

## ABSTRACT

In the field of data mining, due to the burgeoning development of computer and database technologies, the datasets are generated with more than thousands of features (variables/attributes). The presence of large number of features makes it a complex task to perform classification. Therefore, it is important to reduce the number of features and determine the most relevant features. Feature Selection is an important preprocessing technique in data mining and it is the process of selecting the relevant features from the datasets. The objectives of the feature selection technique are to reduce the number of features and to improve the classification accuracy. The selection of relevant features and the improvement of classification accuracy is a challenging task of the data mining applications. The main aim of this research work is to select the relevant features and to enhance the classification accuracy.

Three contributions have been achieved in the research. They are sequential backward selection algorithm for generating neighbourhood generation, Simulated Annealing algorithm for selecting the feature subset and Relief algorithm for ranking the attributes. Each one is different from others based on the objective. But all the three algorithms are assessed based on the accuracy of the classification. These three approaches are combined and a new novel algorithm is proposed known as simulated Relief. The proposed algorithm is implemented in the cotton data set. The efficiency and effectiveness of proposed algorithm is evaluated with Cotton Disease data set provided by Central Cotton Research Station at Coimbatore. WEKA tool and Microsoft excel contribute the data manipulation task for the research process. The research is concluded by its finding that the classification accuracy of the proposed simulated Relief algorithm is higher than the Relief algorithm.