3rd National Conference on Green Technology and Science for Sustainable Development
© 2020 IJSRST | Volume 5 | Issue 6 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
International Journal of Scientific Research in Science and Technology

Genetic Algorithm for Detection of Cancer

Sujata R. Jadhao, Akangsha Gawai, Prof. Chhajed
Information Technology, Anuradha Engineering college, Chikhli, Maharashtra, India

ABSTRACT

One of the major disease is cancer. Several deaths occur due to overdue of detection of this disease. At present, several researchers trying to find the appropriate system to detect the diseases early, so as to find medical treatment of it. The Genetic Algorithm(GA) is one of the optimization method. The first part of this shebang briefly traces their theory , explain essential concept and deliberate theoretical aspects. The second part center at an detailed implementation of GA. Its discuss the population, selection, crossover and mutation also implementation of image in GA to detect cancer tumor. Genetic Algorithm is used to produce new generation by using crossover operation and also to do some modification on new generation. It gives good detection rate.

Keywords : Cancer, Genetic Algorithm, Crossover, Mutation.

I. INTRODUCTION

Cancer is one of the most risky diseases in the world.
Once of the leading causes of death is Cancer.
In this section introduced the Genetic Algorithm relative to Chromosome, Initial Population, Fitness Function and Genetic Operators.
Correct Diagnosis and early detection of Cancer can increase Survival rate of Patient
Finally, in the last section presented the Conclusion and Direction for future work.

II. CANCER

Cancer is the uncontrolled growth of abnormal cells anywhere in the body. Cancer is defined as a group of diseases in which the cells of body outset Uncontrolled and Abnormal growth of cells. A cancerous tumor is malignant[1].

Malignant means it can grow and spread to other parts of the body also. These abnormal cells are denomination cancer cells, malignant cells, or tumor cells. These cells can to trickle normal body tissues. Many cancers and the abnormal cells that reconcile the cancer tissue are forrader identified by the name of the tissue that the abnormal cells originated from (for example, breast cancer, lung cancer). Cancer is not narrow to humans; animals and other living organisms can get cancer. Below is a schematic that shows normal cell partition and how when a cell is damaged or altered without repair to its system, the cell usually dies. Also shown is what occurs when such detriment or unrepaired cells do not die and become cancer cells and show uncontrolled partition and growth a mass of cancer cells exploitation. Anything that may cause a normal body cell to develop singularly on the postcard can cause cancer; general categories of cancer-related or causative agents are as follows: chemical or toxic blend exposures, ionizing radiation, some pathogens, and human genetics.

A. How Cancer Grow and Spreads?

Each gene contain fix of instructions that telling to cell what function to perform as well as how grow and division Frequently, cancer cells can break away from this original mass of closet, travel through the blood and lymph systems, and lodge in other organs where they can again repeat the uncontrolled growth. The cancer type or more limited is the cancer in the body. Scaffold methods differ from cancer to cancer and need to be individually discussed with your health care provider.

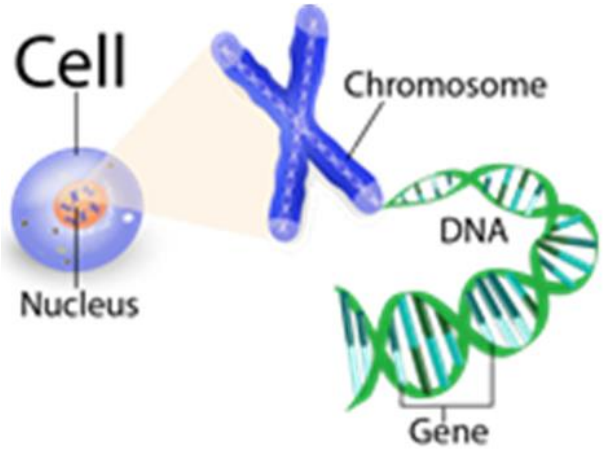

Fig 1. Development of Cancer

Treatment etiquette vary according to the type and stage of the cancer. Most treatment etiquette are designed to fit the individual patient's disease. However, most treatments include at least one of the following and may include all: surgery, chemotherapy, and radiation therapy[5].

