Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 4, April 2014, pg.368 – 388

RESEARCH ARTICLE

An Innovative Smart Soft Computing Methodology towards Disease (Cancer, Heart Disease, Arthritis) Detection in an Earlier Stage and in a Smarter Way

Mr. Tanupriya Choudhury[1], Prof. (Dr.) Vivek Kumar[2]
Dr. Darshika Nigam[3], Vasudha Vashisht[4]

[1]Research Scholar (CSE), Jagannath University, Jaipur, India
[2]Principal, Delhi College of Technology and Management, Palwal, India
[3]Department of Bio Chemistry, SLS, Dr. B.R.Ambedker University, Agra, India
[4]Assistant Professor, Lingaya's University, Faridabad, India

[1] tanupriya86@gmail.com; [2] profvivekkumar@gmail.com;
[3] drdarshikanigam@rediffmail.com; [4] ervasudha@gmail.com

Abstract: ***Cancer, Heart disease, arthritis are the most common diseases found in the majority of the populations in recent years. Medical diagnosis is enormously essential but complex task that should be accomplished exactly and proficiently. Although momentous progress has been made in the diagnosis and treatment of these diseases, further investigation is still desired. These disease diagnosis are a challenging task which can offer automatic prediction about the disease of patient so that further treatment can be made informal. Due to this fact, disease diagnosis has received enormous interest globally among medical community. In this paper soft computing played an important role in diagnosis of these diseases with improved effectiveness and suitable accuracy. It gives a detailed view of a innovative earlier detection system for these diseases with the help of soft computing methods and proper attributes reference value to predict the diseases analytically. Different clinical values for attributes and biomarkers have been taken as an input and match these with the reference values to predict the diseases accurately. Diseases datasets are also analyzed here using soft computing approach. The outcome of this would help doctors, scientists, pharmacists in understanding the characteristic and association of attributes which is responsible for these diseases and provide proper diagnosis method and in discovering new drugs.***

I. INTRODUCTION

The term soft computing was invented by L. A. Zadeh in1992. Soft computing can be well-defined as a set of "imprecise" computing methods, which are intelligent to model and analyse very complex problems. For these complex problems, classical and precise scientific methods were unable to yield low-cost, systematic, or comprehensive solutions [2]. In Zadeh's own words, "Soft computing is an emerging approach to computing which parallel the remarkable ability of the human mind to reason and learn in an environment of uncertainty and imprecision"[1]. Fuzzy logic, artificial neural networks, evolutionary algorithms (genetic algorithm, genetic programming, machine learning and evolutionary strategies),bio-informatics, simulated annealing, and probabilistic reasoning are fundamental computing elements of soft computing [3]. Among the various soft computing constituents, fuzzy logic (FL), artificial neural networks (ANNs) and evolutionary algorithms (EAs) are considered as the fundamental practices of soft computing. Every of these procedures has their own asset. Just like bio-informatics is a part of evolutionary algorithms and deals with algorithms, databases and information systems, web technologies, artificial intelligence and soft computing, information and computation theory, software engineering, data mining, image processing, modelling and simulation, signal processing, discrete mathematics, control and system theory, circuit theory, and statistics. Bioinformatics generates new knowledge as well as the computational tools to create that knowledge [4]. Cancer research, heart disease research, artharitis detection is a field of bio-informatics, where we can use classification, clustering algorithmic methods and soft computing methods for better estimate and detect in a smarter way in an earlier stage.

II. SOFT COMPUTING & DATA MINING METHODS

Soft Computing is a branch of artificial computational intelligence that employs a variety of statistical, probabilistic and optimization techniques that allows computers to "learn" from past examples and to detect hard-to-discern patterns from large, noisy or complex data sets. This capability is particularly well-suited to medical applications, especially those that depend on complex proteomic and genomic measurements.

As a result, computational intelligence is frequently used in cancer diagnosis and detection. More recently soft computing has been applied to cancer prognosis and prediction. A number of trends are there, including a growing dependence on protein biomarkers and microarray data, a strong bias towards applications in prostate and breast cancer, and a heavy reliance on "older" technologies such artificial neural networks (ANNs) instead of more recently developed or more easily interpretable soft computing techniques.

A number of published studies also appear to lack an appropriate level of validation or testing. Among the better designed and validated studies it is clear that soft computing techniques can be used to substantially (15– 25%) improve the accuracy of predicting cancer susceptibility, recurrence and mortality. At a more fundamental level, it is also evident that computational intelligence is also helping to improve basic understanding of cancer development and progression.

The techniques of soft computing may include neural network, fuzzy set theory, genetic algorithm and simulated annealing etc. The below stated table describes the strength of the soft computing techniques.

TABLE 1: VARIOUS METHODOLOGIES FOR SOFT COMPUTING

S.Nm.	Methodology	Strength
1	Neural Networks	Learning & Adaption
2	Fuzzy Logic and fuzzy set theory	Knowledge representation via fuzzy if-then rules
3	Genetic algorithm and simulated annealing	Systematic random search
4	Conventional AI	Symbolic manipulation

TABLE 2: VARIOUS SOFT COMPUTING TECHNIQUES IN DIAGNOSTICS OF DISEASES [36][37][38]

Sl. No.	SC Techniques used	Diseases cure/detection/recognition
1	Fuzzy logic	Neural system disorder
2	Medical imaging (bio inspired soft computing)	Cancer, arteriosclerosis, epilepsy, Alzheimer, Parkinson
3	Object-oriented expert system	Diagnosis of fungal diseases of date palm
4	Decision support systems	Diagnosis of disease states and corresponding herbal prescriptions
5	Neural networks, image processing	Oral cysts
6	Artificial neural network	Neonatal disease diagnosis
7	Decision support system	Congenital heart disease diagnosis based on signs and symptoms
8	Fuzzy knowledge base	Glaucoma monitoring
9	Clustering techniques	To distinguish the data set to two primary clusters i.e. diseased and disease free
10	Classification techniques	To classify a sample at first as diseased or free from disease and subsequently if diseased then particular type of the disease

Classification & clustering [29][30] is a method in which Objects are considered by one or more features Classification is a task which assigns objects to classes or groups on the basis of measurements made on the objects

- Have labels for some points
- Want a "rule" that will accurately assign labels to new points
- Supervised learning

Clustering is to group observations that are "similar" based on predefined criteria.

- No labels
- Group points into clusters based on how "near" they are to one another
- Identify structure in data
- Unsupervised learning

Clustering is a partition of data into groups of similar substances. Each collection, called a cluster, contains of objects that are similar between themselves and different compared to objects of other collections.

TABLE3: METHODS FOR SOFT COMPUTING & DATA MINING COMPARISON

Sr. Nm.	Soft Computing & Data Mining methods	Description	Reference
1.	Decision Trees (DT's)	A decision tree is a tree where each non-terminal node signifies a test or decision on the measured data item. Choice of a certain branch be contingent upon the outcome of the test. To classify a particular data item, we start at the root node and follow the assertions down until reach a terminal node (or leaf). A decision is made when a terminal node is approached.	[4]
2.	Support Vector Machine (SVM)	In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier.	[5]
3.	Genetic Algorithms (GAs) / Evolutionary Programming (EP)	Genetic algorithms and evolutionary programming are algorithmic optimization strategies that are motivated by the principles experiential in natural evolution of a collection of potential problem solutions that compete with each other, the best explanations are selected and combined with each other. In doing so, one expects that the overall goodness of the solution set will become better and better, similar to the process of evolution of a population of organisms. Genetic algorithms and evolutionary programming are used in data	[6]

