Iranian Journal of Operations Research
Vol. 13, No.2, 2022, pp. 21-43
DOI:

Predicting the recovery of COVID-19 patients using recurrent neural network and Markov chain

Mehrdad Fadaei PellehShahi[1], Sohrab Kordrostami[2]*, Amir Hossein Refahi Sheikhani[3], Marzieh Faridi Masouleh[4], Soheil Shokri[5]

In this paper, a new method is presented using a combination of deep learning method, specifically recursive neural network, and Markov chain. The aim is to obtain more realistic results with lower cost in predicting COVID-19 patients. For this purpose, the BestFirst algorithm is used for the search section, and the Cfssubseteval algorithm is implemented for evaluating the features in the data preprocessing section. The proposed method is simulated using the real data of COVID-19 patients who were hospitalized in treatment centers of Tehran treatment management affiliated to the Social Security Organization of Iran in 2020. The obtained results were compared with three valid advanced methods. The results showed that the proposed method significantly reduces the amount of memory resource usage and CPU usage time compared to similar methods, and at the same time, the accuracy also increases significantly.

Keywords: *Prediction, Covid-19, Recovery, Markov Chain, Recurrent Neural Network*.

Manuscript was received on 22/06/2022, revised on 06/03/2022 accepted for publication on 08/05/2023.

1. Introduction

Prediction and estimation are very important in the business of any organization, and accurate predictions result in higher productivity, more cost savings, increased quantity and quality of profits, and better services to stakeholders.

By examining the degree to which deep learning addresses common modeling challenges in support of decision making, the followings are provided. First, previous studies have investigated the effectiveness of deep learning based on customer behavior data at the individual level and, by predicting the risky behaviors of traders, have focused on microfinance, which is an important area of operational research. Experimental results show that deep learning predicts more accurately than machine learning methods. Second, the ability of deep learning to automatically learn the information features of operational data has demonstrated [87].

Process analysis involves a complex layer of data analysis based on the traditional concept of process mining [82]. Compared to process mining, process analysis represents a more important issue in using process-generated or process-related data to provide practical insights into business processes. Process analysis uses a wide range of data and includes not only the processing of the records, but also event-related records [67], supply-related records, decision-related or content-

[1] Iranian Social Security Organization, m.fadaei.1357@gmail.com
*Corresponding Author
[2] Department of Mathematics, Lahijan Branch, Islamic Azad University, Lahijan, Iran, kordrostami@liau.ac.ir
[3] Department of Mathematics, Lahijan Branch, Islamic Azad University, Lahijan, Iran, ah_refahi@liau.ac.ir
[4] Computer and Information Technology Department, Ahrar Institute of Technology and Higher Education, Rasht, Iran, m.faridi@ahrar.ac.ir
[5] Department of Mathematics, Lahijan Branch, Islamic Azad University, Lahijan, Iran, soheilshokri@gmail.com

related records [71] and questions about the analyses such as resource optimization and sample prioritization [39]. Process mining techniques make it possible to extract useful information from event logs and information based on the date of Business Plan (BP) [1].

Deep learning has recently gained considerable attention and is essentially an application of neural networks [68]. Recent innovations in the algorithms which provide new structures of neural networks as well as computing hardware innovations, in particular, access to the Graphics processing unit (GPU), have directed great attention towards neural networks and the term “deep learning” has, accordingly, been widely used [47]. Recurrent Neural Networks, in particular, Long Short-Term Memory (LSTM), have led to unexpected advances in solving complex sequences of modeling in various domains such as image understanding, speech recognition, and natural language processing [47, 68].

In [54], a novel prediction model that predicts the number of new confirmed cases is presented. The proposed model uses a set of statistical based techniques in a supervised machine learning process. The model is tested on Egypt as well as the top 10 ranked countries for COVID-19 till end of September 2020. The results of the proposed model are compared against the Bayesian Ridge regression model. In [54], the rank of Egypt based on the number of confirmed cases and the rate of change is calculated. It is found that, Egypt’s rank is 43 around the world based on the number of infections while, its rank based on the change in rate is 143. These calculations are performed using the WHO dataset till end of September 2020.

Semi-structured social insurance business processes are defined by the heterogeneous data available. Corona disease data from the Social Security can be divided into two components as shown in Figure 1.

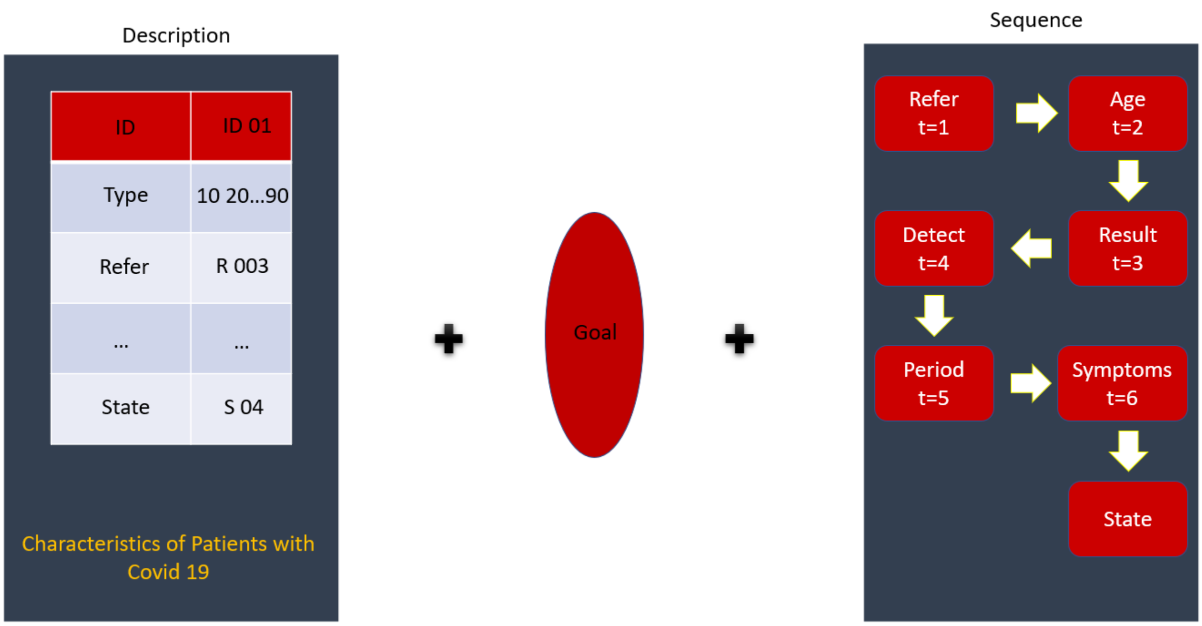

Figure 1. An example of a heterogeneous data set in Covid 19

Semi-structured business characteristics do not change over time and transactions that occur over the life of the insured are recorded. Heterogeneous social insurance data are divided into two main levels: data level and structure level. How to use this information effectively in the social insurance industry is a very important question.

There are two general and specific challenges to developing a right predicting system for social insurance. The main general challenge is the sources, and the main challenge of a recurrent neural network, is the high proliferation of memory nodes. In the present study, it is suggested that the Markov chain method be used to select subsequent nodes. The study tries to present a method capable of producing the results closer to the real world with little time and processing overheads, given the previous records of an event and the prediction systems of social insurance.

The present study consists of the following sections: In Section 2, a summary of some basic concepts for predicting processes and execution time in semi-structured businesses. In Section 3, the research literature is examined by process prediction, and in particular, the prediction of the next event in a process. In section 4, the methods are presented in this study. In Section 5, the implementation and evaluation of the empirical results are reported in detail; then the proposed method is compared with the three advanced methods on the real dataset of the Tehran treatment management from the Iranian social security organization. Finally, in section 6, the findings are discussed and the challenges are identified in this area again.

2. Basic concepts

Social and business insurance differ in details and mechanisms, but they are all designed to deal with unforeseen risks and share a common purpose. Therefore, accurate prediction is very important to the insurance business and especially social insurance.

Semi-structured processes are widely created in industries such as government industry, insurance, banking and health care [8]. Some examples of these processes are car insurance claims, prescription drug administration and patient management in the hospital. These processes depart from the traditional, predetermined, structured and sequential processes because their life cycles are not entirely driven by a process model [89].

In an environment where file processing requires the analysis of large volumes of dynamic data and deadlines for quick execution, file management is challenging and error-prone even for experts [45].

Bankruptcy prediction models estimate default probabilities based on the ratios of accounting variables (e.g., total assets/total liabilities) [19]. In the context of deep learning, these ratios represent a low-level representation. Using balance sheet figures as input, the lower layers in deep neural networks can relate variables to each other and calculate information ratios using the data-driven method. Higher levels of data representation may include this trend in financial ratios or interdependencies between ratios. Specific representation is calculated independently. A hierarchical combination of representations of different complexities enables deep neural networks to learn abstract concepts such as delinquent borrower. Representation learning also increases the ability of a model to extract patterns that are not well represented in the training data, which is a problem for machine learning models [57].Event prediction is a task to predict future events based on events that have already occurred [15, 45].

