International Journal of Scientific Research in Engineering and Management (IJSREM)

Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

COLON CANCER IDENTIFICATION USING CNN INTEGRATED PARTICLE SWARM OPTIMIZATION

MADHAN.S , JANARTHANAN.R ,AMIRTHA.C , SINDHUKAVI.R

Department of Computer Science and Engineering - University College of Engineering Thirukkuvalai, Nagappattinam , Tamilnadu , India.
(A Constituent College of Anna University::Chennai and Approved by AICTE, New Delhi)

Abstract

The Colon cancer prompts to more number of death as of late. The diagnosis of colon cancer as early is protected to treat the patient. To distinguish and treat this type of malignancy, Colonoscopy is applied normally. Several risk prediction models for colon cancer have been created and approved in various populations but colon cancer effecting the young adults. In this research, we projected a Supervised Learning Technique for detecting colon cancer in high dimensional information. One of the most important and very popular tool for performing the machine learning tasks that Includes novelty detection, classification or regression is Particle Swarm Optimization . PSO algorithm simulates animal's social behaviour , including insects , herds , birds and fishes.A more precise classification of medical data can effectively predict the development of colon cancer, according to the findings of numerous research Rapid technical breakthroughs in the fields of image processing and machine learning have resulted in the development of a slew of quick computer-assisted diagnostic methods.The proposed system utilizes Modified PSO algorithm to locate colon cancer with higher precision rate.

Keywords

Colon Cancer, Particle Swarm Optimization(PSO), Machine Learning , Convolutional Neural Network (CNN),Colon Cancer prediction.

Introduction

Colon Cancer could be a substantial public unhealthiness and also the global incidence of this cancer has risen quickly with increment . World health Organization (WHO) GLOBOCAN database study 2020 reported 106,180 new cases of carcinoma.Worldwide , carcinoma is that the 3rd most diagnosed cancer. And estimated 1,880,725 people where diagnosed with carcinoma in 2020. The WHO reported carcinoma is that the 3rd leading reason behind cancer-related deaths in men and in women , and therefore the second most typical reason behind cancer deaths when numbers for men and ladies are combined . It's expected to cause about 52,580 deaths during 2022 .In this context, feature selection is commonly considered as a necessary preprocess step to research these data, as this method can reduce the dimensionality of the datasets and sometimes conducts to raised analyses [6].Two models of feature selection exist counting on whether the choice is including a learning scheme or not. The first one, filter model, which carries out the feature subset selection and also the classification in two separate phases,uses a measure that's simple and fast to compute

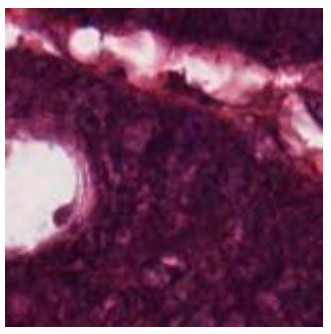
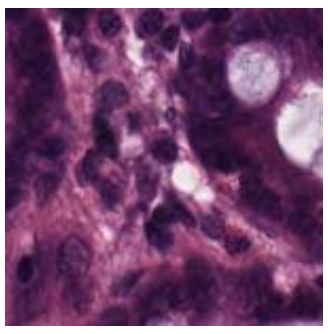
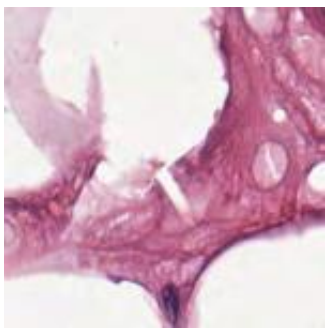
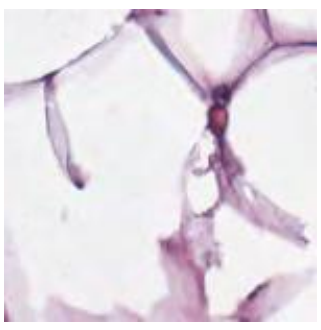
Data collection

In this phase and open source disease dataset has been taken UCI repository . This dataset contains both the
photographs of carcinoma affected region. This dataset is in an unprocessed form . The dataset need to be preprocessed . during

this dataset contains 5000 images of carcinoma affect region. This are accustomed training purpose
.

