IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

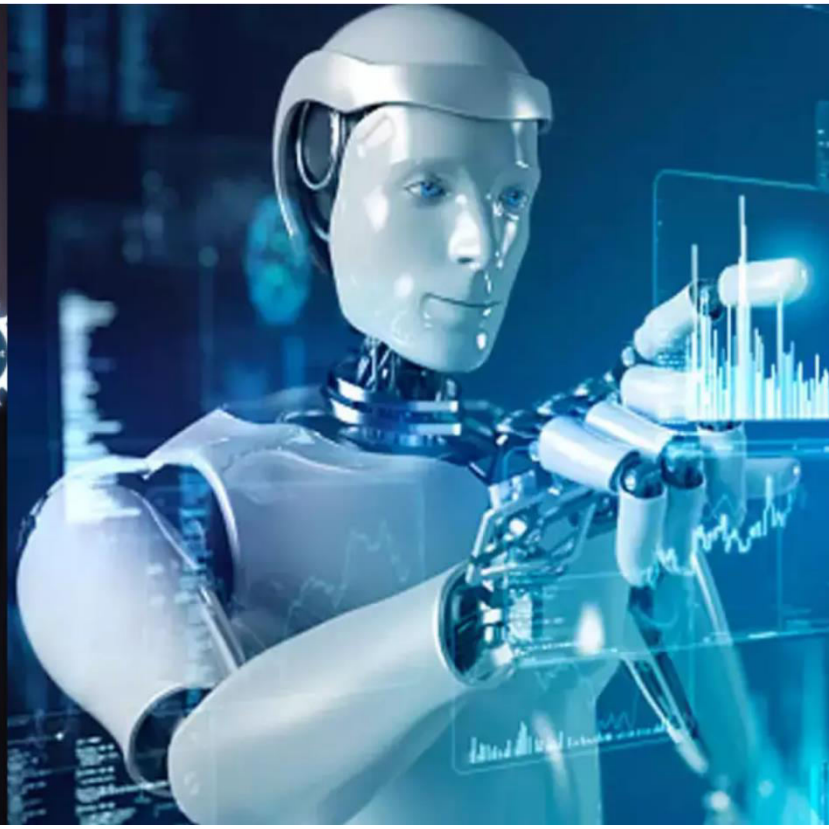
ISSN(O): 2319-8753
ISSN(P): 2347-6710



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.699

Volume 14, Issue 12, December 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412140|

Detection of Colorectal Cancer

Dr Sathyavathi S, Shobana, Bertilla Madonna C P , Rohith R , Vishruta S R

Assistant Professor, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India

Assistant Professor, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India

P.G Student, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India

P.G Student, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India

P.G Student, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India

ABSTRACT: Abstract— Colorectal cancer (CRC) is the second leading cause of cancer death worldwide in 2020 (9.4%). A high-quality colonoscopy is decisive for the maximum detection of precancerous lesions. Hence, this focuses on the detection of colorectal cancer. Histopathological images of colorectal are used for the detection of cancer. There are two classes of data: benign and malignant. 4500 images are taken with 2 classes, where 80% is used for training, 10% is used for validation, and the remaining 10% is used for testing. This detection is achieved by deep learning techniques with python using Tensor flow method. Deep Learning technique helps in acquiring more accurate diagnosis.

KEYWORDS: component, formatting, style, styling, insert

I. INTRODUCTION

Colon cancer is a type of cancer that begins in the large intestine (colon). The colon is the final part of the digestive tract. Colon cancer typically affects older adults, though it can happen at any age. It usually begins as small, noncancerous (benign) clumps of cells called polyps that form on the inside of the colon. Over time some of these polyps can become colon cancers. Polyps may be small and produce few, if any, symptoms. For this reason, doctors recommend regular screening tests to help prevent colon cancer by identifying and removing polyps before they turn into cancer. If colon cancer develops, many treatments are available to help control it, including surgery, radiation therapy and drug treatments, such as chemotherapy, targeted therapy and immunotherapy. Colon cancer is sometimes called colorectal cancer, which is a term that combines colon cancer and rectal cancer, which begins in the rectum. Doctors aren't certain what causes most colon cancers. In general, colon cancer begins when healthy cells in the colon develop changes (mutations) in their DNA. A cell's DNA contains a set of instructions that tell a cell what to do. Healthy cells grow and divide in an orderly way to keep your body functioning normally. But when a cell's DNA is damaged and becomes cancerous, cells continue to divide — even when new cells aren't needed. As the cells accumulate, they form a tumor. With time, the cancer cells can grow to invade and destroy normal tissue nearby. And cancerous cells can travel to other parts of the body to form deposits there (metastasis).

