

Submitted: 1 Jan, 2018; Accepted: 1 Jan, 2018; Publish: 13 March, 2018

Pediatric Irritable Bowel Syndrome Prediction Using 2 - Tier Ensemble Classifier

Maninder Kaur¹, Ashish Jat² and Reuel Ajith³

¹Thapar Institute of Engineering and Technology,
Patiala, India
maninderdhiran78@gmail.com

²Thapar Institute of Engineering and Technology,
Patiala, India
erashishjat@gmail.com

³Faculty of Medicine, Vilnius University,
Vilnius, Lithuania
reuel.ajith@gmail.com

Abstract: Irritable Bowel Syndrome (IBS) is a chronic, painful digestive disease that adversely affects child's wellbeing and health. It is generally in the form of frequent abdominal twinge in child that can make life of child pathetic. In such a situation, a disease prediction model can be of huge assistance in recognizing high-risk individuals. In this work, a novel bi-level ensemble model is developed for prediction of IBS. The work employs the Wrapper method for determining the optimal set of features. In this approach of ensemble, five different models of machine learning are combined in a group of three with different permutation to further improve the results. The majority voting technique is employed to get the outcome of meta-classifiers. The novel model achieves accuracy of 92.754 % with 67:33 ratio of training test balanced class data splitter. The results revealed that the correctness of proposed model is increased in contrast to the single model accuracy. This is the first initiative to predict IBS among children. It is possible that early forecasting of chronic diseases can impact a huge number of individuals and lessen the prevalence and expenditure of these diseases.

Keywords: IBS Prediction, Machine Learning, Ensemble approach, Feature Selection

I. Introduction

Functional gastrointestinal disorders (FGIDs) are a diverse collection of incessant conditions of the gastrointestinal tract that are considered vital to public wellbeing. These disorders being extraordinarily widespread can provoke a major societal and financial burden. Among these disorders, Irritable bowel syndrome (IBS) is the majority widespread FGID syndrome that drastically affects the larger intestine of human body [1]. The IBS affected patient may suffer from bloating, diarrhea, cramping, gas and/or constipation. Although

the exact origin of IBS is still unknown, doctors believe that stress and genetic pre disposition may add to the onset of IBS. Despite of being illustrated 150 years before, IBS still remains as a medical confront in the today's era [2, 3]. About one in five people can be affected by IBS at some point in their lives, detrimentally impacting their quality of life [4]. The dominance of IBS differs according to nation and measure used to characterize IBS. IBS is not linked with the growth of severe disease or increased death rates but nevertheless every sufferer with IBS is habitually checked for more serious ailments such as gastrointestinal carcinoma or inflammatory bowel disease. IBS is more commonly recognized in kids [5]. Teenagers who have a history of persistent abdominal pain are at increased threat of IBS throughout the period of adolescence and young adulthood.

There can be many causes of IBS within the child's body like GI motor problems, mental health problems, brain-gut signal problems, small intestinal bacterial overgrowth, hypersensitivity, bacterial gastroenteritis and genetics [6]. IBS is chronically recurring abdominal pain and uneasiness linked with distorted bowel behavior. The Center for Disease Control and Prevention (cdc) claims the unremitting diseases to be among the most avertable diseases [7]. In such situations, a disease prediction algorithm can be of huge assistance in recognizing high-risk individuals. It is possible that early forecasting of chronic diseases can impact a huge number of individuals and lessen the prevalence and expenditure of these diseases.

In recent times, a number of researchers have utilized machine learning (ML) methods for disease prediction [8], [9], [10]. Cruz and Wishart [11] gave an outline of a few machine learning methodologies used for cancer prognosis, including decision trees, neural networks, and nearest neighbor classi-

fiers. Conroy et al. [12] built a hazard model on basic clinical and demographic data to forecast possibility of cardiovascular disease. Khalilia et al.[13] applied random forest ML classifier for predicting disease using highly imbalanced data. Even though IBS is among the most widespread disorders in primary care practices and gastroenterology, not much work so far has been done to predict the possibility of this syndrome in recent past. In the current work, an effort has been made to build an efficient ensemble model that predicts IBS ahead of time in view of the present significant medical status of the patient. Diverse classification techniques have been executed in the automated model and ensemble of best models has been performed to increase the performance of model. These models are implemented on standalone machine. Without compromising on the diagnosis quality, these models have a tendency to lessen the number of tests required by the sufferer.