There are many medication for cancers but patients are strongly advive to discuss these before use with their cancer doctors. As cancer cells partition, a tumour will develop and grow. Cancer cells have the same needs as usual cells. They need a blood supply to synchronize oxygen and nutrients to grow and survive. When a tumour is very small, it can easily grow, and it get oxygen and nutrients from nearly blood vessels. But as a tumour accrue, it needs more blood to bring oxygen and other nutrients to the cancer cells. So cancer cells dispatch signals for a tumour to make new blood vessels.

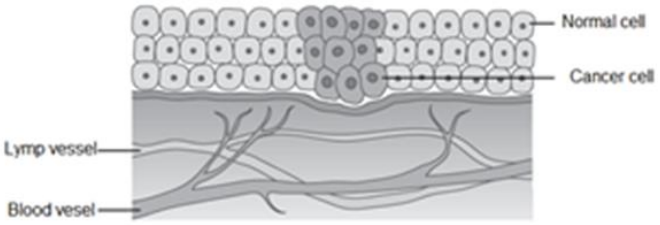

Fig 2 . Sample tissue of normal cell

As a tumour gets bigger, cancer cells can spread to surrounding tissues and structure by pushing on normal tissue beside the tumours. Cancer cells also spread from where it first started to other parts of the body.

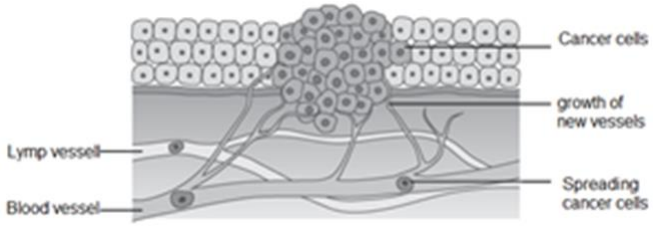

Fig 3. Sample tissue of spreading of cancer

B. Cancer Symptoms and Causes

Cancer symptoms and signs depend on the specific type and grade of cancer.

Although there are many tests to screen and presumptively diagnose cancer, a biopsy sample of suspected cancer tissue.

- ✓ Fatigue
- ✓ Abnormal bleeding
- ✓ Prolong Cough
- ✓ Weight Changes (may loss or gain of weigh)
- ✓ Skin Changes (may redness, yellowness or also be darking)
- ✓ Alcohol

Lump (thickening of skin)

The following is major causes and is not all-inclusive as specific causes are routinely added as research advances:

Chemical or toxic compound exposures: Benzene, nickel tobacco or cigarette smoke.

Ionizing radiation: Uranium, ultraviolet rays from sunlight, radiation from alpha, beta, gamma, and X-ray-emitting sources.

Pathogens: Human papillomavirus (HPV), hepatitis viruses B and C.

Genetics: A number of specific cancers have been linked to human genes and are as follows: breast, ovarian, colorectal, prostate, skin and melanoma; the specific genes and other details are beyond the scope of this general article so the reader is referred to the National Cancer Institute for more details about genetics and cancer [3].

It is important to point out that most everyone has risk factors for cancer and is exposed to cancer-causing substances (for example, sunlight, secondary cigarette smoke, and X-rays) during their lifetime, but many individuals do not develop cancer. In addition, many people have the genes that are linked to cancer but do not develop it.

III. GENETIC ALGORITHM

GA belongs to Optimization Technique based on the principle of Genetics and Natural Selection. GA is based on Fitness Function [10].

Genetic Algorithm is used as a method of gene selection is base for SVM and ANN to classify Cancer status of patient. Genetic Algorithm is used as a method of ensign (genes) selection for the support vector machine and artificial neural network to classify cancer status of a patient. In a system is developed to test the cancer risk level. Classifiers namely Support Vector Machine (SVM), Artificial Neural Network (ANN) applied for the detection of cancer to find the severity of disease.

A genetic algorithm is mimicry the process of natural process used to generate useful solutions to optimization and search problems.

Holland method considers mutation and also utilizes genetic recombination that is crossover to find the optimum solution. Crossover and mutation are two basic parts of operators of Genetic Algorithm. The Performance of Genetic Algorithm very depend on them.

a) Optimization

Optimization is the process of making something better output.In this process taking of Set of Input for making of particular Process on the set of inputs individual population and then using optimization produses the Set of Output.

