International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

A Survey on Cancer in Chhattisgarh Using Fuzzy Logic and Genetic Algorithm

Indrani Bhattacharya[1], Sana Tak[2]
[1]M.Tech Student, Kalinga University, Kotni, Naya Raipur, Chhattisgarh, India
[2]Assisstant Professor, Kalinga University, Kotni, Naya Raipur, Chhattisgarh, India

ABSTRACT

Cancer is the most common cause of death these days. Although, it is no longer incurable. With the help of new technologies, new medicines, immunization therapy, with the help of ayurvedic medicines, and proper food, Cancer can be cured if detected in early stage. Although, Cancer is genetical and in some cases hereditary, but still environmental factors are also responsible for causing Cancer. Working environment, exposure to chemicals for certain period of time, food habits, lifestyle increases the chances of cancer. This paper is written to study the different environmental reasons that might cause cancer in the people of Chhattisgarh. To make the study we have used Fuzzy Logic and Genetic Algorithm.

Keywords: Chhattisgarh, Women, Artificial Intelligence, Algorithm.

I. INTRODUCTION

Chhattisgarh is a newly formed State. It was formed on November 1, 2000. Chhattisgarh is very rich in mining coal and steel industries. Although it increases the job opportunity for the people who are living near these industries. But along with that comes lot of harmfull Cancerous diseases because of those toxic gases released by those Industries. Specially people living in Villages like Sarasmal, Kosampali, Dongamahua, Libra, Kodkel and surrounding villages in the Raigarh district of Chhattisgarh. Bhilai, Urla, korba, Raigarh, Durg, Siltara and bilaspur are the main regions where the the industrial areas have emerged. Wastes coming out from these Industries affect the environment. Chhattisgarh has so many rivers and all of them, some them are flowing through these Industrial areas as well.

To ascertain the presence of these Chemicals and in what amount they are present in their daily life and till what time we have used the techniques of Fuzzy Logic and Genetic Algorithm.

Fuzzy Logic, is a way giving instruction to computer and make it work according to human thinking.

Genetic Algorithm makes the Optimization process further clear and more accurate.

An overview of the pollutants found in air, water, soil, fly ash and sediment samples of Raigarh is as shown below :-

Table 1

No.	**Sample**	**No of samples**	**Analyzed for**	**Results**
1.	Air	4	PM2.5 Toxic heavy metals (Arsenic; Lead; Nickel; Manganese; Silicon; Aluminium; Calcium; Cadmium)	The aggregate levels of toxic particles in PM2.5 in Sarasmal, Kosampali and Dongamahua are higher than the permissible standard set by the Indian MoEF standard. PM2.5 in one location (Sakta Sitapur) adjacent to the above three villages is strikingly high and exceeds all standards – WHO, USEPA and Indian Standard. Toxic metals like Arsenic, Manganese, Nickel and Silicon all exceed the permissible standards.
2.	Water	7	Aluminum; Arsenic; Boron; Cadmium; Chromium; Lead; Manganese; Selenium; Total Dissolved Solids	All toxic metals have been found in the samples; Arsenic and Manganese levels are strikingly high. The levels of Aluminum, Boron, Cadmium and Selenium – all are above permissible standards.
3.	Soil	9	All above; in addition, Antimony and Vanadium.	All are above permissible standards – Vanadium, Chromium and Nickel in almost all samples, followed by Arsenic, Antimony, Cadmium and Lead.
4	Fly ash	2	Aluminum; Iron; Titanium; Zinc; also, Calcium and Magnesium.	Aluminum; Iron; Titanium; Zinc levels are substantially high in soil due to fly ash deposit; Calcium and Magnesium are also found to be high but in variable

				amounts
5	Sediment	6	Arsenic; Cadmium; Chromium; Lead; and, Zinc	Chromium found in all samples; Arsenic, Cadmium in most samples and are exceed the permissible levels.

Health and safety are the most neglected concerns in Indian steel sector. Blast furnace gas is produced at a melted iron-making unit called blast furnace. This waste gas is poisonous, containing a mixture of carbon monoxide (26 per cent) and hydrogen (4 per cent); nitrogen and carbon dioxide make up the rest.

