

TO STUDY THE LIKELIHOOD OF CANCER BASED ON TODAY'S LIFESTYLE -A SURVEY

Chetna
Assistant Professor,
CSE, Chandigarh Engineering
College
Landran

Abstract— Purpose – The objective of this paper is to review the impact of lifestyle which include habits and exposure of the ethnic population to any chemical compounds which leads to cancer.

Design/methodology/approach – The review has been done to study the impact of lifestyle on the occurrence of cancer. Different authors use different technique. Urbanization, increased life expectation and embracing western lifestyles activate the occurrence of cancer disease. Sometimes geographically conditions also propel the occurrence. So many researchers tried to recognize these risk factors so that it can be avoided to decrease the cancer mortality rate to minimum extent. Early detection of this disease is also crucial for the survival.

Findings – The results obtained show the impact of smoking, diet and exposure to chemical substance on the occurrence of cancer. The starting age of smoking and the quantity of cigarette smoke each day also has a huge impact on how fast a person gets this deadly disease. In some cases, results also show the statistical difference between the ethnic group health status with that of control group. By appropriately detecting the risk factors using data mining tools, these risks can be avoided or at least minimized to avoid further expansion.

Originality/value – The findings are to detect the cancer with different lifestyle and reach to conclusions. Further it can be extended to use latest data mining tools to detect the actual cause of cancer with much accuracy.

Keywords— Data mining, control group, tandoor, cooking bread, taluk

I. INTRODUCTION

As Countries tend to become more industrialized and people become more aged, the frequencies of occurring diseases generally increase. The disease which occurs because of this reason is called lifestyle diseases or non-communicable diseases. These diseases include heart disease, hypertension, cancer, obesity, type II diabetes, mental disorders and many other diseases (Meng *et al.*, 2013; Amin *et al.*, 2013). They are different from other infectious diseases which are dispersive or contagious in nature which are called communicable diseases. In accordance with the World Health Organization(WHO)

also, the increased wave of spreading non communicable diseases in developing as well as developed countries is because of dietary and lifestyle changes(WHO,2003).

Cancer is a potentially critical ailment which is the resultant of abnormal cell functioning which causes extensive masses of anomalous cells whose escalation raze normal tissues and can extend to crucial organs followed by even patient's death. It is still challenging for the doctor's to detect cancer in the field of medicine. Detecting cancer in earlier stage is still difficult although various tests are available for predicting this disease. If it is detected at an earlier stage, this disease is curable. In medical field where data motivated analytical research is a big hit, new research methods are always recognized to further move ahead in this domain too. Difference in demography and lifestyle inequalities is much more clear in occurring diseases especially cancer cases.

Although Cancer was earlier treated clinically and biologically but during recent years analytical studies have become more common. There is good progress in diagnosing and treatment of cancer disease, still it poses a considerable menace to our society. This is the most frequent reasons for deaths after heart diseases in the whole world (Jemal *et al.*,2007). Nearly 23% deaths in USA and 7% deaths in India are because of cancer only. The expected world's population will be 7.5 billion by 2020 in which prediction is that, 15.0 million will be diagnosed to be cancer cases (Brayand *et al.*,2006) . In India these number will be approximately about 2.5 million in which 8, 00,000 will be fresh cases and 5,50,000 will be the number of death cases per annum . With the survey of historical data it has been found that the figure of cancer cases has drastically expanded over the time and is continue to do so with 96.4% for males and 88.2% for females out of 100,000 cases will be analyzed by the end of the last century (Rao *et al.*,1998; Nandakumar *et al.*,2001;Dinshaw *et al.*, 1999).

Investigation was done on 70% of cancer cases during the last decade and was successful with existence of only some

patients. In the Indian sub-continent also, the degree of cancer patient is increasing due to impartial living standards and lack of proper medical care (Ali *et al.*,2001). The risk of developing cancer among people is due to variation in the ecological factors (Kotnis *et al.*,2005) such as smoking cigarette (Wynder *et al.*,1974) or following some particular dietary patterns. For example people in Australia highly suffer from skin cancer because of their high vulnerability to the sun and due to the ease of availability of tobacco and majority of the persons are having the habit of chewing it, people living in India develops oral cancer 25% times the average rate. In fact the India has the maximum number of cases nearly about 75,000 to 80,000 new cases added every year. Due to the highest consumption of raw fish, maximum number of stomach cancer patients are found in Japan (Hanan Farouk Aly,2012).

