Iranian Journal of Operations Research
Vol. 11, No. 2, 2020, pp. 48-64
DOI:

Predicting the recovery of COVID-19 patients using recursive deep learning

M. Fadaei PellehShahi[1], S. Kordrostami[2]*, A. H. Refahi Sheikhani[3], M. Faridi Masouleh[4] , S. Shokri[5]

In this study, an alternative method is proposed based on recursive deep learning with limited steps and prepossessing, in which the data is divided into A unit classes in order to change a long short term memory and solve the existing challenges. The goal is to obtain predictive results that are closer to real world in COVID-19 patients. To achieve this goal, four existing challenges including the heterogeneous data, the imbalanced data distribution in predicted classes, the low allocation rate of data to a class and the existence of many features in a process have been resolved. The proposed method is simulated using the real data of COVID-19 patients hospitalized in treatment centers of Tehran treatment management affiliated to the Social Security Organization of Iran in 2020, which has led to recovery or death. The obtained results are compared against three valid advanced methods, and are showed that the amount of memory resources usage and CPU usage time are slightly increased compared to similar methods and the accuracy is increased by an average of 12%.

Keywords: *Long Short Term Memory, Recurrent Deep Learning, Prediction, COVID-19, Neural Network.*
Manuscript was received on 07/23/2020, revised on 10/15/2020 and accepted for publication on 11/27/2020.

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 [60].
Event prediction is a task to predict future events based on events that have already occurred [15]. 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. A higher levels of data

[1] Ph.D. Student of Computer Science, Department of Mathematics, Lahijan Branch, Islamic Azad University, Lahijan, Iran, Mehrdad.fadaei95@stumail.liau.ac.ir.
* 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] Department of Computer and Information Technology, 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.

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 training data, which is a problem for machine learning models [2]. Deep learning methods have provided excellent results in programs such as computer vision, language processing, and so on [50]. The result of effective feature learning based on deep learning in [28] creates applications that rely on unstructured data.

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 [22].

Using social security operations arising from different data of artificial intelligence techniques is still very difficult for most social insurance organizations. In the following, four key challenges that place much greater emphasis on the application of artificial intelligence and deep learning techniques in social insurance operations are listed:

- **Existence of many features in an event that makes it difficult to record and use it as a systematic set in an artificial intelligence structure**

Social insurance organizations collect a wide range of information from insured persons and available services to control risk and provide better services. More than a dozen features can be provided in the datasets, common dimensions and specifications of the insured person, and service-related items.

In social insurance datasets, data quality and density can be quite evident. Selecting the most effective features of the goal of a social insurance business along with effective connections for optimal data processing is very challenging. For example, the ratio of age and insurance records may be an important factor in liability risk analysis, while the amount of premium paid is one of the most important features involved in calculable risk analysis. The effective use of few features that have the strongest connections among the high-dimensional features is a very important factor in the analysis of social insurance data.

- **Heterogeneous data structure**

Heterogeneity is one of the key features in social insurance data in order to meet different needs all over a social insurance organization. Different heterogeneous social insurance data can be divided into two components as shown in Figure 1. The main features of social insurance do not change over time, and the type of requests and services that occur over the life cycle of an insured person are recorded.

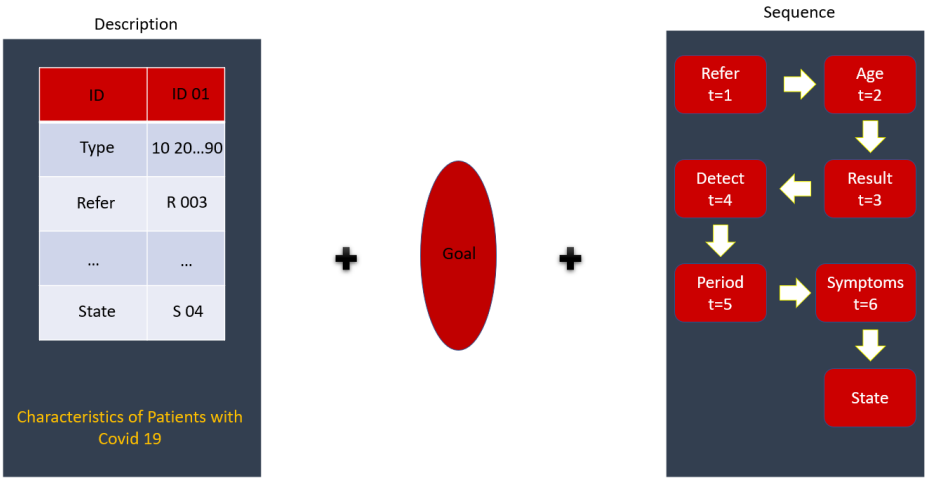

Figure 1. An example of a heterogeneous data set in COVID-19

Heterogeneity of social insurance data is divided into two main levels: data-level and structure-level [40, 19, 9]. Data-level heterogeneity involves different types of data (for example, data structures that contain different types of data such as integer and character), while structure-level heterogeneity involve a combination of different types of data formats and data sources (such as a complex data structure that combines static features in order to describe features and dynamic services for time series activities). Heterogeneous data usually occupy different positions in the data space, so using a deep learning method or a prediction to extract patterns from heterogeneous data does not provide useful comparisons [30, 48, 16]. How to use this heterogeneous information effectively in social insurance is a very important and practical question.

- **Low rate of assigning each data to a class**

Compared to other financial business processes such as banking, in which customer bank account transactions are usually performed in a large number and customer behaviors are tracked by creating long sequences, the frequency of requests and services provided to insured persons in social insurance is much lower. In social insurance, the request or physical presence or direct remote affairs may be made only once or twice a year, and the provision of services in a life cycle such as the requests of the insured persons also occurs at a low frequency. However, it may be overlooked that tracking a request or providing services is important for social insurance operations, and that the insured person's contact with social insurance organizations indicates an important behavior such as paying premiums, requesting services, and so on. Therefore, in-depth analysis of data related to the transactions of the insured persons provides a more accurate insight into social insurance.

