Machine Learning Based Medicinal Care in Cloud

Nandita Sengupta#1, Ramya chinnasamy*2

[#1] Assistant Professor, [*2]Lecturer& Department of Information Technology
& University College of Bahrain
Saar, Bahrain

Abstract- *The cloud computing is the emerging technology with benefits like reduced investment and maintenance cost, increased scalability, availability and reliability. With the characteristics of ubiquitous access, on-demand access, pay-per-use service and resiliency, the cloud computing is applied everywhere. Due to non-availability of specialists in primary health centers in developing and under developing countries, patients are not treated properly. In this paper, treatment model for treating the patients in primary health center has been suggested. This model is built on machine learning system, where IaaS, SaaS, PaaS based cloud computing contribute a lot. Data like symptoms and prescribed medicines for different diseases are collected by different hospitals and various research centers. Same data will be stored in cloud and for similar types of symptoms; medicines can be prescribed whenever required. Thus, in our suggested treatment model, machine learning will help in treating the patients with stored knowledge of specialists through cloud computing.*

Keywords - *Cloud computing, Data, Knowledge, Machine Learning, Specialists, Symptoms;*

I. INTRODUCTION

Cloud computing is a pervasive open system connected by heterogeneous networks. It gets the name as a metaphor for the internet. Using internet technologies, scalable and elastic IT-enabled capabilities are delivered as a service to external customers in cloud computing. Standard services are delivered through the internet as pay-per-use or metered usage. It is an IT-enabled capability that provides remotely provisioned scalable and measured resources. The following are some of the drivers behind cloud computing.

A. Capacity Planning

Capacity planning is the process of analyzing and forecasting future demands of an organization's IT resources such as a hardware resource, software resource or network equipment. The on- premise information system becomes either under-provisioned or over-provisioned due to the variance between the capacity and the demand for an IT resource. All the capacity planning strategies such as lead, lag and match strategy become inefficient at some point. When the organization adopts cloud computing, the cloud provider considers the overhead of capacity planning. The organization can spend money on other core investment instead of spending on IT resources.

B. Cost Reduction

An IT-based automation solution of any organization is always limited by the infrastructure's processing power. An expansion of automation solution needs more investment on IT infrastructure. An on-premise organization faces infrastructure-related operating overhead such as technical personnel, upgrades and patches for software, utility bills for power and cooling, security measures to protect infrastructure resources, administrative and accounts staff. An IT department potentially becomes critical player who drains the organization's profit and overall evolution. If the company moves from traditional on-premise IT related resources to cloud resources, the cost overhead discussed in this section can be minimised or eliminated as cloud computing provides pay-per-use or metered usage.[1]-[4]

C. Organizational Agility

Businesses have to adopt changes whenever any new version of a resource arises in the market. The business's responsiveness to changes in the resources is termed as organizational agility. The traditional on-premise organization has to invest more to make the organization agile. If the organization adopts cloud computing, this overhead is resolved by the cloud provider.

D. Technological Innovation

Some of the technological innovations that make an actual foundation of cloud computing are clustering, virtualization and grid computing.

II. LITERATURE REVIEW

A cloud is a distinct IT environment that provides elastic and scalable IT resources. Cloud is a term that is a metaphor for internet that provides remote access to a set of decentralized IT resources. A cloud has a finite boundary in cloud computing. A cloud is privately owned and offers pay-per-use IT resources. Fig. 1 shows the symbol of a cloud. An IT resource in cloud computing is any IT-related artefact that can

either be software based or hardware based that can be accessed remotely as pay-per-use.IT resource includes a physical server, network equipment, virtual server, storage device and a software. Fig. 2 depicts different IT resources.An on-premise IT resource is any IT-resource hosted within the boundary of the organization. Cloud provider is the party that provides cloud-based IT resources. Cloud consumer is the party that uses the cloud-based IT resources provided by the cloud provider. A cloud service consumer is the temporary runtime role assigned by the software program when it accesses the cloud. The characteristics of the cloud include on-demand usage, ubiquitous access, resilient support, scalable resources, measured usage and multitenant. The cloud provides a lot of benefits such as reduced investment and proportional cost, increased scalability, increased availability and reliability. We can apply cloud computing in any information system.[5]-[8]

