AMSE Association for the Advancement of Modelling & Simulation Techniques in Enterprises
IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

**Advances in Modelling and Analysis B**
Vol. 61, No. 3, September, 2018, pp. 145-150
Journal homepage:http://iieta.org/Journals/AMA/AMA_B

# A case study approach for Brest cancer prediction using feature selection method based on AOC and SVM

Dudekula Mahammad Rafi[1*], Chettiar R. Bharathi[2]

[1] Research Scholar, Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu. & Asst. Professor, Vivekananda Institute of Engineering & Technology, JNTU University, Hyderabad 501301, India
[2] Assoc.Prof, Department of ECE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai 600054, Tamil Nadu, India

Corresponding Author Email: nagendra.csbu@gmail.com

https://doi.org/10.18280/ama_b.610307

**Received:** 4 July 2018
**Accepted:** 19 August 2018

***Keywords:***
*feature selection method,breast cancer prediction,aoc, svm,case study, genetic methods*

**ABSTRACT**

Data mining assumes an essential part in Health care. It truly predicts the ailment in light of examined information. Finding in the medicinal field is an entangled undertaking that ought to be performed with exactness and proficiency. A determination performed by a doctor for a solitary patient may vary fundamentally if the same is analysed by alternate doctors or by similar doctors at various occasions to that solitary patient. Presently a days, robotized therapeutic examination are utilized to assist specialists with predicting ailments at a quick pace. Especially, Cancer is one of the main sources of death around the world. Early identification and avoidance of tumour assumes an imperative part in decreasing passing's caused by disease. Recognizable proof of hereditary and ecological elements is critical in creating novel techniques to identify and anticipate disease. We propose subterranean insect state enhancement based SVM to group the therapeutic information with a specific end goal to get more precise outcomes than existing strategies. Also, here in this paper we introduces a contextual investigation of utilizing information mining methods in the examination of determination Breast Cancer illness. This exploration grandstands the information mining procedure and strategies to change a lot of patient information into helpful data and possibly profitable examples to help comprehend disease results.

## 1. INTRODUCTION

As per the overview led in 2017, in only us there are 252,710 instances of bosom disease. So the quantity of bosom growth all around the globe will be an esteem extremely immense. In US the passing rate of ladies because of bosom malignancy is higher contrasted with other disease classifications. After skin malignancy bosom disease is most usually found among ladies. In India it is discovered that bosom tumor is currently generally found in more youthful age bunches too. It is the most well-known sort of tumor in Indian urban territories and second regular in rustic zones. Every one of these actualities focuses hand towards the significance of relieving bosom tumor at a beginning time. For doing that some instrument is really expected to foresee the malignancy in light of the past information and the present information.

Highlight Selection is a fundamental piece of learning disclosure. FS is utilized to enhance the characterization exactness and lessen the computational time of order calculations. FS is separated into the managed and unsupervised classes. At the point when class names of the information are accessible we utilize regulated element determination, generally unsupervised element choice is proper. In numerous information mining applications, class names are obscure, along these lines showing the hugeness of unsupervised component determination [2]. Regarding highlight choice strategies, they fall into channel and wrapper classifications. In channel show, highlights are assessed in light of the general attributes of the information without depending on any mining calculations. Despite what might be expected, wrapper demonstrate requires one mining calculation and utilizations its execution to decide the decency of capabilities [3].

At that point highlight choice is done to upgrade tumor information order. For the classifier mix two primary methodologies are utilized: classifiers fusipon and classifiers determination. In the primary strategy, all classifiers in the group add to the choice of the MC framework, e.g. through entirety or dominant part voting. In the second approach, a solitary classifier is chosen from the troupe and its choice is dealt with as the choice of the MC framework [11]. SVM is additionally a classifier. The SVM endeavors to decide a tradeoff between limiting the preparation set mistake and boosting the edge keeping in mind the end goal to accomplish the best speculation capacity and stay impervious to over fitting. Because of its invaluable nature, SVM has been connected to an extensive variety of characterization assignments. Specifically, SVM has been appeared to perform exceptionally well on numerous therapeutic finding undertakings. Notwithstanding, there is as yet a requirement for enhancing the SVM classifier's execution [12].

