

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Adaptive Neural Network Classifier-Based Analysis of Big Data in Health Care

Manaswini Pradhan

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.77225>

Abstract

Because of the massive volume, variety, and continuous updating of medical data, the efficient processing of medical data and the real-time response of the treatment recommendation has become an important issue. Fortunately, parallel computing and cloud computing provide powerful capabilities to cope with large-scale data. Therefore, in this paper, a FCM based Map-Reduce programming model is proposed for the parallel computing using AANN approach. The FCM based Map-Reduce, clusters the large medical datasets into smaller groups of certain similarity and assigns each data cluster to one Mapper, where the training of neural networks are done by the optimal selection of the interconnection weights by Whale Optimization Algorithm (WOA). Finally, the Reducer reduces all the AANN classifiers obtained from the Mappers for identifying the normal and abnormal classes of the newer medical records promptly and accurately. The proposed methodology is implemented in the working platform of JAVA using CloudSim simulator.

Keywords: fuzzy C-means clustering (FCM), adaptive artificial neural networks (AANN), whale optimization algorithm (WOA), parallel computing, Map-Reduce model

1. Introduction

Big Data has been characterized by its three properties i.e. 1. Volume, 2. Velocity and 3. Variety. Volume refers to the huge amount of data being generated by several sources. Velocity refers to the rate at which this data is being generated and the variety refers to the different type of the data being used [1]. Now a days with so much of data all around the world, the trend in healthcare is shifting from cure to prevention. Hospitals and healthcare systems are good repositories of big data (like patient records, test reports, medical images etc.) that can be

utilized to cut the cost in healthcare, to improve reliability and efficiency, and to provide better cure to patients.

Healthcare applications require large amounts of computational and communication resources, and involve dynamic access to large amounts of data within and outside the health organization leading to the need for networked healthcare [2]. Data Analysis has always in demand in all the industry as it gives the approximate prediction of how the market is growing [3]. Although the innovations are in the healthcare field, there are some issues that need to be solved, particularly the heterogeneous data fusion and the open platform for data access and analysis [4].

Today, the healthcare industry is turning to big data technology to improve and manage medial systems. For this purpose, healthcare companies and organizations are leveraging big data in health informatics [5]. The analysis of big data carried out through different ways. Machine learning algorithm helps in analysis of big data very efficiently [3].

Feature selection is an important preprocessing technique used before data mining so that it can reduce the computational complexity of the learning algorithm and remove irrelevant/redundant features to remove noise [16]. Decision Tree is a predictive model of classification, which can be viewed as a Tree like structure [6]. It is simple and gives a fast and accurate result.

Neural Network is one of the other machine learning algorithms which showed a lot of modification. Neural Network is an adaptive learning model which adjusts the weight of the connecting links between its neuron [15]. K-Nearest Neighbor model of classification is one of the simplest classification algorithm which work on the classifying the data set based on the nearest neighbor of the existing class label of already trained mode [7]. Naïve Bayesian Classifier has a very good accuracy in classification for large set of data [8].

Clustering algorithm makes the groups or clusters of homogenous data. It is an unsupervised learning technique. In Partitioned Clustering the number of cluster was defined beforehand. In Hierarchical Clustering we do not need to define the number of clusters in advance [9, 10]. In both of the above approaches the stopping criterion is usually the number of clusters to be achieved; once the required number is achieved the algorithm can be stopped. Different methods are used for the analysis of Big Data in Health Care has been discussed below.

2. Literature survey

Abdulsalam Yassine et al. [11] have proposed a model that utilizes smart home big data as a means of learning and discovering human activity patterns for health care applications. They proposed the use of frequent pattern mining, cluster analysis and prediction to measure and analyze energy usage changes sparked by occupants' behavior.

Md. Mofijul Islam et al. [12] have proposed a mobility- and resource aware joint virtual-machine migration model for heterogeneous mobile cloud computing systems to improve the performance of mobile Smart health care applications in a Smart City environment.

Mohammad-Parsa Hosseini et al. [13] focused on an autonomic edge computing framework for processing of big data as part of a decision support system for surgical candidacy, an optimized model for estimation of the epileptogenic network, and an unsupervised feature extraction model.

Bernhard Schölkopf et al. [14] have designed a class of support vector algorithms for regression and classification.

Chandra et al. [15] have proposed a approach for using MLP to handle Big data. There was high computational cost and time involved in using MLP for classification of Big data having large number of features. This is a promising technique for handling big data and is the idea extracted for the present research work.

Huan Liu et al. [16] have introduced a concepts and algorithms of feature selection, surveys existing feature selection algorithms for classification and clustering, groups and compares different algorithms with a categorizing framework based on search strategies, evaluation criteria, and data mining tasks.

Malika Bendeche et al. [17] have proposed a distributed clustering approach to deal efficiently with both phases; generation of local results and generation of global models by aggregation.

3. Proposed FCM Map-Reduce based adaptive neural network classifier for handling big data in health care

The large amounts of data, driven by record keeping, compliance & regulatory requirements, and patient care will historically render for the healthcare industry. While most data is saved in hard copy form, the current trend is towards quick digitization of these large amounts of data. Driven by mandatory requirements and the potential to develop the quality of healthcare delivery meanwhile minimizing the costs, these massive quantities of data known as 'big data' hold the promise of supporting a wide range of medical and healthcare functions, admitting between others clinical decision support, disease surveillance, and population health management. Some troubles that exist in big data analysis in health care are, i) to succeed, big data analytics in healthcare requires to be packaged so it is menu driven, user-friendly and transparent. ii) The lag among data collection and processing has to be addressed. iii) The crucial managerial issues of ownership, governance and standards have to be conceived. iv) Continuous data acquisition and data cleansing is another issue.

In the increasingly quick generation of large amounts of data, and across several areas of science, technological and conceptual advances are resulting. The collection and organization of data, the volume, variety, and velocity of current 'big data' production inaugurates novel opportunities and challenges in both scale and complexity these are always admitted on research. Also, in health care sector, the dealing of big data has currently get an interesting research topic, as since there are wide amount of medical data's available in cloud storage.

Moreover, the huge number of data records within very large datasets that comprise an extremely high amount of information is conceived to be a very critical issue. Thus processing with sequential algorithm results in greater computational cost in terms of memory space and time complexities. Hence, for discovering the above mentioned issues, a parallel architecture is required to be demonstrated.

In order to minimize the computational complexity and the memory requirement while leading large healthcare data, it is suggested to have a parallel adaptive artificial neural network (AANN) technique applying Map-Reduce programming model for health care analysis from big data in cloud environment. The introduction of abnormality in the medical data records applying the proposed Map-Reduce based Adaptive Artificial Neural Network classification method by the trained data these are determined by suggested approach. The medical data from the cloud has to be first clustered in order to distinguish the similar classes of data associated to any one particular health disorder for better classification of data. Here, the clustering of similar sets of data is done with the help of Fuzzy C means clustering algorithm, so as to develop the classification performance. The dataset separated as sub clusters were afforded to Map-Reducer framework, where AANN is implemented in parallel. Once the clustering is done the normal and abnormal classes of medical data are then learned applying the proposed map decrease programming model based AANN. By training the AANN models, it can be capable to predict for newer data as well.

The map minimized programming model comprises of two phases: (1) Mapper phase and (2) Reducer phase. Data belonging to each cluster are mapped applying separate mappers. Each mapper based AANN receives one training item (i.e. any one data cluster) and then calculates the weights of the network applying the training item in the suggested parallel prediction model. Here, to develop the precision of classification of the data, the proposed AANN method applies the concept of optimization, where the weight factors are maximized by applying Whale Optimization Algorithm. The Reducer separates the test medical record in order to distinguish the health condition established on the mapped data.

