

INTISARI

INTEGRASI *ENSEMBLE LEARNING* DAN *EXPLAINABLE AI (XAI)* PADA DETEKSI KANKER PAYUDARA MENGGUNAKAN DATASET WDBC (*BREAST CANCER WISCONSIN DIAGNOSTIC*)

Oleh

HUMAIRAH

23/528786/PPA/06689

Diagnosis kanker payudara memerlukan sistem *machine learning* (ML) yang tidak hanya akurat, tetapi juga robust terhadap ketidaksempurnaan data klinis serta transparan bagi tenaga medis. Penelitian ini mengusulkan sebuah kerangka kerja yang berorientasi klinis dengan mengintegrasikan *ensemble learning*, analisis *robustness*, dan *Explainable Artificial Intelligence (XAI)* untuk prediksi kanker payudara menggunakan dataset Wisconsin Diagnostic Breast Cancer (WDBC). *Pipeline* yang dikembangkan terdiri atas modul *OutlierImputer*, penskalaan Min–Max, seleksi fitur menggunakan *SelectKBest*, *Sequential Feature Selection*, dan metode *embedded* berbasis *Random Forest*, serta empat model dasar, yaitu SVM, Random Forest, GBM, dan XGBoost, yang dikombinasikan melalui skema *soft voting* dan *stacking*. Dengan menggunakan skema pembagian data *stratified train test* 80/20, kombinasi model XGBoost dengan subset fitur SFS sebanyak 10 fitur (SFS-10) menghasilkan performa terbaik dengan akurasi sebesar 98,25%, presisi 100%, recall 95,24%, dan F1-score 97,56%, yang menunjukkan kemampuan *rule in* yang sangat kuat. Analisis *robustness* pada skenario *Missing Completely at Random* (MCAR) dan *Missing at Random* (MAR) hingga tingkat *missingness* sebesar 15% menunjukkan bahwa akurasi tetap berada di atas 95% dan presisi untuk kelas *malignant* tetap mendekati 100%, sementara nilai *recall* mengalami penurunan secara bertahap namun masih berada pada rentang 88–92%. Untuk memastikan keterpahaman klinis, tiga metode XAI lokal, yaitu LIME, SHAP, dan Anchors, divalidasi menggunakan instrumen yang diadaptasi dari CLIX-M. Hasil validasi menunjukkan bahwa dokter menilai SHAP dan LIME dalam kategori “sangat baik” dengan skor masing-masing sebesar 88% dan 85%, sedangkan Anchors memperoleh kategori “baik” dengan skor 77%, serta penjelasan yang dihasilkan dinilai relevan dan dapat diimplementasikan dalam konteks klinis. Temuan ini menunjukkan bahwa kerangka kerja *ensemble XAI* yang diusulkan merupakan fondasi yang menjanjikan bagi pengembangan sistem pendukung keputusan diagnosis kanker payudara yang transparan dan dapat diterima secara klinis.

Kata Kunci : Kanker payudara, Diagnosis berbantuan komputer, Diagnosis medis, Kecerdasan buatan yang dapat dijelaskan, Seleksi fitur.

ABSTRACT

INTEGRATION OF ENSEMBLE LEARNING AND EXPLAINABLE AI (XAI) TO BREAST CANCER PREDICTION USING THE WISCONSIN DIAGNOSTIC BREAST CANCER (WDBC) DATASET

By

HUMAIRAH

23/528786/PPA/06689

Breast cancer diagnosis requires machine learning (ML) systems that are not only accurate, but also robust to imperfections in clinical data and transparent to physicians. This research proposes a clinically oriented framework that integrates ensemble learning, robustness analysis, and explainable AI (XAI) for breast cancer prediction using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. The pipeline consists of an OutlierImputer, Min–Max scaling, feature selection (SelectKBest, Sequential Feature Selection, and embedded Random Forest), and four base models (SVM, Random Forest, GBM, XGBoost) combined with soft voting and stacking schemes. Under an 80/20 stratified train–test split, the combination of XGBoost with a 10-feature SFS subset (SFS-10) achieves optimal performance with an accuracy of 98.25%, precision of 100%, recall of 95.24%, and an F1-score of 97.56%, indicating a strong rule-in capability. Robustness analysis under Missing Completely at Random (MCAR) and Missing at Random (MAR) mechanisms up to 15% missingness shows that accuracy remains above 95% and precision for the malignant class remains close to 100%, while recall decreases gradually but stays in the 88–92% range. To ensure clinical interpretability, three local XAI methods (LIME, SHAP, and Anchors) are validated using an instrument adapted from CLIX-M. Physicians rate SHAP and LIME as “very good” (88% and 85%, respectively), while Anchors is rated “good” (77%), and consider the resulting explanations relevant and implementable. These findings indicate that the proposed ensemble–XAI framework is a promising foundation for a transparent and clinically acceptable decision support system for breast cancer diagnosis.

Keywords : Breast cancer, Computer aided diagnosis, Medical diagnosis, Explainable artificial intelligence, Feature selection.