CHAPTER 1

1.1. Introduc on

Decision making in medical domain is a special one as there is the question of life and death. So, accurate diagnosis is solicited for further treatment planning. Medical diagnosis is the process of attempting to determine or identify a possible disease or disorder from an available set of symptoms and laboratory findings. It is an expertise as it involves a complex trouble-shooting procedure and its solution requires skill as well as experience. Due to the workload pressure, uncertainty in information, ambiguous nature of the problem domain; experts use some shortcut techniques (heuristics) without exploring every pros and cons. Everyone are prone to take mental shortcuts while thinking through difficult problems and physicians are no exception.

The challenge of reaching the correct diagnosis becomes harder without all the necessary clinical information. According to an analysis of AIM (Archives of Internal Medicine) in July'2005, one in three diagnostic errors, results solely from physician's cognitive mistakes. Moreover, there are different sources of uncertainties which affect medical diagnosis are shown in Table – I. These sources of uncertainties may lead to the following difficulties:

- Different diseases may have some common symptoms and/or secondary symptoms may be prominent which hide the original leaving the diagnosis process difficult.
- The patients may also have more than one medical problem making the diagnosis process difficult and confusing.
- Conflicting diagnosis due to different interpretations of symptoms; experts' opinion differ.
- Much of the observations are subject to error and the error correction method requires many assumptions that do not always hold in practice.

Table 1.1. Sources of Inexact Informa on

1. Problem domain : a) Lack of precise numeric aspiration levels;
b) Lack of appropriate or available well defined algorithms.

2. Pa ents: a) Lack of adequate data;
b) Inconsistency of data;
c) Subjective reply;
d) Reply in fuzzy terms;
e) Fumbling answer;
f) No information.

3. Parents / Guardians : a) Lack of adequate data;
b) Inconsistency of data;
c) Subjective reply;
d) Reply in fuzzy terms;
e) Fumbling answer indicating lack of confidence;
f) Ignorance.

4. Doctors: Matching of similar rather than identical situations: model in equivalence.

5. Laboratory tests / technicians: Imprecision in measurement during clinical / pathological / radiological tests : instrumental / technician's error; impurity with the chemicals / glassware used.

6. Symptoms: a) Most of the symptoms are valid for more than one adverse situation;
b) Information hiding: a secondary symptom may be more prominent than a primary symptom.

7. **Laboratory** results are not available.

The large variety and complexity of the medical diagnosis leads to the need of automated intelligent decision support systems. Starting from MYCIN in the early 70s', a good number of such automated intelligent systems have been developed with varying degrees of success in the medical floor. In the recent years soft computing techniques are being exploited for developing intelligent systems for biomedical problems including diagnosis as these techniques are tolerant to imprecision, uncertainty, vagueness and partial truth and close to human like decision making.

Soft computing is a consortium of different methodologies like fuzzy logic, artificial neural networks, roughest, and evolutionary algorithms. We used these techniques to some biomedical problems such as breast cancer, hepatitis disease, cardiac arrhythmia, chronic kidney disease etc. Regarding the hybrid combinations, various combinations have been attempted for better results as well as generality and simplicity of the methods. Broadly, two approaches – logical and black box approaches have been attempted. The logical approaches include supervised learning, unsupervised learning, feature selection and theoretical studies. The black box approaches consist of ANN and evolutionary computing. Regarding the issue of uncertainty we attempted to explore rough set theoretic approach. The issue of missing values has not been taken care of by most of the previous works; they exclude the instances having missing values. We tried to address the missing value problem using arithmetic mean, regression and multiple imputation techniques. The issues related to feature selection and reduction are vital in medical decision making. So, this issue has also been explored with various approaches. The issues like applicability of different techniques and tools, innovations of new methodologies, comparative studies etc. have also been considered. Attempts have been made to compare the results with 'golden standard'.

