# APPLYING AN UNSUPERVISED MACHINE LEARNING APPROACH TO DETECT DIETARY HABITS OF BREAST CANCER PATIENTS IN BANGLADESH

**Mst. Farzana Akter[1], Shahnaj Sultana Sathi[1], Ayesha Akter[2] and Mohammad Ohid Ullah[1]\***

[1]Department of Statistics, Shahjalal University of Science and Technology, Sylhet, Bangladesh
[2]Department of Statistics, Dhaka College, Dhaka, Bangladesh

## ARTICLE INFO

Corresponding Author:
Mohammad Ohid Ullah
Email: tohid-sta@sust.edu

## ABSTRACT

*Purpose:* The study aims to examine the association between food habits/lifestyle and breast cancer using an unsupervised machine approach.

*Method:* The dataset was collected from the hospitals of eight divisional cities in Bangladesh using a semi-structured questionnaire. Descriptive statistical tools and an unsupervised machine learning approach- Factor Analysis were used to analyze the data.

*Results:* The highest numbers of breast cancer patients were observed in the Sylhet division, followed by the Dhaka and Khulna divisions. It is noted that, overall, left breast cancer patients outnumber right breast cancer patients. We found that betel nuts, beverages, beef/mutton, etc. are high commonalities, which indicates that these food habits are highly associated with breast cancer. Moreover, most of the patients can't bear the cost of treatment.

*Conclusions:* It is concluded that most breast cancer patients are used to taking betel nuts and beverages that may cause this disease. Therefore, we should avoid unhealthy and junk foods.

## INTRODUCTION

Breast cancer is one of the most frequent cancers occurring among women and food habits influence 35% of cancer cases (Kotepui, 2016). The incidence rate of breast cancer is increasing gradually and also the mortality rate and with this huge amount of economic loss per year worldwide. The rate of breast cancer cases is higher in developed countries, but the rate of survival is lower in developing countries and this condition is because of different environmental, economic and lifestyle-related factors (Taghavi et al., 2012). In Bangladesh, from 2005 to 2010, 5255 breast cancer cases were detected as stated by the National Institute of Cancer Research and Hospital (NICRH) and 56% of breast cancer patients were from the reproductive age group (Hossain et al., 2014). Cancer is the most severe disease worldwide nowadays and its incidence rate is increasing day by day as a result, the prediction that cases of cancer would be doubled in the immediate future will become tangible. According to Bangladesh Maternal Mortality and Health Care Survey, 2016, cancer are responsible for 24% of total deaths of women aged 15-49 years which is more than the percentage found in the

report of the same survey of 2010 and According to WHO's (World Health Organization) International Agency for Research on Cancer (IARC). Breast cancer is one of the most occurring cancers among Bangladeshi women (Bangladesh Maternal Mortality and Health Care Survey, 2016).

Breast cancer is one of the most common cancers and the number of breast cancer cases is high in all countries, especially in developed countries. Although the incidence rate of breast cancer is high in developed countries, deaths due to this are high in developing countries, and WHO reported in 2015 that the Breast cancer incidence rate was 19.4 in East Africa while 89.7 in West Europe per 100,000 people (Ghoncheh et al., 2016). Nowadays medical sector has developed through the improvement of technology and the number of years a person is expected to live is increased because of modern medical systems. Increasing life longevity is highly associated with breast cancers outbreak rates and its mortality rates (Gu, X. et al.,2018). The frequency of occurring breast cancer was almost 23 per 100000 women in Bangladesh (Uddin et al., 2013). Indeed, the specific cause behind any cancer is not known but some possible factors that influence cancers are discovered by many studies. These risk factors are different for different areas and are changed concerning the lifestyle of various countries.

Women must know the way of reducing the risk of breast cancer and women must know the factors which are related to this. Some of those factors are lifestyle-related and some factors cannot be changed. The risk factors are early menstruation and late menopause (Khatib & Modjtabai,2006), Diet (Kotepui,2016), Family history of breast cancer (Collaborative Group on Hormonal Factors in Breast Cancer,2001), Infertility (Cetin et al., 2008; Meggiorini et al., 2012), Being overweight/obese (Picon-Ruiz M et al., 2017), Smoking and alcohol consumption (Jones ME et al., 2017), Not breastfeeding (Stuebe A,2009), Trauma (Rigby JE et al.,2002), Not physically (McTiernan A, 2003). It reduces the chance of affecting weight, hormones, the balance of the system, etc.