Dataset

We use pixel base dataset in this project . totally 5000 snaps are used to training purpose and it refers the seven types of colon cancer like Tumor , Stroma , Complex , Limphio , Debris , Mucosa , Adipose and these type of affected region's snaps are used to train the machine learning for the purpose of increase the accurate level.To emulate these cognitive abilities, computer vision algorithms make heft use of collections of images called datasets. A dataset in computer vision may be a curated set of digital photographs that developers use to check, train and evaluate the performance of their algorithms

Proposed System

The initial stage is data pre-processing.This process took part in noise reduction and image reconstruction. In case any noise recorded after pre-processing includes additional filters. For this, median filtering approach is presented. The process of random noise reduction maintains the edges of images in an acceptable way. By sorting pixel vale from neighbourhood and it would replace middle pixel value. The image delivered from above stage is delivered to post-processing. This includes a step by step procedure to detect cancer.The modified PSO algorithm should function upon trained dataset.Finally, colon cancer is identified.

Literature survey

1.Enrique Alba, José García-Nieto, Laetitia Jourdan, El-Ghazali Talbi. "Gene selection in cancer classification using Particle Swarm Optimization/Sopprt Vector Machine & Genetic Algorithm/Support Vector Machine hybrid algorithms" Congress on Evolutionary Computation, Singapore : Singapore (2007)

During this work we compare the utilization of a Particle Swarm Optimization (PSO) and a Genetic Algorithm (GA)(both augmented with Support Vector Machines SVM) for the classification of high dimensional Micro array Data. Both algorithms are used for finding small samples of informative genes amongst thousands of them. A SVM classifier with ten-fold cross-validation is applied in order to calculate and evaluate the provided solutions.A first grant is to prove that PSOSVM is able to find interesting genes and to give classification competitive performance. It's not usually, a new version of PSO, called Geometric PSO, is empirically evaluated for the 1st

time in this work using a binary representation in Hamming value. In this sense, a comparison of this approach with a replacement GASVM and also with other existing methods of literature is provided. A second important contribution consists within the actual discovery of recent and challenging results on six public datasets identifying significant within the development of a spread of cancers (leukemia, breast, colon, ovarian, prostate,and lung)

2.Lingyun Gao, Mingquan Ye, and Changrong Wu. "Cancer Classification Based on SVM Optimized by PSO and Artificial Bee Colony" Molecules 2017 Dec; 22(12)

Intelligent optimization algorithms have advantages in handling complex nonlinear problems among good flexibility and flexibility. during this paper, the FCBF (Fast Correlation-Based Feature selection) method is employed to filter irrelevant and redundant features so as to enhance the standard of cancer classification. Then, we perform classification supported SVM (Support Vector Machine) optimized by PSO (Particle Swarm Optimization) combined with ABC (Artificial Bee Colony) approaches, which is represented as PA-SVM.. By comparison with other classification methods, the results demonstrate the effectiveness and therefore the robustness of the proposed PA-SVM method in handling various sorts of data for cancer classification.

3.Maolong Xi, Jun Sun, Li Liu, Fangyun Fan, and Xiaojun Wu. "Cancer Feature Selection and Classification Using a BQ-Behaved Particle Swarm Optimization and Support Vector Machine" Computational and Mathematical Methods in Medicine 2016

This paper focuses on the feature gene selection for cancer classification, which employs an optimization algorithm to pick a subset of the genes. We propose a binary quantum-behaved particle swarm optimization (BQPSO) for cancer feature gene selection, coupling support vector machine (SVM) for cancer classification. First, the proposed BQPSO algorithm is described, which may be a discretized version of original QPSO for binary 0-1 optimization problems. Then, we present the principle and procedure for cancer feature gene selection and cancer classification supported BQPSO and SVM with leave-one-out cross validation (LOOCV). Finally, the BQPSO coupling SVM (BQPSO/S VM), binary PSO coupling SVM (BPSO/SVM), and genetic algorithm coupling SVM (GA/SVM) are tested for feature gene selection and cancer classification on five micro array data sets, namely, Leukemia, Prostate, Colon, Lung, and Lymphoma. The experimental results show that BQPSO/SVM has significant advantages in accuracy, robustness, and therefore the number of feature genes selected compared with the opposite two algorithms.

