

# Smart Risk Prediction Tools of Appendicitis Patients: A Machine Learning Approach

Fuyad Al Masud <sup>1,2</sup> , Md. Rejaul Islam Royel <sup>1</sup> , Md. Mizanur Hasan Khan Sajal <sup>1</sup> , Sohely Jahan <sup>3</sup> , Bikash Kumar Paul <sup>1,2,4</sup> , Kawsar Ahmed <sup>2,4,\*</sup> 

<sup>1</sup> Department of Software Engineering, Daffodil International University, Sukrabad, Dhanmondi, Dhaka 1205, Bangladesh

<sup>2</sup> Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University (MBSTU), Santosh, Tangail, 1902, Bangladesh

<sup>3</sup> Department of Computer Science and Engineering, University of Barisal, Barishal 8200, Bangladesh

<sup>4</sup> Group of Bio-photomatiç, Mawlana Bhashani Science and Technology University (MBSTU), Santosh, Tangail 1902, Bangladesh

\* Correspondence: [kawsar.ict@mbstu.ac.bd](mailto:kawsar.ict@mbstu.ac.bd); [k.ahmed.bd@ieee.org](mailto:k.ahmed.bd@ieee.org); [kawsarit08050@gmail.com](mailto:kawsarit08050@gmail.com);

Scopus Author ID 34879354400

Received: 5.06.2020; Revised: 30.06.2020; Accepted: 1.07.2020; Published: 4.07.2020

**Abstract:** Appendicitis is a common disease or sickness that can cause serious complications. A person's appendix gets infected and painful due to appendicitis. In this study, an android based application has been developed by incorporating medical data received from the patient affected with appendicitis. A total of 200 subject's data, including case and control group, has been examined and correlated with the common risk factors like fever, fever runs, appetite, abdominal pain, pain qualification, vomiting, rate of nausea, migration pain clinical symptom, which may suggest strongly significant to have appendicitis. Feature selection technique (correlation, information gain, gain ratio, relief, and symmetrical uncertainty) has been used to figure out the best relevant features. A predictive Apriori algorithm has been applied to find out the best rules for appendicitis. From the best rules, a risk score table has been generated and developed a risk flowchart, which will correctly identify 99 patients among 100 affected patients between the risk levels of medium to very high. At long last, this flowchart has used to develop a risk prediction application. Finally, the developed "Predict Appendix" application will be helpful to predict the risk level of appendicitis not only among peoples of Bangladesh but also all over the world and, at the same time, increase awareness.

**Keywords:** Public Health; Data Mining; Machine Learning; Smart Tool; Appendicitis in Bangladesh.

© 2020 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Appendicitis is an inflammation of the appendix, 5-9 cm long, and projects from the colon on the lower right side of the human's abdomen [1]. The incidence rate of appendicitis is 9.4 among 10,000 populations [2]. The lifetime risk of appendicitis is respectively 8.6% for men and 6.7% for women, and median ages (25 years old) man are frequently affected by appendicitis [3]. In Bangladesh, it is the top 49th disease, and 1194 peoples are dead only for this surgical disease.

Without symptoms, an accurate diagnosis is totally difficult. To find those symptoms, imaging technology is used in the last 15 years. Nanotechnology is an emerging field of research [5, 6]. Also, some other technologies like abdominal Computed Tomography [CT] scan, Helical Computed Tomography [HCT], ultrasonography are used to find out the

sensitivity of appendicitis. But identifying the probability of disease using radiation is not effective because it cannot identify pre-operative symptoms of appendicitis when the patients are the child and pregnant women. Severe abdominal pain could be the main reason for acute appendicitis [7]. There are many significant studies were conducted with children and find out those who are 0 to 18 years old are highly affected by appendicitis [8]. In the United States (1993 to 2008) male patients are getting more diagnosed with appendicitis rather than female patients [9].