II. CLIMATE AND HUMAN HEALTH

2.1 Health Impacts

- Increased exposure to toxic chemicals, known or suspected to cause cancer, that are released into the environment following heavy rainfall or flooding and by increased volatilization of chemicals under conditions of increased temperature
- Depletion of stratospheric ozone leads to an increase in UV exposure and temperature, increasing the risk of skin cancer and cataracts. Alternatively, an increase in exposure to UV radiation can lead to elevated levels of Vitamin D, which has been associated with a decreased risk of some types of cancer
- A decline in air quality and rise in concentrations of certain air pollutants increases the risk of lung cancer.

2.2 Mitigation and Adaptation

- Reducing greenhouse gas emissions and other hazardous air pollutants, through energy efficient power generation, lower vehicle miles traveled, and efficient industrial processes can decrease toxic outputs of fossil fuel-based power generation and transportation, such as sulfur oxide, nitrous oxides, heavy metals, and particulate matter
- Increasing and preserving green space in urban settings
- Exploring the use of new technologies that decrease greenhouse gas emissions, such as nanotechnology, biofuels, electronic vehicles, and solar cells. However, it is important to note that the impacts of some mitigation technologies have not been fully explored yet and may have unintended negative health consequences
- Employing sun safety actions, such as increased use of sunscreen and staying covered up in the sun, can reduce the risk of cancer from an increase in UV radiation.

2.3 Research Needs

- Using animal cancer surveillance and investigations as sentinel biomedical models to better understand the environmental factors, mechanisms, and pathways of mammalian cancer risk
- Understanding the impact of increased heavy precipitation and flooding events on the risk of toxic contamination of the environment from storage facilities or runoff from land containing toxic chemicals, including the geographical areas, ecosystems, and populations most likely to be

impacted and the health outcomes that could result

- Understanding how climate changes such as changes in temperature and precipitation affect exposure to toxic chemicals including volatile and semi-volatile compounds and known or suspected human carcinogens
- Elucidating the effects of ambient temperature on UV radiation- induced skin cancers, including the amplification of non- melanoma skin cancers
- Evaluating the potential cancer risks through the entire lifecycle of biofuel production, including risks from novel air pollutants and changes in agricultural practices that may increase exposures to pesticides, herbicides, and other environmental contaminants
- Understanding cancer risks from the lifecycle emissions of carcinogens and untested compounds associated with alternative energy and transportation technologies, particularly electricity storage systems and photovoltaic systems
- Clarifying the lifecycle cancer risks of nuclear energy radiation, including through occupational and environmental exposures
- Developing mechanisms to conserve and explore marine and terrestrial biodiversity in environments likely to yield cancer cures and treatments
- Characterizing and quantifying changes in cancer rates from implementation of specific greenhouse gas mitigation strategies, especially for existing fossil fuel-based energy production and use.

III. TECHNIQUES USED

3.1 FUZZY LOGIC:-

Fuzzy logic starts with and builds on a set of user-supplied human language rules. The fuzzy systems convert these rules to their mathematical equivalents. This simplifies the job of the system designer and the computer, and results in much more accurate representations of the way systems behave in the real world.

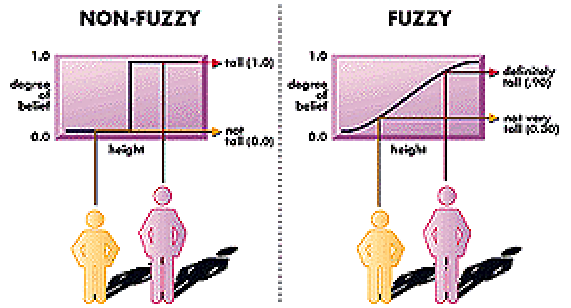
Additional benefits of fuzzy logic include its simplicity and its flexibility. Fuzzy logic can handle problems with imprecise and incomplete data, and it can model nonlinear functions of arbitrary complexity. "If you don't have a good plant model, or if the system is changing, then fuzzy will produce a better solution than conventional control techniques," says Bob Varley, a Senior Systems Engineer at Harris Corp., an aerospace company in Palm Bay, Florida.

