ISSN- 2394-5125 VOL 07, ISSUE 12, 2020

Bigdata and cloud-based health care records monitoring using deep learning technology

[1]Dr BJD KALYANI, Assistant Professor, Department of Computer Science and Engineering, Priyadarshini Institute of Science and Technology for Women, Khammam, India.kjd_kalyani@yahoo.co.in

[2]Y venkateswara Reddy, Associate professor, Dept of ECE, Mallareddy institute of technology, Maisammaguda, Dulapally, Secunderabad, Telangana 500100

[3]K saikumar, JRF, Dept of ECE, Koneru Lakshmaiah Educational Foundation, Green Fields, Vaddeswaram, Andhra Pradesh 522502, saikumarkayam4@gmail.com

ABSTRACT

Big data has real time data-intensive processing that runs on high performance clusters. Big data computing and information sharing are carried out effectively using data pre-processing model in cloud environment. The noise and inconsistent present in the data obtained from various sources are removed with help of pre-processing which minimizes the time taken for computation and improves the rate of information sharing. Therefore, big data involves the process of collecting and sharing the information with better memory consumption. Health care information is provided with certain conditions that make fastest communication using big data for sharing medical data. For the distribution of medical information and data allocation in cloud environment with big data approach, an efficient PSM-PBC model is proposed. Cloud computing presents a cost-effective approach of providing facilities for computation and big data processing. The sharing of information about medical records presented with different management according to the proposed PSM-PBC model includes three processes. Initially, tridiagonal symmetric matrix is constructed in parallel on distributed patients' records with the help of big data applications.The size or complexity of the big data includes transaction and interaction of datasets that exceed regular technical capability in capturing, managing and processing data within reasonable cost. DSV-CP model is proposed for classifying the medical data to offer better communication in cloud environment. Initially, parallel processing mechanism ensures minimum runtime of medical data among different organizers by using linguistic fuzzy rules based on the MapReduce parallel programming model. While reducing the runtime of medical data, a large number of data can be processed within a stipulated time that ensures various aspects in medical field. Then canopy shuffle algorithm is applied to the resultant linguistic fuzzy rules to train a different sample set that accelerates classification accuracy. Similarly, the convergence rate of canopy fuzzy MapReduce algorithm is accelerated. Finally, a hybrid classification model is developed to improve the classification time and search accuracy in parallel manner based on fuzzy knowledge and canopy fuzzy MapReduce algorithm.

Keywords: deep learning, bigdata, cloud, medical records

ISSN- 2394-5125 VOL 07, ISSUE 12, 2020

I. INTRODUCTION

There are various types of data included in medical field such as analyzed report, clinical and genomic information present in different organizers. Privacy preserving is one of the better solutions for sharing the information that are stored and managed in medical data representation. By using big data communication and information sharing, individual medical reports are shared with the other medical organizers without any damage on applications. Let an example of sharing medical data among different medical organizers be considered.Figure 1 represents the medical communication with big data approaches in different medical aspects [1]. The distribution on medical information between different medical organizers is a smart purpose of cloud computing that provides more security.This model provides an efficient computation on big data applications and allocation of information in cloud computing environment. Initially, pre-processing based on IED is performed in DSV-CP model that helps to remove the noise and inconsistent medical data that are obtained from various sources.

After performing the data pre-processing task, DSV-CP model uses support vector prediction classifier to effectively classify big data in cloud. A framework named as parallel LFR-CM is proposed for big data classification and distribution of medical data in cloud environment. At the same time, several numbers of medical data are processed for sharing specific information and for providing more security.

