

## ABSTRAK

Ilmu bioinformatika menggunakan data-data genetis seperti DNA dan asam amino dalam penerapannya. Data-data genetis tersebut dapat digunakan untuk mendapatkan berbagai macam informasi dalam hal pendeteksian penyakit seperti penyakit kanker. Tugas akhir ini membahas tentang pendeteksian jenis kanker dengan menggunakan susunan protein asam amino. Data protein asam amino yang digunakan dalam bentuk sekuen dengan format *fasta file* yang diambil dari database protein UniProt dan NCBI (*National Center for Biotechnology Information*). Penggunaan metode klasifikasi terhadap susunan protein asam amino tersebut dapat membantu deteksi dini penyakit kanker. Metode klasifikasi yang digunakan adalah *Support Vector Machine* (SVM). Metode SVM ini merupakan salah satu metode klasifikasi yang sering digunakan dalam bidang ilmu bioinformatika.

Data sekuen protein asam amino akan diekstraksi cirinya terlebih dahulu untuk kemudian diklasifikasi. Proses ekstraksi ciri dilakukan menggunakan metode FFT (Fast Fourier Transform) dan LPC (Linear Predictive Coding). Pertama-tama data sekuen protein asam amino yang sebelumnya berupa *String* diubah menjadi numerik dengan memberikan nilai EIIP pada sekuen tersebut. Data sekuen yang sudah berbentuk numerik selanjutnya diubah menjadi sinyal dan diekstraksi cirinya menggunakan metode FFT dan LPC. Klasifikasi SVM dilakukan setelah mendapatkan ciri dari data protein asam amino. Pada penelitian ini dilakukan banyak pengujian. Pengujian dilakukan dengan membuat variasi metode pada saat proses ekstraksi ciri dan penggunaan kernel yang terdapat pada metode SVM. Dari seluruh pengujian yang dilakukan, didapatkan hasil akurasi terbesar yaitu 85,03%.

## ABSTRACT

Bioinformatics uses generally data in the implementation like DNA and protein amino acid. The genetically data can be used to get a lot of information in illness detection like cancer. This research tells about of cancer type by using the structure of protein amino acid. The data of protein amino acid is in the form of sequence with fasta file format that can be got from UniProt protein database and NCBI (National Center for Biotechnology Information). The use of classification method to this structure of protein amino acid can help early cancer detection. The classification method used is Support Vector Machine (SVM). This method is one of classification method that is often used in bioinformatics.

The sequence data of protein amino acid will be firstly extracted its feature then classified them. The process of feature extraction done uses FFT and LPC method. The previous sequence data of protein amino acid is String, then it will change into numerical data by giving EIIP value to its sequence. The numerical sequence data will be changed into signal and extracted the feature using FFT and LPC. The next SVM classification is done after getting the feature of protein amino acid data. In this research, it has done many test. The test is done by using variation method when having feature extraction process and kernel use which is in the SVM method. From all test that can be done is got for the biggest accuracy result is 85,03%.