Correlating Fatality Rate to Road Accidents in India: A Case Study using Big Data

Subha Koley[a,b], Saket Srivastava[b], Prasun Ghosal[a]
[a]Indian Institute of Engineering Science and Technology, Shibpur, India
[b]University of Lincoln, UK
Email: subha@ieee.org, ssrivastava@lincoln.ac.uk, p_ghosal@it.iiests.ac.in

***Abstract*— Number of vehicles on Indian roads is increasing at a very high rate every year and the number of road accidents is rising at a similar rate. In 2016, around half a million (reported) people were injured in India due to different types of road accidents and out of them, around 150,000 people were killed. This leads to a very serious concern that there are some major flaws in emergency rescue services in the country. Big Data analysis and different statistical models can identify accident frequencies and patterns in a region, which may be useful to identify accident-prone regions in the country. A centralized database of all possible rescue authorities with their exact location and contact information can be a very important part of a smart accident reporting system and rescue operations. In this paper, we have studied the number of injuries in road accidents and deaths in most of the Indian states and proposed a model correlating them with the number of hospitals and police stations available in those states. This model will help not only to figure out critical accident-prone states in India but also to create a database for an emergency rescue system. The data used for this model has been generated using Google Radar Search and Reverse Geocoding API that can be very much useful to accelerate development f emergency rescue operations needed for Indian road systems and can be replicated easily for other countries.**

Keywords— Road Accidents, Big Data, Radar Search, Reverse Geocoding

I. Introduction

Rise in fatalities due to road accidents is increasing rapidly in India each year. There are a number of factors that have contributed to the rise in road accident related injuries and fatalities in the past few years with nearly half a million injuries being reported in 2016 across the country [1]. In a developing country like India, the absence of a centralized accident reporting and rescue system in a contributing factor towards a high percentage of accidents resulting in fatalities. Based on a recent report, 55 accidents occur in Indian roads every hour and out of those 17 people die [2]. In 2016, 480652 road accidents occurred in India where 494624 people got injured and out of those 150785 (30.48%) couldn't survive [3].

Data shows that there is plenty of hospitals, ambulances and police stations available in the country, but the only problem is, they are not able to reach at the right place at the right time due to lack of communication infrastructure. The response time of emergency services is a critical factor that determines the fatality rate during a road accident. Most of the victims of road accidents are killed due to the late arrival of rescue teams to the locations. It is obvious that the number of accidents will be higher in the urban areas than in rural areas as the number of vehicles are higher. But the number of deaths is lower in urban areas due to the obvious reason that there are more hospitals located in smaller distance and victims are transferred to the hospitals in a shorter time.

In a previously reported work [4], an IoT enabled real-time communication and location tracking system for a vehicular emergency has been proposed having the ability to report accidents immediately to the nearest hospital and police station. Such a system requires a huge database of all hospitals and police stations with an exact location across India for proper working. In this present work, data have been extracted and a database has been created with all hospitals, nursing homes, police stations, sub-police stations from all over India to categorize them with their administrative location. Road accidents and casualty data have also been collected and analyzed to understand the scenario of emergency rescue services in different states of India and classified accordingly.

Rest of the paper is organized in the following way. Some of the related researches and motivation for present work have been discussed in Section II. Section III describes the sources and procedures used to collect data about road accidents and develop possible rescue centers' database. Findings and results have been discussed in Section IV. And Section V finally concludes the paper and describes the future scope of the research.

II. Related Works

Road accident is one of the most unpredictable events that can happen at any time on any road. Researchers have tried to predict the probability of road accident in a certain place/ road by analyzing several statistical data. Data mining approaches have been used to classify and characterize road accidents. Singh et al. [5] have used M5 model tree to develop a predictive model for traffic accidents. Association rule mining and clustering techniques have been used by Kumar et al. [6] to identify circumstances where an accident may occur for each cluster. Researchers have also tried to identify the relation between road accident and associated factors by using Poisson models [7-8]. But the fact of the problem is all these researches only try to predict accident by analyzing the causal factors.

