ABSTRACT

An Intelligent prediction model for the discovery of knowledge from clinical datasets is proposed and evaluated. The model has been specifically designed to be used by physicians as an aid for clinical decision making. The features supported by the model include extracting knowledge from the clinical datasets and representing the extracted knowledge in the form of rules or network parameters. The knowledge extracted was evaluated with the help of an expert in the medical field. Fuzzy logic was used to handle uncertainty present in the clinical. The model can handle clinical data with and without time dimension and support hybrid approaches for mining clinical data. The model has been tailored and evaluated using the datasets on hepatitis and heart disease.

The first component this research work focuses on techniques for extracting the knowledge from hepatitis data. This model has been tailored to include normalization, principal component analysis and fuzzy c-means clustering in the pre-mining subsystem. The mining subsystem was implemented as a neuro-fuzzy classifier which was manually tuned to identify the optimal values for premise and consequent parameters of the network. The extracted knowledge was validated and stored in the knowledge base. When input query was given, the inference system applied the knowledge stored in the knowledge base to predict whether the patient would survive hepatitis.
The second component extracted knowledge from time-series hepatitis data. The pre-mining subsystem processed the data to represent the variations of the data values with respect to the date of examination. The system extracted knowledge from hepatitis data using the techniques association rule mining algorithm, decision tree algorithm and neural network. The extracted knowledge was stored in the knowledge base for predicting whether the patient has hepatitis or not. The results of this system indicate that the decision tree algorithm implemented in this work performs better than the neural network and association rule mining algorithm.

The third component extracted knowledge from heart disease data. The pre-mining subsystem discretized the continuous valued attributes using entropy based discretization. The mining subsystem extracted rules by generating contingency table and performing $\chi^2$ analysis. The extracted rules were validated and stored in the knowledge base. The inference subsystem implemented using Bayesian classifier along with the aid of rules discovered by statistical measure predicted the presence of heart disease.

The fourth component was also used to extract knowledge from heart disease data. The pre-mining subsystem was implemented to fuzzify the continuous valued attributes. Feed forward neural network trained by genetic algorithm was used to extract the knowledge to predict the presence of heart disease along with its severity. The knowledge extracted was validated with the help of a medical expert and stored in the knowledge base. The knowledge in the knowledge base was used to predict the severity of heart disease for a previously unseen data.
A survey had been carried out on the studies carried out to identify correlation among cardiovascular, diabetes, hepatitis and anemia (renal). Some of the studies reveal that hepatitis is more prevalent among hemodialysis patients. Studies reveal that hepatitis is associated with diabetes; some of the studies identify positive correlation between them whereas others a negative correlation. Patient with hepatitis infection was found to have different cardiovascular risks when compared to a non-infected patient. Studies show that diabetic patients are more prone to have cardiovascular risks.