ARTIFICIAL INTELLIGENCE IN THE DIAGNOSIS OF SEVERE APLASTIC ANEMIA: A SYSTEMATIC REVIEW OF EMERGING DIAGNOSTIC MODELS

Eman M. Nagiub[1], Abdulrahman Nasiri[2], Ali Alahmari[3], Mahmoud Aljurf[3], Hazzaa Alzahrani[3], Mostafa F. Mohammed Saleh[3,4]

1- Hematopathology Unit, Department of Clinical Pathology, Faculty of Medicine, Assiut University, Assiut , Egypt.
2-College of Medicine, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia.
3-Adult Hematology, Transplantation and Cellular Therapy Department, Oncology Center, King Faisal Specialist Hospital and Research Center, Riyadh , Saudi Arabia.
4-Clinical Hematology Unit, Department of Internal Medicine, Faculty of Medicine, Assiut University, Assiut, Egypt.

INTRODUCTION

Differentiating severe aplastic anemia (SAA) from other cytopenic disorders, particularly myelodysplastic syndromes (MDS), remains a major diagnostic challenge. Artificial intelligence (AI) has emerged as a promising tool to enhance diagnostic accuracy using multimodal data. Conventional diagnostic approaches rely heavily on invasive bone marrow examination and expert morphological interpretation, which may be limited by inter-observer variability and delayed access in some settings. Recent advances in machine learning and deep learning enable integration of laboratory, imaging, and clinical data, offering a non-invasive and scalable strategy to improve early diagnosis and clinical decision-making.

METHODS

We conducted a systematic review of PubMed-indexed studies (2019–2026) evaluating AI-based diagnostic models in aplastic anemia and bone marrow failure syndromes. Eligible studies included those utilizing machine learning (ML) or deep learning (DL) applied to morphology, laboratory data, imaging, or integrated clinical-genomic variables. Extracted outcomes included diagnostic accuracy, sensitivity, specificity, and area under the curve (AUC).

RESULTS

A total of 10 studies met inclusion criteria, encompassing >6,000 patients across multiple diagnostic platforms (Table1).

Morphology-based DL models demonstrated the highest diagnostic performance, with CNN-based systems achieving sensitivity of 96.2%, specificity of 100%, and AUC of 0.99 in differentiating SAA from MDS.

Bone marrow smear DL models achieved multiclass AUCs up to 0.968. Laboratory-based ML models using complete blood count (CBC) parameters showed robust performance (AUC 0.92–0.936), with platelet count, WBC, and RBC indices as key predictors.

Biomarker-integrated models incorporating thrombopoietin achieved AUC values up to 0.99. Explainable AI models differentiating iron deficiency anemia from SAA achieved accuracy up to 96%.

Imaging-based radiomics using MRI demonstrated significant discriminatory capability, with markedly higher marrow fat fraction in SAA (≈79% vs. 43% in MDS).

Integrated clinical-genomic models achieved accuracies of 87–89%, identifying telomere length and phenotypic features as critical variables.

CONCLUSION

AI-driven diagnostic models demonstrate high accuracy across multiple modalities in distinguishing SAA from overlapping hematologic conditions. These tools offer a non-invasive, scalable approach for early triage and diagnostic support, particularly in resource-limited settings. Prospective validation and integration into clinical workflows are needed to enable real-world implementation.

CONTACT INFORMATION

Mostafa F. Mohammed Saleh, MD

Adult Hematology, Transplantation and Cellular Therapy Department, Oncology Center, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia.

Email: mmostafa@kfshrc.edu.sa

Tel: +966-554740623

Table 1: Summary of AI-Based Studies in the Diagnosis of Severe Aplastic Anemia

Study (Year)	Modality	Data Source	Sample Size	AI Model	Key Variables	Performance	Key Findings
Kimura et al. (2019)	Morphology (PB smear)	Image dataset	>695,000 images (3,261 smears)	CNN + XGBoost	Cell morphology (17 types, 97 features)	Sens 96.2%, Spec 100%, AUC 0.99	Accurate differentiation of MDS vs AA using dysplastic features
Wang et al. (2022)	Morphology (BM smear)	Image dataset	Not specified	CNN	Morphologic patterns	AUC up to 0.968, Acc ~93%	Multiclass classification (AA vs MDS vs AML)
Wang et al. (2026)	Laboratory (CBC)	Clinical dataset	>3,800 patients	LightGBM	PLT, RBC, WBC	AUC 0.92	Effective non-invasive triage tool
Seo et al. (2024)	Laboratory (CBC)	Clinical dataset	Large cohort	ML classifier	CBC indices	AUROC 0.915–0.936	High sensitivity and NPV (>98%)
Zhu et al. (2025)	Biomarker + CBC	Clinical dataset	751 patients	LR, XGBoost, LightGBM	TPO + CBC	AUC up to 0.99 (AA)	TPO significantly improves diagnostic accuracy
Darshan et al. (2025)	Laboratory + XAI	Clinical dataset	500 patients	Ensemble + LightGBM	PLT, MCV, WBC, HGB	Accuracy 96%	Explainable AI improves interpretability
Qi et al. (2024)	Laboratory (CPD)	Clinical dataset	160 patients	Logistic regression	CPD parameters + age	AUC 0.79 (test), 0.72 (external)	Useful in elderly population
Xiang et al. (2023)	Imaging (MRI radiomics)	Imaging dataset	77 patients	SVM	Marrow fat fraction (IDEAL-IQ)	Significant separation ($p<0.001$)	Higher fat fraction in AA (79% vs 43%)
Gutierrez-Rodrigues et al. (2023)	Clinical + Genomic	Multicenter	Not specified	Ensemble ML	Telomere length, clinical features	Accuracy ~89%	Differentiates acquired vs inherited BMF
Massaccesi et al. (2026)	Clinical + Genomic	Clinical dataset	140 patients	RF + K-means	Telomere length, phenotype	Accuracy 87.2%	Reclassification of unclear cases (17%)

Abbreviations: AA = aplastic anemia; MDS = myelodysplastic syndromes; AML = acute myeloid leukemia; CBC = complete blood count; CPD = cell population data; TPO = thrombopoietin; CNN = convolutional neural network; RF = random forest; LR = logistic regression; XAI = explainable AI.