Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 875-882
Received 15 April 2021; Accepted 05 May 2021.

A Detailed Review of Diabetes in Health Care Using Big Data

S.Narayanan
Research Scholar
Vels Institute of Science,Technology&AdvancedStudies (VISTAS),
Chennai, India.
sivnarayana2008@gmail.com

Dr. M.Muthu Selvam
Assistant Professor
Department of Information Technology
Vels Institute of Science,Technology &nd Advanced Studies (VISTAS),
Chennai,India.
muthuselvam.mca@gmail.com

ABSTRACT
Diabetes Mellitusin recent days is commonlytreated as a major health problem, whichdistresses people of all categories in the world. If the disease is not diagnosed at early stage and is not carried out in correct time, It leads to major Health problems like Heart attack, Stroke, Retina problems, Foot ulcer, Amputation and failure of kidneys. Hence all major of human body parts will be damaged one by one. As Type II Diabetes Mellitus is a kind of abnormal disorder. By tracing the medical information concentrated in the recent days of research in the extraction of the significant data, This data is useful for Medical experts for the enhancement of treatment and diagnosis of diabetic disorders. Now a day's enormous amount of various types of data's are coming from variety of sources. This huge amount of data are stored in Bigdata. Hadoop with Map Reduce a Big data tools produces an efficient result. Hence Health care related to Bigdata playsdominant role in medical industry. It provides appropriate information to medical experts and lead them into desired framework.

***Keywords* –** Big Data, Diabetics. Map Reduce, Hadoop, Medical sensors, clustering

I INTRODUCTION

In our Human body Pancreas is one of the most important organ. Pituitary glands in our body control pancreas and produces Insulin has an effect on the metabolism of sugar, fat and protein for day to day Human energy life. The insulin acts as an agent which makes the blood glucose enters into each cells of human body and produces energy. The intolerance level of insulin in human body causes Diabetes. Diabetes mellitus is anon-communicable and Harmful disease affects our human organs one by one slowly, it also changes life style, huge medical expenses. It creates most complications such as hypertension, stroke, kidney, eye vision, foot ulcer and nervous disorder etc. Even though obesity and lack of physical exercises plays a vital role, Heredity factor is also related in this. Diabetes is now a days most common problem in both ofdeveloped and developing countries.

Diabetes is due to either the pancreas not producing enough insulin or the cells of the body not responding properly to the insulin produced by pancreas. There are three main types of diabetes mellitus.

Type1: Results from non production of insulin from pancreas. This type of diabetes is known as Type1 or" insulin - dependent diabetes. Mostly young people below 20 years can get affected. People belonging to type1 diabeteshas to take insulin entire life.

Type2: Due to lack of insulin or cells fail to respond insulin. This type was previously referred to as "non-insulin dependent" disease. The most common cause is excessive body weight and not enough exercise.

Type3: Called Gestational diabetes occurs when pregnant women without previous history of diabetes [15].
Pregnancy during old age may have a risk of developing it.

Obesity one of main reason for type 2 diabetes, by doing exercises and taking proper diet we can control it. By taking prescribed medicine form physician also we can control type 2 diabetes.

The Healthcare system now a days contains enormous data. These data may be homogenous and heterogeneous type. Traditional database system are incapable of storing these data. Bigdata which can store both homogeneous and heterogeneous data. Traditional databases contains only few datasets whereas bigdata contains large datasets. Bigdata analytics can resolve major challenges of healthcare industry. Health care data sources are of[3]classical testing equipments and techniques such as Electro cardiogram(ECG) mammography, Magnetic Resonance Imaging(MRI), electronic medical records(EMR), Ultrasound, CT scanners and many other testing equipment. Healthcare is a data-intensive field, hence the data cannot be handled by traditional system. Moreover health data has become very ubiquitous due to improvements to recording system in healthcare, the participation of patients and their treatment using social networks. Hence the field of bigdata promises a bright prospect in building healthcaresystem [3].