Recent developments in artificial intelligence have provided new opportunities for the insurance industry to create appropriate solutions and services based on customers' new knowledge and implement advanced operations and Business functions. However, insurance data is heterogeneous, and the imbalanced class distribution with low-frequency and high-dimensional has led to four major learning challenges in the real-world business. Traditional machine learning algorithms can mainly be used only for standard datasets, which are usually homogeneous and balanced. By increasing the computing power of modern technologies, especially machine learning and deep learning algorithms, the ongoing topics of image, text and speech recognition have begun to be used in business data analysis. With this trend, insurance operations can benefit greatly from recent developments in artificial intelligence and machine learning. Some insurers use machine learning methods to analyze some types of data with lower cost, and improve profitability in their business. For example, they may use the analyzed outcomes for making a commitment, assistance to employees in arranging large datasets collected by insurance companies to identify high-risk items, and thus, to reduce claims [35].

Conventional process mining algorithms [79, 12, 30, 81, 84] are able to extract information from event logs and can help to create a process model and subsequent event.

2.1. Process Mining

Process mining focuses on finding the differences between the two processes with regard to key performance indicators such as process time [42]. Most companies currently store large amounts of data extracted from their information systems in event logs. The event log is a set of cases in which each item has a track record of relevant events through the process [85]. By storing this information in the event log, companies can track their process implementation in great detail [42]. Process mining [1] is a field in computer engineering that aims at discovering, monitoring and improving business processes [76]. Using process mining techniques in event logs, the process model of a system can be extracted and displayed [77, 83].

One way to gain a deeper insight into a process is to use a compliance review. In the compliance review, a process model has been compared with input data, such as event logs [70, 80].

2.2. Recurrent Neural Networks (RNN) and LSTM

LSTMs are a special kind of RNNs which are capable of learning long-term dependencies [27]. LSTMs are resistant to vanishing gradient problem and are specifically designed to avoid the problem of long-term dependency [36].

RNN is inappropriate to be used in circular cells and this leads to LSTM. A basic LSTM cell is defined as follows: C_{t-1} and h_{t-1} are accepted as a state and the inputs of the initial open cell are accepted at the same level and X_t is accepted as the input of the cells of the previous layers. In turn, C_t and h_t as the state, transfer the outputs to the next open cell and provide h_t as output to the next layer [15].

3. Research Literature Review

Using different data mining techniques to extract patterns and consequently knowledge from different types of databases is necessary and important in process mining. Data mining will be very successful when the challenges or problems are properly detected and sorted [32].

In [45], a probabilistic process model (PPM) is presented that illustrates the probabilities of transition to a semi-structured process sample. A specific PPM sample tries to predict the probability of different results. Some PPMs can also become a Markov chain.

The main contribution [3] is a technique based on DBNs that is capable of context-sensitive process prediction. Our artifact contributes to the discussion on how to use event log data and its associated contexts for prediction by introducing the concept of symptom and background variables. By instantiating our technique, we also contribute to practice by offering the possibility to make use of the artifact for real-world data sets. Their research uses a problem-centered technique, as it commences with the *Identify & Motivate Problem* phase. They designed a PPM technique (CECA-DBN) that is both process-aware and context-sensitive based on established research from the area of DBNs. Their dif- ferentiation of types of context based on them having a cause (background) or effect (symptom) relationship to the process flow, is novel in the PPM field. Through their benchmark on established data sets, they showed that CECA-DBN can improve the predictive quality of probabilistic models by including additional context information [3].

In [59], a systematic literature review is carried out to capture the state-of-the-art deep learning methods for process prediction. In total, 32 different approaches are compared against carefully selected criteria to identify strengths and weaknesses and reveal research gaps for future research. The main focus [59] review lays on a qualitative comparison of existing implementations. In particular, the literature is classified along the dimensions: neural network type, prediction type, input features and encoding methods.

In [57], we investigated the efficiency of a one-way language model approach among fully attention-based transformer models for predicting future process events of the current process instance being executed.

The study provided by lwendi et al. [31] aims to fill the void of the traditional healthcare system, using machine learning (ML) algorithms to simultaneously process healthcare and travel data along with other parameters of COVID-19 positive patients, in Wuhan, to predict the most likely outcome of a patient, based on their symptoms, travel history, and the delay in reporting the case by identifying patterns from previous patient data. Our contribution includes:

• Processing of healthcare and travel data using machine learning algorithms in place of the traditional healthcare system to identify COVID infected person.

• This work compared multiple algorithms that are available for processing patient data and identified the Boosted Random Forest as the best method for processing data. Further, it executed a grid search to fine-tune the hyper parameters of the Boosted Random Forest algorithm to improve performance.

• Their work obliterates the need to re-compare existing algorithms for processing COVID-19 patient data [31].

In [16], suggest a method for pathology prognosis based on analysing time series of chest X-ray images. The proposed method is based on both recurrent and convolutional neural networks, and allows to classify patients into two severity classes: positive or negative evolution. The main originality lies in the use of such a combination for COVID19 prognosis.

Association rules mining algorithms [49] have been successful in applications in many scenarios, such as gene expression data mining [34], data mining, and blog analysis [28]. In [45], a Markov prediction model is proposed to create a probabilistic sample-specific model that predicts the probability of a particular event occurring in the specified sample of a running process.

Evermann et al. [15] and Tax et al. [2] have tried to use LSTMs to predict future activities in recording business-related events. In [21], an event-based spiking neural network incorporating learnable delays is presented to predict the time sequence.

In [56], we propose FinDeepBehaviorCluster, a systematic way of utilizing click-stream data for fraud detection and fraud pattern mining. Specifically, time attention based Bi-LSTM is used to learn the embedding of behavior sequence data. In addition, to increase the interpretability of the system, handcrafted features are generated to reflect domain knowledge.

Abidin et al [2] attempt to develop and to compare the performance of the MLP model against the logit model by using a sample of 41 failed SMEs matched with 41 healthy SMEs in the hospitality industry from 2000 to 2016. Results show that the MLP model gives a higher prediction accuracy rate for both the estimation and holdout samples than the logit model. In terms of the failure indicators, both models identify ROA and board size as important determinants that could differentiate between a failed and a non-failed SME. In addition, the MLP model identifies other variables as main indicators of failure, i.e. current ratio, debt to equity ratio and net income to sales. Creditors, regulators and investors should consider the MLP prediction model as it provides a more accurate and reliable assessment of the company’s financial status. An effective failure prediction model could reduce economic losses to the affected parties by providing signals that would enable them to take preventive measures to possible adverse situations.

Peng et al. [61] first performed preprocessing and exploratory analysis based on Weibo Philanthropy samples and then introduced a series of machine-learning algorithms rarely used in medical crowdfunding before. The 10-fold cross-validation is employed in the training stage, and parameters are optimized by grid search for each algorithm. Indicators mean abstract error, mean-squared error and R-squared are applied to evaluate the performance of algorithms. The experimental results show the performance of Classification and Regression tree, Artificial Neural

Network, Xgboost are not much different, outperforming other algorithms, such as K-Nearest Neighbors and Linear Regression.

Singh et al. [72] was carried out to test all of the parameters and to analyse how stock market prediction actually works. There are several companies that are lacking at the moment because they can't anticipate or foresee future problems in order to make the right choices. Different techniques have been utilised in this project work like Linear regression, K-means Clustering, K nearest neighbour, LSTM, etc. Use of algorithms in stock prediction has proven to be essential and has thus marked their use in strong market plans.

The creation of a Markov model of a data-driven semi-structured business process extracted from a set of process implementation cues is presented in [45]. The model can predict the probability of all potential future tasks in an example of a semi-structured business process based on the values of data associated with it, while also considering the information contained in the loops and parallelism in the process.

Recording events provided by the information systems of all the data related to the implementation of processes has been carried out in [53] and has been used to create models that allow monitoring of business model prediction. According to the predictions, crisis management and resolution of problems has been achieved. In this context, a set of 39 methods has been presented for monitoring of business model prediction based on various prediction techniques. These methods have been classified based on knowledge of the methods and techniques used to build the prediction model (regression, etc.).

Asjad Khan et al. have investigated the application of deep learning to solve a number of problems related to the analysis of predictive processes by using a specific type of neural network known as memory-augmented neural network (MANN). In a typical case, a MANN is a recurrent neural network added to an external memory matrix. In this paper, a special type of MANN called Differentiable neural computer has been used [40].

In [50], a RNN-based combination model has been proposed to predict several functions of event sequences. LSTM networks have been adapted to encrypt information of the events and their features separately, and then a combiner has been used to combine them as a hidden representation of information based on the dates of specific sequences. Then, another LSTM layer has been implemented as a decoder to predict the next event and its features simultaneously with a separate output layer for each task, and the prediction of the next event and the production of extension has been evaluated using five real datasets to predict the feature.

In [35], the focus was on cost-sensitive parallel learning framework (CPLF) to enhance insurance operations with a deep learning approach that does not require preprocessing. Their method involves a coherent and new parallel neural network that provides truly homogeneous data. Then a cost-effective custom-designed matrix automatically provides a robust model for classification of learning, and the parameters of both the cost-effective matrix and the hybrid neural network are used alternately, but are jointly optimized during training.

Research in artificial intelligence and deep learning is increasing at a very high rate [47]. In this study, we focus on deep learning programs in social insurance business. To clarify the method of selecting papers, we confirm that deep learning has other applications in the fields of finance [24], modeling of dynamism in order book data [73], and so on. Deep learning has been also used to create financial predictions from textual data [43]. It may be argued that a recurrent neural network (RNN) is the same as a deep neural network because it has temporary recursive deep cells. With the advent of deep learning, recurrent neural networks such as long short-term memory (LSTM) have become more popular and are often referred to as deep neural networks [17]. This is not necessarily the case for their previous ones, some of which are used in finance [29].