Internet of Things (IoT) and Big Data technology can help in speeding up the rescue operations and save some more precious lives. A GPS based location tracking system has been proposed by [9] that can send location information to

predefined number through SMS. Some researches [10] have also proposed a smartphone-based system to detect an accident and report it to the victim's relatives. As mentioned earlier, our previous work was an attempt to develop a smart system that can automatically report road accident to the nearest rescue authority. To make the system efficient and fully automatic a centralized database of all possible rescue centers (hospitals, nursing homes, police station etc) is highly required.

In the present work, we have developed this database and collected all road traffic accident data in India and categorized sates according to the number of accidents and the percentage of deaths from injuries.

III. Data Collection and Database Generation

In this section, we have listed various data sources used in this study to identify the actual scenario of road accidents in India as shown in Table 1.

Table I. Source/ Collection Procedure of Datasets

Sl. No.	Data Type	Source/ Procedure used
1	Location (latitude, longitude) of all hospitals and nursing homes in India	Google Maps Radar Search
2	Location (latitude, longitude) of all police stations and sub-police stations in India	Google Maps Radar Search
3	Administrative location (postal address) of all hospitals and nursing homes	Google Maps Reverse Geocoding
4	Administrative location (postal address) of all police stations and sub-police stations	Google Maps Reverse Geocoding
5	State/UT wise road accident number	Open Government Data (OGD) Platform India
6	State/UT wise number of injuries due to road accidents	Open Government Data (OGD) Platform India
7	State/UT wise number of deaths due to road accidents	Open Government Data (OGD) Platform India
8	State/UT wise population information	Census Data 2011

a) State wise Road Accident Data

It is very difficult to get an actual number of road accidents and casualties' information for a large country like India. There is no such centralized system for accident reporting in our country. Most of the accidents are reported by the local police stations and often contain only a little information about the incident. The detailed information regarding a road accident e.g. exact location of the incident (latitude, longitude), number of victims, types of injuries, the cause of the accident etc. are mostly unavailable in those reports. Therefore, in this study, we have only dealt with the reported road accidents in India in the year 2016. The reported number of road accidents, injuries and fatalities data have been collected from the Open Government Data (OGD) Platform India [11].

b) State Wise Population Data

To calculate per capita road accidents, injuries and deaths we needed state-wise population information for the year of 2016. But last nationwide census was conducted in 2011. Therefore, 2011's census data have been used in this study assuming population growth rate has been the same for all the states of India from 2011 to 2016. The population data for every state is has been collected from Census Data 2011 [12].

c) Locations of all Hospitals and Police Stations in India

To construct efficient and reliable database exact geographical location (latitude and longitude) of all possible rescue centers are required. The central database of hospitals, nursing homes, and police stations could not be used for our system for the following obvious reasons. Only big hospitals' list is available in the central database. But in the event of an emergency, any small hospital or nursing home can help the victims to survive. Administrative addresses and geographical locations (latitude, longitude) of those hospitals are mostly unavailable or incorrect.

Therefore, we started collecting authentic data from the Google Maps platform. But it is not possible to collect information from all over India through the normal Google Maps Places search [13] framework. Normally Google places search returns 20 places at a time. But to extract datasets for a large country like India, a manual searching procedure is not worth it. Google Maps Radar Search [14] is a suitable alternative to perform this task. This search process works at the highest level on Google maps' administrative locations. Radar search can return up to 200 search results at a time. To perform radar search user has to provide a location (latitude, longitude) as the central point of the searching region. Also, the user needs to provide the length of the radius of the searchable region. An API key [15] has been obtained to perform the searching program. Followings are the key parameters for the searching process.