As human being come from past era to present one, there is great increase in the amount of data and to control that rapidly increasing data is even much harder. When the need to extract the amount of valuable knowledge from the ocean of data arises then data mining (Moro *et al.*,2016; Chen *et al.*,2016) come into being . Data Mining can be defined as analyzing data from different perspectives and realistic predictions be made by summarizing it (Goebel *et al.*,1999; Wu *et al.*2002). So it can be said that data mining helps to find relationship among different variables in a very big database (Fouad *et al.*,2015). The objective of this study was to review the work done by different researchers in the field of cancer according to the lifestyle of the patients.

II. CANCER TYPES

Cancer is classified on the basis of the type of cells in which they build up. Cancer which appears from epithelial cells is called carcinomas that form the exterior layer of some internal organs of the skin. Carcinomas cancer is most prevalent type (Figure 1). Other type of cancer is Leukemias that affects the blood and blood forming organs such as the lymphatic system, bone marrow and the spleen. This cancer affects the resistant power of our body. Sarcoma is also one of the kind of cancer which grow in connective tissues. Connective tissues are those that connect or support other kinds of tissues in the body.

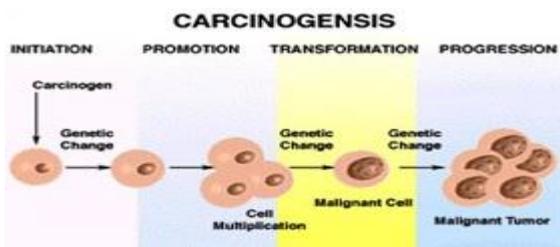


Figure 1: site: www.medscape.com source: cancer control @2007

A. Cancer Growth and its Spread

When cells grow out of control in particular part of the body then cancer develops. Body cells normally nurture, split and die in a sequence manner but in cancer, cells nurture and split without perishing. Instead they replace normal cells and build cancer cells in an abnormal way. Cancer cells keep on growing, form lump and make an added mass of tissue. Exceptions are in leukemia cells where no lumps are formed because here blood forming organs are involved. Carcinomas cancer cells' lump is called malignant tumor and when it grows, it damages nearby tissues too.

Cancer can start in any component of the body and extend to others. Metastasis is a term used to be given when tumor spread to a new site. This is the result when cancer cells blasts out from the tumor and travel through blood to other parts of the human body. In the new area also, the cancer cells continue to expand beyond limits and build up into a malignant tumor.

III. CAUSES OF CANCER

The exact reason of occurring the cancer is yet unknown. Most cancer results from irreversible harm to genes or from variations which occur from internal factors like metabolism, hormones, immune system and the digestion of nutrients within the cells or exposure to ecological factors (Kotnis *et al.*,2005) or external factors. An ecological agent or a chemical that yields cancer is called a carcinogen.

In United States 75 to 80 percent cancer cases are due to environmental factors which include diet, tobacco use, infectious diseases, radiations and chemicals. Lung, mouth, throat, larynx, bladder, kidney, esophagus and pancreas cancer are all caused by intake of cigarettes, cigar, chewing tobacco and snuff (Ahmed *et al.*,2013; Krishnaiah *et al.*,2013). One third of all cancer cases in the whole world are because of smoking alone. Consumption of alcohol beyond the recommended level also enhances the risk of evolving cancer of the mouth, liver, pharynx, larynx, breast and esophagus (Einsele *et al.*, 2015; Senturk *et al.*,2014). Obesity and excess fats have also an association with the enhanced risk of cancers of the breast, endometrium, kidney, colon, esophagus and gallbladder. Coal tars and their derivatives, such as benzene; some hydrocarbons; aniline, a material used to create dyes; and asbestos are the chemicals which have been found to cause cancer: Ultraviolet light from the sun and radiation from other sources leads to skin cancer. Cancer can also be caused by infectious agents. One fifth of the cancer cases are due to chronic viral infections like liver cancer which is caused by hepatitis B Virus(HGB).The bacterium *Helicobacter pylori* have been associated to stomach cancer. The cancer cases which occur due to heredity amounts to about 5 to 10 percent where a faulty gene or damaged DNA inherits in

the next generation. Cancer of colon, kidney, ovary, rectum, esophagus, pancreas can occur hereditary also (Kaladhar *et al.*,2013). Figure 2 depicts the estimated incidence, mortality and 5 year prevalence of cancer for both genders in year 2012.

