A Temporal Mining Framework for Classifying Un-Evenly Spaced Clinical Data

An Approach for Building Effective Clinical Decision-Making System

Nancy Yesudhas Jane[1]; Khanna Harichandran Nehemiah[1]; Kannan Arputharaj[2]

[1]Ramanujan Computing Centre, Anna University, Chennai, India;
[2]Department of Information Science and Technology, Anna University, Chennai, India

Keywords
Clinical time-series data, temporal mining, temporal data acquisition, rough set, neuro-fuzzy

Summary
Background: Clinical time-series data acquired from electronic health records (EHR) are liable to temporal complexities such as irregular observations, missing values and time constrained attributes that make the knowledge discovery process challenging.
Objective: This paper presents a temporal rough set induced neuro-fuzzy (TRiNF) mining framework that handles these complexities and builds an effective clinical decision-making system. TRiNF provides two functionalities namely temporal data acquisition (TDA) and temporal classification.
Method: In TDA, a time-series forecasting model is constructed by adopting an improved double exponential smoothing method. The forecasting model is used in missing value imputation and temporal pattern extraction. The relevant attributes are selected using a temporal pattern based rough set approach. In temporal classification, a classification model is built with the selected attributes using a temporal pattern induced neuro-fuzzy classifier.
Result: For experimentation, this work uses two clinical time series dataset of hepatitis and thrombosis patients. The experimental result shows that with the proposed TRiNF framework, there is a significant reduction in the error rate, thereby obtaining the classification accuracy on an average of 92.59% for hepatitis and 91.69% for thrombosis dataset.
Conclusion: The obtained classification results prove the efficiency of the proposed framework in terms of its improved classification accuracy.

Correspondence to:
Khanna Nehemiah H
Associate Professor
Ramanujan Computing Centre
Anna University
Chennai-600025, India
E-mail: nehemiah@annauniv.edu
Phone No: 91 044 22358013

Appl Clin Inform 2016; 7: 1–21
http://dx.doi.org/10.4338/ACI-2015-08-RA-0102
received: August 19, 2015
accepted: November 8, 2015
published: January 13, 2016
Citation: Jane Nancy Yesudhas, Nehemiah Khanna Harichandran, Arputharaj Kannan. A temporal mining framework for classifying un-evenly spaced clinical data: An approach for building effective clinical decision-making system. Appl Clin Inform 2016; 7: 1–21
http://dx.doi.org/10.4338/ACI-2015-08-RA-0102

Introduction

In the healthcare sector, due to the advancement of medical equipment, the state of patient health is monitored periodically and the results of the laboratory test are captured and maintained as electronic health records (EHR). The clinical time series data that have been obtained from these EHR stores enormous medical knowledge. This medical knowledge describes the temporal relationships among the various clinical observations. There are two ways of extracting medical knowledge: First, from the medical expert and Second, through knowledge discovery methods. The extracted medical knowledge is used to construct clinical decision-support systems to assist clinical activities such as diagnosis, monitoring, prognosis and drug discovery [1–3]. Though clinical data contain useful medical knowledge, they are also liable to temporal complexities such as irregular observations, missing values and time constrained attributes. Clinical time series data is irregular, since the observation of these data does not happen in regular (equal) interval of time and the number of observations done may vary for each patient.

The importance of managing time-oriented concept and knowledge discovery in medicine was investigated in many research studies [4–8]. Bellazzi and Zupan [2] have presented a detailed review about the usage and challenges of predictive data mining in the medical domain. A detailed study about the merits and demerits of various classification methods discussed provides the guidelines required for carrying research studies in clinical data mining. Although there are many existing research works carried out in temporal abstraction, reasoning and mining with clinical data, the presence of temporal complexities in clinical time series data challenges the effectiveness of the knowledge discovery process. The aim of this paper is to build an effective classification model for unevenly spaced clinical time series data. This classification model is used in constructing a clinical decision-making system to classify the stages of the diseases, which helps the physician in decision-making task.

1.1 Outline of the paper

This paper presents a temporal rough set induced neuro-fuzzy (TRiNF) framework that handles the temporal complexities and builds an effective classification model. TRiNF consists of two functionalities namely temporal data acquisition (TDA) and temporal classification. TDA process aims at pre-processing the temporal complexities in clinical time series data. An enhanced double exponential smoothing (DES) method presented by Wright [9] is adopted for constructing a time-series forecasting model. Missing value imputation and temporal pattern extraction are done using the forecasting model. Temporal pattern refers to the trend and state of the clinical attribute. The relevant attributes are selected for classification using a temporal pattern based rough set approach. In the temporal classification process, an effective classification model is built using a temporal pattern induced neuro-fuzzy classifier. The fuzzy sets for the classifier are defined using the trend pattern of each clinical attribute. Experimental results show that the proposed system overcomes the temporal complexities and improves the classification accuracy.

The rest of the paper is organized as follows. Related works are discussed in Section 2. In Section 3 Materials and Methods used in the proposed approach are discussed. Experimental results and discussions are presented in Section 4. Conclusion and scope for future works are presented in Section 5.

2. Related Work

This section reviews the works carried out by the researchers in the field of temporal abstraction, time series classification and handling irregularly sampled data.

2.1 Temporal abstraction

Temporal abstraction aims to transform low level quantitative descriptions of time series data to high level qualitative descriptions. These qualitative descriptions provide a summarized interpre-

tation of all time-stamped variables. Shahar [10] describes the temporal abstraction in clinical domain as the process of interpreting the clinical temporal parameters and events in time stamped data as states and trends. The task of temporal abstraction plays a vital role in temporal reasoning that aids the process of time series data analysis. Combi and Shahar [11] have presented a study about temporal reasoning and temporal data maintenance to develop time-oriented medical systems. Investigations into these two concepts summarize the challenging research works identified in collaborating temporal reasoning and temporal data maintenance. Adlassnig et al. [3] have presented a detailed discussion about relating temporal representation and reasoning to clinical tasks such as monitoring, treatment, etc. The authors examined several concepts related to temporal databases, handling uncertainties in clinical data and reasoning on temporal clinical data for mining. Stacey and Mcgregor [12] have presented a detailed survey on temporal abstraction based clinical data analysis. The authors have discussed several works that illustrate the development of temporal abstraction systems like RESUME [13], TRENDX [14], VIE-VENT [15], ECHO [16], RASTA [17] etc. Tu et al. [18] have proposed a temporal abstraction approach for extracting knowledge from hepatitis dataset collected from Chiba University hospital. The temporal abstraction process extracts the states and trends for each patient for a particular lab test within a specified episode. It is inferred that various machine-learning methods can be applied to the abstracted data in order to extract knowledge that can be used by physicians.

2.2 Time series classification

Time series classification is a task in temporal data mining, which aims at building a trained classification model for time stamped data. Classification of time series data is challenging as they are liable to several temporal intervals and abstracted interpretations in addition to the time stamped data points. In general a time series data can be categorized as univariate or multivariate based on the presence of single variable or multiple variables respectively.

Moskovitch et al. [19] presented a novel framework called KarmaLegoSification (KLS) for classifying multivariate time series data. It includes three processes, namely the symbolic representation of data points, identifying frequent temporal interval relation patterns (TRIP) and classification using the patterns generated as attributes. The experiments that were carried out using benchmark clinical datasets, prove the efficiency of the system in terms of its classification accuracy.

Batal et al. [20] have presented a temporal pattern mining technique named Minimal Predictive Temporal Patterns (MPTP), for performing classification of medical health records. MPTP algorithm combines the pattern selection and frequent pattern mining. . The health records of heparin induced thrombocytopenia (HPT) patients, were used for experimentation. This MPTP framework extracts useful features for classification, which is used in developing a clinical decision-making system. Moerchen [21] has presented an effective unsupervised algorithm to perform mining from the temporal concepts extracted by temporal language time series knowledge representation (TSKR) based on sequential pattern and itemset mining. The usage of TSKR in mining has overcome the limitations of Allens [22] interval relations and is demonstrated using a sport medicine dataset.

Bodyansky et al. [23] have presented a neuro-fuzzy network using the Kolmogorov's superposition theorem named neuro-fuzzy Kolmogorov's network (NFKN). The output layer in NFKN is trained using least square method (LSM) and the hidden layer is trained using the gradient descent method. The NFKN is highly suited for classification since it effectively handles the dimensionality complexity using a two level structure based on KST. However, the training process requires improvement in the convergence behavior and in extending the classification process to support multiple class labels. Petkovic et al. [24] have used an adaptive neuro-fuzzy inference system (ANFIS) network presented by Jang [25] to study the impact of autonomic nervous system (ANS) on the significant heart rate variability (HRV) parameters. For analysis, they have extracted 14 parameters of HRV signal. They have done a detailed investigation to identify the HRV parameters that are affected by the ANS functions. Two ECG datasets were used for the analysis, namely MIT Arrhythmia Database and epilepsy database. A comparative analysis between the ANFIS prediction method and linear regression model with respect to its regression error shows that the performance of ANFIS model is improved over the linear regression model.

Mcnameea et al. [26] presented a neuro-fuzzy inference system (NFIS) to simulate heart rate variations. They have developed a system to predict the changes in a patient's health conditions in the neurological intensive care unit (NICU). They have demonstrated the NFIS model with both observed and simulated data from NICU patients. The experimental results indicate that the NFIS is capable of effectively predicting the changes in heart rate. Khanna et al. [27] 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 store the rules in the knowledge base. For experimentation the authors have used two time series datasets of hepatitis and thrombosis patients [28, 29].

