Expert Systems With Applications 78 (2017) 283–300

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A bio-statistical mining approach for classifying multivariate clinical time series data observed at irregular intervals

Jane Y. Nancy[a], Nehemiah H. Khanna[a,*], Arputharaj Kannan[b]

[a] Ramanujan Computing Centre, Anna University, Chennai-600025, India
[b] Department of Information Science and Technology, Anna University, Chennai-600025, India

ARTICLE INFO

Article history:
Received 26 September 2016
Revised 27 January 2017
Accepted 28 January 2017
Available online 6 February 2017

Keywords:
Clinical time series
Fuzzy
Double exponential smoothing
Tolerance rough set
Data mining
Decision tree

ABSTRACT

In medical information system, the data that describe patient health records are often time stamped. These data are liable to complexities such as missing data, observations at irregular time intervals and large attribute set. Due to these complexities, mining in clinical time-series data, remains a challenging area of research. This paper proposes a bio-statistical mining framework, named statistical tolerance rough set induced decision tree (STRiD), which handles these complexities and builds an effective classification model. The constructed model is used in developing a clinical decision support system (CDSS) to assist the physician in clinical diagnosis. The STRiD framework provides the following functionalities namely temporal pre-processing, attribute selection and classification. In temporal pre-processing, an enhanced fuzzy-inference based double exponential smoothing method is presented to impute the missing values and to derive the temporal patterns for each attribute. In attribute selection, relevant attributes are selected using the tolerance rough set. A classification model is constructed with the selected attributes using temporal pattern induced decision tree classifier. For experimentation, this work uses clinical time series datasets of hepatitis and thrombosis patients. The constructed classification model has proven the effectiveness of the proposed framework with a classification accuracy of 91.5% for hepatitis and 90.65% for thrombosis.

1. Introduction

The advancement in the clinical applications has led to the generation of large time stamped data. These data are stored in electronic health records (EHR), which describes a set of observations on patient health conditions (Orphanou, Stassopoulou, & Keravnou, 2014). Knowledge discovery from EHR data remains a challenging area of research since they are susceptible to complexities such as irregular observations, missing values and large attribute set. The presence of missing value is due to the negligence of the patients or practitioner to record few observations and the error in data recordings from medical equipment. The irregular observations refer to the data observed at unequal time intervals. In clinical domain, data are often considered as irregular since different patients are observed at different time points and a patient's health state is usually observed at irregular time intervals. An attribute in clinical data represents a lab test undergone by a patient. In general, clinical data consists of large attribute set since a patient may have undergone many common lab tests in addition to the disease specific lab tests.

Although, clinical time series data suffers from temporal complexities such as irregular observations, missing values and large attribute set, they also possess useful hidden medical knowledge (Moskovitch, & Shahar, 2015a). This medical knowledge, when extracted can be used for developing a clinical expert system that can assist the physician in clinical decision-making tasks such as diagnosis, monitoring, prognosis and drug discovery (Isola, Carvalho, & Tripathy, 2012). The importance of managing time-oriented concept and knowledge discovery in medicine was investigated in many research studies (Adlassnig, Combi, Das, Keravnou, & Pozzi, 2006; Chittaro, & Montanari, 2000; Keravnou, & Shahar, 2005; Orphanou et al., 2014). To develop time oriented medical systems, temporal reasoning and temporal maintenance are considered as two challenging areas of research (Augusto, 2005; Bouzid, Combi, Fisher, & Ligozat, 2006; Combi & Shahar, 1997). Further, various challenging research areas in temporal representation and reasoning in medicine that are related to several clinical tasks such as monitoring, treatment, diagnosis and prognosis have been

* Corresponding author.
E-mail addresses: nancy@annauniv.edu (J.Y. Nancy), nehemiah@annauniv.edu (N.H. Khanna), kannan@annauniv.edu (A. Kannan).

http://dx.doi.org/10.1016/j.eswa.2017.01.056
0957-4174/

identified and discussed by researchers (Adlassnig et al., 2006). There are many temporal mining methods (Batal, Sacchi, Bellazzi, & Hauskrecht, 2009; Kahn, 1991a, Kahn, 1991b; Moskovitch & Shahar, 2015a; Moskovitch & Shahar, 2015c; Moskovitch, & Shahar, 2015b; Sacchi, Larizza, Combi, & Bellazzi, 2007; Shahar, 1997; Tu Bao et al., 2003; Verduijn et al., 2007) available to extract medical knowledge from clinical time series data, but still the presence of temporal complexities challenges their accuracy in decision-making. The proposed work aims at constructing an efficient classification model for clinical time series data by handling its following complexities namely irregular observations, missing values and large attribute set.

Outline of the paper: This paper 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 preprocessing, attribute selection and classification. In temporal preprocessing, a forecasting model is build using the proposed fuzzy inference double exponential smoothing (FIDES) method. The forecasting model imputes the missing value and extracts the temporal patterns for each attribute. Temporal pattern for a clinical attribute refers to its observed trend and state. The trend describes the overall growth rate of a clinical attribute and state describes its mean or level value. In attribute selection, the significant attributes are selected using temporal pattern based tolerance rough set algorithm. 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 is evaluated using two clinical time series dataset of hepatitis and thrombosis patients. Thus, the presented approach combines fuzzy inference based time series analysis with the knowledge discovery process to improve the effectiveness of decision-making.

The rest of the paper is organized as follows. In Section 2, the related work is presented. Section 3 discusses materials and methods used in the proposed bio-statistical mining approach. Experimental results and discussions are presented in Section 4. Conclusion and scope for future works are presented in Section 5.

2. Related works

This section reviews the works carried out by the researchers in the field of temporal abstraction, time series data classification and forecasting using double exponential smoothing.

2.1. Temporal abstraction

Temporal abstraction is a key task in modelling a time stamped data. In clinical time series data, temporal abstraction represents the process of transforming quantitative values in the clinical attributes to summarized qualitative interpretations such as trend and state. The trend describes the overall growth rate and state describes the mean or mean of the attribute over a period of observation. The process of temporal abstraction has been proven to be effective in discovering knowledge from time series Moskovitch and Shahar (2015c). Shahar (1997) and Shahar and Musen (1993) proposed a knowledge based inference structure for temporal data. The process of knowledge based temporal abstraction (KBTA) is divided into sub tasks that is carried out by four domain independent abstraction mechanisms which depends on four knowledge types namely structural, classification, temporal semantic and temporal dynamic knowledge. This KBTA framework is implemented in RESUME system (Shahar & Musen, 1993), which employs temporal abstraction mechanism like temporal interpolation, point temporal abstraction and temporal inference to automate the temporal abstraction process. The input to the system includes time stamped patient data, events and the domain knowledge base. The RESUME system was used for summarizing patient data in clinical domains like, monitoring of children's growth and management of insulin-dependent diabetes.

Batal et al. (2009) have proposed a segmented time series feature (STF) - Mine algorithm for extracting the discriminative temporal abstraction (TA) patterns from frequent temporal patterns obtained using qualitative temporal abstraction mechanism and Allen's temporal relations (Allen, 1983). The multivariate features obtained are used to learn a classification model. The algorithm has improved the classification accuracy when applied on real heparin-induced thrombocytopenia dataset. Sacchi et al. (2007) and Verduijn et al. (2007) have proposed an algorithm apriori-like, which extracts the temporal association rules based on the relationship between complex temporal patterns from time series data. The temporal patterns are obtained by applying the framework for knowledge based temporal abstraction and Allen's temporal operators are used for obtaining the temporal precedence. The work is experimented on two different kinds of data, first set includes clinical data monitored during hemodialysis sessions and the second consists of DNA microarray gene-expression data for mining genetic regulatory relationships. Verduijn et al., 2007 have compared the performance of two types of temporal abstractions namely qualitative TA and quantitative TA. In the former, domain specific knowledge were used to derive a small set of symbolic representation for the temporal data whereas the later derives numeric abstraction using statistical process. The results indicated that quantitative procedure are informative and improves prediction accuracy than qualitative procedure for ICU data.

Orphanou et al. (2014) have presented a survey on temporal abstraction (TA) and temporal Bayesian networks (TBN) to perform clinical tasks like diagnosis and prognosis in clinical domains. Kahn (1991a) has presented a program named TOPAZ that uses clinical knowledge about time stamped relationships to infer temporal clinical events. TOPAZ summarizes the temporal events in the patient health records that can be used in temporal reasoning tasks. To overcome the temporal reasoning problems faced in developing a clinical expert system an extension to the time-oriented databank model based databases was presented by Kahn (1991b). The structures allow to store and retrieve clinical patient database by enabling them to encode temporal relationships. Ramsay and Silverman (2005) have presented statistical methods for functional data analysis. Horvath and Kokoszka (2012) have presented various inferential methodologies for functional data analysis. The authors have elaborated the analysis of dependent functional data including change point detection, functional time series and spatial statistics with functional data. Tu Bao et al. (2003) have proposed a temporal abstraction approach for mining knowledge from hepatitis database. The hepatitis temporal database used in their work was collected from Chiba university hospital between the years 1982-2001. The authors have investigated hepatitis background knowledge with data analysis and have proposed methods for temporal abstractions in tracking long term and short term changes in the lab examinations. The authors have claimed that various machine-learning methods can be applied on the abstracted data for extracting knowledge that can be used by medical doctors.

2.2. Time series data classification

Classification on time series data is challenging as they possess several temporal intervals and abstracted interpretations in addition to the time stamped data points. Moskovitch and Shahar (2015a) have discovered symbolic frequent time intervals related pattern (TIRP) using an algorithm named KarmaLego. The proposed algorithm comprises a data structure and new method for TIRP candidate generation using the Allen's temporal relations. The au-

thors have experimentally proved the improvement in the execution speed of their karmalego algorithm over the other state-of-art time interval pattern mining techniques like ARMADA, IEMiner and H-DFS. Moskovitch and Shahar (2015b) have proposed a framework, KarmaLegoSification (KLS) for classification of multivariate time series data. Temporal abstraction is carried out to transform time point series into time interval series. In order to avoid finding all of the TIRPs while mining a single entity, modified KarmaLego is proposed which consists of SingleKarma and SingleLego algorithms. Various discretization strategies such as symbolic aggregate approximation (SAX), Equal-Width Discretization (EWD) were used for transforming time series data into time intervals series.

Moskovitch and Shahar (2015c) have presented a supervised discretization technique to improve the accuracy of the classification process. The authors have described the procedures for determining optimal-cut offs in continuous data in order to create discrete symbols. They have derived symbolic time intervals and frequent temporal interval relation pattern (TIRP) using the methods discussed in their previous work (Moskovitch & Shahar, 2015b). The extracted patterns were used to induce a classifier (Moskovitch & Shahar, 2015c). The authors have compared the presented temporal discretization for classification (TD4C) with unsupervised discretization namely EWD, SAX and Knowledge-base (KB) method. It was observed that the performance of TD4C is effective compared to the unsupervised methods.

Kamali, Boostani, and Parsaei (2014) have proposed an approach for quantitative analysis of electromyography (EMG) signals by decomposing it into motor unit action potential trains (MUAPTs). The authors have presented several supervised classification techniques that fuse both time and time-frequency features of a MUAP to determine its class label. The time domain features include rise time, duration, spike duration, peak-to-peak amplitude, area and turns; the frequency domain features include the characteristics extracted from the discrete wavelet transform. Single classifier scheme as well as multi-classifier techniques that use ensemble of support vector machines (SVMs) as base classifiers were experimented. The subsets of features employed for each base classifiers were optimized through wrapper method. The results show that time-frequency features are superior to those of time domain features taken separately.

Ismail, Shabri, and Samsudin (2011) have presented a combination of self-organizing maps and least square support vector machine (SOM-LSSVM) classification method for time series forecasting. This combination has proved to be more flexible than a single least square support vector machine. Radial basis function, which is the kernel function in the LSSVM, proves to provide better efficiency than the other kernel. The study is carried out on two datasets namely Wolf yearly sunspot data and the Monthly unemployed young women data. The experimental results show that SOM-LSSVM outperforms than a single LSSVM based on the criteria of mean absolute error (MAE) and root mean square error (RMSE) and offers a better prediction results on the time series dataset.

Patra and Bruzzone (2014) have proposed a semi supervised novel active learning based on SOM-SVM model. The model proves to be effective in areas where collecting labelled samples are very much expensive. The system trains initially with less number of labelled training samples. The system uses the topological properties of SOM to select most informative samples that satisfy both uncertainty and diversity criteria in each iteration and tries to label them. The confidence of accurate classification of unlabeled samples is computed using the previously trained SVM. The experimental results of the proposed method on remote sensing data set proves to be effective than other active learning algorithms.

Minas, Karaolis, Moutiris, Hadjipanayi, and Pattichis (2010) have proposed a data mining system for the evaluation of risk factors related to heart event. The system uses classification algorithm C4.5 but instead of using single splitting criteria, five different splitting criteria's namely information gain, Gini index, likelihood ratio chi-squared statistics, gain ratio, and distance measure were investigated to segregate the attributes. A heuristics process is used for pruning based on the statistical significance of splits. For classifying myocardial infarction (MI), percutaneous coronary intervention (PCI), and coronary artery bypass graft surgery (CABG) patients, three classification models were developed based on decision trees.

Ding, Trajcevski, Scheuermann, Wang, and Keogh (2008) have presented a comparitive study on time series representation methods and similarity measure approaches. The authors through the experimental evaluations have suggested that the accuracy of elastic measures on large datasets converges with Euclidean distance. Górecki et al (2013) have proposed a distance measure function based on derivative that considers the shape of time series. The authors have used their proposed distance function to perform time series classification with nearest neighbor method. The experimental results prove the effectiveness of the proposed distance function compared to Euclidean distance and dynamic time wrapping.

