Qi *et al.*
BMC Medical Informatics and Decision Making (2024) 24:379
https://doi.org/10.1186/s12911-024-02781-z

BMC Medical Informatics and Decision Making

RESEARCH

Open Access

A potential predictive model based on machine learning and CPD parameters in elderly patients with aplastic anemia and myelodysplastic neoplasms

Yuxiang Qi[1,2†], Xu Liu[2,3†], Zhishan Ding[1], Ying Yu[2] and Zhenchao Zhuang[4*]

Abstract

Background Aplastic anemia (AA) and myelodysplastic neoplasms (MDS) have similar peripheral blood manifestations and are clinically characterized by reduced hematological triad. It is challenging to distinguish and diagnose these two diseases. Hence, utilizing machine learning methods, we employed and validated an algorithm that used cell population data (CPD) parameters to diagnose AA and MDS.

Methods In this study, CPD parameters were obtained from the Beckman Coulter DxH800 analyzer for 160 individuals diagnosed with AA or MDS through a comprehensive retrospective analysis. The individuals were unselectively assigned to a training cohort (77%) and a testing cohort (23%). Additionally, an external validation cohort consisting of eighty-six elderly patients with AA and MDS from two additional centers was established. The discriminative parameters were carefully analyzed through univariate analysis, and the most predictive variables were selected using least absolute shrinkage and selection operator (LASSO) regression. Six machine learning algorithms were utilized to compare the performance of forecasting AA and MDS patients. The area under the curves (AUCs), calibration curves, decision curves analysis (DCA), and shapley additive explanations (SHAP) plots were employed to interpret and assess the model's predictive accuracy, clinical utility, and stability.

Results After the comparative evaluation of various models, the logistic regression model emerged as the most suitable machine learning model for predicting the probability of AA and MDS, which utilized five principal variables (age, MNVLY, SDVLY, MNLALSEGC, and MNCEGC) to accurately estimate the risk of these diseases. The best model delivered an AUC of 0.791 in the testing cohort and had a high specificity (0.850) and positive predictive value (0.818). Furthermore, the calibration curve indicated excellent agreement between actual and predicted probabilities. The DCA curve further supported the clinical utility of our model and offered significant clinical advantages in guiding treatment decisions. Moreover, the model's performance was consistent in an external validation group, with an AUC of 0.719.

Conclusions We developed a novel model that effectively distinguished elderly patients with AA and MDS, which had the potential to provide physicians assistance in early diagnosis and the proper treatment for the elderly.

Keywords Cell population data, Parameters, Aplastic anemia, Myelodysplastic neoplasms, Machine learning

†Yuxiang Qi and Xu Liu contributed equally to this work and shared first authorship.

*Correspondence:
Zhenchao Zhuang
zhuangzzc2015@163.com
Full list of author information is available at the end of the article

Introduction

Aplastic anemia (AA) is a bone marrow hematopoietic failure disorder, mainly manifested by low bone marrow hematopoiesis, decreased whole blood cells, and anemia [1, 2]. Myelodysplastic neoplasms (MDS) are highly heterogeneous myeloid neoplasms manifested by chronic cytopenias, ineffective and dysplastic hematopoiesis. This leads to a decrease in blood cell counts and morphological dysplasia in one or more blood cells. MDS are called the early stages of leukemia, almost 3 out of 10 patients with MDS progress to acute myeloid leukemia (AML) [3]. Routine blood tests are the most common of the clinical laboratory tests. Although changes in blood cell counts are associated with a number of clinical conditions, an abnormal routine blood test may indicate the presence of a hematologic malignancy [4]. These two bone marrow failure disorders have similar peripheral blood manifestations and are clinically characterized by a reduction in the hematological triad, they have different etiologies and involve a variety of clinical and molecular alterations [5–7]. Therefore, how to narrow their boundaries or eventually redefine them altogether poses a major challenge to researchers, and treatment and prognosis of both disorders are greatly influenced by an appropriate differential diagnosis.

The diagnostic process of each disease generates a large amount of laboratory data, making it possible to conduct comprehensive data mining to effectively analyze the diseases. By harnessing the power of big data analysis, we can gain insights into the characteristics, patterns, and trends of these diseases. Machine learning (ML) involves models or algorithms that can allow computers to learn from data and identify individual features of data [8]. As medical data volumes grow, ML can assist doctors to make more accurate diagnoses, predict patient outcomes, and personalize treatment plans. It automates tedious tasks, reducing human labor and enabling doctors to focus on providing better patient care [9]. In hematology, machine learning has been used to improve risk stratification, categorical diagnosis, and prognosis of diseases, as well as mortality prediction and treatment of tumors [10].

For the past few years, with the rapid development of technology, blood analyzers have provided more information in addition to the usual blood image parameters, translating cell morphology and characteristic changes into reportable cell count results and derived study parameters [11]. Hematology analyzer data has been used to predict a series of clinical outcomes, from blood culture results [12], sepsis patients [13, 14], and COVID-19 patients [15]. The application of this novel technology allows reporting of new parameters as well as basic complete blood counts. Leukocytes (neutrophils, monocytes, eosinophils, and lymphocytes) can be classified according to their morphological and functional characteristics using cell population data (CPD). Since CPD is derived from the results obtained from routine blood analysis, it has the advantages of being rapid, economical, and reliable, without the need for additional reagents or processes, and possesses good prospects for clinical application.

In the diseases of the blood system, CPD parameters have shown a great role in various diseases that can cause morphological changes in leukocytes. CPD parameters have shown applicability in the diagnosis and differentiation of various hematological diseases, including multiple myeloma [16], neoplastic hematological diseases [17], thalassemia traits [18], chronic myeloid leukemia [19], with a particular emphasis on its significance in the diagnosis of myelodysplastic syndromes [20–22]. Given the challenges associated with the screening and diagnosis of MDS, the application of the CPD parameter is particularly crucial.

Currently, MDS and AA are mainly diagnosed by hematology, cytomorphology, bone marrow examination, and cytogenetics [2, 23]. However, bone marrow examination is an invasive technique that tests the doctor's skills, and the bone marrow cells may not be observed during the bone marrow biopsy due to inappropriate bone marrow extraction, which brings more pain to the patients. The sensitivity of single immunophenotypic indexes for the differential diagnosis of MDS and AA is too low, which restricts the wide application of FCM in diagnosing myelodysplastic neoplasms, and genetic testing is more costly [24–26]. These facts prompted us to look for routine laboratory tests based on which to diagnose AA and MDS. Apart from Wu et al. [27], few studies have reported models based on machine learning to distinguish AA and MDS. However, their model utilized numerous indicators including blood cell count, blood smear, and marrow smear, which may not be practical in clinical settings. Additionally, the model lacked an external validation set and primarily targeted the general population rather than the elderly.

In the current study, we built a model based on CPD parameters and machine learning to distinguish between AA and MDS in elderly patients. Detection of these parameters in these diseases may contribute to early diagnosis and rapid intervention of the disease, which contributed to improving elderly patients' prognosis. In addition, with further research and optimization, the model was expected to become a powerful tool in clinical practice, and could also provide a reference for other medical-related research.

Methods

Patient involvement

According to the guidelines for the diagnosis and management of adult aplastic anemia [7, 28] as well as

the clinicians' diagnoses, we collected 252 patients (age ≥ 50 years) diagnosed with AA and MDS from May 16, 2022 to August 28, 2023, in Zhejiang Provincial Hospital of Chinese Medicine (Hubin). According to the exclusion criteria, 92 elderly patients were excluded, as shown in Fig. 1. Finally, included in the study were a combined total of 89 cases classified as AA and 71 cases classified as MDS. Furthermore, 86 cases from Zhejiang Provincial Hospital of Chinese Medicine (Qiantang and Xixi) were collected as an external validation cohort from May 16, 2022 to August 28, 2023.

Diagnostic criteria for aplastic anemia: 1. Blood routine examination: Total blood cells (including reticulocytes) uniformly depressed, and the proportion of lymphocytes increased. Meet at least two of the following three criteria items: HGB < 100 g/L; PLT < 50×10^9/L; Neutrophil rejection Opposition value (ANC) < 1.5×10^9/L. 2. Bone marrow aspiration: Erythropoiesis was reduced or absent. Megakaryocytes and granulocytic cells were markedly reduced or absent. The proportion of non-hematopoietic cells (lymphocytes, reticular cells, plasma cells, mast cells, etc.) increased. 3. Bone marrow biopsy (ilium): The biopsy specimen was hypocellular throughout, with reduced hematopoietic tissue, increased non-hematopoietic cells, no increase in reticulin, and no abnormal cells. 4. Congenital and other acquired and secondary BMF were excluded. The diagnosis of myelodysplastic neoplasms was according to the 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms [29]. Exclusion criteria: Patients with other blood diseases or who had undergone bone marrow transplantation or hematopoietic stem cell therapy, or incomplete data were excluded.

