Grenze International Journal of Engineering and Technology, Jan Issue

Big Data based Deep Learning for Predicting the Conditions of Patients from Medical IoT Data in Healthcare

Dr. P. Umaeswari[1], Dr. S. Senthil Kumar[2] and S.S.Gnanesh[3]
[1]Associate Professor, Department of Computer Science and Business Systems, R.M.K. Engineering College, Kavaraipettai-601206, Tamil Nadu, India
Email: pui.csbs@rmkec.ac.in
[2]Professor, Department of Mechanical Engineering., R.M.K. College of Engineering and Technology, Puduvoyal, Tamil Nadu, India
Email: senthilkumar@rmkcet.ac.in
[3]Second Year Student, Department of Electronics & Instrumentation Engineering, R.M.K Engineering College, Kavaraipettai-601206, Tamil Nadu, India
Email: gnan21104.ei@rmkec.ac.in

***Abstract*—Data from the Internet of Things (IoT) is being generated and captured at a rapid pace, much of which is relevant to healthcare systems. These pieces of information are in real time and are not structured. The storage and processing of real-time medical data in Internet of Things applications, on the other hand, continues to be a significant challenge. In this paper, a big data pipeline for processing the data from healthcare IoT devices. The study is about to monitor the mental and physical strength of patients admitted in intensive care units via IoT devices. The big data processing is conducted using Apache Flume tool that collects and transfers large data from IoT data to Hadoop Distributed File System from cloud server. The stability of the patients is conducted using a deep learning framework that finds the features and classifies the stability. The simulation is conducted on the data collected from the patients in real-time environment. The simulation shows accurate prediction of patient stability from large collected data. Thus, it is believed that the proposed method acts a better tool in providing viable solution for prediction of patients' stability from IoT data.**

***Index Terms*— Big Data, Deep Learning, Medical IoT, Healthcare.**

I. INTRODUCTION

In addition to smart cities and healthcare, analytics and the Internet of Things can be applied to agricultural and industrial automation. The Internet of Things (IoT) devices and apps have created a massive amount of data in a wide range of formats. [1]. Big data technologies, on the other hand, have opened up new possibilities for the development of Internet of Things systems. As a result, when merging with Big Data technologies with IoT devices [2], various challenges are confronted by the researchers for data processing and storage [3].
The IoT technology is referred as the process of making common objects, such as refrigerators, cars, and other electric and electrical devices, smart through the use of the internet. The IoT uses computer-based chips and sensors to collect data from devices that are unable to connect directly with the world of internet [4]. Big Data

Grenze ID: 01.GIJET.9.1.508_1

and the Internet of Everything (IoE) are being used to analyse and gain insight into industries as diverse as agriculture, transportation, and logistics, to name a few [5]. Plants in agriculture fields are mainly connected with this IoT based moisture monitoring systems, and the information gathered by these sensors is made available to farmers in real time.

II. Literature Survey

Riggins and Wamba [6] developed a formalized IoT technology, where when the data is acquired from smart devices displays a high volume, rapidity, and variety, Big Data and the IoT come into play. There is a great deal of information that can be gathered from various sensors, both structured and unstructured. Because of the rapidity with which IoT data is being processed, it serves as a constant reminder of the difficulties associated with velocity in Big Data [7]. Variety is one type of IoT-based data, which refers to data that is available in a variety of formats.

The massive amounts of data generated by IoT devices offer a significant challenge for the Internet of Things. Because it is constantly streaming, this Internet of Things-generated data may be managed by Big Data technologies [8]-[10]. IoT sensors often generate unprocessed data, and big data analytics techniques are used to extract relevant insights from this raw data in order to make IoT systems smarter and more efficient. The amount of IoT data, on the other hand, is found to be fundamentally different and hence it is very essential in finding a suitable method for the extraction of data in an accurate way [8] [9].

In this paper, the authors model a big data processing pipeline that is collected from different IoT devices in specific to healthcare environment. The study is about to monitor the mental and physical strength of patients admitted in intensive care units via IoT devices. The big data processing is conducted using Apache Flume tool that collects and transfers large data from IoT data to Hadoop Distributed File System from cloud server.

The integration of the IoT into E-Health systems has prompted a broader shift in healthcare paradigms by providing quick access to data [10]. Gartner in [11] predicts that by the year 2025, the IoT will be networked with 25 billion devices, according to their projections. Gartner defines a strategic partnership as a partnership between a company and its customers. Heart rate monitors, blood sugar monitors, heart rate monitors, mood swing monitors, and body mass monitors at various levels will be included. The entire data is hence generated through the IoT devices and this is processed specifically by the Big Data processing pipeline in order to gain meaningful insights [12].