Here, the schematic diagram of the proposed healthcare application model for the analysis of large datasets is presented in **Figure 1**.

The proposed FCM based Map-Reduce AANN approach comprises of the following phases, 1) Fuzzy C-means (FCM) based Data Grouping 2) Mapper phase involving assigning each data groups to separate Mappers and training Data using Adaptive ANN 3) Reducer Phase consisting of Testing Phase. Each of the steps is detailed in the following sections.

3.1. Phase 1: Fuzzy C-means (FCM) based data grouping

Established on the membership function, Fuzzy C-means (FCM) is a data clustering technique in which each and every data in that group will comes under one cluster. It will group all the data in to particular number of clusters in high dimensional search space. The degrees of the cluster are determined by the membership function in terms of $[0, 1]$ which affords the flexibility that the data point can belong to more than one cluster.

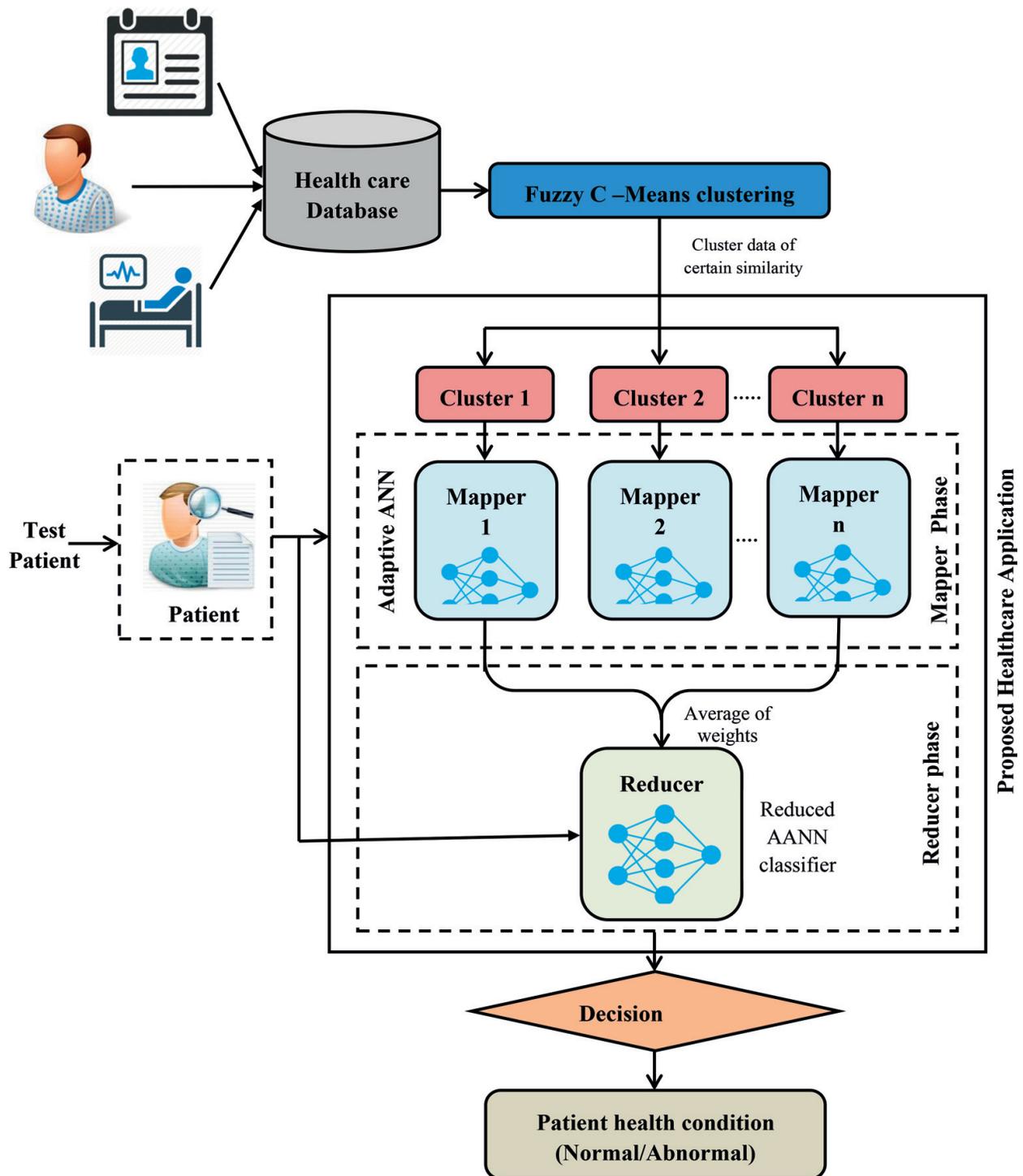


Figure 1. Schematic diagram.

The proposed method applies FCM for clustering the input large data into smaller groups of similar data. The input data will be grouped into number of clusters randomly and the centroids will be rendered for the clusters during Fuzzy c-means clustering. The clusters are

updated established on the membership grade of the data points and the novel centroid is depicted correspondingly at the each iteration.

Moreover, how the clustering with fuzzy c means algorithm is made for a set of input samples is afforded below.

Let us considering the input sample be,

$$D_u (u = 1, 2, \dots, n) \quad (1)$$

The input sample is to be separated into 'v' number of clusters. The clustering cannot be exactly but it will be made by means of the grouping with respective to the grade of membership function.

The objective function of FCM algorithm is effectively explained as follows.

$$ObjecFun = \sum_{u=1}^n \sum_{v=1}^C M_{uv} \|D_u - c_v\|^2 \quad (2)$$

where,

" M_{uv} " is the membership of u^{th} data (D_i) in the v^{th} cluster c_v .

" c_v " is the v^{th} cluster center.

" D_u " is the u^{th} data record.

" n " is the total number of data record.

" C " is the required number of clusters.

" $\|D_u - c_v\|^2$ " is the similarity between u^{th} data record and the center vector of v^{th} cluster.

Now the cluster center calculation is done by Eq. (3),

$$c_v = \frac{\sum_{u=1}^n M_{uv} D_u}{\sum_{u=1}^n M_{uv}} \quad (3)$$

Membership updation is done by Eq. (4),

$$M_{uv} = \frac{1}{\sum_{y=1}^C \left(\frac{\|D_u - c_v\|}{\|D_u - c_y\|} \right)^{\frac{2}{b-1}}} \quad (4)$$

where, 'b' is the fuzziness coefficient. The membership matrix $M^{(x)} = [M_{uv}]$ is calculated for among every iteration. If $\|M^{(x)} - M^{(x-1)}\| < T$ then stop, Where, "x" denotes the current iteration and "T" is the threshold of termination criterion, which is among 0 and 1.

The input data is clustered into data groups of certain similarity for established on the above procedure of FCM. We found the number of cluster set such as $C_1, C_2, C_3, \dots, C_n$ at the end of the FCM process. For the parallel implementation, all the clusters are aligned to divide the mappers.

3.2. Phase 2: Mapper phase

For large scale mobile data process, the mapper is a programming model and a connected implementation. Programmers only required to specify a Map-Reduce job which is composed of Reducer functions and the mapper. A Mapper function receives a key/value pair and generates a set of intermediate key/value pairs. With the same intermediate key, and a Reducer function merges all intermediate values are connected. Here, in parallel, the Mapper receives the clustered data and trains the AANN. Then established on all the Mappers output network model, the Reducers improve an AANN model to predict for unknown/newer data.

