Future Computing and Informatics Journal

Volume 4
Issue 2 *2019 Volume 4, Issue 2*

Article 6

2019

A Comparative Study for Methodologies and Algorithms Used In Colon Cancer Diagnoses and Detection

Mona Mohamed Nasr
Faculty of Computers and Information, m.nasr@helwan.edu.eg

laila mohamed abdelhamid
helwan university, dr.laila.abdelhamed@gmail.com

Naglaa Shehata
Helwan University, nagla_sd@yahoo.com

A Comparative Study for Methodologies and Algorithms Used In Colon Cancer Diagnoses and Detection

Mona Nasr[a], Laila Abdelhamid[b], and Naglaa Shehata[c]
Faculty of Computers and Artificial Intelligence, Helwan University, Egypt

[a]am.nasr@helwan.edu.eg [b]dr.laila.abdelhamed@gmail.com [c]nagla_sd@yahoo.com

A B S T R A C T

Colon cancer is also referred to as colorectal cancer, a kind of cancer that starts with colon damage to the large intestine in the last section of the digestive tract. Elderly people typically suffer from colon cancer, but this may occur at any age.It normally starts as little, noncancerous (benign) mass of cells named polyps that structure within the colon. After a period of time these polyps can turn into advanced malignant tumors that attack the human body and some of these polyps can become colon cancers. So far, no concrete causes have been identified and the complete cancer treatment is very difficult to be detected by doctors in the medical field. Colon cancer often has no symptoms in early stage so detecting it at this stage is curable but colorectal cancer diagnosis in the final stages (stage IV), gives it the opportunity to spread to different pieces of the body, difficult to treat successfully, and the person's chances of survival are much lower. False diagnosis of colorectal cancer which mean wrong treatment for patients with long-term infections and they are suffering from colon cancer this causing the death for these patients. Also, the cancer treatment needs more time and a lot of money. This paper provides a comparative study for methodologies and algorithms used in colon cancer diagnoses and detection this can help for proposing a prediction for risk levels of colon cancer disease using CNN algorithm of the deep learning (Convolutional Neural Networks Algorithm).

Keywords:Colorectal Cancer Deep Learning Algorithms, Convolutional Neural, Network,Polyps

1. Introduction

Cancer is the world's most known highly cause of death on the earth Planet all over the world and the Level of colon and rectal cancer infections is augmenting in developing countries.

As per ongoing studies, cancer growth is ascribed to around 9.6 million dying around the

Published by Arab Journals Platform, 2019

world, which makes it the second most deadly illness.

Cases of cancer infections and deaths worldwide a re rising, the number of deaths has been estimated according to the GLOBOCAN 2018,
where new cancer cases reach to 18.1 cases and the numbers of cancer death reach to 9.6 million deaths. The infection rate of colorectal cancer reaches to 9.2%. The incidence and the number of colon cancer deaths vary from one country to another according to the economic situation of each country and its level of progress. (F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, 2018)
Colon cancer cases differ according to the economic and social situation of each country. In the US, for example, 101,420 new cancer cases are being affected by colon cancer and 60,680 deaths, with a death rate of 1700 cases per day. (P. Rebecca L. Siegel, MPHKimberly D. Miller, MPHAhmedin Jemal, DVM, 2019)

Previous studies say that there are 9.5million cancer demises worldwide and the anticipated future would have 13 million deaths by growth in 2030. (K. S. Sankari and M. Logambal,2018) The prediction of cancer at early stage has a very significant role in dropping deaths originates by cancer. Therefore, Knowing family history, medical records, and factors hereditary to cancer and environmental factors is very important in the judiciary cancer prevention, and avoid its risks.
Cancer is a threat that threatens the entire world and threatens human life. Colon cancer sometimes is known as "bowel cancer" or "colorectal cancer" and defined as

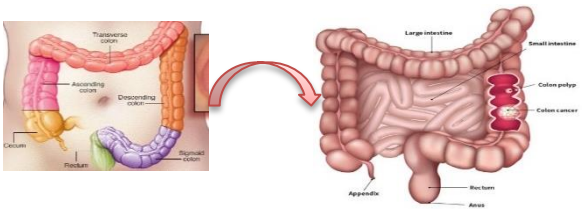

Fig.1. Colon cancer polyps

Abnormal cell growth which can affect other parts of the body and spread in many places.Colon canc er as figure 1 is a malignant development for large intestine (colon) cells.
Most cases beginning with a small non-cancerous or benign group of cells called adenomatous polyps, which is the last component of the digestive tract. If polyps accumulate and increase over time, they may lead to colon cancer. (K. S. Sankari and M. Logambal,2018)
Colon cancer cells cause damage to the healthy tissue near the tumor,
that results in cancer spread, colon wall penetratio n and deteriorating health as shown in figure2
that specifies the difference between polyp-combined colon cancer and regular colon without polyps. Such cancer cells can develop in various ways, invading and killing other healthy tissue around the body. When the tumor spreads in many parts, it is called a malignant tumor and it is difficult to treat and recover from it. Within a rectum near to the anus, a few inches from the large intestine, malignant development begins. Colon cancer is the third largest cancer in the world and has grown every year. The causes of colon cancer are not explicit in most cases. Doctors know that colon will infect healthy human cells by cancer and disrupted from their functions and attacked by other foreign cells on the body in their genetic blueprint, the DNA see figure 2. (K. S. Sankari and M. Logambal,2018)

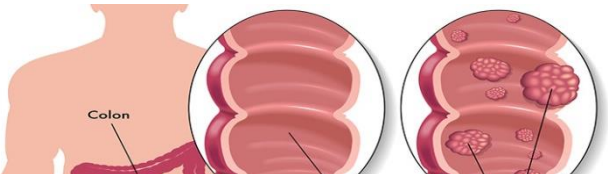

Fig.2. Differences Between The Normal Colon And Cancer Colon (Goldberg,2019)

1.1 Cancer Stage Classifications

Colon cancer consists of four stages, each stage has its symptoms and has its own methods of prevention, so it is necessary to study and know each stage of colon cancer to help in treatment. Figure 3 illustrates how the polyps develop at each stage with the four stages of colon cancer inside o r outside the colon. *Stages of colon cancer (* M. DePietro, 2019*)* (Christina Chun,2019*) as figure 3*

Stage0: This period may be called the stage of non-cancer which indicates that it did not grow outside the mucosa or the colon's inner layer.

Stage1:

The infection in the inner part of the colon enters t he mucos membrane at this stage but the lymph no des have not spread or become infected.

