# Reviewing the Techniques of Disease Detection and Classification from the Challenging Medical Data

Chandru A.S
Asst. Professor,
Department of ISE,
NIEIT, Mysuru, India

Seetharam K, PhD
Principal and Professor,
Basava Academy of Engg,
Bengaluru, India

## ABSTRACT

The area of healthcare sector is now meeting a new challenge of data management. Owing to adoption of advance technology for patient-related services as well as diagnosis, a high-dimensional data is being generated. The biggest problems of such data are manifold e.g. i) they are much bigger in size that is difficult to be stored in physical servers, ii) they are massively growing in size with respect to increase of time, iii) they are of various forms and formats owing to be generated from multiple devices, and iv) there is larger dimensionality of uncertainty too. Owing to all these problems, it is almost impossible to apply the conventional data analysis algorithm for extracting teh knowledge. This paper discusses about the some of the recently adopted technique for analysis such medical data for an effective disease detection and classification with a contribution of exploring the research gap for the existing literatures.

## Keywords

Disease Detection, Disease Classification, Medical Data, Unstructured Data

## 1. INTRODUCTION

With the rise of advance technology, the existing healthcare systems are now adopting all the novel equipments for effective healthcare services. 10 years back the patient records were made on paper, but now the hospital uses a distribute file system for managing the clinical information of the patient. The biggest chain of healthcare system also has some of the updated medical equipments as well as softwares that assist in delivering quality medical services. Hence, the so called medical data is basically captured from Electronic Medical Record (EMR) that consists of different forms of data e.g. i) contact information about patient, ii) photograph of patient, iii) billing information, iv) prescribed medication information, v) information pertaining to surgery or some onsite medical treatment, vi) information related to next appointment, and vii) radiological data. Hence, it can be seen that different types of data are being collected by one hospital just for one patient. Hence, it can be imagined that on big healthcare facility may need to attend around 1000 patients everyday in different departments (Fig.1). Hence, such information increases with high leaps right from gigabytes to terabytes in one day and around petabytes in one month itself. Although conventional RDBMS storage management is still in practice to store some basic information. But it is near to impossible to perform advance datamining over such forms of data. The reasons are many. The first reason being the nature of files and their formats.

There are different formats of each radiological data which is not possible for advance data analytics. A closer look into formation of such data will show that such medical data are of heterogeneous and highly unstructured in nature. Owing to such property, it is not probable to relate the existing data analysis algorithm on such data directly. It is required that such data should be normalized first, which is again a huge task. Another significant problem is quantity of such data. Normally such data is highly massive in size, which leads to higher dependency of cloud storage system. As relying on physical storage system will be highly expensive one.



**Fig 1: Existing System of Data Generation in Healthcare**

At present the study carried out in the area of disease detection and classification from the medical data already exists. But the majority of the existing literatures speak about the simple classification technique using conventional data mining algorithms that cannot be implemented on the existing system. Hence, it is essential to understand first that which is the best technique discussed for disease detection and classification of medical data? Therefore, this paper discusses about the existing system about the techniques used in generation of medical data and reviewing the existing research-based techniques of data analysis. Section II discusses about the concept of medical data generation practiced at present in majority of the healthcare section followed by discussion of issues in it in Section III. Section IV discusses about the existing system. This section is further classified in two units, where the first unit discusses about existing literatures and second unit discusses about the most frequently used techniques for disease detection and classification. Section V discusses about research gap and Section VI summarizes the paper.

## 2. ABOUT MEDICAL DATA GENERATION

There are a few activities inside of the medical services group to exchange persistent restorative records crosswise over health services offices. These records hold the capability of giving general well being authorities the capacity to recognize and observe diseases pretense health threat to their groups. Instruments are expected to misuse the abundance of information and data contained inside of these records. Synthetic medical record information is required for improvement and assessment of new procedures.