Early diagnosis of any surgical diseases will indeed give a good outcome. In recent years significant study will conduct on the pre-operative treatment of appendicitis. Different diagnosis techniques have been compared and found that imaging diagnosis (MRI has around 96% accuracy) is better than any other technique [10]. However, this technique cannot detect the risk level if before it occurred.

A study shows that abdominal pain in the right lower quadrant lump is not available more than 50% of patients, and it is more difficult to diagnosis appendicitis affected children because they do not have any previous medical report [11, 12]. From that perspective, this study was conducted to find out risk factors for appendicitis using pre-operative symptoms by feature selection and predictive Apriori algorithm and, at the end of the study, developed an android based application. This app performs a score-based calculation on the individual user's input and categorizes the person who is in very high, high, medium, or low risk. This app will be very helpful in increasing awareness among people about appendicitis diagnosis in needs.

The rest of the article is designed as follows: Methodology in section 2 describes the feature selection method and generates association rules from a predictive Apriori algorithm to figure out the risk score and finally develop a risk prediction flow char based on the risk score table. Section 3 and 4 describe all results and data analysis and discussion about results and the app. Then finally conclusion takes place in the section of 5.

## 2. Materials and Methods

Feature selection can be divided into three parts as filters, wrappers, or embedded methods. Filter methods are defined as using some actual property of the data to select a feature using the classification algorithm [13]. In this study, we have only discussed the filtered method as like predictive Apriori algorithm and a classifier algorithm to calculate the correlation, information gain, gain ratio, relief, and symmetrical uncertainty. Different filtering methods are described briefly in the following sections.

### 2.1. Correlation-based feature selection.

A correlation-based feature selection technique evaluates the correlation within the subset of features that are extremely correlated with the class by using a greedy search strategy in a mode of Ranker search method [14]. So, if the ranks rate is high, then subsets are extremely correlated with each other. Otherwise, there are no strong relationships among them.

$$\text{Correlation, } C(A|B) = \frac{H(B) - H(B|A)}{H(B)} \quad (01)$$

In the equation 1, here,  $C(A|B)$  is the correlation between A and B and  $H(A)$ ,  $H(B)$  is the entropy of respectively A and B and  $H(B|A)$  is the entropy of B given A [15].

2.2. *Information Gain (IG) based Feature selection.*

Information gain-based feature selection generates an expected number of results (Information) from the classification target attribute [16]. By using this feature, a score is calculated based on how much information is gained by the class. The information gain of feature A is defined as follows

$$\text{Information Gain (A)} = H(B) - H\left(\frac{B}{A}\right) \tag{02}$$

In equation 2, H(B) is the entropy of B and H(B/A) is a conditional entropy of class B given Feature A [17, 18].

2.3. *Gain Ratio (GR) based Feature selection.*

Information Gain Ratio is a ratio and formulated by Information Gain. It will maximize the feature information gain while minimizing the number of its values. The formula is defined as

$$\text{Gain Ratio (A)} = \frac{\text{IG (A)}}{I(A)} \tag{03}$$

In equation (03), the gain ratio (A) defined as the ratio between the information gain and the entropy of feature A [17, 18, 19].

2.4. *Relief Based Feature Selection.*

Relief algorithm first formulated by Kira and Rendered on by instance-based learning [20]. It is an instance-based searching to assign a new weight for an individual feature. The searching procedure is very simple because it searches the nearest neighbors in the training dataset but not in-depth. For each sampled instance, the nearest sample match and not match are found. Those matches and the not-matching ratio will update the weight of individual features and gives a rank [17].

2.5. *Symmetrical Uncertainty (SU) Based Feature Selection.*

The SU can be measured by the following equation:

$$SU = 2\left[\frac{\text{IG (A)}}{H(X) + H(Y)}\right] \tag{04}$$

where, H(X) is the entropy of features X, and H(Y) is the entropy of features [17, 21]. It covers the information gain's bias toward features with more values and tempers its values to the range (0, 1). The value 1 indicates that attributes X and Y are completely hooked, and 0 refers they are independent.

