CLINICAL DECISION SUPPORT SYSTEM ON COPD PREDICTION USING BIG DATA ANALYTICS WITH IMPROVED PATIENT MATCHING

[1]PASUPULATI. SANDHYA, [2]DR.DHANRAJ VERMA
[1]Research Scholar, Dept. of C.S.E, Dr.A.P.J. Abdul Kalam University, Indore- Dewas Bypass Road, Indore, M.P, India
[2]Research Guide, Dept. of C.S.E, Dr.A.P.J. Abdul Kalam University, Indore- Dewas Bypass Road, Indore, M.P, India

ABSTRACT: Big data analytics is a fast developing area that plays a vital part in research and health care practice advancements. Clinical Decision Support systems need patient identification and matching of their information residing in different systems for making better diagnosis and treatments at the right time. The COPD (Chronic Obstructive Pulmonary Disease) was a major cause of mortality and morbidity global outcome in social and financial burdens which was increasing significantly. In this paper Clinical Decision Support (CDS) system on COPD prediction with improved Patient matching utilizing big data analytics is presented. As the healthcare organizations share different documents of patients from different systems such as pharmacy, laboratory, claim systems, etc. they are required to be link with correct patient records for guarantying the better decisions by clinicians and minimized duplicate information and overheads. The Fuzzy Matching algorithm & Map Reduce are introduced in this work for addressing the issue of patient's record matching from various systems to support clinical decision greatly. Then utilizing the data mining application of big data, Decision Tree (DT) model is applied to obtain best approach in the detection of COPD in independent patients. The result analysis show that this system has scalability and flexibility utilizing any fuzzy algorithm and handling the data source exhibits greater accuracy in COPD patient diagnosis with better efficiency.

KEYWORDS: Clinical Decision Support system, Big data analytics, Fuzzy Matching algorithms, COPD.

I. INTRODUCTION

Open questions are there at health debates centre that how data is manipulated and how its value is produced out of it, secure it and share it [1]. However the Big data is a term that has become a buzzword in Information Technology (IT) filed, it is applicable to health and biological based data which is quite complicated naturally and hard to collect, still it is confined. The technology of big data with health services and science, embraced the data science research mixed by the concepts correspond to design, integration, modeling management and development of health information systems [2]. In big data domains, information's were replicate, transferred to numerous numbers of nodes. Further the sensitive data is also saved in error logs, disk caches, configuration files, system logs, etc. Hence the requirement of protection is a significant task of data analysts [3].

Generally modeling the biological mechanism is quite complex and always understanding as an intensive procedure computationally. For drawing the meaning from exponentially enhancing the amount of healthcare information, that should be deal with the perspective of big data utilizing methods are efficient for function the huge quantity of information securely as well as efficiently. Collection and aggregation of anonymous data from world widely dispersed location this makes it feasible for statistically constructing meaningful database depending on where macroscopic reasoning will made rather than focusing solely over associated and individual pathology [4]. The data mining is a large dataset examining process for extracting unknown and hidden patterns, knowledge and relationships are very hard for identifying with classical statistical approaches [5]. In healthcare fields one of the emerging fields is data mining to provide great significance to the opinion of doctor corresponds to how a patient will recover from diseases and deeply understanding the medical information.

Healthcare administration was considering a serious look at big data, identifying complex purpose aspects and problems related to managing huge amounts of information. Big data is not only a big difficultly for medical experts and also high chance in public. Using extremely accessible medical information allows doctors to simulate capability results and stop patients from receiving ineffective medications or provide them with best treatment. In different terms, collecting and utilizing information to further knowledge of path physiological methods can lead important medical developments. Anyhow, the strategic benefits formed by big-data in medical were still slow to happen, as only a few large-scale administrations had introduced some pilot or proof-of-concept operations.

