International Journal of Research in Advent Technology, Special Issue, ICATESM 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Healthcare System for Chronic Obstructive Pulmonary Disease (COPD) Patient

Pranali Hyalij[1], Nikita Jadhav[2], Shivani Bhardwaj[3], Sheetal Kadam[4], Prof. Gokul Patil [5]
Computer Engineering Department [1,2,3,4,5],
Sandip Institute Of Technology And Research Centre Nashik, India [1,2,3,4,5]
Email: pranalihyalij2208@gmail.com[1], jadhavprachivij01@gmail.com[2], mbshivani30@gmail.com[3], kadamsweety13@gmail.com[4], gokul.patil@sitrc.com [5]

Abstract: Chronic Obstructive Pulmonary Disease (COPD) is a serious and progressive respiratory disease that causes lung damage and blocks the airways. People with COPD may experience chronic and progressive breathlessness, cough, sputum or mucus production, wheezing, and chest congestion. The goal of system is diagnose the stage of Chronic Obstructive Pulmonary Disease (COPD) in the patient. .After diagnose the stage of COPD in the patient provide the proper medication to the patient. Healthcare data faced with the challenges such as volume, velocity, variety, and veracity, health systems need to adopt technology capable of collecting, storing, and analyzing this information to produce actionable insight .This research investigates using cloud computing and data mining techniques for improve the performance of diagnose the COPD in the patient. The centralized clinical data repository contains patient's details with reference to unique aadhaar card number. Aadhaar card number helps to know about the treatments taken by the each patient in different hospitals and doctor treated.

Keywords*:* Aadhaar card Number, (COPD), cloud computing, data mining, wheezing, chest congestion.

1. INTRODUCTION

Big data is one of the latest and trending technologies accompanied by the digitalization of health and patient data through advance in information technology which leads to the generation of large volume of data in health care domain. Big Data is a term which is specifically use to handle both structured and unstructured data. A healthcare professional finds it difficultly to identify patients at high risk of having COPD but knowledge of the patients clinical history may make it easy for them. Data mining is the process of examining large datasets to extract hidden and unknown patterns, relationships and knowledge that are very difficult to identify with traditional statistical methods. This research investigates using cloud computing and data mining techniques for improve the performance of diagnose the COPD in the patient. This system provide the suggestions to the the COPD patient in the form of medication.

2. LITARATURE SURVEY

Chronic Obstructive Pulmonary Disease (COPD) is a type of obstructive lung disease which is characterized by long-term breathing problems and poor airflow.[1][8] The main symptoms include shortness of breath and cough with sputum production. Between 1990 & 2010 the number of deaths from COPD decreased slightly from 3.1 million to 2.9 million[156] & became the fourth leading cause of death.[2] In 2012 it became the third leading cause as the number of deaths incresed again to 3.1 million.[157] In different countries, mortality has decreased in men but increased in women.[158] . This is similar due to rates of smoking in women and men becoming almost same.[16] COPD is more common in older people;[9] it affects 34–200 out of 1000 people older than 65 years, depending on the population under review.[9][67]In England, an estimated 0.84 million people (of 50 million) have been diagnosed from COPD; this translates into approximately one person in 59 receiving a diagnosis of COPD at some point in their lives. In the most socioeconomically deprived part of a country, one in 32 people were diagnosed with COPD, compared with one in 98 in the most affluent areas.[159] In the United States approximately 6.3% of the adult population and totaling approximately 15 million people, have been diagnosed with COPD.[160] 25 million people may have COPD if currently undiagnosed cases are included.[161] In 2011, there were approximately 730,000 hospitalizations in the United States for COPD.[162] In the United States, COPD is estimated to be the third leading cause of death in 2011. As of 2015, COPD affected about 174.5 million (2.4%) of the global population.[6] It typically occurs in people over the age of 40.[3] Males and females are affected equally commonly.[3] In 2015, it resulted in 3.2 million deaths, up from 2 million deaths in 1990.[7][12] More than 90% of these deaths occur in the developing world.[3] The number of deaths is projected to increase further because of higher smoking rates in the developing world, and an aging population in many countries.[13] It resulted in an estimated economic cost of $2.1 trillion in 2010.[14]

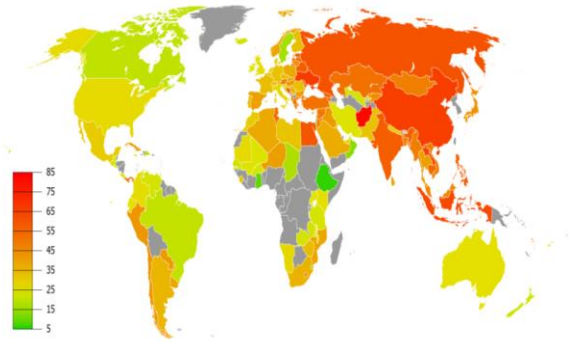

Fig1.1 Percentage of males smoking tobacco as of the late 1990s and early 2000s. Note the scales used for females and males differ.

