Machine Learning and Data Mining Methods in Early Detection of Stomach Cancer Risk

Md. Rejaul Islam Royel[1], Md. Ajmanur Jaman[1], Fuyad Al Masud[1,2], Arzo Ahmed[3], Abdul Muyeed[4], and Kawsar Ahmed[2,5,*]

[1] *Department of Software Engineering (SWE), Daffodil International University, Dhaka, Bangladesh*
[2] *Department of Information and Communication Technology (ICT), Mawlana Bhashani Science and Technology University, Tangail-1902, Bangladesh*
[3] *Department of Statistics, Mawlana Bhashani Science and Technology University, Tangail-1902, Bangladesh*
[4] *Department of Statistics, Jatiya Kabi Kazi Nazrul Islam University, Trishal, Bangladesh*
[5] *Group of Bio-photomatix, Department of Information and Communication Technology (ICT), Mawlana Bhashani Science and Technology University, Tangail-1902, Bangladesh*
[*] *Corresponding author. E-mail: k.ahmed.bd@ieee.org*

Received: May. 01, 2020, Accepted: Jul. 07, 2020

Background: In point of mortality rate, stomach cancer is the fifth leading cancer. There are some risk factors of stomach cancer those are varied with country to country and associated with urbanization and economic development. Diagnosis of stomach cancer is a difficult task, only about 10% of people are diagnosed while it's still in the initial stage. The main objective of this research is to design a tool for early detection of stomach cancer risk level. **Methodology:** Firstly, feature selection techniques are applied to filter the collected data. After that, the best rules technique is used to check the correlations of risk factors with stomach cancer. Besides, the visual relationship among factors and selected cancer are also exhibited. Then, the score is assigned for each factor according to the impact of risk on stomach cancer patients. Finally, the stomach cancer risk level prediction tool is designed. **Results:** After the experiment of 300 subjects' records (150 are affected and 150 are non-affected) with 32 risk factors, we have received 18 significant-top risk factors of stomach cancer. Abdominal Pain, Nausea, Skin Color Turn into pale are respectively found top risk factors of stomach cancer. Furthermore, some other factors related to socio-economic conditions are also indicated to have stomach cancer. **Conclusion:** In conclusion, this study will be helpful to early detection of stomach cancer risk level and to increase the awareness among the people of Bangladesh as well as the rest of the world.

Keywords: Stomach Cancer in Bangladesh; Preoperative Risk Factors; Feature Selection Method; Risk algorithm; Correlation

http://dx.doi.org/10.6180/jase.202102_24(1).0001

1. Introduction

According to GLOBOCAN cancer database estimation, in the last two decades, the death rate of cancer has been reduced. During this period fewer people are dead than before. All disease has occurred in our body due to having a shortage of some kind of antibody and hormones. Also, all disease symptoms are shown in our body mostly in a long period. According to our study, we found that most of the affected people are taking gastric medicine for a long time but they do not think a long time gastric could be a risk factor of Stomach Cancer. The rates of stomach cancer affected patients are alleviated. Stomach cancer is the fifth leading cause of cancer-related to mortality in the world [1]. There are remarkable geographical variations in stomach cancer mortality and high incidence which are highly observed in Asia, Central, South, and Eastern America [2, 3]. There are 20 lakh people affected by cancer and every year approximately 2 lakh people are newly diagnosed with

cancer where the position of stomach cancer is the top five in Bangladesh [4, 5]. It is a positive sight that the ratio has decreased in the last few decades. But it is still high in Asia, Central, South, and Eastern America [6].

Overall, most of the stomach cancer cases were diagnosed based on pathology examination, surgery, Goss tissue specimens, gastroscopy, ultrasound or radiological examinations and some of the basis on physical exam, family history [7]. The physical condition depends on daily activities and fruit habits. Citrus fruit, vegetables, and whole grains, high in antioxidant vitamins and polyphenols [8], may lower stomach cancer risk by protecting the gastric epithelium from inflammatory responses caused by Helicobacter pylori and by reducing endogenous carcinogenic nitrosamine formation [9, 10].

