The increased assimilation of computing into day to day life has lead to the creation of massive datasets in most realms of science and engineering. The pace of data production is further expected to accelerate in the next few years. The volume of data is expanding in two folds: number of records and number of dimensions. Handling high dimensional data is a big challenge for research communities. The challenges include “curse of dimensionality” and inherent sparsity in data space. Clustering is one of the most prominent data mining techniques for analyzing large data. However, performing clustering in high dimensional data is a NP hard problem. Nature inspired algorithm are promising algorithms in providing near optimal solution to NP hard problems. Their stochastic nature improves clustering by recuperating from poor solution initialization and avoiding local optima.

This thesis presents the novel variants of nature inspired algorithms for handling the challenges of partition based and high dimensional clustering. In first approach, flower pollination algorithm (FPA) is integrated with K-Means algorithm (KFPA) to improve clustering in simple problems with small dimension. A variant of KFPA that is modified FPA (MFPA) is developed to enhance the clustering efficiency on small dimensional complex clustering datasets. To further substantiate the observations, MFPA is evaluated on CEC2014 complex numerical benchmark problems against five existing algorithms upto 100 dimensions. However, MFPA did not perform well on these problems. Thus, with a motive of finding the best algorithm, performance analysis of five well known nature inspired algorithms is made and based on it, a new hybrid nature inspired algorithm is developed which is termed as ABC_DE_FP. It is a hybrid of artificial bee colony (ABC), FPA and differential evolution (DE) algorithm. The hybrid algorithm outperformed its contemporary counterparts on high dimensional complex problems. In the next approach binary version of ABC_DE_FP is integrated with subspace clustering (S_FAD) to cluster high dimensional data. S_FAD is successful in finding overlapped subspace clusters of varied densities up to 6400 dimensional real dataset. S_FAD is more appropriate for high dimensional applications where accuracy is the main concern.