II. Computation Methods

The dataset for pediatric IBS consists of 33 features[14], wherein important factors include erythrocyte sedimentation rate (ESR) sequences, albumin, alanine trasedinsainase (ALT) and vitamin D. The properties of the data set are listed in Table 1.

Various studies [15] have shown that Vitamin D plays a significant part in the development and treatment of IBS among adults. From one case study [16] it has been found that vitamin D deficiency was vastly prevalent among teenagers with IBS. One major symptom in IBS is the inflammation in stomach

Table 1: IBS Dataset Features

Features	Information	Features	Information
C1	Class i.e. IBS, control	F17	Log albumin
F1	BoxPlot Code	F18	Albumin score
F2	Study ID number	F19	ALT
F3	ESR(0-20 mm/hr)	F20	Log ALT
F4	Sex	F21	ALT score
F5	Date of Diagnosis	F22	VitaminD (ng/mL)
F6	Duration(yr)	F23	Vit D (nmol/L)
F7	Age	F24	logVitD
F8	Height (cm)	F25	VitDStatus 15ng/mL cutoff
F9	Height (SDS)	F26	VitDStatus 20ng/mL cutoff
F10	Weight (kg)	F27	VitDStatus 30 ng/mL cut off
F11	Weight (SDS)	F28	Vit-D draw date
F12	BMI	F29	Season
F13	BMI (SDS)	F30	Season Code
F14	BMI %ile	F31	Race
F15	BMI2	F32	Race Code
F16	Albumin		

which can be easily diagnosed by relatively simple, inexpensive ESR test [17]. Recent studies[18] have discovered that patients with high level of ALT test cause the liver damage that further increases the chance of IBS in the body. Furthermore, the epidemiologic data reveals that obesity is coupled with chronic gastrointestinal sickness, many of which

are related with FGIDs such as dyspepsia or IBS[19]. If there is fluctuation of weight out of the healthy weight range then it can elevate the risk of IBS. In this work, Weka[20] open source software, licensed under GNU GPL, is utilized which comprises of tools for classification, regression, clustering, data pre-processing, visualization, and association rules. Of the numerous machine learning approaches explored, 10 classification models have been utilized in current work.

A. Feature Selection using Wrapper method

While building model, feature selection technique filters biases, unwanted noise and correlated variables from the dataset. It chooses vital features that may enhance the performance of the model. In the current work, for attribute selection, a wrapper subset evaluation method is utilized. Wrapper methods [21] test each attribute or a collection of attributes and rank the classification on the basis of predetermined metric (such as precision, accuracy, or recall).When using the this method, the Best First approach is used in conjunction. Best First uses a greedy algorithm technique that works backwards by beginning with a full dataset then keep on removing or keeping the feature iteratively based on an evaluation.

B. Methodology used

In the initial steps after procuring data, the dirty or coarse data is replaced using Replace Missing Values filter in Weka[22]. The filter fixes absent values with conceivable data values that are computed from a distribution explicitly designed for each missing value. In next step, Wrapper method is utilized to obtain the subset of significant attributes. This practice boosts the accuracy as well as reduces the computation time of the model. The feature selection process is followed by training the classifiers using given dataset with their optimum tuning parameters. These models are further combined to obtain 2-tier ensemble approach. In this work, 67% of the entire dataset is kept for training purpose and the remaining 33% is utilized for testing the correctness of the classifier.