While dietary habits, such as foods rich in nitrate or nitrite and their derivatives, a high salt diet, a high carbohydrate diet, and a diet low in fresh fruits and vegetables have been associated with these chronological and geographical variations in stomach cancer incidence [11]. Several types of pathological changes in stomach cancer, i.e., atrophic gastritis, intestinal metaplasia, gastric ulcer, and polyps have been suspected as premalignant lesions for stomach cancer [12]. Some previous prospective studies of stomach cancer have been small [13] that limited statistical power to examine risk separately by gender or other potential risk modifiers. Consumption of processed meats, which contain nitrosamine precursors, has been inconsistently associated with stomach cancer [14]. Risk factors for stomach cancer include low intake of vegetables, fruits, alcohol drinking, tobacco, smoking, and high intake of salt [15]. From the above studies, it is clear that stomach cancer depends on fruits habits, socio-economic status, physical conditions as well as activities. With our best of knowledge, there are no studies to show the risk level of a stomach cancer patient by analyzing the risk factors. So, there is still room to design a model for early predicting of stomach cancer risk level.

This study aims to discover significant preoperative risk factors of stomach cancer perspective to Bangladesh and to develop a risk prediction tool. This tool will be able to determine the risk level of any individual random person's stomach cancer.

2. Theory and formula

Data Collection: Collecting data is an important task for any kind of survey-based research. The survey question has been designed for the study of different preoperative risk factors of stomach cancer-based research paper. In this study, the age range of the collected data was greater than or equal to 30 years old. All data are collected from the National Institute of Cancer Research and Hospital and the total collected individual's sample size of data is about 300 where the case group is 150 and the control group is 150.

Analytical Approach: In a machine learning approach, first of all, we have used feature selection technique (attribute evaluator in weka) to find out the most significant risk factors of stomach cancer. Five different filtering techniques (Correlation, Information Gain, Gain Ratio, Relief, and Symmetrical Uncertainty) have been applied with ranker algorithms to get each feature rank within 0 to 1 [16–20]. After that, the average ranking has been applied to get a single accurate rank of individual features. If any features rank is nearest to 1 then it is very much correlated. And nearest to 0 means less correlation with the disease. Then, we have used association rules (by Predictive Apriori Algorithm in R) to extract the hidden pattern of data [19, 21]. It is also an advanced form of Apriori algorithm where we used minimum support 0.01, minimum confidence is 0.80 and the lift is greater than 1 and all related figures are developed by "arulesViz" packages in R [22, 23].

Sub factors Priority Calculation: Each sub-factors risk level is calculated by their supported value; those are analyzed from the Apriori algorithm and we have measured risk criteria by its top supported rules. Top rules are selected by the methods of first come first serve. For "Disease = Yes" rules, we have observed maximum support of 41; minimum support of 14, and carried out the outcome of the risk level as 4 (Low, Moderate, High, and Very High). So, the Support difference is (41 – 14) = 27. Here, we have set the supported value 14 to 20 as "Low Risk", value 21 to 27 as "Moderate Risk", value 28 to 34 as "High Risk" and value 35 to 41 as "Very High Risk" [24–26].

The same procedure has been applied to get a risk level for "Disease = No" rules. But we have to calculate score maximum to minimum. Because no rules supported value indicates these sub-factors are not responsible to have the disease. If the support value is decreased, the risk level is increased. So, the supported value between 11 to 18 is "Very High Risk"; between 19 to 26 is "High Risk"; between 27 to 34 is "Moderate Risk"; and between 35 to 42 is "Low Risk".

Risk Algorithm Design: Designing of a risk prediction algorithm is a major important task in this study. This algorithm is developed by the available value of risk score on table 3 following the articles [24–26]. Here, we get the maximum score of 69.2 and minimum score of 30.35 where the score difference is 69.2 - 30.35 = 38.85 = 39 and the considerate number of risk level is 04. So, each risk level will get a value of 39/4 = 9.75. So, here is observed the score of Very High-Risk Level as (Score $\geq$ 59.45); High Risk

