OPEN ACCESS Freely available online

International Journal of Biomedical Data Mining

ISSN: 2090-4924

Research Article

Potential Risk Factors of Blood Cancer and Association of Significant Factors: A Survey in Bangladesh Using Data Mining Techniques

Farzana Tasnim[1*], Tasniya Ahmed[2], Kawsar Ahmed[3], Fazlul Karim Patwary[4], Karam Newaz[1]

[1]Department of Information Technology, Gono Bishwabidyalay University, Dhaka, Bangladesh; [2]Department of Information Technology, Noakhali Science & Technology University, Nokhali, Bangladesh; [3]Department of Information & Communication Technology, Mawlana Bhasani Science and Technology University, Tangail, Bangladesh; [4]Department of Bioinformatics, Jahangirnagar University, Savar, Dhaka, Bangladesh

ABSTRACT

In recent years, cancer is one of the main causes of death worldwide and in Bangladesh. Therefore, the research interest is associated with blood cancer patients as only a few blood cancer studies are available in Bangladesh. Initially, we collect 340 patient data (blood cancer and non-blood cancer) from BSMMU hospital. Then we implement data mining techniques to rank 30 factors (conducted with questionnaires) that are frequently related to blood cancer. In this research, we use Data mining approaches such as classification, chi-square (χ^2) test, P-value, and association rule mining. This study finds the predictive role of 15 factors among 30 input factors, among them muscle pull, inability to control the bladder, unusual bleeding, fever/raised temperature, weakness usually in legs identified as most potent predictors (p-value<0.001). Subsequently, an association among these significant elements is anticipated using the most popular rule mining algorithm such as Apriori, Predictive Apriori, and Tertius. The experimental result shows that weakness usually in legs=yes, fever/raised temperature=yes, and not being able to control bladder=no is a common rule for blood cancer=yes. Again, unusual bleeding=yes, rapidly becoming more ill=yes and, muscle pull=yes rule is likely to have a significant association. Besides, a frequent pattern of having a fair skin tone with fast breathing and weakness in the legs might be a threat.

Keywords: Blood cancer; Data mining techniques; Association rule mining; Healthcare

INTRODUCTION

Nowadays the most common deathful disease is cancer. In Bangladesh, there is no population based cancer registry. As a result, the real condition of cancer is generally anonymous here. Consistent with World Health Organization, Bangladesh is facing a growing quantity of the latest cancer cases. The number of the latest subjects is projected to grow by way of approximately seventy seven % in 2030. WHO predicts that the wide variety of blood-associated most cancers instances could increase utilizing about forty eight % in much less developed countries with the aid of 2030 in comparison to 2012 [1].

One of the most not unusual causes of demise is cancer, and blood cancer is 1/3 among them. Blood cancers affect the production and feature of our blood cells. Many factors are influencing the occurrence of Blood cancer. These are known as risk factors. This research describes the correlation between significant patterns and risk levels of various factors that improve blood cancer prognosis. It aims to analyze multiple factors associated with blood cancer and determine the risk prediction model of blood cancer. In this research, data mining and statistical analysis take place. The data processing techniques are popular to search hidden relationships and frequent item sets for the prediction of various diseases from the medical field. A limited number of publications exists that address blood cancer using the data mining technique. Some of these are the following:

Correspondence to: Farzana Tasnim, Department of Information Technology, Gono Bishwabidyalay University, Dhaka, Bangladesh, Tel: 8801765524232; E-mail: rekha.tasnim@gmail.com

Received date: December 13, 2021; **Accepted date:** December 27, 2021; **Published date:** January 03, 2022

Citation: Tasnim F, Ahmed T, Ahmed K, Patwary MFK, Newaz MK (2022) Potential Risk Factors of Blood Cancer and Association of Significant Factors: A Survey in Bangladesh Using Data Mining Techniques. Int J Biomed Data Min. 11:446.