Electronic Medical Records (EMRs) are advanced renditions of the paper outlines in clinician workplaces, facilities, and health centers. EMRs contain notes and data gathered by and for the clinicians in that office, facility center, or medical center as well as are normally used by suppliers for determination and treatment, enhance human services quality.
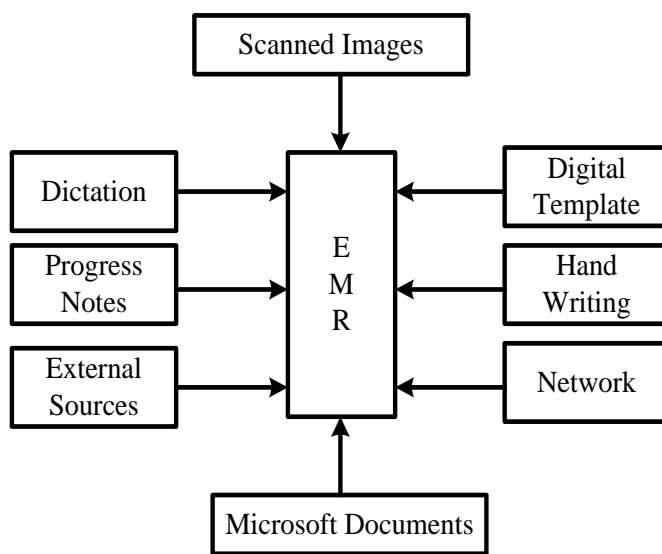


**Fig 2: Generation of Medical Data**

Personal Health Records (PHRs) contain the same sorts of data as EHRs—diagnoses, medicines, inoculations, family therapeutic histories, and supplier contact data—yet are intended to be situated up, gotten to, and oversaw by patients. Patients can utilize PHRs to keep up and deal with their medical data in a confidential, protected, and home observation gadgets, and patients themselves.

Electronic Health Records (EHRs) are assembled to go past standard clinical data accumulated in a supplier's office and are far reaching of a more broad point of view of an understanding's patient's care. EHRs contain data from every one of the clinicians included in a quiet's consideration and every single approved clinician included in an understanding's consideration can get to the data to give consideration to that patient. EHRs likewise impart data to other human services suppliers, for example, research facilities and specialists. EHRs take after patients – to the authority, the health facility, the nursing home, or even across the nation.

An EMR encloses the usual medical and clinical information accumulated in one supplier's office. Electronic Health Records (EHRs) go past the information gathered in the supplier's office and incorporate a more far reaching patient history. Case in point, EHRs are intended to contain and offer data from all suppliers included in a persistent consideration. EHR

information can be made, overseen, and counseled by approved suppliers and staff from crosswise over more than one medicinal services association. Dissimilar to EMRs, EHRs likewise permit an understanding's patient medical record to move with them— to other social insurance suppliers, Health center, healing centers, nursing homes, and even crosswise over states. An EMR is more useful than paper records because it permits suppliers to:

- Track information after some more time.

- Identify patients who are expected for preventive visits and screenings.

- Monitor how patients measure up to specific parameters. For example, immunizations and pulse readings,

- Improve general nature of consideration in a practice.

The data put away in EMRs is not effortlessly imparted to suppliers outside of a practice. An understanding's record may even must be printed out and conveyed via mail to masters and different individuals from the consideration group. EHR adoption can give medical care suppliers:

Precise and complete information around a patient's medical. This engages suppliers to give the best possible thought, whether in the midst of a normal office visit or in a medicinal crisis, by giving the data they have to assess a persistent present condition in the connection of the understanding's health history and different medications.

The capacity to rapidly give care. In an emergency, EHRs give moment access to data around an understanding's therapeutic history, hypersensitivities, and medicines. This can empower suppliers to settle on choices sooner, rather than sitting tight for data from test outcomes. The capacity to better arrange the health they give. This is particularly essential if a patient has a genuine or perpetual restorative condition, for example, diabetes. An approach to impart data to patients and their family parental figures. This implies patients and their families can all the more completely tune in choices about their medicinal services.

The fundamental objective of medical IT is to enhance the quality and security of patient consideration. The guarantee of completely acknowledged EHRs is having a solitary record that incorporates the greater part of a patient's medical data: a record that is up and coming, complete, and precise. This places providers in a superior position to work with their patients to use good judgment. EHRs can likewise signal conceivably unsafe medication communications (to help endorsing specialists investigate options before an issue happens), confirm meds and doses (to guarantee that drug specialists apportion the right medication), and diminish the requirement for possibly hazardous tests and techniques.