		mining to formulate hypotheses about dependencies between variables, in the form of association rules or some other internal formalism.	
4.	Fuzzy Sets	Fuzzy sets form a key methodology for representing and processing uncertainty. Uncertainty arises in many forms in today's databases: imprecision, non-specificity, inconsistency, vagueness, etc. Fuzzy sets exploit uncertainty in an attempt to make system complexity manageable.	[8]
5.	Neural Network	Neural networks (NN) are those systems modelled based on the human brain working. As the human brain consists of millions of neurons that are interconnected by synapses, a neural network is a set of connected input/output units in which each connection has a weight associated with it. The network learns in the learning phase by adjusting the weights so as to be able to predict the correct class label of the input.	[7]
6.	Rough Sets	A rough set is determined by a lower and upper bound of a set. Every member of the lower bound is a certain member of the set. Every non-member of the upper bound is a certain non-member of the set. The upper bound of a rough set is the union between the lower bound and the so-called boundary region.	[8]
7.	Naïve Bayes Classifier	A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". An overview of statistical classifiers is given in the article on pattern recognition.	[46]
8.	K-means clustering	K-means is a well-known partitioning method; objects are classified as belonging to one of k-groups, k chosen a priori. Cluster membership is determined by calculating the centroid for each group and assigning each object to the group with the closest centroid.	[50]
9.	Hierarchical Clustering	Hierarchical clustering is an agglomerative (top down) clustering method. As its name suggests, the idea of this method is to build a hierarchy of clusters, showing relations between the individual members and merging clusters of data based on similarity.	[50]
10.	EM algorithm	The EM algorithm can be used for soft clustering. Intuitively, for clustering, EM is like the k-means algorithm, but examples are probabilistically in classes, and probabilities define the distance metric.	[50]
11.	Microarray Analysis	Gene expression profiling provides tremendous information to help unstitch the complexity of cancer. The selection of the most informative genes from huge noise for cancer classification has taken centre stage, along with predicting the function of such identified genes and the construction of direct gene regulatory networks at different system levels with a tune able parameter	[41]

III. DISEASE DETECTION MODULE (CANCER, HEART DISEASE, ARTHRITIS DETECTION)

The classifier Algorithm:

- Let *D* be a training set of tuples and their associated class labels. As usual, each tuple is represented by an *n*-dimensional attribute vector, $\boldsymbol{X} = (x1, x2, \ldots, xn)$, depicting *n* measurements made on the tuple from *n* attributes, respectively, $A1, A2, \ldots, An$.
- Suppose that there are *m* classes, $C1, C2, \ldots, Cm$. Given a tuple, $\boldsymbol{X}$, the classifier will predict that $\boldsymbol{X}$ belongs to the class having the highest posterior probability, conditioned on $\boldsymbol{X}$. That is, the naïve Bayesian classifier predicts that tuple $\boldsymbol{X}$ belongs to the

  class *Ci* if and only if

$$P(C_i|X) > P(C_j|X) \quad \text{for } 1 \le j \le m, j \ne i.$$

  Thus we maximize $P(Ci \boldsymbol{X})$. The class *Ci* for which $P(Ci| \boldsymbol{X})$. is maximized is called the maximum posteriori hypothesis. By Bayes theorem

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}.$$

- As $P(\boldsymbol{X})$ is constant for all classes, only $P(\boldsymbol{X}|Ci)P(Ci)$ need be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, that is, $P(C1) = P(C2) = ...= P(Cm)$, and we would therefore maximize $P(\boldsymbol{X}|Ci)$. Otherwise, we maximize $P(\boldsymbol{X}|Ci)P(Ci)$. Note that the class prior probabilities may be estimated by $P(Ci)=|Ci,D|$ $|D|$where $|Ci,D|$ is the number of training tuples of class *Ci* in *D*.
- Given data sets with many attributes, it would be extremely computationally expensive to compute $P(\boldsymbol{X}|Ci)$. In order to reduce computation in evaluating $P(\boldsymbol{X}|Ci)$, the naive assumption of class conditional independence is made. This presumes that the values of the attributes are conditionally independent of one another, given the class label of the tuple (i.e., that there are no dependence relationships among the attributes). Thus,

$$\begin{aligned} P(X|C_i) &= \prod_{k=1}^{n} P(x_k|C_i) \\ &= P(x_1|C_i) \times P(x_2|C_i) \times \cdots \times P(x_n|C_i). \end{aligned}$$