Figure 1 depicts the methodology of the proposed scheme for efficient IBS prediction. The data splitter partitions the entire dataset into the ratio of (67:33) of (Train: Test) subsets. The ten base models are trained with training subsets followed by performance assessment. The performance of the model is assessed on the basis of various performance criteria such as recall,precision, area under the curve (AUC) and accuracy. [23, 24]. In this research, a novel combination of five best machine learning models is used to develop an ensemble machine learning approach for determining the presence of IBS in the pediatric patient. The technique generates the proposed model in two phases-

1st Phase - Ten different models are trained with 67% dataset and 33% dataset for IBS prediction.(Tier 1 models) 2nd Phase -Top five models (based on accuracy) are combined in group of three, making 10 different combinations to further improve the performance of prediction model(Tier 2 models).The final output is selected using Majority Voting.

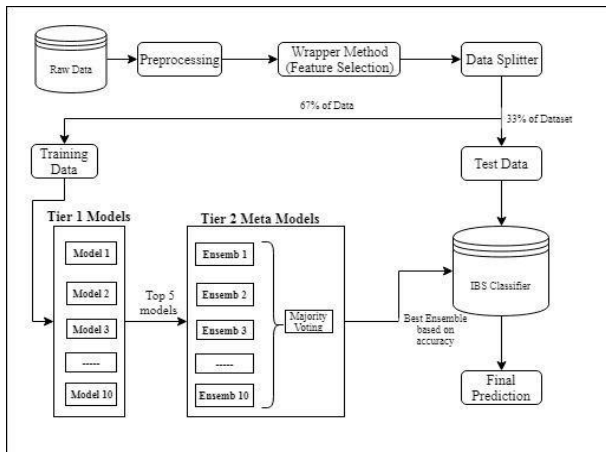


Figure. 1: The methodology of the proposed approach

The flow of proposed methodology is presented in Figure 2. The figure depicts the schematic diagram of the final combination of all five ensemble models. As the data travels through three models, the models splendidly train the data to offer reliable and precise outcomes. Ensemble is employed to manage the worst case of model prediction [25]. The models are learnt on 67% of the data whereas the left 33% is utilized for testing purpose. The degree of exploitation of the

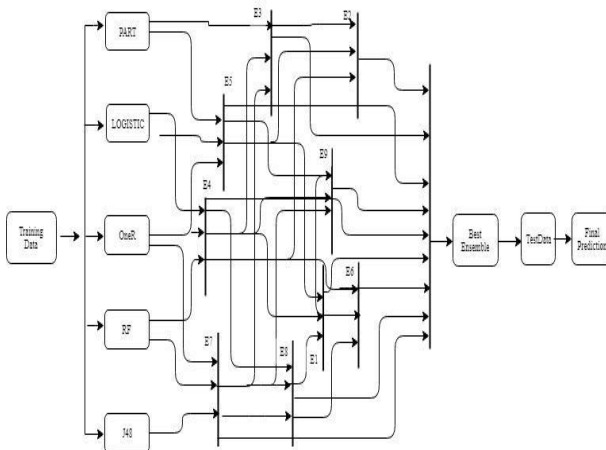


Figure. 2: The workflow of the proposed scheme

potential of combining machine learning algorithms relies on the grouping strategy employed. The Majority Voting (MV) method [22] is the most often used straightforward combiners for binary classification that works by adding up the votes of all the predictions fetched from the learner, building a resultant outcome, utilizing the class labels with majority votes.

C. Evaluation criteria

The intent of this study is to forecast the incidence of IBS by application of different machine learning methodologies. The results from various models and their ensemble are compared and analyzed based on several evaluation criteria

such as accuracy, precision, AUC, and recall as listed below.