The literatures [30, 31] present a mining technique used for building classification model for clinical data. Vijaya et al. [30] have proposed a diagnosis system using fuzzy neuro-genetic approach for predicting the severity of the cardiovascular diseases. The fuzzified continuous input variables are fed as input to the neural network. Genetic algorithm is used to train the neural network. Nahato et al [31] have presented a rough set indiscernibility relation method with back-propagation neural network (RS-BPNN) classifier to extract knowledge from 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. From the experiment, it can be inferred that the RS-BPNN classifier achieves significant improvement in classification accuracy.

2.3 Handling irregularly sampled data

The irregular data refers 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 observed at unequal time intervals. Liu et al. [32] have presented a new hierarchical system framework that builds a temporal model for irregularly sampled time series data to support clinical decision making. The authors have presented algorithms to learn temporal models from the data. Moreover, these models accurately predict future values. The framework uses machine learning and data mining algorithms such as linear dynamical system (LDS) and Gaussian process (GP) [33, 34]. GP models irregular time series data that accurately predict future values [34]. GP makes observations as a function of time and there is no need to mark when the observations were made and whether they are regularly or irregularly spaced. LDS [33] defines a state-space process with linear transitions between two consecutive states taken at discrete time points. However, in most of the real world applications time series is not discrete. Hence, GP is used at lower levels over time windows for modeling irregular time series data. LDS then tracks the transition in the GP process.

Two methods are used for analysis of an irregularly sampled data; namely, direct value interpolation [35–39] and windows-based segmentation [40, 41]. The former assumes that all values are collected regularly with a pre-specified sampling frequency and converts time series with irregular observations to discrete time observation sequences. The later first segments time series to fixed-sized windows. From this summary statistic is calculated. Like LDS, Autoregressive model (AR) is a discrete time series model used to represent a stochastic process. Prediction is carried out by taking an initial sequence using AR or LDS model. The correctness of the system was proved using mean absolute prediction error and absolute percentage error. The framework was used over a univariate time series data. CBC (complete blood count) lab time series data was used and 3.13% average prediction accuracy improvement was achieved when it was compared against the best performing baseline (AR, LDS, GP and other window based segmentation method. The authors have concluded that their work has a limitation that it works only with univariate time series data and it was extended to support multivariate time series data.

Bahadori et al. [42] have presented a Generalized Lasso Granger (GLG) method that discovers the temporal dependencies from irregular time series data. The authors also have presented a review on various techniques used in analyzing irregular time series. The general methods available for analyzing irregular time series are namely the repair approach, Lomb-Scargle Periodogram(LSP), wavelets and Kernel methods [38, 39, 43, 44]. GLG uses kernel functions to simplify the inner prod-

uct for irregular time series. The authors have presented a theoretical analysis and simulated experiments with four synthetic datasets to prove the effectiveness of their proposed work. An application of the GLG method with the datasets of δ18O (a radio isotope of Oxygen) is provided to detect the moisture transfer patterns. GLG is likely to have lower absolute errors because it predicts the actual observations without additional repair error. GLG becomes more accurate when there is a decrease in probability of missing a data. However, the authors have concluded that GLG approach has limitations with respect to the scalability in data analysis.

Ceusters et al. [45] have presented a work that generates Instance Unique Identifier to hold descriptions of relevant facts and assumptions about a patient's medical condition, his treatment and risk factors. The referent tracking system proposed by the authors may be used as an aid to support any application of information systems at the point where EHRs and other existing clinical terminologies integrate together. When information systems dealing with temporal data have to be applied to cases in spatiotemporal reality with respect to patients, their disorders and the particular treatments, then the suggested referent tracking approach will certainly reduce the complexity of the system, in terms of handling ambiguities, inconsistency and noise in the data. Our work provides the core framework for handling clinical temporal data. If the entities in the temporal window are mapped to unique referents, then the computational cost of processing may increase, but the performance of the system from an user-oriented perspective will certainly increase in terms of semantic interoperability of computer systems, patient management, diagnosis and prognosis.

There has been many works in the literature that addresses the task of mining in time series data. However, these methodologies have restrictions to work with multivariate time series data observed at irregular intervals because they are either tuned to support regular time series data or irregular univariate time series data. Clinical observations are often irregular and multivariate. Hence, mining in such clinical time series data is a challenging area of research. Compared to the works discussed in the literature the proposed work is different in the following ways: First, this paper proposes the incorporation of time series forecasting model into the pre-processing of temporal complexities like irregularities, missing values and into the derivation of the temporal patterns such as trend and state for each clinical attribute. Since, clinical data are observed at irregular intervals, an enhanced DES method that supports forecasting in irregular time-series presented by Wright [9] was adopted. Second, the derived temporal patterns for each clinical attribute are used in the attribute selection and classification process instead of using the actual observed value. A temporal rough set induced attribute selection process is presented to identify the relevant attributes. Third, the fuzzification of inputs for the temporal pattern induced neuro-fuzzy classifier is done using membership functions derived from the temporal trend pattern of each clinical attribute.

3. Materials and Method

3.1 Dataset description

For experimentation, this work uses two time series clinical data sets of hepatitis and thrombosis patients. The datasets were released in Principles and Practice of Knowledge Discovery in Databases (PKDD) discovery challenge for a data mining contest [28, 29]. These datasets were collected from Chiba hospital, which contains clinical records stored from 1981 to 2001 and 1980 to 1999 respectively. These datasets were used in our previous work [27]. Currently access to the data sets is unavailable. ▶Table 1 shows the general dataset summary for the hepatitis and thrombosis patients. Hepatitis data set consists of 771 patient's laboratory test reports of Hepatitis B and C. Each patient has undergone 983 laboratory tests. It has to be noted that not all the laboratory tests taken are related to hepatitis. The expert guidance and the dataset descriptions given by Ohsaki [46] were considered and 29 suggested tests have been selected for experimentation with hepatitis dataset. The average missing value percentage in hepatitis dataset is 11. Thrombosis data set consist of laboratory test report pertaining to 1000 patients. Each patient has undergone 564 laboratory tests. The expert's knowledge and dataset descriptions given in [47-49] were considered and 33 suggested tests have been identified for experimentation with thrombosis dataset. The average missing value percentage in thrombosis dataset is 8.

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.

The suggested lab tests of hepatitis and thrombosis patients along with their date of examination is considered as temporal input attribute set for demonstrating the effectiveness of the proposed temporal mining framework.

3.2 Methods

The framework for the proposed system is shown in the ▶Figure 1. The major components of the system are temporal data acquisition and temporal classification. To demonstrate the work two clinical time-series datasets of hepatitis and thrombosis patients were used.

3.2.1 Temporal Data Acquisition (TDA)

In TDA process, the temporal complexities like irregular observations, missing values and time constrained attributes that occur in the clinical time series data are preprocessed. The clinical attribute corresponds to the laboratory test taken on each patient that exhibits temporal patterns namely trend and state. The trend is the overall growth rate of the attribute and is referred as increase, decrease and stable. The state represents the range of the attribute and is referred as low, high and normal.

3.2.1.1 Missing data imputation and temporal pattern extraction

The importance of imputing the missing values in time series and its approaches is discussed by Little et al. and Enders [50, 51]. The following are the few traditional techniques that are commonly used for imputing missing value, namely mean, median imputation, K-nearest neighbor (KNN), hot-deck, maximum likelihood etc., [51]. The proposed TRiNF framework uses a time-series forecasting model to handle the missing value and to extract the temporal patterns by adopting the mathematical framework of an improved DES method presented by Wright [9].

The adopted method builds a forecasting model by computing the growth rate and level for each clinical attribute over a period of observations. The temporal patterns (trend and state) for the clinical attributes are obtained from the computed growth rate and level value. The forecasted value of each clinical attributes observed value is calculated with their previous observed trend and state. In the proposed framework, for each clinical attribute, its missing value at a time-period is imputed using its corresponding forecasted value.

A brief description about the classical DES method presented by Holt [52] is discussed. The classical DES method computes the trend and level for each observation using two smoothing constant parameters α and β respectively. The value for α and β is chosen to be in the range of 0 and 1 and this value remains constant for all the computations. However, a wrong choice of this constant value affects the accuracy of the forecasting results. To overcome these limitations and to extend DES for irregular time series too, Wright [9] suggested that, instead of assigning a constant value to the smoothing constant parameter, dynamic assignment can be done based on the interval spacing's among the observations. Let $Y = \{Y_{t_n}(i), n \in T, i \in A; t_{n+1} > t_n\}$ be an irregular time series, where t_n is the observation time for a patient, T is the set of observation time points , A be attribute set, $Y_{t_n}(i)$ is the value of i^{th} attribute at time t_n, $M_{t_{n+1}}(i)$ is the level for i^{th} attribute at time t_{n+1}, $G_{t_{n+1}}(i)$ is the growth rate for i^{th} attribute at time t_{n+1}, $\alpha_{t_{n+1}}(i)$ and $\beta_{t_{n+1}}(i)$ is the smoothing constant for the level and trend of i^{th} attribute at time t_{n+1}, α_{t_0} and β_{t_0} is the initial smoothing constants for level and trend, M_{t_0} and G_{t_0} represents the initial value of level and trend. The values of M_{t_0} and G_{t_0} is initialized using least square estimation. Here, attribute refers to lab test taken by a patient.

To build a time-series forecasting model for clinical time series data, the following estimations were made using Wright enhanced DES mathematical model

(i) Smoothing constants: In this work for each clinical attribute the smoothing constants α_{t_0}, β_{t_0} $\alpha_{t_{n+1}}(i)$ and $\beta_{t_{n+1}}(i)$, are updated based on the interval days between each observation $(t_{n+1} - t_n)$ as defined in the equation (1), (2), (3) and (4) respectively.

$$\alpha_{t_0} = 1-(1-\alpha)^{\delta} \quad (1)$$

$\beta_{t_0} = 1-(1-\beta)^{\delta}$ (2)

where δ is the average interval spacing.