Healthcare system data comes from both structured data sources (EMR) and insurance provider's databases as well as unstructured forms such as doctor's notes, prescription, IoT devices, Medical sensors, electronic monitors, mobile applications, social media, and research registers. Bigdata Analytics contain several tools hence some static data such as EMR records canbe handled effectively by batch computing platforms such as Hadoop MapReduce, some real time data such as ECG reading or social network can be handled by Apache Spark[3].

Many applications are involved on diabetes in healthcare system that requires algorithms to prepare effective information. The plan and the use of algorithms have become a monotonous business especially in this time of big data. To face new challenges, it is fundamental to find appropriate algorithms for big data.

Received 15 April 2021; Accepted 05 May 2021.

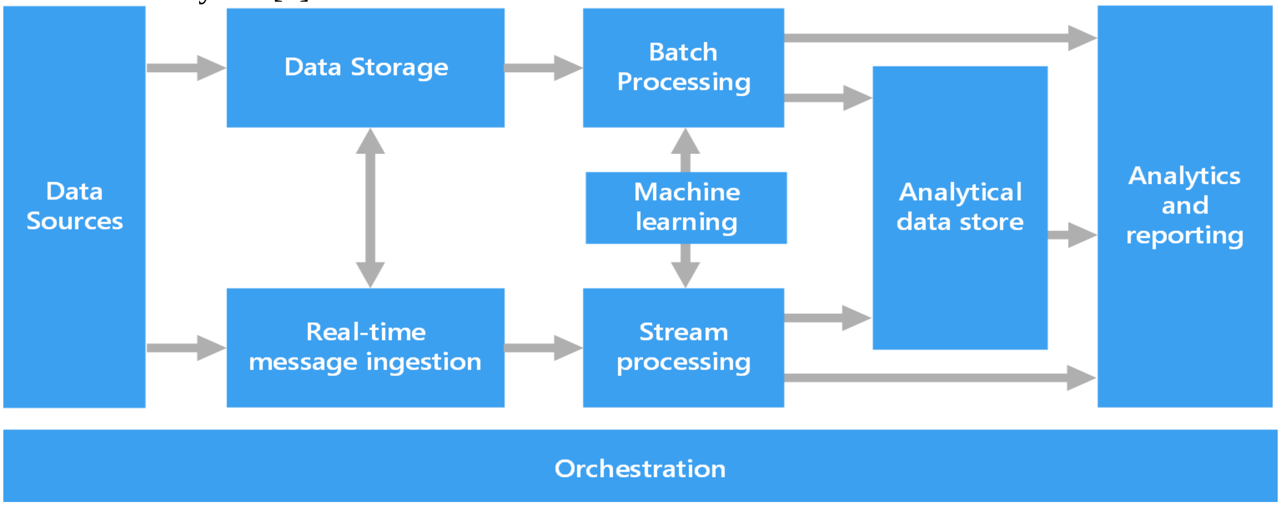
II BIGDATA

Big data is a new technology that is much used in health care and other areas of research. It support heterogeneous data also. It contain both structured and unstructured data. The storage capacity is vast and we can store in exabyte range. The large data set is characterized by '7Vs'.

- Volume - Extreme amount of data
- Velocity – Rapid data capture
- Variety – Various type and nature of data
- Veracity – Ambiguity of data.
- Variability – Interpretation of data
- Value – Importance of data
- Virality – Makes N dimensions of data.

A. *BIGDATA ARCHITECTURE*

To process big information the structure is made to manage insertion, processing and analysis of information that is too large or complex for conventional database system. The limit of which organizations access the information differ according to the abilities related to customer and their tools which they use [2].



Big data structures include some or all of the following components:

- **Data sources**. Huge information are stored more than one data sources.
- **Data storage**. Data for group processing operations is typically stored in a distributed file store that can hold high volumes of large files in various formats.
- **Batch Processing**: The Data files are preferred to be processed by big data solutions which usually utilize long running batch jobs to Filter, Aggregate and also prepare them for an Analyzation. This procedure is carried out if the data sets are very large.