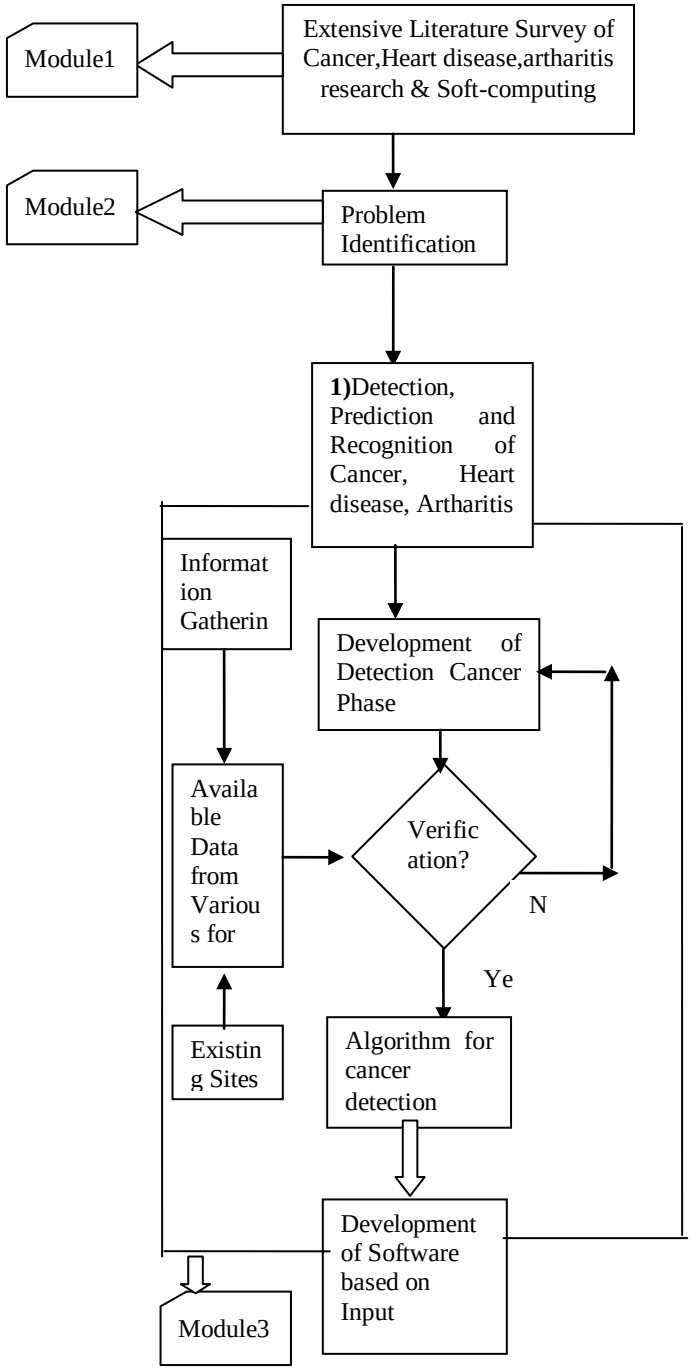


FIG 1: OUTLINE OF THE PROPOSED DISEASE DETECTION MODULE

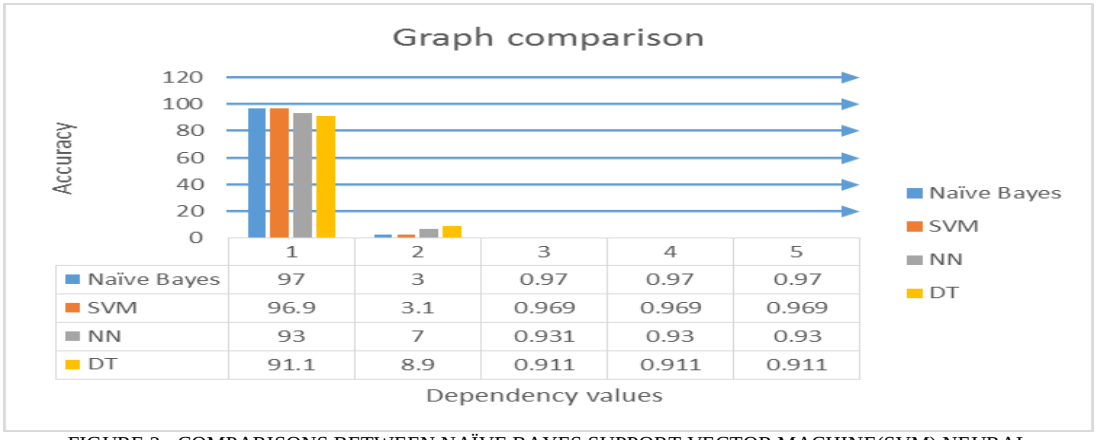
The classifier algorithm has been used as medication of naïve Bayes algorithm. After pre-processing and feature selection phases, the numbers of attribute will be knowingly reduced and are more accurate for the use in construction the classification prototype. For the classification stage, Naïve Bayes is used as the classifier because of its easiness and good enactment in document and simplicity and text classification, as reported and discussed by Chakrabarti et al. [10] & S.L. Ting, W.H. Ip, Albert H.C. Tsang[11].

The advantages of using naive Bayes are:[11][12]

- Naïve Bayes classifier is the simplest instance of a probabilistic classifier.
- The naive Bayes classifier's beauty is in its simplicity, computational efficiency, and good classification performance.
- The output $Pr(C|d)$ of a probabilistic classifier is the probability that a document d belongs to a class C.
- This advantage is especially pronounced when the number of predictors is very large.
- Each document contains terms which are given probabilities based on its number of occurrence within that particular documents.
- With the supervised training, Naïve Bayes can learn the pattern of examining a set of test documents that have been well-categorized and hence comparing the contents in all categories by building a list of words as well as their occurrence.

But this have to be keep in mind that the naive Bayes classifier requires a very large number of records to obtain good results. Second, where a predictor category is not present in the training data, naive Bayes assumes that a new record with that category of the predictor has zero probability. [12]

Not only this, if the usage of naïve Bayes can be compared with other existing algorithm as discussed in [11], it is the best classifier so far.The same algorithm can be tested using the weka data mining tool also.In data mining tool weka different other types of classifiers are also available, the classification results also shown[11] the output of naïve Bayes is the best classifier amongst others if can take large dataset as here same has been used for disease detection.



	1	2	3	4	5
Naïve Bayes	97	3	0.97	0.97	0.97
SVM	96.9	3.1	0.969	0.969	0.969
NN	93	7	0.931	0.93	0.93
DT	91.1	8.9	0.911	0.911	0.911

FIGURE 2 : COMPARISONS BETWEEN NAÏVE BAYES,SUPPORT VECTOR MACHINE(SVM),NEURAL NETWORK(NN),DECISION TREE(DT)(1- CORRECTLY CLASSIFIED INSTANCES,2-INCORRECTLY CLASSIFIED INSTANCE,3-PRECISION,4-RECALL,5-F-MEASURE)

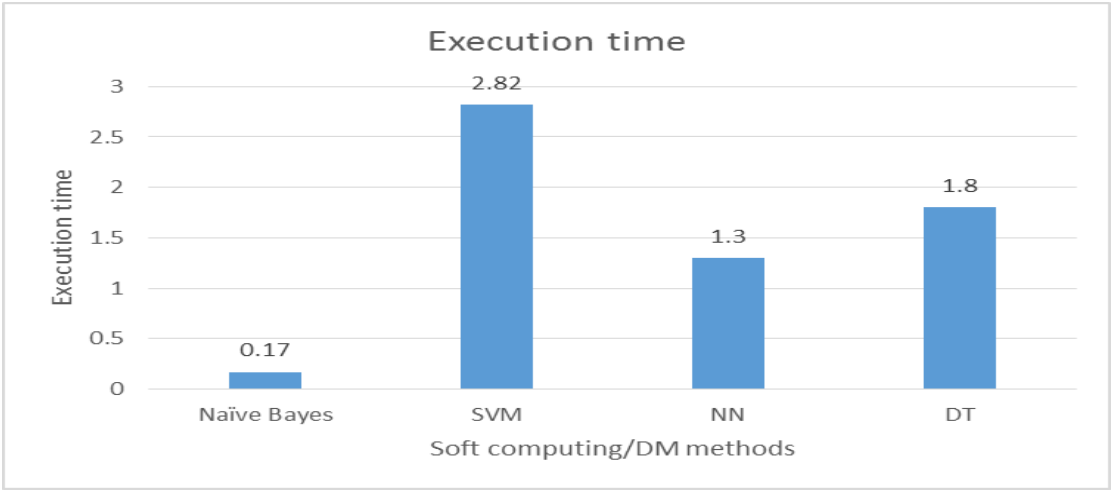


FIGURE3 : EXECUTION TIME COMPARISON BETWEEN NB,SVM,NN,& DT.

IV. HOW THE SYSTEM IS INCORPORATED WITH THE ALGORITHM AS WELL AS ATTRIBUTE

Here to detect the diseases like cancer, heart disease and arthritis the proposed system used naïve Bayes classification algorithm. In cancer, heart disease and arthritis in every module, there are some rudimentary attributes for which the disease happens. To find the attributes deep investigation as well as doctor's ideas has been taken care of.

For the first time here to detect the diseases the prediction module takes bio-markers to find it out it in an earlier stage. As the biomarkers has been taken it gives the actual and accurate output, which is very much useful to detect these diseases in an earlier stage.

And for every attribute there must be a range like if a patient has BP high or low, it must be in a range, if the range crosses then it might be the cause also, for which the attribute measurement value must be taken care of by medicine and all to treatment the disease. And each and every attribute has a reference value. And these values are incorporated in the model, and for best classification naïve Bayes has been used as because it gives the best result amongst other classifiers. After incorporating the values when a patient enters his/her value in the system, it checks methodically and comprehensively the associations between the values what has taken. If the association level for attributes exceed a given value then it shows the patient will be affected by the disease, and from there effortlessly and simply it can be noticed for which value the problem is occurring. Then by medicine it must be taken care of.
Not only this but also in this disease detection module firstly it was taken for less attributes and accuracy has been generated and then more number of attributes has been added and accuracy is showing that where the more number of attributes, the system is more consistent and precise than the other systems.

V. CANCER DETECTION ATTRIBUTES AND GUI OF ATTRIBUTES

Cancer is an abnormal cell-growth occurring in human body and may originate from any of the areas or organs. The disorder can be very dangerous, or even fatal, if ignored for long. It develops in the form of tumors that have a typical tendency to metastasize. Such tumors metastasize or spread to various parts of the body via blood stream[1].

Cancer is characterized by out-of-control cell growth. [2,4]. Research requires detailed study of most common Cancer in men and women including Lung, Prostate, Breast and Oral Cancer and their recognition.

Diagnosis[3][5] of any type of cancer in human being: Diagnosis is concerned with the development of algorithms and techniques that are able to determine whether the behavior of a system is correct. If the system is not functioning correctly, the algorithm should be able to determine, as accurately as possible, which part of the system is failing, and which kind of fault it is facing. The computation is based on observations, which provide information on the current behavior.