$\alpha_{t_{n+1}}(i) = \alpha_{t_n}(i) / \left(\alpha_{t_n}(i)+(1-\alpha)^{(t_{n+1}-t_n)}\right)$ (3)

$\beta_{t_{n+1}}(i) = \beta_{t_n}(i) / \left(\beta_{t_n}(i)+(1-\beta)^{(t_{n+1}-t_n)}\right)$ (4)

(ii)Level Estimate: This calculates the level value of an observed clinical attribute at a specified time say t_{n+1} using the $\alpha_{t_{n+1}}(i)$.

$M_{t_{n+1}}(i) = \alpha_{t_{n+1}}(i) Y_{t_{n+1}}(i) + (1-\alpha_{t_{n+1}}(i))\,(M_{t_n}(i) + (t_{n+1} - t_n) G_{t_n}(i))$ (5)

(iii)Growth Rate Estimate: This calculates the growth rate value of an observed clinical attribute at a specified time say t_{n+1} using the $\beta_{t_{n+1}}(i)$.

$G_{t_{n+1}}(i) = \beta_{t_{n+1}}(i)\,(M_{t_{n+1}}(i)) - M_{t_n}(i)/(t_{n+1} - t_n) + (1-\beta_{t_{n+1}})\,G_{t_n}(i)$ (6)

(iv)Forecasted Value: The forecasted value of an observed clinical attribute at a specified time is calculated from its previous observed level and trend.

$F_{t_{n+2}}(i) = M_{t_{n+1}}(i) + G_{t_{n+1}}(i)$ (7)

The algorithm temporal_preprocessing summarizes the steps followed in missing value imputation and temporal pattern extraction.

Algorithm 1: temporal_preprocessing (Y, α, β, L, N, M_{t_0}, G_{t_0})

Input:Y be an irregular time series, α and β are Smoothing Constants, L is the number of lab test, N is the number of observations, M_{t_0} and G_{t_0} is initial estimate of trend and growth rate.

Output: Level set (M), Growth rate (G), Forecasted Value (F)

1. Initialize the level and trend Smoothing Constants α_{t_0} and β_{t_0} using equation (1) and (2).
2. Initialize the M_{t_0} and G_{t_0} using least square estimation.
3. For i= 1 to L do
4. For j= 1 to N do
5. If $Y_{t_j}(i)$ is missing then
6. $Y_{t_j}(i) = F_{t_j}(i)$ //missing value imputed for i^{th} lab test at time t_j
7. End
8. Compute $\alpha_{t_j}(i)$ and $\beta_{t_j}(i)$ using equation (3) and (4).
9. Compute $M_{t_j}(i)$, $G_{t_j}(i)$, $F_{t_{j+1}}(i)$ using equation (5), (6) and (7).
10. End for
11. Return M_{t_j}, G_{t_j}, $F_{t_{j+1}}$
12. End for

▶Figure 2 shows a worked example for illustrating the steps carried out in temporal_preprocessing algorithm with few samples of data taken from hepatitis dataset. The Exam_date column shows the date of observation for lab test T-BIL taken for the patient whose medical identity (MID) is 1. For T-BIL examination the 147 th observation was taken on 17/12/97 and was assumed missing. TDA process effectively imputes it using the forecasted value derived from its previous 146 th observation growth rate and level value calculated using the adopted improved double exponential method. The temporal patterns trend (T) and state (S) for each clinical attribute is obtained from the growth rate estimate $G_{t_{n+1}}$ and level $(M_{t_{n+1}})$ in TDA process. Before using these patterns in the temporal attribute selection and classification a min-max normalization [53] is used to normalize the trend value in the scale of (-1 to 1).

3.2.1.2 Temporal attribute selection

Clinical time series data are susceptible to a high dimensional set of attributes which represents the lab test taken on each patient.

Since lab test reports taken for most of the patients include common tests which may not be relevant to diagnosis of a particular disease, it is often not necessary to include all the attributes for classification. Therefore, in this paper, identifying and eliminating such irrelevant attributes is considered as a pre-requisite before the classification process. Rough set is a mathematical concept pro-

posed by Pawlak [54] and is widely used in attribute selection. Dash et al. [55] presented a detailed study about various attribute selection techniques proposed for classification tasks.

The importance of using rough set in attribute selection as pre-processing in knowledge discovery has been illustrated in [56, 57]. Chouchoulas et al. [58] have investigated the application of rough set concept in dimensionality reduction. They have illustrated the popular quick reduct algorithm, which identifies the minimal reduct using degree of dependency among the attributes. Pradipta et al. [59] have proposed an optimized way to perform attribute selection based on fuzzy–rough sets by concurrently selecting and extracting the attributes using the perception of its significance. Rough set performs attribute selection with the information extracted from the attributes in the data and there is no need for providing additional information or any domain expert knowledge for the attribute selection process. This is the major advantage of using rough sets in attribute selection process. However, attribute selection in a time series data is a challenging task since the data exhibits temporal pattern (trend) and state that changes over time. In addition, if the time series data is observed at irregular intervals the complexity of applying the traditional attribute selection algorithms increase. So to handle these complexities, this paper presents a temporal pattern based rough set concept for identifying relevant attribute set from irregular time series data. This work performs attribute selection by incorporating the temporal patterns (trend and state) obtained for each clinical attribute in the rough set concepts.

The concept of rough set is described using information system and topological operations known as approximations. An information system is a form of data representation that is utilized by rough set for defining topological operations. In rough sets, an information system is represented as $I=(\mathbb{U},A)$, where $=\{x_1, \ldots, x_i, \ldots x_j \ldots, x_n\}$ is called as an universe which is a nonempty set of finite objects and $A=\{ a_1,a_2. \ldots, a_m \}$ is the knowledge in Universe which is the non-empty finite set of attributes [54]. For clinical time-series data an object in the universe refers to a patient and knowledge refers to the attributes (lab test) of a patient. From the information system rough set generates indiscernibility relation which is defined as the relation between two or more objects with respect to the subset of attributes.

The notations used in this work by the rough sets are described. Let $\mathbb{U}=\{x_1, \ldots, x_i, \ldots x_j \ldots, x_n\}$ is the universe; where x_i denotes the i^{th} object in the universe, $A=\{ a_1, \ldots, a_k, \ldots, a_m\}$ is the Knowledge(attribute set); where a_k denotes the k^{th} attribute in attribute set $a_k \in A$, X is the subset of universe $X \subseteq \mathbb{U}$, B is the subset of knowledge A, $B \subseteq A$, $a_{k\langle(T,S)\rangle}(x_i)$ denotes the trend (T) and state (S) of x_i for the attribute ' a_k '; Q is the decision attribute, $(B_{(T,S)})$ indiscernible relation or equivalence class, $\underline{B}_{(T,S)}X$ temporal lower approximation, $(POS_{B_{(T,S)}}(Q))$ temporal positive region, K temporal degree of dependency, $\varepsilon_{(C,Q)}(R)$ reduct approximation error , C is the condition attribute set, R is the reduced attribute set.

To perform temporal attribute selection this paper defines the temporal pattern based rough equivalence class $(B_{(T,S)})$, temporal lower approximation $\underline{B}_{(T,S)}X$, temporal B-positive region $(POS_{B_{(T,S)}}(Q))$, temporal degree of dependency (K) and reduct approximation error $\varepsilon_{(C,Q)}(R)$ using the equations (8) to (12) respectively.

(i) Temporal pattern based rough equivalence denoted as $(B_{(T,S)})$ partitions the universe which represents an elementary portion of temporal knowledge that can be extracted. This is generated based on the equivalence among objects temporal patterns in the universe $\mathbb{U}$. It is also denoted as $[X]\, B_{(T,S)}$ where $X \subseteq \mathbb{U}$,

$$\text{tempIND}\,(B_{(T,S)}) = \left\{(x_i, x_j) \in \mathbb{U} \,|\, \forall a \in B_{(T,S)}, a_{\langle(T,S)\rangle}(x_i) = a_{\langle((T,S))\rangle}(x_j)\right\} \quad (8)$$

(ii) Temporal lower approximation of X denoted as $\underline{B}_{(T,S)}X$ contains all elements that surely belong to the set X.

$$\underline{B}_{(T,S)}X = \left\{x \,|\, [x]B_{(T,S)} \subseteq X\right\} \quad (9)$$

(iii) Temporal B-positive region of X $(POS_{B_{(T,S)}}(Q))$ contains all the objects of '$\mathbb{U}$' that can be classified to equivalence classes of '$\mathbb{U}/Q$' such that using the information in the attributes 'B'.

$$POS_{B_{(T,S)}}(Q) = \cup_{x \in \frac{\mathbb{U}}{Q}} \underline{B}_{(T,S)} \quad (10)$$

(iv) Temporal degree of dependency (K) or $(\gamma_{B_{(T,S)}}(Q))$ is measure that is used to identify the dependencies in the attributes . It is computed using equation (11).

$$K = \gamma_{B_{(T,S)}}(Q) = \frac{|POS_{B_{(T,S)}}(Q)|}{|\mathbb{U}|} \quad (11)$$

(v) Reduct approximation error $\varepsilon_{(C,Q)}(R)$ is a measure that represents how a reduced attribute (R) approximates the condition attribute set (C) in association to the decision attribute (Q). It is calculated using equation (12) adopted from [56].

$$\varepsilon_{(C,Q)}(R) = 1-(\gamma_{R,\tau}(Q)/\gamma_{C,\tau}(Q)) \quad (12)$$

Where C is the condition attribute, Q is the decision attribute, R is the reduced attribute set $R \subseteq C$. If $\varepsilon_{(C,Q)}(R) = 0$, then R is reduct of C. The minimal level of reduct approximation error is proved to increase the accuracy of the classification process [56].