Applied data processing techniques to get the relations between biopsy characteristics and blood tumors to predict the disease in an early stage, the final results showed that association rules could give the relationship between blood test characteristics and blood tumors. Also, it demonstrated the simplest ability to predict tumor sorts of blood diseases with an accuracy of 79.45%. Proposed a work that ambition to survey distinct computer-aided strategies to phase the blood smear image they have used six hundred images in the experimentation. The model can distinguish between a routine tangential blood coat and an irregular blood coat. Some papers predicted the *leukemia* [2] existence by determining the relationships of blood properties and *leukemia* with gender, age and health status of patients using data mining techniques. They noticed that the DT classifier obtains properties regarding outer attributes such as a city (eastern regions) that are most vulnerable to *leukemia*.

There are many cases where data processing techniques are being applied for the diagnosis of various diseases like heart conditions, diabetes, cancer, etc.

Many people of Bangladesh do not even know that they are affected by blood cancer, and in most cases, patients get treatment at the final stage when heal is not feasible. Therefore, early prediction of blood cancer plays a crucial position in the diagnosis technique and a prevention approach [3].

Most of the previous studies did not specify risk attributes and access association among risk factors and blood cancer. From that context, the purpose of this research is to find out which attributes are an extra concern for predicting blood cancer by evaluating the p-value. Also, discover association rules among attributes that assist in causing blood cancer [4].

In this research, we have used data processing tool SPSS and WEKA 3.8. Physicians and patients can use this developed system to quickly know a person's cancer status prior to screening them for testing cancer.

The rest of the paper is organized as follows: Section 2 describes the detail about the materials and methods for conducting this research. The next section covers experimental results and discussion. Section 4 presents conclusions and future scope. The Acknowledgement and references are presented at the end of this paper.

MATERIALS AND METHODS

This study is an on-going multi-center population-based study to investigate the relationship among factors causing blood cancer [5]. The workflow of this paper is given in Figure 1. Extensive related works, case studies, and discussions with medical experts and hematologists show that there are several factors influencing Blood cancer [6]. These factors are recognized and taken as attributes for this study.

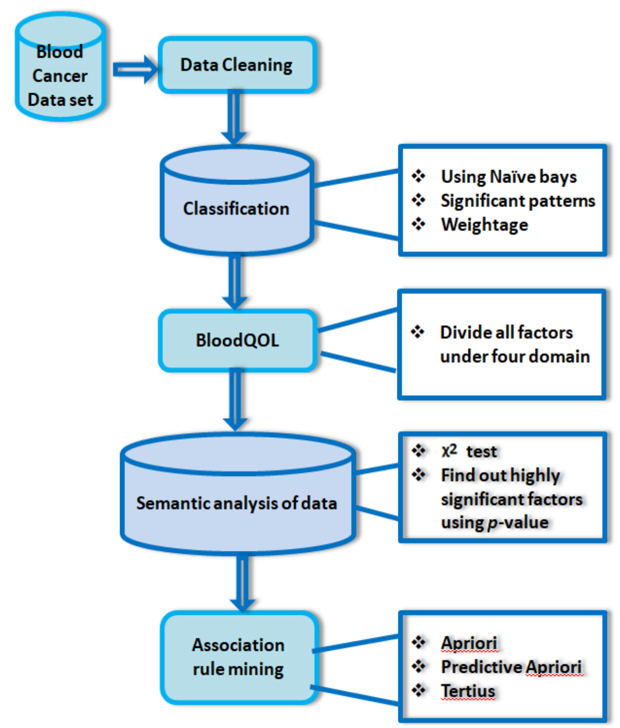

Figure 1: Work flow of this research.

Data Source

The data for this study was collected from the patient admitted to the National Cancer Institute, Bangladesh, and BSMMU. In recorded data, the participants were both blood cancer and non-blood cancer patient. The data consist of more than 30 attributes occupied from previous studies.

Dataset description

Attributes of Blood cancer we have used in this research were summarized following different research work. The questionnaire sample is provided in a supplementary file [7]. A summarized description of the attributes of the blood cancer data set is given in Table 1.

Table 1: Detailed description of blood cancer data set.