Artificial intelligence (AI) in medical domain significantly uses various soft computing techniques, which has been proved to be helpful in performing clinical diagnosis and suggesting treatments. The history of modern computer has successfully elaborated about the process of developing an artificial thinking mechanism electronically, which can take decisions whenever required, if there is a proper supply of information. Artificial intelligence mechanism can detect meaningful relationship in a set of data and is used in many clinical situations i.e. in diagnosing, treatment planning and predicting the results. Consequently, the doctors, scientists and researchers were greatly allured by the potential of such technology. At present, this technology has been mostly utilized in making decisions in the field of medical science. Numerous intelligent attempts and their applications have been made in

medical information technology. The process of storing and analyzing large scale medical information is necessary to assist doctors and clinicians in different fields like diagnosing and managing the disease and providing some expert opinion.

If we consider a patient, who has various symptoms of a disease, comes for medical check-up in a health center; he/she is usually examined by junior doctors who possess less clinical know-about, whereas on the other hand the accumulation of clinical knowledge is generally seen among the senior or experienced doctors. It is possible, that this can be one of the major consequences of medical decision error. The effect of this type of error may be considered as one of the causes of increasing mortality, morbidity and economic loss.

In present scenario, clinicians are working hard with information overload. It is predicted that nearly two million information pieces are used by us in the various decision makings and that biomedical knowledge is doubled after every two decades [1]. However accurate and sufficient information may not always be accessible. Scarcity of domain expert is also a major issue while diagnosing and managing diseases.

Health, nutrition and education are primary traits for the improvement of human resource. The altogether improvement of any area is dependent on these criteria. Unluckily, several reasons, like – unavailability of domain experts, poor funding, unsatisfactory performance of medical professionals, poor medical facilities in rural areas, are the barriers for the development. In this work, the alarming mortality rate, due to breast cancer, cardiac arrhythmia and hepatitis diseases have been discussed in chapter 2. We are inspired to think of such an intelligent system which can provide support to reduce the mortality rate to some extent with the assistance of soft computing techniques. Researchers are occasionally using modeling techniques and knowledge of domain to utilize the various known facts in order to solve complicated problems. This kind of system may be highly beneficial and improve the information retrieval process as well as accurate diagnostic performance, if such methods are implemented in real fields. The gaps between clinicians and patients may be bridged by this work.

Researchers and scientists use different statistical approaches to discover knowledge from huge data set during analysis. The objective is to find out any relationship among different parameters of the data set. By looking at the raw data and searching for such relationship using statistical approaches is very much difficult, whereas learning systems are able to characterize complex relationship efficiently. The diagnoses of diseases, under study, are basically applications of artificial

intelligence medicine domain. Every individual can be served more by a decision support system adopted for their respective interest.

These are the reasons which motivate us to begin with a study on different diseases such as breast cancer, cardiac arrhythmia, hepatitis disease, and kidney disease, which inspired us to develop intelligent models that help clinicians to diagnose these diseases.

1.2. Decision Support System (DSS) in Disease Diagnosis

1.2.1. Needs

During the course of the research work, we have found numerous reasons for which there can be an application of DSS or expert system in the field of medical diagnosis. The reasoning may be extended to the following areas:

- **Feasible and Aiding in Diagnos c Process**

  It may be very much useful if DSS is found feasible and has the ability to aid in various diagnostic purposes during emergency cases.

- **Aiding Diagnosis and Be er Healthcare Management**

  Proper healthcare management and better treatment planning are needed for augmenting survivability. Outcomes of the laboratory findings are also required in time. DSS may also prove its efficiency in this respect.

- **Aiding in the Retrieval Process of Informa on**

  This is highly resourceful in the procedure of retrieval of medical information from large database of problem related to domain in reasonable time. It helps medical professionals in retrieving information related to diseases based on some search criteria, which is very much useful while diagnosing and managing diseases.

- **Helping on Interpreta on of Medical Images**

  Different medical images, such as E.C.G. reports, X-Rays, CT scan, MRI etc., are to be interpreted to diagnose various diseases. DSS may prove its efficiency in doing so.

- **Observant**

  Features may be added to alert patients as well as doctors if any such situation arises for example a laboratory finding may cross the limit. So DSS can be used as a good observer.