2.6. *Association Rules by Predictive Apriori Algorithm.*

Association rules are very useful to extract the hidden pattern of data. The most popular association rules generator algorithm are Apriori, Predictive Apriori, and Tertius [20]. In this study, Predictive Apriori was used to generate the best rules. Both Apriori and Tertius generate association rules by its support level [22]. On the other hand, Predictive Apriori generates association rules by its support, confidence, and lift level. This three-evaluation parameter is very much helpful in narrowing down the association rules. That's why we choose, Predictive Apriori rather than others. It is also an advanced form of Apriori algorithm where the minimum support is considered as 0.01, confidence as 0.5, and lift as 2.

### 3. Results and Discussion

#### 3.1. Dataset.

In this research, we used a data set of appendicitis with 200 records, which were collected from Shaheed Suhrawardy Medical College & Hospital (SSMCH) Dhaka. All collected data is about pre-operative symptoms of case and control group where both groups have 100 records, and All 18 features are gender, drinking status, stale food, urine condition, fever, fever runs, appetite, passing gas, belly pain, abdominal pain, pain qualification, family history, diarrhea, vomiting, rate of nausea, migration pain and disease status.

#### 3.2. Feature selection results.

In table 1 represents some popular filtering techniques that are evaluated by the ranker algorithm and used to find out the relationship of features with appendicitis disease. Where rank score nearly 1 represents high correlation, and 0 refers to no relation to disease. From this perspective, abdominal pain, rate of nausea, fever runs, pain, qualification show high relationship and gender, urine condition, etc. show a pretty low relationship with disease. But it is very difficult to figure out which factor is highly correlated with appendicitis.

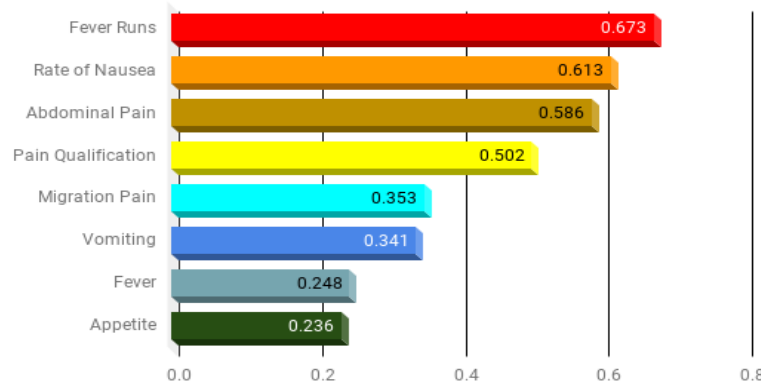
**Table 1.** Feature selection techniques with respect to ranker method.

Features	Correlation	GR	IG	Relief	SU
Abdominal Pain	0.75	0.456561	0.801389	0.4635	0.456545
Rate OF Nausea	0.6787	0.441837	0.697254	0.6775	0.56962
Fever Runs	0.6715	0.478504	0.850899	0.749	0.612543
Migration Pain	0.6231	0.305341	0.303135	0.2275	0.304234
Vomiting	0.6125	0.293426	0.291709	0.213	0.292565
Pain Qualification	0.5783	0.375749	0.456528	0.6115	0.488335
Fever	0.4956	0.18929	0.186207	0.1795	0.187736
Appetite	0.4615	0.160528	0.159786	0.2395	0.160156
Bally Pain	0.4604	0.159029	0.158846	0.0645	0.158938
Passing Gas	0.2376	0.053764	0.041829	0.067	0.047051
Family History	0.2121	0.05857	0.052925	0.0275	0.055604
Diarrhea	0.1361	0.025061	0.013622	0.0605	0.017651
Slate Food	0.1202	0.010478	0.010451	0.0335	0.010464
Age	0.0935	0.03316	0.034759	0.0495	0.033941
Drink Water	0.0326	0.000863	0.000766	0.053	0.000812
Gender	0.0308	0.000713	0.000686	0.048	0.000699
Urine Condition	0.0293	0.003716	0.002817	0.0645	0.003205

It is not clear that which filtering methods are work better on the appendicitis dataset. So, in this perspective average of filters ranks result will be a solution to find out top features see in figure 1. Figure 1 shows the average importance of a single factor and gets the top eight features that are highly correlated with appendicitis. Those are Fever Runs, Rate of Nausea, Abdominal Pain, Pain Qualification, Migration Pain, Vomiting, Fever, and appetite.