Henceforth to defeat the issue a few new techniques figure out how to analyze sickness in a completely information driven way, utilizing multivariate arrangement or relapse to straightforwardly outline imaging information to conclusion. These methods are not limited by momentum information on

infection related radiological examples and frequently have higher demonstrative exactness than more customary quantitative investigation in view of straightforward volume or thickness measures [13]. Anyway the exchange of basically sick patients requires great coordination to give the demonstrative devices and the most suitable treatment for their conditions [14]. Consequently it is vital to characterize the infection on time and along these lines giving medications to the patients to hazard evasion. Once in a while same properties with similar side effects of maladies required diverse medicines. Along these lines exact choices and characterization of information is required [15].

Scientists endeavored to utilize differing techniques to show signs of improvement precision of information order. It is fundamental to perform malignancy information arrangement with a specific end goal to recognize the disease. Subsequently to defeat those issues our proposed strategy is utilized. Here at first the pre-handling will be connected to remove helpful information and to change over reasonable example from crude therapeutic datasets. In the wake of preprocessing for ideal determination of highlights ACO based SVM is utilized. This technique characterizes the tumor information as typical and irregular. Henceforth our proposed strategy precisely orders the tumor information utilizing ideal highlights. The execution of the proposed technique is assessed as far as precision, affectability and specificity. The proposed technique will be executed in MATLAB.

Hazard factors related with bosom growth Every lady needs to recognize what she can do to bring down her danger of bosom tumor. A portion of the variables related with bosom cancer2 :

(1) Being a lady: Just being a lady is the greatest hazard factor for creating bosom tumor. There are around 190,000 new instances of intrusive bosom growth and 60,000 instances of non-obtrusive bosom malignancy this year in American ladies. While men do create bosom disease, under 1% of all new bosom malignancy cases occur in men. Around 2000 instances of bosom disease will be analyzed in American men this year.

(1) Age: As with numerous different illnesses, your danger of bosom malignancy goes up as you get more established. Around two out of three intrusive bosom malignancies are found in ladies 55 or more seasoned.

(2) Family history: Women with close relatives who have been determined to have bosom growth have a higher danger of building up the sickness. On the off chance that you have had one firstdegree female relative (sister, mother, little girl) determined to have bosom malignancy, your hazard is multiplied.

(3) Genetics: About 5% to 10% of bosom malignancies are believed to be innate, caused by irregular qualities go from parent to kid. • Personal history of bosom disease: If you have been determined to have bosom growth, you are three to four times more prone to build up another tumor in the other bosom or an alternate piece of a similar bosom. This hazard is not quite the same as the danger of the first disease returning (called danger of repeat).

(4) Radiation to chest or face before age 30: If you had radiation to the chest to treat another growth (not bosom tumor, for example, Hodgkin's illness or non-Hodgkin's lymphoma, you have a higher-thanaverage danger of bosom malignancy. On the off chance that you had radiation to the face at a pre-adult to treat skin inflammation (something that is never again done), you are at higher danger of creating bosom tumor sometime down the road.

(5) Certain bosom changes: If you have been determined to have certain amiable (not tumor) bosom conditions, you may have a higher danger of bosom growth. There are a few sorts of considerate bosom conditions that influence bosom malignancy chance.

(6) Race/ethnicity: White ladies are somewhat more prone to create bosom tumor than African American, Hispanic, and Asian ladies. Yet, African American ladies will probably grow more forceful, further developed stage bosom disease that is analyzed at a youthful age.

(7) Being overweight: Overweight and fat ladies have a higher danger of being determined to have bosom tumor contrasted with ladies who keep up a sound weight, particularly after menopause. Being overweight likewise can build the danger of the bosom malignancy returning (repeat) in ladies who have had the sickness.

(8) Pregnancy history: Women who have not had a fullterm pregnancy or have their first kid after age 30 have a higher danger of bosom growth contrasted with ladies who conceived an offspring before age 30.

(9) Breastfeeding history: If a lady breastfeeds for longer than one year this may lessen bosom disease hazard.

(10) Menstrual history: Women who began discharging (having periods) more youthful than age 12 have a higher danger of bosom growth further down the road. The same is valid for ladies who experience menopause when they are more seasoned than 55.

