A SURVEY ON MULTI OBJECTIVE OPTIMIZATION BASED ON NSGA-II FOR BREAST CANCER CLASSIFICATION

M.Divyasatyapadma
computer science and engineering
Griet
Hyderabad,India

G.Mallikarjuna Rao
computer science and engineering
Griet
Hyderabad,India

Abstract---Breast cancer is the quotidian disease among women. Due to this reason, carcinoma detection is most focused area by many researchers. The accurate diagnostic detection of the cancerous cells in patients critical and should alter the next treatment and increase the prospect of survivability. A trickle of problems like logistical reasons, which related to imaging exceptional and human error, boom misdiagnosis of carcinoma via way of means of radiotherapist Computer-aided detection systems (CADs) are advanced to overcome those regulations and are studied in lots of imaging modalities for carcinoma detection in current years. And Artificial Neural Network (ANN) classifier is one among the foremost used work classifier for breast cancer classification however it is weak in determining parameters hence we introduce optimization using genetic algorithm along with NSGA 2 global optimization tool for better parameter estimation and to get more accuracy rate. This survey paper focuses on optimization techniques

*Keywords— multi-objective optimization, genetic algorithm, non-dominated, NSGA II*

1. INTRODUCTION

Optimization is a technique which is used in our daily activities from building plan to budgetary markets from design innovation to mass correspondence. If given a problem to be optimized the results would be in such a way that to maximize or minimize something to provide efficient solution and this is simply know as objective function. During the period 1939-1945 the optimization strategies have commenced, beforehand they were wont to optimize the ballistic missiles. Later on mathematical programming has been advanced to recognize the optimization via the equipment of mathematical strategies. Considering the fact that in 1970 the optimization methods has been broadly evolved. There are numerous optimization tools to determine the solution for objective functions, organizations are implementing optimization techniques to escalate their profits and diminish their cost.

CLASSICAL OPTIMIZATION TECHNIQUES

The below figure will give a clear view of classification of optimization techniques ,Generally any engineering hassle can have a big wide variety of answers out of which a few are viable and a few are infeasible. The challenge is to get fine answer out of viable answers. The whole set of viable answers constitutes viable design area and the development towards the optimal design entails the search within the space. And

they are two types of searches one is deterministic and the other is stochastic, in deterministic search, the output of the version is completely decided with the aid of using the parameter values and the preliminary situations. Stochastic search own a few inherent randomness. The identical set of parameter values and preliminary situations will result in an ensemble of various outputs. Whether the quest is deterministic or stochastic it's far viable to enhance the reliability of outcomes and to enhance the reliability a transition rule is used to enhance the result

Nontraditional search has grown to be greater favored in optimization problem, the algorithms which can be utilized in non-traditional optimization are Genetic algorithm, simulated annealing, Particle swarm optimization,
Direct algorithm, Fmincon and so forth.
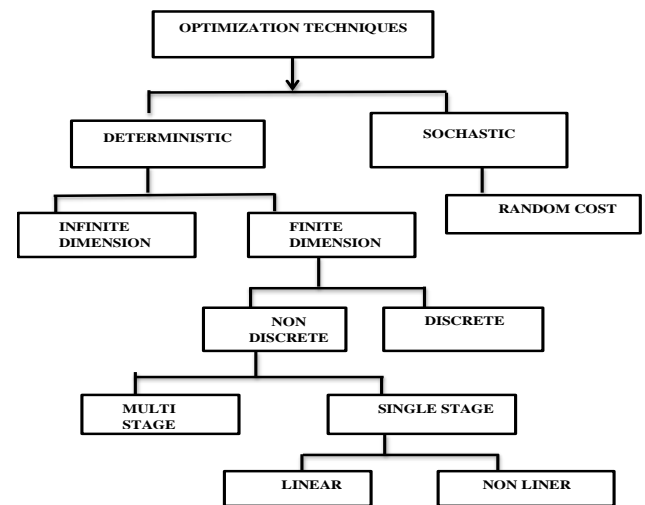
GENETIC ALGORITHM

