Early Diagnosis Of Breast Cancer Using Data Mining And Machine Learning Methods

Esra Sogut[a*], O. Ayhan Erdem[a]

[a]University Of Gazi

*esrasogut@gazi.edu.tr

Abstract:

Breast cancer is the leading cancer type among women. Breast cancer, which has a very high incidence in the world, can also be seen in men. It is a cancer that develops from the cells of the breast tissue and can originate from any part of the tissue. Every year thousands of patients lose their lives from breast cancer.

As in other types of cancer, early detection is also considered the best preventive method in breast cancer. If cancer is detected early before it spreads, the patient's chances of survival can be significantly increased. Patient biopsy is done for early diagnosis and definite diagnosis in breast cancer. Biopsy; It is a process to recognize and certify whether a suspicious area or tissue on the breast is cancerous and is also used to diagnose various diseases.

In this study, data mining and machine learning methods were used for breast cancer detection. With these methods, it is aimed to be able to diagnose disease according to biopsy results. Breast-Cancer-Wisconsin data provided by the UCI Machine Learning Repository was used in the study. The dataset donated by Olvi Mangasarian includes 699 patient information and contains 9 features and 2 class features. The data set was trained using the C4.5 decision tree algorithm with an accuracy of 97.4955%. This training resulted in a classification of 98.5714% accuracy. The resulting data show that the values indicated are a crucial guide in the diagnosis of breast cancer.

Keywords: Breast Cancer, Data Mining, Machine Learning, Dataset