

RINGKASAN

ANALISIS PENGARUH SELEKSI FITUR HASIL PENGGABUNGAN FITUR PADA DETEKSI PRA-KANKER CITRA KOLPOSKOPI DENGAN MODEL SVM DAN XGBOOST

M. Saujana Shafi Kehaulani

Kanker serviks menjadi salah satu penyebab utama kematian akibat kanker pada wanita di seluruh dunia. Deteksi dini menjadi kunci utama dalam penanganan dan pencegahan kanker serviks. Kolposkopi merupakan metode inspeksi visual dalam deteksi dini kanker serviks. Seiring kemajuan teknologi, pengolahan citra dan *machine learning* menjadi solusi potensial untuk mengotomasi dan mempercepat diagnosis. Meski demikian, tantangan terbesar terletak pada pemilihan fitur yang relevan agar model *machine learning* dapat bekerja secara optimal.

Seleksi fitur menjadi salah satu kunci dalam pemilihan *subset* fitur yang optimal. Metode *filter* mengevaluasi fitur berdasarkan karakteristik statistik tanpa bergantung pada model tertentu, sehingga mengurangi risiko *overfitting*. Sebaliknya, metode *wrapper* memilih fitur berdasarkan performa model, memastikan bahwa fitur yang dipilih benar-benar mengoptimalkan akurasi prediksi. Kombinasi kedua metode ini bertujuan untuk mendapatkan *subset* fitur terbaik untuk menghasilkan model dengan performa optimal.

Support Vector Machine (SVM) dan *Extreme Gradient Boosting* (XGBoost) merupakan algoritma yang dikenal memiliki performa tinggi dalam tugas klasifikasi. SVM unggul dalam memisahkan data dengan pola *overfitting* melalui regulasi yang baik. Kombinasi kedua model ini memungkinkan evaluasi yang komprehensif terhadap pengaruh seleksi fitur pada berbagai karakteristik model *machine learning*.

Kata kunci : Kolposkopi, pengolahan citra, seleksi fitur, *machine learning*

SUMMARY

THE EFFECT OF FEATURE SELECTION AND FEATURE FUSION ON COLPOSCOPY IMAGES OF CERVICAL CANCER USING SVM AND XGBOOST METHOD

M. Saujana Shafi Kehaulani

Cervical cancer is one of the leading causes of cancer deaths in women worldwide. One of the keys in the treatment and prevention of cervical cancer is through early detection. Colposcopy is one method of early detection of cervical cancer through visual inspection. As technology develops, image processing and machine learning are a potential solution to automate and speed up diagnosis. However, the biggest challenge lies in ensuring the reliability of machine learning models in identifying lesion optimally.

Feature selection is one of the keys in selecting the optimal feature subset. Filter methods evaluate features based on statistical characteristics without relying on a particular machine learning model, thus reducing the risk of overfitting. In contrast, wrapper methods evaluate features based on model performance, ensuring that the selected features actually contribute to optimize prediction accuracy. The combination of these two methods aims to obtain the best subset of features to produce a model with optimal performance.

Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost) are algorithms known for their high performance in classification tasks. SVM excels at separating data with complex patterns using kernel tricks. XGBoost on the other hand, effectively handles large datasets and prevent overfitting through fine regulation. The combination of these two models enables a comprehensive evaluation of the effect of feature selection on various characteristics of machine learning models.

Keywords : Colposcopy, image processing, feature selection, machine learning