text{sensitivity})$

The results indicated that among the 3 models, the value of F1 measure was the highest for C5.0 model at 81.48 % and hence C5.0 model should have greater ability to make comprehensive forecast and classification for MDS cases. However, the F1 measure value of QUEST model was similar to that of C5.0 model at 81.03 % and the QUEST model also should possess similar differentiating capacity for MDS cases. On the other

TABLE 4: THE ACCURACY COMPARISON OF 3 MODELS

Model	correct	error	Accuracy (%)
C5.0	125	35	78.12
CART	118	42	73.75
QUEST	123	37	76.88

Note: $\chi^2 = 0.897$, $P = 0.638$

TABLE 5: THE PRECISION COMPARISON OF 3 MODELS

Model	TP	FP	Precision (%)
C5.0	77	24	76.24
CART	67	21	76.14
QUEST	79	28	73.83

TP (true positive) represented number of the MDS cases correctly classified; FP (false positive) represented number of the AA cases wrongly classified as MDS. $\chi^2 = 0.205$, $P = 0.903$

TABLE 6: THE SENSITIVITY COMPARISON OF 3 MODELS

Model	TP	FN	Sensitivity (%)
C5.0	77	11	87.50
CART	67	21	76.14
QUEST	79	9	89.77

TP (true positive) represented number of the MDS cases correctly classified; FP (false positive) represented number of the MDS cases wrongly classified as AA. $\chi^2 = 7.161$, $P = 0.028$

TABLE 7: THE SPECIFICITY COMPARISON OF 3 MODELS

Model	TN	FP	Specificity (%)
C5.0	48	24	66.67
CART	51	21	70.83
QUEST	44	28	61.11

TP (true positive) represented number of the AA cases correctly classified; FP (false positive) represented number of the AA cases wrongly classified as MDS. $\chi^2 = 1.531$, $P = 0.465$

hand the F1 measure value of CART model was low at 76.14 %.

As mentioned above, it is essential to differentiate MDS patents from AA patients; in the same way, it is also very important to differentiate AA from MDS patients and in this investigation this problem could be addressed with specificity. Specificity represented the % of negative cases correctly differentiated to the total negative cases. As to C5.0, CART and QUEST models, the specificity were 66.67, 70.83 and 61.11 %, respectively and these differences were not significant as shown in Table 7.

Youden's index, also known as correct index, represented the capacity of the model to find real MDS and AA cases. The range of Youden's index is 0 to 1. The higher the value, the truer the result. Its computational formula is $(\text{sensitivity} + \text{specificity}) - 1$. The Youden's indices for C5.0, CART and QUEST models were 0.54, 0.47 and 0.5, respectively, indicating that the C5.0 model has the highest Youden's index. Through comparing the 3 models C5.0, CART and QUEST, it was observed that the performance of C5.0 decision tree model was the best and thus it could be used for differential diagnosis of MDS and AA. The classification rules of this model are as shown in fig. 1.

If mature lymphocyte ratio was = or < 45.45 % in the bone marrow flow cytometry test, rheumatoid factor was = or < 18.8 IU/ml in immunoassay and myeloblast appeared on bone marrow smear, then the probability for the patient to have MDS was 80 %. If mature lymphocyte ratio was = or < 45.45 % in the bone marrow flow cytometry test, rheumatoid factor was or < 18.8 IU/ml in immunoassay and myeloblasts were not found on bone marrow smear, then probability for the patient to have AA was 69.3 %. If mature lymphocyte ratio was = or < 45.45 % in the bone marrow flow cytometry test, rheumatoid factor was >18.8 IU/ml in the immunoassay and granulocyte ratio was = or < 0.4 %, then probability for the patient to have MDS was 51.3 %. If mature lymphocyte ratio was or < 45.45 % in the bone marrow flow cytometry test, rheumatoid factor was > 18.8 IU/ml in the immunoassay and the granulocyte ratio was > 0.4 %, then probability for the patient to have MDS was 84.4 %.

If mature lymphocyte ratio was >45.45 % in the bone marrow flow cytometry test and myeloblasts appeared on bone marrow smear, then probability for the patient to have MDS was 66.7 %; if mature lymphocyte ratio was >45.45 % in the bone marrow flow cytometry test, myeloblasts were not found on bone marrow smear and rubriblast ratio was = or < 3 %, then the probability for the patient to have AA was 92.8 %; if mature lymphocyte ratio was > 45.45 % upon the bone marrow flow cytometry test, myeloblasts were not found on bone marrow smear and the rubriblast ratio was > 3 %, then the probability for the patient to have AA was 71.7 %.