Jane, Nehemiah, and Arputharaj (2016a) have presented a mining framework temporal rough set induced neuro-fuzzy (TRiNF) for classifying clinical time series data with uneven spacings. TRiNF performs following functionalities namely missing value imputation, temporal pattern derivation, attribute selection and classification. A Wright updated double exponential smoothing (DES) (Wright, 1986) is adopted to extract the attribute's growth rate, mean and forecast value for each observation. Missing values are imputed using forecast value. The growth rate and mean value is mapped to trend and state and is used in attribute selection and classification. The experiments were evaluated using two clinical time series data of hepatitis and thrombosis patients. The authors have proved the system's efficiency in terms of its classification accuracy. The major difference between the work presented in Jane et al. (2016a) and the proposed STRiD framework are as follows: The STRiD framework presents a FIDES forecasting technique for extracting the attribute's growth rate, mean and forecast value, whereas the TRiNF framework presented in Jane et al. (2016a)) uses Wright updated DES. The attribute selection and classification in Jane et al. (2016a) was done using rough set concept and neuro-fuzzy whereas the STRiD framework uses tolerance rough set concept with temporal similarity metric and decision tree with temporal pattern induced gain ratio. To identify the suitable network parameters for training TRiNF classifier there was a huge network realization cost. Thus, the computational complexity in constructing a classification model using TRiNF is higher than the STRiD.

Jane et al. (2016b) have proposed a Q-backpropagated time delay neural network (Q-BTDNN) for diagnosing parkinson disease based on the gait patterns. In Q-BTDNN, the network weights are adjusted using a reinforced Q-learning induced backpropagation strategy. Vijaya et al. (2010) have proposed a diagnosis system using fuzzy neuro-genetic approach for predicting the severity of the cardiovascular diseases. The fuzzified non-discrete input variables are fed as input to the neural network (NN). Genetic algorithm is used to train the neural network and is used to identify the rules that are significant for classification. Nehemiah et al. (2007) have proposed a clinical decision making process using four different mining techniques namely association rule mining, decision tree, neural network and neuro-fuzzy along with the temporal constraints. These approaches extract temporal rules, validate it and stores in the knowledge base. Two time series datasets of hepatitis and thrombosis patients were used for experimentation (Hepatitis Dataset for Discovery Challenge, Thrombosis Dataset for Discovery Challenge). The access to the data is currently unavailable. Hepatitis dataset was used as input dataset to decision tree, neural network and association rule mining. Neuro-fuzzy classification was performed using thrombosis dataset. This process im-

proves the classification accuracy and the number of relevant rules generation.

Nahato, Nehemiah, and Kannan (2015) have presented a rough set indiscernibility relation method with back-propagation neural network (RS-BPNN) classifier to perform knowledge extraction in clinical data. The classifier is experimented with datasets obtained from the University of California at Irvine (UCI) machine learning repository namely Wisconsin breast cancer, hepatitis and Statlog heart disease. The authors have claimed that RS-BPNN classifier shows significant improvement in the classification accuracy. Ruggieri et al. (2002) have presented an effective algorithm for C4.5 based classification for continuous attributes. The authors have identified effective strategies for computing local threshold, which is used for information gain computation to speed-up the classification process. The local threshold is computed by sorting the whole training set and applying binary search on the data to identify the threshold. Xinmeng et al. (2012) proposed a new method to build decision trees by using maximum similarity of attributes as the splitting criteria. The splitting criterion is determined based on maximum similarity for constructing the decision tree. The algorithm also does pruning, which is not done in ID3 algorithm making it more accurate than ID3.

2.3. Forecasting using double exponential smoothing

Forecasting in irregular time series have been studied in various research studies. Holt (1957, 2004) proposed a double smoothing method, which is used in forecasting the future behaviour of a variable in a time series based on its past behaviour. Holt's method is also referred as exponential smoothing with trend (Hyndman, Koehler, Ord, Snyder, 2008). This is an extension to the single exponential smoothing (SES) algorithm. SES was originally designed for time series with no seasonal or trend pattern, but Holt method performs exponential smoothing for time series with trend component. In Holt's method for each new observation the estimates of mean and slope is adjusted. SES performs forecasting in data with no trend or seasonal patterns, but if there is a linear trend, Holt's method is preferred.

Winters (1960) proposed a dynamic method for forecasting time series with triple exponential smoothing. The author extended the classical Holt method by incorporating seasonal component. The Holt-Winters method, which was popularly used in forecasting method uses exponentially weighted moving average approach to estimate the mean and slope of the time series with trend and seasonal patterns. However, these classical methods by Holt and Winters are strongly influenced by the presence of irregularities in the time series. The two major limitations in traditional DES are the presence of data irregularities and the choice of optimal value for smoothing constant parameters (α, β) used in trend and mean estimation. The values of these smoothing constant parameters directly influence the forecasting results and wrong choice of its value increases the forecasting error. There are several works in the literatures that address these limitations (Wright, 1986; Cipra & Hanzak, 2008)

Wright (1986) proposed an enhanced Holt method for the special case of handling time series observed at irregular time intervals. The author has incorporated few enhancements in the mathematical model used for estimating mean and slope component to handle the irregularities. For experimentation, Wright used several examples of irregular time series, one of them was the time series of world record times in one mile run. However, the presence of time-close observations affected Wright's version of Holt method. Later, Cipra and Hanzak (2008) updated the calculation for smoothing constant to perform forecasting in irregular time series. However, the time-close observation problems, which is defined as the smallest interval spacing between the observations remains challenging. The proposed work effectively handles this time-close problem and improves forecasting accuracy using fuzzy inference systems (FIS).

Tsaur (2003) have presented a forecasting model named fuzzy double exponential smoothing method that combines fuzzy polynomial trend analysis and the fuzzy exponential smoothing model. The model was experimented with forecasting in internet users data of Taiwan and was shown that it provides effective forecasting by choosing optimal smoothing constant. However, the work in Tsaur (2003) does not address the problem of forecasting with irregular time series. The proposed FIDES method overcomes the data irregularities by capturing the trend and state of each clinical attribute and computes forecasted value for each observation based on their interval spacings. In FIDES, the smoothing constant values are updated using the designed fuzzy inference system.

Comparing to the works discussed in literature, the proposed work is different in following ways: First, the proposed work uses FIDES method to enhance the DES method (Holt, 1957; Holt, 2004) by incorporating fuzzy inference system in smoothing constant estimations. The obtained forecasted model is used for missing value imputation and temporal pattern derivation. The FIDES chooses optimal smoothing constant value for trend and estimation based on interval spacing's among each observation. Second, the attribute selection and classification process is performed based on the attributes temporal pattern and not on the actual observed value. In attribute selection, the significant attribute set is formed using a temporal pattern based tolerance rough set approach. The classification model is built using a decision tree classifier that uses a temporal pattern induced gain ratio as splitting criteria.

3. Methods

The framework for the proposed system is shown in Fig. 1. The system comprises of following components namely temporal data pre-processing, temporal attribute selection and classification.

3.1. Temporal data pre-processing

Temporal data pre-processing aims at handling data irregularities, missing value imputation, temporal pattern derivation and attribute selection. In temporal pre-processing, an enhanced fuzzy inference double exponential smoothing (FIDES) method is presented to impute the missing values and to derive the temporal patterns for each attribute. The relevant attributes are selected using the tolerance rough set concept based on the temporal patterns.

3.1.1. Missing data imputation and temporal pattern extraction using FIDES

To perform missing data imputation and temporal pattern extraction, this work presents a fuzzy inference double exponential smoothing (FIDES) method, which adopts the basic idea of DES proposed by Holt (1957, 2004). The two common strategies for handling missing values are ignoring (deletion) and imputation (Enders, 2010). There are several missing value imputation techniques namely mean, median, nearest neighbour, hot-deck, maximum likelihood, regression (Dempster, Laird, & Rubin, 1977; Little & Rubin, 2002; Enders, 2010). The applicability of these missing imputation techniques in non time series data differs from time series data, due to the presence of temporal patterns like trend, seasonal, cyclic and irregular variations in time-series data. Several research works has been carried out to illustrate the importance of imputing the missing values in time series data (Little & Rubin, 2002). FIDES derives the forecast value, trend and mean value for each clinical observation. The forecasted value is used to impute the missing value. This section first provides the mathemati-

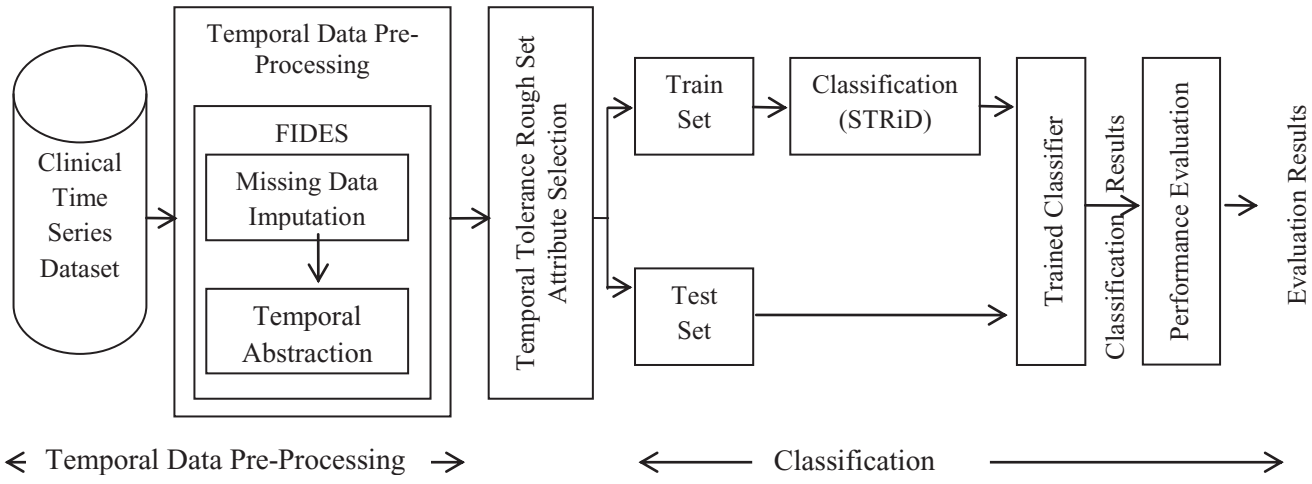

Fig. 1. Proposed framework- STRiD.

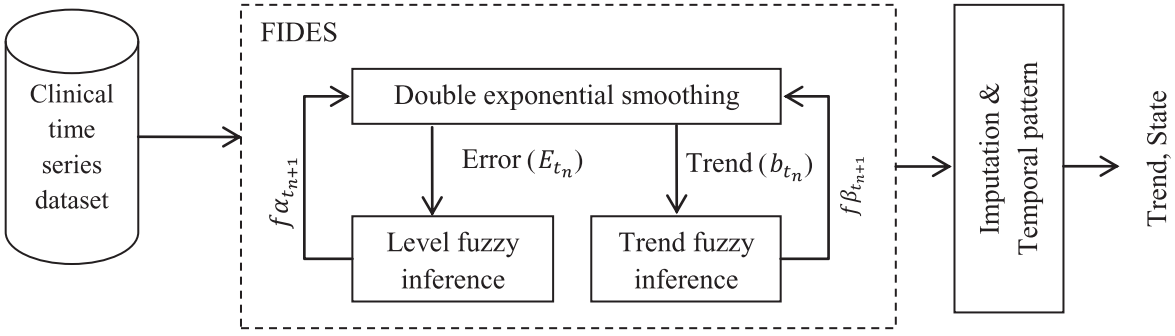

Fig. 2. Fuzzy inference double exponential smoothing (FIDES) system.

cal outline of traditional DES for regular time series and the proposed FIDES for irregular time series.

Let $\{rY_t,\ t \in Ts\}$, be a regular time series with local linear trend, where t is the time, Ts is the set of observations, α is the smoothing constant for the mean estimate $(0 \le \alpha \le 1)$, β is the smoothing constant for trend estimate $(0 \le \beta \le 1)$, rL_{t+1} is the mean at time $t+1$, rL_t is the mean at time t, $rb_{t+1}$is the trend at time $t+1$ and rb_t is the trend at time t. The mean and trend estimate in DES is calculated using Eqs. (1) and (2) respectively. The forecast value is calculated using Eq. (3).

$$rL_{t+1} = \alpha\ rY_{t+1} + (1-\alpha)(rL_t + rb_t) \tag{1}$$

$$rb_{t+1} = \beta(rL_{t+1} - rL_t) + (1-\beta)rb_t \tag{2}$$

$$rF_{t+1} = rL_t\ + rb_t \tag{3}$$

The values of α and β are taken with response to weight adjustments for recent changes, error rate and smoothing effect. The value of α and β is chosen to be in the range [0,1]. In traditional DES, this value remains constant in all the mean and trend estimations. The wrong choice of this value affects the forecasting results. To overcome this, the proposed FIDES method uses a fuzzy inference system (FIS) to dynamically adjust the smoothing constant parameter value α and β for each observation based on the interval spacing among them. The block diagram for the proposed FIDES method is shown in Fig. 2. The major units of FIDES include two fuzzy inference systems namely mean FIS and trend FIS. The mean FIS and trend FIS is responsible for adjusting the smoothing factor α and β for mean and trend estimation based on the interval spacing's in irregular time series. A Mamdani type fuzzy model is used in designing the mean and trend FIS (Jang, Sun, & Mizutani, 1996).

The process of designing a fuzzy inference system includes the following steps: first, fuzzification; second, fuzzy rule base; third, implication and aggregation and fourth, defuzzification. In the fuzzification process, fuzzy sets are defined for the variables and the degree to which each value belongs to a fuzzy set is determined using membership function. The mean and trend fuzzy inference system in the FIDES consists of two input and one output FIS structure. For the mean FIS, the two input variables are error, duration and the output variable is mean parameter. For the trend FIS, the two input variables are trend, duration and the output variable is trend parameter. The input variables are fuzzified using trapezoidal membership function and output variables uses triangular membership function.

Let A represent the fuzzy set, x represents the universe. The trapezoidal function is defined in Eq. (4).

$$\mu_A\ (x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a \le\ x \le b \\ 1, & b \le\ x \le c \\ \dfrac{d-x}{d-c}, & c \le\ x \le d \\ 0, & d \le x \end{cases} \tag{4}$$

where (a,b,c,d) are the four parameters that represents the x coordinates of a trapezoid.