The purpose of developing decision support system (DSS) in medical diagnosis is to assist doctors in making decisions accurately while diagnosing and managing diseases based on available information. Correct diagnosis of diseases and effective treatments are necessary for providing good quality healthcare service. Poor decision in medical domain may lead to fatal situation, which is undesirable. Even the most technically sound health institution in India does not use data mining process to extract knowledge in diagnosing diseases though DSS may be helpful in this area. A complete automated intelligent system may help in making decisions and may provide better healthcare management. Moreover to diagnose a disease accurately by an expert a huge amount of data needed to be analyzed in a time constraint, which is sometimes unmanageable. In this context an automated system may be helpful. Implementation of automated intelligent system in medical domain increases the speed as well as reliability of prediction of diseases. Medical professionals are overburdened and consequently suffering from huge mental stress in making decisions. DSS may prove itself as an expert while handling complicated cases.

DSS may provide good quality healthcare service which is cost effective. Up-to-date clinical knowledge as well as efficiency may satisfy expectations of patients. It certainly improves the efficiency of healthcare centers, quality, speed, safety, patient's care in reduced cost.

1.2.2. Usefulness of DSS

Some of the usefulness of DSS in medical domain are noted below:

- Patient's care can be improved.
- Efficiency of the health service provider will be increased.
- Difference in interpretations of symptoms can be avoided, which results in diagnosis without any conflict.

- Expert's decision of diagnosis and advice, based on updated information, can be provided automatically.
- Better health service may be provided with reduced cost.
- Can be implemented easily in all areas without much investment, so people of large scale will be benefited.
- Patients may interact with the system.
- DSS can be used to impart training and medical knowledge.

1.3. Objec ves of the Study

1.3.1. Objec ves

The objective of the research work is to develop intelligent models to diagnose diseases such as breast cancer, cardiac arrhythmia, and hepatitis disease using soft computing techniques. Hybridization of data mining approaches with artificial intelligence to improve the performance of models. Data mining techniques, soft computing, genetic algorithm, rough set theory, artificial neural network, and any combinations of these are to be tried in developing automated system for prediction of the said diseases efficiently.

Steps in data preprocessing should be taken care of for the development of any model. Different processes for dimensionality reduction by selecting important features from a data set are to be tried to extract important features that would be worthwhile while final decision is made by doctors.

Another objective of our study is to analyze different techniques to handle missing values. Missing values in a data set should be taken care of otherwise it may cause distorted analysis or may produce biased results. Techniques, which are to be studied, are:

i) List-wise deletion

ii) Pair-wise deletion

iii) Arithmetic mean imputation

iv) Regression imputation

v) Multiple imputations.

Decision has to be taken in selecting benchmarking parameters to measure the predictive performance of an intelligent model. Benchmarking parameters such as precision, correct classification accuracy, sensitivity, specificity that is the whole confusion matrix, area under ROC curve, Kappa statistic, lift, cumulative gain, and probability threshold are to be studied for choosing the appropriate parameters to judge the models to be proposed.

Finally, comparisons of developed models with the previously proposed model have to be done in order to judge the performance in terms of bench marking parameters. Lowest performance of the intelligent model for medical diagnosis should have to be noted as it is an important factor especially in this domain.

1.3.2. Area of the Study

Attempts have to be made to develop intelligent automated systems to diagnose following diseases:

a) Breast Cancer disease

b) Cardiac Arrhythmia

c) Hepatitis Disease

Any attempt towards the development of such system would be helpful in our society. Based on the historical profiles, symptoms and laboratory findings, knowledgebase can be formed, which may help clinicians extraordinarily in terms of efficiency and performance in decision making.

Different feature selection techniques and classifiers have to be attempted to deal with the disease diagnosis problems. Performances of models are to be measured in terms of some benchmarking parameters such as correct classification accuracy, sensitivity, specificity, and are under ROC curve. The highest, lowest, and average behaviors of the methods are to be recorded as these are important parameters, to be judged for a DSS before using it.

Outliers in the data set are to be taken care of to avoid any adverse effect during analysis of data. Some steps are taken after consulting physician to manage outliers in hepatitis disease data set [2].