1.1. Literature Survey

- Files must be in MS Word only and should be formatted for direct printing, using the CRC MS Word provided. Figures and tables should be embedded to assess the availability and completeness of healthcare quality indicators for breast and colorectal cancers in the KHCC cancer registry. Statistical analysis techniques were likely employed to analyze the collected data and assess the availability and usability of hospital-based cancer registry data in measuring the selected quality outcome indicators. The study likely concludes with a discussion of the implications of the findings and recommendations for improving the availability and usability of hospital-based cancer registry data for quality assessment purposes. [5]
- The collected data was pre-processed and engineered to extract relevant features. Multiple machine learning classifiers were evaluated for their effectiveness in predicting cancer recurrence. Commonly used classifiers such as logistic regression, support vector machine (SVM), gradient boosting, and multi-layer perceptron (MLP) were likely considered. The selected machine learning models were trained using the pre-processed data. The selected machine learning models were trained using the pre-processed data.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

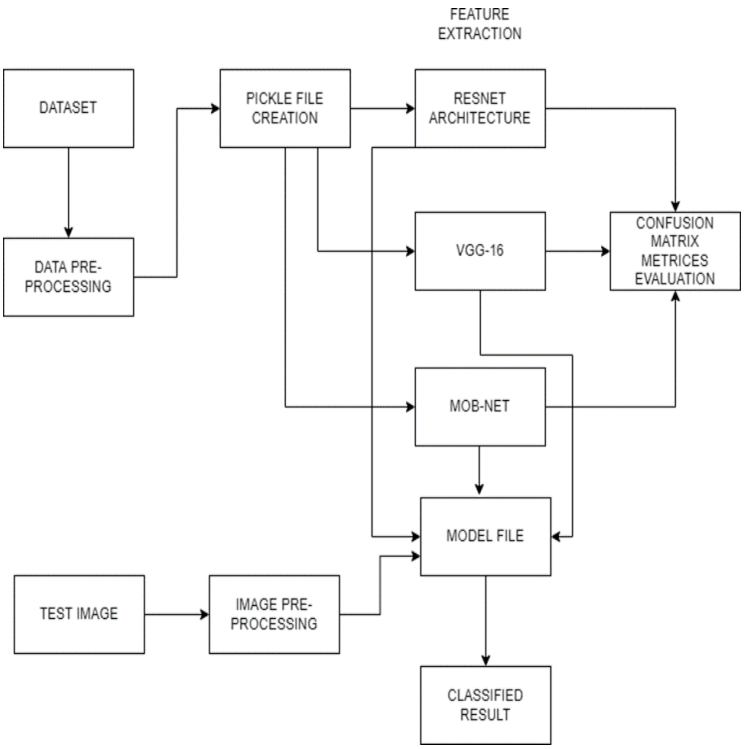
Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412140|

- Real-time assistance during live colonoscopy procedures can greatly enhance the efficiency, accuracy, and safety of the process. During a colonoscopy, a flexible tube with a camera attached is inserted into the colon to examine its interior. Live imaging and video feed from the colonoscope provide real-time visualization of the colon's walls and any abnormalities present. [8]
- To study this kind of cancer, this research attempts to explore the relations between the genes responsible for colon cancer via Data Mining classification techniques [9].
- The method proposed segment the glands automatically by using intensity-based thresholding and organizational properties for classification.