Stage2:

Colon cancer begins to form at this level and the c ancer reaches the mucous membrane and colon me mbrane and spreads to the adjacent lymph nodes a nd tissues but it does not fully expand.

There are three internal phases in this phase: 2a, 2b and 2c. The lesion did not split into the lymph hubs in Step 2b, but into the instinctive peritoneum emerged in the outer colon layer, and this is the film that carries Abdominal organs, and in arrangement 2 c be Ales Tan did not spread to nearby lymph hubs However, in addition to development through the colon's outer layer, neighboring organs or structures have evolved.

Stage 3: In this step, colon cancer risks and threat to people's lives this stage consists of 3 internal phases 3a, 3b, or 3c. phase 3a is a tumor that formed into the colon's muscle layers, located in adjacent lymph hubs, which did not spread to the hubs or distal organs.

During the colon's extraordinary layers, it joins the instinctive peritoneum or assaults Specific organ or structures, and the tumor has formed in arrangement 3c outside the layer of muscles, to place it in four or more neighboring lymph nodes but not far apart.

Stage 4: This stage is the advanced stage of infection that is hard to treat, hard to heal, and a damage human's life. It is composed of two 4s and 4b phases. Phase 4a is the most advanced stage in a colorectal cancer that is spread into two or more areas, such as the lunges and liver, while Phase 4a spreads to a distant area, such as the liver and lungs.

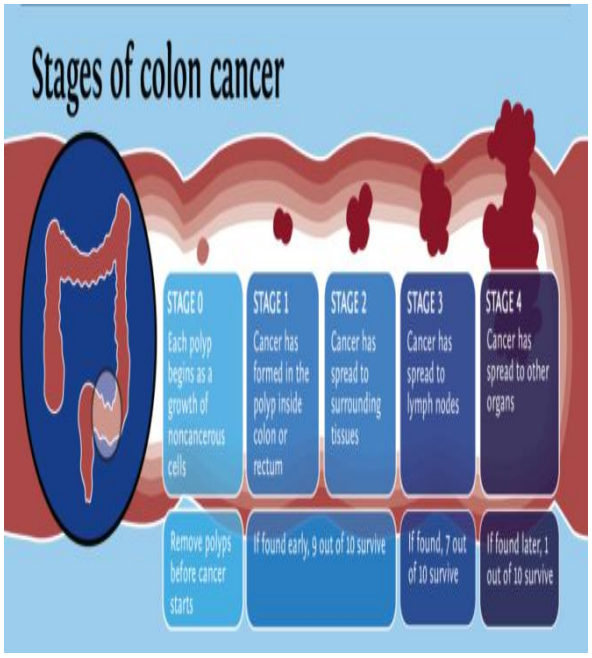

Fig. 3 Colon Cancer Awareness (M. Melissa Buffalo,2019)

Colon cancer is graded as small or large in relation to staging, and The pathologist calculates a number from 1 to 4, depending on the amount of cells that look like healthy cells, while analyzing cancer cells under a microscope. Although, low-grade cancers appear to rise steadily compared with high-grade cancers; this is better diagnosed for people with low-grade colon cancer. (D. J. Ahnen *et al.*,2014)

Examinations to verify the stage of colon cancer

Clinicians use a colonoscopy procedure for the diagnosis of colon cancer, so colonoscopy means an operator has a wide narrow tube on the inside of the colon with a small camera. If colon cancer is found, further testing may be needed, such as a cT scan, X-ray or MRI, to assess the size or extent of the tumor and to assess if it has spread beyond the colon.

Published by Arab Journals Platform, 2019

Some cases may not to be fully determined for any cancer stage of disease until after surgery of the colon and after the surgery the doctor can check for the main tumor along with the removed lymph nodes to help in the determination of the stage of illness

1.2 Colon cancer threat factors

There are many Factors that can raise the risk of colon cancer. Can explain these factors according to (K. S. Sankari and M. Logambal,2018) (N. Singh, S. Kumar, and S. Bhadauria, 2016):

• **Age**. The majority of those suffering from colon cancer are more than fifty years old, but colon cancer can be diagnosed at any age. Colonic cancer in people less than 50 now has risen so, at present many youth can infect with colon cancer. (F. Bénard, A. N. Barkun, M. Martel, and D. Von Renteln,2018)

• **Inflammatory intestinal conditions**. Continuous colon inflammatory illness, for example, ulcerative colitis and Crohn's infection, can establish the risk of colon malignancy.

• **An individual history of colon disease or polyps**. In the event that a person has previously had colon problems from benign tumors or polyps infections, this exposes him to the risk of colon cancer disease.

• **Acquired disorders that expansion colon cancer disease chance**. Many changes in quality have been made over years of family age may fundamentally cause colon cancer infection. A few aspects lead to a small degree of colon cancer.

The most well-known acquired disorders that expansion Genetic colorectal cancer is a risk in the family because it causes family venous polyps, lynch conditions and FAP (family adenomatous polyposis).

• African-US race. Among other races, there is a greater chance of colon cancer for African Americans humans.

• The family's history of development of colon cancer. Once you have a family which has cancer in there colon, you may experience malignant colon development.

There are several theories regarding the history of colon cancer in the family that are said to be a very important factor.

• A stationary way of life. Latent Individuals can create colon malignant growth. Getting standard physical action may decrease the danger of colon cancer infection.

• Obesity. fat people have a common chance of colon malignancy and an elevated risk of colon disease in the basin when separated from people find average weight.

Diabetes The risk of colon cancer growth is increasing in diabetes patients or insulin opposites.

• High fat, Low fiber eating routine. Colon malignant growth and rectal malignant growth illness may be connected with a common Western eating schedule, which is low fiber, high fat, and calories. Individuals who eat eats less carbs high in red meat and handled meat are very dangerous on their health**.**

• Smoking Individuals smoking can have an increased chance of infection with colon cancer.

Figure 4 shows that the researcher collects the risk factors of the colon cancer are then divided into modifiable risk factors and un modifiable risk factors in both prior studies and medical science.

• Alcohol. Excessive alcohol consumption raises the risk of colon cancer infection.

• Cancer radiation treatment The risk of colon disease rises through radiation therapy in the stomach region in order to treat past malignancies.

