

## ABSTRACT

Author: Koby Bond  
Title: Medical Data Classification: A Machine Learning Approach  
Thesis Advisor: Alaa Sheta  
Department: Department of Computer Science  
Year: 2021

Data classification is a major problem in the data mining field. It can be defined as the process of splitting (i.e., categorizing) data into appropriate groups (i.e., classes) based on their common characteristics. The challenge of classifying medical data is also a significant data mining problem explored in various real-world applications by numerous researchers. In this research work, we compared different data classification techniques and their predictive accuracy on several datasets. The classification techniques include the Support Vector Machine, Artificial Neural Networks, and Decision Trees. The classifiers were implemented on the Weka data mining tool and were compared using accuracy measures such as Precision, Recall, and F-Measure. The performance of the classifiers was analyzed to choose the best algorithm based on the conditions of the datasets which were chosen from the UCI Machine Learning Repository. These datasets include Pima Indians Diabetes, Heart Failure Clinical Records, and the Breast Cancer Coimbra datasets. The results of the experiment show that the Artificial Neural Network-based classifier was the most accurate classification model in all cases. For the Pima Indians dataset, the classifier indicated a 75.13% accuracy rate. For the Breast Cancer dataset, the classifier indicated a 70.69% accuracy rate. Lastly, for the Heart Failure dataset, the classifier indicated a 76.92% accuracy rate, respectfully when classifying instances.