These days, there were clear spaces in the field of big data analytics for health bio-data. Data-driven services were still required for data versatility, volume, speed and accuracy across the entire data value chain of medical statistics. There is a real chance to generate value out of big data in medicine with the aim of revolutionizing integrated and personalized medical services. A Decision Support system for medical care was discovered method used by the caretaker for patient-centered result assessment. Typically methods include patient, public and caretaker input into the assessment method. Anyhow, present process utilized in medical care requires direct patient involvement as an integral and transformative area of methodological techniques.
A healthcare professional determined its difficulty towards identifying the patients at higher damage of having COPD but patients clinical history understanding might made it simple for them. Data mining is being utilized by researchers in various diseases medical diagnosis like diabetes, cancer, heart store and disorders and approaches like
DT, Naïve Bayes, bagging algorithm, neural network, kernel density, automated defined groups and SVM (Support Vector Machine) exhibits various accuracy levels [6]. This work studied the patient's record matching issue from different systems and presented a solution through Big Data Analytic methods such as Map Reduce and Fuzzy Matching algorithm to make a better clinical decision.
Then a methodology is introduced for diagnosing the COPD with enhanced Accuracy using data mining techniques of Decision tree. Centralized clinical data repository consists of patient particulars with respect to their individual Aadhar number that supports for knowing to obtain treatment of every patient in various hospitals and treated doctor.

II. LITERATURE SURVEY

Big data in medical were electronic health data sets that were very large and difficult and tough to handle with traditional software, hardware and common information management tools [7]. Big data in a medical care system is huge not only because of its size but also because of the kinds of data types and the rate at which it needs to be processed. All data related to patient medical is formed with "big data". It consist of clinical information from CPOE and clinical decision support systems (doctors written documents and prescriptions, medical imaging, pharmacy, laboratory and remaining information), patient data in Electronic Patient Records

(EPRs). So far the use of computer-aided diagnosis and decision-supportive systems in the field of intensive care medicine has been restricted. Still, the amount of information generated for each individual patient is enormous, extremely the efficiency of any physician and hence underscoring the requirement for experienced examination. The difficulty of systems and the appearance of several internal connections require the use of big data analytics methods. Usual or unusual variations of needed indications and earlier deterioration have not yet been identified with the means accessible in clinical practice at present.

Currently several approaches are employed to monitor the health systems. The healthcare data digitization brings new possibilities to payers and providers for improving health care results, quality of care and to reduce cost. The advanced technologies and tools are being utilized over digital information of healthcare systems will produce valuable insights. In addition organizations should analyze external and internal information of patient for measuring outcomes and risk most accurately. Whereas, several payers and providers has been working for increasing transparency of data for generating new insight knowledge. Utilizing the health information exchange the integrated delivery networks are formed by several payers and providers. Certain larger pharmaceutical organizations are trying to de-identify the data from their clinical trials and to protect the privacy of patients by making the information only available to qualified researchers which are outside the organizations.

Cheng Hongbing, Rong Chunming, Hwang Kai, Wang Weihong, Li Yanyan et. al. [8] presented secured Big data storage and sharing technique to the tenants of cloud. An alternative technique was presented that separated big data into sequencing portions and stores them between multiple cloud providers. This approach secures data mapping instead of information itself. Because of high protection costs through the big data encryption, this will protects mapping data without this the information is meaningless. The obtained results proved the same. Joseph M. Woodside et. al. [9] stated that inefficient vendors will detect and identify poor in taking way of life choices of members and agreement with preventing care programs. Intensives will give to individuals like gift cards, cash that are taken as most recommended change in health care fields. Kiyana Zolfaghar, Naren Meadem, Ankur teredesai, Senjuti Basu Roy, Si-Chi Chin et al. [10] suggested prediction technique to congestive heart failure incidents risk readmissions, and utilized Mahout Framework. The training data is applied to Hadoop file systems; raw information can be preprocessed and transformed to classifiable information that is encoded format known as a vector which can applied as input to the framework of Mahout through RF (Random Forest) algorithm.

Developed analytical approaches are applied to massive amount of available (presently not analyzed) patient health and medical information for deeply understanding the results that are applied at care point. Ideally this information can inform to every physician and his/her patients in decision making process and utilized for identifying the suitable option to that specific patient. It was explained that data mining is larger than conventional relational data for real-time structured and unstructured data. A Big data approach is still smaller for network system monitoring. In addition to the robustness and scale features of the aforementioned techniques, the present analysis also involves data collection and integration works. It adapts data mining applications to big data analytics infrastructure, at the same time producing scalability through a cloud-based distributed evaluating plan.

