DOI: http://dx.doi.org/10.26483/ijarcs.v8i9.5034

ISSN No. 0976-5697

Volume 8, No. 9, November-December 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A NOVEL FRAMEWORK FOR CLASSIFICATION OF HEART ATTACK RECORDS FROM ABC OPTIMIZED CLOUD STORAGE

M.Shobeetha
Assitant professor, Dept. of. Computer science,
Bharathidasan University Constituent College(w)
Orthanadu, Thanjavur, Tamilnadu, India,

Dr.A.Shaik Abdul Khadir
Associate professor, Dept. of. Computer Science,
Khadir Mohideen College,
Adirampattinam, Thanjavur, Tamilnadu, India,

Abstract: Cardiac Attacks make worse and threatens the middle aged persons in the world. There are lot of researches are undergoing to make control the cardiac heart attacks. It is difficult to predict the causes and possibilities of heart attacks based on the historical patient records. In data mining, the existing techniques like Decision Tree Classification, Naïve Bayesian Algorithm are used to predict the possibilities of heart attacks and their symptoms. However it works well for the structured data. It will not evaluate the unstructured heterogeneous data from various resources. In this proposed work, the analysis performance should be improved to obtain the accurate results. To produce a mechanism for perform predictive analysis about the medical treatments through unstructured heterogeneous data using K Means clustering and SVM Classification. Here the system retrieves the data from the cloud storage. To optimize the data retrieval, here the framework implemented with the Artificial Bee Colony Optimization Algorithm. This helps to obtain the classified results for the preventive measures through the factors from historical data from various cloud resources. To improve the improvements in the analysis based on the mechanism implementation results should be compared with the existing algorithms using UCI dataset of heart attacks in weka tool.

Keywords: K Means Clustering, Decision Tree Classification, Naïve Bayesian Classification, Support Vector Machine classification.

1. INTRODUCTION

Today's Medical Treatments qualities are increasing day by day. More number of advancements has been come for make treatments to the patients. It is difficult to understand the patient health based on their fitness reports. Each patient is unique by their records. They should be treated based on their fitness records not based on the common treatment for heart attacks.

The prediction and analysis of the heart disease through the data mining techniques works good rather than the human intervention. In existing, the more number of techniques and algorithms are proposed for this above said without human intervention. In this works, the existing techniques used are Hierarchal Clustering and Naïve Bayes Classification. [1] These techniques are used to obtain the data analysis through the structured data resources. However these techniques make analysis and decision treatment prediction, it seems to be effective but need some fulfillments in obtaining the data. The data from the various heterogeneous data resources should be used to obtain the results from the huge data.

In existing [1] [2] [3] [4] proposed the framework based on Naïve Bayes Classification and compares this obtained result to existing to this algorithm to prove the effectiveness of the Navie bayes Algorithm. Rupali and Patil[1] were proposed a novel framework as Decision Support in the Heart Disease Prediction System using the data mining techniques such as Naïve Bayes Classification. Garima Singh et.al. [4] proposed this same features of algorithm as a online consultation project.

Nevertheless, it provides the prediction results, to improve the effectiveness of the prediction analysis; a new mechanism should be proposed to obtain the effectiveness in data prediction with heterogeneous data. Here the classification can be done through heterogeneous data retrieved from various cloud data sources. To improve the cloud data retrieval, the Artificial Bee colony Optimization techniques used in our proposed mechanism. In existing [9] [6], the Ant Colony Optimization and Particle Swam Optimization techniques were used for the cloud data retrieval optimization. However, the proposed method should obtain somewhat improved optimization results.

The major contributions of this paper are as follows:

1. To cluster the heterogeneous unstructured data, our proposed mechanism should use K Means Clustering Algorithm.
2. To obtain the improved prediction analysis results through the Cluster based Support vector Machine Classification.
3. Finally, the data retrieval process from the cloud resource should be optimized through the Artificial Bee Colony (ABC) optimization techniques.

The following sections will clearly describes the works related with our mechanism and working principle of our propose mechanism and implementation results.

