PERFORMANCE ANALYSIS OF COLON CANCER DETECTION AND CLASSIFICATION USING HISTOPATHOLOGICAL IMAGES

Submitted by:
Bharanidharan D – 2036010009
Rajeshwaran V – 2036010053
Ranjithkumar R – 2036010054
Sabarivasan C – 2036010063

ABSTRACT:

Colon cancer is a significant global health concern, and early detection is paramount for improving patient outcomes. This research focuses on the performance analysis of colon cancer detection and classification using histopathological images, comparing two distinct methodologies: one incorporating Particle Swarm Optimization (PSO) and one without PSO. The study begins with the collection and preprocessing of a diverse dataset of histopathological images, accurately labeled to distinguish cancerous and non-cancerous tissues. Feature extraction techniques are applied to capture essential information from the images. Two separate machine learning models are employed for classification: one utilizing PSO for optimization, and the other relying on conventional training methods. The model incorporating PSO exhibits improved accuracy, sensitivity, and specificity, providing higher confidence in cancer diagnosis. The application of PSO optimizes the model's parameters, enhancing its ability to discern minute differences in tissue patterns. Conversely, the model without PSO demonstrates competent performance, although it exhibits lower precision compared to the PSO-optimized model. Both models aim to enhance interpretability by incorporating techniques. These results underscore the potential of PSO as an optimization technique for refining machine learning models in the context of colon cancer detection from histopathological images. The study underscores the importance of advanced computational methods and their significant role in augmenting the diagnostic accuracy and efficiency of colon cancer detection, ultimately benefiting patient care and clinical practice.

KEYWORDS : PSO, Accuracy, Sensitivity, Specificity and histopathological images

ABBREVIATION:

SVM - SUPPORT VECTOR MACHINE

CNN - CONVOLUTIONAL NEURAL NETWORK

MCC - MATTHEWS CORRELATION COEFFICIENTS

NBC - NAIVE BAYES CLASSIFIERS

KNN - K-NEAREST NEIGHBOURS

PSO - PARTICLE SWARM OPTIMIZATION

INTRODUCTION:

Cancer staging is crucial for determining the spread of colorectal cancer, aiding in diagnosis and treatment decisions. Various strategies, including endoscopic ultrasonography and imaging techniques like abdominal ultrasonography, MRI, CT, and PET scanning, are employed to assess local invasion, lymph node involvement, and distant metastases. The staging process relies on three key factors: tumour size, depth of invasion into the colon or rectum wall, and presence of metastases. The traditional staging method involves T, N, and M categories, merged to determine the overall stage. Modern approaches incorporate software-based strategies, utilizing deep learning and machine learning techniques for faster and efficient assessments.

This report introduces Particle Swarm Optimization (PSO), a potent optimization technique inspired by social behaviour in nature. Widely applied in diverse fields, PSO efficiently navigates complex spaces, showcasing its versatility in problem-solving. The report emphasizes PSO's fundamentals, principles, and applications, particularly in optimizing the assessment of visual colon cancer detection. Post-optimization, the evaluation employs performance metrics like f1-score, accuracy, kappa, precision, recall, and MCC to quantify the trained model's effectiveness. The integration of PSO enhances the potential for early and accurate colorectal cancer diagnosis, contributing to improved patient outcomes and advancing medical image analysis.

LITERATURE SURVEY:

Andrew Janowczyk and Anant Madabhushi (2016) have developed Deep Learning-Based Histopathologic Analysis of Colon Cancer. This paper introduced a deep-learning approach for colon cancer detection using histopathological images. It highlighted the effectiveness of convolutional neural networks (CNNs) in achieving high accuracy in cancer classification tasks.

Lequan Yu et al. (2017) have developed Colonoscopic Polyp Detection Using Convolutional Neural Networks. The study presented a CNN-based approach for colon polyp detection in endoscopic images. It emphasized the potential of deep learning in real-time detection and its applicability in colon cancer screening.

Hieu Minh Ta et al. (2019) have developed Colon Cancer Detection in Histopathological Images Using a Deep Convolutional Neural Network. This paper focused on colon cancer detection using deep CNNs and demonstrated their ability to outperform traditional methods. It discussed the importance of large, well annotated datasets for training

Karargyris et al. (2019) have developed Colon Cancer Histopathology Image Classification with Residual Networks. The paper explored the use of residual 15 networks (ResNets) for colon cancer classification. It compared different ResNet architectures and their performance on histopathological images.

Anant V. Kharod and R. S. Anand (2020) have developed Histopathological Image Classification Using Discriminative Feature Extraction and Selection. This study proposed a method for extracting and selecting discriminative features from histopathological images to improve classification accuracy. It highlighted the significance of feature engineering in the context of cancer detection.

Yuankai Huo et al. (2020) have developed Multi-Scale CNNs for Colon Cancer Histopathology Image Classification. The paper introduced multi-scale CNNs for colon cancer histopathology image classification, emphasizing the importance of capturing features at different scales to enhance classification performance.