3.3. Phase 2(a): assigning each data groups to separate Mappers

In the Mapper phase, the clustered data is now processed. Mapper receives several items of the training sets (i.e. data items from the cluster groups) and accomplishes many mapper tasks. Each Mapper receives one training item (i.e. data items from one cluster group) and then performs AANN learning/training task by maximizing the weight values in the network applying this training item; so as to develop the learning efficiency. Through the AANN algorithm, their outputs are the trained network models resulted. The Mapper process (i.e. the AANN training procedure) is accomplished repeatedly until the expected precision is attained.

3.4. Phase 2(b): training data clusters using parallel adaptive ANN

Artificial neural network is otherwise named as Neural Network (NN). For calculation, it contains of an interconnected collecting of artificial neurons and procedures data applying a connectionist approach. Here a feed forward neural network (FFNN) inaugurated by this work. The data moves in just a single direction, forward, from the input layers, through the hidden layers, and to the output layers by this system. There are no cycles or circles in the system. The information handling can stretch out over numerous (layers of) units, yet no criticism associations are available, that is, an association reaching out from outputs of units to contributions of units in a similar layer or past layers is not present. There are associations among the processing elements (PEs) in every layer that have a weight (parameter) connected with them. Amid preparing this weight is balanced. The proposed adaptive ANN renders the optimal training network aligned by optimally selecting the interconnection weights among the hidden and output layers applying Whale Optimization Algorithm.

Input information is displayed to the system and proliferated through the system until it attains the output layer in FFNN. A predicted output is delivered by this forward procedure. The desired output is subtracted from the actual output and error esteem for the systems is ascertained. The error function can be characterized as:

$$e(w) = \sum (desired - actual\ output)^2 \quad (5)$$

For altering weights, a couple of traditional researches has applied Backpropagation learning algorithm. In reverse through the system, the calculation begins with the weights among the output layer PE's and the last hidden layer PE's and works. Once back propagation has fulfilled, the forward procedure begins once more, and this cycle proceeds until the error among is predicted and actual output are reduced. Rather than back propagation algorithm, the Whale Optimization algorithm is displayed because it can acquire valuable output than back propagation calculation.

The proposed Adaptive Artificial Neural Network model is given in below **Figure 2**.

a. Whale optimization approach

Recently a novel optimization algorithm named whale optimization algorithm (Mirjalili 2016) has been introduced to metaheuristic algorithm by Mirjalili and Lewis. As highly intelligent animals with motion, the whales are conceived. The WOA is inspired by the unique hunting behavior of humpback whales. The humpback whales prefer to hunt krills or small fishes which are close to the surface of sea at normally. Humpback whales use a special unique hunting method named bubble net feeding method. In this method they swim around the prey and produce distinctive bubbles along a circle or 9-shaped path. The mathematical model of WOA is described in the following sections a) Encircling prey b) Bubble net hunting method

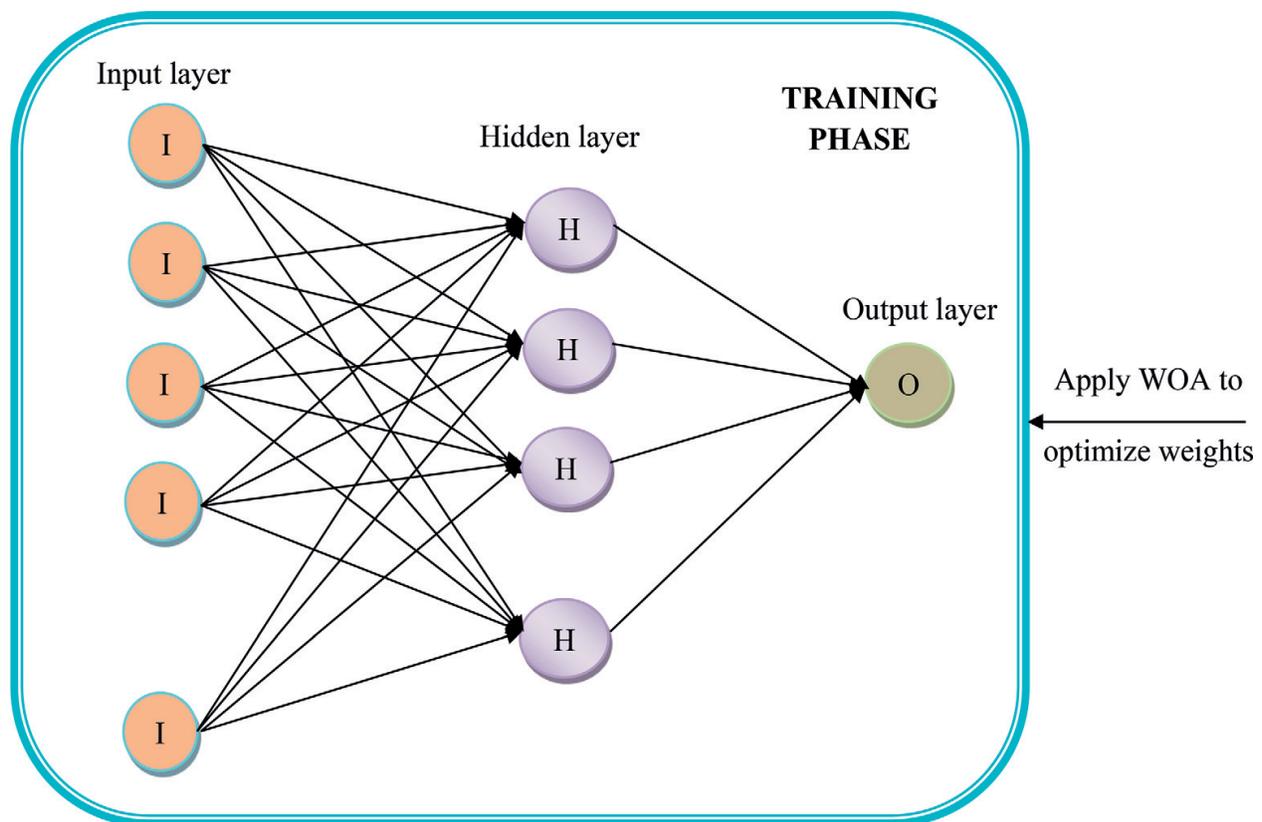


Figure 2. Proposed adaptive artificial neural network.

and c) Search the prey. The steps admitted in the proposed Whale optimization algorithm for rendering the optimal network structure by maximizing the interconnection weights of the neurons are afforded as follows,

Step 1: Initialization.

The algorithm is showed by arbitrarily generating the solutions (i.e. the interconnection weight values) that communicates to the result. Here the neural network structure comprising the interconnection weights among the hidden layers and the output layers are referred by the random value in the search space is afforded as:

$$W_m = \{w_{m,1}, w_{m,2}, \dots, w_{m,h}, \dots, w_{m,R}\} \quad (6)$$

where, $\{w_{m,1}, w_{m,2}, \dots, w_{m,h}\}$ represents the set of weights among the input and hidden layer and $\{w_{m,h}, \dots, w_{m,R}\}$ represents the set of weights among the input and hidden layer. Also, each solution, $W_m = \{w_{m,1}, w_{m,2}, \dots, w_{m,h}, \dots, w_{m,R}\}$ is a R -dimensional vector where R being the number of optimization parameters. And also start the coefficient vectors of whale such as, \vec{f} , \vec{F} , and \vec{G} .

Step 2: Fitness Calculation.

Determine the fitness of the input solutions on the basis of the Eq. (7). To get the best weight values, the fitness value of the solutions are computed. It's revealed in below,

$$Fit_{W_m} = \min(MSE) \quad (7)$$

The minimum of mean square error (MSE) determines that, how correct the network predicted targets are (i.e. high classification accuracy) in above equation. Hence, for further development, the initial solution with minimum error is chosen as best solution and checked.