- **Imbalanced distribution of data in predicted classes**

Social insurance operations focus on minority classes rather than balanced classifications with traditional technologies. Definable risks in social insurance are usually events that occur with low probability, which means that the business objectives analyzed usually focus on the minority classification in social insurance. Compared to the traditional balanced classification, we have an imbalanced distribution of data in the predicted classes [17, 44].

In order to take advantage of artificial intelligence, social insurance requires an effective machine learning framework to overcome the challenges and create meaningful relationships in social insurance operations. Compared to traditional machine learning, deep learning approaches have some advantages and are able to extract features and nonlinear relationships without relying on econometric assumptions and human expertise [58, 61, 37, 59, 38].

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 [55], 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.
The rest of the paper is organized as follows: A summary of some basic concepts in process prediction and deep learning in semi-structured businesses is presented in Section 2. Section 3 presents the research background on process prediction, and in particular the prediction of subsequent events in a process. Section 4 describes the methods used in this study. Details of the implementation and evaluation of the experimental results are reported in Section 5, where the proposed method is compared with three advanced methods on the real dataset of the Tehran treatment management from Iranian social security organization. Finally, Section 6 gives the findings and identifies open challenges in this field.

2. Basic concepts

Social insurances, pension funds, as well as commercial insurances, have many differences in terms of details, mechanisms and business models, but all of them have been formed to deal with the crises ahead and have common end goals. Therefore, the correct prediction is very important in insurance businesses, especially social insurances, and will play a very important role in the sustainability of the business model. In this section, some useful concepts in this field are explained.

2.1. Deep learning

The goal of deep learning is to learn different levels of data representation. Higher levels represent more abstract concepts. A deep architecture with multiple abstract layers and its ability to learn distributed representations offers many advantages over conventional shallow machine learning methods [60].
Machine learning methods learn a functional relationship between variables that determine the relationship between observation and a predicted objective. The high diversity of this function complicates the machine learning method and may lead to poor generalization. Learning theory demonstrates that to represent a functional relationship, a learning machine with depth k exponentially needs more computational units than a machine with depth k + 1 [32]. The depth of machine learning methods is usually as follows:
Depth 1 includes linear and logistic regression.
Depth 2 contains decision trees, artificial neural networks with one hidden layer, and support vector machines
Depth 5 to 10 incorporates the visual system in the human brain [53].
The concept of depth in many relevant empirical findings, for example artificial neural networks or support vector machines, includes explanations and outcomes that work better than simple regression models [26]. Increasing depth allows these methods to implicitly learn a higher level of data representation. Additional levels facilitate the generalization to new combinations of features that are less shown in the training data. This increased capacity also allows the learning machine to obtain more changes in the objective function that accurately separates classes. In addition, the number of computational units that a model can use is severely limited by the number of training examples. As a result, when there are different interests in the objective function, shallow architectures require large number of computational units to fit the function. Therefore, they need more training examples than deep models [2].

2.2. Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM)

LSTM is a special type of RNN that is capable of learning long term dependencies . LSTMs are resistant to vanishing gradient and are specifically designed to avoid long-term dependencies [24].
RNN is not suitable for the use of sigmoid cells, and this leads to LSTM. A basic LSTM cell is defined as follows: It receives Ct-1 and ht-1 respectively as state and the input data from the primary opened cell at the same level, and receives Xt as input from cells of the previous layers. In turn, it transfers Ct and ht to the sequence opened cell and provides ht as the output to the next layer [10].

3. Research background

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 is very successful when the challenges or problems are properly detected and sorted [34].
Most prediction methods deal with process results rather than predicting the next event in the process. Most of the reviewed results are the time left to complete a case. Dongen et al. presented an approach on event repetitions, event time, and case data by using incremental regression [8]. Pandey et al. applied a hidden Markov model in event and runtime sequences [39]. Like van der Aalst et al.'s approach is based on an annotated transition system [57]. Folino et al. have presented two cases that use cluster tree and Finite State Machines (FSM) to predict the time remaining for a running case [13, 14]. Schwegmann et al. have reported the development of a software tool that utilizes Complex Event Processing (CEP) for event sequences and has been trained to predict their future behavior [51]. Rogge-Solti et al. have used two approaches of stochastic Petri net simulation for the same purpose [45, 46]. Bevacqua et al. have put forward a clustering and regression-based prediction approach on case data [3]. Bolt et al. applied the clustered approach to partial and complementary cases [4]. Polato et al. have proposed two approaches based on annotated transition systems, as well as support vector regression and naïve Bayes classifiers [42, 43].
Orman et al. [10] have developed a model for predicting the behavior of running processes that uses deep learning and, in particular, recurrent neural networks with LSTM cells to predict the next event in a running process, using the comparison of natural language processing.
A technique based on DBNs is propoed in [5] that is capable of context-sensitive process prediction. Their 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 their technique, they 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 ofDBNs. 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 [5].
In [35], 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 review [35] laid on a qualitative comparison of existing implementations. In particular, the literature is classified along the dimensions, including neural network type, prediction type, input features and encoding methods.
In [33], the efficiency of a one-way language model approach among fully attention-based transformer models is investigated for predicting future process events of the current process instance being executed.

[29] 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. Their contribution includes:

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

• Their 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 [29].

In [12], a method is suggested 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 COVID-19 prognosis.

Algorithms for extracting correlation rules [27] have been successful for applications in many scenarios, such as extracting gene expression data [23], data mining and blog analysis [20]. In [31], Fin Deep Behavior Cluster has been proposed as 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. [1] attempt to develop and compare the performance of the MLP model against the logit model 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. [41] 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 was employed in the training stage, and parameters were 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.

Izadi et al. [21] applied fuzzy Delphi to extract the necessary data for classification plans due to the impotance of the investigation of the problem of lack of required data.

Singh et al. [52] tested all parameters under consideration and analysed 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. Using algorithms in stock prediction has proven to be essential and has thus marked their application in strong market plans.

Only five approaches [56, 6] are related to subsequent event prediction, many of which use explicit representation of process models such as Hidden Morkov Model (HMM) and Probabilistic Finite Automatons (PFA). 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 [25].

Semi-structured processes are widely created in industries such as government industry, insurance, banking and health care [7]. 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 [62].

