Artificial Intelligence-Based Prediction of Graft-Versus-Host Disease After Allogeneic Hematopoietic Stem Cell Transplantation: Clinical Promise or Premature Adoption?

Karabulut, Zeynep Tuğba[1]

[1]*Adana City Hospital, Department of Hematology*

Introduction: Acute and chronic graft-versus-host disease (GVHD) remain among the leading causes of morbidity and mortality after allogeneic hematopoietic stem cell transplantation (allo-HSCT), affecting a substantial proportion of recipients despite advances in donor selection and prophylaxis regimens. In recent years, machine learning (ML) models incorporating clinical, genetic, and laboratory data have been proposed to predict GVHD risk before or shortly after transplantation, raising the question of whether such tools can meaningfully inform prophylaxis and monitoring strategies.

Aim: This review critically evaluates current evidence on AI/ML-based GVHD prediction models, discussing their reported performance, methodological limitations, and readiness for clinical implementation.

Methods: A narrative synthesis of recent studies (2020–2025) applying machine learning algorithms (random forest, gradient boosting, neural networks, and ensemble models) to GVHD prediction was conducted. Studies were evaluated based on model performance metrics, data sources (clinical, genetic, transcriptomic, and electronic health record data), validation strategies, and clinical applicability.

Findings: Reported models show moderate-to-high discriminative performance, with AUC values reaching approximately 0.70–0.80 in pediatric cohorts using combined clinical and genetic predictors. Larger studies integrating longitudinal electronic health record data have shown improved prediction of both acute and chronic GVHD compared with models relying solely on baseline transplant variables. Transcriptomic approaches using bone marrow gene expression have further identified candidate signaling pathways associated with GVHD risk, suggesting potential biological insight beyond prediction alone. However, most studies remain retrospective, single-center, and based on relatively small cohorts, with limited external validation across diverse patient populations and transplant protocols.

Conclusion: While AI-based GVHD prediction models demonstrate promising discriminative ability and may eventually support individualized prophylaxis decisions, the current evidence base remains insufficient for routine clinical adoption. Larger multicenter cohorts, prospective validation, and transparent integration into clinical decision workflows are needed before these tools can be considered reliable adjuncts to clinical judgment in GVHD risk stratification.