

Colorectal Cancer Detection Using Pre-Trained Ensemble Algorithms

S.Raghavendra¹, U.Venkat Charan², P.Akshara³, E.Hari Krishna⁴, L.Arjun⁵

¹Assist prof, Dept of CSE

^{2, 3, 4, 5} UG Student, Dept of CSE

^{1, 2}-CMR Technical Campus Hyderabad, Telangana, India

Abstract- *Colorectal cancer is the second leading cause of cancer-related deaths globally and the most common cancer after lung and breast cancer. If caught early, many lives can be saved. On the other hand, conventional diagnosis requires doctors to manually analyze histopathological images, which can be a lengthy process, expensive, and sometimes even lead to human mistakes. This creates a strong need for smarter and more reliable systems that can support medical professionals in their work.*

This paper presents a colorectal cancer detection system built on pre-trained deep learning models. Our system employs popular architectures such as Inspection V3, Resnet50, and EfficientNetB0 that are not only capable of identifying key features in medical images but also help in enhancing the accuracy of the predictions. The model was built and evaluated using the LC25000 dataset, which comprises histopathological images of both cancerous and non-cancerous tissues.

We do not need to depend on a single model but instead use an ensemble approach, where the strengths of multiple models are combined into one. This helps us achieve better results in terms of accuracy, precision, recall, and F1-score compared to traditional methods such as Random Forest and Naive Bayes. Our findings also indicate that the system can detect cancer faster and more effectively.

Overall, this approach provides a simple, reliable, and cost-effective solution to assist doctors in diagnosing colorectal cancer. By supporting early detection, it helps improve treatment outcomes and gives patients a better chance of survival.

I. INTRODUCTION

Colorectal cancer is one of the main causes of death worldwide. Early detection is crucial because it helps people live longer and receive better treatment. Usually, doctors check medical images to find cancer, but this process takes a lot of time, and mistakes can sometimes occur. Therefore, we need automated systems that can detect cancer quickly and

accurately. With the help of deep learning, medical images can be analyzed effectively.

Conventional Neural Networks (CNNs) are very good at understanding images, so they are useful for detecting cancer. In our study, we used three models, InspectionV3, Resnet50, and EfficientnetB0, to classify images as tumors or normal. We also used transfer learning to improve the performance and reduce the workload. We prepared the images by resizing and normalizing them to obtain better accuracy. Among all the models, inspection V3 yielded the best results.

Moreover, this project has a simple and easy-to-use chatbot interaction, where users can upload their images and get instant results. If cancer is detected, the system also helps users book a doctor appointment, making it a complete and practical healthcare solution.

Overall, this project aims to build a reliable and accurate system that can automatically detect colorectal cancer and help in early detection, which leads to better treatment and improved patient outcomes.

II. PROBLEM DEFINATION

A. Need for Automated Cancer Detection

Early detection of colorectal cancer is important because it helps patients live longer and reduces serious problems. However, detecting cancer in medical images is difficult. It requires expert doctors accuracy and a lot of time. Therefore, it can be a heavy and costly process. In many hospitals, there are more patients but fewer specialists available. This can delay diagnosis and affect treatment success.

One way to solve this problem is by using an automated cancer detection system that uses deep learning to analyze images quickly and accurately. These systems reduce the need for manual work and human errors. They can handle large amounts of data very efficiently and provide consistent

and reliable results. They can also detect cancer at a very early stage, which is difficult to find through manual checking.

By helping doctors with accurate results and fast analysis, these systems improve decision-making and make diagnosis faster, helping to provide better care for patients.

B. Limitations of Traditional Diagnosis Methods

Traditional diagnostic methods mostly depend on doctors manually checking medical images. These methods can work, but they have the same problem. Checking many images requires a considerable amount of time and effort, which can delay diagnosis. In addition, human mistakes can occur as small details may be missed. Different doctors may also provide different results, which reduces reliability. Early cancer detection is challenging because the signs are not always clear.

