

# Ovarian Cancer Prediction using Bacterial Foraging Optimization

Narmada.G.S<sup>1</sup>, Pavithra.C.J<sup>1</sup>, Prema.P<sup>1</sup>, T. Guhan<sup>2</sup>

<sup>1</sup>Final Year CSE, <sup>2</sup>Asst. Prof / CSE  
Sri Ramakrishna Engineering College, Coimbatore, India.

---

**Abstract:** Ovarian cancer is a cancer that forms in or on an ovary. It results in abnormal cells that have the ability to invade or spread to other parts of the body. Earlier diagnosis of Ovarian Cancer saves enormous lives, failing which may lead to other severe problems causing sudden fatal end. Its cure rate and prediction depends mainly on the early detection and diagnosis of the disease. The ovarian cancer can be detected in the early stage by the statistical approach to interpret the change in levels in women's blood of the protein CA125, which is linked to ovarian cancer. Now the computer aided method combined with the medical field to provide the better treatment for the cancer patient by using data mining techniques. This is mainly used to predict the levels of the ovarian cancer. This can be predicted with the help of dataset which contains the gene names which affect the ovary and the sample tested values. Thus the aim of the paper is to provide the optimal solution to predict the ovarian cancer.

**Keywords:** ovarian cancer, optimal feature selection, predicting levels.

---

## 1. INTRODUCTION

Women who have a strong family history of breast or ovarian cancer are more likely to be affected by cancer of the ovaries, fallopian tube, or peritoneal cavity. This is thought to be due to a mutation in one of the genes that are involved in the regulation of cell growth and replication in these areas, which can be inherited from the parents. It is estimated the 10-15% of ovarian, fallopian tube, or peritoneal cancers are associated with an inherited genetic mutation. The remaining majority of cases of cancer are linked to a genetic mutation that is acquired by the individual in their lifetime.

### Genes associated with ovarian cancer:

The Breast Cancer 1 (BRCA1) and Breast Cancer 2 (BRCA2) genes have been identified as genes that are linked to an increased risk of the development of both breast cancer and ovarian cancer. Everybody possesses these genes in their body because they play an important role in the regulation of cell growth in the breasts and ovaries, but a mutation in one or both of these genes increases the likelihood that an individual will be affected by breast or ovarian cancer. A woman with a mutation in the BRCA2 gene has a lifetime risk of 10-20% of developing ovarian cancer. This is approximately ten times higher than the risk of an average woman, which is 1-2%.

Some of the genes that have been linked to an increased risk of ovarian cancer include:

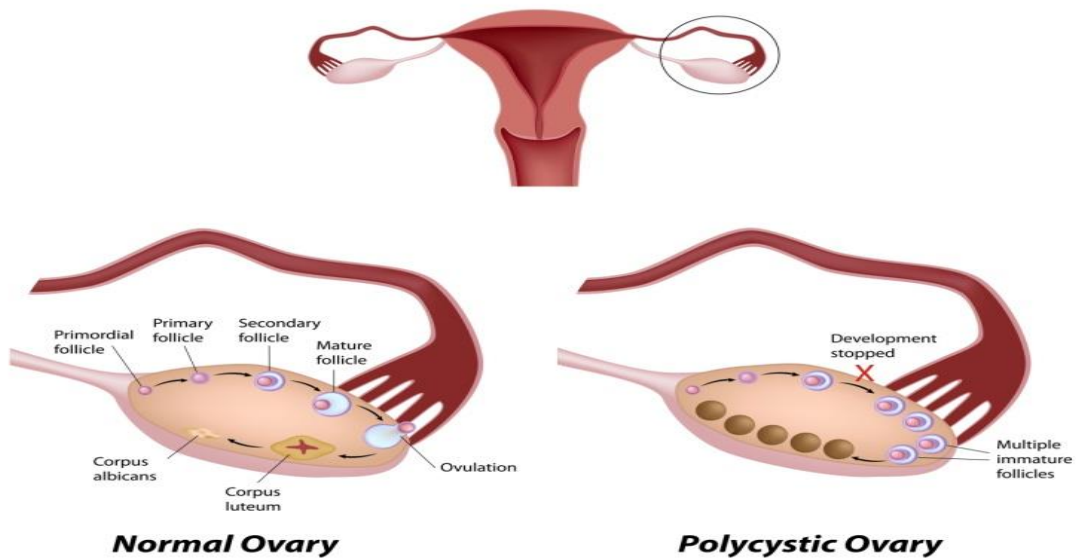
- CDH1: mutation is linked to a raised risk of ovarian and breast cancer.
- MLH1 gene: mutation is linked to a raised risk of both Lynch syndrome and ovarian cancer.
- MLH2 gene: mutation is linked to a raised risk of both Lynch syndrome and ovarian cancer.
- PALB2 gene