Step 3: Update position of current solutions towards the best

A. Encircling prey

The position of prey (i.e. the current best solution) is distinguished by humpback whale and then it encircles the prey. Towards the best search operator the other search operators will consequently attempt to update their positions when the best search agent is characterized. The updation method is determined by the below equations:

$$\vec{N} = \left| \vec{G} \cdot \vec{W}_m^{best}(x) - \vec{W}_m(x) \right| \quad (8)$$

$$\vec{W}_m(x+1) = \vec{W}_m^{best}(x) - \vec{F} \cdot \vec{N} \quad (9)$$

where ' $\vec{W}_m(x+1)$ ' denotes the newer solutions for next iteration, \vec{F} and \vec{G} denotes a Coefficient vector, \vec{W}_m^{best} denotes a position vector for best solution, $\vec{W}_m(x)$ represents a current position Vector and $||$ represents an absolute value.

The vectors \vec{F} and \vec{G} are calculated as follows:

$$\vec{F} = 2 \cdot \vec{f} \cdot \vec{k} - \vec{f} \quad (10)$$

$$\vec{G} = 2 \cdot \vec{k} \quad (11)$$

where, \vec{f} is linearly reduced from 2 to 0 during the course of iterations (in both exploration and exploitation phases), $\vec{k} \in (0, 1)$.

B. Bubble-net attacking method (exploitation phase)

To model the bubble-net behavior of humpback whales mathematically two approaches developed are a) Shrinking encircling mechanism and b) Spiral updating position

a. Shrinking encircling mechanism

The value of \vec{f} in the Eq. (10) is reduced to attain this behavior. Note that \vec{f} is applied to reduce the variation range of, \vec{F} . In other words, where \vec{f} is minimized from 2 to 0. The novel position of a search agent can be determined anywhere by setting the random value, \vec{F} from $[-1, 1]$.

b. Spiral updating position

A spiral equation among the position of whale and prey is produced to mimic the helix-shaped movement of humpback whales is as follows:

$$\vec{W}_m(x+1) = N' \cdot \exp^{pq} \cdot \cos(2 \prod q) + \vec{W}_m^{best}(x) \quad (12)$$

where $N' = \left| \vec{W}_B^{best}(x) - \vec{W}_m(x) \right|$ and denotes the distance of the B^{th} whale (which is the best solution obtained so far) to the prey, q is the random value from, $[-1, 1]$, p denotes the shape of the logarithmic spiral and it is a constant value. During maximization the position of whales is updated by assuming a probability of 50% by choosing either the spiral model or shrinking encircling mechanism to model this simultaneous behavior. The mathematical model is afforded by Eq. (13).

$$\vec{W}_m(x+1) = \begin{cases} \vec{W}_m^{best}(x) - \vec{F} \cdot \vec{N}, & \text{if } L < 0.5 \\ N' \cdot \exp^{pq} \cdot \cos(2 \prod q) + \vec{W}_m^{best}(x), & \text{if } L \geq 0.5 \end{cases} \quad (13)$$

where, $L \in [0, 1]$. The humpback whales search for prey randomly to form bubble net.

c. Search for prey (exploration phase)

To search for prey in exploration phase, the same search approach applied in the exploitation phase established on the variation of the \vec{F} vector can be applied. In fact, allowing to the position of each other humpback whales search randomly. Therefore, to force search agent to move so far from a reference whale we use \vec{F} with the random values greater than 1 or less

than -1 . In exploitation phase, the position of the search agent is updated. This mechanism and $|\vec{F}| > 1$ emphasize exploration permit the WOA algorithm to perform a global search. The mathematical model is afforded below:

$$\vec{N} = \left| \vec{G} \cdot \vec{W}_m^{rand}(x) - \vec{W}_m(x) \right| \quad (14)$$

$$\vec{W}_m(x+1) = \vec{W}_m^{rand}(x) - \vec{F} \cdot \vec{N} \quad (15)$$

where, $\vec{W}_m^{rand}(x)$ is a current population random position vector. Search agents update their positions at each iteration with respect to either the best solution found so far or a randomly selected search agent. In order to render exploration and exploitation the parameter ' f ' is reduced from 2 to 0, respectively. A random search agent is selected when $|\vec{F}| > 1$, while the best solution is chosen when $|\vec{F}| < 1$ for the position of the search agents for updating. Depending on the value of ' L ', WOA is able to switch development either a circular or spiral movement.

The solutions are updated established on the best search agent between the current solutions found from the fitness evaluation step during each iteration. Again, at each time of generating newer weight values, it is aligned to the network and the fitness is determined and established on the back propagation error (i.e. the min MSE), which is the fitness function.

Step 4: Termination criteria.

Once the optimal weights are generated for all the networks of the Mappers, the training of the networks is finished. Now the AANN becomes a classifier and it can be generalized to predict for newer data also. The output mapped networks are then forwarded to Reducer phase.

To create the optimal network structure, the WOA algorithm is finished when best weight values are found. Also, the satisfaction of a termination criterion is confirmed when the mean square error is decreased to the needed limit or when the maximum iteration is attained.

Once the optimal weights are rendered for all the networks of the Mappers, the training of the networks is completed. Now the AANN gets a classifier and it can be generalized to predict for newer data also. The output mapped networks are then forwarded to Reducer phase.

3.5. Phase 3: Reducer phase

A Reducer accepts the data element of each Mapper for each Reducer task. With the same intermediate key, and a Reducer function merges all intermediate values connected. Established on the requirement, the Reducers can be customized. The proposed healthcare analysis model needs only one Reducer for improving a classifier model that must separate the patient's medical records. Here, the Reducer task is to form a robust classifier model from the parallelly trained network models. Since the Reducer results in only one classifier network model, it will average all the maximized weight results for each interconnection links found for each training item and find the final optimal weights of the classifier. Here the analyzing

data's (i.e. the unknown/newer data) are separated in the minimized AANN classifier model found from the Reducer phase.

4. Result and discussion

This section comprises result and discussion about the proposed parallel AANN (Adaptive Artificial Neural Network) technique for health care analysis from big data in cloud environment. The proposed algorithm is accomplished through JAVA software and the experimentation is carried out applying a system of having 4 GB RAM and 2.10 GHz Intel i-3 processor.

For estimating the performance of the proposed FCM based accuracy, Map-Reduce model, time, memory, precision, and recall are taken into an account and equated with the existing k-means based Map-Reduce and DBSCAN model. The experimental results for the suggested FCM based Map-Reduce model and other being k-means based Map-Reduce model and DBSCAN are tested in this section. The prediction efficiency is evaluated established on differentiating the number of records and number of mappers.

4.1. Performance analysis

The performance judgment of the proposed FCM based Map-Reduce model to predict the inauguration of abnormality in the medical data records is established in this section and equated with accomplishing k-means based Map-Reduce and DBSCAN method. The efficiency of our proposed method is evaluated in terms of time, memory, precision, recall and accuracy established on number of records and number of mappers.

4.2. Case (1): Performance analysis based on varying data size

4.2.1. Time

In the medical data records, an effective method should minimize the time needed to predict the inauguration of abnormality. The proposed FCM based Map-Reduce model decreases the time while equating with other accomplishing k-means based Map-Reduce and DBSCAN method.

Figure 3 establishes the time needed for prediction of abnormality applying our proposed FCM based Map-Reduce model and k-means based Map-Reduce and DBSCAN model while the number of records rises. This clearly establishes that our proposed FCM based Map-Reduce model decreases the time needed for prediction of abnormality while equating with other accomplishing k-means based Map-Reduce and DBSCAN method.

4.2.2. Memory

An effective method should decrease the requirement of memory. The proposed FCM based Map-Reduce model decreases the requirement of memory while equating with other accomplishing k-means based Map-Reduce and DBSCAN method.