III. CDS SYSTEM ON COPD PREDICTION USING BIG DATA ANALYTICS

The documents of patients were spread around geographies and various treatment facilities which can contain different technologies. Co-mingle data from two or more persons increasing protection issues lead to false positive health document matching. The false negative health document matching fails to attach several document to equal individuals resulted in fragmented, not completed HER that will compromised outcomes. The big data is leveraged for improving detection of specific patient records in medical files and/or EMPI files stored in labs, pharmacies, hospitals and health plans, as outcome exchange of health information will facilitated and greatly enhanced. The matching solution based on big data can work on underlying Map Reduce and Fuzzy Matching algorithms which are accessible in open source.

The fuzzy matching algorithm performs the matching at various ranges and critical relation among various datasets is discovered. Here utilization of Big Data analytics with Fuzzy logic and MapReduce is presented for matching the numerous patient records and presented approach overcome many of challenges of earlier systems. This solution can be discussed with an instance of suggestive data that can be scalable to any volume of data flow to improve patient matches. For example 25 patient records are consolidated from 4 various systems later cleansing, integration and transformation e.g. date of birth (DoB) is transformed to similar format. After the consolidation if a patient has many similar records then challenge is matching his/her records for removing de-duplication. In the nest stage subject oriented records are collected from resulted match pattern for designing the COPD dataset. In latter stages data is preprocessed since data has various challenges such duplicate data, no quality data and data is not clean. The entire process of the CDS system on COPD prediction is shown in figure (1).

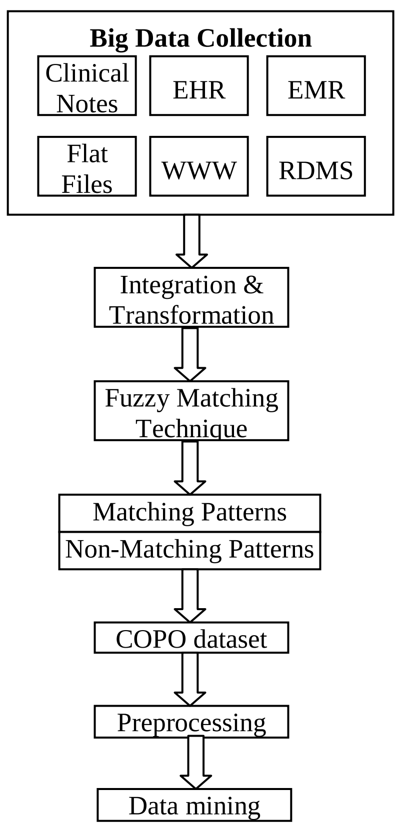

Fig. 1: ENTIRE PROCESS OF CDS SYSTEM ON COPD PREDICTION

3.1 Big data Collection

Many medical systems continue their self documents by their self platforms. These kinds of systems distribute data very rarely. Patient data is hold in a disintegrated way. In healthcare systems it is impossible to hold entire data over single legacy platform because of captured data nature. In this domain data is highly fragmented. Few of the major data pools are EHR (Electronic Health Records), behavior of the patient, pharmaceutical data, flat files, provider clinical Notes, sentiment data and WWW (World Wide Web) and EMR (Electronic Medical Records). The data are resides in various kind of systems like Excel, Pdfs, Flat files, RDBMS, etc. Absence of communication will show a bad effect over patient's health. After information is gathered it needs to be processed or converted into a form appropriate for later processing. A service-oriented design method with web services is an example of transforming information. Data is in raw state and services are utilized to derive, obtain and transform the information. In data warehousing, data is obtained from different sources and kept prepared for method. Data from different sources was cleaned and ready for different steps like extracting, transforming and loading.

3.2 Fuzzy Matching Technique

The issue of integrating patient information from different systems could be rectified by utilizing Big Data Analytic methods such as Fuzzy Matching & MapReduce. This fuzzy matching model performs matching at several phases and finds difficult relationships among various datasets. This will be utilized to detect de-duplicates from data based on detected keys. MapReduce is also utilized to compare patient information. Multiple features of a patient can be allocated various weights. For any set of structures, the distance among the related features is evaluated. Attribute-wise distances were combined with total features of a patient document to observe the distance among patients documents. By utilizing MapReduce, data about patient detection will be exploded and many documents will be formed with keys ranging from last name, date of birth, pin code, etc. This analysis examines the issue of comparing patient documents from different systems and suggests an outcome by using big data analytic methods such as Fuzzy Matching algorithms (Levenshtein) & MapReduce for best clinical decision support. The main advantages of the suggested system were scalability, low cost, use of any fuzzy algorithm and regulation of any data source.