4.A. S. M. Shafi, M. M. Imran Molla, Julakha Jahan Jui & Mohammad Motiur Rahman. "Finding of colon cancer based on proteomics dataset using machine learning as a feature selection and classification techniques" SN Applied Sciences, volume 2, (2020)

Micro array data is an increasingly important tool for providing information on organic phenomenon for analysis and interpretation. Researchers try to utilize the littlest possible set of relevant organic phenomenon profiles in most organic phenomenon studies to boost tumor identification accuracy. This research aims to research and predicts carcinoma data employing a machine learning approach and have selection technique supported a random forest classifier. More particularly, our proposed method can reduce the burden of high dimensional data and permit faster calculations by combining the "Mean Decrease Accuracy" and "Mean Decrease Gini" as feature selection methods into a renowned classifier namely Random Forest, with the aim of accelerating the prediction model's accuracy level. additionally, we've also shown a comparative model analysis with selection of features and model without selection of features. The large experiment results have exhibit that the proposed model with feature selection is favorable.

5.Jing-Jing Wan, Bo-Lun Chen, Yi-Xiu Kong, Xing-Gang Ma & Yong-Tao Yu. "An initialy Intestinal Cancer Prediction Algorithm Based on DBN" Scientific Reports volume 9, (2019)

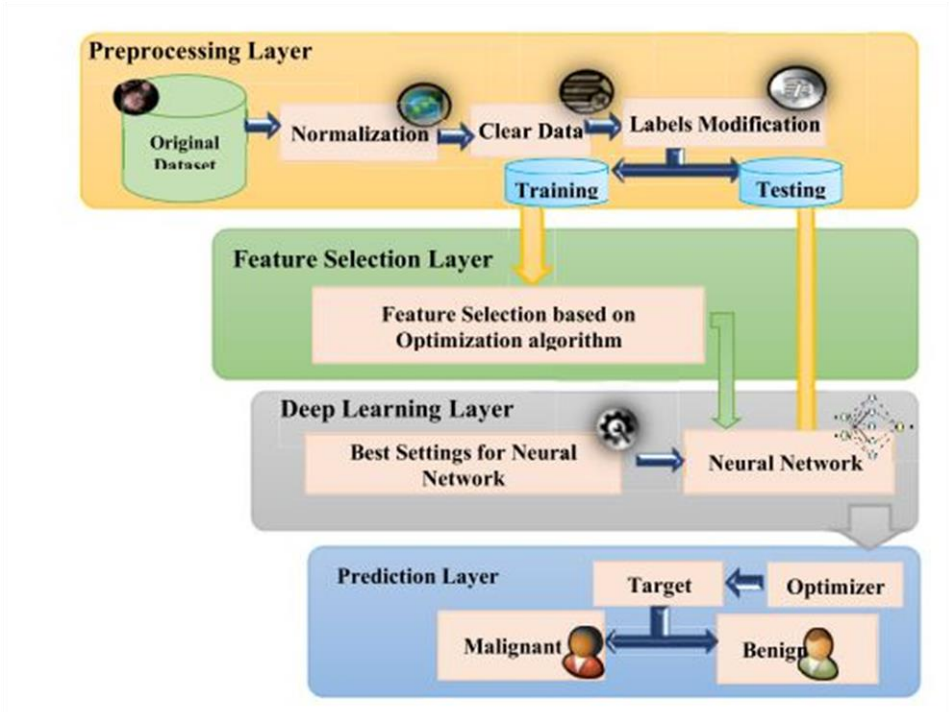
The incidence of colorectal cancer (colorectal cancer, CRC) in China has increased in recent years, and its fatality rate has become one amongst the best among all cancers. CRC also increasingly affects people's health and quality of life, and also the workloads of medical doctors have further increased because of the shortage of sufficient medical resources in China. The goal of this study was to construct an automatic expert system employing a deep learning technique to predict the probability of early stage CRC supported the patient's case report and therefore the patient's attributes. Compared with previous prediction methods, which are either supported sophisticated examinations or have high computational complexity, this method is shown to produce valuable information like suggesting potentially important early signs to help in early diagnosis, early treatment and prevention of CRC, hence helping medical doctors reduce the workloads of endoscopies and other treatments.

