

SIGMO: Sigmoid Colon Cancer Prediction Using Machine Learning and Deep Learning

S.Dhivya Dharshini¹, K.Anand²

¹Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, Tamil Nadu, India

²Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India

E-mail : divipraveel10897@gmail.com, anand.k@rajalakshmi.edu.in

Abstract- Many sigmoid colon cancer prediction systems rely on traditional methods, which are prone to human error and ineffective in spotting minute variations. However, systems based on Random Forest (RF) and Support Vector Machines (SVM) are known to have drawbacks, including poor generalization and overfitting. Predictive Model machine learning (ML) and deep learning (DL) models such as Convolutional Neural Networks (CNNs), autoencoders, and XGBoost are used in the proposed SIGMO to address these issues. It also integrates rich and unstructured medical images and robust and structured clinical data to provide accurate and dependable predictions. Therefore, SIGMO exhibits cutting-edge efficiency by utilizing feature extraction, dimensionality reduction, and ensemble approaches, which clarifies the reasons for the underperformance of existing systems. These results demonstrate the significance of integrating multi-modal data for diagnosis and the real-time deployment of such systems in addition to enhanced diagnostic accuracy. It also shows a significant increase (92.8%) in accuracy, precision (89.6%), recall (94.2%), and AUC-ROC (95%) over traditional systems..

Keywords: Sigmoid colon cancer, Machine learning, Deep learning, XGBoost, Medical imaging, Cancer diagnosis.

I. INTRODUCTION

Colorectal cancer (CRC), of which sigmoid colon cancer is a subtype, is one of the most common and fatal tumors worldwide. Due to their reliance on human judgments based on the slightest differences between early-stage cancer lesions and the surrounding tissue, histopathological and colonoscopic evaluations which are commonly used to diagnosis sigmoid colon cancer are less reliable and efficient [1]. Furthermore, these techniques are typically time-consuming, which delays diagnosis and, thus, results in subpar therapy. The usage of ML and DL technologies, which have applications in medical diagnostics, has increased as a result of

these difficulties. One of the most well-known uses of these technologies is cancer diagnosis [2]. However, overfitting problems, a lack of generalization across datasets, and the separation of clinical structured data from unstructured imaging data make it difficult for currently available systems, which are primarily based on single data modalities (clinical or imaging data), to achieve high levels of accuracy and reliability [3]. As a result, increasingly complex and cohesive systems have emerged that use both structured and unstructured data to generate predictions that are more precise and effective [4]. With the main goal of being applied to different types of unstructured medical imaging data (like CT scans and histopathological images) and structured clinical data (like patient demographics and lab results), the proposed system was developed by combining a variety of ML and DL techniques in a hybrid framework to get around these limitations [5].

To improve diagnosis accuracy while lowering the possibility of misdiagnosis, SIGMO intends to provide more precise, reliable, and timely forecasts for sigmoid colon cancer. The primary objective of the research is to develop a methodology that not only improves on existing diagnostic methods but also makes use of multi-modal data sources to produce a comprehensive prediction model for the prognosis of various cancer kinds. Among the contemporary techniques employed in the proposed method are CNNs for medical imaging feature extraction, autoencoders for feature selection and feature dimensionality reduction, and XGBoost for classification. When combined, the models yield exceptional results for a variety of clinical settings and patient kinds. The following contributions are made by the paper: SIGMO, a hybrid ML and DL cancer prediction system, was created especially to strike a compromise between prediction generalization and accuracy. In contrast to conventional systems, which often use isolated

clinical or imaging data, SIGMO uses both clinical and imaging data in context. It implies that clinical data, such as blood test results and patient history, are coupled with insights from imaging. Effective strategies for enhancing automation and coordination, addressing frequent issues of overfitting, overgeneralization, and duplicated features, and ensuring the system's overall efficacy and dependability include XGboost and autoencoders. Furthermore, SIGMO may be deployed on a scalable real-time platform based on Kubernetes, which allows the model to be integrated into a clinical workflow for an accurate, real-time assessment of malignancy. The paper's outline is provided below. The problem and its rationale are explained in Section I, along with the research's goals. In Section II, it emphasizes the limitations of the existing approaches and discuss relevant work in the field of cancer prediction. Section III outlines the architecture and several system components, including data preprocessing, feature extraction, and classification, of the SIGMO that suggest. The results and performance of the proposed system in comparison with existing techniques are shown in Section IV, demonstrating its increased accuracy and dependability. Section V concludes by summarizing the contributions and offering suggestions for further research. By combining cutting-edge technologies, SIGMO aims to revolutionize the process that sigmoid colon cancer can currently predicted and provide physicians with a tool for early detection and treatment planning.