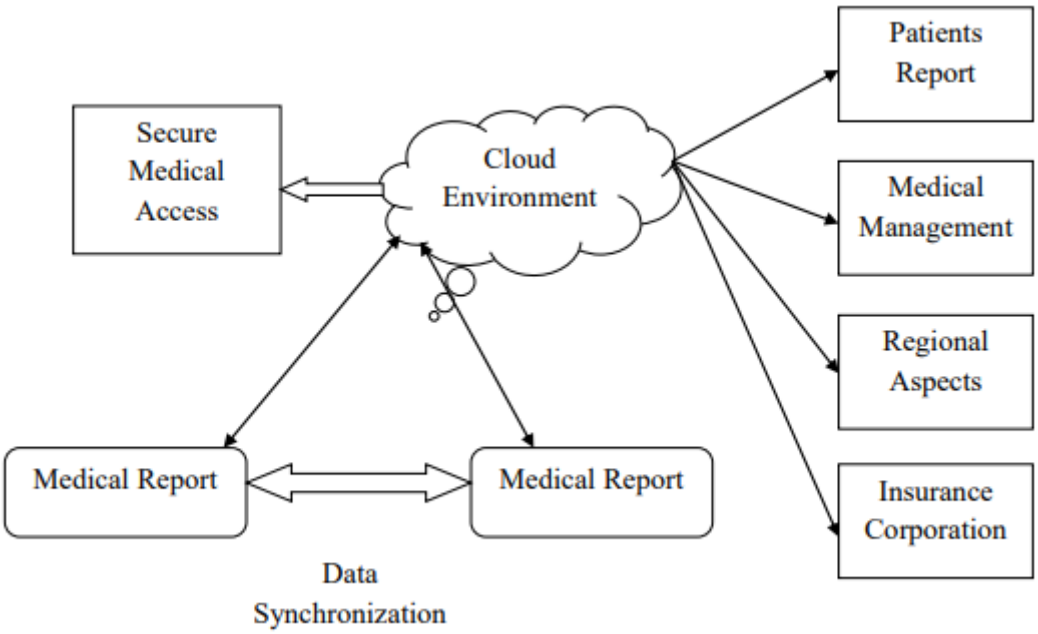

Figure 1 Big Data Communication using Medical Organizers in Medical Field

There are various types of medical data located in medical organizer for communication. Hence Cross-validated bayes classifier model is developed in the proposed PSM-PBC which evaluates real-value diagonal search data for its corresponding query results given by medical report. Thus, the result obtained from each user's medical request increases the prediction rate for identifying the medical data about medical users.

ISSN- 2394-5125 VOL 07, ISSUE 12, 2020

Finally, MapReduce function is enhanced with bayes classes which present predictive analytics about big data for better data computation and information sharing[2]. The MapReduce function in PSM-PBC model processes a large amount of big data in a parallel manner. While removing the noise and inconsistency present in medical data, computation time and space complexity are reduced based on different medical user request. Sharing of data is classified using support vector prediction classifier with parallel hyperplanes that improve the classification accuracy, i.e., user request information on big data. Finally, the proposed DSV-CP model accurately predicts the user request information on big data with the classified data [3].

To analyze the performance of big data computation and information sharing in Cloud environment, the three proposed methods, namely PSM-PBC, DSV-CP and LFR-CM are implemented using Java language with Amazon EC2 cloud Stanford Large Network dataset collection. The experiments are conducted on HDFS two-layer namespace and it offers distinct resource configurations for several virtual machine instances. Each virtual machine instance type is configured with a particular amount of memory, CPUs and local storage [4]. Nodes in the Stanford Large Network dataset collection in Amazon network represent products and edges that link commonly co-purchased products. The proposed models are also equipped with two quad core 2.33-2.66 GHz Xeon processors (8 cores total), 7 GB RAM, and 1690 GB local disk storage. The information sharing with big data is done in an efficient manner in cloud computing environment with HDFS two-layer namespace [5]. CloudSim simulator is used to measure the experimented parameters and the results are presented for different sizes of big data considering the user requests [6].

II. PERFORMANCE ANALYSIS OF THE PARAMETERS BASED PSM-PBC MODEL, DSV-CP MODEL AND LFR-CM FRAMEWORK

The performance of the proposed PSM-PBC model, DSV-CP model and LFR-CM framework is estimated and compared with the existing method named as MRPR method. The results are analyzed based on the parameter in terms of the following:

1. Classification accuracy
2. Prediction rate
3. Search accuracy

Performance Analysis of Classification Accuracy The classification accuracy measures the number of correct classifications regarding the total number of big data instances in training dataset classified. The classification accuracy ‘Ai ’ of an individual instance in training dataset ‘i’ depends on a number of data correctly classified and it is measured in terms of percentage (%) is evaluated by the formula as in Equation (1).

$$A_{i=} \frac{DCC}{n} *100 \qquad (1)$$

ISSN- 2394-5125 VOL 07, ISSUE 12, 2020

In Equation (1), 'DCC' signifies the number of Data Correctly Classified and 'n' is the total number of data considered for evaluation for measuring the Classification accuracy [7]. When the classification accuracy is higher, the method is said to be more efficient.