Users of deep neural network use data sets from more than 3.5 billion observations, along with 272 variables related to loan characteristics and local economic factors, to estimate and support portfolio management. For this purpose, they model the probabilities of transferring individual loans between the existing state and the arrears payment state as well as fraud or non-payment, using recurrent networks including up to 7 layers [74].

Another common goal of process prediction is the binary evaluation of its result, i.e., whether a process instance will fail or not. In the early 2000s, it was first addressed by Castellanos et al., and Grigori et al., who used decision trees, resources, and case data [24, 22, 23]. The decision tree was also used in Conforti et al.'s approach [6]. Kang et al. provide two different approaches, one using a Support Vector Machine (SVM) [37] and the other one on clustering and local outlier data detection [38] to predict process failures. Maggie et al. used decision trees to predict LTL violations [52]. Leontjeva et al. used random forests for the same purpose [48]. These two techniques have been combined in Francescomarino et al.'s approach [14]. Metzger and Folino et al. both have dealt with binary results; while the former applies the neural networks, constraint satisfaction, and service quality aggregation [55] , the latter relies on clustering and regression [18].

Only five approaches [19, 45, 46, 75, 5] are related to subsequent event prediction, many of which use an explicit representation of process models such as Hidden Morkov Model (HMM) and Probabilistic Finite Automatons (PFA).

Each prediction problem requires the creation of a new decision tree in relation with the training dataset to predict the output class [65, 63]. In [44], an initial survey was performed to predict the probabilities of future tasks in a semi-structured business process instance assuming that the process was loop-free and not parallel.

Some possible approaches aiming at modeling and extracting business processes are Markov models [7], models presented by stochastic process modeling techniques for finite state machine production [9], and random graph-based models [25, 26]. All of these techniques focus on extracting, modeling, and simulating business processes. It is different from probabilistic prediction of different results based on the value of the document contents in a running instance of a distinct business process in which decisions on subsequent tasks are guided by the contents of the document. Several probablistic models have recently been presented in researches related to business process management to predict the next step in a business process [69].

Numerous decision tree techniques or similar tools have been proposed for predicting business processes. Using the available tools in SPSS, the specific association between the practical processing of a case and the direct features associated with a case has been investigated [78].

Probabilistic process models have been proposed as a way of tracking the progress of one process and answering questions such as the probabilities that a sub-process will perform according to the status of the other sub-processes [62]. Colored Petri Nets [33] are presented as the representative of alternatives to the probabilistic process model. Colored Petri Nets (CPNs) can model a general class of random processes [33, 88]. Rozinat et al. have combined decision extraction results with an extracted process model to create CPNs for business process modeling and simulation [66]. While the focus of [66] is on end-to-end sample simulation using CPNs to answer "what if" questions and provide information such as runtime and waiting time for process samples.

Integrated probabilistic process models [62, 66] with theoretical decision models [58, 86] have captured a lot of attention to create integrated decision support systems [51]. A decision support system empowers ordinary users in both decision analysis and domain knowledge. It has, for example, combined decision analysis with investment valuation techniques and stock market knowledge [64]. A detailed discussion of case-based reasoning, rule-based reasoning, and combinatorial methods was provided in [51] to support the decision-maker. Liu et al. [51] conducted a comprehensive review of the available prediction techniques that integrated Case Based Reasoning (CBR) [13] and Rule Based Reasoning (RBR) techniques with reasoning methods such as Bayesian Belief Networks to develop Integrated Decision Support Systems (IDSS).

There are several differences between predictive monitoring and other prediction tasks that are mentioned below. First of all, process-aware methods are clearly distinct. Second, real-time predictive monitoring is performed during the execution of process instances over a specified period. This means that the prediction is made at a particular point in the implementation, called the control point [55, 18].

One way to find out the reason of bottlenecks is finding links between different process features [42, 10, 11].

Many authors have suggested techniques for finding relevance in an event log. Multiple approaches are implemented to predict the time remaining of a process based on a partial tracking [45, 79]. Other authors have focused on using previous features to predict business performance [6, 20 and 41]. The other approach focuses on the prediction of violations due to trade constraints [52]. Based on these approaches, the ProM plugin was created. This plugin is called feature prediction and connects these ideas to a plugin [10, 11].

As mentioned in the research literature, traditional approaches to solve the next event prediction problem include the use of mode transfer models, HMM and PFA models [39].

4. The proposed method

In general, the proposed method uses the recurrent neural network method to solve the prediction problem. However, the Markov chain is used to predict subsequent nodes and eliminate additional processes in order to reduce the nodes and layers of the network and eliminate the cost problems and to improve or at least maintain the accuracy of the computation results.

In this section, the proposed method will be described in four parts. In the first section, the preprocessing data will be addressed to customize the data to obtain the proper quality data. The proposed recurrent neural network will be examined. Then the method of creating the Markov chain is addressed based on the problem conditions, and finally a general overview of the proposed method is provided by combining the different sections.

4.1. Preprocessing data

In the preprocessing section, the number of attributes and the data normalization will be reduced.

4.1.1. Reduction of the number of data attributes

- Search method: The search method specifies how much data is recognised as a subset of a feature. Moreover, in the search method, the method of search for an agent is also specified among the created sub-domains.
- Attribute evaluator: The type of comparison and finding similarities and differences between attribute values is determined by the comparator. A comparator might, for instance, consider the number of iterations as similarities, or another comparator might consider the intervals between the data of an attribute as the main factor for similarity values.

The proposed method applies the BESTFIRST algorithm to the search section and the Cfssubseteval algorithm for the feature comparison section.

If figure 2 is considered as the value of CFSSUBSETEVAL determined by the number of unique data in each attribute and the attribute represents its value, the most valuable subset of attributes that is achieved can be obtained by using BESTFIRST method.

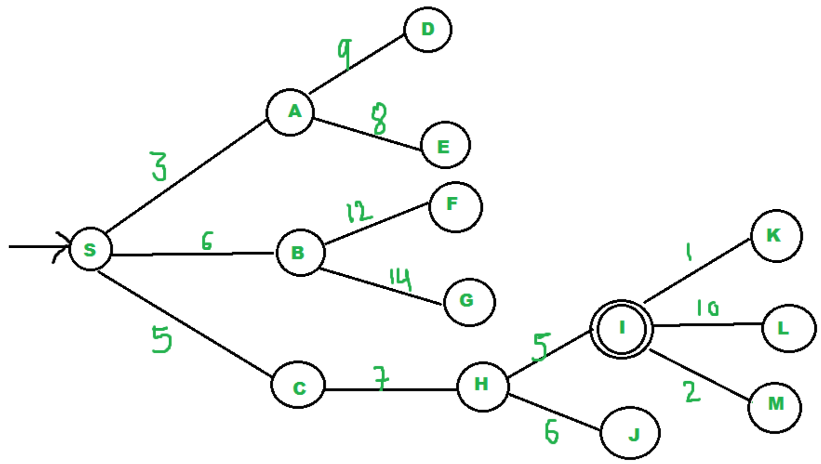

Figure 2. A set of features investigated using CFSSUBSETEVAL and BESTFIRST [87]

Although it can be seen that the best set of attributes of *S*, *C*, *H*, *I*, *L* is the most valuable set of attributes, this method can produce the best results in a relatively short time and relatively little complexity.

4.1.2. Data normalization

The first step to create a usable data is to convert all available values into numerical values. After converting all strings into numbers, the numerical intervals must be converted into a constant value for all attributes to eliminate the effect of very large data in computation. Therefore, relation 1 will be used.

$$X_i = \frac{D_i - Min}{Max - Min} \tag{1}$$

This relation maps the data between 0 and 1 where X_i is the normalized numerical value and D_i is the original data value in the dataset. Min and Max represent the minimum and maximum data available for that attribute in the database, respectively.

4.2. Recurrent neural network

The proposed recurrent neural network will be a many to many online network. In this type of multilayer network, each level is connected online both to its next layer node and the current layer node. In other words, each layer is associated with several nodes of the other layer in a multipoint-to-multipoint communication, and the results of each layer affect those of the other layers online. The network will simultaneously process each input and produce new output. Like all RNN networks, it has an internal state and looks at that state as well as new inputs at each step for making decisions. This state is updated with each new entry.