2. RELATED WORKS

Rupali and Patil [1] finished their work with existing algorithms and compares the results over the proceeding algorithms. In [6], Mythili et.al proposed a Heart Disease Prediction Model using SVM-Decision Trees through the rule based model. They recommended to this process using the neural networks to improve the results.

Hayashi H et.al [7] and Talayeh et.al [8] shows the results for the SVM Classification. They utilized this algorithm based on perspective of patient health records. In [4] the authors proposed with the Gait Analysis to obtain the results for the Lumbar Spinal Stenosis. The author in [5]

proposed Multilevel weighted SVM for Classification with the missing values in the data set.

For optimization many successful exiting works are presented using the Ant Colony Optimization and Particle swarm Optimization. In existing works [9] [10] [11], the authors proposed load balancing algorithm policies to schedule the tasks. It deals with the load balancing to distribute the load equally to enhance the data retrieval and achieve scalability.

In this proposed work, the Artificial Bee Colony optimization algorithms it enhanced with the heuristics task scheduling. The related works for the ABC in [12] [13] [14], the authors proposed to design with ABC algorithm and improve the task scheduling and load balancing policies.

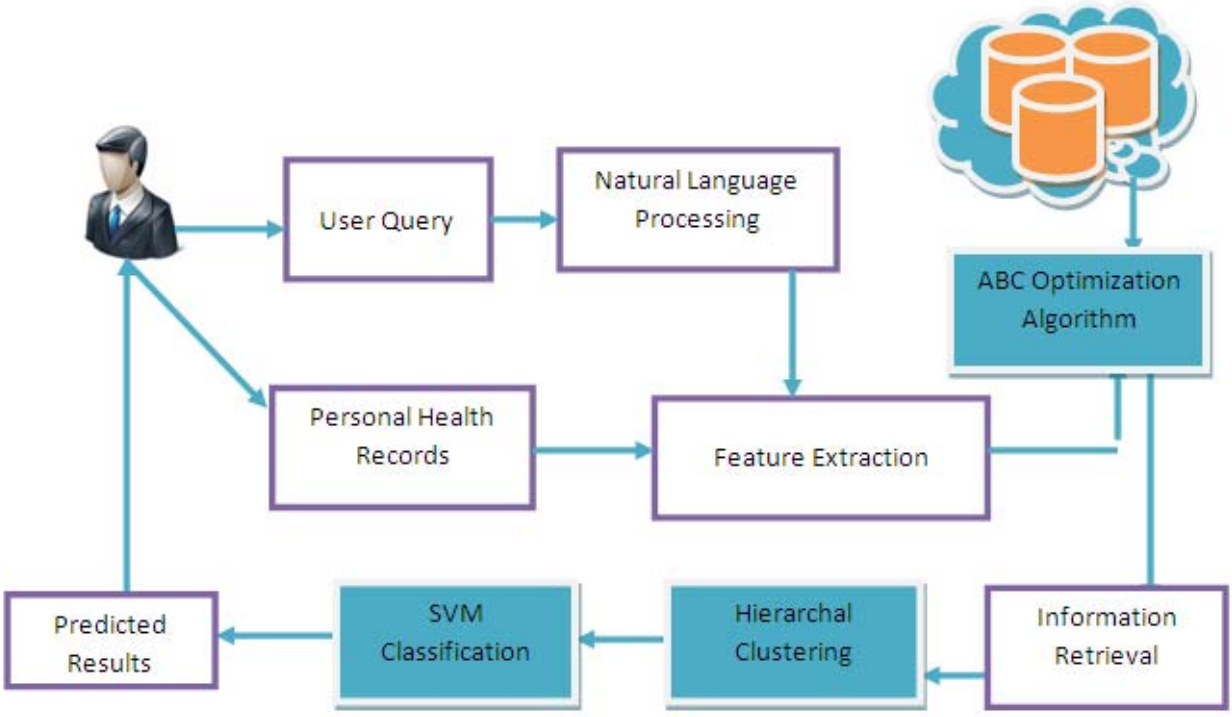

Fig. 3.1 Architecture for Proposed Mechanism

3. PROPOSED MECHANISM

The figure Fig.3.1 illustrates the architecture of proposed mechanism. In this architecture, the steps and procedure for the mechanism of the proposed work were described diagrammatically. The user should provide the patient health record as test data and the query should also be given to analyze the symptoms and causes of the heart attacks. The obtained queries should be processed to get the stem words through the Natural Language Processing. Natural Language Processing is a process technique for extract the features as stem words for the data retrieval. It contains some of the inner processes such as Lexical Analyzer, POS Tagger, Word Sense Disambiguation and Stemming.