Data collection

Detailed information about these patients' baseline population characteristics (age, gender, and comorbidities) and CPD parameters were carefully collected from their electronic medical records. Following the completion of the enrollment

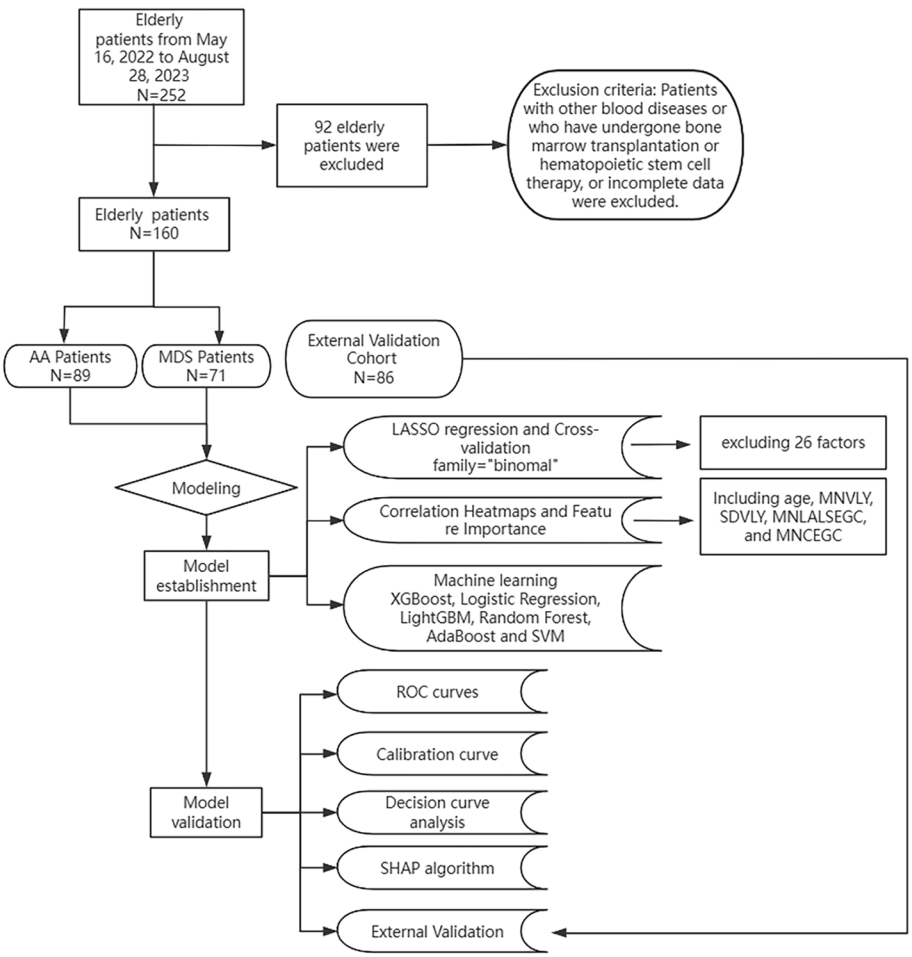

Fig. 1 The flow chart demonstrated the participants encompassed within in the study

process, a total of 160 elderly individuals were subjected to random assignment, with 77% being allocated to the training cohort and the remaining 23% to the testing cohort. By establishing a random seed, the current investigation could guarantee the replicability of the stochastic procedure, enabling precise replication of research findings as necessary. The hyperparameters of the best model were chosen using grid search and cross-validated ten times. In ten-fold cross-validation, the dataset was divided into ten equal-sized sections. One of the ten sections was used for testing and the remaining nine sections were used for training. Ten-fold cross-validation was looped ten times throughout the process. Utilizing CPD parameters, six machine learning models were built up in the training cohort and subsequently validated in the testing cohort. The final filtered optimal model was validated using an external validation cohort.

Statistical analysis

All analyses were done through SPSS 26.0 and the platform. Frequencies and percentages were used to indicate qualitative variables, while mean ± standard deviation or median and interquartile range (IQR) were used to indicate quantitative variables. Count data were analyzed using the chi-square test, and measurement data were analyzed using the independent samples *t*-test or Wilcoxon test.

Through univariate analysis, indicators with significant disparities between AA and MDS groups were screened, and we further utilized the least absolute shrinkage and selection operator (LASSO) regression to pick out the factors that were more relevant to AA and MDS. Via random seeds, 77% of patients were allocated to form the training cohort, whereas the remaining 23% of patients were allocated to the testing cohort. Calibration and Decision plots were utilized to visually evaluate the model, while the area under curve (AUC) was employed for assessing calibration. The intricate feature ranking was interpreted via shapley additive explanations (SHAP) plots. $P < 0.05$ is considered statistical significance.

Machine learning

The building and training of machine learning models were accomplished through the platform Deepwise & Beckman Coulter DxAI, and the reason for choosing these models was that these are the more common machine learning models. This is a mature platform, publishing a lot of high-score literature. The platform is capable of automatically selecting machine learning models and generating an analysis page online.

Results

Demographic characteristics

The demographic characteristics of elderly patients were summarized in Table 1. This study encompassed 89 (55.625%) classified as AA and 71 (44.375%) classified as MDS. There were 39 males (43.820%) and 50 females (56.180%) in the AA group, while 44 males (61.972%) and 27 females (38.028%) were in the MDS group. As shown in Table 1, the gender difference between the two groups was statistically significant ($P = 0.022 < 0.05$). The MDS group exhibited a significantly higher mean age compared to the AA group ($P < 0.001$). The median age of the MDS group was 69.000, as compared to the AA group's median age of 61.000. The most common comorbidity in AA patients was hypertension (29.213%), followed by diabetes (16.854%). Additionally, infectious fever, hypoproteinemia, tumors, and coronary heart disease were present in 12.360%, 5.618%, 5.618%, and 2.247% of AA patients, respectively. The most common comorbidity in MDS patients was hypertension (28.169%), followed by diabetes (15.493%). Additionally, tumors, infectious fever, hypoproteinemia, and coronary heart disease were present in 14.085%, 9.859%, 7.042%, and 4.225% of MDS patients, respectively.

In the external validation cohort, this study encompassed 57 individuals (66.279%) diagnosed as AA and 29 (33.721%) diagnosed as MDS. Within the AA cohort, 24 patients (42.105%) were male, and 33 (57.895%) were female. The median age of the cohort was 60 years old. The most common comorbidity in patients was hypertension (14.035%), followed by diabetes (8.772%). Additionally, tumors, infectious fever, coronary heart disease, and hypoproteinemia were present in 7.018%, 5.263%, 1.754%, and 1.754% of AA patients, respectively. Within the MDS cohort, 15 patients (51.724%) were male, and 14 (48.276%) were female. The median age of the cohort was 65 years old. The most common comorbidity in patients was hypertension (24.138%), followed by tumors (20.690%). Additionally, diabetes, infectious fever, and hypoproteinemia were present in 13.793%, 3.448%, and 3.448% of AA patients, respectively. There were no patients with coronary heart disease in this group.

Comparison of CPD parameters between AA and MDS patients

The results were shown in Table 2, which indicated that there were significant differences in gender, age, SDVNE, MNCNE, MNMALSNE, SDMALSNE, MNUMALSNE, SDUMALSNE, MNLMALSNE, SDLMALSNE, MNLALSNE, SDAL2NE, MNVLY, SDVLY, SDCLY, SDMALSLY, SDUMALSLY, MNLMALSLY, SDLMALSLY, SDLALSLY, MNAL2LY, SDAL2LY, MNVMO, SDVMO, SDCMO, SDLMALSMO, MNLALSMO, SDAL2MO, SDAL2EGC, MNLALSEGC, SDLMALSEGC, MNLMALSEGC, MNUMALSEGC, SDMALSEGC, MNMALSEGC, SDCEGC, MNCEGC, SDVEGC, MNVEGC, MNLMALSEO, SDUMALSEO, MNUMALSEO, SDMALSEO, MNMALSEO,

Table 1 Baseline Characteristics of AA and MDS Patients

Characteristic	AA Patients ($n=89$)	MDS Patients ($n=71$)	Validation Cohort ($n=86$)		*P*-value
			AA Patients ($n=57$)	MDS Patients ($n=29$)	
Gender					
Male[(n, %)]	39 (43.820)	44 (61.972)	24 (42.105)	15 (51.724)	0.022
Female[(n, %)]	50 (56.180)	27 (38.028)	33 (57.895)	14 (48.276)	
Age [year (median IQR)]	61.000 (56.000–68.500)	69.000 (61.000–73.000)	60.000 (55.000–69.000)	65.000 (59.000–69.000)	< 0.001
Basic disease					
Yes[a][(n, %)]	50 (56.180)	43 (60.563)	18 (31.579)	17 (58.621)	0.577
No[(n, %)]	39 (43.820)	28 (39.437)	39 (68.421)	12 (41.379)	
Comorbidities [(n, %)]					
Tumor[b]	5 (5.618)	10 (14.085)	4 (7.018)	6 (20.690)	0.068
Hypertension	26 (29.213)	20 (28.169)	8 (14.035)	7 (24.138)	0.885
Diabetes	15 (16.854)	11 (15.493)	5 (8.772)	4 (13.793)	0.817
Infectious fever	11 (12.360)	7 (9.859)	3 (5.263)	1 (3.448)	0.619
Hypoproteinemia	5 (5.618)	5 (7.042)	1 (1.754)	1 (3.448)	0.712
Coronary heart disease	2 (2.247)	3 (4.225)	1 (1.754)	0 (0.000)	0.475

[a] Patients with 1 of the following: tumor, hypertension, diabetes, infectious fever, hypoproteinemia, or coronary heart disease

[b] Any type of tumor

and MNCEO between the two groups ($P<0.05$). This meant that there were some morphological changes in neutrophils, lymphocytes, monocytes, early granulated cells, and eosinophils in the blood of patients with AA and MDS.

