Design Model for Information Classification in Big Data Environment

[1]Prachi Paliwal, [2]Gourav Sitlani
[1,2]RGPV University, Bhopal, Madhya Pradesh, India

Abstract
Deep learning is a standard machine learning approach which has accomplished a hug of advancement with completely traditional machine learning in the domain of areas. The current issue centers around how to extract and classify the greatest, huge scale real dataset. Consequently, the task manages proposing a framework that can efficiently contribute a to a extremely proficient and precise capable prediction model. In the human health services domain: A point to point analysis of the patient records and patient live condition. Change and improvement of the existing frameworks by appending or replacing manual work with the utilization of huge data and information based artificial intelligence forms. Applying deep learning to these domains has been one among the prevalent points of research. The model given by us has the capability of maintaining high a long term accuracy. In case of an NDLCM network, a stacked NDLCM layer makes possible learning a high level temporal feature without any need of fine tuning and preprocessing that may otherwise be important in case of another technique. In this proposed paper, we construct a deep learning computational model which uses the normal back propagation neural network training set that helps to build a precise prediction model.

Keywords
Deep Learning, Deep Neural Network, Internet of Things, IoT Big Data.

I. Introduction

Presently, there are many IOT applications available that depend on big data which emerged in different vertical spaces as well as domains like wellbeing, transportation, shrewd home, brilliant city, horticulture, instruction and so forth. For all these applications, the main feature is an insightful learning instrument for forecast (i.e., grouping or relapse), or bunching. Deep Learning (DL) has been one of the most effective and used IoT applications of late. Both IoT and big data have two-way relationship between each other. IoT is the most essential maker of numerous information whereas it's the most important focus for all huge informations in order to increase various procedures as well as administrations of IoT. Mainly, its information examination gives basic knowledge to general public IoT information that are not same as general huge information. In order to serve better, it is very important to check out all the requirements of IoT information and look for the properties of healthcare information in a hospital environment.
IoT information shows different kind of accompanying qualities such as :

1. Large-Scale big data information classification patient usage Data hospital environmental data
2. A heap of information catching gadgets are dispersed and sent for IoT applications and produce surges of information ceaselessly for an immense volume
3. Heterogeneity–Now, there are many IoT information procurement gadgets which help you to accumulate different data by gathering about information heterogeneity.
4. Time and space connection–there are various sensor gadgets which have been added to particular area in many parts of IoT applications where all lines have a specific area as well as time span for every information.
5. High commotion information– There are few information that might be liable to mistakes & clamor amid procurement and transmission.In such cases, we need to use predicted error for prediction model in order to build up deep learning model which can detect whether sample is abnormal and this model named as NDLCM.
6. We use three kinds of data sets, each contains time dependencies as long-term, short-term or very weak. NDLCM (Novel Deep Learning Computational Model) is a powerful and sturdy model - as given by an experimental. The remainder of this paper is organized as follows :

- **Section II:** Deals with a review of related work.
- **Section III:** Description of challenges and solutions for peculiarity detection proposed methodology.
- **Section IV:** Elaboration on the NDLCM
- **Section V:** Details about the Deep learning based on back propagation neural network model for Anomaly Detection in IoT (Internet of Things) Data and at the same time analysis of the results.
- **Section VI:** The conclusion.

II. Related Work

To the preeminent of our research and analysis, there breathes a colossal amount of effort and an article in the literature which depends on for not only surveying the particular domain among IoT data and deep learning computational model but also on technique and approach and deep learning methods in IoT. They show a few works that are common to data mining also deep learning model that has been used in IoT environments.
P. Li, et al [1] In this paper, proposed a deep convolutional computation model based on hierarchical feature learning on big data in IoT. In order to reveal the hidden correlations over different modalities of the big data, the concept revolving around the tensor-based model is adopted to accurately represent each and every single heterogeneous object. A. Gupta et al[2]this work demonstrates the framework behind the big data classification. This framework is observed to be based on the combination of deep learning and, namely Apache Spark, which is particularly specified as a single structure. Using the Cascade Learning for stabilizing the relationship between deep learning and, namely Apache Spark. This is a method of fusion based analysis that utilizes a three-tier combination for increasing the accuracy and precision to an agreeable degree. M. Mohammadi et al[3] in this work studies the characteristics of IoT data in order to determine the problem and other limitations involved in the method of deep learning technique. In particular, focusing on two categories, first specified IoT big data and to find out the requirement for analysis. This work also aims to propose a deep learning model for applying the IoT application, use the open source framework for the effective implementation of the deep learning architectures. A large number of research studies proposed an information classification system using supervised Learning and Artificial Neural Networks based