AA often shows bone marrow hematopoietic hypofunction and pancytopenia, but these symptoms also occur in the hypoplastic myelodysplastic syndrome (hypo-MDS). Treatment and prognosis of the 2 diseases are quite different, as median lifetime of AA can be longer than 174 mo, but that of hypo-MDS is often only 22-23 mo. Thus, it is significant to differentiate the 2 diseases^[16]. However, as there are many patients in clinical practice at present, it faces an acute shortage of clinicians, especially experts and with the new and young clinicians generally lack clinical experience, the rate of misdiagnosis of MDS and AA tends to be high so that many patients might not receive appropriate treatment in time.

In this study, correct and effective diagnosis rules could be extracted through introducing decision tree methodology in data classification and analyzing previous cases to conclude the classification rules. In

order to ensure accuracy of the case diagnosis used, all the cases selected in this investigation were from Chinese Academy of Medical Sciences Hematology Hospital after expert consultation and definite diagnosis. As a top hematology hospital in China, Chinese Academy of Medical Sciences Hematology Hospital gathers many experts in related fields, who could ensure the accuracy and authority of disease diagnosis; besides, as patients with blood diseases all over the country, especially those face diagnostic difficulties are referred to this hospital, one could ensure the case sources to be extensive and representative.

This study adopted 3 common decision tree models, C5.0, CART and QUEST to model, research and analyse the case data. After modelling, all the 3 models were found to have relatively high forecast accuracy. Although differences of the forecast accuracy among the 3 models bear no statistical significance, upon comparison it showed that accuracy rate of C5.0 model is the highest, which was 78.12 %. With the model built on the basis of C5.0 algorithm, mature lymphocyte ratio obtained from bone marrow flow cytometry test was selected as root node variable, the patients diagnosed to have a small mature lymphocyte ratio own a big probability to have MDS, which was close to the study reported by Yamazaki *et al.*^[17]. As AA patients suffer hematopoietic cell non-function, non-hematopoietic cell ratio becomes higher, but as MDS patients' normal hematopoietic function is limited, which often leads to dyshaematopoiesis non-hematopoietic cell ratio won't rise obviously. At present, if the patient showed dyshaematopoiesis is often a reference for clinicians to differentiate the 2 diseases. But as some of the AA patients also have dyshaematopoiesis, especially red dyshaematopoiesis, then combining the mature lymphocyte ratio obtained from decision tree bone marrow flow cytometry test and other related indices, it would provide stronger assurance for the clinicians to make correct diagnosis. Myeloblasts in the bone marrow smear also regarded as an important variable for modelling. If MDS patients displayed dyshaematopoiesis, quantity of myeloblasts often rise. Some studies also found that AA patients do not show bone marrow myeloblast; but in the MDS patients' bone marrow, the ratio of metamyelocyte to myeloblast is obviously higher than that of the AA patients and multi-system dyshaematopoiesis occur^[18-20]. In clinical practice, myeloblast is also used as an important index to differentiate these 2 diseases, myeloblast ratio is also an important reference for MDS sub-classification and decision rules also remind clinicians to pay attention to

the test result of myeloblasts, so as to better differentiate the 2 diseases. It can be concluded from the rules that culture of progenitor cell colony unit also makes for diagnosis of these 2 diseases. Among the patients whose granulocyte-monocyte colony volume is less than 24, AA patients take over 88 %. Thus, culture and test of progenitor cell colony unit has great clinical significance^[21]. Bone marrow investigation is common for diagnosing hematopoietic system diseases and decision tree analysis results could remind clinicians to consider the culture of progenitor cells in bone marrow as a relatively regular test and then a better differentiation of the 2 diseases can be made with the results of progenitor cell colony culture.

CART model also selected mature lymphocyte ratio obtained from the bone marrow flow cytometry test as root node for modelling and rheumatoid factor is also selected as an important branch node of the decision tree. Rheumatoid factor is an important immune indication factor, which also reminds that pathogenesis of MDS and AA correlates to immune factor and the exhaustion of bone marrow hemopoietic cell also might be caused by immune cell abnormality. This is just the reason why this index although appeared irrelevant has great value in differential diagnosis of these 2 diseases. At present, rheumatoid arthritis is also found to be concurrent with these 2 diseases in clinical practice^[22,23]. Among the patients whose rheumatoid factor is higher than 18.75, MDS patients constitute a large share. Clinicians could pay attention to these immune indices and value of these indices often can be used to differentiate these 2 diseases, so that clinicians can use these as a reference index to differentiate these 2 diseases comprehensively.

QUEST model also selected mature lymphocyte ratio obtained from the bone marrow flow cytometry test as the root node for modelling in a manner similar to the above 2 decision tree models. Granulocyte-monocyte colony forming unit is selected as the second-layer branch node variable of the decision tree and blood platelet volume is also an important variable of this model. As to these 2 leaf nodes on the branch of blood platelets, the leaf nodes whose blood platelet volume is $\leq 18.222 \times 10^9/l$ represented MDS patients and the leaf nodes whose blood platelet volume is $>18.222 \times 10^9/l$ represented AA patients. Yan *et al.* also have emphasized the role of blood platelets in the differential diagnosis of these 2 diseases. As a necessary item in clinical examination, routine blood count could reflect hematopoietic functions and hematology status to assist clinicians to determine severity of many diseases.