Research in Iceland on 2139 breast cancer cases found that the risk of developing breast cancer was high for the left breast and also higher for those who had a first-degree relative with cancer compared with those who had no such relative (Tulinius H et al.,1990). A case-control study in India (Kaushal M et al., 2010) was performed to find the role of four environmental factors (smoking and chewing of tobacco, alcohol, and betel quid chewing) and Polymorphisms in enzymes in the risk of breast cancer. From the analysis of the data of 117 cancer cases and 174 cases without cancer, betel-quid chewing was marked as the main risk factor, and women who chewed betel-quid had five times more risk of developing breast cancer. So far we realized most of the studies in Bangladesh were only focused to obtain the percentages of different factors that influenced breast cancer. Therefore, we are intended to know the food habits or lifestyle-related factors that are associated with breast cancer in Bangladesh.

## METHODOLOGY

To know the overall situation of breast cancer patients in Bangladesh, we used data of only breast cancer patients from the dataset that was collected for a cross-sectional study titled "Extent of lifestyle and heritability effects on cancer in Bangladesh" in January 2019. The study was funded by the University Research Centre, Shahjalal University of Science and Technology (SUST) and approved by its ethical committee. The study randomly collected information on cancer patients (n = 384) from public/private hospitals in eight divisions of Bangladesh. A structured questionnaire was used, which included patients' personal

information, personal cancer-related information, food habits, and family history of cancer, cost-related information, and lifestyle to collect information from cancer patients.

To identify the lifestyle factors associated with breast cancer, we extracted the data of 73 breast cancer patients from a total of 384 cancer patients. Descriptive statistical tools and an unsupervised machine learning approach- Factor Analysis, were applied to the data. We used the SPSS program for this study.

## RESULTS AND DISCUSSION

The mean age of the study patients was 44.47 (±10.29) years. Among 73 breast cancer patients, 87.67% of them were housewives, 84.93% had less than 10 years of formal education and 60.27% of them had cancer on their left breast. A total of 78.08% and 79.45% of breast cancer patients had no family history of cancer and no other disease before cancer respectively while 71.23% of them couldn't bear the cost of their cancer treatment.
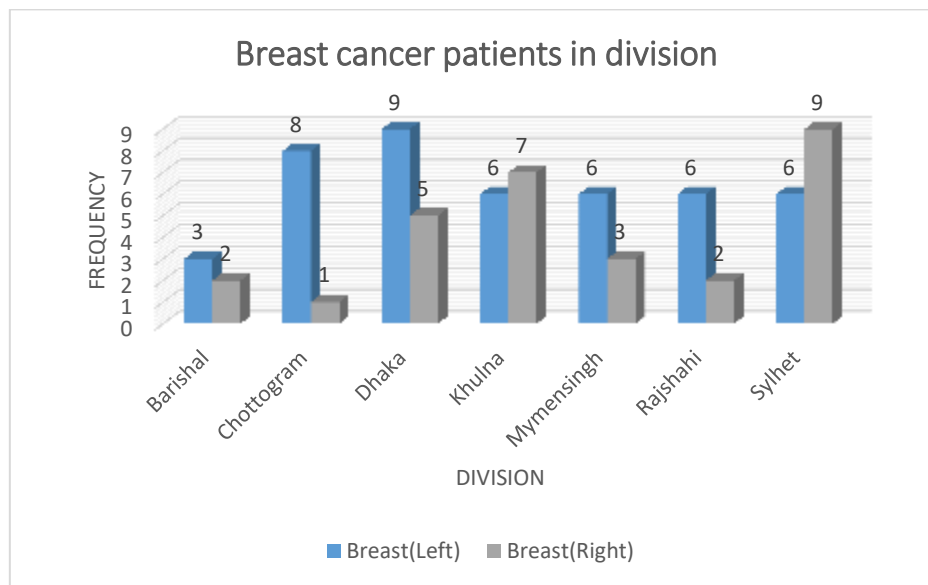


**Figure1**: Number of Breast Cancer patients in each Division

The number of breast cancer patients in the Sylhet division is the highest followed by Dhaka and Khulna divisions. It is also illustrated that in hospitals of Khulna and Sylhet the number of right breast cancer patients was higher than the left breast cancer while in Chattogram, Dhaka, Mymensingh, Rajshahi and Barishal the number of Left breast cancer patients were higher [**Fig. 1**]. After applying Factor Analysis (FA) on the variables of Lifestyle/Food habits, we observed that three components explained almost 70% of the total variation.
The association among the dietary habits of the patients is depicted in the following loading plot, or component plot (for component 1 and component 2).
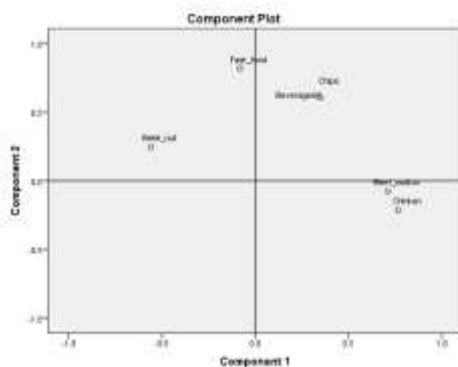
**Figure 2**: Loading plot for component 1 and component 2.