Sl.no	Attribute	Type	Values of attribute
1	Age	Numeric	10-70 year
2	Gender	Nominal	Male, Female
3	Living place	Nominal	Town, Village

4	Weight	Numeric	40-70 kg
5	Height	Numeric	152-177 cm
6	Being Overweight	Nominal	Yes, No Based on BMI
7	Skin color	Nominal	Fair, Medium, Dark
8	Occupation	Nominal	Student, Business,
			Teacher, Housewife
			Shopkeeper
			Job holder, Farmer
9	Job Environment	Nominal	Sunny,
			Shadow, Both
10	Family History	Nominal	Yes, No
11	Smoking	Nominal	Yes, No
12	Bone/Joint pain	Nominal	Yes, No
13	Pain when move/cough	Nominal	Yes, No
14	Difficulty in passing urine	Nominal	Yes, No
15	Weakness usually in legs	Nominal	Yes, No
16	Fever/raised temperature	Nominal	Yes, No
17	Unexpected weight loss	Nominal	Yes, No
18	Unusual Bleeding	Nominal	Yes, No
19	Fast breathing	Nominal	Yes, No
20	Stressed behavior	Nominal	Yes, No
21	Rapidly becoming more ill	Nominal	Yes, No
22	Night Sweats	Nominal	Yes, No
23	Not being able to control bladder	Nominal	Yes, No
24	Muscle pull	Nominal	Yes, No
25	Mouth sores/	Nominal	Yes, No
	gum edema		
26	Type 2 diabetes	Nominal	Yes, No
27	Eating vegetable	Numeric	No of meal per week
28	Previous cancer	Nominal	Yes, No
	treatment		
29	Blood cancer	class attribute	Yes, No

Data cleaning

Sometimes, collected data from different places contain double or more information of the same person or missing any input values. That may produce a wrong investigation result. So data cleaning is a fundamental step to make a proper analysis of collected records. It decreases the memory and normalizes the values that signify information in the form. In our study, we discard 2.9% missing values and 2.98% duplicate values [8].

Classification

Naive Bayes is the most effective model to predict patients with cancer disease. Naive Bayes classifier is a widely used framework for classification based on a simple theorem of probability known as Bayes theorem [9]. The Naive Bayes algorithm assesses the chances of the frequencies and combinations of values of a given data set. The probability with vector:

$x = x1, x2, xn$ is

$$P(h1|xi) = P(xi|h1).P(h1) / \{P(xi|h1).P(h1) + P(xi|h2).P(h2)\text{....} \quad (1)$$

Here, $P(h1|xi)$is a subsequent probability, while $P(h1)$ is the prior probability linked with hypothesis $h1$.

$$P\,(xi) \;=\; \sum_{k=0}^{n} P\,(xi|hi)\,P\,(hi) \quad \text{..................} \quad (2)$$

Thus,

$$P(h1|xi) = P(xi|h1).P(h1) /\; P\,(xi) \quad \text{.............} \quad (3)$$

The following form is used to discover the major regular pattern.

$$s_{\omega(i)} = \sum_{i=0}^{n} m * t * n \quad \text{..............................} \quad (4)$$

$s_{\omega(i)}$ Is a Significant Frequent Pattern; m be the number of data, t be the number of attributes, and n be the number of times occurrences of value.

We have used Naive Bayes classification because it is easy to use, and it requires a little quantity of training data to approximate the parameters (means and variances of the variables) needed for categorization.

The classification result shows that the instances are classified correctly, and there is no incorrectly classified instance.

Now we can go for further processing.

Blood QOL

According to the recent statistical analysis trend named Quality of Life, we categorize data into many classes where every category has exclusive features.

Through the Blood specific Quality of Life (Blood QOL) technique (consist of 31 factors), the data set obtained from the classification is categorized into four domains: personal history domain, Habit domain, family history domain, and disease domain [10]. The factors under these domains are mentioned in Table 2.

Table 2: Socio-demographic quality of life.

Domain	Attribute
Personal History	Name, Age, Gender, height, weight, Being over-weight, Living place, skin color, occupation, job environment
Family History	Family History
Habits	Smoking, Eating vegetable
Disease	Bone/Joint pain, pain when move or cough, Difficulty in passing urine, Previous cancer treatment, Night sweats, weakness usually in legs, Unexplained weight loss, Fever or raised temperature, Unusual bleeding, Fast breathing, Stressed behavior, Rapidly becoming more ill, Night sweats, Not being able to control the bladder, Muscle pull, Mouth sores, Type 2 diabetes.

Significant patterns find out

The significance of each attribute under these domains shown in Table 2 is calculated using Equation (5).

