COMMUNICABLE AND NON -COMMUNICABLE DISEASES IN INDIA: A COMPARATIVE ANALYSIS

Ajit Kumar[1], Dr. Harsh Pratap Singh[2]

M. Tech. Scholar, Department of Computer Science and Engineering, SSSUTMS, Sehore, M.P., India[1]
Associate Professor, Department of Computer Science and Engineering, SSSUTMS, Sehore, M.P., India

Abstract

Communicable Disease (CD) are the fundamental reason of death around the globe for years following by non-communicable disease (NCD) causing significant issues in health care industry. Among all NCDs and CDs, Cancers (NCD) and Tuberculosis (CD) constitute the real reason for dreariness and mortality in creating nations, including India. The particular point of this investigation is to evaluate the commonness of communicable (Tuberculosis) and non communicable disease (Cancers) in urban and provincial territories of Katra and to dissect the relative scourge utilizing a comparable framework. The information was broke down measurably utilizing Chi-Square test and Odds proportion with 95% certainty interim. All investigations were performed with MATLAB 2012a programming. The finding recommends that there is an expanding pervasiveness of Communicable and non communicable diseases in India because of way of life changes and urbanization. These are the difficulties that are to be handled in new thousand year.

Keywords: *Communicable diseases, Non-communicable diseases, Rural, Urban.*

1. Introduction

India is one of the well known nations around the globe for social exercises. Infections can be comprehensively arranged into two sorts: Communicable and Non Communicable. Communicable Disease (CD) were the primary driver of death around the globe for quite a long time took after by non-communicable disease (NCD) causing real issues in industrialized nations. Communicable disease that is transmitted through direct contact with a tainted individual or by implication through a vector additionally called contagious disease while non-communicable disease (NCD) is a medical condition or disease that isn't caused by irresistible operators (non-irresistible or non-communicable). NCDs can allude to interminable infections which keep going for drawn out stretches of time and advance gradually. In some cases, NCDs result in fast passings, for example, found in specific disease, for example, immune system sicknesses, heart ailments, stroke, tumors, diabetes, endless kidney ailment, osteoporosis, Alzheimer's ailment, waterfalls, and others. While now and again alluded to as synonymous with "constant ailments", The spread of non-communicable disease(NCDs) primarily coronary illness, stroke, diabetes, tumors, and endless respiratory ailment speaks to a worldwide emergency; in all nations and in all salary gatherings, men, ladies, will represent 80% of the worldwide weight of ailment, causing seven out of each ten passing's in creating countries, contrasted and not as much as half today [1] .Globally, around 57 million individuals kicked the bucket and kids are in danger of these sicknesses [2]. Pestilences of non-transmittable maladies (NCD) are directly rising or quickening in most creating nations [3]. In 2008, 33 million (58%) of the passing's were because of perpetual (non-transmittable) maladies (chiefly cardiovascular infection, diabetes, growth, and endless respiratory ailments) [4]. In 2004, 4.8 million (59.4 percent) of the evaluated 8.1 million Indian passings were because of NCDs [5]. Communicable Disease (CDs) like tuberculosis (TB), cholera, meningitis, hepatitis, intestinal sickness, dengue, yellow fever, AIDS, Ebola, SARS and others in parallel likewise keep on being the real reason for mortality in creating nations [6]. Indeed, even Asian nations like India have a noteworthy general wellbeing test of developing extent of non communicable and communicable disease in the present century. In India, among all NCDs and CDs, Cancers (NCD) [7] and Tuberculosis (CD) constitute the significant reason for bleakness and mortality. In this paper, we utilize the term interminable non-communicable disease to allude to major ceaseless issue, for example, Cancer and communicable disease to allude Tuberculosis; different scatters are not shrouded in this paper. A correlation was made of the age-particular frequency rates in Katra area at 2-year interims from 2015 to 2016

2. Data Mining Techniques

Data mining technique is defined as the process of discovering interesting patterns and knowledge from the large amounts of data. Technique refers to analyzing data from different viewpoints and abstracting it to get the necessary information. DM technique provides an important means for extracting valuable medical rules hidden in medical data and acts as an important role in clinical diagnosis.

2.1 Classification

Classification is the process of predicting output based on some given input data. The goal of classification is to accurately predict the target class for each case in the data. In order to predict the data, it processes the training set and predictive set. It first develop relationships between the attributes of training data set .Then it is provided with the predictive data set, which contains similar attributes but with different data values, Then it analyze the given data and produce prediction by placing the different data sets in different classes based on the relationship of attributes [9][10]. Classification uses predictive rules expressed in the form of IF-THEN rules where the first part (IF part) consist of conjunction of conditions and the second part(THEN part) predict a certain prediction attribute value that satisfy the first part.