1.3 Deep Learning

Deep learning (DL) is a part of machine learning and it considered ML class, ML is a small artificial intelligence subset, and deep learning allows machines to respond as a human does in natural obstacles. DL was viewed as a summary approach to learning that can tackle a wide array of problems in different areas of application this meaning DL isn't task explicit. Deep learning is utilized when the issue size is unreasonably huge for explicit constrained thinking capacities (estimation page positions, coordinating advertisements to Facebook, notion examination). Deep learning is a fundamental concept behind

driverless cars, allowing them to interpret a stop sign or identify a person on foot. In buyer gadget s like phones, tablets, TV and handsfree speaker s it can be regulated. Deep learning is getting loads of consideration recently and in light of current circumstances. It's accomplishing results that were unrealistic previously (S. Tanwar and J. Jotheeswaran,2019).
Deep learning can be defined with two definitions:

Definition for the concept:

"Deep learning is a computer program that can identify what something is"
Definition for the Technical meaning:
"Deep learning is a class of machine learning algorithms in the form of a neural network that uses a cascade of layers (tiers) of processing units to extract features from data and make predictive guesses about new data".
The expression "Deep" normally alludes to the numbers of concealed layers in the neural system. Conventional neural systems just contain 2-3 shrouded layers, while profound systems can have upwards of 150.

Deep learning models build in millions of data and use that data to build structures of neural information that gain includes the information straight away without the need for manual item exploring. The related features are not pre-trained; they are discovered while the device tests on a variety of images. This computerized utilizing of highlights makes "deep learning models" incredibly accurate, for example, for PC vision companies. (Z. Alom et al.,2019)
Most DL techniques utilize neural network system structures, which is the reason "deep learning models" are frequently alluded to DL neural systems. One of the most well-known sorts of DL neural systems is known as convolutional neural systems (CNN or ConvNet).
A CNN blends learned highlights with input infor mation and uses 2D convolution layers to make thi s device ideal for handling 2D information, e.g. im ages.

CNNs dispense with the requirement for manual component extraction, so you don't have to distinguish highlights used to arrange pictures. The CNN works by selecting highlights straightforwardly from pictures.
Deep learning usually teaches a PC model makes sense of how to perform portrayal tasks really from pictures, substance, or sound. The models of deep learning models can achieve high accuracy in reaction and behavior to outperform human execution. Models are set up by using a colossal information plan of checked data and neural framework structures that contain various layers.
Figure 5 shows that Deep learning differs from other algorithms because it saves Time & Performance. Deep learning performance increases with the increase of data.

In a word, accuracy DL achieves affirmation accuracy at more imperative levels than at later. This allows satisfying the desires of the users in a way that greatly exceeds the human element and is very necessary in smart systems such as cars that do not contain a driver. On-going advances in profound learning have improved to where profound learning outmanoeuvres individuals in specific endeavours like grouping objects in pictures.
A key preferred position of deep learning systems is that they frequently keep on improving as the size of your data increments. (D. Bychkov, 2018)

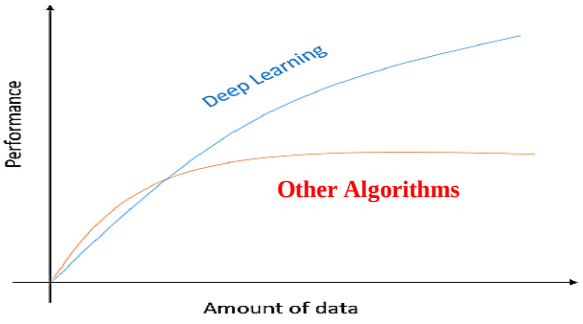

Fig. 4 Deep Learning Performance With Big Data (Jason Brownlee, 2016)

Published by Arab Journals Platform, 2019

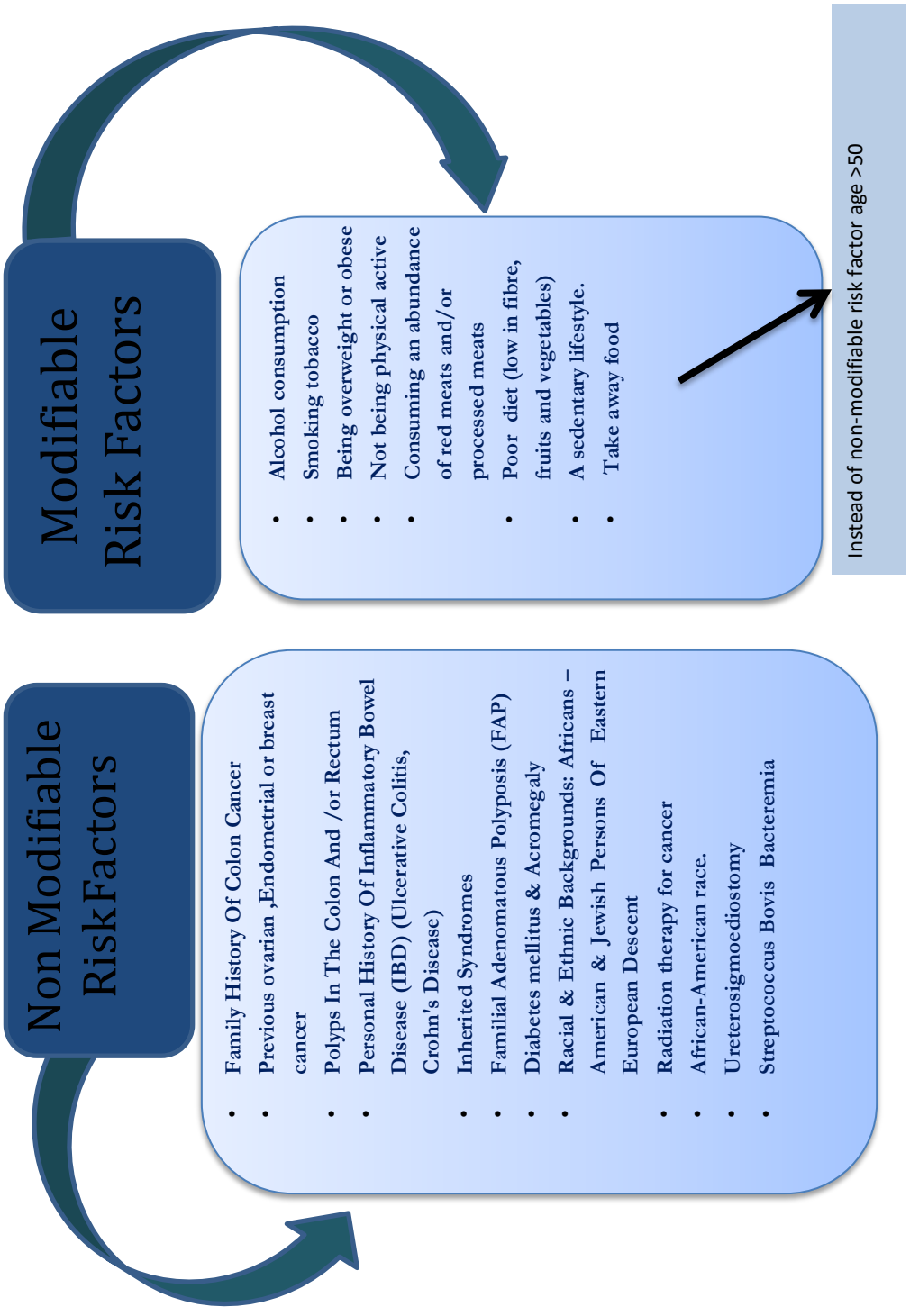

Fig. 5 Colon Cancer Risk Factors Made by the researcher

1.3.1 Deep Learning Algorithms

"A lot of algorithms are part of deep research. In the following paragraphs, we will discuss three common deep learning algorithms, such as: "Deep Neural Networks", "Convolutional Neural Networks", and "Recurrent Neural Networks Algorithms