Screening for optimal predictors by LASSO regression

In the current study, we collected a total of 71 indicators from elderly patients classified as AA and MDS. After excluding non-significant indicators, 45 features were retained for LASSO regression analysis to screen the optimal predictors that have correlated with two diseases. The results via LASSO regression showed that age, MNLMALSNE, MNVLY, SDVLY, SDCLY, SDVMO, MNLALSEGC, SDCEGC, and MNCEGC were considered to be relevant factors with AA and MDS (Fig. 2). Furthermore, using the nine indicators chosen via LASSO regression, the current study examined heat maps of correlations and importance rankings between these indicators.

AUCs of nine indicators

In Fig. 3, the ROC and AUCs were presented, highlighting the significant differences in diverse indicators between the two groups in forecasting AA and MDS. Among these CPD parameters, MNLMALSNE was the most efficient (AUC = 0.760). SDCLY was then followed (AUC = 0.758).

Feature importance and correlation heatmap of CPD parameters

Upon analyzing the importance of diverse indicators, the current study eventually chose five predictors rooted in the count of elderly individuals afflicted with AA and MDS. The feature importance between the nine filtered indicators was shown in Fig. 4A. The most valuable of these nine indicators was age. Additionally, MNVLY, SDVLY, MNLALSEGC, and MNCEGC were followed, respectively. In turn, the interrelationships among the five indicators were analyzed. It was believed that the correlation between the two indicators < 0.7 would not interfere

Table 2 Comparison of CPD parameters between AA and MDS patients

Variable Category	AA Patients ($n=89$)	MDS Patients ($n=71$)	*P*-value
MNVNE	148.000(142.000–156.000)	150.000(142.000–166.000)	0.219
SDVNE	18.540(17.370–21.810)	21.300(18.420–27.130)	< 0.001
MNCNE	142.000(140.000–146.000)	140.000(136.000–143.000)	< 0.001
SDCNE	5.280(4.800–6.500)	5.570(4.950–6.500)	0.229
MNMALSNE	139.000(133.000–145.000)	127.000(119.000–136.000)	< 0.001
SDMALSNE	11.880(10.800–13.770)	12.630(11.510–13.780)	0.030
MNUMALSNE	138.000(130.000–143.000)	130.000(120.000–135.000)	< 0.001
SDUMALSNE	12.270(11.490–13.820)	13.240(11.830–16.550)	0.010
MNLMALSNE	135.000(126.000–141.000)	122.000(112.000–129.000)	< 0.001
SDLMALSNE	13.710(12.600–16.890)	15.380(14.020–17.800)	0.006
MNLALSNE	171.000(155.000–179.000)	151.000(142.000–167.000)	< 0.001
SDLALSNE	30.580(28.030–39.140)	33.000(29.110–37.340)	0.184
MNAL2NE	141.000(136.000–147.000)	143.000(136.000–149.000)	0.450
SDAL2NE	12.630(11.130–14.920)	15.130(12.980–18.090)	< 0.001
MNVLY	89.281 ± 4.851	93.394 ± 5.875	< 0.001
SDVLY	14.260(13.420–15.370)	15.570(14.140–17.730)	< 0.001
MNCLY	110.000(108.000–112.000)	111.000(109.000–112.000)	0.493
SDCLY	7.160(6.370–8.620)	9.760(7.800–13.310)	< 0.001
MNMALSLY	65.730 ± 6.475	65.648 ± 5.780	0.934
SDMALSLY	15.860(14.430–17.330)	17.540(15.940–19.220)	0.001
MNUMALSLY	62.101 ± 10.179	64.592 ± 8.649	0.105
SDUMALSLY	19.720(18.230–22.050)	20.730(18.890–22.650)	0.047
MNLMALSLY	61.180 ± 4.882	59.352 ± 4.753	0.019
SDLMALSLY	17.410(16.550–19.070)	19.190(17.770–20.950)	< 0.001
MNLALSLY	41.034 ± 3.449	41.507 ± 2.940	0.362
SDLALSLY	11.150(10.330–12.020)	12.190(11.020–13.460)	< 0.001
MNAL2LY	74.000(70.000–78.000)	77.000(72.000–82.000)	0.009
SDAL2LY	13.110(11.730–15.040)	14.350(13.060–16.100)	0.001
MNVMO	173.000(168.000–179.000)	177.000(169.000–190.000)	0.015
SDVMO	20.290(18.120–22.600)	22.720(20.180–27.450)	< 0.001
MNCMO	120.000(118.000–122.000)	119.000(116.000–122.000)	0.071
SDCMO	4.970(4.560–5.870)	5.870(4.730–7.770)	0.019
MNMALSMO	87.000(83.000–89.000)	86.000(81.000–90.000)	0.277
SDMALSMO	11.310(10.520–12.360)	11.910(10.830–14.650)	0.060
MNUMALSMO	95.000(90.000–99.000)	95.000(89.000–99.000)	0.461
SDUMALSMO	12.690(11.550–14.470)	13.060(11.820–15.490)	0.217
MNLMALSMO	74.000(72.000–77.000)	73.000(69.000–77.000)	0.100
SDLMALSMO	13.880(12.910–15.490)	14.840(13.220–16.710)	0.041
MNLALSMO	93.000(87.000–98.000)	85.000(78.000–94.000)	< 0.001
SDLALSMO	24.030(21.280–27.280)	23.660(21.070–28.600)	0.639
MNAL2MO	125.000(121.000–131.000)	127.000(118.000–135.000)	0.520
SDAL2MO	16.290(13.970–19.800)	18.650(15.480–21.810)	0.006
SDAL2EGC	17.380(13.530–20.180)	19.490(15.250–24.280)	0.017
MNAL2EGC	143.930(141.000–149.000)	146.000(140.000–158.000)	0.175
SDLALSEGC	22.470(15.880–25.789)	22.210(18.500–24.780)	0.793
MNLALSEGC	133.820(126.000–140.000)	118.000(105.000–131.000)	< 0.001
SDLMALSEGC	8.380(6.360–9.480)	9.360(8.490–10.780)	< 0.001
MNLMALSEGC	128.000(125.280–136.000)	122.000(113.000–127.000)	< 0.001
SDUMALSEGC	10.060(8.900–11.960)	10.940(9.300–13.540)	0.056

Table 2 (continued)

Variable Category	AA Patients ($n=89$)	MDS Patients ($n=71$)	*P*-value
MNUMALSEGC	146.030(141.000–151.000)	143.000(131.000–150.000)	0.012
SDMALSEGC	6.285(4.850–7.200)	7.630(5.870–8.820)	< 0.001
MNMALSEGC	138.000(137.670–146.000)	133.000(124.000–140.000)	< 0.001
SDCEGC	2.050(1.720–2.290)	2.470(2.100–3.140)	< 0.001
MNCEGC	133.000(130.750–136.000)	129.000(127.000–133.000)	< 0.001
SDVEGC	25.940(21.380–31.040)	29.850(23.750–36.100)	0.013
MNVEGC	159.000(148.000–168.000)	163.000(153.000–181.000)	0.007
SDAL2EO	10.620(9.060–14.270)	11.040(9.800–14.540)	0.190
MNAL2EO	123.000(119.000–129.000)	127.000(119.000–135.000)	0.145
SDLALSEO	41.650(38.000–44.720)	40.610(33.930–44.510)	0.241
MNLALSEO	165.000(160.000–175.000)	167.000(156.000–176.000)	0.959
SDLMALSEO	10.490(8.810–11.700)	11.320(9.400–13.590)	0.061
MNLMALSEO	184.000(178.000–188.000)	179.000(169.000–184.000)	< 0.001
SDUMALSEO	10.300(8.660–13.880)	11.170(9.990–15.330)	0.047
MNUMALSEO	211.000(204.000–216.000)	202.000(190.000–211.000)	< 0.001
SDMALSEO	8.520(7.470–10.410)	9.940(8.640–11.840)	0.006
MNMALSEO	200.000(193.000–203.000)	193.000(179.000–199.000)	< 0.001
SDCEO	4.560(3.630–16.850)	5.120(4.160–9.310)	0.476
MNCEO	146.000(142.730–151.000)	144.000(141.000–147.000)	0.006
SDVEO	18.260(16.190–23.410)	19.680(16.800–23.080)	0.236
MNVEO	157.000(153.000–164.000)	159.000(149.000–166.000)	0.988

The first two alphabets (MN/SD) in each parameter refer to mean or standard deviation, and the last two alphabets (NE/LY/MO/EGC/EO) refer to the cell type (neutrophils, lymphocytes, monocytes, early granulated cells or eosinophils respectively), middle alphabets (V/C/MALS/UMALS/LMALS/LALS/AL2) refer to the cell property measured by the analyzer (volume, conductivity, median angle light scatter, upper median angle light scatter, lower median angle light scatter, low angle light scatter, or axial light loss respectively)

with each other. As presented in Fig. 4B, age, MNVLY, SDVLY, MNLALSEGC, and MNCEGC exhibited a low correlation, potentially preventing insufficient generalization of the model to new data from other sources.

