ORIGINAL RESEARCH

A diagnostic prediction model for colorectal cancer in elderlies via internet of medical things

Parvaneh Asghari[1]

Abstract Internet of Medical Things (IoMT) and embedded systems have improved the healthcare systems by enabling remote monitoring the patients' health conditions anywhere and anytime especially during novel COVID-19 pandemic. In this paper, an IoT-based predicting model is proposed to predict colorectal cancer (CRC) in elderlies. It provides a CRC predicting model for the involved medical team to continuously trace an elderly's biological indicators using smart wearable embedded systems and medical IoT devices. In this model, vital medical data is collected by IoMT devices and sensors, then analytical information is derived via machine learning (ML) methods for early CRC diagnosis and elderly's health parameters changes. The experimental results confirm that the suggested model meets the proper accuracy of predicting the CRC in aged people.

Keywords Internet of Medical Things (IoMT) · Health monitoring system · Colorectal cancer · Data mining

1 Introduction

The fast growing population of aged people and high related healthcare overheads has become a serious challenge in our today's life [1]. Therefore, the fast declining health condition of elderlies motivates them to apply online health monitoring systems to prevent common threatening diseases such as cancers [2]. Recent technologies such as wireless sensor networks (WSN), Internet of Thing (IoT) and Internet of Medical Things (IoMT) contribute significantly to develop the smart health monitoring systems and applications for early diagnosis of non-contagious diseases such as cancers [3]. IoMT as Medical sensor-based IoT smart devices provide remote elderlies' health monitoring while they are at home, in hospital, or being anywhere else. Medical IoT sensors attached to the patient and transfer continuously vital health data to the medical team from anywhere and anytime [4].

With a high prevalence of novel COVID-19, early detection and prediction of CRC can contribute widely to decrease its heavy burden on communities comprising of aged and disabled people through making on-time and accurate clinical decisions on CRC detection for recommending on-time necessary treatments to them [5]. Most of the studies concentrated on the prediction of general chronic diseases such as diabetes mellitus, heart failures, kidney disease in IoT contexts by ML methods [6]. However, early detection of cancers as a serious non-contagious disease via IoT technology and data mining methods have been remained as an important problem due to the poor performance of clinical prediction models and the high cost of diagnosis procedures. This issue has led us to provide an IoT-based diagnostic model for early detection of CRC in elderlies based on monitoring the most important influencing indicators. The suggested model applies medical IoT sensors to constantly monitor the elderly's overall health function condition to detect the symptoms of CRC. Our IoT-based CRC prediction model is developed for detecting the high risk conditions and identifying changes in elderlies' biological indicators via data mining methods over the steady observing. The main contributions of this paper include: (i) proposing an IoT-based monitoring model for on-line collecting colorectal health data from

✉ Parvaneh Asghari
p_asghari@iauctb.ac.ir

1 Department of Computer Engineering, Central-Tehran Branch, Islamic Azad University, Tehran, Iran

anywhere and anytime for elderlies, (ii) providing an abstract model for early CRC diagnosis over following the influencing parameters including essential colorectal vital signs and clinical health status, and finally iii) evaluating the elderlies' biological alterations over data mining methods to predict the CRC cases and the level of its severity in elderlies.

The rest of this paper is organized as follows. Section 2 presents a brief survey and analysis of the current related works in this field. In Sect. 3, the proposed model is explained in detail that includes the IoT-based model for collecting the required data and the method for prediction of CRC based on the extracted associative rules. Section 4 demonstrates the experimental outcomes of the suggested method over statistical examination using some existing data classification methods to compare their outcomes. Section 5 concludes the proposed method for CRC prediction and provides future research directions in this area.

2 Related work

Relevant studies on CRC prediction has been less focused in digital healthcare systems, especially in IoT environments [1]. In this section, some of the recent related studies are surveyed. Also, a comparison between them is presented at the end of this section in Table 1.