Each prediction problem requires the creation of a new decision tree in relation with the training dataset to predict the output class [47]. Some possible approaches aiming at modeling and extracting business processes are Markov models, models presented by stochastic process modeling techniques for finite state machine production, and random graph-based models [18].

In [22], 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 [50, 28]. Deep recurrent networks are a generalization of three-layer networks that have been widely used in previous works [36]. The architecture of deep neural network proposed in [36] is different. It includes multiple layers in different types of units and relies on unsupervised pretraining to extract predictive features. Pretraining factors offer different benefits and have been effective in financial programs [45].

Extended deep learning methods have been implemented in predicting the behavior of retail investors in the broad market in [60]. Yang et al. examined the effect of deep learning on management support, focused on predicting financial risk-related behaviors, and developed a DNN-based risk management system [60].

Fadaei Pellehshahi et al. [11] presented a method to predict insurance systems. A combination of deep learning technique, specifically the recursive neural network, and Markov chain have been applied to the problem of predicting outcomes in an insurance process. The proposed method was also simulated with real data of the Social Security Organization. The method increased the use of memory resources compared to the Markov method to some extent, but the CPU usage time was decreased significantly compared to the Markov and recursive neural network methods, and also, the accuracy and efficiency were improved using their method.

Sayed et al. [49] presented a model to predict different levels of severity risks for the COVID-19 patients using machine learning techniques. Their approach was based upon X-ray images. Shastri et al. [53] introduced deep learning based comparative analysis of COVID-19 cases in India and USA and the data sets of veified and death cases of COVID-19 were addressed.

Nevertheless, in this research, predicting the recovery or death of COVID-19 patients is investigated using a substitute approach. Actually, an approach based on recurrent deep learning with limited steps and preprocessing are provided while the data are decomposed into A part classes to alter a long short-term memory and clarify existing challenges.

4. Methodology

The presented method in this study is based on recursive deep learning with limited steps and preprocessing. The method is proposed to divide the data into A part classes to change a long short-term memory and solve existing challenges. The social insurance industry faces four major challenges in using artificial intelligence:

- Heterogeneous data
- Imbalanced distribution of data in the predicted classes
- Low rate of assigning each data to a class

- Existence of many features in an event that makes it difficult to record and use it as a systematic set in an artificial intelligence

Given that this insurance data has little reproducibility, which means that, for example, it may only be taken from customers or business owners once a year, the challenges of this type of data will become more apparent.

Traditional machine learning algorithms usually show justifiable performance only in standard datasets and are more suitable for uniform and balanced data. Utilizing deep learning facilitates this process to some extent, given the potential of these methods and some of these challenges, but designing a problem-solving method with a deep learning network requires care with the resources consumed and the efficiency of using them. In this paper, we present a method based on recursive deep learning with limited steps and preprocessing.

The proposed recurrent neural network is a type of multilayer online networks. In this type of multilayer network, each level communicates online with nodes of both L-th layer and (L+1)-th layer. This network processes and delivers new output at the same time as it receives each input. Like all recurrent neural networks, it has an internal state, and at each step, in addition to the new input, it also refers to that state for decision making. This state will be updated with each new input.

To solve the existing challenges, we need to make changes in the structure of a recurrent network. Since the input data have an imbalance in their structure, a pretraining phase must be performed on the data before the data enters the training process. This event (pretraining) takes place in each LSTM. This phase is added to LSTM for two reasons:

- Increasing the separation rate in order to increase the difference in the rate of assigning each data to a class
- Reducing the amount of resource consumption by reducing the calculations

If we can find a model in which each interval of input data is describable, then we can say that the two goals for the pretraining phase have been achieved.

The first step in the proposed method for changing an LSTM is to divide the data into A part classes. The size of class Ai depends on the size of the data and should not be greater than 20% of the total data. In this paper, by dividing the data into equal classes, we calculate the length of the class Ai and create a descriptive binomial distribution model of all the expected features for the data in the class Ai with the least possible error in pretraining. The sum of these distributions is actually a polynomial of P variables (P is the number of features considered for a data) that provides a description of how the data will change based on its model. Figure 2 shows an example of the values available for the data in an assumed class Ai.

	P_1	P_2	P_3	....	P_p
D_1	X_1	--	X_2	....	--
D_2	--	X_3	X_4	....	X_5
....	....	....	....	....	....
D_n	--	X_1	---	.....	x_p

Figure 2. A schematic of the data in the class Ai, emphasizing that this data may not be the same in the number of constructing features

Then, a state change function will encrypt any data in Ai by using Equation 1 as follows:

$$y = \sigma(w\tilde{x} + b) \quad (1)$$

In this equation, y is the value of the given state change, w is the weight considered for the recurrent neural network in a cell (computational weight of each bilstm) and b is its bias value. Also,

$\tilde{x} = |x - Binomial(x)|$ is the value obtained from the binomial distribution error used in the step before pretraining in the LSTM.

An estimator function will be responsible for changing the state of y value to reach the desired estimation value in the proposed model for class Ai. The estimator function can be seen in Equation 2.

$$z = \sigma(\tilde{w}y + \tilde{b}) \quad (2)$$

In this equation, $\tilde{w}$ and $\tilde{b}$ are the values of the weight and bias of the estimator relation. Once the value of z is obtained, we replace this value with x, which is the original data, and again, we create the binomial distribution of features by the new x. In Figure 3, the replacement cycle of these values in a process is illustrated.

