International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org
Vol. 09 Issue 09, September-2019,
ISSN (O): 2249-3905, ISSN(P): 2349-6525 | Impact Factor: 7.196 |

Deep learning and big data analysis for predictive modeling of diabetes patient data

V S Aravind Nambala[1], Dr. Rashi Agarwal[2]
Department of Computer Science & Engineering
[1,2]Shri Venkateshwara University, Uttar Pradesh, (India)

ABSTRACT

There is a lot of data coming out of the health care industry these days. This data must be collected, stored, and processed in order to get information from it and put that knowledge to good use.Diabetes Mellitus is a non-communicable disease (NCD) that affects a large number of people. DM has become a major health problem in developing countries like India. The DM is one of the most serious syndromes, with long-term repercussions and a wide range of health issues following it. We need a method to save and analyse diabetic data and detect possible hazards with the help of technology. "Data mining techniques, machine learning algorithms, and statistics" are all used in predictive analysis in order to acquire insights and anticipate future dangers from current and historical data sets. Hadoop/Map Reduce predictive analysis is used in this article to forecast diabetes kinds, complications, and treatment options in a Hadoop/Map Reduce context. Using this approach to treat and care for patients gives improved outcomes, such as cost-effectiveness and accessibility, according to the report.

Keywords:*Healthcare, Hadoop, Machine Learning, Predictive Modeling*

1. Introduction

PCs have made significant advancements in technology, which have resulted in vast amounts of data being generated. The healthcare industry is home to a colossal amount of private information. To obtain the most value from this information, it must be handled with extreme care. Despite the fact that diabetes mellitus has been identified as one of the world's leading health threats, new predictive models are needed to accurately identify the disease. The challenge with this initiative is that it is difficult to diagnose whether or not someone has diabetes, whether it is positive or negative. There are numerous explanations for this**(kumar et al., 2015).**Different persons may exhibit a variety of indications. Consequently, it is difficult to tell if they possess it or not. True results will only be known after conducting a thorough examination. That's why an analysis of large diabetic data sets is needed to uncover some previously unknown facts that could lead to the development of a predictive model. Machine learning algorithms are used to build the prediction models. It's a type of artificial intelligence in which computers may pick up new skills on their own, without any outside guidance. Machine learning is all about creating computers that can learn and adapt to new or previously unforeseen information **(Iyer et al., 2015).**Unsupervised and supervised machine learning algorithms are the two most common types of algorithms.

1.1. Types of Diabetes

Insulin-dependent diabetes mellitus (IDDM) or juvenile-onset diabetes is the medical term for **Type 1 Diabetes.**It is most common in those under the age of 30 and can affect both children and adults; however the vast majority of these instances included youngsters **(Kalyankar et al., 2017)**

Diabetes mellitus in adults, often known as NIDDM or adult-onset diabetes, **Type 2 diabetes** is more likely in those who are older, obese, have a family history of the disease, have had gestational diabetes, have impaired glucose tolerance, are physically inactive, or are members of a minority ethnic group**(Rajesh & Sangeetha, 2012).**

The third most common kind of diabetes is **Gestational Diabetes,**pregnant women who have never been diagnosed with diabetes

Cystic fibrosis-related diabetes and steroid diabetes are both caused by genetic abnormalities in insulin secretion in humans, as are **congenital diabetes** and diabetes caused by high doses of glucocorticoids.

2. Methodology

This approach uses a sequential pattern to identify Sentence Element Inference. An indicative extraction pattern is one that uses a sequential pattern to recognise what the user wants to say and extract structured analysis from distant storage media with perfect information (IEP). Auto-Disease Inferences can be developed using Prediction Algorithms**(Vaishnav & Patel, 2015).**Initial tokens collected from the given unstructured utterances are used to begin the prediction process. A question collection is queried for any questions that contain a certain token or token pair, and those questions are then treated as structured questions for the sake of scoring. The specified unstructured statements are used to construct all feasible sequential patterns, which are then evaluated based on Pattern Evaluation for their reliability score **(Dai & Ji, 2014).**

