An Insight into Public Healthcare Data of Punjab State: A Data Mining Approach

Parveen Singh[1], Vibhakar Mansotra[2]
[1] *Department of Computer Science & IT, University of Jammu, Jammu, India*
imparveen@yahoo.com
[2] *Professor, Department of Computer Science & IT, University of Jammu, Jammu, India*
vibhakar20@yahoo.co.in

Abstract **-** In the Information Technology driven society, knowledge assumes maximum significance. Health care sector in particular requires an effective Knowledge Management System in order to make sense of the data generated. There are challenges like creation, dissemination and preservation of this knowledge, which can be solved with the use of Database Systems, Data Warehousing and the Knowledge management Technologies. Healthcare has so far relied on the ability of its individual practitioners to turn the available data into effective Knowledge. This information has increased to a point of glut and yet the multidimensional Analysis rendered to it by the various ICT tools have only served so far, increases the complexity, without furthering insights into the cases. In India, with the passage of time, health Care system has improved a lot, but health care data is still not used for the policy formulation and health planners and management personals mainly rely on the experiences to take decisions. In this paper, we explored the facility level data of Punjab state and applied Knowledge discovery and Data Mining techniques ,using SPSS Modeler, to effectively leverage the available knowledge management resources for policy formulation and effective utilization of resources available. We explored the structure of public health care institutions prevalent in with the distribution of patients in different districts using public health care data available with HMIS database and observed that distribution pattern of heath care facility is not uniform in the state. We put forward suggestions, which can help in improving the overall performance of public health care institutions and patient services.

Keywords**:-***Data mining, HMIS, Health centres, Primary health centres, Community Health centers, Compound annual growth rate.*

I. INTRODUCTION

Public health is the science of treatment and prevention of disease, prolonging life and promoting human health through promotion of healthy lifestyles, research for diseases, injury prevention and detection and control of infectious diseases of entire populations. These populations can be as small as a local neighborhood, or as big as an entire country or region of the world [1]. Public health care services generally refer to cost effective health care system funded by government and available to support the health care need of all members of the population [2].

In India, Healthcare is mostly governed by state government and central governments and in addition private organisation are playing good role in healthcare system. India is very big country in terms of geographical variations and population. The Public healthcare sector in India is expected to grow at Compound annual growth rate (CAGR) of 22.9 per cent during 2015-20 to US$ 280 billion and with the rise in income level, health awareness, increased precedence of lifestyle diseases and improved access to insurance seems to be important contributors to this growth [3]. In addition to Public Healthcare system, the private sector also has been playing a crucial and vibrant role in India's healthcare industry. And it is very surprising fact that private health care system in india accounts for almost 74 per cent of the country's total healthcare expenditure[5].

In Health care sector, Telemedicine is a fast-emerging trend in India and very reputed hospitals Like Apollo, AIIMS, Narayana Hrudayalaya and Escort have adopted telemedicine services and the telemedicine market in India is valued at US$ 7.5 million currently and is expected to grow at a CAGR of 20 per cent to reach US$ 18.7 million by 2017. Further, presence of world-class hospitals and skilled medical professionals has strengthened India's position as a preferred destination for medical tourism. During January-November 2016, 82 health technology companies have raised about US$ 80 million.

Keeping in facts in mind, the Government of India aims to develop India as a global healthcare hub. It has created the National Health Mission (NHM) for providing effective healthcare to both the urban and rural population. The Government is also providing policy support in the form of reduced excise and customs duty, and exemption in service tax, to support growth in healthcare.

Investment in healthcare infrastructure is set to rise, benefiting both 'hard' (hospitals) and 'soft' (R&D, education) infrastructure [4]

Public Healthcare services are catering the population of different Locations through public health centre located in rural and urban area of the country and all health centres operating in 29 states of india, maintains record of their primary activities and update the same on HMIS portal launched in 2008.Indian government initiates different programmes in Health care systems which are monitored by NRHM for review and Future Planning [4]
Data mining has very vital role in analysis of this huge volume of data captured from different Health care centres located at various parts of India.