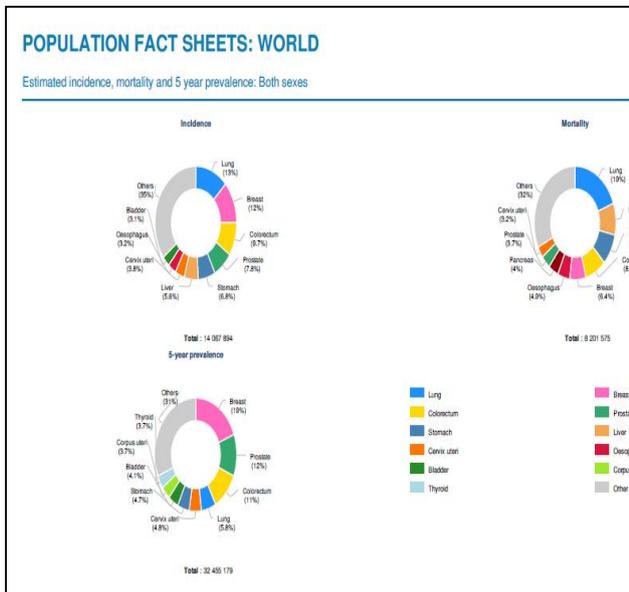


Figure 2 source: GLOOCAN 2012

Figure 3 is showing the incidence and mortality rate of lung cancer for both the genders worldwide.

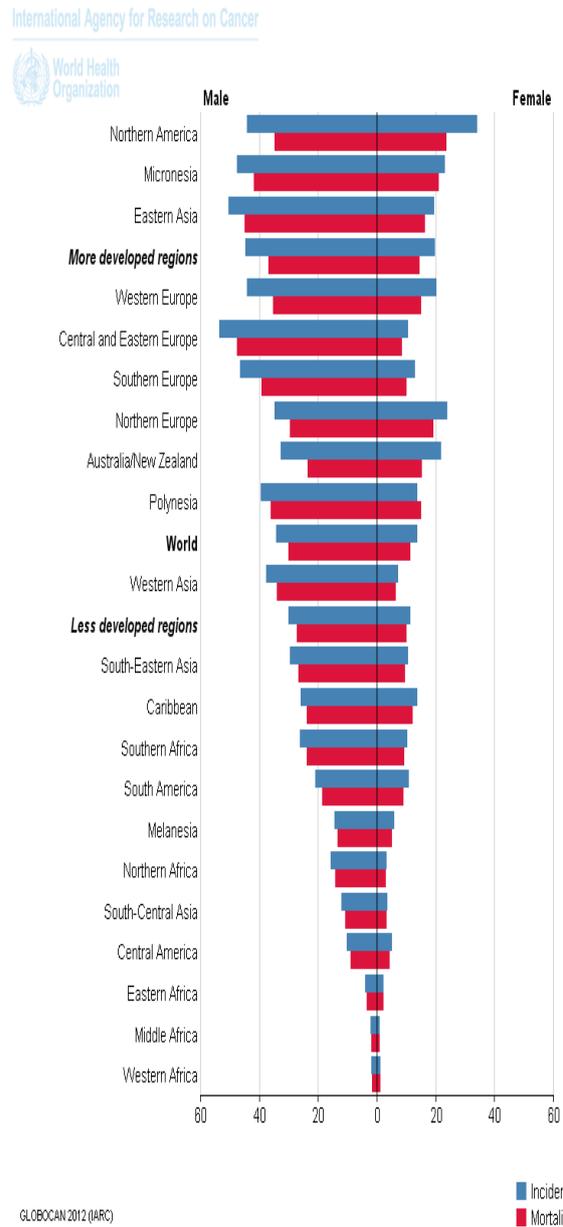


Figure 3: Standardized-Age Incidence Rates of Lung Cancer by Sex and World Area. Source: GLOBOCAN 2012.

Their survey results clearly states that the present status of smoking is now comparatively less among Mexican American Women. This is achieved by preventing smoking among this ethnic group, which results in decrease of lung cancer mortality rate in America among the target group.

Sebahattin Celik *et al.* (2014) in their paper “The Relationship between Eating and Lifestyle Habits and Cancer in Van Lake Region: Another Endemic Region for Esophageal and Gastric Cancers” intended to estimate the association between esophageal and gastric cancer which is generally seen in Van Lake region and their traditional consumption habits. They also created control group

having same gender, age and socio-economic features equivalent to patient group. Their study came to conclusion that Cheese consumption and smoking were higher in patient group. So heavy smoking, cooking bread on tandoor and high intake of herby cheese were found to be significant risk factors for gastric and esophageal cancers.

Table 3-shows the Comparison between patient and control groups in terms of herby cheese and tea consumption

Table 3

	Patient group Mean(%95- CI)*	Control group Mean(95%- CI)*	P
Amount of Herby Cheese(gram/day)	238.05(217.85-258.26)	85.00(68.71-101.29)	<0.001
Tea(cc/day)	929.38(841.49-1017.28)	1045.70(966.04-1125.36)	0.055

*Mean and 95% confidence interval

Table 3 above states that the patient group usually has higher intake of herby cheese whereas there is no significant difference exists as far as Tea is concerned. Figure 4 concludes the effect of smoking, tandoor exposure and herby cheese consumption per meal between the patient group and control group.

