Volume 4, Issue 9, September 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Feature Selection by Genetic Algorithm and SVM Classification for Cancer Detection

Tanzeem Khan Mansoori
M.Tech. scholar CSE
LNCTS Bhopal, India

Amrit Suman
Dept. of CSE
LNCTS Bhopal, India

Dr. Sadhana K. Mishra
Dept. of CSE
LNCTS Bhopal, India

Abstract— The feature selection is used for select a set of input variables by removing useless features from dataset. Feature selection strategies are often decomposed into three groups. View of the substantial range of existing feature selection algorithms, the necessity arises to count on criteria which enable to adequately decide in certain situations which algorithm should be used. This work proposes algorithms for cancer cell classification by applying genetic algorithm feature selection and then classify these cells into healthy and cancer cells. Genetic algorithm is a better method for feature selection and parameters optimization. These selected best features are used by SVM classifier on the training dataset for classify cells. Cancer Diagnosis is very important and difficult task to detect at early stage

Keywords— Feature selection, SVM, Diagnosis, Cancer detection, Genetic Algorithm Include at least 5 keywords or phrases

I. INTRODUCTION

Many practical pattern classification tasks need learning an appropriate classification function that assigns a given input pattern, usually represented by a vector of attribute values to a finite set of categories. Feature selection is employed to identify a powerfully predictive subset of fields in the database and reduce the number of fields given to the mining method. By extracting the maximum amount of data from a given data set while using smallest number of features. It is able to save important computation time and build models that generalize much better for unseen data points. According to Yang and Honavar [1], the selection of features are used to represent patterns which are given to a classifier affects many pattern classification aspects, including the accuracy of the learned classification rule, the time required for learning a classification function, the number of examples required for learning, and the cost associated with features.

Genetic algorithms can generate both optimal feature set and SVM parameters at the time. Our analysis objective is to optimize the parameters and feature set, without any lose in the SVM classification accuracy. The proposed technique performs feature selection and parameters setting in an evolutionary way. Feature set selection algorithms are often classified into two categories: the filter approach and the wrapper approach [2, 3]. The wrapper approach to feature set selection is used in paper because of accuracy. Within the literature, few algorithms are proposed for SVM feature selection [4, 5, 6, 7, 8]. Some other GA-based feature selection strategies were also proposed [9]. However, these papers only focused on feature selection, not on the parameters optimization for the SVM classifier. [10] Proposed a GA-based feature selection approach which used theoretical bounds on the generalization error for SVMs.

SVM is used to classify data with totally different class labels by determining a set of support vectors which are members of the set of training inputs that define a hyper plane in the feature space. SVM offer a standard mechanism that same as the hyper plane surface to the training values by use of a kernel function. The user might choose a kernel function (e.g. linear, polynomial, or sigmoid) for the SVM throughout the training process that selects support vectors along the surface of this function.

While using SVM, two issues are confronted: the way to opt the optimum input feature set for SVM, and the way to set effective kernel parameters. These two issues are crucial, because the appropriate kernel parameters influenced by selection of feature subset and vice versa [10]. Therefore, for getting the optimum feature set and SVM parameters should occur at the same time.

II. RELATED WORK

The paper [20] has presented a study of medical data processing and data mining that is involving the use of eleven feature selection strategies and three fuzzy modeling strategies; such strategies aren't all available in a commercial data processing and data mining package. Feature selection methodology supported association rules

The paper [21] a hybrid methodology for identification of erythemato-squamous diseases supported Association Rules (AR) and Neural Network (NN). Feature extraction is t the key for pattern recognition and classification. If the features aren't chosen well, the best classifier can perform poorly. A feature extractor should reduce the feature vector to a lower dimension that contains most of the useful data from the initial vector. So, AR is employed for reducing the dimension of erythemato-squamous diseases dataset and NN is employed for intelligent classification.

The paper [22] provided a CD-MFS algorithm which is based on mimetic evolutionary idea that uses accurate set of fuzzy if-then rules that can classify gene expression data. It begins with low quality rules, and results in high quality rule set. This algorithm classifies cancerous and benign tumors efficiently and has acceptable accuracy.