II. PUBLIC HEALTHCARE IN PUNJAB

Punjab is one of the biggest states of India with total population of 29078483 till 2016. Its Public Health care facilties includes Sub Centres (N_SC): 2951, Primary Health Centres (N-PHC):427, Community Health Centres (N_CHC):150, Sub District Hospital(N_SDH) 41 and District Hospital(N_DH): 22 .Table 1 shows status of Public Health Facilities in Punjab till 2016 and it reflects that maximum number of Sub Centre hospital and Primary Health Centre are in Patiala followed by Ludhiana and Hoshiarpur. The National Rural Health Mission has been established with the vision to provide accessible, affordable and quality health care to the population with special focus on vulnerable sections living in rural areas. NHRM deals with establishment of health infrastructure in public domain and with careful examination of existing facilities from the data available, future Planning can be done to strengthen the infrastructure.

Table: 1 Status of Public Health Facilties in Punjab -2016

S. No	District	Population	N_SC	N_PHC	N_CHC	N_SDH	N_DH
1	Amritsar	2610519	213	47	20	2	3
2	Barnala	624187	80	13	12	1	1
3	Bathinda	1455348	167	26	19	3	2
4	Faridkot	647226	67	11	7	2	2
5	Fatehgarh	629046	86	16	14	3	1
6	Fazilka	1114929	115	23	14	1	1
7	Firozpur	1011794	132	23	11	1	1
8	Gurdaspur	1699770	235	33	27	1	1
9	Hoshiarpur	1662981	272	36	26	4	2
10	Jalandhar	2299156	226	38	25	3	1
11	Kapurthala	854398	122	15	16	4	1
12	Ludhiana	3667114	316	50	30	6	2
13	Mansa	806795	106	15	11	2	1
14	Moga	1043666	139	25	14	0	1
15	Mohali	1042494	110	23	10	2	1
16	Muktsar	945300	108	23	13	3	1
17	Nawanshar	641777	100	18	11	1	1
18	Pathankot	709159	74	9	8	1	1
19	Patiala	1986916	328	47	25	8	2
20	Rupnagar	717575	90	14	9	1	2
21	Sangrur	1734824	208	36	20	4	1
22	Tarn Taran	1173509	158	21	20	2	1

III. DATA COLLECTION FOR ANALYSIS PURPOSE

For Analysis Purpose, accuracy of desired data is very important. Marshall and Rossman describe data analysis as the process of bringing order, structure and meaning to the mass of collected data. It is described as messy, ambiguous and time-consuming, but also as a creative and fascinating process [5]. Data used in this paper has been collected from Ministry of Health & Family Welfare ,Government of India , HMIS Portal data of Public Healthcare system of Punjab State for the Year 2015-16 and we have applied data Cleaning Techniques for the missing attributes and values so as to prepare the data for the analysis. In addition, we have also used latest data of

year 2015-16 from the facts sheets uploaded on HMIS portal. For Exploratory analysis, we have used data mining software, SPSS Modeler, a very powerful tool, which gives detailed analysis , by providing numerous graphs and provides better view of all details desired by Researchers [2]

IV. DATA ANALYSIS

The SPSS Modeler tool has been used for Analysis of data obtained from HMIS Portal of Punjab for the Year 2015-16.The Data obtained from Portal was transformed into EXCEL files and after applying cleaning Techniques, analysis has been carried on various parameters which includes Distribution of OPD verses District wise Population, IPD verses District wise Population, Distribution of IPD verses OPD, Distribution of Sub Centre's district wise. The detailed analysis of each parameter is as under:

a) Distribution of OPD verses District wise Population

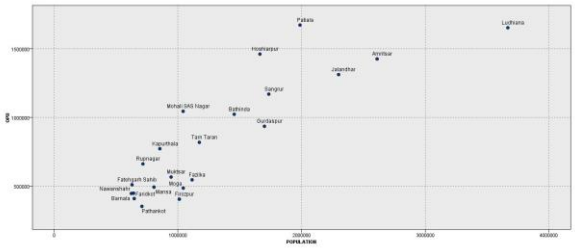

Fig.1: Graph between OPD w.r.t district wise Population

The careful Examination of graphs reveals that Ludhiana district is highly populous district so number of OPDs is also high among all districts where as Amritsar, Jalandahr are highly populous than Patiala and Hoshiarpur but number of OPD cases are less in these districts and it indicates the persons belonging to these districts prefers some other districts for routine checking. NRHM can review the infrastructure available at these places to avoid unnecessary movement of Patients from one district to another.

