## Feature Selection Using Particle Swarm Optimization for Optimal Email Spam Classification

N.A.S. Vinoth, A. Rajesh

Vels Institute of Science, Technology and Advanced Studies.

\*Corresponding author

## Abstract

**Background:** Spam, or unwanted commercial bulk emails, has become a major problem on the internet in recent years. The spammer is the one who sends the spam mails. This person collects email addresses from a variety of sources, including websites, chat groups, and viruses. Spam prohibits users from getting the most out of their time, storage space, and network bandwidth. Spam mail has a negative impact on email servers' memory space, communication bandwidth, CPU power, and user time due to the large volume of spam mails streaming across computer networks.

**Objectives:** One of the most critical process for the success of classification, data mining, or machine learning applications is feature subset selection. The primary goal of feature subset selection is to decrease the problem's dimensionality while maintaining the most discriminatory data required for accurate classification.

**Methodology**: Particle Swarm Optimization, is a popular and effective global search strategy. The primary space is the search space in which PSO was used to investigate and pick a subset of principle components or principal characteristics. As a result, each particle represents a different subset of the primary components. The particles swarm is randomly initialised, then updated in the search space or primary space to find the best subset of characteristics.

**Results and discussion:** This section discusses the results of proposed method of feature selection with the raw datasets having 38 features in which 8 features where selected. This method provided a smaller subset of features by which it can able to improve the performance of the classifier and minimize its architecture's complexity.

**Conclusions and future work:** The majority of researchers did not consider the computational cost while deciding which algorithm to utilise for feature Selection. Their main focus is on accuracy performance. In future, the computational cost should also be considered to get betterment in this work.