concept Social data information classification using Supervised Learning: this is not possible for applying all supervised learning algorithm for information classification in big social data domain are computationally as costly as Deep Neural Network. Previous work [4] illustrates that classification algorithms accomplish a higher rate of accuracy for prediction rates with textual content and minor accuracy rates with measurable content. Classification of the text information by the utilization of Artificial Neural Networks (ANN) number of researcher work previously on that [5-6] ANN mostly applying for sentiment classification on Twitter. More precisely, [7] use ANN with n-gram analyses for feature extraction. Authors developed DAN2 (a Dynamic Architecture for Artificial Neural Networks) using a feed forward approach with input, hidden and output layer. Though, number of hidden layers in not fixed a priori as in the existing research.

III. Proposed Methodology

This research work has been primarily designed with the focus offending the right design an novel deep learning computational model: NDLCM for the purpose of classification of information in the big data environment and also, to frame a predictive model which can significantly aid in the process of improving healthcare domain and improve the learning knowledge in terms of the detection of anomalies and also various strategies. We aim to evaluate (Multilayer perceptron, Naive Bayes) different algorithms, models and statistical approaches, in order to discover the best current algorithm based on model, in order to estimate post performance and then subsequently create predictions based on the quantity of suitable attributes with back propagation neural network (BPNN) and Deep learning Neural Networks (DLNN). The model that we have proposed is Simple, and a model which can lead to precise and consistent predictions of post-performance. Choosing the accurate features and considerate health care domain-specific characteristics of data leads to enhanced results.

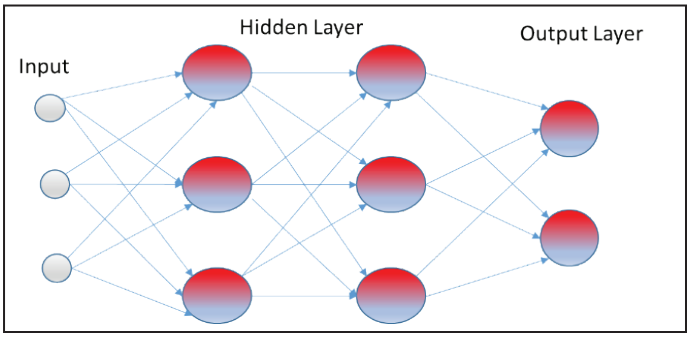

Fig. 1: Back Propagation Prediction Model

Our proposed Deep learning model can improve data analyses and accomplish a higher rate of accurate results than simple BPNN network due to the improved quantity of hidden units and number of network layers that disseminate data points by weighing them in each layer.

IV. Using the Novel Deep Learning Computational Model: NDLCM

The neural network is an extensively interconnected network of modest neurons which can regulate, and simulate the response of the biological nervous system to real world objects. A BP (Back Propagation,) neural network has the capability to learn, memorize, associate, convince, simplify and excerpt features, tolerance faults and meditation. This is utilized for extract complex interactions among input and output, even when the relationship themself. Recently, BP neural network has been widely used to resolve the problems of identification and prediction. At the same time, BP neural network as a new deep learning computational model: NDLCM prediction method can a help in approximation of nonlinear quantities with a high prediction accuracy. Complete the use of the training and testing interactions amongst before and advanced, the past explanations as BP neural networks input and the imminent value as the BP network's output, which build a prediction model.