- The Radius of the search
- Location of central point (latitude, longitude) of the searching region
- Searching place type
- Google Maps API key

To accomplish searching from all over India we firstly had to find out the central points (reference points) for searching and searching radius length. To identify the suitable length of the radius, we perform some arbitrary checking from different parts of India and found that in most of the regions there are less than 200 hospitals, police stations situated within a radius of 50 km region. Therefore, we used 50 km as our searching radius.

Next, to find out reference points for searching, we started segmenting the map with 100 km2 squares and used the corner points and the central point of every square, as reference points for searching as shown in Figure I. After collecting all results, duplicate values have been refined, and the points outside Indian map have also been removed. As a result, 921 points have been identified from all over India that can cover almost every part of the country.

Finally using Google Maps API key, Radar Search has been performed on those reference points with a radius of 50 km and searching key as "hospital". After completion of the program, again duplicate values have been removed and finally, we got a total of 105252 hospitals, nursing homes' geographic locations (latitude, longitude). The same procedure has been used with search keyword "police" and got 29390 police stations and sub-police stations' locations in India.

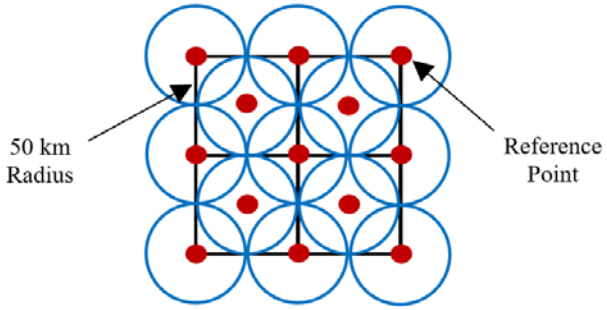

Figure I. Selection of reference points and search radius

d) Extracting Administrative Address of Rescue Centres

Google's Radar Search is a very useful platform to get a large number of places' location, but it only returns very few information about the places. Therefore, in the previous searching process we only able to collect the locations of the rescue centers. But to classify them region wise and present them in the human understandable format those places locations must need to be converted into administrative (physical) address format. Google Maps Reverse Geocoding [16] algorithm has been used to convert geographic location (latitude, longitude) to administrative address (state, district/ city, locality etc). We have collected the address of locations according to three levels viz. Administrative Area Level 1 (State/UT), Administrative Area Level 2 (District) and Locality. Python programming language has been used to program Radar search and Reverse Geocoding.

IV. Results and Discussion

This study has been conducted for 27 states in India excluding Punjab, Mizoram and the Union Territories because of unavailability of authentic data.

The collected data on road accidents are not normalized. So, it is very important to normalize those datasets before any comparison because the larger states should contain a higher number of hospitals and police station than the smaller states. To normalize the datasets, we have used the population of every state and calculated the corresponding number of rescue centers per 1 lakh people and calculated fatality rates from accidental injuries.

a) Findings from Accident Datat

It is obvious that the number of deaths will be lesser if there are hospitals near accident locations or police/ ambulance vehicles are near. The result also shows that the percentage of deaths in road injuries are higher in the rural areas.

Table II. Classification of States with Fatality Rate

Fatality Percentage	No. of States	Category
<25	9	1
25 – 50	13	2
50 – 75	2	3
>75	3	4

According to fatality rates, the states have been classified into four categories as shown in Table II. Nine states have been found with fatality rates less than 25% (Category 1), thirteen states with fatality rates between 25% to 50% (Category 2), two states with fatality rates between 50% to 75% (Category 3) and three states with fatality rates more than 75% (Category 4). Manipur (8.48), Kerala (9.72) and Jammu and Kashmir (12.45) have the lowest rate of death from accidental injuries whereas Bihar (86.73), Jharkhand (79.80) and Uttar Pradesh (76.98) have the highest rate. Detailed list of the states has been shown in Table III and graphically represented in Figure IV.