## 3. ISSUES IN ANALYZING MEDICAL DATA

Many surveys all over the world say that patients were willing to participate and support in the research work but they first want to let to know information use of their medical records. The patients were worried that their data can be used for insurance and medical purposes. They were also having concern about distribution of their sensitive information widely. These concerns have raised international efforts to increase the protection afford for data in medical records. Several federal legislative proposals were developed in order to address the issues. In recent past Indian researchers in their retrospective studies to be reviewed by using must require medical data. The Indian Council of Medical Research (ICMR) has provided the

guidelines to waiver study of minimal risk or in an emergency. Such waivers will get permit by MCI (Medical Council of India) if patient's identity is not revealed. However such proposals have to clear by ERB's, as there were few ERB's in India, which will slow down the research. The consent with the patient indirect or direct contact will interrupt in the research process. This may leads gap in the privacy with patients. Such kind of contact will cause former patient like social, psychological or other harm. The implemented strategies are increased the work load for researcher and ERBs, due to this dropout rate of proposed studies have increased, when investigators were unable to meet requirements. The researchers feel that the charitable organizations and government agencies are wasting public money by ERBs, when studies required doing with the help of multiple ethics review. Most of the publication Indian institutes were related with medical record review. Few of these institutes have ERBS and this type of research is not related to review of ethics.

## 3.1 Medical Data Heterogeneity
Medical data raw are heterogeneous and voluminous. Medical data will be obtained from several images, laboratory data, and interviews with patients, physician's interpretations and observations. Above components will bear on the prognosis, patients treatment, diagnosis and these will not be ignored. Many areas of Heterogeneity of medical data may be given by

## 3.2 Volume and Complexity of Medical Data
Medical data raw are heterogeneous and voluminous. Medical data will be obtained from several images, laboratory data, and interviews with patients, physician's interpretations and observations. Above components will bear on the prognosis, patients treatment, diagnosis and these will be used in data mining research and in many medical procedures used imaging as required diagnostic tool. Hence there is method to be developed for mining in image databases (much difficult than numerical database mining). For an example PET, EEG or ECG signals and MRI, image technology c produce GB (Giga Byte) of data in a day. The patient's organ image always will be given by the physician or other hospital information. This kind of heterogeneity analysis needs new tools and storage data capacity. It is difficult for unaided human in gigabyte record process. While handling with images will be easy for humans to recognize and grasp the basic data trends and calculate the rational decision. If the information is not in comprehensive format, the information will becomes less helpful.

## 3.3 Physicians' Interpretation
Signals, images or other clinical data, written in English (Unstructured), physical interpretation is difficult in standardization and mine. Some specialists with same discipline did not agree with unambiguous term used to describe patient's condition. They used to describe several names for single disease. They use other grammatical constructions for relationships in medical entries. It is presented that computer translations will take part in solution to process physicians interpretation. The machine / computer translation includes analysis (source language), transfer and generation (target language). The natural language has large set of expression, which leads endless job. The systems with current translation will analyze and translate only 10 words, which leads ambiguity.

## 3.4 Sensitivity and Specificity Analysis
All treatments and diagnoses in medicine were imprecise and they subjects to rate of error. The paradigm of medicine in measuring this type of error is specialty and sensitivity analysis. One have to distinguish between diagnosis and test in medicine.

Test is used to characterize patients many medical conditions. Diagnosis is used to test many syntheses and observations which points patho-physiologic process in patient. The diagnoses and tests were inexpensive (new test), where the hypotheses are expensive test. The comparison of prescribed any new test and other test value gives accuracy of test. The accuracy can be given as,

$$Accuracy = (TP+TN)/(TP+TN+FP+FN)$$

$$Sensitivity = TP/(TP+FN)$$

$$Specificity = TN/TN+FP$$

Where, TP & TN is true positive and true negative respectively. FP & FN is false positive and false negative respectively.

## 3.5 Poor Mathematical Characterization
Poor mathematical characterization is unique data mining feature where medical data structure are characterized poorly mathematically, compared to areas of physical science. The physical scientists take all the data which can put equations reasonable, formulas, models and equations used to reflect data among them. The structures of medicine contains images and word-descriptions, with few formal vocabulary constrains, image composition or relationships allowed in basic concepts. The fundamental medicine entries like ischemia, neoplasia or inflammation are real to physicians, entries like length, force or mass to physical scientists. But the medicine doesn't have formal structure for comparison where data miner will organize egression model, sequence analysis and clustering information.