They investigated the features of lifestyle and eating habits of Van Lake region and reach to a conclusion that there exist a negative relationship between their habits and esophageal and gastric cancers and taking collective measures towards these diseases can only lower the frequency and mortality rates of this deadly disease.

Arnis Kirshners *et al.* (2015) in their paper “Gastric Cancer Risk Analysis in Unhealthy Habits Data with Classification Algorithms” initially took 136 attributes which after preprocessing reduced to finally 20 attributes and 742 records. Their experimental evaluation determined that C4.5 algorithm as the most accurate algorithm which proves that the infection risk which leads to H. pylori development is working environment and physical activity. The ethnic group which shows positive results worked in an environment where they come into direct contact with chemical compounds as well as heavy workload. Another factor is lack of physical activity which also leads to the development of these bacteria. Analysis of their results also shows that H. pylori infection is also promoted by other habits like: smoking, consuming alcohol, medical history and high intake of salt. Figure 5 shows analysis of the obtained results.

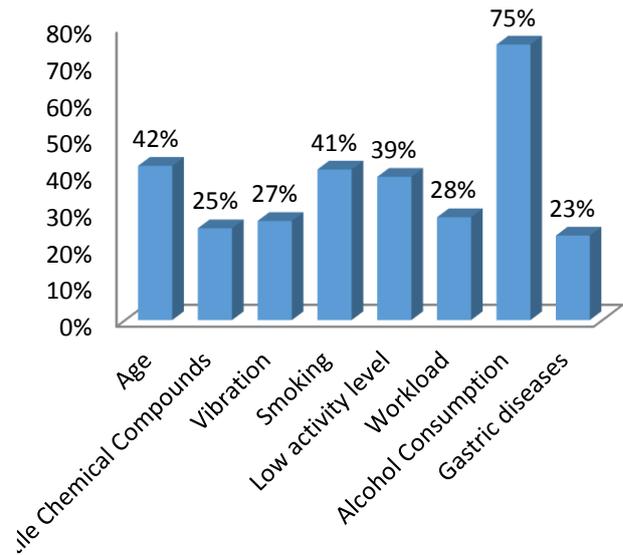


Figure 5: Analysis of the obtained results.

These statistics states that the maximum number of respondents are those who consume alcohol besides that the persons’ who face hazardous work and volatile compounds at work also be recommended for H. pylori test which could facilitate early detection and prevention of gastric diseases and cancer(Ramachandran *et al.*,2014).

The purpose of Thangaraju P *et al.* (2014) in their paper “Mining Cancer Data for Smokers and Non-Smokers by Using Data Mining Techniques” is to assess the risk factor of lung cancer and they concluded that the risk of lung cancer depends on for how long one smokes and how much number of cigarette the person smokes per day (Chen *et al.*,1996; Deoskar *et al.*,2013).

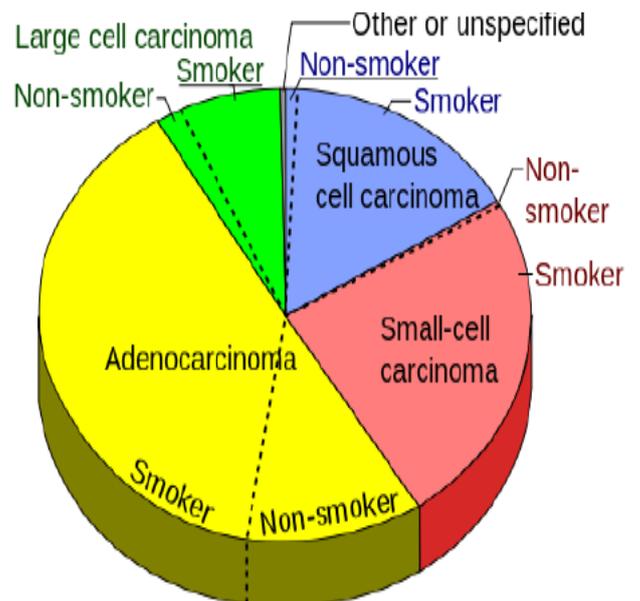


Figure 6: Categorization of lung cancer based on cell carcinoma

They classify the lung cancer into four categories and depicted the prevalence of cancer among smokers and non smokers (Figure 6).

Adenocarcinoma: This lung cancer is very common. Most cases are associated with smokers.

Squamous cell carcinoma: These cancers initiate in the flat cells that lie inside the airways of lungs. They are also most prevalent in smokers and are found near bronchus.