Comparative evaluation of six machine learning models

The AUCs of six machine learning models for tenfold cross-validation on the training cohort were presented in Table 3. We focused on the AUC performance of each machine learning model on the validation cohort to determine the optimal model. According to the AUCs in the testing cohort, the highest AUC among the six ML algorithms was achieved by logistic regression (AUC = 0.827). The second-highest AUC was resented by random forest (AUC = 0.787). The third-highest AUC was offered by support vector machines (SVM) (AUC = 0.780). In addition, adaptive boosting (AdaBoost) demonstrated the lowest manifestation (AUC = 0.705) and was excluded. The result showed that the logistic regression model excelled in predicting performance compared to the other five models.

Machine learning models establishment and assessment

Drawing from the data presented in Fig. 5 and Table 4, it was evident that the logistic regression model possessed a robust discriminatory capability in distinguishing aplastic anemia and myelodysplastic neoplasms. The model exhibited an AUC of 0.791 in the testing cohort (Fig. 5B), with specificity and positive predictive value exceeding 80% (Table 4). In addition, Fig. 5C indicated excellent calibration of the model. The calibration curve exhibited a good agreement between the actual probability and the predicted probability. The DCA curve highlighted the clinical benefits of the model, indicating its strong performance in clinical settings (Fig. 5D).

Figure 6A depicted the correlation between SHAP values of the five most pertinent features we identified. As Fig. 6B illustrated, the logistic regression model's interpretation of feature ranking, as determined by the SHAP algorithm, was presented. It was explained that the most powerful features for predicting outcomes of elderly patients were MNLALSEGC, MNVLY, age, SDVLY, and MNCEGC. These characteristics had the greatest impact on predicting patient outcomes and should be considered

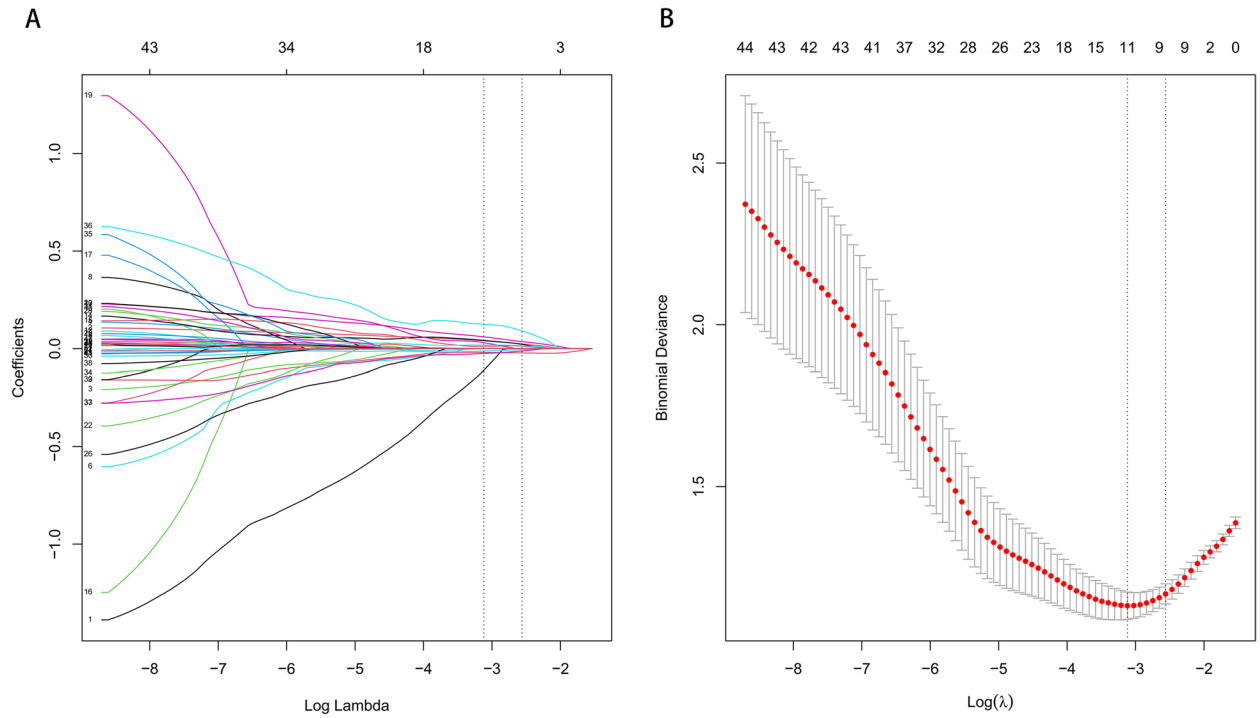

Fig. 2 Screening the optimal predictors via LASSO regression. **A** Regression coefficient path plot in LASSO regression. Diverse colored lines indicate that different variables will gradually become zero, and the later they become zero, the more important the indicator. **B** The cross-validation curve of LASSO regression. The minimum standard is on the left line and the 1-SE standard is on the right line. In the current study, we selected 9 non-zero predictors according to the 1-SE standard. SE, the standard error

in the evaluation and treatment of elderly patients. By utilizing SHAP force plots, the study offered a visual representation of the SHAP value of a single indicator, demonstrating its impact on modifying the baseline predicted value, whether positive or negative. Figure 6C and D showed the individual force plots for MDS patients and AA patients, respectively. The figures offered a visual representation of the impact of each feature on modifying the model's predicted value for each patient group. The features that contribute positively, denoted in red, propel the model's score upward, whereas those that contribute negatively, denoted in blue, pull the model's score downward. The length of the arrow offered a visual representation of the magnitude of its impact on the prediction. As the arrow lengthens, the greater the influence on the prediction of MDS.

External validation

Eighty-six elderly patients were recruited from two additional centers to serve as an external validation cohort. As shown in Fig. 7, the AUC of the logistic regression model was 0.719 when validated using the external validation cohort. This suggested that the model based on CPD parameters had high value in practical applications.

Discussion

Distinguishing between the various features of aplastic anemia and myelodysplastic neoplasms is critical clinically, as it affects patient drug therapy and outcomes [6]. It has been reported that the risk of progressing to AML patients with MDS was much higher than those with AA [30, 31].

Nowadays, the use of machine learning methods to help clinicians process laboratory results can avoid the influence of empirical differences between clinicians on diagnostic results. Novel leukocyte CPD parameters are emerging as potential markers in diverse clinical settings. These parameters have been initially applied to disease identification, such as in COVID-19 [32] and sepsis [14]. Based on VCS technology, morphological analysis of leukocyte subtypes is performed, cell volume (V) is measured by DC impedance to obtain accurate cell size, and electrical conductivity (C) of internal components of each cell is characterized by radio frequency transmittance. The light scattering (S) beam of cytoplasmic particle size and nuclear structure is measured using a laser [33].

The application of CPD parameters has several advantages in the clinic. These parameters are generated during a routine complete blood count (CBC) analysis,

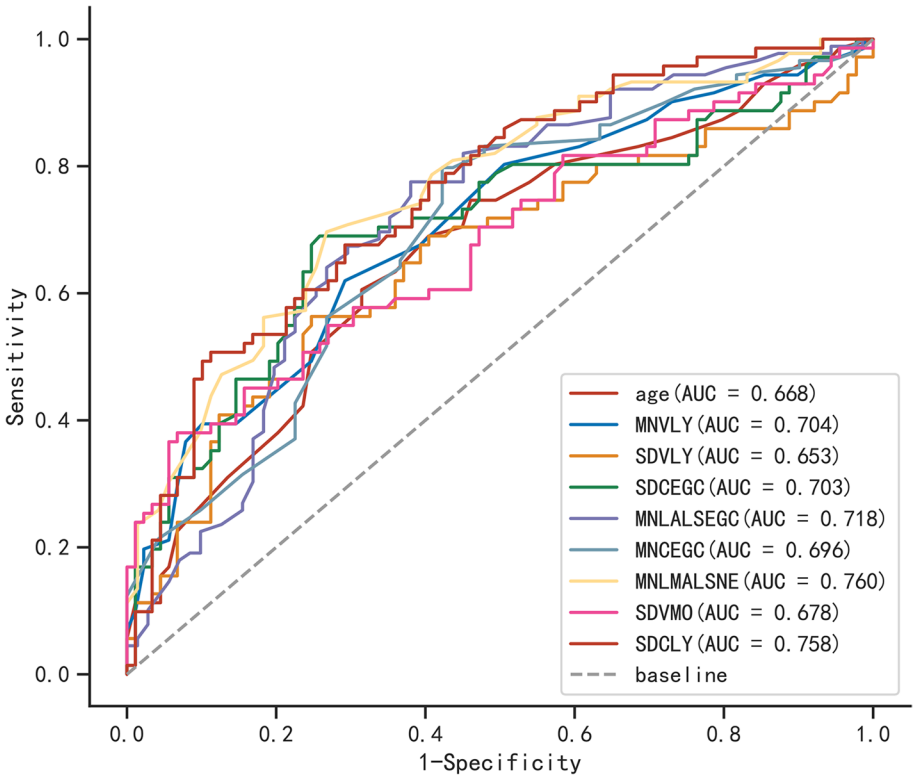

Fig. 3 The ROC curves of AA and MDS were independently predicted by 9 predictors between the two groups