Figure 6 shows that deep learning has 3 algorithms used in many fields

1. Deep Neural Networks for Improved Traditional Algorithms providing lift for classification and forecasting models. For example,
Finance: Improved fraud detection by finding more complex trends**.**

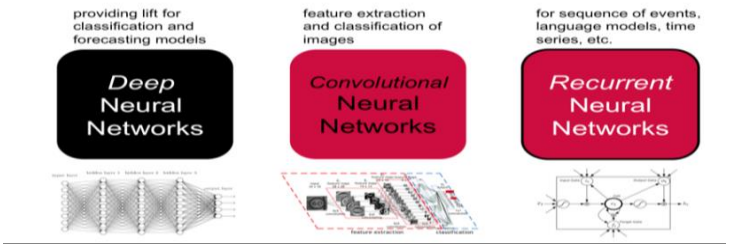

Fig.6 Deep Learning Algorithms (Ridhigrg,2019)

Fabrication: Increased imperfection recognition dependent on more profound peculiarity discovery.
2. Convolutional Neural Networks for images It is commonly used with images to extract feature and classification from images. For example,
Healthcare: *the identification of diseases through x-rays, scans, etc*
Satellite images: ability to mark areas and classifying items.
Retail: study of video to assess transport in the traffic**.**
Automotive: the identification of roads and their obstacles.

3. Recurrent Neural Networks for sequenced data
It is focus on the sequence of events, language models, time series, etc. For example,

Customer satisfaction: translation of voice information to content for NLP examination.
Account Predicting: conduct based by means of time arrangement examination (additionally improved suggestion frameworks.
Internet based life: constant interpretation of social and item discussion posts
Photograph subtitling: scan files of pictures for new bits of knowledge

2. Methodologies and Algorithms used in Colon Cancer

In the last few years a Medical health care service has been increasing expand consideration and prevalence. Atomic, bio-therapeutic methods, Restaurant photography and patient medical history, a lot of medicinal information are created each day because of advances in technology. Such therapeutic data are organized into many private individually, similarly as Open Databases after digitalization, after therapeutic information, such as electronic records, test reports and so on from clinical services to individuals. (R. Mia, 2018)
Data mining is used in various diagnosis techniques such as tumour detection and the analysis of protein structure, quality arrangement, cancer characterization dependent on microarray information, bunching of quality articulation information, factual model of protein-protein interaction and so on. (S. S. R. DEVENTHIRAN1,2019)
Data mining techniques were used extensively in the medical field and they had many advantages and also have weaknesses. This will be explained in the next lines during this research.
Table 1 shows Data mining algorithms used in previous healthcare researches, particularly in the area of colon cancer

Published by Arab Journals Platform, 2019

Table 1 Data Mining Methodology in Healthcare (Colon Cancer)

Author /year	Proposed methodology	Results
(Chih-Hung Jen, Chien-Chih Wang, Bernard C Jiang, Yan-Hua Chu, and Ming-Shu Che, 2012)	Create an early warning method for classifying chronic conditions, using K-NN and the Linear Discriminate Analysis (LDA).	The findings indicate that major factors in the classification of materials and screening are both unequivocally successful at a revised rate of 80%.
(Edi Winarko Rusdah,2013)	The numerous data mining methods for the diagnosis of tuberculosis were reviewed.	98.7% is granted for SVM's DM methodthat gives the highest exactness, followed by 98.4% for bagging method and 98.3% for random forest method.
(D. S. V. G. K. Kaladhar, B. K. Pottumuthu, P. V. N. Rao, V. Vadlamudi, A. K. Chaitanya, and R. H. Reddy, 2013)	Logistics, ADTree, Kstar, Ibk , Random Forest, and NNge Algorithms have been shown to be the best suited algorithms for classification analysis and Hierarchical Clustering approach is used to classify clusters using Colon Cancer dataset	Through using weka 3.6 application the Classification for the data set of the colon cancer showing that Kstar, ADTree, Logistics, Ibk, Random Forest , and NNge Algorithms gives 100% correctly classified cases, followed by Navie Bayes and PART with 97.22%, then Simple Cart and ZeroR showed at least 50% of correctly classified cases.
(I. Roseline Jecintha and V. Poonguzhali,2018)	"Effectiveness of data mining based cancer prediction system" using J48, ID3, Naïve Bayes.	The three algorithms were compared and the highest accuracy is given by ID3.
(R. Al-bahrani, A. Agrawal, and A. Choudhary,2013)	Establish precise survival predictive models for colon malignant development growth, incorporate the colon malignant growth information accessible from the SEER programme.	"Prediction accuracies of 90.38%, 88.01%, and 85.13%and an AUC of 0.96, 0.95, and 0.92 were obtained for the one year, 2 year and 5 year colon cancer survival prediction using the ensemble voting classification scheme" Further techniques to manage imbalanced data are required.
(I. Roseline Jecintha and V. Poonguzhali,2018)	Introduce Data Mining Techniques For Diagnosis And Cancer Prediction. Introduce techniques with two data mining methods The first includes Association rule mining and the second classification techniques such as FP-growth cancer diagnostics algorithm.	The results show that the decision tree is regarded as the best measure and predictor for Wisconsin data collection among the majority of the data mining classification and soft computing approaches.
(N. Singh, S. Kumar, and S. Bhadauria, 2016)	Data Mining Techniques for Early Detection of colon Cancer	Study 20 Risk factors considered and considered for colon cancer growth assessments of populations including age, sexual orientation, legacy, the use of hostile medicines, smoking, food prone, body movement, heftiness, tobacco, hereditary hazard, condition, and mental damage, consumption of red meat, balance diet, high blood pressure, Coronary cancer, excess drugs, radiation and chronic lung cancer.
(K. S. Sankari and M. Logambal,2018)	Using data mining techniques to make a prediction for Colon Cancer	This research gathers 13 colon cancer lifestyle factors. This study indicates colon cancer effects but no solution is suggested or no proposed methodology introduced.
(N. Kharrat, M. Assidi, M. Abu-elmagd, and P. N. Pushparaj,2019)	The beginning of Colorectal cancer through study and link between Data mining to human gut Microbiota with Fusobacterium spp.	Proof that the non-invalid lower limits of CRC's causal effects were tested by 4 species, including Citrobacterium, fusobacterium, Slaxkia, and Microbacterium. These results suggest that bacteria (microbial markers) in particular work with privately transformed microbiota to induce or influence CRC movement. These results are based on the theory.