In this disease detection system, the algorithm has been used is modified naïve Bayes algorithm, earlier it has been used for less attributes, but here the measurements and attributes are greater than 50 for cancer and heart disease also, and the prediction system using 16 attributes for arthritis. And for cancer detection for the first time it has been using 10 bio-markers which must help health-care professionals to find it out the problem for or the cause of cancer, or it must be detect in earlier stage. Biomarkers are often restrained and estimated to scrutinize normal biological methods, pathogenic methods, or pharmacologic retorts to a therapeutic involvement. Biomarkers are used in many technical fields. A biomarker may be a molecule concealed by a tumor or a precise response of the body to the presence of cancer. Genetic, epigenetic, proteomic, and imaging biomarkers can be used for cancer diagnosis, prognosis, and epidemiology. Ideally, such biomarkers can be assayed in non-invasively collected bio fluids like blood or serum. As an example here the bio-markers for cancers are CA_125,CA 19-9 (cancer antigen 19-9) or GICA (gastrointestinal cancer antigen),CA_153,CA_2729,CA_72_4 etc.

TABLE 4: COMPARISONS OF VARIOUS CANCER DETECTION METHODS [14]

S. No	Cancer Type	Technique	Algorithms used	Results
1	Oral Cysts	Neural networks, Image Processing	Contrast stretching, Radial Basis function	Severity of cysts is measured. For each dental Image accuracy is calculated for classification of Cysts
2	Brain Tumor	Neuro Fuzzy	Fuzzy c-means clustering algorithm	Detected brain tumour at an earlier stage

3	Oral cancer	Image Processing	Active contour model(snakes)	Segmentation of oral lesion is obtained in single band images from true color images
4	Breast Cancer	CAD, Image processing	Super resolution technique	Detected cancer at very early stage
5	Breast Cancer	Image processing	Watershed segmentation	Detected cancer tumors at an early stage
6	Oral Cancer	Image Processing	Supervised segmentation, image Feature extraction	Oral cancer reoccurrence is predicted automatically
7	Oral cancer	Wavelet, neural networks	Multilayered feed forward neural network	The feature vectore are extracted from each contiguous 64*64 blocks by wavelet decomposition
8	Oral cancer	Wavelet, Data mining, neural network	Bayesian classification, support vector machines	48 gabor wavelet features& 9 wavelet features are extracted

TABLE 6: CANCER ATTRIBUTES

#no	Attribute	Description	Values
1	Age	Age of Patients	Continuous
2	Gender	Gender of patients	Continuous
3	Stage	Stage of cancer	Binary
4	Treatment	Treatment taken or not	Binary
5	Dtime	Date and time for the appointment	Continuous
6	Date on study	The study for the patient	Binary
7	Weight index	The weight inbex as BMI for Patient	Continuous
8	Pf	Positive family history	Binary
9	Hx	hexosaminidase enzyme	Continuous
10	Famhist	Has anyone in your family had colorectal cancer (no [0]; yes [1])	Binary
11	Fbs	(Fasting blood sugar .120mg/dl) 0,1 & (1=true; 0=false)	Binary
12	Dig	Earlier diagnosis taken or not	Binary
13	Exang	exercise induced angina	Binary
14	Sbp	A systolic blood pressure of 120-139 means you have normal blood pressure	Continuous
15	Dbp	A normal diastolic blood pressure number is 80 or less. A diastolic blood pressure between 80 and 89 is normal but higher than idea	Binary
16	Ekg	Electrocardiogram test	Continuous
17	Sg	Value in blood	Continuous
18	Hz	low rates of repetitive stimulation yes or no	Binary
19	Index of stage	Begin or malignant	Binary
20	Acid_Phophates	The different values for acid phosphates in blood	Continuous
21	Bone-specific alkaline phosphatase(B-AKP)	Mean age was 69.70 years (50-83 years). Average B-AKP value was 29.28 ng/ml (9-56 ng/ml). Bone scan was positive in 31 patients. Among these 31, 26 patients had positive B-AKP and five negative. AP patients had negative bone scan. From this AP patients, B-AKP was positive in 17 and negative in 63. The sensitivity of the test was 83.8%. The specificity was 78%. Positive predictive value was 60% and negative predictive value was 92%..	Binary
22	Clonality	gene rearrangements (ie, leukemia-specific sequences)	Continuous
23	Autonomy	autonomy is due to a gradual increase in the numbers of cells having relatively autonomous thyroid hormone synthesis	Binary
24	Anaplasia	Anaplastic tumors have a high mitotic rate and lymph vascular invasion. It rapidly invades surrounding tissue	Binary

25	Metastases	The upper limit of normal for alkaline phosphatase and carcinoembryonic antigen did not represent the optimal levels for use in predicting liver metastases. However, with alkaline phosphatase greater than 135 I.U., and/or carcinoembryonic antigen greater than 10 ng/ml, sensitivity was 88%: 23 of 26 patients with liver metastases fulfilled either or both criteria. The false-positive rate was 12%. Liver scanning, alone, demonstrated metastases in only 69% of 35 patients with liver metastases. The combination of alkaline phosphatase and carcinoembryonic antigen can be used economically to screen for liver metastases	Continuous
26	Alpha_Feto_Protein	A typical normal range is 0.5 to 2.0 or 2.5 MoM	Continuous
27	CEA (Carcinoembryonic Antigen)	2.5 ng/ml is stated to be the upper limit of normal for plasma CEA levels. Values in excess of 2.5 ng/ml may be found in association with cancers	Continuous
28	PSA(prostate-specific antigen)	Men below age 50: PSA less than 2.5,Men 50 - 59 years: PSA level less than 3.5,Men 60 - 69 years: PSA level less than 4.5,Men older than 70 years: PSA level less than 6.5	Continuous
29	CA_125	Normal values of CA125 in serum range from 0 to 35 U/ml. CA125 is also expressed by a number of tissues of both cancerous and noncancerous origin	Continuous
30	CA 19-9 (cancer antigen 19-9) or GICA (gastrointestinal cancer antigen)	The normal blood levels of CA 19-9 are below 37 U/ml and with this reference level a false positive rate of 20 per cent has been reported in pancreatitis	Continuous
31	CA_153	CA15-3 appears to be a marker for individualizing therapy in patients with breast cancer, where patients with high CA 15-3 show good response to aggressive	Continuous
32	CA_2729	CA27.29. A cut-off value of 30 U/ml resulted in a specificity of 62% and a sensitivity of 81%. (http://www.ncbi.nlm.nih.gov/pubmed/11326676)	Continuous
33	CA_72_4	Normal levels of CA72-4 were considered to be <3.8 U/mL. (http://www.hindawi.com/journals/dm/2013/984641/)	Binary
34	(Human chorionic gonadotropin (hCG)	The normal range for men is between 0-5 mIU/mL (http://www.tc-cancer.com/tumormarkers.html)	Binary
35	NSE (neuroenzym-specific-enolase)	the NSE level : in limited forms, 50% show high levels; in disseminated cancers, 100% show high levels. (http://www.cancersafe.com/screening/index.asp) , < or =15 ng/mL Serum markers are not specific for malignancy, and values may vary by method. (http://www.mayomedicallaboratories.com/test-catalog/Clinical+and+Interpretive/80913)	Binary
36	POA	Pancreatic oncofetal antigen (POA) measured in the sera of patients with cancer of the pancreas to determine the two positive rates of these cancer markers.	Continuous
37	Lactate_Dehydrogenase	A typical range is 105 - 333 IU/L (international units per liter)	Continuous
38	5-HIAA	Up to 45 mg/24 h is considered normal by some laboratories	Continuous
39	Catecholamines (METANEPHRINE ,NORMETANEPHRINE,TOTAL METANEPHRINE)	In plasma (blood) or through a 24-hour urine collection. Care should be taken to rule out other causes of adrenergic (adrenalin-like) excess like hypoglycemia, stress, exercise, and drugs affecting the catecholamines like stimulants, methyldopa, dopamine agonists, or ganglion blocking antihypertensives.	Continuous
40	Urinary_Steroids	Normal range is 1.001 to 1.035	Continuous
41	Calcitonin	hypercalcitoninemia (O10 pg/ml) of 4.87%, including 0.26% for medullary cancers and 0.12% for CCH. It is noteworthy that the great majority of these cases of hypercalcitoninemia (77%) are due to CT levels between 10 and 20 pg/ml.	Continuous
42	Thyroglobulins	THYROGLOBULIN, TUMOR MARKER, > or =16 years: < or =33 ng/mL, Athyrotic individuals normally have human thyroglobulin values < or =2 ng/mL.,THYROGLOBULIN ANTIBODY SCREEN,<22 IU/mL ,Reference values apply to all ages	Continuous