Optimization method used to finding to the input problems in such a way that to get “best” output to that problem. Genetic Algorithm is a heuristic search and optimization technique based on principle present in natural evolution.

b) Fitness Function

The Fitness Function produces the next generation of states.

Fitness Function simply define as a function which considers as possible solution to the problem as input and produces some output. The fitness function simply counts the numbers of matches between the target phrase and the give chromosome. The fitness function determine how fit an individual is. It gives a fitness Score to each individual. The probability that an individual will be selected for reproduction is based on its fitness function score.

Only those individual for possible solution can be selected which are fit enough.

Following is the formal wording of the fitness function used to evaluate each chromosome:

Fitness function formula

where the size_of_chromosome is the count of genes that are used as biomarkers in the chromosome and mean_size_of_initial_population is the average number of genes that are used as biomarkers in the starting population of the GA, across all chromosomes. This special form of fitness function pushes GA to select solutions that contain a small number of genes and are as accurate as possible. We performance with different weights in order to boost one metric over the other but we obtained similar results, just with slower convergence. The F-measure metric has been preferred as it is geometric average of the precision and recall, both being of great interest when designing diagnostics tests

c) Natural Selection

Natural Selection is the process that results in the Adaptation of individual with its Environment, by means of selectively reproducing changes in its Genotype or Genetic constitution.

A Genetic Algorithm is a search heuristic that is backbone by Charles Darwin's theory of natural evolution.

The processes of natural choice strarts with the selection of fittest individuals from population (That is the current generation taken to reproducing new generation). They produce offspring which inherit the characteristics of the patents and will be added to the next generation. If parents have preferable fitness, their offspring will be better than parents and have a better chance at surviving. This process keep a generation with the fittest individuals will be found.

This opiniomn can be applied for a search problem. We consider a set of solutions for problem and select the set of best ones out of them.

Five phases are considered in a genetic algorithm.

1. Initial population
2. Fitness function
3. Selection
4. Crossover
5. Mutation

IV. GENETIC ALGORITHM REPRODUCTION

Reproduction in Genetic Algorithm is done by applying the Genetic Algorithm Sophisticated Operators. An individual is characterized by a set of parameters (variables) known as gene. Gene are joined into a string to form a Chromosome (solution).

In genetic algorithm, the set of genes of an individual is represented using a string, in terms of an alphabet. Usually, binary significance are used (String of 1s and 0s).

1) Selection
2) Crossover
3) Mutation

A. Selection: The intellection of selection phase is to select the fittest individuals and let them pass the genes to the next generation. Selection of current generation to reproduce the next generation is used. They are choice based on their fitness function score. Two pairs are select at randomly to reproduction. Two pairs of individuals (parents) are selected based on their fitness score. Individuals with high fitness have more chance to be selected for reproduction in the genetic algorithm [10].

B. Crossover: The crossover process involved more than one parent and produce one or more offspring using genetic materials of parents. Crossover is the most effect phase in a genetic algorithm.

a. Single point crossover –

In this container one crossover possition is choiced, binary string from beginning of chromosome to the crossover point is copied from one parent, and the rest is copied from the second parent 11001011+11011111 = 11001111

b. Two point crossover –

Here two crossover section are choiced, binary string from starting of chromosome to the first crossover point is replicated from one parent, the part from the first to the second crossover point is replicated from

the second parent and the remaining is replicated from the starting point of parent 11001011 + 11011111 = 11011111

c. Uniform crossover –

In this method bits are randomly replicated from the starting or from the second parent 11001011 + 11011101 = 11011111. For each and every pair of parents to be overpower, a crossover point is chosen at random from within the genes. Offsprings (new generation) are created by exchanging the genes of parents among themselves until the crossover point is reached. The new offspring are added to the population of the next process of genetic algorithm that the mutation.

C. Mutation: In certain new offspring formed, some of their genes can be subjected to a mutation with a low random probability. This implies that some of the bits in the bit string can be spasm. In Mutation happens that have features (gene) which are present in children but are not present in parents. Mutation occurs to maintain diversity within the population and prevent premature convergence [10].