2.1.1 Decision Tree

Decision tree is similar to flow chart in which every non-leaf node denote a test on a particular attribute and every branch represent a outcome of the test. Root node is the topmost node in the decision tree. For example, with the help of readmission tree, we can decide whether a patient needs to be readmitted or not. Using Decision Tree, a decision maker can choose best alternative and traversal from root to leaf indicates unique class separation based on maximum information gain [11]. Decision tree are self explanatory and easy to follow. Set of rules can also be constructed with the help of decision tree. Decision Tree can be considered as nonparametric method because there is no need to make assumptions regarding distribution of space and structure of classifier. Decision tree have several disadvantages. These are: Most of the algorithm like ID# and C4.5 require target attributes to have discrete values as decision tree use divide and conquer strategy. More the complex relationship among attributes lesser is the performance.

2.1.2 Support Vector Machines

Vladimir Vapnik first introduced idea of Support Vector Machine [12]. Its accuracy is better than all other available techniques. It was first introduced for binary classification problems; but it can be further extended to multi class problems. It creates hyper-planes to separate data points.
It can be implemented in 2 ways:
1. Mathematical programming
2. Using kernel functions

With the help of training data sets, non linear functions can be easily mapped to high dimensional space. This can only be possible using kernel functions like Gaussian, sigmoid etc.

2.1.3 Artificial Neural Network

Artificial neural networks models have been studied for many years in the hope of achieving human like performance in several fields. In Neural Networks, basic elements are neurons or nodes. These neurons are interconnected and within the network they worked together in parallel in order to produce the output functions. From existing observations they are capable to produce new observations even in those situations where some neurons or nodes within the network fails or go down due to their capability of working in parallel. An activation number is associated to each neuron and a weight is assigned to each edge within a neural network. In order to perform the tasks of classification and pattern recognition neural network is mainly used [13]. ANN is based on the biological neural networks in the human brain and described as a connectionist model.
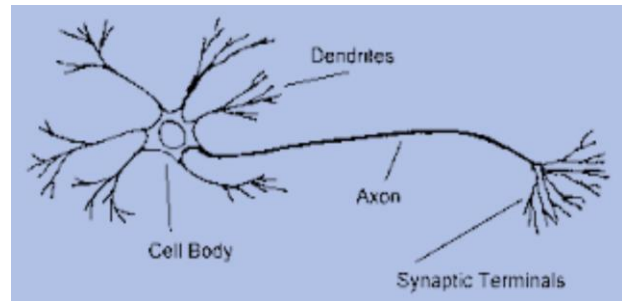

Fig.1 Biological Neuron

It is based on the neuron, a cell that processes information in the human brain [14]. The neuron cell body contains the nucleus, and has two types of branches, the axon and the dendrites. The axon transmits signals or impulses to other neurons while the dendrites receive incoming signals or impulses from other neurons. Every neuron is connected and communicates through the short trains of pulses [16]. The nodes are the artificial neuron and the directed edges represented the connection between output neurons and the input neurons. In training phase, the internal weights of the neural network are adjusted according to the transactions used in the learning process. For each training transaction the neural network receives in addition the expected output. This allows modification of weight.
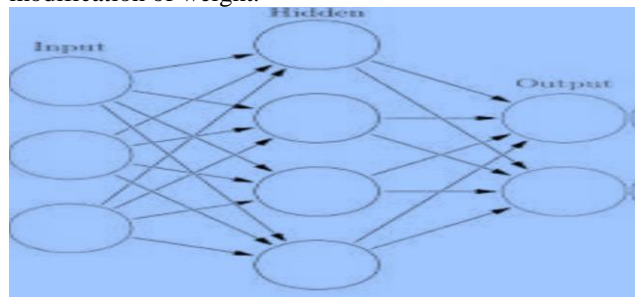

Fig.3 Artificial Neural Networks

2.2 ARM (Association Rule Mining)

The aim of ARM is to identify the useful rules from the large amounts of data. Association rule mining has following logical process and appeal.

- Logical process: Interesting rules are determined in terms of support and confidence.

1) Support: It reflects the usefulness of discovered rules.

2) Confidence: It reflects certainty of discovered rules.

- Appeal: Association rules are considered as interesting if they satisfy both minimum support threshold and minimum confidence threshold.

These thresholds set by users or domain experts. Additional analysis can be performed to discover interesting statistical correlations between associated items.

