Systematic Analysis of Health Risks of Smoking Women UsingDatamining Techniques

Kanimozhi V A[1]*, Dr. Krishnapriya V[2]

[1]*Assistant Professor, Department of Computer Science (CT), PSG College of Arts & Science, Coimbatore, Tamilnadu, India.
[2]Professor & Head, Department of Computer Science with Cognitive System, Sri Ramakrishna Collge of Arts and Science, Coimbatore, Tamilnadu, India.

***Corresponding Author: -** Kanimozhi V A
*Assistant Professor, Department of Computer Science (CT), PSG College of Arts & Science, Coimbatore, Tamilnadu, India.

Abstract
An increasing number of women are becoming victims of their smoking habit. The marketing experts promote smoking as a way of remaining slim in a culture obsessed with thinness. The woman who smokes today is a heavier smoker, on average, with the percentage of women smoking more than 25 cigarettes per day. Women start smoking at younger and younger ages. 84% of women smokers who are now 28-37 years began to smoke before age 20 as compared with 42% of those now 58-67 years. Today more young women than young men smoke. This proposed work provides a comprehensive analysis of various health risks associated with smoking women. And also, a detailed comparison has given on numerous health risks among smoking and non-smoking women. Various data mining techniques has proved the adverse impact of smoking habit in women health. This analysis enlightened data mining techniques utilized to demonstrate the influence of smoking habit in health risks of women.

Keywords: Data mining, smoking, smoking women, Health risks.

INTRODUCTION TO SMOKING

Smoking is the most preventable reason ofdeath in our country, yet more than 140,000women die each year from smoking-relatedcauses. The highest rate of smoking (27%)occurs among women between 25 and 44.Despite all the warnings today's teens haveheard about the dangers of smoking, the sadreality is that almost all of the new smokerstoday are teenagers; over 1.5 million teenagegirls smoke cigarettes.
Women smokers suffer all the consequences of smoking that men do such as increased risk of various cancers (lung, mouth, esophagus, kidney, pancreas, kidney, and bladder) and respiratory diseases, and also they face numerous unique challenges compared to men.

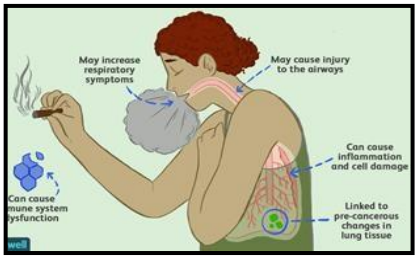

Figure. 1. Smoking Women Health Risks

1. WHY DO WOMEN SMOKE

Various data mining techniques has proved the adverse impact of smoking habit in women health. The causes of smoking in women arenumerous. Increased stress, so-called metro culture and many more reasons why womensmoke.

- A stress buster
- It looks cool
- The feel of being an independent woman
- In the pursuit of weight loss

Some believe the misconception that smoking helps them reduce their weight which on the contrary damages their skin and figure. [4]

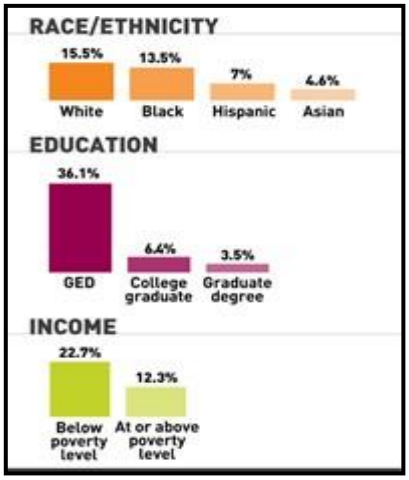

Figure. 2. Smoking Rates of Women In 2016

A) Heavy Smokers versus Light Smokers Bruno Samways dos Santos et.al.,[19]conducted a test on on four datasets extractedfrom the National Survey on Drug Use andHealth (NSDUH, USA) database. The bestresults were obtained with Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Logistic Regression (LR), whichhad an accuracy of approximately 85% andrecall of up to 77%.