Recall

It is the proportion of rightly categorized sufferer to the count of sufferer with predicted IBS. It is calculated as given in eq 1.

$$Recall = \frac{tp}{(tp + fn)} \quad (1)$$

Precision

It is described in terms of the proportion of accurately recognized IBS sufferer to the count of people having IBS. Precision is computed as in eq 2.

$$Precision = \frac{tp}{(tp + fp)} \quad (2)$$

Accuracy

It quantifies the precision of the learner. The accuracy is computed as in eq3.

$$Accuracy = \frac{tp + tn}{TotalData} * 100 \quad (3)$$

F measure

It evaluates the test's accuracy in terms of both recall and precision. F measure is claculated as in eq 4

$$F_Score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (4)$$

tp signifies true positive : IBS patients categorized as having IBS

tn signifies true negative: IBS patients categorized as hale and hearty

fp signifies false positive : Fit individuals predicted as suffering from IBS.

fn signifies false negative: People with IBS wrongly classified as fit

ROC

The correctness of prediction models can also be anticipated by finding the total area under ROC curve (AUC) measures the quality of classifier. The AUC value ranges from 0 to 1. The model achieving more value of AUC in contrast to rest of the models is observed as effective model and the model with better performance has AUC closer to 1.

III. Results and Discussion

On Standalone Machine, ten different models are implemented in WEKA 3.9.2 under Intel Corei5 processor @ 2.80 GHz, 4.00 GB RAM, on a 64 bit Windows 7 operating system. The whole steps are divided into mainly six stages

- Feature selection
- Training test split (balanced classes)
- Classification using 10 machine learning algorithms

- Selecting top 5 algorithms based on accuracy
- Ensemble the top 5 algorithms (3 in group)
- Employing majority voting and selecting the best ensemble approach based on accuracy

The CSV file of original dataset is transformed to Attribute-Relation File Format (ARFF) files for WEKA 3.8. In the initial stage, the Wrapper method for feature selection is applied on the given dataset which automatically selects most relevant attributes, discarding 21 features from the dataset. Based on these selected features and target class, IBS prediction is formulated as mentioned in eq 5:

$$C1 = f(F3, F4, F7, F10, F14, F15, F16, F19, F25, F29) \tag{5}$$

Thereafter, the data is partitioned into training and test data such that the ratio of patients suffering from IBS to healthy patient remains the same in both test and training data as that of original dataset. The ratio of features healthy to sufferer is 2.11 in original dataset. After splitting, the training and test dataset contains 114 and 57 features respectively bearing the same ratio.

The data splitting phase follows the application of 10 machine learning classifiers [26, 27, 28]. Table 2 shows the models used in present study along with their tuning parameters Table 3 depicts the values of various performance met-

Table 2: The models with tuning parameters

Classifier	Parameter Name	Parameter Value
Multilayer perceptron	Learning Rate	0.3
	Momentum	0.2
	Hidden Layers	A
Logistic	Max Its	-1
	ridge	1.00E-08
Nave Bayes	Use Supervised Discretization	FALSE
	Use Kernel Estimator	FALSE
PART	confidence actor	0.25
	Min Num Obj	2
	Num Folds	3
J48	Confidence Factor	0.25
	Min Num Obj	2
	Num Folds	3
SMO	Filter Type	Normalize training data
	kernel	PolyKernel
Random forest	toleranceParameter	0.001
	Break Ties Randomly	FALSE
	Compute Attribute Importance	FALSE
jRip	Num Iteration	100
	Folds	3
	Min No	2
One R	Optimizations	2
	Min Bucket Size	6
IBK	Batch Size	100
	Cross Validate	FALSE
	Neighbour Search Algorithm	Linear NN Search
	Window Size	0

rics obtained from 10 different machine learning models. It is clear from the Table that J48 and OneR models achieved

the higher accuracy i.e. 88.4058% and 89.8551% respectively, as compared to other eight models. Furthermore, the same models are able to be exceedingly precise in filtering healthy instances from the dataset with a precision of 0.897 and 0.901 respectively. Apart from these two models, the accuracy achieved by Random forest is 86.96% and PART is 84.06%. The SMO model achieved the least accuracy of 71.01% as compared to other models. The OneR classification model reveals highest recall value of 0.899 whereas SMO and Nave Bayes have least recall value i.e. 0.710. It can be observed from the table that the three classifiers random forest, j48 and OneR showed AUC value greater than 0.85. Furthermore, the AUC value is minimum for SMO and IBK (i.e., 0.642) and maximum for random forest (i.e., 0.933) classifier model. The results presented in Table IV reflect the predictive power of the OneR model.