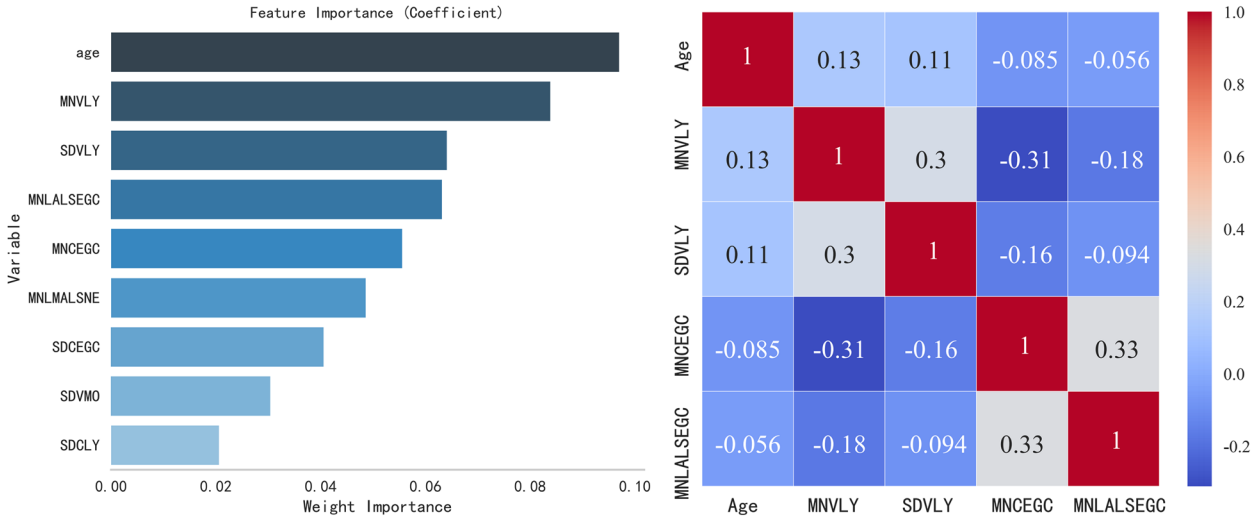

Fig. 4 **A** The weight importance of nine filtered indicators. **B** Heat map of correlation of top five indicators. The correlation degree is from low (blue) to high (red)

eliminating the need for additional samples. CPD parameters are more objective and accurate than manual difference counts due to the automatic assessment of thousands of white blood cells, making them suitable as an additional marker at a lower cost than other laboratory tests [34, 35]. Therefore, we emphasized the importance of CPD parameters not only for the rapid screening of diseases but also as a simple method for its

Table 3 Comparative evaluation of six machine learning models for ten-fold resampling-validation

Classifier	Cohorts	AUC	Cut off	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value	F1
XGBoost	Training cohort	1.000	0.794	0.992	1.000	1.000	1.000	0.986	1.000
	Validation cohort	0.759	0.794	0.705	0.679	0.842	0.765	0.629	0.712
Logistic Regression	Training cohort	0.842	0.555	0.778	0.672	0.878	0.813	0.759	0.735
	Validation cohort	0.827	0.555	0.762	0.750	0.866	0.773	0.770	0.748
LightGBM	Training cohort	0.984	0.495	0.937	0.938	0.952	0.939	0.937	0.938
	Validation cohort	0.770	0.495	0.697	0.728	0.780	0.708	0.714	0.712
Random Forest	Training cohort	1.000	0.545	0.988	1.000	0.997	0.998	0.981	0.999
	Validation cohort	0.787	0.545	0.678	0.750	0.747	0.808	0.620	0.769
AdaBoost	Training cohort	0.994	0.499	0.958	0.978	0.956	0.948	0.967	0.962
	Validation cohort	0.705	0.499	0.676	0.705	0.718	0.618	0.734	0.644
SVM	Training cohort	0.819	0.428	0.766	0.780	0.770	0.723	0.806	0.749
	Validation cohort	0.780	0.428	0.719	0.716	0.825	0.698	0.747	0.700

XGBoost eXtreme gradient boosting, *LightGBM* light gradient boosting machine, *AdaBoost* Adaptive boosting, *SVM* Support vector machines

rapid performance, we could acquire a wealth of valuable data from circulating blood for the description of hematologic diseases.

The present study showed that it was possible to distinguish between AA and MDS using white blood cell population data parameters. Age, MNVLY, SDVLY, MNLALSEGC, and MNCEGC were identified to build up the model. In actual clinical practice, the analysis of a single feature was frequently inadequate to capture the entire nature of the disease. Consequently, our model considered the above five indicators as a whole rather than making diagnostic predictions based on individual features in order to distinguish between AA and MDS. The model, as presented in Fig. 5, demonstrated high discrimination and calibration, indicating a strong performance and higher clinical utility. Furthermore, the model performed effectively in both the testing cohort (AUC = 0.791) and the external validation cohort (AUC = 0.719). These results indicated that the model had significant value in accurately and stably classifying the probability of AA and MDS occurring in elderly patients on an individual basis. Our model had the potential to offer doctors a user-friendly and highly effective tool for discriminating between AA and MDS in clinical practice.

Recently, several machine learning algorithms for predicting MDS have been developed. Park et al. [36] have created a model by using cell population data, the model had an AUC of 0.891. Pozdnyakova et al. [21] have created a model by using CBC parameters, the model had an AUC of 0.860. In our logistic regression model, the AUC in the testing cohort was 0.791 which was lower than their model. However, our model performs better in specificity, achieving a value of 0.850, while their specificities are 0.790 and 0.720, respectively. Additionally, it is worth noting that their models have not undergone further validation using a test cohort or external validation cohort.

Aplastic anemia (AA) affects 7.4 people per million per year, with a higher prevalence in China than in the West [28]. The disease's occurrence also varies with age, with the highest frequency observed among individuals over the age of 60. The gender ratio is also different: AA patients over 60 years of age are predominantly female (60.000%) [37, 38], which is similar to our study (56.180%). From a biological perspective, it has been observed that elderly individuals with AA often exhibit a higher frequency of mutations that are potentially linked to adverse outcomes. Furthermore, several studies have identified age > 60 as an independent risk factor for mortality in aplastic anemia [39]. However, the prevalence of MDS is estimated to be only as high as 75 per 100,000 over the age of 65 [40]. The immune system undergoes morphological or functional changes with aging, as evidenced by a decrease in autoimmune cells and a higher prevalence of autoantibodies, with AA and MDS being increasingly diagnosed in the elderly [41, 42]. In the present study, age also shows great weight and importance in these indicators (Fig. 4A and Fig. 6B).