Table 1
Parameter values for the fuzzy sets.

FIS	Variable	Type	Range	Membership Function	Fuzzy Sets	Parameters
Trend	Trend	Input	[−1 1]	Trapezoidal	Decrease	[−1 −1 −0.7 −0.2]
					Stable	[−0.3 −0.2 0.2] 0.3]
					Increase	[0.1 0.5 1 1]
	Duration	Input	[0 100]	Trapezoidal	Minimum	[0 0 9 40]
					Average	[16 39 45 60]
					Maximum	[50 92 100 100]
	Trend parameter	Input	[0 1]	Triangular	Low	[0 0 0.4]
					Medium	[0.1 0.5 0.9]
					High	[0.6 1 1]
Level	Error rate	Input	[0 100]	Trapezoidal	Low	[0 0 4 13]
					Medium	[10 26 34 46]
					High	[44 66 100 100]
	Duration	Input	[0 100]	Trapezoidal	Minimum	[0 0 9 40]
					Average	[16 39 45 60]
					Maximum	[50 92 100 100]
	Level parameter	Input	[0 1]	Triangular	Low	[0 0 0.4]
					Medium	[0.1 0.5 0.9]
					High	[0.6 1 1]

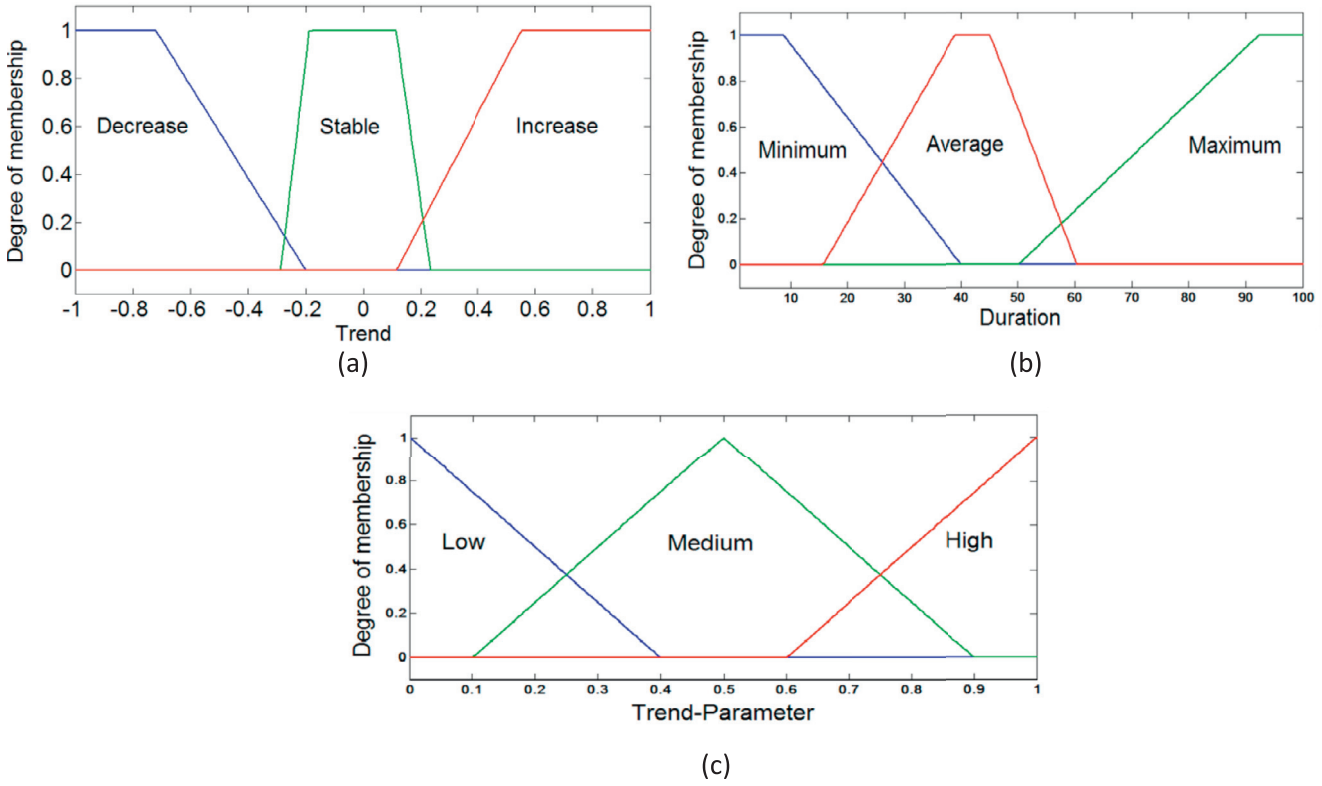

Fig. 3. (a): Membership function plot for input 'Trend', (b): Membership function plot for input 'Duration', (c): Membership function plot for output 'trend-parameter'.

The triangular function is defined in Eq. (5).

$$\mu_A(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ \dfrac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases} \tag{5}$$

where (a,b,c) are the three parameters that represents the x coordinates of a triangle. The number and range of the fuzzy sets are selected based on the expert's opinion. Table 1 summarizes the parameter values used for defining the membership functions of each fuzzy set in FIS.

The membership plot for the trend FIS is shown in Fig. 3. The input variable *trend* defines three fuzzy sets namely Decrease $\mu_{\langle Decrease\rangle}(x)$, Stable $\mu_{\langle Stable\rangle}(x)$ and Increase $\mu_{\langle Increase\rangle}(x)$ using trapezoidal membership function, shown in Fig. 3a. The parameters for $\mu_{\langle Decrease\rangle}(x)$ is [-1 -1 -0.7 -0.2], $\mu_{\langle Stable\rangle}(x)$ is [-0.3 -0.2 0.2 0.3], $\mu_{\langle Increase\rangle}(x)$ is [0.1 0.5 1 1]. The input variable *duration* defines three fuzzy sets namely $\mu_{\langle minimum\rangle}(x)$, $\mu_{\langle average\rangle}(x)$, $\mu_{\langle maximum\rangle}(x)$, shown in Fig. 3b. The parameters for $\mu_{\langle minimum\rangle}(x)$ is [0 0 9 40], $\mu_{\langle average\rangle}(x)$ is [16 39 45 60], $\mu_{\langle maximum\rangle}(x)$ is [50 92 100 100].

The output variable *trend_parameter* is described using three triangular membership function namely $\mu_{\langle low\rangle}(x)$, $\mu_{\langle medium\rangle}(x)$,$\mu_{\langle high\rangle}(x)$ with the parameters [0 0 0.4], [0.1 0.5 0.9] and [0.6 1 1], respectively, shown in Fig. 3.c. In mean FIS, fuzzy sets are defined using two input variables *error_rate*, *duration* and one output variable *mean_parameter*. The membership plot for the mean FIS is shown in Fig. 4. For the *error_rate* three membership functions of type trapezoidal (trapmf) namely

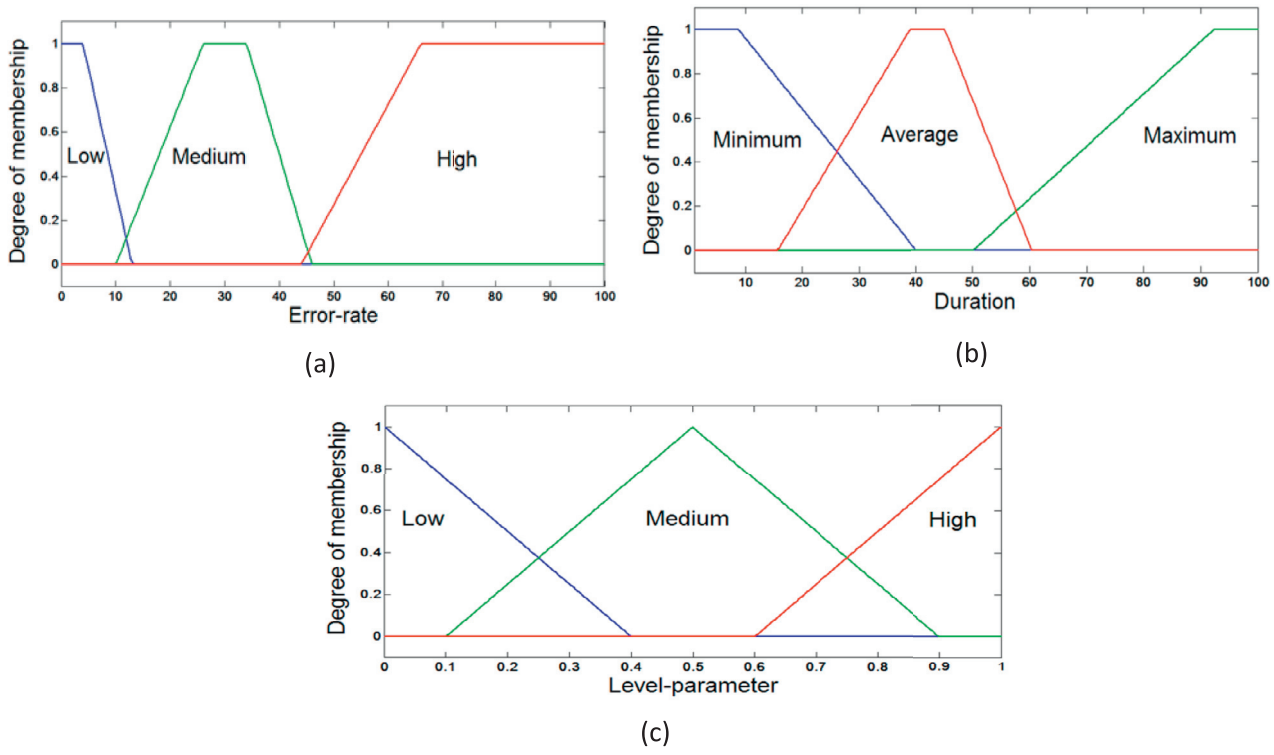

Fig. 4. (a): Membership function plot for input 'Error-rate', (b): Membership function plot for input 'Duration', (c): Membership function plot for output 'level-parameter'.

$\mu_{\langle low\rangle}(x), \mu_{\langle medium\rangle}(x), \mu_{\langle high\rangle}(x)$ is defined with the parameters [0 0 4 13], [10 26 34 46] and [44 66 100 100], respectively, shown in Fig. 4a. The input variable *duration* defines three fuzzy sets namely $\mu_{\langle minimum\rangle}(x)$, $\mu_{\langle average\rangle}(x), \mu_{\langle maximum\rangle}(x)$ that uses trapezoidal (trapmf) membership function is shown in Fig. 4b. The output variable *mean_parameter* is described using three triangular membership function namely $\mu_{\langle low\rangle}(x), \mu_{\langle medium\rangle}(x), \mu_{\langle high\rangle}(x)$ with the parameters [0 0 0.4], [0.1 0.5 0.9] and [0.6 1 1], respectively, shown in Fig. 4c. The fuzzy rules for the FIS are obtained from the domain experts.

Let P, Q, R, S and T be the universes, which represents *error, duration, trend, mean _parameter and trend_parameter* respectively. Let M, N, E, C and Z are the fuzzy sets of the universes where M= $\{\langle decrease\rangle, \langle stable\rangle, \langle increase\rangle\}$, N=$\{\langle minimum\rangle, \langle average\rangle, \langle maximum\rangle\}$, E =$\{\langle low\rangle, \langle medium\rangle, \langle high\rangle\}$ represents fuzzy sets of R, Q and *P* respectively and C= $\{\langle low\rangle, \langle medium\rangle, \langle high\rangle\}$, Z=$\{\langle low\rangle, \langle medium\rangle, \langle high\rangle\}$ represents fuzzy sets of S and T respectively. The fuzzy rule format, If-Then rules with two antecedent and one consequent for the mean FIS and trend FIS is described as follows:

Mean_FIS: If Error (P) is E_i AND Duration (Q) is N_i THEN mean _parameter (S) is C_i

Trend_FIS: If Trend (R) is M_i AND Duration (Q) is N_i THEN trend_parameter (T) is Z_i

These fuzzy rules are converted into fuzzy relation (FR) using Mamdani fuzzy implication functions (Jang et al., 1996). For mean FIS, the fuzzy rule "E X N $\rightarrow$ *C*" is transformed into fuzzy relation (FR) using mamdani fuzzy implication function defined in Eq. (6).

$$FR(E,N,C) = (E\ X\ N)X\ C = \int_{P\ X\ Q\ X\ S} \mu_E\ (P) \Lambda\ \mu_N\ (Q) \Lambda\ \mu_C\ (S)/(P\ X\ Q\ X\ S) \quad (6)$$

For trend FIS, the fuzzy rule "M X N $\rightarrow$ *Z*" is transformed into fuzzy relation using Mamdani fuzzy implication function defined in Eq. (7).

$$FR(E,N,Z) = (M\ X\ N)X\ Z = \int_{R\ X\ Q\ X\ T} \mu_M\ (R) \Lambda\ \mu_N\ (Q) \Lambda\ \mu_Z\ (T)/(R\ X\ Q\ X\ T) \quad (7)$$

where $\mu_E(P)$, $\mu_M(R)$, $\mu_N(Q)$, $\mu_C(S), \mu_Z(T)$ denotes fuzzy membership functions for *error, trend, duration, mean parameter and trend parameter.* This work uses generalized modus ponens (GMP) for reasoning as defined in Eqs. (8) and (9).

$$c' = \left(E'X\ N'\right) \circ (E\ X\ N \rightarrow C) \quad (8)$$

$$z' = \left(M'X\ N'\right) \circ (M\ X\ N \rightarrow z) \quad (9)$$

where E′, N′ and M′ are the obtained error, duration and trend values, c' is the derived consequent value for the corresponding E′ and N′ . z' is the derived consequent value for the corresponding M′ and N′. The inferred fuzzy rules are aggregated using max aggregation operation. To obtain the final crisp value for the smoothing mean parameter ($f\alpha_{t_{n+1}}$) and trend parameter ($f\beta_{t_{n+1}}$) a centroid defuzzification is adopted. Centroid defuzzification returns the center of area under the curve (Jang et al., 1996). This is defined using Eqs. (10) and (11).