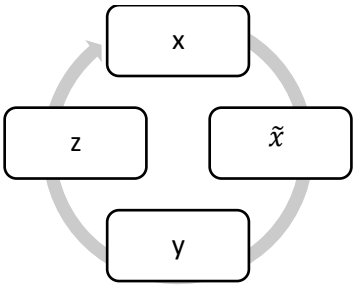

Figure 3. Replacement cycle of variables in pretraining

This leads to the adjustment and creation of a model to describe the changes in the data. To accurately adjust the variables, *w* using Equation 3 is calculated.

$$w = \frac{-1}{N}\sum_{i=1}^{N}\sum_{j=1}^{P} \log(z_{ij} + (1 - x_{ij}))\log(1 - z_{ij}) \quad (3)$$

In fact, this equation is an adaptation of the average slope change rate for the two equations. In Equation 3, N is the number of data examined, z_{ij} is the value of estimated data for i-th data in the j-th property, and x_{ij} is the original value of data at each step of the training. Also, the new value of b is obtained from Equation 4:

$$b_{new} = \frac{w_{new}z}{wx + b} \quad (4)$$

In fact, w and b are two polynomials with P variables. This process is done until the expected error value of an LSTM is less than the defined value.
After the pre-training of each LSTM, Equation 5 can be used as follows in order to estimate the assignment rate of i-th data to j-th class.

$$L_{ij} = \frac{e^{w_i x + b}}{\sum_{j=1}^{N} e^{w_j x + b_j}} \quad (5)$$

In this equation, w_i is the weight and b is the bias in the network for data x in its latest state, w_j is the weight and b_j is the bias for data x in each of the Lstm of the relevant layer. After reviewing the data and estimating the outcome, the values of estimation error E_i for each of the Lstm are recalculated by using Equation 6.

$$E_i = \sum_{j=1}^{N} \log\frac{L_{jd}}{L_{jr}} \quad (6)$$

In this equation, L_{ir} is the estimation for the class of which j-th data is actually a member, and L_{jd} is the estimation for the class that the network has recognized that j-th data is a member of

it. The network will continue to operate as long as the available weights for each Lstm can guide the Ei value toward a range commensurate with the work objectives. Depending on the deviation of Ei, the initial values of w and b used in related adaptive network will be adjusted and modified. Actually, if we want to show the proposed method as a process, we can use Figure 4.

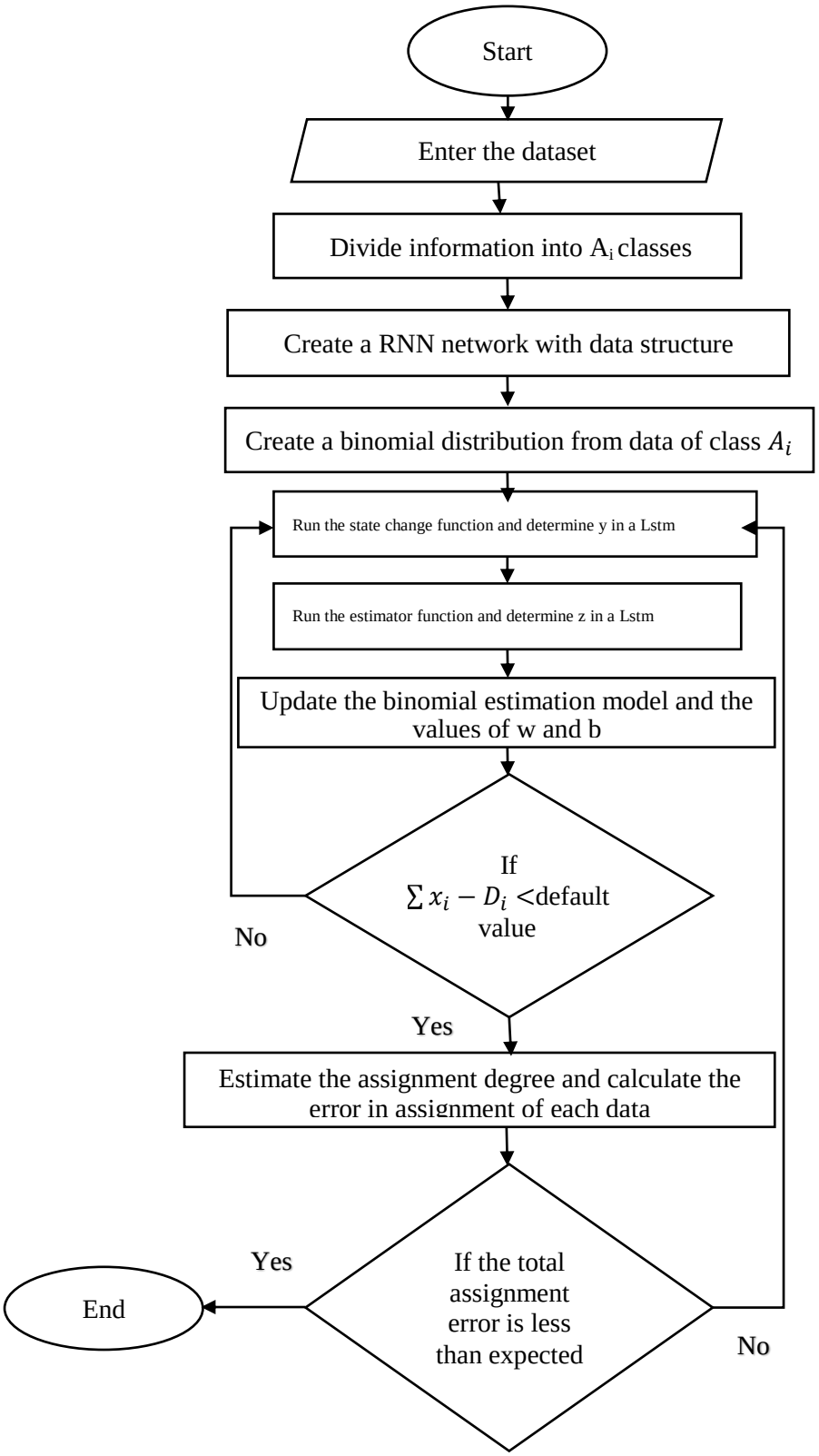

Figure 4. Algorithm of the proposed method in improving the separability of imbalanced data

5. Implementation, results and discussion

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 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.

Nevertheless, we confronted with some limitations such as the extraction of some data related to COVID-19 due to privacy and confidentiality. Also, we can refer to the limited availability of some resources related to the topic under investigation as another limitation.

Using the dataset and based on the mentioned materials, the proposed method was compared with [60], [22] and [10] in terms of the prediction accuracy, the time required for predicting and the amount of memory resources usage.