It was observed that patients who are consuming beef/mutton also consume chicken. Similarly, those who are used to taking fast food also take chips/chocolate/ice cream and beverages. This indicates that beef/mutton are associated with chicken, and Fast food is highly associated with Chips and Beverages. In Factor Analysis (FA) or Principal Component Analysis (PCA), commonality indicates the proportion of common variance found in a particular variable and its value lies between 0 and 1. The communalities of each variable of the dietary habits were presented in **Table 1**.

**Table 1: Communalities of different variables of lifestyle/Food habits**

| Dietary habits/Lifestyle | Communalities | Rank of dietary habits/lifestyle |
|---|---|---|
| Betel-nuts | 0.815 | 1st |
| Beverages | 0.752 | 2nd |
| Beef/mutton | 0.721 | 3rd |
| Fast food | 0.682 | 4th |
| Chicken | 0.669 | 5th |
| Chips/Chocolate/ice-cream | 0.552 | 6th |

From the communalities of each variable, we found that among the variables of "Lifestyle/Food habits" betel-nuts explained 81% of the variation, followed by beverage (75%), beef/mutton (72%), and Fast food (68%). It was indicated that breast cancer patients are used to taking betel-nuts and bevarges.

Breast cancer is one of the common cancer diseases occurring worldwide and it is at number two in the list of cancer developing among people. In 2018, over 2 million new cases of breast cancer were found worldwide. Among the top 25 countries having the highest rates of breast cancer, Belgium was in first place followed by Luxembourg and Netherlands (Bray F et al., 2018; Breast cancer statistics, 2018). A research study on breast cancer and cervical cancer in 187 countries provided the information that among women who died because of breast cancer of developing nations in 2010, 16% were from the reproductive age (15-49 years) (Forouzanfar MH et al.,2011). The rate of occurrence of breast cancer has been rising rapidly in Asian countries and the incidence rate reached its peak among Asian women at their forties while it peaked among postmenopausal women of the United States and Europe at their sixties(Leong SPL et al.,2010). In this study, we found that the age of women with breast cancer was from 34 to 54 years with an average of 44 years and it had similarities with the findings of other studies (Kuo WH et al., 2006). There is a myth that if you have first-degree relative with breast cancer you will develop breast cancer at some point in your life. But this is not true because the fact is that having a first or second-degree relative with breast

cancer only increases the risk of developing that cancer and most of them will never have breast cancer. Total 89% of women having breast cancer had no first-degree relative with breast cancer and below 10% of women with breast cancer may be affected by alterations in genes (Collaborative Group on Hormonal Factors in Breast Cancer,2001; Joy JE et al.,2005). Our findings were similar to those studies as about 78% of our study patients had no relatives with any type of cancer.

A study on Turkish women found that the risk of developing breast cancer was higher in working women than in housewives (Yilmaz M et al., 2011) but our findings conflicted with this outcome. We found that almost 88% of breast cancer patients were housewives. In Bangladesh, most of the women who are housewives are either illiterate or have less than 10 years of formal education and most of the patients cannot bear the treatment cost. Having proper education can help one to decrease the risk of having breast cancer although there might be many other factors because of which the incidence rate of breast cancer is high for housewives (Morton WE,1995). Recently a fact about breast cancer has been found which is that breast cancer mostly happens to left breast than the right one according to some unilateral breast cancer cases and among 2011 one-sided breast cancer cases, 53% of them were left-breast cancer patients (Tulinius H et al., 1990). Our study also depicted the same scenario as among 73 breast cancer patients 60% of them had left breast cancer. The left breast is more likely to have cancer than the right breast and also the left side of the body is more susceptible to have another type of cancer although the reasons behind this are yet unknown. One possible theory is that the left breast tends to be bigger than the right breast and so cancer gets more tissue to be developed in the left breast. A study in India on environmental factors responsible for breast cancer found that chewing betel quid increased the chance of having breast cancer by five times (Kaushal M et al., 2010).