Probabilistic and Fuzzy matching techniques are required for effective matching. This fuzzy matching model matches patient (ID, Name & Address) across various datasets (Hospitals silos) through phonetics, approximate spellings and synonyms. This is an advanced mathematical procedure and determines the similarities among information, facts and datasets. This same result was neither every time true nor false or hundred percent sure, results a word "Fuzzy". In this technique any length, any kind of data and from any place is compared for finding the non-exact matching. A probability score is generated to each piece of examined data using fuzzy matching technique for determining the accuracy of matching. Similarity is checked while calculating the distance among the information pieces. If the distance is more than the two, datasets are more dissimilar.

The enormity and speed of data is considered to match patient records, a MapReduce algorithm is presented implementation of Levenshtein Distance algorithm and distance among the terms can be low number for single characters modifies required for making single term similar as different term. These modifies can be substitutions or deletions and insertions. MapReduce was a programming technique based on distributing calculations over huge volume of information. It included an implementation structure to larger scale data processing on commodity server clusters. Originally it is designed by Google and build on the known familiar rules in distributed and parallel processing since many years. Since MapReduce had liked wide spread of assumption through implanting an open source named Hadoop and it was developed by Yahoo (currently it is an Apache project).

MapReduce is utilized for matching data of patients. The patient multiple attributes are DoB, Name, City, Address, Pin code, etc. The person identity differentiates the person from others but various part of this identity will differentiates to a lesser or greater extent. For instance persons are differentiated greatly with their names instead of DoB as several individuals are born on similar date than the persons with similar name. Every feature detected had combined weight. The features of patient are allocated with disparate weights (i.e. Last Name compare counts are greater than the count of First Name, DoB match counts are more than City match counts). As per the requirements of healthcare, these weights can be configurable. To most set of structures, distance can be evaluated among the related attributes using presented solution. Distances are calculated attribute wisely on the entity attributes for determining distance in between two entities. Information about the detection of a patient was exploded and various documents were produced with keys from Name, Gender, DoB, Address 1 and Address 2. Classes over these, weighted distance is calculated in reduce phase to observe if any two records are matched for considering as duplicates. For identical records distance is 0. Distance is closer to 0 to the records which is same another for very few typographical errors. Various documents can have higher distance values. Weighted distance can be calculated through above mentioned patient records executing later the approach in Hadoop MapReduce. If the distance is less then more chances will be there for finding the patient records.

3.3 Preprocessing

The preprocessing steps are given below:

Step 1: the Cleansing procedure filtered the duplicate and unwanted data and reduces dataset size.

Step 2: The cleaned data integration is accessible as unstructured and structured information for data storage.

Step 3: Later, the completion of above mentioned steps, the size of data can be reduced through the extraction of essential attributes and redundant data will be eliminated.

Step 4: transformation of information to scaled values to fit in a slight range. Next data is examined and formed as centralized data, with respect to Aadhar card number of patient i.e., 12 digit unique identity number is given by UIDAI (Unique Identification Authority of India). The Aadhar inclusion provides stronger authentication to e-health services. In addition, the number of intervals that patient taken the medication in various hospitals will be tracked in selected way.

3.4 Data Mining

After the completion of data collection, preprocessing steps in methodology stage 1 a COPD dataset is created for COPD disease prediction. Utilized dataset containing 8 attributes for accuracy and classification of COPD disease. In data mining tools, method of classification handle with issue identification while observing the COPD disease characteristics between the patients and predicts or diagnosis that algorithm exhibits better performance on WEKA's statistical result basis. The classification model obtains a list of classified instances (training set) and is utilized to train the algorithms. The classification of testing data takes place with trained algorithms basing on extracted patterns and rules from training set. The utilized accuracies are given as:

1. Exactly classified accuracy represents the test accuracy percentage which is classified exactly.
2. Incorrectly classified accuracy indicates the test accuracy percentage which is classified incorrectly.
3. Mean Absolute Error indicates number of error for analyzing the accuracy of classification.
4. Time defines the required time for building a model to a disease prediction.
5. The ROC (Receiver Operating Characteristics) represents the test performance guide for diagnostic test classification accuracies basing on fail (0.50 – 0.60), poor (0.60-0.70), fair (0.60-0.70), good (0.80-0.90), excellent (0.90-1).