The accuracy of diagnosis is determined by measuring the number of correct diagnoses on all test data. It should be noted that the test data are selected randomly from 20% of the total project data by using k-fold strategy. To eliminate the effect of the data, the studied methods were performed 10 times and the average of the outcomes of all runs were selected for the final outcome. The unit of time required for processing is second. The amount of processing resources consumed are calculated based on the CPU involvement in the processing process in terms of (unit of process/unit of time), and the amount of memory resources consumed from RAM is calculated in KB.

The data were processed using a CPU with 4 processing units and a maximum processing frequency of 2.2 MHz. Also, the simulation environment was created in MATLAB 2018 software.

The total number of LSTMs created is more than 715,000 LSTMs, which is equal to the total number of input numeric units. Each LSTM has 2 outputs and 3 inputs.

Each test was performed 10 times in order to measure each point of the axis of the diagrams presented in this section, and after removing the best and worst outcomes, the average amounts of outcomes reported in the tables were recorded. In addition, the default values of some variables affecting the simulations are shown in Table 1.

Table 1. The values of used variables with initialization

Row	Variable	Description	Default Value
1	μ	Adjuster of increased or decreased epochs	0.01
2	w	Computational weight of each biLSTM	Initial value of 1
3	b	The bias of each biLSTM	Random initial value
4	C_t	Output/input status	Initial value of 1
5	h_t	Output/input	Initial value of 1
6	R_t	Maximum number of comparisons in a step to obtain the best C_t	100
7	W	The weight of the cost, resources and accuracy factors	0.33
8	Removing unit	Maximum removal of units from the layer at the beginning of processing	10
9	Hidden Layers Unit	Maximum number of units in hidden layers	100
10	number of iterations	Maximum number of iterations without change in biLstm	10

11	Epoch	Maximum epochs	100
12	batch size	Minimum batch size	4

After performing the simulation with the mentioned conditions, the amount of time consumed, memory resources and the accuracy of the proposed method and also the comparison with the three valid advanced methods are shown in Figures 5, 6 and 7.

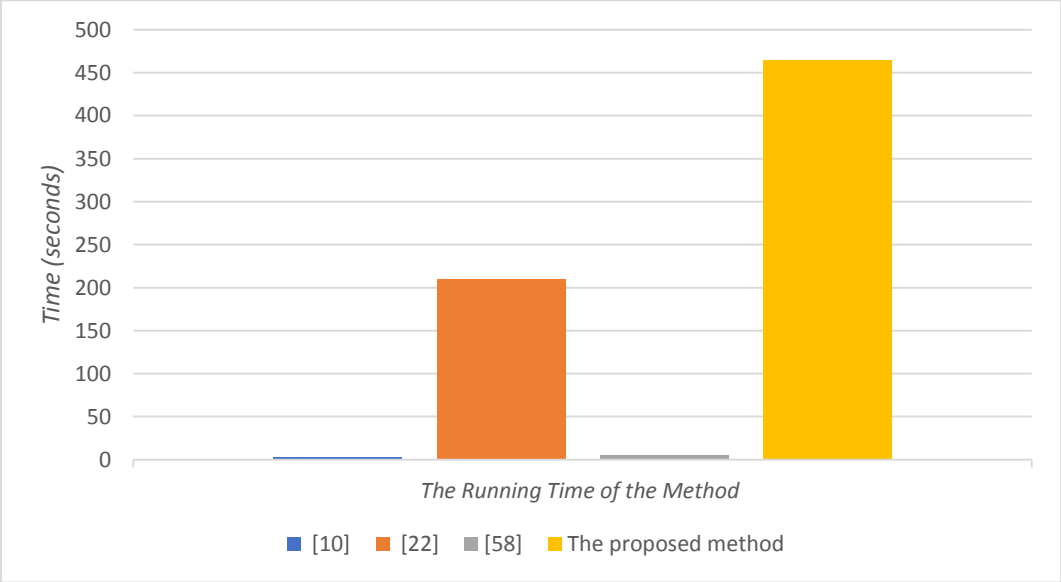

Figure 5. The amount of time consumed in the poposed approach and the methods compared

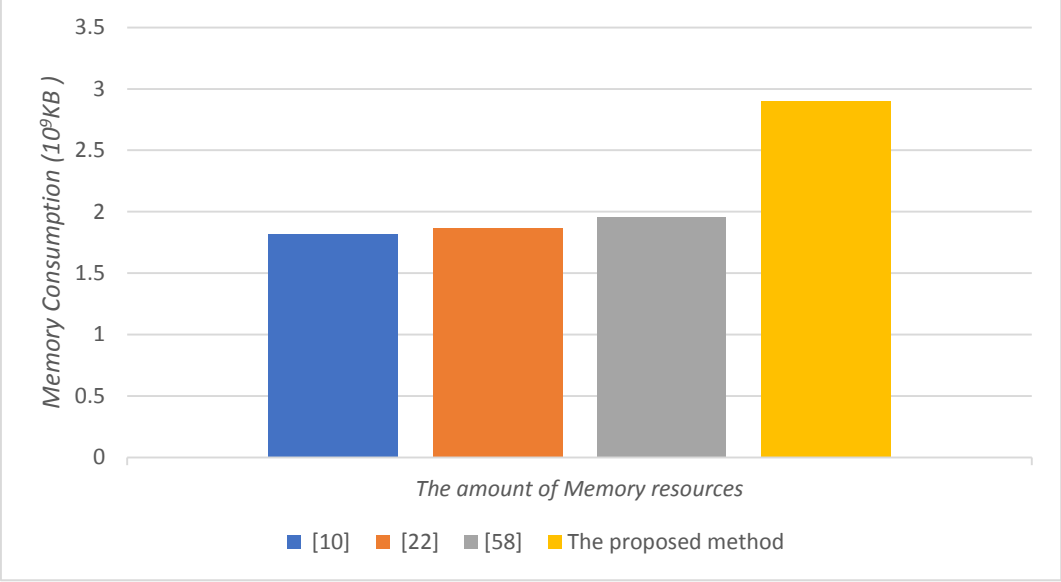

Figure 6. The amount of memory resources (RAM)