An approach for prediction of advanced-stage CRC survival or death of patients was proposed in [1]. The authors considered the survival prediction procedure as a two-phase process. At the primary phase, a tree-based classification algorithm is offered for grouping the survivability of advanced-stage CRC patients. In the second phase, a selective regression technique was provided for lifetime prediction. The authors claimed that their offered prediction model reached a more accurate prediction in comparison to the one-phase regression methods. In [2], a supervised ML model for CRC prediction based on high dimensional gene data was proposed. In the offered supervised model, the authors presented a three layered light neural network model using a group of factors. The authors claimed that their model which employs the Monte Carlo algorithm presented a firm efficiency in comparison to the other methods. An approach using data mining methods deployed on a huge volume of medical information records to provide a model for CRC prediction in [3]. The decision tree learner was applied as the classification method. The performance of the proposed model was evaluated based on sensitivity and specificity factors. In [7], a predictive model based on ML techniques classifies patients based on blood counts representing a greater probability of CRC and deserving colonoscopy operation. The efficiency of the offered mode was evaluated based on True Positive/ False Positive (TP/FP) and True Negative/ False Negative (TN/FN) parameters. Also, a predictive model for CRC detection was proposed in [4] based on electronic Medical Records (EMRs). Some ML methods such as Random Forests (FR), Support Vector Machine (SVM), classification and regression tree (CART) algorithms were applied in this model. The parameters of TP/FP and TN/FN were used for assessing the proposed model. In [5], an approach was presented for colorectal surgical difficulties using Electronic Health Records (EHRs) as heterogeneous clinical data, based on ML

Table 1 Comparison of related works

Reference	Platform IoT/Cloud/Internet-based	Data mining me thod	Performance Factors				
			Accuracy	Sensitivity	Specificity	Time Consumption	TP/FP and TN/.FN
[1]	Not mentioned	Tree-based selective ensemble regression method	✕	✕	✕	✕	✔
[2]	Not mentioned	Neural network	✕	✕	✕	✔	✕
[3]	Not mentioned	Decision tree	✕	✔	✔	✕	✔
[7]	Not mentioned	Not mentioned	✕	✕	✕	✕	✔
[4]	Not mentioned	Random forests (FR), support vector machine (SVM), classification and regression tree (CART)	✕	✕	✕	✕	✔
[5]	Not mentioned	Support vector machine (SNM)	✕	✕	✕	✕	✔
[8]	Internet-based	Not mentioned	✕	✕	✕	✕	✕
[9]	Internet-based	Not mentioned	✕	✕	✕	✕	✕
Our model	Cloud-based IoT	Multi-layer perceptron (MLP), J48, sequential minimal optimization (SMO), naïve bayes	✔	✔	✔	✔	✔

methods. Both linear and non-linear forms of the SVM technique were applied in the proposed model. Using patient vital signs, blood samples values, and EHRs enhanced the performance of the proposed model. Recently, an Internet-based virtual humans (VH) approach has been presented in [8] for CRC screening. The authors provided the strategies for CRC screening which are derived from two aspects for having users' perception including the visual design and the information about the user intentions to follow additional health data. An experiential analysis compared the impact of the information medium comprising the animated and static VH, and also the text-based involvement on user objectives. It was revealed that VH-based interference can effect users' determinations outside of a laboratory environment. In [9], several potential predictors for CRC were recognized in a selected population of aged people using the interface among the screening database and primary care records. The authors observed that the considered predictors can be considered in a refined risk prediction model merging the newer quantitative FIT for bowel cancer screening. Table 1 presents a summary of the employed platform, applied data mining method and measurement factors for performance evaluation in the reviewed papers. As observed in Table 1, none of the surveyed papers employed the cloud or IoT as the powerful technologies in their approaches for predicting the critical condition of CRC for monitoring the potential patients.

3 Proposed model for CRC prediction

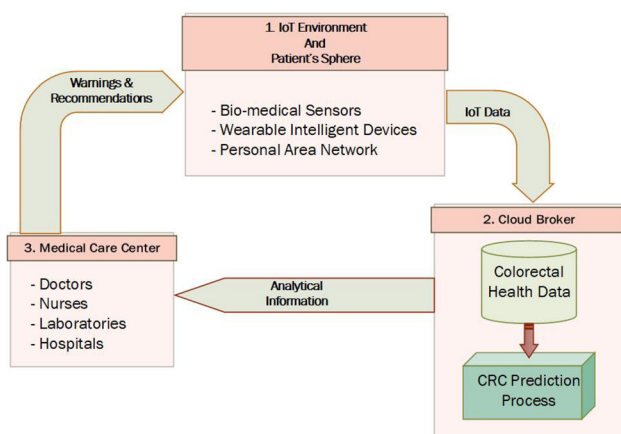The proposed abstract model of CRC prediction in Fig. 1 comprises of three main components as follows:

Fig. 1 The proposed screening model for early diagnosis of CRC in elderlies in IoT platform

1. In the IoT environment and patients' sphere, besides the required personal information which has to be given by the patient, the other on-line required data for CRC prediction are collected by bio-medical sensors and medical IoT devices which all of them are considered as behavioral data. Two other factors of Hemoglobin (Hgb) and Fecal Immunochemical Test (FIT) which are known as vital signs for CRC prediction are also collected in this component.
2. In the Cloud broker component, data preprocessing and CRC prediction processes are performed using the colorectal health data which have been collected in the previous component.
3. In the medical care component, the analytical outcomes derived from the CRC prediction process are passed to the involved medical team for recommending the required prescriptions and instructions to the patients.

The proposed model consists of four phases including that describe in the next subsections.

3.1 Determining essential data for CRC prediction

Most of the previous studies have considered some clinical signs for CRC possibility consist of complete blood count (CBC) [7]. However, there are some other signs for the CRC detection in aged people which are vital to assess [5, 10]. The main clinical indicators for CRC detection in elderlies and their definitions besides normal limits are described in Table 2. As shown in Table 2, the required parameters that influence on colorectal health status consist of three types of data including personal data, clinical data and clinical IoT data which have been proved by the world health organization WHO [11, 12]. The first group of required data consists of gender, personal history and family history of CRC [5]. This group is considered as personal data that has to be entered by the patient in the system once. The second group of required data that categorized as clinical data that must be entered daily in the system, consists of Constipation, Diarrhea, Rectal bleeding, Abdominal Pain and Abdominal tenderness. The third group of essential data for CRC prediction is the clinical IoT data which have to be collected via bio-medical sensors and MIoT devices that provide continuously gathering the essential data consist of Hemoglobin (Hgb), Fecal Immunochemical Test (FIT) and Wight loss which are sensed and collected online and constantly.

3.2 Data collection

To monitor the elderly's colorectal health factors, the IoT data have to be collected via MIoT smart devices. The

1426

Int. j. inf. tecnol. (August 2021) 13(4):1423–1429

Table 2 The main elderly's clinical indicators

No	Type of the data	Feature	Data Type	Unit
1	Personal data	Gender	Nominal	Male, Female
2	Personal data	Personal history	Nominal	Yes, no
3	Personal data	Family history	Nominal	Yes, no
4	Clinical data	Constipation	Nominal	Yes, no
5	Clinical data	Diarrhea	Nominal	Yes, no
6	Clinical data	Rectal bleeding	Nominal	Yes, no
7	Clinical data	Abdominal Pain	Nominal	Yes, no
8	Clinical data	Abdominal tenderness	Nominal	Yes, no
9	Clinical data	Abnormal rectal form	Nominal	Yes, no
10	Clinical IoT data	Wight loss	Nominal	Yes, no
11	Clinical IoT data	Hemoglobin (Hgb)	Numerical	Gms
12	Clinical IoT data	(FIT)	Nominal	Yes, no

detected indicators contain the behavioral or clinical data besides the vital signs which sensed through the installed bio-medical sensors on body via Body Area Network (BAN) to get the essential factors based on Table 2 which were describedin Sect. 3.1. The medical IoT smart devices continuously get and collect the colorectal health indicators in order to detect the potential CRC in patient [6].

3.3 Data preprocessing

The data preprocessing phase is required to clean the sensed medical IoT data from the inconsistencies and the noises for the data mining phase. Furthermore, the feature selection procedures are uses for reducing the dimension of data for easier classification processes in the CRC prediction phase [6].

3.4 CRC prediction

In this study, 400 Iranian aged people have been considered as the samples for evaluating the proposed method. According to the symptoms for the elderlies' colorectal health status, the CRC possibility is classified in four cases including (1) The white status that means No-CRC, (2) the yellow status that means Pre-CRC which means the patient needs more follow up, (3) the orange status that means CRC detected with medium probability, and (4) the red status that means CRC detected with high probability. These different cases are resulted from the monitored values of the colorectal health indicators based on Table 2, and the derived rules for case diagnosis based on the Table 3.