You can create a fuzzy system to match any set of input-output data. The Fuzzy Logic Toolbox makes this particularly easy by supplying adaptive techniques such as adaptive neuro-fuzzy inference systems (ANFIS) and fuzzy subtractive clustering.

Fuzzy logic models, called fuzzy inference systems, consist of a number of conditional "if-then" rules. For the designer who understands the system, these rules are easy to write, and as many rules as necessary can be supplied to describe the system adequately (although typically only a moderate number of rules are needed).

In fuzzy logic, unlike standard conditional logic, the truth of any statement is a matter of degree. (How cold is it? How high should we set the heat?) We are familiar with inference rules of the form p -> q (p implies q). With fuzzy logic, it's possible to say (.5* p) -> (.5 * q). For example, for the rule if (weather is cold) then (heat is on), both variables, cold and on, map to ranges of values. Fuzzy inference systems rely on membership functions to explain to the computer how to calculate the correct value between 0 and 1. The degree to which any fuzzy statement is true is denoted by a value between 0 and 1.

Not only do the rule-based approach and flexible membership function scheme make fuzzy systems straightforward to create, but they also simplify the design of systems and ensure that you can easily update and maintain the system over time.



3.1.1 Fuzzy Sets

Fuzzy Set Theory was formalized by Professor Lofti Zadeh at the University of California in 1965. What Zadeh proposed is very much a paradigm shift that first gained acceptance in the Far East and its successful application has ensured its adoption around the world.

A paradigm is a set of rules and regulations which defines boundaries and tells us what to do to be successful in solving problems within these boundaries. For example the use of transistors instead of vacuum tubes is a paradigm shift - likewise the development of Fuzzy Set Theory from conventional bivalent set theory is a paradigm shift.

Bivalent Set Theory can be somewhat limiting if we wish to describe a 'humanistic' problem mathematically. For example, Fig 1 below illustrates bivalent sets to characterise the temperature of a room.

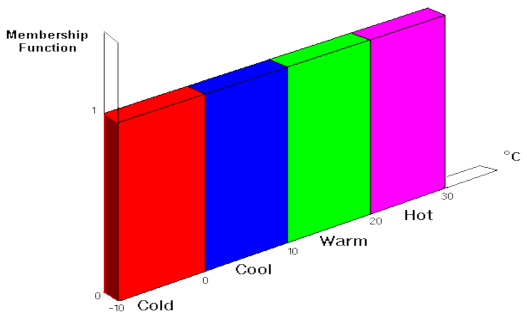


Fig. 1 : Bivalent Sets to Characterize the Temp. of a room.

The most obvious limiting feature of bivalent sets that can be seen clearly from the diagram is that they are mutually exclusive - it is not possible to have membership of more than one set (opinion would widely vary as to whether 50 degrees Fahrenheit is 'cold' or 'cool' hence the expert knowledge we need to define our system is mathematically at odds with the humanistic world). Clearly, it is not accurate to define a transition from a quantity such as 'warm' to 'hot' by the application of one degree Fahrenheit of heat. In the real world a smooth (unnoticeable) drift from warm to hot would occur.

This natural phenomenon can be described more accurately by Fuzzy Set Theory. Fig.2 below shows how fuzzy sets quantifying the same information can describe this natural drift.

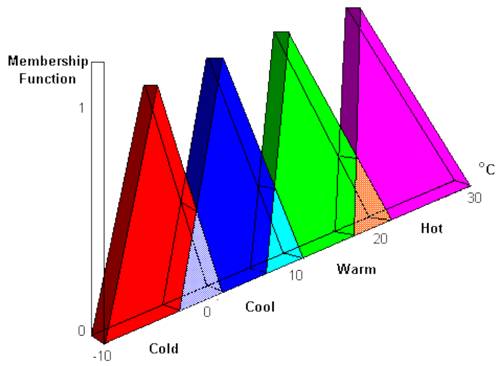


Fig. 2 - Fuzzy Sets to characterize the Temp. of a room.

3.1.2 Fuzzy Set Operations

Definitions.

Universe of Discourse

The Universe of Discourse is the range of all possible values for an input to a fuzzy system.