Machine Learning

A more precise classification of medical data can effectively predict the development of colorectal cancer, according to the findings of numerous research. As a result, utilizing recent Machine Learning progress, it is critical to build diagnosis procedures that are both successful and low-cost.In this paper we used particle swarm optimization (PSO), it is a computational method that optimizes a controversy by iteratively trying to boost a candidate solution with relevancy a given measure of quality

Particle Swarm Optimization (PSO)

flyspeck mass optimization is a stochastic optimization fashion grounded on the movement and intelligent of masses. PSO is used to find the outside or minimum of a function defined on a multidimensional vector space. Such a big quantum of data ca n't be reused by the experts in a short time to make opinion, prognostic and treatment schedules in quick time. PSO searches for the optimal result by streamlining the haste and the position of the flyspeck



Architecture

In our frame, we've proposed a procedure that's separated into colorful stages as appeared in below

1. Preprocessing subcaste
2. point selection subcaste
3. Deep literacy subcaste

4. vaticination subcaste

In the preprocessing subcaste, the original dataset will need to be regularized for reduce the duplicated dataset and also the cleared dataset will go to the coming subcaste of point selection, in this stage the dataset is named essential point by the flyspeck mass optimization, and also the essential features is classified by the convolutional neural network to get the result, in this last subcaste the result will be shown. Relative Study relative analysis of point selection applying pso with bracket algorithms

Comparative Study

Comparative analysis of feature selection applying pso with classification algorithms:

Algorithm	Dataset	Instances	Attributes	Accuracy(%)
	Wisconsin Breast Cancer	699	11	95.85
	Pima Indians Diabetes	768	8	91.27
	Heart-Statlog	270	13	89.96
	Breast Tissue	106	10	63.16
	Haberman's Survival	306	3	72.53
PSO based fast K-means algorithm	Hepatitis	155	20	88.3
PSO based fast K-means algorithm	KDD cup 2008 for Breast cancer data	102924	117	99.39
PSO+ SVM	WDBC	569	30	95.61
DNPSO and DP AC	Dermatology	366	33	92.61
PSO-SFS and PSO SFS-SBS	Liver	346	6	96.4
PSO + Neural Network Feed Forward Back Propagation	Cleveland Heart Disease	296	13	91.94
PSO+GA	Wiscon	699	11	90.1
	sin Breast Cancer (Original)			
PSO+ Boosted C5.0	Breast Cancer Wisconsin	569	31	96.38
PSO+ Convolutional neural network	Colon cancer pixel dataset	5000	11	99.89

