

Analysis of Breast Cancer Disease Classification (C50) Using the K-Nearest Neighbor (KNN) Method at Baladhika Husada Jember Level III Hospital.

Mudafiq Riyan Pratama, S.Kom., M.Kom. (Pembimbing)

Najwa Gina Safitri

Health Information Management Study Program

Department of Health

ABSTRACT

Breast cancer is a disease with high incidence and mortality rates in Indonesia. Utilizing medical record data through data mining methods can assist in the classification of breast cancer stages. This study aimed to classify breast cancer stages (C50) using the K-Nearest Neighbor (KNN) method at Baladhika Husada Level III Hospital, Jember. Research variables included age, family history, history of prior breast disease, body mass index (BMI), breast pain, breast lumps, axillary lumps, changes in breast shape and size, nipple discharge, nipple changes, breast skin abnormalities, local status, and tumor size. The results showed that the best model was achieved with $K=5$ and a 90:10 data split ratio, yielding an accuracy of 87.50%, precision of 86.36%, and recall of 86.36%. Characteristic analysis revealed that patients with advanced-stage cancer were more likely to be aged ≥ 50 years; have a family history and history of prior breast disease; experience breast pain, axillary lumps, changes in breast shape and size, nipple discharge, nipple changes, and breast skin abnormalities; present with bilateral local status; and have a tumor size > 5 cm. Feature ablation results indicated that breast skin abnormalities was the variable contributing most significantly to model performance. Based on these findings, the K-Nearest Neighbor (KNN) method is capable of classifying breast cancer stages with good performance. Furthermore, patients with advanced-stage cancer tend to exhibit more complex clinical characteristics compared to those with early-stage cancer.

Keywords: Breast cancer, classification, K-Nearest Neighbor, data mining.