Table 2 The Advantage And Disadvantages Of Data Mining Techniques Used In Healthcare (S. C. Pandey, 2016 **)**

Technique name	Advantages	Disadvantages
Neural Networks	1. It can handle noisy data properly for training. 2. It is capable of producing complex relationships between input and output. 3. Various neural networks can be used for clustering and prototype creation.	1. It does not work well with thousands of input features and for complex problems. 2. Local minima. 3. Over fitting. 4. It is difficult to understand the model built by the neural networks and requires high processing time.
Decision Trees	1. It can handle all types of variables, variables with missing values as well and it is easy to interpret. 2. For constructing decision trees one does not need to know about the domain. Even it can handle numerical and categorical data. 3. It can process high dimension data easily and it minimizes ambiguity of complex decisions and assigns exact values to the outputs.	1. For the numeric dataset, it generates complex decision trees. 2. It is an unstable classifier, i.e., the performance of a classifier depends on the dataset. 3. It is restricted to one output attribute and generates categorical data. 4. The Performance of decision trees is not affected by co linearity and linear- reparability problems.
Rough Sets	1. It does not need any additional knowledge about data like Probability in statistics. 2. Identifies relationships that would not be easily found using statistical methods. 3. From data, it produces sets of decision rules.	1. Some new discretization methods are required for quantitative attributes. Even more research is needed in this field. 2. Studies of a new approach to missing data are also needed.
Genetic Algorithms	1. Here the fitness function is a flexible expression of modeling criteria.	1. Finding a fitness function is critical.
Fuzzy sets	1. Unsupervised approach. 2. Converges approach.	1. Larger computational time. 2. Sensitivity to speed, local minima. 3. Sensitivity to noise and one expects zero or low noise level.
Support Vector Machines	1. It Provides better accuracy in comparison to other classifiers and it is effective in high dimensional spaces. 2. It is effective in cases where the number of dimensions is greater than the number of samples. 3. It easily handles complex nonlinear data points and over fitting is not a problem like in other cases. 4. It is memory efficient because it uses a subset of training sets in support vectors. 5. It is versatile because different kernel functions can be specified for the decision functions	1. It gives poor performances when the number of features is much greater than the number of samples. 2. It is computationally expensive and even the training process takes more than in comparison to other methods. 3. Selection of right kernel function is a problem because for every dataset different kernel function shows different results. 4. SVM was developed to solve the problems of binary class. 5. It does not provide probability estimates directly.
Bayesian Networks	1. It is fast and accurate for huge datasets as well. 2. It makes computations easier.	1. In some cases, where there is dependency among variables, it does not give accurate results.

Published by Arab Journals Platform, 2019

Using clear techniques to find this information, data mining algorithms process the data and concentrate imperative focuses from the big data collection. Various techniques including such mixed, characterized, generalized, bunching, affiliation, growth, organizing design and Those studies on colon cancer have used many techniques and methodologies such as data mining, machine learning and deep learning. perception of data are also not limited to meta-regulated mining. (S. S. R. DEVENTHIRAN1,2019)
Many studies write about colon cancer, its diagnosis, its risk factors, and survival after the disease with the different proposed methodology.

The Most Common Obstacles in Healthcare during Applying Data Mining Methodologies and Algorithms

- The data concerning raw health is massive and heterogeneous.
- Incomplete, inaccurate, incompatible or non-standard data, such as pieces of information saved from different data sources in similar formats.
- Interpretations of pictures, signals, or other scientific data by physicians are written in an unstructured language.
- How to extend powerful algorithms for comparing the contents of two versions of knowledge. ((This Difficulties Need powerful algorithms and data structures) (B. S. Srinivas, 2014

Many researches have begun moving towards Machining Learning after 2015 and Deep Learning addresses many advantages of using these techniques and their supremacy over Data Mining.
As described earlier, the use of DM healthcare is making more progress and producing more accurate results, but It is essential that take advantage from all data volumes, the findings do not change. Deep learning came as an optimal solution to the barriers of data mining, as it takes advantage of all existing data, no matter how massive the data is, The larger the volume of data, the more accurate and better results this will be shown in the following table.

Table 3 shows the advantage and disadvantages of colon cancer used methodology by machine learning and deep learning**.**

3. CONCLUSION AND FUTURE SCOPE

Comparative studies of different methodologies and algorithms used in colon cancer diagnoses and detection are presented in this paper, collect these studies from various researches and journals published in different sources.

The various approaches, techniques, and algorithms are
Proposed by several researchers for recognizing the colon cancer disease, showing the risk factors and the four stages of the cancer. Awareness of colon cancer, the stages of disease progression and the probability of survival and the likelihood of death of a patient is extremely important. Factors leading to colon cancer have been explained and categorized into modifiable and non-modifiable risk factors with 12 risk factors that cannot be changed and 8 changeable risk factors. The importance of deep learning and its used algorithms are determined for example, "deep neural network", "convolutional networks", and "recurrent neural network "so; in the future researches can use these algorithms to predict colon cancer before it happens.
The advantages and disadvantages of all previous studies have been clarified using Data Mining or machine Learning and Deep Learning so that in the next plan we can develop a methodology using Deep Learning that takes into account all the drawbacks faced by previous studies and provide a strong methodology that can predict colon cancer before it occurs.
The classification has been achieved by the various classical machine learning algorithms and the accuracy is appropriate. Only a few researchers have proposed the in-depth learning technique for classifying and segmenting colon cancer and artificial intelligent application will be discussed in future study.