III. Analysis of ClassificationAccuracy

	Classification Accuracy (in %)			
Size of Big Data (in GB)	**Existing MRPR**	**Proposed MRPR**	**Proposed PSM-PBC**	**Proposed DSV-CP**
20	**54.68**	**62.31**	**69.12**	**74.3**
40	**56.73**	**63.8**	**71.54**	**76.12**
60	**58.14**	**65.62**	**73.41**	**77.89**
80	**59.34**	**66.12**	**74.87**	**79.12**
100	**61.53**	**68.97**	**76.12**	**82.34**
120	**63.19**	**69.34**	**78.95**	**83.78**
140	**65.82**	**72.87**	**79.81**	**85.67**
160	**67.91**	**73.16**	**80.23**	**86.12**
180	**69.98**	**75.63**	**83.68**	**88.47**
200	**72.43**	**77.98**	**85.48**	**89.36**

Table1ResultsforClassificationAccuracy

Table1revealstheresultsofclassificationaccuracy withrespect todifferentdatasizesintherangeof20GBto200GB,anditismeasured in termsofpercentage.Whentheclassificationaccuracyishigher,themethod issaidtobemoreefficient [8].Table1showsthecomparison ofclassification accuracyoftheproposedmethods,namelyPSM-PBC model,DSV-CP model andLFR-CMframeworkwiththeexistingMRPRmethod.

ISSN- 2394-5125 VOL 07, ISSUE 12, 2020

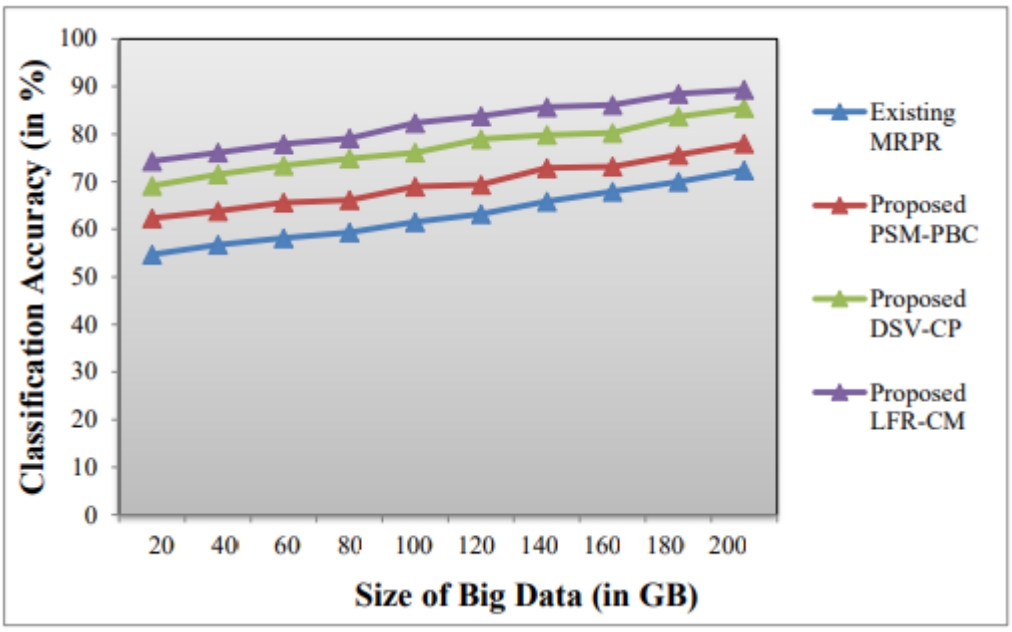

Figure2 PerformanceAnalysisofClassificationAccuracy

Figure 2 illustrates the classification accuracy of the PSM-PBC model, DSV-CP model and LFR-CM framework with the existing MRPR method for visual comparison based on the varied big data size. For example, the classification accuracy with the data size of 200GB using the PSM-PBC model is 62.31%, DSV-CP model is 69.12% and LFR-CM framework consists of 74.3%. Similarly, the existing MRPR method consists of 54.68% of classification accuracy. The classification accuracy is increased in the proposed LFR-CM framework because of the application of canopy shuffle MapReduce algorithm [9].