Table 4, shows the tabular presentation of the obtained rules based on the observed symptoms to show the colorectal health status which categorized with white, yellow, orange and red boxes. After detecting the status of the patient, the procedure of CRC prediction is performed based on the evaluation of the indicators comprising of FIT factor, personal history and family history. Figure 2 shows the procedure of the offered scheme for CRC prediction. In the case of No-CRC detection, the patients should be followed up for the potential CRC screening in the future.

4 Experimental Results

For evaluating the effectiveness of the CRC prediction method, five main parameters of precision (Pr.), accuracy (Acc.), f-score, recall (Re.) and execution time are considered. Table 5 refers to the mentioned factors and their descriptions and equations concerning the confusion metrics which are usually employed for measuring the performance of ML classifiers [6].

The confusion metrics contains the instances with four features comprising: (1) *True Positive* (TP) that shows the cases that have been correctly categorized, (2) *False Negative* (FN) specifies the CRC instances that have been detected as normal. (3) *False Positive* (FP) indicates the normal instances that have been grouped as CRC cases and (4) *True Negative* (TN) defines the normal instances that have been classified truly [6].

In the offered model, some classification techniques are used for CRC prediction on collected instances. The outcomes of the classification process achieved by Weka 3.6 with different ML classification techniques such as J48, Sequential Minimal Optimization (SMO), Multi-Layer Perceptron (MLP) and Naïve Bayes (NB) methods. The collected dataset is tested by the training step in the classifier. Also, the k-fold cross validation technique is employed in which the dataset is separated into k mutual-exclusive folds of close equal size in a random way for k times training and testing of the classification process. The accuracy factor is considered as the total number of right

Table 3 Obtained rules for CRC prediction

Extracted rules for colorectal health status and CRC prediction

CASE 1: (Diarrhea = Yes OR Constipation = Yes) And the other parameters = No then STATUS = WHITE;

CASE 2: (Abdominal pain = Yes OR Abdominal Tenderness = Ye OR Weight loss = Yes OR abnormal rectal form = Yes) OR (Abdominal Tenderness = Ye AND Constipation = Yes OR (Abdominal pain = Yes AND Constipation = Yes OR (Abdominal pain = Yes AND Diarrhea = Yes OR (Diarrhea = Yes OR Constipation = Yes) OR (Abdominal pain = Yes AND Abdominal Tenderness = Yes) then STATUS = YELLOW;

CASE 3: (Hgb < 11 AND Gender = Female) OR (Hgb < 12 AND Gender = Male)) OR (Rectal bleeding = Yes)) OR ((Hgb < 11 AND Gender = Female) OR (Hgb < 12 AND Gender = Male)) AND (Rectal bleeding = Yes)) OR ((Hgb < 11 AND Gender = Female) OR Hgb < 12 AND Gender = Male)) AND (Diarrhea = Yes)) OR ((Hgb < 11 AND Gender = Female) OR (Hgb < 12 AND Gender = Male)) AND (Constipation = Yes)) OR ((Hgb < 11 AND Gender = Female) OR (Hgb < 12 AND Gender = Male)) AND (Weight loss = Yes) OR ((Hgb < 11 AND Gender = Female) OR (Hgb < 12 AND Gender = Male)) AND (Abdominal pain = Yes)) OR (Abdominal Tenderness = Yes AND Diarrhea = Yes) OR (Abdominal Tenderness = Yes AND Rectal bleeding = Yes) OR (Abdominal Tenderness = Yes AND Rectal bleeding = Yes) OR (Abnormal Rectal form = Yes AND Constipation = Yes) OR (Abnormal Rectal form = Yes AND Abdominal pain = Yes) OR (Abdominal pain = Yes AND Rectal bleeding = Yes) OR (Abdominal pain = Yes AND Weight loss = Yes) OR (Weight loss = Yes AND Constipation = Yes) OR (Weight loss = Yes AND Diarrhea = Yes) OR (Weight loss = Yes AND Rectal bleeding = Yes) OR (Rectal bleeding = Yes AND Constipation = Yes) OR (Rectal bleeding = Yes AND Diarrhea = Yes) then STATUS = ORANGE;