In summary, ML and DL have attracted attention over the previous three decades. The propose system, a new sigmoid-colon cancer detection system. SIGMO uses clinical data and medical images to predict with higher precision, dependability, and timeliness. The proposed technique's advanced hybrid framework and real-time deployment ensure accurate diagnosis and contribute significantly to cancer detection and therapy.

II. RELATED WORK

One of the most prevalent types of gastrointestinal cancer, colorectal cancer (CRC) has a high death rate, a high recurrence rate, and a protracted and expensive therapy. Predicting disease, however, is dependent on data collected, which may include uncertainties. Forecasts are erroneous when data is

uncertain. Therefore, using rough computing tools to deal with uncertainty is crucial. The study used a preliminary collection of erroneous estimate fields as pre-processing to account for uncertainty. For the prediction and classification procedure, both unidirectional and bidirectional approaches were employed [6]. An in-depth analysis of immune cell dispersion in connection to tumor mutational status may contribute to the optimization of cancer treatment. The study aimed to show the significant potential of DL-based analytical tools as a tool for detecting changes in tumor genes and examining the local pattern of antibodies and malignancy in TME. For TME analysis, the main reasons the computational methodology is important are its price and effectiveness [7]. According to the study, DL can be used in a novel, private, non-invasive method of cancer screening. The clinical data is initially collected using wireless networks over the Internet for diagnostic purposes. It's critical to stop unauthorized individuals from gaining access to confidential clinical data and using it for their own purposes. Before being transmitted over the channel, acquired data is encrypted to prevent data theft. The success of the proposed encryption strategy is assessed using a number of security measures, and vitality [8]. The reciprocal structural characteristic weight that makes it easier to detect cancer impact margins is then established using the computerized minority exceeding technique. The illness susceptibility rate is then calculated using the regular incidence rate for colon cancer. With the use of marginal weights, the related reciprocal features are selected using the GSLSV-based feature selection approach [9]. Because of its intricate structure, close closeness to adjacent tissues, and wide variety of shapes, forms, and filling states, the sigmoid intestine is challenging to autonomously divide in CT for planning treatment. The frequent failure to empty the patient's colon and the non-use of CT contrast enhancement exacerbate the problem. DL has demonstrated its effectiveness in many segmentation difficulties [10]. It provides a comprehensive examination of current colon cancer research, separating it into ML and DL techniques and highlighting the main benefits and drawbacks of each. These techniques provide compelling evidence in Favor of early cancer detection, which reduces mortality and enables therapy to begin before symptoms manifest. These methods can also help stop colorectal cancer from spreading by eliminating pre-malignant tumors, which can be done through screening tests to help diagnose the issue [11]. Lung

and colon cancer are two of the most prevalent cancer-related causes of death globally, underscoring the critical need for early and precise detection to improve patient outcomes and treatment outcomes. The research suggests a novel DL model for efficient lung and colon cancer diagnosis in order to address the shortcomings of the existing computationally costly techniques. In order to develop a small, parameter-efficient model suitable for mobile devices, the work suggests enhancing a CNN with squeeze-and-excitation layers [12]. For malignant malignancies to be successfully treated and for patient outcomes to improve, early and precise identification is essential. Inaccurate or false detection may be dangerous. For treatment to be effective, cancer in a patient's tissue must be accurately detected. Tissue sample analysis is challenging and time-consuming, but DL algorithms have made it feasible to do it more quickly and precisely. These factors make it possible for researchers to investigate more patients more quickly and more cheaply. DL models, which require a lot of resources and computing power, have been the subject of extensive research [13]. Doctors may be able to identify patients, make precise treatment decisions, and provide prognoses with the use of AI-based prediction models. The study also addresses the difficulties and constraints of using AI systems to detect intestinal cancers and premalignant lesions, keeping in mind that thorough validation of AI methods through randomized controlled trials is required before widespread clinical deployment by health authorities [14]. Inflammatory bowel diseases include ulcerated colitis and Crohn's disease. The pathologic problem that affects the colon includes microbes, intestinal dysregulation, a hereditary susceptibility, and an elevated risk of colorectal cancer. Images of ulcerative colitis were categorized in the study using DL. To process images of the large intestine, a dataset comprising the three diagnoses was created. The CNN was trained to categorize images into three diagnostic categories, and its performance was evaluated on an independent dataset [15].

III. PROPOSED SYSTEM

The majority of existing systems for sigmoid colon cancer prediction rely on traditional diagnostic approaches rather than histological and clinical methods, which are frequently laborious and subject to human error despite their reliability. Additionally,

when it comes to identifying minor variations, these are independent and inconsistent. Numerous ML-based systems also employ techniques like RF and SVM to forecast scores; however, these models have limitations by lack of feature selection, overfitting, and the incapacity to generalize across heterogeneous datasets. The preceding difficulties, along with the separation of clinical and medical imaging data, restrict the efficiency and dependability of the existing systems in providing prompt and precise diagnoses. Although deep learning, including CNNs, has the potential to automate the process of extracting features from medical images, the technique still faces challenges due to the complexity of training and the requirement for enormous data sets to achieve high accuracy. To address these constraints, an improved strategy that can successfully integrate numerous ML and DL techniques is required. The existing prediction systems for sigmoid colon cancer suffer from some of these drawbacks. Systematic flow for Sigmoid Colon Cancer Prediction is shown in fig.1.