A temporal rough attribute selection procedure is presented in this work for selecting relevant attributes, which is an extension of quick reduct algorithm [57]. In quick reduct, attributes observed values are used in forming equivalence class, lower approximation, positive region and degree of dependency whereas in the presented approach temporal patterns derived for each attributes are used in these computations. Before starting the attribute selection process the normalized trend value is grouped under three categories. The positive value (>0) in the trend shows an increase in the trend denoted as "I" and negative value (<0) shows a decrease in the trend denoted as "D", zero value indicates that it is stable with respect to the patient date of test denoted as "S". This trend representation is considered to be an important factor in identifying the short term and long term changes in the lab tests. The state represents the levelled (or mean) value for the lab test at a particular observed point. The normal range and descriptions for each clinical lab test (attribute) is obtained from panel of experts in clinical domain. Based on these, we have discretized the state into "Low" (L), "Normal" (N) and "High" (H). The following algorithm 2 illustrates the steps used in selecting relevant attributes.

Algorithm 2: temporal_Rough_attributeSelection

Input: Clinical attribute set A={ a_1,a_2.....a_m}, Trend (T) and State (S) for the corresponding clinical attributes.

Output: Reduced attribute set R.

1. Generate temporal tolerance class for the attributes in {A} using equation (8).
2. Determine lower approximation, positive region based on temporal patterns using equation (9) and (10).
3. Compute significance of each attribute using temporal degree of dependency using equation (11).
4. Select the significant attribute a∈ A with high degree of dependency, include it in reduced attribute set (R).
5. Remove the attribute 'a' from A.
6. If there are attributes from (A) to form a subset with attributes in (R) then
7. Generate a superfluous set (SS) which contains candidate attribute subset of A & R
8. Repeat step 2–6 for every attribute subset in SS.
9. End if
10. Compute the reduct approximation error using equation (12).
11. Return reduced attribute set (R) and its reduct approximation error.

▶Table 2 shows the subset of normalized and discretized temporal pattern derived from TDA process for the lab examination (T-BIL, GPT) of five patients selected at random. To add a clear explanation to the proposed algorithm a worked example with a subset of Hepatitis data is provided.

The temporal classes generated for the attributes T-BIL, GPT and decision attribute Hepatitis are as follows,

$\mathbb{U}/IND\,(T-BIL_{(T,S)}) = \{\{1, 2\}, \{3\}, \{4\}, \{5\}\}$,

$\mathbb{U}/IND\,(GPT_{(T,S)}) = \{\{1, 2, 4\}, \{3\}, \{5\}\}$,

$\mathbb{U}/IND\,(Hepatitis) = \{\{1, 2, 5\}, \{3, 4\}\}$

The lower approximations for the decision attribute Hepatitis based on derived temporal patterns of lab examination T-BIL and GPT is calculated as follows,

$T-BIL_{(T,S)}\{1, 2, 5\} = \{1, 2\}$, $T-BIL_{(T,S)}\{3, 4\} = \{\}$

$GPT_{(T,S)}\{1, 2,5\} = \{1, 2\}$, $GPT_{(T,S)}\{3,4\} = \{\}$

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

$POS_{T-BIL_{(T,S)}}(Hepatitis) = \cup_{x\in\mathbb{U}/Hepatitis} T-BIL_{(T,S)}X = T-BIL_{(T,S)}\{1, 2, 5\} \cup T-BIL_{(T,S)}\{3, 4\} = \{1,2\}$

$$POS_{GPT_{(T_{pn},S_{pn})}}(\text{Hepatitis}) = \cup_{x \in \mathbb{U}/\text{Hepatitis}} T - CHO_{(T_{pn},S_{pn})}X$$

[Formel:]= $GPT_{(T,S)}\{1, 2,5\} \cup GPT_{(T,S)}\{3,4\} = \{1, 2\}$

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 T-BIL and GPT is computed as follows,

$K = POS_{T-BIL_{(T,S)}}(\text{Hepatitis})/|\mathbb{U}| = \{1, 2\} / \{1, 2, 3, 4, 5\} = 2/5$

$K = POS_{GPT_{(T,S)}}(\text{Hepatitis})/|\mathbb{U}| = \{1, 2\} / \{1, 2, 3, 4, 5\} = 2/5$

The attributes with highest dependency is selected and subset of these attributes is 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 are no attributes left to form new subsets. The significant reduced set returns the identified relevant attributes.

3.2.2 Temporal classification

The selected attributes and its temporal trend pattern obtained from the TDA process is used in the temporal classification process. A temporal pattern induced neuro-fuzzy classifier which adopts a five-layer feed forward back propagated neuro-fuzzy network structure [25] is used to build a temporal classification model. The sugeno-type ANFIS network model is considered. ▶Figure 3 shows the neuro-fuzzy network structure and the fuzzy membership graph for the trend pattern of clinical attributes derived using guassian functions. Each input node in the layer 1 corresponds to the selected clinical attribute (lab test). The nodes in layers 2, 3 and 5 represent nodes that are used for propagating and fixing the firing strength of the rule, whereas the nodes in layer 1 and layer 2 have parameters to be learnt. The network is trained using back propagation learning with Levenberg-Marquardt optimization [25, 53].

The fuzzy rules are generated by partitioning the input space using CART algorithm [60]. Fuzzy membership values are defined for each input clinical attribute using the temporal trend pattern obtained for each clinical lab test in TDA process. Let $V=\{v_1,v_2,\ldots..v_n\}$ represents the set of input nodes in the network; where n is the number of input nodes, FS= {"Increase", "Decrease", "Stable"} represents the fuzzy set. The fuzzy set is defined by membership function represented as $\mu\langle\text{increase}\rangle^{(v_i)}$, $\mu\langle\text{Decrease}\rangle^{(v_i)}$, $\mu\langle\text{Stable}\rangle^{(v_i)}$ where $v_i \in V$, i=1,2..n. For example, if the input node representing for lab test GPT shows a gradual increase, the trend identified in TDA process is denoted by positive value over a period of observation. Similarly, a negative value of trend denotes a decrease and zero denotes that it is stable. Categorizing these transitions in trend as increase (I), decrease (D) or stable (S) is considered to be a trend pattern for the attribute GPT. Fuzzification layer derives the membership value for each attributes trend pattern. The fuzzy rule layer forms the antecedent part of the fuzzy rule and the firing strength of each rule is computed using T-Norm operation [25]. The defuzzification layer uses a least square method for mapping the antecedent with the appropriate consequent.The output is computed by taking summation.

4. Experiment Results and Discussions

The work proposed was initiated with an experiment by applying the TDA and classification (TRiNF) method to the hepatitis and thrombosis datasets. This section provides a detailed discussion about the experimental results and observations. The raw data was randomly divided into two sets train and test which contains 75% and 25% of patient's respectively. In TDA, for each clinical attribute its growth rate, level and forecasted values were computed over a period of time. The growth rate and level at each observed time was used to compute the forecast value. This forecasted value was used to impute the missing values of the corresponding attribute. To construct a forecasting model, missingness is randomly incorporated for the known data points during training. The TDA process is allowed to forecast the value and the error rates were calculated. A subset of result from TDA process derived for a patient with medical identity 1 for the lab test (T-BIL) taken in the year 1984 is shown in the ▶Table 3. The initial values for growth rate and level are calculated from least square estimation. For this record the initial values for Level (M_{t0}) and growth rate (G_{t0}) are assigned with value of 0.8239 and –0.0008 respectively. The positive value in the growth rate column

(G_t) of ▶Table 3 shows an increase in the trend and negative value shows a decrease in the trend with respect to the date of lab test of the patient. This increase or decrease in the trend indicates the short term and long term changes in the lab test.

Initially the smoothing constant 'α' and 'β' is taken to be 0.2 and 0.4 respectively and based on the interval spacing between the observations the smoothing factors are adjusted using the equation (1), (2), (3) and (4). The smoothing factor 'α' for level and 'β' for trend used in the smoothing factor calculation as specified in equation (3) and (4) is any value chosen between 0 and 1. However, when 'α' closer to 1 it denotes that more weight is given to the recent observations. If stable predictions with smoothed random variation are desired then a small value of 'α' is desire. 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. The performance measures such as MSE (Mean Squared Error), MAD (Mean Absolute Deviation), error rate, MAPE (Mean Absolute Percentage Error) are derived [61]. A 10-fold cross validation method was used to obtain the performance of the forecasting model. The ▶Figure 4 shows the value of MSE for different combinations of smoothing constant 'α' and 'β' for the hepatitis dataset patient record with MID 1. The changes in the values of 'α' , 'β' and the variations in MSE over different observation time points are shown in ▶Figure 4. From this figure, it can be inferred that by adjusting its smoothing constant 'α' and 'β' over different observations there is a decrease in MSE value. A statistical paired t-test [62] was carried out to check whether there is a significant improvement in the performance of presented DES based imputation technique over other imputation techniques such as mean, median imputation, K-nearest neighbor (KNN), hot-deck (HD), maximum likelihood (ML) with significant level of 0.05. For hepatitis data set, the ϱ-value obtained for DES based imputation over mean, median, HD and ML was found to be less than 0.05. For thrombosis data set, the ϱ-value obtained for DES based imputation over mean, median, HD and ML was found to be less than 0.05. The ϱ-value obtained for presented DES based imputation was less than 0.05, so a reject in null hypothesis is considered which means DES based imputation over mean, median, HD and ML provides effective performance results.

In the attribute selection process, a temporal equivalence class is generated using rough set based on the trend and state of each clinical attribute. The temporal degree of dependency for each attribute is computed using the equation defined in (11). The attribute with the highest degree of dependency is taken to be a first candidate in the selected attribute set.