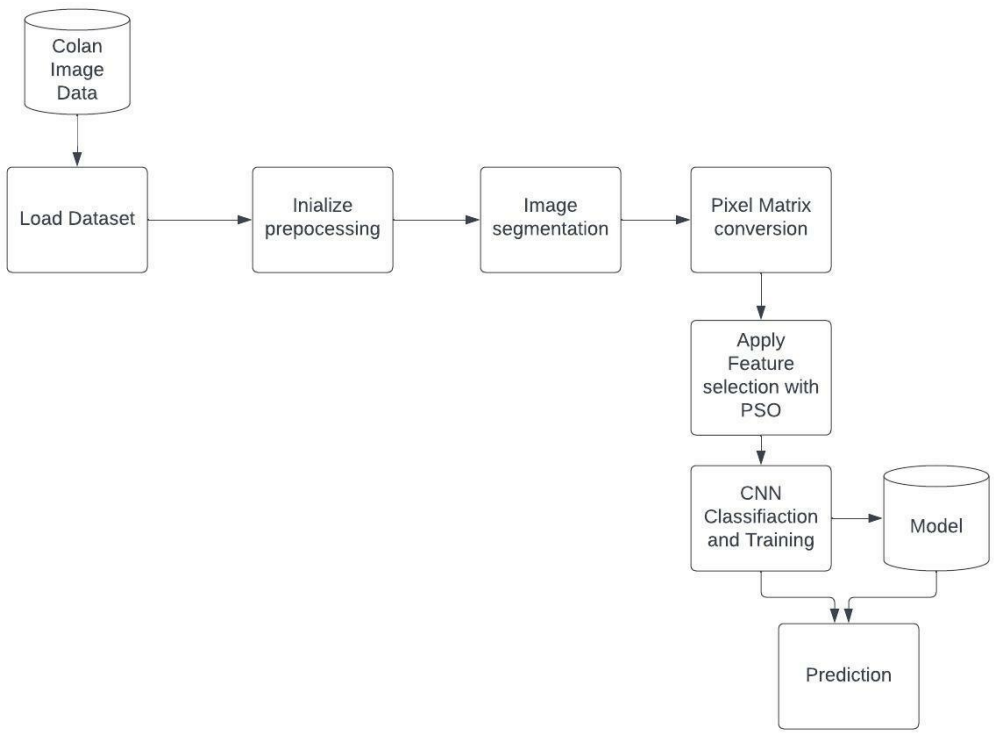
Colorful experimenters are performed comparison on PSO point selection to prompt better bracket delicacy using colorful classifiers. This section compares comparison summary of varied experimenters. Table 1 provides comparison summary of colorful experimenter's conclusion. From the comparison summary, experimenter's opinion on better medical data bracket system isn't unswerving. Further we'd like to study rainfall " medical data bracket " styles performance depends on test dataset. By assaying the results achieved from Table 1, we set up that the rigorousness of the point subset named using the suggested PSO grounded bracket algorithms and have

selection applying PSO also applying any bracket

algorithm is on top of rigor of other styles. likewise, it's putatively reveals that PSO and its combined classifiers had attained advanced bracket for the medical dataset with great number of attributes. either, this approach applied on KDD mug 2008 for melanoma data and Central Nervous System has produced better delicacy while these datasets with gigantic number of cases. therefore PSO grounded colorful bracket models are quantitatively and qualitatively doable.

System overview

Way related from image processing to descry the colon cancer and chance of what kind of complaint that is.