There are various search techniques, few techniques include calculus based, guided random and enumerative were each of this are classified into different algorithms and methods, one such algorithm is genetic algorithm which is derived from evolutionary algorithm which belongs to guided random search technique.
in 1975 prof.Holland of university of Michigan has publicize that genetic algorithms are search algorithms in which optimization dispose on natural genetics and selection, genetic algorithm can successfully include biology, computer science, image processing, pattern recognition and neural networks. In this paper I want to discuss basic concepts of genetic algorithm

Search space: The vicinity for all viable answers is called search space wherefore it is frequently recognized by the point of fixing a drag, however, search space are frequently complex and one doesn't recognize in which to appear for the solution or in which to start off from and that is frequently in which genetic set of rules is beneficial. Algorithm is commenced with a hard and fast of answers known as populations. From one population the solutions are taken and used to generate new population and this solution are selected based on the fitness value the solutions with highest fitness value have more chances to reproduce. This is rehashed till a few circumstances for improvement of first-class association is fulfilled. Going to the Selection process there are numerous techniques which a a genetic set of rules can use to choose the people to be copied over into the subsequent generation

.

Elitist selection: In every single generation the fittest members are destined to be chosen.

Roulette wheel selection: In this selection is done by designating a fixed point on the wheel. The province of wheel which comes in front of this fixed point is picked out, the fittest individual has largest stake of the roulette wheel and weakest has the smallest stake. Therefore the odds of choosing an individual depends directly on its fitness

Tournament selection: Subgroups of individual are looked over the bigger populace, and individuals from every subgroup contend with each other. In each particular subgroup a single individual is replicated

Rank selection: First and foremost the population is tabulated and later every chromosome is assigned with fitness based on this ranking presently, all the chromosomes get an opportunity to be chosen. In any case, it can prompt more slow assembly, in light of the fact that the best chromosomes don't vary such a great amount from other one

Steady-State Selection: each generation are chosen a couple (great - with high fitness) chromosomes. At that point a few chromosomes are evacuated and the novel offspring is set in their place the rest of population pull through new generation

2. Crossover: is an approach were the individuals are opted to produce the offspring. The widely known sort is single point crossover. A point on both parents' chromosomes is picked randomly, and designated a crossover point. Bits to the right of that point are swapped between the two parent chromosomes. This results in two offspring, each carrying some genetic information from both parents

Chromosome1	11011 | 00100110110
Chromosome2	11011 | 11000011110
Offspring 1	11011 | 11000011110
Offspring 2	11011 | 00100110110

3. Mutation: the new populace are generated after selectin and crossover, were as few are straightforwardly replicated the further are produced by crossover. To assure that the individuals are not identical we consider mutation, the individuals who are selected for the mutation is replaced with a new value and ensures the diversity in the population

BEFORE: 11011100110111
AFTER: 110110001101110

I might want to specify about my past venture in which I utilized Wisconsin breast cancer information for breast cancer classification using KNN (k-closest neighbor calculation) and the precision got by KNN is 95%. KNN is the most effortless supervised machine learning algorithm which will be wont to solve both classification and regression problems.

Let's see a brief review about the project that I have done and steps followed to acquire the accuracy are Data Pre-processing it is used to remove the Outliers once the outliners are removed Applied PCA(Principal component analysis) to shrink the variables and Identify the required Variables and remove the least contributed variables

Next step is splitting the data into train and test datasets in which 70% is for training and 30% is for testing. The next important step is to identify the k which is done by using cross validation score once

the k value is identified we fit the model and later uses Euclidean distance to group the nearest data the next important step is to predict the model using test data once it is done the last step is to find the accuracy using confusion matrix.

Derisma, Meza Silvana, Imelda in 2018 has addressed that neural networks are weak in determining parameter values, to overcome this issue the author proposed to use genetic algorithm for optimizing the neural networks ;as known that genetic algorithm is an optimization method ,this observation leads to the conclusion that genetic based neural network has the increasing rate of accuracy value and decreasing rate of error compared to neural networks alone, they used WBCD and rapid miner for experimentation[1]