b) Distribution of IPD verses District wise Population

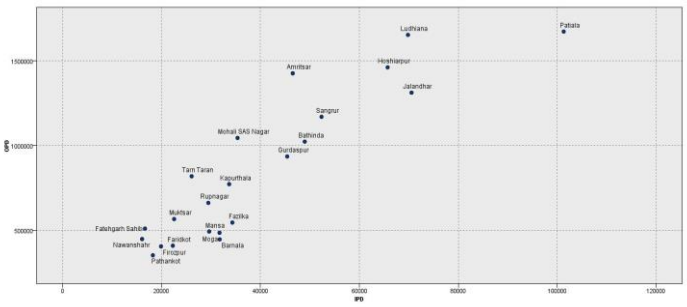

Fig.2: Graph between IPD w.r.t Population distrcitwise

It has been observed that Population of Patiala is lesser than Ludhiana, Amritsar and Jalandahr but number of IPD cases is much in Patiala than other Districts and approximately two Lac Patients are handled by Patiala whereas districts like TarnTaran, Barnala, Pathankot, Mohali, Rupnagar, Nawanshar deals with less Patients. It indicates distribution of Patients is not proper among districts. [3]

c) Distribution of IPD verses OPD Patients

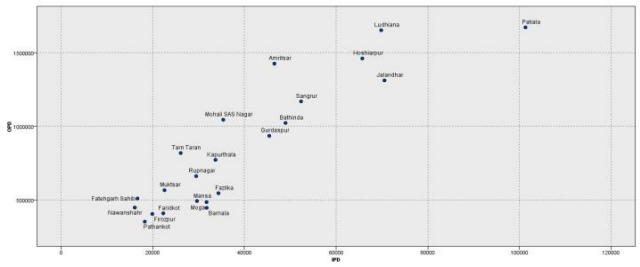

Fig.3: Graph between IPD w.r.t IPD District wise

On observation of patterns of OPD and IPD patients in Fig 3, we found that number of outpatient and inpatients are mostly concentrated in Patiala followed by Ludhiana, Hoshiarpur and Jalandhar and more than 3.5 Lacs patients are coming to Ludhiana for OPD and 2 Lacs patients are coming to IPD in Patiala. It means Ludhiana and Patiala are dealing with much more Patients in comparison to other districts [4]

d) Distribution of Pattern of Public health Institutions

The Public Health care system in India is divided in three Sections namely Primary, Secondary and Tertiary Level, Primary level includes SCs, and PHCs. Secondary Level includes CHCs or SDH and District Hospital. In Punjab, the variation in health care organizations is given below:

e) Community Health centre in Punjab

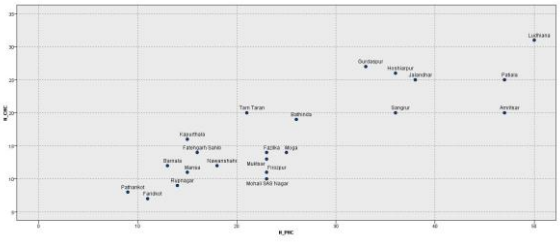

Fig.4: Graph between distributions of PHC/CHC across Districts

Community Health centre are basic centre of Health care system and initially all patients are referred to CHCs and

sometimes also patients from PHCs are provided referral health care from CHCs. in Punjab, There have been hardly any change in number of CHCs from the year 2014-2016 and community centres in Punjab has been not rationally distributed as Ludhiana and Patiala has highest numbers and Pathankot has lowest number.

