"Medical-Symptom-Based Intelligent Diseases Classification and Recognition"

By: Wesam Adel Al-Mhani
ID: 201525038

Supervisor:
Dr. Ahmad Azzazi

# Abstract

Diagnosis is difficult, but it is critical in medicine and disease diagnosis, and has important implications for patient care. In the medical field, the size and complexity of patient data collected and stored, which include data obtained from clinical studies or additional patient records, radiographs and test results, is increasingly extensive and contains hidden information that needs to be processed and extracted.

In the past decade, data mining has become an important tool for processing data and extracting knowledge, as well as for discovering patterns to generate important information that helps to make decisions in the areas where they were used.

Data mining is also an important technique for research in many fields. It extracts useful patterns from a large database that contains information that may be hidden and difficult to obtain. It also plays a major role in health care and in predicting the disease.

Data mining is primarily used to predict disease, so many techniques have been developed and used to predict disease, including data processing, classification, assembly, composition, and sequencing patterns. The provision of treatment schemes that are compatible with a disease according to different types of symptoms is important, while most classification methods may be the ineffective inaccurate classification of diseases that need multiple processing stages.

This research analyses the performance of classification techniques; And propose two techniques to test the accuracy, precision, recall and the time taken to build model for three diseases which are heart disease, hepatitis and diabetes, the research have receivedpatient and sample tests from the Computer Learning Depot Center of the University of California (UCI). Then compared the results.

The first technique was made by ANN (Artificial Neural Network) algorithm, and the second by the same algorithm but after optimization using GA (Genetic Algorithm).

Two types of tools have been used, the first one is Weka, and all the measurements were taken after and before optimization by the same tool. To improve the data set one by one through the genetic algorithm (GA), this research used the second tool, Rapid miner.

After obtaining the new Dataset after optimization, this research was presented again onthe Weka and checked the measurements again.

This research has done a test for more than ten times on the heart disease dataset, then chose the best result by changing parameters each time separately and the best results were taking. Then the best test parameter was applied to other diseases. Finally, this research made a comparison with all results, as well as another comparison with previousstudies was made.