Temporal degree of dependency is calculated for the generated subset and the highest degree of dependency is considered to be 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. ▶Figure 5 shows the first level degree of dependency computed for the clinical attributes (lab tests) from the hepatitis dataset using temporal rough sets. In ▶Figure 5, the attribute GPT for hepatitis patients has the highest degree of dependency. Hence, GPT is considered to be the first candidate in the selected dimension set. Subsets are generated in the combination of GPT with remaining attributes.

▶Table 4 shows the results of attribute selection. For hepatitis patients from total attributes of 29, the temporal pattern based rough set forms most significant attribute set with 25 attributes. Finally, for thrombosis patients the temporal rough set forms the most significant attribute set with 32 attributes from the total attributes of 33. The reduct approximation error $\varepsilon(R)$ for hepatitis and thrombosis dataset is computed using the equation (12). For hepatitis dataset reduct approximation error $\varepsilon(R_{Hepatitis})$ is 0.147 and for thrombosis $\varepsilon(R_{Thrombosis})$ is 0.168. This approximation error illustrates how a reduced attribute set approximates the condition attribute set in association to the decision attribute. The lower reduct approximation error ensures the improvement in the classification accuracies [56]. For evaluation, we have used different combinations of reduced attribute set using Rosetta toolkit [63] and found that the approximation error derived using the identified reduced attribute set is minimal. This ensures that there will be no loss in the information while selecting the relevant attributes which improves the classifier performance.

The presented attribute selection process has selected 25 attributes out of 29 attributes from hepatitis dataset and 32 attributes out of 33 attributes from thrombosis dataset. The number of reduced attributes varies, but still it shows a significant improvement in classification results since it removes the irrelevant attributes before classification without compromising any loss in the information. The

selected significant attributes are considered in the classification process. In this work a sugeno type inference model with 25 input parameters for hepatitis dataset and 32 input parameters for thrombosis dataset is used.

For hepatitis datasets 269 fuzzy rules and for thrombosis 217 fuzzy rules were formed by partitioning the input space using CART [60]. The trend pattern for every input parameter (lab_test) is used to form fuzzy sets and membership functions. The training parameters for neural network type classifiers were decided after experimenting the data with different network sizes, activation functions and learning algorithms back propagation with optimization functions like Levenberg-Marquardt, Gradient descent, Gradient descent with momentum and Scaled Conjugate Gradient etc., that was available in neural network toolbox in MATLAB 2013 [64]. From the experimental settings, the network structure with one hidden layer, 20 hidden nodes, sigmoid activation trained with Levenberg-Marquardt learning optimization function for 160 epochs gives an effective RMSE value. These parameters were considered for training the neural network.

The classification results were compared with classical fuzzy neural network (FN), neural network (NN), Decision tree inductions C4.5, ID3, Support Vector Machine (SVM), Naïve Bayes (NB), K-nearest neighbour (KNN) [25, 65-67]. In this work a back-propagation algorithm was used to train the FN and NN. Decision tree induction is a classification technique. ID3 and C4.5 are the decision tree classifier algorithms taken for comparison [66, 67]. The main difference among these algorithms is in the splitting criteria they choose for identifying test attribute during the decision tree construction. The FN is the fuzzy rule based neural network classifier that trains the network based on the rules extracted from an expert or learning methodologies [65]. The classification results were evaluated with the following performance measures: accuracy, sensitivity, specificity, error rate, precision, positive likelihood ratio (PLR), Negative likelihood ratio (NLR), positive predictive value (PPV), negative predictive value (NPV) [53, 68]. The obtained result shows that the classification accuracy rate is increased in TRiNF on an average of 92.59% for hepatitis and 91.69% for thrombosis patients. The Wilcoxon rank sum test and paired t-test presented by Wilcoxon [69] and *Zimmerman* et al. [62] was carried out with significant level of 0.05 to identify whether there was any significant improvement in the classification accuracy of TRiNF compared with classical FN, NN, C4.5 and ID3 methods. For TRiNF, the ϱ value of less than 0.05 is obtained for hepatitis and thrombosis dataset, which proves that the classification result of TRiNF is improved compared to FN, NN, C4.5 and ID3 classification methods. Since the distribution of clinical time series data is near normal there is no difference between the Wilcoxon rank sum test and Paired t-test.

It has been observed that the temporal data acquisition and temporal rough set induction in the neuro-fuzzy construction on an average had improved the performance of the TRiNF classification system compared to FN, NN, C4.5, ID3, SVM, NB, KNN, KLS and NKFN. ▶Table 5 shows the comparisons for classification results of TRiNF, FN, NN, C4.5, ID3, SVM, NB, KNN, KLS and NKFN based on performance measures namely classification accuracy, sensitivity, specificity, precision and error rate for hepatitis and thrombosis patients.

In the ▶Table 5, values in the parenthesis refer to the classification results obtained without applying proposed TDA process. The state-of -art method discussed in the literature [19, 23] uses the traditional classifiers to perform classification. The authors of these literatures have demonstrated their work with different sets of data. Hence, the classification results obtained from their experimentation cannot be used directly in comparison study with the proposed work. Therefore, to prove the efficiency of TRiNF with those state-of-art methods we have implemented and tested them with our hepatitis and thrombosis data. Thus, the classification results for the related studies mentioned in the ▶Table 5 were derived after testing them with the hepatitis and thrombosis dataset.

5. Conclusions

Due to the presence of temporal complexities such as irregular observations, missing values and large time constrained attributes in clinical time series data, knowledge discovery from these data is considered as a challenging area of research. In this paper, a temporal rough set induced neuro-fuzzy (TRiNF) classification framework that constructs a classification model for effective decision making process is presented. An improved DES method proposed by Wright [9] is adopted to handle

temporal complexities and to derive temporal patterns for each clinical attribute in the time series datasets. The temporal patterns are used further in attribute selection and classification process. A temporal pattern induced rough set is proposed for performing attribute selection to identify relevant attributes for building the TRiNF classification model. The trained TRiNF model effectively classifies the stages of disease. The observed mining results show that, the proposed TRiNF has handled the temporal complexities and increased the classification accuracy on an average of 92.59% for hepatitis and 91.69% for thrombosis patients compared to the FN, NN, C4.5, ID3, SVM, NB, KNN, KLS and NKFN classifiers. There are many interesting aspects for future research. Since clinical time series data are often considered to be irregular, extracting temporal patterns from the clinical variables is a challenging task. Research studies can be carried out to efficiently handle the temporal complexities in clinical data, thereby improving the classification accuracy.

Clinical Relevance Statement

Knowledge discovery from clinical time series data becomes challenging due to its temporal complexities. The proposed temporal mining framework effectively handles the complexities such as missing values, time constrained attributes and irregular observations to build a classification model. This model can assist the clinicIan in clinical decision making.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Protection of Human Subjects and Animals in Research

The primary objective of this work is to propose a clinical decision-making system for assisting the physician in diagnosing a particular disease. The experimentation of the work is carried out using medical datasets (hepatitis and thrombosis) that were collected from patients admitted in Chiba hospital from 1980 to 2001. The authors acknowledge and ensure that there is respect for all human subjects, the personal details and rights of the subjects are confidential. The authors acknowledge that the work has followed all the medical ethical and legal standards for research involving human subjects.

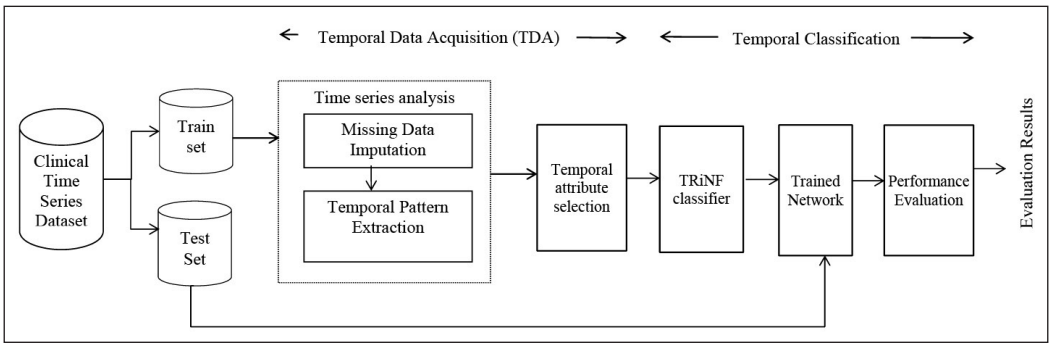
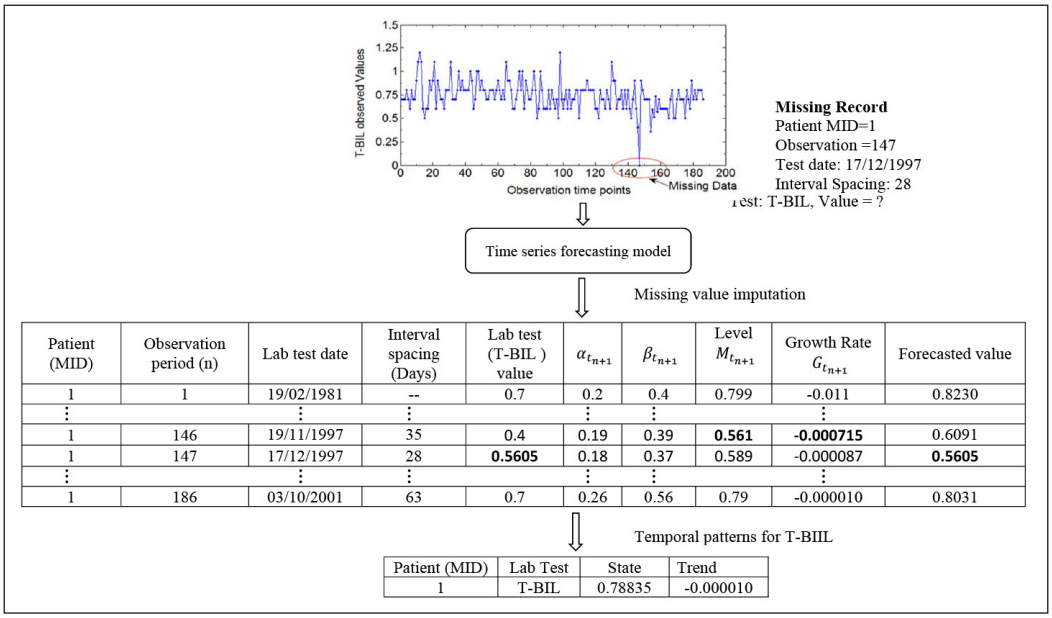