- PTEN gene: mutation is linked to a raised risk of Cowden syndrome and ovarian cancer.
- STK11 gene: mutation is linked to a raised risk of Peutz-Jeghers syndrome and ovarian cancer.
- TP53 gene: somatic mutation is present in almost half of all cases of ovarian cancer.

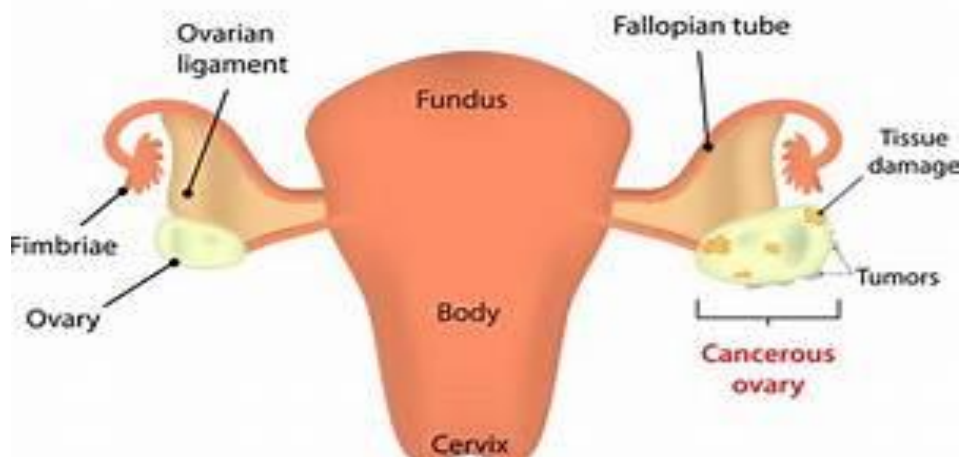
## 2. LITERATURE SURVEY

Improvements in cancer detection and treatment have led to consistent declines in mortality from many cancers. However, many patients present for treatment at a point where more invasive treatment is required and treatment outcomes are less than optimal. One factor that has been consistently shown to be associated with late diagnosis and treatment is delay in seeking help for symptoms. This paper reviews the literature on women’s awareness of cancer symptoms and aims to identify knowledge gaps that need to be addressed in order to improve help-seeking behaviors. The discovery of substantial gaps in awareness suggest a need for improved community education regarding cancer symptoms.

## 3. NORMAL OVARY AND AFFECTED OVARY



## Ovarian Cancer



#### A. Abbreviations and Acronyms

BFO - Bacterial Foraging Optimization.

KNN - k- Nearest Neighbor.

BRCA1 , BRCA2 - Breast Cancer gene.

#### *Existing Idea*

Ovarian cancer is a most lethal gynecological cancer which is increasing day by day in developing countries. Therefore, identification of genetic factors including mutations in BRCA1 and BRCA2 gene. The algorithm used for the prediction of ovarian cancer are: K-means clustering algorithm for identifying the relevant data, Significant Frequent Pattern were discovered using Apriori tid and Decision tree.

The drawbacks of this system are:

- Predicting accuracy is to a small extent in the existing system.
- Stages of cancer was not predicted in this system.

#### **Proposed System**

More than 8 out of 10 (80%) ovarian cancers occur in women above the age of 40. Cancer may occur in any part of the body and may spread to several other parts.

Only early detection of cancer at the benign stage and prevention from spreading to other parts in malignant stage could save a person's life.

In the proposed system the algorithms used are,

- Clustering can be done with the dataset by using k means algorithm
- Classify the generated data from the clustered data using k nearest neighbor.
- Predicting the optimal feature selection for the classified data using bacterial foraging optimization.

#### **Dataset Description:**

The dataset consists of gene names and sample tested values.