Some sub-factors are strongly responsible for having appendicitis disease. To examine those sub-factors, here, the top ten association rules are generated, which are responsible for having appendicitis, and those are demonstrated in table 2. Where, it is examined that if the value of fever runs is More than above, Abdominal Pain is Yes, Vomiting is Yes, Pain

Qualification is High, appetite is Yes, and so on then it is extremely supported to have appendicitis disease.



**Figure 1.** Top risk factors for appendicitis.

**Table 2.** Best Rules for YesDisease by Predictive Apriori Algorithm.

Serial	Rules	Result	Support
1	{'More than aboveFeverRuns'}	YesDisease	0.2537
2	{'More than aboveFeverRuns', 'YesAbdominalpain'}	YesDisease	0.2388
3	{'HighPainQualification', 'YesVomiting', 'YesAbdominalpain'}	YesDisease	0.1990
4	{'HighPainQualification', 'YesFever', 'YesVomiting'}	YesDisease	0.1791
5	{'YesFever', 'YesAppetite', 'HighPainQualification'}	YesDisease	0.1741
6	{'Little BitRateOFNowsha', 'More than aboveFeverRuns', 'YesAbdominalpain'}	YesDisease	0.1642
7	{'3 to 5 timesFeverRuns', 'YesAppetite', 'YesAbdominalpain'}	YesDisease	0.1393
8	{'YesFever', '3 to 5 timesFeverRuns', 'YesAppetite'}	YesDisease	0.1393
9	{'HighPainQualification', 'More than aboveFeverRuns', 'YesVomiting'}	YesDisease	0.1393
10	{'HighPainQualification', 'YesMigrationPain', 'More than aboveFeverRuns'}	YesDisease	0.1194

Also, the top ten association rules which have no appendicitis are investigated in this study, which is exhibited in Table 3. Where, it is examined that if the value of fever Runs is No, Abdominal Pain is No, Vomiting is No, Pain Qualification is Normal, Appetite is No, and so on then, it is highly supported to have no appendicitis disease.

**Table 3.** Best Rules for NoDisease by Predictive Apriori Algorithm.

Serial	Rules	Result	Support
1	{'NoAbdominalpain', 'NoneFeverRuns'}	NoDisease	0.4279
2	{'NoAbdominalpain', 'NonePainQualification'}	NoDisease	0.4080
3	{'NoMigrationPain', 'NoneFeverRuns'}	NoDisease	0.4080
4	{'NoVomiting', 'NoneFeverRuns'}	NoDisease	0.3930
5	{'NoAppetite', 'NoneRateOFNowsha'}	NoDisease	0.3234
6	{'NoneFeverRuns', 'NoFever'}	NoDisease	0.3234
7	{'LowPainQualification', 'NoneFeverRuns'}	NoDisease	0.0398
8	{'NoVomiting', 'NoneRateOFNowsha', 'Less than 2FeverRuns'}	NoDisease	0.0199
9	{'NoAppetite', 'NoneRateOFNowsha', 'Less than 2FeverRuns'}	NoDisease	0.0199
10	{'NormalRateOFNowsha', 'NoneFeverRuns'}	NoDisease	0.0100

### 3.3. The Score Calculation.

Each sub-category risk weight or score is demonstrated in Table 4, where the score is assigned by their supported ratio by following Table 2 and Table 3, respectively. The score (Highest is 4, and the lowest is 0.5) of sub-categories were formulated by their supporting rate. The rate of the score of nausea was subcategorized into 4. The score was assigned based on the level of risk stress and association with having appendicitis. Where, if the nausea rate is less than 2 then the score is two, when between 3 to 5 times, then score is assigned as 3, and for

more than above, the score is 4. The lowest score is 0.5 for the user when nausea is none. A similar calculation score has been conducted for other variables.