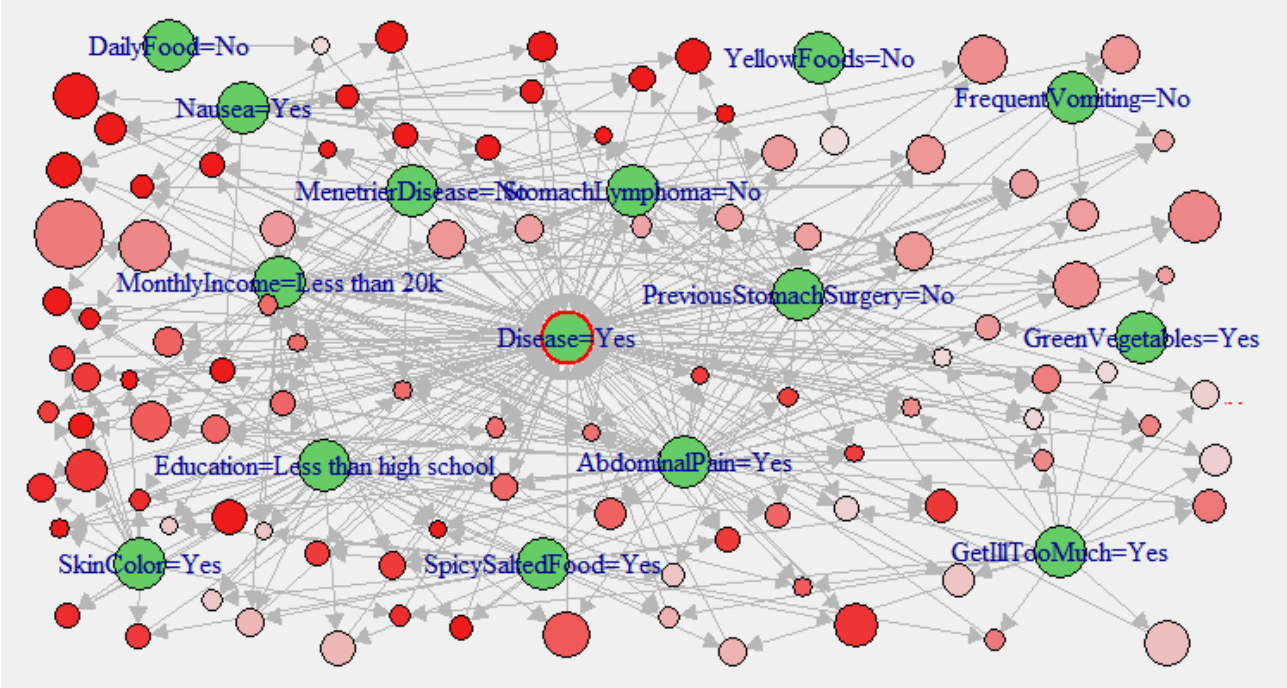

Fig. 1. Visual relationship among factors for Disease = Yes.

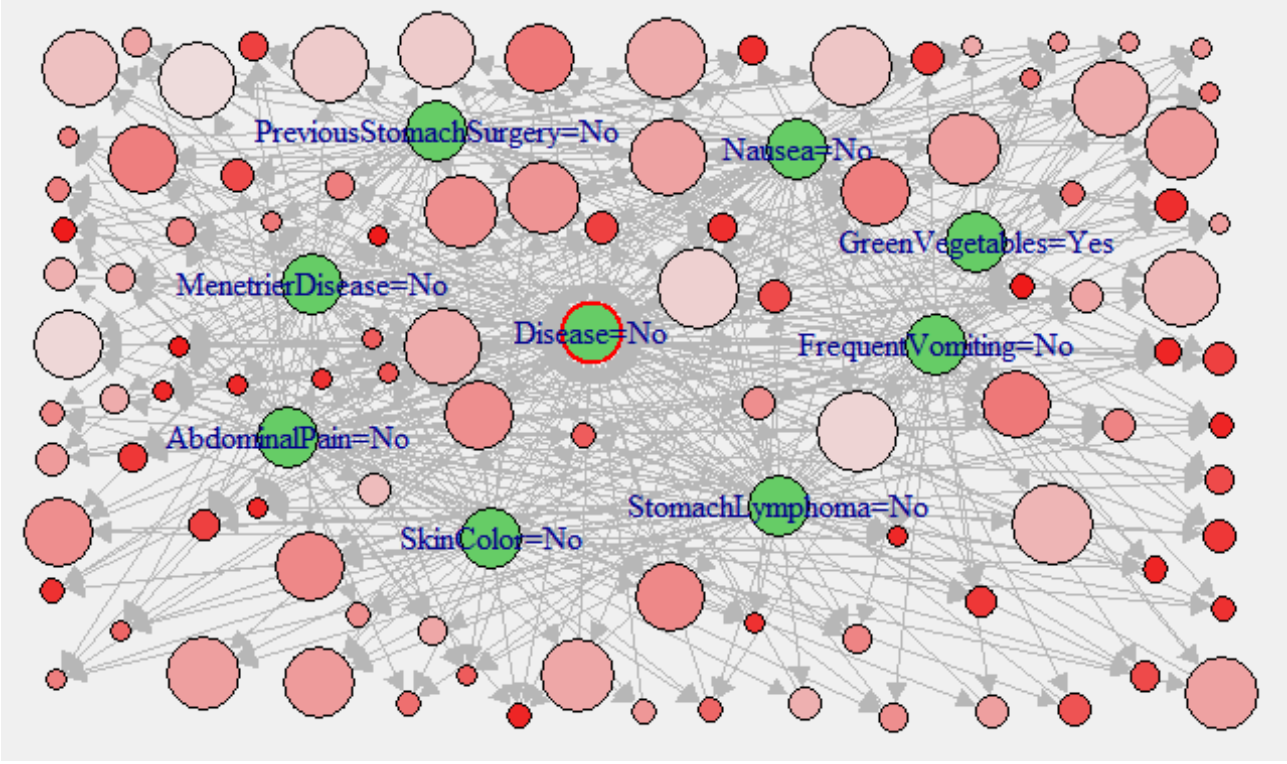

Fig. 2. Visual relationship among factors for Disease = No.