Fuzzy Set

A Fuzzy Set is any set that allows its members to have different grades of membership (membership function) in the interval [0,1].

Support

The Support of a fuzzy set F is the crisp set of all points in the Universe of Discourse U such that the membership function of F is non-zero.

Crossover point

The Crossover point of a fuzzy set is the element in U at which its membership function is 0.5.

Fuzzy Singleton

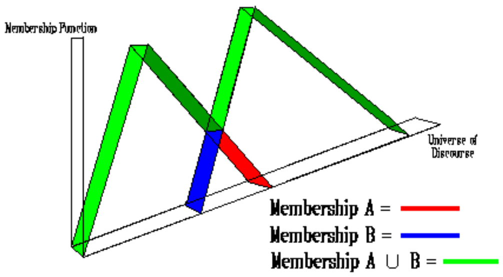
A Fuzzy singleton is a fuzzy set whose support is a single point in U with a membership function of one.

Fuzzy Set Operations.

1) **Union**

The membership function of the Union of two fuzzy sets A and B with membership functions μ_A and μ_B respectively is defined as the maximum of the two individual membership functions

$$\mu_{A\cup B} = \max(\mu_A, \mu_B)$$

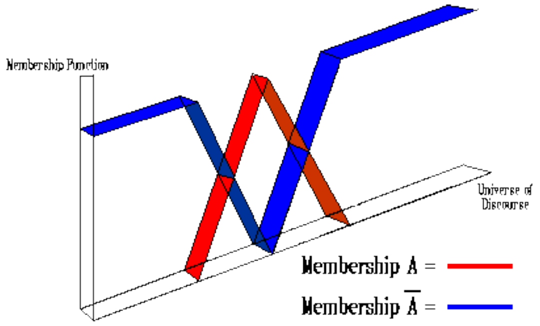


The Union operation in Fuzzy set theory is the equivalent of the OR operation in Boolean algebra.

2) **Complement**

The membership function of the Complement of a Fuzzy set A with membership function μ_A is defined as

$$\mu_{\overline{A}} = 1 - \mu_A$$



The following rules which are common in classical set theory also apply to Fuzzy set theory.

3) **De Morgans law**

$$\overline{(A\cap B)} = \overline{A}\cap\overline{B} \,,\; \overline{(A\cup B)} = \overline{A}\cap\overline{B}$$

4) **Associativity**

$$(A\cap B)\cap C = A\cap(B\cap C)$$
$$(A\cup B)\cup C = A\cup(B\cup C)$$

5) **Commutativity**

$$A\cap B = B\cap A \,,\; A\cup B = B\cup A$$

6) **Distributivity**

$$A\cap(B\cup C) = (A\cap B)\cup(A\cap C)$$
$$A\cup(B\cap C) = (A\cup B)\cap(A\cup C)$$

3.1.3 DESIGN GOALS

Control of the environment for large computing systems is often a far greater challenge than for rooms inhabited by people. Not only do the systems themselves generate heat, but they are often specified by their manufacturers to be maintained in as little as a plus-or-minus 1 degree (Fahrenheit) range. Humidity is also a challenge, causing, for example, corrosion and jamming of associated mechanical systems at high humidity levels and the enhanced possibility of static discharge with low levels. Humidity control is often specified to be 50%

relative humidity, with a maximum swing of plus-or-minus 3% per hour.

In addition, the design of a precision environmental control system also faces nonlinearities, caused by such system behavior as airflow delay and dead times, uneven airflow distribution patterns, and duct work layouts. Uncertainties in system parameters are often present, for example, room size and shape, location of heat-producing equipment, thermal mass of equipment, walls, and amount and timing of external air introduction.

Recognizing these challenges, Liebert undertook the design of a control system requiring (in general terms):

- ✓ Precision temperature and humidity control;
- ✓ Minimization of cycling times (i.e., the opening and closing of the damper and turning on and off of the compressor), thereby increasing reliability and component life, and also resulting in increased energy efficiency;
- ✓ Straightforward and therefore inexpensive control electronics.