Table III. Category of States according to Fatality Rate

State	No. of Accidents	No. of Injury	Persons Killed	% of Death	Cate gory
Andhra Pradesh	24888	30051	8541	28.42	2
Arunachal Pradesh	249	391	149	38.11	2
Assam	7435	6127	2572	41.98	2
Bihar	8222	5651	4901	86.73	4
Chhattisgarh	13580	12955	3908	30.17	2
Goa	4304	2026	336	16.58	1
Gujarat	21859	19949	8136	40.78	2
Haryana	11234	10531	5024	47.71	2
Himachal Pradesh	3168	5764	1271	22.05	1

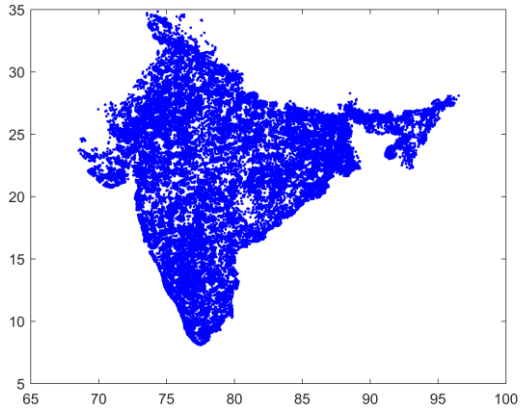

Figure II. Locations of all hospitals and nursing homes in India

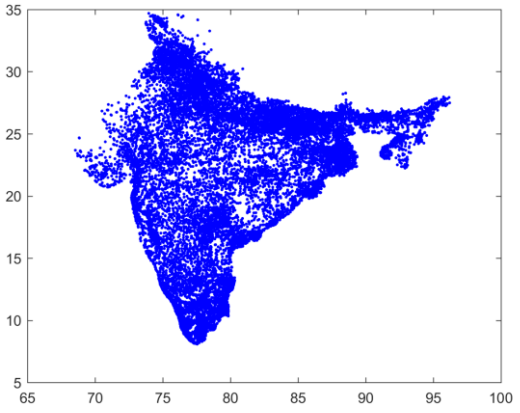

Figure III. Locations of all police station, sub-police stations in India

Jammu & Kashmir	5501	7692	958	12.45	1
Jharkhand	4932	3793	3027	79.80	4
Karnataka	44403	54556	11133	20.41	1
Kerala	39420	44108	4287	9.72	1
Madhya Pradesh	53972	57873	9646	16.67	1
Maharashtra	39878	35884	12935	36.05	2
Manipur	538	955	81	8.48	1
Meghalaya	620	354	150	42.37	2
Nagaland	75	120	46	38.33	2
Odisha	10532	11312	4463	39.45	2
Rajasthan	23066	24103	10465	43.42	2
Sikkim	210	263	85	32.32	2
Tamil Nadu	71431	82163	17218	20.96	1
Telangana	22811	24217	7219	29.81	2
Tripura	557	853	173	20.28	1
Uttarakhand	1591	1735	962	55.45	3
Uttar Pradesh	35612	25096	19320	76.98	4
West Bengal	13580	11859	6544	55.18	3

b) Findings from Rescue Centers' Data

From the hospitals and police station database it is found that Himachal Pradesh (19.94), Rajasthan (14.72) and Uttarakhand (14.66) have the highest number of hospitals and nursing homes per 1 lakh people whereas Jammu and Kashmir (0.49), Arunachal Pradesh (4.12) and Uttar Pradesh (5.00) have the lowest. In the other hand Nagaland (6.16), Manipur (5.40) and Himachal Pradesh (4.91) have the highest number of police stations per 1 lakh population whereas Jammu and Kashmir (0.14), West Bengal (0.89) and Arunachal Pradesh (1.01) have the lowest number. A total number of hospitals, nursing homes, and police stations and the numbers per 100000 people have been provided in Table IV. Correlation between a number of hospitals, police stations and the percentage of death is presented in Figure V.