## 4. RELATED WORK
Anomaly identification refers to discovering irregularity with the desired medical data pattern, which is called discovery, abnormality mining, and noising mining. This type of detection have numerous applications, such as restorative diagnostics medical data inconsistency recognition, misuse of credit card detection, equipment deficiency recognition and structural and unstructured data identification, system interruption discovery and a novel topic of content mining. As of now, this anomaly identification based on clustering, nearest neighbor method, statistical anomaly recognition, anomaly identification based on data theory and spectral theory of anomaly identification methods are used.

## 4.1 Existing Literatures
Chen et al. [1] has designed an innovative and efficient healthcare system to monitor and secure the huge medical data continuously. He proposed a novel healthcare communication system using Hadoop technology. The result indicates that the novel Hadoop based healthcare architectures are highly effective to develop wireless healthcare data systems. He also compared this result with 3 aspects like, huge data storage, lightweight digital sign, and relational and interactional database.

There are numerous open issues for Big Data management and analysis, specifically in the computational science and medical services fields. The investigation in omics filed is moving from a speculation headed to information driven methodology. Merelli et al. [2] has researched on efficient investigation and storage of Big Data. Bioinformatics can be seen as the "glue" for every one of these procedures. An awareness High Performance Computing (HPC) keys in Bioinformatics, huge data examination for computational science.

Zhang et al. [3] has introduced a novel classification methods based on a One-class kernel Principle Component Analysis (KPCA). It has been developed for different types of medical datasets. An efficient scheme was verified by using a breast

cancer data set as well as retinal image set. The found outcomes are associated with other traditional techniques; the proposed system gives better results compared to other techniques.

Behadada et al. [4] has proposed a novel technique to build semi-automatic fuzzy based rules to get a powerful and very accurate insight into cardiac arrhythmia. The experimented result shows clearly a marked improvement compared with the conventional algorithms and it obtains high accuracy of rate 93% and Interpretability of 0.646. Additionally, our novel method proposals more clearness and elasticity in the system finding and enable the decision making method.

Wang et al. [5] presented an anomaly identification and huge medical data analysis. Due the sampling theory as well as storage of huge medical data and inadequacy of uncompressed data, a new theory called compressive sensing was introduced and utilized in the anomaly identification techniques. This identification principle based on wavelet transform and statistic signal processing and control theory was assumed. The investigations for identifying humans behind a brick wall and gypsum built on UW radar signal accepted. It shows, the projected detection method could successfully recognize the reality of human being over structured and unstructured medical data.

Ahmed et al. [6] suggest a competent skin cancer prediction model based on data analysis and achieve an efficient method for feature extraction of the sigficant pattern form different medical data warehouse for prediction of skin cancer. This paper predict skin cancer with very lost cost, and it theatres a vital part in the medicinal diagnosis process. This novel method is experimented using Lotus notes. This method achieves an efficient and successful prediction of skin cancer.

Ungurean et al. [7] was performed a preliminary experiments on prediction of patients with parkinson's diseases and also he was proposed the uses of FLAME grouping techniques on speech signal and it also acquired from the patients. This method has been optimized for CBEA based processors in order to use intensive computing resources. The results shows that the FLAME methodology be integrated into a medical huge data systems and also it used algorithms for analysis of other types of clinical symptoms in order to achieve a better prediction of suspected PD patients.

Shouman et al. [8] investigates applying K-Nearest Neighbor (KNN) to help healthcare professionals in the diagnosis of heart diseases. It is one of the successful data mining method used in classification problems. It will enhance its accuracy in the diagnosis of heart disease patients. The result shows that applying KNN achieves higher accuracy than neural network ensembles in the diagnosis of heart diseases patients.