CASE 4: (Diarrhea = Yes AND Abnormal Rectal form = Yes) OR ((Hgb < 11 AND Gender = Female) OR (Hgb < 12 AND Gender = Male)) AND(Abdominal tenderness = Yes)) OR (Abnormal Rectal form = Yes AND Diarrhea = Yes) OR (Abnormal Rectal form = Yes AND Rectal bleeding = Yes) OR (Abnormal Rectal form = Yes AND Weight loss = Yes) OR (Abnormal Rectal form = Yes AND Abdominal tenderness = Yes) OR (Abdominal tenderness = Yes AND Weight loss = Yes) then STATUS = RED;

classifications, which is divided by the instances in the dataset in cross-validation method.

The results from different classification methods are presented in Figs. 3, 4, 5, 6 and 7 based on some different cross folds values. Typically k-fold cross validation technique is a common procedure in which the folds are provided from the same portion of labels for creating dissimilar datasets. Here, for performance evaluation of the applied classifiers, stratified k-fold cross-validation is employed with values between 1 and 20 for k-folds. As shown in Fig. 3 to Fig. 7, the MLP technique achieved the minimum execution time with a major difference comparing to the other algorithms. The MLP showed an accuracy of 97% with a precision and recall of 94 and 99% and f-score of 96.4%. Also, the lowest exclusion time of 120 ms for MLP classifier shows that it has the highest quality regarding the results. Hence, it can be resulted that in CRC prediction, the MLP classifier has the maximum performance among the others.

5 Conclusion and future work

Regarding the challenges of elderlies' health worries of chronic diseases such as cancers and growing the expenditures of cancer therapy, mainly in developing countries, we present an abstract IoT-based model for colorectal cancer (CRC) predicting and early CRC diagnosis to support the medical teams for on-time decision making about the patients. Since people in such countries hardly access to hospitals and medical care centers for following their colorectal health status, we propose a CRC monitoring and prediction model via applying the benefits of sensor technology in the IoT context. Furthermore, we suggest a

Table 4 Tabular presentation of indicators evaluation for CRC prediction

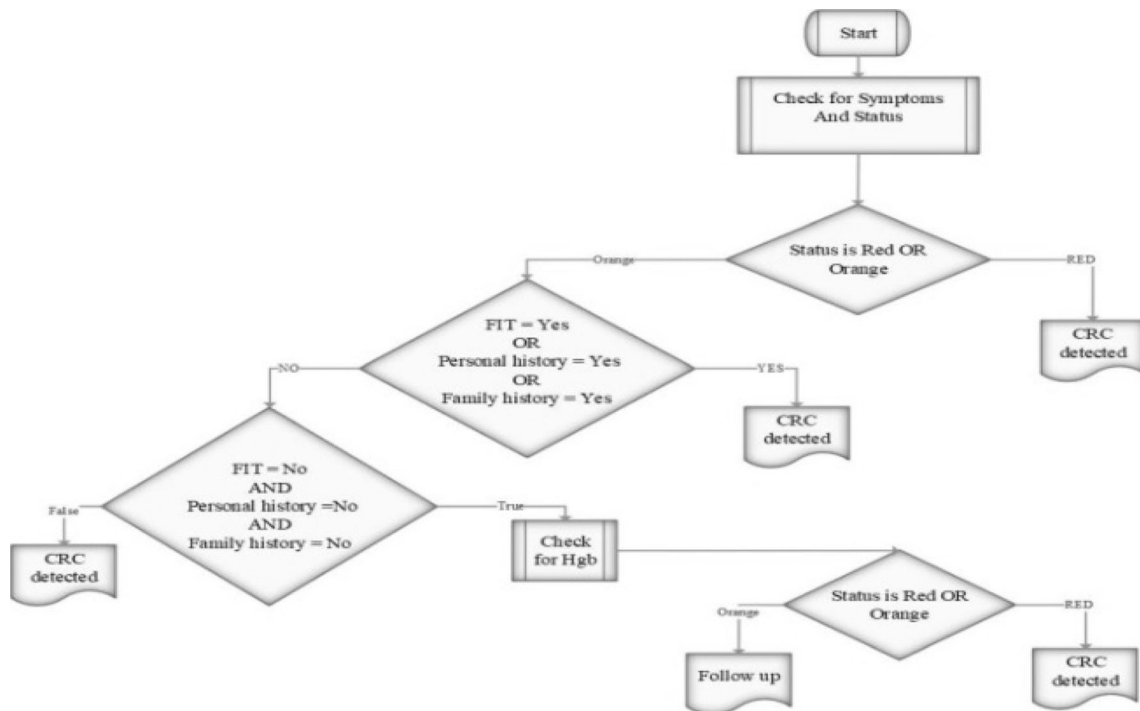|	Hgb < 11 in Female) OR (Hgb < 12 in Male	Abnormal Rectal form	Abdominal tenderness	Abdominal Pain	Weight loss	Rectal bleeding	Diarrhea	Constipation
Each factor alone	ORANGE	YELLOW	YELLOW	YELLOW	YELLOW	ORANGE		
Constipation	ORANGE		YELLOW	YELLOW	YELLOW		YELLOW	
Diarrhea	ORANGE	RED	ORANGE	YELLOW			YELLOW	
Rectal bleeding	ORANGE	RED	ORANGE	YELLOW	YELLOW	RED		
Weight loss	ORANGE	RED	RED	YELLOW				
Abdominal Pain	ORANGE		YELLOW	YELLOW				
Abdominal tenderness	RED	RED	YELLOW	ORANGE	YELLOW			