Table IV. Number of Hospitals and Police Stations

State/UT	No. of Hospitals	Hospitals /100000 People	No. of Police Stations	Police Stations /100000 People
Andhra Pradesh	6065	12.28	1414	2.86
Arunachal Pradesh	57	4.12	14	1.01
Assam	2610	8.37	787	2.52
Bihar	5774	5.56	1284	1.24
Chhattisgarh	3174	12.43	591	2.31
Goa	197	13.51	67	4.60
Gujarat	5956	9.86	1348	2.23
Haryana	2193	8.65	878	3.46
Himachal Pradesh	1369	19.94	337	4.91
Jammu & Kashmir	62	0.49	17	0.14
Jharkhand	2707	8.21	628	1.90
Karnataka	8280	13.54	1624	2.66
Kerala	2612	7.82	1052	3.15
Madhya Pradesh	6927	9.54	1684	2.32
Maharashtra	11533	10.26	2338	2.08
Manipur	388	14.26	147	5.40
Meghalaya	269	9.08	115	3.88
Nagaland	203	10.25	122	6.16
Odisha	3898	9.29	1007	2.40
Rajasthan	10098	14.72	1815	2.64
Sikkim	81	13.33	13	2.14
Tamil Nadu	7067	9.80	2451	3.40
Telangana	3988	11.30	1246	3.53
Tripura	326	8.88	120	3.27
Uttarakhand	1483	14.66	398	3.93
Uttar Pradesh	9990	5.00	3826	1.91
West Bengal	5162	5.65	812	0.89

From the above results, it is very clear that the number of hospitals and police stations has not that much impact on the fatality rates of road accidents. Lack of communication

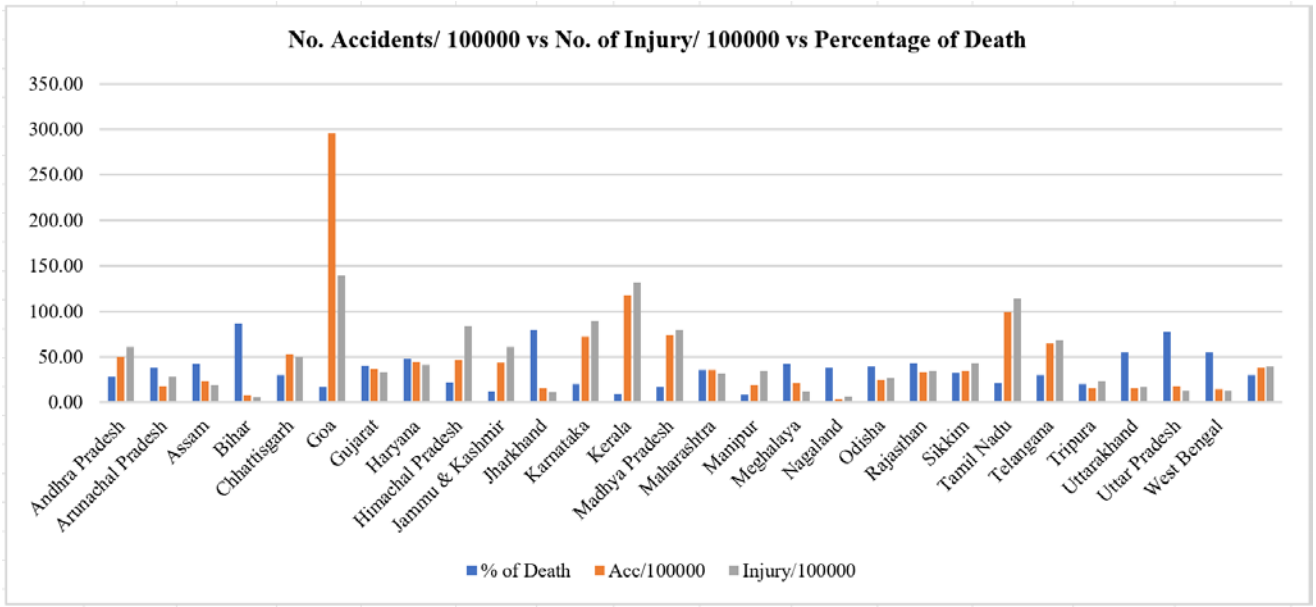

Figure IV. Number of Accidents/ 100000 people vs Number of Injury/ 100000 people vs Percentage of Death