$$f\alpha_{t_{n+1}} = \int_S \frac{\mu_C\ (S) S\ dS}{\mu_C\ (S) dS} \quad (10)$$

$$f\beta_{t_{n+1}} = \int_T \frac{\mu_z\ (T) T\ dT}{\mu_Z\ (T) dT} \quad (11)$$

FIDES computes the mean, trend and forecasted value for each clinical observation using Eqs. (12)–(14). Let $Y = \{Y_{t_n}, n \in TS; t_{n+1} > t_n\}$ is an irregular time series, where t_n is the time, TS is the set of observations.

$$L_{t_{n+1}} = f\alpha_{t_{n+1}} Y_{t_{n+1}} + \left(1 - f\alpha_{t_{n+1}}\right)\left(L_{t_n} + b_{t_n}\right) \quad (12)$$

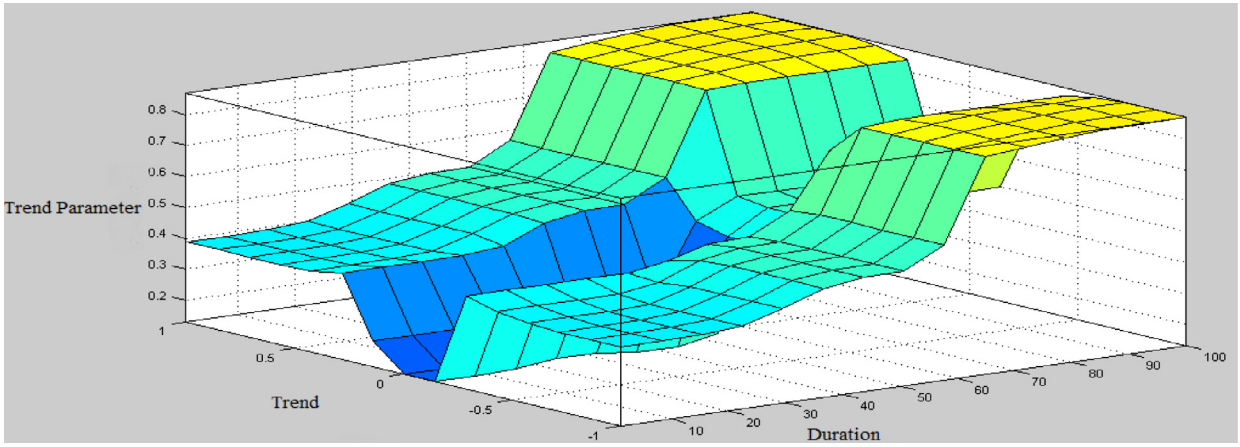

Fig. 5. Surface plot for trend FIS.

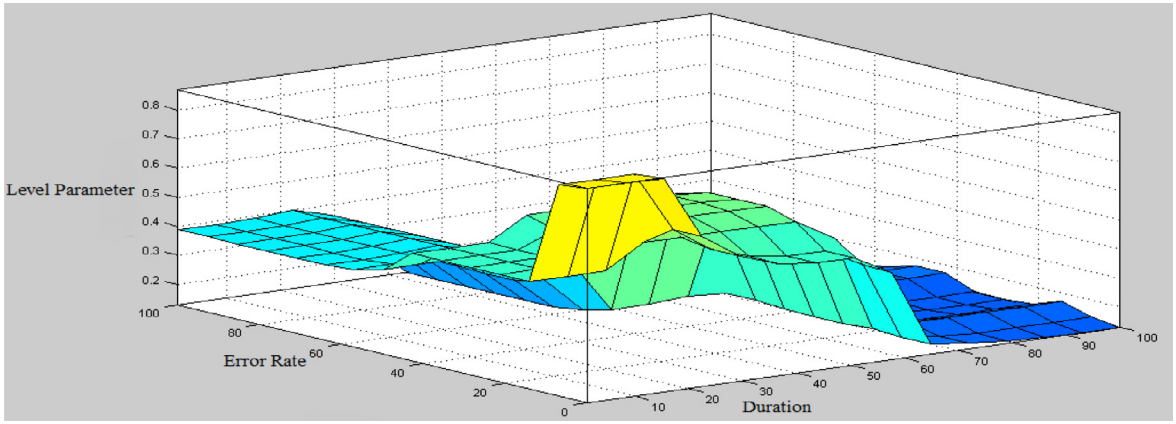

Fig. 6. Surface plot for mean FIS.

$$b_{t_{n+1}} = f\beta_{t_{n+1}}\left(L_{t_{n+1}} - L_{t_n}\right) + \left(1 - f\beta_{t_{n+1}}\right)b_{t_n} \tag{13}$$

$$F_{t_{n+1}} = L_{t_n} + b_{t_n} \tag{14}$$

where $f\alpha_{t_{n+1}}$ is the smoothing constant for the mean $(0 \le \alpha \le 1)$ at time t_{n+1}, $f\beta_{t_{n+1}}$ is the smoothing constant for the trend $(0 \le \beta \le 1)$ at time t_{n+1}, $L_{t_{n+1}}$ is the mean estimate, $b_{t_{n+1}}$ is the trend estimate and $F_{t_{n+1}}$ is the forecasted value at time t_{n+1}. For each patient observation, FIDES assigns values to the smoothing parameters based on the error rate, trend and duration.

The surface plot for the designed trend FIS is shown in Fig. 5. It can be inferred from Fig. 5 that as the duration increases and there is high trend fluctuations the smoothing trend parameter value is assigned value close to 1 whereas when there is minimum duration and lower trend fluctuations the smoothing trend parameter value is assigned value close to 0.

Fig. 6 shows the surface plot for the designed Mean FIS. The value of *mean _parameter* is obtained based on the error rate and duration of each clinical observation. From Fig. 6, it can be inferred that as the duration increases and there is moderate error rate the smoothing mean parameter value is assigned value close to 1 whereas when there is minimum duration and lower error rate the smoothing mean parameter value is assigned value close to 0.

The error measures such as MSE (mean squared error), MAD (mean absolute deviation), error rate, MAPE (mean absolute percentage error) are computed to evaluate the performance of the forecasting model (Cooray, 2008).

The following algorithm (Algorithm 1) illustrates the steps followed in temporal preprocessing of the clinical data.

To derive the temporal pattern for each patient lab tests, the initial estimates for mean (L_{t_0}) and trend (b_{t_0}) is computed using least square trend estimation. The forecasted value for corresponding observation is calculated using Eq. (14). This forecasted value is

Algorithm 1 Time Series Data Pre-Processing.

Input: Dataset (Y), Smoothing constants α and β, N is the Number of lab test, P is the patient medical identity (MID)
Output: Mean set (L), Trend (b), Error (E), Forecasted Value (F)
//Perform the following steps for patient (MID) =P
1. FOR lab test 1 to N do
2. Get the initial estimates L_{t_0} and b_{t_0} by using least squares estimation on lab test .
3. Initialize the mean and trend parameter $f\alpha_{t_0}$ and $f\beta_{t_0}$ $f\alpha_{t_0} = 1 - (1 - \alpha)^{\left(\frac{q}{\delta max}\right)}, f\beta_{t_0} = 1 - (1 - \beta)^{\left(\frac{q}{\delta max}\right)}$, Where q and δ_{max} is average and maximum interval spacing for the lab test.
4. If Y_{t_1} is missing then assign $Y_{t_1} = L_{t_0} + b_{t_0}$ where Y_{t_1} is an observed value at time t_1.
5. Compute L_{t_1}, b_{t_1} and F_{t_1} .
6. WHILE (there is observations for the lab test)
7. Compute the smoothing constant for mean and trend $f\alpha_{t_n}$ and $f\beta_{t_n}$ using Eqs. (10) and (11).
8. If Y_{t_n} is missing then assign $Y_{t_n} = F_{t_n}$
9. Calculate the mean $(L_{t_{n+1}})$ and trend estimate $(b_{t_{n+1}})$ and forecasted value $F_{t_{n+1}}$ using Eqs. (12), (13) and (14).
10. END WHILE
11. Calculate Error Rate, MSE, MAD and MAPE.
12. END FOR

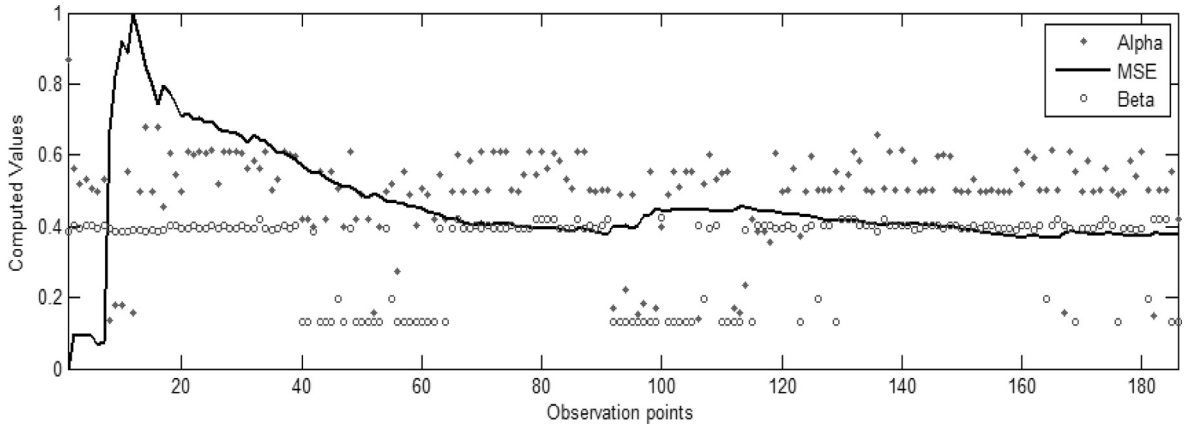

Fig. 7. Plot to depict MSE for different alpha and beta - hepatitis patient record subset (patient_ MID = 1, lab examination = ALB).

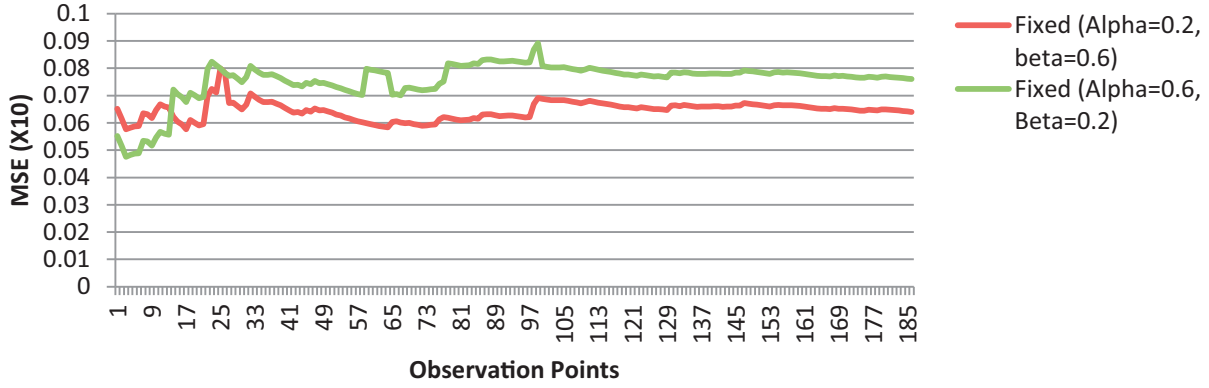

Fig. 8. Plot to depict MSE for fixed Alpha and Beta: Hepatitis Patient Record Subset (patient_ MID = 1, lab examination = ALB).

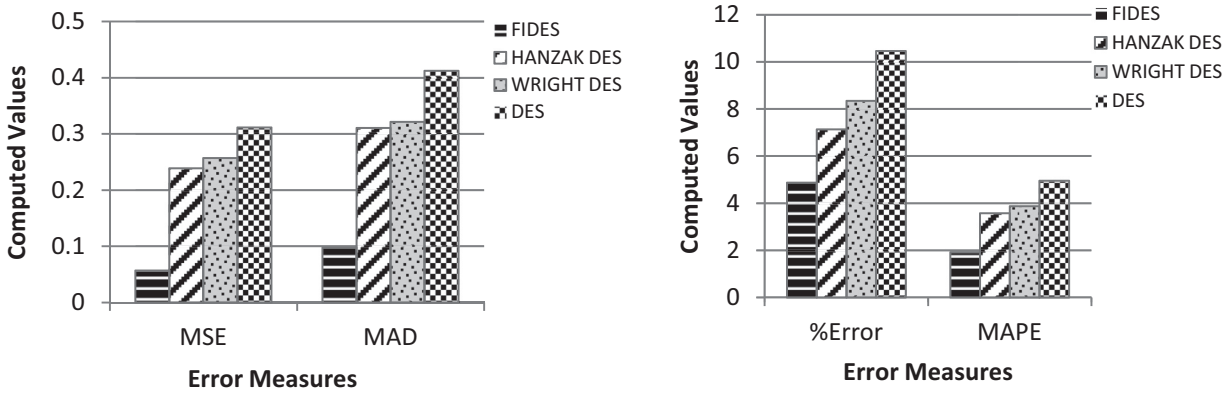

Fig. 9. Comparison of performance measures (MSE, MAD, % Error, MAPE) on FIDES, DES, Hanzak and Wright updated DES.

used to impute missing value in the observed time series data. It has been inferred through experimental results that the presented FIDES has effectively handled irregularities in the time series by adjusting the smoothing constant parameter values ($\alpha_{t_{n+1}}$, $\beta_{t_{n+1}}$) dynamically using fuzzy inferences. The forecasting accuracy rate was increased compared to the other improved versions of DES. The derived temporal pattern (trend and mean) for each patient's lab test are used in attribute selection and classification process. For each patient's clinical observation, the trend refers to the rate of change (increase or decrease) over period and the mean refers to the state.