Therefore, the entire generated data from the IoT devices in healthcare systems has increased dramatically in recent years as a result of technology such as mobile health and biosensors. This includes EHRs, as well as the findings of electronic patient records, among other things (ePRO). Using big data in case of healthcare systems has the goal of reducing computer time by processing massive amounts of data on a regular basis [10].

Big data has the potential to predict diseases, allowing us to save lives by averting needless deaths before they occur. An individual profile can be built more quickly and accurately using data from several sources, such as insurance and medical records, that are acquired and analysed in a big data pipeline [13]. As a result of the IoT, patient data may be acquired via cloud platforms and then processed utilising big data processing technologies, making big data healthcare a highly profitable endeavor [14]-[16].

Big Data platforms and analytics environments are becoming increasingly vital in the IoT is to generate significant advantages and improve decision-making processes.

III. Proposed Method

This research presents an integration of IoT data collection and big data pipeline for processing and retrieval of data to diagnose the condition of patients. Patient monitoring equipment, such as Internet of Things devices, will be used to keep tabs on the emotional and physical well-being of patients in intensive care units.

Using Apache Flume, large amounts of IoT data may be gathered and sent from a cloud server to the Hadoop Distributed File System (HDFS), where it can be processed. Deep learning architecture that recognizes and categorizes the qualities of the patient is used to assess the stability of the patient.

To begin, there are three phases to implementing a cloud-based and IoT-based healthcare system.

- The initial phase involves the collection of data from Internet of Things devices, the UCI repository, and medical records.
- In the second phase, patient records are stored in a secure cloud database, which is password-protected.

In the third step, the disease is diagnosed and the severity of the disease is predicted. This step makes an attempt to estimate the severity of the disease based on information from medical records and a sub-module known as

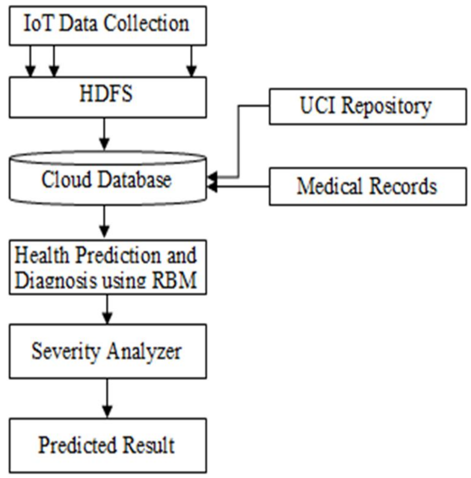

Figure 1. Big Data and IoT in Healthcare

the severity analyzer, among other sources. Every one of these three stages will be thoroughly discussed in the three subsections that follow this one.

A. Data Collection

The IoT sensors in case of a healthcare monitoring enables the integration of various data points is being developed. In this phase, wearable IoT devices with sensors are utilised to collect personal information about each individual patient, which will be used in following phases. These wearable IoT devices are found to be attached with the physical human body in order to collect medical information or data on a regular basis about the unique patient.

The vast majority of the time, these devices compare the patient medical data to a predetermined set of numbers or parameters.

When any of the patient medical data exceeds the normal range in any of the device essential parameters, this device sends an alarm message and the relevant numbers to the doctors and the data collection module. A medical diagnosis system is presented that transfers clinical data to the cloud utilizing 5G mobile networks, rather than wired networks. This dataset additionally makes use of values from the UCI repository in order to map to data supplied by IoT devices. Furthermore, it is used for the mapping of individual patient data with the medical records in healthcare management system are hence utilized for this purpose.

B. Cloud Data Storage

Decisions were made in the proposed system based on three categories of medical data that were collected. Devices connected to the IoT in this environment collect medical data from a variety of individuals in a number of locations for analysis. It is possible to store an extremely large amount of data in Hadoop scalability. The aim of the proposed model is to store and retrieve the data but it poses limitations on data security to ensure the protection of medical information.

C. Disease Severity Prediction and its Diagnosis

In order to detect diseases and assess their severity, a deep learning model is used as a classifier and it is discussed below:

D. Restricted Boltzmann Machine (RBM)

The study uses RBM as its classifier that has two layers: an input layer and an output layer that contains latent variables. An RBM is composed of two layers. Underlayer 1 is responsible for disseminating the training data. To optimize the weights of the network, the backpropagation and gradient descent algorithms are utilized in conjunction with the training.

The classifier aims at the maximization of product of the visible layer probability. When applying this strategy, the latent variables are determined through the use of forward feeding. Because of this, the fact that RBM acts just like arithmetic equations is quite easy to understand.