(11) Using HRT (hormone substitution treatment): Current or later past clients of HRT have a higher danger of being determined to have bosom disease. • Drinking liquor: Research reliably demonstrates that drinking mixed refreshments – brew, wine, and alcohol – expands a lady's danger of hormone-receptorpositive bosom malignancy.

(12) Having thick bosoms: Research has demonstrated that thick bosoms can be six times more prone to create tumor and can make it harder for mammograms to recognize bosom growth.

(13) Lack of activity: Research demonstrates a connection between practicing consistently at a direct or extreme level for 4 to 7 h for every week and a lower danger of bosom disease.

(14) Smoking: Smoking causes various illnesses and is connected to a higher danger of bosom growth in more youthful, premenopausal ladies. Research additionally has demonstrated that there might be connect between substantial second-hand smoke introduction and bosom growth chance in postmenopausal ladies.

(15) Low of vitamin D levels: Research recommends that ladies with low levels of vitamin D have a higher danger of bosom malignancy. Vitamin D may assume a part in controlling typical bosom cell development and might have the capacity to prevent bosom malignancy cells from developing.

(16) Light introduction during the evening: The aftereffects of a few investigations recommend that ladies who work during the evening – assembly line laborers, specialists, medical attendants, and cops, for instance – have a higher danger of bosom growth contrasted with ladies who work amid the day.

(17) DES (diethylstilbestrol) presentation: Women who took DES themselves have a marginally higher danger of bosom growth. Ladies who were presented to DES while their moms were pregnant with them additionally may have marginally higher danger of bosom malignancy sometime

down the road.

(18) Eating undesirable sustenance: Diet is believed to be in any event mostly in charge of around 30% to 40% all things considered. No nourishment or eating routine can keep you from getting bosom disease.

(19) Exposure to synthetic compounds in beautifying agents: Research unequivocally proposes that at certain presentation levels, a portion of the synthetic concoctions in beautifiers may add to the improvement of malignancy in individuals.

(20) Exposure to synthetic compounds in nourishment: There is a genuine worry that pesticides, anti-infection agents, and hormones utilized on yields and animals may cause medical issues in individuals, incorporating an expansion in bosom tumor chance. There are likewise worries about mercury in fish and mechanical synthetic compounds in sustenance and nourishment bundling.

(21) Exposure to synthetic concoctions for yards and greenery enclosures: Research firmly recommends that at certain introduction levels, a portion of the synthetics in grass and garden items may cause disease in individuals.

## 2. PROPOSED WORK

### 2.1 Cancer prediction

The general research procedure for this examination was adjusted in view of the information revelation process and is delineated in Figure 1. The information securing stage was the principal period of this philosophy, in which we acquired the pertinent information for the investigation. The second stage was the information preprocessing stage, in which the gathered data was coordinated, cleaned, and changed with the end goal that the datasets were reasonable for order expectation. After this, still in the second stage, we completed component extraction. The information from the preprocessing stage (Phase 2) were then extended to Phase 3 for order forecast. In Phase 4, the upgraded forecast calculation, in view of the AOC and SVM. In the last period of the exploration, we played out a relative investigation of the models without include determination and the models that utilized element choice.

2.1.1 Data acquisition

In the data selection phase, we collected breast cancerdata from the UCI public database [60]. The BreastCancer Wisconsin Breast Cancer Prognostic Dataset has198 instances and 34 attributes.

2.1.2 Data preprocessing

(1) Data cleaning

The integrated database went through the data cleaning process,in which we removed improper data entries, such asthose that provided an irrelevant answer, in the database.To smooth noisy data, the tuples with improper data entry were eliminated or filled with the most probable value, as thisis one of the most popular strategies to counter this issue. Additionally, the find and replace function was used to handleinconsistency in the format of data from the survey.

Figure-1 shows theCase study of Cancer data classification, here initially take the cancer data set and apply the attribute selection with the help AOC and then proceed for AOC based SVM for exact taking of features and also apply the prediction of cancer for the data set.