To perform well organized and efficient classification of medical data in various medical organizers, the canopy shuffle MapReduce algorithm in LFR-CM framework is used with the highest weight, i.e., rule weight, as the most robust one. This helps in improving the classification accuracy with regard to correct classification using LFR-CM framework. Therefore, the classification accuracy is improved using LFR-CM framework by 31.05% compared to the existing MRPR method. Similarly, the other proposed PSMPBC model and DSV-CP model improve the classification accuracy by 10.67% and 23.06% when compared with the existing MRPR method [10] [11]. Hence the proposed LFR-CM framework provides better classification on medical field with the improved accuracy

IV. PERFORMANCE ANALYSIS OF PREDICTION RATE

Prediction rate is presented for predicting the user requests based on the current data and historical facts. It is the practice of extracting information from the existing user queries in order to share the results with the other users and to predict future outcomes and trends.

$$Prediction\,Rate = \frac{Current\,data(size) + Historical\ \ facts(size)}{Size\,of\ \ BigData} * 100 \qquad (2)$$

ISSN- 2394-5125 VOL 07, ISSUE 12, 2020

From Equation (6.2), the prediction rate is determined by summing up the current data size obtained and the size of historical facts with respect to the size of big data. When the prediction rate is high, the method is said to be more efficient [12]. The prediction rate using the proposed PSM-PBC model, DSV-CP model and LFR-CM framework is compared with the existing MRPR method as in Table 2. The framework is also considered with different big data sizes in the range of 200 GB to 2000 GB for experimental purpose using Java language

Size of Big Data (in GB)	Existing MRPR	Proposed MRPR	Proposed PSM-PBC	Proposed DSV-CP
20	55.36	65.23	71.58	59.31
40	58.45	66.82	73.17	62.75
60	59.12	68.47	73.87	64.31
80	62.13	69.38	76.12	66.27
100	62.96	72.12	77.98	67.81
120	64.89	73.89	79.34	69.37
140	65.12	76.21	81.68	70.12
160	69.32	78.38	85.34	74.93
180	71.24	81.64	86.57	76.81
200	72.06	84.62	89.43	79.34

Table2 Results for Prediction Rate

The prediction rate is measured in terms of percentage. From the Table 2, it is obvious that the prediction rate using the proposed DSV-CP model is higher when compared with the other methods [13]. Figure3 visually compares the prediction rate using the proposed PSM-PBC model, DSV-CP model and LFR-CM framework with the existing MRPR method based on the varied big data size. For example, the prediction rate with the data size of 200 GB using the PSM-PBC model is 65.23%, DSVCP model is 71.58% and LFR-CM framework consists of 59.31%. Similarly, the existing MRPR method consists of 55.36% of prediction rate. The prediction rate is improved in the proposed DSV-CP model because of the application of support vector prediction classifiers.