The paper [23] presents a genetic programming based methodology to classify diabetes data. To facilitate the selection of features and for evaluating the effectiveness of diabetes features various methodologies have been used in this research.

Ovarian cancer diagnosis is a vital study because early detection and accuracy staging are the keys to increase the survival rate of the patient. In papers, [24] propose a novel hybrid intelligent system, that derives simple yet convincing fuzzy inference rules to diagnose ovarian cancer and determine its stage according to the level of seriousness.

III. FEATURE SELECTION

Feature selection algorithms fall under two categories: feature ranking and set selection. Feature ranking eliminates all features that do not achieve an adequate score and rank the features by metric. Set selection searches for the optimal set from the set of possible features. Feature selection (known as set selection) is a method used in machine learning, wherein for application of learning algorithm subsets of the available features are selected from data. The most effective set contains the smallest range of dimensions that contributes to accuracy; one discards the remaining unimportant dimensions. This is a stage of preprocessing which is very important and is one of two ways by which curse of dimensionality is avoided (the other is feature extraction) [11]. These are two approaches:

Forward selection: Begin with no variables and add them one by one, at each step adding the one that decreases the error, until any more addition does not considerably decrease the error.

Backward selection: Begin with all the variables and eliminate them one by one, at each step removing the one that decreases the error (or will increase it slightly); until any more removal increase the error considerably.

There are two main models for feature selections that are filtering and wrapper model [12]. The filtering approach filter features independently from the induction algorithm after receiving a set of features. The wrapper model searches for best feature subsets, and evaluate them by using n-fold cross-validation on the training data. With any other induction algorithm this scheme may be used in conjunction and which can be used for evaluating feature subsets on the validation set. The simple greedy algorithm can be used for searching feature subsets. Greedy algorithms such as backward elimination or forward selection, or more complex ones that can both delete and add features at each step.

Since the wrapper model needs more computation, filtering is the common kind of feature selection. This is often very true within the domain of textual information retrieval, where use of bag-of-words model results large number of features. It was found that CHI, information gain (IG) and document frequency (DF) are most effective (reducing the feature set by 90-98% with no performance penalty, or small performance increase as a result of removal of noise). Contrary to a popular belief in data retrieval that common terms are less informative and document frequency that prefers frequent terms (except for stop words) was found to be quite effective for text categorization.

Advantages of feature selection

- It reduces the dimensionality of the feature space, to limit the requirement of storage and increase speed of algorithm;
- It removes the redundant, noisy or irrelevant data.
- The running time of learning algorithms are speeding up by immediate effects of data analysis tasks.
- Data quality improves.
- The accuracy of the resulting model increases.
- Feature set reduction, to avoid wasting of resources in the next round of information collection or throughout utilization.
- Performance improvement to achieve in predictive accuracy.
- Data understanding, to achieve information about process that generated the info or just visualizes the info.

Feature selection is analogous to data preprocessing technique. It is approach which is used to identify subset of features which are mostly related with target model The aim of feature selection is to increase the level of accuracy, reduce dimensionality; shorter training time and enhances generalization by reducing over fitting. Feature selection techniques are a subset of general field of feature extraction. Feature extraction use to creates new features from functions of the original features, whereas feature selection returns a set of the options. Feature selection techniques return a subset of features. Feature selection is used in domain where there are few samples (or data points) and comparatively many features.

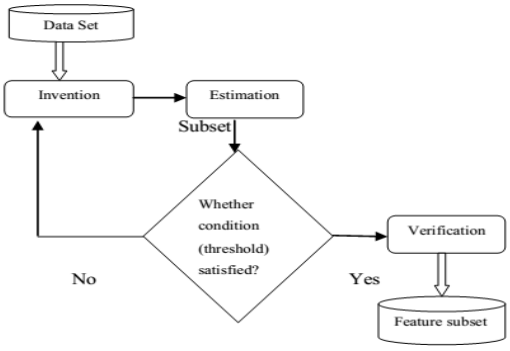

Fig. 1 Steps involved in feature selection

Steps in an every Feature selection Method:

- Invention Procedure: Manufacture candidate set from original feature set.
- Estimation Function: Estimate the candidate set.
- Evaluation: Compare with user outlined threshold value.
- Verification Method: Check out whether or not the set is valid.