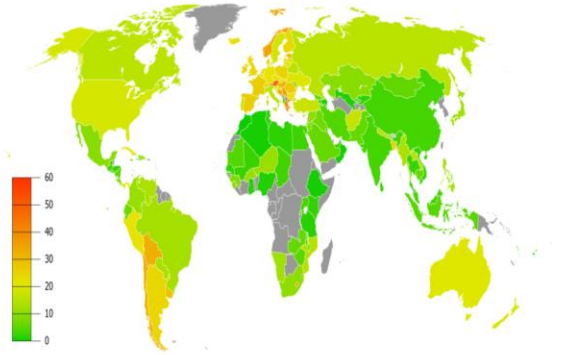

Fig 1.2 Percentage of females smoking tobacco as of the late 1990s early 2000s.

3. ARCHITECTURE

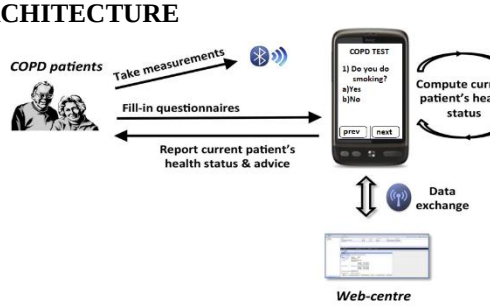

Fig 2.1 Architecture diagram of COPD system

The fig 2.1 shows Architecture diagram of COPD system in which firstly patient will login into the system and then he/she will fill-in questionnaire form and then system will process the given data with the help of cloud and web-center we are using KNN algorithm for prediction of COPD disease.

4. METHODOLOGY

The proposed research work is carried out in three phases as follows:

Phase I: Data collection and preprocessing

Figure 3.1 shows that the data is collected from different data sources and the data types includes survey form, physician's notes, diseases registries,paper documents and different form of medical reports. In next stage we collected records from different healthcare systems (hospitals, physicians, health centers, clinicians etc.) in different format such as .csv and tables .Then in next stage data is design in the form of COPD Data set and then data is preprocessed as data has many challenges like data is unstructured, duplication of data and the irrelevant data. Followings are the steps for data Preprocessing:

• Step 1: The Cleansing process includes filtering of unstructured and duplicated data; which reduces the size of the dataset.

• Step 2: Integrating the cleaned data which is in structured format in data warehouse.

• Step 3: After the first two steps, the data size is reduced by extracting necessary attributes which are required and eliminating duplicate data.

• Step 4: Transforming data to scaled values to fill within a smaller range.

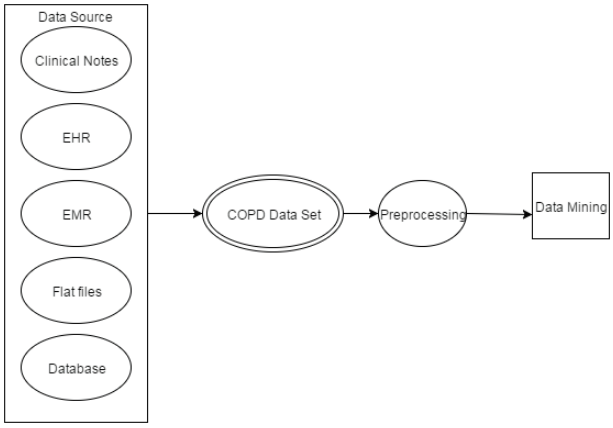

Fig 3.1 Data colleting and Preprocessing of COPD Dataset

Then later the data is analyzed and made as a centralized data with reference to the patient Aadhaar card number which is 12 digit unique identification number provided by UIDAI (Unique Identification Authority of India– .A Government Body established in 1999 under Planning Commission of India). In e-health services the inclusion of Aadhaar will provide a strong authentication. The number of times that each patient takes treatment in

different hospital can also be known and tracked in choice manner.