43	AMAS(anti malignant antibody in serum)	IF the AMAS test is normal, there is a better than 99% chance that the doctor will not find cancer(Normal 1, Abnormal 0)	Binary
44	thalach	maximum heart rate achieved (yes/no)	Binary
45	thalrest	Resting heart rate	Binary
46	diuretic	diuretic used used during exercise ECG: 1 = yes; 0 = no)	Binary
47	prop	Beta blocker used during exercise ECG: 1 = yes; 0 = no	Binary
48	nitr	nitrates used during exercise ECG: 1 = yes; 0 = no	Binary
49	lv-x1	negative control lentivirus vectors (1) (5.98% ± 0.7%)	Continuous
50	lv-x2	negative control lentivirus vectors (2) (12.32% ± 0.9%)	Continuous
51	Lv-X3	negative control lentivirus vectors (3) (13.52% ± 2.2%)	Continuous
52	cathef	cardiac catheterization	Binary
53	Thalreal	Defect type	Continuous

FIG4 : SCREENSHOT FOR CANCER DISEASE DETECTION

VI. HEART DISEASE DETECTION ATTRIBUTES AND GUI OF ATTRIBUTES

Heart disease diagnosis is a multifaceted task which needs much knowledge and information. Outdated way of predicting Heart disease is doctor's examination or number of medical tests. Nowadays, Health care industry contains huge amount of heath care data, which contains hidden information.

This hidden information is useful for making operative decisions. Computer based information along with advanced soft computing & Data mining techniques are used for appropriate results. Neural network is widely used tool for predicting Heart disease diagnosis. A Heart Disease Prediction system is developed using various soft computing techniques especially neural networks (naive Bayes classification modest algorithm). The system predicts the likelihood of patient getting a Heart disease. For prediction, the system uses sex, blood pressure, cholesterol like 50 medical parameters. Using this system, the output would be shown like that the patient would be affected by heart disease or not in an earlier stage. The parameters are contains with binary as well as continuous values. It's a mainly a java based desktop application and in backend MySQL databases are the. The system will enter and incorporates information from various patients and fill up the required fields, some field may have missing values also. The proposed system has an algorithm which has been trained by naïve Bayes algorithm with the help of reference systematic values, if the patient information in exceeding the minimum or maximum or threshold value, it will give the

output for which input it is showing the problem, then from there by the help of healthcare professionals, it may be taken care of to sustain the value which is showing error.

The method primarily based on the information collected from precedent experiences and from current circumstances, which visualizes something as it may occur in future, is known as prediction. The degree of success differs every day, in the process of problem solving on basis of prediction. Neural networks are one among the widely recognized Artificial Intelligence (AI) machine learning models, and a great deal has already been written about them. A general conviction is that the number of parameters in the network needs to be associated with the number of data points and the expressive power of the network. The process utilizes a multi-layer perceptron (MLP) with back-propagation (BP) algorithm to train the selected significant patterns.

TABLE 5: HEART DISEASE ATTRIBUTES

#no	Attribute	Description	Values
1	Age	Patient's Age	Continuous
2	Gender	Patient's Gender	Continuous
3	Fatigue	Feeling very tiredness due to overflow of blood	Binary
4	Shortness of breath	Unable to take a normal breath	Binary
5	Headache	A pain in the head or neck due to defect in the heart	Binary
6	Sweating	Due to tiredness (if a defect present in heart)	Binary
7	Cold	Symptom of Chest Infection	Binary
8	Problem of feeding	Children can face problem while feeding due to shortness of breath	Binary
9	Less weight	If a defect present in the heart, children loss normal weight.	Binary
10	Easily tiring	Getting tiredness during exercise due to defect in the heart.	Binary
11	Cough	Comes due to chest infection (which brings up phlegm).	Binary
12	Chest pain	Getting pain in the chest due to abnormal flow of blood	Binary
13	Clubbing	The proliferation of soft tissue around the ends of fingers	Binary
14	Hypertention	Due to high blood pressure (which occurs due to the mixing of blood).	Binary
15	Palpitations	Periods of rapid and irregular heartbeats (Due to overflow of blood).	Binary
16	Fever	Infection caused by bacteria	Binary
17	Dyspnea	Shortness of breath	Binary
18	Dizziness	Due to low blood pressure, cardiac arrhythmias	Binary
19	Vomits	Common symptoms for Acute Myocardial Infection (AMI).	Binary
20	Diarrhea	Symptom for a Congenital Heart Defect.	Binary
21	Chest Infection	Infection Is a bacterial or viral infection of the airways leading down into the lungs	Binary
22	Syncope	Is a sudden or temporary loss of consciousness or Fainting	Binary
23	Trauma	Impaired blood flow to the lungs	Binary
24	Build up of blood	Which occur as symptom of a defect in the heart	Binary
25	Systolic B.P	Maximum pressure in the arteries during the cardiac cycles	Continuous
26	Diastolic B.P	Refers to the pressure that is exerted on the walls of the various ,around the body in between heart beat when the heart is released.	Continuous
27	Heart beat	Beat A cardiac cycle of the heart.	Continuous
28	Cynosis	Is a severe condition indicates lack of oxygen in the blood supply causes	Binary
29	Edima	Swelling of organs or body tissue.	Binary
30	Thrill	Chest wall vibrations of sufficient intensity to be 0 or 1 recognized by tactile sensation	Binary
31	Cardiac failure	Is a condition in which the heart function as a pump to deliver oxygen rich blood to the body is inadequate to meet the body's need.	Binary

32	Regurgitation	Is characterized by diastolic reflex of blood from the aorta into the ventricle.	Binary
33	Systolic Murmur	Heard when the heart is squeezing and pumping blood out of the heart.	Binary
34	Diastolic Murmur	Heard when the heart is relaxing and filling with blood.	Binary
35	Both Murmurs	Heard during the entire heartbeat signs of a heart defect	Binary
36	Anemia	Anemia is a common in acquired heart failures and affects prognosis.	Binary
37	Smoking	Smoking habits	Binary
38	Over weight	Patient weight trade-off with height	Continuous
39	Alcohol Intake	Patient intake of alcohol greater than a general level	Binary
40	High salt diet	Salt consumption with food	Continuous
41	Fat Diet	Dietary food with fat diet	Binary
42	Exercise	Daily exercise or not	Binary
43	Hereditary	Genetic or not	Binary
44	Bad Cholesterol	Cholesterol value	Binary
45	Blood pressure	High or low	Binary
46	Blood sugar	High or general or low	Binary
47	Heart rate	Value for heart rate	Continuous
48	Chol	Serum cholesterol	Continuous
49	Restecg	Resting electrographic results	Continuous
50	Slope	Slope of the peak exercise ST segment	Continuous
51	Thal real	Defect type	Continuous
52	Ca (number of major vessels colored by floursopy)	Number of major vessels colored by floursopy	Binary

FIG5 : SCREENSHOT FOR HEART DISEASE DETECTION SYSTEM

VII. ARTHRITIS DETECTION ATTRIBUTES AND GUI OF ATTRIBUTES

Arthritis is a major encumbrance for people living with this disease and for their families. It is also a heavy fee to the health care system. Early detection and rapid dealing of arthritis can sojourn joint harm, improve ability to perform daily activities, and condense pain. Regrettably, a large number of people with early arthritis do not receive the right treatment.