1.2 Proposed Method

The framework will continuously use large sets of related data to colo-rectal cancer patient collected from multiple sources like medical sources, lifestyle and demography. The datasets are taken in the form of histo-pathological images. The detection of colorectal cancer using conventional technique is time consuming and less accuracy, so we apply deep learning techniques to predict the colorectal cancer. The CNN algorithm has been employed to carry out the prediction. Compared with traditional methods, the proposed method has achieved about 17% of improvement on the classification accuracy, suggesting that neural network is a powerful tool for the diagnosis of cancerous diseases. Based on our work, the same or similar methods can be used to diagnose other cancer related diseases providing an intelligent healthcare system. Data pre-processing refers to all transformations on the raw data before it is fed to the machine learning or deep learning algorithm. For instance, training a convolutional neural network on raw images will probably lead to bad classification performances. The pre- processing is also important to speed up training such as clustering and scaling technique. Real- world data is often incomplete, inconsistent, and lacking in certain behaviours or trends, and is likely to contain many errors. Data pre-processing is a proven method of resolving such issues. Feature selection is an effective strategy to optimize the predictive performance of machine learning algorithms. In this study, we have tested a number of feature selection techniques including SelectKBest. Sequential Forward Selection Sequential Backward Selection, and Recursive Feature Elimination implemented in the sklearn package. The models with the best performance resulted from the SelectKbest, which is a univariate correlation feature selection method.



SOFTWARE DESCRIPTION

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412140|

1.3 SOFTWARE DESCRIPTION

Python in Visual Studio Code editor is one of the applications of Anaconda Navigator in our project as it is an easy to learn, powerful programming language. The provided datasets are taken from Kaggle which is an open source, where the MRI images of brain is obtained. It is classified and diagnosed using this programming languageGeneral guidelines for the preparation of your text

II. WORKING PROCESS

The working structure is simple, it consists of three major steps, they are

1. Data Pre-Processing
2. Computation of data via Neural network using CNN
3. Computation of confusion matrix and cancer prediction.

The working structure of the project demanded learning concepts of CNN and going through research papers to get a comprehensive understanding of the activities involved in processing, training and testing a neural network.

1. Data Pre-Processing

This section deals with image pre-processing in specific, as tedious procedures were undertaken to merge these individual slices into 2D images so that we can feed them into a 2D Convolutional neural network. Using the libraries matplotlib and OpenCV, visualize and resize the images to the specified format suitable based on the computational speeds and capacity of the system.

2. CNN Computation

The convolutional neural system, or CNN for short, is a specific sort of neural system model intended for working with two-dimensional picture information, in spite of the fact that they can be utilized with one-dimensional and three-dimensional information. To perform convolution activity, a channel (A small matrix) is utilized whose size can be determined. This channel moves everywhere throughout the picture network and its work is to duplicate its qualities by the first pixel esteems. Every one of these increases are summarized to one number toward the end. The channel moves further and further on its right side by n units (can fluctuate) playing out a comparative activity. In the wake of going over every one of the positions, a framework is gotten which is a lot smaller in size than the information grid.

Convolutional Layer has the following components:

- Filters
- Activation maps
- Parameter sharing
- Layer-specific hyperparameters

III. EDGE DETECTION

To detect the horizontal and vertical images with the help of matrix, let’s consider greyscale image, with 6 x 6 matrix and a filter of 3x3 applied to it.

After the above calculations of matrix, we get the matrix as show in in fig:6. To calculate it, we take the initial 3 X 3 framework from the 6 X 6 picture and increase it with the channel. Lets consider the following matrix of 4x4 order and the calculation takes place as: for example 3*1 + 0 + 1*-1 + 1*1 + 5*0 + 8*-1 + 2*1 + 7*0 + 2*-1 = - 5. To compute the second component of the 4 X 4 order, we will move the channel one step ahead to the right side of the original Grey scale matrix and so on

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412140|

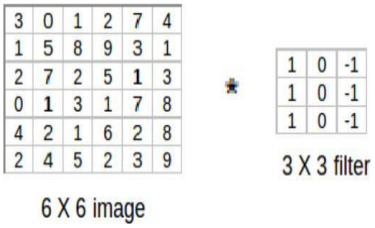


Fig 4.3: Scanning of the image pattern

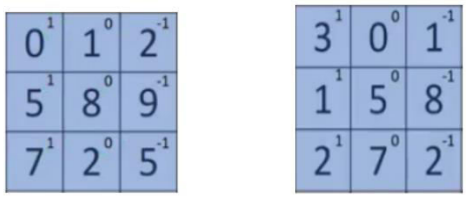


Fig 4.4: Matrix calculation in convolutional layer

-5	-4	0	8
-10	-2	2	3
0	-2	-4	-7
-3	-2	-3	-16

Fig 4.5 Convolving over entire image

The way to detect the vertical edge in the image is to look for the pixel values as, if the pixel values are greater, then brightness at that part of the image will be more and if the value is less it will be dark.