- **Real Time Message Ingestion**: There is a simple Data store process where a technique is engaged to capture and store Real time messages for Stream processing. It's a necessary procedure if Real time sources are included in the Solutions.
- **Stream Processing**: The real Time messages had to be processed by the Solutions after Capturing. For an analyzation, the data will undergo multiple processing such as Filtering, aggregating and also some other preparations. Once the process is complete, the Streamed data will be written to an output sink.
- **Analytical Data Store**: The Data are usually prepared for an analysis by some Data Solutions, which after processing tends to serve those data in a structured format for various queries. The Data can be queried by Analytical Tools. Serving the data for such Analysis can be done by various Analytical Data stores such as Kimball Style relational data warehouse.
- **Analysisand Reporting**: By the Process of Analysis and Reporting, some Big Data Solutions intend to provide a clear Insight into the data.
- **Orchestration:** Some of the Big Data Solutions have an Orchestrated methodology of Data processing operations that repeat and aptly included in Workflows, which again change the source data. Provided, they will move the data between Sources and Sink.

B. BIGDATA ALGORITHMS AND ANALYTICS TOOLSALGORITHMS

The existing work on Bigdata to process huge data several numerous technologies are used. The major popular techniques to handle for diabetes patients are mentioned below.

- **Classification**: Classification the process of identification of similar values on the basis of previously grouped values

- **Prediction**: Prediction involves based on given input what all the organs are going to affect.And alert the patient

- **Nearest Neighbor**: By using already Existing values in the Records on par with the Values meant to be predicted and nearest in Order, certain and particular Values are predicted. This method of prediction is the Nearest Neighbor.

- **Clustering**: Clustering involves collection of records of diabetic patients which are similar by discovering the distance between them in multidimensional space [2].

C. BIGDATA ANALYTICSTOOLS

Several free tools are available now a day to help for processing bigdata. Some of them are

- **Hadoop**:Processing of huge datasets across group of networked computersUsing simple encoding models.

- **MapReduce:**This is a programming model tool engaged by Hadoop.

- **MangoDB**:A combination platform document focused on database software system.

- **Cassandra**:It offers no space for failure and is one of the most dependable bigdata tools

III. REVIEW OF LITERATURE

Several researches were completed to illustrate the concept of bigdata analytics on diabetes, indicates this is frequently a concept to analyze, methodically extractinformation from large datasets that are too much big or complex information of health care system.

Several studies on the Health care system using bigdata can be determined and discussed by simply Kalyankar, Dharwadkar, This author machine learning algorithms [14]In Hadoop mapreduce environment on pima Indian's dataset , to findout missing values and find out patterns in it. They found out C4.5 algorithm best and gives result in less time and suggests that implemented algorithms are able to impute missing values and to recognize pattern from the dataset [14].

K.S Praveen kumar, ,DrR.Gunasundari did a prediction work by implementing different parameter like classification accuracy and classification error, on bigdata tool Hadoop, Map reduce framework by using C4.5 algorithm and they found this produce good result.

Kavinprrasadarjunan, ManivelSivasakhti predict that by combining two different supervised machine learning algorithma SVM and K-meand produces better result of 94.9%. This application propose an effective technique for earlier detection of diabetic disease.

Mohamed Azeemsarwar, Nazeerkamal.They did a comparison by implementing six different machine learning algorithms on UCI machine learning repository [17], PIMA Indians dataset. These algorithm are KNN, NB, SVM, DT, LR, RF and concluded that SVM and KNN are appropriated for predicting diabetics disease.

Ayman Mir, N Dhage by using WEKA tool they conduct research on PIMA indians dataset to predict diabetics disease by employing NB, SVM, RF AND Simple CART algorithm and observed that Support Vector Machine (SVM) performed best in prediction of the disease having maximum accuracy [18].