Blood QOL (%)=Response of individual Factor/Total Response of BloodQOL Factors *100% (5).

From each domain, we have deleted the attributes whose BloodQOL (%) result was low. The deleted attributes were Name, Height, Weight, Eating vegetables, previous cancer treatment, and Night sweats. The remaining attributes were considered significant.

Find out highly significant data

To discover the factors of blood cancers that are highly significant is the objective of this analysis. The P-value of various attributes has been calculated using the χ^2 test formula. A p-value of smaller than .005 is measured as highly significant [11].

This analysis has been implemented by SPSS version 16.0 using the following formula (6).

$$\chi 2 = \Sigma(O - E)^2 \,/\, E \ldots\ldots\ldots\ldots \quad (6)$$

Here O =Observed occurrence and E = Expected occurrence

Association rule mining

In recent years, one of the very popular data mining techniques is the association rule, which is applied to discover the frequent pattern among different attributes in medical data. This research uses three trendy algorithms to derive frequent itemsets: Apriori, Predictive Apriori, and Tertius. For the Apriori algorithm we evaluated rules based on confidence, for the Predictive algorithm we considered accuracy and confirmation for the Tertius algorithm [12].

EXPERIMENTAL RESULT AND DISCUSSION

Most Significant Attributes

To calculate the correlation among QOL factors, we use the Chi-square (χ^2) test on the class attribute. Tables 3, 4, 5, & 6 represent the frequency distribution of the attributes. An attribute with p-value<=0.005 can be considered as significant.

The sign* indicates the factors are statistically significant. Table 3 shows that in the personal domain, the job environment factor has the highest impact in Blood QOL as it has the smallest p-value (.001); skin color has second-highest, and then age, occupation, respectively. Since the p-value of the remaining attributes of Table 3 is more extensive than 0.005, those variables are considered non-significant [13].

Table 3: P-value of personal domain risk factor attributes.

Attribute		Blood Cancer status		P-value
		Affected N (%)	Unaffected N (%)	
Age	10-25	35(21.9)	40(25)	0.003*
	26-40	90(56.2)	35(21.9)	
	41-55	30(18.8)	25(15.6)	
	Above 55	5(3.1)	60(37.5)	
Gender	Male	75(46.9)	110(68.8)	0.064
	Female	85(53.1)	50(31.2)	
Living Place	Town	85(53.1)	115(71.9)	0.098
	Village	75(46.9)	45(28.1)	
Being Over Weight	No	140(87.5)	150(98)	0.057
	Yes	20(12.5)	10(2)	
Skin Color	Fair	110(68.8)	50(31.2)	0.002*
	Medium	40(25)	45(28.1)	
	Dark	10(6.2)	65(40.6)	
Occupation	Student	30(18.8)	20(12.5)	0.005*
	Business	25(15.6)	20(12.5)	
	Teacher	15(9.4)	5(3.1)	
	Shopkeeper	10(6.2)	30(18.8)	
	Housewife	55(34.4)	5(3.1)	
	Job holder	20(12.5)	40(25)	

	Farmer	5(3.1)	40(25)	
Job Environment	Sunny	30(18.8)	60(37.5)	0.001*
	Shadow	75(46.9)	100(62.5)	
	Both Sunny and Shadow	55(34.4)	0(0.0)	

As for the disease domain that is indicated in Table 4, the most significant features were Bone/Joint pain (p-value .003), Weakness in legs (p-value<.001), Fever/raised temperature (p-value<.001), Unusual Bleeding (p-value<.001), Fast breathing (p-value =.005) rapidly becoming more ill (p-value=.002), Not being able to control the bladder (p-value<.001), and muscle pull (p-value<.001).

Table 4: P-value of "disease domain" risk factor attributes.