In short, Liebert wanted to precisely control with simple hardware a nonlinear system with significant uncertainties. Several traditional linear approaches were considered but proved inadequate. A fuzzy logic approach was investigated and ultimately implemented. Design specifics - The LogiCool control system has six fuzzy inputs, three fuzzy outputs, and 144 principles (rules). It runs on a Motorola 6803 microprocessor, and is programmed in C.

LogiCool's fuzzy input variables are: e_temperature, the temperature relative to a setpoint; delta_T/delta_t, the rate of temperature change; e_humidity, the humidity relative to a setpoint; delta_H/delta_t, the rate of humidity change; and two proprietary variables associated with the action of the controllers.

Fuzzy outputs control: 1) amount of cooling, 2) amount of dehumidification, and 3) heat. Outputs can also be treated as fedback input variables, and time delays are treated as fuzzy outputs as well. Each fuzzy variable is assigned seven membership functions as values, with the traditional Large_Negative, Medium_Negative, Small_Negative, Near_Zero, Small_ Positive, Medium_Positive, and Large_Positive as labels. Ranges for the values of each variable are proprietary.

An example of a temperature control principle, using the as ...then ... (rather than the if ... then ...) syntax, is:

as temperature relative to set point is small_positive and temperature rate of change is medium_positive then amount of cooling is small_positive;

The Liebert design also incorporates time delays into their principles. The following demonstrates both this as well as the use of a fuzzy output as a feedback variable.

as temperature relative to setpoint is small_negative and amount of cooling is small_positive then wait delay to cooling change is medium_positive;

A fuzzy OR operator (maximizer) is used as the defuzzification technique, avoiding the complicated calculations associated with a centroid approach. Liebert has found that with the large number of principles, a more elaborate approach is unnecessary. Inputs are sampled, the principle-base accessed, and outputs are updated once a second. The "long" inter-sample delay allows the 6803, a simple eight-bit microprocessor, to implement this rather large fuzzy system.

3.1.4 FUZZY LOGIC OBJECTIONS

It would be remarkable if a theory as far-reaching as fuzzy systems did not arouse some objections in the professional community. While there have been generic complaints about the "fuzziness" of the process of assigning values to linguistic terms, perhaps the most cogent criticisms come from Haack . A formal logician, Haack argues that there are only two areas in which fuzzy logic could possibly be demonstrated to be "needed," and then maintains that in each case it can be shown that fuzzy logic is not necessary.

The first area Haack defines is that of the nature of Truth and Falsity: if it could be shown, she maintains, that these are fuzzy values and not discrete ones, then a need for fuzzy logic would have been demonstrated. The other area she identifies is that of fuzzy systems' utility: if it could be demonstrated that generalizing classic logic to encompass fuzzy logic would aid in calculations of a given sort, then again a need for fuzzy logic would exist.

In regards to the first statement, Haack argues that True and False are discrete terms. For example, "The sky is blue" is either true or false; any fuzziness to the statement arises from an imprecise definition of terms, not out of the nature of Truth. As far as fuzzy systems' utility is concerned, she maintains that no area of data manipulation is made easier through the introduction of fuzzy calculus; if anything, she says, the calculations become more complex. Therefore, she asserts, fuzzy logic is unnecessary.

Fox has responded to her objections, indicating that there are three areas in which fuzzy logic can be of benefit: as a "requisite" apparatus (to describe real-world relationships which are inherently fuzzy); as a "prescriptive" apparatus (because some data is fuzzy, and therefore requires a fuzzy calculus); and as a "descriptive" apparatus (because some inferencing systems are inherently fuzzy).

His most powerful arguments come, however, from the notion that fuzzy and classic logics need not be seen as competitive, but complementary. He argues that many of Haack's objections stem from a lack of semantic clarity, and that ultimately fuzzy statements may be translatable into phrases which classical logicians would find palatable.

Lastly, Fox argues that despite the objections of classical logicians, fuzzy logic has found its way into the world of practical applications, and has proved very successful there. He maintains, pragmatically, that this is sufficient reason for continuing to develop the field.

3.2 Genetic Algorithms

Genetic Algorithms (GAs) are adaptive heuristic search algorithms that belong to the larger part of evolutionary algorithms. Genetic algorithms are based on the ideas of natural selection and genetics. These are intelligent exploitation of random search provided with historical data to direct the search into the region of better performance in solution space.