The RBM is a joint distribution modelled using different input $x = (x_1,..., x_D)$ with the class of hidden layer (i.e. target) $y \in \{1,...,C\}$ using a stochastic units in hidden layer $h = (h_1,...,h_H)$. Therefore, the energy function is hence modelled as below:

$$E(y,x,h) = -\left(h^T Wx + b^T x + c^T x + d^T e_y - h^T U e_y\right)$$

The parameters $\Theta = (W, b, c, d, U)$ with ey is considered as the representation of C in y. The probabilities are therefore assigned from the energy function for the y, x and h values as below:

$$P(y,x,h) = \frac{\exp(-E(y,x,h))}{Z}$$

where

Z - normalization constant.

It is assumed that the elements present in the input x is of a binary valued one. Therefore, the entire conditional distributions are assumed to be simple and then the visible layer is conditioned as below:

$$P(h \mid y,x) = sigm\left(c_j + U_{jy} + \sum_i W_{ji} x_i\right)$$

where sigm(a) - logistic sigmoid function and is given as $(1+\exp(-a))^{-1}$. When conditioning on the hidden layer, we have

$$P(y \mid h) = \frac{\exp\left(d_y + \sum_j U_{jy} h_j\right)}{\sum_{y^*} \exp\left(d_{y^*} + \sum_j U_{jy^*} h_j\right)}$$

Then compute $p(y|x)$ to perform classification. Therefore

$$\sum_{h_{1,\dots H} \in \{0,1\}} h_j = \exp\left(d_y + \sum_j softplus\left(c_j + U_{jy} + \sum_i W_{ji} x_i\right)\right)$$

where

softplus(a) = log(1+exp(a)) and it is expressed as below:

$$P(y \mid h) = \frac{\exp(-F(y,x))}{\sum_{y^*} \exp(-F(y,x))}$$

where

$F(y,x)$ - free energy required to compute the classification of instances.

IV. Results and Discussions

The requested task was completed through the use of python programming on a high end computing system that is aimed at processing the big data. The dataset for various diseases are collected from UCI repository [17]. The classification of data into normal and pathological categories, as well as the severity of each category, is a critical component of this endeavour. As proven, the proposed RBM classifier can be used in conjunction with other classification approaches, such as SVM, ANN, LSTM and RNN, to improve accuracy and efficiency. Furthermore, we have restricted our research to only diabetes in this study. Despite the fact that the RBM criteria we utilized in the proposed classifier can also be used to predict other potentially deadly diseases, we chose to focus on cancer.

A. Metrics

In this study, we looked at the accuracy and sensitivity all of which are usual evaluation criteria, as well as the reliability. It is possible to calculate the true positive (TP), true negative (TN), false negative (FN), and false positive (FP) values, among other things.

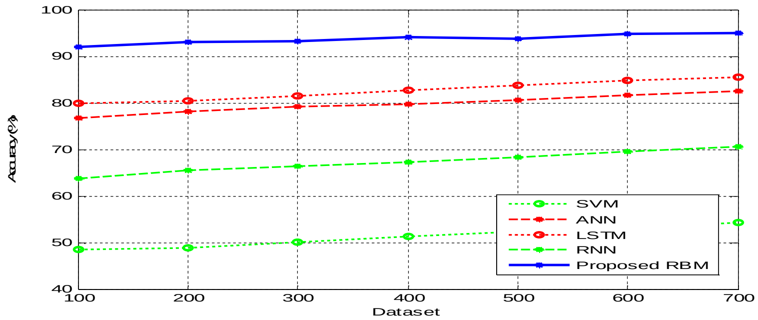

Figure 2: Accuracy

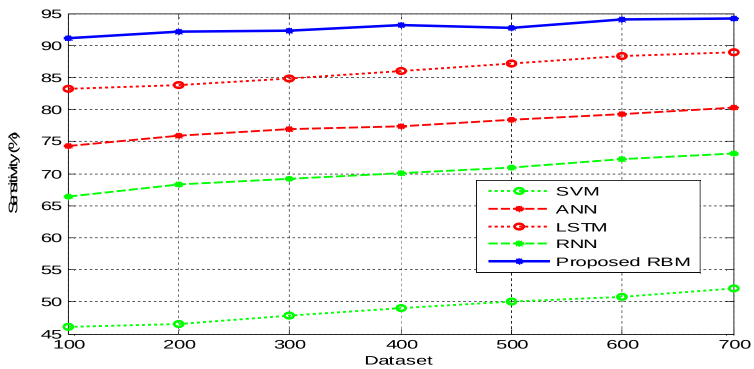

Figure 3: Sensitivity

Additionally, RNN temporal collaboration using time-series medical data yields promising outcomes in the area of early disease prediction and detection. In a study of decision-making in health informatics, the LSTM model trained on raw data outperformed the RBM network.

