CHAPTER 10

CONCLUSION AND FUTURE WORKS

This chapter presents the conclusion and scope for the future works pertaining to this research work.

10.1 CONCLUSION

Clinical data obtained from Electronic Health Records (EHR’s) are time-stamped and are liable to temporal complexities such as missing data, irregular observations and time constrained attributes. These temporal complexities challenge the mining process in clinical time series data. This research work presents mining frameworks that aims in building an effective classification model for unevenly spaced clinical time series data. The constructed classification model can be used in developing a Clinical Decision Making System (CDMS) that aid the physician in clinical diagnosis. CDMS can assist the physician to monitor and diagnose the state of patient health that is captured in electronic health records (EHR’s). The following are the research contributions: TRiNF classifier, Q-BTDNN classifier, missing value imputation using TRiBS for classification, STRiD classifier and FeAB segmentation for TDNN classification.

Contribution 1: This contribution presents a Temporal Rough Set induced Neuro-Fuzzy (TRiNF) framework that handles the temporal complexities and builds an effective classification model. TRiNF consists of two functionalities namely Temporal Data Acquisition (TDA) and temporal classification. TDA process aims at pre-processing the temporal complexities in clinical time series data. An enhanced Double Exponential Smoothing (DES) method presented by Wright (1986) has been adopted for constructing a time series forecasting model. Missing value imputation and temporal pattern extraction are done using the forecasting model. The relevant attributes are selected for classification using a temporal pattern based rough set approach. In the temporal classification process, an effective classification model is built using a temporal pattern induced neuro-fuzzy classifier. The fuzzy sets for the classifier are defined using the trend pattern of each clinical attribute. Clinical time series Hepatitis and Thrombosis datasets have been used for experimentation. The results show that the proposed system overcomes the temporal complexities and obtains the classification accuracy of 92.59% for Hepatitis dataset and 91.69 % for Thrombosis dataset.

Contribution 2: This contribution aims in developing a Clinical Decision Making System (CDMS) that uses an effective classification model for diagnosing the severity of gait disturbances in Parkinson Disease (PD). A Q-Back Propagated Time Delay Neural Network (Q-BTDNN) classifier has been presented for building a classification model. Q-BTDNN is a dynamic feed forward time-delay neural network (TDNN) in which the learning is done using the proposed Q-Learning induced Backpropagation (Q-BP) technique. This research contribution has been experimented with time stamped gait database acquired through wearable sensors. The experimental result, prove the effectiveness of the presented Q-BTDNN classifier in terms of its improved classification accuracies of 91.49%, 92.19% and 90.91% for the three PD study datasets.

Contribution 3: This contribution presents a TRiBS framework for imputing missing values in an unevenly spaced clinical time series data. TRiBS adopts and improves the classical IDW using two key concepts namely tolerance rough set analysis and Particle Swarm Optimization (PSO). Tolerance rough set analysis identifies similar records, which forms the significant set. The PSO technique finds the influence factor value for fixing the weights of known data included in the significant set. The significant set and the influence factor were used in IDW process to derive the interpolated values. These interpolated values impute the missing values in clinical time series dataset. For experimentation, time series data of Hepatitis and Thrombosis patients were used. Classification on TRiBS imputed data using classifiers such as neural network, decision tree and SVM shows an accuracy of 83.57 %, 81.14 % and 78.89 % respectively for missing rate of 10% with Hepatitis dataset and 80.15 %, 77.91 % and 76.19 % respectively for missing rate of 15% with Thrombosis dataset.

Contribution 4: This contribution presents a Statistical Tolerance Rough Set induced Decision Tree (STRiD) mining framework that builds a classification model for an irregularly observed clinical time series data. The proposed framework contributes work in three stages of knowledge discovery namely temporal pre-processing, attribute selection 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 value and extracts the temporal patterns for each attribute. In attribute reduction, the significant attributes are selected using temporal pattern based tolerance rough set concept. A classification model is built with the selected attributes using decision tree classifier with temporal pattern induced gain ratio as splitting criteria. The proposed work has been experimented using two clinical time series dataset of Hepatitis and Thrombosis patients. The STRiD approach combines fuzzy inference based time series analysis with the knowledge discovery process to improve the effectiveness of decision-making. The results show that the proposed system overcomes the temporal complexities and obtains the classification accuracy of 91.5% for Hepatitis and 90.65% for Thrombosis.

Contribution 5: This research contribution presents a FeAB segmentation approach for TDNN classification. The novelty of the contribution lies in two aspects. First, in effectively using the forecasted results obtained from Hanzak updated DES in bottom-up segmentation and second, incorporating the temporal summarized segments with TDNN to classify unevenly spaced data. Although, classifiers like TDNN have dynamic response to time series data, the challenge arises due to the irregularities in clinical data. This research work consists of two stages, namely temporal summarization and classification. In temporal summarization stage, this work performs two activities, namely smoothing and segmentation. In smoothing, for each clinical laboratory test the growth rate, level and forecasted value of its observed time points are computed using an enhanced DES method presented by Cipra & Hanzak (2008).