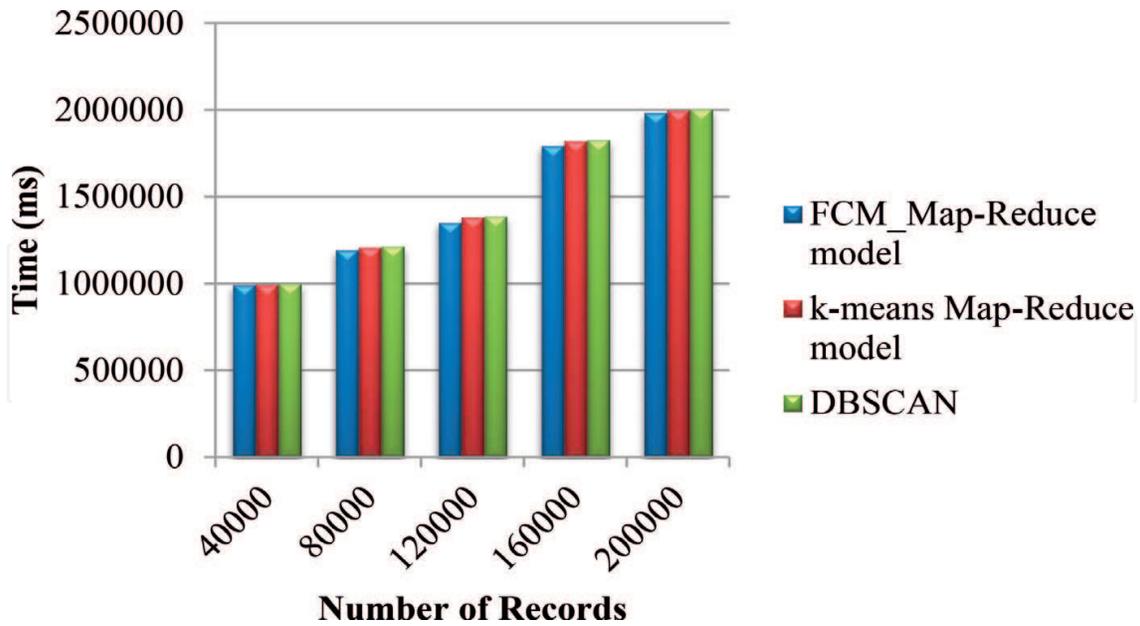


Figure 3. Time taken by FCM based Map-Reduce; k-means based Map-Reduce model for prediction.

Figure 4 demonstrates the memory needed for our proposed FCM based Map-Reduce model and k-means based Map-Reduce and DBSCAN model while the number of records increases. This clearly establishes that our proposed FCM based Map-Reduce model decreases the memory requirement while equating with other accomplishing k-means based Map-Reduce and DBSCAN method.

4.2.3. Precision

The method with high precision will be more efficient. The proposed FCM based Map-Reduce model maximizes the precision while equating with other being k-means based Map-Reduce and DBSCAN method.

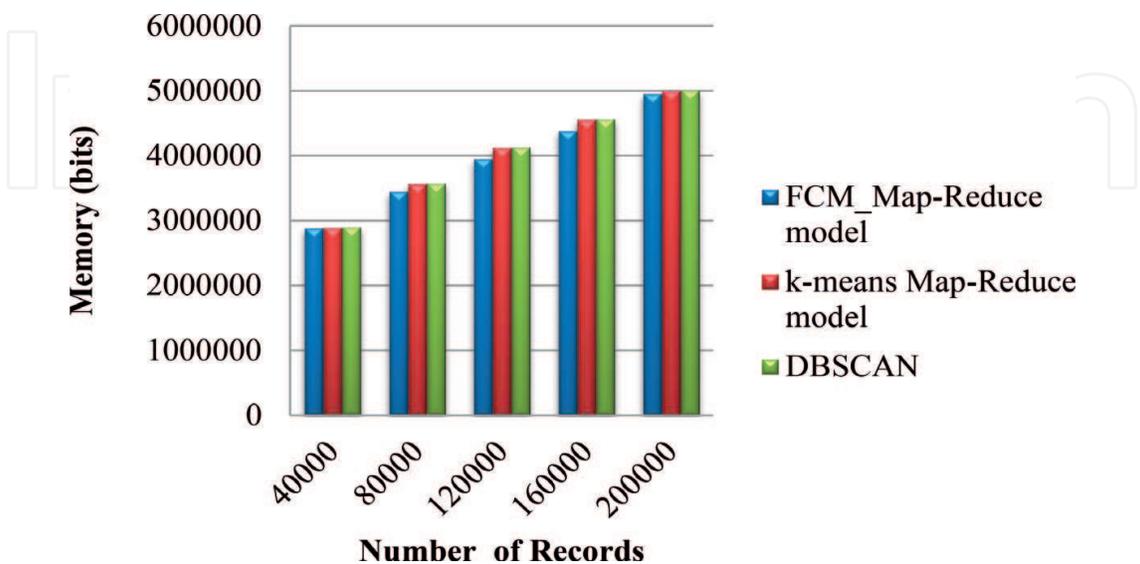


Figure 4. Memory requirement for FCM based Map-Reduce and k-means based Map-Reduce model.

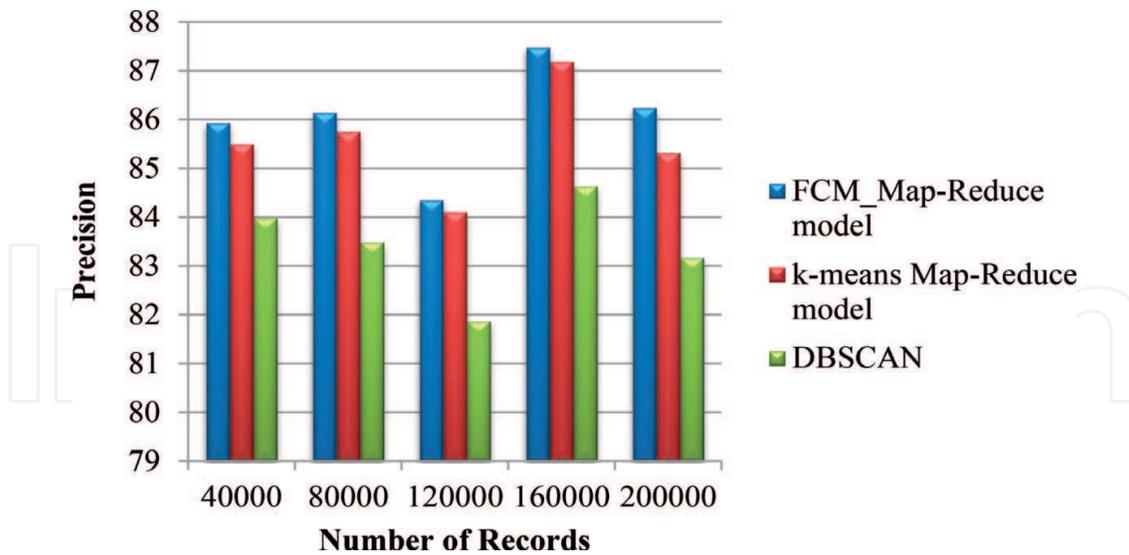


Figure 5. Precision for FCM based Map-Reduce; k-means based Map-Reduce and DBSCAN model.

Figure 5 establishes the precision level for our proposed FCM based Map-Reduce model and k-means based Map-Reduce and DBSCAN model while the number of records rises. This clearly demonstrates that our proposed FCM based Map-Reduce model rises the precision level while equating with other being k-means based Map-Reduce and DBSCAN method.