First the data is preprocessed and filtered. File was provided in arff format. Then results of classification accuracy can be analyzed by choosing J48 algorithm with 10 cross validation. The "visited" node represents the visits of a patient to hospital. In this work COPD diagnosis is done on number of times that a patient went for check up to the hospital and was shown in clinical data and signs in each check up are also recorded. During the first visit patient has taken treatment and in next check up doctor is advice to decrease smoking; if indications are not decreased then COPD test was suggested for the patient.

IV. RESULTS

In this paper Big Data analytics technique is presented for the solution to match huge volumes of patient records using MapReduce along with Fuzzy logic is outlined with the instance of indicative data as shown in Figure (2). It shows the 25 records of patients from 4 different systems. After consolidation this partial dataset shows how same patient record will appear many times.

Patient ID	Patient Name	Gender	DOB	Address1	Address2
1	Richard Gomez	M	11/16/1996	2934 Encino Pt	San Antonio TX 78259
2	Jim Dobbs	M	6/4/1938	2756 Bulls Bay Hwy	Jacksonville FL 32220
3	Robert Lewis	M	12/24/1971	194 Buckboard Dr	Augusta GA 30907
4	Rick Gomez	M	11/16/1969	2934 Encino Point	San Antonio TX 78259
5	Dharam Patel	M	8/11/1937	84 Prospect Hill Dr	Tewksbury MA 01876
6	Richard Gomez	M	11/6/1996	2935 Encino Pt	San Antonio TX 78259
7	Jim Dobbs	M	4/6/1938	2756 Bulls Bay Hwy	Jacksonville FL 32220
8	Ellis Cornwell	M	11/16/1996	4031 Laurel Lane	Midland TX 79701
9	Elizabeth Lange	F	6/4/1938	1447 Orchard Street	Minneapolis MN 55401
10	Salvador Keels	M	12/24/1971	339 Elliot Avenue	Seattle WA 98115
11	Richard Gomaz	M	11/9/1996	29350 Encino Pt	S. Antonio TX 78259
12	Edith Morgan	F	12/31/1948	2145 Stout Street	Carlisle PA 17013
13	Robert Dorman	M	8/11/1937	80 Snowbird Lane	Omaha NE 68144
14	Nicole Zorn	F	11/2/1986	1559 Shingleton Road	Grand Rapids MI 49503
15	Greg Hill	M	7/5/1987	4384 Elkview Drive	Stuart FL 34994
16	John Boyd	M	4/4/1990	1849 Liberty Street	Plano TX 75074
17	Kenneth Austin	M	7/19/1939	2497 Rockford Road	Mcgill NE 89318
18	Crystal Walker	F	12/28/1937	2437 Church Street	New York NY 10017
19	Mildred Cross	F	6/16/1956	1227 Fowler Avenue	Norcross GE 30093
20	Donald Sorensen	M	10/7/1963	3323 Oak Avenue	Hickory Hills FL 60457
21	Nancy Hughes	F	7/19/1938	60 Losh Lane	Hickory PA 15340
22	Valerie Brown	F	10/4/1975	1867 Ward Road	El Paso TX 79902
23	Sandra Philpps	F	1/4/1971	248 Railroad Street	Jacksonville FL 32207
24	Boyd John	M	4/4/1990	1849 Liberty Street	Plano TX 75074
25	Rich S. Gomez	M	11/16/1996	2934 Encino Point	San Antonio TX 78259

Fig. 2: SAMPLE PATIENT DATA FOR PATIENT MATCHING EXPERIMENT

For example, weighted Levenshtein distance is calculated for Patient ID 1, as shown in figure (3) with other patients. Spend the attention on documents of ID1 & 6 and distance among them is

0.5. They seem to be same except certain typographical errors. In the same way, the records of patient ID 1&4, 1&11 and 1& 25 has less distance and examined as duplicates (represents in green color). Remaining documents have longer distances (represented as blue color) and does not categorize as duplicate documents compared to patient ID1. The thresholds are termed automatically for combing the documents of patient and to assign the records for manual reviewing e.g. the records which have less than 3 distance scores will be merged automatically. For manual review the distance between 5 & 3 is marked. Rest is ignored as nano matches. If 1 point reduction is take place in threshold then there would be multifold increase in number of records which may required for reviewing the massive datasets of patients manually.