D. Termination: The Genetic Algorithm finished if the population has converged (does not produce offspring which are significantly different from the previous generation). Then it is called that the genetic algorithm has provided a set of solutions to our problem.

V. FLOWCHART OF GENETIC ALGORITHM

The existing Genetic Algorithm are founded upon the following main principles:

Reproduction

Fitness

Crossover

Mutation

$$Fitness\_function = 1+F\_measure - \frac{Size\_of\_chromosome}{mean\_size\_of\_initial\_population}$$

Fig 4. Genetic Algorithm Flowchart

There are many flavours of GAs in circulation, changing in implementation of these three parameters, but in essence the algorithms [6] all follow a standard procedure:

1. Start with a randomly generated population of strings (candidate solutions to a problem). (These "solutions" are not to be confutation with "answers" to the quetion, think of them as possible characteristics that the system would employ in order to reach the answer).
2. Calculate the fitness function of each string in the population.
3. Repeat the following steps pending n new strings have been created:

• Select a pair of parent strings from the current population, the technique of selection being an increasing function of fitness. choice is done "with replacement" meaning that the same string can be selected more than once to become a parent.

• With the crossover processes, cross over the pair at a randomly chosen point to form two new strings. If no crossover occures, form two new strings that are exact copies of their respective parents.

• Mutate the two new strings at each locus with the mutation process, and place the resulting strings in the new population.

4. Changes the current population with the new population.
5. Go to step 2.

VI. FUTURE

Future Plane can optimize of classification process of tissue sample by using Filling Missing gene into data list Base on optimization process can find out a new Genetic algorithm that enable to early detection and higher accuracy percentage.

Genetic Algorithm have linger for above fourty years already, and they are still go-to algorithm in engineering and optimization problems in which computational time isn't a big issue. Genetic

Algorithm provide global solution to optimization problems, which aren't easy to come by in the swivel or optimization algorithms. New developments and circulation of genetic algorithms are taking basic these days, and assume that they will continue to be refined and explored to improve compute times and accuracy. The Genetic Algorithm and their extended cousins are going to play an important role in optimization for years to come. Quantum computing is likely to speed up their adoption in machine learning and real-time algorithm, as they become more computationally feasible [8].

VII. IMAGE CLASSIFICATION STAGES (IMPLEMENT GA TO DETECT CANCER TUMOUR

Image Classification Stages
1. Preprocessing
2. Clustering
3. Genetic Algorithm based classification

The image can be any jpg, or other any other valid medical effigy format. Once the input is taken from the management, it can have some impurities respective to colorization, brightness etc. The first work is to convert this image to a normalized image. Here the approximate image means converting the image according to some defined standard. Some enhancements over the image is performed to the grayscale image during this phase. As the preprocessing phase get completed, the next work is to define this image as the initial population set. Now in the second phase, the clustering algorithm is been implemented over the image to perform a certain classification. The work of clustering is to identify the image areas that can have maximum chances of tumor. In this work, we have used a fuzzy C means clustering. According to this clustering approach, we initially defines the number of clusters along with center of these clusters. Once the clusters are defined, the next work is to identify the euclidean distance of each pixel from these center points. According to these distance measure, the pixels are placed in the specific clusters. After the clustering process, we will consider the cluster that represents the center area of the image as the new population set. As this area has the maximum chances of tumor occurrence. In the final stage, the genetic is implemented on this population set. The genetic process begins with some input specification in terms of population set and the number of iterations processed by the algorithm. After these all specification, the genetics is initiated and is processed by the algorithm by its continuous stages of selection, crossover, mutation etc. The selection stage is about the selection any two random pixels for the comparative analysis. On this pixels, the crossover is been performed to select the next elected pixel and it is followed by the mutation process as the election or the rejection of the particular pixel. It can also perform some changes if required. As the genetics process is finished, it will return a valid threshold value respective to which the decision regarding the pixel selection as the tumor area is been performed. This selected pixel area is presented as detected tumor in the brain image[13].