II.EXISTING METHOD

Md. Milon Islam, Hasib Iqbal, Md. Rezwanul Haque, and Md. Kamrul Hasanhas in 2017 presented the prediction of breast cancer using supervised machine learning techniques such as support vector machine and K-Nearest Neighbors the dataset used is WBDC data set from UCI machine learning repository they used10-fold cross validation to get accurate outcome. This experiment obtained the accuracy of 98.57% by support vector machine and 97.14% by k-nearest neighbors [2]

Shweta kharya in 2012 addressed multiple data mining techniques suchlike decision tree, naive Bayes classifier , logistic regression ,Bayesian networks, support vector machine on the most commonly used database i.e. WBCD and the outcome suggested that decision tree is the best predictor with 93.62% of accuracy [3]

Mehmet Fatih Akay in 2009 proposed breast cancer diagnosis based on SVM combined with feature selection experiments have been conducted on Wisconsin breast cancer dataset (WBCD) which is widely used among researchers, SVM with feature selection yielded the highest classification accuracy with five features and also sensitivity, specificity, positive and negative predictive values, receiver operating characteristic curves and confusion matrix are used to evaluate the performance. The accuracy obtained for various training- test portions are (98.53% for 50-50%, 99.02% for 70-30%, and 99.51% for 80–20%) [4]

Meriem amrane,saliha oukid,Ikram gagaoua, tolga
ENSAR in 2018 initiated a comparison between two machine learning techniques which provides highest accuracy and effective diagnosis the two classifiers they used are Naive Bayes (NB) classifier and K-nearest neighbor (KNN) the accuracy is evaluated using cross validation the conclusion of this experiment is shown that highest accuracy with 97.51% is obtained by KNN compared to NB classifier with 96.19% [14]

M.Gayathri, C.P.Sumathi, And T.Santhanam in 2013 put down the survey on different machine learning and data mining methods for the classification of breast cancer this survey helps to know about number of papers implemented on breast cancer diagnosis [15]

Dr. K. Usha Rani in 2010 have discussed a uniformity of different parallelization methods like exemplar-parallelism(EP), block parallelism ,neuron parallelism for better classification of breast cancer diagnosis. The experiment conducted is referred as single and multi-layer neural network models.to train the networks author used back propagation algorithm and MLP was embraced to produce an accuracy of 92% [16]

Arpit Bhardwaj, Aruna Tiwari, DharmilChandarana and Darshil Babel in 2014 proposed genetically optimized neural networks (GONN) algorithm in this paper they evolve neural network genetically to optimize its

structure for classification and also introduced new crossover and mutation operation which differ from normal genetic algorithm they used GONN algorithm to classify breast cancer as benign or malignant this algorithm achieved better accuracy than back propagation neural network and a Support Vector Machine [17]

Yasmeen M. George, Bassant Mohamed Elbagoury, Hala H. Zayed and Mohamed I. Roushdy in 2012 used four classification models namely multilayer perceptron (MLP) using back-propagation algorithm, probabilistic neural networks (PNN), learning vector quantization (LVQ) and support vector machine (SVM).to develop an excellent diagnosis system for breast cancer classification, and also to classify whether the tumor is benign or malignant in fine needle aspiration cytology FNAC is the popular component in breast lesion it is very easy to use ,accurate and affordable The optimum network for classification of breast cancer cells was found using probabilistic neural networks. This is followed in order by support vector machine, learning vector quantization and multilayer perceptron. The results showed that the predictive ability of probabilistic neural networks and support vector machine are stronger than the others in all evaluated datasets, the dataset used is Breast Cancer Wisconsin (Diagnostic) dataset using matlab and this experiment has proved that the classification is capable to produce 99.7 % sensitivity and specificity for their datasets. [19]

The table below contains few different methods used for breast cancer diagnosis by using artificial neural networks and data mining approaches

no	Year	Author/r eference	Algorithm technique	Accu racy
1	2009	Mehmet faith akay[4]	SVM with feature selection	99.51 %
2	2012	Yasmeen M.Georg e[19]	MLP,PNN,LVQ,S VM	99.7 % SVM
3	2012	Shweta kharya [3]	ANN,NB,LR,SV M.DECISION TREE	93.62 % DT
4	2014	Arpit bharadw aj[17]	GONN	97.08 %
5	2017	MD.milo n islam[2]	KNN&SVM	98.57 % SVM 95.65 % KNN
6	2018	Meriem &ikram[14]	NAVIE BAYES &KNN	97.51 %KN N 96.19 % NB