IV. GENETIC ALGORITHM

Genetic algorithms (GA), is a general adaptive optimization search methodology which supports an analogy of Darwinian natural selection and genetics biological systems, could be a promising alternative to standard heuristic search. GA work with a collection of candidate solutions referred to as a population. Based on the Darwinian principle of 'survival of the fittest', the GA gains the optimum solution when a series of repetitive computations are applied. GA generates successive populations of alternate solutions which are represented by chromosomes, i.e. an answer to the problem, till acceptable results are obtained related to the characteristics of exploitation and exploration search. GA will cope with large search areas efficiently, and therefore has less likelihood to induce local optimal solution than other algorithms.

A fitness function assesses the standard of a solution in the analysis step. The crossover and mutation functions are the units that impact the fitness value. For reproduction, chromosomes are selected by evaluating the fitness value. The fitter chromosomes have higher chance to be elected into the recombination pool using the roulette wheel or the tournament selection methods.

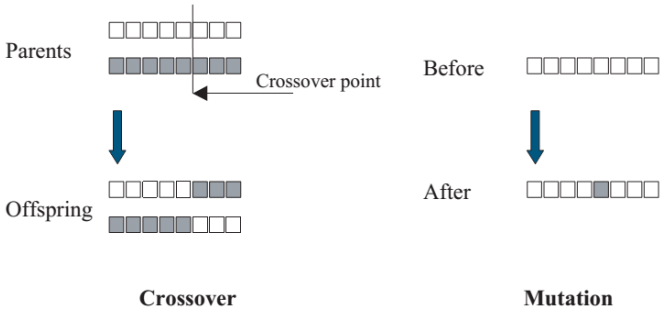

Fig 2. Genetic crossover and mutation operation.

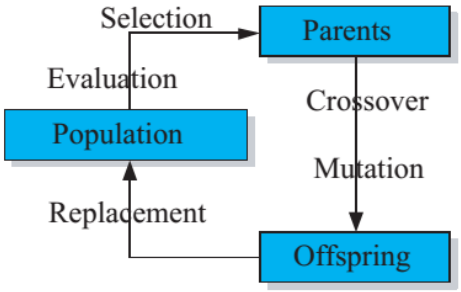

Fig 3. Evolutionary cycle.

Fig. 2 illustrates the genetic operators of crossover and mutation. Crossover, the critical genetic operator that enables new solution regions within the search space to be explored, it is a random mechanism for exchanging genes between two chromosomes using the one or two points crossover, or homologue crossover. In mutation the genes could often be altered, i.e. in binary code genes change genes code from 0 to 1 or vice versa. Offspring replaces the previous population using the diversity replacement strategy and forms a replaced or new population in the next generation. The evolutionary {biological process} process operates several generations till termination conditions satisfy. Fig. 2 depicts the GA evolutionary {biological process} process mentioned above [13]

Main steps of a GA:

1. Construction of the first generation
2. Selection

While stopping criteria not met do

3. Crossover
4. Mutation
5. Selection

End

A. Encoding solutions

The first step to perform GA is that the encoding of solutions. Indeed, in GA, each and every potential solution totally represented by a numerical vector. The historical encoding is bit strings however real encoding has a lot of uses and has sure benefits. In our case, every solution must contain the labels of peaks and the corresponding thresholds values [14].

B. Initial population

Like in any step by step optimization drawback, the information of good beginning parameters advantages the convergence speed of the algorithmic. But such kind of information is rarely available. This leads to cover large part of solution space due to generation of random initial population. Therefore, to make the exploration of the solution space easier, a very heterogeneous initial population is suitable. In our case, parameters to be set are the following, for each of the Spop individual of the initial population: the number of peaks α, the set of peaks used and corresponding thresholds of peaks. Moreover, a peak having massive amplitude among the various spectra is likely to be more interesting for discrimination [15]. That's why the probability to pick out one peak within the initial population is proportional to range of its intensity values among all the spectra. Then, for each individual, the peaks are chosen according to their amplitude and random numbers of peaks are computed. This selection leads to disadvantage very small peaks which might be remaining noise.