The features considered most relevant for both classes were identified and discussed. TheCIGAGE variable, which is the age when a woman started smoking daily, for example, shows that each year that a woman lives withoutsmoking on a daily basis reduces the chances of her being classified as a heavy smoker by afactor of 0.911. [19].

2. Comparison of smoking women versus non-smoking women health risks

A) Lungs Damage

- The risk for dying of lung cancer is 20 times higher among women who smoke two or more packs of cigarettes per day than among women who do not smoke.[1] This risk is 25 times greater for men and 25.7 times greater for women [2]
- Women exposed to cigarette smoke are twice as likely to develop lung cancer as men.
- Smoking is directly to linked to 80 percent of Chronic Obstructive Pulmonary Disease (COPD) deaths in women each year. [1]. American Lung Association report that smoking causes 80 percent of COPD deaths. Also, The Centre for Disease Control and Prevention (CDC) report that roughly 9 out of 10Trusted Source lung cancer deaths is linked to smoking. [2]

B) Heart Disease

- Smoking cigarettes can damage theheart, blood vessels, and blood cells. Research shows a direct link between smoking and developing peripheral artery disease (PAD). Even those who used to smoke face a higher risk than women who never smoked.
- Researchers in Denmark have found a

50 percent greater risk of a heart attack in women smokers over men smokers. This difference may be due to the interaction of estrogen with the chemicals found in cigarettes.[2]

C) Aging

- Smoking ages women. Smokers have more facial wrinkles, gum disease, dental decay, and halitosis (bad breath). One studyfound more women smokers had gone grey by age 40. That risk doubled by age 50. [1]

D) Anxiety Disorder

- The Surgeon General's Report concluded that smokers are more likely to be depressed than non-smokers and that womenwith anxiety disorders are more likely tosmoke.[1]

E) Fertility Diseases

- Smoking cigarettes can damage a female's reproductive system and make it more difficult to get pregnant. This may be because tobacco and the other chemicals in cigarettes affect hormone levels. [2]
- Cigarette smoking has many adverse reproductive and early childhood effects, including an increased risk for infertility, preterm delivery, stillbirth, low birth weight and sudden infant death syndrome (SIDS). Women smokers often have symptoms ofmenopause about three years earlier than non- smokers.[1]

F) Pregnancy Complications

- Women who smoke have more difficulty becoming pregnant and have a higher risk of never becoming pregnant.[6]
- Chemicals in tobacco are passed frompregnant mothers through the bloodstream tothe fetus. These toxic chemicals present serious risks to the unborn child as well as the mother. According to "Our Bodies, Ourselves for the New Century," by the Boston Women's Health Book Collective:
- Smoking during pregnancy is associated with preterm delivery, low birth weight, premature rupture of membranes, placenta previa, miscarriage, and neonatal death. new- borns whose mothers smoked during pregnancy have the

same nicotine levels in their bloodstreams as adults who smoke, and they go through withdrawal during their first days of life.[3]

- Smoking during pregnancy can cause tissue damage in the unborn baby, particularly in the lung and brain, and some studies suggests a link between maternal smoking and cleft lip.
- Studies also suggest a relationship between tobacco and miscarriage. Carbon monoxide in tobacco smoke can keep the developing baby from getting enough oxygen. Tobacco smoke also contains other chemicals that can harm unborn babies.[6]

G) Osteoporosis

- Smoking causes a significant increase in the risk of bone loss and osteoporosis. Women who smoke, one pack of cigarettes a day, often experience a loss of bone density equalling five to 10 percent more than non-smokers by the time they reach menopause. [3]
- Postmenopausal women who smoke have lower bone density than women who never smoked. Women who smoke have a bigger risk for hip fracture than never smokers.[1] Also, Pelvic inflammatory disease occurs with 33% more frequency in smokers than in non- smokers. [3]

H) Breast Cancer

- The American Cancer Society published the results of a study in 1994 which indicated that breast cancer patients who smoke may increase their risk of dying at least 25 percent—a risk that increases with the number of cigarettes smoked per day.
- The possible risk of fatal breast cancer rises up to 75 percent for women who smoke two packs or more per day.[3]