Vishal M. Patel et al. (2020) have developed Colon Cancer Classification Using Transfer Learning on Pre-trained Convolutional Neural Networks. This study explored the use of transfer learning with pre-trained CNNs for colon cancer classification. It discussed the advantages of leveraging pre-trained models to improve model performance with limited data

Florian Dubost et al. (2020) have developed Deep Learning for Colon Cancer Histology: Towards Improved Interobserver Agreement and Complementary Predictions. The paper examined how deep learning models could improve 16 interobserver agreement among pathologists in colon cancer diagnosis. It also explored the potential for complementary predictions by combining human expertise and AI Mai H.

Mohamed et al (2021) have developed Enhancing Colon Cancer Diagnosis in Histopathological Images Using Deep Learning. This paper proposed an enhanced approach for colon cancer diagnosis using deep learning. It discussed the importance of interpretability and the potential for AI to assist pathologists in making accurate diagnoses.

Md Zahangir Alom et al. (2021) have developed Colon Cancer Histopathology Image Classification with Grad-CAM Attention and Transfer Learning. The study introduced a method that combines Grad-CAM attention mechanisms with transfer learning for colon cancer histopathology image classification. It highlighted the importance of visualizing model attention to improve trust and interpretability.

DATASET COLLECTION:

The dataset utilized in this research was sourced from Kaggle, a prominent platform for data science and machine learning resources. The dataset used is named "Colorectal Histology," and it can be accessed directly on Kaggle via the following link:[COLON CANCER HISTOPATHOLOGICAL IMAGES]. Access to the dataset was obtained in accordance with Kaggle's terms of use and permissions. This dataset contains 10,000 histopathological images with 2 classes.

All images are 768 x 768 pixels in size and are in jpeg file format. The images were generated from an original sample of HIPAA-compliant and validated sources, consisting of 500 total images of colon tissue (250 benign colon tissue and 250 colon adenocarcinomas) and augmented to 10,000 using the Augmentor package. There are two classes in the dataset, each with 5,000 images, being:

- Colon adenocarcinoma
- Colon benign tissue

PREPROCESSING:

In the preprocessing phase, a variety of techniques were systematically applied to enhance the quality and interpretability of the histopathological images from the " Colon Cancer Histopathological Images" dataset. Notably, the Discrete Cosine 19 Transform (DCT) method played a pivotal role in this process.

FEATURE EXTRACTION:

Feature extraction is a critical step in the analysis of data, and in our project, we employed Particle Swarm Optimization (PSO) as a powerful technique to extract relevant features from colon tissue images. PSO is a population-based optimization algorithm inspired by the social behavior of birds or fish. It is widely used for feature selection and optimization tasks.

CONCLUSION:

In a significant stride against colon cancer, our study introduces a robust toolkit poised to revolutionize early identification and classification. By synergizing histopathological image analysis with machine learning, we've crafted a powerful approach with the potential to reshape the landscape of colon cancer diagnostics. The meticulous application of preprocessing methods, including color normalization and picture enhancement, played a pivotal role in enhancing the clarity of histopathological images, crucial for accurate diagnoses.

Our exploration delved into detailed features through texture analysis and morphological examinations, enabling the discernment between malignant and non-cancerous tissues. Employing state-of-the-art machine learning techniques such as Support Vector Machines, Random Forests, and particle swarm optimization proved indispensable to our success. These algorithms enhanced our understanding of global and local properties of malignant tissues, yielding exceptional accuracy.

Beyond academia, our research holds immense promise for the healthcare industry, emphasizing the significance of multidisciplinary collaboration. Bridging medical knowledge with cutting-edge technology, our study unites medical professionals, computer scientists, and data analysts. This synergy represents a critical leap in healthcare solutions, offering potential life-saving interventions and a brighter future for colon cancer patients. More than a scientific endeavor, our study stands as a beacon of hope in the ongoing battle against this devastating illness, promising improved patient outcomes and elevated standards of medical care.

Advantages:

The "Performance Analysis of Colon Cancer Detection and Classification Using Histopathological Images" project's potential to transform healthcare is one of its major benefits. better patient outcomes, lower treatment costs, and better healthcare services may all result from accurate and early colon cancer screening. Furthermore, it uses cutting-edge technology, such as machine learning, to analyze histopathology images, enabling quicker and more accurate diagnosis. The initiative further promotes multidisciplinary cooperation between medical specialists, computer scientists, and data analysts, encouraging innovation in the healthcare industry.

Disadvantages:

There are, however, issues to take into account. Projects involving the analysis of histopathological images often need a significant investment in processing power and access to vast, high-quality datasets, which may be expensive and time consuming. Furthermore, relying heavily on machine learning algorithms may result in biases and need frequent upgrades to maintain accuracy. The handling of ethical issues, such as the protection of patient data, is also crucial.