Large cell carcinoma: These cancers tend to grow rapidly and quite similar to small cell carcinoma.

Small cell carcinoma: These start in the bronchi and tend to spread widely.

They concluded that Naïve Bayes algorithm gave better performance with an accuracy of 83.4% as compared to other techniques and proved that smokers tend to have more risk of getting lung cancer as compared to non smokers.

Hanan Farouk Aly (2012) in his paper “Dietary habits and relation to cancer disease in different population” states that in China there is a great benefits of eating allium vegetables in protection against stomach cancer. Even if the total number of stomach cancer cases is reducing, but it remains the second important cause of cancer death throughout the world. Although the sample size is small but there are proofs that allium vegetables have shielding benefits against cancer in China. Relationship between consumption of allium vegetables and stomach cancer has been determined in their research. Variables were taken like education, number of years of smoking, salt intake, alcohol consumption; body mass index and intake of fresh fruits and vegetables and inverse relationship of onion consumption with cancer were observed. In Shanghai also, negative relationship was noticed between the regular consumption of onion or garlic and cancer. The same relationship was also observed in Qingdao between the risk of stomach cancer and garlic stalks. These results in their findings confirmed the defending benefits of allium vegetables (onions and garlic) against stomach cancer.

Harathi Parasur Babu *et al.* (2015) in their paper “Spatial data mining using association rules and fuzzy logic for autonomous exploration of geo-referenced cancer data in Western Tamilnadu, India” stated that data mining with association rules can be used in medical databases for recognizing the hidden disease and their causative agents. The database which they used is divided into taluks where each taluk is related with demography related cancer data and its respective lifestyle.

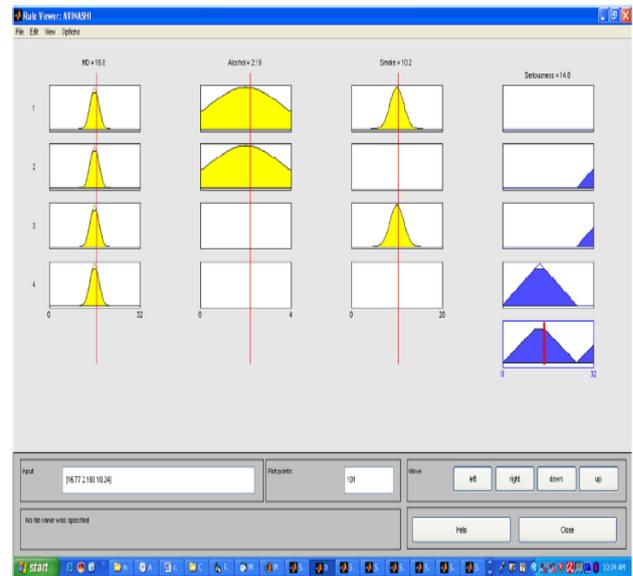


Figure 7: Different Lifestyle's Normal distribution curve and seriousness in Avinashi taluk

IV. DISCUSSION ON THE FINDINGS

The main objective of their study (Carrie D. Llewellyn *et al.*, 2003) was to analyze the major risk factors for oral cancer in public who are of age 45 years or younger. They had taken 116 cases of samples who were suffering from oral cancer in the United Kingdom. Although intake of tobacco and alcohol have been diagnosed earlier in the causation of oral cancer in older patients (Jeske *et al.*, 2016), Authors attempted their effort to take young patients as their sample. Their sample includes 63% of females and 79% of males who were either existing or ex-smokers. They concluded that there has to be long period of exposure for malignant transformation to occur in younger patients. Greater risk were evident from their findings for female smoking > 21 years are more prone to risk with (OR=2.5; 95% CI: 1.1—6.2) (Table 4).

An enhanced risk of (OR=1.7; 95% CI: 0.4—7.0) in female cannabis smokers (regulated for tobacco and alcohol habits) were also shown. The most regular use of illegal drug that prevails among both males and females in England was cannabis at that time. Regardless of these results, little proofs are there on the risk of cannabis smoking for the occurrence of oral cancer, although non-malignant oral situations have been proved to be more prevailing in cannabis smokers. Taking supari or Betel quid/ pan were also accountable to aggregate the risk with value of (OR=7.6; 95% CI: 0.8—7.4) (Table 4), although the numbers of persons reporting this practice were extremely less. A noticeable variation was found in more males for alcohol consumption who used to drink more than recommended amount of alcohol daily (Einsele *et al.*, 2015). However, enhanced (although non-significant) risks

were shown for women too who regularly used to drink alcohol before the age of 18 years.