Attempts are to be made to implement appropriate technique to handle missing values in the data set. For example, in UCI hepatitis disease data set [3] there are

number of missing values. Attempts have to be made to implement suitable technique to handle the situation effectively.

1.4. Overview of Breast Cancer Disease

Formation of malignant cells in breast cells causes breast cancer. Cancer cells in the breast of a patient suffering from breast cancer become abnormal and multiply uncontrollably causes formation of tumor (mass of abnormal tissue). Delayed treatment may cause spread of malignant cells to the other parts of body. Breast cancer can be developed in male also, which is rare. Less than 1% of all breast cancer cases are found in male. Previously, the mortality rate of breast cancer cases in men was high due to lack of knowledge or awareness, and delayed diagnosis.

Tumors in breast are of two types – 'benign', which is non-cancerous and 'malignant', which is cancerous. Benign tumors are not dangerous for tissues surrounding it. Patients are suggested by doctors to leave it undisturbed rather than removing. The tumor is operated only when it causes pain or other problematic symptoms. On the other hand, malignant tumors are dangerous and also invasive. They invade and affect the cells surrounding it. Different test are there to find out the malignancy of tumors.

Some facts about breast cancer are as follows:

- In United State one in every eight women is diagnosed with breast cancer in her lifetime.
- It is the second leading cause of death in women after lung cancer.
- More than 200,000 new breast cancer cases for both sexes are recorded in US in the year 2012 [4].
- Early diagnosis of the disease results in survival of more than 2.5 million breast cancer patients.
- According to World Health Organization (WHO) it is the most common problem in woman worldwide.

Depending on the position where the cancer starts in breast, there are two kinds of breast cancer – i) ductal and ii) lobular. Ductal cancer is when the cancer starts in ducts that carry milk to the nipple. Lobular cancer is when the cancer begins in milk producing gland (lobule). Other types of breast cancer are rare. Spreading of breast cancer is due to the mixing of cancer cells into blood and lymph system. Sign and symptoms of breast cancer are as follows:

- A lump in the breast that appears different from surroundings.
- Change in the size and shape of breast.
- Change in the shape of nipple.
- Peeling and scaling in areola or breast skin.
- Reddish appearance on breast skin.

According to doctors hormonal, lifestyle, and environmental factors are the main reasons of causing breast cancer. Moreover 5% to 10% breast cancer cases are identified as genetic. Different risk factors of causing breast cancer are:

- Women are at high risk than male to suffer from this disease.
- If already infected with cancer in one breast, the chance of growing cancer in other breast is at high risk.
- The presence of the disease in the family history increases the risk of happening it.
- Radiation therapy received to chest at an early age may cause breast cancer.
- Obesity increases the risk of breast cancer.
- For women, if period starts at an early age (before age of 12 years), the risk of breast cancer may be increased.
- Menopause at an older age also causes breast cancer.
- Hormonal therapy after menopause is also an important reason of happening breast cancer.

Some factors such as healthy diet, maintaining proper weight, avoiding hormonal therapy, breast feeding, self examination of breast, avoiding consumption of alcohol may reduce the risk of developing breast cancer. Early diagnosis of breast cancer is needed for survival. An automated system may be helpful to mitigate the problem.

1.5. Overview of Cardiac Arrhythmia

Sinoatrial node also known as sinus node or simply SA node can be considered as ‘natural pacemaker’ as it initiates electrical impulses, which passes through heart

causing heart beat. Normally heart rate of one ranges from 60 to 100 beats per minute. There are four chambers in heart. The two upper chambers are known as atria (atrium in singular) and two lower chambers are known as ventricles. There are also four valves in heart, which are able to open and close in order to allow blood to flow in one direction during contraction of heart. Electrical impulses generated from SA node follows a particular path, which causes the heart to pump properly in regular manner. Any abnormality in the sequence of electrical impulses leads to abnormal heart beats known as cardiac arrhythmia. Cardiac arrhythmia is of two types:

- Bradycardia – if the heart rate is slow (less than 60 beats per minute.
- Tachycardia – if the heart rate is fast (greater than 100 beats per minute).