The output of each proposed RNN layer will be obtained through relation 2:

$$a_j^l = \sigma\left(\sum_i^N w_i^{l-1} a_i^{l-1} + b_j^l\right) \tag{2}$$

In this relation, a_j^l is the output of cell *j* in the *L* layer. *N* is the total number of cells in the *L-1* layer and a_i^{l-1} is the response of i^{th}-cell of the *L-1* layer. b_j^l is also the bias value of j^{th} cell of *L* layer. The function is also obtained from relation 3:

$$\sigma(x) = \frac{1}{1 + exp\,(-x)} \quad (3)$$

Figure 3, shows an overview of the recurrent neural network:

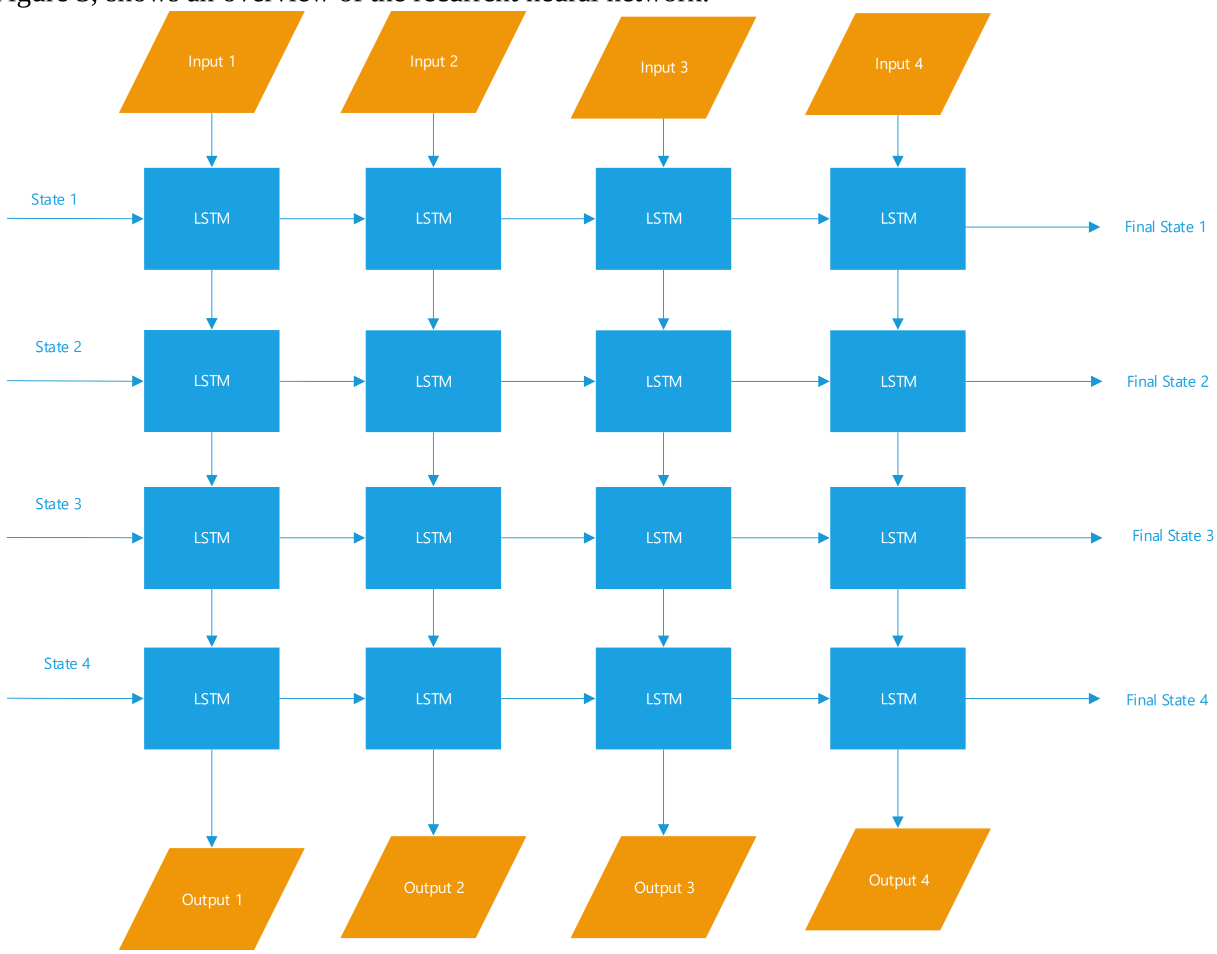

Figure 3. an overview of the proposed RNN network

This structure uses LSTM.

Two functions are available for each input of the LSTM. These two functions can be seen in relations 4 and 5:

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (4)$$

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (5)$$

In these two relations, the functions of f_t and i_t are Data Forget gate and Data Input gate, respectively. f_t allows the LSTM to destroy C_{t-1} inputs. i_t also enables the LSTM to select C_{t-1}

inputs as needed. In these two relations, variables W (weight) and b (bias), are two learnable variables and will be updated according to the final results.

The set of possible values for the current situation or $\overline{C_t}$ is calculated by using relation 6.

$$\overline{C_t} = tanh(w_c[h_{t-1}, x_t] + b_c) \tag{6}$$

The value t actually specifies the sum of all the cells before the current cell in the previous layers and the current layer.

Selecting the best set of C_t from all of the set of subgroups of $\overline{C_t}$ is a challenge that reduces the computation volume in the RNN by optimal selection. This method of selecting C_t can virtually remove some of the LSTMs from the computational cycle, and eventually, reduce LSTMs or even RNN network layers by removing these nodes.

4.2.1. Markov Chain

As mentioned earlier, choosing a suitable C_t for an RNN is a major challenge and the Markov chain is the proposed method to solve this problem.

The proposed Markov chain consists of 5 stages.

1. Receiving the input set
2. Creating a decision tree for each node in the process model
3. Calculating the probability of a process occurring in a particular *PPM* and creating the Markov chain
4. If a PPM contains a parallel path, then the probability of the transition states can be calculated by using transmission states in the Markov chain.
5. If the model has a parallel path, then the probability of future tasks is calculated by using the concept of the first transition time, and it is assumed that the next task depends only on the current task. Otherwise, Markov methods, including the probability of the absorption period or the first transmission will be considered for calculating the probability of performing the subsequent tasks.

A PPM will be displayed in relation 7.

$$PPM = (G, l_{input}, l_{output}, \alpha, \beta) \tag{7}$$

Figure 4, provides an overview of the Markov chain in this study.

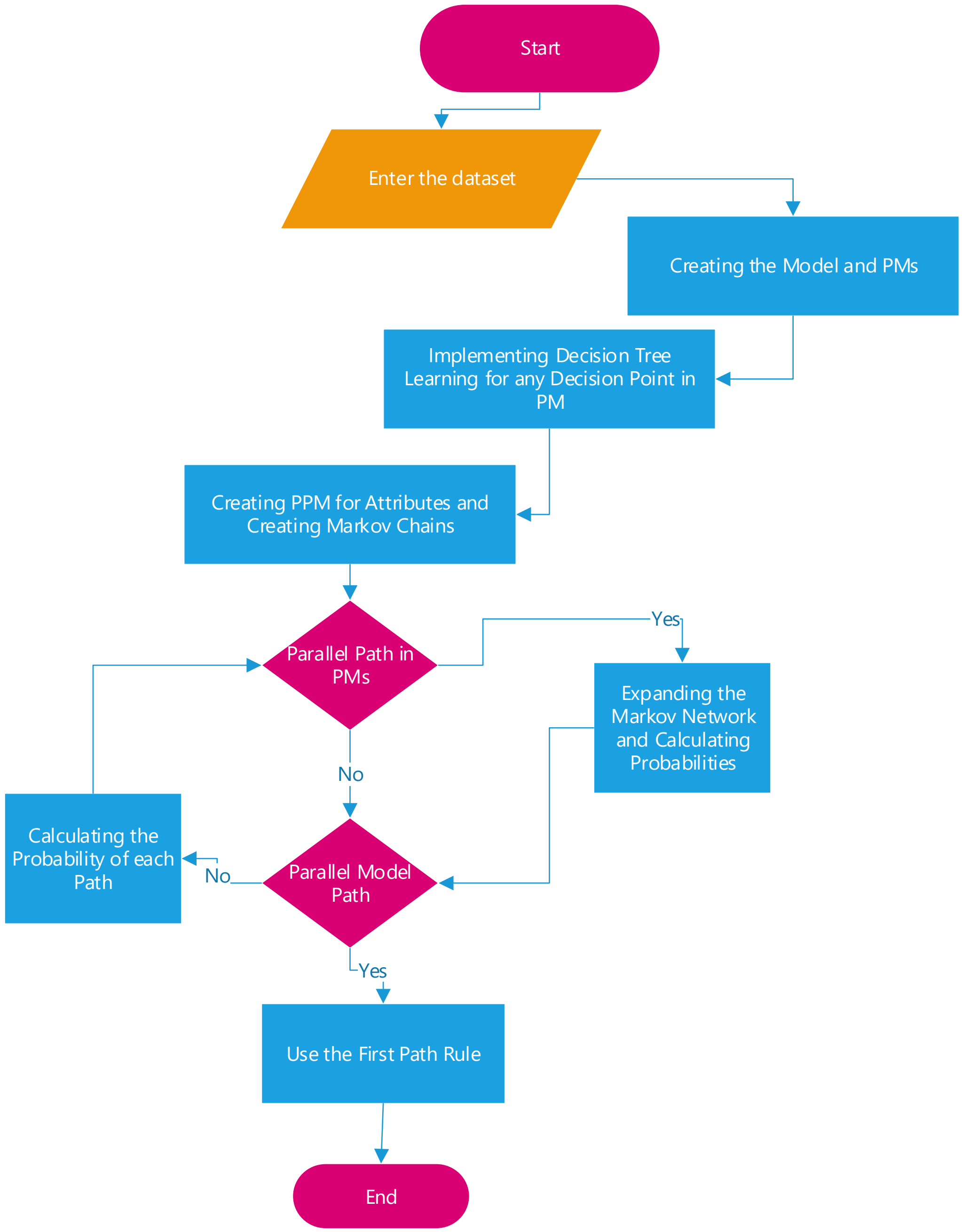

Figure 4. Markov chain formed by the proposed method

4.2.2. Prediction of the Resulting Values

By obtaining C_t, the value of h_t can be calculated by using the relations 8 and 9, which is actually the value of Ct.