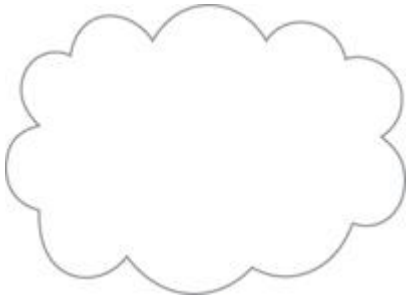

Fig 1: The Cloud

Fig 2: IT resources

III. MACHINE LEARNING

Machine learning is the process of making computers to act without being explicitly programmed. The machine learning can be broadly classified into Supervised learning and Unsupervised learning.

A. Supervised Learning:

Given the training sets with input features and the output, the machine will learn to predict the output for the incoming input features. If the output value is discrete then it is known as classification problem. On the other hand, if the output is continuous, then it is known as regression problem

B. Unsupervised Learning

Unsupervised Learning is a type of machine learning technique where the machine will find the pattern in the input and cluster it accordingly.
Machine learning can be applied to any kind of data and can get the predicted result or can find the pattern in the data. Many algorithms exist in both supervised and unsupervised learning category.

C. Logistic Regression

Logistic regression is a statistical technique for binary classification problem. Logistic regression algorithm is a supervised learning method that can be applied to the data with discrete output values. The different features of input with output will be considered as training data set. The machine will be trained with the given training data set and will find out the hypothesis to predict the output for the future input. The main aspect of logistic regression technique is logistic function that can be defined as
transformed = 1 / (1 + e^-x) (1)
where e is the numerical constant Euler's number and x is the input.
Similar to linear regression ,the output for the given input features is calculated using hypothesis
$h\theta(x) = \theta Tx$ (2)
$\theta \epsilon \{\theta 1, \theta 2, \ldots, \theta n\}$
where $h\theta(x)$ is the predicted output, θ is the parameter to be estimated and x is the input.
To calculate the prediction, we have to assign some initial values to θ and calculate the predicted output. Find out the error value between the actual output and the predicted value. If it is not matching then change the value of θ and try to minimize the error value between the predicted output and the actual output. in general the parameters of the logistic regression model for each training instance can be found using two steps.[9]-[11]
1.using current values of parameters, calculate a prediction.
2.based on the error in the prediction, calculate the new parameter.
Repeat step 1 and 2 until the model is accurate enough or some number of iterations.
The predicted output is then converted into probability using logistic function which will give the discrete output as shown below.

p(class=0) = 1 / (1 + e^(-output)) (3)
The pictorial representation of the logistic regression algorithm is given the Fig. 3.[12],[13]

IV. Social Preventive Diseases and Medicine

A preventive health care is the one that provides measures to prevent a disease instead of treating it. Most of the people from tropical countries were affected by preventable, curable diseases such as Malaria, Cholera, Acute lower respiratory infections and Tuberculosis. A statistics from WHO(World Health Organization) shows that 6.1 million people had died due to social, preventable and curable diseases.