2.3 Probabilistic Learning Method (Bayesian Classifier)

For probabilistic learning method Bayesian classification is used. Bayes theorem of statistics plays a very important role in it. While in medical domain attributes such as patient symptoms and their health state are correlated with each other but Naïve Bayes Classifier assumes that all attributes are independent with each other. This is the major disadvantage with Naïve Bayes Classifier. If attributes are independent with each other then Naïve Bayesian classifier has shown great performance in terms of accuracy. There are two types of probabilities

- Posterior Probability [P(H/X)]
- Prior Probability [P(H)]

Where X is data tuple and H is some hypothesis. According to Bayes' Theorem,

$$P(H/X)= P(X/H)P(H) / P(X)$$

Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

2.4 k-Nearest Neighbors

In pattern recognition, the k-Nearest Neighbors algorithm (or k-NN for short) is a non – parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature spaces. The output depends on whether k-NN is used for classification or regression: In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. K-NN has a number of applications in different areas such as health datasets,

image field, cluster analysis, pattern recognition, online marketing etc. There are various advantages of KNN classifiers.[15] These are: ease, efficacy, intuitiveness and competitive classification performance in many domains. If the training data is large then it is effective and it is robust to noisy training data. A main disadvantage of KNN classifiers is the large memory requirement needed to store the whole sample. If there is a big sample then its response time on a sequential computer will also large.[16]
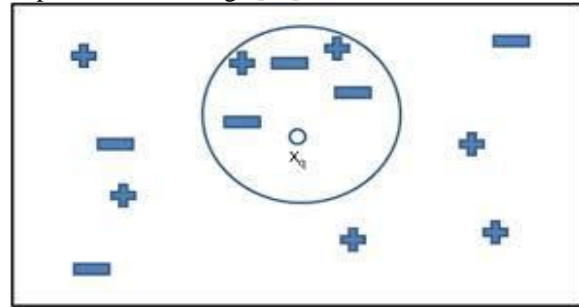
3. Materials and Methods

All the recently analyzed cases in the years 2019 and 2020 were incorporated as the examination cases. Among 3000 patients' biopsy information in the lab, just cases having Cancers (CD) and Tuberculosis (NCD) were incorporated into this examination. The information was a clinical examination information containing Biops number, Disease types, age, sex, Locality, Height, Mark segment, cut area, Histopathological examination and Microscopic examination. Among these factors just illness sort, age, sex, area was considered for the investigation. The reference date for recording data was the date of analysis for cases.

Statistical Methods

The data was analyzed statistically using Chi-Square test and Odds ratio with 95% confidence interval. All analyses were performed with MATLAB version 2012a software.

4. Results and Discussion

Non-communicable diseases

Among the different NCDs like cardiovascular disease, Diabetes and Cancers, just diseases were utilized for this investigation. Disease that are utilized as a part of the study incorporates Breast, Cervical, Colon, Bladder, Lung, Oral, Prostrate, Rectal, Skin, Uterine and Stomach Cancer. Among 3000 patients biopsied in the year 2019 to 2020, 303 patients were seen influenced with these malignancies. The country region (52.1% Vs 47.9%) has expanded danger of NCDs: patients with Breast Cancer (33%), Cervical Cancer (35%), Oral Cancer (8.9%), Prostrate Cancer (6.3%), Skin Cancer (4%) and Other

Cancers (12.9%) like Bladder, Colon, Lung, Rectal, Stomach, Uterine (P=0.07, Table 1). The pervasiveness of the non communicable disease in urban and rural area is diagrammatically spoken to in the graph (Chart 1) given below.

Table 1: Non communicable diseases in urban and rural areas of Katra (2015-2016)

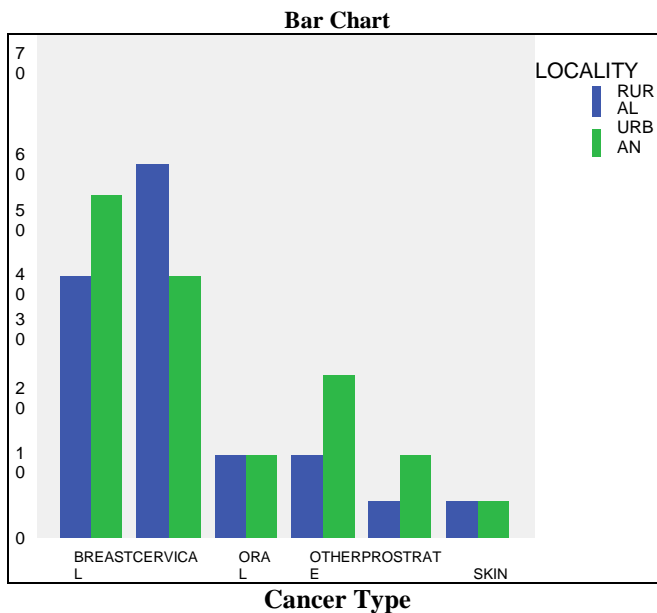| Cancer Type	Patients, n (%)		df	χ2	P
	Rural	**Urban**			
Breast	43 (14.19)	57 (18.81)	5	10.1	0.07
Cervical	62 (21.7)	44 (15.4)			
Oral	14 (1.24)	13 (1.15)			
Prostrate	6(0.77)	13 (1.67)			
Skin	6(0.24)	6 (0.24)			
Other	14 (1.8)	25 (3.2)			

Bar Chart

Chart 1: Prevalence of the non-communicable diseases in urban and rural areas

Non-communicable diseases
Among 3000 patients biopsied in our laboratory 35 patients were seen affected with Tuberculosis. There was a significant urban-rural difference in having Tuberculosis in the last 2 years (71.4% vs 28.6%, respectively; *P* < 0.001) (Table 2).

