Implementation of Particle Swarm Optimization - SVM on Gene Expression Data for Lung Cancer Identification in Smoker Person

Muhammad Aufa Anasin, Fhira Nhita, Isman Kurniawan,

School of Computing, Telkom University, Bandung

aufaanasin@student.telkomuniversity.ac.id, fhiranhita@telkomuniversity.ac.id, ismankrn@telkomuniversity.ac.id

Abstract

The risk of lung cancer at ten years is quite variative among the participant in the Computed Tomography (CT) study cases. From 15% of 68 years old that smokes two packs per day and continuously smoke, to 0.8% for 51 years-old women who smoke one pack per day for 28 years before quited for smoking nine years before tobacco smoking is most dangerous with 80%-90% of the cigarette smokers. The symptoms can be continuous if the disease is uncontrolled with therapy. However, it will get another sign [3]. The alternative method for cancer treatment is chemotherapy. Unfortunately, it may have a negative effect on the patient. Identification of cancer can use machine learning since it has an automatic algorithm—the study used for microarray data and SVM from machine learning. A microarray dataset allows an examination of levels of gene expression in specific cell samples. For the feature selection, we choose Particle Swarm Optimization. The result gathered from Gene Expression Data with Particle Swarm Optimization – Support Vector Machine validated using accuracy, precision, recall, and F1-score. Best validation comes from RBF model that contains 520 features with Accuracy, recall, and F1-score 1.00, 1.00, and 1.00.

Keywords: Lung cancer, SVM, Microarray, Particle Swarm Optimization Algorithm, Cancer Detection