Fig. 1 Proposed framework- TRiNF



Patient (MID)	Observation period (n)	Lab test date	Interval spacing (Days)	Lab test (T-BIL) value	$\alpha_{t_{n+1}}$	$\beta_{t_{n+1}}$	Level $M_{t_{n+1}}$	Growth Rate $G_{t_{n+1}}$	Forecasted value
1	1	19/02/1981	--	0.7	0.2	0.4	0.799	-0.011	0.8230
⋮		⋮	⋮		⋮	⋮		⋮	
1	146	19/11/1997	35	0.4	0.19	0.39	**0.561**	**-0.000715**	0.6091
1	147	17/12/1997	28	**0.5605**	0.18	0.37	0.589	-0.000087	**0.5605**
⋮		⋮	⋮		⋮	⋮		⋮	
1	186	03/10/2001	63	0.7	0.26	0.56	0.79	-0.000010	0.8031

Temporal patterns for T-BIIL

Patient (MID)	Lab Test	State	Trend
1	T-BIL	0.78835	-0.000010

Fig. 2 Illustration of missing data imputation and temporal pattern extraction for Hepatitis patient (MID=1)

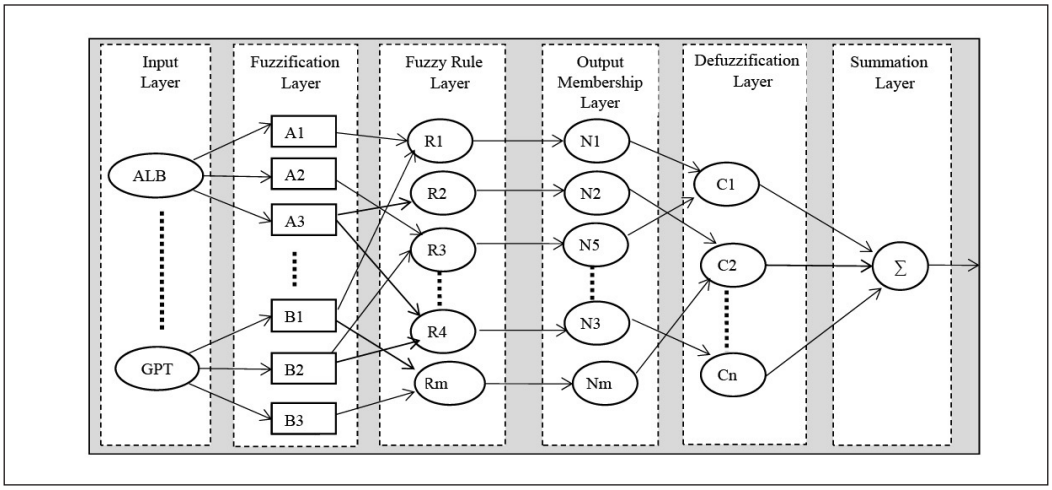


Fig. 3 Network training process: (a) Network Structure and (b) Membership function plot for trend patterns of lab test (ALB, GOT, ZTT, ...GPT

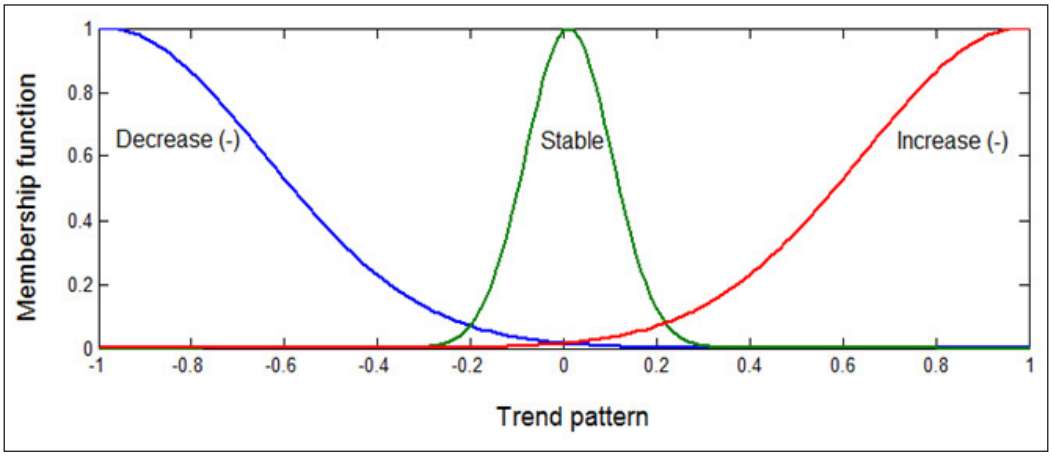


Fig. 3 Continued

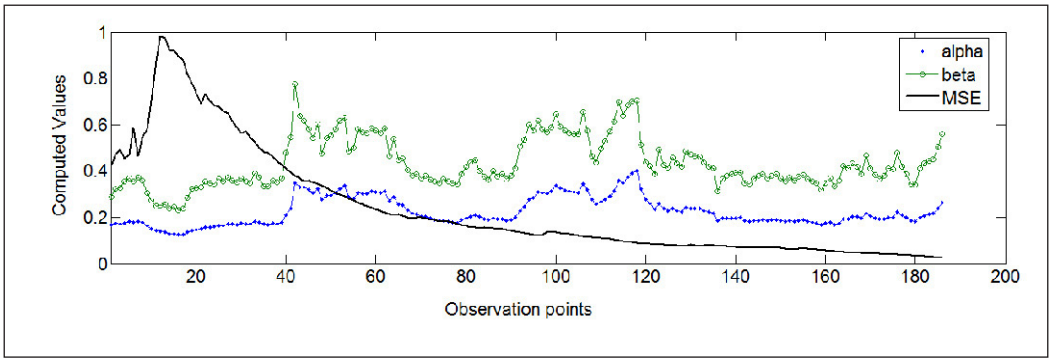


Fig. 4 Plot to depict MSE for different alpha and beta: hepatitis patient (patient_ MID=1, lab examination = T-BIL)

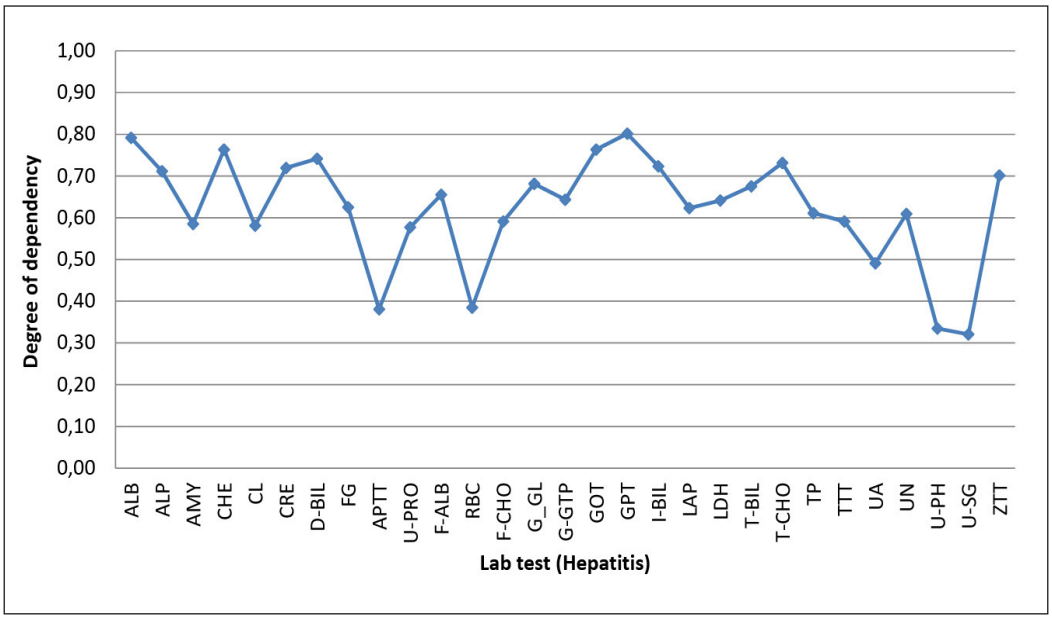


Fig. 5 First level degree of dependency

Table 1 Dataset summary

Dataset	Total Records	Expert Suggested Lab test	Total Patients	Average Missing value (%)
Hepatitis	1565876	29	771	11
Thrombosis	57543	33	1000	8

Table 2 Normalized temporal patterns for subset of lab examination (T-BIL, GPT)

Patients (random selection)	T-BIL		GPT		Class Hepatitis
	State (S_{p_n})	Trend (T_{p_n})	State (S_{p_n})	Trend (T_{p_n})	
1	H	D	H	I	B
2	H	D	H	I	B
3	L	I	L	I	C
4	N	I	H	I	C
5	L	D	N	D	B

Table 3 Results of wright updated DES method for hepatitis patient_ Mid=1, Lab Test = T-BIL, Year= 1981