Gene name: includes the different gene names which mainly causes ovarian cancer.

Samples tested1: which has sample minimum test values.

Samples tested2: which has sample maximum test values.

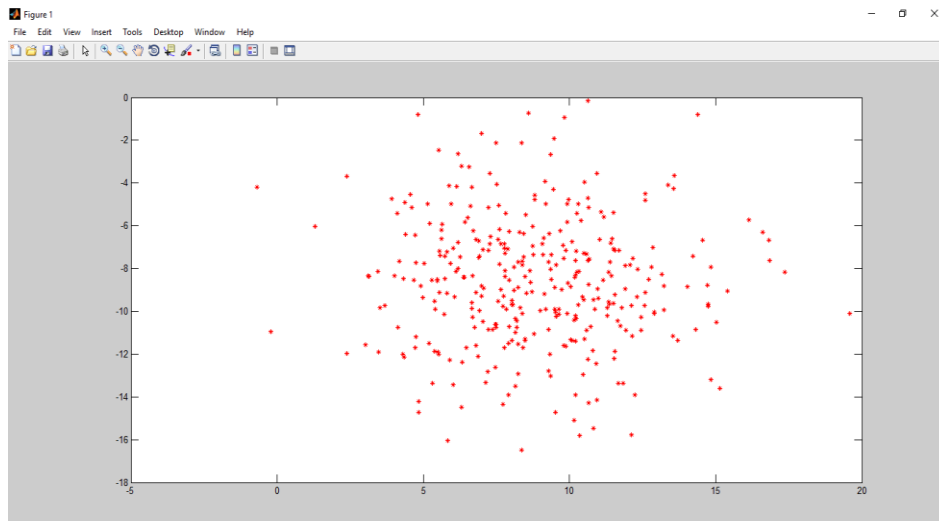
## **4. CLUSTERING USING KMEANS**

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into different cells. The algorithm works iteratively to assign each data point to one of  $K$  groups based on the features that are provided. Data points are clustered based on feature similarity. The results of the  $K$ -means clustering algorithm are:

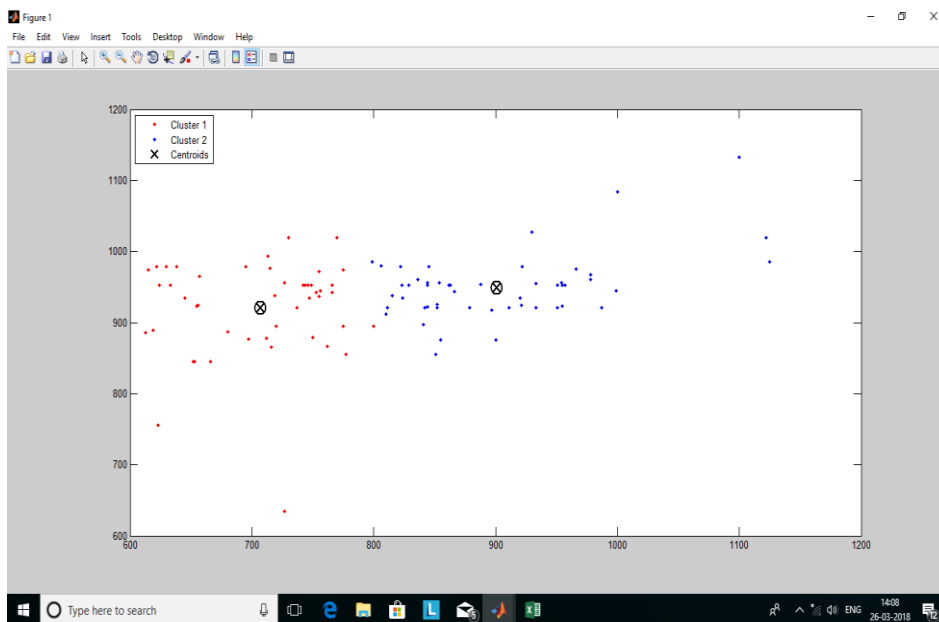
#### **K-Means Clustering Algorithm:**

K-MEANS clustering aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into different cells. In cluster analysis, the k-means algorithm can be used to partition the input data set into  $k$  partitions (clusters). However, the pure k-means algorithm is not very flexible, and as such is of limited use. In particular, the parameter  $k$  is known to be hard to choose when external constraints are not provided. Another limitation of the algorithm is that it cannot be used with arbitrary distance functions or on non-numerical data.