Decision trees remind that the blood platelet volume is helpful for differential diagnosis of these two diseases and clinicians should make better use of the routine blood examination results to assist to differentiate the occurrence of these 2 diseases^[24].

Discriminant factors selected by the 3 decision tree models are different, as each model has its own method to compute the selected nodes and pruning methods since the decision trees after built completely are also different. However, as all the 3 decision tree models have selected the variables, including mature lymphocyte ratio obtained from the bone marrow flow cytometry test and quantity of myeloblasts, it indicates that bone marrow flow cytometry test and myeloblast count are very important for the differential diagnosis of these 2 diseases. In recent years, multi-parameter flow cytometry has improved the diagnosis accuracy of blood diseases to a greater extent. Studies showed that flow cytometry, immunophenotyping are more sensitive than morphological examination on examining bone marrow abnormalities and it requires not that much as morphological examination on sample preparation^[25,26]. At present, with clinical test methods becoming more abundant and perfect, it brings greater convenience and assistance for clinicians to diagnose diseases. The decision tree modelling results of this study conclusively remind clinicians that while differentiating these 2 diseases MDS and AA, to pay greater attention to the mature lymphocyte ratio obtained from the bone marrow flow cytometry test first, and then differentiate these 2 diseases combining the indices, such as primary cell count, progenitor cell culture results, clinically common dyshaematopoiesis analysis at present and chromosome detection.

As compared to AA, it is easier for MDS to transform to more malignant leukemia, this study sets MDS as positive and AA as negative for data classification^[15]. While evaluating the indexes of positive cases, differences in forecast accuracy among the 3 models are not significantly different. Sensitivity of QUEST model is close to that of C5.0 model, obviously higher than CART model and the differences in sensitivity among the 3 models have statistically significant differences and this indicated that QUEST model and C5.0 model showed a very high classification accuracy for positive cases and they are relatively excellent decision-assisting models. However, comprehensive evaluation not only helps clinicians to know about these single indices, but also from F1 measure, a value that combines the two indices, the C5.0 model

has a higher F1 measure, which also demonstrated the excellence of this model from another point of view. As to the specificity that presents classification accuracy of negative cases, the differences of specificity among the 3 models are not significant. As to the index that shows overall classification performance of sub-class cases-Youden's index, C5.0 model also has the highest value, which further demonstrated its prominent advantage. Generally speaking, forecast accuracy of all the 3 models is higher than 70 % and the differences of forecast accuracy, precision, specificity and several other indexes bear no statistical significance. But from the comprehensive index that reflects overall classification performance, C5.0 model has the highest value, and thus it is the single best model. Although the decision tree models cannot totally help clinicians to diagnose diseases, their high classification accuracy would bring much guidance and reference to the clinicians, especially to some inexperienced young ones. Overall, the C5.0 model is the best in interpreting the results clinically with great precision and accuracy.

Acknowledgements

This study was funded by Hebei Provincial Natural Science Foundation (H2017209172).

REFERENCES

- Killick SB, Bown N, Cavenagh J, Dokal I, Foukaneli T, Hill A, *et al.* Guidelines for the diagnosis and management of adult aplastic anemia. *Br J Haematol* 2015;172:187-207.
- Huang J, Deng MF, Chen YL, Tang Y, Huang Z. Diagnosis and differential diagnosis between hypoplastic myelodysplastic syndrome (Hypo-MDS) and aplastic anemia. *Chin J Health Lab Tec* 2014;2371-3.
- Hast R, Eriksson M, Widell S, Arvidsson I, Bemell P. Neutrophil dysplasia is not a specific feature of the abnormal chromosomal clone in myelodysplastic syndromes. *Leukemia Res* 1999;23:579-84.
- Rashid A, Khurshid M, Shaikh U, Adil S. Chromosomal abnormalities in primary myelodysplastic syndrome. *J Coll Physicians Surg Pak* 2014;24:632-5.
- Wu L, Shi W, Li X, Chang C, Xu F, He Q, *et al.* High expression of the human equilibrative nucleoside transporter 1 gene predicts a good response to decitabine in patients with myelodysplastic syndrome. *J Transl Med* 2016;14:66.
- de Souza DC, Fernandez Cde S, Camargo A, Apa AG, Sobral da Costa E, Bouzas LF, *et al.* Cytogenetic as an Important Tool for Diagnosis and Prognosis for Patients with Hypocellular Primary Myelodysplastic Syndrome. *Biomed Res Int* 2014;2014:542395.
- Huang J, Ge M, Lu S, Shi J, Yu W, Li X, *et al.* Impaired Autophagy in Adult Bone Marrow CD34+Cells of Patients with Aplastic Anemia: Possible Pathogenic Significance. *Plos One* 2016;11:1-17.
- Jiang F, Wang YY, Cen JN, Chen ZX, Liang JY, Liu DD, *et al.*