ISSN- 2394-5125 VOL 07, ISSUE 12, 2020

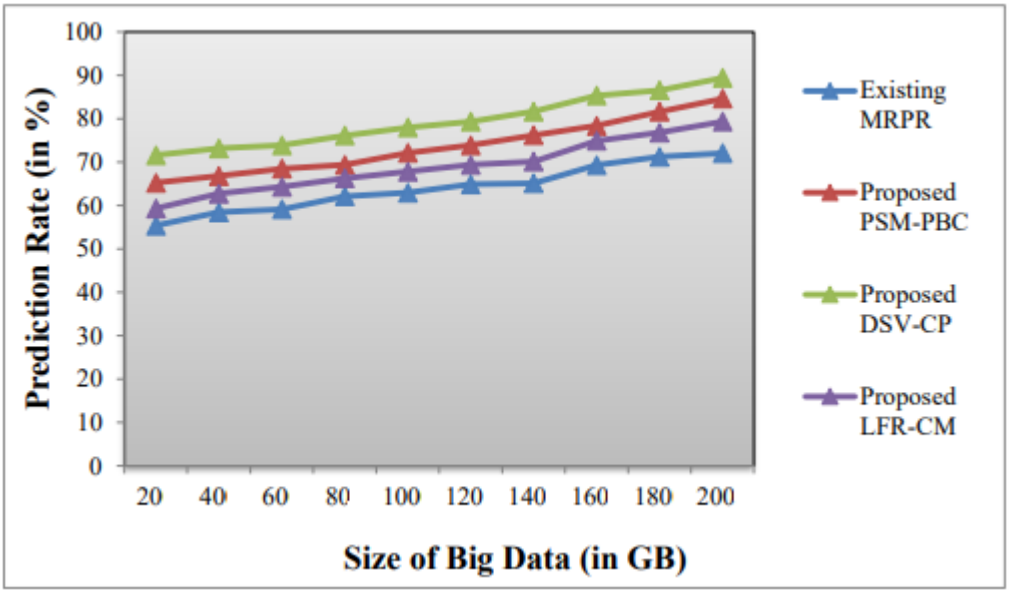

Figure 3 Performance Analysis of Prediction Rate

According to the medical data representation, support vector prediction classifier uses parallel hyperplanes to restrict hyperplane that passes through the origin by selecting optimal hyperplane to classify the big data. Therefore, the prediction rate is improved using DSV-CP model by 24.22% when compared with the existing MRPR method [14]. Similarly, the other proposed PSM-PBC model and LFR-CM framework improve the prediction rate by 15.02% and 7.82% than that of MRPR method. Hence the proposed DSV-CP model is said to provide better prediction rate on medical data.

V. PERFORMANCE ANALYSIS OF SEARCH ACCURACY

Search accuracy on big data refers to the measure of total amount of correctly identified patterns, i.e., medical reports made by the particular patient, with the size of data. It is measured in terms of percentage.

$$Search\ Accuracy(\%) = \frac{Correctly\ identified\ patterns}{Size\ of\ Big\ Data} * 100 \qquad (3)$$

Size of Big Data (in GB)	Existing MRPR	Proposed MRPR	Proposed PSM-PBC	Proposed DSV-CP
20	52.31	66.89	63.71	56.4
40	53.68	69.25	65.84	57.89
60	55.33	71.23	68.12	59.12
80	56.87	72.64	69.84	61.24
100	58.71	74.85	71.12	62.96
120	60.58	76.98	74.86	65.47

ISSN- 2394-5125 VOL 07, ISSUE 12, 2020

140	**62.12**	**77.158**	**76.37**	**67.24**
160	**64.53**	**79.52**	**78.43**	**69.32**
180	**66.47**	**81.32**	**82.34**	**72.27**
200	**68.93**	**82.35**	**83.81**	**73.86**

Table 3 Results for Search Accuracy

Table 3 shows the results of the performance of search accuracy on big data using the proposed PSM-PBC model, DSV-CP model and LFRCM framework in addition to the existing method MRPR. The big data size ranges from 20 GB to 200 GB and the results confirm that with the increase in the size of big data, the search accuracy also gets increased.