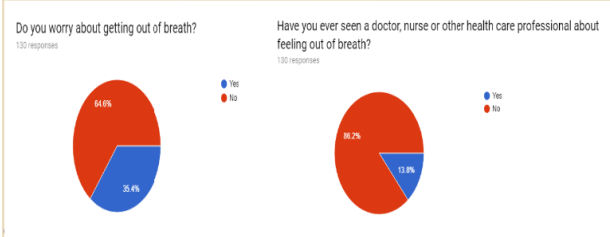

Fig3.2.1 Outcomes of Survey form for Breathlessness

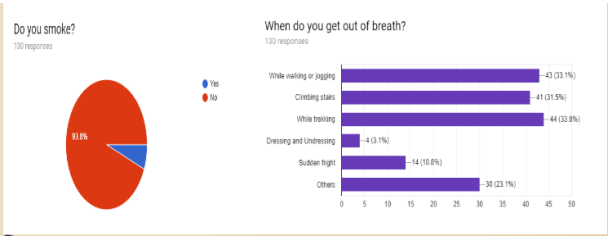

Fig3.2.2 Outcomes of Survey form for Breathlessness and Smoking

Fig3.2.1 shows the outcome of survey form in which 35.4% people are worried about getting out of breath.
Fig3.2.2 shows the outcome of survey form in which 93.8% people don't smoke and bar shows different activities in which people get out of breath where 33% people get out of breath while jogging & trekking and 3% while dressing and undressing.

Phase II: Preparing Dataset and Classifying

Dataset is a collection of data and it contains different forms like tables, pie-charts, bar graphs, etc. where each attribute of data represents variable and each instance has its own description. For prediction of COPD disease we have created COPD data set after following the steps of data collection and preprocessing in phase 1 of methodology. The COPD dataset used by us contains 8 attributes for COPD disease classification and accuracy. Figure 3.3 shows a description of COPD dataset.
Fig 3.3 Screenshot view of COPD Dataset and by using data mining tools, the classification process deals with identifying the problem by observing the characteristics of COPD disease among the patients and diagnoses or predicts COPD stage in patient

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Name	Surname	DOB	Age	Gender	Height,cm	Weight,kg	BMI	Smoker	Pack-Year	FVC-pre	FVC-post	FEV1-pre	FEV1-post	FEV1/FVC %-pre	FEV1/FVC %-post
2	Darshan	Abad	05-Oct-90	27	Male	178	87	27.46	No	No	3.68	4.37	3.49	4.11	94.8	94.1
3	M.	Abamadevi	15-May-78	39	Female	152	56	24.24	No	No	3.28	3.34	2.7	2.92	82.3	87.4
4	Manda	Abhale	30-Jan-76	41	Female	163	48	18.07	No	No	1.14	1.35	0.69	0.84	60.5	62.2
5	Yogesh	Abhang	23-Oct-92	25	Male	163	58	21.83	No	No	3.26	3.43	2.98	3.09	91.4	90.1
6	Nagesh	Abhonkar	12-Dec-42	74	Male	167	75	26.89	Ex	0.25	1.69	2.34	1.41	1.62	83.4	69.2
7	Gigi	Abraham	15-Feb-71	46	Female	165	88	32.32	No	No	2.85	3.02	2.56	2.62	89.8	86.8
8	Kailas	Adgonkar	11-Oct-78	38	Male	182	90	27.17	No	No	3.32	3.25	2.64	2.77	79.5	85.2
9	Amit	Adhangale	29-Nov-80	37	Male	162	62	23.62	No	No	3.42	3.26	2.67	2.61	78.1	80.1
10	Sangeeta	Adhav	18-Nov-63	54	Female	150	55	24.44	No	No	1.16	1.29	0.93	1.13	80.2	87.6
11	Pallavi	Adke	31-Jul-85	31	Female	154	61	25.72	No	No	2.7	2.69	2.35	2.29	87	85.1
12	Nupur	Admane	10-Oct-97	19	Female	161	53	20.45	No	No	3.16	3.23	2.85	3.01	90.2	93.2
13	Bharatbhai	Adroja	01-Jun-74	43	Male	169	65	22.76	No	No	2.19	2.57	2.01	2.52	91.8	98.1
14	Shakuntala	Adsare	01-Jun-60	58	Female	154	64	26.9	No	No	1.97	2.1	1.43	1.55	72.6	73.8
15	Sita	Agarwal	20-Feb-56	61	Female	155	82	34.13	No	No	0.79	0.93	0.41	0.5	64.5	53.8
16	Ajinath	Aghav	12-Jul-83	35	Male	169	81	28.36	No	No	3.94	3.95	3.4	3.47	86.3	87.8
17	Ashabai	Aghav	01-Jun-60	56	Female	158	63	25.24	No	No	2.07	2.1	1.96	1.99	94.7	94.8
18	Chitra	Agnihotri	13-Mar-67	51	Female	146	75	35.18	No	No	1.95	1.91	1.92	1.91	98.5	100
19	Shlini	Agnihotri	22-Nov-56	60	Female	154	61	25.72	No	No	0.94	1.29	0.38	0.54	40.4	41.9
20	Anand	Agrawal	17-Jul-82	35	Male	160	87	33.98	Yes	0.15	3.2	3.33	2.71	2.93	84.7	88

