ABSTRAC

Ovarian cancer is one of the gynaecological cancers with a high mortality rate due to the difficulty of early detection, mainly due to non-specific and often asymptomatic early symptoms. Conventional detection methods such as ultrasonography and CA-125 biomarker test require high cost and access to adequate health facilities, making it difficult to reach especially in areas with limited medical resources. This study aims to develop an ovarian cancer pre-screening system based on Gaussian Naive Bayes (GNB) algorithm as the main method, by comparing its performance against CatBoost and Multi-Layer Perceptron (MLP) algorithms as benchmarks.

The research methodology includes clinical data acquisition in the form of ovarian cancer symptom questionnaires collected from medical personnel at Hasan Sadikin Hospital Bandung with a total of 240 patient data samples. The system was developed using data preprocessing with StandardScaler normalisation technique and dataset balancing using SMOTE. GNB algorithm was chosen as the baseline due to its ability to handle categorical data and independence assumption that is suitable for medical screening. The model was tested using accuracy, sensitivity, specificity, F1-score, and AUC-ROC metrics. The evaluation results showed GNB achieved 57.75% accuracy with 14.29% sensitivity and 100% specificity. In comparison, CatBoost achieved the best performance with 85.92% accuracy, 97.14% sensitivity, and 75% specificity, while MLP achieved 76.6% accuracy. The system was successfully integrated in a widely accessible Android application, providing recommendations based on the results of analysing the symptoms entered by users.

Keywords: ovarian cancer, Gaussian Naive Bayes, CatBoost, Multi-Layer Perceptron, pre-screening, Android application, early detection, machine learning, classification system.