4.2.4. Recall

The method with high recall is said to be more effective. The proposed FCM based Map-Reduce model rises the recall while equating with other being k-means based Map-Reduce and DBSCAN method.

While the number of records rises, the **Figure 6** establishes the recall for the proposed FCM based Map-Reduce model and k-means based Map-Reduce and DBSCAN model. This clearly

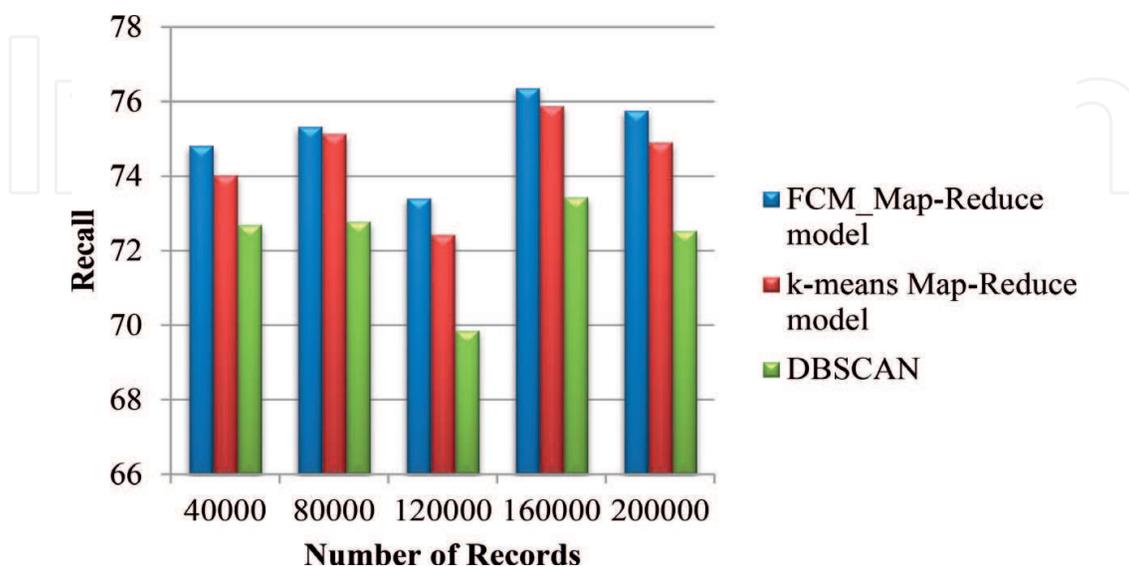


Figure 6. Recall for FCM based Map-Reduce; k-means based Map-Reduce and DBSCAN model.

establishes that the proposed FCM based Map-Reduce model rises the recall while equating with other being k-means based Map-Reduce and DBSCAN method.

4.2.5. Accuracy

The method with high accuracy is said to be more effective. With other being k-means based Map-Reduce and DBSCAN method, for the proposed FCM based Map-Reduce model raises the accuracy while comparing.

Figure 7 establishes the accuracy for the proposed FCM based Map-Reduce model and k-means based Map-Reduce and DBSCAN model while the number of records rises. This clearly demonstrates that the proposed FCM based Map-Reduce model raises the accuracy while equating with other accomplishing k-means based Map-Reduce and DBSCAN method.

4.3. Case (2): performance analysis by varying number of Mapper

4.3.1. Time

In the medical data records, an effective method should decrease the time needed to predict the presence of abnormality. The proposed FCM based Map-Reduce model decreases the time while equating with other accomplishing k-means based Map-Reduce and DBSCAN method.

Figure 8 establishes the time needed for prediction of abnormality applying the proposed FCM based Map-Reduce model and k-means based Map-Reduce model while the number of mapper rises. This clearly demonstrates that the proposed FCM based Map-Reduce model decreases the time needed for prediction of abnormality while equating with other accomplishing k-means based Map-Reduce method.

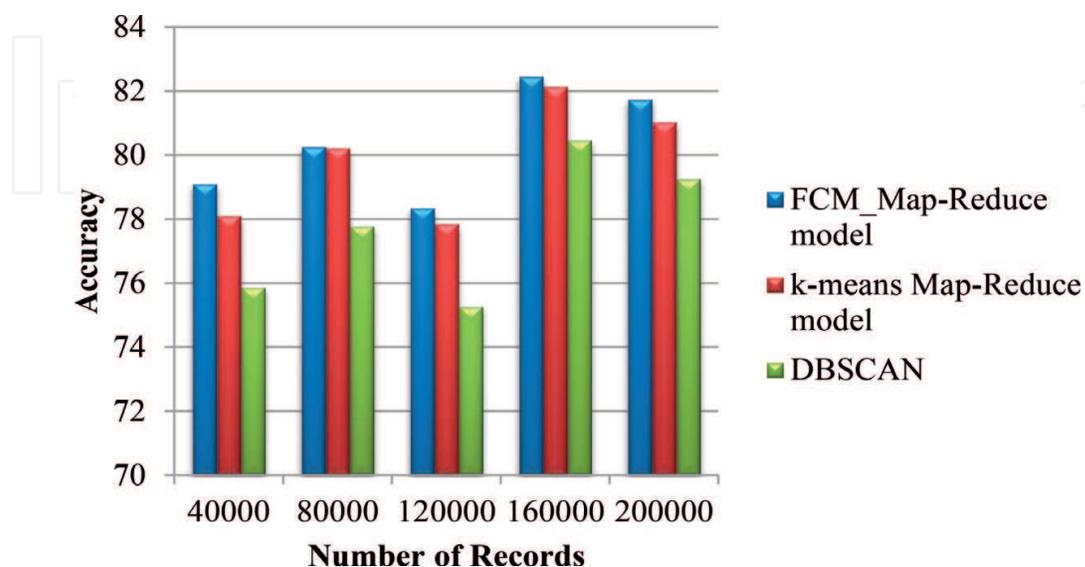


Figure 7. Accuracy for FCM based Map-Reduce; k-means based Map-Reduce and DBSCAN model.

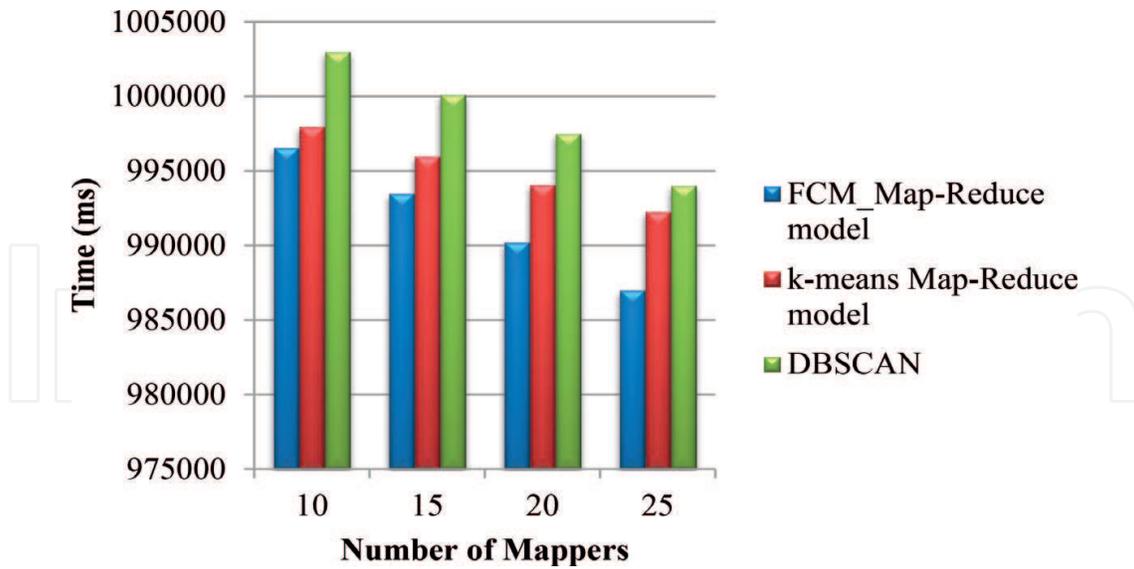


Figure 8. Time taken by FCM based Map-Reduce; k-means based Map-Reduce model for prediction.