infrastructure is one of the major causes for that. Data shows that there are plenty of hospitals available in most of the states but the fatality rate for some of those states is surprisingly high.

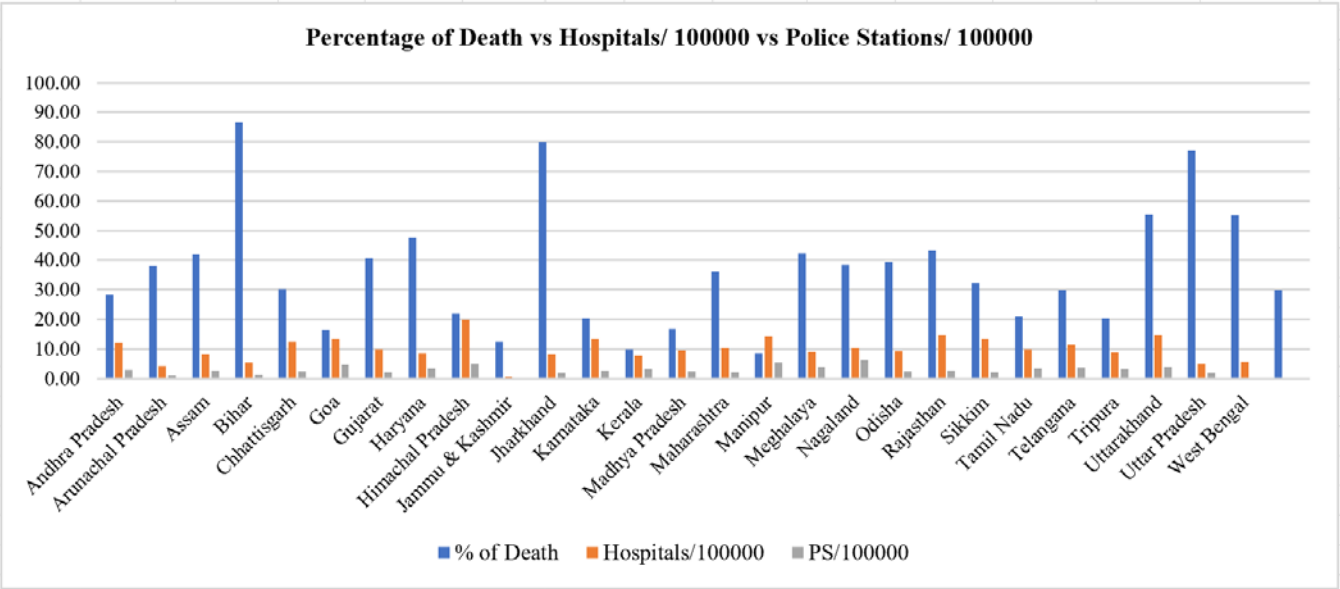

Figure V. Percentage of Death vs Hospitals/ 100000 people vs Police Stations/ 100000 people

V. CONCLUSION

Every day around 413 persons are killed from 1317 road accidents in India. Following traffic rules and regulations, road safety guidelines must reduce the number of road accidents. But after the occurrence of any crash, immediate rescue is the only option to reduce some deaths from the injuries. Data clearly shows that the number of hospitals and police stations is not the only factor to reduce accidental deaths. Immediate reporting of those accidents to the nearest hospitals and police station can be a vital factor to minimize accidental deaths.