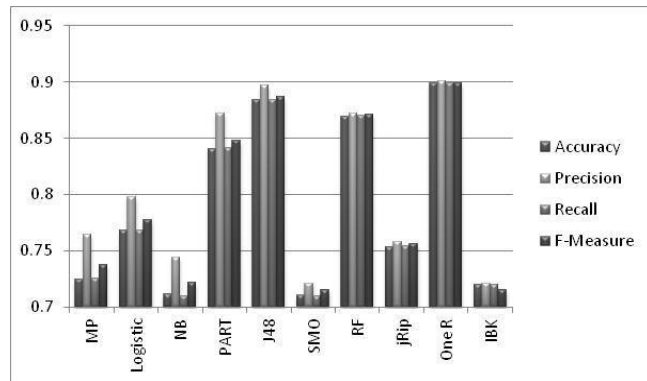


Figure. 3: The methodology of the proposed approach

The figure 3 represents the prediction accuracies, recall, precision and F measure for 10 classifiers in hand. Herein One R classifier is shown to be the best overall performer and SMO to be the worst.

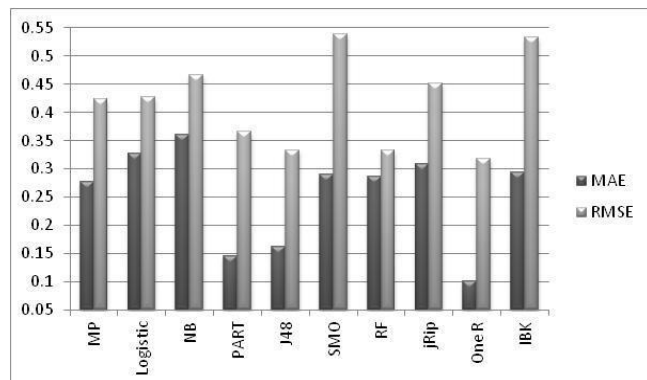


Figure. 4: The methodology of the proposed approach

The Root Mean Square Error (RMSE) value calculates the dissimilarity between values forecasted by an estimator and the values observed [29]. Mean absolute error (MAE), signifies the mean of the dissimilarity in forecasted and actual value for all test cases. The lesser the values of

Table 3: Performance evaluations of ten classifiers for 67% training data and 33% test data

Models	Correctly Classified	Incorrectly Classified	TP Rate	FP Rate	Precision	Recall	F Score	MCC	ROC Area	PRC Area	MAE	RMSE
MLP	72.464	27.536	0.725	0.313	0.764	0.725	0.737	0.376	0.829	0.862	0.277	0.423
NB	71.138	28.862	0.710	0.354	0.744	0.710	0.722	0.328	0.686	0.761	0.361	0.465
RF	86.957	13.044	0.870	0.190	0.872	0.870	0.871	0.668	0.933	0.947	0.287	0.332
jRip	75.362	24.638	0.754	0.375	0.758	0.754	0.756	0.373	0.694	0.723	0.309	0.450
IBK	72.012	27.989	0.720	0.426	0.721	0.720	0.715	0.275	0.642	0.684	0.294	0.533
Logistic	76.812	23.188	0.768	0.262	0.798	0.768	0.777	0.467	0.803	0.823	0.326	0.427
One R	89.855	10.145	0.899	0.144	0.901	0.899	0.899	0.742	0.877	0.866	0.101	0.319
SMO	71.015	28.986	0.710	0.426	0.721	0.710	0.715	0.275	0.642	0.684	0.290	0.538
PART	84.058	15.942	0.841	0.128	0.872	0.841	0.848	0.651	0.900	0.886	0.145	0.365
J48	88.406	11.594	0.884	0.113	0.897	0.884	0.887	0.727	0.894	0.879	0.161	0.333