As the key link of immune defense, white blood cells will change their shape, internal structure, and function, and their CPD parameters can show this change in sensitivities as an immunoreactive change in the body in a pathological state. In the process of exploring the relationship between CPD parameters and disease, we observed notable disparities in the distribution profiles

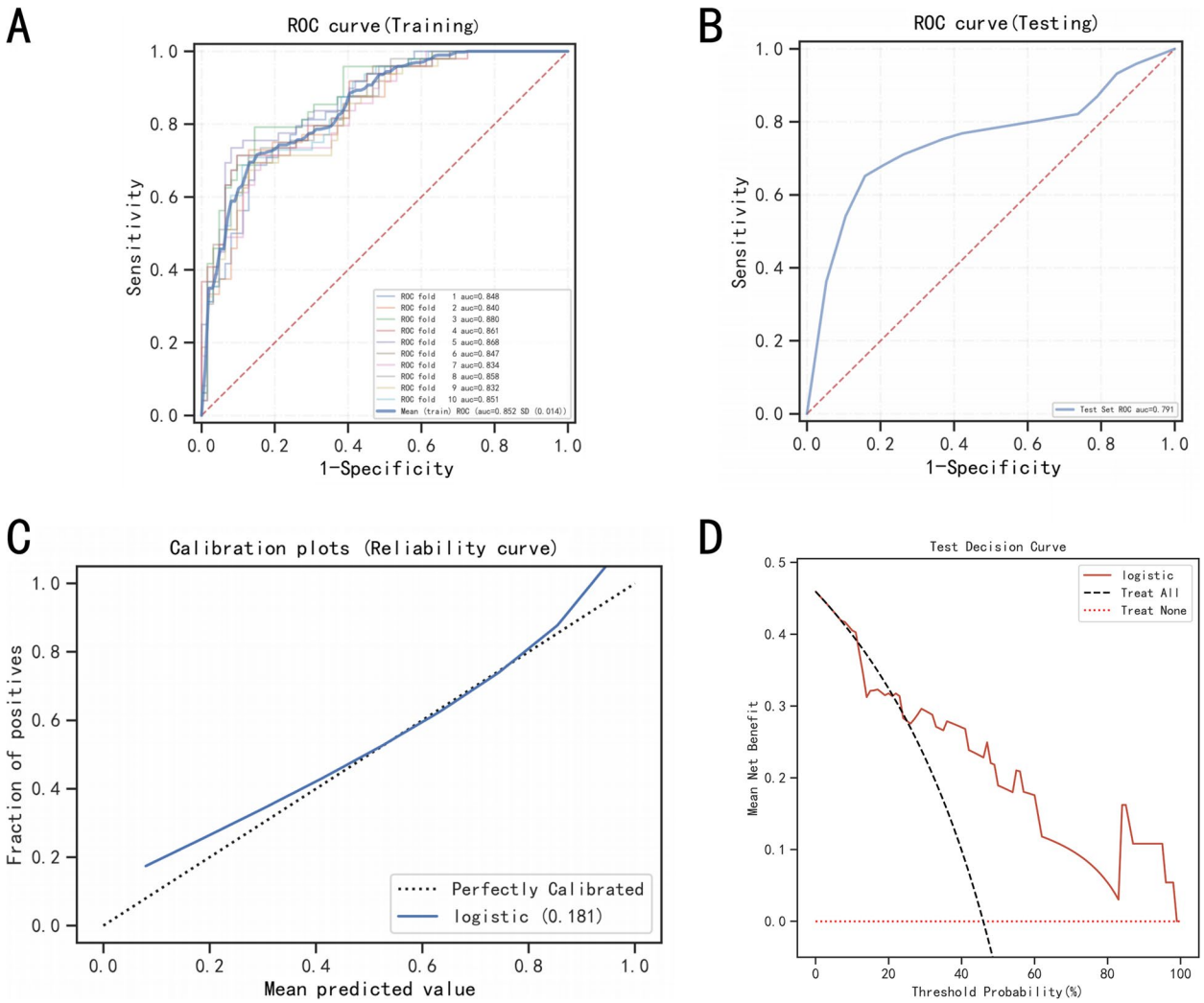

Fig. 5 The performance of six machine learning models. **A** ROC curve of the training cohort; **B** ROC curve of the testing cohort; **C** Calibration curve; (D) Decision curve analysis

Table 4 Evaluation of the optimal logistic regression model for ten-fold cross-validation

Cohorts	AUC	Cut off	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value	F1
Training cohort	0.852	0.525	0.796	0.718	0.873	0.812	0.787	0.762
Testing cohort	0.791	0.532	0.73	0.706	0.850	0.818	0.692	0.758

of CPD parameters among patients with AA and MDS (Table 2).

Since granulocyte dysplasia is a high-visibility feature of MDS, neutrophil-associated parameters have been extensively studied and are commonly utilized to discern granulomatous dysplasia [42, 43]. In addition to age, among the neutrophil-associated CPD parameters, this study observed an increased SDVNE, SDUMALSNE, SDLMALSNE, SDAL2NE, and

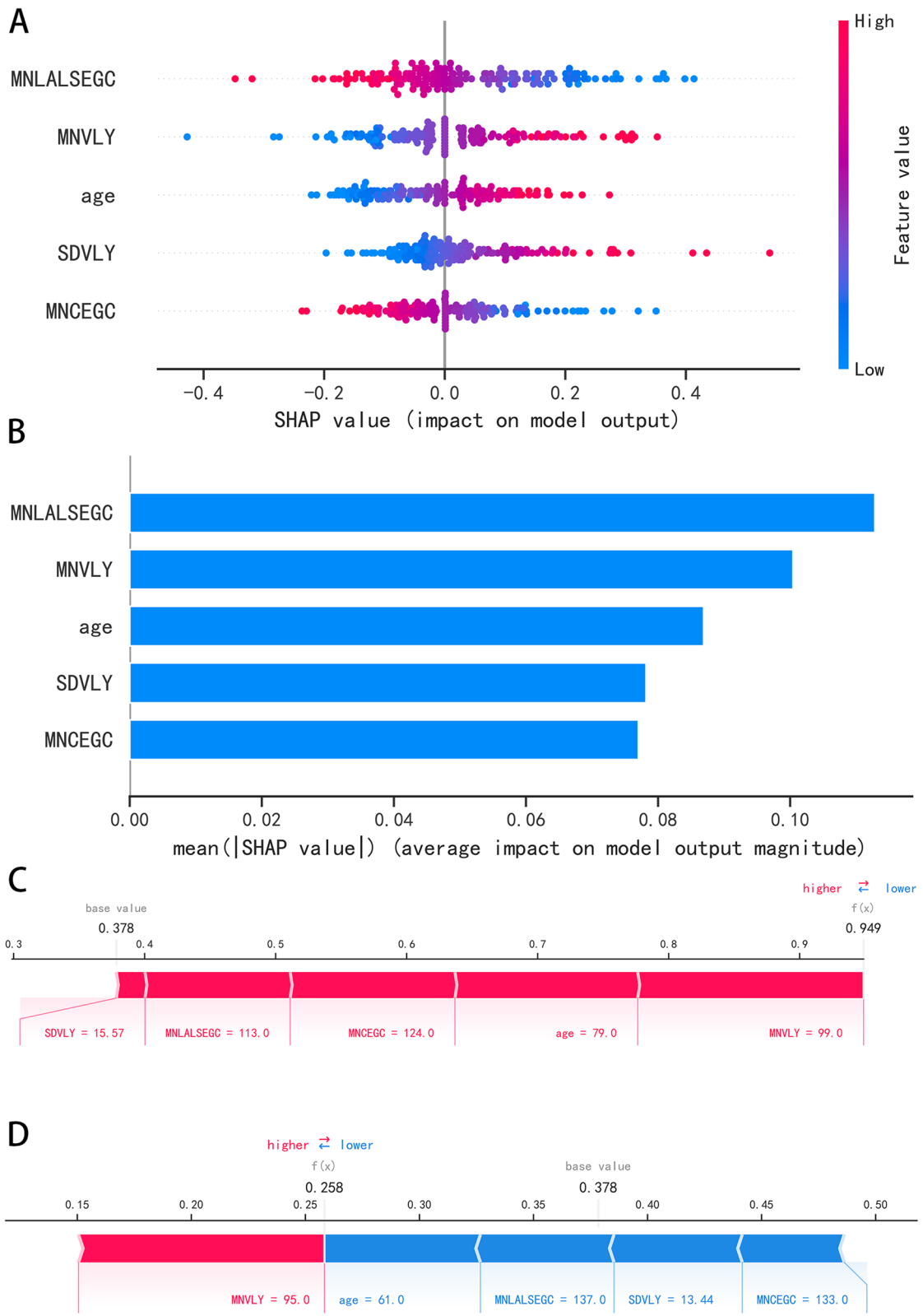

Fig. 6 Model explainability via the SHAP algorithm. **A** The horizontal SHAP value represents the influence on the prediction result, and the vertical coordinate is each indicator, the contribution degree is from low (blue) to high (red). **B** The importance ranking of independent variables. **C** The SHAP force plot of patients with myelodysplastic neoplasms. **D** The SHAP force plot of patients with aplastic anemia