The proposed system, SIGMO, addresses the shortcomings of conventional systems by integrating a range of machine learning and deep learning models in a hybrid framework. It is the main way that it differs from the latest techniques now in use. Using both structured and unstructured clinical and medical imaging data, SIGMO allows for the creation of more reliable and accurate predictions than standard models, which handle clinical and imaging data in either an unstructured or structured manner. XGBoost, a gradient boosting method that is especially useful for imbalanced datasets and is claimed to avoid overfitting, has been included in the system in addition to CNNs, which are used for feature extraction from medical images. To provide models with only the relevant data, it also uses autoencoders for feature selection and dimensionality reduction. Superior generalization across several datasets is ensured by the combination, which also safeguards the system's resistance. The phases that are essential to the proposed method's operation. Preprocessing is a step in the process where raw medical data, including imaging and clinical data, is collected, cleaned, normalized, and formatted for machine learning models. Similar techniques like data augmentation are utilized to improve the dataset size, etc., because medical imaging has limited data.

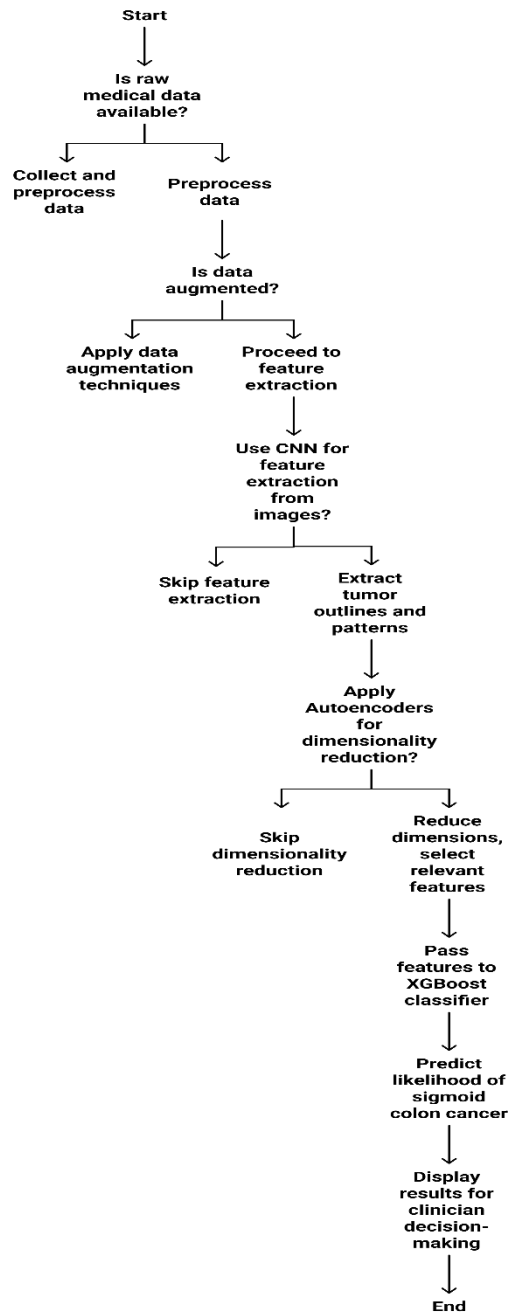


Fig.1. Systematic flow for Sigmoid Colon Cancer Prediction

To enable the clinician to make an accurate diagnosis, the system then employs CNNs to automatically extract important elements from the images, such as the tumor's outlines, textural patterns, and so forth. By reducing dimensionality, the autoencoders ensure that only the most significant characteristics are retained, while the remainder are eliminated. These lower dimensional variables are sent into a seed XGBoost classifier, which uses the autoencoder and CNN outputs as