4.3.2. Memory

An effective method should decrease the requirement of memory. While equating with other accomplishing k-means based Map-Reduce and DBSCAN method, for the proposed FCM based Map-Reduce model decreases the requirement of memory.

While the number of mapper rises, **Figure 9** establishes the memory needed for the proposed FCM based Map-Reduce model and k-means based Map-Reduce model and k-means base model. This clearly demonstrates that the proposed FCM based Map-Reduce model decreases the memory requirement while equating with other accomplishing k-means based Map-Reduce and DBSCAN method.

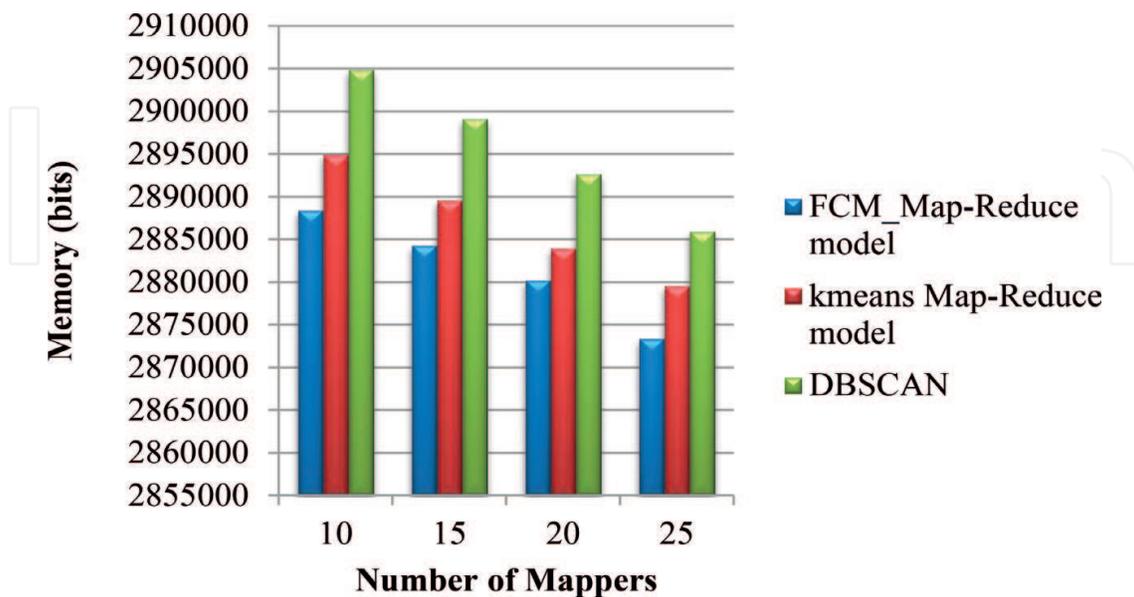


Figure 9. Memory requirement for FCM based Map-Reduce; k-means based Map-Reduce and DBSCAN model.

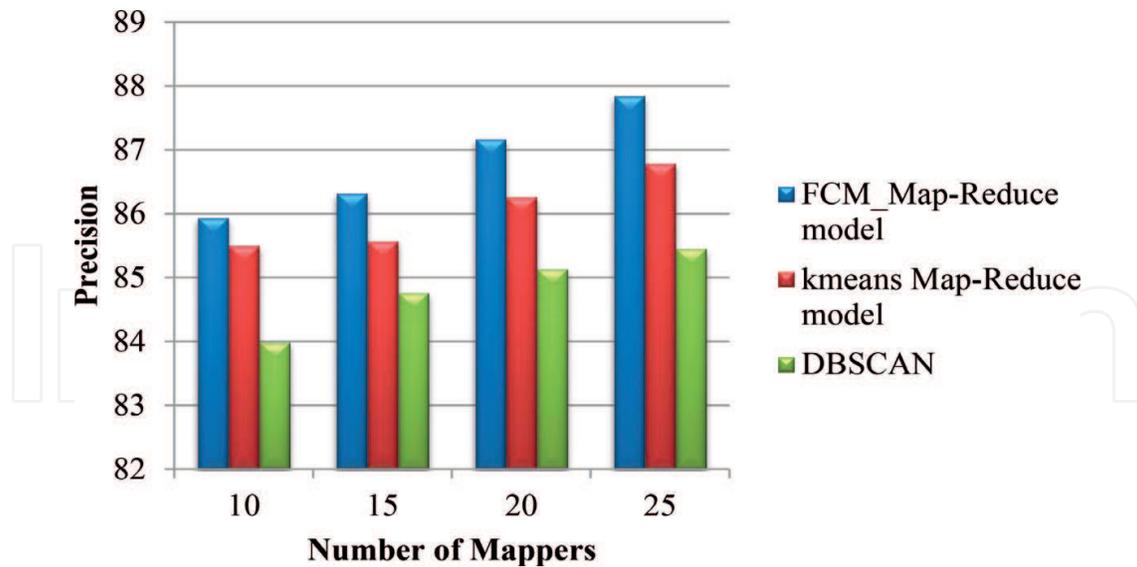


Figure 10. Precision for FCM based Map-Reduce; k-means based Map-Reduce and DBSCAN model.

4.3.3. Precision

The method with high precision will be more effective. The proposed FCM based Map-Reduce model raises the precision while equating with other accomplishing k-means based Map-Reduce method.

While the number of mapper rises, Figure 10 establishes the precision level for the proposed FCM based Map-Reduce model and k-means based Map-Reduce model. While equating with other accomplishing k-means based, Map-Reduce method this clearly demonstrates that the proposed FCM based Map-Reduce model rises the precision level.

4.3.4. Recall

The method with high recall is said to be more effective. Our proposed FCM based Map-Reduce model raises the recall while comparing with other accomplishing k-means based Map-Reduce method.

Figure 11 establishes the recall for the proposed FCM based Map-Reduce model and k-means based Map-Reduce and DBSCAN model while the number of mapper rises. This clearly demonstrates that the proposed FCM based Map-Reduce model raises the recall while equating with other accomplishing k-means based Map-Reduce and DBSCAN method.

4.3.5. Accuracy

The method with high accuracy is said to be more effective. With other being k-means based Map-Reduce and DBSCAN method, for the proposed FCM based Map-Reduce model raises the accuracy while comparing.