3. Role of datamining techniques in the analysis of smoking women health risks

Data Mining is a non-trivial method for finding valid, novel, possible, useful and ultimately understandable patterns in data. It puts together a range of tools and techniques that can be applied to the processed data in order to discover hidden patterns [16]

Data mining in medical helps to extract useful knowledge and provide scientific decisionmaking for the diagnosis and treatment of disease from the medical database. Applying data mining techniques in medical data to extract meaningful patterns and knowledge is called Medical Data mining. [17] Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs.

Data mining provides methods and technologies for the exploration of expertise in the clinical data warehouse. Data mining methodologies play an unavoidable role in the medical industry. [16] For example, data mining can help the healthcare industry in fraud detection and abuse, customer relationship management, effective patient care, and best practices, affordable healthcare services. [15]

Following are the various kinds of techniques utilized for characterization and forecast of information for different issues. [18]

- Classification,
- K-Nearest Neighbor,
- Decision Tree (DT),
- Support Vector Machine (SVM),
- Neural Network (NN),
- Bayesian Methods,
- Regression,
- Clustering,
- Partitioned Clustering,
- Hierarchical Clustering,
- Density based Clustering,
- Association rule mining,
- Random Forest

S. No	Category	Authors	Title	Techniques used	Year	Accuracy
1	Smoking lead s to heartdisease	S. R. Rathodand C. Y. Patil	Performance Assessment of Ensemble Learning Model for Prediction of Cardiac Disease Among SmokersBased onHRV Features	Ensemble machinelearning methodswith boostingtechnique	2021	95.2%
		Wael K. Al Delaimy,John E Manson,Caren G Solomon	Smokingand risk of coronaryheart disease among women with type 2 diabetes mellitus	Multivariate analysistechniques	2002	Relative risk = 95%
2	Smoking leads to Aging	Arzu Baloglu,Evrim Zeynep Unlu	Data Mining Analysison Healthy Aging inTurkey	Spearmen rank correlation	2020	15.9% correlation between general appearance and smoking& alcohol consumption
3	Smoking leads to Breast Cancer	Mohammad Esmaeil Akbari, Soheila Sayad, Saed Sayad, Maryam Khayam zadeh, Leila Shojaee, Zeynab Shormeji, and Mojt aba Amiri	Breast Cancer Status in Iran: Statistical Analysisof 3010 Cases between 1998 and 2014	Univariate analysis	2017	-
		Ronak Sumbaly, N. Vishnusri and S. Jeyalatha	Diagnosis of Breast Cancer using DecisionTree Data Mining Technique	J48 –Decision Tree	2014	94.5%
4	Smokingleads to Lun g Cancer	Thangaraju P, BarkaviG, Karthikeyan T	Mining Lung Cancer Data forSmokersand Non- Smokersby UsingData Mining Techniques	NaïveBayes	2014	83.4 %
5	Smoking leads to Osteoporosis	Walid Moudani, Ahmad Shahin, Fadi Chakik, Dima Rajab	Intelligent Predictive Osteoporosis System	J48 –DecisionTree	2011	99.4 %
6	Smoking leads to Pret erm birth	Hsiang-Yang Chen, Chao-Hua Chuang,Yao- Jung Yang, Tung-Pi Wu	Exploring the risk factors of preterm birth using data mining	Decisiontree C5.0	2010	-

Table 1. Role of data mining techniques inthe analysis of smoking women health risks

a) SMOKING LEADS TO LUNG CANCER: Thankaraju et.al.,[7]

People who smoke have the greatest risk of lung cancer. The risk of lung cancerincreases with the length of time and number ofcigarettes they have smoked. If they quitsmoking, even after smoking for many years, they can significantly reduce his/her chances of developing lung cancer. In this work, author apply classification techniques on a dataset of lung cancer patients based on smoking and non-smoking people. Smoking is the biggest riskfactor of lung cancer. The more years and larger number of cigarettes smoked the greater the risk of developing lung cancer. [7]