C. Fitness values

As discrimination of the various spectra is the objective, the fitness values have taken into account the good classification rate, τ achieved by every potential solution. Moreover, generally, a committee created of various peaks is seemingly to perform a good discrimination than a small one. However, using of several decision stumps could lead to over fit the training set and lose generality [14, 15]. So, a parsimony term concerning the number of peaks, α, in the committee is added. As there is no priori information of the optimum range of decisions stumps required, the parsimony term, ρ(a) are going to be defined as a linear function of the committee size (small sizes are favored):

$\rho(a) = \alpha a + \beta$

The fitness function is given by-

$\text{Fitness} = \tau + c \times \rho(a)$

In practice, τ (the good classification rate) is the first goal of the optimization, therefore it'll be favored. As τ and ρ (a) belong to [0, 1], c must be a real number in [0, 1].

D. Selection step

This step supported the fitness values. According the fitness value individuals are ranked and highest rank is given to the best one. Then, to keep one solution in the next generation, its probability is

P (selecting kth stratified solution) = $\partial + \mu \times k$

Where ∂ and μ are chosen so

$$\sum_{k=1}^{S_{pop}} \partial + \mu \times k = 1$$

E. Crossover step

The objective of this step is to gather interesting features (peaks and thresholds) of several solutions in new individuals by making combination of the previously retained solutions. It's necessary to notice that this step is independent from the optimization, that's to say, a crossover can produce good and worse solutions equally. Only the selection step is used to eliminate bad solutions.

F. Mutation step

This step brings the necessary hazard for efficiently explore the solution space. Any point of this space can be reached, it is assured. Moreover, if a local optimum is obtained, then a too quick convergence to this local optimum will be avoided by mutation. The mutation rate (proportion of the solutions which will undergo mutation), πm is defined for every generation. At the beginning it is set to a maximum value then decreases to allow convergence and it increases again to avoid local optima, finally. At the beginning of the algorithmic, usually having probability of one mutation per individual (in this stage, the proportion of the various mutations is more of important), hence it's set to 0.9 [14, 15]. Mutation consists in changing values in the vectors corresponding to the solutions that have been chosen for undergoing mutation. In our context, mutations will be divided into three types:

- Peak elimination: Randomly a peak is chosen and removed from solution (i.e. the committee),
- Peak addition: A new peak is chosen, added and then the optimal threshold is associated,
- Threshold relocation: Randomly one of the thresholds is removed and replaced by another one.

V. SUPPORT VECTOR MACHINE

Support Vector Machine is constructed on the structural risk minimization principle to seek a decision surface that may separate the data points into two categories with a maximal margin between them. The selection of the correct kernel function is the main challenge when using a SVM. It might have completely different forms like Radial Basis function (RBF) kernel and polynomial kernel. The advantage of the SVM is its capability of learning in sparse, high dimensional spaces with only a few training examples by minimizing the empirical error and the complexness of the classifier at same time. . WEKA uses the Sequential Minimal Optimization (SMO) algorithm for SVM. The Support Vector Machines (SVM) form a group of methods stemming from the structural risk minimization principle, with the linear support vector classifier as its most basic member. The SVM aims at creating a decision hyper plane that maximizes the margin, i.e., the distance from the hyper plane to the nearest examples from each of the classes This

allows for formulating the classifier training as a constrained optimization problem. Importantly, the objective function is unimodal, contrary to e.g. neural networks, and thus can be optimized effectively to global optimum. In the simplest case, compounds from different classes can be separated by linear hyper plane; such hyper plane is defined solely by its nearest compounds from the training set. Such compounds are referred to as support vectors, giving the name to the whole method. In most cases, however, no linear separation is possible. To take account of this problem, slack variables are introduced. These variables are associated with the misclassified compounds and, in conjunction with the margin, are subject to optimization. Thus, even though the erroneous classification cannot be avoided, it is penalized. Since the misclassification of compounds strongly influences the decision hyper plane, the misclassified compounds also become support vectors.