Factor		Cancer Status		P-value
		Affected	Unaffected	
		N (%)	N (%)	
Bone/Joint pain	No	110(68.8)	110(68.8)	0.003*
	Yes	10(6.2)	50(31.2)	
Pain when move/cough	No	100(62.5)	95(59.4)	0.5
	Yes	60(37.5)	65(40.6)	
Difficulty in passing urine	No	135(84.4)	140(87.5)	0.5
	Yes	25(15.6)	20(12.5)	
Weakness usually in legs	No	10(6.2)	80(50)	0.000*
	Yes	150(93.8)	80(50)	
Fever/raised temperature	No	20(12.5)	90(56.2)	0.000*
	Yes	140(87.5)	70(43.8)	
Unexpected weight loss	No	40(25.0)	75(46.9)	0.059
	Yes	120(75.0)	85(53.1)	
Unusual Bleeding	No	90(56.2)	160(100)	0.000*
	Yes	70(43.8)	0(0.0)	
Fast breathing	No	125(78.1)	120(43.8)	0.005*
	Yes	35(21.9)	90(56.2)	
Stressed behavior	No	90(56.2)	95(59.4)	0.5
	Yes	70(43.8)	65(40.6)	
Rapidly becoming more ill	No	75(46.9)	110(84.4)	0.002*
	Yes	85(53.1)	50(15.6)	

Night Sweats	No	75(46.9)	110(68.8)	0.064
	Yes	85(53.1)	50(31.2)	
Not being able to control bladder	No	155(96.9)	90(56.2)	0.000*
	Yes	5(3.1)	70(43.8)	
Muscle pull	No	75(46.9)	155(96.9)	0.000*
	Yes	85(53.1)	5(3.1)	
Mouth sores/gum edema	No	80(50.0)	125(78.1)	0.018
	Yes	80(50.0)	35(21.9)	
Type 2 diabetes	No	150(93.8)	125(78.1)	0.074
	Yes	10(6.2)	35(21.9)	

In the family history domain shown in Table 5, the only attribute named family history was found (p-value=.005) considerable.

Table 5: P-value of "history domain" risk factor attributes.

Attribute		Cancer Status		P-value
		Affected	Unaffected	
		N (%)	N (%)	
Family History	No	150(93.8)	130(87.5)	0.005*
	Yes	10(6.2)	20(12.5)	

This analysis result supported those attributes because the p-value was less than .005. The other elements from these factors were discarded.

From the habit domain, we found no important factor. So the result table was not shown in this paper [14].

Association rules among significant factors

This segment demonstrates the consequence of applying the association rule on blood cancer data set. At this point, we prepared a dataset consisting of only major attributes. And converted numerical attributes into nominal where needed. To find correlation among these noteworthy factors we applied three association rule algorithm i,e. Apriori, Predictive Apriori, and Tertius in WEKA 3.8 and found several rules for each algorithm. The peak rules were elected based on confidence, accuracy, and confirmation respectively. Here, terms are separated in the rules using the symbol'∪' and '∩' that stands for 'or' and 'and' correspondingly [15].

Table 6: Association rule mining for blood cancer by the apriori algorithm.

Serial	Rules	Result(Blood Cancer)	Confi dence
	Rules for class attribute Yes		
1	Weakness usually in legs=Yes ∩ Fever/raised temperature=Yes ∩ Not being able to control bladder=No	Yes	1

2	Family history=No ∩ Fever/raised temperature=Yes ∩ Fast breathing=No ∩ Not being able to control bladder=No	Yes	1
3	Bone/Joint pain=Yes ∩ Weakness usually in legs=Yes ∩ Not being able to control bladder=No	Yes	1
4	Skin color (Fair/Medium/Dark)=Fair ∩ Weakness usually in legs = Yes ∩ Fever/raised temperature = Yes	Yes	1
5	Age=Young ∩ Weakness usually in legs=Yes ∩ Not being able to control bladder=No	Yes	1
	Rules for class attribute No		
1	Bone/Joint pain=No ∩ Fever/raised temperature=No ∩Unusual bleeding=No ∩Muscle pull=No	No	1
2	Fever/raised temperature=No ∩ Unusual bleeding=No ∩ Muscle pull=No	No	1
3	Unusual bleeding=No ∩Fast breathing=Yes ∩ Rapidly becoming more ill=No	No	.95
4	Fast breathing=Yes ∩ Rapidly becoming more ill=No ∩Muscle pull=No	No	.95
5	Job environment (Sun/Shadow)=shadow ∩Unusual bleeding = No ∩Rapidly becoming more ill = No ∩ Muscle pull=No	No	.95

Apriori algorithm has been applied in weka 3.8 for the class attribute (yes and no) with minimum support of .25 and minimum confidence of 90%. The rules with a confidence level of at least 95% were chosen. Here, only the rules containing class attributes on the right-hand side (RHS) were reported. We found 6 yes rules with 100% confidence, also found 6 no rules, where 3 rules have 100% confidence and the other 3 rules have 95% confidence see Table 6. Then, we executed the Predictive Apriori algorithm and considered the association rules that only predict the class level [16]. We reported a total of 12 rules for this algorithm within the accuracy highest 99.44% and lowest 97.02% see Table 7.