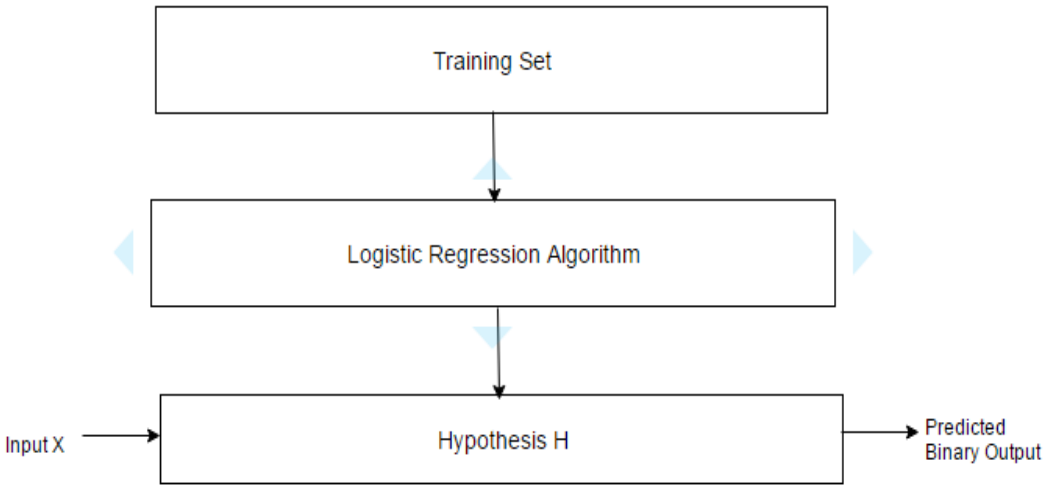

Fig 3: Supervised Learning Algorithm

A. Specialist in Social Preventive Medicine

Specialist in Social Preventive Medicine is a person who has studied and practising exclusively about Social Preventive Medicine. These specialists are very less as the seat is minimal. According to Indian Journal of Medical Research, most of the people suffered from social preventive disease were died due to lack of specialist available in rural places. Since most of the specialist people available in the city, the people from agrestic places were unable to get specialist treatment.

B. Primary Health Centre

The primary health centre is a hospital with basic amenities in rural places. Mostly fresh medical graduates or those who are working on their internship period, health workers and supervisors were available to help diseased people. But there is a great gap between knowledge level of health workers and the specialists. The specialist may visit the rural primary health centres once in a while. They have to analyse the patient record in a short span of time and have to prescribe medicine that needs frequent monitoring. Present data storage methods in primary health centres do not provide facilities for continuous monitoring. For example, those who are suffering from tuberculosis needs to undergo liver function test and X-ray every month. The resultant report of the blood test and X-ray should be reviewed by the specialist and not by the health supervisor.[14]

V. SPM Data Storage in Primary Health Centre

The current scenario of storing data of a patient in primary health centre involves either paper-based record or a standalone computer storage. Both paper-based and standalone system of storing patient data do not allow continuous remote monitoring. Fig. 4 depicts the pictorial representation of paper-based records. Fig. 5 shows a data storage on a standalone computer. Storing patient data on a centralized server is expensive as it requires more capital expenditure and operational overhead. Resource sharing is also difficult in the case of centralized server storage.[15]

To obtain resilient storage, we need to provide redundant implementation of resources which, on the other hand, makes the entire system more expensive. Since all the primary health centres have been established and maintained by the government, the authorities find it difficult to allocate more funds on storage of medical records. Sharing of files to more than one user is also difficult in centralized storage.