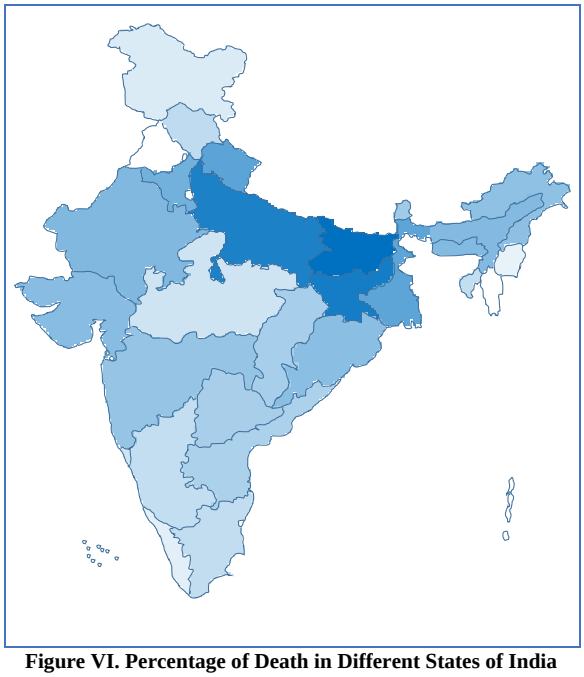

Figure VI. Percentage of Death in Different States of India

Current research provides some important facts about the status of emergency rescue services in different states of India that could be helpful to strategize and improve their rescue services. The database of all the possible rescue centers in India can be a precious part for developing a smart, reliable and fully automatic accident reporting and monitoring system.

In future we'll try to extend this research with more localize data e.g. district or city wise and use the database to develop a centralized accident reporting system.

VI. ACKNOWLEDGEMENTS

Authors would like to acknowledge the Commonwealth Service Commission, UK for providing the grant under Split-site PhD Fellowship scheme. Authors also appreciate the technical help provided by Arnab Nath and Akash Banerjee, IIEST, Shibpur, India

REFERENCES

[1] Press Information Bureau, Government of India, Ministry of Road Transport & Highways, http://pib.nic.in/newsite/PrintRelease.aspx?relid=170577, Last accessed on 31st August, 2018.

[2] "Road Accidents in India – 2016", Government of India Ministry of Road Transport & Highways, Transport Research Wing, "http://pibphoto.nic.in/documents/rlink/2017/sep/i20179601.pptx", Last accessed on 31st August, 2018.

[3] "Statistics of Road Accidents in India From 2013 to 2016", Open Government Data (OGD) Platform India, Govt. of India, "https://data.gov.in/resources/statistics-road-accidents-india-2013-2016", Last accessed on 31st August, 2018.

[4] S. Koley and P. Ghosal, "An IoT Enabled Real-Time Communication and Location Tracking System for Vehicular Emergency," in 2017 IEEE Computer Society Annual Symposium on VLSI, 2017, pp. 671–676.

[5] G. Singh, S. N. Sachdeva and M. Pal, “M5 model tree based predictive modeling of road accidents on non-urban sections of highways in India”, Accident Analysis and Prevention, Volume 96, November 2016, Pages 108-117.

[6] S. Kumar and D. Toshniwal, “Analysing road accident data using association rule mining”, 2015 International Conference on Computing, Communication and Security (ICCCS), 2015.

[7] B. Jones, L. Janssen, and F. Mannering, “Analysis of the Frequency and Duration of Freeway Accidents in Seattle”, Accident Analysis and Prevention, Elsevier, vol. 23, 1991.

[8] S. P. Miaou, and H. Lum, “Modeling Vehicle Accidents and Highway Geometric Design Relationships”, Accident Analysis and Prevention, Elsevier, vol. 25, 1993.

[9] J. Maleki, E. Foroutan, and M. Rajabi, “Intelligent Alarm System for Road Collision,” Journal of Earth Science and Engineering, vol. 1, no. 3, 2011.