Jaafar et al. [9] was studied an automated techniques to distinguish and grade the seriousness of hard exudates is proposed. In this work, another choice bolster system for programmed location and evaluating of HEs has been proposed. The recognition procedure is in light of top-down picture division and nearby thresholding by a mix of edge discovery and area developing. The aftereffects of hard exudates discovery procedure were accepted in light of clinician hand-marked information (ground truth) with a general affectability of 93.2%. The unrivaled execution of this method recommends that it could be utilized for a PC supported mass screening of retinal ailments. Impact of parameter changes has been considered and the outcomes recommend that the conduct of the proposed technique is very productive.

Oyana et al. [10] was Studied over and gives an extra proof on this algorithms, that was intended to expand the general productivity of the first k-means clustering strategy—the Fast, Efficient, and Scalable k-means algorithm (FES-k-means). This method was tried utilizing two genuine datasets and one manufactured dataset. The advantages of this technique are that it creates groups like the first k-means system at a much speedier rate as indicated by runtime examination information; and it gives proficient investigation of extensive geospatial information with suggestions for infection mechanisms discovery.

Huang et al. [11] studies the additional feature choice and SVM recursive elimination (SVM-RFE) to explore the arrangement exactness of multiclass issues for Dermatology and Zoo databases. It will support vector machine (SVM) has great execution on grouping and expectation and is generally utilized on medical analysis. The exploratory results demonstrate that the order precision can be more than 95% after SVM-RFE highlight determination and Taguchi parameter streamlining for Dermatology and Zoo databases.

Huang et al. [12] have studied on the utilization of artificial neural system (ANN) models for anticipating myofascial torment manage after desiccated needling and to differentiate the farsighted capacity of ANNs and that of bolster vector machine (SVM) and numerous direct relapse (MLR). Diverged from the MLR and SVM models, the ANN demonstrate for the most part had humbler mean square slip (MSE) and mean preeminent rate botch (MAPE) values in the arranging dataset and testing dataset. Most ANN models had MAPE qualities running from 3.4% to 4.6% and most had selective standard exactness. Separated and the MLR and SVM models, the ANN demonstrate in this study was more right in suspecting patient reported BPI scores and had superior general execution records.

Adetiba et al. [13] complete an exploratory correlation of manufactured neural system (ANN) and bolster vector machine (SVM) troupes and their "nonensemble" variations for lung growth forecast. The histogram of situated angle (HOG) and neighborhood paired example (LBP) cutting edge highlight extraction plans were connected to concentrate delegate genomic elements from the encoded arrangements of nucleotides. The ANN outfit and HOG best fit the preparation dataset of this study with a precision of 95.90% and mean square blunder of 0.0159. The consequence of the ANN group and HOG genomic elements is promising for robotized screening and early discovery of lung disease.

Sharaf et al. [14] investigated and analyzed data are required to better understand the causes of the differences between young males and females and to plan new strategies to fight the major pernicious form of skin cancer. It uses ten time intervals (12 months each) and in order to do the comparison between the PLANN and Mani's method we used 10 inputs for the ten time intervals in PLANN instead of one so that we can compare the output of the PLANN with the second method.

Gunavathi et al. [15] was exhibited a Comparative Analysis of Swarm Intelligence Techniques for Feature Selection in Cancer Classification. The proposed work is connected on 10 diverse benchmark datasets and inspected with SI systems. The trial results demonstrate that the outcomes got from k-NN classifier through SFLLF highlight determination strategy beat PSO, CS, and SFL.

Alshamlan etal. [16] suggest the first endeavor at applying ABC calculation in investigating a microarray quality expression profile. What's more, we propose an imaginative element choice calculation, least excess most extreme importance (mRMR), and

consolidate it with an ABC calculation, mRMR-ABC, to choose useful qualities from microarray profile. The test results demonstrate that the proposed mRMR-ABC calculation accomplishes exact grouping execution utilizing little number of prescient qualities when tried utilizing both datasets and contrasted with already recommended routines. This demonstrates that mRMR-ABC is a promising methodology for understanding quality choice and disease order issues.

Volna et al. [17] manages ECG forecast in view of neural systems characterization of diverse sorts of time courses of ECG signs. This work actualized two location frameworks that have been made with utilization of neural systems. Every single trial result from both of the proposed classifiers are commonly looked at in the conclusion. The proposed work tried different things with the new technique for time arrangement straightforward expectation taking into account fluffy change with semantic IF-THEN guidelines. Preparatory results show fascinating results in view of the novel ability of this methodology bringing normal dialect elucidation of specific expectation, that is, the properties of time arrangement.