f) Primary Health Centre in Punjab

Primary Health centre are major element of public health care system where qualified doctor have been engaged for treatment of patients. In Punjab, again we find that there has been no change in Number of Primary Health Centre from 2013-2016.again we find that Patiala has highest number of Sub Centre and little less PHC in comparison to Ludhiana but highest than other districts. Again, PHC has maximum concentration in Ludhiana, Patiala, Hoshiarpur, Jalandahr and Amritsar.[4]

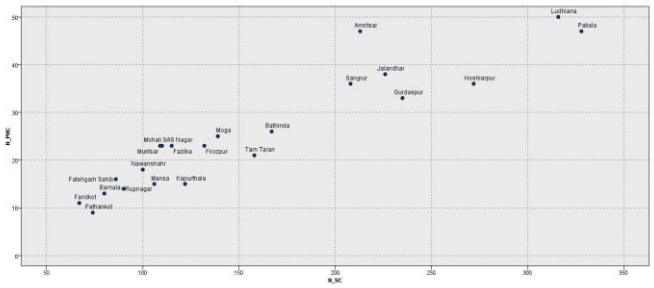

Fig.5: Graph between distributions of PHC/SC across Districts

g) Sub Centres and District Hospitals

Sub centers also acts as interface with community at grass root level and in Punjab number of Sub centres has been remain same from 2014 to 2016.

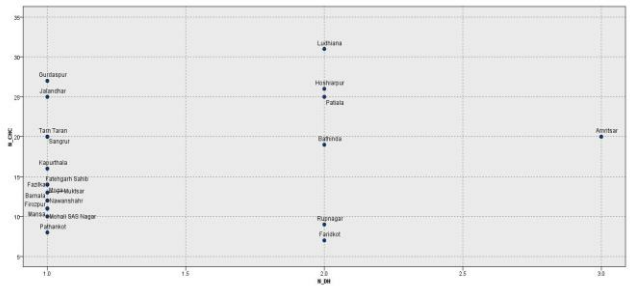

Fig.6: High Priority Districts in Punjab

h) High Priority Districts:

After careful Examinations, we find there are 22 Districts in Punjab with Ludhiana as most populous districts with maximum number of Primary Health Centre, Community Health Centre, Sub District Hospitals and Districts Hospitals but lesser Sub Centre than Patiala and in addition we found that Number of Sub Centres in Rupnagar, Pathankot, Fatehgarh Sahib, Barnala are much lesser in comparison to other districts even though population of these district is higher than others. In addition, we found that Patiala has been given much priority even though Jalandahr and Amritsar has much population in comparison to Patiala [6].

V. CONCLUSION

It has been seen that distribution of public healthcare institutions is not rational, with the maximum concentration seen around districts nearby capital of the state, Patiala. As the secondary and tertiary levels of public healthcare institutions are more available in Ludhiana, Patiala, patients approach to them easily.

Because of huge volume of data available in HMIS Portal and NRHM Portal, there is wider scope of research and analysis in the Direction of Anomaly Detection, Outlier Detections as well as data mining can be applied for better cluster generations and development of classification based models for further analysis.

VI. REFERENCES

[1]. P. Singh and V. Mansotra, "Data mining based Tools and Techniques in Public Health Care management:- A study", in INDIACom2017, New Delhi, 2017, pp. 425-29.

[2]. U. Fayyad, G. P Shapiro and P. Smyth, "The KDD process of extracting useful knowledge form volumes of data communication", ACM, vol. 39, no. 11, pp. 27-34, 1996.

[3]. M. Durairajl, "Data Mining Applications in Healthcare sector: A Study", published in international Journal of scientific & Technology research, vol.2, no.10, pp.3-8, 2013.

[4]. K Rangral, "Comparative Study of Data Mining tools", in International Journal of advanced Research in Computer Science and Software engineering, volume 4, no. 6,pp.45-51,2014

[5]. M. Herland, "A review of data mining using big data in health informatics", in Springer Journal of Big Data , vol. 1,no.1,2014.

[6]. P. Ahmad," Techniques of Data Mining in Healthcare: A Review", in International journal of Computer Applications ,vol,20,no.15, pp. 1- 6, 2015.