**Table 4.** Score table for each factor sub-category.

Attribute	Sub-category	Score
Fever Runs	Less than 2	2.00
	3 to 5 times	3.00
	More than above	4.00
	None	0.50
Rate of Nausea	Normal	2.00
	Little Bit	3.00
	So much	3.00
	Not tolerable	1.00
	None	0.50
Abdominal Pain	Yes	4.00
	No	1.00
Pain Qualification	Low	1.00
	Medium	3.00
	High	3.00
	None	0.50
Migration Pain	Yes	4.00
	No	0.50
Vomiting	Yes	4.00
	No	1.00
Fever	Yes	4.00
	No	1.00
Appetite	Yes	4.00
	No	1.00

The total score was 30, and the lowest was 6. The difference between the two variables was 24 (like  $30 - 6 = 24$ ), and the interval was 6 (like  $24 / 4 = 6$ ) because the total number of risk levels was 4.

So, the category of the risk score was surrounding its 6 intervals. Figure 2 shows the risk level of patients. If the total score is greater than or equal to 24 then the risk is “Very High,” and in the same way the score is greater than or equal to 18 then the risk is “High”, the score is greater than or equal to 12 then the risk is “Medium” and otherwise is “Low-risk”.



**Figure 2.** Flow Chart of Risk Prediction Algorithm.

### 3.4. Android Application Features.

The data and score tree were integrated into the Android application named “Predict Appendix”. First of all, Figure 3(a) and 3(b) are representing the starting page information, and Figure 3(c) is the registration page where users can register with self-information such as username, user division, and contact or email address. After completed successful registration, the user will be received some questions according to appendicitis, which is shown in Figure 3(d). The question has consisted of user fever runs, rate of nausea, abdominal pain, pain qualification, migration pain, vomiting, and fever. The user provides a specific tickmark to all questions about the relevant known question. After that, the “Predict Appendix” apps will be calculated the score and categorizes the user risk level that is shown in Figure 3(e).

After completing the risk identity, users can find out nearest hospital information into his/her division shows in Figure 3(f). Finally, the application is synchronized with a server, and all data was recorded in a private server.

### 3.5. Project Environment and Necessary Tools.

- Java
- XML
- Android Studio 3.2.1
- Server Storage (Firebase)
- Some Dependency library

### 3.6. Test Report of Predicted Algorithm.

After developing the application, a question is arising first: How accurately it classified the risk level of patients? To answer this question, a test has been conducted on case group peoples (100 appendicitis patients) because it is clear that every patient is in some amount of high risk, and they should not have a low risk. Test results have shown in Table 5 with the risk-based analysis among 4 categories.