## 4.2 Frequently Used Techniques
This segment will talk about the as often as possible utilized procedures of sickness location, order, and development examination on the medicinal information.

### 4.2.1 Support Vector Machine
SVMs are another strategy suitable for paired order assignments, which is identified with and contains components of non-parametric connected insights, neural systems and machine learning. Like traditional systems, SVMs additionally order an organization as dissolvable or wiped out as per its score esteem, which is a component of chose monetary proportions. Yet, this capacity is neither direct nor parametric. SVMs are another promising non-direct, non-parametric arrangement strategy, which effectively demonstrated great results in the medicinal diagnostics, optical character acknowledgment, and electric burden anticipating and different fields.
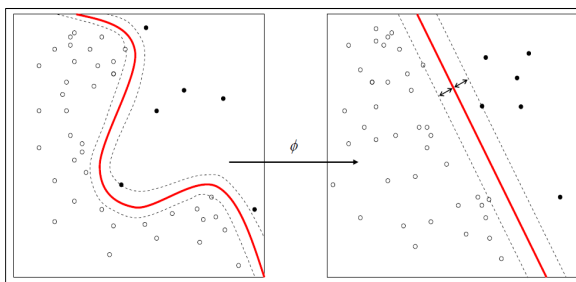


**Fig 3: "Non-Linear and Linear Classification" in SVM**

### 4.2.2 Artificial Neural Network
(ANNs) are created taking into account mind structure. Like the cerebrum, ANNs can perceive examples, oversee information and learn [3]. They are made by artificial neurons (Fig.4), which actualize the substance of natural neurons.
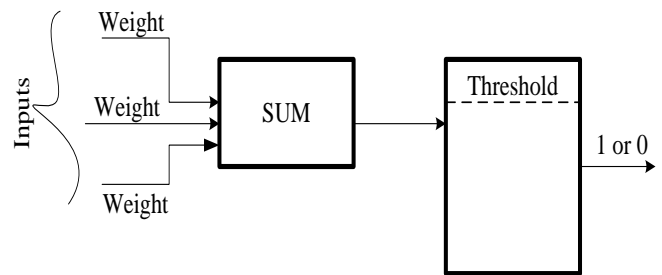


**Fig 4: Block Diagram of ANNs**

ANNs is turning into a standout amongst the most famous distinct options for traditional factual displaying. It is really imagined as a progressed summed up direct model. We have seen different utilizations of ANN used in diverse investigative subjects like designing, financial matters, environment, and wellbeing, among others.

### 4.2.3 Genetic Algorithm
Genetic Algorithm is an enhancement systems enlivened by regular determination and characteristic hereditary qualities. Dissimilar to numerous inquiry calculations, which perform a neighborhood, ravenous pursuit, GA is a stochastic general hunt technique, fit for adequately investigating vast inquiry spaces. A hereditary calculation is essentially made out of three administrators: proliferation, hybrid, and change. As a first stride of GA, an introductory populace of people is created indiscriminately or heuristically. The people in the hereditary space are called chromosome. The chromosome is an accumulation of qualities where qualities can for the most part be spoken to by distinctive systems like twofold encoding, worth encoding, change encoding and tree encoding. Quality is the essential building square of the chromosome. Locus is the position of specific quality in the chromosome.
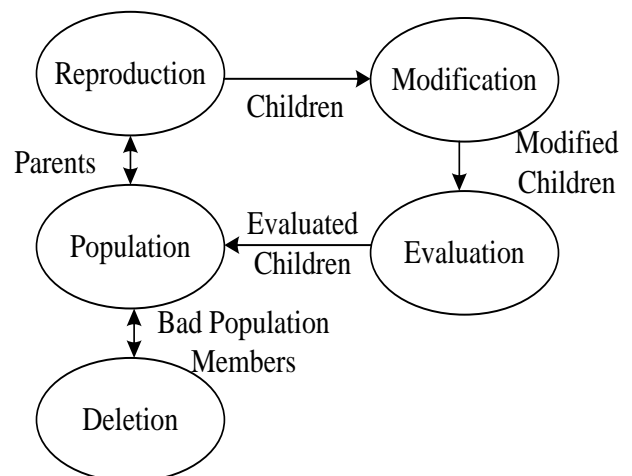