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{8}$$

$$h_t = o_t tanh\,(C_t) \tag{9}$$

In relation 9, o_t shows the output gate and allows the *LSTM* cell to selectively transfer a part of the new state to the subsequent cells.

In Figure 5, a general overview of the actions in *LSTM* for predicting the situation can be seen.

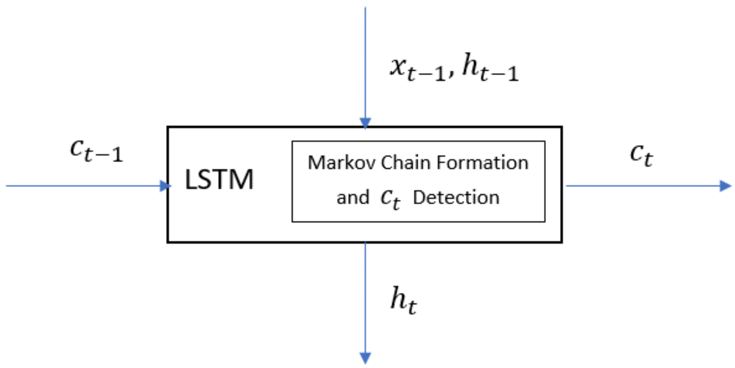

Figure 5. General Status of Detection Process in LSTM

4.3. Theoretical conclusions of the proposed method

The number of subdomains in BESTFIRT algorithm will be 2^n at worst state where *n* is the number of attributes.

Also, the number of necessary comparisons, at the worst state, for the algorithm of relation 10 is calculated as follows:

$$C = \sum_{k=1}^{n} \binom{n}{k} \tag{10}$$

where *k* is the number of the members of each subset at level *K*.

The BESTFIRST algorithm is a relatively fast algorithm in terms of time because it is a time complexity of less than O(n^2).

As it was seen earlier, the maximum number of comparisons for this algorithm will be *N* which is the number of samples in a dataset or records in a dataset.

The time complexity of this algorithm is O(n), which indicates that the CFSSUBSETSVAL algorithm will be considered as a relatively fast algorithm.

The Recurrent Neural Network (RNN) will have a *q* * *p* row. Given *q* as the number of input samples and *p* as the number of network layers and taking into account

$$n = Max\,(q, p) \tag{11}$$

in the worst case, $O(n^2)$ is the time that the method will spend.

Given the structure of the Markov chain creation, this algorithm, at its worst, will have a time equivalent to $O(n^2)$ because of *n* parallel paths created in the model.

Based on the earlier issues, in this section, it can be said that the time spent in the algorithm is ultimately a member of $O(n^2)$.

Since there is no return and stack effect in the proposed method, it can be stated that the amount of memory consumption is also a linear function.

Finally, it can be concluded that the proposed method requires processing resources at the level of a home PC at an intermediate data level and eliminates the resource challenge by using two methods of estimating the time complexity and memory complexity.

5. Implementation and Results

In this section, detection accuracy is determined by measuring the rate of correct diagnosis on all Iranian social security organization data. It is noteworthy that the test data are randomly selected from 10% of the total project data through k-fold strategy. The test data is processed in a CPU with 4 processing units with a maximum processing frequency of 2.2 MHz. The simulation environment is also created in MATLAB simulation software.

In order to test our hypotheses for imbalanced data, 3094 random data were collected from Covid-19 patients hospitalized in treatment centers of Tehran treatment management affiliated to Social Security Organization of Iran in 2020, which has led to recovery or death. The dataset includes string and numeric data as follows:

String data: gender, type of insurance, start date of hospitalization, end date of hospitalization, inpatient department, initial diagnosis, result of primary sampling, referral type, and final status of the patient.

Numeric data: ID, age, degree of fever, level of consciousness, respiratory distress status, cough, muscle pain.

The research dataset consists of 3 columns of general information about the patients, 13 columns related to the symptoms of disease and other specialized data of COVID-19 decease, and finally, one column as the result of treatment with the content of recovery or death of patients. The dataset has a maximum of 4 features after preprocessing operations.

Using the dataset and based on the mentioned materials, the proposed method was compared with [4], [1] and [6] in terms of the prediction accuracy, the time required for predicting and the amount of memory resource usage.

As inappropriate data is one of the challenges in predicting processes, we will pre-process the data to fit the data to obtain good quality data. In the proposed method, we will have two steps:

• Reduction of the number of data attributes: Remove attributes that were not effective in the resultant feedback and reduce them.

• Data normalization: All data in the research dataset are mapped to the [0 1] interval after normalization.

The total number of LSTMs generated is more than 715,000 LSTMs, which is equal to the total number of input units. Each LSTM has 2 outputs and 3 inputs as shown in Figure 3.

Each test was performed 10 times to measure each axis of each of the graphs in this section, and after the best and worst results were removed, the average reported results were recorded in the tables.

After performing the simulation with the mentioned conditions, amount of time consumed, memory resources and accuracy of the proposed method and the comparison with the three valid advanced methods are shown in Figures 6, 7 and 8.

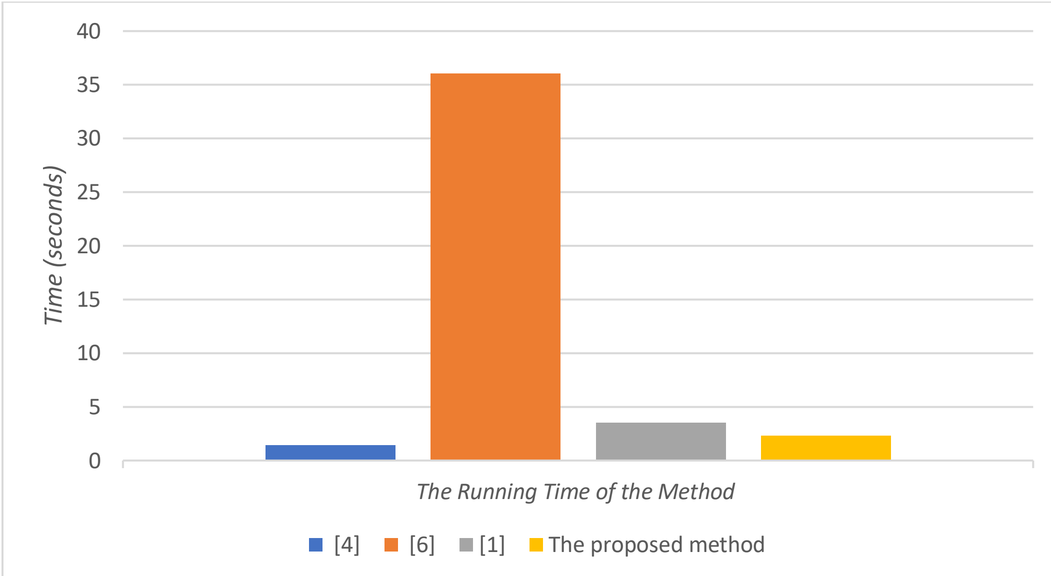

Figure 6. The amount of time consumed in the 4 compared methods

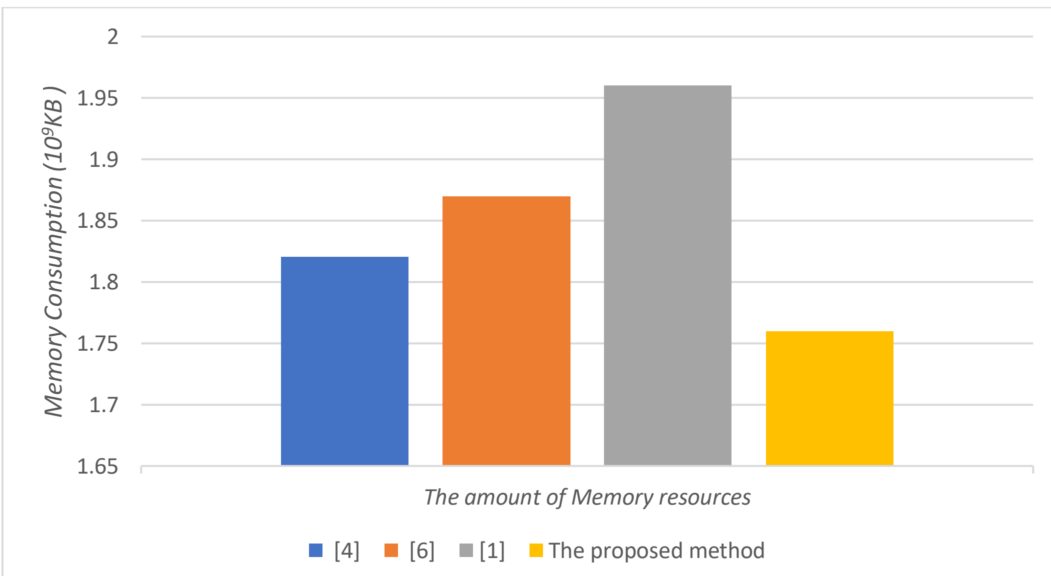

Figure 7. The amount of memory resources (RAM)

