

ABSTRAK

Penyakit sirosis merupakan masalah kesehatan global yang signifikan, ditandai dengan kerusakan hati kronis yang menyebabkan fibrosis dan nodul, dengan gejala awal yang sering tidak spesifik sehingga menyulitkan deteksi dini. Keterlambatan diagnosis dapat memperburuk kondisi, menyebabkan komplikasi serius seperti kegagalan hati atau kanker hati. Penelitian ini bertujuan untuk membandingkan performa model machine learning, yaitu *Support Vector Machine*, *Random Forest*, dan *Multilayer Perceptron* dalam mengklasifikasikan stadium penyakit sirosis dengan penyetelan *hyperparameter* menggunakan GridSearchCV untuk meningkatkan akurasi deteksi dini.

Penelitian ini menggunakan data sekunder dari platform Kaggle yang berisi 13 variabel melalui pemilihan variabel seperti usia, *albumin*, *bilirubin*, *alkaline phosphatase*, *cholesterol*, SGOT, *prothrombin*, *platelets*, *spiders*, *hepatomegaly*, *ascites*, *edema* dan *stage*. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Tahapan *preprocessing* meliputi pemilihan fitur, *label encoding* untuk mengubah data kategorikal menjadi numerik, dan normalisasi menggunakan metode *MinMax*. Ketiga model dikembangkan dengan melakukan penyetelan *hyperparameter* melalui GridSearchCV untuk menemukan kombinasi parameter terbaik. Evaluasi model dilakukan menggunakan metrik akurasi, *precision*, *recall*, *F1-Score*, serta akurasi rata-rata *cross-validation* untuk memastikan performa model.

Hasil penelitian menunjukkan bahwa *Random Forest* menghasilkan performa terbaik dengan akurasi sebesar 95,84%, *precision* 95,87%, *recall* 95,84%, *F1-Score* 95,84%, serta akurasi rata-rata sebesar 95,52%. *Multilayer Perceptron* mencapai akurasi 90,30%, *precision* 90,37%, *recall* 90,30%, *F1-Score* 90,30%, dan akurasi rata-rata 90,2%. Sementara itu, *Support Vector Machine* menunjukkan performa paling rendah di antara ketiganya dengan akurasi sebesar 89,56%, *precision* 89,59%, *recall* 89,56%, *F1-Score* 89,57%, dan akurasi rata-rata 89,02%. Hasil ini menunjukkan bahwa *Random Forest* terbukti paling efektif dalam mengklasifikasikan stadium sirosis dengan tingkat kesalahan minimal.

Kata Kunci: Sirosis, Support Vector Machine, Random Forest, Multilayer Perceptron

ABSTRACT

Cirrhosis is a significant global health issue characterized by chronic liver damage that leads to fibrosis and nodules, with early symptoms often being nonspecific, making early detection difficult. Delayed diagnosis can worsen the condition, leading to serious complications such as liver failure or liver cancer. This study aims to compare the performance of machine learning models, namely Support Vector Machine, Random Forest, and Multilayer Perceptron, in classifying the stages of cirrhosis disease with hyperparameter tuning using GridSearchCV to improve early detection accuracy.

This research utilizes secondary data from the Kaggle platform, containing 13 variables through feature selection such as age, albumin, bilirubin, alkaline phosphatase, cholesterol, SGOT, prothrombin, platelets, spiders, hepatomegaly, ascites, edema, and stage. The data is divided into 80% for training and 20% for testing. Preprocessing steps include feature selection, label encoding to convert categorical data into numerical data, and normalization using the MinMax method. The three models are developed by performing hyperparameter tuning through GridSearchCV to find the best parameter combinations. Model evaluation is conducted using metrics such as accuracy, precision, recall, F1-Score, and average cross-validation accuracy to ensure model performance.

The results show that Random Forest achieved the best performance with an accuracy of 95.84%, precision of 95.87%, recall of 95.84%, F1-Score of 95.84%, and an average accuracy of 95.52%. Multilayer Perceptron reached an accuracy of 90.30%, precision of 90.37%, recall of 90.30%, F1-Score of 90.30%, and an average accuracy of 90.2%. Meanwhile, Support Vector Machine demonstrated the lowest performance among the three, with an accuracy of 89.56%, precision of 89.59%, recall of 89.56%, F1-Score of 89.57%, and an average accuracy of 89.02%. These results indicate that Random Forest is the most effective in classifying the stages of cirrhosis with minimal error rates.

Keyword: *Cirrhosis, Support Vector Machine, Random Forest, Multilayer Perceptron*