Heart Disease Predictions Using Numerous Classification Techniques and Dynamic LSTM Model

Conference Paper · July 2019

DOI: 10.1109/ICCES45898.2019.9002054

2 authors, including:

Aishwarya Mishra
Malaviya National Institute of Technology Jaipur
6 PUBLICATIONS **16** CITATIONS

SEE PROFILE

Proceedings of the Fourth International Conference on Communication and Electronics Systems (ICCES 2019)
IEEE Conference Record # 45898; IEEE Xplore ISBN: 978-1-7281-1261-9

Heart Disease Predictions Using Numerous Classification Techniques and Dynamic LSTM Model

Aishwarya Mishra
Information Technology and Computer Applications.
MMM University of Technology,
Gorakhpur (U.P)-273010, India
aishwaryamishra17@yahoo.com

Dayashankar Singh
Information Technology and Computer Applications
MMM University of Technology,
Gorakhpur (U.P)-273010, India
dsscs@mmmut.ac.in

***Abstract*— Now-a-days, a lot of clinical data are used for utilizing the intellectual advancements of clinical basis leadership. In the area of health sector, the need for improving the nature of patient's lives and decreasing the expenses as well as measures the work engaged with their everyday medicinal services condition is very important. An Electronic health record (EHRs) is a data record set which resides patient diagnostic lab records with different parameters, physician/consultants' histories, also in the accounts of different sections of the hospital. Earlier, in the case of heart disease prediction, we have attained a bulk unorganized set of data since the time series of the EHR system. By dividing and taking out these time-sensitive EHRs information, we can distinguish/perceive the aggregate factors among all the obtained set of clinical data for our research purpose. Despite the fact that it is troublesome undertaking to utilize the current EHR information legitimately, on the grounds that they might be repetitive, not in a standardized structure, and some of the data missed/blanked. Consequently, this paper has been proposed a viable and solid design of the Dynamic LSTM model for coronary illness forecast. The main aim of doing this paper has been used to predict heart sickness utilizing an alternate information mining arrangement system. In this paper, we have designed a Dynamic LSTM using classification techniques and generated a training and test model which produce a more accurate result as compare to previously LSTM risk algorithm. One of the advantages of the proposed algorithm is it can work on dynamic dataset records and capable to produce a more efficient classified result for heart disease prediction.**

Keywords—Electronic health record, Deep learning, Dynamic LSTM

I. INTRODUCTION

Heart disease/ Coronary illness is a popular term that implies that the heart isn't used regularly. Children can be brought into the world with heart-related problems. This is called inherent heart disorder [1]. Many of the heart/coronary related diseases can be acquired. Coronary illness, which incorporates ischemic coronary illness, coronary heart disappointment and different sicknesses of the heart, is the main source of all around. It represents the greater part of all s because of cardiovascular illness.

Coronary illness can cause angina (chest torment), heart assaults (myocardial localized necrosis) and unexpected heart failure. A heart assault is typically the principal indication of coronary illness. A Heart disease ordered into three classes:

- Coronary artery disease – This disease occurs when veins are not passing on blood properly in the heart.
- Congestive heart failure – This disease occurs when a heart is not pumping properly.
- Bad heart rhythms – This is a very rare heart disease. This disease defines the electronic pulse like the movement of the heart.

CVDs are the number 1 explanation behind at around the globe. A bigger number of individuals die every year from CVDs than any other reason. A normal 17.9 million people passed on from CVDs in 2016, addressing 31% of each and every overall demise. Of these deaths, 85% is a result of heart strike and stroke. Out of the 17 million startling misfortunes (more youthful than 70) on account of non-transmittable diseases in 2015, 82% are in low-and focus pay countries, and 37% are realized by CVDs. Most cardiovascular ailments can be evaded by keeping an eye on social danger factors, for instance, tobacco use, bothersome eating routine and weight, physical inaction and destructive use of alcohol using people wide philosophies [5][6].