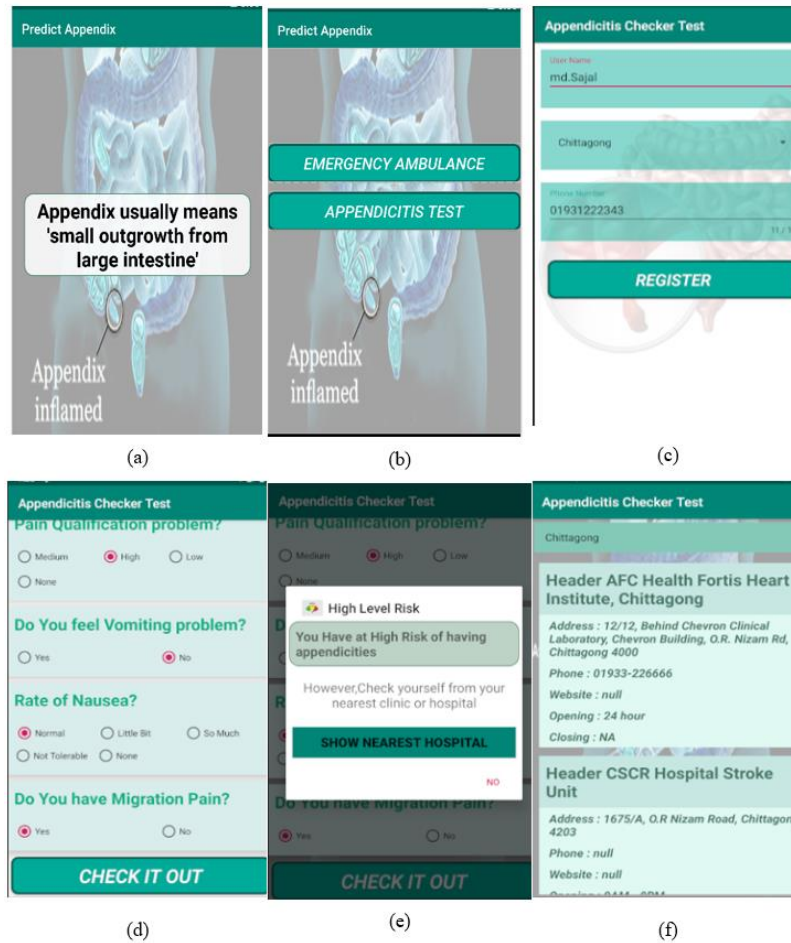
**Table 5.** Appendix Risk-based Analysis.

Risk Level	Risk Number	Percentage
Low	1	1%
Medium	4	4%
High	32	32%
Very Risk	74	74%
Total	100	100%

A total number of 74 patients were exploited by very high-level appendix disease, 32 patients as high level, and 4 patients as medium level and only 1 patient have been classified as low risk of appendix disease. From this perspective, it is clear that the developed risk algorithm can classify correctly (about 99%) of patients.

The findings of this study yielded some practical significance. First of all, appendicitis is common for all human beings. Someone could be affected with appendicitis in the earlier or middle stage of life or not affected by his/her entire life anymore [23]. Here all experimental results are tested with the dependent variable “disease status”. From the result analysis section, feature selection identified the top eight independent features among eighteen features those are (Fever Runs, Rate of Nausea, Abdominal Pain, Pain Qualification, Migration Pain, Vomiting, Fever, and Appetite) highly significant with appendicitis disease. It is depicted that,

if once have a fever and it more than five times in a day, have little nausea bit or so much and, have vomiting, have abdominal pain, and the pain qualification is medium or high, have migration pain, and have an appetite then he/she is in the high risk of having appendicitis.



**Figure 3.** Different steps of checking appendicitis risk level and suitable nearest hospital name.

The diagnosis of appendicitis could be best by doing ultrasonography, computed tomography (CT), and laparoscopic appendectomy [24]. This testing procedure is a timely and costly matter for low-income people. Also, uneducated peoples are not aware of their health condition. Sometimes, it could be late to take a proper diagnosis for appendicitis, and some people get died without treatment [4]. In this case, the developed app “Predict Appendix” could be a solution. It will be very helpful to make a proper decision. It can predict a person’s appendicitis risk level in a few seconds. And also, if someone gets a medium risk level to a very high-risk level, then it will suggest users go to the nearest hospital and take your treatment. So, this study will be very helpful for those peoples who want to identify appendicitis at the initial stage.

#### 4. Conclusions

Appendicitis is a common disease for all age people. One could be affected with appendicitis at any time anywhere. Presently, vast numbers of peoples are utilizing a smartphone to do their day by day needs. Smartphone-based disease prediction application is always a smart solution to take the correct decision in a certain moment. Our developed application “Predict Appendix” could be helpful to easily predict the risk level of appendicitis



and at the same time, increase awareness not only among peoples of Bangladesh but also all over the world.