The diet rich in fresh fruits and vegetables use to provide shield against oral cancer has been proved from the statistics too (Vidhu *et al.*, 2016). Similarly antioxidant diet which is derived from fresh fruits and vegetable too plays the same role against oral cancer. So the strongest dietary protection that comes from doctors too has been associated to eat plenty of fresh fruits and vegetables to fight against oral cancer which should be free from any tobacco, alcohol or social class. The authors' (Carrie D. Llewellyn *et al.*, 2003) findings too showed inverse relationship between the oral cancer risk and intake of fresh fruits and vegetables. Table 5 strengthens their findings too however there is more significant reduction in the female cases as compared to males. Women consuming good amount of fresh fruits and vegetables every day in the ten years prior to cancer diagnosis show significant reduction to malignant disease as compared to females who consumes two or less portions daily in their childhood. The statistics results show that childhood diet can be an important factor to study than diet in adulthood in younger patients.

So focus on two areas should be there to reduce morbidity and mortality of patients suffering from oral cancer, these areas are: primary avoidance by minimizing risk factors and early recognition. Generally people prefer visiting their physician rather than a dentist but physician are less expected to detect the patient for oral cancer. This is the moral duty of all the primary-care providers for counseling patients about habits that make them more prone to develop this type of cancer. Habits include tobacco use or excessive alcohol consumption, and if there is some suspicious oral lesion refer patients to an appropriate consultant for further treatment.

Oral cancer gets least amount of attention despite being the most accessible and visible part of human body. It accounts to 30 percent of mortality rate in India, adding to approximately 1 million per year. Because of the lack of understanding and alertness, oral cancers, as a group, are also not cured to as much in the post treatment or therapy phase. India is definitely not in protected region because of the easy access to tobacco or other narcotics (as per the Global Adult tobacco Survey) in the country; however, death rate because of oral cancer is preventable. Early detection, accurate treatment plan and its execution, followed by proper post treatment care, can help the person to restore his regular lifestyle in most cases.

V. CONCLUSION

Due to the data disparity, it is a complex task to extract useful knowledge from such collections. Data mining is therefore highly effective method to sort out the huge amount of data related problems by the researchers

(Goebel *et al.*, 1999; David *et al.*, 2013). Data mining using association rule combined with fuzzy logic is mostly functional in medicine especially in cancer treatments (Delen *et al.*, 2009; Priyanga *et al.*, 2013). During medical activities, large amount of data is generated and to get a comprehensive and systematic amount of data, data mining is highly used. To fulfill the requirements of a wider and deeper level of management, Data mining is the best solution.

In conclusion, this paper is to review the risk factors associated with lifestyle habits which leads to cancer. The World Health Organization (WHO) affirmed that as far as cancer is concerned at least one-third cases are preventable and if they are diagnosis early about 40% of them can be cured. Some of the most common risks are heavy smoking, alcohol consumption, low socioeconomic status, intake of heavy salty foods, consuming very hot drinks, high frequency of *HP* and indoor cooking bread on tandoor. Results also revealed that there is six times increase in the total cases if there will be no control on tobacco smoking and consumption of alcohol.

In the world, many persons are suffering from cancer. Most of the persons even do not know that they are suffering from this disease. After completely affected, this disease has no remedy. So the prediction of having the cancer at an early stage plays a significant role in the diagnosis process (Kharya *et al.*, 2012). If the prediction quality is good, it is believed that various data mining tools (Vital *et al.*, 2015; Naik *et al.*, 2016) are beneficial not only for the patients but also help doctors in making decision with patient consent. This is done by making them understand the risks involved in particular procedure of treatment depending upon each patient's condition.

There is a clear indication that the amount of cancer patients in India is also increasing every year. Various factors which are responsible for cancer needs to be eradicated. Depending on the region and severity of disease, several methods are recommended like surgery, chemotherapy or radiation therapy. These are given in solitude or in combination manner. Proper care which involves oncologist, clinical nutritionist and physiotherapists all are required for finest outcome. Cessation of smoking is an important part of treatment that may involve medication for tobacco de-addiction. India being growing towards developing to the developed country, one need to focus more on this issue. The awareness must be created among the public about this deadly disease and its prevention. Diet and living style must be taken care of by each and every Indian, to curb the spreading of this deadly disease. Effort is to hold back these above mentioned risk factors, in order to decrease the cancer mortality rate to minimum extent (Kaladhar *et al.*, 2011). One must discourage adolescents from adopting with noxious lifestyles and inappropriate habits which may lead to harmful results; instead they must be encouraged to

eat plenty of fresh fruits and green leafy vegetables (five to six portions per day).