III.PROPOSED METHOD

I would like to present a general review on NSGA-II optimization technique. NSGA-II is a depiction of multi objective genetic algorithm(MOGA) around the timeline of 1993-1995 a number of EAs are propound to solve multiobjective optimization problems, NSGA, NPGA, MOGA are the algorithms which demonstrated the necessary additional operators for converting simple EA to MOEA.NSGA-II is part of MOGA which is well liked and definitive as observed ANN, GP, Fuzzy, regression analysis and some other methods are used to predict the machine performance were as coming to NSGA-II it is used to meet the requirements of the machining process in finding set of solutions based on combination of suitable variables. The common features that included in NSGA, NPGA, and MOGA are fitness of population is allocated based on nondominated sorting and maintaining the diversity among solutions
The primary motivation behind my task is to discover the exactness gotten by NSGA-II and neural systems and furthermore to see whether the shortcoming of hereditary calculation for example (calculation time) will decreased or not.

VI.RELATED WORK: RESULTS FOR PSEUDO CODE
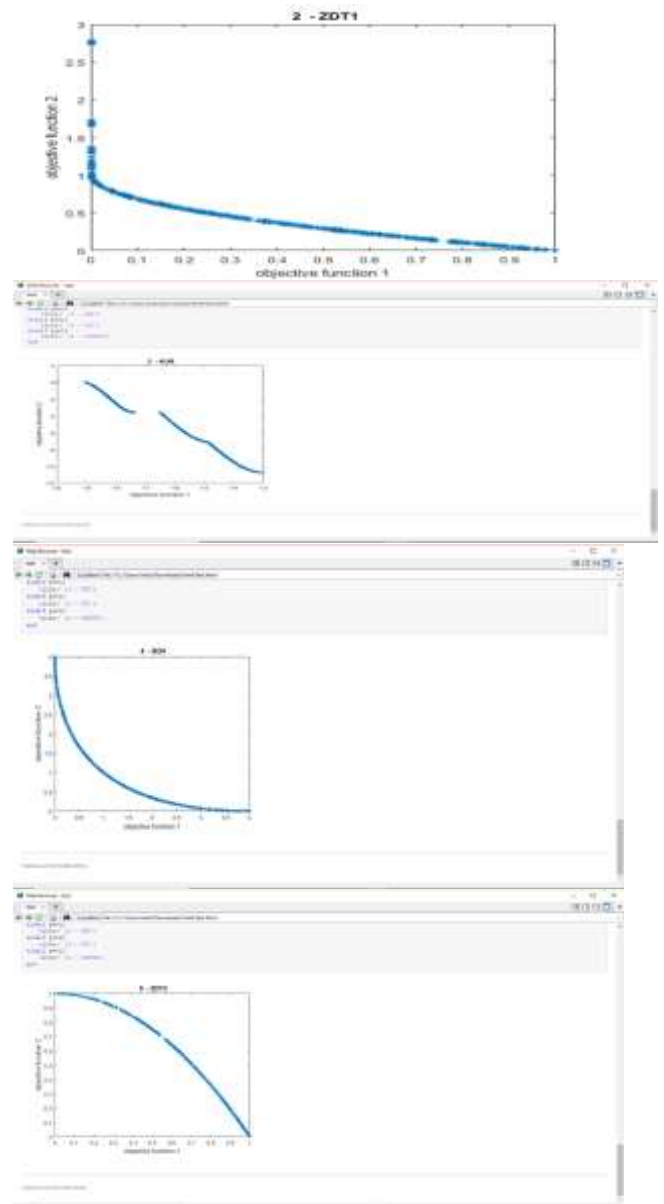
V.CONCLUSION

As the genetic algorithm is proved as capable to enhance the predication accuracy on carcinoma disease by this literature survey we have acknowledged about various techniques, methods and frameworks associated with deadly problem of breast cancer we tried the dataset with simple KNN approach and found the accuracy acquired is 95%.