Here experiment was successfully performed with several data miningclassification techniques and it is found that the Naive Bayes algorithm gives a better performance over the supplied data set with the accuracy of 83.4%. It is believed that the data mining can significantly help in the Lung Cancer research and ultimately improve the quality of health care of Lung Cancer patients. Equations. [7]

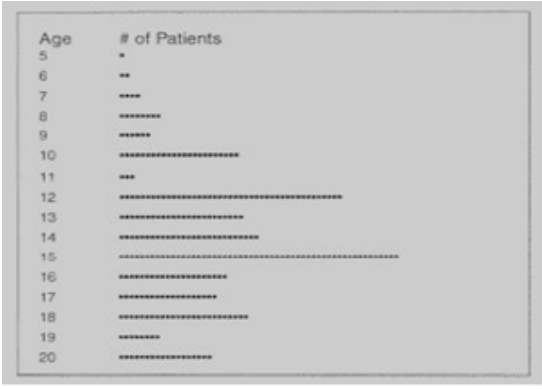

Figure. 3 No of Patients Who Begun ToSmoke At Each Age

Above figure illustrates a majority of thepatients began to smoke between the ages of 10-
20. Percentage-wise, there were more women who began to smoke before the age of 10 than men, and again in the category of those who began after 30, there are more women than men.[7]

b) SMOKING LEADS TO AGING: Arzu Baloglu et.al.,[5]

Obstacles in the path to successful agingare connected with behavioural determinants and genetic determinants. Author gathered information related to behavioural determinants (smoking, alcohol use, diet, BMI index, physical activity, sleep patterns, education level), genetic determinants (diabetes, anemia, hemochromatosis, breast cancer, allergy, mental illnesses), socio-psychological determinants (marital status, interest in arts, and music, having a child, happiness index, relationship with friends & family, loneliness, financial situation). Results can be used to understandmajor risk factors among determinants.[5]

A survey is conducted by distributing questionnaires to people having age 65+. All participants are residents in Istanbul. The study is conducted with 68 participants, of which 42 are female and 26 are male. Survey questions cover participants' behavioural determinants (smoking/alcohol consumption, physical activity, level of education), genetic determinants (genetic disease history in the family).[5]

Normality Tests, then Spearman Rank Correlation Tests, Chi-Square Tests, andKruskal Wallis Tests were conducted, respectively. There is a 15.9% correlation between general appearance and smoking & alcohol consumption in a negative direction and45.8% positive correlation supports healthy aging, physical activity and happiness. Smoking and drinking alcohol has a negative effect on a healthy appearance. Alcohol consumption and smoking associate with cancers, chronic liver disease, cardiovasculardisease, alcohol poisoning, and lung disease, allof which can be prevented by lifestyle change.[5]

c) SMOKING LEADS TO PRETERM BIRTH: Hsiang-Yang Chen et.al., [8]

The goal of this work was to explore therisk factors of preterm using data mining with neural network and decision tree C5.0. Using the nest case-control study design, a total of910 mother–child dyads were recruited from14,551 in the original data. Thousands of variables are examined in this data including basic characteristics, medical history, environment, and occupation factors of parents,and variables related to infants.[8]

The results indicate that multiple birth, hemorrhage during pregnancy, age, disease, previous preterm history, body weight before pregnancy and height of pregnant women, and paternal life style risk factors related to drinking and smoking are the important riskfactors of preterm birth. Preterm birth is one of the leading causes of diseases and death among newborns. In addition, preterm infants often suffer long-term health problems, including lung diseases, vision and hearing impairments,and learning disabilities.[8]

d) SMOKING LEADS TO HEART DISEASE: Wael K Al-Delaimy et,al.,[12]

Main object of this work is to assess the relationship between cigarette smoking and risk of CHD among women with type 2 diabetesmellitus in the Nurses' Health Study co hort. Author documented 458 incident cases of CHD (200fatal CHD-related cases and 258 nonfatal myocardial in- farctions) during 20 years (68227person-years) of follow-up. [12]