- Autophagy Activity of CD34(+) Cells in MDS Patients and Its Clinical Significance. *J Exp Hematol* 2016;24:779-83.
- Aa VDL, Alhan C, Béné MC, Della-Porta MG, Dräger AM, Feuillard J, *et al.* Standardization of flow cytometry in myelodysplastic syndromes: report from the first European Leukemia Networking conference on flow cytometry in myelodysplastic syndromes. *Haematologica*, 2009;94:1124-34.
- Stoia M, Kurtanek Z, Oancea S. Reliability of a decision-tree model in predicting occupational lead poisoning in a group of highly exposed workers. *Am J Ind Med* 2016;59:575-82.
- Wu JH, Wei W, Zhang L, Wang J, Damaševičius R, Li J, *et al.* Risk Assessment of Hypertension in Steel Workers Based on LVQ and Fisher-SVM Deep Excavation. *IEEE Access* 2019;7:23109-19.
- Heydari M, Teimouri M, Heshmati Z, Alavinia SM. Comparison of various classification algorithms in the diagnosis of type 2 diabetes in Iran. *Int J Diabetes Dev Ctries* 2016;36:167-73.
- Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 2009;114:937-51.
- Zhang Z. *Blood disease diagnosis and curative effect standard.* Beijing Science Press 2007.
- Koh Y, Lee HR, Song EY, Kim HK, Kim I, Park S, *et al.* Hypoplastic myelodysplastic syndrome (h-MDS) is a distinctive clinical entity with poorer prognosis and frequent karyotypic and FISH abnormalities compared to aplastic anemia. *Leuk Res* 2010;3:1344-50.
- Elghetany MT, Hudnall SD, Gardner FH. Peripheral blood picture in primary hypocellular refractory anemia and idiopathic acquired aplastic anemia: an additional tool for differential diagnosis. *Haematologica* 1997;82:21-4.
- Yamazaki H, Nakao S. Border between aplastic anemia and myelodysplastic syndrome. *Int J Hematol* 2013;97:558-63.
- Cazzola M, Malcovati L. Myelodysplastic syndromes coping with ineffective hematopoiesis. *N Engl J Med* 2005;352:536-8.
- Visconte V, Lindsley RC, Berlyne D. Aplastic Anemia & MDS International Foundation (AA&MDSIF): Bone Marrow Failure Disease Scientific Symposium 2014. *Leuk Res* 2015;39:110-13.
- Nakanishi R, Ishida M, Hodohara K, Okuno H, Yoshii M, Horinouchi A, *et al.* Occurrence of Epstein-Barr virus-associated plasmacytic lymph proliferative disorder after antithymocyte globulin therapy for aplastic anemia: a case report with review of the literature. *Int J Clin Exp Pathol* 2014;7:1748-54.
- Wang H, Yan H, Wang Z, Zhu L, Liu J, Guo Z. Cotransplantation of allogeneic mesenchymal and hematopoietic stem cells in children with aplastic anemia. *Pediatrics*, 2012;129:e1612.
- Tsuji T, Yamasaki H, Tsuda H. Successful treatment with cyclosporine for myelodysplastic syndrome with erythroid hypoplasia following rheumatoid arthritis. *Jpn J Clin Oncol* 2010;51:427-32.
- Stewart BL, Storer B, Storek J, Deeg HJ, Storb R, Hansen JA, *et al.* Duration of immunosuppressive treatment for chronic graft-versus-host disease. *Blood* 2005;104:3501-6.
- Yan ZY, Tian X, Li Y, Yang MR, Zhang S, Wang XM, *et al.* Changes of CD34(+) and CD71(+)/CD45(-) cell levels in bone marrow of MDS and AA patients. *Exp Hematol* 2014;22:382-6.

25. Huang M, Li J, Zhao G, Sui X, Zhao X, Xu H. Immuno phenotype of myeloid granulocytes: a pilot study for distinguishing myelodysplastic syndrome and aplastic anemia by flow cytometry. *Int J Lab Hematol* 2010;32:275-81.
26. Stetler-Stevenson M, Arthur DC, Jabbour N, Xie XY, Mollidrem J, Barrett AJ, *et al.* Diagnostic utility of flow cytometric immunophenotyping in myelodysplastic syndrome. *Blood* 2001;98:979-87.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms

**This article was originally published in a special issue, "Biomedical research applications in Pharmaceutical Sciences"
Indian J Pharm Sci 2020:82(2)Spl issue3;50-58**