[10] C. Thompson, J. White, B. Dougherty, A. Albright, and D. C. Schmidt, “Using Smartphones and Wireless Mobile Sensor Networks to Detect Car Accidents and Provide Situational Awareness to Emergency Responders,” in ICST Conf., June, 2010.

[11] Open Government Data (OGD) Platform India, “https://data.gov.in/”, Last accessed on 31st August, 2018.

[12] Office of the Registrar General & Census Commissioner, India, Ministry of Home Affairs, Government of India, “http://www.censusindia.gov.in/”, Last accessed on 31st August, 2018.

[13] “Place Searches”, Google Maps Platform, “https://developers.google.com/places/web-service/search”, Last accessed on 23rd August, 2018.

[14] “Radar Search”, Google Maps Platform, “https://developers.google.com/maps/documentation/javascript/examples/place-radar-search”, Last accessed on 11th June, 2018.

[15] “Places API”, Google Maps Platform, “https://developers.google.com/places/web-service/get-api-key”, Last accessed on 23rd August, 2018.

[16] “Reverse Geocoding (Address Lookup)”, Google Maps Platform, “https://developers.google.com/maps/documentation/geocoding/intro”, Last accessed on 23rd August, 2018.

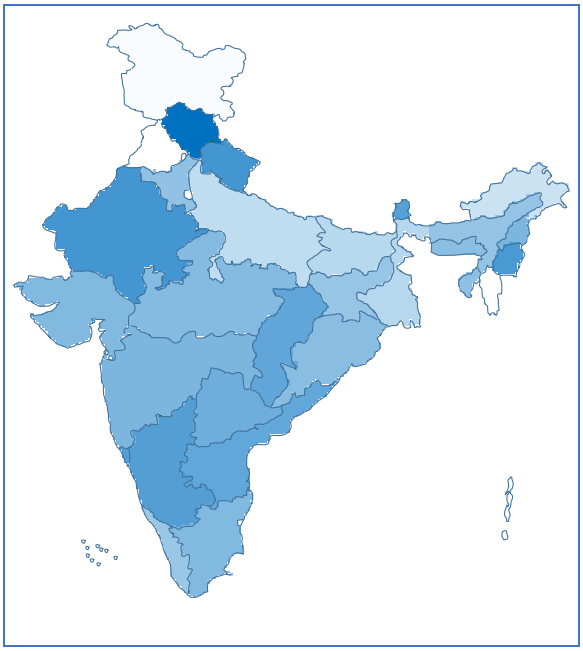

Figure VII. Hospitals in Different States of India per 100000 People

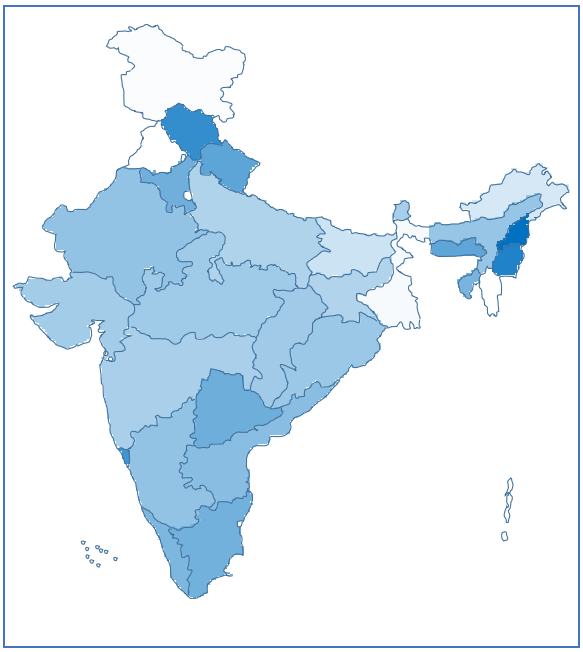

Figure VIII. Police Stations in Different States of India per 100000 People