## Funding

This research received no external funding.

## Acknowledgments

The authors are grateful who have participated in this research work.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Sellars, H.; Boorman, P. Acute appendicitis. *Surgery (Oxford)* **2017**, *35*, 432-438, <https://doi.org/10.1016/j.mpsur.2017.06.002>.
2. Finnerty, B.M.; Wu, X.; Giambone, G.P.; Gaber-Baylis, L.K.; Zabih, R.; Bhat, A.; Zarnegar, R.; Pomp, A.; Fleischut, P.; Afaneh, C. Conversion-to-open in laparoscopic appendectomy: A cohort analysis of risk factors and outcomes. *International Journal of Surgery* **2017**, *40*, 169-175, <https://doi.org/10.1016/j.ijso.2017.03.016>.
3. Ahmed, H.O.; Muhedin, R.; Boujan, A.; Aziz, A.H.S.; muhamad Abdulla, A.; Hardi, R.A.; Abdulla, A.A.; Sidiq, T.A. A five-year longitudinal observational study in morbidity and mortality of negative appendectomy in Sulaimani teaching Hospital/Kurdistan Region/Iraq. *Scientific Reports* **2020**, *10*(1), 1-7., <https://doi.org/10.1038/s41598-020-58847-1>
4. Subhanandaraj, T.T.; Raghavan, K.T.; Narayanan, R. Antibacterial and antibiofilm activity of probiotic based silver nanoparticles is a green approach in biomedical applications, *Letters in Applied NanoBioScience* **2020**, *9*, 988-994, <https://doi.org/10.33263/LIANBS92.988994>
5. Bonadio, W.; Shahid, S.; Vardi, L.; Buckingham, C.; Kornblatt, A.; Free, C.; Homel, P. A pre-operative clinical scoring system to distinguish perforation risk with pediatric appendicitis. *Journal of pediatric surgery* **2018**, *53*, 441-445, <https://doi.org/10.1016/j.jpedsurg.2017.05.017>.
6. Faisal, N. Recent trends and advances in 3d printing for biomedical applications. *Letters in Applied NanoBioScience* **2020**, *9*, 1126-1135, <https://doi.org/10.33263/LIANBS92.11261135>
7. Hosseini, M.; Tizmaghz, A.; Shabestanipour, G.; Aein, A. and Otaghvar, H.A. The Frequency of Different Clinical Presentation of Appendicitis, Complications and Prognosis in Elderly. *Annual Research & Review in Biology* **2014**, 4381-4388, <https://doi.org/10.9734/ARRB/2014/11717>.
8. Omling, E.; Salö, M.; Saluja, S.; Bergbrant, S.; Olsson, L.; Persson, A.; Björk, J.; Hagander, L. Nationwide study of appendicitis in children. *British Journal of Surgery* **2019**, *106*, 1623–1631 <https://doi.org/10.1002/bjs.11298>
9. Buckius, M.T.; McGrath, B.; Monk, J.; Grim, R.; Bell, T. and Ahuja, V.; 2012. Changing epidemiology of acute appendicitis in the United States: study period 1993–2008. *Journal of Surgical Research* **2012**, *175*, 185-190, <https://doi.org/10.1016/j.jss.2011.07.017>.
10. Haringa, J.B.; Bracken, R.L.; Davis, J.C.; Mao, L.; Kitchin, D.R.; Robbins, J.B.; Ziemlewicz, T.J.; Pickhardt, P.J.; Reeder, S.B.; Repplinger, M.D. Prospective evaluation of MRI compared with CT for the etiology of abdominal pain in emergency department patients with concern for appendicitis. *Journal of Magnetic Resonance Imaging* **2019**, *50*(5), 1651-1658, <https://doi.org/10.1002/jmri.26728>
11. Lo, H.C.; Chien, W.K.; Huang, C.S.; Wu, S.M.; Huang, H.C.; Chen, R.J.; Luo, C.C. Does age affect the outcomes and management of pediatric appendicitis in Taiwan?. *Formosan Journal of Surgery* **2016**, *49*, 174-177, <https://doi.org/10.1016/j.fjs.2016.03.009>.
12. Chen, C.L.; Chao, H.C.; Kong, M.S.; Chen, S.Y. Risk factors for prolonged hospitalization in pediatric appendicitis patients with medical treatment. *Pediatrics & Neonatology* **2017**, *58*, 223-228, <https://doi.org/10.1016/j.pedneo.2016.02.011>.
13. Kou, G.; Yang, P.; Peng, Y.; Xiao, F.; Chen, Y.; Alsaadi, F.E. Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Applied Soft Computing* **2020**, *86*, p.105836., <https://doi.org/10.1016/j.asoc.2019.105836>
14. Oh, D.; Ahn, Y.C.; Park, H.C.; Lim, D.H.; Han, Y. Prediction of radiation pneumonitis following high-dose thoracic radiation therapy by 3 Gy/fraction for non-small cell lung cancer: analysis of clinical and dosimetric factors. *Japanese journal of clinical oncology* **2009**, *39*, 151-157, <https://doi.org/10.1093/jcco/hyn158>.

