

KLASIFIKASI PENENTUAN JENIS KANKER PAYUDARA DAN PARU-PARU BERDASARKAN SUSUNAN PROTEIN DENGAN ALGORITMA *DECISION TREE*

ABSTRAK

Angka kematian di Indonesia karena penyakit kanker menurut WHO pada tahun 2014 mencapai 195.300 orang dengan kematian terbanyak pada laki-laki sebanyak 103.100 dengan penyakit kanker paru-paru pada tingkat pertama sebesar 21,8% dan pada wanita sebanyak 92.200 dengan penyakit kanker payudara pada tingkat pertama sebesar 21,4%. Ilmu informatika dapat membantu menganalisa masalah ini dengan menggunakan cabang ilmu bioinformatika, yang merupakan penerapan teknis komputasional untuk mengolah dan menganalisa informasi biologi seperti data DNA, RNA dan Protein.

Penelitian ini membahas tentang pendeteksian jenis kanker payudara dan paru-paru berdasarkan susunan protein berupa asam amino. Data protein yang digunakan berformat *.fasta* yang diambil dari database protein yaitu *UniProt* dan NCBI (*National Center for Biotechnology Information*). Metode klasifikasi digunakan untuk mendeteksi secara dini penyakit kanker berdasarkan susunan protein. Algoritma yang digunakan pada penelitian ini adalah *decision tree*, yaitu salah satu algoritma yang digunakan dalam bidang bioinformatika.

Sebelum tahap klasifikasi dilakukan tahap pre-processing data sekuen protein bertipe *String* ditransformasikan terlebih dahulu menggunakan EIIP (*Electron-Ion Interaction Potential*) based protein value. Data sekuen protein yang telah ditransformasikan menjadi numerik selanjutnya diekstraksi ciri menjadi sinyal *frequency based* dengan menggunakan FFT (*Fast Fourier Transform*) dan LPC (*Linear Prediction Coding*) dan turunan dari LPC dengan order 8 dan 12.

Klasifikasi dengan *decision tree* dilakukan setelah proses ekstraksi ciri selesai. Pada penelitian ini dilakukan pengujian kombinasi *feature* LPC pada saat proses ekstraksi ciri. Dari seluruh pengujian (16 pengujian) yang dilakukan, didapatkan hasil akurasi terbesar yaitu 79,85% dan waktu yang diperlukan untuk membuat tree 29,09 detik dengan akurasi rata-rata sebesar 74,82% pada *feature* LPC order 8.

**THE CLASSIFICATION OF BREAST CANCER AND LUNG CANCER
TYPE BASED ON PROTEIN STRUCTURE
USING *DECISION TREE* ALGORITHM**

ABSTRACT

The death rate from cancer in Indonesia by WHO in 2014 reached 195.300 people with the number of death in men at 103.100 from lung cancer at the first level by 21.8% and in women at 92.200 from breast cancer at the first level by 21.4%. Science of Informatics can help analyze this problem by using a branch of bioinformatics, which is the computational technique application to process and analyze the data of biological information such as DNA, RNA and Protein.

This research discusses the detection of breast cancer and lung cancer based on the composition of proteins in the form of amino acids. The protein data used the format of *.fasta* derived from the protein database *UniProt* and NCBI (*National Center for Biotechnology Information*). The classification method that used for early detection of cancer based on protein structure. The algorithm used in this study is a *decision tree*, which is one of the algorithms used in bioinformatics.

Before classification phase, pre-processing phase of protein sequence data of *String* type are transformed beforehand using EIIP (*Electron-Ion Interaction Potential*) based on protein value conducted. Protein sequence data that have been transformed into numeric, then the feature extracted to become frequency based signal by using FFT (*Fast Fourier Transform*) and LPC (*Linear Prediction Coding*) and a derivative of LPC with order of 8 and 12.

Classification by *decision tree* is happened after the feature extraction process is complete. In this research, testing the combination of LPC feature during the process of feature extraction conducted. From all the tests (16 tests) were conducted, showed the greatest accuracy is 79.85% and time to build tree 29.09 second with an average accuracy of 74.82% in order 8 of LPC feature.