Table 7: Association rule mining for blood cancer by the predictive apriori algorithm.

Serial	Rules	Result (Blood Cancer)	Accuracy %
	Rules for class attribute Yes		
1	Weakness usually in legs=Yes ∩ Fever/raised temperature=Yes ∩ Not being able to control bladder=No	Yes	99.44%

2	Bone/Joint pain=Yes ∩ Not being able to control bladder=No	Yes	99.42%
3	Age=Young ∩ Weakness usually in legs=Yes ∩ Not being able to control bladder=No	Yes	99.38%
4	Skin color(Fair/Medium/ Dark)=Fair ∩Weakness usually in legs=Yes ∩ Fast breathing=Yes	Yes	98.56%
5	Age=Young ∩ Skin color (Fair/ Medium/Dark)=Fair ∩ Weakness usually in legs=Yes	Yes	99.33%
	Rules for class attribute No		
1	Bone/Joint pain=No ∩ Unusual bleeding=No ∩ Rapidly becoming more ill=No	No	99.38%
2	Fever/raised temperature=No ∩ Unusual bleeding=No ∩ Fast breathing=No	No	99.33%
3	Age=Young ∩ Weakness usually in legs=Yes ∩ Not being able to control bladder=No	No	99.30%
4	Family history=No ∩ Weakness usually in legs=Yes ∩ Fever/raised temperature=No ∩ Unusual bleeding=No ∩Muscle pull=No	No	99.05%
5	Skin color (Fair/Medium/ Dark)=Fair ∩ Bone/Joint pain=No ∩ Fever/raised temperature=No ∩ Rapidly becoming more ill=Yes	No	98.56%

Finally, we implemented the Tertius association rule mining algorithm. This algorithm generated so many association rules. We kept rules holding class level on LHS or RHS and which are applicable for plenty of attributes based on confirmation. In this case, the highest confirmation was 84.94% see Table 8.

Table 8: Association rule mining for Blood cancer by the tertius algorithm.

Serial	Rules	Result (Blood Cancer)	confirmation
	Rules for class attribute Yes		
1	Weakness usually in legs=Yes ∩ Not being able to control bladder=No ∪ Age=Younger	Yes	83.08%
2	Unusual bleeding=Yes ∪ Rapidly becoming more ill=Yes ∪ Muscle pull=Yes	Yes	75.31%
3	Job environment (Sun / Shadow)=Both ∪ Unusual	Yes	75.08%

	bleeding=Yes ∪ Rapidly becoming more ill=Yes		
4	Occupation=Job ∪ Unusual bleeding=Yes ∪ Muscle pull=Yes		72.66%
5	Skin color(Fair/Medium/Dark)=Fair or Job environment (Sun/Shadow)=Both ∪ Unusual bleeding=Yes	Yes	63.64%
	Rules for class attribute No		
1	Skin color(Fair/Medium/Dark)=Dark ∪ Fever/raised temperature=No ∪ Not being able to control bladder=Yes	No	84.94%
2	Skin color (Fair/Medium/Dark)=Dark ∪ Fever/raised temperature=No ∪ Age=Old	No	81.59%
3	Weakness usually in legs=No ∪ Fever/raised temperature=No ∪ Not being able to control bladder=Yes	No	79.19%
4	Skin color(Fair/Medium/Dark)=Dark ∪ Weakness usually in legs=No ∪ Fever/raised temperature=No	No	72.04%
5	Job environment (Sun / Shadow)=Sun ∪ Weakness usually in legs=No ∪ Fever/raised temperature=No	No	66.06%

Discussion

Our first experiment was to find significant attributes based on the p-value. The result of this study indicates the predictive role of 13 factors amongst 31 input attributes in the model. Here, we find that always fever/raised temperature, muscle pull, not being able to control the bladder, unusual bleeding, weakness usually in legs has greater significance as their p-value is<0.001. Other important influence elements are job environment, skin color, rapidly becoming more ill, bone/joint pain, and age. The p-value of these is<=0.003 Remaining considerable attributes are family history, fast breathing, and, occupation. Top risk factors are represented in a bar chart in Figure 2.