Support vectors and margins in linearly separable (a) and non-separable (b) problems. In non-separable case, negative margins are encountered and their magnitude is subject to optimization along with the magnitude of the positive margins.

Originally developed by Vladimir Vapnik [16], Support Vector Machines (SVMs) are a machine learning technique for supervised classification that have gained both popularity and momentum.

Support Vector Machines are linear classifiers based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. Additionally, SVMs are "a classification method that determines the maximum-margin hyperplane" [17]. This terminology is defined shortly. It should be noted that Support Vector Machines compete with Neural Networks as tools for solving pattern recognition problems.

Support Vector Machines learn their classification through a training data set of the form

$$\{\vec{x}_i, y_i\}, \vec{x}_i \in R^n, y_i \in \{-1, 1\}, i = 1 \ldots l$$

The l instances of the training data each contain an n-dimensional vector ~x that describes the features of that instance and a label y that classifies the instance as belonging to one of two categories, 1 or -1 ('positive' or 'negative') respectively. Given sufficient training examples the Support Vector Machine is then able to classify previously unseen examples (instances of data), those with no predefined label, into one of the two categories [18].

In the case of basic linear classification a Support Vector Machine creates a maximum margin hyper plane that lies in a potentially transformed input space. Given binary choice training examples (labeled either 'positive' or 'negative'), a maximum-margin hyper plane divides the 'positive' and 'negative' examples, such that the distance from between the respective class (in this case 'positive' or 'negative'), to the hyper plane is maximized. This is termed maximizing the margin.

This can also be viewed from a geometric standpoint. The Support Vector Machine attempts to construct a decision surface that bisects Rn such that all instances belonging to the positive class appear on one side of the surface with all instances belonging to the negative class appearing on the other. See Figure 4. While this approach is not new to the classification field, where SVMs set themselves apart is in their implementation.

In order to obtain a maximum margin between a class and the decision surface we must define a convex hull for that class and maximize the margin in respect to that hull.

This is because the closest approach of a particular class to the decision surface may not be at a specific point, but a linear combination of points.

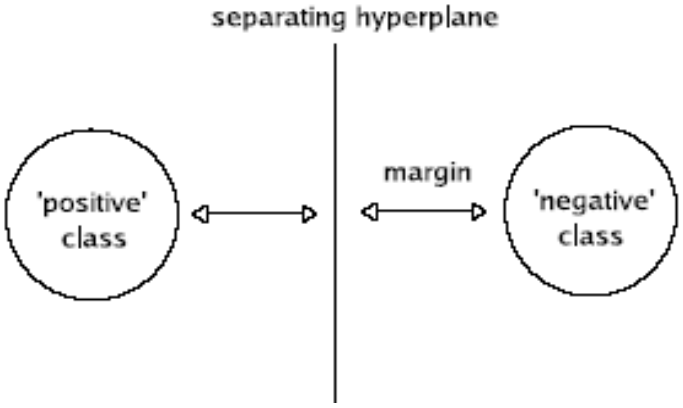

Fig. 4.Margin between classes

Formally, a convex hull is defined as "a set of points S in n dimensions that is the intersection of all convex sets containing S" [19]. For N points p1, ..., pN, the convex hull C is then given by the expression:

$$C \equiv \left\{ \sum_{j=1}^{N} \lambda_j p_i : \lambda_j > 0 \quad \text{for all} \quad j \quad \text{and} \quad \sum_{j=1}^{N} \lambda_j = 1 \right\}$$

As previously stated, Support Vector Machines are linear classifiers that construct decision surfaces (hyper planes) between the convex hulls of classes. However, through the use of kernel functions, SVMs can find hyper planes in an extended attribute space which is equivalent to finding a non-linear separating surface in the original attribute space. This allows non-linear classification.

VI. PROPOSED WORK

A. Feature Selection

Feature selection is the process of finding a subset of the original variables, with the aim to reduce and eliminate the noise dimension. The main idea of feature selection is to choose a subset of input variables by eliminating features with little or no predictive information.