While the number of mapper increases, the Figure 12 establishes the accuracy for the proposed FCM based Map-Reduce model and k-means based Map-Reduce and DBSCAN model. This

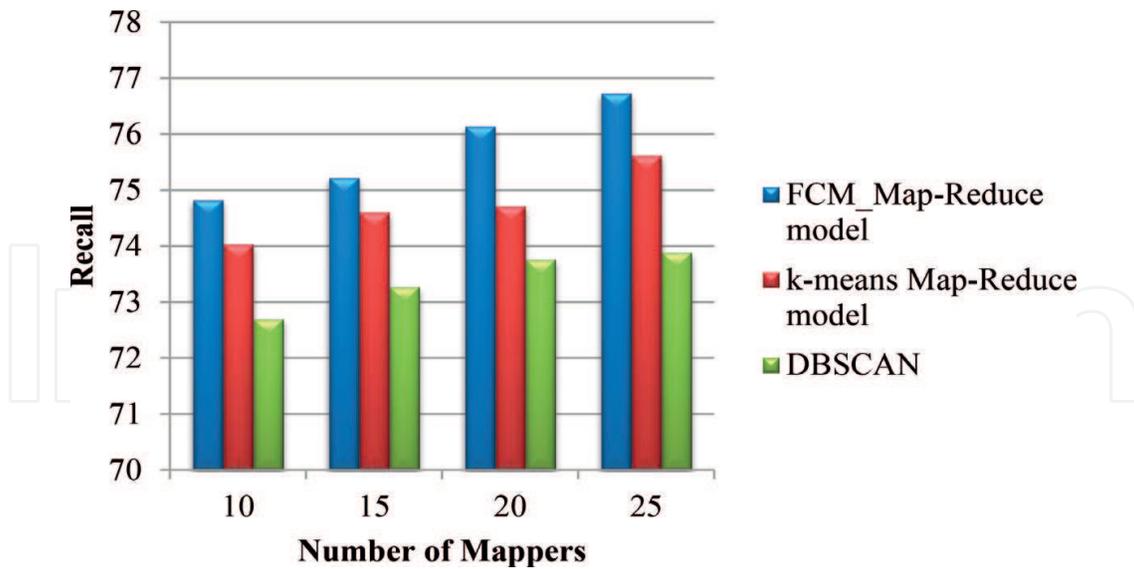


Figure 11. Recall for FCM based Map-Reduce; k-means based Map-Reduce and DBSCAN model.

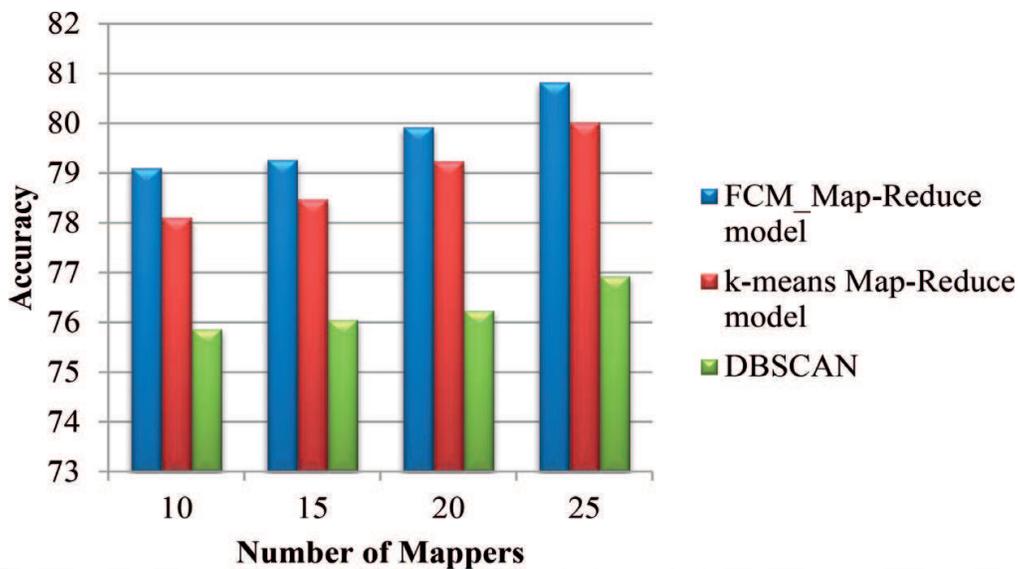


Figure 12. Accuracy for FCM based Map-Reduce; k-means based Map-Reduce and DBSCAN model.

clearly demonstrates that the proposed FCM based Map-Reduce model raises the accuracy while equating with other accomplishing k-means based Map-Reduce and DBSCAN method.

5. Conclusion

The presented research method have improved a FCM based Mapreduce programming model for the implementation parallel calculating applying Adaptive Artificial Neural Network approach for the prediction of abnormality of medical records. The proposed FCM

based Mapreduce model is equated with the accomplishing k-means based Mapreduce and DBSCAN model and tested in terms of different evaluates like time, memory, precision, recall and accuracy by differentiating the data size as well as the number of mappers. It can be seen from the results that, all the values found for the proposed method is better when equated to the being method. Moreover, the time and memory requirements are very much minimized when the number of mappers is raised. This establishes the efficiency of proposed model and so the proposed application can be applicable for handling large healthcare databases in cloud environment.

Author details

Manaswini Pradhan^{1,2*}

*Address all correspondence to: mrs.manaswini.pradhan@gmail.com

1 Visiting Researcher School of Computer Science and Software Engineering, East China Normal University, Shanghai, China

2 P.G. Department of ICT, Fakir Mohan University, India

References

- [1] Suthaharan S. Big data classification: Problems and challenges in network intrusion prediction with machine learning. *ACM Sigmetrics Performance Evaluation Review*. 2014; **41**(4):70-73
- [2] Tawalbeh L'a A, Mehmood R, Benkhelifa E, Song H. Mobile cloud computing model and big data analysis for healthcare applications. *IEEE Access*. 2016;**4**(99):1-12
- [3] Ramesh D, Suraj P, Saini L. Big data analytics in healthcare: A survey approach. In: *Microelectronics, Computing and Communications (MicroCom)*. 2016
- [4] Chandola V, Sukumar S, Schryver J. Knowledge discovery from massive healthcare claims data. In: *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. 2013. pp. 1312-1320
- [5] Fang R, Pouyanfar S, Yang Y, Chen S-C, Iyengar SS. Computational health informatics in the big data age: A survey. *ACM Computing Surveys (CSUR)*. July 2016;**49**(1):1-36
- [6] Wang Dingxian, Xiao Liu, and Mengdi Wang. "A DT-SVM strategy for stock futures prediction with big data." *Computational Science and Engineering (CSE)*, 2013 IEEE 16th International Conference on. IEEE, 2013
- [7] Bijalwan V et al. KNN based machine learning approach for text and document mining. *International Journal of Database Theory and Application*. 2014;**7**(1):61-70

- [8] Tomar D, Agarwal S. A survey on data mining approaches for healthcare. *International Journal of Bio-Science and Bio-Technology*. 2013;5(5):241-266
- [9] Fahad A et al. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *Emerging Topics in Computing, IEEE Transactions on*. 2014;2(3):267-279
- [10] Belciug Smaranda et al. "Clustering-based approach for detecting breast cancer recurrence." *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on*. IEEE, 2010
- [11] Yassine A, Singh S, Alamri A. Mining human activity patterns from smart home big data for healthcare applications. *IEEE Access*. 2017;5:13131-13141
- [12] Islam MM, Razzaque R, Hassan MM, Nagy W, Song B. Mobile cloud-based big healthcare data processing in smart cities. *IEEE Access*. 2017;5:11887-11899
- [13] Hosseini M-P, Tran TX, Pompili D, Elisevich K, Soltanian-Zadeh H. Deep learning with edge computing for localization of epileptogenicity using multimodal rs-fMRI and EEG big data. In: *Autonomic Computing (ICAC)*. 2017
- [14] Schölkopf B et al. New support vector algorithms. *Neural Computation*. 2000;12(5):1207-1245
- [15] Chandra B, Ratnesh K Sharma. "Fast learning for big data applications using parameterized multilayer perceptron." *Big Data (Big Data), 2014 IEEE International Conference on*. IEEE, 2014
- [16] Liu H, Yu L. Toward integrating feature selection algorithms for classification and clustering. *Knowledge and Data Engineering, IEEE Transactions on*. 2005;17(4):491-502
- [17] Belciug S, Gorunescu F, Salem A-B, Gorunescu M. Clustering-based approach for detecting breast cancer recurrence. In: *Intelligent Systems Design and Applications (ISDA)*. 2010