Another problem is that these methods depend on highly trained doctors who may not always be available, especially in remote areas. In addition, handling large amounts of data manually is neither practical nor efficient. Doctors can also become fatigued when working continuously, which can lead to more mistakes. These problems show that automated systems are needed to improve accuracy and detect cancer faster.

C. System Problem Addressed by the Proposed System

The main problem that this project attempts to solve is the lack of a fast and automated system for detecting colorectal cancer. Traditional methods are slow and depend heavily on expert doctors. This can delay diagnosis and affect the patient's outcome. Therefore, these systems used deep learning to analyze medical images and classify them as tumors or normal.

This system uses pre-trained CNN models, such as InceptionV3, ResNet50, and EfficientNetB0 to improve accuracy and reduce manual work. It can process images quickly and provide reliable results through a simple interface. The system also includes a chatbot that allows users to easily obtain predictions. It also helps in booking a doctor's appointment, making the system more user-friendly.

In summary, this system is a reliable tool for early disease detection. It reduces the time required for diagnosis and makes healthcare services more effective. It also helps reach more people and improve overall healthcare.

III. RELATED WORK

The detection of colorectal cancer using medical images has improved significantly with machine learning and deep learning. Early methods used simple algorithms, such as SVM, Decision Tree and Random Forest, which used manually selected features, such as texture, color and shape to classify images. They have good results, but they cannot detect complex patterns in medical images.

Conventional neural networks (CNNs) have significantly improved image classification. In medical imaging, models such as VGG16, ResNet50, and InceptionV3 are widely used. These models can automatically identify important features that provide more accurate results than traditional methods. In addition, using pre-trained models with transfer learning helps achieve high performance even with small datasets.

Recently, ensemble learning has become important, where multiple CNN models are combined to improve accuracy and reliability. Among these models, InceptionV3 performed the best in detecting colorectal cancer images. Overall, deep learning methods are better than traditional methods in terms of effectiveness, scalability, and reliability.

IV. METHODOLOGY

This colorectal cancer detection system uses deep learning to analyze medical images more accurately and efficiently. The system takes colorectal images as input, identifies important features, and provides reliable results. It also has a user-friendly interface that makes it easy for users.



First, the datasets contain images of tumors (cancer) and normal cases with labels. These images were obtained from public medical sources and were used to train and test the system. However, the images may differ in quality, size, and format; therefore, they must be handled carefully.

After collecting the data, we preprocessed the images for training. These include resizing the image, normalizing the pixel values, and removing noise or unwanted parts. These

steps help make the data clean, consistent, and suitable for deep learning models.

After preprocessing, the data were divided into two parts: training and testing data, usually in an 80:20 ratio. The training data were used to train the model, and the testing data were used to check its performance. Feature extraction is then performed using pre-trained CNN models such as InceptionV3, ResNet50, and EfficientNetB0 with the help of transfer learning.

After that the system checks the image features and decides whether it is tumor or normal. The models performance is measured using accuracy, precision, recall and F1-score. The system also have a chatbot interaction which guides user steps by step to upload images and that results. If cancer is detected the system provides an option to book a doctor's appointment for a quick medical help.

Overall the method includes image preprocessing deep learning features extraction and Smart Prediction. All these steps together create an efficient and easy to use system for detecting colorectal cancer.

V. QUANTITATIVE COMPARISON WITH EXISTING METHODS.

The proposed method is compared with existing machine learning and deep learning methods using matrices like accuracy, precision, recall and F1-score. Traditional methods like random forest naïve bayes give lower accuracy usually between 85% and 92%. This is because they depend on manual feature extraction and cannot handle complex images data well. In contrast deep learning models InceptionV3, ResNet50 and EfficientNetB0 achieve higher accuracy around 97% to 99%. This is because they automatically learn feature from images.