Table 4: Performance Evaluations of Ten Ensemble Models

Ensemble Classifier	Correctly Classified	Incorrectly Classified	TP Rate	FP Rate	Precision	Recall	F Score	MCC	ROC Area	PRC Area	MAE	RMSE
E1	89.855	10.145	0.899	0.144	0.901	0.899	0.899	0.742	0.877	0.866	0.101	0.319
E2	91.304	8.696	0.913	0.139	0.913	0.913	0.913	0.775	0.887	0.879	0.087	0.295
E3	92.754	7.246	0.928	0.133	0.927	0.928	0.927	0.809	0.897	0.894	0.073	0.269
E4	91.304	8.696	0.913	0.139	0.913	0.913	0.913	0.775	0.887	0.879	0.087	0.295
E5	92.754	7.246	0.928	0.133	0.927	0.928	0.927	0.809	0.897	0.894	0.073	0.269
E6	91.304	8.696	0.913	0.103	0.918	0.913	0.914	0.784	0.905	0.889	0.087	0.295
E7	89.855	10.145	0.899	0.144	0.901	0.899	0.899	0.742	0.877	0.866	0.101	0.319
E8	91.304	8.696	0.913	0.139	0.913	0.913	0.913	0.775	0.887	0.879	0.087	0.295
E9	88.406	11.594	0.116	0.345	0.884	0.113	0.897	0.884	0.887	0.727	0.886	0.865
E10	91.304	8.696	0.913	0.103	0.918	0.913	0.914	0.784	0.905	0.889	0.087	0.295

RMSE and MAE, the better is the performance of model. RMSE and MAE values of 0 specify a perfect fit. From Figure 4, it is concluded that the OneR model gives minimum testing RMSE of 0.3185 whereas SMO gives the worst value for RMSE i.e 0.5384.

As each category of classifiers has potentials and shortcomings, consequently top resultant models (based on accuracy) are combined to generate ensemble classifiers[30]. Ensembles use a number of different algorithms in order to achieve a superior predictive performance than could be achieved by using any of the individual learners. Based on the results of accuracy, top five models namely J48, PART, Random Forest, One R and Logistic are selected for the development of ensemble models by combining three models in each ensemble for IBS disease prediction [31].

The following are 10 ensemble models from the top five models chosen based on their accuracy values

- E1(J48, Logistic, One R)
- E2(Random Forest, PART, Logistic)
- E3(One R, Random Forest, PART)
- E4(One R, Random Forest, Logistic)
- E5(Logistic, One R, PART)
- E6(J48,Random Forest, PART)
- E7(J48,Random Forest, OneR)
- E8(Logistic, J48, Random Forest)
- E9(J48, PART, One R)
- E10(J48,Logistic, PART)

The current work employs majority voting for finalizing the outcome of the ensemble classifiers. Table 4 shows the performance parameter values obtained from ten ensembles (three in group), after application of majority voting scheme on three classification results. It is clear from the Table that both Ensemble 3 (E3) and Ensemble 5(E5) obtained the highest accuracy of 92.7536 % as compared to other models. As majority voting rule for ensemble is utilized to arrive at a