Level as (Score $\geq$ 49.70); Moderate Risk Level as (Score $\geq$ 39.95); and Low Risk Level as (Score < 39.95) [24–26].

3. Result discussions

Recently, a diagnosis and prognosis based study has been conducted to predict risk factors of any disease using med-

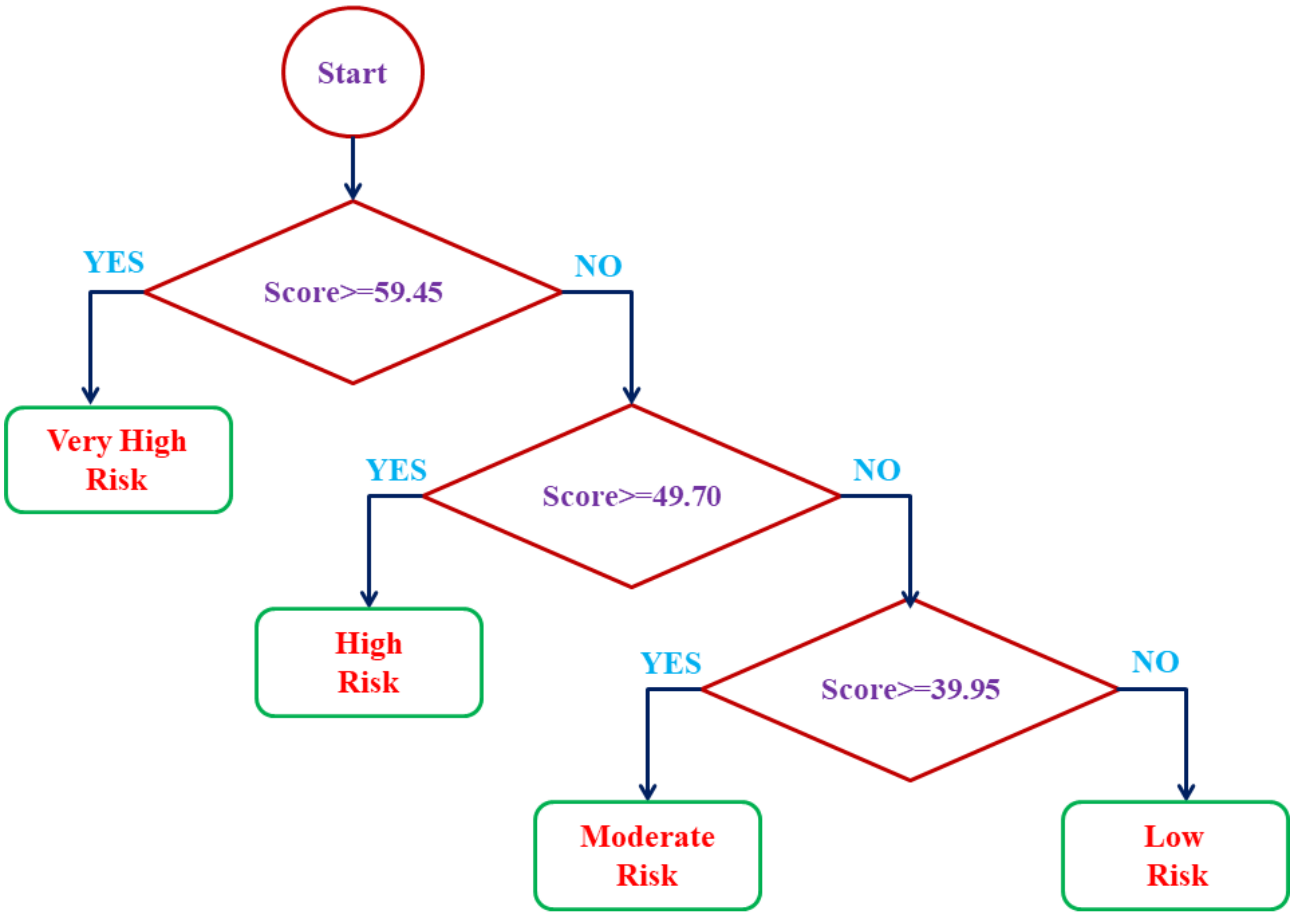

Fig. 3. Stomach Cancer Risk Prediction Algorithm Flowchart.