To perform the following process for information classification Data Accession: Data is the key benefit in the enterprise of future intelligent wireless networks. In order to acquire spectrum data, the radio first senses its environment by gathering raw data from numerous spectrum bands. The raw data contains a number of samples of the received wireless signal. Data processing: Data pre processing is worried through the analysis and management of the composed spectrum data with the intention to reach at possibly respectable wireless data illustrations. The raw illustrations of that set prepared hooked on data vectors block are pipelined as contribution for signal processing (SP) tools that investigate, process and convert the data to reach at simple data representations such as frequency, amplitude, phase and spectrum, In numerous Deep learning applications the best of features is even-handed as significant, if not additional significant than the select of the deep learning model. Information Classification: The Information Classification processing block supports intelligence abilities to evaluate the environmental radio setting by detecting the occurrence of wireless signals. This might be the type of the emitters that are applying the spectrum (spectrum contact scheme, variation format, wireless technology), kind of interfering, detecting an accessible spectrum band, etc. Using that, our process to learn the spectrum and to apply the wireless networks Deep Learning model might Performa key role in routinely classifying wireless signals as a step. Decision processing: The predictions premeditated by the deep learning model are applied as input for the decision component. The decision might be communicated to an interfering moderation strategy for a definite time period. In other communication circumstances such as spectrum parameter, the decision might transmit to a spectrum policy or spectrum acquiescence implementation useful to a distinguished source of harmful interference.

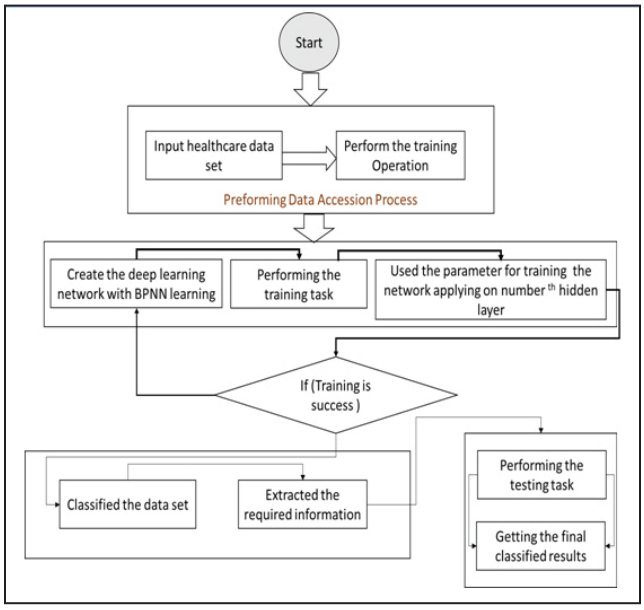

Fig. 2: Novel Deep Learning Computational Model

V. Deep learning based on back propagation neural network model for Anomaly Detection in IoT Data

IoT data has high noise topographies along with combination of multi dimensionality as well as dynamic features. Significant objectives of preprocessing of IoT were reducing dimensionality of the data without noise. If there is Noisy data it will raises the complexity of anomaly detection, as well as problematic on the specific IoT Data. As well as atypical data was not accessible or sparse, it will create a problem to learn sequence classification model which are normal as well as abnormal.

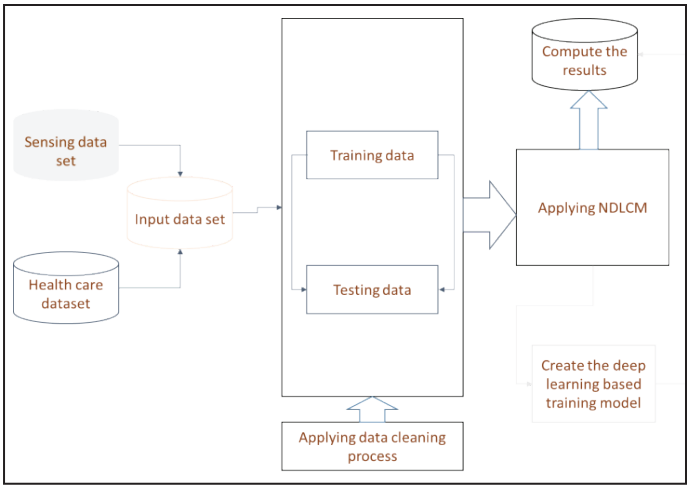

Fig. 3: Show The Proposed Model Working in Term of Training and Testing Based Results Prediction Model

Results Analysis: When simulation running on Windows 1064-bit system CPU Intel-I3 computer, the code is executed on i386 R platform. Thereby simulation as well as verification of model was acknowledged on basis of the health care data set, and which comprises of 5000 sets. Two types of learning sets were required in order to permit the built model of the Anomaly Detection.