3.1.2. Temporal interpreted tolerance rough set attribute selection

In clinical domain, a patient undergoes several common laboratory tests before diagnosed for a particular disease. Thus, clinical time series data contains high dimensional set of attributes which includes the common laboratory test and disease specific laboratory test taken on each patient. This work reduces these irrelevant common laboratory tests before classification process to improve the accuracy of decision-making system for diagnosing a particular disease (Dash & Liu, 1997). Rough set is a mathematical concept proposed by Pawlak (1982). Rough set has been popularly used in attribute selection process. The major advantage of

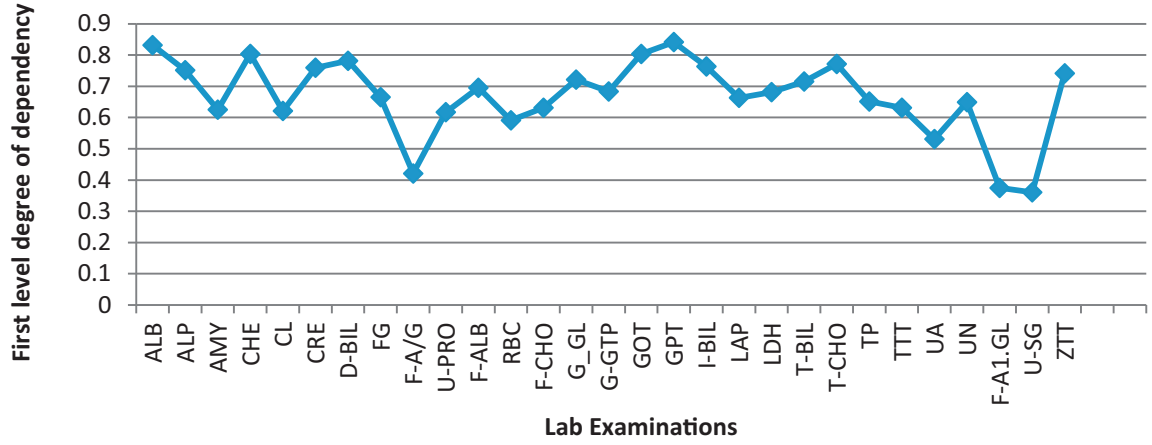

Fig. 10. Temporal first level degree of dependency- hepatitis.

using rough set in attribute selection is because it does not require any prior knowledge or information about the data (Komorowski, Pawlak, Polkowski, & Skowron, 2002; Zhong & Skwron, 2001).

Attribute selection using rough set is performed by forming equivalence classes and then finding significance of the attributes based on the degree of dependency (Pawlak, 1982; Komorowski et al., 2002). The generation of equivalence class becomes complex when the attributes hold real value data. Therefore, to address the real value data problem, Jensen and Shen (2007) have used the concept of tolerance rough set. Tolerance rough set forms tolerance classes based on the similarity in attributes, which represents an elementary portion of extracted knowledge. In this work, a temporal interpreted tolerance rough set approach is presented for attribute selection. The basic logic from tolerance quick reduct proposed by Jensen and Shen (2007) was adapted and enhanced by incorporating temporal pattern extracted from FIDES process. A temporal degree of dependency is computed to identify the significance of each attribute.

An information system in rough set is represented as $I = (\mathbb{U}, A)$, where $\mathbb{U} = \{x_1, ..., x_i, ..x_j., x_n\}$ is called as an universe which is a nonempty set of finite objects, x_i refers to the i^{th} object and A={a_1, . . ., a_k, . a_m} is the knowledge in an universe which is the nonempty finite set of attributes; a_k refers to the k^{th} object (Pawlak, 1982).

The tolerance classes of a relation are generated by the similarity relation of an object. Lower and upper approximations are the two main operations that are used in characterizing this knowledge. Lower approximation of set contains all elements that surely belong to the set. Upper approximation of set contains all elements that possibly belong to set. Based on the approximations three regions are characterized namely positive, negative and boundary region. Positive region contains all the objects of $'\mathbb{U}'$ that can be classified to equivalence classes of $'\mathbb{U}/Q'$ using the information in the attribute set B, Q is the decision attribute. Negative region contains all the objects, which is certainly non-member of X. Boundary region contains all objects, which is possibly member of X. The degree of dependency of the attributes is an important factor to be considered in the attribute selection process.

The traditional tolerance quick reduct approach determines the relevant attribute set based on the attributes values, whereas the proposed attribute selection method determines the attribute relevance and significance based on the temporal pattern mean (L) and trend (b) for each attribute. A temporal similarity measure computes the similarity for each attribute based on its temporal pattern. This temporal based similarity measures defines the lower approximations to construct positive regions. Finally, the significance of the attribute is measured using temporal tolerance based degree of dependency. The computed significance attributes are used to predict the effectiveness of the attribute in the classification process.

The following computations are done to identify the significance of an attribute: temporal tolerance similarity measure, temporal tolerance class relation, temporal tolerance lower approximation, temporal tolerance positive region, temporal tolerance degree of dependency and reduct approximation error.

Let xi, xj be the objects in the universe $\mathbb{U}$, a is the attribute, $a \in A$, $a_{\langle b\rangle}(\mathrm{xi})$ and $a_{\langle b\rangle}(\mathrm{xj})$ are the trends of xi and xj for the attribute a, $a_{\langle L\rangle}(\mathrm{xi})$ and $a_{\langle L\rangle}(\mathrm{xj})$ are the means of xi and xj for the attribute a, τ is the tolerance value (0 to 1), $a_{\langle L\rangle \max}$ and $a_{\langle L\rangle \min}$ is are the maximum and minimum mean value, $a_{\langle b\rangle \max}$ and $a_{\langle b\rangle \min}$ is are the maximum and minimum trend value, $\bar{a}$ represents the mean, B is the attribute set; where $B \subseteq A$, X is the subset of the universe $X \subseteq \mathbb{U}$, $\mathrm{a} < (L, b), \tau >$ is the mean and trend for attribute a, C is the condition attribute, Q is the decision attribute and R is the reduced attribute set $R \subseteq C$.

The temporal tolerance similarity measure and temporal tolerance class relation is defined in Eqs. (15) and (16).

$$\mathrm{SIM}_{\mathrm{a}\langle(L,b),\tau\rangle}(\mathrm{xi},\ \mathrm{xj}) = 1 - \bar{a}\left\{\left(\frac{a_{\langle L\rangle}(\mathrm{xi}) - a_{\langle L\rangle}(\mathrm{xj})}{a_{\langle L\rangle max} - a_{\langle L\rangle min}}\right),\ \left(\frac{a_{\langle b\rangle}(\mathrm{xi}) - a_{\langle b\rangle}(\mathrm{xj})}{a_{\langle b\rangle max} - a_{\langle b\rangle min}}\right)\right\} \tag{15}$$

$$\mathrm{tempTOL}\tau(\mathrm{B}) = \{(xi,\ xj) \epsilon\ \mathrm{U}\ |\forall\ \mathrm{a}\epsilon\ \mathrm{B},\ (\mathrm{xi},\ \mathrm{xj}) \in \mathrm{SIM}_{\mathrm{a}\langle(L_{p_n},b_{p_n}),\tau\rangle}(\mathrm{xi},\ \mathrm{xj}) \tag{16}$$

These selected attributes are called as an approximate reduct from the whole expert suggested attributes. The temporal tolerance lower approximation, temporal tolerance positive region and temporal tolerance degree of dependency are calculated using Eqs. (17)–(19).

$$\underline{B_\tau}X = \{x | SIM_{B\langle(L,b),\tau\rangle}(x) \subseteq X\} \tag{17}$$

$$POS_{B,\tau}(Q) = \cup_{\mathrm{x}\in \mathrm{U}/\mathrm{Q}}\ \underline{B_\tau}X \tag{18}$$

$$K = \gamma_{B,\tau}(Q) = |POS_{B,\tau}(Q)|/|\mathbb{U}| \tag{19}$$

The reduct approximation error $\varepsilon_{(C,\ Q)}(R)$ is calculated using Eq. (20) adopted from Pawlak (1982).

$$\varepsilon_{(C,Q)}(R) = 1 - (\gamma_{R,\tau}(Q)/\gamma_{C,\tau}(Q)) \tag{20}$$

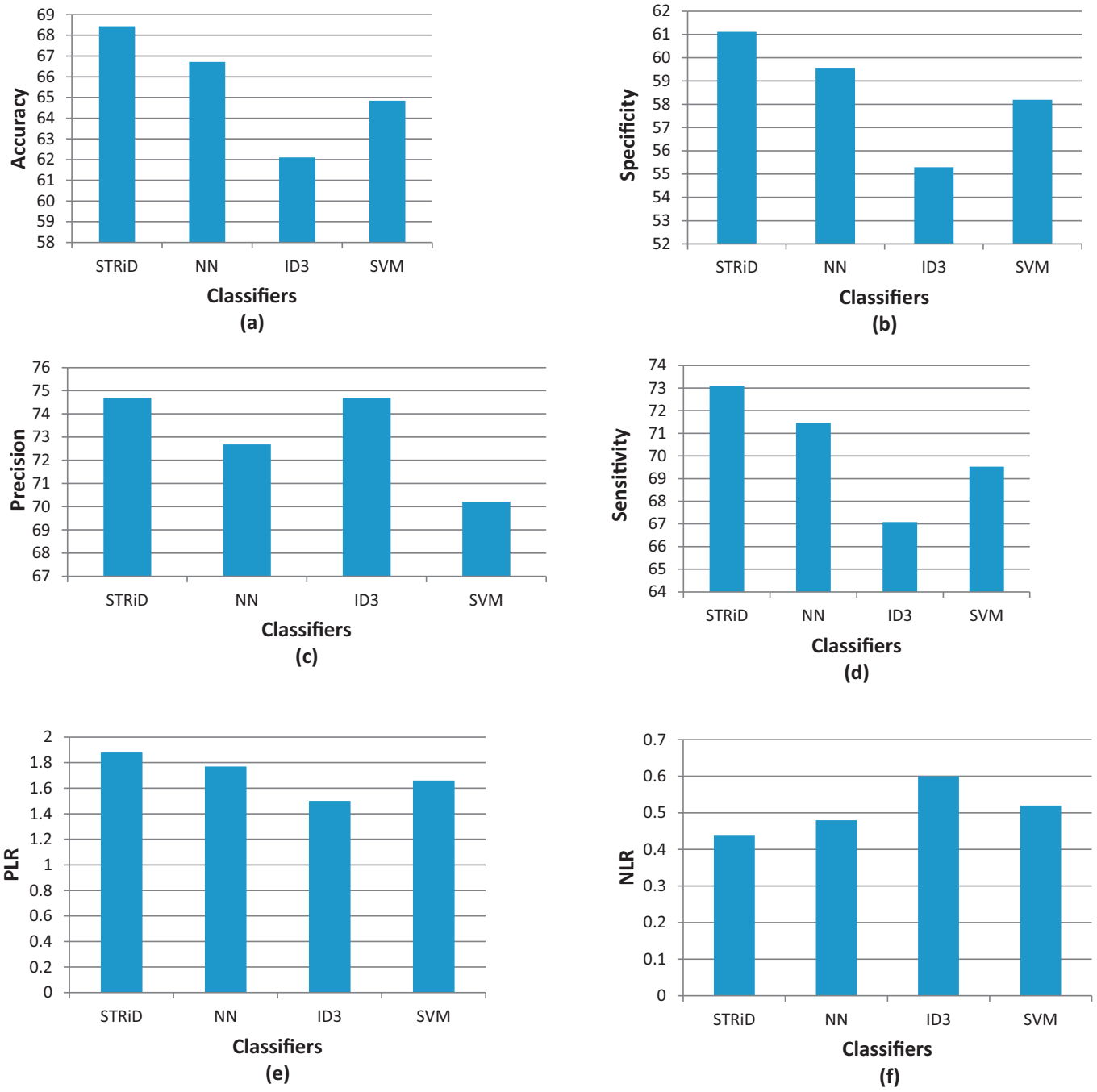

Fig. 11. Classification accuracy computed before without applying imputation but with attribute selection for Hepatitis dataset (a) Accuracy (b) Specificity (c) Precision (d) Sensitivity (e) PLR (f) NLR.

If $\varepsilon_{(C,Q)}(R) = 0$, then R is reduct of C. The algorithm temporal interpreted tolerance rough set attribute selection depicts the steps followed in the presented attribute selection process. This contribution extends tolerance quick reduct (Jensen & Shen, 2007) procedure to support the temporal constraints by using temporal similarity measure shown in Eq. (15). In tolerance rough computations for each clinical attribute its temporal pattern mean and trend obtained from smoothing process is considered rather than its actual observed value.

The Table 2 shows the subset of temporal pattern derived from FIDES process for the lab examination (ALB, T-CHO) of five patients selected at random. The process of computing temporal tolerance degree of dependency using the Temporal_ Tolerance_ RoughSet_Attribute Selection algorithm (Algorithm 2) for the sample data in Table 2 is illustrated below.

Table 2
Temporal interpretations for lab test (ALB, T-CHO), patient (1-5).

Patients	ALB		T-CHO		Class (Hepatitis)
	Level	Trend	Level	Trend	
1	5.2413	-0.0591	216.446	-6.8603	B
2	4.9989	-0.0273	235.476	1.0928	C
3	5.2773	0.02724	203.732	-7.8906	C
4	5.0581	0.0075	236.345	1.7707	C
5	4.8704	-0.0504	208.713	-8.7278	B

Algorithm 2 Temporal Tolerance RoughSet_Attribute Selection.

Input: Dataset (X), τ-tolerance factor, A represents the attribute set
Output: Reduced attribute set R.