REFERENCES

- Ahmed, K., Abdullah-Al-Emran, A. A. E., Jesmin, T., Mukti, R. F., Rahman, M., & Ahmed, F. (2013). "Early Detection of Lung Cancer Risk using Data Mining", *Asian pacific journal of cancer prevention*, vol.14,no.1, pp. 595-598.
- Ali, I., Wani, W. A., & Saleem, K. (2011). "Cancer Scenario in India with Future Perspectives", *Cancer therapy*, vol. 8, no. 1,pp. 56-70.
- Aly, H. F. (2012)." Dietary habits and relation to cancer disease in different population", *Archives in Cancer Research*, vol. 1, no. 1.
- Amin, S. U., Agarwal, K., & Beg, R. (2013). "Genetic Neural Network based Data Mining in Prediction of Heart Disease using Risk Factors" ,*Information & Communication Technologies (ICT)*,pp. 1227-1231.
- Babu, H. P., Selvaraj, J., Ramachandran, S., Marimuthu, P. D. and Somanathan, B. (2015). " Spatial Data Mining using Association Rules and Fuzzy Logic for Autonomous Exploration of Geo-referenced Cancer Data in Western Tamilnadu, India", *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 4,no. 1,pp. 1-11.
- Brayand, F., Moller B. (2006). "Predicting the Future Burden of Cancer", *Nat Rev Cancer* 6, pp. 63–74.
- Celik, S., Yilmaz, E. M., Özden, F., Kotan, C., & Okut, H. (2015)." The relationship between eating and lifestyle habits and cancer in Van Lake region: another endemic region for esophageal and gastric cancers", *Journal of cancer epidemiology*, vol. 2015, no. 254823
- Centers for Disease Control and Prevention. Preventing and controlling oral and pharyngeal cancer. Recommendations from a national strategic planning conference. MMWR 1998; 47 (No. RR-14): 1–12. External Resources Pubmed/Medline (NLM) Cambridge Scientific Abstracts (CSA).
- Cerquitelli, T., Chiusano, S. and Xiao, X. (2016). "Exploiting Clustering Algorithms in a Multiple-Level Fashion: A Comparative Study in the Medical Care Scenario", *Expert Systems with Applications*, vol. 55,pp. 297-312.
- Chen, L. F. and Tsai, C. T. (2016)." Data Mining Framework based on Rough Set Theory to Improve Location Selection Decisions: A Case Study of a Restaurant Chain", *Tourism Management*, vol. 53, pp. 197-206.
- Chen, M. S., Han, J., & Yu, P. S. (1996). "Data Mining: An Overview from a Database Perspective",*IEEE Transactions on Knowledge and data Engineering*, vol. 8, no. 6, pp. 866-883.
- David, S. K., Saeb, A. T. and Al Rubeaan, K. (2013). "Comparative Analysis of Data Mining Tools and Classification Techniques using WEKA in Medical Bioinformatics", *Computer Engineering and Intelligent Systems*, vol. 4, no. 13,pp. 28-38.
- Delen, D. (2009). "Analysis of cancer data: a data mining approach", *Expert Systems*, vol. 26, no. 1, pp. 100-112.
- Deoskar, P., Singh, D., & Singh, A. (2013). "Mining Lung Cancer Data and Other Diseases Data using Data Mining Techniques: A Survey", *International journal of computer engineering & technology (IJCET)*. Journal Impact Factor, vol. 4,no.2, pp. 508-516.
- Dinshaw, KA., Rao DN., Ganesh B. (1999).Tata Memorial Hospital Cancer Registry Annual Report, Mumbai, India.
- Einsele, F., Sadeghi, L., Ingold, R., & Jenzer, H. (2015). "A study about discovery of critical food consumption patterns linked with lifestyle diseases using data mining methods", *In Proceedings of the International Conference on Health Informatics*, Lisbon, Portugal, pp. 239-245.
- Fouad, M. M., Oweis, N. E., Gaber, T., Ahmed, M., and Snasel, V. (2015). "Data Mining and Fusion Techniques for WSNs as a Source of the Big Data", *Procedia Computer Science*, vol. 65, pp. 778-786.
- Goebel, M. and Gruenwald, L. (1999). "A Survey of Data Mining and Knowledge Discovery Software Tools ", *ACM SIGKDD Explorations Newsletter*, vol. 1, no. 1, pp. 20-33.
- Holck, S. E., Warren, C. W., Rochat, R. W., & Smith, J. C. (1982)." Lung cancer mortality and smoking habits: Mexican-American women", *American journal of public health*, vol. 72, no. 1,pp. 38-42.
- <http://www.who.int/cancer/prevention/en/>
- Jemal, A., Siegel R., Ward E., Murray T., Xu J. and Thun MJ. (2007). "Cancer Statistics", *CA Cancer J Clin* 57, pp. 43-66.
- Jeske, D. R., Longmate, J., Katheria, V. and Hurria, A. (2016). "Mining Branching Rules from Past Survey Data with an Illustration Using a Geriatric Assessment Survey for Older Adults with Cancer", *Algorithms Research*, vol. 9,no. 2, pp.33.
- Kaladhar, D. S. V. G. K., Chandana B. and P. B. Kumar(2011). "Predicting Cancer Survivability using Classification Algorithms", *LMT*, vol. 34,no. 65.7, pp. 96-106.
- Kaladhar,D. S. V. G. K., Pottumuthu, B. K., Rao, P. V. N., Vadlamudi,V., Chaitanya,A. K. and R. H. Reddy.(2013). "The Elements of Statistical Learning in Colon Cancer Datasets: Data Mining, Inference and Prediction", *Algorithms Research*, vol. 2, no. 1, pp.8-17.
- Kharya, S. (2012). "Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease."arXiv preprint arXiv:1205.1923.
- Kirshners, A., Polaka, I. and Aleksejeva, L. (2015). "Gastric Cancer Risk Analysis in Unhealthy Habits Data with Classification Algorithms". *Information Technology and Management Science*, vol. 18,no. 1, pp. 97-102.
- Kotnis, A., Sarin R and Mulherkar R. (2005). "Genotype, Phenotype and Cancer: Role of Low Penetrance Genes and Environment in Tumor Susceptibility", *J Biosci* 30, pp. 93-102.