Table 3 The advantages and disadvantages of using machine learning and deep learning techniques for colon cancer

Author Name	Date	Proposed Methodology	Advantages	Disadvantages
P. Hajela	2018	Survey on Deep Learning for the Detection and Segmentation of Cells	Begin to discuss the various approaches and techniques already used for early cancer detection and testing in the field.	Multiple predictive problems cannot be overcome in a single algorithm. A lack of real data may be viewed in relation to a "perfect" model and sometimes contributes to the program being "overfitting" or "underfitting"
A. row. F. Crawford-williams, S.march , m.j.ireland	2018	A Colorectal Cancer Systematic Analysis and geographical Differences in Australian Clinical Care	Assess details of the regional variations in the clinical and colorectal cancer diagnosis and treatment.	The analysis is limited because it creates minimal studies with the use of incoherent methods in these studies. Direct comparisons between studies are difficult due to variability of the population samples and the use of various regional classifications.
G. Urban *et al.*	2018	Locate and classify real time polyps with 96% colonoscopy screening accuracy through deep learning	Deep CNN designs and trains for detecting polyps with 8,641 hand-marked images obtained from colono-screening of more than 2,000 patients. Checked models for a total period of 5 hours on 20 colonoscopy videos.	With a standard desktop machine with a contemporary graphics processing unit the CNN program detected and localized polyps well within real time limits. This system can enhance ADR and decrease colorectal cancer period, but requires validation in large multicentre trials. There are no known potential effects of CNN on colonoscopy inspection behaviour.
D. Bychkov	2018	"Deep tissue analysis based on experience predicts colorectal cancer outcome."	Enter convolutional and recurrent systems to build a deep framework to predict the effects of colorectal cancer on tumor test images. The most recent approach is that patient outcomes are explicitly expected	The other algorithm does not measure random pictures compared with LSTM and requires additional steps to organize highlights from similar tiles and to construct an international TMA location descriptor. In this case, encoded 32 components by Enhanced Fisher Vector 47 (IFV), the descriptor compacting the variables in 32 components with the basic variable analysis.
N. Dimitriou, O. Arandjelovi, D. J. Harrison, and P. D. Caie	2018	Introduce machine learning framework to improve accuracy of colorectal cancer second stage prognosis	Introduce data driven framework which utilizes countless various kinds of features, Rapidly obtained in patients in the second phase (AUROC = 0:94), the immunophonic appearance is above that of ebb and Flow Principles, e.g., pT orchestrate (AUROC = 0:65) and shown on a partner of 173 patients infected by colorectal dieses.	Assessment corpus contains just stage II for the colon malignant growth patients The numbers of patients is little to give real and precise outcomes. There are a lot of colon cancer risk factors missed and not considered
J. Gründner, H. Prokosch, and M. Stürzl	2018	Prediction of clinical findings from colorectal cancer by machine learning	Created prediction models with accuracies above 0.70 using a fully automated process, which predicted relevant outcomes like chemotherapy response and survival.	The main problems identified were the availability of data and choosing the right performance measure to select the best model. The outcomes that were predicted with the highest accuracies were Relapse and RCT response (Yes/No), as well as survival and disease-free survival. The amount of available data are very limited and is restricted factor, as the process of element choice feature involved the data being split into three sets rather than two.
F. Ponzio, E. Macii, E. Ficarra, and S. Di Cataldo	2018	Innovative research by large convulational networks on colorectal cancer classifications	Suggest a comprehensive learning cycle targeted at adenocarcinomas and benign lasions isolated from healthy tissues	CNN is well qualified in a wide number of CRC annotated samples, with a very computerized intensive training approach

Table 3 (Continued)

Author Name	Date	Proposed Methodology	Advantages	Disadvantages
P. Hajela	2018	Survey on Deep Learning for the Detection and Segmentation of Cells	Begin to discuss the various approaches and techniques already used for early cancer detection and testing in the field.	Multiple predictive problems cannot be overcome in a single algorithm. A lack of real data may be viewed in relation to a "perfect" model and sometimes contributes to the program being "overfitting" or "underfitting"
A. row. F. Crawford-williams, S.march , m.j.ireland	2018	A Colorectal Cancer Systematic Analysis and geographical Differences in Australian Clinical Care	Assess details of the regional variations in the clinical and colorectal cancer diagnosis and treatment.	The analysis is limited because it creates minimal studies with the use of incoherent methods in these studies. Direct comparisons between studies are difficult due to variability of the population samples and the use of various regional classifications.
G. Urban *et al.*	2018	Locate and classify real time polyps with 96% colonoscopy screening accuracy through deep learning	Deep CNN designs and trains for detecting polyps with 8,641 hand-marked images obtained from colono-screening of more than 2,000 patients. Checked models for a total period of 5 hours on 20 colonoscopy videos.	With a standard desktop machine with a contemporary graphics processing unit the CNN program detected and localized polyps well within real time limits. This system can enhance ADR and decrease colorectal cancer period, but requires validation in large multicentre trials. There are no known potential effects of CNN on colonoscopy inspection behaviour. The videos anonymised excluded patient history details. Indications (screening against monitoring) can vary CNN output.
D. Bychkov	2018	"Deep tissue analysis based on experience predicts colorectal cancer outcome."	Enter convolutional and recurrent systems to build a deep framework to predict the effects of colorectal cancer on tumor test images. The most recent approach is that patient outcomes are explicitly expected	The other algorithm does not measure random pictures compared with LSTM and requires additional steps to organize highlights from similar tiles and to construct an international TMA location descriptor. In this case, encoded 32 components by Enhanced Fisher Vector 47 (IFV), the descriptor compacting the variables in 32 components with the basic variable analysis.
N. Dimitriou, O. Arandjelovi, D. J. Harrison, and P. D. Caie	2018	Introduce machine learning framework to improve accuracy of colorectal cancer second stage prognosis	Introduce data driven framework which utilizes countless various kinds of features, Rapidly obtained in patients in the second phase (AUROC = 0:94), the immunophonic appearance is above that of ebb and Flow Principles, e.g., pT orchestrate (AUROC = 0:65) and shown on a partner of 173 patients infected by colorectal dieses.	Assessment corpus contains just stage II for the colon malignant growth patients The numbers of patients is little to give real and precise outcomes. There are a lot of colon cancer risk factors missed and not considered
J. Gründner, H. Prokosch, and M. Stürzl	2018	Prediction of clinical findings from colorectal cancer by machine learning	Created prediction models with accuracies above 0.70 using a fully automated process, which predicted relevant outcomes like chemotherapy response and survival.	The main problems identified were the availability of data and choosing the right performance measure to select the best model. The outcomes that were predicted with the highest accuracies were Relapse and RCT response (Yes/No), as well as survival and disease-free survival. The amount of available data are very limited and is restricted factor, as the process of element choice feature involved the data being split into three sets rather than two. A focus on subgroups of the dataset also reduced the data available for some experiments.
F. Ponzio, E. Macii, E. Ficarra, and S. Di Cataldo	2018	Innovative research by large convulational networks on colorectal cancer classifications	Suggest a comprehensive learning cycle targeted at adenocarcinomas and benign lasions isolated from healthy tissues	CNN is well qualified in a wide number of CRC annotated samples, with a very computerized intensive training approach being disadvantageous (approx. 90 percent in the test).