1. Compute temporal similarity measure for each attribute using Eq. (15).
2. Generate temporal tolerance class using Eq. (16).
3. Determine lower approximation, positive region based on temporal similarity measure using Eqs. (15), (17) and (18).
4. Compute significance of each attribute using tolerance based degree of dependency based on temporal similarity measure using Eqs. (15) and (19).
5. Select the significant attribute with high degree of dependency; include it in reduced attribute set (R).
6. Check if there is attributes from (A) to form subset with attributes in (R), If there is no attributes to form subset go to step 8 otherwise go to step 7.
7. Form possible significant candidate attribute subset (S). Repeat step 2-6 for thesignificant set (S).
8. Compute the reduct approximation error using Eq. (20).
9. Return reduced attribute set (R).

The temporal classes generated for the attributes ALB, T-CHO and decision attribute Hepatitis are as follows:

$$U/SIM_{ALB_{\langle (L,b),\tau\rangle}} = \{\{1, 2, 5\}, \{3, 4\}\}$$
$$U/SIM_{T-CHO_{\langle (L,b),\tau\rangle}} = \{\{1, 3, 5\}, \{2, 4\}\}$$
$$U/SIM_{Hepatitis,\ \tau} = \{\{1, 5\}, \{2, 3, 4\}\}$$

The lower approximations for the decision attribute Hepatitis based on derived temporal pattern of lab examination ALB and T-CHO is calculated as follows,

$$\underline{ALB_{\langle (L,b),\tau\rangle}}\{1, 5\} = \{x|\ SIM_{ALB_{\langle (L,b),\tau\rangle}}(x) \subseteq \{1, 5\}\} = \{\}$$
$$\underline{ALB_{\langle (L,b),\tau\rangle}}\{2, 3, 4\} = \{x|\ SIM_{ALB_{\langle (L,b),\tau\rangle}}(x) \subseteq \{2, 3, 4\}\} = \{3, 4\}$$
$$\underline{T-CHO_{\langle (L,b),\tau\rangle}}\{1, 5\} = \{x|\ SIM_{T-CHO_{\langle (L,b),\tau\rangle}}(x) \subseteq \{1, 5\}\} = \{\}$$
$$\underline{T-CHO_{\langle (L,b),\tau\rangle,C}}\{2, 3, 4\} = \{x|\ SIM_{T-CHO_{\langle (L,b),\tau\rangle}}(x) \subseteq \{2, 3, 4\}\} = \{2, 4\}$$

The positive region for the obtained approximations of lab examination ALB and T-CHO is constructed as follows,

$$POS_{ALB_{\langle (L,b),\tau\rangle}}(\text{Hepatitis}) = \cup_{x\in U/\text{Hepatitis}}\underline{ALB}_{\langle (L,b),\tau\rangle,B}X$$
$$= \underline{ALB_{\langle (L,b),\tau\rangle,B}}\{1, 5\} \cup \underline{ALB_{\langle (L,b),\tau\rangle,B}}\{2, 3, 4\} = \{3, 4\}$$
$$POS_{T-CHO_{\langle (L,b),\tau\rangle}}(\text{Hepatitis}) = \cup_{x\in U/\text{Hepatitis}}\underline{T-CHO_{\langle (L,b),\tau\rangle,C}}$$
$$= \underline{T-CHO_{\langle (L,b),\tau\rangle,T}}\{1, 5\} \cup \underline{T-CHO_{\langle (L,b),\tau\rangle,C}}\{2, 3, 4\} = \{2, 4\}$$

Like-wise positive regions are calculated for all the other attributes. Finally based on positive regions, the tolerance degree of dependency (K) with temporal similarity measure for lab examination ALB and T-CHO is computed as follows,

$$K = \left|POS_{ALB_{\langle (L,b),\tau\rangle}}(\text{Hepatitis})\right|/|U| = \{3, 4\}/\{1, 2, 3, 4, 5\} = 2/5$$

$$K = \left|POS_{T-CHO_{\langle (L,b),\tau\rangle,C}}(\text{Hepatitis})\right|/|U| = \{2, 4\}/\{1, 2, 3, 4, 5\} = 2/5$$

The attributes with highest dependency is selected and subset of these attributes are formed. The above steps are repeated and the attribute subset with high dependency is selected to be in the significant reduced set. This process continues until there is no attributes to form new subsets. The significant reduced set returns the identified relevant attributes.

3.2. Temporal classification

In the proposed STRiD framework, a decision tree classifier is used to build a temporal classification model. The selected attributes from the attribute selection process and its temporal pattern are used in the classification process. This work uses a C4.5 classifier, which uses gain ratio as splitting criteria for constructing the decision tree. However, the proposed framework computes the gain ratio using the temporal pattern of each clinical attribute instead of its actual observed quantitative values. The trend value is normalized in the range [-1 to 1] for each attribute using min-max normalization (Han & Kamber, 2001). The positive value (>0) in the trend column shows an increase in the trend and negative value (< 0) shows a decrease in the trend, zero value indicates that it is stable with respect to the patient date of examination. Accordingly, the trend (T) of a clinical attribute is represented as increasing (I+), decreasing (D-) or stable (S). The mean of a clinical attribute represents the state which is categorised as low (L), high (H), normal (N) based on its range specified by a medical expert.

In the decision tree, the test attribute in each node is represented with three elements namely name, state and trend of the attribute, denoted as <attribute name, trend (T), state (S)>. The attribute name refers to the lab test, state refers to the mean value of lab test as low (L), high (H), normal (N) and trend refers to the growth rate which is represented as increasing (I+), decreasing (D-) and stable (S) over a period. During the decision tree construction the test attribute is selected using the trend and state of the attribute and not on the actual observed value.

Let A be the investigated attribute, p_i be the probability of class i in dataset D, m is the number of class values, $|D_{j_{\langle T,S\rangle}}|$ is the number of records with attribute j^{th} trend T and state S in dataset D, $|D|$ represents total number of observations in dataset D, $\langle T, S\rangle$ represents the temporal pattern trend and state, $D_{j_{\langle T,S\rangle}}$ represents subset of D that contains attribute j^{th} trend T and state S, v is the attributes temporal pattern. To classify the tuple in the dataset an entropy measure, $Info(D)$ used by Quinlal (1993) which is represented in Eq. (21) is adopted.

$$Info(D) = -\sum_{i=1}^{m} p_i log_2(p_i) \tag{21}$$

The quantitative values of each clinical attribute are summarized into two temporal pattern namely trend and state. The temporal interpreted attribute split information, $Tinfo_{A\langle T,S\rangle}(D)$ is computed using Eq. (22). This is computed for each attribute to classify the dataset D. In this work, the attribute split information and the gain ratio has been computed using the temporal pattern and not on the actual observed quantitative values.

$$Tinfo_{A\langle T,S\rangle}(D) = \sum_{j=1}^{v}\left|\frac{D_{j_{\langle T,S\rangle}}}{|D|}\right| \times Info(D_{j\langle T,S\rangle}) \tag{22}$$

Finally, the gained information is computed using temporal gain ratio, $TGain(A)$ represented in Eq. (23).

$$TGain(A) = Info(D) - Tinfo_{A\langle T,S\rangle}(D) \tag{23}$$

This attribute selection measure is computed for all the attributes and the attribute with highest temporal gain ratio is selected as test attributes (or partitioning attribute). Based on the selected test attribute the subsets of datasets are formed and this process is continued recursively until there is no attribute left for partitioning. Classification results such as sensitivity, accuracy, specificity, error rate and precision is derived from the constructed tree.

4. Experimental results and discussions

This section discusses the experimental settings and evaluation criteria used to assess the effectiveness of the proposed framework with hepatitis and thrombosis dataset.

Table 3
Dataset summary.

Dataset	Total Records	Lab tests	Suggested Lab test (by Expert)	Total Patients	Average Missing value (%)
Hepatitis	1,565,876	983	29	771	11
Thrombosis	57,543	564	33	1000	8

Table 4
Results of FIDES, Hanzak updated DES Cipra and Hanzak (2008) and classical DES Holt (1957, 2004) for subset of hepatitis patient (patient_ MID = 1, lab examination = ALB, year = 1981).

Interval (days)	Value (Y_{t_n})	FIDES			Hanzak updated DES Cipra and Hanzak (2008)			Classical DES Holt (1957, 2004)		
		Level (L_{t_n})	Trend (b_{t_n})	Forecast (F_{t_n})	Level (L_{t_n})	Trend (b_{t_n})	Forecast (F_{t_n})	Level (L_{t_n})	Trend (b_{t_n})	Forecast (F_{t_n})
––	––	5.3500	−0.0200	––	5.3500	-0.0200	––	5.400	−0.020	
1	5.4	5.3579	−0.0193	5.3300	5.3440	−0.0144	5.3300	5.390	−0.016	5.380
35	5.2	5.3146	−0.0089	5.3386	5.1993	−0.0041	5.3296	5.287	−0.051	5.374
28	5.2	5.2890	−0.0056	5.3057	5.1998	0.0000	5.1951	5.218	−0.058	5.236
35	5.4	5.3040	−0.0028	5.2834	5.3999	0.0057	5.1998	5.280	−0.010	5.160
35	5.2	5.2829	−0.0019	5.3012	5.2002	−0.0057	5.4056	5.235	−0.024	5.270
27	5.3	5.2848	−0.0012	5.2811	5.2994	0.0037	5.1945	5.255	−0.006	5.211
35	5.4	5.3046	−0.0005	5.2835	5.4000	0.0029	5.3031	5.325	0.024	5.249
28	4.7	5.1968	−0.0016	5.3041	4.7015	−0.0249	5.4029	5.024	−0.106	5.349
14	4.8	5.1316	−0.0021	5.1952	4.7811	0.0056	4.6766	4.859	−0.129	4.918
14	5.5	5.1866	−0.0011	5.1295	5.4719	0.0493	4.7867	5.115	0.025	4.730
14	5.4	5.2165	−0.0005	5.1855	5.4335	−0.0027	5.5212	5.270	0.077	5.139
21	4.8	5.1581	−0.0011	5.2160	4.8055	−0.0299	5.4308	5.073	−0.033	5.346
21	5.1	5.1495	−0.0009	5.1570	5.0915	0.0136	4.7756	5.070	−0.021	5.041

4.1. Dataset description

The proposed work uses two clinical time series dataset for experimentation. The first dataset contains the results of laboratory examinations taken on hepatitis patients. The second dataset contains the results of laboratory examinations taken on thrombosis patients. These two dataset was collected from Chiba university hospital between the years 1982 to 2001 and 1980 to 1990 respectively. The datasets were made available through a data mining competition Principles and Practice of Knowledge Discovery in Databases (PKDD) discovery challenge (Hepatitis Dataset for Discovery Challenge, Thrombosis Dataset for Discovery Challenge). These datasets were used in our previous work (Nehemiah et al., 2007). The access to the dataset is currently unavailable.

Hepatitis dataset contains laboratory examination reports of 771 hepatitis B and C patients. In hepatitis dataset, each patient has undergone 983 laboratory tests. It has to be noted that not all the laboratory tests relates to hepatitis. Hence, by combining the expert guidance and the dataset descriptions given with the dataset and by Ohsaki, Sato, Yokoi, and Yamaguchi (2002), 29 suggested tests have been selected for experimentation. The average missing value percentage in hepatitis dataset is 11.

Thrombosis dataset consists of 1000 patient's laboratory examination reports. Each patient has undergone 564 laboratory tests. On combining the expert's knowledge and dataset descriptions given with the dataset (Jensen, 2001; Tsumoto, 1999; Zytkow & Gupta, 2001), 33 suggested tests have been selected for experimentation. The average missing value percentage in thrombosis dataset is 8. Table 3 shows the general description of the dataset for hepatitis and thrombosis datasets.

All the data included in the hepatitis and thrombosis dataset is associated with an unevenly spaced (irregular) time stamp. The suggested lab tests for hepatitis and thrombosis patients along with their date of examination is considered as temporal input attribute set for demonstrating the proposed bio-statistical mining approach. The observations of a few patients were not recorded properly and their EHR reports contain more than 30% of incomplete data and hence those patients were not included for experimentation. Thus, for Hepatitis datasets 499 records and for thrombosis data sets 770 records only were considered in further experimentation.

4.2. Results and discussions

This section discusses the experimental results and discussion used to assess the effectiveness of the proposed framework with Hepatitis and Thrombosis dataset. The work proposed has been initiated with an experiment by applying STRiD framework to the Hepatitis and Thrombosis datasets. This section provides a detail discussion about the experimental settings and the evaluation process. The clinical data was randomly divided into two sets train and test which contains 75% and 25% of patient's respectively. This process is repeated for 10 runs. For every repeated runs, a 10-fold cross validation strategy is used to evaluate the training phase with training set and remaining set is kept to see how well the trained model generalizes to unseen data.

To perform missing value imputation and temporal pattern extraction, FIDES dynamically adjusts the smoothing constant α and β in the mean and trend estimation using fuzzy inference system. The smoothing factor (α and β) for mean and trend is any value chosen between 0 and 1. However, when 'α' is closer to 1 it denotes that more weight is given to the recent observations. If stable predictions with smoothed random variation is desired then a small value of 'α' is desired. If a rapid response to a real change in the pattern of observations is desired, a large value of 'α' is appropriate. Similarly when 'β' is closer to 1 the trend estimate is updated with respect to forecast error. If 'β' is closer to 0 the trend estimate is updated constantly. Initially the smoothing constant 'α' and 'β' was taken to be 0.2 and 0.4 respectively and based on the interval spacing between the observations and the fuzzy inference the smoothing factors are adjusted.

Table 5
Performance measures FIDES, Hanzak updated DES Cipra and Hanzak (2008) Vs classical DES Holt (1957, 2004) for subset of Hepatitis and Thrombosis.