Fig. 2 The procedure of CRC prediction

Table 5 The performance parameters for measuring the efficiency of the CRC prediction method

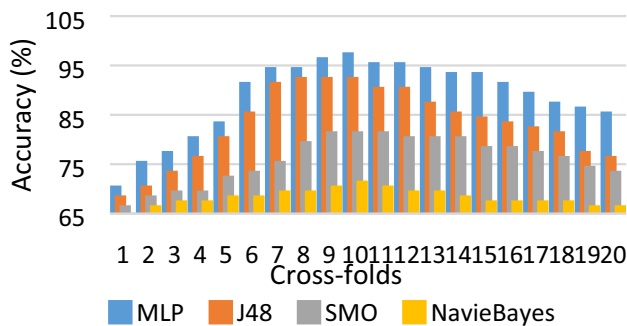| Factors	Acc Equation 1	Pr Equation 2	Re Equation 3	F-score Equation 4
Formula	$\dfrac{TP + TN}{TP + TN + FP + FN}$	$\dfrac{TP}{TP + FP}$	$\dfrac{TP}{TP + FN}$	$\dfrac{2 \times Pr \times Re.}{Pr. + Re.}$

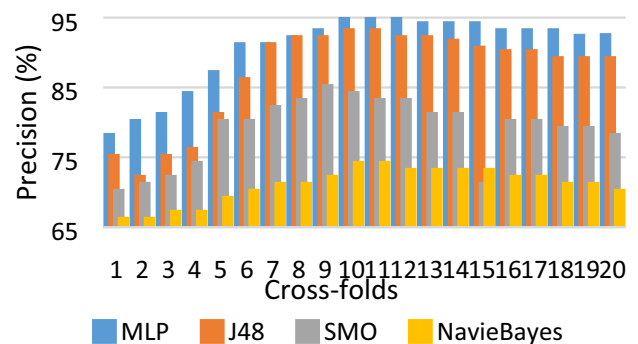**Fig. 3** Accuracy of each algorithm with 20 cross validation fold