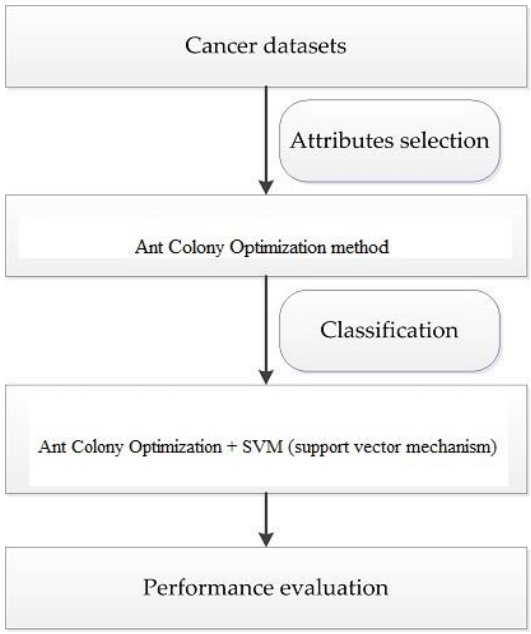



**Figure 1.** A case study of cancer data classification

Feature choice is the way toward finding a subset of features, from the first feature set, ideally. Ant Colony Optimization technique investigates to locate the ideal feature subset utilizing a few cycles. The primary goal of the proposed technique is to limit the excess between them by choosing a subset of features. In this technique most minimal closeness features is chosen by every subterranean insect to the past chose features. In this manner, if a feature is chosen by the vast majority of the ants, this shows the feature has the most minimal comparability to alternate features. The feature gets the best measure of pheromone, and the odds of its determination by alternate ants will be expanded in the following cycles. At last, by utilizing the comparability between features the best chosen features will have high pheromone esteems. In this way, the proposed technique chooses the best features with least repetition. The feature significance prompts the minimization of the repetition which will be figured in light of the closeness inside the features. The means to perform ACO feature choice is portrayed underneath. In this strategy, before the feature determination technique begins, the inquiry space must be communicated as an appropriate frame for ACO. In this manner, the hunt space is communicated as a completely associated undirected weighted chart, G = < F , E > where F = {F1 , F2 ,… , Fn } demonstrates an arrangement of all features in that each feature signifies a hub in the diagram, E = {(Fi , Fj ): Fi , Fj ∈ F} shows the chart limit. The association of the limit (Fi , Fj ) ∈ E will be set to the connection esteem amongst Fi and Fj .

Stage 1: The reason for a closeness related with the limit between features I and j. The comparability between any two features is recognized by figuring the ideal estimation of the cosine closeness among them. The cosine comparability between features An and B is ascertained utilizing the accompanying condition.

$$sim(A,B) = \left| \frac{\sum_i^p = 1(a,b_i)}{\left(\sqrt{\sum_i^p = 1^{a^2}}\right)\left(\sqrt{\sum_i^p = 1^{b^2}}\right)} \right|$$

Here An and B demonstrates two features with p-dimensional vector (A={a1 , a2 ,..., a p } B ={b1 , b2 ,..,bp}).The similitude estimation of two features spoke to as 0 and 1 where 1 shows comparative features and 0 speaks to non-comparable features. On the off chance that the computed likeness esteem between two features is more noteworthy than zero the features are comparable. By utilizing ACO technique in the feature choice issue, "heuristic information" and "attractive quality" must be characterized by ACO calculation. In the proposed technique, the heuristic information is portrayed as the opposite of the closeness inside the features. An allure measure τi∀i = 1… n, ecalled "pheromone", which is identified with the features and will be revived by ants routinely.

Stage 2: The proposed technique is made out of numerous emphasess. The gathering of pheromone allotted to every hub before cycles begin. In each emphasis, NAnt ants are settled haphazardly on the different hubs. Persistently as indicated by a probabilistic "state progress manage" every ant crosses the hubs monotonously until the point that a cycle halting tenet is fulfilled. Halting tenet is depicted as the number occasions the hubs that ought to be chosen by every ant.

$$j = \underset{u \in J_i^k}{\arg} \max\left\{[\tau_u][\eta(F_i,F_u)]^{\beta}\right\}, if\ q \le q_0$$