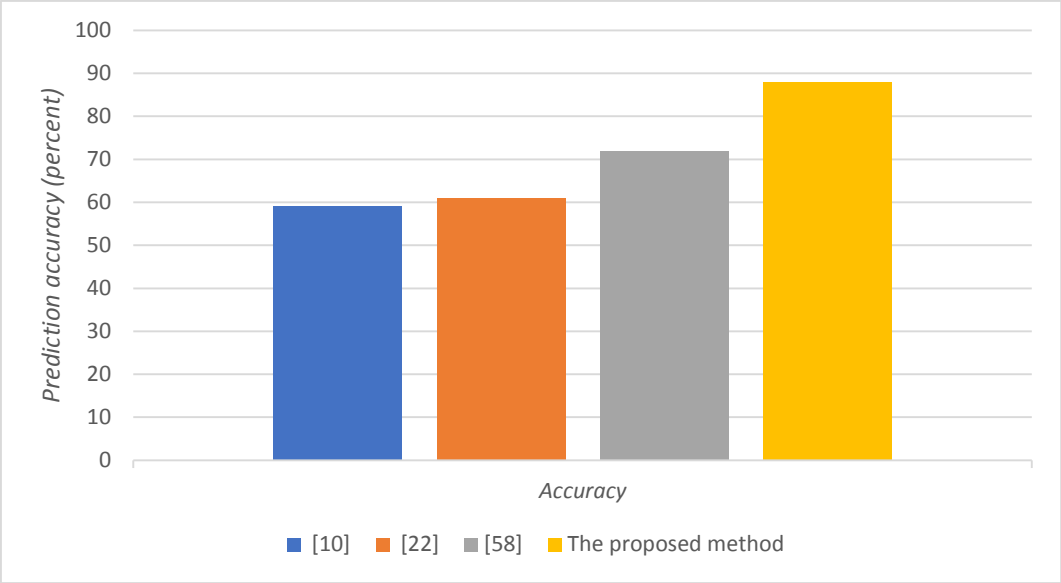

Figure 7. The accuracy of the methods

Figures 5, 6 and 7 show that by applying the proposed method the accuracy, the amount of time consumed and RAM increase in comparison with other approaches under examination.
Due to the high accuracy of the method, in hospitalized patients, patients at higher risk can be identified much earlier and favorable conditions can be provided for those patients to reduce mortality and, accordingly, increase recovery.

6. Conclusions

This paper has presented a new achievement for predicting the recovery or death of COVID-19 patients. To solve the existing challenges in the problem of predicting outcomes in an insurance process, including many features, heterogeneous data, imbalanced distribution and low rate of assigning data, a method based on recurrent deep learning with limited steps and preprocessing was presented. In this method, the data were divided into A part classes to change a long short-term memory and solve existing challenges.
In the multilayer network used in this study, each level communicates online with nodes of both L-th layer and (L+1)-th layer, and processes and delivers new output at the same time as it receives each input. Similar to all recurrent neural networks, it has an internal state, and at each step of decision making, it refers to that state in addition to the new input, so that the state is updated with each new input.
Since the input data from heterogeneous data set in COVID-19 has an imbalance in its structure, a pretraining is performed on the data before it enters the training process. The pretraining has been performed on the data in each LSTM. The research data set included general information and clinical symptoms of a person with COVID-19, covered by Social Security Insurance in Iran, and finally a column related to the outcome of treatment by patient recovery or death. Using the proposed method, symptoms with higher risk can be identified and the maximum effort and ability of the treatment staff can be used to control the symptoms and ultimately improve patients.
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 hospitalized in treatment centers of Tehran treatment management affiliated to the Social Security Organization of Iran in 2020, which has led to recovery or death. The obtained results were compared against three valid advanced methods, and showed that the amount of memory resources usage and CPU usage time are slightly increased compared to similar methods. However, the use of the proposed method leads to a significant increase in accuracy.

In the proposed method, depending on the number of data sets used in training, the error reduction trend occurs in a smaller number of data. This error reduction trend, despite the existence of unbalanced data, shows that this method requires less data than similar methods. In other words, the method has better performance and flexibility in data size in training phase than other scientific methods, and the limitation of the method is only for data with 0 and 1 results.
According to the results obtained in this study, suggestions for future research are as follows:

- developing a model that can predict data with outcomes other than 0 and 1.
- Providing a model that can create a suitable pattern and make the relevant predictions just by receiving different business data.
- Using other artificial intelligence methods to improve accuracy, given that predictive accuracy is very important in business processes.
- Proposed algorithms can also be provided to reduce implementation costs.
- Research on methods that cover a larger number of features are also suggested.

Acknowledgment

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

References

1. Abidin, J.Z., Hiauabdullah, N.A., Lee, K. and Khaw, H. (2021), Predicting Business Failure for Malaysia SMEs in the Hospitality Industry, *proceedings of the conference on International Issues in Business and Economics Research*, http:// creativecommons. org/ licenses/ by-nc/ 4.0/0.
2. Bengio, Y. (2009), Learning deep architectures for AI, Foundations and trends R in Machine Learning, 2 (1), 1–127.
3. Bevacqua, A., Carnuccio, M., Folino, F., Guarascio, M. and Pontieri, L. (2014), A datadriven prediction framework for analyzing and monitoring business process performances. *Hammoudi, S., Cordeiro, J., Maciaszek, L.A., Filipe, J. (eds.) ICEIS 2013*. *LNBIP*, vol. 190, 100–117. Springer, Cham. doi:10.1007/978-3-319-09492-2 7.
4. Bolt, A. and Sep´ulveda, M. (2014), Process remaining time prediction using query catalogs. *Lohmann, N., Song, M., Wohed, P. (eds.) BPM 2013*. *LNBIP*, vol. 171, 54–65. Springer, Cham. doi:10.1007/978-3-319-06257-0 5.
5. Brunk, J., Stierlr, M., Papke, L., Revoredo, K., Matzner, M. and Becker, J. (2020), Cause vs. Effect in Context-Sensitive Prediction of Business ProcessInstances, *Information systems*. Doi.org/10.1016/j.is.2020.101635.
6. 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, Dˇzeroski, S., Panov, P., Kocev, D., Todorovski, L. (eds.) DS 2014. *LNCS (LNAI)*, vol. 8777, 49–61. Springer, Cham doi:10.1007/978-3-319-11812-3 5.
7. Critical Capabilities for Composite Content Management Applications. Gartner, report. (2010).
8. Dongen, B.F., Crooy, R.A. and Aalst, W.M.P. (2008), Cycle time prediction: When will this case finally be finished? *Meersman, R., Tari, Z. (eds.) OTM 2008. LNCS*, vol.5331, 319–336. Springer, Heidelberg, doi:10.1007/978-3-540-88871-0 22.
9. Emmert-Streib, F., de Matos Simoes, R., Glazko, G., McDade, S., Haibe-Kains, B., Holzinger, A., Dehmer, M. and Campbell, F. C. (2014), Functional and genetic analysis of the colon cancer network, *BMC bioinformatics*, vol. 15, no. 6, p. S6.
10. Evermann, J., Rehse, J.R. and Fettke, P. (2016), A deep learning approach for predicting process behavior at runtime, I*nternational Conference on Business Process Management, Springer*, 327-338.

