

Machine Learning-Based Early Screening of Myelodysplastic Syndromes Using Complete Blood Count

Guoqing Zhu¹, Juan Zhang²

¹Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin, China

²Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu, China

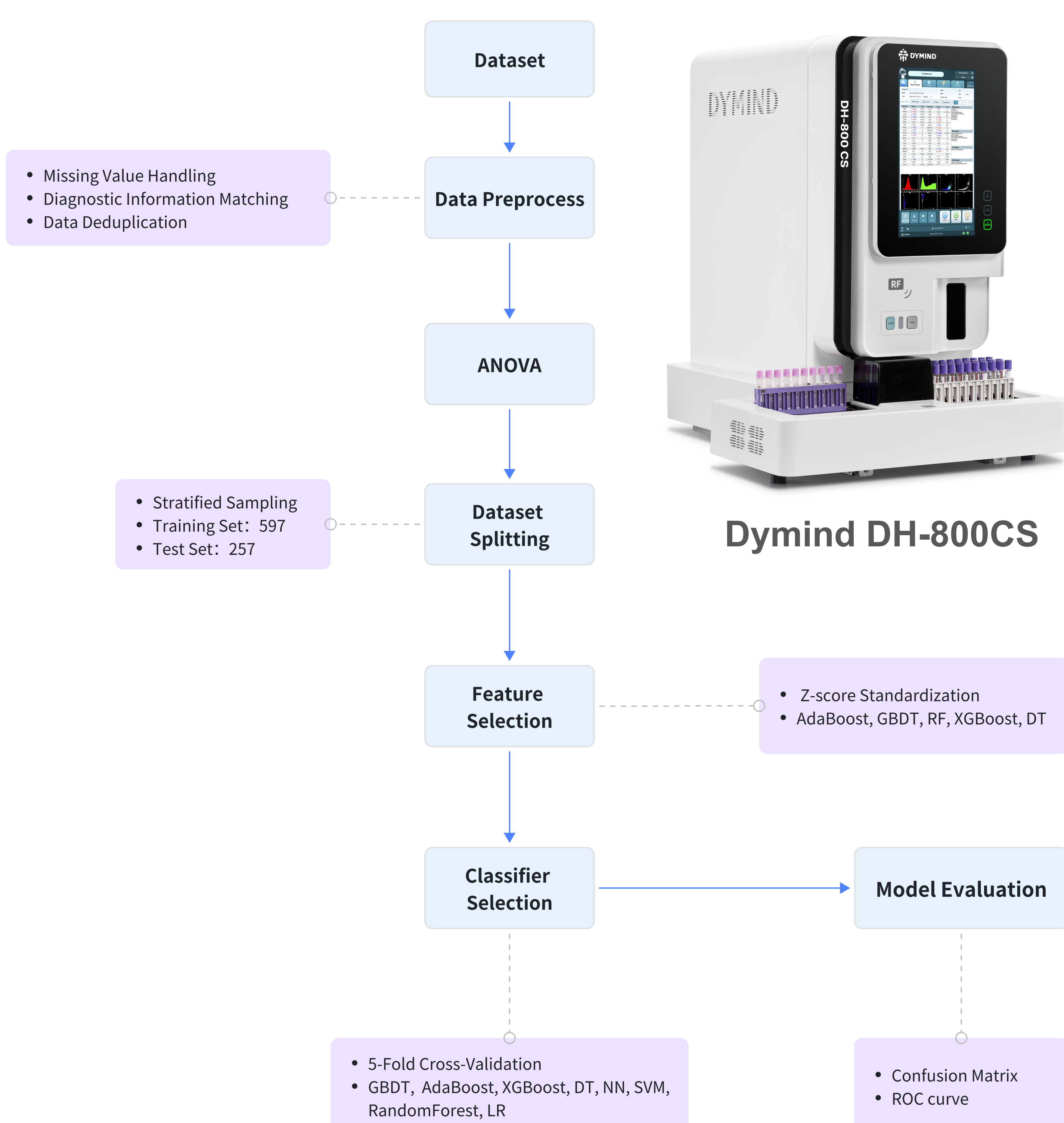
✉ zhuguoqing2026@163.com

INTRODUCTION

Myelodysplastic syndromes (MDS) are hematopoietic stem cell-originated disorders with high progression risk to acute myeloid leukemia. Current diagnosis relies on invasive bone marrow aspiration and cytogenetic testing, which are time-consuming, costly, and susceptible to interference from cytopenia caused by inflammation or immune disorders. There is an urgent need for a cost-effective, rapid laboratory screening tool to differentiate MDS from other cytopenia. This study aims to develop an MDS risk prediction model using routine blood test parameters to reduce unnecessary invasive procedures and improve early screening.