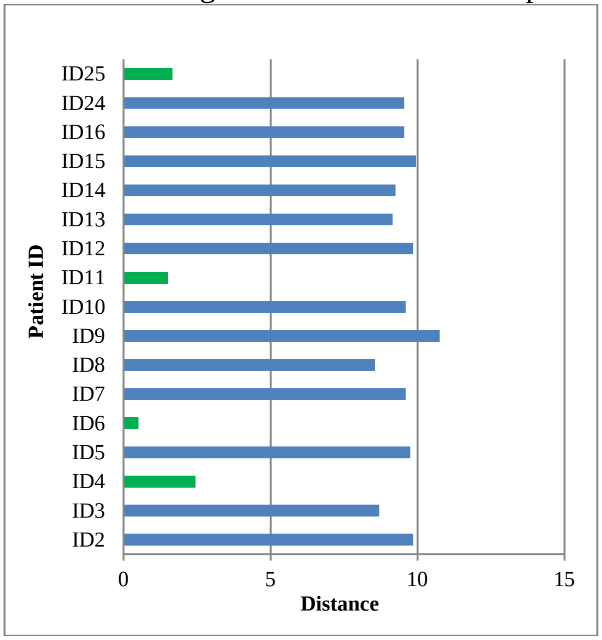

Fig. 3: LEVENSHTEIN DISTANCE OF PATIENT ID 1 FROM OTHER PATIENTS

However from the front end the Bid Data programs night looks like bulky investment, utilization of Big Data analytics inside a hospital can leads to reduction in operation costs by the detection of in capabilities in patient recognition while decreasing the risk for wrong treatment and diagnosis by improving the adequate of patient and improving the Clinical Decision Support Systems. In Pharamacy, real-time unified view of a patient can support for identifying and reducing the customer instances containing various contacts with pharma contacting channels to the equal transaction. This solution can be worked for any volume of data with high velocity as in several Healthcare systems. As an open source structure, this result will be low-cost considered too out of box costly results. Its flexibility is different benefits for handling any kind of information e.g. Pdfs, Excels, Flat files, RDBMS, etc.

In this work I48 tree is utilized for deciding the value of target basing on different dataset attributes for predicting the ML model and classifying its accuracy. In addition J48 tree is utilized on COPD disease dataset. The classifier output is analyzed after the execution of this algorithm, the output provides various statistics basing on 10 cross validation for predicting the each dataset instances. The Fig. 4 represents the explorer interface during the prediction of COPD disease utilizing data mining tool Weka and using COPD dataset in arff file format with their graphical view.

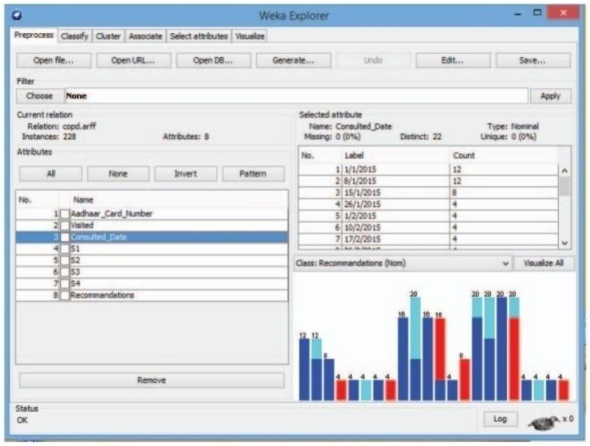

Fig. 4: VIEW OF ARFF COPD DATASET FILE IN WEKA EXPLORER INTERFACE

The experimental results represent the effectiveness and efficiency of J48 algorithm during COPD prediction. The Table 1 indicates the achieved classification accuracy of J48 algorithm is 92% and area of ROC is 1.

Table 1: STATISTICAL ACCURACY OF J48 TREE

Class / Parameter	Medication	Test for COPD	Reduce Smoking
TP Rate	0.857	1	1
FP Rate	0	0	0.106
Precision	1	1	0.667
Recall	0.587	1	1
F-Measure	0.923	1	0.8
ROC Area	0.957	1	0.93

The measurement of ROC area is approximately more than 0.8 to all classes which means the process of classification is succeeded in training the dataset. Hence the predicted instances are similar to the training set; this proved the presented classification approach. The J48 algorithm is utilized to extract and retrieve the data for appearing as unseen data. Extracted data is utilized for achieving the quality in several fields.