However the proposed ensemble model improves the performance even more by combining different pre-trained models. It achieves very high accuracy around 99.3% with high precision Recall and F1-score. So, it perform better than both traditional methods and single deep learning models. This is because it can learn better reduce overfitting and understand features more effectively. Overall these comparison shows the ensemble method is more accurate reliable and effective for reducing colorectal cancer.

VI. PROPOSED SYSTEM

The system is designed to make colorectal cancer detection simple, fast and accurate using technology. Instead

of human checking it uses a deep learning models to automatically find features from medical images. User can easily upload an image and the system will process it automatically without any complicated steps. The image is then sent to train models like InceptionV3, ResNet50 and EfficientNetB0. These models analyze the image and detect tumor patterns or signs of cancer. Finally the system gives the result quickly which helps in faster diagnosis without waiting for manual checking.

In addition the system has a chatbot to help user interact easily. It ask the user to upload image and show the result in a clear and simple way. So the system is user friendly and even people without technical knowledge can use it. Another important feature is doctor booking. If cancer is detected the system shows nearby doctor and allow user to book an appointment immediately. This helps users take quick action instead of delaying treatment.

All the results predictions and recommendation are clearly shown on the users screen. This system uses deep learning for accurate detection and also provides user-friendly support along with healthcare services. So it becomes powerful and easy to use tools for tumor detection.

VII. IMPLEMENTATION DETAILS

Our colorectal cancer detection system is mainly a software basis system. It uses medical image processing and deep learning techniques to improve accuracy and speed. This system does not need any extra hardware like sensors to give good results.



Users simply upload a medical image through an easy interface which is the starting point of the system. Then the image goes through preprocessing like resizing and normalization using tools such as NumPy and OpenCV. This step makes sure that all images are in the same format and ready for analysis. Sometimes small improvements are also made to make the image clearer.

After preprocessing the image is sent to feature extraction. The system uses deep learning models like InceptionV3 and EfficientNetB0 which are already trained using TensorFlow and Keras. These models analyze the image

and identify cancer features similar to how a doctor examines a medical image.

After the data goes to prediction module. The system decides whether the image is “tumor detected” or “normal”. This helps save a lot of time compared to manual diagnosis. Then, the result is shown through a simple interface or a chatbot. If cancer is detected the system also gave an option to book a doctor’s appointment making the process easy and quick. Overall the system is efficient easy to use and gives a fast and accurate results.

VIII. PERFORMANCE METRICS

To check if the system is working well we use different performance measures. These measure help us understand how correctly and quickly the system classifies images as Tumor or Normal. By looking at these values we can know how reliable and efficient the system is. The first metric is accuracy which shows how the correct predictions are. It is numbered of correct results compared to the total number of images. Higher accuracy means the system is good at identifying both cancer and normal cases.

Another important matrices is precision which tells us out of all images predicted as tumor how many are actually tumor. This help reduce false alarms and avoid marking normal images as cancer. Next is recall also called sensitive which shows how well the system finds all real tumor cases. Higher recall means fewer cancer cases are missed which is very important in healthcare. Finally we use the F1 score which combines precision and recall Indias single value. It helps us balance false positives and false negatives under gives an overall performance of the system.

IX. DISCUSSION

The result of the system shows how powerful deep learning is classifying medical images the system can clearly classify images. As Tumor or Normal. Among all the model tested inceptionV3 gave the highest accuracy showing it is very good at finding complex features. Compared to traditional methods like SVM and Random Forests deep learning models perform better. They give high precision most stable results and more reliable predictions.

Also the system is very effective and easy to use. It has a clean interface and chatbot which makes it a simple for user to upload images and get results. These reduces the time of diagnosis and also reduces manual work. Another useful feature is doctor appointment booking which contains the

prediction to real medical treatment. Overall these features make the system very useful for healthcare.

In conclusion the system is accurate, fast and efficient for colorectal cancer. It reduces human work helps in early detection and improves healthcare decisions. Overall the system has strong potential to improve healthcare survives and support better patient care.