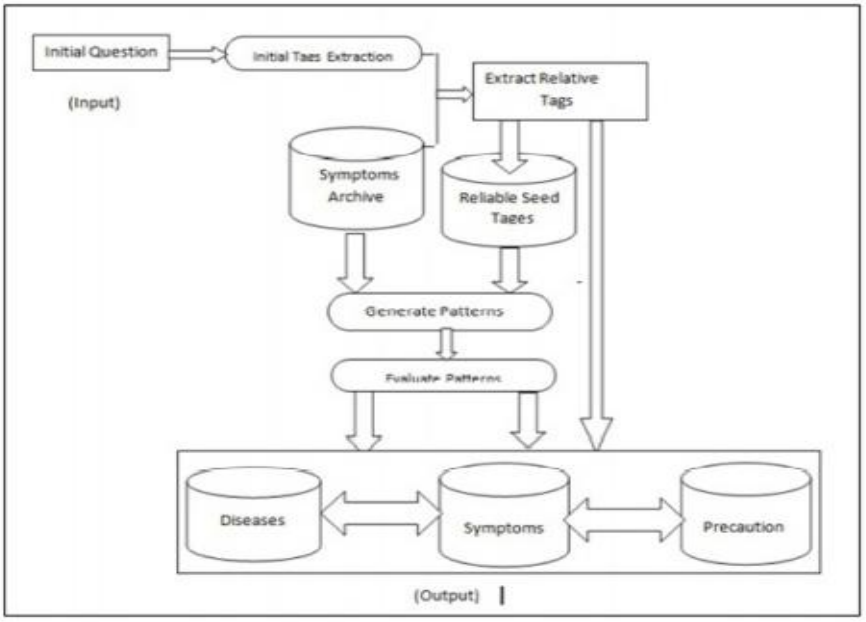

Figure1 Architecture of Proposed System

2.1 Algorithms

2.1.1. Two minimum standards and Two dividers

- "Start
- Compute Data length symbolized by DLen.
- NCore is an indicator for how many cores a system has.
- Compute block size symbolized as BSize.
- *BSize=DLen%NCore==0?DLen/Ncore:DLen/NCore ;*
- Make sure that each chunk's start(Sx) and end indices are set (Ex)
- Stop"

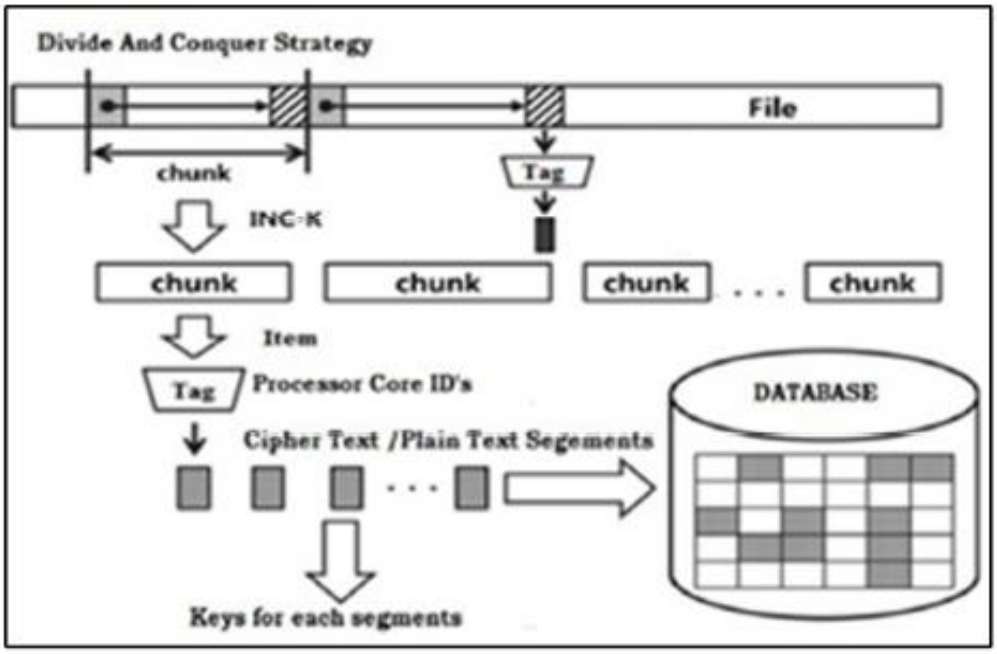

Figure 2

2.1.2(Multilayer Perceptron) MLP Algorithm

- "Start
- Randomly select the weights, and then select a learning rate
- until system is designed
- Forward-pass across net (with fixed weights) to generate output (s)
- Feedbacksrequested
- compounded the weights
- Addition
- Squashed nonlinear startingfunction
- Each neuron in the following layer receives the output.
- Until network outputs have been generated, repeat the following steps."