The randomly generated cluster is formed by using the dataset



The clusters thus formed by using k-means clustering



Clustering variable: Samples tested 1, samples tested 2.

## 5. CLASSIFICATION

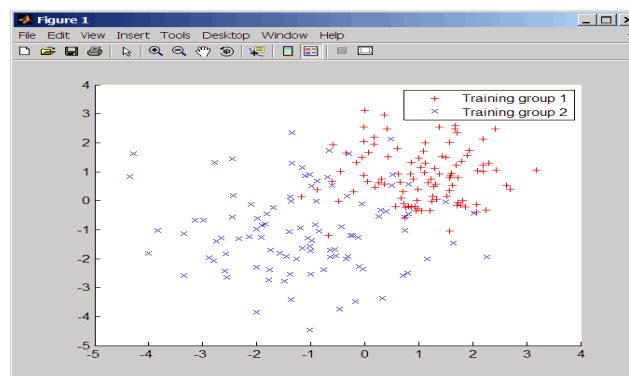
Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. The simplest type of classification problem is binary classification. In binary classification, the target attribute has only two possible values. Multiclass targets have more than two values. Classification models are tested by comparing the predicted values to known target values in a set of test data. The historical data for a classification project is typically divided into two data sets: one for building the model; the other for testing the model. Classification is used to find out in which group each data instance is related within a given dataset. It is used for classifying data into different classes according to some constraints. Several major kinds of classification algorithms including C4.5, ID3, k-nearest neighbor classifier, Naive Bayes, SVM, and ANN are used for classification. Generally a classification technique follows three approaches Statistical, Machine Learning and Neural Network for classification. The classification algorithm used in this project is k-nearest neighbor.

**K-NEAREST NEIGHBOR CLASSIFIER:**

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If  $K = 1$ , then the case is simply assigned to the class of its nearest neighbor.

*1) Knn Algorithm Pseudocode:*

1. Calculate " $d(x, x_i)$ "  $i=1, 2, \dots, n$ ; where  $d$  denotes the Euclidean distance between the points.
2. Arrange the calculated  $n$  Euclidean distances in non-decreasing order.
3. Let  $k$  be a +ve integer, take the first  $k$  distances from this sorted list.
4. Find those  $k$ -points corresponding to these  $k$ -distances.
5. Let  $k_i$  denotes the number of points belonging to the  $i^{\text{th}}$  class among  $k$  points i.e.  $k \geq 0$
6. If  $k_i > k_j \forall i \neq j$  then put  $x$  in class  $i$ .

**Screenshot of classification:****6. CONCLUSION**

Thus the levels of ovarian cancer can be predicted with the data mining techniques based on the sample tested values. This can be further enhanced to produce the optimal feature selection using bacterial foraging optimization algorithm.

**Further Enhancement:****Bacterial Foraging optimization:**

- Bacterial foraging optimization algorithm (BFOA) has been widely accepted as a global optimization algorithm of current interest for optimization and control.
- BFO possesses a poor convergence behavior over complex optimization problems as compared to other nature-inspired optimization techniques.

**REFERENCES**

- [1] ArindamRoy, BipulShyampurkayastha, SunitaSarkar, 2013, "Application of particle swarm optimization in data mining clustering: A survey", International Journal of computer applications, volume 65-no.25.
- [2] DeveshC.Jinwala, PawanR.Bhaladhare, 2014, "A clustering approach for the diversity model in privacy preserving data mining using fractionalcalculus-Bacterial Foraging Optimization Algorithm", Hindawi Publishing Corporation, Article ID 396529.
- [3] MdShariful Islam et al, 2016, "Early detection of ovarian cancer(braca 1 and braca 2mutation)risk prediction for low income country using data mining technology", 6<sup>th</sup> International Conference on Bioinformatics and Systems Biology.
- [4] T.Bhuvaneshwari, Ph.D, N.Girija, Ph.D, P.Ramachandran, Ph.D 2014, "Early detection and prevention of cancer using data mining techniques", International Journal of Computer applications, Volume 97-no.13.