Patients	Hepatitis: LAB_EXAMINATION = ALB, Patient = 1-5											
	FIDES				Hanzak updated DES Cipra and Hanzak (2008)				Classical DES Holt (1957, 2004)			
	MSE	MAD	Error(%)	MAPE	MSE	MAD	Error(%)	MAPE	MSE	MAD	Error (%)	MAPE
1	0.038	0.155	2.600	3.054	0.070	0.182	2.572	3.651	0.045	0.171	3.190	3.376
2	0.014	0.093	3.185	1.777	0.109	0.248	4.004	4.888	0.084	0.175	6.035	4.840
3	0.013	0.095	1.939	1.831	0.070	0.182	2.572	3.651	0.080	0.191	4.034	4.766
4	0.016	0.111	2.945	2.110	0.105	0.253	2.061	4.970	0.080	0.140	5.712	4.845
5	0.013	0.098	0.880	1.864	0.005	0.070	1.296	1.296	0.062	0.210	2.917	4.112
	Thrombosis: LAB_EXAMINATION =WBC, Patient =1-5											
Patients	FIDES				Hanzak updated DES Cipra and Hanzak (2008)				Classical DES Holt (1957, 2004)			
	MSE	MAD	Error(%)	MAPE	MSE	MAD	Error(%)	MAPE	MSE	MAD	Error (%)	MAPE
1	0.078	0.223	1.528	4.358	0.105	0.225	2.245	4.370	0.072	0.174	8.237	4.537
2	0.029	0.133	0.711	2.666	0.064	0.199	1.855	3.897	0.063	0.201	2.466	3.939
3	0.026	0.126	3.254	2.552	0.109	0.248	4.004	4.888	0.072	0.189	6.812	4.481
4	0.067	0.202	0.152	3.941	0.064	0.201	6.071	4.005	0.069	0.191	0.746	4.326
5	0.028	0.132	1.057	2.658	0.059	0.195	2.071	3.923	0.049	0.189	3.437	3.669

Table 6
Results of attribute selection-Temporal tolerance rough sets.

Dataset	Expert suggested attributes	Attribute selection Temporal tolerance rough sets		Reduct approximation error $\varepsilon(R)$
		Relevant attributes	Selected attribute list	
Hepatitis	29	24	ALB, ALP, AMY, CHE, CL, CRE, D-BIL, F-A/G, F-ALB, F-CHO, G_GL, G-GTP, GOT, GPT, I-BIL, LAP, LDH, T-BIL, T-CHO, TP, TTT, UA, UN, ZTT	0.158
Thrombosis	33	27	aCL IgG, ANA, aCL IgA, KCT, LAC, CPK, RVVT, IGM, RBC, HGB, HCT, PLT, PT, APTT, FG, A2PI, U-PRO, IGG, IGA, SC170, CRP, RNP, SM, SSA, SSB, CENTROMEA, CRE	0.179

The forecasted value derived using FIDES to impute the missing data. Table 4 shows the mean, trend and forecasted value calculated using the proposed FIDES method for the subset of data taken for Hepatitis patient (patient_medical identity (MID) =1, lab examination = ALB, year = 1981). The patient included in the subset totally had 181 observations. A comparison was done with Hanzak updated DES (2008), Wright updated DES (1986) and classical DES method (2004). The initial estimates for mean and trend $L_{t_0}$and $b_{t_0}$is computed using least square trend estimation. The positive value in the trend column of Table 4 shows an increase in the trend and negative value shows a decrease in the trend with respect to the patient date of examination. This increase or decrease in the trend indicates the short term and long term changes in the lab test.

The performance measures such as MSE (Mean Squared Error), MAD (Mean Absolute Deviation), error rate, MAPE (Mean Absolute Percentage Error) are used to compare the results of FIDES, Hanzak updated DES and classical Holt method (Holt, 1957; Holt, 2004; Cipra & Hanzak, 2008; Cooray, 2008).The obtained comparison results for Hepatitis and Thrombosis is shown in Table 5. It can be inferred from the Table 5 that the performance of FIDES was effective compared to the Hanzak updated DES and classical DES.

A 10-fold-cross-validation (Han & Kamber, 2001) has been used for evaluating the classification model. The performance measures are derived from the number of classifications from the training and validation accuracies.

Fig. 7 shows the MSE value for the various smoothing constant parameter values dynamically updated by FIDES for subset of records from Hepatitis Patient (Patient_ MID = 1, lab examination = ALB). From Fig. 7it can be inferred that the dynamic updation of 'α' and 'β' by FIDES reduces the MSE value. Also, it can be observed in the Fig. 7 that as the value of the smoothing constant parameters gets adjusted there is a reduction in the MSE value . Fig. 8 illustrates the MSE value observed for the fixed value of alpha and beta. The MSE value observed for fixed value of alpha and beta is higher compared to the MSE obtained through smoothing constant parameter values which is dynamically updated by FIDES.

The Wilcoxon rank sum test has been carried out with significant level (ρ-value < 0.05) to identify whether there is any significant reduction in the forecasting errors of FIDES compared to classical DES (Wilcoxon, 1945), Hanzak and Wright updated DES methods. It has been observed that the obtained -p-value for FIDES is less than 0.05 which shows that there is a significant reduction in the forecasting errors thereby leading to improvement in the forecasting accuracy rate. Fig. 9 represents the graphical comparison for the proposed FIDES over other versions of DES.

The forecasting error rate is reduced approximately by 2.5% on an average in the FIDES compared to classical DES, Wright and Hanzak updated DES methods. To check whether there is a significant improvement in the performance of presented FIDES based imputation technique over other imputation techniques such as mean, median imputation, K-nearest neighbor (KNN), hot-deck (HD), maximum likelihood (ML) this work performed a statistical paired t-test (Zimmerman, 1997) with significant value of 0.05. For Hepatitis data set, the p-value obtained for FIDES based imputation over mean, median, HD and ML was found to be less than 0.05. For Thrombosis data set, the p-value obtained for FIDES based imputation over mean, median, HD and ML was found to be less than 0.05. The p-value obtained for presented FIDES based imputa-

Table 7
Performance comparison on classification results of STRiD, NN, C4.5 and ID3.

Datasets	Classifiers	Performance measures					
		Accuracy	Sensitivity	Specificity	Precision	PLR	NLR
Hepatitis	STRiD	**91.50**	91.61	91.29	95.18	10.51	0.09
		(62.10)	(67.08)	(55.29)	(67.75)	(1.5)	(0.6)
	NN	72.33	76.96	65.2	77.33	2.21	0.35
		(61.82)	(66.58)	(55.05)	(67.75)	(1.48)	(0.61)
	ID3	68.16	73.78	57.79	76.32	1.75	0.45
		(57.78)	(65.62)	(46.26)	(64.22)	(1.22)	(0.74)
	SVM	70.17	77.36	58.89	74.72	1.88	0.38
		(60.81)	(66.92)	(52.68)	(65.27)	(1.41)	(0.63)
Thrombosis	STRiD	**90.65**	93.93	84.17	92.13	5.93	0.07
		(61.82)	(70.52)	(45.52)	(70.80)	(1.29)	(0.65)
	NN	70.25	79.89	49.38	77.39	1.58	0.41
		(60.78)	(69.88)	(45.04)	(68.75)	(1.27)	(0.67)
	ID3	66.62	78.21	43.36	73.49	1.38	0.5
		(56.75)	(66.53)	(39.05)	(66.40)	(1.09)	(0.86)
	SVM	71.04	80.61	50.41	77.80	1.63	0.38
		(58.57)	(68.74)	(41.46)	(66.40)	(1.17)	(0.75)

() value in parentheses represents the classification results without applying proposed imputation and attribute selection

tion was less than 0.05, so a reject in null hypothesis is considered which means FIDES based imputation over mean, median, HD and ML provides effective performance results.

For attribute selection using rough set, instead of equivalence classes this work generates temporal tolerance classes using similarity measures. Since this work considers similarity measure along with the time dimension, objects are distributed to more than one tolerance class. The temporal tolerance degree of dependency is calculated for each clinical attribute. The attribute with the highest degree of dependency is taken to be a first candidate in the selected attribute set. This is known as temporal first mean degree of dependency. Fig. 10 shows the plot for the temporal first mean degree of dependency for Hepatitis patients. Fig. 10 shows that the attribute GPT has highest degree of dependency. Hence, GPT is considered to be the first candidate in the selected dimension set. The subsets containing this attribute (GPT) in combination with the other attributes are generated in the next level. Temporal tolerance degree of dependency is calculated for the generated subsets, the subset with highest degree of dependency is identified and it is included as the second candidate in the selected attribute set. This process continues till all the possible attribute subset combinations with respect to the selected attribute set are extracted and processed.

Table 6 shows the results of attribute selection process. For Hepatitis patients from total attributes of 29, the temporal tolerance induced rough set forms most significant attribute set with 24 attributes. The identified 24 attributes are ALB, ALP, AMY, CHE, CL, CRE, D-BIL, F-A/G, F-ALB, F-CHO, G_GL, G-GTP, GOT, GPT, I-BIL, LAP, LDH, T-BIL, T-CHO, TP, TTT, UA, UN, ZTT is considered to be as relevant attribute and is used in the classification process.

For Thrombosis patients the temporal tolerance induced rough set forms the most significant attribute set with 27 attributes from the total attributes of 33. The identified 27 attributes are aCLIgG, ANA, aCL IgA, KCT, LAC, CPK, RVVT, IGM, RBC, HGB, HCT, PLT, PT, APTT, FG, A2PI, U-PRO, IGG, IGA, SC170, CRP, RNP, SM, SSA, SSB, CENTROMEA, CRE. The selected attributes which is identified as the most significant attributes is considered for classification. The reduct approximation error $\varepsilon(R)$ for Hepatitis Thrombosis dataset is computed using Eq. (20). For Hepatitis dataset reduct approximation error $\varepsilon(R_{Hepatitis})$ is 0.158 and for Thrombosis $\varepsilon(R_{Thrombosis})$ is 0.179.This error measure represents the significance of the attribute set selected with respect to the decision attribute.

Table 7, shows the comparisons for classification results obtained using STRiD, Neural Network (NN), ID3, Support Vector Machine (SVM) based on performance measures namely classification accuracy, true positive, true negative, false positive, false negative, sensitivity and specificity for Hepatitis and Thrombosis patients. It has been observed that the temporal tolerance rough set induction in the decision tree construction on an average has improved the performance of the classification system compared to the NN, ID3 system. In the Table 7, values in the parenthesis refer to the classification results obtained without applying proposed imputation and attribute selection. The classification results were evaluated with the following performance measures: accuracy, sensitivity, specificity, error rate, precision, Positive Likelihood Ratio (PLR), Negative Likelihood Ratio (NLR) (Han & Kamber, 2001). Fig. 11 and Fig. 12 illustrate the classification accuracy computed without applying imputation but with attribute selection for Hepatitis and Thrombosis dataset. Fig. 11 and Fig. 12 illustrate that there is a significant improvement in the classification accuracy, when the irrelevant features are removed before classification process.

Table 8
Comparative analysis based on area under curve (AUC) measures.

Dataset	Method	AUC
Hepatitis	STRiD	0.93 (0.74)
	NN	0.85 (0.66)
	ID3	0.71 (0.53)
	SVM	0.79 (0.61)
Thrombosis	STRiD	0.91 (0.77)
	NN	0.83 (0.61)
	ID3	0.69 (0.55)
	SVM	0.75 (0.59)

() value in parentheses represents the classification results without applying proposed imputation and attribute selection

Table 8 shows a comparative analysis done based on area under curve (AUC) measures. The Wilcoxon rank sum test and paired t-test with significant value of 0.05 is used to identify whether there was any significant improvement in the classification accuracy of STRiD compared with classical NN, ID3 and SVM methods. Wilcoxon and t-test compares the classifiers by examining the null hypothesis that the results of the two classifiers are equal (Demsar, 2006). Wilcoxon compares the mean ranks of the classifiers. For STRiD, the p-value of less than 0.05 is obtained for Hepatitis and Thrombosis dataset, which proves that the classification result of STRiD is improved compared to NN, ID3 and SVM classification methods. Since the distribution of clinical time series data

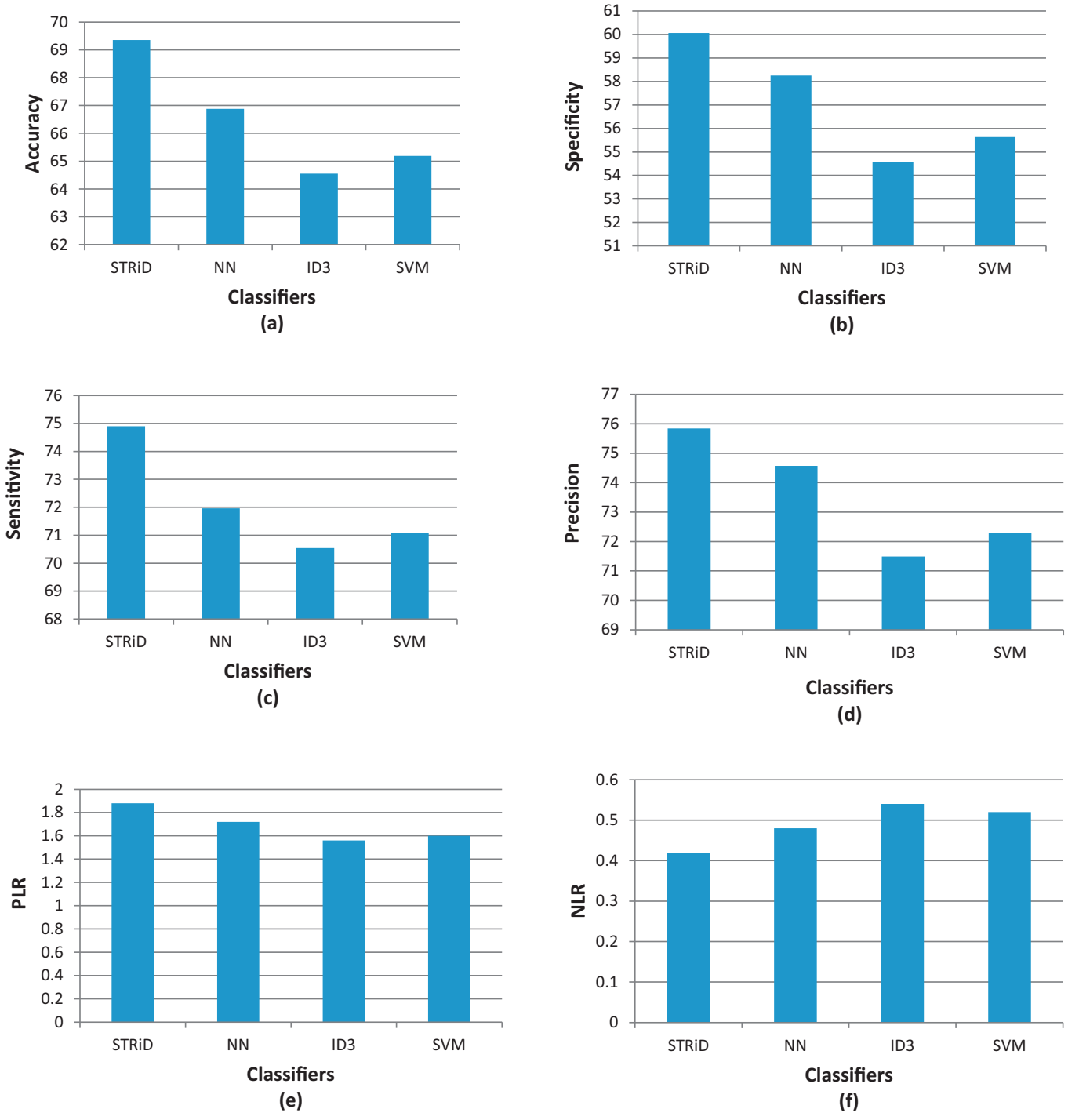

Fig. 12. Classification accuracy computed before without applying imputation but with attribute selection for Thrombosis dataset (a) Accuracy (b) Specificity (c) Precision (d) Sensitivity (e) PLR (f) NLR.