3. Procedure of proposed system

- To begin, a user must enter his or her basic information into the system, after which an auto-generated password will be sent to the user's email address.
- By inputting their username and password, the user will be able to access the system.
- After a successful login, the user is prompted to provide symptoms.
- The user will be presented with a list of symptoms that are similar to the user's own.
- If a person has any of the symptoms listed, they can choose from the options provided.

- Click "I have none of the aforementioned symptoms" if none of the indicated symptoms apply by user. "
- Anticipated disease will be shown on screen and will be predicted.
- Doctor will recommend a course of action based on the patient's condition.

3.1Modules:

3.1.1. Registration

To begin using the system, a new user must first create an account by providing the required personal information. Auto-generated passwords will be sent to the user's email address after they register.

Figure 3 Login Page

3.1.2. Login

Upon entering the required credentials (Username and Password), the system will take the user to the system's home page if everything is in order, else an error notice will appear.

3.1.3. Prediction Structure

The Disease Prediction structure:

- Diabetic complications can be anticipated with this method.
- It will be anticipate the type of diabetes i.e. TYPE1,and TYPE 2.
- The highest proportion of correct predictions will be provided by the system.

3.1.4. Hadoop

For storing and running programmes on commodity computer clusters, it is a free and open-source software framework. Massive storage space, computing power, and the ability to handle almost unlimited processes or jobs all in one machine make it ideal for large-scale data centres.

4.Benefits

1. Machine learning is autonomous and makes its own decisions.

2. With HDFS, large amounts of data may be accessed quickly and easily.

3. Processing time might be slashed by.

4. Accuracy is matched by excellent performance.

5. RESULTS

After a successful login, the administrator can inspect the user database and access the user database for any necessary modifications from the user login page. After logging in, users can go for a health checkup or review their profile and make changes if necessary. Once they've finished, they can log out. There is a list of symptoms that can be used to diagnose diabetes. Fasting and postprandial blood sugar levels are displayed on the diabetes monitor, allowing users to maintain tabs on their health. Users can find the list of the most common signs and symptoms of diabetes. Proper data entry and a click on "predict" are required for prediction. The test data provided by the user is passed on to the classification rule by the user. The test results are categorised as either diabetes or non-diabetic based on the data presented.

6.CONCLUSION

Implementing Big Data Analytics in Hadoop gives a systematic means of improving results, such as universal access to healthcare at reasonable costs. One of the most serious health threats in India is the prevalence of non-communicable diseases, such as diabetes. This analysis will help the diabetic patient better comprehend the possible issues that may arise from their condition. Big data analytics will be used in this project to examine the treatment of diabetes in the healthcare business. Designing a diabetic treatment prediction analysis system could provide better data and analytics that lead to better healthcare outcomes. Anyone living in a remote area can now receive high-quality healthcare at a low cost thanks to location-aware healthcare services. Treatment can be done if it is discovered before it occurs. To anticipate the prevalence of diabetes and the risk levels associated with it in the future, pattern matching will be used.

Reference

Dai, W., & Ji, W. (2014). A mapreduce implementation of C4. 5 decision tree algorithm. *International Journal of Database Theory and Application*, *7*(1), 49–60.

Iyer, A., S, J., & Sumbaly, R. (2015). Diagnosis of Diabetes Using Classification Mining Techniques. *International Journal of Data Mining & Knowledge Management Process*, *5*(1), 01–14. https://doi.org/10.5121/ijdkp.2015.5101

Kalyankar, G. D., Poojara, S. R., & Dharwadkar, N. V. (2017). Predictive analysis of diabetic patient data using machine learning and Hadoop. *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 619–624. https://doi.org/10.1109/I-SMAC.2017.8058253

kumar, N. M. S., Eswari, T., Sampath, P., & Lavanya, S. (2015). Predictive Methodology for Diabetic Data Analysis in Big Data. *Procedia Computer Science*, *50*, 203–208. https://doi.org/https://doi.org/10.1016/j.procs.2015.04.069

Rajesh, K., & Sangeetha, V. (2012). Application of data mining methods and techniques for diabetes diagnosis. *International Journal of Engineering and Innovative Technology (IJEIT)*, *2*(3), 224–229.

Vaishnav, R. L., & Patel, K. M. (2015). Analysis of various techniques to handling missing value in dataset. *International Journal of Innovative and Emerging Research in Engineering*, *2*(2), 191–195.