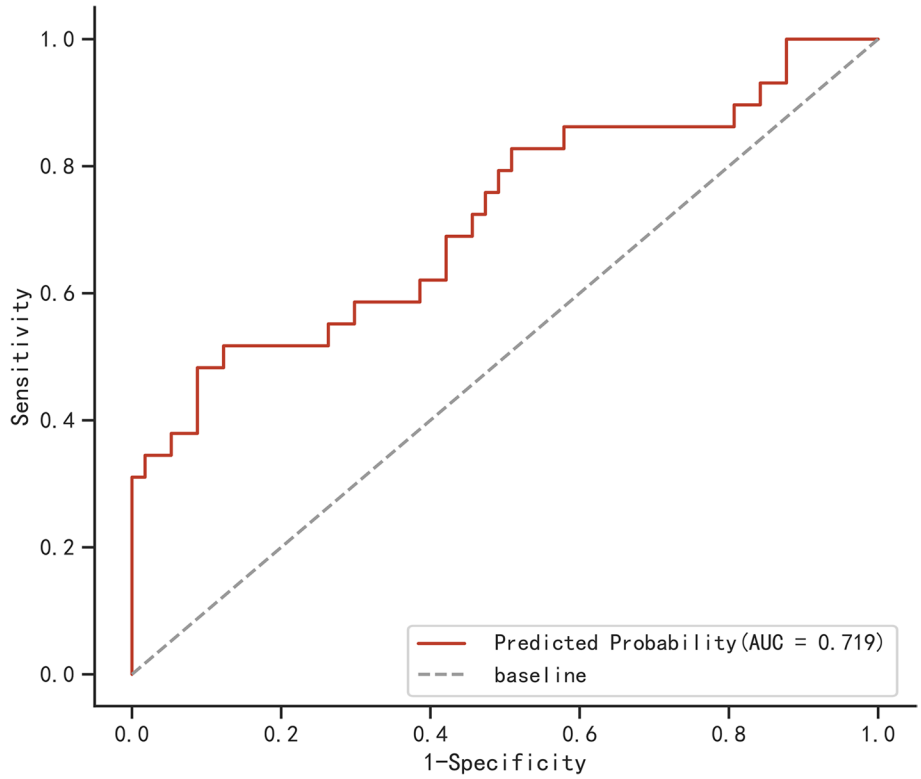

Fig. 7 ROC for the external validation cohort

decreased MNCNE, MNMALSNE, MNUMALSNE, MNLMALSNE, MNLALSNE in MDS patients. Among the lymphocyte-associated CPD parameters, the markedly increased variation in MNVLY, SDVLY, SDCLY, SDMALSLY, SDLMALSLY, SDLALSLY, MNAL2LY, SDAL2LY and decreased in MNLMALSLY in MDS patients. Interestingly, in the two groups, almost all lymphocyte CPD parameters changed, which was also consistent with the two lymphocyte-associated parameters (MNVLY and SDVLY) in the LASSO regression. This finding was consistent with the previous findings [43].

Almost all of the CPD parameters changed in the early granulated cells, which was also consistent with the two CPD parameters in the LASSO regression (MNLALSEGC and MNCEGC). These could be dysplastic features of myelodysplastic neoplasms.

In the MDS group, the degree of heterogeneity of SD measurements was increased, and other studies that have utilized various hematology analyzers to explore CPD in MDS patients have also reported heterogeneity in cellular characteristics [21, 36, 43]. In spite of this, it is unclear what the mechanisms behind this matter are.

However, the current study had some limitations. Firstly, the sample size was relatively small, consisting of only 160 elderly individuals with a diagnosis of AA and MDS. It might lead to biases in the model when generalized. In future research, larger studies with more diverse patient populations and data from multiple centers are needed to further validate the model and assess its performance in real-world clinical settings. Secondly, this model was only validated using Chinese patients. Future studies should include patients from diverse countries and ethnic backgrounds to confirm the generalizability of the model. Additionally, there may be some inevitable bias in clinicians' assessments of disease severity, which could introduce subjective elements. Finally, this research only focused on investigational CPD parameters between AA and MDS patients, without considering other potential biomarkers. In terms of future research directions, we aim to explore ways to optimize the model, such as incorporating new biomarkers (like reticulocytes, bone marrow blast percentage, or other routine blood parameters) or refining the existing algorithm, to improve its accuracy and reliability.

Conclusions

In conclusion, a recognition machine learning model based on CPD parameters was constructed to predict which AA and MDS the patient was. Five filtered indicators were utilized to develop the ML models. The logistic regression model excelled in predicting performance

compared to the other five models (XGBoost, AdaBoost, SVM, LightGBM, and random forest). This model exhibited excellent discrimination and calibration, making it well-suited for clinical application. The model may be a powerful tool in scenarios where timely and accurate diagnosis is critical but resources are limited. This could enable early screening for cytopenic patients (AA or MDS) and guide clinical decision-making, especially in lower-level hospitals.

Abbreviations

MN	Mean
SD	Standard deviation
MNVNE	Mean of volume of neutrophils
SDVNE	SD of volume of neutrophils
MNCNE	Mean of conductivity of neutrophils
SDCNE	SD of conductivity of neutrophils
MNMALSNE	Mean of median angle light scatter of neutrophils
SDMALSNE	SD of median angle light scatter of neutrophils
MNUMALSNE	Mean of upper median angle light scatter of neutrophils
SDUMALSNE	SD of upper median angle light scatter of neutrophils
MNLMALSNE	Mean of lower median angle light scatter of neutrophils
SDLMALSNE	SD of lower median angle light scatter of neutrophil
MNLALSNE	Mean of low angle light scatter of neutrophils
SDLALSNE	SD of low angle light scatter of neutrophils
MNAL2NE	Mean of axial light loss of neutrophils
SDAL2NE	SD of axial light loss of neutrophils
MNVLY	Mean of volume of lymphocytes
SDVLY	SD of volume of lymphocytes
MNCLY	Mean of conductivity of lymphocytes
SDCLY	SD of conductivity of lymphocytes
MNMALSLY	Mean of median angle light scatter of lymphocytes
SDMALSLY	SD of median angle light scatter of lymphocytes
MNUMALSLY	Mean of upper median angle light scatter of lymphocytes
SDUMALSLY	SD of upper median angle light scatter of lymphocytes
MNLMALSLY	Mean of lower median angle light scatter of lymphocytes
SDLMALSLY	SD of lower median angle light scatter of lymphocytes
MNLALSLY	Mean of low angle light scatter of lymphocytes
SDLALSLY	SD of low angle light scatter of lymphocytes
MNAL2LY	Mean of axial light loss of lymphocytes
SDAL2LY	SD of axial light loss of lymphocytes
MNVMO	Mean of volume of monocytes
SDVMO	SD of volume of monocytes
MNCMO	Mean of conductivity of monocytes
SDCMO	SD of conductivity of monocytes
MNMALSMO	Mean of median angle light scatter of monocytes
SDMALSMO	SD of median angle light scatter of monocytes
MNUMALSMO	Mean of upper median angle light scatter of monocytes
SDUMALSMO	SD of upper median angle light scatter of monocytes
MNLMALSMO	Mean of lower median angle light scatter of monocytes
SDLMALSMO	SD of lower median angle light scatter of monocytes
MNLALSMO	Mean of low angle light scatter of monocytes
SDLALSMO	SD of low angle light scatter of monocytes
MNAL2MO	Mean of axial light loss of monocytes
SDAL2MO	SD of axial light loss of monocytes
SDAL2EGC	SD of axial light loss of early granulated cells
MNAL2EGC	Mean of axial light loss of early granulated cells
SDLALSEGC	SD of low angle light scatter of early granulated cells
MNLALSEGC	Mean of low angle light scatter of early granulated cells
SDLMALSEGC	SD of lower median angle light scatter of early granulated cells
MNLMALSEGC	Mean of lower median angle light scatter of early granulated cells
SDUMALSEGC	SD of upper median angle light scatter of early granulated cells
MNUMALSEGC	Mean of upper median angle light scatter of early granulated cells
SDMALSEGC	SD of median angle light scatter of early granulated cells
MNMALSEGC	Mean of median angle light scatter of early granulated cells
SDCEGC	SD of conductivity of early granulated cells
MNCEGC	Mean of conductivity of early granulated cells
SDVEGC	SD of volume of early granulated cells
MNVEGC	Mean of volume of early granulated cells
SDAL2EO	SD of axial light loss of eosinophils
MNAL2EO	Mean of axial light loss of eosinophils
SDLALSEO	SD of low angle light scatter of eosinophils
MNLALSEO	Mean of lower angle light scatter of eosinophils
SDLMALSEO	SD of lower median angle light scatter of eosinophils
MNLMALSEO	Mean of lower median angle light scatter of eosinophils
SDUMALSEO	SD of upper median angle light scatter of eosinophils
MNUMALSEO	Mean of upper median angle light scatter of eosinophils
SDMALSEO	SD of median angle light scatter of eosinophils
MNMALSEO	Mean of median angle light scatter of eosinophils
SDCEO	SD of conductivity of eosinophils
MNCEO	Mean of conductivity of eosinophils
SDVEO	SD of volume of eosinophils
MNVEO	Mean of volume of eosinophils

Acknowledgements
Not applicable.

Authors' contributions
ZZ: Investigation, Writing– review & editing. YQ and XL: Formal Analysis, Writing– original draft. ZD: Data curation, Methodology, Writing– review & editing. YY: Project administration, Writing– review & editing. All authors reviewed the manuscript.

Funding
This work was supported by the Zhejiang Province Traditional Chinese Medicine Science and Technology Plan (2023ZL054).

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics statement and consent to participate
This study was approved by the Ethics Committee of the First Affiliated Hospital of Zhejiang Chinese Medical University with approval number 2024-KLS-348–01. Written informed consent to participate was obtained from all of the participants in the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]School of Medical Technology and Information Engineering, Zhejiang Chinese Medical University, Hangzhou, China. [2]Department of Laboratory Medicine, The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), Hangzhou, China. [3]The First School of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou, China. [4]Adicon Clinical Laboratories, Hangzhou 310023, Zhejiang, China.