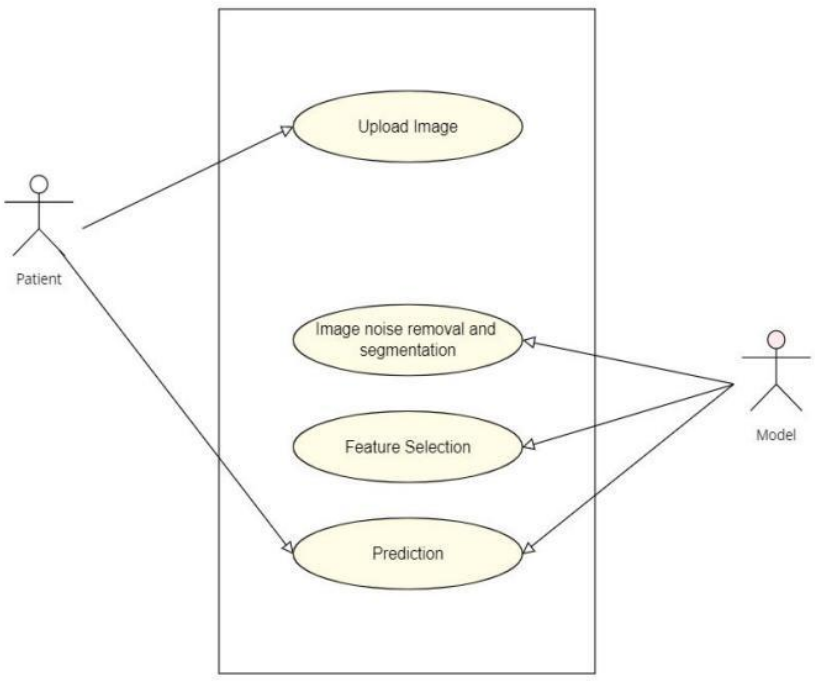


1. Need to preprocess the un-preprocessed data set or the original data set, And also normalization the original dataset
2. Markers the dataset after the normalization process for the coming step of the training and testing purpose.

3. The diagnosing and treating of complaint in a health care unit originally starts with collection of image of mortal body portion.
4. The procesed involved in it are pre-processing, filtering, point birth and bracket.
5. Next the captured images are re-sized to make the computational device to reuse on it.
6. Other than boundary reduction, the necessary features collections among named image is to be done.

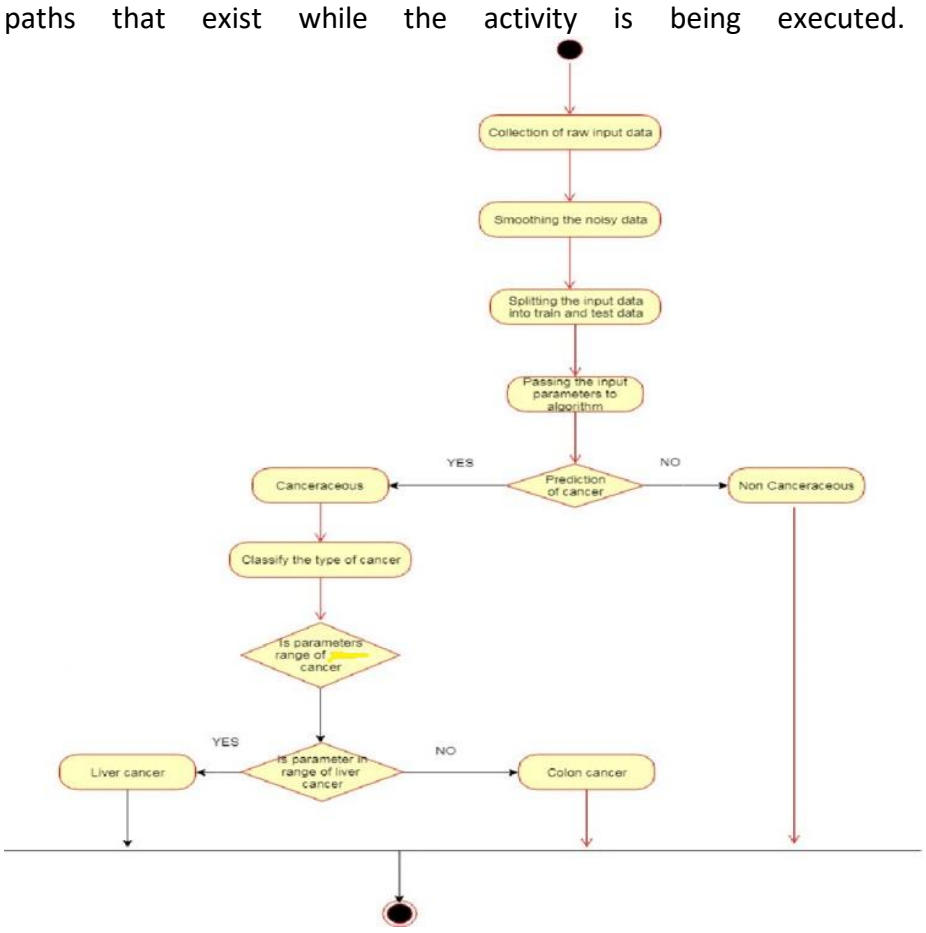
Use case diagram

A use case diagram may be a dynamic or behaviour diagram in UML. Use case diagrams model is describe the functionality of a system using actors and use cases. Use cases are a collection of actions, services, and functions that the system must perform.