Feature selection is a technique to transform high-dimensional data into lower dimensions. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features. Dimensionality reduction can be achieved by either eliminating data closely related with other data in the set, or combining data to make a smaller set of features.

1) Genetic Algorithm

- Genetic algorithms (GA), a general adaptive optimization search methodology based on a direct analogy to Darwinian natural selection and genetics in biological systems, is a promising alternative to conventional heuristic methods.
- GA work with a set of candidate solutions called a population. Based on the Darwinian principle of 'survival of the fittest, the GA obtains the optimal solution after a series of iterative computations.
- GA generates successive populations of alternate solutions that are represented by a chromosome, i.e. a solution to the problem, until acceptable results are obtained.
- GA can deal with large search spaces efficiently; A fitness function assesses the quality of a solution in the evaluation step. The crossover and mutation functions are the main operators that randomly impact the fitness value.

B. SVM Classification

Best features selected by GA are train and test for cancer cell detection.

C. Proposed Algorithm

Step 1: Start
Step 2: Read ovarian cancer dataset.
Step 3: Select best feature.
Step 4: Set the number of desired features.
Step 5: Set the fitness function Biogafit.
Step 6: call the Genetic Algorithm
Step 6.1: Construction of the first generation
Step 6.2: Selection
While stopping criteria not met do
Step 6.3: Crossover
Step 6.4: Mutation
Step 6.5: Selection
End
Step 7: Apply SVM classification
Step 7.1: Loading data from file.
Step 7.2: Initializing and generating SVM using Sub Clustering.
Step 7.3: Viewing SVM structure.
Step 8: Calculation of error and accuracy.

VII. RESULTS AND COMPARISON

A. Results

Diagnosis results divides the whole data into two sets- malignant (cancerous) and benign (non-cancerous). Ovarian Cancer and Premalignant Pancreatic Cancer databases are used for training and testing the ANFIS. We have got different results which are shown below-

B. Result using Ovarian Cancer dataset

The results given below shows in tabular format in which firstly the overall ovarian cancer dataset features is reduced refer table 1, in which large dataset of size 15154*216 is reduced to 20*216.The selected 20 features is listed in table 2. Then the reduced dataset is loaded in SVM to classify results the rules and the result is given in accuracy in percentage in table 3.

TABLE I : GENETIC ALGORITHM FOR OVARIAN CANCER

DATASET	NO. of Attributes	No. of instances	NO. of Classes
Ovarian Cancer (without GA)	15154	216	2 (Benign,Cancer)
Ovarian Cancer (with GA)	20	216	2 (Benign,Cancer)

TABLE II: BEST 20 FEATURE SET OF OVARIAN CANCER

DATA SET	NO. of Classes	NO. of Selected Attributes	Best Feature Set
Ovarain Cancer (with GA)	2 (Benign,Cancer)	20	F4186,F4831,F2475,F1540,F2592,F8463,F46,F3947,F1515,F4713,F71,F970,F8362,F1686,F54,F7601,F7511,F885,F7413,F116

TABLE III:ACCURACY FOR DIFFERENT FEATURE SET OF OVARIAN CANCER

S. No.	No. of Feature Selected	Accuracy (in %)
1	2	91.67
2	5	97.22
3	8	97.22
4	10	94.44
5	16	98.15
6	20	99.07
7	30	98.15
8	50	98.15
9	75	96.30
10	99	96.30

The paper [20] has presented a study of medical data processing and data mining that is involving the use of eleven feature selection strategies and three fuzzy modeling strategies; such strategies aren't all available in a commercial data processing and data mining package. The objective is to determine that which combination of feature selection and fuzzy modeling strategies has the best performance for a given dataset. Two medical datasets and one industrial dataset were tested with fivefold stratified cross-validation. All the combination of feature selection and fuzzy modeling strategies were applied.

C. Result using Premalignant Pancreatic Cancer dataset

For feature selection the coding is done using genetic algorithm commands in command window and for classification SVM tool of MATLAB is used.