METHODS

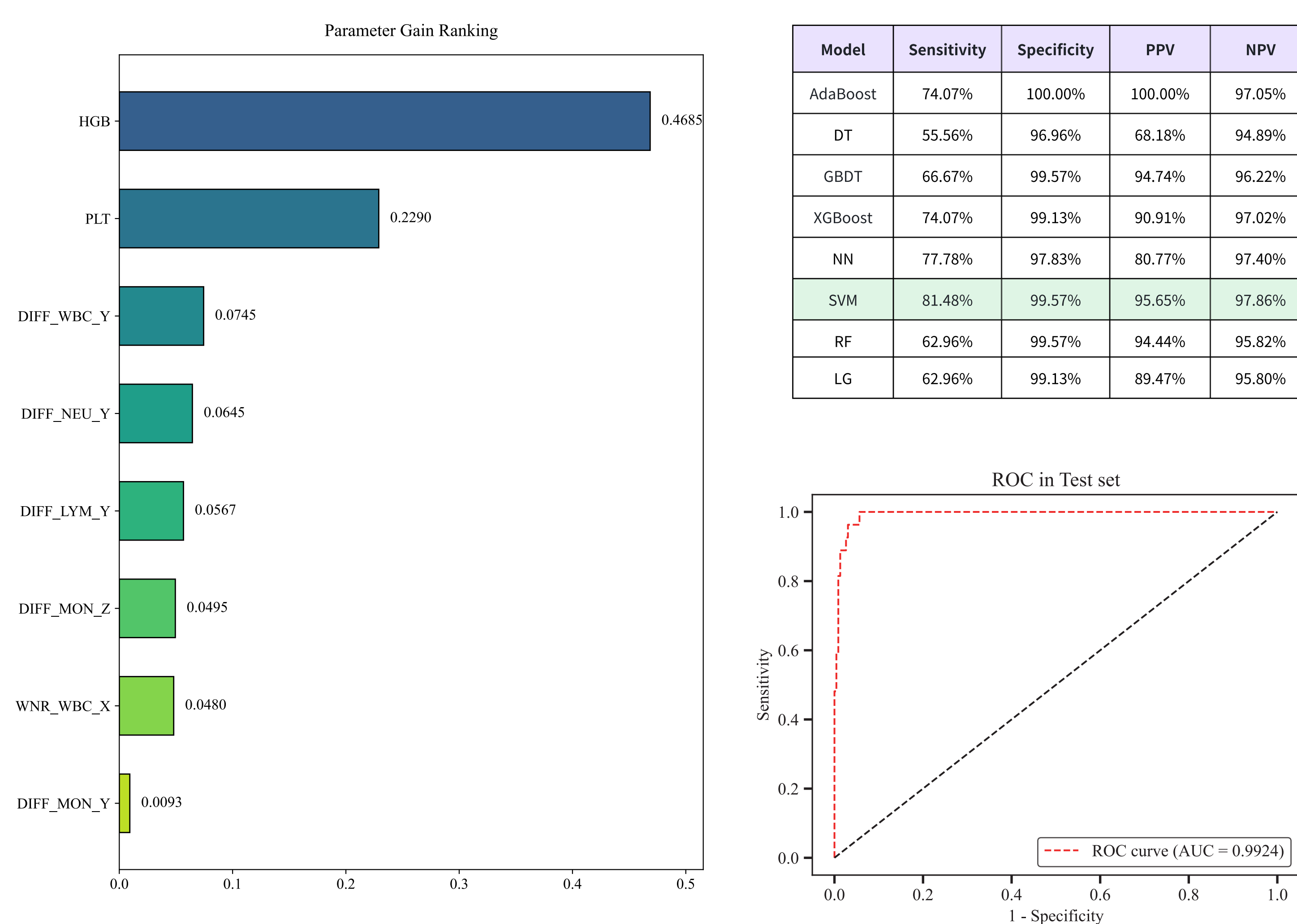
This retrospective cohort study included 854 samples: 80 confirmed MDS patients, 412 cytopenia patients without MDS morphological evidence, and 362 age-matched healthy controls. All diagnoses followed international standards. Complete blood count parameters were analyzed using the Dymind DH-800 serial hematology analyzer. Multiple machine learning algorithms (decision trees, random forests, XGBoost, SVM). Model performance was evaluated by AUC.



RESULTS

Feature selection analysis using the Gain algorithm identified eight key parameters, including two conventional hematological parameters (hemoglobin, HGB; platelet count, PLT) and six leukocyte scatter plot-derived parameters (DIFF_NEU_Y, WNR_WBC_X, DIFF_LYM_Y, DIFF_WBC_Y, DIFF_MON_Y, DIFF_MON_Z). These parameters are ranked in descending order of their Gain values, reflecting their respective contributions to MDS risk prediction. The support vector machine (SVM) demonstrated superior diagnostic performance among tested algorithms.

In an independent validation set (27 MDS, 120 cytopenia, 110 healthy controls), the SVM model achieved 81.48% sensitivity, 99.57% specificity, and an AUC of 0.98 for MDS warning. This performance significantly surpassed the Horiba-MDS scoring system (65% sensitivity, 90.6% specificity, AUC 0.841), showing clear advantages in differentiating MDS from non-malignant cytopenias.



CONCLUSION

We successfully developed an MDS screening model based on routine parameters from the Dymind DH-800 serial hematology analyzer. The high specificity (99.57%) and good sensitivity (81.48%) of the model support its feasibility as a clinical warning tool. This digital solution enables efficient differentiation of MDS from other cytopenias, promoting a shift from invasive to non-invasive screening paradigms.