Different internal reasons of causing cardiac arrhythmia are given below:

- Under certain conditions almost all heart tissues can start electrical activity to generate beats.
- Interruption of electrical impulses in the transmission pathway.
- Sometimes SA node generates irregular electrical impulses causing irregularities in heart beats.

Cardiac arrhythmia may affect the heart's ability to pump blood properly, which in turn causes improper function of lungs, brain other organs of human body. Different external risk factors causing cardiac arrhythmia are as below:

- consumption of alcohol
- excessive smoking
- mental stress
- high blood pressure
- diabetes
- thyroid problem may also cause heart to beat irregularly

Some of the sign and symptoms of cardiac arrhythmia are:

- chest pain
- feeling tired

- shortness of breathe
- loss of consciousness
- sweating
- palpitations
- sudden weakness
- difficulties in doing exercise

Some facts about cardiac arrhythmia are as follows:

- It is one of the major problems worldwide. About 33.5 million people in world (20.9 million male and 12.6 million female) are suffering from atrial fibrillation in the year 2010 [5].
- In North America and Europe 1-2% of population are suffering from atrial fibrillation [6], [7].
- Study on atrial fibrillation (AF) in India was reported first time in the year 1995 in which they examined 984 people of Himalayan village. They observed only 0.1% prevalence of AF, which is too low [8]. This is due to the fact that the participants were of good health, examined by single ECG, only 6 participants are of age greater than 65 years.
- Another study made in West Birmingham Atrial Fibrillation project, reported AF prevalence of 0.6% in Indian perspective [9].
- According to REALIZE AF study and IHRS AF registry the average age of patients with AF are 60 and 54 years respectively [10].

Cardiac arrhythmia can be diagnosed by analyzing ECG (echocardiogram), EKG (electrocardiogram), chest X-ray images, holter monitor, tilt table test. Some other tests are also available to examine the occurrence of arrhythmia. Severe cardiac arrhythmia may cause heart attack or even death. For this reason early detection of heart disease and proper medication is absolutely necessary to prevent any bad happenings. An automated intelligent system may be helpful for this purpose.

1.6. Overview of Hepa s Disease

Hepatitis is a disease, which causes an inflammation of liver. It may be of both infectious or of noninfectious type. Consumption of alcohol, drugs and some diseases are responsible for causing noninfectious type of hepatitis. Viruses which are causing infectious type of hepatitis are – hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis D virus (HDV), and hepatitis E virus (HEV). In United States HAV, HBV and HCV are the main reasons of hepatitis.

HAV is highly contagious and spreads due to poor sanitation system, unhygienic environment, contamination of water and food, contact with fecal material etc. The virus is excreted in the stool of an infected person, which in turn may contaminate water, food or other products on ground. Hepatitis – A is transmitted through fecal – oral path. Mucosal exposure to body fluid or blood of one infected with Hepatitis – B is the main reason of causing the disease. Only supportive treatment can be provided to prevent major liver damage by reducing Hepatitis B viral replication. HCV is transmitted parenterally, generally through injection drug use. 20% of patients infected with hepatitis – C develop cirrhosis of lever and 5% of those will develop lever cancer. [11].

Recently, viral hepatitis is a global burden. In United States routine vaccination of children reduces incidence of hepatitis – A from 21.1 cases per 100,000 population to 2.5 cases per 100,000 population in the year 2003 [12]. According to CDC, in US number of new patients infected with HBV, and HCV in the year 2012, are 19,000, and 21,869 respectively [12].

The highest prevalence of hepatitis – E is found in South and East Asia regions. 60% of the global incidence of hepatitis – E is in this region. In India, number of patients infected with viral hepatitis in the years 2012 and 2013 are 119,000 and 290,000 respectively, which is really alarming [13].