IV. POOLING LAYER

Spatial Pooling (additionally called subsampling or down sampling) diminishes the dimensionality of each element map yet holds the most significant data. Spatial Pooling can be of various kinds: Max, Average, Sum and so on. If there should be an occurrence of Max Pooling, we characterize a spatial neighbourhood (for instance, a 2×2 window) and take the biggest component from the redressed highlight map inside that window. Rather than taking the biggest component we could likewise take the normal (Average Pooling) or total of all components in that window. By and by, Max Pooling has been appeared to work better.

As shown in figure below, Max pooling take the biggest component from the rectified feature map. Taking the biggest component could likewise take the normal pooling. Entirety of all components in the element map call as sum pooling.

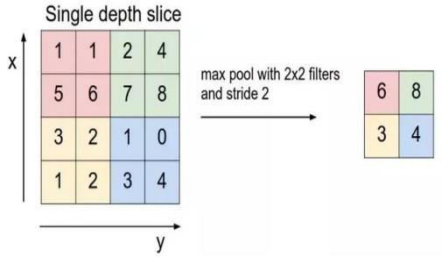


Fig 4.6: Max Pooling

4.1 FULLY CONNECTED LAYERS

In Fully Connected layers, each neuron from one layer is associated with each neuron in another layer. Chiefly, FC acts like as the conventional l neural system, Multi-Layer Perceptron (MLP). Nonetheless, the main distinction is that the information sources would be in the shape and structure made by the past phases of a CNN.[7].As shown in figure below, feature map matrix is converted into vector as(x1,x2,…xn) with FC layer where we combine these to create a model. And then we use activation function to classify the output.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412140|

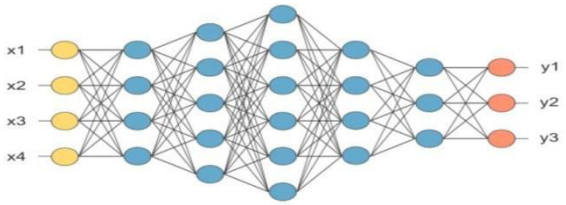


Fig 4.7: Fully connected layer

Computation of confusion matrix and cancer prediction

The confusion matrix is a way to determine whether the model is good or not. Here I implemented confusion matrix for predicted labels vs actual labels. The 0's represent no cancer and the 1's represent cancer. Ideally the false positives and the false negatives value should be low.

4.2 FEATURE EXTRACTION

The process of feature extraction is useful when you need to reduce the number of resources needed for processing without losing important or relevant information. Feature extraction can also reduce the amount of redundant data for a given analysis. Also, the reduction of the data and the machine's efforts in building variable combinations (features) facilitate the speed of learning and generalization steps in the machine learning process.

4.3 CNN

In deep learning, a CNN is a type of deep neural networks, which deals with the set of data to extract information about that data. Like images, sounds or videos etc. can be used in the CNN for the data extraction. There are mainly three things in CNN. First one is local receptive field and then shared weight and biases and the last one is activation and pooling. In CNN, first the neural networks are trained using a heavy set of data so that the CNN can extract the feature of given input. When the input is given, first image preprocessing is done then the feature extraction occurs on the basis of set of data stored and then the classification of data is done and output is shown as the result.

The CNN can deal with that input only for what the neural network is trained and the data is saved.
They are used in image and video recognition, recommender systems, image classification, medical image analysis, and natural language processing

4.4 DATASET

One major advantage of using CNNs over NNs is that you do not need to flatten the input images to 1D as they are capable of working with image data in 2D. This helps in retaining the "spatial" properties of images.