Table 3 (Continued)

Author Name	Date	Proposed Methodology	Advantages	Disadvantages
Kopelman Y1, Gal O1, Jacob H2, Siersema P3, Cohen A4, Eliakim R4, Zaltshendler M4 and Zur D*4	2019	Polyp identification in colonoscopy by using both deep learning methods and image processing.	"The goal of high automated colonoscopic imaging polyp detection rate (≥ 95%) with high specificity (≥ 98%) will be achieved by exposing the system to greater amounts of data during the training phase and adding more computer vision and logic capabilities to the system."	Need million data to get an accurate result Consume larger computational time Hundreds or thousands of input features are not working well.
Xingzhi Yue, Neofytos Dimitriou, Peter D. Caie, David J. Harrison, and Ognjen Arandjelovi_c	2019	"Use Machine Learning and Inferred Phenotype Profiles to collect Colorectal Cancer Outcome Prediction numbers from hematoxylin & eosin Whole Slide Images".	Current structure based on machine for the prediction of colorectal malignancies results from full digitized diaphragms of histopathology with hematoxyl and eosin. The efficacy of the technique is demonstrated by the use of a specific knowledge set and a definite examination of its different elements, which verifies the capacity to identify and understand exceptional, discriminatory and clinically important substances.	All existing CNN proposals need careful review of tissue images by a licensed pathologist, and the rates are restricting.
J. Malik, S. Kiranyaz, S. Kunhoth, T. Ince, S. Al-maadeed, and R. Hamila	2019	Comparative analysis for CRC detection in histology photos	To learn more predictive models for the target (e.g., biomedical) mission, research transfer learning methods that apply knowledge gained from solving a source (e.g. non-medical problem). Proposed a compact, adaptive CNN architecture that even on scarce and low resolution information can be trained from scratch. In addition, quantitative comparative analyses are carried out amongst the traditional methods, learning-based methods are transferred and the preferred adaptive approach is implemented for a specific cancer screening and limited histology imagery is established.	Similar to traditional machine learning, CNN approaches are relatively stronger. Two CNN models have shown a substantially better accuracy and reliable cancer detection outcomes when the design of a scalable and lightweight SNC from scratch has been drastically different.
M. Shapcott, K. J. Hewitt, and N. Rajpoot	2019	Using a deep learning algorithm with large samples of histologic colon cancer	Using a deep-learning cell Recognition Algorithm to identify colon cancer pictures in the Cancer Genome Atlas (TCGA). Better sample output without loss of accuracy.	Through this study, various characteristics of cellularity could be calculated using deep learning. Such characteristics include maturity in the tumour, which is closely linked to aggressive cancer in a single cell or in a small group of up to five stroma cells.
Samuel Li	2019	ML forecast techniques for the survival of colorectal cancer	Ml Techniques used To test models for Hispanics, Whites and mixed patients and to predict 2 years of survival, and to underline the appropriate rankings. The models built for individual ethnicities are more accurate and are characteristic of important item rankings when prepared in various populations.	There may not be enough data found in the SEER data set to boost our present models ' performance significantly. Future research involves expanding our attention to ethnic groups outside the Hispanic, White, and other cancer groups.
Jakob Nikolas Kathe	2019	Survival prediction from colon cancer by using deep learning algorithms in slides of histology	The research examined whether the CNNs extract prognosticators directly from those broad images.	This research must be reviewed in advance before use of the clinical routine. Another drawback is that in this study, an observer blinding gathered tumor areas manually from photographs in whole slide histology. This manual process could be replaced in a fully automatic workflow. Prediction of colorectal cancer historical survival using PLOS.

Acknowledgements

I sincerely thank the supervisors for their guidance and their constructive comments to me. I also thank my parents a lot for their support and motivation to me

REFERENCES

A. M. Godkhindi, "Automated Detection of Polyps in CT Colonography images using Deep Learning Algorithms in Colon Cancer Diagnosis," *2017 Int. Conf. Energy, Commun. Data Anal. Soft Comput.*, pp. 1722–1728, 2017.

A. row. F. Crawford-williams, S.march , m.j.ireland, "Geographical Variations in the Clinical Management of Colorectal Cancer in Australia : A Systematic Review," vol. 8, no. May, 2018.

A. Silva, T. Oliveira, J. Neves, and P. Novais, "Treating Colon Cancer Survivability Prediction as a Classification Problem," Advances in Distributed Computing and Artificial Intelligence Journal, vol. 5, pp. 37–50, 2016.

A.Godkhindi, P.Dayananda, and C. N. Sowmyarani, "A Literature Survey on Computer-Aided Diagnosis in Detection and Classification of Polyp in Colon Cancer using CT Colonography," pp. 94–102, 2016.

A.Hadjipetrou, D. Anyfantakis, and C. G. Galanakis, "Colorectal cancer, screening and primary care: a mini literature review," World J. Gastroenterol., no. September, 2017.

B. S. Srinivas, "Data Mining Issues and Challenges in Healthcare Domian," *International Journal of Engineering Research & Technology (IJERT*)., vol. 3, no. 1, pp. 857–861, 2014.

Chih-Hung Jen, Chien-Chih Wang, Bernard C Jiang, Yan-Hua Chu, and Ming-Shu Chen. "Application of classification techniques on development an early-warning system for chronic illnesses". *Expert Systems with Applications*, 39(10):8852–8858, 2012.

Christina Chun, "stages of colon cancer," 2019. [Online]. Available: https://www.healthline.com/health/colorectal-cancer/stages-of-colon-cancer.

D. Bychkov et al., "Deep learning based tissue analysis predicts outcome in colorectal cancer," Scientific Reports., no. August 2017, pp. 1–11, 2018.

D. J. Ahnen *et al.*, "The Increasing Incidence of Young-Onset Colorectal Cancer: A Call to Action," *Mayo Clin. Proc.*, vol. 89, no. 2, pp. 216–224, 2014.

D. S. V. G. K. Kaladhar, B. K. Pottumuthu, P. V. N. Rao, V. Vadlamudi, A. K. Chaitanya, and R. H. Reddy, "The Elements of Statistical Learning in Colon Cancer Datasets : Data Mining , Inference and Prediction," Algorithms Res. J., vol. 2, no. 1, pp. 8–17, 2013.