Fig 4: Paper-Based Medical Records

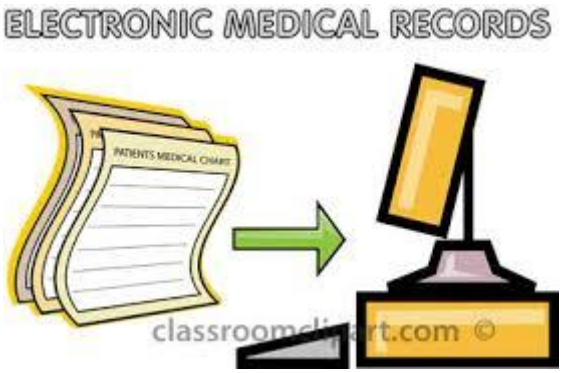

Fig 5: Electronic Medical Records

VI. Proposed Storage of SPM Medical Records in Cloud

To overcome the challenges posed by the conventional medical data storage and processing, we propose a cloud-based storage and processing of social Preventive Medicine health care records.Fig.6 depicts the pictorial representation of a cloud based storage and processing of SPM medical records. In this method, the medical record of a patient with any socially preventable disease will be stored in the cloud storage. The specialist can access the records from cloud easily on his/her computer or smart phone. For example, the specialist can view the liver function test of a patient with tuberculosis and can direct the increase or decrease in dosage of the medicine. The Specialist can also change the medicine if the medicine used by the patient affects the patient's liver. The specialist can also view the X-ray of a tuberculosis patient and can find out the improvement of the lungs after completing a course of the medicine. Specialists perform all the above processes at their leisure time. Most of the specialist do this free of cost as they want to do some medical service without compromising a generous life. If a particular specialist is unavailable, other specialists who are in the cloud can view the medical record along with a current report and can prescribe the medicine.

In this method, if the data volume is huge, then it is difficult for the specialist to view the patient's details whose report is aberrant with preference. For example, the patient under medication for the treatment of tuberculosis needs to perform the liver function test every fortnight for the first 8 weeks since starting the medication. If the result is abnormal, then there is a chance of patient to get into severe hepatitis and fulminant liver failure. Specialist care should be given to the patient with abnormal report as soon as possible.

If we apply supervised machine learning algorithm to the SPM data in the cloud, then we can classify the data into two category. The data of patient with normal range of test results and the data of patients with abnormal result. The specialist can view the details of the patient having abnormal test results and can prescribe the medicine accordingly. This process will avoid delay and can save lots of lives. The patient with normal range of test result can be treated with medical practitioner at primary health center itself. Fig.7 shows the pictorial representation of proposed logistic regression machine learning on SPM medical records in the cloud.