In parallel to this NLP, using the patient health records, the feature can be extracted. Here we use 14 attributes as discussed in section IV Implementation Results. Thus the 14 attributes are used as input and using that the feature can be extracted as key for the data retrieval. The optimized cloud retrieval can be done through the inputs as stem words identified and processed test data. The optimization can be done through the effective algorithm such as Artificial Bee Colony Optimization algorithm. Thus the algorithm improves the performance of data retrieval. The retrieved heterogeneous data from various cloud resources can be sent to the next process as clustering to cluster the data items based on the meta data using Hierarchal Clustering because this mechanism can works with the Cluster based Support Vector Machine Algorithm to classify the predicted treatment.

The pseudo code for the proposed architecture can be obtained through the combination of two or more algorithms such as K Means Clustering and SVM Classification. The initial data entities can be obtained through the optimized entities from proposed ABC algorithm.

Algorithm : CB –SVM Frame work

Input : $E=\{e_1,e_2,e_3,\ldots,e_n\}$ entities
k- number of clusters
Ouput: $C=\{c_1,c_2,\ldots C_n\}$ cluster sets
$CL=\{cl_1,cl_2,\ldots,cl_n\}$
$R=\{R1,R2,\ldots,Rn\}$ classification set
Algorithm:
[1] obtain the entities E and no. of clusters
[2] foreach $c_i \in C$ do
[3] $c_i \leftarrow e_j \in E$
[4] end
[5] foreach $e_i \in E$
[6] $cl(e_i) \leftarrow E$
[7] end
[8] set ch=false; itr=0;

[9] repeat
[10] foreach e_i Є E do
[11] UpdateClusters(c_i)
[12] end

[13] foreach e_i Є E do
[14] minDist ← argminDistance(E);
[15] if(minDist ≠ cl(E) then
[16] ch=true;
[17] end
[18] end
[19] until ch=true and itr<k
[20] while there are interruption points do
[21] Find a candidate
[22] candidateSV=candidateSV ∪ candidate
[23] if any $\alpha_p < 0$ £ additions of S then
[24] candidaeSV = candidataSV / P
[25] repeat until such points pruned
[26] end if
[27] end while

The classification algorithm can be enhanced through the optimization algorithm. Here the framework can be optimized through the ABC optimization algorithm rather than the existing works its performance improvements are adequate in nature.

Algorithm : ABC Optimized CB-SVM

Input : D= Initial training Dataset
T= Test Data
Ouput: E={e_1,e_2,e_3,…,e_n}entities

Algorithm:
[1] obtain the initial dataset D as vector variables $\underset{x_m}{\rightarrow}$
[2] set criteria for fitness f $\underset{x_m}{\rightarrow}$= no
[3] check fitness of initial dataset
[4] repeat
[5] employee bee phase
[6] onlooker bee phase
[7] scout bee phase
[8] calculate fitness measure f
[9] criteria matching if exists then f= yes
[10] end until f=yes
[11] return E

In above algorithm prescribes the phases of the ABC algorithm to accomplish the optimization technique in CB-SVM. The optimization can be done through the three predefined phases. Thus the phases used to retrieve the entities from the training transaction data sources. There are three types of bees in this algorithm. The employee deals with the food sources or instances, the employed bees within the hive used to choose the a food source can be observed through the onlooker bees and the scout bees searching for food sources randomly. In initialization phase, the scout bees set the control parameters and also it initialized the population of food sources as $\underset{x_m}{\rightarrow}$ are initialized as m=1,2,…,SN, SN : dataset size), $\underset{x_m}{\rightarrow}$ is the solution vector to the optimization problem, each $\underset{x_m}{\rightarrow}$ contains n variables, $x_{mi}, i = 1,2 \ldots n.$ are to be optimized obtained from the dataset D so as to minimize the objective function.