Table 2: Tuberculosis in Urban- rural communities (2019-2020)

Cancer Type	Tuberculosis Patients, n (%)		Difference (95% CI)	df	t	P
	Rural	**Urban**				
Age	10 (28.6)	25 (71.4)	30.743(25.58to 35.91)	34	12.102	<0.001

Comparison of communicable and non communicable diseases
There was a considerable urban-rural difference in burden to Communicable and non-communicable diseases in the years 2015-2016 (54.1% vs 45.9%, respectively; *P* = 0.03) (Table 3)

5. Conclusion
Our discoveries propose that the observed differences amongst urban and rural ladies could be generously decreased by changing the way of life or legitimate discovery in the beginning times. It is likewise vital to instruct general society and human services experts in rural and urban regions with a specific end goal to advance early identification. Generally, our study concludes that people residential of Katra are more presented to both communicable and non communicable disease contrasted with those from 10 towns encompassing Katra. This research paper also discusses some data mining techniques which help us to early diagnosis of both the communicable and non-communicable disease. These data mining techniques can be most effective for diagnosis of these diseases by ensemble the two or more approach. So in future work, uses the hybrid approach for diagnosis of the disease by using more attributes.

References
[1] A. Boutayeb, The double burden of communicable and non-communicable diseases in developing countries, Transactions of the Royal Society of Tropical Medicine and Hygiene., 2006, 100, 191.
[2] Ala Alwan. Global status report on noncommunicable diseases 2010. World Health Organization, 2011.
[3] Murray CJL, Lopez AD. Global Health Statistics: Global Burden of Disease and Injury Series. Volumes I and II. Boston: Harvard School of Public Health, 1996.
[4] A. Alwan, D. R. MacLean, L. M. Riley, E. Tursan d'Espaignet, C. D. Mathers, G. A. Stevens, D. Bettcher. Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. Lancet., 2010, 376: 1861.
[5] A. Mahal, A. Karan, M. Engelgau. Existing Evidence on the Economic Impact of NCDs in India and its Limitations. HNP Discussion paper., 2010.

[6] World Health Organization. Cholera: the prototype "global" disease. In: Global defence against the infectious disease threat. Geneva: WHO, 2003 (WHO/CDS/2003.15).

[7] T. Seth, A. Kotwal, R. Thakur, P. Singh, V. Kochupillai. Common cancers in India: knowledge, attitudes and behaviours of urban slum dwellers in New Delhi. Public Health., 2005, 119: 87.

[8] M. S. Pednekar, P. C. Gupta. Prospective study of smoking and tuberculosis in India. Preventive Medicine, 2007, 44: 496.

[9] S. Palaniappan, and Rafiah Awang "Intelligent heart disease prediction system using data mining techniques." IEEE Conference on Computer Systems and Applications, 2008.

[10] K. B. Srinivas, Kavihta Rani, and A. Govrdhan "Applications of data mining techniques in healthcare and prediction of heart attacks." International Journal on Computer Science and Engineering (IJCSE) Vol. 2, No. 02, 2010, pp. 250-255.

[11] Parvez Ahmed, Saqib Qamar, and Syed Qasim Afser Rizvi. "Techniques of Data Mining In Healthcare: A Review." International Journal of Computer Applications Vol. 120, No.15, 2015.

[12] Vladimir Vapnik. "The support vector method of function estimation."Non-linear Modeling. Springer US, 1998, pp.55-85.

[13] M. H. Dunham, "Data mining introductory and advanced topics", Upper Saddle River, NJ: Pearson Education, Inc., (2003).

[14] A. K. Jain "Artificial neural network : a tutorial[Online].

[15] Bramer, M "Principles of data mining", Springer 2007

[16] Alpaydin, E. "Voting over Multiple Condensed Nearest Neighbors. Artificial Intelligence Review", p. 115–132. 1997.

[17] Farrukh Aslam Khan et al. "Detection and Prediction of Diabetes using Data Mining: A Comprehensive Review", IEEE ACCESS 2019.

[18] Batuhan Bakırarar et al. "Prediction of new prescription requirements for diabetes patients using big data technologies", Journal of Health Research Vol. 36 No. 2, 2022 pp. 334-344 Emerald Publishing Limited, 2022