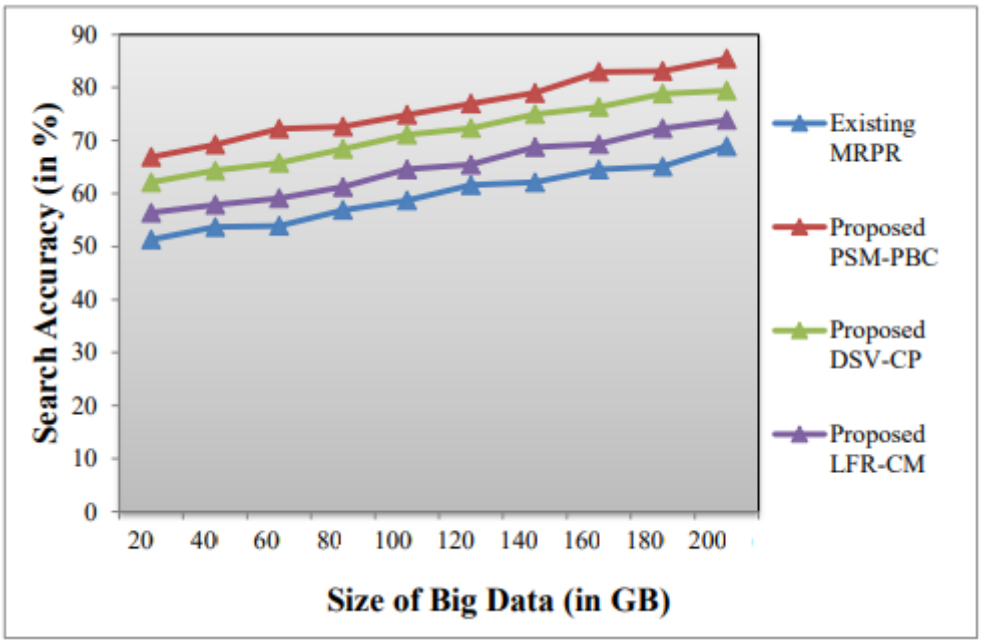

Figure 4 Performance Analysis of Search Accuracy

Figure 4 illustrates the search accuracy based on the size of big data that ranges from 20 GB to 200 GB used in cloud paradigm. The proposed PSM-PBC model, DSV-CP model and LFR-CM framework are compared with the existing method MRPR for experimental conditions [15]. The proposed PSM-PBC model is found to perform relatively well when compared with the other methods due to the use of tridiagonal symmetric matrix. Search accuracy in medical field is applied by using tridiagonal symmetric matrix to present a large number of reports about the patients. Tridiagonal symmetric matrix is applied to identify patterns and correlations in a significant manner that is made parallel across distributed cloud [16]. This improves the search accuracy which fits the user requirement, and as a result, it increases accuracy. Furthermore, using Householder transformation in PSM-PBC model, 'M*N' matrix is reduced to tridiagonal pattern, and therefore improves the search accuracy. Hence the search accuracy is improved by 25.72% using PSM-PBC model when it is compared to the existing MRPR method. Similarly, the other proposed DSV-CP model and LFR-CM framework improve search accuracy by 22.50% and 7.70% when compared with the existing

ISSN- 2394-5125 VOL 07, ISSUE 12, 2020

MRPR method. Hence the proposed PSM-PBC model provides better search accuracy on medical representation.

VI. SUMMARY

The proposed PSM-PBC model, DSV-CP model and LFR-CM framework are perfectly evaluated in this chapter. Theoretical analysis and experimental result reveal that the proposed methods are designed for achieving efficient big data computation and information sharing in cloud computing environment. Initially, PSM-PBC model avoids the computationally expensive power and space complexity problem in cloud environment. Tridiagonal symmetric matrix model shares the information in cloud environment to increase coarser construction with search accuracy on big data. The MapReduce function on the Bayes classes provide efficient predictive analytics about big data for efficient computation and information sharing in parallel manner. Then DSV-CP model performs data pre-processing task to effectively classify big data in cloud. It significantly reduces the misclassification errors, therefore improves the search accuracy and prediction accuracy of the user request information on big data. The data pre-processing task efficiently removes the noise and inconsistent data in the dataset which, in turn, reduces the computation time and space complexity effectively. LFR-CM framework is also designed for big data classification and information sharing in cloud environment. The framework uses linguistic fuzzy rules and procures database in an adhoc manner using triangular membership function in cloud environment using fuzzy logic. By applying canopy shuffle algorithm, different sample sets are trained, which ensure minimum runtime for big data classification. Finally, the big data classification is performed in a hybrid manner with MapReduce to improve the classification time and accuracy in cloud environment

REFERENCES

[1] . Yakushev, A & Mityagin, S 2014, ‘Social networks mining for analysis and modeling drugs usage’, Procedia Computer Science, vol. 29, pp. 2462-2471. Yang, T, Qian, K, Lo, DC-T, Xie, Y, Shi, Y & Tao, L 2016,

[2] . J. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, and H. Bustince, “Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system,” Applied Soft Computing Journal,2013, Article in Press.