ical data. This type of study has drawn the attention of new researchers. Analyzing risk factors and developing algorithms are a common trend at present. Total of 300 Bangladeshi residences including case and control group people's data are analyzed in the result section. Total 32 preoperative factors are selected and the analysis process takes place at the four different tables including Feature selection, Disease= Yes rules, Disease = No rules, and Risk Score table. The examined analysis is discussed below stepwise.

Table 1 represents some most popular feature selection techniques (Correlation, Gain Ratio, Information Gain, Relief, and Symmetrical Uncertainty) results which are mostly used to filter medical data. All attribute evaluators are filtered with the ranker method and the expected average rank of 1.00 is highly correlated with target variable (Disease) and 0.00 refers there is no relationship among them. From this perspective, Abdominal Pain (0.764), Nausea (0.526), and Skin Color (0.445) exhibit a perfectly significant correlation with stomach cancer. Besides, Get Ill Too Much (0.299), Frequent Vomiting (0.13), BMI (0.218), Yellow Fruits (0.217), Education Level (0.191), Stomach Lymphoma (0.159), Spicy and Salted Food (0.153), Green Vegetables (0.149), Tobacco Status (0.144), Menetrier Disease (0.134), Previous Stomach Surgery (0.126), Daily Food (0.122), Age (0.09), Monthly Income (0.89) and Gastric Medicine (0.08) show significant correlation with stomach cancer.

Considering the positive disease patient, the top fifteen rules are considered to have stomach cancer. Where, it is examined that "Age = 50 to 59", "Skin Color = Yes", "BMI = Underweight", "BMI =severely underweight", "Nausea = Yes", "Education = Less Than High School", "Tobacco Status = Yes Excessive" and so on are extremely supported risk level to have stomach cancer. In addition, "Daily Food =No", "Stomach Lymphoma=No", "Menetrier Disease=No", "Previous Stomach Surgery = No" show low-risk level perspective to Bangladeshi patients. But those rules are also significant risk factors of stomach cancer in

Table 1. Feature selection techniques with respect to ranker method.

Features	Correlation	Gain Ratio	Information Gain	Relief	Symmetrical	Uncertainty Average Rank
Abdominal Pain	0.764	0.764	0.764	0.764	0.764	0.764
Nausea	0.712	0.505	0.466	0.459	0.485	0.526
Skin Color Turn into Pale	0.679	0.437	0.402	0.288	0.419	0.445
Get Ill Too Much	0.554	0.246	0.238	0.170	0.242	0.290
Eat Yellow Foods	0.442	0.160	0.149	0.180	0.154	0.217
Frequent Vomiting	0.420	0.277	0.169	0.075	0.210	0.230
Spicy Salted Food	0.349	0.093	0.090	0.143	0.092	0.153
Green Vegetables	0.334	0.111	0.085	0.122	0.096	0.150
Daily Food	0.328	0.082	0.079	0.041	0.080	0.122
Tobacco Status	0.325	0.080	0.111	0.110	0.093	0.144
Education Level	0.323	0.129	0.189	0.159	0.154	0.191
Stomach Lymphoma	0.308	0.218	0.093	0.045	0.130	0.159
Menetrier Disease	0.291	0.172	0.075	0.030	0.104	0.134
Previous Stomach Surgery	0.260	0.195	0.066	0.008	0.099	0.126
Gastric Medicine	0.221	0.041	0.036	0.061	0.038	0.080
Age	0.211	0.050	0.083	0.046	0.062	0.090
Monthly Income	0.211	0.049	0.063	0.056	0.055	0.087
BMI	0.185	0.176	0.341	0.156	0.232	0.218
Living Area	0.182	0.026	0.033	0.066	0.029	0.067
Family Member	0.168	0.026	0.035	0.021	0.030	0.056
Working Status	0.164	0.034	0.057	0.067	0.043	0.073
Physical Activity	0.145	0.023	0.030	0.031	0.026	0.051
Gender	0.140	0.015	0.014	0.065	0.015	0.050
Blood Vomiting	0.130	0.138	0.017	0.000	0.030	0.063
Tarry Stools	0.124	0.053	0.012	0.002	0.020	0.042
Diabetes	0.102	0.020	0.008	0.014	0.011	0.031
Blood Group	0.087	0.024	0.046	0.101	0.032	0.058
Another Cancer	0.058	0.104	0.003	0.013	0.006	0.037
Alcohol Status	0.048	0.064	0.011	0.007	0.019	0.030
Poor appetite	0.037	0.005	0.001	-0.002	0.002	0.009
Family History	0.012	0.006	0.004	0.041	0.005	0.014
Breast Cancer Status	0.000	0.000	0.000	-0.001	0.000	0.000

other studies [27–30].