Qian Wang, N.Davis[5]propose an effective prediction algorithm for diabetics mellitus on Imbalanced data with missing values(DMP-MI) and they adopted a method ADASYN adaptive synthetic sampling method, they conduct research using this method on PIMA indians data set and found random forest(RF)classifier is used to generate prediction, they proposed DMP-MI algorithm has outperformed other algorithms on accuracy and other classifier performance indicators and has shown great potential for diabetics prediction.

MinyechilAlehegn research on Machine learning algo such as SVM, Naine Net, Decision stump, and proposed ensemble method (PEM) and they proposed PEM provides high accuracy. The research was done on UCI repository.

Shakuntalajatav and Vivek Sharma has analyzed prediction system for diabetics, kidney, Liver disease [24] based on support vector machine (SVM) and random forest (RF). [9]The performance of these technique is compared based on precision, recall, accuracy, f_measure as well as time. They found the result shows the accuracy of 99.35%, 99.37% and 99.14% on diabetes, kidney and liver disease respectively.

P.Sureshkumar and S.Pranaviconduct research to analyze and compare different machine learning algorithms to identify best predicting algorithm based on various metrics such as accuracy, kappa, precision, recall, sensitivity and specificity using Random Forest, SVM, K-NN, CART and LDA and found that RF is giving more accurate predictions compared to other algorithms[6].

B.Suvarnamukhi, M.Seshashayee perform research on bigdata using machine learning technique for [21] prediction, the diabetics' prediction was carried out by Extreme Learning Machine classifier (ELM). The ELM was carried out by varying classifiers, prediction accuracy, precision, recall and time consumption and they found that ELM provides better efficiency.

K.Saravananathan conduct research on diabetic dataset using [20] popular clustering algorithms k-means and fuzzy C-Means algorithms. The performance of this algorithm is tested based on its execution time. The execution time of the algorithms to form clusters is compared for for different executions and shows that k-Means is better than Fuzzy C-Means [10].

Raghavendra S, Santosh Kumar work exploits machine learning technique such as LR, ANN, SVM,RF, [7]NN with 10-fold cross validation (CV) for classification and prediction of diabetes with future selection methods (FSM) using R platform on PIMA Indian data set from UCI repository. From experimental result it is identify that classification accuracy was occured84.25% whereas, with reduced set attributes an accuracy of 85.24% is achieved using NN with 10-fold CV techniques compared to others.

Basharat Naqvi, Arshad Ali conduct research for prediction of diabetes in healthcare industry. By using Rapidminer platform Algorithm [17] such as random forest, decision stump, random tree and ID3 were applied. This research work presents description of chosen classification models and data set. The comparison of five classification techniques on chosen data set was performed. And finally the work shows decision tree is the best technique for prediction of disease in diabetic patient.

All the above researches have been successful in analyzing the diabetic dataset and they developed good result. But all the methods or tools are considered only on few parameters and the dataset contain structured, unstructured or semistructuredinformation. The projected system is planned to work with more parameters also produce good result

IV. CONCLUSION

The Health care system with big data environment provides a massive support to diabetes disease in recent digital environment. Now a days the data coming from different areas of different types. Traditional system does not support these data. .Prediction and classification algorithms in Machine learning on big data produces efficient result. Bigdata and bigdata analytics support latest techniques and works efficiently for several algorithms and yield good results. The goal of the system is to predict and provide good treatment for diabetics in efficient manner. Prevention of diabetics taken place when it is diagnosed in advance.

References:

1. Gaspard Harerimana, Beakcheol Jang, Jong Wook Kim, Hung Kook Park. "Health Big Data Analytics: A Technology Survey", IEEE Access, 2018
2. P.Suresh Kumar, S.Pranavi. "Performance analysis of machine learning algorithms on diabetes dataset using big data analytics", 2017 International Conference on Infocom Technologies and Unmanned Systems(Trends and Future Directions)(ICTUS),2017
3. Qian Wang, Weijia Cao, JiaweiGuo, JiadongRen, Yongqiang Cheng, Darryl N. Davis. "DMP_MI; An Effective Diabetes Mellitus Classification Algorithm on imbalanced Data With Missing Values", IEEE Access, 2019
4. J.M.M. Rumbold, M.O'Kane, N.Philip, B.K.Pierscionek. "Big Data and Diabetes: the applications of Big Data for diabetes care now and in the future",Diabetic Medicine, 2019
5. GauriD.Kalyankar, ShivanandaR.Poojara, Nagaraj V. Dharwadkar. "Predictive analysis of diabetic patient data using machine learning and Hadoop", 2017 International Conference on ISMAC(IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2017
6. "Intelligence in Big Data Technologies-Beyond the Hype", Springer Science and Business Media LLC, 2021
7. "International Conference on Innovative Computing and Communications", Springer Science and Business Media LLC, 2021
8. M.M. Rumbold, M.O' Kane, N.Philip and B.K.Pierscionek , "Big Data and diabetes: the applications of Big Data for diabetes care now and in the future" DIABETIC Medicine, DOI10.1111/dme.14044
9. GauriD.Kalyankar, ShivanandaR.Poonjara and NagarajV.Dharwadkar, "Predictive Analysis of Diabetic Patient Data Using Machine Learning and Hadoop" International conference on I-SMAC 2017.
10. Muhammad AzeemSarwar, Nasir Kamal, Wajeeha Hamid, and Munam Ali nShah, "Prediction of Diabetes Using Machine Learning Algorithms in Healthcare" 24th International Conference on Automation & Computing Newcastle University, Newcastle upon Tyne, UK, 6-7 september 2018
11. Ayman Mir, sudhir N. Dhage, "Diabetes Disease Prediction using Machine Learning on Big Data of Healthcare", 978-1-5386-5257-2/18/$31.00@2018 IEEE
12. QIANWANG, JIAWEI GUO AND DARRYL N.DAVIS , "DMP-MI: An Effective diabetes Mellitus Classification Algorithm on Imbalanced Data With Missing Values" publication july 19,2019, date of current version August 19,2019 IEEE Access
13. MinyechilAlehen, Rahul Joshi&DrPreetiMulay, "Analysis and Prediction of Diabetes Mellitus using Machine Learning Algorithm" International Journal of pure and Applied Mathematics, ISSN:1314-3395
14. Raghavendra s, Santosh Kumar J,Raghavendra B.K, "Performance Evaluation of Machine Learning Techniques in Diabetes Prediction" International Journal of Engineering and Advanced Technology(IJCET)ISSN: 2249-958, Feburary 2019
15. Basharat Naqvi, Arshad ALI, Muhammad Atif, "Prediction Techniques for Diagnosis of Diabetic Disease:AComparitive Study", IJCSNS Internal Journalof Computer Science and Network Security, August 2018
16. GASPPARD HARERIMANA, BEAKCHEOL JANG, HUNG KOOK PARK "Health Big Data Analytics" A Technology Survey" 10.1109/ACCESS 2018.2878254

17. K.S Praveen kumar, DR R Gunasundari "Diabetes Mellitus prediction in Big Data using Hadoop/Map Reduce Frame Work", IJARCEE Dec 2018.
18. DR. R.Vijayakumar, KavinprrasadArjunan, ManivelSivasakthi, Karthikeyan Lakshmanan, "Diabetes Prediction By MachineLearning over BigData from Healthcare Communities",IRJET Apr 2019.
19. Shakuntalajatav and Vivek Sharma ,"An Algorithm for Predictive Data Kining Approach in Medical Diagnosis", IJCST Feb 2018
20. P.Suresh Kumar, S.Pranavi, "Performance Analysis of Machine Learning Algorithms on Diabetes Dataset using Diabetes Dataset using Big Data Analytics".
21. B.Suvarnamuki, M.Seshashayee, "Big Data Processing System for Diabetes Prediction using Machine Learning Technique", IJITEE,ISSN : 2278-3075, October 2019.
22. K.Saravananathan, DrT,Velmurugan, "Cluster based performance analysis for Diabetic data", International Journal of pure and applied mathematics, 2018.