4.5 PRE-PROCESSING STEPS

The pre-processing steps were conducted with resizing, patch, and augmentation steps. The first pre-processing step normalizes the size of the input images. Almost all the radiographs were rectangles of different heights and too large (median value of matrix size ≥1,800). Accordingly, we resized all images to a standardized 224×224 pixel square, through a combination of preserving their aspect ratios and using zero-padding. The investigation of deep learning efficiency depends on the input data; therefore, in the second processing step, input images were pre- processed by using a patch (a cropped part of each image). Finally, data augmentation was conducted for just the training dataset, using mirror images that were reversed left to right and rotated −30, −10, 10, and 30 degrees.

4.6 IMAGE LABELING AND DATASET DISTRIBUTIONS:

All subjects were independently labelled twice. Labelling was first evaluated with the original images on a picture archiving communication system (PACS) and secondly with the resized images that were used for the actual learning data. Datasets were defined as the internal dataset and temporal dataset, with the temporal dataset used to evaluate the test. The internal dataset was randomly split into training (70%), validation (15%), and test (15%) subsets.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412140|

V. RESULTED OUTPUT

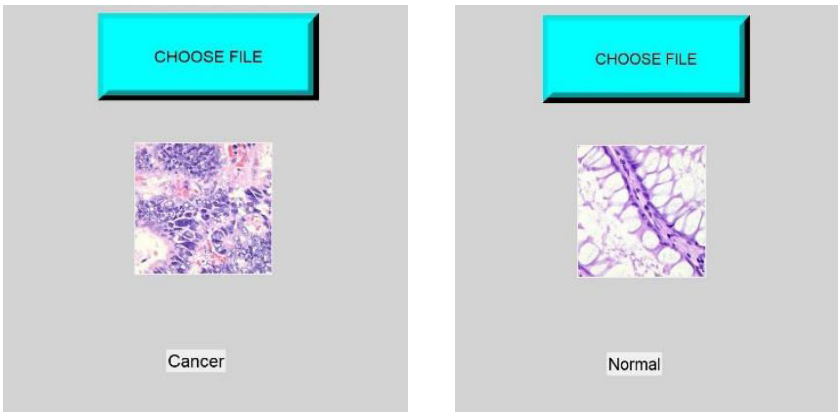


Fig 5.1 a): Output of Normal Cell Fig 5.1 b): Output of cancerous Cell

5.1 ACCURACY GRAPH

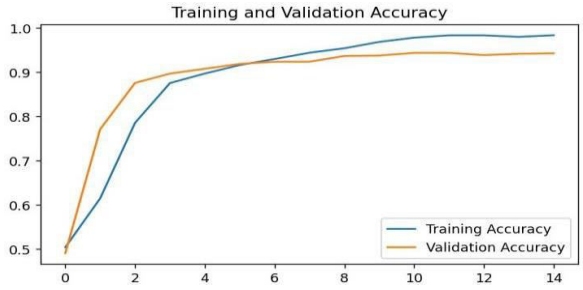


Fig 5.2 a): Training and Validation Accuracy

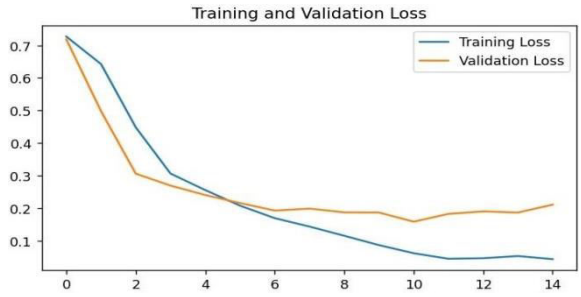


Fig 5.2 b): Training and Validation Loss

5.3 Confusion Matrix

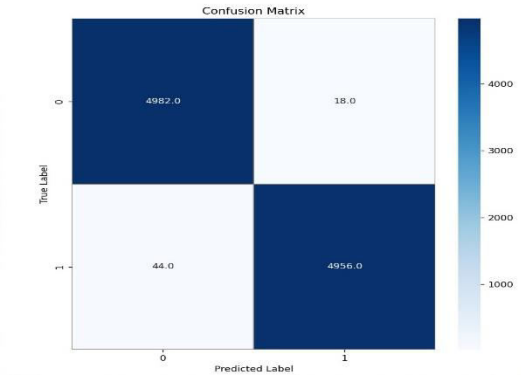


Fig 5.3: Confusion Matrix between True Label and Prediction Label

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412140|

VI. CONCLUSION

The implementation of a modified 5-layered CNN demonstrates promising results in detecting colorectal cancer patients from normal patients, achieving an impressive accuracy of 97%. This outperforms traditional neural network systems in terms of accuracy, highlighting the efficacy of the deep learning approach for this task. The high accuracy achieved by the modified CNN underscores its potential as a valuable tool for colorectal cancer classification. By leveraging deep learning methods, it becomes feasible to differentiate between normal and malignant cases with a high degree of accuracy. This capability holds significant implications for early detection, diagnosis, and ultimately, the treatment of colorectal cancer. In conclusion, the utilization of a modified 5-layered CNN represents a promising advancement in the field of colorectal cancer detection. Its lightweight architecture coupled with high performance makes it a viable option for clinical applications aimed at improving patient outcomes through early and accurate identification of colorectal cancer cases.