Figure 1 is very helpful to understand the relationship among some top factors those are responsible for having a disease. Here, bubble size is representing the support and color is representing confidence. If bubble size goes to bigger than its support level will increase and if the color shows dark red then it shows pretty high confidence in this relationship. Here we get, if someone gets abdominal pain, have stomach lymphoma, having nausea, education level is less than high school, monthly income is less than 20K Bangladeshi Money, get ill too much, do not take daily food properly, do not eat yellow fruits and vegetables every day, also habited to eat spicy and salted food, and overall subjects skin color turn into pale then he/she could be affected with stomach cancer. And some factors like no frequent vomiting, no previous stomach surgery are also indicated to have stomach cancer.

Considering the negative disease patient, the top fifteen rules are considered to have no Stomach Cancer. Where it is examined that "Education = University Graduate", "Abdominal Pain = No", "BMI = Overweight", "Daily Food = Yes", "Abdominal Pain = No", "Menetrier Disease = No", "Spicy Salted Food = No and so on are highly supported to have no appendicitis disease. Also, "Tobacco Status = No", "Age = 30 to 49", "Stomach Lymphoma = No", "Skin Color = No" and so on represent low supported value to

have stomach cancer.

Figure 2 shows the visual relationship among some top factors that are gathering evidence to have no disease. At the same conditional parameters as Figure 1 as bubble size represents the support and color represents the confidence. Bigger bubble size indicates high support level and color dark red shows pretty high confidence among those relationships. Here we get, if someone will not have any stomach lymphoma, menetrier disease, nausea, frequent vomiting, abdominal pain, not changed skin color, and eats fresh green vegetables every day then he/she will be remaining safe from stomach cancer.

From the above analysis, 18 factors are selected as risk factors of stomach cancer. Table 2 represents the overall subcategories score in a single table. It has a total of 18 factors with 46 sub-factors individuals scored. First of all, the initial score is calculated. Next, the average score, elegant score are calculated. Then finally, we get the final score. Each sub-category score is defined by their importance or impact on the disease. Like, the Lowest score of Stomach Lymphoma = No is 0.5 and the Highest score of Abdominal Pain = Yes is 6.15. Finally, this table helps to design a risk level prediction tool.

Figure 3 is a conditional flowchart. The proposed tool will be worked based on this flowchart. To calculate the score, you have to fill up the conditions of the top 18 risk factors. Then, the score will be generated following the Table 2. It is clearly shown that if any individual subjects risk score is Score >= 59.45 then he is in Very High Risk, if Score >= 49.70 then he/she is in "High Risk", if Score is >= 39.95 then he/she is in "Moderate Risk" other wiles subject is in "Low" risk to have Stomach Cancer.

Finally, there are some common risk factors of stomach cancer that are associated with urbanization and economic development and salt, including salt-preserved foods, are probable causes of this cancer. Old age, female gender, and poor daily living are also the factors most frequently linked with the morbidities of stomach cancer. For advanced stomach cancer, surgery is a must for a cure and in the initial stage its success rate is 50% and it could be rectified. This study will be helpful to early detection of stomach cancer risk and to increase the awareness among lower developing countries like Bangladeshi people.

In limitation, the present study has been conducted on some common risk factors of stomach cancer perspective to Bangladesh. This study is mainly conducted in a government hospital. A large volume of data in different places all over the world may vary the finding of the study. In the future, data could be collected from all reputed private and public hospitals/clinics to enhance proper understanding.

4. Conclusions

According to the study, we found that long time gastric medicine could be a risk factor of Stomach Cancer. Perspective to Bangladesh, the most common risk factors Abdominal Pain, Nausea, Skin Color Turn into Pale, Get Ill too Much, Frequent Vomiting, BMI, Yellow Fruits, Education Level, Stomach Lymphoma, Take Spicy and Salted Food, Green Vegetables, Tobacco Status, Menetrier Disease, Daily Food in Time, Older Age, Monthly Income, Gender, Living Area, Blood Group, Physical Activity are associated with stomach cancer. Where, Abdominal Pain, Nausea, and Skin Color Turn into Pale are the top most risk factors of Stomach Cancer, those are also significantly found risk factors in only some studies. Notice that from the last two decades Stomach Cancer death rates are decreasing. It could be prevented if one takes action in the initial stage of cancer. We will recommend all people try to take proper nutrition, vitamin A, E and C, >80g vegetable and fruits every day, perform physical exercise regularly, and try to avoid tobacco. Vitamin, nutrition, fruits, and vegetables are very much protective against stomach cancer but tobacco decries this prevention power. Our implemented tool will be helpful for the early detection of stomach cancer risk level. If it is possible to use everybody, the low-income country like Bangladesh will be kept safe from stomach cancer and capable of saving a huge amount of money for the future.