X. CONCLUSION

We tested the system at different stages and it gives consistent results. These deep learning system is designed for colorectal cancer detection. It helps in analyzing images detecting cancer at an early stage and classifying images as humor are normal. By using pre-trained CNNmodels the system reduces manual work and improves accuracy. It can also give quick prediction which helps in early and timely detection of cancer.

In addition the chatbot and appointment features helps in better healthcare management. Overall the system shows how modern technology can be used in medical diagnosis. It also helps doctors to give treatment that is effective and on time.

XI. RESULT AND FUTURE SCOPE

The result include traing dataset and prediction:



Figure 1: Training dataset used for model development



Figure 2: Prediction output of colorectal cancer detection

Using machine learning and deep learning these colorectal cancer detection systems work very well. The training data set includes medical images which are used to

train models like InceptionV3, ResNet50, and EfficientNetB0. When a new image is given the system predicts whether it is cancer or not with very high precision.

Among all the models InceptionV3 perform the best in terms of accuracy, precision, recall and F1 score. The system achieves more than 99% accuracy showing it can detect colorectal cancer very reliable especially at early stages. It reduces manual work and makes diagnosis faster. It also helps doctors in making better decisions. Overall the systems can make a big impact in cancer diagnosis. However some improvements are still needed to make it more effective in real world use.

In the future the system can be improved in many ways. One way is to use more advanced deep learning models and larger more diverse datasets. This will improve accuracy and help the system understand different types and stages of cancer.

REFERENCES

- [1] O. Singh and K. K. Singh, "An approach to classify lung and colon cancer of histopathology images using deep feature extraction and an ensemble method", *International Journal of Information Technology*, vol. 15, no. 8, pp. 4149-4160, 2023.
- [2] Dildar, S. Akram, M. Irfan et al., "Skin cancer detection: a review using deep learning techniques", *International journal of environmental research and public health*, vol. 18, no. 10, pp. 5479, 2021.
- [3] S. Garg and S. Garg, "Prediction of lung and colon cancer through analysis of histopathological images by utilizing Pre-trained CNN models with visualization of class activation and saliency maps", *Proceedings of the 3rd Artificial Intelligence and Cloud Computing Conference*, pp. 38-45, 2020.
- [4] Y. Su, D. Li and X. Chen, "Lung nodule detection based on faster RCNN framework", *Computer Methods and Programs in Biomedicine*, 2021.
- [5] S. Ramani, J. Natarajan, M. Karuppiah, S. K. H. Islam, M. M. Hassan and G. Fortino, "Lung cancer prediction using higher-order recurrent neural network based on glowworm swarm optimization", *Neural Computing and Applications*, vol. 32, pp. 4373-4386, 2020.
- [6] R. S. Yadav, "Data analysis of COVID-2019 epidemic using machine learning methods: a case study of India", *International Journal of Information Technology*, vol. 12, no. 4, pp. 1321-1330, 2020.
- [7] T. Babu, D. Gupta, T. Singh and S. Hameed, "Colon cancer prediction on different magnified colon biopsy images", *Proc. 10th Int. Conf. Adv. Comput. (ICoAC)*, pp. 277-280, Dec. 2018.
- [8] Da, R. V. Medeiros, S. A. Peixoto, S. P. P. da Silva and P. P. R. Filho, "Lung nodule classification via deep transfer learning in CT lung images", *2018 IEEE 31st International Symposium on Computerbased Medical Systems (CBMS)*, pp. 244-249, 2018.
- [9] Zhang, Z. Yang, L. Gong, S. Jiang and L. Wang, "Classification of benign and malignant lung nodules from CT images based on hybrid features", *Physics in Medicine & Biology*, vol. 64, no. 12, 2019.
- [10] M. Suren, P. W. C. Prasad, A. Alsadoon, A. K. Singh and A. Elchouemi, "Lung cancer detection using CT scan images", *Procedia Computer Science*, vol. 125, pp. 107-114, 2018.