**Fig 5: Genetic Algorithm**

### 4.2.4 Fuzzy Logic
Fuzzy logic is an Artificial Intelligence strategy which can imitate human personality as far as rough thinking as opposed to being careful. A Fuzzy Set has values with incomplete enrollment alongside the fresh values. Fluffy Sets are helpful in setting up conditions which are loose in definition through halfway enrollment values. Components in fluffy set can cover, so a given fresh esteem can have a place with various fluffy sets with

distinctive enrollment degrees in every set. The multifaceted nature of therapeutic practice makes conventional quantitative methodologies of investigation unseemly. In prescription, the absence of data, and its imprecision, and, ordinarily, opposing nature are normal actualities. Fluffy set hypothesis and fluffy rationale are an exceedingly suitable and pertinent premise for creating information based frameworks in pharmaceutical for undertakings, for example, the understanding of sets of restorative discoveries, disorder separation in eastern prescription, analysis of illnesses in Western medication, blended determination of incorporated western and eastern solution, the ideal choice of therapeutic medications coordinating western and eastern drug, and for ongoing observing of patient information.
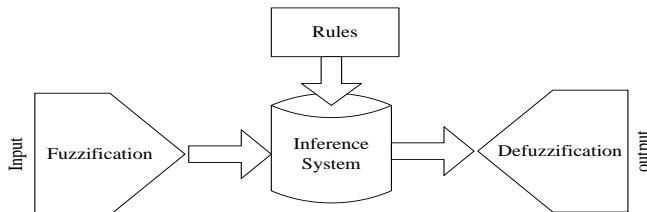


**Fig 6: Fuzzy Interface System**

## 5. RESEARCH GAP

Generation of effective information from high-dimensional medical data leads to a challenges in performing precise analysis towards effective diagnosis. After reviewing some of the recent literatures in disease detection and classification for medical data, we have come to identify following research gap,

- It has been seen that majority of the techniques have adopted Support vector machine for diagnosis of the disease from the medical data. However, the outcomes of such study show less transparency. Moreover, usage of multiclass SVM classifier is less to found in the existing system for performing classification techniques.

- Neural network has also been the technique of an effective disease detection and diagnosis. Although it shows consistent improved results in its each epoch, but owing to heterogeneity in the medical signals (or inputs), the regression results is found with significant errors. Moreover the diagnosis precision depends on the amount of dataset. More the training data exists, more accuracy. Hence, the system is more time consuming just to arrive at the correct outcome of diagnosis.

- Lack of an effective optimization technique is another research gaps. Some of the studies have used genetic algorithm but still an effective optimization is missing owing to selection of improper fitness function and inefficient chromosome block selection. Moreover none of these techniques implemented till date ensure global optimum. The consistency of optimization is quite poor in this technique, hence, the outcome may sound good theoretically but not in real-time.

- A closer look into the implementation techniques being carried out in fuzzy logic for an effective analysis of medical data will show that the effectiveness of an outcome depends largely on the fuzzy inference rules. Such inference rules are normally logical rules created by user. Hence, such rules are quite effective for simple problems of medical data diagnosis but not for complicated data diagnosis. For an example, it is feasible to set a rule to perform detection, but in order

to perform classification to large degree; it requires higher flexibility in rule design. Hence in case of queried data not matching with rules will be treated as false positives.

## 6. CONCLUSION

This paper has discussed about existing techniques of data analysis of the medical or healthcare sector. The paper discusses that with the increasing usage of advance technologies, the existing system offers greater deal of flexibility to capture, record, and manage various clinical data. Unfortunately, such clinical data keeps on rising almost every day resulting in performing data analysis. This paper discusses about the existing techniques found most frequently by the researcher and ultimately finds a research gap. Our future direction of the work will be to address the existing research gap by introducing a novel model of data analysis for medical data.