Liver helps in purifying blood, producing some vital substances such as protein, vitamin etc., storing sugars, fats, vitamins. Inflamed liver is not able to do these functions properly. As a result of these different signs, symptoms may occur which are listed below:

- Felling tired or weak
- Fever

- Loss of appetite
- Vomiting
- Nausea
- Jaundice

People, those are at high risk of infecting with viral hepatitis are:

- Healthcare service providers
- People of Asia and Pacific Island
- People working at sewage and water treatment plant
- Intravenous drug users
- HIV patients

Viral hepatitis is diagnosed by testing blood of infected patients. Patients are advised to test blood for lever enzymes, antibodies of hepatitis virus and other parameters to detect the disease. Decision has to taken by evaluating these test reports. A decision support system may be helpful in this regard.

1.7. Bene t Expected from the Proposed Study

It is evident that automated intelligent systems to diagnose diseases such as breast cancer, cardiac arrhythmia, hepatitis disease, and kidney disease would be helpful to assist physicians in taking decisions. The proposed systems may be helpful in the remote and rural areas, where the health centre is suffering from acute crisis of domain expert. The proposed system may also be helpful for physicians, who are engaged in district or sub-divisional hospitals or in private practices. Another advantage of using such systems is that it can be implemented without investing much as the proposed system is cost effective. A computer or a laptop with power supply is sufficient to form 'mobile healthcare unit'. The proposed model may also be useful to medical students.

Domain experts are overloaded with the increasing population as well as increasing patients with increasing diseases. These automated systems may reduce their burden and at the same time efficiency will be increased.

These systems may be helpful for early diagnosis of diseases and better treatment planning, which are keys for survival of patients.

Any knowledge of data mining, artificial intelligence and decision support system are not required to use the system. A person with minimum knowledge of running programs in computer can use the system efficiently.

Reduced sets of features for different diseases may reduce the cost of laboratory tests, which is also an advantage of using the proposed systems.

Recently, Ministry of Human Resource Development, Government of India is encouraging researchers to develop such decision support systems by investing through various funding agencies [14]. Nowadays every medical practitioner is familiar with the use of computer and is able to adopt such system easily. Currently, doctors, biomedical researchers and computer professionals are working together in developing projects. Some of the medical equipments, used by medical professionals, are embedded with expert system.

Advancement in technology, especially in the field of information technology demands the use of such systems which are helpful to both doctors as well as general people of the society.

1.8. Summary of the Research Work

Whole research work is presented in nine chapters. Chapter wise contents are as listed below:

i) **Medical Scenario on Breast Cancer, Cardiac Arrhythmia and Hepa s Diseases – Chapter 2:**

In this chapter medical knowledge on diseases such as breast cancer, cardiac arrhythmia and hepatitis disease are discussed. Incidence rate, mortality rate, type of the said diseases, symptoms, reasons, diagnosis processes and health care services are also discussed. Various statistical reports in global and Indian perspective are summarized.

ii) **So Compu ng Techniques in Medical Diagnosis – Chapter 3:**

Different soft computing techniques that are used in medical diagnosis are presented in this chapter. Applications of these techniques and related works are studied and given in tabular form. Three of such

techniques, namely Fuzzy Expert System, Genetic Algorithm and Artificial Neural Network are the areas of interest. Learning algorithms such as Incremental Backpropagation Learning Network and Levenberg-Marquardt algorithms are also discussed in this chapter.

iii) Preprocessing of Data and Handling Missing Values – Chapter 4:

Types of data and necessity of data preprocessing are discussed in this chapter. Different data preprocessing steps such as data cleaning, data integration, data transformation, and data reduction are described. Different techniques used for dimensionality reduction of data such as Correlation based Feature Subset Selection (CFS), Rough Set Theory (RST), Principal Component Analysis (PCA), and Genetic Algorithm (GA) are also presented. This chapter deals with the importance of handling missing values, the techniques used to handle missing values in the data set and their advantages and disadvantages. Missing value management procedures are analyzed to implement suitable approach for a problem while designing an intelligent system.

iv) Performance Predic on Parameters – Chapter 5:

In this chapter, various benchmarking parameters to predict the performance of an intelligent automated system such as precession, correct classification accuracy, sensitivity, specificity, Kappa statistic, lift, cumulative gain and probability threshold are studied. Some statistical and experimental approaches have been made to list of such parameters to judge the performance of an intelligent model in the domain of medical diagnosis.