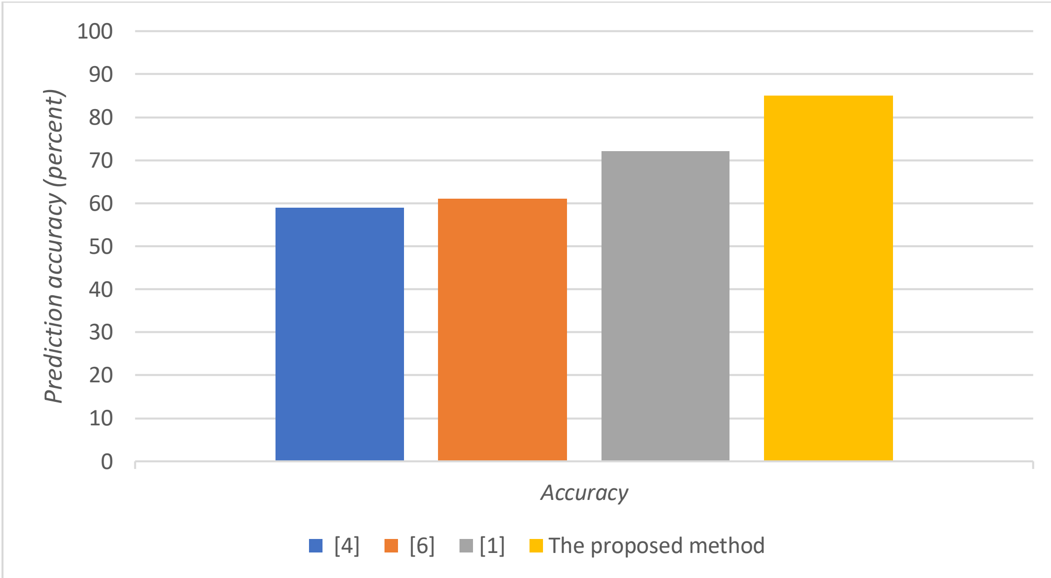

Figure 8. The accuracy of the methods

Figures 6, 7 and 8 show that the use of the proposed method has increased the accuracy.

By implementing and analyzing the results of the proposed method and method [60] with the data of patients suffering from Covid-19, comparing this method in the same conditions shown in Figures 9, 10 and 11, that is, with a significant reduction in memory resources and execution time of method, we attain the approximate precision of the method [60].

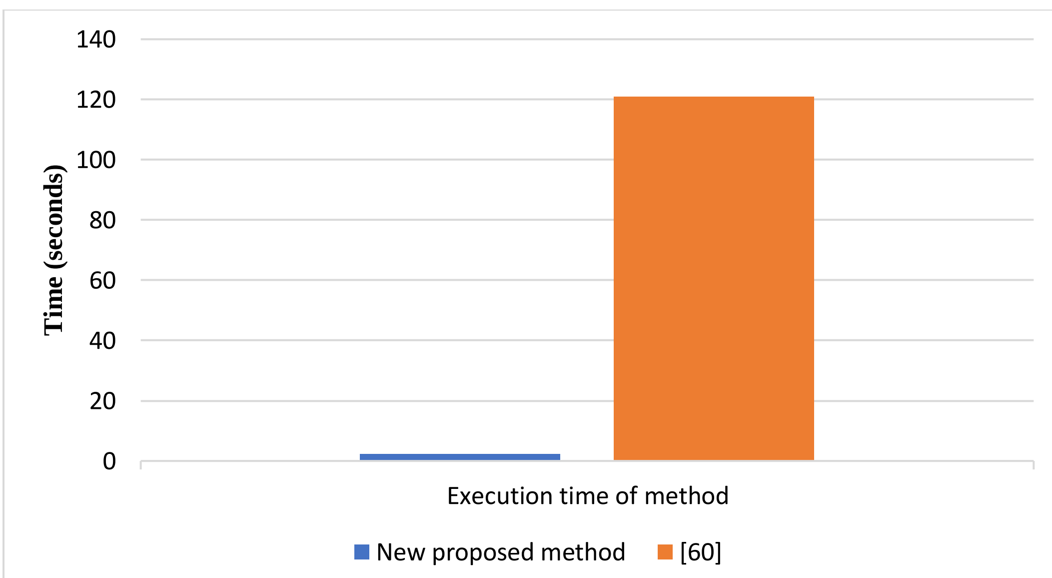

Figure 9. Comparison of consumed time in methods

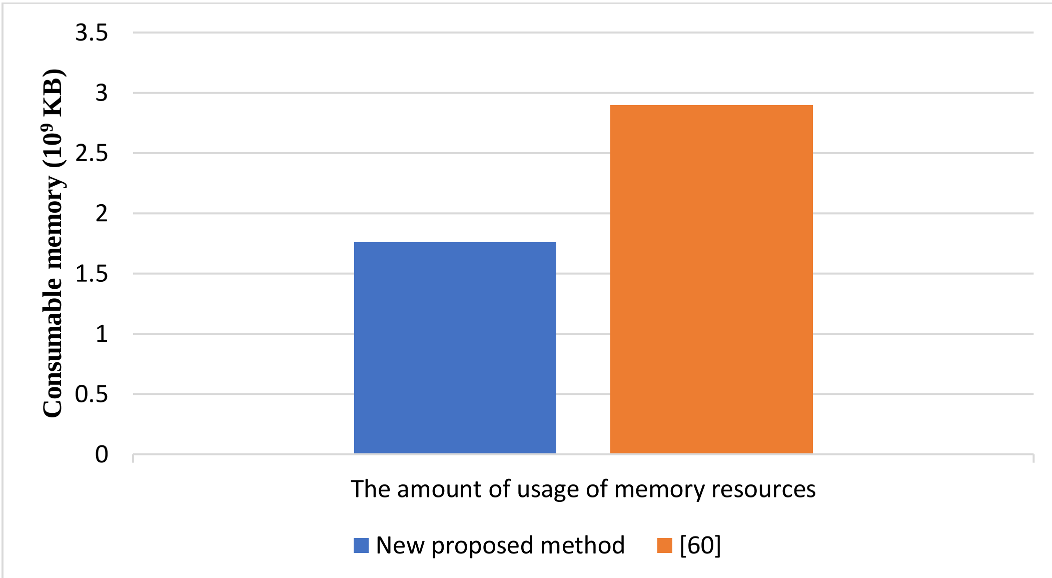

Figure 10. Comparison of the use of memory resources (RAM) in methods

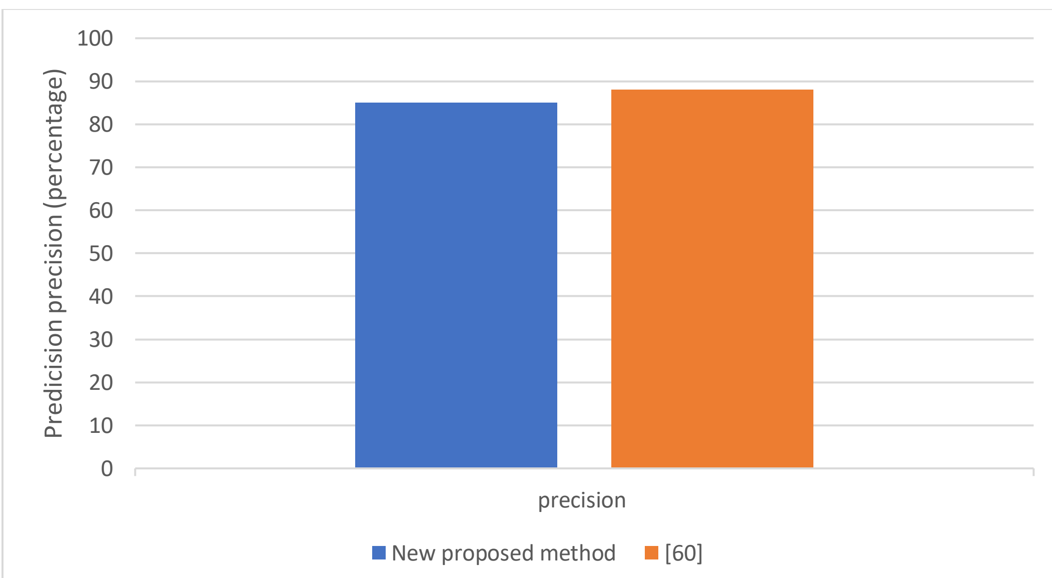

Figure 11. Comparison of precision in methods

6. Discussion and conclusion

This paper presented a new achievement for predicting the recovery or death of COVID-19 patients. Using the combination of deep learning, the recurrent neural network and Markov chain were applied to predict the results of heterogeneous data set in Covid 19. Also, the proposed method is simulated with real data of the Tehran treatment management from the Iranian Social Security Organization and the related graphs were observed.

Comparing a method against previous methods is common in order to show the performance of it, so the proposed method is simulated using the real data of COVID-19 patients who were hospitalized in treatment centers of Tehran treatment management affiliated to the Social Security Organization of Iran in 2020. The obtained results were compared with three valid advanced methods. The results showed that the proposed method significantly reduces the amount of memory resource usage and CPU usage time compared to similar methods, and at the same time, the accuracy also increases significantly.

In order to show the performance of each method, it is necessary to compare it with the previous methods. Therefore, the results were compared with three valid advanced methods. Accuracy rates were estimated up to 85% in this study and these show that they have a higher performance than other similar methods.