is near normal there is no difference between the Wilcoxon rank sum test and Paired t-test. Bergmann and Hommel dynamic correction (adjustment) of p-values presented by Garcia and Herrera (2008) is used for comparing the classifiers. The adjusted p-values (APV) are computed for the classifiers STRiD, NN, ID3 and SVM based on their classification accuracies.It can be inferred that the APV computer for the proposed STRiD framework shows a significant improvement in the classification accuracies compared to NN, ID3 and SVM. Table 8 shows a comparative analysis done based on area under curve (AUC) measures.

5. Conclusion and scope for future works

Clinical time series data are susceptible to complexities such as data incompletion, irregularities and large attribute sets. This paper presents a bio-statistical mining approach that effectively handles these complexities and builds a classification model for unevenly spaced clinical time series data. The proposed approach contributes work in three stages of knowledge discovery namely temporal preprocessing, attribute selection and classification. A FIDES forecasting method is proposed to handle the data irregularities, impute missing values and extract temporal patterns. FIDES uses fuzzy

inference in DES and updates the smoothing constants value for mean and trend estimation dynamically. From the experimental results, it is inferred that FIDES decreases the error rate on an average to 2.5% compared to that of other improved versions of DES method. Temporal pattern obtained for each clinical attribute is further used in attribute selection and classification. The presented temporal interpreted tolerance rough set algorithm performs attribute selection and identifies the significance value for each attribute using temporal tolerance degree of dependency. The presented STRiD framework performs decision tree induction by using temporal pattern in gain ratio computation. The observed mining results shows that the proposed STRiD achieves the classification accuracy on an average by 91.5% for hepatitis and 90.65% for thrombosis.

There are many interesting aspects for future research. This work can be applied to other unevenly spaced time series data with few minor domain specific changes. Since clinical time series data is often considered irregular, extracting temporal pattern from the clinical variables is a challenging task. Research studies can be carried out to efficiently handle these temporal complexities in the clinical data thereby improving the classification accuracy.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.eswa.2017.01.056.

References

Adlassnig, K. P., Combi, C., Das, A. K., Keravnou, E. T., & Pozzi, G. (2006). Temporal representation and reasoning in medicine: Research directions and challenges. *Artificial Intelligence in Medicine, 38*(2), 101–113.

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM, 26*(11), 832–843.

Augusto, J. C. (2005). Temporal reasoning for decision support in Medicine. *Artificial Intelligence in Medicine, 33*(1), 1–24.

Batal, I., Sacchi, L., Bellazzi, R., & Hauskrecht, M. (2009). Multivariate time series classification with temporal abstractions. In *Artificial intelligence research society conference* (pp. 344–349). AAAI Press.

Bouzid, M., Combi, C., Fisher, M. D., & Ligozat, G. (2006). Guest editorial: Temporal representation and reasoning. *Annals of Mathematics and Artificial Intelligence, 46*(3), 231–234.

Chittaro, L., & Montanari, A. (2000). Temporal representation and reasoning in artificial intelligence: Issues and approaches. *Annals of Mathematics and Artificial Intelligence, 28*, 47–106.

Cipra, T., & Hanzak, T. (2008). Exponential smoothing for irregular time series. *Kybernetika, 44*, 385–399.

Combi, C., & Shahar, Y. (1997). Temporal reasoning and temporal data maintenance in medicine: Issues and challenges. *Computers in Biology and Medicine, 27*(5), 353–368.

Cooray, T. M. J. A. (2008). *Applied Time Series: Analysis and Forecasting*. UK: Alpha Science International Limited Oxford.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis: An International Journal, 1*(3), 131–156.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B, 39*, 1–38.

Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7*, 1–30.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., & Keogh, E. (2008). Querying and mining of time series data: Experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment, 1*(2), 1542–1552.

Enders, C. (2010). *Applied missing data analysis*. New York: Guilford.

Garcia, S., & Herrera, F. (2008). An extension on statistical comparisons of classifiers over multiple data sets for all pairwise comparisons. *Journal of Machine Learning Research, 9*, 2677–2694.

Gorecki, T., & Luczak, M. (2013). Using derivatives in time series classification. *Data Mining and Knowledge Discovery, 26*(2), 310–331.

Han, J., & Kamber, M. (2001). *Data Mining, Concepts and Techniques* (2nd edition). San Francisco, CA: Morgan Kaufmann.

Hepatitis Dataset for Discovery Challenge. European conference on machine learning and principles and practice of knowledge discovery in databases, 16th ECML+9th PKDD Conference, October 3-10, Porto, Portugal, http://lisp.vse.cz/challenge/ecmlpkdd2005/ (Accessed: 1 Jan 2006).

Holt, C. C. (1957). *Forecasting trends and seasonal by exponentially weighted averages*. Pittsburgh: Carnegie Institute of Technology ONR memorandum no. 52.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting, 20*, 5–10.

Horvath, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. New York: Springer.

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Springer Science & Business Media.

Isola, R., Carvalho, R., & Tripathy, A. K. (2012). Knowledge discovery in medical systems using differential diagnosis, LAMSTAR and K-NN. *IEEE Transactions on Information Technology in Biomedicine, 16*(6), 1287–1295.

Jane, Y. N., Nehemiah, H. K., & Arputharaj, K. (2016). A temporal mining framework for classifying un-evenly spaced clinical data. *Applied Clinical Informatics, 17*(1), 1–21.

Jane, Y. N., Nehemiah, H. K., & Arputharaj, K. (2016). A Q-backpropagated time delay neural network for diagnosing severity of gait disturbances in Parkinson's disease. In *Journal of Biomedical Informatics: 60* (pp. 169–176). Elsevier.

Jang, J. S. R., Sun, C. T., & Mizutani, E. (1996). *Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence*. Prentice-Hall, Inc..

Jensen, R., & Shen, Qiang (2007). Tolerance-based and fuzzy-rough feature selection. *Fuzzy Systems Conference*, 1–6.

Jensen, S. (2001). *Mining medical data for predictive and sequential patterns discovery challenge on thrombosis data* Freiburg.

Kahn, M. G. (1991a). Combining physiologic models and symbolic methods to interpret time-varying patient data. *Methods of Information in Medicine, 30*, 167–178.

Kahn, M. G. (1991b). Extensions to the time-oriented database model to support temporal reasoning in medical expert systems. *Methods of Information in Medicine, 30*, 4–14.

Keravnou, E., & Shahar, Y. (2005). Temporal reasoning in medicine. In M. D. Fisher, D. M. Gabbay, & L. Vila (Eds.), *handbook of temporal reasoning in artificical intelligence* (pp. 587–653). Amsterdam. The Netherlands: Elsevier.

Komorowski, J., Pawlak, Z., Polkowski, L., & Skowron (2002). *Rough sets perspective on data and knowledge, in handbook of data mining and knowledge discovery* (pp. 134–149). Newyork: Oxford University press.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd edition). NewYork: Wiley.

Minas, A., Karaolis, J., Moutiris, A., Hadjipanayi, D., & Pattichis, C. S. (2010). Assessment of the risk factors of coronary heart events based on data mining with decision trees. *IEEE Transactions on Information Technology in Biomedicine, 14*, 559–566.

Moskovitch, R., & Shahar, Y. (2015a). Classification of multivariate time series via temporal abstraction and time intervals mining. *Knowledge and Information Systems, 45*(1), 35–74.

Moskovitch, R., & Shahar, Y. (2015b). Classification-driven temporal discretization of multivariate time series. *Knowledge and Information Systems, 29*, 871–913.

Moskovitch, R., & Shahar, Y. (2015c). Fast time intervals mining using the transitivity of temporal relations. *Knowledge and Information Systems, 42*(1), 21–48.

Nahato, K. B., Nehemiah, H. K., & Kannan, A. (2015). Knowledge mining from clinical datasets using rough sets and backpropagation neural network. *Computational and Mathematical Methods in Medicine*.

Nehemiah, H. K., Kannan, A., Vijaya, K., Jane, Y. N., & Merin, J. B. (2007). Intelligent rule mining for decision making from clinical data sets. *ICGST International Journal on Bioinformatics and Medical Engineering, 7*(1), 37–45.

Zhong, Ning, & Andrzej Skowron (2001). A rough set-based knowledge discovery process. *International Journal of Applied Mathematics Computer Science, 11*(3), 603–619.

Ohsaki, M., Sato, Y., Yokoi, H., & Yamaguchi, T. (2002). A rule discovery support system for sequential medical data, in the case study of a chronic hepatitis dataset, international workshop on active mining. In *IEEE International conference on data mining ICDM* (pp. 97–102).

Orphanou, K., Stassopoulou, A., & Keravnou, E. (2014). Temporal abstraction and temporal Bayesian networks in clinical domains: A survey. *Artificial Intelligence in Medicine, 60*(3), 133–149.

Patra, S., & Bruzzone, L. (2014). A novel SOM-SVM-based active learning technique for remote sensing image classification. *IEEE Transactions on Geoscience And Remote Sensing, 52*(11), 6899–6910.

Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Sciences, 11*, 341–356.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Los Altos: Morgan Kaufmann.

Tsaur, R. C. (2003). Forecasting by fuzzy double exponential smoothing model. *International journal of computer mathematics, 80*(11), 1351–1361.

Ramsay, J., & Silverman, B. (2005). *Functional data analysis*. New York: Springer.

Ruggieri, S. (2002). Efficient C4.5. *Knowledge and Data Engineering Transactions, 14*, 438–444.

Sacchi, L., Larizza, C., Combi, C., & Bellazzi, R. (2007). Data mining with temporal abstractions: Learning rules from time series. In *Data Mining and Knowledge Discovery: 15* (pp. 217–247). Springer.

Shahar, Y., & Musen, M. A. (1993). RESUME: A temporal-abstraction system for patient monitoring. *Computers and Biomedical Research, 26*(3), 255–273.

Shahar, Y. (1997). A framework for knowledge-based temporal abstraction. *Artificial Intelligence, 90*, 79–133.

Ismail, S., Shabri, A., & Samsudin, R. (2011). A hybrid model of self-organizing maps (SOM) and least square support vector machine (LSSVM) for time-series forecasting. *Expert Systems with Applications*, 10574–10578.

Kamali, T., Boostani, R., & Parsaei, H. (2014). A multi-classifier approach to MUAP classification for diagnosis of neuromuscular disorders. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22*(1), 191–200.

Thrombosis Dataset for Discovery Challenge (1999). *3rd European conference on principles and practice of knowledge discovery in databases, September 15-18* http://lisp.vse.cz/pkdd99/ (Accessed: 1 Jan 2006).

Tsumoto, S. (1999). In P. Berka (Ed.), *Guide to the medical data set* (pp. 45–47). Prague: PKDD 99 Workshop Notes on Discovery Challenge.

Tu Bao, H., Trong Dung, N., Saori, K., Si Quang, L., DungDuc, N., Hideto, Y., & Katsuhiko, T. (2003). Mining hepatitis data with temporal abstraction. In *KDD '03 Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 369–377).

Verduijn, M., Sacchi, Peeka, L., Bellazzi, N., Evert de Jonge, R., Bas, & de Mol, A. J. M. (2007). Temporal abstraction for feature extraction: A comparative case study in prediction from intensive care monitoring data. *Artificial Intelligence in Medicine, 41*, 1–12.

Vijaya, K., Nehemiah, H. K., Kannan, A., & BhuvaneswariAmma, N. G. (2010). Fuzzy neuro genetic approach for predicting the risk of cardiovascular diseases. *International Journal of Data Mining, Modelling and Management, 2*(4), 388–402.

Wilcoxon, F. (1945). Individual comparisons by ranking method. *Biometrics, 1*, 80–83.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 324–342.

Wright, D. J. (1986). Forecasting data published at irregular time intervals using extension of Holt's method. *Management Science, 32*, 499–510.

Xinmeng, Z., & Shengyi, J. (2012). A splitting criteria based on similarity in decision tree learning. *Journal of Software, 7*, 1775–1782.

Zimmerman, & Donald, W. (1997). A note on interpretation of the paired-samples t test. *Journal of Educational and Behavioral Statistics, 22*(3), 349–360.

Zytkow, J., & Gupta, S. (2001). In P. Berka (Ed.), *Guide to medical data on collagen disease and thrombosis*. Freiburg: PKDD 2001 Discovery Challenge on Thrombosis Data.