According to the analyst, RBM algorithms use a range of distinct features for the aforesaid data in order to provide an accurate output (Figure 2 and 3). When using supervised learning, the type of data used should be discriminative in nature. To summarize, we can state that the RBM algorithm outperforms other algorithms and that it may be applied in medical facilities such as diagnostic labs and hospitals that are integrated with IoT infrastructure in order to deliver faster diagnostic findings than other algorithms.

V. Conclusions

This research presents a RBM for the classification of the data in the big data pipeline collected from IoT medical devices. Deep learning architecture that recognizes and categorizes the qualities of the patient is used to assess the stability of the patient. Simulated outcomes are derived from real-time data collected from patients. The simulation demonstrates that massive volumes of collected data may be used to accurately predict patient stability in the future. Consequently, it is believed that the proposed method is a more effective tool for forecasting the stability of patients based on data collected through the IoT.

References

[1] Ahmed, I., Ahmad, M., Jeon, G., & Piccialli, F. (2021). A framework for pandemic prediction using big data analytics. *Big Data Research*, *25*, 100190.

[2] Silahtaroğlu, G., & Yılmaztürk, N. (2021). Data analysis in health and big data: a machine learning medical diagnosis model based on patients' complaints. *Communications in Statistics-Theory and Methods*, *50*(7), 1547-1556.

[3] Aldahiri, A., Alrashed, B., & Hussain, W. (2021). Trends in using iot with machine learning in health prediction system. *Forecasting*, *3*(1), 181-206.

[4] Vamsi, B., Doppala, B. P., Thirupathi Rao, N., & Bhattacharyya, D. (2021). Comparative Analysis of Prevalent Disease by Preprocessing Techniques Using Big Data and Machine Learning: An Extensive Review. *Machine Intelligence and Soft Computing*, 27-38.

[5] Li, W., Chai, Y., Khan, F., Jan, S. R. U., Verma, S., Menon, V. G., & Li, X. (2021). A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. *Mobile Networks and Applications*, *26*(1), 234-252.

[6] Riggins, F. J., & Wamba, S. F. (2015, January). Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics. In *2015 48th Hawaii international conference on system sciences* (pp. 1531-1540). IEEE.

[7] Majnarić, L. T., Babič, F., O'Sullivan, S., & Holzinger, A. (2021). AI and big data in healthcare: towards a more comprehensive research framework for multimorbidity. *Journal of Clinical Medicine*, *10*(4), 766.

[8] Ye, Y., Shi, J., Zhu, D., Su, L., Huang, J., & Huang, Y. (2021). Management of medical and health big data based on integrated learning-based health care system: A review and comparative analysis. *Computer Methods and Programs in Biomedicine*, *209*, 106293.

[9] Lu, H., Uddin, S., Hajati, F., Moni, M. A., & Khushi, M. (2022). A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. *Applied Intelligence*, *52*(3), 2411-2422.

[10] Dey, N., Hassanien, A. E., Bhatt, C., Ashour, A., & Satapathy, S. C. (Eds.). (2018). *Internet of things and big data analytics toward next-generation intelligence* (Vol. 35). Berlin: Springer.

[11] Rivera, J., & van der Meulen, R. (2014). Gartner says the internet of things will transform the data center. *Retrieved August*, *5*, 2014.

[12] Manogaran, G., Varatharajan, R., Lopez, D., Kumar, P. M., Sundarasekar, R., & Thota, C. (2018). A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system. *Future Generation Computer Systems*, *82*, 375-387.

[13] Dimitrov, D. V. (2016). Medical internet of things and big data in healthcare. *Healthcare informatics research*, *22*(3), 156-163.

[14] Rashid, M., Goyal, V., Parah, S. A., & Singh, H. (2019). Drug prediction in healthcare using big data and machine learning. In *Hidden link prediction in stochastic social networks* (pp. 79-92). IGI Global.

[15] Rashid, M., Hamid, A., & Parah, S. A. (2019). Analysis of streaming data using big data and hybrid machine learning approach. In *Handbook of Multimedia Information Security: Techniques and Applications* (pp. 629-643). Springer, Cham.

[16] Rashid, M., Singh, H., Goyal, V., Ahmad, N., & Mogla, N. (2022). Efficient big data-based storage and processing model in Internet of Things for improving accuracy fault detection in industrial processes. In *Research Anthology on Big Data Analytics, Architectures, and Applications* (pp. 945-957). IGI Global.

[17] UCI machine learning repository-heart disease data set: https://archive.ics.uci.edu/ml/datasets/heart+disease