According to the results and contents of this study, the following is suggested for further studies:

- It can be tried to create a model that can predict data with the results other than zero and one.
- Some studies can be carried out on the methods which cover more attributes and, then the results can be compared with each other.
- Proposed algorithms can also be provided to reduce implementation costs.
- Prediction accuracy is very important in semi-structured processes. Therefore, proposing other methods to increase accuracy can have many applications.
- Research on methods that cover a larger number of features are also suggested.

Conflicts of Interest

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

Acknowledgment

We thank the moral support provided by the Education and Research Unit of the Human Resources Management of Iran's Social Security Organization.

References

[1] Aalst, W. V. D. (2016), Process mining: data science in action. *Springer-Verlag Berlin Heidelberg*, 167-177.

[2] Abidin, J. Z., Abdullah, N. A. H. and Khaw, K. L. H. (2021, January), Predicting business failure for Malaysia SMEs in the hospitality industry. In *Conference on International Issues in Business and Economics Research (CIIBER 2019)* (pp. 67-73). Atlantis Press.

[3] Brunk, J., Stierle, M., Papke, L., Revoredo, K., Matzner, M. and Becker, J. (2020), Cause vs. Effect in Context-Sensitive Prediction of Business Process Instances. *arXiv preprint arXiv:2007.07549*.

[4] Castellanos, M., Salazar, N., Casati, F., Dayal, U. and Shan, M. C. (2005), Predictive business operations management. In *Databases in Networked Information Systems: 4th International Workshop, DNIS 2005, Aizu-Wakamatsu, Japan, March 28-30, 2005. Proceedings 4* (pp. 1-14). Springer Berlin Heidelberg.

[5] Ceci, M., Lanotte, P. F., Fumarola, F., Cavallo, D. P. and Malerba, D. (2014), Completion time and next activity prediction of processes using sequential pattern mining. In *Discovery Science: 17th International Conference, DS 2014, Bled, Slovenia, October 8-10, 2014. Proceedings 17* (pp. 49-61). Springer International Publishing.

[6] Conforti, R., De Leoni, M., La Rosa, M. and Van Der Aalst, W. M. (2013), Supporting risk-informed decisions during business process execution. In *Advanced Information Systems Engineering: 25th International Conference, CAiSE 2013, Valencia, Spain, June 17-21, 2013. Proceedings 25* (pp. 116-132). Springer Berlin Heidelberg.

[7] Cook, J. E. and Wolf, A. L. (1998), Discovering models of software processes from event-based data. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, *7*(3), 215-249.

[8] Critical Capabilities for Composite Content Management Applications. Gartner, report. (2010)

[9] Datta, A. (1998). Automating the discovery of as-is business process models: Probabilistic and algorithmic approaches. *Information Systems Research*, *9*(3), 275-301.

[10] De Leoni, M. and Van der Aalst, W. M. (2014, September), The FeaturePrediction Package in ProM: Correlating Business Process Characteristics. In *BPM (Demos)* (p. 26).

[11] De Leoni, M., van der Aalst, W. M. and Dees, M. (2016), A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. *Information Systems*, *56*, 235-257.

[12] De Medeiros, A. K. A., Weijters, A. J. and van der Aalst, W. M. (2007), Genetic process mining: an experimental evaluation. *Data mining and knowledge discovery*, *14*, 245-304.

[13] De Paz, J. F., Bajo, J., González, A., Rodríguez, S. and Corchado, J. M. (2012), Combining case-based reasoning systems and support vector regression to evaluate the atmosphere–ocean interaction. *Knowledge and information systems*, *30*, 155-177.

[14] Di Francescomarino, C., Dumas, M., Federici, M., Ghidini, C., Maggi, F. M. and Rizzi, W. (2016), Predictive business process monitoring framework with hyperparameter optimization. In *Advanced Information Systems Engineering: 28th International Conference, CAiSE 2016, Ljubljana, Slovenia, June 13-17, 2016. Proceedings 28* (pp. 361-376). Springer International Publishing.

[15] Evermann, J., Rehse, J. R. and Fettke, P. (2016), A deep learning approach for predicting process behaviour at runtime. *Lecture Notes in Business Information Processing*, *281*, 327-338.

[16] Fakhfakh, M., Bouaziz, B., Gargouri, F. and Chaari, L. (2020), ProgNet: Covid-19 prognosis using recurrent and convolutional neural networks. *medRxiv*, 2020-05.

[17] Fischer, T. and Krauss, C. (2018), Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research*, *270*(2), 654-669.

[18] Folino, F., Guarascio, M. and Pontieri, L. (2015, September), A prediction framework for proactively monitoring aggregate process-performance indicators. In *2015 IEEE 19th International Enterprise Distributed Object Computing Conference* (pp. 128-133). IEEE.

[19] Geng, R., Bose, I. and Chen, X. (2015), Prediction of financial distress: An empirical study of listed Chinese companies using data mining. *European Journal of Operational Research*, *241*(1), 236-247.

[20] Ghattas, J., Soffer, P. and Peleg, M. (2014), Improving business process decision making based on past experience. *Decision Support Systems*, *59*, 93-107.

[21] Gibson, T. A., Henderson, J. A. and Wiles, J. (2014, July), Predicting temporal sequences using an event-based spiking neural network incorporating learnable delays. In *2014 International Joint Conference on Neural Networks (IJCNN)* (pp. 3213-3220). IEEE.

[22] Grigori, D., Casati, F., Castellanos, M., Dayal, U., Sayal, M. and Shan, M. C. (2004), Business process intelligence. *Computers in industry*, *53*(3), 321-343.
[23] Grigori, D., Casati, F., Dayal, U. and Shan, M. C. (2001, September), Improving business process quality through exception understanding, prediction, and prevention. In *VLDB* (Vol. 1, pp. 159-168).
[24] Heaton, J. B., Polson, N. G. and Witte, J. H. (2017), Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry*, *33*(1), 3-12.
[25] Herbst, J. (2000), A machine learning approach to workflow management. In *Machine Learning: ECML 2000: 11th European Conference on Machine Learning Barcelona, Catalonia, Spain, May 31–June 2, 2000 Proceedings 11* (pp. 183-194). Springer Berlin Heidelberg.
[26] Herbst, J. and Karagiannis, D. (1998, August), Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models. In *Proceedings of the 9th International Workshop on Database and Expert Systems Applications* (p. 745).
[27] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.
[28] Huang, X. and An, A. (2002, December), Discovery of interesting association rules from livelink web log data. In *2002 IEEE International Conference on Data Mining, 2002. Proceedings.* (pp. 763-766). IEEE.
[29] Huck, N. (2009), Pairs selection and outranking: An application to the S&P 100 index. *European Journal of Operational Research*, *196*(2), 819-825.
[30] Huser, V. (2012), Process mining: Discovery, conformance and enhancement of business processes, 45, 1018-1019.
[31] Iwendi, C., Bashir, A. K., Peshkar, A., Sujatha, R., Chatterjee, J. M., Pasupuleti, S., ... and Jo, O. (2020), COVID-19 patient health prediction using boosted random forest algorithm. *Frontiers in public health*, *8*, 357.
[32] Jassim, M. A., and Abdulwahid, S. N. (2021, March), Data Mining preparation: Process, Techniques and Major Issues in Data Analysis. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1090, No. 1, p. 012053). IOP Publishing.
[33] Jensen, K. (1997). *Colored Petri nets: basic concepts, analysis methods and practical use* (Vol. 3). Springer Science & Business Media.
[34] Jiang, X. R., and Gruenwald, L. (2005), Microarray gene expression data association rules mining based on BSC-tree and FIS-tree. *Data & Knowledge Engineering*, *53*(1), 3-29.
[35] Jiang, X., Pan, S., Long, G., Xiong, F., Jiang, J. and Zhang, C. (2018), Cost-sensitive parallel learning framework for insurance intelligence operation. *IEEE Transactions on Industrial Electronics*, *66*(12), 9713-9723.
[36] Jozefowicz, R., Zaremba, W. and Sutskever, I. (2015, June), An empirical exploration of recurrent network architectures. In *International conference on machine learning* (pp. 2342-2350). PMLR.
[37] Kang, B., Kim, D. and Kang, S. H, (2011), Periodic performance prediction for real-time business process monitoring. *Industrial Management* and *Data Systems*, 112(1), 4–23.
[38] Kang, B., Kim, D. and Kang, S. H. (2012), Real-time business process monitoring method for prediction of abnormal termination using KNNI-based LOF prediction. *Expert Systems with Applications*, *39*(5), 6061-6068.
[39] Khan, A., Le, H., Do, K., Tran, T., Ghose, A., Dam, H. and Sindhgatta, R. (2018), DeepProcess: Supporting business process execution using a MANN-based recommender system. *arXiv preprint arXiv:1802.00938*.

[40] Khan, A., Le, H., Do, K., Tran, T., Ghose, A., Dam, H. and Sindhgatta, R. (2018), Memory-augmented neural networks for predictive process analytics. *arXiv preprint arXiv:1802.00938*.

[41] Kim, A., Obregon, J. and Jung, J. Y. (2014), Constructing decision trees from process logs for performer recommendation. In *Business Process Management Workshops: BPM 2013 International Workshops, Beijing, China, August 26, 2013, Revised Papers 11* (pp. 224-236). Springer International Publishing.

[42] Klees, P. H. and van Zelst-Tutor, S. J. (2017), The Application of Comparative Process Mining in Logistics.

[43] Kraus, M. and Feuerriegel, S. (2017), Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems*, *104*, 38-48.

