

ABSTRACT:

The advances in the scientific community have provided us innumerable instruments for a better safe and sound tomorrow. We owe a lot to the scientific community whose persistent efforts have given us vaccinations and preventions against many unknown diseases of the days of yore. The discovery of the genome, the DNA model by Watson and Crick has been an important landmark in the history of biological sciences. Further Computer Science revolutionized bioinformatics with data mining entering its arena. The various clustering algorithms put up an end to the curiosity in the minds of the scientists about the specific genes responsible for certain diseases. It has indeed provided a solution to outsourcing the various diseases with proper diagnosis of the deadly diseases. In this project we have put up a sincere effort to study some of the clustering algorithms and their strengths and weaknesses.

In this work, we propose two new algorithms for efficient clustering of gene expression data. The first algorithm, namely a link-based clustering algorithm, uses the concept of common neighbors. The second algorithm extends the former proposed algorithm by using a fuzzy approach for detecting intersecting and overlapping clusters. We have also compared both the algorithms to the existing popular approaches and found our algorithms to give good results in terms of z-score measure of cluster validity and p-value measure.