There is a dose-response relationship between current smoking status and risk of CHD among diabetic women. Compared with never smokers, the relative risks (RRs) for CHDwere1.21 (95%confidenceinterval [CI], 0.97- 1.51) for past smokers, 1.66 (95% CI, 1.10-
2.52) for current smokers of 1 to 14 cigarettes per day, and 2.68 (95%CI, 2.07-3.48) forcurrent smokers of 15 or more cigarettes per day in multivariate analyses (P<.001fortrend). The multivariate RR of CHD among diabetic women who had stopped smoking formore than 10 years was similar to that among diabetic women who were never smokers (RR, 1.01; 95% CI, 0.73-1.38). [12]

In secondary analyses involving diabeticand nondiabetic women, the multivariate- adjusted RR of CHD for those with diabeteswho currently smoked (≥15 cigarettes per day)com-paredwiththosewhoneversmokedwas7.67 (95%CI, 5. 88-10.01). So Cigarette smoking isstrongly associated with an increased risk ofCHD among women with type 2 diabetesmellitus. Furthermore, quitting smoking seemsto decrease this excess risk substantially;women with diabetes should b is trogly advisedagainst smoking. [12]

S.R. Rathod et,al.,[9]

Smoking impacts the pattern of heart rate variability (HRV); HRV therefore acts as a predictor of cardiac diseases (CD). In this study,to predict CD non-invasively among smokers, ensemble machine learning methods have been used. A single model is created based on ensemble voting classifier with a combined boosting technique to improve the accuracy of predictive model. [9]

The final ensemble model shows an accuracy of 95.20%, precision of 97.27%, sensitivity of 92.35%, specificity of 98.07%, F1score of 0.95, AUC of 0.961, MCE of 0.0479, kappa statistics value of 0.9041, and MSE of 0.2189. The obtained accuracy by using the proposed method is the highest value achieved so far for the prediction of CD among smokers using HRV data. [9]

e) SMOKING LEADS TO BREASTCANCER: Mohammad Esmaeil Akbari et,al.,[10]

Smoking can increase the risk of breast cancer occurrence. Unfortunately, the number of women smokers is increasing. According to the Iranian Atlas of Women, 4.3% of Iranian women were smokers in 2004. This has increased

to 6.9% in 2010. In this study, 6.6%of the studied women were smokers. This shows that the percentage of women with breastcancer who are smokers. [10]

Ronak Sumbaly et,al.,[13]

The causes of breast cancer are not fully known. Researchers believe that these risk factors increase (or decrease) the changes ofdeveloping breast cancer. Since breast cancer isa complex disease it is likely to be caused by a combination of risk factors. Some of the factorsassociated with breast cancer – can't be changed (Non-preventable) like age, genetic factor, heredity. While making choices can change other factors (Preventable) like overweight, lack of exercises and smoking.But smoking is linked with higher risk of breastcancer in younger and premenopausalwomen.[13]

f) SMOKING LEADS TO OSTEOPOROSIS: Walid Moudani et,al., [11]

Smoking inhibits the activity of osteoblasts (cells responsible of formation), and is an independent risk factor for osteoporosis. Smoking also results in increased breakdown of exogenous estrogen, lower body weight and earlier menopause, all of which contribute to lower bone mineral density. [11]

5. CONCLUSION

India has around 12.1 million female smokers and about 250 million women are dailysmokers. Smoking habit is giving numerous dangerous impacts on women health. Due tosmoking, lungs will get severely damaged. In addition to this risk of lung cancer, women who smoke will have risky effects in reproductive organs. An antiestrogen effect of smoking may provide the explanation for why smoking women reach menopause 1-2 years earlier than non-smokers. Smoking also reduces fertility, increases the possibility of spontaneous abortion of chromosomally normal foetuses. A growing body of evidence also shows that smoking during pregnancy has long-term effects on children. Women should be strongly advised to quit smoking.