REFERENCES

[1]Derisma, Meza Silvana, Imelda (2018), Optimization of Neural Network with Genetic Algorithm for Breast Cancer Classification, International Conference on Information Technology Systems and Innovation, 978-1-5386-5692-1, 22-25

[2]Md. Milon Islam, Hasib Iqbal, Md. Rezwanul Haque, and Md. Kamrul Hasan (2017). Prediction of Breast Cancer Using Support Vector Machine and K-Nearest Neighbors, IEEE Region 10 Humanitarian Technology Conference, 21 – 23

[3]Shweta Kharya (2012)"Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease" International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2,

[4]Mehmet Fatih Akay (2009), support vector machine and with feature selection for breast cancer diagnosis, Experts system with applications 36, 3240-3247

[5]Aalaei,S.,Shahraki,H., Rowhanimanesh, A., & Eslami, S. (2016). Feature selection using a genetic algorithm for breast cancer diagnosis: Experiment on three different datasets. Iranian Journal of Basic Medical Sciences, 19(5), 476-482.

[6]Asri, H., Mousannif; H., A1 Moatassime, H., & Noel, T. (2016). Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. Procedia Computer Science, 83(Fams), 1064-1069.

[7] Ch. Shravya, K. Pravalika, Shaik Subhani(2019), Prediction of Breast Cancer Using Supervised Machine Learning Techniques, International Journal of Innovative Technology and Exploring Engineering, 2278-3075, Volume-8 Issue-6

[8]N. Srinivas and K. Deb (1994), "Muiltiobjective Optimization Using Nondominated Sorting in Genetic Algorithms," Evol. Comput. vol. 2, no. 3, pp. 221--248

[9]A. Zhou, B.-Y. Qu, H. Li, S.-Z. Zhao, P. N. Suganthan, and Q. Zhang (2011), "Multiobjective evolutionary algorithms: A survey of the state of the art," Swarm Evol. Comput. vol. 1, no. 1, pp. 32–49,

[10]E.Zitzler and L.Thiele (1998), "An evolutionary algorithm for multiobjective optimization: The strength Pareto approach," TIKRep, vol. 43, no. 43.

[11]K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan (2002), "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II," IEEE Trans. Evol. Comput. vol. 6, no. 2, pp. 182–197,

[12]Kalyanmoy Deb (2002) "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II" IEEE Transactions on Evolutionary Computation, vol. 6, no. 2

[13] Farid Jauharia , Wayan Firdaus Mahmudyb , Achmad Basukic Multi-Objective Optimization For Proportional Tuition Fees Assessment Using Non-Dominated Sorting Genetic Algorithm II (NSGA II)

[14]Meriem AMRANE,saliha OUKID,Ikram GAGAOUA,tolga ENSAR(2018) , "Breast Cancer Classification Using Machine Learning",78-1-5386-5153-3/18

[15]M.Gayathri, C.P.Sumathi,And T.Santhanam (2013)" Breast Cancer Diagnosis Using Machine Learning Algorithms –A Survey" International Journal Of Distributed And Parallel Systems (IJDPS) Vol.4, No.3

[16]. Dr. K. Usha Rani(2010) "Parallel Approach for Diagnosis of Breast Cancer using Neural Network Technique" International Journal of Computer Applications International Journal of Computer Applications (0975 – 8887), Volume 10– No.3

[17]Arpit Bhardwaj, Aruna Tiwari, Dharmil Chandarana and Darshil Babel(2014)"A Genetically Optimized Neural Network for Classification of Breast Cancer Disease",978-1-4799-5838-2/14

[18]Kemal Polat , Salih Güneş(2007),"Breast cancer diagnosis using least square support vector machine",Digital Signal Processing 17 694-701

[19]. Yasmeen M. George, Bassant Mohamed Elbagoury, Hala H. Zayed and Mohamed I. Roushdy(2012) ,"Breast Fine Needle Tumor Classification using Neural Networks"- 2012- International Journal of Computer Science IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 2

[20]Basavaraj Hiremath & SC Prasannakumar(2015) "Automated Evaluation of Breast Cancer Detection Using SVM Classifier"(IJCSEITR) ISSN (P): 2249-6831; ISSN(E): 2249-7943 ,Vol. 5, Issue 1, 11-20