Peoples/Individuals with a cardiovascular ailment or who are at high cardiovascular hazard (because of the nearness of at least one hazard factors, for example, hypertension, diabetes, hyperlipidemia or officially settled malady) need early identification and the board utilizing directing and prescriptions, as fitting. Coronary illness and stroke together added to 28·1% of all-out deaths in India year 2016 — contrasted and 15·2% in 1990. Coronary illness added 17·8% of all-out deaths and stroke contributed 7·1% of complete deaths. The extent of deaths and incapacity from coronary illness was essentially higher in men than in ladies, however, was comparable among people for stroke. Death due to cardiovascular diseases increases from 13 lakh in 1990 to 28 lakhs in the year 2016.

So, diagnoses & prediction of heart disease earlier is compulsory and treatments for individuals/peoples who are probably going to have a coronary illness and help them have a longer life.

A favored technique to determine the issues of precise finding and the conveyance of focused treatments is the continuous execution of complete physical assessments [4]. Be that as it may, total and continuous physical assessments would prompt information over-burden. So, in this paper, we design a novel approach namely Dynamic LSTM model with classification techniques. The main purpose of this algorithm to produce accurate results for finding out heart disease prediction.

A Data mining alludes to extricating or mining information from a lot of information. As it were, Data mining is the science, workmanship, and innovation of finding huge and complex assemblages of information so as to find helpful examples. To accomplish an objective, the proposed calculation deals with information mining characterization strategies. Characterization is an established issue in AI and information mining [1][23]. In other words, the order issue is worried about producing a portrayal or a model for each class from the given informational index. "Characterization" is a standout amongst the most vital information mining strategies. It is utilized to foresee bunch/class enrollment for information cases.

The remainder of the paper is broken down into six sections. Section II gives a review of the previous work in this area and its different aspects. The Background and related work covered in this section. Section III identifies the problems with the existing approach and Section IV presents proposed work. The experimental results define in Section V. This paper concludes with Section VI, which contains a summary and future work for further research.

II. BACKGROUND AND RELATED WORK

Ji.Zhang et.al [1]design a bagging based troupe display is used to anticipate the patient's condition one day ahead of time for creating the last suggestion. A blend of three classifiers artificial neural network, least squares-support vector machine, and naïve Bayes are utilized to develop a group structure. This proposed framework is a promising device for examining time arrangement medicinal information and giving precise and solid suggestions to patients experiencing endless heart infections.

Fen Miao et.al [2] proposed an improved RSF (iRSF) with a novel split standard and a halting basis for distinguishing progressively exact indicators that can isolate survivors and non-survivors and, in this way, improve segregation capacity. Initial, a weighted log-rank test was utilized to part the hub and can be connected to non-relative risk circumstances to improve the test for a scope of elective theories.

Afef Mdhaffar et.al [4] presents a novel wellbeing examination approach for heart failure prediction/ expectation. It depends on the utilization of complex occasion preparing (CEP) innovation, joined with measurable methodologies. A CEP motor procedure approaching wellbeing information by executing edge-based investigation rules. Rather than having to physically set up limits, our novel factual calculation naturally figures and updates edges as indicated by recorded chronicled information.

Chrysa et.al [5] The heart phonocardiogram is breaking down by utilizing Ensemble Empirical Mode Decomposition (EEMD) joined with kurtosis highlights to find the proximity of S1, S2 and concentrate them from the recorded information, framing proposed HSS conspire, to be specific HSS-EEMD/K.

Bo Jin et.al [3] proposes a successful and strong design for heart disease prediction. The primary commitment of paper is to anticipate heart failure(attack) by utilizing a neural network (i.e., to foresee the likelihood of cardiovascular sickness dependent on patient's electronic restorative data). Specifically, they utilized one-hot encoding and word vectors to demonstrate the analysis occasions and anticipated heart disappointment occasions utilizing the fundamental standards of a long short-term memory model.

Assessments dependent on a genuine informational collection show the promising utility and adequacy of the proposed engineering in the expectation of the danger of heart disappointment.

III. PROBLEMS WITH THE EXISTING APPROACH

Coronary illness or CVDs are the number 1 reason for at around the world for death. A bigger number of people die every year from CVDs than from some other reason.