11. Fadaei Pellehshahi, M., Kordrostami, S., Refahi Sheikhani, A.H. and Faridi Masouleh, M. (2021), Predicting business processes of the social insurance using recurrent neural network and Markov chain, *Journal of Modelling in Management*, Vol. ahead-of-print No. ahead-of-print. https:// doi.org/10.1108/JM2-04-2021-0105.
12. Fakhfakh, M., Gargouri, F., Bouaziz, B. and Chaari, L. (2020), Prognet:covid-19 prognosis using recurrent and convolutional neural networks, ResearchGate, doi: 10.1101/ 2020.05.06.20092874.
13. Folino, F., Guarascio, M. and Pontieri, L. (2013), Context-aware predictions on business processes: an ensemble-based solution, *Appice, A., Ceci,M., Loglisci, C., Manco,G., Masciari, E., Ras, Z.W. (eds.) NFMCP 2012. LNCS (LNAI)*, vol. 7765, 215–229. Springer, Heidelberg. doi:10.1007/978-3-642-37382-4 15.
14. Folino, F., Guarascio, M. and Pontieri, L. (2012), Discovering context-aware models for predicting business process performances, *Meersman, R., et al. (eds.) OTM* 2012. LNCS, vol. 7565, 287–304**.** Springer**,** Heidelberg. doi:10.1007/978-3-642-33606-5 18.
15. Geng, R., Bose, I. and Chen, X. (2015), Prediction of financial distress: An empirical study of listed Chinese companies using data mining, E*uropean Journal of Operational Research*, 241 (1), 236–247.
16. Globerson, A., Chechik, G., Pereira, F. and Tishby, N. (2006), Embedding heterogeneous data using statistical models, *in Proceedings of The National Conference on Artificial Intelligence*, vol. 21, no. 2, 1605.
17. He, H. and Garcia, E. A. (2009), Learning from imbalanced data, *IEEE Transactions on knowledge and data engineering*, vol. 21, no. 9, 1263–1284.
18. Herbst, J. (2000), A machine learning approach to workflow management. *ECML*, 183–194.
19. Hu, R., Yu, C. P., Fung, S.-F., Pan, S., Wang, H. and Long, G. (2017), Universal network representation for heterogeneous information networks, *in 2017 International Joint Conference on Neural Networks (IJCNN)*, 388–395.
20. Huang, X. and An, A. (2002), Discovery of interesting association rules from livelink web log data, *Data Mining. ICDM 2003. Proceedings. 2002 IEEE International Conference on, IEEE*, 763-766.
21. Izadi, B., Ranjbarian, B., Ketabi, S. and Nassiri-Mofakham, F. (2013), Multi-Group Classification Using Interval Linear Programming, *Iranian Journal of Operations Research*, Vol. 4, No. 1, 2013, pp. 55-74.
22. Jiang, X., Pan, S., Long, G. Xiong, F., Jiang, J. and Zhang, C. (2018), Cost-sensitive parallel learning framework for insurance intelligence operation, *Transactions on Industrial Electronics*, 1–11.
23. Jiang, X. R. and Le, G. (2005), Microarray gene expression data association rules mining based on bsc-tree and _s-tree, *Data & Knowledge Engineering*, 53, 3-29.
24. Jozefowicz, R., Zaremba, W. and Sutskever, I. (2015), An empirical exploration of recurrent network architectures, *International Conference on Machine Learning*, 2342-2350.
25. Jozefowicz, R., Zaremba, W. and Sutskever, I. (2015), An empirical exploration of recurrent network architectures, *International Conference on Machine Learning*, 2342-2350.
26. Lessmann, S., Baesens, B., Seow, H.V. and Thomas, L. C. (2015), Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research, *European Journal of Operational Research*, 247 (1) , 124–136.
27. Letham, B., Rudin, C. and Madigan, D. (2013), Sequential event prediction, *Machine learning*, 93, 357-380.
28. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y. and Alsaadi, F. E. (2017), A survey of deep neural network architectures and their applications, *Neurocomputing*, 234 , 11–26.
29. lwendi, C., Bashir, A.K., Peshkar, A., Sujatha, R., Chatterjee, J.M., Pasupuleti, S., Mishra, R., Pillai, S. and Jo, O. (2020), Covid-19 patient healt orediction using boosted