They are commonly used to generate high-quality solutions for optimisation problems and search problems.

Genetic algorithms simulate the process of natural selection which means those species who can adapt to changes in their environment are able to survive and reproduce and go to next generation. In simple words, they simulate “survival of the fittest” among individual of consecutive generation for solving a problem. **Each generation consist of a population of individuals** and each individual represents a point in search space and possible solution. Each individual is represented as a string of character/integer/float/bits. This string is analogous to the Chromosome.

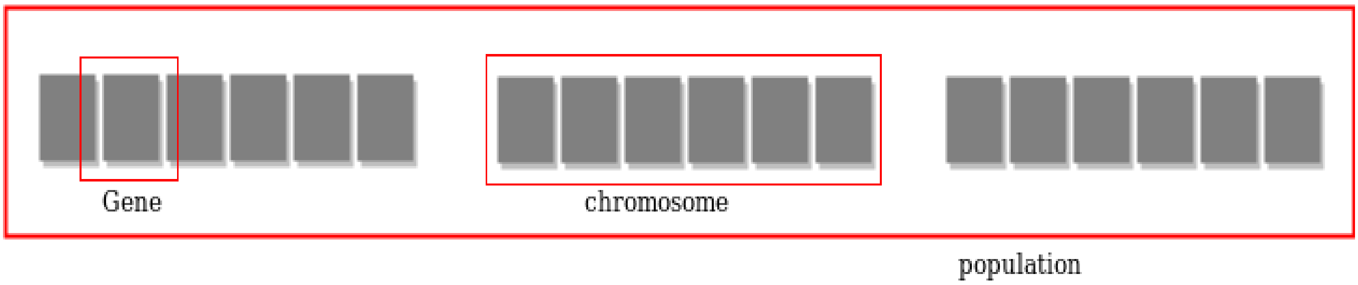
3.2.1 FOUNDATION OF GENETIC ALGORITHMS

Genetic algorithms are based on an analogy with genetic structure and behavior of chromosome of the population. Following is the foundation of GAs based on this analogy –

1. Individual in population compete for resources and mate
2. Those individuals who are successful (fittest) then mate to create more offspring than others
3. Genes from "fittest" parent propagate throughout the generation, that is sometimes parents create offspring, which is better than either parent.
4. Thus, each successive generation is more suited for their environment.

3.2.2 SEARCH SPACE

The population of individuals are maintained within search space. Each individual represent a solution in search space for given problem. Each individual is coded as a finite length vector (analogous to chromosome) of components. These variable components are analogous to Genes. Thus a chromosome (individual) is composed of several genes (variable components).



Fitness Score

A Fitness score is given to each individual, which **shows the ability of an individual to "compete"**. The individual having optimal fitness score (or near optimal) are sought.The GAs maintains the population of n individuals (chromosome/solutions) along with their fitness scores.The individuals having better fitness scores are given more chance to reproduce than others. The individuals with better fitness scores are selected who mate and produce **better offspring** by combining chromosomes of parents. The population size is static so the room has to be created for new arrivals. So, some individuals die and get replaced by new arrivals eventually creating new generation when all the mating opportunity of the old population is exhausted. It is hoped that over successive generations better solutions will arrive while least fit die.