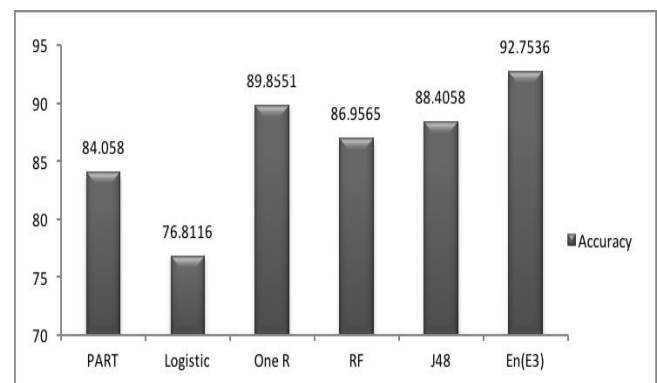


Figure 5: The methodology of the proposed approach

decision. It is clear from the results that the common classifiers of E3 and E5 i.e. One R and PART have the potential to significantly improve the accuracy in predicting IBS disease. Moreover, the Ensemble 9 (E9) combining J48, PART and One R gives lowest accuracy of 88.4058% as compared to other ensembles. In terms of the AUC performance criteria, the ensembles E1, E2 and E4 all have AUC greater than 90% , meanwhile E10, E11 and E14 ensemble are giving good

performance in terms of recall and precision. It can be concluded that Ensemble methods enhanced the performance in comparison to their base methods with the exception of E9 combining J48, PART and One R.

Figure 5 depicts the comparative analysis of top five base models with the ensemble approach in terms of average accuracy. The ensemble approach gives the highest accuracy value of 92.7536% to detect the presence of IBS. The proposed approach is multilevel ensemble model, a group of three base models that outperforms in contrast to the top five classification models. It is revealed from the results that the ensemble model has improved accuracy in contrast to the base model accuracy.

IV. Conclusions

In the current work, a supervised 2 tier-ensemble approach for prediction of IBS among children is proposed. The work presents a cost effective way of diagnosing IBS in children with a high degree of efficiency. The children dataset consists of features that are easily diagnosable and are cost effective. The capability of the model to predict IBS relies on right training features chosen using Wrapper method. The work combines three models out of top 5 classifiers with majority voting to improve the performance of classification. The model built using ensemble approach is evaluated using accuracy, recall, precision and ROC performance evaluators. The empirical relevance of diversity estimates is assessed with regards to combining classification models by majority voting. The model is 92.754 % accurate in predicting whether a child is suffering from IBS or not. The present study is first step towards early forecasting of IBS using the ML-based classification model that can impact a huge number of individuals and lessen the prevalence and expenditure of these diseases.

References

- [1] Rebecca M Lovell and Alexander C Ford. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. *Clinical Gastroenterology and Hepatology*, 10(7):712–721, 2012.
- [2] W Cumming. Electrogalvinism in a particular affliction of mucous membrane of the bowels. *London Med Gaz*, 59:969–73, 1948.
- [3] Brenda J Horwitz and Robert S Fisher. The irritable bowel syndrome. *New England Journal of Medicine*, 344(24):1846–1850, 2001.
- [4] George F Longstreth, Alisa Wilson, Kevin Knight, John Wong, Chiun-Fang Chiou, Victoria Barghout, Feride Frech, and Joshua J Ofman. Irritable bowel syndrome, health care use, and costs: a us managed care perspective. *The American journal of gastroenterology*, 98(3):600–607, 2003.
- [5] James W Varni, Mariella M Lane, Tasha M Burwinckle, Eve N Fontaine, Nader N Youssef, Jeffrey B Schwimmer, Perrie E Pardee, John F Pohl, and David J Easley. Health-related quality of life in pediatric patients with irritable bowel syndrome:: A comparative analysis. *Journal of Developmental & Behavioral Pediatrics*, 27(6):451–458, 2006.
- [6] Wael El-Matary, Christine Spray, and Bhupinder Sandhu. Irritable bowel syndrome: the commonest cause of recurrent abdominal pain in children. *European journal of pediatrics*, 163(10):584–588, 2004.
- [7] Chronic disease overview. <http://www.cdc.gov/chronicdisease/overview/>, 2016.
- [8] N. S. Philip A. Abraham and P. Saratchandran. Modeling chaotic behavior of stock indices using intelligent paradigms. *International Journal of Neural, Parallel & Scientific Computations*, 11:143–160, 2003.
- [9] A. Abraham H. Izakian and V. Snasel. Fuzzy clustering using hybrid fuzzy c-means and fuzzy particle swarm optimization. *World Congress on Nature and Biologically Inspired Computing (NaBIC 2009)*, India, IEEE Press, ISBN: 978-1-4244-5612-3, pages 1690–1694, 2009.
- [10] Z.X. Chi Y.Y. Hao X. Yue, A. Abraham and H.W. Mo. Artificial immune system inspired behavior based anti-spam filter. *Soft Computing: Soft Computing - A Fusion of Foundations, Methodologies and Applications*.
- [11] Joseph A Cruz and David S Wishart. Applications of machine learning in cancer prediction and prognosis. *Cancer informatics*, 2:117693510600200030, 2006.
- [12] et al. Conroy, R. M. Estimation of ten-year risk of fatal cardiovascular disease in europe: the score project. *European heart journal*, 24.(11):987–1003., 2009.
- [13] Mohammed Khalilia, Sounak Chakraborty, and Mihail Popescu. Predicting disease risks from highly imbalanced data using random forest. *BMC medical informatics and decision making*, 11(1):51, 2011.
- [14] http://escholarship.umassmed.edu/pediatrics_data/4/.
- [15] Yasir Khayyat and Suzan Attar. Vitamin d deficiency in patients with irritable bowel syndrome: does it exist? *Oman medical journal*, 30(2):115, 2015.
- [16] Benjamin Udoka Nwosu, Louise Maranda, and Ninfa Candela. Vitamin d status in pediatric irritable bowel syndrome. *PloS one*, 12(2):e0172183, 2017.
- [17] Ivana Plavšić, Goran Hauser, Mladenka Tkalčić, Sanda Pletikosić, and Nermin Salkić. Diagnosis of irritable bowel syndrome: role of potential biomarkers. *Gastroenterology research and practice*, 2015, 2015.
- [18] Seung-Hwa Lee, Kyu-Nam Kim, Kwang-Min Kim, and Nam-Seok Joo. Irritable bowel syndrome may be associated with elevated alanine aminotransferase and metabolic syndrome. *Yonsei medical journal*, 57(1):146–152, 2016.