The following notation might be used for the initialization purpose

$$x_{mi} = l_i + rand(0,1) * (u_i - l_i)$$

where l_i and u_i are the lower and upper bound of the parameter x_{mi}, respectively.

Im employee bee phase the new food sources $(\underset{v_m}{\rightarrow})$ are collected from the neighbourhood food source $(\underset{x_m}{\rightarrow})$ as training dataset. This can be obtained through the evaluation of the profitability as fitness. The neighbour food source $(\underset{v_m}{\rightarrow})$ calculated using the below equation.

$$v_{mi} = x_{mi} + \emptyset_{mi}(x_{mi} - x_{ki})$$

Where x_k is a randomly selected food source, I is a randomly chosen parameter index and $\emptyset_{mi}$ is a raondom number within the range [-c,c]. After defining the new food source v_{mi}, its fitness

$$fit_m(\overrightarrow{x_m}) = \begin{cases} \dfrac{1}{1 + f_m(\overrightarrow{x_m})} & if\ f_m(\overrightarrow{x_m}) \geq 0 \\ 1 + abs\left(f_m(\overrightarrow{x_m})\right) & if\ f_m(\overrightarrow{x_m}) < 0 \end{cases}$$

where $f_m(\overrightarrow{x_m})$ is the objective function values of solution $(\overrightarrow{x_m})$

.Thus the results should be evaluated for the obtaining the comparison results with the existing works.

4. IMPLEMENTATION RESULTS

The dataset used here may obtained from UCI online dataset repository. The patient data set is maintained and compiled from heart attack patient data collected from medical practitioners from University Hospital, Hungarian Zurich and Switzerland. Only 14 attributes from the database are considered in the existing works among 76 attributes in the Cheveland Dataset for the predictions required for the heart disease. Here we consider only 10 attributes to improve the classification of missing values. The following attributes with nominal values are considered: Age, Sex, cp(chest pain), trestbps(resting blood pressure), chol(serum cholestoral), restecg(resting electrocardiographic), thalach(maximum heart rate), oldpeak, slope, num.
The attributes should be analyzed with the Weka tool 3.8.1. In this tool, the two main processes should be done with the input dataset. First, the processed data should be clustered by calling the Hierarchal Clustering and apply classification on the obtained clusters. The obtained results are produced in the tabular columns as follows for comparison.

Table 4.1 Comparison results for Naïve Bayes and SVM Classification.

No. of Datasets	Testing Data	Naïve Bayes	CB-SVM
50	5	87.89	92.15
100	10	87.23	91.82
250	25	86.25	90.79
500	40	85.59	89.25

The above result shows that the accuracy in % of two different algorithms for classification. Existing algorithm Naïve Bayes Algorithm shows that the less accuracy when compared to the new proposed algorithm CB-SVM. The iterative set of input and dataset are obtained from the user to get various set of accuracy results.

The above percentage of instances can be classified when it can be obtained from the tool using the cardiac patient details dataset. With this the confusion matrix can also be validated through the similarity matching.

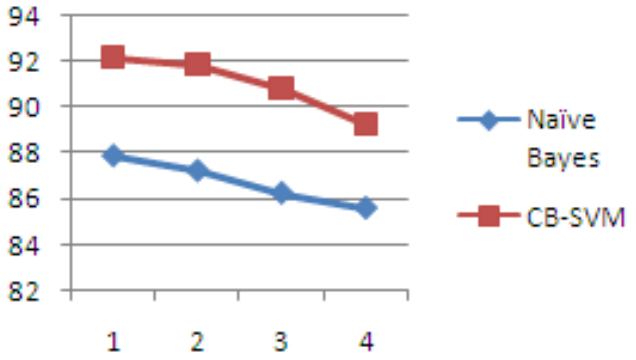

Fig.4.1 Graphical representation of results obtained

Three data mining goals we discussed in the introduction section were defined based on exploration of the heart disease dataset and objectives of this research. They were evaluated against the selected model. Results show that the selected model had achieved the stated goals, suggesting that the model could be used in heart disease diagnostic process.

