ABSTRAK

Dodi Putra Artawan, I Wayan (2026), Framework Hybrid Feature Selection Berbasis PCA–Adaptive Correlation Threshold untuk Optimasi dan Generalisasi Klasifikasi Pada Dataset Breast Cancer Wisconsin

Tesis ini disetujui dan diperiksa oleh: Pembimbing I: Dr. I Gede Aris Gunadi, S.Si. M.Kom. dan Pembimbing II: Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs.

Kata Kunci**:** Machine Learning, Breast Cancer Wisconsin Dataset, hybrid feature selection, Principal Component Analysis (PCA), Adaptive Correlation Threshold (ACT).

Perkembangan machine learning di bidang medis menuntut sistem diagnosis yang cepat, akurat, dan objektif, khususnya untuk kanker payudara. Namun, sebagian besar studi sebelumnya masih menggunakan seleksi fitur statis yang kurang adaptif terhadap berbagai karakteristik data. Dataset Breast Cancer Wisconsin (Diagnostic) yang berdimensi tinggi (569 sampel, 30 fitur) memiliki korelasi antarfitur yang tinggi sehingga berisiko menimbulkan redundansi, inefisiensi komputasi, dan overfitting. Penelitian ini mengusulkan kerangka hybrid feature selection yang memadukan Principal Component Analysis (PCA) untuk reduksi dimensi dan Adaptive Correlation Threshold (ACT) untuk secara adaptif mengeliminasi fitur yang berkorelasi tinggi. Evaluasi dilakukan menggunakan Support Vector Machine, Random Forest, k-Nearest Neighbor, dan Logistic Regression dengan skema 10-fold cross-validation, serta metrik accuracy, precision, recall, F1-score, dan AUC-ROC. Hasil eksperimen menunjukkan bahwa framework hybrid PCA-ACT secara signifikan mereduksi fitur dari 30 fitur menjadi 8 komponen, menghasilkan tingkat reduksi dimensi sebesar 73,33% serta meningkatkan rasio kepadatan sampel-ke-fitur menjadi 56,88 : 1 (+275% dari baseline), dengan waktu pelatihan yang efisien sebesar 0,0151 detik dan waktu inferensi sebesar 0,0011 detik per 100 sampel. Pengembangan selanjutnya dapat diarahkan pada dataset yang lebih besar, arsitektur deep learning, serta integrasi Explainable AI untuk meningkatkan transparansi dan interpretabilitas klinis.

ABSTRACT

Dodi Putra Artawan, I Wayan (2026), A PCA–Adaptive Correlation Threshold-Based Hybrid Feature Selection Framework for Optimization and Generalization of Classification on the Breast Cancer Wisconsin Dataset

This thesis has been approved and examined by: Supervisor I: Dr. I Gede Aris Gunadi, S.Si., M.Kom. and Supervisor II: Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs.

Keywords: Machine Learning, Breast Cancer Wisconsin Dataset, hybrid feature selection, Principal Component Analysis (PCA), Adaptive Correlation Threshold (ACT).

Advancements in medical machine learning necessitate diagnostic systems that are fast, accurate, and objective, particularly for breast cancer. However, most prior studies have relied on static feature selection methods that lack adaptability to diverse data characteristics. The high-dimensional Breast Cancer Wisconsin (Diagnostic) dataset (569 samples, 30 features) exhibits strong inter-feature correlations, posing risks of redundancy, computational inefficiency, and overfitting. This study proposes a hybrid feature selection framework combining Principal Component Analysis (PCA) for dimensionality reduction with Adaptive Correlation Threshold (ACT) to adaptively eliminate highly correlated features. Evaluation was conducted using Support Vector Machines, Random Forests, k-Nearest Neighbors, and Logistic Regression within a 10-fold cross-validation scheme, utilizing accuracy, precision, recall, F1-score, and AUC-ROC as performance metrics. Experimental results demonstrate that the hybrid PCA-ACT framework significantly reduced the feature set from 30 features to 8 components—achieving a 73.33% reduction rate and improving the sample-to-feature density ratio to 56.88:1 (a 275% increase over the baseline)—while maintaining efficient training and inference times of 0.0151 seconds and 0.0011 seconds per 100 samples, respectively. Future work could focus on larger datasets, deep learning architectures, and Explainable AI integration to enhance clinical transparency and interpretability.