VIII. CONCLUSION

Cancer is one of the most risky diseases in the world. Accurate Diagnosis and early detection of Cancer can raise the survival rate. The system includes feeding the Genetic Algorithm with an initial selection of population and then selects the best choice to generate the new springer by using crossover operation. Then some mutation enables management to do some development on offspring. The future plans we can optimize of the classification process of tissue samples by using filling missing gene data list. Foundation on optimisation process we can find out a new genetic algorithm that enables us to early detection and higher accuracy percentage. In addition, the algorithm provides a list of play the most important role in this disease as a by-product. However, due to stochastic

nature of the Genetic Algorithm, we decided to statistically measure the importance of the genes by measuring their appearances on “good” solutions generated by the Genetic Algorithm in all our experiments, which is also suggested procedure if one should use the method. Results were verified by the biological makes of these genes in specific biological beano in human organism. The image classification stages implement genetic algorithm to detect cancer tumour.

IX. REFERENCES

[1]. M. V. A. Gajdhane and L. Deshpande, "Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques."

[2]. O. Abdoun, J. Abouchabaka, and C. Tajani, "Analyzing the Performance of Mutation Operators to Solve the Travelling Salesman Problem," arXiv preprint arXiv:1203.3099, 2012.

[3]. S. Peng, Q. Xu, X. B. Ling, X. Peng, W. Du, and L. Chen, "Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines," FEBS letters, vol. 555, pp. 358-362, 2003.

[4]. W. F. Baile, R. Buckman, R. Lenzi, G. Glober, E. A. Beale, and A. P. Kudelka, "SPIKES—a six-step protocol for delivering bad news: application to the patient with cancer," The oncologist, vol. 5, pp. 302-311, 2000.

[5]. M. S. AL-TARAWNEH, "Lung Cancer Detection Using Image Processing Techniques," Leonardo Electronic Journal of Practices and Technologies, vol. 11, pp. 147-58, 2012.

[6]. J. Schneider, N. Bitterlich, H.-G. Velcovsky, H. Morr, N. Katz, and E. Eigenbrodt, "Fuzzy logic-based tumor-marker profiles improved sensitivity in the diagnosis of lung cancer," International journal of clinical oncology, vol. 7, pp. 145-151, 2002.

[7]. A. Giatromanolaki, M. Koukourakis, E. Sivridis, H. Turley, K. Talks, F. Pezzella, K. Gatter, and A. Harris, "Relation of hypoxia inducible factor 1α and 2α in operable non-small cell lung cancer to angiogenic/molecular profile of tumours and survival," British journal of cancer, vol. 85, p. 881, 2001

[8]. Banzhaf, Wolfgang; Nordin, Peter; Keller, Robert; Francone, Frank,"Genetic Programming: An Introduction, Morgan Kaufmann, San Francisco, CA, 1998.

[9]. S. Thrun. Learning to Play the Game of Chess. In G. Tesauro, D. Touretzky, and T. Leen, editors, Advances in Neural Information Processing Systems (NIPS) 7, Cambridge, MA, 1995. MIT Press. [

[10]. J.H. Holland (1975),"Adaptation in Natural and Artificial Systems", University of Michigan Press, Ann Arbor, Michigan; re- issued by MIT Press 1992.

[11]. D.E. Goldberg,"Optimal initial population size for binarycoded genetic algorithms.TCGA Report 85001, University of Alabama, Tuscaloosa, 1985.

[12]. K.A. De Jong,"An analysis of the behavior of a class of genetic adaptive systems", Doctoral dissertation, University of Michigan, Ann Arbor, Michigan, 1975.

[13]. H.J. Bremermann, J. Rogson, S. Salaff,"Global properties of evolution processes", In H.H. Pattee (ed.), Natural Automata and Useful Simulations, pp. 3–42, 1964.

[14]. Lorenzen P., Joshi S., Gerig G., Bullitt E.,“Tumor-Induced Structural and Radiometric Asymmetry in Brain Images”, Proc the IEEE workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), 1, pp. 488-501, 2001.

[15]. N. Nandha Gopal , Dr. M. Karnan,“Diagnose Brain Tumor Through MRI Using Image Processing Clustering Algorithms Such As Fuzzy C Means Along With Intelligent Optimization Techniques”, 2010 IEEE.