In our proposed approach we have increased the accuracy and efficiency by replacing the static LSTM which was used in the traditional approaches by dynamic LSTM. Our results clearly depict that the dynamic LSTM have higher accuracy than static LSTM. A few investigations/studies in the writing have been led utilizing outfit approaches, with some achievement, in the medicinal area [1][4]. However, none of them drew closer or tackled the issue that we manage in our examination to give exact suggestions to coronary illness persons to take or cure further medical test in for coming days.

IV. PROPOSED ALGORITHM

We contrast to the above-mentioned problems with the existing approach, we develop a classification technique with dynamic LSTM Model framework for heart disease prediction earlier stage. Our proposed model working on the basic LSTM model [3].
To handle Electronic health records data, we proposed a new algorithm namely Dynamic LSTM model (Long Short-

Term Memory Model). In our algorithm we use a training data set and test data set. This algorithm gives better result compare to the above mention algorithms [1]. We are shown details of the algorithm in this section.

We utilize Classification methods of data mining. Classification is the handling of finding a lot of models (or capacities) those define and recognize information classes. A classification technique has two types of data set namely training and test data. First, one is used to implement a model or classifier and the second dataset is used for testing of implemented model. There are some popular classification methods like:

A. CART :

CART represents Classification and Regression Trees. CART is one of the famous techniques for structure choice tree in the AI people group [13][28]. It was created by [Breiman and his team. At 1984] and is portrayed by the way that it develops binary trees, specifically every interior hub/node must be precisely 2 active edges, every one of which presently endeavors to part in a similar way as the root hub/node. The parts are chosen to utilize the Gini index.

B. Bayesian classifier:

A Bayesian classifier is a measurable classifier. This classifier is used for prediction with probabilities.

C. Fuzzy logic:

Fuzzy logic is a set of the rule-based classification system. It improves all the problems of crisp and rule-based classifiers.

D. Association rule mining :

This is an imperative & active field of data mining researcher. One strategy for association rule mining, called associative classification/characterization, comprises of two stages. In the primary advance, affiliation guidelines are produced utilizing an adjusted form of the standard affiliation rule mining calculation known as A prior. The second step builds a classifier dependent on the affiliation rules found.

E. LSTM Model:

Hochreiter and his team design an LSTM model demonstration. It's a unique RNN method. To fulfill a requirement of long-term memory, the RNN capture of condition/situation of the current concealed layer. The result of the calculation comes in the form of an exponential increment and it shows time, cost of a model. That is a reason RNN is not applicable for long term memory calculations.

F. Dynamic LSTM Model:

A collaborative approach is an effective model, which combines all the multiple parent classifiers to solve a problem and increasing performance of every base classifiers. The working principle is taken from LSTM model and deep learning. There are two different datasets namely training data set and test data set. This model is also in working condition if data varies [3].

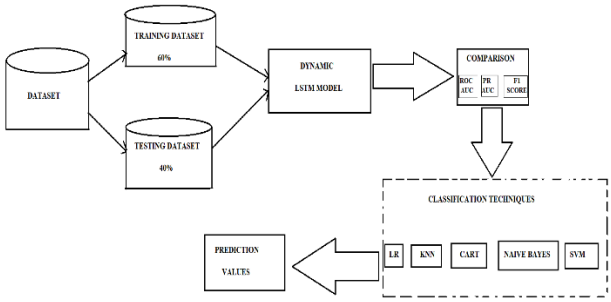

Fig. 1. Framework of proposed dynamic LSTM model.

V. Result Analysis

We used a real-life data collection "HEART-DISEASE dataset" occupied from UCI repository. This data composed from four following different locations: Cardiology University Hospital. The characteristics of the multivariate attributes of the dataset are shown in Table 1.