processor for implementing and testing the proposed

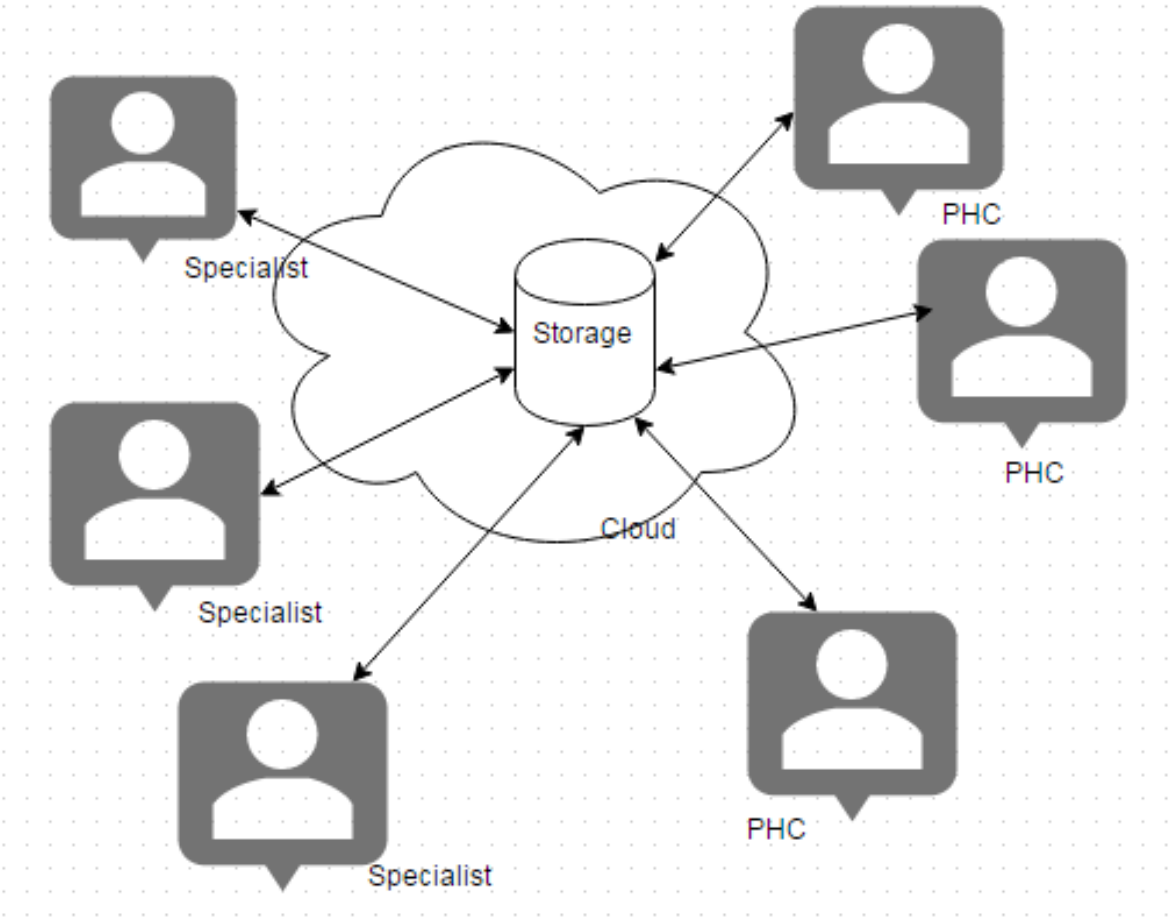

Fig 6: A cloud-based storage of SPM medical records

VII. EXPERIMENTAL RESULTS

We have used Octave4.2.1 environment for the experimentation. Octave 4.2.1 is an open source high level language for numerical computing. Octave has broad range of tools for solving numerical linear algebra problems. The scientific programming language with powerful mathematics oriented syntax and having built-in visualization tools makes octave suitable to create prototype easily. Synthetic data set of 100 records have been considered for training. We have used system with 6GB RAM and 2.50GHz

algorithm. The data is stored in inputTrainingSet1.txt file.The Alanine transaminase (ALT) Aspartate transaminase (AST) of the liver function test during the medication for the treatment of tuberculosis has been taken as input. The normal range of ALT is between 14 to 75 IU/L and of AST is between 14 to 62 IU/L.The logistic regression learning has been applied on the training set and the system learned the