ACTIVITY DIAGRAM

An activity diagram could be a behavioural diagram i.e. it depicts the behaviour of a system. An activity diagram portrays the control ensue a start point to a finish point showing the varied decision paths that exist while the activity is being executed.



CONCLUSION

Computer-aided diagnosis systems may become one amongst the foremost useful applications of deep learning. A technology that improves the detection of tumors for colorectal cancer will be easily adapted for detecting many other styles of diseases. Given the shortage of pathologists, this kind of software could also be ideal for screening cases to pick the most effective regions for pathologists to target when making diagnoses of cancer. It may be integrated into genomic analytic pipelines to scale back pathologist workload and improve turnaround times of genomic profiling.

REFERENCE

1. World Health Organization https//www.who.int/
2. DarrenR. Brenner, HannahK. Weir, AlainA. Demers, LarryF. Ellison, Cheryl Louzado, Amanda Shaw, Donna Turner, RyanR. Woods and LeahM. Smith; for the Canadian Cancer Statistics Advisory Committee. Projected estimates of cancer in Canada in 2020, CMAJ March 02, 2020 192(9) E199- E205; DOI https//doi.org/10.1503/cmaj.191292
3. Public Health Agency of Canada; Statistics Canada; Canadian Cancer Society; parochial/ territorial cancer registries. Release notice- Canadian Cancer Statistics 2019. Health Promot Chronic Dis Prev Can. 2019 Sep; 39(8- 9) 255. doi10.24095/ hpcdp.39.8/9.04. PMID 31517469; PMCID PMC6756131
4. Global Cancer Observatory(GCO) Cancer Today, Cancer Factsheets. https//gco.iarc.fr/ moment/ data/ factsheets/ populations/ 124- canada- fact-sheets.pdf
5. Canadian Cancer Society 6 statistics that reveal the impact of cancer in Canada for 2020. https//www.cancer.ca/en/about-us/our-stories/6-statistics-about-cancer-in-canada-for2020/?region=bc
6. Canadian Cancer Society Childhood cancer statistics. https//www.cancer.ca/en/cancer-information/cancer-101/childhood-cancer-statistics/?region=on
7. Stages of Cancer, https//www.webmd.com/cancer/cancer-stages#1
8. Shankar Prinja, Nidhi Gupta and Ramesh Verma." laundering in clinical trials Review of Survival analysis ways." Indian J Community Med 35,no. 2(2010) 217.
9. TG Clark, MJ Bradburn, SB Love and DG Altman. “ Survival Analysis introductory generalities and first analysis. ” British Journal of Cancer 89, 2003 232
. Swaminathan R and Brenner H, Statistical styles for cancer survival analysis,Vol. 2, 2011. https//survcan.iarc.fr/survivalchap2.php
11. Abadi, Alireza, Parvin Yavari, Monireh Dehghani- Arani, Hamid Alavi- Majd, Erfan Ghasemi, Farzaneh Amanpour, and Chris Bajdik." Cox models survival analysis supported melanoma treatments." Iranian journal of cancer forestallment 7,no. 3(2014) 124.
12. Manish Devgan, Gaurav Malik, Deepak Kumar Sharma, Machine literacy & Big data, Semisupervised Learning, 2020, doi10.1002/ 9781119654834. ch10
13. Xiaojin Zhu; Andrew Goldberg, preface toSemi-Supervised Learning, Morgan & Claypool, 2009, doi10.2200/ S00196ED1V01Y200906AIM006.
14. Ecloudvalley digital technology co Ltd “ Types of machine literacy styles ” https//www.ecloudvalley.com/mlintroduction/
15. Abdulaziz Aflakseir and Parinaz Abbasi. Health Beliefs as Predictors of melanoma Webbing geste in an exceedingly Group of womanish workers in Shiraz, Iranian Journal of Cancer Prevention, 2012; 3124- 129
16. Canadian Partnership Against Cancer Canadian Strategy for Cancer Control. https//www.partnershipagainstcancer.ca/cancer-strategy/
17. Thongpim, Nattawut, Chidchanok Choksuchat, Tanan Bejrananda, and Sureena Matayong." On Predicting Survival openings for glandular melanoma by COX Retrogression in PSU Cases Data." In 2020 17th International Conference on Electrical Engineering/ Electronics, Computer, Telecommunications, and data Technology(ECTI- CON),pp. 775- 778. IEEE, 2020
. Chai, Hua, Zi- na Li,De-yu Meng, Liang- yong Xia, and Yong Liang." A new semi- supervised literacy model combined with cox and sp- posterior models in cancer survival analysis." Scientific reports 7,no. 1(2017) 1- 12.