According to the report of the World Health Organization, 14,900 cases of breast cancer are found in Bangladesh and 48% of them died annually (World Health Organization, 2019). Unfortunately, the rate of new cases of breast cancer increases day by day and so is the number of deaths in Bangladesh. To reduce the number of new breast cancer cases the fundamental knowledge about risk factors of breast cancer, its early warning signs, different types of screening approaches, and therapeutic procedures applied to someone after being diagnosed with breast cancer is necessary among women aged 15-49 years (Akter & Ullah, 2022). It has been discovered that obesity is related to the higher risk of deaths with breast carcinoma among women of the premenopausal and postmenopausal period and physical exercise reduced the chance of having carcinoma (Picon-Ruiz M et al., 2017). On the other hand, we all know that a healthy diet can improve the health condition of a person, but an unhealthy and irregular diet plan can disturb the regular process of the body as well as develop different kinds of diseases in the body (Sierra-Johnson J et al., 2008; Pot GK et al., 2014; Wennberg M et al., 2016). In Bangladesh, there is a lack of knowledge about diet among people and young people are following irregular diets just only to reduce weight but they do not have proper knowledge about it.

## CONCLUSION

The study aimed to find the dietary habits that are associated with breast cancer in Bangladesh. Based on the data analysis, it is concluded that betel nuts and beverages are the most frequent dietary habits of breast cancer patients. It is recommended that avoiding junk foods may reduce the risk of breast cancer and its associated diseases. Every woman should have proper knowledge about breast cancer, as the incidence rate is growing steadily. Women having proper knowledge about risk factors and early warning signs of breast cancer can detect breast cancer early, and early detection may possibly increase the chance of survival

and lead to cost-effective treatment. The study is limited in terms of sample size. Further studies can focus on other statistical tools with larger sample sizes.

## ACKNOWLEDGEMENTS

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## FUNDING

## REFERENCES

Akter MF, Ullah MO (2022). Awareness levels of breast cancer among female university and medical college students in Sylhet city of Bangladesh. *Cancer Reports*;e1608. https://doi.org/10.1002/cnr2.1608

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal For Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Cetin, I., Cozzi, V., & Antonazzo, P. (2008). Infertility as a cancer risk factor - a review. *Placenta*, *29 Suppl B*, 169–177. https://doi.org/10.1016/j.placenta.2008.08.007

Collaborative Group on Hormonal Factors in Breast Cancer. (2001). Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet (London, England)*, *358*(9291), 1389-99. https://doi.org/10.1016/S0140-6736(01)06524-2

Forouzanfar, M. H., Foreman, K. J., Delossantos, A. M., Lozano, R., Lopez, A. D., Murray, C. J., & Naghavi, M. (2011). Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. *Lancet (London, England)*, *378*(9801), 1461–1484. https://doi.org/10.1016/S0140-6736(11)61351-2

Gu, X. et al. (2018). Interactions between life expectancy and the incidence and mortality rates of cancer in China: a population-based cluster analysis. *Cancer communications,* 38(1), 44. https://doi.org/10.1186/s40880-018-0308-x.

Ghoncheh, M., Pournamdar, Z., & Salehiniya, H. (2016). Incidence and Mortality and Epidemiology of Breast Cancer in the World. *Asian Pacific Journal Of Cancer Prevention : APJCP*, *17*(S3), 43–46. https://doi.org/10.7314/apjcp.2016.17.s3.43

Hossain, M. S., Ferdous, S., & Karim-Kos, H. E. (2014). Breast cancer in South Asia: a Bangladeshi perspective. *Cancer Epidemiology*, *38*(5), 465–470. https://doi.org/10.1016/j.canep.2014.08.004

Jones, M. E., Schoemaker, M. J., Wright, L. B., Ashworth, A., & Swerdlow, A. J. (2017). Smoking and risk of breast cancer in the Generations Study cohort. *Breast cancer Research : BCR*, *19*(1), 118. https://doi.org/10.1186/s13058-017-0908-4

Joy JE et al. (2005). Understanding Breast Cancer Risk. National Academies Press (US). Kotepui

Kotepui M. (2016). Diet and risk of breast cancer. *Contemporary oncology (Poznan, Poland)*, *20*(1), 13–19. https://doi.org/10.5114/wo.2014.40560

Kaushal, M., Mishra, A. K., Raju, B. S., Ihsan, R., Chakraborty, A., Sharma, J., Zomawia, E., Verma, Y., Kataki, A., Kapur, S., & Saxena, S. (2010). Betel quid chewing as an environmental risk factor for breast cancer. *Mutation research*, *703*(2), 143–148. https://doi.org/10.1016/j.mrgentox.2010.08.011