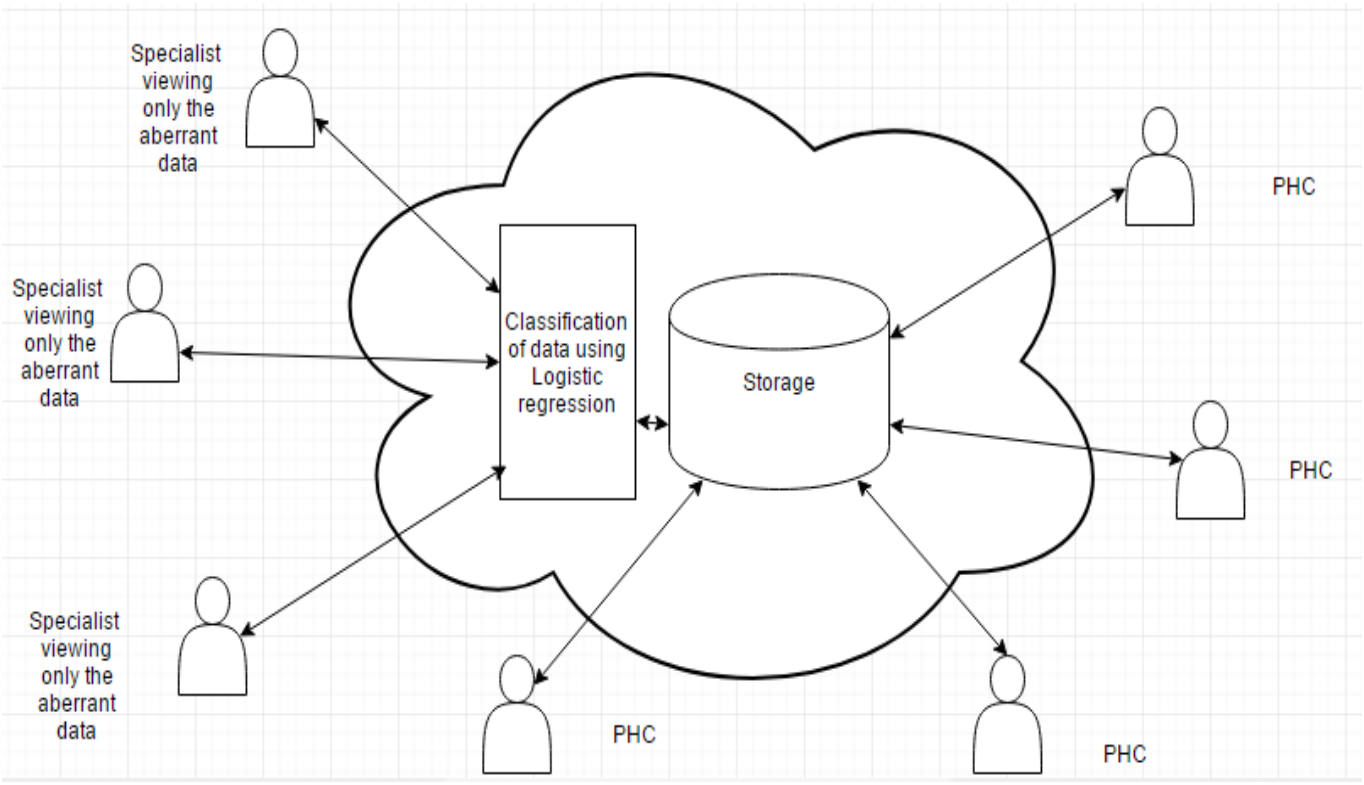

Fig 7: Machine learning on SPM Records with cloud computing

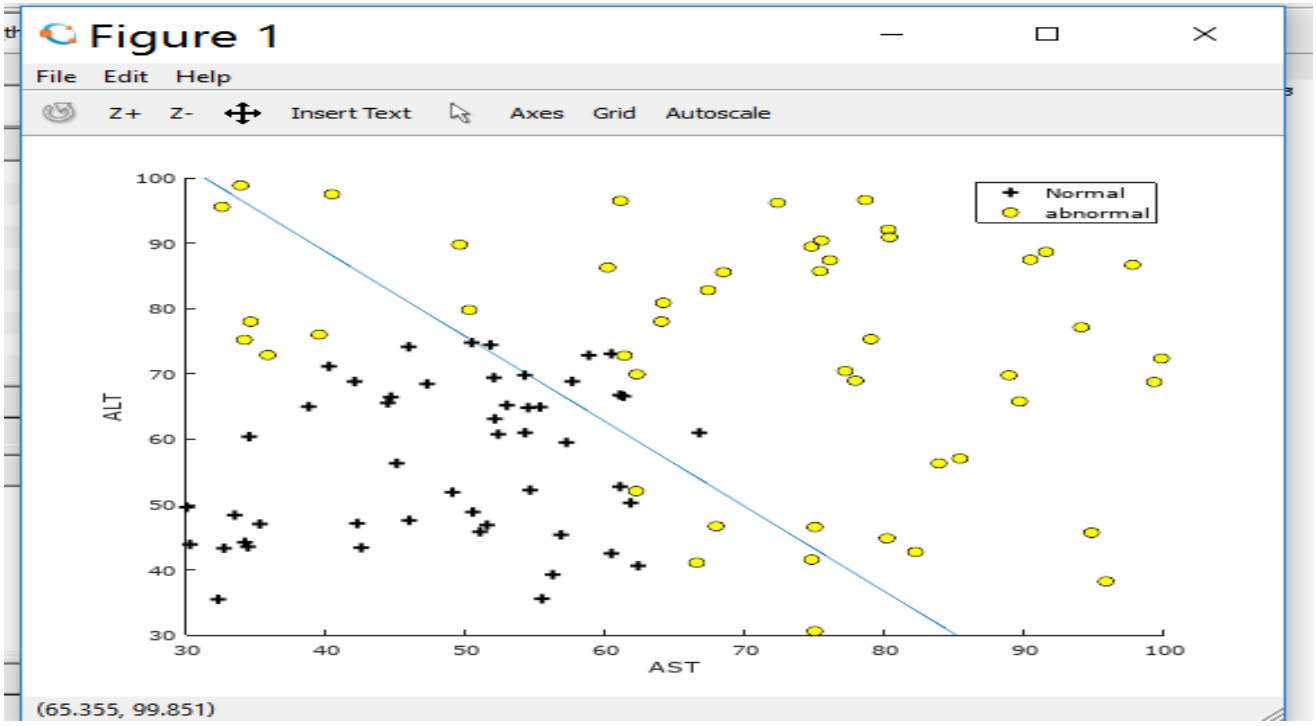

Fig 8: Scatter diagram for the training set

```
Command Window
Plotting data with + indicating (y = 1) examples and o indicating (y = 0) examples.

Program paused. Press enter to continue.
Cost at initial theta (zeros): 0.693147
Gradient at initial theta (zeros):
 -0.020000
 4.379162
 3.302573

Program paused. Press enter to continue.
Cost at theta found by fminunc: 0.343869
theta:
 13.534547
 -0.125092
 -0.096135

Program paused. Press enter to continue.
For a patient with AST 42 and ALT63, we predict normal probability of 0.902399

Train Accuracy: 83.000000
```

Fig 9:Output for the given input AST 42 and ALT 63

behavior. Fig. 8 is the graphical representation of the training set. Fig .9 shows the cost, gradient at initial theta, the predicted output cost, normal probability and the training accuracy.[16]-[23]

VIII. CONCLUSION

Applying machine learning on Social preventive medicinal health care in cloud computing can save a lot of lives as it uses the classification machine learning technique and ubiquitous characteristics of cloud. With the above proposal, we can relieve the ordeal of the people affected by socially preventable diseases with reduced cost.

IX. FUTURE WORK

We have applied and tested the logistic regression algorithm for liver function test of tuberculosis patients. we have used synthetic data to train the system. In future, we will apply the training algorithm for all the social and communicable diseases and will collect the real sample and improve the prediction.

REFERENCES

[1] Velte, "Cloud Computing-A Practical Approach," Tata McGraw-Hill Edition

[2] Barrie Sosinsky ,"Cloud Computing Bible, "Wiley Publishing Inc.

[3] Thomas Erl / Ricardo Puttini / ZaighamMahmood, "Cloud Computing:Concepts, Technology &Architecture," Prentice Hall

[4] Cloud computing: Benefits, risks and recommendations for information security,www.enisa.europa.eu/actyrm/files/deliverables/cloud-computing-risk-assessment/atdownload/fullReport/

[5] IDC, "IDC Ranking of issues of Cloud Computing model," http://blogs.idc.com/ie/?p=730/.