The tool is shaped using information gathered in a literature review and a series of consults with a panel of experts. The panel included consumers, health care workers, rheumatologists, and family physicians. The tool is designed to help family doctors and healthcare workers identify people with possible diagnosis of early Arthritis. This tool was thoroughly tested to ensure that it is reliable and valid. It is hoped that this tool will lead to earlier detection of arthritis, lower healthcare costs and, most importantly, better quality of life for people with arthritis.

TABLE 7: ARTHRITIS ATTRIBUTES INFORMATION:

#no	**Attribute**	**Description**	**Values**
1	Joints	Joints are ok or not	Binary
2	Skin	Patient have any skin disease or not	Binary
3	Lungs	Lungs transparency & functionality	Binary
4	Kidneys	Kidney functionality	Binary
5	Heart Rate	The patient's heart rate	Continuous
6	blood vessels	Blood vessels and blood circulation is ok or not	Binary
7	Ocular	Surface type disease in eye	Binary
8	Hepatic	Rate of hepatitis	Continuous
9	Hematological	Anemia of Inflammation rate	Continuous
10	Osteoporosis	Fracture prevention measures	Continuous
11	Lymphoma	Cancer affected or not	Binary
12	Age	The patients age	Continuous
13	Gender	Gender as male or female	Binary
14	Neurological	If the person have neurological disorder or not.	Binary
15	Constitutional symtoms	Constitutional symptoms refers to a group of symptoms that can affect many different systems of the body. Examples include weight loss, fevers, fatigue, and malaise	Continuous
16	Hx	hexosaminidase enzyme	Continuous

FIG6 : SCREENSHOT FOR ARTHRITIS DISEASE DETECTION

VIII. RESULTS & DISCUSSION & PERFORMANCE EVALUATION FACTORS

CONFUSION MATRIX

		Condition positive	Condition negative	
Test outcome	Test outcome positive	**True positive**	**False positive** (Type I error)	Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Test outcome positive}}$
	Test outcome negative	**False negative** (Type II error)	**True negative**	Negative predictive value = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Test outcome negative}}$
		Sensitivity = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	Specificity = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	**Accuracy**

$$\text{accuracy} = \frac{\text{number of true positives} + \text{number of true negatives}}{\text{number of true positives} + \text{false positives} + \text{false negatives} + \text{true negatives}}$$

$$\text{precision} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{false positives}}$$

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$\text{True negative rate} = \frac{tn}{tn + fp}$$

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

The proposed implemented system examine the potential use of classification based data mining techniques such as Naive Bayes to massive volume of Cancer, Heart and arthritis Data. For data pre-processing and Data prediction naive bayes classification are used. The experimental results show that our naive bayes classification can achieve more accuracy when using more than 50 attributes in our datasets which is perform rule based classification over dataset and prediction of result can be performed based on naive bayes rule classified values. Finally various factors will be evaluated from this prediction results such as precision, recall, accuracy and etc...

A confusion matrix (Kohavi and Provost, 1998) contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two class classifier.

The entries in the confusion matrix have the following meaning in the context of our study:

- a is the number of correct predictions that an instance is negative,
- b is the number of incorrect predictions that an instance is positive,
- c is the number of incorrect of predictions that an instance negative, and
- d is the number of correct predictions that an instance is positive.

Confusion Matrix (Heart Disease Prediction For 30 attributes)		Predicted	
		Negative	Positive
Actual	Negative	a= 270	b = 0
	Positive	c = 15	d = 348

TABLE8 : CONFUSION MATRIX FOR HEART DISEASE PREDICTION FOR 30 ATTRIBUTES

Confusion Matrix (Heart Disease Prediction For 52 attributes)		Predicted	
		Negative	Positive
Actual	Negative	a= 353	b = 0
	Positive	c = 5	d = 547

TABLE9 : CONFUSION MATRIX FOR HEART DISEASE PREDICTION FOR 52 ATTRIBUTES

Several standard terms have been defined for the 2 class matrix:

- The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the equation:

  AC= (a+d)/(a+b+c+d)= 97.63 when parameters are 30 and 99.44 when parameters are 52 (Heart Disease)

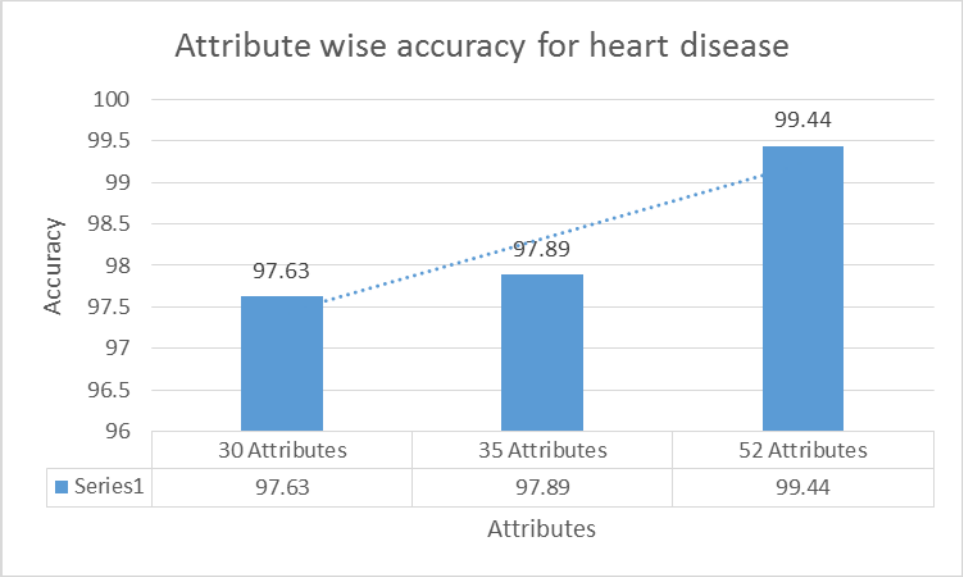


FIG 7: ATTRIBUTE WISE ACCURACY FOR HEART DISEASE PREDICTION SYSTEM

Confusion Matrix (Cancer Prediction For 30 attributes)		Predicted	
		Negative	Positive
Actual	Negative	a= 235	b = 0
	Positive	c = 25	d = 305

TABLE 10: CONFUSION MATRIX FOR CANCER DISEASE PREDICTION FOR 30 ATTRIBUTES

Confusion Matrix (Cancer Prediction For 53 attributes)		Predicted	
		Negative	Positive
Actual	Negative	a= 290	b = 0
	Positive	c = 10	d = 365

TABLE 11 : CONFUSION MATRIX FOR CANCER DISEASE PREDICTION FOR 53 ATTRIBUTES

- The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the equation:

  AC= (a+d)/(a+b+c+d)= 95.5 when parameters are 30 and 98.4 when parameters are 53 (Cancer Disease)