Our next study was finding an association between these elements. By using three association rule mining algorithms we find the rules shown in Table 6, 7 and 8. From Tables we observe that, Apriori and Predictive Aprioi algorithm produces more accuracy for class=Yes. On the other hand Tertius algorithm gives more accurate result for class=no instead of class= yes. In general, we see almost the same rules from these three algorithms. However, weakness usually in legs=yes, fever/raised temperature=yes, and not being able to control bladder=no is a common rule for blood cancer=yes.

Again, unusual bleeding= yes, rapidly becoming more ill=yes and, muscle pull=yes rule is likely to have a significant association.

Besides, having a fair skin tone with fast breathing and weakness in legs might be a threat.

From Tables 6, 7 and 8 it is confirmed that people in the age group younger (10-25) and young (26-40) are at higher risk for blood cancer. In Table 5, we see family history as an inversely significant attribute.

Here, in association rules, we see that having no family history is correlated with both blood cancer yes and blood cancer no class [17].

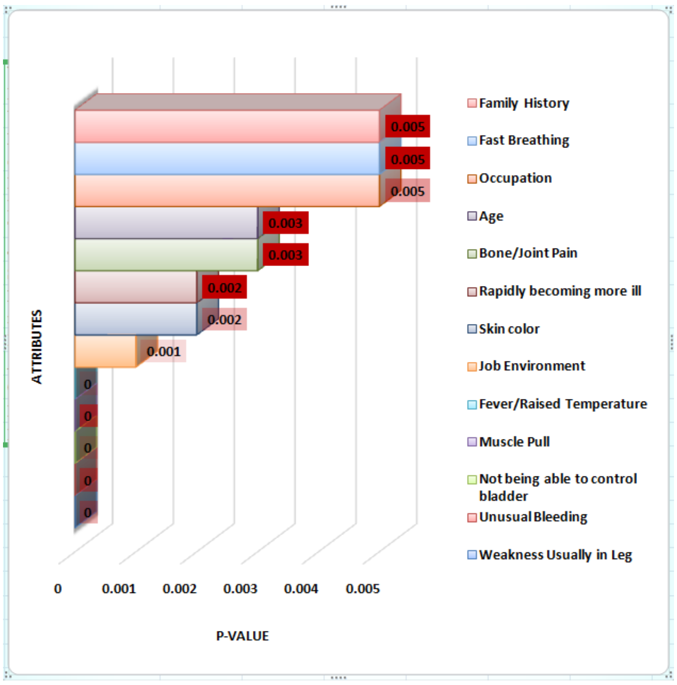

Figure 2: Bar chart of most significant attributes along with the p-value.

While comparing our outcome to older studies, we must point out that those studies find different classification algorithms' accuracy levels. In contrast, our study casts a new light on the significance of attributes. Also relation among those attributes. Some of the limitations of this study are that we conducted a face to face interview to collect data. It may include some wrong information by mistake. The frequency of data may change for large volumes of data.

CONCLUSION AND FUTURE SCOPE

In recent years, the quantity of blood cancer patients is growing at a rapid pace in Bangladesh. The most efficient way to reduce death by cancer is to detect it earlier. This paper has worked on blood cancer and different data mining tools and techniques which are becoming popular day by day. This paper projected a multi-layered data mining method combining The Naive Bayes classifier, BloodQOL, and association rule to guess blood cancer risk. It will be beneficial for the early detection of blood cancer, which is better than treatment and can diminish cancer passing. This aspect of the research suggested that age, skin color, occupation, job environment, bone/joint pain, weakness usually in legs, fever or raised temperature, unusual bleeding, fast breathing, rapidly becoming more ill, muscle pull, etc. are strongly linked with blood cancer. This research also suggests an association between these factors. We can use this finding to raise consciousness among natives about these factors that may cause blood cancer. It will be useful for fast prohibition and superior to treat.