Stage 3: The state change run expects to choose features utilizing most elevated pheromone esteems and littlest likenesses to beforehand chosen features. Feature counter exhibit stores the features that is been chosen by any ant. Stage 4: Then, toward the finish of the emphasis, by applying a "worldwide refreshing guideline" the quantity of pheromone for every single hub is refreshed. In light of its feature counter esteem the quantity of pheromone for each hub is processed. The ants have a tendency to furnish more pheromone to hubs with higher feature counter qualities. And additionally, a bit of the pheromone dissolves on all hubs. Stage 5: The undertaking is rehashed till a given measure of cycles are accomplished. Next, the features are gathered of their pheromone esteems in bring down request. At that point, the best chosen features with most noteworthy pheromone esteems are chosen as the last feature subset. The "state progress manage" depends on a succession of the heuristic information and the vertex pheromone esteems as demonstrated as follows: When the ant k is put on the feature I, the up and coming feature j can be choosed by a voracious strategy or in a probabilistic technique. In the ravenous technique, the forthcoming feature is chosen in light of the accompanying condition:

$$P_k(i,j) = \frac{[\tau_j][\eta(F_i,F_j)]^{\beta}}{\sum_{u \in J_i^k}[\tau_u][\eta(F_i,F_j)]^{\beta}}, if\ j \in J_i^k\ if\ q > q_0$$

where k I j is the unvisited feature set, τu is the pheromone allocated to the feature u, η (Fi, Fu) = 1 (, ) F Fi u sim is the opposite of the closeness between features I and u, β is a parameter that is utilized in the substance of pheromone versus likeness (β > 0), q0 ∈ [0,1] is a constant parameter, and q is an irregular incentive meanwhile [0,1]. In the probabilistic strategy, the forthcoming feature j will be chosen in view of the likelihood Pk (i,j) which is depicted as takes after:

State progress administer in light of the parameters q and q0 which is an arrangements amongst Exploitation and Exploration. In the event that q ≤ q0 , then ants pick the ideal feature in the ravenous way, or there will be consequences, each feature has a plausibility of being chosen by its likelihood esteem that is ascertained utilizing investigation. The motivation behind the probabilistic strategy is to abstain from being caught inside a neighborhood ideal. The association of both the probabilistic and the covetous techniques are designated "pseudo-arbitrary corresponding rule"6,7. The "worldwide refreshing principle" is utilized to whole hubs toward the finish of the ant's navigate utilizing the given condition:

$$\tau_i(t+1) = (1-q)\tau_i(t) + \frac{FC[i]}{\sum_j^n = 1 FC[j]}$$

where n is the number of unique features, and are the total number of pheromone values of feature i at times t and t+1, thus, ρ is a pheromone evaporation p.

## 3. EXPERIMENTAL EVALUATION:

### 3.1 Evaluation metrics

**Table 1.** Attribute information of the dataset

```
RangeIndex: 569 entries, 0 to 568
Data columns (total 31 columns):
diagnosis                569 non-null object
radius_mean              569 non-null float64
texture_mean             569 non-null float64
perimeter_mean           569 non-null float64
area_mean                569 non-null float64
smoothness_mean          569 non-null float64
compactness_mean         569 non-null float64
concavity_mean           569 non-null float64
concave points_mean      569 non-null float64
symmetry_mean            569 non-null float64
fractal_dimension_mean   569 non-null float64
radius_se                569 non-null float64
texture_se               569 non-null float64
perimeter_se             569 non-null float64
area_se                  569 non-null float64
smoothness_se            569 non-null float64
compactness_se           569 non-null float64
concavity_se             569 non-null float64
concave points_se        569 non-null float64
symmetry_se              569 non-null float64
fractal_dimension_se     569 non-null float64
radius_worst             569 non-null float64
texture_worst            569 non-null float64
perimeter_worst          569 non-null float64
area_worst               569 non-null float64
smoothness_worst         569 non-null float64
compactness_worst        569 non-null float64
concavity_worst          569 non-null float64
concave points_worst     569 non-null float64
symmetry_worst           569 non-null float64
fractal_dimension_worst  569 non-null float64
dtypes: float64(30), object(1)
memory usage: 137.9+ KB
```

Keeping in mind the end goal to survey the effectiveness of our proposed technique different assessment measurements are used. The measurements comprises of gathering of regards that contains typical essential assessing strategies. The assessment measurements utilized here contains True Positive, True Negative, False Positive and False Negative, Sensitivity,

Specificity and Accuracy.

$$Sensitivity = \frac{T(P)}{T(P)+F(N)}$$

$$Specificity = \frac{T(N)}{F(P)+T(N)}$$

$$Accuracy = \frac{T(P)+T(N)}{T(P)+F(N)+F(P)+T(N)}$$

For this exploration, which concentrated on the strategies and methods beforehand examined, the examination utilized the accessible dataset given by the UC Irvine machine learning storehouse, gained from the Wisconsin Prognostic Breast Cancer sub-catalog with 198 occurrences [60]. A portrayal of the dataset's characteristics and areas is given in Table 1.