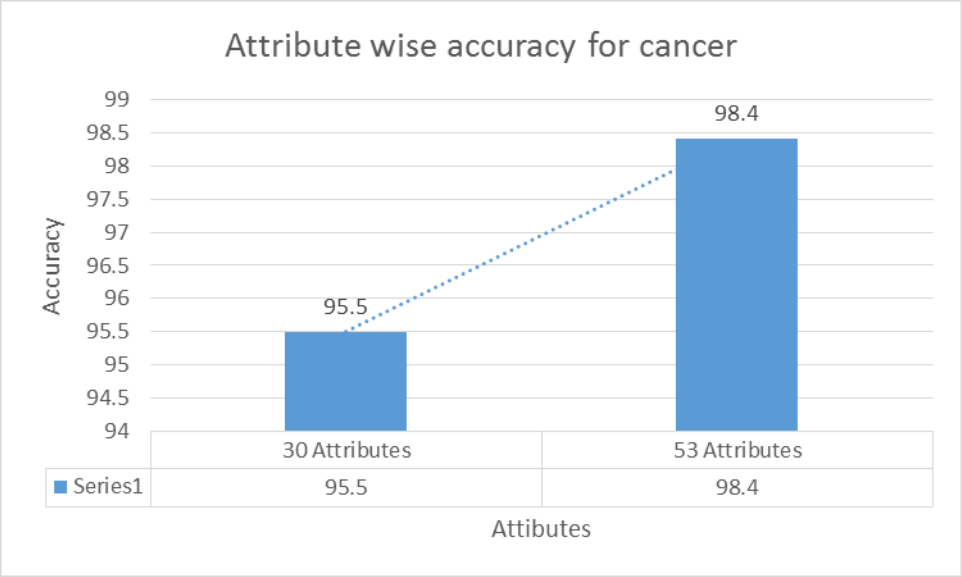


FIG 8: ATTRIBUTE WISE ACCURACY FOR GENERAL CANCER DISEASE PREDICTION SYSTEM

Confusion Matrix (Arthritis Prediction For 10 attributes)		Predicted	
		Negative	Positive
Actual	Negative	a= 117	b = 0
	Positive	c = 20	d = 203

TABLE 12: CONFUSION MATRIX FOR ARTHRITIS DISEASE PREDICTION FOR 10 ATTRIBUTES

Confusion Matrix (Arthritis Prediction For 16 attributes)		Predicted	
		Negative	Positive
Actual	Negative	a= 152	b = 0
	Positive	c = 8	d = 255

TABLE 13 : CONFUSION MATRIX FOR ARTHRITIS DISEASE PREDICTION FOR 16 ATTRIBUTES

- The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the equation:
  AC= (a+d)/(a+b+c+d)= 94.11 when parameters are 10 and 98.07 when parameters are 16 (Arthritis Disease)

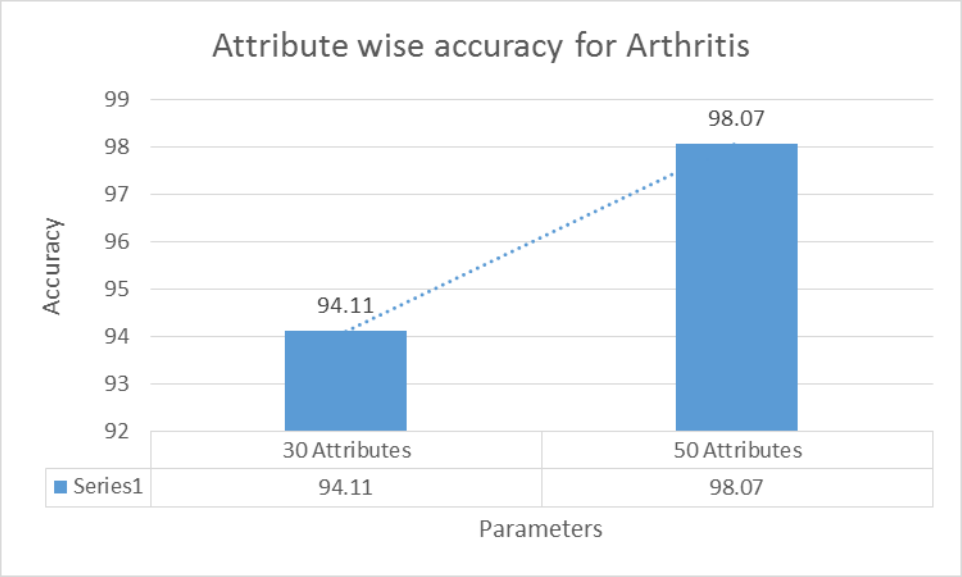


FIG 9 : ATTRIBUTE WISE ACCURACY FOR ARTHRITIS DISEASE PREDICTION SYSTEM

IX. CONCLUSIONS & FUTURE WORK

In this paper the problem of obliging and succincting different algorithms of soft-computing. The focus is on using different algorithms for prophesying combinations of several target attributes. The proposed method have presented an intelligent and effective cancer ,heart disease and arthritis prediction/detection methods using soft computing & data mining methods.

This proposed method briefly examine the potential use of classification based data mining techniques such as Naive Bayes to massive volume of Cancer, Heart and arthritis Data. For data pre-processing and Data prediction naive Bayes classification are used. Experimental results show that our naive Bayes classification can achieve more accuracy when using more than 50 attributes in our datasets which is perform rule based classification over dataset and prediction of result can be performed based on naive Bayes rule classified values. Finally various factors will be evaluated from this prediction results such as precision, recall, accuracy and etc.

Firstly, this method provided an efficient approach for the extraction of significant patterns from the disease data warehouses for the efficient prediction of cancer, heart disease, arthritis Based on the calculated substantial weightage, the numerous patterns having value greater than a predefined threshold were chosen for the valuable prediction of cancer, heart disease & arthritis. The goals are defined based on business intelligence and data exploration. The goals are to be evaluated against the trained models. The second conclusion is that the accuracy of the Bayesian Classification further improves the prediction system to reduce the actual data size to get the optimal subset of attribute sufficient for disease prediction. All these models could answer complex queries in predicting these diseases.

The tool is shaped using information gathered in a literature review and a series of consults with a panel of experts. The panel included consumers, health care workers, rheumatologists, and family physicians. The tool is designed to help family doctors and healthcare workers identify people with possible diagnosis of early cancer, heart disease, arthritis. This tool was thoroughly tested to ensure that it is reliable and valid. It is hoped that this tool will lead to earlier detection of cancer, heart disease, arthritis, lower healthcare costs and, most importantly, better quality of life for people with cancer, heart disease, arthritis.

In future work, this can further improved and prolonged. For predicting diseases like cancer more precisely lung cancer & oral cancer significantly 10 Attributes are listed. Besides the 10 listed in medical literature it can also be incorporated using other soft computing Techniques, e.g., Genetic algorithm, Clustering and Association Rules and microarray genetic expressions. Continuous data can also be used instead of just categorical data. Also Text Mining to mine the vast amount of unstructured data available in healthcare databases. The proposed work can be further improved and lengthened for the automation of disease (cancer, heart disease, arthritis) prediction with the help of genetic algorithm and microarray gene expression. Real data from Health care organizations and agencies needs to be collected and all the available techniques will be compared for the optimum accuracy.