Regardless, future research could continue to explore different machine learning feather selection methods to find out significant attributes.

REFERENCES

1. Acharya V, Kumar P. Detection of Acute Lymphoblastic Leukemia Using Image Segmentation and Data Mining Algorithms. Med Biol Eng Comput. 2019;57(8):1783-1811.
2. Ahmed K, Jahan P, Nadia I, Ahmed F. Assessment of Menopausal Symptoms among Early and Late Menopausal Midlife Bangladeshi Women and Their Impact on the Quality of Life. J Menopausal Med. 2016;22(SUPPL-1):39-46.
3. Bouckaert RR, Eibe Frank, Mark Hall, Richard Kirkby. WEKA Manual for Version 3-7-10. 2013.
4. Daqqa KA, Maghari AY, Al-Sarraj WF. Prediction and Diagnosis of Leukemia Using Classification Algorithms. In 2017 8th International Conference on Information Technology (ICIT), Amman, Jordan: IEEE. 2017;638-43.
5. Gultepe Y, Sabah Rashed. The Use of Data Mining Techniques in Heart Disease Prediction. Int J Comput Sci Mob Computing. 2019;8(4):136-41.
6. Hossain MS, Iqbal MS, Khan MA, Rabbani MG, Khatun H, Munira S, et al. Diagnosed Hematological Malignancies in Bangladesh-A Retrospective Analysis of over 5000 Cases from 10 Specialized Hospitals. BMC Cancer. 2014;14(1):1-7.
7. Hossain MS, Begum M, Mian MM, Ferdous S, Kabir S, Sarker HK, et al. Epidemiology of Childhood and Adolescent Cancer in Bangladesh, 2001-2014. BMC Cancer. 2016;16(1):1-8.
8. Kaur P, Pruthi Y, Bhatia V, Singh J. Empirical Analysis of Cervical and Breast Cancer Prediction Systems Using Classification. Int J Educ Manag Eng. 2019;9(3):1-15.
9. Layth M, Alkaragole Z, Kurnaz AS. Comparison of Data Mining Techniques for Predicting Diabetes or Prediabetes by Risk Factors. Int J Comput Sci Mob Computing. 2019;8(3):61-71.
10. Lee K, Kim SG, Kim D. "Potential Risk Factors for Haematological Cancers in Semiconductor Workers." Occup Med. 2015;65(7):585-589.
11. Li D, Yang D, Zhang J, Zhang X. AR-ANN: Incorporating Association Rule Mining in Artificial Neural Network for Thyroid Disease Knowledge Discovery and Diagnosis. IAENG Int J Comput Appl. 2020;47(1):25-36.
12. El-Halees AM, Shurrab AH. Blood Tumor Prediction Using Data Mining Techniques. Health Informatics Int J. 2017;6(2):23-30.
13. Nahar J, Tickle KS, Ali AS, Chen YP. Significant Cancer Prevention Factor Extraction: An Association Rule Discovery Approach. J Med Syst. 2011;35(3):353-367.
14. Raihan M, Mondal S, More A, Sagor MO, Sikder G, Majumder MA, et al. Smartphone Based Ischemic Heart Disease (Heart Attack) Risk Prediction Using Clinical Data and Data Mining Approaches, a Prototype Design. 19th International Conference on Computer and Information Technology, ICCIT 2016. 2017;299-303.
15. Samarakoon YM, Gunawardena NS, Pathirana A, Perera MN, Hewage SA. Prediction of Colorectal Cancer Risk among Adults in a Lower Middle-Income Country. J Gastrointest Oncol. 2019; 10(3): 445-452.
16. Sarker IH, Kayes ASM. ABC-RuleMiner: User Behavioral Rule-Based Machine Learning Method for Context-Aware Intelligent Services. J Netw Comput. 2020.
17. Satyam S, Gupta DL, Prasad BR. Comparative Study of Recent Trends on Cancer Disease Prediction Using Data Mining Techniques. Int J Database Theory Appl. 2016;9(9):107-118.
18. Stein CJ, Colditz GA. Modifiable Risk Factors for Cancer. Br J Cancer. 2004;90(2):299-303.