19. Wang, Qingyong, Liang- Yong Xia, Hua Chai, and Yun Zhou."Semi-supervised literacy with ensemble tone- training for cancer bracket." In 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Dispatches, Cloud & Big Data Computing, Internet of individualities and Smart City Innovation(SmartWorld/ SCALCOM/ UIC/ ATC/ CBDCom/ IOP/ SCI),pp. 796- 803. IEEE, 2018.
20. Qiu, JohnX., Shang Gao, Mohammed Alawad, Noah Schaefferkoetter, Folami Alamudun, HongJun Yoon, Xiao- Cheng Wu, and Georgia Tourassi."Semi-Supervised Information birth for Cancer Pathology Reports." In 2019 IEEE EMBS International Conference on Biomedical & Health Informatics(BHI),pp. 1- 4. IEEE, 2019.
21. Kabir, Md Faisal, and SimoneA. Ludwig." Bracket Models and Survival Analysis for adenocarcinoma Using RNA Sequencing and Clinical Data." In 2019 IEEE International Conference on Big Data(Big Data),pp. 2736- 2745. IEEE, 2019.
22. Huang, HaiHui, and Yong Liang." A new Cox commensurable hazards model for high- dimensional genomic data in cancer prognostic." IEEE/ ACM Deals on Computational Biology and Bioinformatics(2019).
23. Nezhad, Milad Zafar, Najibesadat Sadati, Kai Yang, and Dongxiao Zhu." A deep active survival analysis approach for perfection treatment recommendations operation of adenocarcinoma." Expert Systems with operations 115(2019) 16- 26.
24. Katzman, JaredL., Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger." DeepSurv individualized treatment recommender system employing a Cox commensurable hazards deep neural network." BMC medical exploration methodology 18,no. 1(2018) 24.
25. Hirozawa, Tatsuki, Takeshi Yamada, and Hayato Ohwada." New survival vaticination system for terminal cases supported machine literacy." In 2018 IEEE International Conference on Bioinformatics and Biomedicine(BIBM),pp. 2756- 2758. IEEE, 2018.
26. Yang, Cheng- Hong, Sin- Hua Moi, Fu Ou- Yang, Li- Yeh Chuang, Ming- Feng Hou, and Yu- Da Lin." relating threat Position related to a Cancer for Overall Survival by Deep Learning- Grounded CoxPH." IEEE Access 7(2019) 67708- 67717.
27. Barsainya, Aditya, Anusha Sairam, and AnnapurnaP. Patil." Analysis and vaticination of Survival after Colorectal Chemotherapy using Machine Learning Models." In 2018 transnational Conference on Advances in Computing, Dispatches, and Informatics(ICACCI),pp. 862- 865. IEEE, 2018.
28. Cai, Zhuqing, Zhuliang Yu, Haiyu Zhou, and Zhenghui Gu." The early stage melanoma prognostic vaticination model supported support vector machine." In 2018 IEEE 23rd International Conference on Digital Signal Processing(DSP),pp. 1- 4. IEEE, 2018.

29. Imani, Farhad, Ruimin Chen, Conrad Tucker, and Hui Yang." Random Forest Modeling for Survival Analysis of Cancer Recurrences." In 2019 IEEE 15th International Conference on robotization Science and Engineering(CASE),pp. 399- 404. IEEE,.