Fig 3.3 Screenshot view of COPD Dataset

The classification learning algorithm takes a set of classified examples (training set) and use it for training the algorithms. With the trained algorithms, classification of the testing dataset takes place based on the patterns and rules extracted from the training set.
The data is first preprocessed in phase I and then filtered. The file input is given in csv format and then analyzes the classification accurately by selecting the KNN algorithm.

Phase III: Implementation of KNN algorithm.

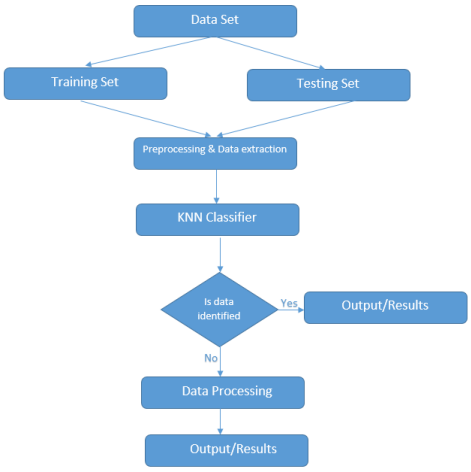

Figure 3.4 KNN Flow

KNN is one of the supervised machine learning algorithm and it is based on feature similarity and it can be used for both classification and regression predictive problems using KNN classifier.
It classifies a data point based on how its neighbors are classifies.
Here in this research we are diagnosing COPD on the number of times the patient has visited the hospital which is present in clinical history and the symptoms during each visit is also noted. During first visit patient

is given medication and on second visit told to reduce smoking and if symptoms not reduced the patient is recommended for COPD test.

5. CONCLUSION

This research investigates using cloud computing and data mining techniques for diagnose the COPD in the patient. The centralized clinical data repository contains patient's details with reference to unique aadhaar card number. This system predict the COPD in the patient and provide the Medication to the patient.

REFERENCES:

[1] Nathell L, Nathell M, Malmberg P, Larsson K (December 2007). "COPD diagnosis related to different guidelines and spirometry techniques". Respiratory Research. **8** (1): 89. doi:10.1186/1465-9921-8-89. PMC 2217523. PMID 18053200.

[2] Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguez-Roisin R, van Weel C, Zielinski J (September 2007). "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary". American Journal of Respiratory and Critical Care Medicine. **176** (6): 532–55. doi:10.1164/rccm.200703-456SO. PMID 17507545.

[3] GBD 2015 Disease and Injury Incidence and Prevalence Collaborators (October 2016). "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015". Lancet. **388** (10053): 1545–1602. doi:10.1016/S0140-6736(16)31678-6. PMC 5055577. PMID 27733282.

[4] GBD 2015 Mortality and Causes of Death Collaborators (October 2016). "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015". Lancet. **388** (10053): 1459–1544. doi:10.1016/S0140-6736(16)31012-1. PMC 5388903. PMID 27733281.

[5] Roversi S, Corbetta L, Clini E (5 May 2017). "GOLD 2017 recommendations for COPD patients: toward a more personalized approach". COPD Research and Practice. **3**. doi:10.1186/s40749-017-0024-y.

[6] Vestbo J (2013). "Definition and Overview". Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care Medicine. **187**. Global Initiative for Chronic Obstructive Lung Disease. pp. 1–7. doi:10.1164/rccm.201204-0596PP. PMID 22878278.