REFERENCES

1.Lotfi A. Zadeh, Fuzzy Logic, Neural Networks And Soft Computing. Onepage Course Announcement Of CS 294-4, University Of California At Berkeley,1992.
2.Yanbo Huanga, Et.Al.,Development Of Soft Computing And Applications In Agricultural And Biological Engineering, Computers And Electronics In Agriculture, 71 (2010), Pp. 107–127.
3.L.A. Zadeh, Fuzzy Logic, Neural Networks, And Soft Computing,Comm. ACM, Vol. 37, Pp. 77-84, 1994.
4.Mr. Tanupriya Choudhury, Dr. Vivek Kumar, Dr. Darshika Nigam, Cancer Research Through The Help Of Soft Computing Techniques: A Survey, Ijcsmc, Vol. 2, Issue. 4, April 2013, Pg.467 – 477
5.Abdelaal Ahmed Mohamed Medhat And Farouq Wael Muhamed, "Using Data Mining For Assessing Diagnosis Of Breast Cancer," In Proc. International Multiconfrence On Computer Science And Information Technology, 2010, Pp. 11-17.
6.Chang Pin Wei And Liou Ming Der, "Comparison Of Three Data Mining Techniques With Genetic Algorithm In Analysis Of Breast Cancer Data".[Online]. Available:Http://Www.Ym.Edu.Tw/~Dmliou/Paper/Compar_Threedata.Pdf
7.Jamarani S. M. H., Behnam H. And Rezairad G. A., "Multiwavelet Based Neural Network For Breast Cancer Diagnosis", GVIP 05 Conference, 2005, Pp. 19-21..
8.Gandhi Rajiv K., Karnan Marcus And Kannan S., "Classification Rule Construction Using Particle Swarm Optimization Algorithm For Breast Cancer Datasets," Signal Acquisition And Processing. ICSAP, International Conference, 2010, Pp. 233 – 237.
9.S. Chakrabarti, S. Roy, And M.V. Soundalgekar, "Fast And Accurate Text Classification Via Multiple Linear Discriminant Projection", The VLDB Journal The International Journal On Very Large Data Bases, 2003, Pp. 170–185.
10. S.L. Ting, W.H. Ip, Albert H.C. Tsang," International Journal Of Software Engineering And Its Applications, Vol. 5, No. 3, July, 2011"
11. Galit Shmueli; Nitin R. Patel; Peter" Data Mining For Business Intelligence: Concepts, Techniques, And Applications In Microsoft Office Excel® With Xlminer®, Second Edition"
12. Http://Www.Cancer8.Com/
13. Http://Www.Wcrf.Org/Cancer_Facts/5-Most-Common-Cancers.Php
14. Http://Www.Cancer.Gov/Cancertopics/Cancerlibrary/What-Is-Cancer
15. Http://Www.Medicalnewstoday.Com/Info/Cancer-Oncology/
16. Http://Www.Wcrf.Org/Cancer_Facts/5-Most-Common-Cancers.Php
17. Http://Www.Scribd.Com/Doc/9522057/An-Overview-Of-Soft-Computing-With-Its-Techniques-And-Application
18. Http://Dl.Acm.Org/Citation.Cfm?Id=1539227
19. Http://Www.Iconip2012.Org/S02.Pdf
20. Http://Www.Biomedcentral.Com/1471-2105/6/148
21. Http://News.Harvard.Edu/Gazette/Tag/Cluster-Of-Cancer-Cells/
22. Http://Www.Healthcommunities.Com/Cancer-Treatment-And-Care/Cancer-Staging.Shtml
23. Http://En.Wikipedia.Org/Wiki/Oncology#Progress_And_Research
24. Http://Www.Cancer.Net/Patient/All+About+Cancer/Cancer.Net+Feature+Articles/Treatments%2C+Tests%2C+And+Procedures/Understanding+Cancer+Research+Studies%2C+Part+II
25. K. Anuradha And K.Sankaranarayanan "International Journal Of Advances In Engineering & Technology, March 2012" Vol. 3, Issue 1, Pp. 84-91
26. M. E. Futschik, A. Reeve, And N. Kasabov, "Evolving Connectionist Systems For Knowledge Discovery From Gene Expression Data Of Cancer Tissue," Artif. Intell. Med., Vol. 28, Pp. 165–189, 2003
27. H. Midelfart, A. Lægreid, And J. Komorowski, Classification Of Gene Expression Data In An Ontology, Vol. 2199. Lecture Notes In Computer Science, Berlin, Germany: Springer-Verlag, 2001, Pp. 186–194
28. M. Banerjee, S. Mitra, And H. Banka, "Evolutionary-Rough Feature Selection In Gene Expression Data," IEEE Trans. Syst., Man, Cybern. C, Appl.

29. Ahmad M. Sarhan, "Cancer Classification Based On Micro Array Gene Expression Data Using DCT And ANN", Journal Of Theoretical And Applied Information Technology, Vol. 6, No. 2, Pp. 208-216, 2009
30. Bharathi And Natarajan, "Cancer Classification Of Bioinformatics Data Using ANOVA", International Journal Of Computer Theory And Engineering, Vol. 2, No. 3, Pp. 369-373, June 2010
31. Bo Li, Chun-Hou Zheng, De-Shuang Huang, Lei Zhang And Kyungsook Han, ""Gene Expression Data Classification Using Locally Linear Discriminant Embedding", Computers In Biology And Medicine, Vol. 40, Pp. 802–810, 2010
32. Xiaosheng Wang And Osamu Gotoh, "A Robust Gene Selection Method For Micro Array-Based Cancer Classification", Journal Of Cancer Informatics, Vol. 9, Pp. 15-30, 2010
33. Mallika And Saravanan, "An SVM Based Classification Method For Cancer Data Using Minimum Micro Array Gene Expressions", World Academy Of Science, Engineering And Technology, Vol. 62, No. 99, Pp. 543-547, 2010
34. K. Deb And A. Raji Reddy, "Reliable Classification Of Two-Class Cancer Data Using Evolutionary Algorithms," Biosystems, Vol. 72, Pp. 111–129, 2003.
35. H. Midelfart, J. Komorowski, K. Nørsett, F. Yadetie, A. K. Sandvik,And A. Lægreid, "Learning Rough Set Classifiers From Gene Expression And Clinical Data," Fundamenta Inf., Vol. 53, Pp. 155– 183, 2002.
36. F. Chu, W. Xie, And L. Wang, "Gene Selection And Cancer Classification Using A Fuzzy Neural Network," In Proc. 2004 Annu. Meet. North Amer. Fuzzy Information Processing Soc. (NAFIPS), Vol. 2, Pp. 555–559
37. S. Mitra, "An Evolutionary Rough Partitive Clustering," Pattern Recognit. Lett., Vol. 25, Pp. 1439– 1449, 2004
38. S. Bicciato,M. Pandin, G. Didon`E, Andc.Dibello, "Pattern Identification And Classification In Gene Expression Data Using An Autoassociative Neural Network Model," Biotechnol. Bioeng., Vol. 81, Pp. 594–606, 2003.
39. Http://En.Wikipedia.Org/Wiki/Bioinformatics
40. Www.Broadinstitute.Org/...2006/.../Classification_And_Clustering.Ppt
41. Http://Www.Ims.Nus.Edu.Sg/Programs/Microarray/Files/Singdiscv2.Pdf
42. Http://Cscjournals.Org/Csc/Manuscript/Journals/IJBB/Volume5/Issue4/IJBB-133.Pdf
43. An Extensive Survey On Artificial Neural. Network Based Cancer Prediction Using. SOFTCOMPUTING Approach. Manaswini Pradhan, Dr. Ranjit Kumar Sahu
44. Http://Ijcta.Com/Documents/Volumes/Vol2issue5/Ijcta2011020550.Pdf
45. Http://Www.Iula.Upf.Edu/Materials/040701wanner.Pdf
46. Sami Halawani And Rafat Parveen,International Journal Of Research And Reviews In Computer Science (IJRRCS), Vol. 3, No. 1, February 2012, ISSN: 2079-2557
47. Http://Www.Softcomputing.Es/Metaspace/Portal/7/274
48. Http://Www.Medwelljournals.Com/Abstract/?Doi=Ijscomp.2009.201.207
49. Http://Www.Scribd.Com/Chris_Cao_11/D/82513248-A-Proposed-Model-Of-Softcomputing-In-Diagnosing-Diseases-And-Prescribing-Herbal-Prescriptions-By-Oriental-Medicine
50. Www.La-Press.Com/Redirect_File.Php?...Soft-Computing-Cancer-Classification
51. Megha Gupta,Naveen Agarwal "Classification Techniques Analysis", NCCI 2010 -National Conference On Computational Instrumentation
52. Http://Www.Mrc-Lmb.Cam.Ac.Uk/Genomes/Madanm/Microarray/Chapter-Final.Pdf
53. Http://Www.Ijcaonline.Org/Volume19/Number5/Pxc3873087.Pdf
54. Www.Sciacademypublisher.Com/Journals/Index.Php/IJRRCS/.../581
55. Http://Metronu.Ulb.Ac.Be/Imacs/Papers/T3-I-114-0597.Pdf