TABLE I. MULTIVARIATE ATTRIBUTES OF DATASET

Attribute name	Attribute type
Age	Numeric
Sex	Binary
Cp	Numeric
Trestbps	Numeric
Chol	Numeric
Fbs	Binary
Restecg	Binary
Thalach	Numeric
Exang	Numeric
Oldpeak	Numeric
Slope	Numeric
Ca	Numeric
Thal	Numeric
Target	Numeric
Bglf	Numeric
Bgl pp	Numeric
Bglr	Numeric

A property of this dataset is that it is a multivariate and generally used for classification purpose. This dataset contains categorical, integer and real attributes. Basically, there are 14 different attributes that were used earlier and 3 more attributes namely Bglf (blood glucose level fasting), Bgl pp (blood glucose level PP) and Bglr (blood glucose level regular) are added to find accurate results.

Table III clearly shows that when dynamic LSTM model is chosen with numerous classification techniques such as LR, KNN, CART, NB and SVM gave the better results. Graphs are plotted in between the Baseline and the prediction values. The corresponding graphs are as given below.

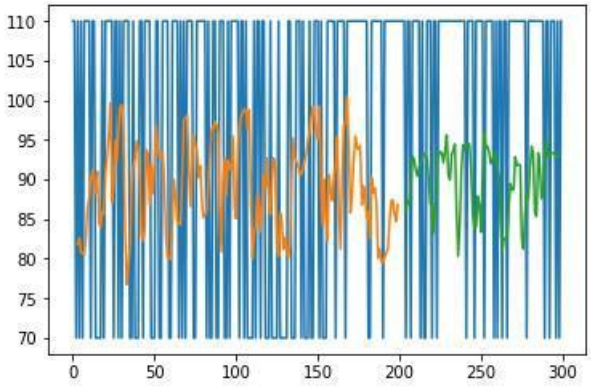

Fig. 2. The layout of Heart disease dataset in histogram format

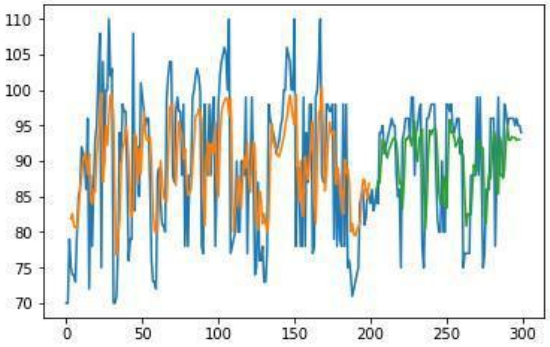

Fig. 3. Dynamic LSTM model Trained on regression Design of Heart Disease Prediction Problem.

Classification techniques with dynamic LSTM model are implemented on Python 3.6 version. This model is tested on different parameters and classification techniques:

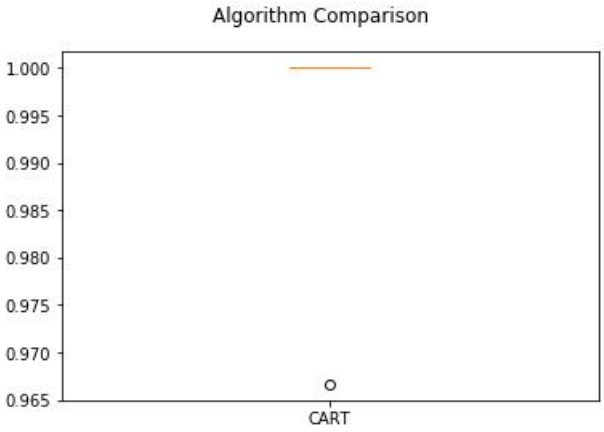

Fig. 4. Classification of Heart disease by CART technique

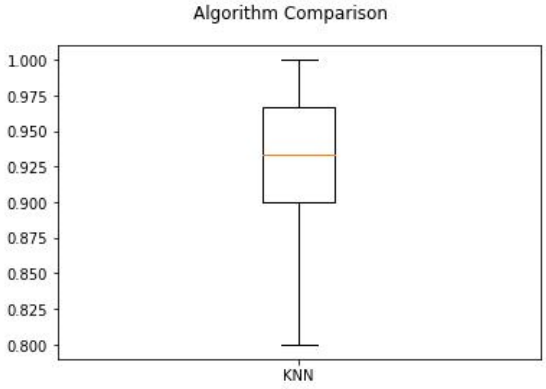

Fig. 5. Classification of Heart disease by KNN technique

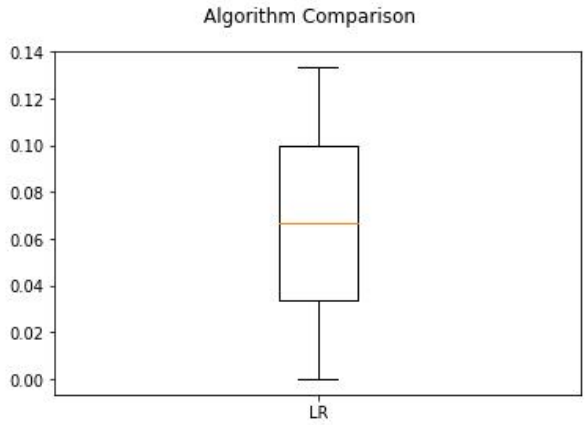

Fig. 6. Classification of Heart disease by LR technique

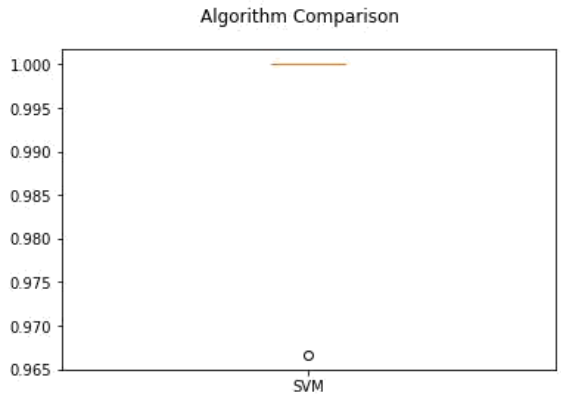

Fig. 7. Classification of Heart disease by Support vector machine

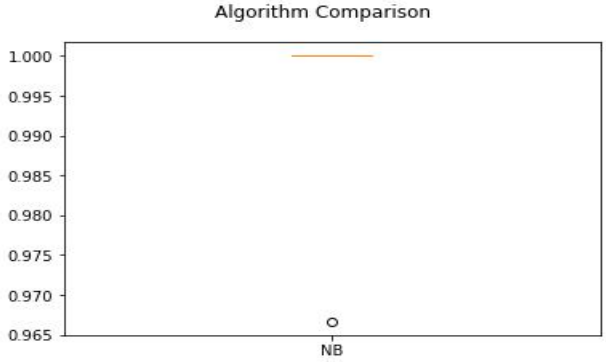

Fig. 8. Classification of Heart disease dataset by Naïve Bayes method

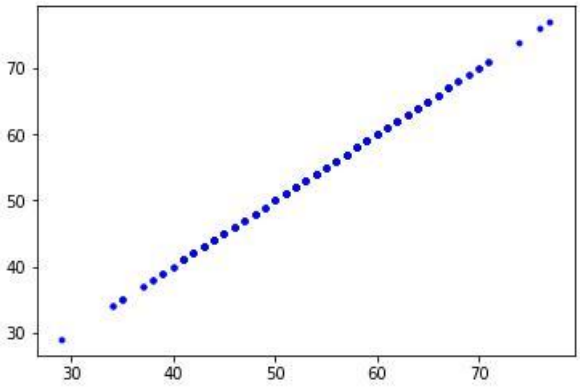

Fig. 9. A line plots of the series creates & shows a clear growing learning.