REFERENCES

1. Amid Abu Hmaidan,Elsie Boutou,Khaled Jamal,Amal Al-Omari,'Availability and Usability of the Hospital-based Cancer Registry Data for Measuring the Quality Outcome Indicators of Healthcare Provided to Breast and Colorectal Cancer Patients at King Hussein Cancer Center',2018 1st International Conference on Cancer Care Informatics (CCI)
2. Yanke Li,Fuqiang Zhang,Chengzhong Xing,'Screening of Pathogenic Genes for Colorectal Cancer and Deep Learning in the Diagnosis of Colorectal Cancer',IEEE Access
3. Yi-Jie Huang,Qi Dou,Zi-Xian Wang,Li-Zhi Liu,Li-Sheng Wang,Hao Chen,Pheng-Ann Heng,Rui-Hua Xu,'HL-FCN: Hybrid loss guided FCN for colorectal cancer segmentation',2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
4. Danliang Ho,Dawn Qingqing Chong,Brenda Tay,Iain BeeHuat Tan,Mehul Motani,'Prognosticating Colorectal Cancer Recurrence using Machine Learning Techniques',2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM)
5. Tian Li,Chunqiu Zheng,Le Zhang,Ziyuan Zhou,Rong Li,'Exploring the risk dietary factors for the colorectal cancer',2015 IEEE International Conference on Progress in Informatics and Computing (PIC)
6. Mohammed Othman,Faten F Khatbat,Tarik Al Amsy,'Exploring colorectal cancer genes through data mining techniques',2018 Advances in Science and Engineering Technology International Conferences (ASET)
7. Namita Sengar,Neeraj Mishra,Malay Kishore Dutta,Jiri Prinosil,Radim Burget,'Grading of colorectal cancer using histology images',2016 39th International Conference on Telecommunications and Signal Processing (TSP)
8. Hendrick Gao-Min Lim,Yuan-Chii Gladys Lee,'A Cross-Platform Comparison of Affymetrix, Agilent, and Illumina Microarray Reveals Functional Genomics in Colorectal Cancer Progression',2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)
9. Joohyung Lee,Ji Eun Oh,Min Ju Kim,Bo Yun Hur,Dae Kyung Sohn,'Reducing the Model Variance of a Rectal Cancer Segmentation Network',IEEE Access 27
10. Chong Li,Xiaolei Xie,Jinlin Li,Nan Kong,'Improving colorectal polyp classification based on physical examination data â€" A ensemble learning approach',2017 13th IEEE Pedro Conference on Automation Science and Engineering (CASE)
11. G.M. Condelipes,Pedro M. Fontes,Katerina Nikolaidou,Vanda Marques,Eduardo J.S. BrÃ¡s,Marta B. Afonso,CecÃlia M.P. Rodrigues,JoÃ£o GonÃ§alves,V. Chu,JoÃ£o Pedro Conde,'Development Of a Microfluidic Colorectal Cancer Cell Culture System with Integrated Optical Sensors for Rapid Phage Selection',2020 IEEE SENSORS 28



INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details