ABSTRACT

The innovation in biomedical devices for healthcare has drastically increased the emergence of temporal data that describe the state of patient's health. These data vary over time and are stored as Electronic Health Records (EHR's). The data in EHR's are liable to several complexities such as missing values, irregular observations and large time-constrained attribute set. The presence of these complexities challenges the knowledge mining in clinical time series data acquired from EHR. This research work has proposed mining frameworks that aim in building an effective classification model for clinical time series data. The constructed model has been used in developing Clinical Decision Making System (CDMS) for assisting the physician in clinical diagnosis.

This research work presents five contributions that perform knowledge mining process in clinical time series data. The first contribution presents a Temporal Rough Set induced Neuro-Fuzzy (TRiNF) mining framework that builds a classification model for unevenly-spaced clinical time series data. TRiNF provides two functionalities namely temporal data acquisition and temporal classification. In temporal data acquisition, a time series forecasting model has been constructed by adopting an improved double exponential smoothing method. The forecasting model has been used in missing value imputation and temporal pattern extraction. The relevant attributes were selected using temporal pattern based rough set approach. In temporal classification, a temporal pattern induced neuro-fuzzy classifier constructs classification model with the selected attributes.

The second contribution presents a Q-Back Propagated Time-Delay Neural Network (Q-BTDNN) classifier that builds a temporal classification model. Q-BTDNN is a feed forward Time Delay Neural Network (TDNN) in which the training is done using the Q-Learning induced Back-Propagation

(Q-BP) technique. Q-BP functions in an incremental way by combining the relative advantages of both the reinforced Q-learning and backpropagation. The classification model is used to develop a CDMS that aids the physician in diagnosing the severity of gait disturbances in Parkinson's disease affected patients.

The third contribution presents a Tolerance Rough Set Induced Bio-Statistical (TRiBS) framework that imputes missing values in an unevenly spaced clinical time series data. The proposed framework adopts an Inverse Distance Weight Interpolation Technique (IDW) and enhances it using the concept of tolerance rough set and Particle Swarm Optimization (PSO). The classical IDW interpolation suffers from two major limitations while interpolating an unknown data point: first, in selecting the known data points and second, choosing an optimal influence factor. TRiBS framework overcomes this first limitation of IDW using tolerance rough set and the second using PSO. Experimental evaluation with TRiBS imputed data and the classifiers such as neural networks, support vector machine and decision tree have shown an improvement in the classification accuracy.

The fourth contribution presents a bio-statistical mining framework, named Statistical Tolerance Rough Set induced Decision Tree (STRiD) that handles the temporal complexities in clinical time series data and builds an effective classification model for assisting the physician in medical diagnosis. The proposed framework contributes work in two stages of knowledge mining, namely temporal pre-processing and classification. In temporal pre-processing, a forecasting model is built using the proposed Fuzzy Inference Double Exponential Smoothing (FIDES). The forecasting model imputes the missing data and extracts the temporal patterns for each attribute. The significant attributes are selected using temporal pattern based tolerance rough set algorithm. In classification, a decision tree classifier with temporal pattern induced gain ratio as splitting criteria has been used to build a classification model.

The fifth contribution presents a Forecast-Error Approximation based Bottom-Up (FeAB) segmentation approach for effectively classifying unevenly spaced clinical time-series data using time delay neural network. The proposed approach includes two functionalities namely temporal data summarization and classification. In temporal data summarization, an irregularly observed clinical time-series is divided into sequence of temporal interpreted segments using the proposed FeAB segmentation. FeAB adopts a double exponential smoothing technique to derive the growth rate, mean and forecast-error for each clinical observation. FeAB performs segmentation using a bottom-up segmentation strategy. The obtained forecast-error is used to compute the merge-cost for FeAB segmentation. In classification, a classification model has been constructed for the segmented time series using time delay neural network classifier.

For experimentation the first, third, fourth and fifth contributions (TRiNF classifier, TRiBS framework, STRiD classifier and FeAB segmentation for TDNN classification) have used time series Hepatitis and Thrombosis dataset. The second contribution (Q-BTDNN classifier) has used Parkinson's disease dataset. The experimental results have shown improvement in the classification accuracies. The classification model obtained from the proposed works can be used to develop a CDMS. This CDMS can assist the physicians to perform the following clinical activities: to diagnose and monitor the severity of Parkinson's disease based on the gait disturbances, to diagnose Hepatitis B and Hepatitis C and to diagnose and monitor the severity of thrombosis in collagen diseases. Moreover, the proposed mining frameworks can be extended to develop CDMS for diagnosing other diseases with minor domain specific changes.