The results are tabulated in Table 4 and Table 5.The overall pancreatic cancer dataset features are reduced from a large dataset of size 6771×181 to 16×181.The selected features are listed in Table 5.

TABLE IV: GENETIC ALGORITHM FOR PREMALIGNANT PANCREATIC CANCER

Data Set	No. of Attributes	No. of instances	No. of Classes
pancreatic Cancer (without GA)	6771	181	2 (Benign, Cancer)
Pancreatic Cancer (with GA)	16	181	2 (Benign, Cancer)

TABLE V:BEST 16 FEATURE SET FOR PREMALIGNANT PANCREATIC CANCER

Data Set	No. of Classes	No. of Selected Attributes	Best Feature Set
pancreatic Cancer (with GA)	2 (Benign, Cancer)	16	F1759,F1504,F1417,F3578,F2010,F2332,F4517,F5117,F1854,F1371, F2108,F1904,F361,F4227,F5378,F1791

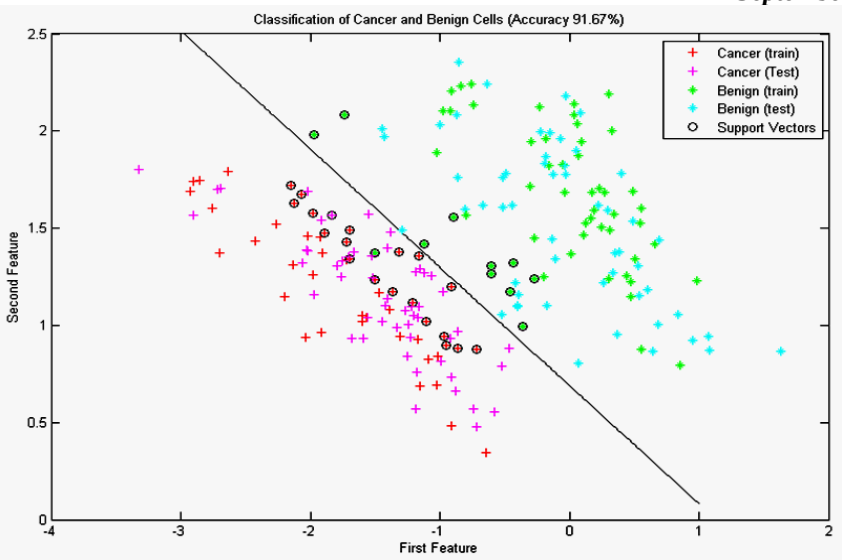

Fig 5. SVM Classification of two features as best feature from GA algorithm.

In the fig. 5 the graph is representing SVM Classification of two features as best feature from GA algorithm. It is the linear classification of benign and cancer cells. During training the data sets, classification of benign cells are represented by green color and cancer cells are represented by red color. And during testing, classifications of benign cells are represented by blue color and cancer cells are represented by pink color.

VIII. CONCLUSION

It is necessary for cancer patient to detect disease at very initial stage to survive and for better treatment. For this purpose feature selection play important role for detecting cancer at initial stage to diagnose at correct time. Feature selection process involves particular relation definition for different feature how they lead responsible for cancer. Thus Feature selection Algorithm (FSA) should be economic with respect to time and cost of detection and also need to be reliable as well. Lots of algorithms are given for this purpose which is based on totally different methods to explore various possibilities of finding rule as relation of different features with cancer disease. Feature selection technique is used to improve accuracy of classifier, reduce dataset and remove irrelevant data. This work proposed feature selection by GA and then classifies cancer cell using SVM classifier by training and testing SVM structure. By this structure new cell can be classify by observing only best feature values for cell and may be conclude as given cell is cancer cell or benign cell. Results show the effectiveness of proposed work this method is gives high accuracy for feature selecting 20 features.

REFERENCES

[1] Yang, J., & Honavar, V. (1998). Feature subset selection using a genetic algorithm.IEEE Intelligent Systems, 13(2), 44–49.

[2] John, G., Kohavi, R., & Peger, K. (1994). Irrelevant features and the subset selection problem. Proceedings of the 11th international conference on machine learning, San Mateo, CApp. 121–129.