[3] . J. Sanz, D. Bernardo, F. Herrera, H. Bustince, and H. Hagras, “A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data,” IEEE Transactions on Fuzzy Systems, 2014.

[4] . M. Lichman, “UCI machine learning repository,” 2013.

[5] . Li, Z, Yang, C, Jin, B, Yu, M, Liu, K, Sun, M & Zhan, M 2015, ‘Enabling big geoscience data analytics with a cloud-based, mapreduce-enabled and service-oriented workflow framework’, PLoS ONE, vol. 10, no. 3, pp. e0116781-e0116781.

[6] . J. Gantz and D. Reinsel, “THE DIGITAL UNIVERSE IN 2020: Big data, bigger digital shadows, and biggest growth in the far east,” 2012.

[7] . Wang, R, He, YL, Chow, CY, Ou, FF & Zhang, J 2015, ‘Learning ELM-Tree from big data based on uncertainty reduction’, Fuzzy Sets and Systems, vol. 258, pp. 79-100.

ISSN- 2394-5125 VOL 07, ISSUE 12, 2020

[8] . Souliotis, K, Kani, C, Papageorgiou, M, Lionis, D & Gourgoulianis, K 2016, 'Using big data to assess prescribing patterns in Greece: The case of chronic obstructive pulmonary disease', vol. 11, no. 5, pp. e0154960-e0154960.

[9] . Baro, E, Degoul, S, Beuscart, Rg & Chazard, E 2015, 'Toward a literature-driven definition of big data in healthcare', BioMed Research International, vol. 2015.

[10] . Kchaou, H, Kechaou, Z & Alimi, AM 2015, 'Towards an Offloading Framework based on Big Data Analytics in Mobile Cloud Computing Environments', Procedia Computer Science, vol. 53, pp. 292-297.

[11] . Kashyap, H, Ahmed, HA, Hoque, N, Roy, S & Bhattacharyya, DK 2014, 'Big Data Analytics in Bioinformatics: A Machine Learning Perspective', Journal of Latex Class Files, vol. 13, no. 9, pp. 1-20.

[12] . Youssef, AE 2014, 'A Framework for Secure Healthcare Systems Based on Big Data Analytics in Mobile Cloud Computing Environments', International Journal of Ambient Systems and Applications (IJASA), vol. 2, no. 2.

[13] . Dinov, ID, Heavner, B, Tang, M, Glusman, G, Chard, K, Darcy, M, Madduri, R, Pa, J, Spino, C, Kesselman, C & others 2016, 'Predictive Big Data Analytics: A Study of Parkinson's Disease Using Large, Complex, Heterogeneous, Incongruent, Multi-Source and Incomplete Observations', PLoS ONE, vol. 11, no. 8, pp. e0157077-e0157077.

[14] . Satagopam, V, Gu, W, Eifes, S, Gawron, P, Ostaszewski, M, Gebel, S, Barbosa-Silva, A, Balling, R & Schneider, R 2016, 'Integration and Visualization of Translational Medicine Data for Better Understanding of Human Diseases', Big data, vol. 4, no. 2, pp. 97-108.

[15] . Fisch, D, Kalkowski, E & Sick, B 2014, 'Knowledge fusion for probabilistic generative classifiers with data mining applications', IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 3, pp. 652-666.

[16] . Peralta, D, del Río, S, Ramírez-Gallego, S, Triguero, I, Benitez, JM & Herrera, F 2015, 'Evolutionary feature selection for big data classification: A mapreduce approach', Mathematical Problems inEngineering, vol. 2015.

[17] . Inukollu, V, Arsi, S & Ravuri, Sr. 2014, 'Security Issues Associated With Big Data in Cloud Computing', International Journal of Network Security & Its Applications, vol. 6, no. 3, pp. 45-56.

[18] . Tcheng, DK, Nayak, AK, Fowlkes, CC & Punyasena, SW 2016, 'Visual recognition software for binary classification and its application to spruce pollen identification', PLoS ONE, vol. 11, no. 2.