

## ABSTRACT

“Without health life is not life; it is only a state of languor and suffering – an image of death”. So care should be taken on health. It is a great challenge to provide proper health care services to citizens of all parts in India especially in rural and remote areas. Poor health care service in underserved areas lead to think of an alternative automated medical diagnosis system for early of diseases. Early detection of diseases helps in planning future treatment immediately, which in turn augments survivability. In this context intelligent automated decision support system may be useful.

An intelligent automated decision support system may assist health professionals in making decisions while diagnosing and managing diseases based on the stored knowledgebase. Sometimes medical diagnosis process becomes complicated due to inexact information. Moreover increasing number of diseases and their complexities cause the process of disease diagnosis too complex. A domain expert has to process huge data while making decision within time constraint. Decision support system, a widely used branch of artificial intelligence, makes it possible to help in solving complex problems within the time limit.

Some computing techniques, an assemble of artificial neural network, genetic algorithm, rough set and evolutionary algorithms, helps in designing or developing automated intelligent systems which helps health professionals in making decision while diagnosing health problems. Some computing techniques possess the tolerance of vague or uncertain information. For this reason the techniques can be regarded as well suited in the field of medical diagnosis as uncertainty is prevalent in medical diagnosis.

First part of the thesis have explored the present scenario of diseases, such as breast cancer, cardiac arrhythmia, and hepatic diseases, in global and Indian perspective to explain the need of alternative medical diagnosis system. The work focuses on the said diseases using UCI data sets for corresponding diseases. After collecting data various steps of data preprocessing have been attempted to prepare good quality data. Data sets, used in our study are free from outliers. Multiple imputations using EMB approach has been used to handle missing values in the data set.

Various benchmarking parameters to predict the performance of an intelligent model are studied to decide the list of parameters to be evaluated to judge the performance of the proposed intelligent systems. Considering the strength and

limitations of the performance prediction parameters, four parameters namely, accuracy, sensitivity, specificity, and area under ROC curve, are used to measure the performance of the proposed intelligent systems.

Attempts have been made to discover reduced features from input data sets that explain the data sets much using correlation-based feature subset selection (CFS), rough set (RS), genetic algorithm (GA), and principal component analysis (PCA). Incremental back propagation learning network (IBPLN) and Levenberg-Marquardt algorithms are used as classifiers.

Combinations which have been attempted to diagnose breast cancer are: CFS + LM, RS + LM, PCA + IBPLN, and PCA + LM i.e. features are selected using CFS, RS, and PCA and LM, and IBPLN algorithms are used for classification. The combination, PCA + LM, shows better performance in terms of accuracy, sensitivity, specificity, and area under ROC curve.

Four combinations namely, GA + IBPLN, GA + LM, CFS + IBPLN, and CFS + LM, have been attempted to design intelligent models to predict cardiac arrhythmia. Features are selected using GA and CFS techniques and IBPLN and LM algorithms are used as classifier. The combination, CFS + LM, provides better performance in terms of the benchmarking parameters evaluated to measure the predictive performance.

Multiple imputations using EMB approach has been implemented to handle missing values in the hepatitis disease dataset and five completed data sets have been generated. Features are extracted from the imputed data sets using CFS and RS. Two different approaches have been attempted. Firstly, all the imputed files have been analyzed using classifiers IBPLN and LM algorithms and the results are combined to obtain the final result. Secondly, five imputed files are integrated into a single file and this file has been analyzed using IBPLN and LM algorithms to obtain the final result. The second approach shows better performance for the combination RS + LM.

Experimental results are compared in terms of correct classification accuracy, sensitivity, specificity, and area under ROC curve. Attempts have been made to compare the outcomes with 'golden standard'. Feature subsets for the said diseases are proposed that would be worthwhile to physicians while diagnosing diseases. It has been suggested to consider the highest, lowest, and average behavior of the decision support system especially in medical domain to judge the performance of the model before using it.

At present, the growing field of intelligent system demands the use of decision support system as one of the most strategic and important tools for any

organization. Though the work successfully accomplish its objectives, still we say that the proposed intelligent models are not supposed to replace knowledge of domain expert, rather each of them can be considered as an effective and reliable tool to assist health professionals in making clinical decisions.