Acknowledgments

The authors would like to acknowledge the chief oncology director and all clinical staff at NICRH in Bangladesh for their valuable support in conducting this study.

Table 2. Feature selection techniques with respect to ranker method.

Attribute	Final Score	Attribute	
Age		**Tobacco Status**	
30 to 49	2.75	No	1.5
50 to 59	3.3	Yes sometimes	3.05
60 to 70	3.3	Yes excessive	4.1
Above 70	4.35	**Skin color turn into Pale**	
BMI		No	2.5
Normal	1.05	Yes	6.05
Obese	1.95	**Abdominal Pain**	
Overweight	3.45	No	3.15
Severely Underweight	4.7	Yes	6.15
Underweight	4.7	**Nausea**	
Education		No	3.55
Less than high school	3.5	Yes	6.1
High school or College	2.4	**Frequent vomiting**	
University graduate	1.85	No	0.5
Doctoral Degree	0.8	Yes	2
Monthly Income		**Previous stomach surgery**	
Less than 20K	3.2	No	1.95
20K - 30K	2.15	Yes	1.45
30K - 45K	1.1	**Stomach Lymphoma**	
Above 45K	1.1	No	0.5
Daily Food in time		Yes	1.7
No	2.9	**Menetrier Disease**	
Yes	1.5	No	1.5
Spicy and Salted food		Yes	3
No	1	**Yellow fruits**	
Yes	3.15	No	3.9
Green Vegetables		Yes	1.5
No	3.6	**Gastric Medicine**	
Yes	1.5	Yes	3.05
Get ill too much		No	2
No	2.5		
Yes	5.8		

References

[1] Guoqin Yu, Nan Hu, Lemin Wang, Chaoyu Wang, Xiao You Han, Mike Humphry, Jacques Ravel, Christian C. Abnet, Philip R. Taylor, and Alisa M. Goldstein. Gastric microbiota features associated with cancer risk factors and clinical outcomes: A pilot study in gastric cardia cancer patients from Shanxi, China. *International Journal of Cancer*, 141(1):45–51, jul 2017.

[2] D. Pottier. Cancer Incidence in Five Continents. Population-at-risk. *IARC scientific publications*, (120):174–177, 1992.

[3] M Kurihara, K Aoki, and S Hisamichi. *Cancer Mortality Statistics in the World: 1950-1985*. 1989.

[4] Timothy L Cover. Helicobacter pylori diversity and gastric cancer risk, 2016.

[5] Syed Akram Hussain and Richard Sullivan. Cancer control in Bangladesh, 2013.

[6] Lis Ellison-Loschmann, Andrew Sporle, Marine Corbin, Soo Cheng, Pauline Harawira, Michelle Gray, Tracey Whaanga, Parry Guilford, Jonathan Koea, and Neil Pearce. Risk of stomach cancer in Aotearoa/New Zealand: A Māori population based case-control study. *PLoS ONE*, 12(7), jul 2017.

[7] Susan Greenhalgh. Diet, Life-style and Mortality in China: A Study of the Characteristics of 65 Chinese

Counties. *Population Studies*, 46(1):177–178, 1992.

[8] Leonard Marquart. Whole Grain Intake and Cancer: A Review of the Literature, jan 1995.

[9] Massimo Rugge, Matteo Fassan, and David Y. Graham. Epidemiology of gastric cancer. In *Gastric Cancer: Principles and Practice*, pages 23–34. 2015.

[10] Naohito Yamaguchi and Tadao Kakizoe. Synergistic interaction between Helicobacter pylori gastritis and diet in gastric cancer, 2001.

[11] Christopher P Howson, Tomohiko Hiyama, and Ernst L Wynder. The decline in gastric cancer: Epidemiology of an unplanned triumph. *Epidemiologic Reviews*, 8(1):1–27, 1986.

[12] David Schottenfeld and Joseph F. Fraumeni. *Cancer Epidemiology and Prevention*, volume 9780195149. 2009.