E. Power, A. Simon, D. Juszczyk, S. Hiom, and J. Wardle, "Assessing awareness of colorectal cancer symptoms : Measure development and results from a population survey in the UK," *BMC Cancer*, vol. 11, no. 1, p. 366, 2011.

E. Ribeiro and A. Uhl, "Colonic Polyp Classification with Convolutional Neural Networks," 2016 IEEE ,29th International Symposium on Computer-Based Medical Systems Colonic.

Edi Winarko Rusdah. "Review on data mining methods for tuberculosis diagnosis". Information Systems, 2:4, 2013.

F. Bénard, A. N. Barkun, M. Martel, and D. Von Renteln, "Systematic review of colorectal cancer screening guidelines for average-risk adults : Summarizing the current global recommendations," vol. 24, no. 1, pp. 124–138, 2018.

F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018.

F. Ponzio, E. Macii, E. Ficarra, and S. Di Cataldo, "Colorectal Cancer Classification using Deep Convolutional Networks An Experimental Study," vol. 2, no. Biostec, pp. 58–66, 2018.

G. Urban *et al.*, "Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy," *Gastroenterology*, vol. 155, no. 4, p. 1069–1078.e8, 2018.

Goldberg, "Colon Cancer Information Overview," 2019. [Online]. Available: https://www.mycoloncancercoach.org/Colon-Cancer-101/Overview.

I. Roseline Jecintha and V. Poonguzhali, "Study on Data Mining Techniques for Cancer Prediction System," *Int. J. Data Min. Tech. Appl.*, no. 60, pp. 60–63, 2018.

J. Gründner, H. Prokosch, and M. Stürzl, "Predicting Clinical Outcomes in Colorectal Cancer Using Machine Learning," *Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth*
A. Ugon et al. (Eds.), vol. 0, pp. 1–5, 2018.

J. Malik, S. Kiranyaz, S. Kunhoth, T. Ince, S. Al-maadeed, and R. Hamila, "Colorectal cancer diagnosis from histology images : A comparative study," pp. 1–12, 2019.

J. Nikolas *et al.*, "Predicting survival from colorectal cancer histology slides using deep learning : A retrospective multicenter study," pp. 1–22, 2019.

Jason Brownlee, "What is Deep Learning?," *https://machinelearningmastery.com/what-is-deep-learning/*, 2016.

K. S. Sankari and M. Logambal, "Predicting Colon Cancer Using Data Mining Techniques," International Journal of Computer Science Trends and Technology (IJCST)., vol. 6, no. 2, pp. 97–99, 2018.

M. DePietro, "Stages of Colon Cancer," 2019. [Online]. Available: https://www.healthline.com/health/colorectal-cancer/stages-of-colon-cancer.

M. Melissa Buffalo, "End Colon Cancer in Indian Country," AICAF American indian cancer Foundation , 2019. Available: https://www.americanindiancancer.org/aicaf-project/colorectal-cancer-awareness/

M. Shapcott, K. J. Hewitt, and N. Rajpoot, "Deep Learning With Sampling in Colon Cancer Histology," *Frontiers in Bioengineering and Biotechnology.*, vol. 7, no. March, 2019.

N. Dimitriou, O. Arandjelovi, D. J. Harrison, and P. D. Caie, " A principled machine learning framework improves accuracy of stage II colorectal cancer prognosis," no. March, pp. 1–9, 2018.

N. Kharrat, M. Assidi, M. Abu-elmagd, and P. N. Pushparaj, "Data mining analysis of human gut microbiota links Fusobacterium spp . with colorectal cancer onset," Biomed. informatics Soc., no. July, 2019.

N. Singh, S. Kumar, and S. Bhadauria, "Early Detection of Cancer Using Data Mining," International Journal of Applied Mathematical Sciences ., vol. 9, no. 1, pp. 47–52, 2016.

P. Rebecca L. Siegel, MPHKimberly D. Miller, MPHAhmedin Jemal, DVM, "Cancer Statistics , 2019," CA CANCER J CLIN, vol. 69, no. 1, pp. 7–34, 2019.

P.Hajela, "Deep Learning for Cancer Cell Detection and Segmentation : A Survey," Conf. Pap., no. November, 2018.

R. Al-bahrani, A. Agrawal, and A. Choudhary, "Colon cancer survival prediction using ensemble data mining on SEER data," in *IEEE International Conference on Big Data*, 2013, pp. 9–16.

R. Mia, "A Comprehensive Study of Data Mining Techniques in Healthcare , Medical , and Bioinformatics," *International Journal of communication and computer Technologies* ., no. February, 2018.

Ridhigrg, "the Elementary Study of Deep Learning Algorithms," 2019.
[Online]. Available: https://www.houseofbots.com/news-detail/11747-1-here-is-the-elementary-study-of-deep-learning-algorithms.

S. C. Pandey, "Data Mining Techniques for Medical Data : A Review," in *International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016 Data*, 2017, no. November 2016.

S. Li, "Personalized Colorectal Cancer Survivability Prediction with Machine Learning Methods *," *Natl. Sci. Found. REU Progr. Res.*, p. 5, 2019.

S. Nadeem and A. Kaufman, "Depth Reconstruction and Computer-Aided Polyp Detection in Optical Colonoscopy Video Frames **," pp. 1–12, 2016.

S. Tanwar and J. Jotheeswaran, "Survey on Deep Learning for Medical Imaging," Journal of Applied Science and Computations (JASC)., vol. v, no. January, pp. 1608–16020, 2019.

S.S.R. DEVENTHIRAN1, "EXPLORING COLORECTAL CANCER GENES THROUGH DATA MINING TECHNIQUES," *Int. Res. J. Eng. Technol.*, vol. 6, no. 04 | Apr 2019, pp. 4770–4773, 2019.

X. Yue, N. Dimitriou, P. D. Caie, and D. J. Harrison, "Colorectal Cancer Outcome Prediction from H & E Whole Slide Images using Machine Learning and Automatically Inferred Phenotype Profiles," Proc. 11th Int. Conf. Bioinforma. Comput. Biol., vol. 60, pp. 139–149, 2019.

Y. Kopelman et al., "Automated Polyp Detection System in Colonoscopy Using Deep Learning and Image Processing Techniques," J. Gastroenterol. Its Complicat., vol. 3, no. 1, pp. 1–7, 2019.

Z. Alom et al., "A State-of-the-Art Survey on Deep Learning Theory and Architectures," electronics, pp. 1–67, 2019.