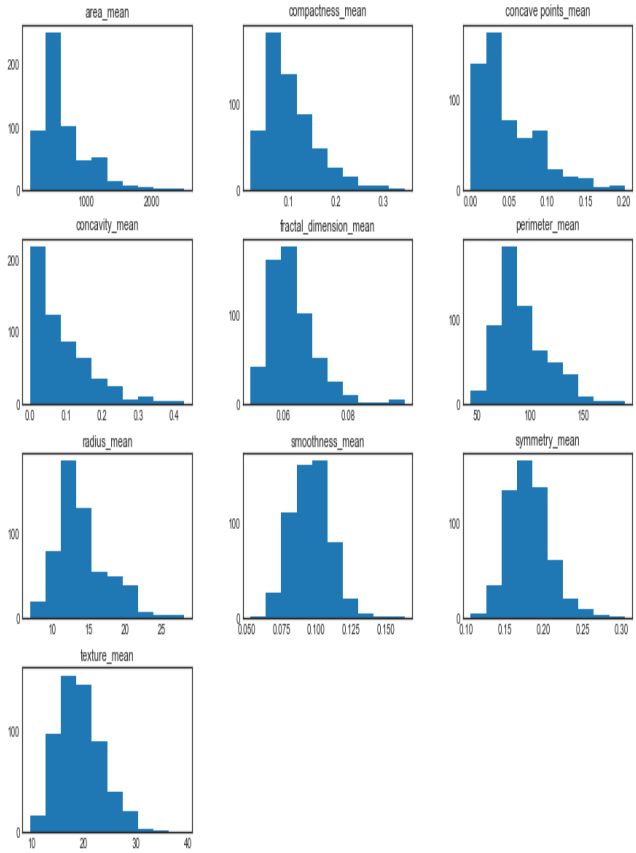



**Figure 2.** Breast cancer identification

The first 30 features were computed from a digitized image of a fine needle aspirate (FNA) of a breast mass sample. They describe characteristics of the cell nuclei present in the image. Ten real-valued features were computed for each cell nucleus:

For each picture, the mean, standard mistake, and ``worst'' or biggest (the mean of the three biggest qualities) of these features were processed, bringing about 30 features. For example, field 4 is Mean Radius, field 14 is Radius SE, and field 24 is Worst Radius. Qualities for features 4-33 were re-coded with four significant digits. There were four instances of missing qualities for the Lymph hub status characteristic. The class dissemination was found to incorporate 151 non-repeats and 47 repeats. Each record speaks to catch up information for one bosom growth case. These are back to back patients treated by Dr. Wolberg since 1984 [64] and involve just those cases exhibiting obtrusive bosom malignancy and no proof of distant metastases at the season of analysis. Alternate traits are recorded in Table 1, which condenses every one of the 35 qualities of the dataset.