V. CONCLUSION

This paper presented a Clinical Decision Support system that provides an enhanced Patient matching process to large volumes and high speed data on COPD prediction using Big Data Analytics. This performance investigated enduring and effective Big Data analytics with MapReduce and Fuzzy algorithm (Levenshtein Distance) for matching the massive records of patients to make better clinical decision. Then COPD dataset is analyzed in this paper using J48 classification algorithm and different class of stages are predicted based on attributes. Once data is retrieved from various resources then the classification algorithms are applied for categorizing the data after performing preprocessing. Healthcare systems that can internally follow this system will ensure their financial stability, long and short term clinical success and ultimately their survival.

VI. REFERENCES

[1] Muhib Anwar Lambay;S. Pakkir Mohideen, "Big Data Analytics for Healthcare Recommendation Systems", International Conference on System, Computation, Automation and Networking (ICSCAN), 2020

[2] Rohit Ranchal, Paul Bastide, Xu Wang, Aris Gkoulalas-Divanis, Maneesh Mehra, Senthil Bakthavachalam, Hui Lei, Ajay Mohindra, "Disrupting Healthcare Silos: Addressing Data Volume, Velocity and Variety With a Cloud-Native Healthcare Data Ingestion Service", IEEE Journal of Biomedical and Health Informatics, Volume: 24, Issue: 11, 2020

[3] Fiaidhi, Jinan, and Sabah Mohammed. "Thick Data: A New Qualitative Analytics for Identifying Customer Insights." IEEE IT Professional, Vol 21, No. 3 (2019): 4-13

[4] P. Appel, V. F. de Santana, L. G. Moyano, M. Ito, and C. S. Pinhanez, "A Social Network Analysis Framework for Modeling Health Insurance Claims Data," arXiv preprint arXiv:1802.07116, 2018.

[5] J. Wang, M. Qiu and B. Guo, "Enabling real-time information service on telehealth system over cloud-based big data platform", J. Syst. Archit., vol. 72, pp. 69-79, Jan. 2017.

[6] Min Chen, Yixue Hao, Kai Hwang, Lu Wang, Lin Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities", IEEE Access, Volume: 5, 2017

[7] M. Shamim Hossain, Ghulam Muhammad, "Healthcare Big Data Voice Pathology Assessment Framework", IEEE Access, Volume: 4, 2016

[8] Cheng Hongbing, Rong Chunming, Hwang Kai, Wang Weihong, Li Yanyan, "Secure Big Data Storage and Sharing Scheme for Cloud Tenants", IEEE Security Schemes and Solution, June 2015, pp. 106- 115

[9] Joseph M. Woodside. Virtual Health Management, 2014 11th International Conference on Information Technology: New Generations 978-1-4799-3187-3/14.

[10] Kiyana Zolfaghar, Naren Meadem, Ankur teredesai, Senjuti Basu Roy, Si-Chi Chin, "Big Data Solutions for Predicting Risk-of-Readmission for Congestive Heart Failure Patients", 2013 IEEE International Conference on Big Data, 978-1-4799-1293-3/13.

[11] Zheng-hang Yan, Yang Liu, "The multidirectional information fusion mobile health management technology", 2013 First International Symposium on Future Information and Communication Technologies for Ubiquitous HealthCare (Ubi-HealthTech), 2013

[12] Meeyoung Park, Hariprasad Sampath kumar, Bo Luo, Xue-wen Chen, "Content-based assessment of the credibility of online healthcare information", 2013 IEEE International Conference on Big Data, 2013

[13] mill BG, Curtis LH, Fonarow GC, Heidenreich PA, Yancy CW, Peterson ED, Hernandez AF. Incremental value of clinical data beyond claims data in predicting 30-day outcomes after heart failure hospitalization. Cardiovasc Qual Outcomes, 4(4):60-67, 2011.

[14] P. Liu, L. Lei, J. Yin, W. Zhang, W. Naijun and E. El-Darzi, "Healthcare data mining: Prediction inpatient length of stay", Proc. 3rd Int. IEEE Conf. Intell. Syst., pp. 832-837, Sep. 2006.

[15] S. D. Pearson, S. F. Kleefield, J. R. Soukop, E. F. Cook and T. H. Lee, "Critical pathways intervention to reduce length of hospital stay", Amer. J. Med., vol. 110, no. 3, pp. 175-180, Feb. 2001.