The time series is then divided into a sequence of segments by the proposed Forecast-Error Approximation based Bottom-Up (FeAB) segmentation method. The growth rate and the mean value for each clinical observation in a segment are aggregated by taking its mean. The segments of each attributes are given as input to the time delay neural networks. For experimentation, time series data of Hepatitis and Thrombosis patients were used. Experimental results show that the proposed framework overcomes the complexity of uneven spacing in time series data and obtains the classification accuracy of 91.98% for Hepatitis dataset and 91.17 % for Thrombosis dataset.

In this research work, CDMSs have been developed to diagnose Parkinson's disease, Hepatitis disease and Thrombosis in collagen diseases.

The developed CDMSs perform following clinical activities: diagnose the severity of Parkinson’s disease based on the gait disturbances, diagnose Hepatitis disease (B and C) and diagnose the severity of Thrombosis in collagen diseases. The proposed mining frameworks can be extended to develop CDMS for diagnosing other diseases with minor domain specific changes.

This research work has presented work in different stages of mining unevenly spaced clinical time series data such as missing value imputation, attribute selection, segmentation and classification. Though, all the five contributions presented in this research aims in building classification models, there are significant differences among the contributions with respect to the methodologies presented. The summarization and differences among the research contributions is outlined below: Contribution 1 and Contribution 4 presents two frameworks namely TRiNF and STRiD framework, that performs activities such as missing value imputation, temporal pattern extraction, attribute selection and classification. The major difference among these frameworks lies in the methodologies incorporated in performing these activities.

Contribution 1 performs missing value imputation and temporal pattern extraction using an enhanced DES forecasting method presented by Wright (1986), whereas in contribution 4 an enhanced DES method named FIDES is proposed to perform missing value imputation and temporal pattern extraction. FIDES uses fuzzy logic to adjust the smoothing constant parameters for trend and level computations. The extracted temporal patterns are used in attribute selection and classification process in both the contributions. However, in contribution 1 relevant attributes are selected using rough set approach based on equivalence class generation, whereas in contribution 4 relevant attributes are selected using tolerance rough set approach based on tolerance class generation. The tolerance rough set approach handles the continuous valued attribute effectively during the process of attribute selection. In contribution 1, neuro-fuzzy classifier is used for building classification model and in contribution 4, decision tree with temporal gain ratio as splitting criteria is used for building classification model.

Contribution 3 presents a TRiBS framework to perform missing value imputation and highlights its influence in the classification process. TRiBS framework enhances IDW interpolation using the concept of tolerance rough set and PSO. Thus, altogether in this research study three approaches for handling missing values in unevenly spaced clinical time series data is presented in the Contribution 1 (TRiNF), Contribution 4 (STRiD) and Contribution 3 (TRiBS) respectively. To handle missing values, Contribution 1 and Contribution 4 uses forecasting approach, whereas Contribution 3 presents an enhanced interpolation approach.

Contribution 2 and Contribution 5 both uses TDNN for constructing classification model. However, contribution 2 and contribution 5 differs in its key activities. In Contribution 2, a Q-BTDNN framework is presented that highlights the importance of using the proposed reinforced Q-BP algorithm in training time delay neural network for building the classification model. Contribution 5 proposes a FeAB segmentation approach that highlights the effectiveness of forecast induced bottom-up segmentation by reducing the length of time series to segments for TDNN classification. In contribution 5, forecasting results are used in segmentation process whereas in Contribution 1 and 4 forecasting results are used for missing value imputation and temporal pattern extraction.

Though, the materials and methods used in all the five presented frameworks differs, the objective of them is to build a clinical decision support system for assisting physician in clinical decision making. The presented frameworks can be extended to support other decision making activities with minor domain specific changes.

10.2 FUTURE WORKS

There are many interesting aspects for future research. The data acquired from bio-medical devices are time-stamped and mostly multivariate. These data describes the reports of laboratory or physical examination undergone by the patients that are generally observed at irregular intervals. Though, there are several works in the literature that performs knowledge mining from time series data. The presence of irregularities (uneven spacing) and multivariate descriptions in clinical time series data always challenges the knowledge mining process that pave way for many research studies such as temporal reasoning, temporal abstraction, handling missing values, time series representations, segmentation and dimensionality reduction. The techniques presented in these research contributions have been experimented with clinical time series data. However, these techniques can be extended to support other irregular time series data with few minor domain specific changes.

Also, as an extension to these proposed mining frameworks, research works that aims at improving the classification accuracies and optimizing the temporal classification process can be carried out. Moreover, due to an enormous growth in field of electronic devices there is a high modernization in biomedical devices. These biomedical devices allow the clinician to monitor the patient health conditions periodically which has led to the generation of large set of clinical observations. Mining knowledge from such kind of huge time stamped data remains as a challenging area of research. In recent years, wearable sensors have been used to monitor the patient's health conditions. These sensors generate huge data that are time-stamped.

Developing a CDMS for these time stamped data remains challenging due to its enormous size. Research work that aims in developing CDMS for clinical time series data is highly recommended in medical field to assist the clinician in clinical decision making.