Received: 29 February 2024 Accepted: 22 November 2024
Published online: 18 December 2024

References

1. Young NS. Aplastic anemia. N Engl J Med. 2018;379(17):1643–56.
2. DeZern AE, Churpek JE. Approach to the diagnosis of aplastic anemia. Blood Adv. 2021;5(12):2660–71.
3. Cazzola M. Myelodysplastic Syndromes. N Engl J Med. 2020;383(14):1358–74.
4. Kim SY, Park Y, Kim H, et al. Discriminating myelodysplastic syndrome and other myeloid malignancies from non-clonal disorders

by multiparametric analysis of automated cell data. Clin Chim Acta. 2018;480:56–64.
5. Bennett JM, Orazi A. Diagnostic criteria to distinguish hypocellular acute myeloid leukemia from hypocellular myelodysplastic syndromes and aplastic anemia: recommendations for a standardized approach. Haematologica. 2009;94(2):264–8.
6. Tanaka TN, Bejar R. MDS overlap disorders and diagnostic boundaries. Blood. 2019;133(10):1086–95.
7. Kulasekararaj A, Cavenagh J, Dokal I, et al. Guidelines for the diagnosis and management of adult aplastic anaemia: a British Society for Haematology Guideline. Br J Haematol. 2024;204(3):784–804.
8. Yang HS, Rhoads DD, Sepulveda J, et al. Building the Model. Arch Pathol Lab Med. 2023;147(7):826–36.
9. Prelaj A, Miskovic V, Zanitti M, et al. Artificial intelligence for predictive biomarker discovery in immuno-oncology: a systematic review. Ann Oncol. 2024;35(1):29–65.
10. Alhajahjeh A, Nazha A. Unlocking the potential of artificial intelligence in acute myeloid leukemia and myelodysplastic syndromes. Curr Hematol Malig Rep. 2024;19(1):9–17.
11 Harte JV, NíChoileáin C, Grieve C, et al. A panhaemocytometric approach to COVID-19: the importance of cell population data on Sysmex XN-series analysers in severe disease. Clin Chem Lab Med. 2023;61(3):e43–7.
12. Lien F, Lin HS, Wu YT, et al. Bacteremia detection from complete blood count and differential leukocyte count with machine learning: complementary and competitive with C-reactive protein and procalcitonin tests. BMC Infect Dis. 2022;22(1):287.
13. Huang YH, Chen CJ, Shao SC, et al. Comparison of the diagnostic accuracies of monocyte distribution width, procalcitonin, and C-reactive protein for sepsis: a systematic review and meta-analysis. Crit Care Med. 2023;51(5):e106–14.
14. Hausfater P, Robert Boter N, Morales Indiano C, et al. Monocyte distribution width (MDW) performance as an early sepsis indicator in the emergency department: comparison with CRP and procalcitonin in a multicenter international European prospective study. Critical care (London, England). 2021;25(1):227.
15. Famiglini L, Campagner A, Carobene A, et al. A robust and parsimonious machine learning method to predict ICU admission of COVID-19 patients. Med Biol Eng Compu. 2022;30:1–13.
16. Cai J, Liu Z, Wang Y, et al. Construction of the prediction model for multiple myeloma based on machine learning. Int J Lab Hematol. 2024;46(5):918–26.
17. Ambayya A, Sathar J, Hassan R. Neoteric Algorithm Using Cell Population Data (VCS Parameters) as a Rapid Screening Tool for Haematological Disorders. Diagnostics (Basel, Switzerland). 2021;11(9):1652.
18. Ambayya A, Sahibon S, Yang TW, et al. A Novel Algorithm Using Cell Population Data (VCS Parameters) as a Screening Discriminant between Alpha and Beta Thalassemia Traits. Diagnostics (Basel, Switzerland). 2021;11(11):2163.
19. Gaspar BL, Sharma P, Varma N, et al. Unique characteristics of leukocyte volume, conductivity and scatter in chronic myeloid leukemia. Biomedical journal. 2019;42(2):93–8.
20. Zhu J, Lemaire P, Mathis S, et al. Machine learning-based improvement of MDS-CBC score brings platelets into the limelight to optimize smear review in the hematology laboratory. BMC Cancer. 2022;22(1):972.
21. Pozdnyakova O, Niculescu RS, Kroll T, et al. Beyond the routine CBC: machine learning and statistical analyses identify research CBC parameter associations with myelodysplastic syndromes and specific underlying pathogenic variants. J Clin Pathol. 2023;76(9):624–31.
22. Ravalet N, Foucault A, Picou F, et al. Automated Early Detection of Myelodysplastic Syndrome within the General Population Using the Research Parameters of Beckman-Coulter DxH 800 Hematology Analyzer. Cancers. 2021;13(3):389.
23. Weinberg OK, Hasserjian RP. The current approach to the diagnosis of myelodysplastic syndromes. Semin Hematol. 2019;56(1):15–21.
24. Plander M, Kányási M, Szendrei T, et al. Flow cytometry in the differential diagnosis of myelodysplastic neoplasm with low blasts and cytopenia of other causes. Pathol Oncol Res. 2024;30:1611811.
25. Li MY, Xu YY, Kang HY, et al. Quantitative detection of id4 gene aberrant methylation in the differentiation of myelodysplastic syndrome from aplastic anemia. Chin Med J. 2015;128(15):2019–25.
26. Park SH, Jeong J, Lee SH, et al. Comparison of High Sensitivity and Conventional Flow Cytometry for Diagnosing Overt Paroxysmal Nocturnal Hemoglobinuria and Detecting Minor Paroxysmal Nocturnal Hemoglobinuria Clones. Ann Lab Med. 2019;39(2):150–7.
27. Wu J, Zhang L, Yin S, et al. Differential Diagnosis Model of Hypocellular Myelodysplastic Syndrome and Aplastic Anemia Based on the Medical Big Data Platform. Complexity. 2018;2018:1–12.
28. Hematology Branch of Chinese Medical Association Red Blood Cell Disease (Anemia) Group. the interpretation of guidelines for the diagnosis and management of aplastic anemia in China (2022). Ch J Hematol. 2022;43(11):881–8.
29 Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. Leukemia. 2022;36(7):1703–19.
30. Sun L, Babushok DV. Secondary myelodysplastic syndrome and leukemia in acquired aplastic anemia and paroxysmal nocturnal hemoglobinuria. Blood. 2020;136(1):36–49.
31. Sankar D, Oviya IR. Multidisciplinary approaches to study anaemia with special mention on aplastic anaemia (Review). Int J Mol Med. 2024;54(5):95.
32. Gómez-Rojas S, Segura GP, Ollé J, et al. A machine learning tool for the diagnosis of SARS-CoV-2 infection from hemogram parameters. J Cell Mol Med. 2023;27(22):3423–30.
33. Zhu J, Clauser S, Freynet N, et al. Automated detection of dysplasia: data mining from our hematology analyzers. Diagnostics. 2022;12(7):1556.
34. Chabot-Richards DS, George TI. White blood cell counts: reference methodology. Clin Lab Med. 2015;35(1):11–24.
35. Kim H, Hur M, Yi JH, et al. Detection of blasts using flags and cell population data rules on Beckman Coulter DxH 900 hematology analyzer in patients with hematologic diseases. Clin Chem Lab Med. 2024;62(5):958–66.
36. Park SH, Kim HK, Jeong J, et al. Research use only and cell population data items obtained from the Beckman Coulter DxH800 automated hematology analyzer are useful in discriminating MDS patients from those with cytopenia without MDS. J Hematop. 2023;16(3):143–54.
37. Contejean A, Resche-Rigon M, Tamburini J, et al. Aplastic anemia in the elderly: a nationwide survey on behalf of the French Reference Center for Aplastic Anemia. Haematologica. 2019;104(2):256–62.
38. Montané E, Ibáñez L, Vidal X, et al. Epidemiology of aplastic anemia: a prospective multicenter study. Haematologica. 2008;93(4):518–23.
39. Yoshizato T, Dumitriu B, Hosokawa K, et al. Somatic Mutations and Clonal Hematopoiesis in Aplastic Anemia. N Engl J Med. 2015;373(1):35–47.
40. Adrianzen-Herrera D, Sparks AD, Singh R, et al. Impact of preexisting autoimmune disease on myelodysplastic syndromes outcomes: a population analysis. Blood Adv. 2023;7(22):6913–22.
41. Fattizzo B, Levati GV, Giannotta JA, et al. Low-Risk myelodysplastic syndrome revisited: morphological, autoimmune, and molecular features as predictors of outcome in a single center experience. Front Oncol. 2022;12:795955.
42. Barcellini W, Fattizzo B, Cortelezzi A. Autoimmune hemolytic anemia, autoimmune neutropenia and aplastic anemia in the elderly. Eur J Intern Med. 2018;58:77–83.
43. Shestakova A, Nael A, Nora V, et al. Automated leukocyte parameters are useful in the assessment of myelodysplastic syndromes. Cytometry B Clin Cytom. 2021;100(3):299–311.

Publisher's Note