v) Breast Cancer Disease Diagnosis – Chapter 6:

This chapter presents various attempts that have been made to develop intelligent models to diagnose breast cancer. Soft computing techniques and their combinations are implemented. This includes correlation based feature subset selection, rough set, principal component analysis, incremental backpropagation learning network and Levenberg-Marquardt algorithm. Finally, experimental results are produced and results are compared with previous studies. Observations and proposals are also presented.

vi) Cardiac Arrhythmia Diagnosis – Chapter 7:

This chapter presents different approaches that have been made to develop automated system to diagnose cardiac arrhythmia. Soft computing techniques are implemented. Different combinations are tried. Genetic algorithm and correlation based feature subset selection are used to generate reduced feature subset. Incremental backpropagation learning network and Levenberg-Marquardt algorithms are used as classifiers. Experimental results, observations and suggestions are also presented.

vii) Hepa s Disease Diagnosis – Chapter 8:

In this chapter multiple imputations using EMB algorithm is implemented to handle missing values. Different approaches are attempted to develop automated intelligent model to diagnose hepatitis disease. Correlation-based feature subset selection and rough set are used for feature extraction. Incremental backpropagation learning network and Levenberg-Marquardt algorithms are used for classification. As EMB algorithm generates more than one imputed data sets from the data set with missing values, integration of the model has to be done to obtain final result. Two different approaches are tried to do this. Experimental results are discussed and compared with the previous similar studies. Finally, observations and proposal are noted.

viii) Accomplishments of the Objec ves and Future Scope – Chapter 9:

This is the final chapter of our research work. It presents how we achieve our goal as discussed in Chapter 1. Our future extension of study is also included in this chapter. We conclude our discussion after mentioning the future scope of our study.

References

1. Wyatt JC, "Use and Sources of Medical Knowledge", Lancet, Vol. 338, pp. 1368 – 1373, 1991.
2. Dr. Amitava Basu, MD, Pathologist, India (personal communication).
3. http://archieve.ics.uci.edu/ml/datasets/Hepatitis.
4. http://www.cancer.org/Cancer/BreastCancer/OverviewGuide.
5. Chugh S. S. et al., "Worldwide Epidemiology of Atrial Fibrillation: A Global Burden of Disease 2010 Study", Circulation, Vol. 129(8), pp. 837 – 847, 2014.
6. Go AS, Hylek EM, et al., "Prevalence of Diagnosed Atrial Fibrillation in Adults: National Implications for Rhythm Management and Stroke Prevention: the Anticoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study", JAMA, Vol. 285, pp. 2370 – 2375, 2001.
7. Camm A. J., et al., Guidelines for the Management of Atrial Fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC)", Euro Heart J, Vol. 31, pp. 2363 -2429, 2010.
8. Kaushal SS, et al., "Electrocardiographic Manifestation of Healthy Residents of a Tribal Himalayan Village", J Assoc Phys India, Vol. 43, pp. 15 -16, 1995.
9. Lip G. Y. H., Bawden L., Hodson R., Rutland E., "Atrial Fibrillation Amongst the Indo – Asian General PracticePopulation – The West Birmingham Atrial Fibrillation Project", Int .J. Cardiol, Vol. 65, pp. 187 – 192, 1998.
10. IHRS AF Registry, Presented in Annual IHRS meeting 2012, Mumbai.
11. Dan C., et al., "Viral Hepatitis: New U.S. Screening Recommendations Assessment Tools and Treatments", A. J. N., Vol. 115(7), pp. 26 – 35, 2015.
12. Centre for Disease Control and Prevention, Surveillance for Viral Hepatitis – United States, 2012, Atlanta, 2014. http://www.cdc.gov/hepatitis/statistics/2012Surveillance/PDFs/2012HepSurveillanceRpt.pdf.
13. Chauhan L. S., From the Desk of the Director, NCDC, NCDC Newsletter, Vol. 3(1), pp. 1 – 12, 2014.
14. "Directory of Extramural Research and Development Projects", Approved for funding by selected Central Govt. Agencies / Departments during 1994 – 95.