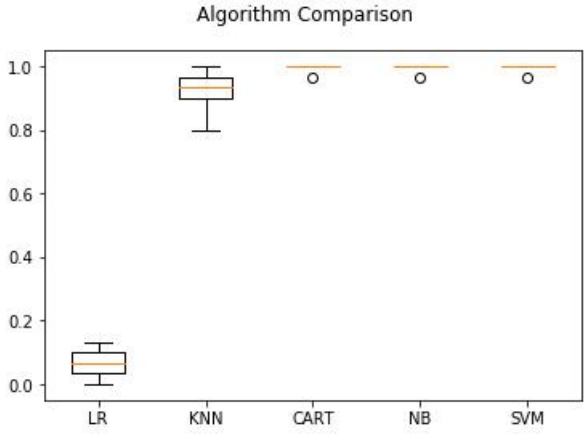

Fig. 10. Collective comparison of classification tehcniques

TABLE II. RESULT OF TEST DATASET

Classification	**Accuracy**
(Test predicts, Train predicts)	(85.97968, 87.23571)

TABLE III. RESULT ANALYSIS BASED ON K-FOLD METHOD

Classification techniques	**Mean Distribution Accuracy**	**Standard Deviation Accuracy**
LR	0.070000	0.043333
KNN	0.926667	0.059255
CART	0.996667	0.010000
NB	0.996667	0.010000
SVM	0.996667	0.010000

TABLE IV. COMPARE LSTM AND DYNAMIC LSTM

LSTM		**DYNAMIC LSTM**	
ROC- AUC	0.6827	LR	0.043333
PR- AUC	0.2678	KNN	0.059255
F1 - SCORE	0.2188	CART	0.010000

VI. Conclusion & Future work

In this section, we conclude the paper with overall findings as well as suggestions for future research work. The paper proposes a design of original methodology for finding accurate result and helps in earlier diagnoses of heart disease predictions. The proposed model gives more efficient and accurate result as compare to previous diagnoses models and algorithms. This classifier varies according to the different data records. One of the advantages of this model is that it can

work on a real, numerical and categorical dataset. In future work, this model can be used in diagnosing of different diseases prediction.