This paper provides a systematic analysis of various data mining techniques which were utilized in the health risks associated with smoking women. Here data mining techniques demonstrated the severe effect of smoking habit to women health. Algorithms like Naïve bayes, decision tree and various ensemble methods proved that thesmoking habit leads to various diseases like lung cancer, breast cancer, heart diseases, osteoporosis and etc.,.This proposed work gave an comprehensive exploration of utilization of data mining techniques in smoking women health risks analysis.

REFERENCES

[1] https://www.nationaljewish.org/conditions/health-information/smoking-and-tobacco-use/health/women-and-smoking
[2] https://www.medicalnewstoday.com/articles/324644#weakened-immune-system
[3] https://www.verywellmind.com/smoking-womens-health-perspective- 3520429#osteoporosis-and-smoking
[4] https://www.indushealthplus.com/smoking-effects-on-womens-health.html
[5] Arzu Baloglu, Evrim Zeynep Unlu, "Data Mining Analysis on Healthy Aging in Turkey", International Journal of Health and Medical Sciences, volume 6, 2020.
[6] https://www.cdc.gov/tobacco/basic_info
[7] rmation/health_effects/pregnancy/
[8] Thangaraju P1, Barkavi G2, Karthikeyan T3, "Mining Lung Cancer Data forSmokers and Non- Smokers by Using Data Mining Techniques", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, July 2014
[9] Hsiang-Yang Chen, Chao-Hua Chuang, Yao-Jung Yang, Tung-Pi Wu, "Exploring the risk factors of preterm birth using data mining",Expert Systems with Applications 38 (2011)
[10] S. R. Rathod and C. Y. Patil,"Performance Assessment of Ensemble Learning Model for Prediction of Cardiac Disease Among Smokers Based on HRV Features", International Journal of Biomedical and Clinical Engineering (IJBCE) ,2021
[11] Mohammad Esmaeil Akbari, Soheila Sayad, Saed Sayad, MaryamKhayamzadeh, Leila Shojaee,1 ZeynabShormeji, and Mojtaba Amiri, "Breast Cancer Status in Iran: Statistical Analysis of 3010 Cases between 1998 and 2014", International Journal of Breast Cancer, 2017.
[12] Walid Moudani, Ahmad Shahin, Fadi Chakik, Dima Rajab, "Intelligent Predictive Osteoporosis System", International Journal of Computer Applications (0975 – 8887) Volume 32– No.5, October 2011
[13] Wael K. Al Delaimy, John E Manson, Caren G Solomon, "Smoking and risk of coronary heart disease among women with type 2 diabetes mellitus", ARCH INTERN MED,vol 162, 2002
[14] Ronak Sumbaly, N. Vishnusri and S. Jeyalatha, "Diagnosis of Breast Cancer using Decision Tree Data Mining Technique",International Journal of Computer Applications,2014.
[15] https://www.medindia.net/patients/lifestyleandwellness/smoking-among-women-why- do-smoke.htm
[16] https://www.healthcatalyst.com/data-mining-in-healthcare
[17] Kanimozhi .V .A, Vibinchandar .S, "Impact of Artificial Intelligence in Feature Selection Methods of Naïve Bayes Prediction Model", Solid State Technology Volume: 63Issue: 5, 2020.
[18] Dr. T. Karthikeyan, V.A.Kanimozhi,"Deep Learning Approach for Prediction of Heart Disease Using Data mining Classification Algorithm Deep Belief Network", International Journal of Advanced Research in Science, Engineering and Technology, ISSN: 2350-0328,Vol. 4, Issue 1, January 2017.
[19] Vibinchandar .S, Dr. Krishnapriya .V, "Survey On Various Prediction Models For Survival Of Breast Cancer

Patients Using WarmBoot Random Forest Classifier", EuropeanJournal of Molecular & Clinical Medicine, ISSN 2515-8260, Volume 07, Issue 09, 2020.

[20] Bruno Samways dos Santos , MariaTeresinha Arns Steiner, Rafael HenriquePalma Lima, "Proposal of a method to classify female smokers based on data mining techniques", Computers & IndustrialEngineering, Volume 170, August © 2022 Elsevier.