UNIVERSITAS
GADJAH MADA
Sistem Pendukung Keputusan Klinis (CDSS) pada Gangguan Trombosit Berbasis Complete Blood Count (CBC)
Menggunakan Machine Learning dan Explainable AI
Muhammad Wahyu Tri Saputra, Ir. Syukron Abu Ishaq Alfarozi, S.T., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Gangguan trombosit seperti *Immune Thrombocytopenia* (ITP), Demam Berdarah Dengue (DBD), dan Trombositosis Esensial memiliki profil hematologi yang dapat tumpang tindih, terutama antara ITP dan DBD pada fase trombositopenia akut. Hal ini menyulitkan diagnosis banding berbasis pemeriksaan *Complete Blood Count* (CBC) standar di fasilitas kesehatan tingkat pertama yang kekurangan tenaga spesialis. Penelitian ini bertujuan mengembangkan *Clinical Decision Support System* (CDSS) berbasis arsitektur *Tri-Brid* untuk mengklasifikasikan Normal, ITP, DBD, dan Trombositosis menggunakan parameter CBC standar. Sistem mengintegrasikan probabilitas *machine learning* (Pilar I), skor gejala klinis (Pilar II), aturan WHO 2009 dan IWG 2009 (Pilar III), serta *Explainable Artificial Intelligence* (XAI) berbasis SHAP.

Dataset terdiri atas 393 data primer CBC dari rekam medis retrospektif RSA UGM. Studi ablasi membandingkan 36 kombinasi *pipeline* yang mencakup empat metode seleksi fitur, tiga metode penskalaan, dan tiga algoritma klasifikasi dengan optimasi *GridSearchCV* pada empat skenario prapemrosesan, yaitu *baseline* tanpa imputasi dan penyeimbangan kelas, *KNNImputer*, *KNNImputer* dengan SMOTE, dan *KNNImputer* dengan SMOTE-ENN. Pada *test set* yang identik, SMOTE hanya menaikkan rata-rata MCC sebesar 0,0071 (tidak signifikan), sedangkan ENN menurunkannya sebesar 0,0351. Dibandingkan *listwise deletion*, *KNNImputer* berasosiasi dengan kenaikan rata-rata MCC sebesar 0,0755, tetapi selisih ini mencampur efek imputasi dengan perbedaan data uji. *Pipeline* yang direkomendasikan, yaitu StandardScaler, *KNNImputer*, *Mutual Information*, SMOTE, dan *Logistic Regression*, memperoleh MCC sebesar 0,9584 (IK 95% 0,899–1,000) pada *hold-out test set* dan rata-rata MCC validasi silang sebesar 0,8605. Analisis SHAP pada model final menunjukkan PLT sebagai fitur pembeda utama (total *mean absolute SHAP* 11,62; 41,4% dari total kontribusi), sedangkan PLR dan MCH berkontribusi terutama pada prediksi ITP, serta MCV dan `ABS_EOS` pada prediksi DBD. Validasi oleh satu dokter spesialis patologi klinik pada model purwarupa, tanpa informasi gejala, menghasilkan *Initial Concordance* sebesar 70% dan skor CLIX-M sebesar 3,00/4,00 (75,0%). Sistem kemudian diimplementasikan sebagai purwarupa CDSS berbasis *web* dengan visualisasi SHAP.

Hasil penelitian menunjukkan bahwa integrasi CBC standar, *machine learning*, dan XAI dalam arsitektur *Tri-Brid* dapat menjadi pendekatan pendukung untuk diferensiasi gangguan trombosit pada fasilitas kesehatan tingkat pertama.

Kata kunci : CDSS, gangguan trombosit, *Complete Blood Count*, *machine learning*, SHAP

UNIVERSITAS
GADJAH MADA
Sistem Pendukung Keputusan Klinis (CDSS) pada Gangguan Trombosit Berbasis Complete Blood Count (CBC)
Menggunakan Machine Learning dan Explainable AI
Muhammad Wahyu Tri Saputra, Ir. Syukron Abu Ishaq Alfarozi, S.T., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Platelet disorders such as Immune Thrombocytopenia *(ITP), Dengue Hemorrhagic Fever (DHF), and Essential Thrombocythemia may exhibit overlapping hematological profiles, particularly ITP and DHF during the acute thrombocytopenic phase. This condition complicates diagnostic differentiation based on standard* Complete Blood Count *(CBC) examinations in primary healthcare facilities with limited access to specialists. This study aimed to develop a* Clinical Decision Support System *(CDSS) based on a* Tri-Brid *architecture to classify Normal, ITP, DHF, and Thrombocytosis cases using standard CBC parameters. The system integrates* machine learning *probabilities (Pillar I), clinical symptom scores (Pillar II), WHO 2009 and IWG 2009 rules (Pillar III), and SHAP-based* Explainable Artificial Intelligence *(XAI).*

The dataset *consisted of 393 primary CBC data records obtained from retrospective medical records at RSA UGM. The ablation study compared 36* pipeline *combinations comprising four feature selection methods, three scaling methods, and three classification algorithms optimized using* GridSearchCV *under four preprocessing scenarios: a* baseline *without imputation or class balancing,* KNNImputer*,* KNNImputer *with SMOTE, and* KNNImputer *with SMOTE-ENN. On an identical* test set*, SMOTE added only 0.0071 to the mean MCC (not significant), whereas ENN reduced it by 0.0351.* KNNImputer*, compared with listwise deletion, was associated with a 0.0755 increase in mean MCC, but this difference combines the imputation effect with a difference in test data. The recommended* pipeline*, consisting of StandardScaler,* KNNImputer*, Mutual Information, SMOTE, and* Logistic Regression*, achieved an MCC of 0.9584 (95% CI 0.899–1.000) on the hold-out* test set *and a mean cross-validation MCC of 0.8605. SHAP analysis of the final model identified PLT as the primary discriminative feature (total* mean absolute SHAP *11.62; 41.4% of total contribution), while PLR and MCH contributed mainly to ITP prediction, and MCV and* `ABS_EOS` *to DHF prediction. Validation by a clinical pathologist on the prototype model, without symptom information, yielded an* Initial Concordance *of 70% and a CLIX-M score of 3.00/4.00 (75.0%). The system was subsequently implemented as a* web*-based CDSS prototype featuring SHAP visualization.*

The results indicate that the integration of standard CBC parameters, machine learning*, and XAI within a* Tri-Brid *architecture can serve as a decision-support approach for differentiating platelet disorders in primary healthcare facilities.*

Keywords : CDSS, platelet disorders, Complete Blood Count, machine learning, SHAP