Patient (MID)	Observation d Date	Interval (days)	Value Y_{tn}	Level M_t	Trend G_t	Forecast F_t	% Error	MSE	MAD_t	$MAPE_t$
		------	-----	0.8239	-0.0008	-----				
1	19/02/1981	1	0.7	0.7984	-0.0107	0.8230	17.5779	0.0151	0.1230	17.5779
1	26/03/1981	35	0.7	0.6994	-0.0028	0.7877	12.5355	0.0114	0.1054	15.0567
1	23/04/1981	28	0.7	0.6998	0.0000	0.6966	0.4838	0.0076	0.0714	10.1990
1	28/05/1981	35	0.8	0.8000	0.0029	0.6999	12.5175	0.0082	0.0786	10.7787
1	02/07/1981	35	0.7	0.7001	-0.0029	0.8028	14.6886	0.0087	0.0834	11.5606
1	29/07/1981	27	0.6	0.6001	-0.0037	0.6972	16.2046	0.0088	0.0857	12.3346
1	02/09/1981	35	0.8	0.7999	0.0057	0.5964	25.4561	0.0135	0.1026	14.2091
1	30/09/1981	28	0.7	0.7005	-0.0035	0.8056	15.0821	0.0132	0.1029	14.3183
1	14/10/1981	14	0.7	0.6979	-0.0002	0.6970	0.4353	0.0117	0.0918	12.7757
1	28/10/1981	14	0.9	0.8910	0.0138	0.6977	22.4736	0.0146	0.1029	13.7455
1	11/11/1981	14	1.1	1.0993	0.0149	0.9048	17.7459	0.0168	0.1113	14.1092
1	02/12/1981	21	1.2	1.2020	0.0049	1.1142	7.1525	0.0160	0.1092	13.5294
1	23/12/1981	21	1.1	1.1019	-0.0048	1.2069	9.7195	0.0156	0.1090	13.2364

Table 4 Results of temporal attribute selection

Dataset	Attribute Selection -Temporal Rough Sets		
	Relevant Attributes	Selected Attributes	Reduct Aproximation error
Hepatitis	25	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,FG	0.147
Thrombosis	32	aCL IgG, ANA, aCL IgA, KCT, LAC, CPK, GLU, WBC, RBC, HGB, HCT, PLT, PT, APTT, FG, A2PI, U-PRO, IGG, IGA, SC170, CRP, RNP, SM, SSA, SSB, CENTROMEA, DNA, RVVT, RA,, RF, IGM, CRE.	0.168

Table 5 Comparison of the classification results

Data-sets	Classi-fiers	Performance measures								
		Accuracy	Sensitivity	Specificity	Error Rate	Precision	PLR	NLR	PPV	NPV
Hepatitis	TRiNF	**92.59** (78.36)	**93.75** (81.45)	**91.00** (74.55)	**7.41** (21.64)	**93.43** (79.72)	10.41 (3.20)	0.07 (0.25)	0.93 (0.80)	0.91 (0.77)
	FN [65]	88.38 (72.34)	89.12 (74.64)	87.38 (69.51)	11.62 (27.66)	90.39 (75.18)	7.06 (2.45)	0.12 (0.36)	0.90 (0.75)	0.86 (0.69)
	NN [53]	81.76 (70.54)	83.16 (73.33)	79.91 (67.25)	18.24 (29.46)	84.64 (72.53)	4.14 (2.24)	0.21 (0.4)	0.85 (0.73)	0.78 (0.68)
	C4.5 [67]	80.16 (69.34)	81.91 (71.28)	77.88 (66.82)	19.84 (30.66)	82.8 (73.63)	3.70 (2.15)	0.23 (0.43)	0.83 (0.74)	0.77 (0.64)
	ID3 [66]	73.55 (62.93)	76.10 (65.85)	70.48 (59.07)	26.45 (37.07)	75.55 (68.00)	2.58 (1.61)	0.34 (0.58)	0.76 (0.68)	0.71 (0.57)
	SVM [53]	79.16 (69.74)	83.05 (70.17)	73.53 (69.12)	20.84 (30.26)	81.94 (76.67)	3.14 (2.27)	0.23 (0.43)	0.82 (0.77)	0.75 (0.62)
	NB [53]	80.96 (70.94)	85.08 (70.51)	75.00 (71.57)	19.04 (29.06)	83.11 (78.20)	3.40 (2.48)	0.20 (0.41)	0.83 (0.78)	0.78 (0.63)
	KNN [53]	76.55 (67.33)	81.02 (66.78)	70.10 (68.14)	23.45 (32.67)	79.67 (75.19)	2.71 (2.1)	0.27 (0.49)	0.80 (0.75)	0.72 (0.59)
	KLS [19] *	90.78 (77.15)	91.19 (80)	90.20 (73.04)	9.22 (22.85)	93.08 (81.10)	9.3 (2.97)	0.1 (0.27)	0.93 (0.81)	0.88 (0.72)
	NFKN [23] *	91.18 (77.56)	91.53 (80.34)	90.69 (73.53)	8.82 (22.44)	93.43 (81.44)	9.83 (3.04)	0.09 (0.27)	0.93 (0.81)	0.88 (0.72)
Thrombo-sis	TRiNF	**91.69** (77.14)	**93.04** (81.91)	**89.53** (69.67)	**8.31** (22.86)	**93.43** (80.88)	8.88 (2.70)	0.08 (0.26)	0.93 (0.81)	0.89 (0.71)
	FN [65]	87.27 (74.03)	90.08 (78.60)	82.77 (66.78)	12.73 (25.97)	89.33 (78.94)	5.23 (2.37)	0.12 (0.32)	0.89 (0.79)	0.84 (0.66)
	NN [53]	80.65 (69.74)	84.31 (74.36)	73.99 (62.42)	19.35 (30.26)	85.51 (75.81)	3.24 (1.98)	0.21 (0.41)	0.86 (0.76)	0.72 (0.61)
	C4.5 [67]	79.35 (67.53)	84.29 (72.55)	70.71 (59.67)	20.65 (32.47)	83.43 (73.81)	2.88 (1.8)	0.22 (0.46)	0.83 (0.74)	0.72 (0.58)
	ID3 [66]	70.26 (61.17)	76.60 (69.04)	59.58 (48.29)	29.74 (38.83)	76.13 (68.61)	1.9 (1.34)	0.39 (0.64)	0.76 (0.69)	0.6 (0.49)
	SVM [53]	81.56 (71.30)	77.86 (71.90)	86.00 (70.57)	18.44 (28.70)	86.97 (74.57)	5.56 (2.44)	0.26 (0.4)	0.87 (0.75)	0.76 (0.68)
	NB [53]	82.73 (71.82)	79.76 (73.10)	86.29 (70.29)	17.27 (28.18)	87.47 (74.70)	5.82 (2.46)	0.23 (0.38)	0.87 (0.75)	0.78 (0.69)
	KNN [53]	79.61 (69.61)	74.05 (65.95)	86.29 (74.0)	20.39 (30.39)	86.63 (75.27)	5.40 (2.54)	0.30 (0.46)	0.87 (0.75)	0.73 (0.64)
	KLS [19] *	90.13 (76.62)	88.81 (73.57)	91.71 (80.29)	9.87 (23.38)	92.79 (81.75)	10.72 (3.73)	0.12 (0.33)	0.93 (0.82)	0.87 (0.72)
	NFKN [23] *	89.22 (75.06)	92.86 (74.05)	84.86 (76.29)	10.78 (24.94)	88.04 (78.93)	6.13 (3.72)	0.08 (0.74)	0.88 (0.79)	0.91 (0.71)

( ) value inside the parentheses represents the classification results without TDA process
*state of art methods