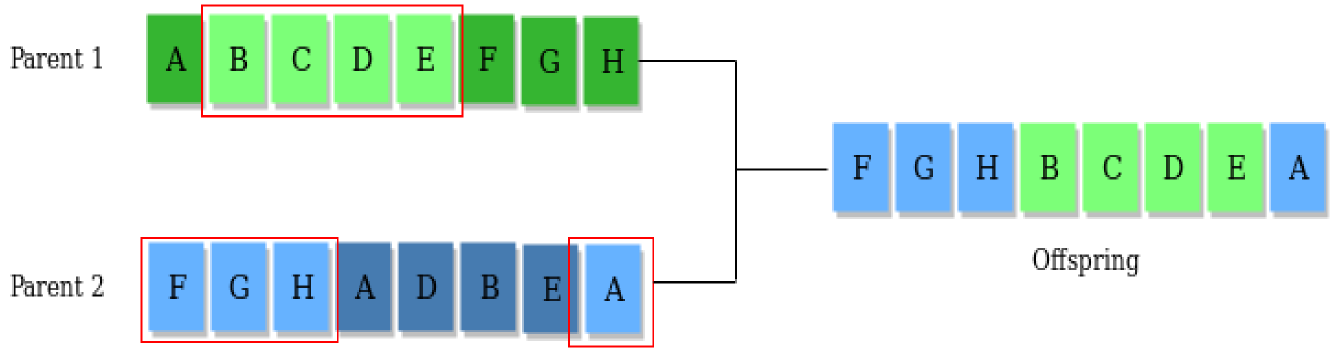
Each new generation has on average more "better genes" than the individual (solution) of previous generations. Thus each new generations have better **"partial solutions"** than previous generations. Once the offsprings produced having no significant difference than offspring produced by previous populations, the population is converged. The algorithm is said to be converged to a set of solutions for the problem.

3.2.3 OPERATORS OF GENETIC ALGORITHMS

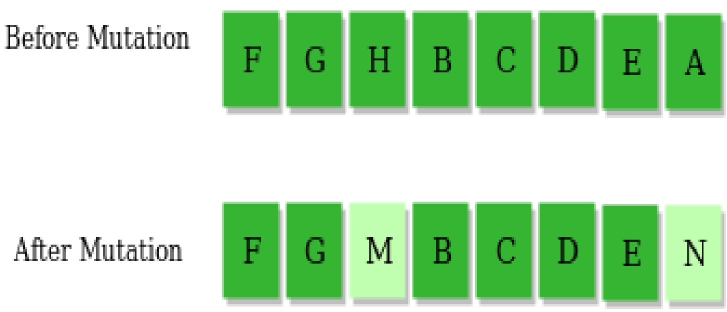
Once the initial generation is created, the algorithm evolve the generation using following operators –

1) Selection Operator: The idea is to give preference to the individuals with good fitness scores and allow them to pass their genes to the successive generations.

2) Crossover Operator: This represents mating between individuals. Two individuals are selected using selection operator and crossover sites are chosen randomly. Then the genes at these crossover sites are exchanged thus creating a completely new individual (offspring). For example –



3) Mutation Operator: The key idea is to insert random genes in offspring to maintain the diversity in population to avoid the premature convergence. For example –



The whole algorithm can be summarized as –

1) Randomly initialize populations
2) Determine fitness of population
3) Until convergence repeat:
   a) Select parents from population
   b) Crossover and generate new population
   c) Perform mutation on new population
   d) Calculate fitness for new population

IV. CONCLUSION

This research will help us getting an idea how these deadly gases and deadly chemicals affects the health of people. These harmful chemicals is responsible for the occurrence of the deadly disease called cancer. This paper helps in making which area are mostly affected by these chemicals and too what extent. This paper helps us in making an estimation what kind of Cancers can mostly occur in these areas. So, that strong precautionary measures can be taken by the Health Care Centre of those Areas or the nearby areas.

V. REFERENCES

[1]. Abir Alharbi and Fairouz Tchier, "A Fuzzy-Genetic Algorithm Method for the Breast Cancer Diagnosis Problem Research Collaboration".
[2]. Ioana SCROBOTĂ, Grigore BĂCIUȚ, Adriana Gabriela FILIP,Bianca TODOR, Florin BLAGA, andMihaela Felicia BĂCIUȚ, "Application of Fuzzy Logic in Oral Cancer Risk Assessment".
[3]. Indrani Nath and Abir Bhattacharya "Selecting an Appropriate Baby Formula Milk and Baby Cereal Food Using Fuzzy LogicMethods".
[4]. Ms Rinchin Dr Prabir Chatterjee Dr Manan Ganguli Dr Smarajit Jana, "The Health and Environmental Impact of Coal Mining in Chhattisgarh".
[5]. Schneider J1, Peltri G, Bitterlich N, Neu K, Velcovsky HG, Morr H, Katz N, Eigenbrodt E., "Fuzzy logic-based tumor marker profiles including a new marker tumor M2-PK improved sensitivity to the detection of progression in lung cancer patients."