REFERENCES

1. Ji Zhan, R. Lafta, X.Tao, Y. LI," Coupling a Fast Fourier Transformation With a Machine Learning Ensemble Model to Support Recommendations for Heart Disease Patients in a Telehealth Environment", IEEE Access, vol.5,pp.10674-10685,2017
2. Fen Miao, Y.P.Cai, Y.X.Zhan," Predictive Modeling of Hospital Mortality for Patients With Heart Failure by Using an Improved Random Survival Forest" IEEE Access, vol.6,pp.7244-7254,2018 .
3. Bo Jin, C.Che , Z. Liu," Predicting the Risk of Heart Failure With EHR Sequential Data Modeling" IEEE Access, vol.6,pp.9256-9261,2018.
4. A.Mdhaffar, I.B.Rodriguez, K.Charfi," CEP4HFP: Complex Event Processing for Heart Failure Prediction", IEEE Transaction, vol.16,pp.1536-1241,2017
5. C.D.Papadaniil, L.J.Hadjileontiadis, "Efficient Heart Sound Segmentation and Extraction Using Ensemble Empirical Mode Decomposition and Kurtosis Features", IEEE Journal of Biomedical and health Informatic,2013
6. F. M. Bianchi, E. De Santis, A. Rizzi, and A. Sadeghian, "Short-term electric burden estimating utilizing reverberation state systems and PCA decomposition," IEEE Access, vol. 3, pp. 1931_1943, 2015.
7. Pati, B. Kumar, D. Manjhi, and K. K. Shukla, "An examination among ARIMA, BP-NN, and MOGA-NN for programming clone development forecast," IEEE Access, vol. 5, pp. 11841_11851, 2017.
8. Y.- T. Su, Y. Lu, M. Chen, and An.- A. Liu, "Spatiotemporal joint mitosis discovery utilizing CNN-LSTM arrange in time-slip by stage differentiate microscopy pictures," IEEE Access, vol. 5, pp. 18033_18041, 2017.
9. G. Zhu, L. Zhang, P. Shen, and J. Melody, "Multimodal motion acknowledgment utilizing 3-D convolution and convolutional LSTM," IEEE Access, vol. 5, pp. 4517-4524, 2017.
10. G. P. Zhang, "Time arrangement anticipating utilizing a crossover ARIMA and neural system show," Neurocomputing, vol. 50, pp. 159_175, Jan. 2003.
11. G. M. Jenkins and A. S. Alavi, "Some parts of demonstrating and estimating multivariate time arrangement," J. Time Ser. Butt-centric., vol. 2, no. 1, pp. 1_47, 1981.
12. R. G. Dark colored, "Exponential smoothing for anticipating request," Oper. Res., vol. 5, no. 1, pp. 145_145, 1957.
13. G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, Linear Nonstationary Models Time Series Analysis), the fourth ed. Hoboken, NJ, USA: Wiley, 1976, pp. 93_136.
14. S. Hochreiter and J. Schmidhuber, "Long momentary memory," Neural Comput., vol. 9, no. 8, pp. 1735_1780, 1997.
15. F. A. Gers and J. Schmidhuber, "Recurrent nets that time and tally," in Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Netw. (IJCNN), vol. 3. Jul. 2000, pp. 189_194.
16. Chung, C. Gulcehre, K. Cho, and Y. Bengio. (2014). "Observational assessment of gated intermittent neural systems.
17. G. E. Hinton, "Learning conveyed portrayals of ideas," in Proc. eighth Annu. Conf. Cognit. Sci. Soc., vol. 1. 1986, pp. 47_61.
18. Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, "A neural probabilistic language show," J. Mach. Learn. Res., vol. 3, pp. 1137_1155, Feb. 2003.
19. C. Buckley and E. M. Voorhees, "Retrieval assessment with deficient data," in Proc. SIGIR, 2004, pp. 25_32.
20. N. Tangri et al., "A prescient model for the movement of endless kidney malady to kidney disappointment," JAMA, vol. 305, no. 15, pp. 1553_1559, 2011.
21. R. Sukkar, E. Katz, Y. Zhang, D. Raunig, and B. T. Wyman, "Disease movement demonstrating utilizing shrouded Markov models," in Proc. Annu. Int. Conf. IEEE Eng. Medications. Biol. Soc., Aug./Sep. 2012, pp. 2845_2848.
22. J. Zhou, J. Liu, V. A. Narayan, and J. Ye, "Modeling illness expert gression through perform various tasks learning," NeuroImage, vol. 78, pp. 233_248, Sep. 2013.
23. Y.- Y. Liu, H. Ishikawa, M. Chen, G. Wollstein, J. S. Schuman, and J. M. Rehg, "Longitudinal displaying of glaucoma movement utilizing 2-dimensional constant time shrouded Markov show," in Proc. Int. Conf. Prescription. Picture Comput. Comput.- Assisted Intervention. (MICCAI), Nagoya, Japan, 2013, pp. 444_451.
24. P. Schulam and S. Saria, "A system for individualizing expectations of malady directions by misusing multi-goals structure," in Proc. Adv. Neural Inf. Procedure. Syst. (NIPS), Montreal, QC, Canada, 2015, pp. 748_756.
25. X. Wang, D. Sontag, and F. Wang, "Unsupervised learning of sickness movement models," in Proc. Knowl. Revelation Data Mining (KDD), New York, NY, USA, 2014, pp. 85_94.
26. E. Choi, N. Du, R. Chen, L. Melody, and J. Sun, "Constructing malady system and fleeting movement show by means of setting touchy Hawkes process," in Proc. Int. Conf. Information Mining (ICDM), Atlantic City, NJ, USA, Nov. 2015, pp. 721_726.
27. N. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, and T. Ploetz," PD sickness state evaluation in naturalistic situations utilizing profound learning," in Proc. AAAI, Austin, TX, USA, 2015, pp. 1742_1748.
28. Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzell. (2016). "Learning to determine to have LSTM repetitive neural systems."
29. Karpathy and L. Fei-Fei, "Deep visual-semantic arrangements for creating picture depictions," in Proc. PC. Vis. Example Recognit. (CVPR), Boston, MA, USA, 2015, pp. 3128_3137.
30. Cho et al., "Learning phrase portrayals utilizing RNN encoder-decoder for measurable machine interpretation," in Proc. Observational Methods Natural Lang. Procedure. (EMNLP), Doha, Qatar, 2014, pp. 1724_1734.