Kuo, W. H., Yen, A. M., Lee, P. H., Hou, M. F., Chen, S. C., Chen, K. M., Chen, T. H., & Chang, K. J. (2006). Incidence and risk factors associated with bilateral breast cancer in area with early age diagnosis but low incidence of primary breast cancer: analysis of 10-year longitudinal cohort in Taiwan. *Breast cancer research and treatment*, *99*(2), 221–228. https://doi.org/10.1007/s10549-006-9194-z

Leong, S. P., Shen, Z. Z., Liu, T. J., Agarwal, G., Tajima, T., Paik, N. S., Sandelin, K., Derossis, A., Cody, H., & Foulkes, W. D. (2010). Is breast cancer the same disease in Asian and Western countries?. *World journal of surgery*, *34*(10), 2308–2324. https://doi.org/10.1007/s00268-010-0683-1

Meggiorini, M. L., Cipolla, V., Rech, F., Labi, L., Vestri, A., & de Felice, C. (2012). Mammographic features in infertile women as a potential risk for breast cancer: a preliminary study. *European journal of gynaecological oncology*, *33*(1), 51–55.

McTiernan A. (2003). Behavioral risk factors in breast cancer: can risk be modified?. *The oncologist*, *8*(4), 326–334. https://doi.org/10.1634/theoncologist.8-4-326

Morton W. E. (1995). Major differences in breast cancer risks among occupations. *Journal of occupational and environmental medicine*, *37*(3), 328–335. https://doi.org/10.1097/00043764-199503000-00010

Picon-Ruiz, M., Morata-Tarifa, C., Valle-Goffin, J. J., Friedman, E. R., & Slingerland, J. M. (2017). Obesity and adverse breast cancer risk and outcome: Mechanistic insights and strategies for intervention. *CA: a cancer journal for clinicians*, *67*(5), 378–397. https://doi.org/10.3322/caac.21405

Pot, G. K., Hardy, R., & Stephen, A. M. (2014). Irregular consumption of energy intake in meals is associated with a higher cardiometabolic risk in adults of a British birth cohort. *International journal of obesity (2005)*, *38*(12), 1518–1524. https://doi.org/10.1038/ijo.2014.51

Rigby, J. E., Morris, J. A., Lavelle, J., Stewart, M., & Gatrell, A. C. (2002). Can physical trauma cause breast cancer?. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP)*, *11*(3), 307–311. https://doi.org/10.1097/00008469-200206000-00014

Stuebe A. (2009). The risks of not breastfeeding for mothers and infants. *Reviews in obstetrics & gynecology*, *2*(4), 222–231. http://www.ncbi.nlm.nih.gov/pmc/articles/pmc2812877/

Sierra-Johnson, J., Undén, A. L., Linestrand, M., Rosell, M., Sjogren, P., Kolak, M., De Faire, U., Fisher, R. M., & Hellénius, M. L. (2008). Eating meals irregularly: a novel environmental risk factor for the metabolic syndrome. *Obesity (Silver Spring, Md.)*, *16*(6), 1302–1307. https://doi.org/10.1038/oby.2008.203

Taghavi, A., Fazeli, Z., Vahedi, M., Baghestani, A. R., Pourhoseingholi, A., Barzegar, F., & Pourhoseingholi, M. A. (2012). Increased trend of breast cancer mortality in Iran. *Asian Pacific journal of cancer prevention : APJCP*, *13*(1), 367–370. https://doi.org/10.7314/apjcp.2012.13.1.367

Tulinius, H., Sigvaldason, H., & Olafsdóttir, G. (1990). Left and right sided breast

cancer. *Pathology, research and practice*, *186*(1), 92–94. https://doi.org/10.1016/S0344-0338(11)81015-0

Uddin, A. F., Khan, Z. J., Islam, J., & Mahmud, A. (2013). Cancer care scenario in Bangladesh. *South Asian journal of cancer*, *2*(2), 102–104. https://doi.org/10.4103/2278-330X.110510

Wennberg, M., Gustafsson, P. E., Wennberg, P., & Hammarström, A. (2016). Irregular eating of meals in adolescence and the metabolic syndrome in adulthood: results from a 27-year prospective cohort. *Public health nutrition*, *19*(4), 667–673. https://doi.org/10.1017/S1368980015001445

Yilmaz, M., Guler, G., Bekar, M., & Guler, N. (2011). Risk of breast cancer, health beliefs and screening behaviour among Turkish academic women and housewives. *Asian Pacific journal of cancer prevention : APJCP*, *12*(3), 817–822. https://doi.org/10.7314/APJCP.2012.13.11.5823