The taxonomy of performance assessment index expending precision (P) rate and recall(R) rate to assess:

P= exact number which are allocated to the class/concrete number of points to the class.

R= exact number of assigned to the class/actual number of points should be allocated to the class.

Deep learning techniques depends on the back propagation neural network which are having advantage for completion of post appointment attributes. For quality data analysis, appropriate performance movements were used in order to regulate. To know whether the post was well comprehended or not, controlled learning methods were used from clustering significances.

Result of hidden layer unit's number:

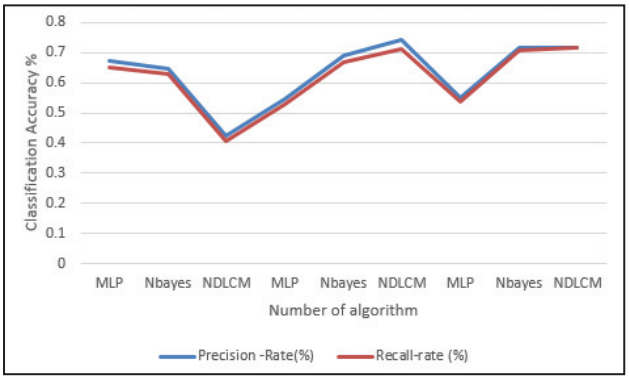

Fig. 4: Comparative Analysis Different Number of Algorithm in Term of Precision and Recall

70% of the optimization and the accuracy will be better by 20% in subsequently optimization, and the accuracy growths with the quantity of iterations. Thus it reveals that, modification which was acceptable and can successfully advance the accuracy of NDLCM model.Similarly, interpretation was another important feature of the model presentation. Increasing the novel data set to test NDLCM model, which helps in order to verify the capability of processing of different data size samples. The data sets are differentiated with three different hidden layers. The classification accuracy was shown in the fig. 5.

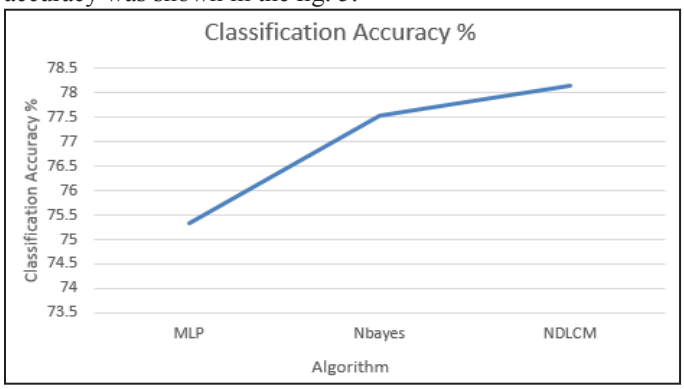

Fig. 5: Classification Accuracy in Term of Different Algorithm

According to the simulation, results reveals that with increase in the number of samples, then the accuracy of three collections of the experimental consequences was in rising trend, but accuracy of network with 3 hidden layers was considerably high than the other two groups. Hence it can be resulted that suitable accumulation of number of layers of network will help in expanding and simplification of the performance of our proposed model. Then, our objective was able to classify is the sample was abnormal, precision and recall were significant metrics in which our target is to evaluate our model. It is supposed that novel deep learning computational model as well as BPNN model were less effective for anomaly detection of the data set. Perhaps, this was attributable to the features of data set which was being incomprehensible & value being in frequently fluctuating range. This was similarly created in term of time for the calculations. So BPNN model algorithm was improved at the precision rate of anomaly detection when compared with that of the novel deep learning computational model.