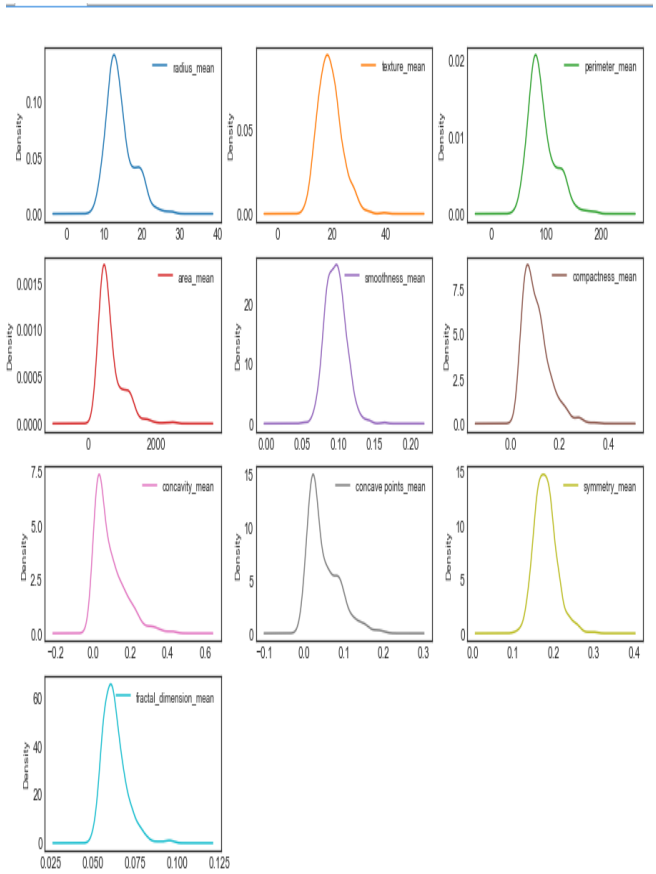
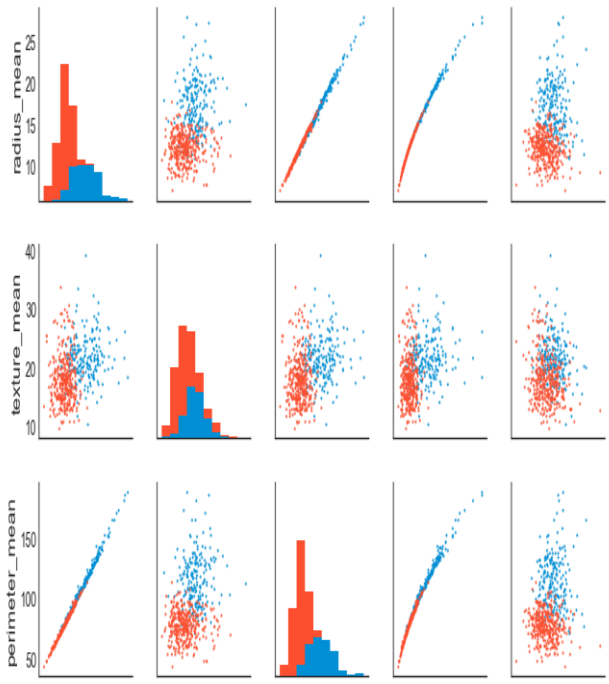



**Figure 3.** Growth of cancer cell



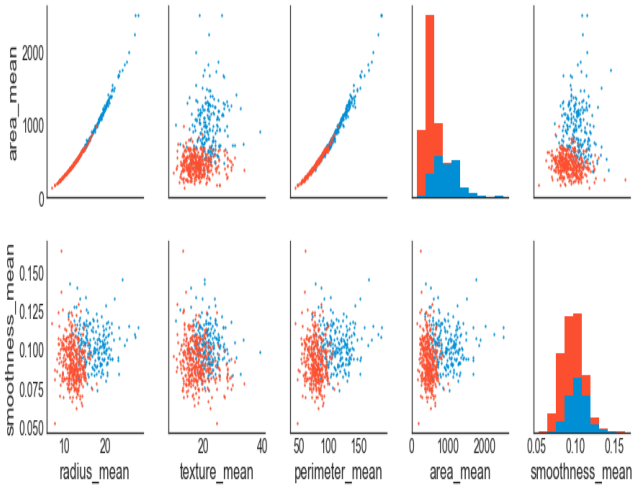



**Figure 4.** Cancer cell clusters

## 4. RESULTS OF ACO BASED FEATURE SELECTION

Ant Colony Optimization based feature determination procedure was utilized to choose the features in bosom tumor datasets. The consequences of feature Selection utilizing Ant Colony Optimization in Breast Cancer datasets are appeared in Figure 2. The test results demonstrate that the feature decrease of WDBC Breast tumor dataset is 55.55%. Ant Colony Optimization based feature was connected to the bosom and liver tumor datasets to locate the ideal features from the first feature set. Bosom growth dataset trait names meant by {A0,A1,A2,A3,A4,A5,A6,A7,A8} and chose features are {A1,A2,A3,A8}. At that point, the entire features and chose features are connected in SVM order system. The level of feature decrease for WDBC dataset is 55.55%. Bolster Vector Machine is utilized for order. The grouping is connected for the entire features and diminished feature subset. The exactness of the SVM arrangement for entire feature set for WDBC dataset is 94.42% and for the decreased feature subset is 96.56%. There is a 2.12% of change in precision.

