

EVALUATING AN AI MODEL'S HEMATOLOGY CONSULTATION PERFORMANCE FOR HOSPITALIZED PATIENTS WITH BLEEDING DISORDERS

G. Sottilotta¹, F. Misiti², G. Praticò¹, D. Megalizzi¹, F. Luise³

¹UOSD Microcitemie - Emostasi e Trombosi. Grande Ospedale Metropolitano di Reggio Calabria, Italy; ²Associazione Emofiliaci di Reggio Calabria, Italy;

³Laboratorio Analisi. Grande Ospedale Metropolitano di Reggio Calabria, Italy

Introduction. Artificial intelligence (AI) is increasingly proposed as a decision-support tool in clinical medicine. Decisions in hemostasis and thrombosis often require nuanced interpretation of bleeding history, laboratory results, and procedural risk. This study evaluated whether a general-purpose AI could provide consultation-level guidance comparable to an expert hematologist for inpatients across diverse hospital services.

Objectives. The objective was to assess the concordance between AI-generated hematology recommendations and those issued by real-life hematology consultants for inpatient questions concerning diagnostic workup, prophylaxis, or on-demand therapy for bleeding and coagulation disorders.

Materials and Methods. We retrospectively selected 18 prior consultations deemed suitable for head-to-head comparison. The cases involved 18 patients (12 male, 6 female) aged 5 to 82 years (mean 47.4). For each case, the original clinical question and context were provided to the AI, which then generated a de novo response. Concordance between the AI and the consultant's recommendation was scored from 0 to 100 against the clinician's answer. A score of 90-100 indicated virtually overlapping recommendations; 75-89 indicated very close concordance with minor or prudential differences; 60-74 indicated partial concordance with a similar rationale but different

therapy; 40-59 indicated a clinically important difference; and below 40 indicated a substantially incorrect framing of the problem. The AI also produced a self-rating after comparing its own output to the clinician's response.

Results. across the 18 consultations, the mean AI self-assessed score versus the clinician standard was 78.4. This indicates overall very close concordance, with predominantly minor or caution-driven deviations. Performance was consistent across surgical and medical services. There was frequent alignment on perioperative hemostatic strategies, factor replacement thresholds, and laboratory triage, such as targeted factor assays, von Willebrand evaluation, and mixing studies when appropriate. Discrepancies most often reflected therapeutic conservatism, such as higher suggested replacement targets or extended monitoring, or differences in drug selection within a class rather than opposing clinical logic. In some cases, partial concordance arose from similar diagnostic reasoning but divergence in dosing intensity or duration. No cases were rated as grossly misframed relative to the human consultant. Patient characteristics and AI concordance are provided in table 1.

Conclusions. In this real-world consultation set spanning multiple specialties and age groups, an AI model's recommendations for bleeding and coagulation questions were, on average, very close to expert hematology guidance

2.07 Novel therapeutic products and approaches (including gene therapy)

Table1. Patient Characteristics and AI Concordance Results	
Total Patients (n)	18
Sex	12 Male (66.7%), 6 Female (33.3%)
Age Range	5 – 82 years
Mean Age	47.4 years
Consultation Services (Origin)	Neurosurgery (6), Otorhinolaryngology (5), Gynecology (3), Urology (2), Internal Medicine (1), Orthopedics (1)
AI Concordance Score	Mean Score: 78.4 (Range: 45-98)
Bleeding Disorder / Factor Deficiency	Number of patients
Factor VII Deficiency	4
Combined Factor VII & Factor V Deficiency	3
Combined Factor VII & Factor XIII Deficiency	5
Combined Factor X & XI Deficiency	1
von Willebrand Disease	1
Factor XII Deficiency	1
Hypofibrinogenemia	1
Isolated Prolonged Prothrombin Time (PT)	1
Isolated Prolonged Partial Thromboplastin Time (aPTT)	1