

# Optimalisasi Parameter pada Preprocessing dan Klasifikasi Spektra Raman Untuk Spesimen Jaringan Kanker Prostat Menggunakan Algoritma XGBoost = Optimization of Parameters in The Preprocessing and Classification of Raman Spectra for Prostate Cancer Tissue Specimens Using The Xgboost Algorithm

Ahmad Rafianto, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920540773&lokasi=lokal>

---

## Abstrak

Pada tahun 2018, kanker prostat merupakan penyakit ganas kedua terbanyak pada pria secara global setelah kanker paru-paru. IDC-P merupakan varian agresif dari kanker prostat yang sering sering disalahinterpretasikan dengan proliferasi intraduktal seperti HGPIN yang dapat mempengaruhi perawatan pasien laki-laki pengidap kanker prostat. Teknik analisis spektroskopi Raman merupakan teknik molekuler berprospek untuk menganalisa jaringan biologis yang telah banyak digunakan dalam mencoba untuk mengidentifikasi berbagai macam jenis kanker. Sayangnya, spektroskopi Raman menghasilkan sinyal yang lemah dan mudah dipengaruhi oleh noise dan latar belakang floresens. Penelitian ini bertujuan untuk mengembangkan pipeline yang mencakup tahapan prapengolahan dan klasifikasi terhadap spektra Raman dari spesies IDC-P dan HGPIN untuk mendapatkan hasil metrik evaluasi yang optimal. Tujuan ini dicapai dengan menemukan nilai parameter optimal pada tahapan prapengolahan (smoothing, baseline correction, normalisasi), dan klasifikasi untuk menghasilkan hasil klasifikasi yang terbaik. Dihasilkan sebuah pipeline yang mencakup tahapan prapengolahan dan klasifikasi dengan kemampuan untuk menghasilkan hasil evaluasi metrik yang tinggi untuk metrik evaluasi F1 Test, accuracy, dan AUC-ROC masing-masing bernilai : 98.8%, 97.9%, dan 98.8%. Berdasarkan hasil analisis ANOVA, ditemukan bahwa perbedaan pada parameter window length dan polynomial order pada tahapan Savitzky-Golay smoothing tidak memiliki signifikansi terhadap hasil evaluasi metrik ( $p > 0.05$ ). Sebaliknya, metode baseline correction beserta nilai polynomial degree yang berbeda cenderung memberikan signifikansi ke hasil evaluasi metrik ( $p < 0.05$ ). .....

In 2018, prostate cancer was the second most common malignant disease in men globally, following lung cancer. IDC-P is an aggressive variant of prostate cancer often misinterpreted as intraductal proliferation like HGPIN, which can impact the treatment of male patients with prostate cancer. Raman spectroscopy analysis is a promising molecular technique for analyzing biological tissues and has been extensively used in attempts to identify various types of cancer. Unfortunately, Raman spectroscopy produces weak signals that are easily influenced by noise and fluorescent backgrounds. This research aims to develop a pipeline that includes preprocessing and classification stages for Raman spectra of IDC-P and HGPIN species to achieve optimal evaluation metric results. This goal is achieved by finding the optimal parameter values in preprocessing stages (smoothing, baseline correction, normalization) and classification to produce the best classification results. A pipeline was created that includes preprocessing and classification stages capable of producing high evaluation metric results for the F1 Test, accuracy, and AUC-ROC metrics, respectively valued at 98.8%, 97.9%, and 98.8%. Based on ANOVA analysis, it was found that differences in the 'window length' and 'polynomial order' parameters in the Savitzky-Golay smoothing stage do not significantly affect the evaluation metric results ( $p > 0.05$ ). Conversely, the baseline correction method and different 'polynomial degree' values tend to significantly impact the evaluation

metric results ( $p < 0.05$ ).