## 5. CONCLUSION

Breast cancer is the most well-known danger among ladies, representing about 1 of every 3 cancers analyzed among ladies in the United States, and it is the second driving reason for cancer demise among ladies. Breast Cancer happens as a consequences of strange development of cells in the breast tissue, usually alluded to as a Tumor. A tumor does not mean cancer - tumors can be favorable (not cancerous), pre-malignant (pre-cancerous), or malignant (cancerous). Tests, for example, MRI, mammogram, ultrasound and biopsy are usually used to analyze breast cancer performed. Since this construct a model that can characterize a breast cancer tumor utilizing two preparing characterization: 1= Malignant (Cancerous) – Present 0= Benign (Not Cancerous) – Absent.

.

## REFERENCES


[1] Siegel R, Ma J, Zou Z, Jemal A. (2014). Cancer statistics, 2014.' CA,Cancer J. Clinicians 64(1): 9-29.

[2] Bhardwaj A, Tiwari A. (2015). Breast cancer diagnosis using Geneticallyoptimized neural network model. Expert Syst. Appl. 42(10): 4611-4620.

[3] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. (2011). Global cancer statistics, CA, A Cancer J. Clinicians 61(2): 69-90.

[4] Youlden DR, Cramb SM, Dunn NA, Müller JM, Pyke CM, Baade PD. (2012). The descriptive epidemiology of female breast cancer: Aninternational comparison of screening, incidence, survival and mortality. Cancer Epidemiol. 36(3): 237-248.

[5] Saghiret ENS. (2007). Trends in epidemiology and management of breastcancer in developing Arab countries: A literature and registry analysis. Int. J. Surg. 5(4): 225-233.

[6] Ravichandran K, Al-Zahrani AS. (2009). Association of reproductivefactors with the incidence of breast cancer in Gulf cooperation council countries. East Mediterr. Health J 15(3): 612-621.

[7] Thompson D, Easton D. (2004). The genetic epidemiology of breast cancergenes. J. Mammary Gland. Biol. Neoplasia 9(3). 221-236.

[8] Bray F, McCarron P, Parkin DM. (2004). The changing global patterns offemale breast cancer incidence and mortality. Breast Cancer Res. 6(6): 229-239.

[9] Parkin DM, Bray F, Ferlay J, Pisani P. (2005). Global cancer statistics. CA, Cancer J. Clinicians 55(2): 74-108.

[10] McPherson K, Steel CM, Dixon KM. (2000). Breast cancer_Epidemiology, risk factors, and genetics. BMJ 321(7261): 624-628.

[11] Perera NM, Gui GP. (2003). Multi-ethnic differences in breast cancer:Current concepts and future directions. Int. J. Cancer 106(4): 463-467.

[12] Ezzat AA, Ibrahim EM, Raja MA, Al-Sobhi S, Rostom A, Stuart RK. (1999). Locally advanced breast cancer in Saudi Arabia: High frequencyof stage III in a young population. Med. Oncol 16(2): 95-103.

[13] Ibrahim EM, Ezzat AA, Rahal MM, Raja MM, Ajarim DS. (2005). Adjuvant chemotherapy in 780 patients with early breast cancer: 10-yeardata from Saudi Arabia," Med. Oncol 22(4): 343-352.

[14] Elkumet N. (2007). Being 40 or younger is an independent risk factor forrelapse in operable breast cancer patients: The Saudi Arabia experience, BMC Cancer 7: 222.

[15] Najjar H, Easson A. (2010). Age at diagnosis of breast cancer in Arabnations. Int. J. Surg 8(6): 448-452.

[16] Farr A, Wuerstlein R, Heiduschka A, Singer CF, Harbeck N. (2013). Modern risk assessment for individualizing treatment concepts in early-stage breast cancer. Rev. Obstetrics Gynecol 6(3-4): 165-173.

[17] Fayyad U, Piatetsky-Shapiro G, Smyth P. (1996). From data mining toknowledge discovery in databases. AI Mag 17(3): 37.