random forest algorithm, *Frontiers in Public Health*, 8:357.doi: 10.3389/ fpubh. 2020.00357.
30. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J. (2013), Distributed representations of words and phrases and their compositionality, *in Advances in neural information processing systems*, 3111–3119.
31. Min, W., Liang, W., Yin, H., Wang, Z., Li, M. and Lal, A. (2021), Explainable Deep Behavioral Sequence Clustering for Transaction Fraud Detection, *arxiv: 2101*, 04285V1 [CS.LG] 12.
32. Montufar, G. F., Pascanu, R., Cho, K. and Bengio, Y. (2014), On the number of linear regions of deep neural networks, *Advances in Neural Information Processing Systems*, 2924–2932.
33. Moon, J., Park, G. and Jeong, J. (2021), POP-ON: Prediction of Process Using One-Way Language Model Based on NLP Approach, *Applied Sciences*. https://doi. org/10.3390/ app11020864.
34. Mustafa Abdalrassual Jassim and Sarah N. Abdulwahid (2021), Data Mining preparation: Process, Techniques and Major Issues in Data Analysis, *IOP Conf. Ser*:. *Mater. Sci. Eng.* 1090 012053.
35. Neu, A., lahann, J. and fettke, P. (2021), A systematic literature review on state-of-the-art deep learning methods for process prediction, *arxiv:2101*,09320V2[CS.LG]26.
36. Oztekin, A., Kizilaslan, R., Freund, S. and Iseri, A. 2016, A data analytic approach to forecasting daily stock returns in an emerging market, *European Journal of Operational Research*, 253 (3), 697–710.
37. Pan, S., Hu, R., Long, G., Jiang, J., Yao, L. and Zhang, C. (2018), Adversarially regularized graph autoencoder, *arXiv preprint arXiv*:1802.04407.
38. Pan, S., Wu, J., Zhu, X., Zhang, C. and Wang, Y. (2016), Tri-party deep network representation, *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 1895–1901.
39. Pandey, S., Nepal, S. and Chen, S. (2011), A test-bed for the evaluation of business process prediction techniques, *7th International Conference on Collaborative Computing:Networking, Applications and Worksharing, CollaborateCom*, Orlando,FL, USA, 382–391.
40. Pavlidis, P., Weston, J., Cai, J. and Grundy, W. N. (2001), Gene functional classification from heterogeneous data, *in Proceedings of the fifth annual international conference on Computational biology*, 249– 255.
41. Peng, N., Zhou, X., Niu, B. and Feng, Y. (2021), Predicting Fundraising Performance in Medical Crowdfunding Campaigns Using Machine Learning, *electronics*, https://doi.org/ 10.3390/ electronics 10020143.
42. Polato, M., Sperduti, A., Burattin, A. and de Leoni, M. (2014), Data-aware remaining time prediction of business process instances. *International Joint Conference on Neural Networks, IJCNN 2014*, Beijing, China, July 6–11, 2014, 816–823.
43. Polato, M., Sperduti, A., Burattin, A., de Leoni, M. (2016), Time and activity sequence prediction of business process instances. *CoRR abs*/1602.07566.
44. Provost, F. (2000), Machine learning from imbalanced data sets 101, *in Proceedings of the AAAI 2000 workshop on imbalanced data sets*, 1–3.
45. Rogge-Solti, A. and Weske, M. (2013), Prediction of remaining service execution time using stochastic petri nets with arbitrary firing delays. *Basu, S., Pautasso, C., Zhang,L., Fu, X. (eds.) ICSOC 2013*. LNCS, vol. 8274, pp. 389–403. Springer, Heidelberg. doi:10.1007/978-3-642-45005-1 27.
46. Rogge-Solti, A. and Weske, M. (2015), Prediction of business process durations using nonmarkovian stochastic petri nets. *Inf. Syst.* 54, 1–14.
47. Ross, S. (2003), Introduction to probability models, 8th edn, Chap. 4.
48. Rudolph, M., Ruiz, F., Mandt, S. and Blei, D. (2016), Exponential family embeddings, *in Advances in Neural Information Processing Systems*, 478–486.

49. Sayed, S. A. F., Elkorany, A. M., & Mohammad, S. S. (2021). Applying Different Machine Learning Techniques for Prediction of COVID-19 Severity. *IEEE Access*, *9*, 135697-135707.
50. Schmidhuber, J. (2015), Deep learning in neural networks: An overview, *Neural Networks*, 61 ,85–117.
51. Schwegmann, B., Matzner, M. and Janiesch, C. (2013), preCEP: facilitating predictive eventdriven process analytics, *Brocke, J., Hekkala, R., Ram, S., Rossi, M. (eds.) DESRIST 2013. LNCS*, vol. 7939, 448–455. Springer, Heidelberg. doi: 10. 1007/ 978-3-642-38827-9 36.
52. Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U. and Poggio, T. (2007), A quantitative theory of immediate visual recognition, *Progress in brain research*, 165, 33–56.
53. Shastri, S., Singh, K., Kumar, S., Kour, P., & Mansotra, V. (2020). Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. *Chaos, Solitons & Fractals*, 140, 110227.
54. 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, *IOP conf.series:materials Science and Engineering*, doi: 10. 1088/ 1757-899x/ 1022/1/ 012098.
55. Tamer Mazen, Sh. (2020) A nivel machine learning based model for covid-19 prediction, *International Journal of Advanced Computer Science and Applications*, Vol.11, No.11**.**
56. Unuvar, M., Lakshmanan, G.T. and Doganata, Y.N. (2016), Leveraging path information to generate predictions for parallel business processes, *Knowl. Inf. Syst*, 47(2), 433–461.
57. van der Aalst, W.M.P., Schonenberg, M.H. and Song, M. (2011), Time prediction based on process mining, *Inf. Syst.* 36(2), 450–475.
58. Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C. and Xu, W. (2016), Cnn-rnn: A unified framework for multi-label image classification, *in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2285–2294. IEEE.
59. Wang, C., Pan, S., Long, G., Zhu, X. and Jiang, J. (2017), MGAE: Marginalized graph autoencoder for graph clustering, *in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 889–898.
60. Yang, Y., Kolesnikova, A., S. Lessmann, Ma, T., Sung, M.-C. and Johnson, J. E. V. (2018), Can deep learning predict risky retail investors? A case study in Financial risk behavior forecasting, *arXiv:1812.06175*.[Online]. Available: https://arxiv. org/abs/1812. 06175.
61. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. and Hovy, E. (2016), Hierarchical attention networks for document classification, *in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1480– 1489.
62. Zhu, WD., Becker, B., Boudreaux, J., Baman, S., Gomez, D., Marin, M. and Vaughan A. (2000), Advanced casemanagement with IBMcasemanager, *IBMredbooks*, http://www. redbooks. ibm. com/ abstracts/sg247929.html?Open.