References

1. Augusto JC. Temporal reasoning for decision support in medicine. Artificial Intelligence in Medicine 2005; 33(1): 1–24.
2. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. International Journal of Medical Informatics 2008; 77 (2): 81–97.
3. Adlassnig KP, Combi C, Das AK, Keravnou ET, Pozzi G. Temporal representation and reasoning in medicine: Research directions and challenges. Artificial Intelligence in Medicine 2006; 38 (2): 101–113.
4. Carrault G, Cordier MO, Quiniou R, Wang F. Temporal abstraction and inductive logic programming for arrhythmia recognition from electrocardiograms. Artificial Intelligence in Medicine 2003; 28 (3): 231–63.
5. Keravnou E, Shahar Y. Temporal Reasoning in Medicine, in Fisher, M.D.,et al (Eds), Handbook of Temporal Reasoning in Artificial Intelligence. Elsevier 2005: 587–653.
6. Kalia O, Athena S, Elpida K. Temporal abstraction and temporal Bayesian networks in clinical domains: A survey. Artificial Intelligence in Medicine 2014; 60 (3): 133–149.
7. Chittaro L, Montanari A. Temporal representation and reasoning in artificial intelligence: issues and approaches. Annals of Mathematics and Artificial Intelligence 2000; 28(1–4): 47–106.
8. Miguel RA, Paulo F. Purificación Cariñena: Discovering Metric Temporal Constraint Networks On Temporal Databases. Artificial Intelligence in Medicine 2013; 58 (3): 139–154.
9. Wright DJ. Forecasting data published at irregular time intervals using extension of Holt's method. Management Science 1986; 32 (4): 499–510.
10. Shahar Y. A framework for knowledge-based temporal abstraction. Artificial intelligence 1997; 90 (1): 79–133.
11. Combi C, Shahar Y. Temporal reasoning and temporal data maintenance in medicine: Issues and challenges. Computers in Biology and Medicine 1997; 27(5): 353–368.
12. Stacey M, McGregor C. Temporal abstraction in intelligent clinical data analysis: A survey. Artificial Intelligence in Medicine 2006; 39: 1–24.
13. Shahar Y, Musen MA. Knowledge-based temporal abstraction in clinical domains. Artificial Intelligence in Medicine 1996; 8: 267–298.
14. Haimowitz IJ, Kohane IS. Managing temporal worlds for medical trend diagnosis. Artificial Intelligence in Medicine 1996; 8: 299–321.
15. Miksch S, Horn W, Popow C, Paky F. Utilizing temporal data abstraction for data validation and therapy planning for artificially ventilated new born infants. Artificial Intelligence in Medicine 1996; 8: 543–576.
16. Semrl A. Real time monitoring of dense continuous data. In: Presented at computers in anaesthesia and intensive care: Knowledge based information management; 1999.
17. O'Connor MJ, Grosso WE, Tu SW, Musen MA. RASTA: a distributed temporal abstraction system to facilitate knowledge-driven monitoring of clinical databases. MedInfo2001 2001; 10: 508–512.
18. Tu Bao H, Trong Dung N, Saori K, Si Quang L, DungDuc N, Hideto Y and Katsuhiko T. Mining Hepatitis Data with Temporal Abstraction. KDD '03 Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining 2003: 69–377.
19. Moskovitch R, Shahar Y. Classification of multivariate time series via temporal abstraction and time intervals mining. Knowledge and Information Systems 2014; 42(1): 1–40.
20. Iyad Batal, Hamed Valizadegan, Gregory F. Cooper, Milos Hauskrecht. A Temporal Pattern Mining Approach for Classifying Electronic Health Record Data. ACM Trans Intelligent System Technologies 2013; 4 (4).
21. Moerchen F. Algorithms for time series knowledge mining. Proceedings of the international conference on Knowledge Discovery and Data mining (SIGKDD) 2006: 668–673.
22. Allen JF, Maintaining knowledge about temporal intervals. Communications of the ACM. 1983; 26 (11): 832–843.
23. Bodyanskiy Y, Kolodyazhniy V, Otto P. Neuro-Fuzzy Kolmogorov's Network for Time Series Prediction and Pattern Classification, in Ulrich Furbach, ed. 'KI'. Springer 2005; 3698: 191–202.
24. Petkovic D, Ojbasic ZC and Lukic S. Adaptive neuro fuzzy selection of heart rate variability parameters affected by autonomic nervous system. Expert Systems with Applications 2013; 11: 4490–4495.
25. Jang JSR, Sun CT, Mizutani E. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. Prentice Hall, Inc. 1996.
26. McNameea RL, Sunb M , Sclabassi R. A neuro-fuzzy inference system for modeling and prediction of heart rate variability in the neuro-intensive care unit.Computers in Biology and Medicine 2005; 35: 875–89.

27. Khanna Nehemiah H, Kannan A, Vijaya K, Nancy Jane Y and Brindha Merin J. Intelligent Rule Mining for Decision Making from Clinical Data Sets. ICGST International Journal on Bioinformatics and Medical Engineering 2007; 7: 37–45.
28. 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).
29. Thrombosis Dataset for Discovery Challenge, 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases, September 15–18, 1999, Prague, Czech Republic .http://lisp.vse.cz/pkdd99/ (Accessed: 1 Jan 2006).
30. Vijaya K, Khanna Nehemiah H, Kannan A and Bhuvaneswari Amma NG. Fuzzy Neuro Genetic Approach for Predicting the Risk of Cardiovascular Diseases. International Journal of Data Mining, Modelling and Management 2010; 2(4): 388–402.
31. Kindie Biredagn Nahato, Khanna Nehemiah Harichandran, Kannan Arputharaj. Knowledge Mining from Clinical Datasets Using Rough Sets and Backpropagation Neural Network. Computational and Mathematical Methods in Medicine 2015.
32. Liu, Zitao, and Milos Hauskrecht. Clinical time series prediction: Toward a hierarchical dynamical system framework. Artificial intelligence in medicine 2014.
33. Kalman RE. Mathematical description of linear dynamical systems. J Soc IndAppl Math Ser A: Control 1963; 1(2): 152–192.
34. Rasmussen CE, Williams CKI. Gaussian processes for machine learning.Cambridge, MA, USA: MIT Press 2006.
35. Pandit SM, Wu S-M. Time series and system analysis with applications. NewYork, USA: Wiley 1983; 56(21): 389–405
36. Adorf HM. Interpolation of irregularly sampled data series – a survey. In: Astro-nomical Data Analysis Software and Systems IV 1995; 77: 460–463.
37. Dezhbakhsh H, Levy D. Periodic properties of interpolated time series. EconLett. 1994;44(3):221–8.
38. Kreindler DM, Lumsden CJ. The effects of the irregular sample and missing data in time series analysis. Nonlinear Dyn Psychol Life Sci 2006; 10(2): 187–214.
39. Rehfeld K, Marwan N, Heitzig J, Kurths J. Comparison of correlation analysis techniques for irregularly sampled time series. Nonlinear Process Geophys 2011; 18(3): 389–404.
40. Chu C-SJ. Time series segmentation: a sliding window approach Inf Sci 1995; 85(1): 147–173.
41. Keogh E, Chakrabarti K, Pazzani M, Mehrotra S. Dimensionality reductionfor fast similarity search in large time series databases. Knowl Inf Syst 2001; 3(3): 263–286.
42. Bahadori, Mohammad Taha, and Yan Liu. Granger Causality Analysis in Irregular Time Series. SDM 2012.
43. Cuevas-Tello J C, Tino P, Raychaudhury S, Yao S, and Harva M. Uncovering delayed patterns in noisy and irregularly sampled time series: an astronomy application. Pattern Recognition 2009; 43(3): 36.
44. Scargle J D. Studies in astronomical time series analysis. I – Modeling random processes in the time domain. The Astrophysical Journal Supplement Series 1981 45(1): 1–71.
45. Ceusters W, Smith B.Strategies for referent tracking in electronic health records. J Biomed Inform 2006; 39(3): 362–378.
46. Ohsaki M, Sato Y, Yokoi H and Yamaguchi T. A Rule Discovery Support System for Sequential Medical Data, In the Case Study of a Chronic Hepatitis Dataset, International Workshop on Active Mining. IEEE International Conference on Data Mining ICDM, Maebashi 2002; 97–102.
47. Jensen S. Mining Medical Data for Predictive and Sequential Patterns Discovery Challenge on Thrombosis Data. Freiburg 2001.
48. Zytkow J, Tsumoto S and Takabayashi K. Medical (Thrombosis) Data Description. In: (Siebes A. &Berka P. eds.) PKDD2000 Discovery Challenge. Lyon. 2000.
49. Zytkow J and Gupta S. Guide to Medical Data on Collagen Disease and Thrombosis. In: (Berka P. ed.) PKDD2001 Discovery Challenge on Thrombosis Data. Freiburg. 2000.
50. Little RJA, Rubin DB. Statistical analysis with missing data. Wiley. 2nd edition. 2002.
51. Enders C. Applied missing data analysis. Guilford, New York. 2010.
52. Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting 2004; 20 (1): 5–10.
53. Han J, Kamber M. Data Mining, Concepts and Techniques 2nd ed., Morgan Kaufmann Publishers Inc. San Francisco. CA. USA. 2001.
54. Pawlak Z. Roughsets. International J Comp Inf Sci 1982; 11 (5): 341–356.
55. Dash M,Liu H. Attribute Selection for Classification. Intelligent Data Analysis An International Journal 1997; 1(1–4): 131–156.

56. Komorowski J, Zdzislaw Pawlak , Lech Polkowski , Andrzej Skowron. Rough Sets Perspective on Data and Knowledge. in Z. Pawlak, et al (Eds). The Handbook of Data Mining and Knowledge Discovery. 1999: 134–149.
57. Jensen, Richard and Qiang Shen. Rough set based attribute selection: A review. Rough Computing: Theories, Technologies and Applications 2007: 70–107.
58. Chouchoulas A and Shen Q. Rough set-aided keyword reduction for text categorisation. Applied Artificial Intelligence 2001; 15 (9): 843–873.
59. Pradipta M,ParthaGarai. Fuzzy-Rough Simultaneous Attribute Selection and Attribute Extraction Algorithm. IEEE Trans. Cybernetics 2013; 43(4): 1166–1177.
60. Breiman L, Friedman J, Olshen R , Stone C. Classification and Regression Trees. Wadsworth International Group 1984.
61. Cooray TMJA. Applied Time Series: Analysis and Forecasting. Alpha Science International Limited Oxford 2008.
62. Zimmerman, Donald W. A Note on Interpretation of the Paired-Samples t Test. Journal of Educational and Behavioral Statistics 1997; 22 (3): 349–360.
63. Ohrn A, Komorowski J, Skowron A, Synak P, The Design and Implementation of a Knowledge Discovery Toolkit Based on Rough Sets – The Rosetta system. In: Polkowski and Skowron 1998:76–399.
64. Demuth H, Beale M . Neural network toolbox for use with MATLAB. Users guide. http://www.mathworks.com Edition. The Math Works Inc 2013.
65. Gabrys B and Burgiela A. General fuzzy min-max neural network for clustering and classification. IEEE Trans. Neural Networks 2000; 11 (3): 769–783.
66. Quinlan JR. Induction of decision trees. Machine Learning 1986; 1 (1): 81–106.
67. Quinlan JR. C4.5: Programs for Machine Learning. Morgan Kaufmann Los Altos 1993.
68. Xiao-Hua Zhou, Nancy A. Obuchowski, Donna K. McClish , Statistical Methods in Diagnostic Medicine, 2nd Edition,New York: Wiley,March 2011.
69. Wilcoxon F. Individual comparisons by ranking methods. Biometrics 1945; 1(6): 80–83.