5. CONCLUSION

Here in conclusion, as our discussions made in the related works, we identified some of the combinational and more complex models to increase the accuracy of predicting the early onset of heart diseases. This paper proposes a mechanism with the procedural combinations of Hierarchal Clustering, Support Vector Machine Classification and optimized cloud retrieval through the Artificial Bee Colony Optimization Algorithm to arrive at an accurate prediction of heart disease. Using the Cleveland Heart Disease database from UCI Dataset Repository, this paper provides guidelines to train and test the system with the proposed mechanism. Further, this paper proposes a comparative study of the multiple results, which include sensitivity, specificity, and accuracy.

REFERENCES

[1] Rupali and R.Patil, "Heart Disease Prediction System using Naïve Bayes and Jelinek-mercer smoothing", International Journal of Advanced Research in Computer and Communication Engineering, Vol.3, No.5, pp. 6787-6792, May 2014.

[2] Hlaudi D Masethe and Mosima a Masethe, "Prediction of Heart Disease using Classification Algorithms", Proceedings of the world congress on Engineering and Computer Science 2014, 22-24 October 2014.

[3] Santhosh Kumar and G.Sahoo, "Classification of Heart Disease using Naïve Bayes and Genetic Algorithm", Computational Intelligence in Data Mining, Vol.2, pp.269-282, December 2014.

[4] Garima Singh, Kiran Bagwe, Shivani Shanbhag, Shraddha Singh, Sulochana Devi, "Heart disease prediction using Naïve Bayes", International research Journal of Engineering and Technology, Vol.04, No.03, March-2017.

[5] W Kuanquan, W Lu, W Dianhui and X Lisheng, "SVM Classification for Discriminating Cardiovascular Disease Patients from Non-cardiovascular Disease Controls using Pulse Waveform Variability Analaysis", Australian Joint Conference on Artificial Intelligence, pp 109-119.

[6] Mythili T, Dev Mukherji, Nikita Padalia and Abhiram Naidu, "A Heart Disease Prediction Model using SVM Decision Trees-Logistics Regression", International Journal of Computer Aplications, Vol.68, No.16, April 2013.

[7] Hayashi H, Toribatake Y, Murakami H, Yoneyama T, Watanabe T and Tsuchiya H, "Gait Analysis using a Support Vector Machine for Lumbar Spinal Stenosis", US National Library of Medicine, Vol.38, No.11, Nov 2015.

[8] Talayeh R, Oleg R, Llya S and N Marko, "Multilevel Weighted Support Vector Machine for Classification on Healthcare Data with Missing Values", US National Library of Medicine, Vol.11, No.5, May 2016.

[9] Kin Li, Gaochao Xu, G Zhao, Y Dong and D Wang, "Cloud Task Scheduling Based on Load Balancing Ant Colony Optimization", Chinagrid Conference Sixth Annual, October 2011.

[10] Medhal A, A El-sisi, Arab R Keshik, F.A.Torkey, "Cloud task scheduling based on ant colony optimization", International Conf. on Computer Engineering and Systems

[11] zrachi Verma, Sonika shrivastava ad R.K.Pateriya, "Enhancing Load Balancing in Cloud Computing by Ant Colony Optimization Method", Int. Journal of Compuer Engineering in Research Trends", Vol. 4, No.6, June 2017, pp.277-284.

[12] S.Saravanan, V.Venkatachalam and S.Then Malligai, "Optimization of SLA Violation in Cloud Computing using Artificial Bee Colony", Int. Journal of Advances in Engineering, Vol.1, No.3, pp.410-414, March 2015.

[13] Warangkhana Kimpan and Bokhatai Kruekaew, "Heuristic Task Scheduling with Artificial Bee Colony algorithm fro virtual Machines", IEEE Conf. on Soft Computing an Intelligent Systems, Aug 2016

[14] Shaleen Shukula and Rutvik Mehta, "Artificial Bee Colony Algorithm on Big Data to find out required Data Sources", International Journal of Research Culture Society, Vol.1, No.3, May 2017.