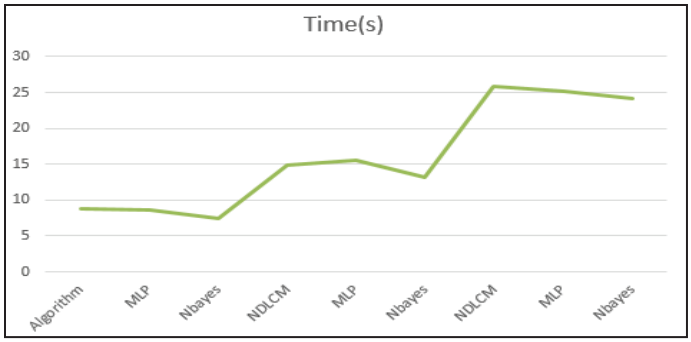

Fig. 6: Bar Graph Showing Maximum Accuracy Reached Through Reverence to Specific Time

VI. Conclusion

The benefits of the Proposed model are it will be used to serve as the key to in order to improve prediction accuracy with the bulk data sets and also used for the classification. Even though it was a time consuming, data preparation, pruning and correct selection of the model will be aid in order to understand various

domain specific features of the data set and therefore helps to donate towards results with accuracy of greater rate. In the present work NDLCM prediction was used as a measurable data, which was used to demonstrate too lower rate of accuracy and results were accomplished in this present paper. Our research revealed that the identical deep neural networks won't contribute to the higher accuracy results as well as it was time consuming in terms of processing and classification of the information which deals with the health care data.

References

[1] P. Li, Z. Chen, L. T. Yang, Q. Zhang, M. J. Deen,"Deep Convolutional Computation Model for Feature Learning on Big Data in Internet of Things", In IEEE Transactions on Industrial Informatics, Vol. 14, No. 2, pp. 790-798, 2018.

[2] A. Gupta, H. K. Thakur, R. Shrivastava, P. Kumar, S. Nag,"A Big Data Analysis Framework Using Apache Spark and Deep Learning," 2017 IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, LA, 2017, pp. 9-16.

[3] M. Mohammadi, A. Al-Fuqaha, S. Sorour, M. Guizani, "Deep Learning for IoT Big Data and Streaming Analytics: A Survey," In IEEE Communications Surveys & Tutorials.

[4] B. T. Ong, K. Sugiura, K. Zettsu,"Dynamic pre-training of Deep Recurrent Neural Networks for predicting environmental monitoring data", 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, 2014, pp. 760-765.

[5] Q. Zhang, L. T. Yang, Z. Chen, P. Li, F. Bu,"An Adaptive Droupout Deep Computation Model for Industrial IoT Big Data Learning with Crowdsourcing to Cloud Computing," in IEEE Transactions on Industrial Informatics.

[6] G. Cheng, Z. Li, J. Han, X. Yao, L. Guo,"Exploring Hierarchical Convolutional Features for Hyperspectral Image Classification", In IEEE Transactions on Geoscience and Remote Sensing.

[7] Gu Yingkui,Cheng Zixin, Zhu Fanlong,"Rolling bearing fault feature fusion based on PCA and SVM[J]. China Mechanical Engineering, 2015,26(20): 2778-2883.

[8] L. Jing, L. Yankai, C. Shanshan, W. Zhijie, J. Haipeng, "Application of deep learning in the hydraulic equipment fault diagnosis," 2018 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE), Hsinchu, 2018, pp. 12-16.

[9] Rui Zhao, Ruqiang Yan, Zhenghua Chen, Kezhi Mao, Peng Wang, Robert X. Gao,"Deep Learning and Its Applications to Machine Health Monitoring: A Survey", Journal of Latex Class Files, Vol. 14, No. 8, 2015.

[10] Eva K. Lee,"Machine Learning Framework for Classification in Medicine and Biology", W.-J. van Hoeve and J.N. Hooker (Eds.): CPAIOR 2009, LNCS 5547, pp. 1–7, 2009.

[11] Y. Demchenko, P. Grosso, C. De Laat, P. Membrey, "Addressing big data issues in scientific data infrastructure", In Collaboration Technologies and Systems (CTS), 2013 International Conference on.

[12] M. Strohbach, H. Ziekow, V. Gazis, N. Akiva,"Towards a big data analytics framework for iot and smart city applications", in Modeling and processing for next-generation big-data technologies. Springer, 2015, pp. 257–282.

[13] G. E. Hinton, R. R. Salakhutdinov,"Reducing the dimensionality of data with neural networks," Science, Vol. 313, No. 5786, pp. 504–507, 2006.

[14] L. Deng,"A tutorial survey of architectures, algorithms, and applications for deep learning", APSIPA Transactions on Signal and Information Processing, Vol. 3, pp. 1–29, 2014.