[44] Lakshmanan, G. T., Duan, S., Keyser, P. T., Curbera, F. and Khalaf, R. (2011), Predictive analytics for semi-structured case oriented business processes. In *Business Process Management Workshops: BPM 2010 International Workshops and Education Track, Hoboken, NJ, USA, September 13-15, 2010, Revised Selected Papers 8* (pp. 640-651). Springer Berlin Heidelberg.

[45] Lakshmanan, G. T., Shamsi, D., Doganata, Y. N., Unuvar, M. and Khalaf, R. (2015), A markov prediction model for data-driven semi-structured business processes. *Knowledge and Information Systems*, *42*, 97-126.

[46] Le, M., Gabrys, B., and Nauck, D. (2012). A hybrid model for business process event prediction. In *Research and Development in Intelligent Systems XXIX: Incorporating Applications and Innovations in Intelligent Systems XX Proceedings of AI-2012, The Thirty-second SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence* (pp. 179-192). Springer London.

[47] LeCun, Y., Bengio, Y. and Hinton, G. (2015), Deep learning. *nature*, *521*(7553), 436-444.

[48] Leontjeva, A., Conforti, R., Di Francescomarino, C., Dumas, M. and Maggi, F. M. (2015), Complex symbolic sequence encodings for predictive monitoring of business processes. In *Business Process Management: 13th International Conference, BPM 2015, Innsbruck, Austria, August 31--September 3, 2015, Proceedings 13* (pp. 297-313). Springer International Publishing.

[49] Letham, B., Rudin, C. and Madigan, D. (2013), Sequential event prediction. *Machine learning*, *93*, 357-380.

[50] Lin, L., Wen, L. and Wang, J. (2019, May), Mm-pred: A deep predictive model for multi-attribute event sequence. In *Proceedings of the 2019 SIAM international conference on data mining* (pp. 118-126). Society for Industrial and Applied Mathematics.

[51] Liu, S., Duffy, A. H., Whitfield, R. I. and Boyle, I. M. (2010), Integration of decision support systems to improve decision support performance. *Knowledge and Information Systems*, *22*, 261-286.

[52] Maggi, F. M., Di Francescomarino, C., Dumas, M. and Ghidini, C. (2014), Predictive monitoring of business processes. In *Advanced Information Systems Engineering: 26th International Conference, CAiSE 2014, Thessaloniki, Greece, June 16-20, 2014. Proceedings 26* (pp. 457-472). Springer International Publishing.

[53] Márquez-Chamorro, A. E., Resinas, M., and Ruiz-Cortés, A. (2017). Predictive monitoring of business processes: a survey. *IEEE Transactions on Services Computing*, *11*(6), 962-977.

[54] Mazen, T. S. (2020), A Novel Machine Learning based Model for COVID-19 Prediction. *International Journal of Advanced Computer Science and Applications*, *11*(11).

[55] Metzger, A., Leitner, P., Ivanović, D., Schmieders, E., Franklin, R., Carro, M., ... and Pohl, K. (2015), Comparing and combining predictive business process monitoring

techniques. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *45*(2), 276-290.

[56] Min, W., Liang, W., Yin, H., Wang, Z., Li, M. and Lal, A. (2021), Explainable deep behavioral sequence clustering for transaction fraud detection. *arXiv preprint arXiv:2101.04285*.

[57] Moon, J., Park, G.. and Jeong, J. (2021), Pop-on: Prediction of process using one-way language model based on nlp approach. *Applied Sciences*, *11*(2), 864.

[58] Natarajan, S., Tadepalli, P. and Fern, A. (2011), A relational hierarchical model for decision-theoretic assistance. Knowl Inf Syst (KAIS),1–21.

[59] Neu, D. A., Lahann, J. and Fettke, P. (2021), A systematic literature review on state-of-the-art deep learning methods for process prediction.

[60] PellehShahi, M. F., Kordrostami, S., Sheikhani, A. R. and Faridi, M. (2020), Predicting the recovery of COVID-19 patients using recursive deep learning. *Iranian Journal of Operations Research*, *11*(2), 48-64.

[61] Peng, N., Zhou, X., Niu, B., and Feng, Y. (2021), Predicting fundraising performance in medical crowdfunding campaigns using machine learning. *Electronics*, *10*(2), 143.

[62] Pfeffer, A. (2005, July), Functional specification of probabilistic process models. In *AAAI* (pp. 663-669).

[63] Pinsky, M. and Karlin, S. (1998), *An introduction to stochastic modeling*. Academic press.

[64] Poh, K. L. (2000), An intelligent decision support system for investment analysis. *Knowledge and Information Systems*, *2*, 340-358.

[65] Ross, S. M. (2003), Introduction to probability models.

[66] Rozinat, A., Wynn, M. T., van der Aalst, W. M., ter Hofstede, A. H. and Fidge, C. J. (2009), Workflow simulation for operational decision support. *Data & Knowledge Engineering*, *68*(9), 834-850.

[67] Santiputri, M., Ghose, A. K. and Dam, H. K. (2017), Mining task post-conditions: Automating the acquisition of process semantics. *Data & Knowledge Engineering*, *109*, 112-125.

[68] Schmidhuber, J. (2015), Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

[69] Schonenberg, H., Weber, B., Van Dongen, B. and Van der Aalst, W. (2008), Supporting flexible processes through recommendations based on history. In *Business Process Management: 6th International Conference, BPM 2008, Milan, Italy, September 2-4, 2008. Proceedings 6* (pp. 51-66). Springer Berlin Heidelberg.

[70] Sebu, M. L. and Ciocârlie, H. (2016, May), Optimizing business processes by learning from monitoring results. In *2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI)* (pp. 43-50). IEEE.

[71] Sindhgatta, R., Ghose, A. and Dam, H. K. (2016), Context-aware analysis of past process executions to aid resource allocation decisions. In *Advanced Information Systems Engineering: 28th International Conference, CAiSE 2016, Ljubljana, Slovenia, June 13-17, 2016. Proceedings 28* (pp. 575-589). Springer International Publishing.

[72] Singh, A., Gupta, P., and Thakur, N. (2021). An empirical research and comprehensive analysis of stock market prediction using machine learning and deep learning techniques. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1022, No. 1, p. 012098). IOP Publishing.

[73] Sirignano, J. A. (2019), Deep learning for limit order books. *Quantitative Finance*, *19*(4), 549-570.

[74] Sirignano, J., Sadhwani, A. and Giesecke, K. (2016), Deep learning for mortgage risk. *arXiv preprint arXiv:1607.02470*.

[75] Unuvar, M., Lakshmanan, G. T. and Doganata, Y. N. (2016), Leveraging path information to generate predictions for parallel business processes. *Knowledge and Information Systems*, *47*, 433-461.

[76] Van Der Aalst, W. (2011), *Process mining: discovery, conformance and enhancement of business processes* (Vol. 2). Heidelberg: Springer.

[77] Van der Aalst, W. M. (2011), Using Process Mining to Bridge the Gap between BI and BPM. *Computer*, *44*(12), 77-80.

[78] Van Der Aalst, W. M., Reijers, H. A., Weijters, A. J., van Dongen, B. F., De Medeiros, A. A., Song, M. and Verbeek, H. M. W. (2007), Business process mining: An industrial application. *Information systems*, *32*(5), 713-732.

[79] Van der Aalst, W. M., Schonenberg, M. H. and Song, M. (2011), Time prediction based on process mining. *Information systems*, *36*(2), 450-475.

[80] Van der Aalst, W., Adriansyah, A. and Van Dongen, B. (2012), Replaying history on process models for conformance checking and performance analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(2), 182-192.

[81] van der Aalst, W., Adriansyah, A., de Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., ... and Wynn, M. (2012), Process Mining Manifesto. *LECTURE NOTES IN BUSINESS INFORMATION PROCESSING*, *99*, 169-194.

[82] Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., ... and Wynn, M. (2012), Process mining manifesto. In *Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I 9* (pp. 169-194). Springer Berlin Heidelberg.

[83] Van der Aalst, W., Weijters, T. and Maruster, L. (2004), Workflow mining: Discovering process models from event logs. *IEEE transactions on knowledge and data engineering*, *16*(9), 1128-1142.

[84] Van Dongen, B. F., and Van der Aalst, W. M. (2004). Multi-phase process mining: Building instance graphs. In *Conceptual Modeling–ER 2004: 23rd International Conference on Conceptual Modeling, Shanghai, China, November 8-12, 2004. Proceedings 23* (pp. 362-376). Springer Berlin Heidelberg.

[85] Van Mourik, J. A. J. (2015), Comparing business processes using event data a visual approach. *University of Technology*.

[86] Vanderfeesten, I., Reijers, H. A. and Van der Aalst, W. M. (2011), Product-based workflow support. *Information Systems*, *36*(2), 517-535.

[87] Yang, Y., Kolesnikova, A., Lessmann, S., Ma, T., Sung, M. C. and Johnson, J. E. (2018), Can deep learning predict risky retail investors? A case study in financial risk behavior forecasting. *arXiv preprint arXiv:1812.06175*.

[88] Zenie, A. (1985). Colored stochastic petri nets. In: Proceedings of the international workshop on timed petri nets, IEEE Computer Society Press, Torino, 262–271.

[89] Zhu, W. D., Becker, B., Boudreaux, J., Baman, S., Gomez, D., Marin, M. and Vaughan, A. (2000), Advanced case management with IBM case manager, IBM redbooks.