- Krishnaiah, V., Narsimha, D. G., & Chandra, D. N. S. (2013). "Diagnosis of Lung Cancer Prediction System using Data Mining Classification Techniques", *International Journal of Computer Science and Information Technologies*, vol.4, no.1, pp. 39-45.
- Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q. and Liu, Q. (2013). "Comparison of Three Data Mining Models for Predicting Diabetes or Prediabetes by Risk Factors", *The Kaohsiung journal of medical sciences*, vol. 29, no. 2, pp. 93-99.
- Moro, S., Rita, P., & Vala, B. (2016). "Predicting Social Media Performance Metrics and Evaluation of the Impact on Brand Building: A Data Mining Approach", *Journal of Business Research*, vol. 69, no. 9, pp. 3341-3351.
- Naik, A. and Samant, L. (2016). "Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime", *Procedia Computer Science*, vol. 85, pp. 662-668.
- Nandakumar, A. (2001). "Consolidated Report of the Population based Cancer Registries", *Incidence and Distribution of Cancer*, pp. 1990-1996.
- Priyanga, A., & Prakasam, S. (2013). "Effectiveness of Data Mining Based Cancer Prediction System (DMBCPS)", *International Journal of Computer Applications*, vol. 83, no. 10.
- Ramachandran, P., Girija, N., & Bhuvaneshwari, T. (2014). "Early Detection and Prevention of Cancer using Data Mining Techniques", *International Journal of Computer Applications*, vol. 97, no. 13.
- Rao, DN., Ganesh B. (1998) "Estimate of Cancer Incidence in India in 1991", *Indian J Cancer* 35, pp. 10-8.
- Rathore, S., Hussain, M., Ali, A., & Khan, A. (2013). "A Recent Survey on Colon Cancer Detection Techniques", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 10, no. 3, pp.545-563.
- Senturk, Z. K., & Kara, R. (2014). "Breast Cancer Diagnosis via Data Mining: Performance Analysis of Seven different algorithms", *Computer Science & Engineering*, vol. 4, no. 1, pp. 35
- Thangaraju, P., Karthikeyan, T. and Barkavi, G. (2014). "Mining Lung Cancer Data for Smokers and Non-Smokers by Using Data Mining Techniques", *International Journal of Advanced Research in Computer and Communication Engineering* vol. 3.
- Vidhu, M. R. and Kiruthika, M. S. (2016). "A New Feature Selection Method for Oral Cancer Using Data Mining Techniques", *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no.1.
- Vital, T. P., Raju, G. P., Sreeramamurthy, K. And Charan, V. V. (2015). "A Probabilistic Neural Network Approach for Classification of Datasets Collected from North Coastal Districts of AP, India Using MatLab", *Procedia Computer Science*, vol. 48, pp. 715-721.
- Wu, H. C. and Lu, C. N. (2002). "A Data Mining Approach for Spatial Modeling in Small Area Load Forecast", *IEEE Transactions on Power Systems*, vol. 17, no. 2, pp. 516-521.
- Wynder, EL., Covey LS, Mabuchi K (1974). "Current Smoking Habits by Selected Background Variables: Their Effect on Future Disease Trends", *Am J Epidemiol* 100, pp. 168-177.