- [19] William D Heizer, Susannah Southern, and Susan McGovern. The role of diet in symptoms of irritable bowel syndrome in adults: a narrative review. *Journal of the American Dietetic Association*, 109(7):1204–1214, 2009.
- [20] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
- [21] Xiaowei Xue, Min Yao, and Zhaohui Wu. A novel ensemble-based wrapper method for feature selection using extreme learning machine and genetic algorithm. *Knowledge and Information Systems*, 57(2):389–412, 2018.
- [22] Jorge García-Gutiérrez, Daniel Mateos-García, Mariano García, and José C Riquelme-Santos. An evolutionary-weighted majority voting and support vector machines applied to contextual classification of lidar and imagery data fusion. *Neurocomputing*, 163:17–24, 2015.
- [23] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- [24] David Martin Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. 2011.
- [25] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.
- [26] Taha M Abraham A. Gharib TF, Nassar H. An efficient algorithm for incremental mining of temporal association rules. *Data & Knowledge Engineering*, 69(8):800–815., 2010.
- [27] Grosan C. Tesema TB, Abraham A. Rule mining and classification of road traffic accidents using adaptive regression trees. *International Journal of Simulation*, 6(10-11):80–94, 2005.
- [28] Chen Y Abraham A. Peng L, Yang B. Data gravitation based classification. *Information Sciences.*, 179(6):809–19, 2009.
- [29] Tianfeng Chai and Roland R Draxler. Root mean square error (rmse) or mean absolute error (mae)?—arguments against avoiding rmse in the literature. *Geoscientific model development*, 7(3):1247–1250, 2014.
- [30] James Large, Jason Lines, and Anthony Bagnall. The heterogeneous ensembles of standard classification algorithms (hesca): the whole is greater than the sum of its parts. *arXiv preprint arXiv:1710.09220*, 2017.
- [31] Vincent Labatut and Hocine Cherifi. Accuracy measures for the comparison of classifiers. *arXiv preprint arXiv:1207.3790*, 2012.