## 7. REFERENCES

[1] Chen, H., and Fu, Z.2015. Research Article Hadoop-Based Healthcare Information System Design and Wireless Security Communication Implementation. Hindawi Publishing Corporation Mobile Information Systems, Article ID. 852173, pp. 9

[2] Merelli, I., Sánchez, H. P., Gesing, S., and Agostino, D. D.2014. Review Article: Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives. Hindawi Publishing Corporation Bio Med Research International, Article ID 134023, pp. 13

[3] Zhang, Y., Zhang, B., Coenen, F., Xiao, J., and Lu, W.2014. One-class kernel subspace ensemble for medical image classification. Journal on Advances in Signal Processing, Vol.17

[4] Behadada, O., Trovati, M., Chikh, M. A., and Bessis, N.2015. Big data - based extraction of fuzzy partition rules for heart arrhythmia detection: a semi - automated approach. Concurrency and Computation: Practice and Experience

[5] Wang, W., Lu, D., Zhou, X., Zhang, B., and Mu, J.2013. Statistical wavelet-based anomaly detection in big data with compressive sensing. EURASIP Journal on Wireless Communications and Networking, No. 1, pp. 1-6

[6] Ahmed, K., Jesmin, T., Rahman, M. Z.2013. Early Prevention and Detection of Skin Cancer Risk using Data Mining. International Journal of Computer Applications (0975 – 8887), Vol. 62, No.4

[7] Ungurean, I., Gaitan, N-C.2012. Speech Analysis for Medical Predictions based on Cell Broadband Engine, 20th Europian Signal Processing Conference (EUSIPCO)

[8] Shouman, M., Turner, T., and Stocker, R.2012. Applying k-Nearest Neighbor in Diagnosing Heart Disease Patients. International Journal of Information and Education Technology, Vol. 2, No. 3

[9] Jaafar, H. F., Nandi, A. K., and Nuaimy, W. A.2011. Automated Detection and Grading of Hard Exudates from Retinal Funds Images. 19th European Signal Processing Conference (EUSIPCO)

[10] Oyana, T.J.2010.Research Article A New-Fangled FES-k - Means Clustering Algorithm for Disease Discovery and

Visual Analytics. Hindawi Publishing Corporation EURASIP Journal on Bioinformatics and Systems Biology, Article ID 746021, pp. 14

[11] Huang, M-L., Hung, Y-H., Lee, W. M., Li, R. K., and Jiang, B-R.2014. Research Article SVM-RFE Based Feature Selection and Taguchi Parameters Optimization for Multiclass SVM Classifier. Hindawi Publishing Corporation e Scientific World Journal, Article ID 795624, pp. 10

[12] Huang, Y-T., Neoh, C-A., Lin, S-Y., and Shi, H-Y.2013. Research Article Comparisons of Prediction Models of Myofascial Pain Control after Dry Needling: A Prospective Study. Hindawi Publishing Corporation Evidence-Based Complementary and Alternative Medicine, Article ID 478202, pp. 8

[13] Adetiba, E., and Olugbara, O. O.2015. Research Article Lung Cancer Prediction Using Neural Network Ensemble with Histogram of Oriented Gradient Genomic Features. Hindawi Publishing Corporation e Scientific World Journal. Article ID 786013, pp. 17

[14] Sharaf, T., and Tsokos, C. P.2015. Research Article Two Artificial Neural Networks for Modeling Discrete Survival Time of Censored Data. Hindawi Publishing Corporation Advances in Artificial Intelligence. Article ID 270165, pp. 7

[15] Gunavathi, C., and Premalatha, K.2014. Research Article A Comparative Analysis of Swarm Intelligence Techniques for Feature Selection in Cancer Classification. Hindawi Publishing Corporation e Scientific World Journal, Article ID 693831, pp. 12

[16] Alshamlan, H., Badr, G., and Alohali, Y.2015. Research Article mRMR-ABC: A Hybrid Gene Selection Algorithm for Cancer Classification Using Microarray Gene Expression Profiling. *Hindawi* Publishing Corporation BioMed Research International, Article ID 604910, pp. 15.

[17] Volna, E., Kotyrba, M., and Habiballa, H.2015. Research Article ECG Prediction Based on Classification via Neural Networks and Linguistic Fuzzy Logic Forecaster. Hindawi Publishing Corporation e Scientific World Journal, Article ID 205749, pp. 10.