ABSTRACT

Decision making in medical domain is a challenging and responsible task. Medical diagnosis of a medical practitioner is an intuitive and complex process based on knowledge, experience and discernment. Besides, the knowledgebase of diagnosis is large, and consists of heterogeneous and ambiguous data. New ways and means need to be explored to complement and enhance, if not replace, the traditional methods of diagnosis and prognosis to equip the medical personnel to meet the critical nature of the domain. Statistical, computational, and machine learning tools have been used by scientific community and researchers in the area of clinical data processing. Data mining is a data analysis technique that explores data to discover knowledge. Artificial Intelligence techniques and expert systems in particular, can construct prospective tools that can make medical diagnosis and risk prediction based on clinical data.

Acute Myocardial Infarction (AMI) and in general, Coronary Artery Disease (CAD), is one of the most serious threats to life and a major public health problem in India. Family history, life style, food habits, physical inactivity and aging are some among many risk factors that contribute towards CAD. A research work based on CAD is relevant and need of the hour in this context to predict the risk of having AMI, when a person has one or more of the risk factors of CAD. Early detection of CAD as well as the prediction of entire risk of CHD helps to choose a life pattern to reduce the modifiable risk factors. However, it is challenging because of the intrinsic nature of the clinical data and manifold manifestation of the same disease.

In the proposed work, the goal is to design and develop “Neuro-Fuzzy Medical Decision Support System” (NFMDSS), a hybrid system that
integrates soft computing techniques, namely, decision trees, fuzzy logic and neural networks for the prediction of acute myocardial infarction. Integration has been incorporated to exploit the advantages of individual techniques and to address the pitfalls of one technique with the benefits of others. Thus, it combines the learning capabilities of neural networks and human-like reasoning of fuzzy systems.

The method followed is a case-control study method in which the data consisting of the risk factors of people with CAD are compared with that of people without CAD. The proposed work identifies a set of 12 risk factors for 301 instances (CAD patients), of which 244 are men and 57 are women, as case data whose age ranges between 21 and 75 collected from Cardiac Critical Care Units of hospitals in Kerala. Another set of data of 301 individuals with the same risk factors with matching age and gender criteria, but not detected with CAD ever, and collected from similar geographical region, is taken as control data. The risk factors identified are Age, Smoking, Obesity in terms of body mass index, Systolic Blood Pressure, Diastolic Blood Pressure, HDL Cholesterol, LDL Cholesterol, Total Cholesterol, Lipoprotein(a), Triglycerides, Homocysteine and Diabetes Mellitus. After having performed necessary pre-processing on the data, the case data is given to NFMDSS architecture for training so that learning takes place.

The architecture is typically a modified version of adaptive neuro-fuzzy inference system (ANFIS), a group of adaptive networks that are functionally comparable to fuzzy inference systems. ANFIS utilizes a hybrid learning algorithm to tune the parameters of a Sugeno-type (TSK) fuzzy inference system (FIS). The incorporation of fuzzy decision tree is employed in the model to overcome the dimensionality problem of ANFIS. The system has a multi-layer feedforward neural network with supervised backpropagation algorithm incorporated into it.
The architecture is implemented in three phases. In the first phase, the decision tree generates a set of fuzzy rules by implementing the fuzzy concept at the classification level of the decision tree. Second and third phases implement the NFMDSS architecture as a neural networks system in a two-stage learning process, where the first stage is the neural network whose consequent parts are expressed as a linear combination of the membership functions of the antecedent part, and second is the network whose consequent parts are expressed as functions of the input variables. A six-layered supervised network at each layer including the output layer, in which the consequent function is the linear combination of gaussian membership functions, is used as the structure for the first-stage learning. As the network uses the backpropagation algorithm, the antecedent and consequent parameters are updated using error-correction by the gradient-descent method by iteratively using the error between actual output and the target output generate a set of optimal rules.

A network with similar architecture, but where the consequent parts are functions of the input variables, a first-order TSK model, is presented in the second stage learning to optimize the system, namely, the parameters as well as rules with respect to the actual data. The second stage learning corrects the structure of the model and fine-tunes the system. The two-stage learning process is performed as a model-corrector mechanism for the model optimization though not entirely involved in direct error reduction approach. At the end of second-stage learning, after 25 iterations, the system achieves a convergence rate of 95% with training data and with the test data.

The simulation tool used for the experimental implementation is MATLAB 7.9. Three different sets of data are presented to the system to validate the model using three standard validation techniques, namely, cross validation with control data, \(k\)-fold cross validation with training data, and
validation with training data with dummy values. In addition, the performance of the proposed system is compared with that of other models, namely, ANFIS and FCNL. The results of the validations and the analysis of the comparisons show that reliable, precise and consistent performance is achieved through proposed hybrid system. Pre-diagnosis and early disease detection with the help of the proposed system may give physicians time to educate patients about preventive measures and offer therapies, thus reducing the chance of early death.