15. Palma-Mendoza, R.J.; de-Marcos, L.; Rodriguez, D.; Alonso-Betanzos, A. Distributed correlation-based feature selection in spark. *Information Sciences* **2019**, *496*, 287-299, <https://doi.org/10.1016/j.ins.2018.10.052>
16. Zhang, G.; Hou, J.; Wang, J.; Yan, C.; Luo, J. Feature Selection for Microarray Data Classification Using Hybrid Information Gain and a Modified Binary Krill Herd Algorithm. *Interdisciplinary Sciences, Computational Life Sciences* **2020**, <https://doi.org/10.1007/s12539-020-00372-w>
17. Cilia, N.D.; De Stefano, C.; Fontanella, F.; di Freca, A.S. A ranking-based feature selection approach for handwritten character recognition. *Pattern Recognition Letters* **2019**, *121*, 77-86, <https://doi.org/10.1016/j.patrec.2018.04.007>.
18. Win, T.Z.; Kham, N.S.M. Information gain measured feature selection to reduce high dimensional data. *Seventeenth International Conference on Computer Applications* **2019**, 68-73, <http://onlineresource.ucsy.edu.mm/handle/123456789/1167>
19. Karegowda, A.G.; Manjunath, A.S.; Jayaram, M.A. Comparative study of attribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management* **2010**, *2*, 271-277.
20. Urbanowicz, R.J.; Meeker, M.; La Cava, W.; Olson, R.S.; Moore, J.H. Relief-based feature selection: Introduction and review. *Journal of biomedical informatics* **2018**, *85*, 189-203, <https://doi.org/10.1016/j.jbi.2018.07.014>.
21. Nashwan, M.S.; Shahid, S. Symmetrical uncertainty and random forest for the evaluation of gridded precipitation and temperature data. *Atmospheric Research* **2019**, *230*, p.104632, <https://doi.org/10.1016/j.atmosres.2019.104632>
22. Scheffer, T. Finding association rules that trade support optimally against confidence. In: *European conference on principles of data mining and knowledge discovery*. Springer, Berlin, Heidelberg. **2001**; pp. 424-435, [https://doi.org/10.1007/3-540-44794-6\\_35](https://doi.org/10.1007/3-540-44794-6_35).
23. Li, H.M.; Yeh, L.R.; Huang, Y.K.; Hsieh, M.Y.; Yu, K.H.; Kuo, C.F. Familial risk of appendicitis: a nationwide population study. *The Journal of pediatrics* **2018**, *203*, 330-335, <https://doi.org/10.1016/j.jpeds.2018.07.071>.
24. Ishikawa, H. Diagnosis and treatment of acute appendicitis. *Japan medical association journal* **2003**, *46*, 217-221.