[3] Kohavi, R., & John, G. (1997). Wrappers for feature subset selection.Artificial Intelligence, 97(1–2), 273–324.

[4] Bradley, P. S., & Mangasarian, O. L. (1998). Feature selection via concave minimization and support vector machines. In Proceedings of the 13th international conference on machine learning (pp. 82–90). San Francisco, CA.

[5] Bradley, P. S., Mangasarian, O. L., & Street, W. N. (1998). Feature selection via mathematical programming. INFORMS Journal on Computing, 10, 209–217

[6] Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., & Vapnik, V. (2001). Feature selection for SVM. In S. A. Solla, T. K. Leen, & K.-R. Muller, Advances in neural information processing systems(Vol. 13) (pp.668–674). Cambridge, MA: MIT Press.

[7] Guyon, I., Weston, J., Barnhill, S., & Bapnik, V. (2002). Gene selection for cancer classification using support vector machines. Machine Learning, 46(1–3), 389–422

[8] Mao, K. Z. (2004). Feature subset selection for support vector machines through discriminative function pruning analysis. IEEE Transactions on Systems, Man, and Cybernetics, 34(1), 60–67

[9] Raymer, M. L., Punch, W. F., Goodman, E. D., Kuhn, L. A., & Jain, A. K. (2000). Dimensionality reduction using genetic algorithms. IEEE Transactions on Evolutionary Computation, 4(2), 164–171.

[10] Fro¨hlich, H., & Chapelle, O. (2003). Feature selection for support vector machines by means of genetic algorithms. Proceedings of the 15th IEEE international conference on tools with artificial intelligence, Sacramento, CA, US App. 142–148.

[11] Fabrizio Sebastiani, "Machine Learning in Automated Text Categorization", ACM Computing Surveys, Vol.34, No. 1, pp. 1–47, 2002.

[12] YZhao and G Karypis, "Hierarchical clustering algorithms for document datasets", Data Mining and Knowledge Discovery, pp. 141-168, 2005.
[13] Davis, L. (1991).Handbook of genetic algorithms. New York: Van Nostrand Reinhold.
[14] Christelle Reynes, Robert Sabatier, Nicolas Molinari, Sylvain Lehmann," A new genetic algorithm in proteomics: Feature selection for SELDI-TOF data", in Elsevier Computational Statistics and Data Analysis 52 (2008) 4380–4394.
[15] M. Dash, H. Liu," Feature Selection for Classification", in Elsevier Intelligent Data Analysis 1 (1997) 131–156.
[16] Vladimir N. Vapnik. The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA, 1995.
[17] George Quievryn. Mutplus - glossary.
[18] Douglas Clarke. An Investigation into Applying Support Vector Machines to Pixel Classification in Image Processing. Honours Thesis, School of Computer Science and Software Engineering, Monash University, 2003.
[19] Eric W. Weisstein. Convex hull. from mathworld–a wolfram web resource, 1999-2005.
[20] Sean N. Ghazavi, Thunshun W. Liao, "Medical data mining by fuzzy modeling with selected features", Artificial Intelligence in Medicine 43, Pages 195—206, Elsevier, 2008.
[21] Murat Karabataka, M. Cevdet Ince, "A new feature selection method based on association rules for diagnosis of erythemato-squamous diseases", Expert Systems with Applications, 36 Pages 12500–12505, Elsevier, 2009
[22] A. Zibakhsh, M. Saniee Abadeh, "Gene selection for cancer tumor detection using a novel memetic algorithm with a multi-view fitness function", Engineering Applications of Artificial Intelligence 26, Pages 1274–1281, Elsevier, 2013.
[23] Muhammad Waqar Aslama, Zhechen Zhu, Asoke Kumar Nandi, "Feature generation using genetic programming with comparative partner selection for diabetes classification", Expert Systems with Applications 40, Pages 5402–5412, Elsevier, 2013
[24] Di Wanga, Chai Queka, Geok See Ng, "Ovarian cancer diagnosis using a hybrid intelligent system with simple yet convincing rules", Applied Soft Computing 20, Pages 25–39, Elsevier, 2014.