[6] Mohd Muntjir1 , Mohd. Junedul haque ,"Cloud Database Infrastructure: Database System Transference in Cloud Computing Management and Security ",International Journal of Computer Trends and Technology (IJCTT) – Volume 47 Number 1 May 2017

[7] Tariq Alwada , Omar Al-Zitawi , Jalal Omer Atoum," Cloud Computing: Privacy, Mobility and Resources Utilization" , International Journal of Computer Trends and Technology (IJCTT) – Volume 41 Number 1 – November 2016

[8] Ibrahim Abaker Targio Hashem , Ibrar Yaqoob , Nor Badrul Anuar ,2014," THE RISE OF BIG DATA ON CLOUD COMPUTING: REVIEW AND OPEN RESEARCH ISSUES", ELSEVIER,Information Systems 47 (2015) 98–115

[9] https://www.coursera.org/learn/machine-learning/lecture/1VkCb/supervised-learning

[10] S. Sivagowry, M. Durairaj," A Study on the accessible techniques to classify and predict the risk of Cardio Vascular Disease", International Journal of Computer Trends and Technology (IJCTT) – Volume 32 Number 1 - February 2016

[11] K. Vivekanand, A. Mohan, K.V.S. Sarma," Study of the structure and functioning of referral mechanism of patients receiving treatment and records linkage under Revised National Tuberculosis Control Programme (RNTCP) of Government of India", Indian Journal of Tuberculosis, Volume 64, Issue 2 pp. 69-150 ,April 2017.

[12] Gao Huang, Shiji Song, Jatinder N. D. Gupta , Cheng Wu,"Semi-Supervised and Unsupervised Extreme Learning Machines", IEEE Transactions on Cybernetics,Volume: 44 Issue: 12, 12 March 2014.

[13] Chandra Sekhar Reddy , Bhanu Sreekar Reddy Karumuri," Age Classification with Motif Shape Patterns on Local Binary Pattern", International Journal of Computer Trends and Technology (IJCTT) – Volume 39 Number 3 - September2016

[14] https://extranet.who.int/sree/Reports?op=Replet&name=/WHO_HQ_Reports/G2/PROD/EXT/TBCountryProfile&ISO2=IN&outtype=html

[15] https://data.gov.in/catalog/number-persons-100000-usual-household-residents-suffereing-any-tuberculosis

[16] S.P.Washimkar , S.D.Chede," Comparative Analysis of Different Classifiers to Detect the Disease in Brain MRI Images", International Journal of Computer Trends and

Technology (IJCTT) – Volume 39 Number 1 - September2016

[17] Balaji Krishnapuram, Lawrence Carin, Ma´rio A.T. Figueiredo, Alexander J. Hartemink,"Sparse Multinomial Logistic Regression:Fast Algorithms and Generalization Bounds" IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 27, Issue: 6,June 2005.

[18] Kuntal Barua , Prasun Chakrabarti,"Detection and Classification for Blood Cancer – A Survey ",International Journal of Computer Trends and Technology (IJCTT) – Volume 36 Number 2 - June 2016 ISSN: 2231-2803

[19] Ana M. Gonzalez , Francisco J. Azuaje, Jose L. Ramirez, Jose F. da Silveira , Jose R. DorronsoroMachine Learning Techniques for the Automated Classification of Adhesin-Like Proteins in the Human Protozoan Parasite Trypanosoma cruzi", IEEE/ACM Transactions on Computational Biology and Bioinformatics Volume: 6 Issue: 4,December 2008.

[20] Dong Huang , Ricardo Cabral, Fernando De la Torre," Robust Regression", IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 38, Issue: 2, Feb. 1 2016)

[21] G.K.Bhaskar,C.Imthyaz Sheriff,G.K.Sandhia" Detection and Exploitation of Heart bleed Vulnerability in Open SSL", International Journal of Computer Trends and Technology (IJCTT) – Volume 34 Number 2 - April 2016 ISSN: 2231-2803

[22] Easton Li Xu, Xiaoning Qian, Tie Liu, Shuguang Cui,"Pairwise interaction analysis of logistic regression models", IEEE Global Conference on Signal and Information Processing (GlobalSIP), Pages: 187 – 191, 2016.

[23] Yaoqing Yang, Pulkit Grover, Soummya Kar,"Fault-tolerant distributed logistic regression using unreliable components", 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Pages: 940 – 947,2016.