[13] Paul Terry, Olof Nyrén, and Jonathan Yuen. Protective effect of fruits and vegetables on stomach cancer in a cohort of Swedish twins. *International Journal of Cancer*, 76(1):35–37, 1998.

[14] World Cancer Research Fund and American Institute for Cancer Research. Food, Nutrition and the Prevention of Cancer: A Global Perspective. In *Washington DC: American Institute for Cancer Research*, 1997.

[15] World Cancer Research Fund. Diet, nutrition and the prevention of cancer: a global perspective. In *World Cancer Research Fund, Washington, USA*, 1997.

[16] Asha Gowda Karegowda, A S Manjunath, and M A Jayaram. Comparative Study of Attribute Selection Using Gain Ratio and Correlation Based Feature Selection. *International Journal of Information Technology and Knowledge Management*, 2(2):271–277, 2010.

[17] Bangsuk Jantawan and Cheng-fa Tsai. A Comparison of Filter and Wrapper Approaches with Data Mining Techniques for. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(6):4501–4508, 2014.

[18] RJ Urbanowicz, M Meeker, W La Cava, RS Olson, and JH Moore. Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*, 2018.

[19] Nicole Dalia Cilia, Claudio De Stefano, Francesco Fontanella, and Alessandra Scotto di Freca. A ranking-based feature selection approach for handwritten character recognition. *Pattern Recognition Letters*, 121:77–86, 2019.

[20] Joseph A. Cruz and David S. Wishart. Applications of machine learning in cancer prediction and prognosis, 2006.

[21] Illhoi Yoo, Patricia Alafaireet, Miroslav Marinov, Keila Pena-Hernandez, Rajitha Gopidi, Jia Fu Chang, and Lei Hua. Data mining in healthcare and biomedicine: A survey of the literature. *Journal of Medical Systems*, 36(4):2431–2448, aug 2012.

[22] Seyed Abbas Mahmoodi, Kamal Mirzaie, and Seyed Mostafa Mahmoudi. A new algorithm to extract hidden rules of gastric cancer data based on ontology. *SpringerPlus*, 5(1), dec 2016.

[23] Michael Hahsler and Sudheer Chelluboina. Visualizing Association Rules: Introduction to the R-extension Package arulesViz. *R project module*, pages 1–24, 2011.

[24] Kawsar Ahmed, Md. Ahsan Habib, Tasnuba Jesmin, Md. Zamilur Rahman, and Md. Badrul Alam Miah. Prediction of Breast Cancer Risk Level with Risk Factors in Perspective to Bangladeshi Women using Data Mining. *International Journal of Computer Applications*, 82(4):36–41, 2013.

[25] Tasnuba Jesmin, Kawsar Ahmed, Md. Zamilur Rahman, and Md. Badrul Alam Miah. Brain Cancer Risk Prediction Tool using Data Mining. *International Journal of Computer Applications*, 61(12):22–27, 2013.

[26] Kawsar Ahmed, Abdullah Al-Emran, Tasnuba Jesmin, Roushney Fatima Mukti, Md Zamilur Rahman, and Farzana Ahmed. Early detection of lung cancer risk using data mining. *Asian Pacific Journal of Cancer Prevention*, 14(1):595–598, 2013.

[27] Hye Min Jung, Jin Seok Lee, David R. Lairson, and Yoon Kim. The effect of national cancer screening on disparity reduction in cancer stage at diagnosis by income level. *PLoS ONE*, 10(8), aug 2015.

[28] Paule Latino-Martel, Vanessa Cottet, Nathalie Druesne-Pecollo, Fabrice H.F. Pierre, Marina Touillaud, Mathilde Touvier, Marie Paule Vasson, Mélanie Deschasaux, Julie Le Merdy, Emilie Barrandon, and Raphaëlle Ancellin. Alcoholic beverages, obesity, physical activity and other nutritional factors, and cancer risk: A review of the evidence, 2016.

[29] Monica S. Sierra, Patricia Cueva, Luis Eduardo Bravo, and David Forman. Stomach cancer burden in Central and South America. *Cancer Epidemiology*, 44:S62–S73, 2016.

[30] Chikara Kunisaki, Hiroaki Miyata, Hiroyuki Konno, Zenichiro Saze, Norimichi Hirahara, Hirotoshi Kikuchi, Go Wakabayashi, Mitsukazu Gotoh, and Masaki Mori. Modeling preoperative risk factors for potentially lethal morbidities using a nationwide Japanese web-based database of patients undergoing distal gastrectomy for gastric cancer. *Gastric Cancer*, 20(3):496–507, 2017.