Fig. 4 Precision of each algorithm with 20 cross validation fold

feature set with influencing factors on CRC that involve influencing clinical symptoms and vital signs which had been resulted regarding the physicians' clinical observations and previous recent studies on CRC prediction. Our abstract model contains four steps, including determining the influencing indicators on CRC, collecting the IoT data using MIoT devices, preprocessing the collected data, and CRC predication. The applied classification methods included MLP, J48, SMO and Naïve Bayes. The experimental results showed that the MLP method achieved the highest performance comparing to the other methods in all cases. In our future work, we intend to apply more samples to evaluate the usefulness of our proposed model. Also, we plan to develop our abstract diagnostic model in a real IoT platform to evaluate the performance of it.
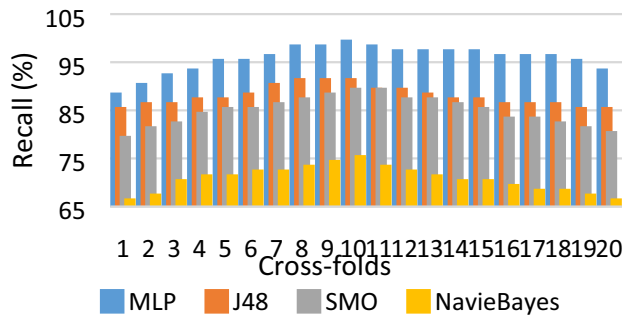
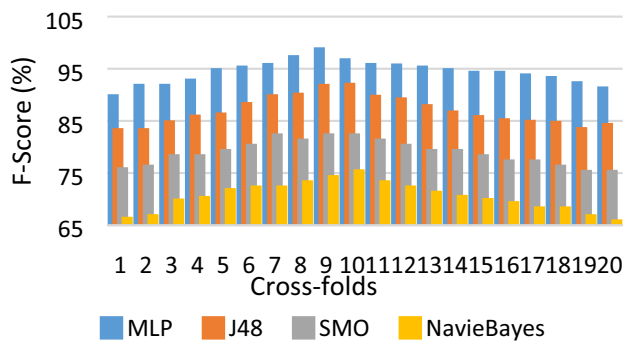**Fig. 5** Recall of each algorithm with 20 cross validation fold

Fig. 6 F-Score of each algorithm with 20 cross validation fold
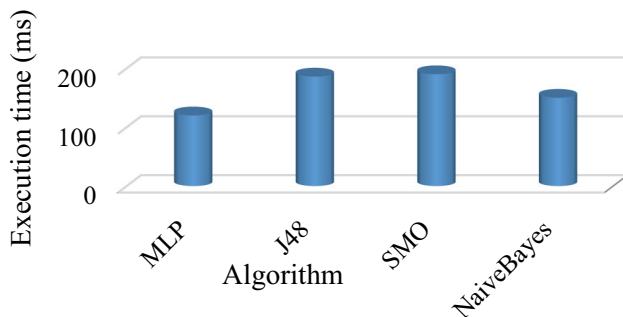

Fig. 7 Average execution time for existing algorithms

References

1. Wang Y, Wang D, Ye X, Wang Y, Yin Y, Jin Y (2019) A tree ensemble-based two-stage model for advanced-stage colorectal cancer survival prediction. Inf Sci 474:106–124

2. Chen H, Zhao H, Shen J, Zhou R, Zhou Q (2015) Supervised machine learning model for high dimensional gene data in colon cancer detection. 2015 IEEE International Congress on Big Data, pp 134–141

3. Hoogendoorn M, Moons LM, Numans ME, Sips R-J (2014) Utilizing data mining for predictive modeling of colorectal cancer using electronic medical records. International Conference on Brain Informatics and Health, pp 132–141

4. Kop R, Hoogendoorn M, Moons LM, Numans ME, ten Teije A (2015) On the advantage of using dedicated data mining techniques to predict colorectal cancer. Conference on Artificial Intelligence in Medicine in Europe, pp 133–142

5. Soguero-Ruiz C, Hindberg K, Mora-Jiménez I, Rojo-Álvarez JL, Skrøvseth SO, Godtliebsen F et al (2016) Predicting colorectal surgical complications using heterogeneous clinical data and kernel methods. J Biomed Inform 61:87–96

6. Asghari P, Rahmani AM, Haj Seyyed Javadi H (2019) A medical monitoring scheme and health-medical service composition model in cloud-based IoT platform. Transac Emerg Telecommun Technol 30:e3637

7. Hornbrook MC, Goshen R, Choman E, O'Keeffe-Rosetti M, Kinar Y, Liles EG et al (2017) Early colorectal cancer detected by machine learning model using gender, age, and complete blood count data. Dig Dis Sci 62:2719–2727

8. Zalake M, Tavassoli F, Griffin L, Krieger J, Lok B (2019) Internet-based tailored virtual human health intervention to promote colorectal cancer screening: design guidelines from two user studies. Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, pp 73–80

9. Cooper JA, Ryan R, Parsons N, Stinton C, Marshall T, Taylor-Phillips S (2020) The use of electronic healthcare records for colorectal cancer screening referral decisions and risk prediction model development. BMC Gastroenterol 20:1–16