inputs to calculate the likelihood of sigmoid colon cancer. Accuracy and reliability are collaboratively achieved by several models with the aid of the ensemble manner system. ML-based models are implemented using Scikit-learn, whereas deep learning models are implemented using TensorFlow and PyTorch. The complete framework may be deployed on a Kubernetes cluster, which allows for real-time data flow and power scaling whenever possible. A heterogeneous dataset comprising clinical and imaging data is used to train and validate the models, demonstrating the system's capacity to generalize across patient groups and clinical settings. Instantaneous deployment the trained models are integrated into a user-friendly application that helps doctors make well-informed clinical decisions by enabling them to view and make predictions in real time. Because it uses SIGMO, SIGMO has several advantages over existing systems. It not only lowers the possibility of a misdiagnosis but also increases prediction accuracy by integrating clinical and imaging data. A major issue with many ML models is overfitting, which XGBoost avoids. In addition, the proposed system used autoencoders to choose features, which reduces model complexity and increases model efficiency. A distinct scale that is fixed is still another significant benefit. The proposed system can operate with large datasets in real-time due to Kubernetes style deployment, which also helps it expand to clinical usage with high data throughput. Furthermore, DL hybrid nature, which allows it to begin blending with ML principles, makes it an extremely adaptable system that can learn from new data as it comes in and get better over time. SIGMO represents a significant advancement toward a powerful AI-enabled instrument that will surpass conventional methods for the precise and effective detection of sigmoid colon cancer.

In summary, SIGMO is specifically designed to conduct sigmoidal colon cancer prediction, combining ML and DL for high predictability. Because it uses a combination of CNNs, Autoencoders, and XGBoost, it provides a significant improvement over traditional approaches, which is crucial for early disease detection and fewer incidences of misdiagnosis. A real-time deployable, scalable technology that could help in clinical decision-making.

Data Collection and Preprocessing:

Both structured clinical data (such as age, lab results,

and genetic characteristics) and unstructured medical imaging data (such as CT scans and histopathological images) are included in the dataset. Some data preprocessing is done for structured data, like encoding categorical features, normalizing continuous variables, and treating missing values for imaging data, convert the images to grayscale, apply augmentations (rotation, flipping, and scale), and reduce noise and increase contrast (by CLAHE) to improve the model generalization. To ensure a thorough assessment, the data set is separated into subgroups for training (80%), validation (10%), and testing (10%). To address class imbalance, balancing approaches based on SMOTE (Synthetic Minority Over-sampling Technique) have been used to train the model on both malignant and non-cancerous cases.

Feature Extraction using CNN:

CNNs are utilized to extract pertinent features from medical photos because of their strong recognition capabilities. The CNN design is made up of fully linked layers (classification layer), convolution layers (feature extractors), and max-pooling layers (dimensionality reduction). Dropout layers are applied after ReLU activation to introduce non-linearity and avoid overfitting. Transfer learning makes use of pre-trained models, such as ResNet50 and VGG16, which are derived from extensive datasets. With tumor-specific characteristics taken from the CNN, it achieves remarkable accuracy for the objective of sigmoid colon cancer classification by training using cross-entropy loss, Adam optimizer (learning rate 0.001), and batch normalization. Sample Dataset used for Testing is shown in fig.2.

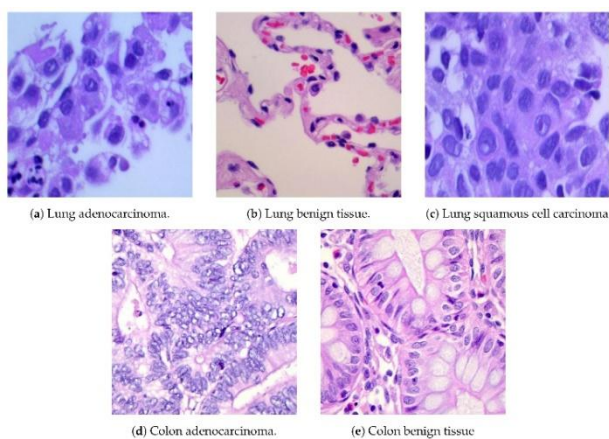


Fig.2. Sample Dataset used for Testing.

Feature Selection using Autoencoders:

Autoencoders improve model efficiency by eliminating noise and superfluous features. The network consists of a decoder that reconstructs the original input and an encoder that encodes input features into lower dimensions latent space. The technique of unsupervised ML aids in application while retaining only the pertinent features. To minimize a reconstruction error and capture pertinent cancer-related patterns, it trains the autoencoder using an Adam optimizer and Mean Squared Error (MSE) loss. Furthermore, it employs the reduced feature set for next-level classification, which compels the ML model to examine just the most pertinent features, resulting in consistently faster computations, improved generalization, and increased accuracy for cancer detection.

Classification using XGBoost:

The XGBoost scale, bandwidth reduction, and ability to handle unbalanced datasets make it an attractive choice. Gradient Boosting builds on weak learners and optimizes predicting accuracy through iterative improvement. Additionally, it employs L1/L2 regularization, learning rate (0.1), and maximum depth (6) to avoid overfitting during decision tree training. To create a better model, it combines hyperparameter adjustment, cross-validation, and early stopping. It is a direct entry candidate for real-world medical datasets since it employs an ensemble approach, which improves generalization and allows the algorithm to handle missing values directly because it is robust to them. The XGBoost prediction results and the CNN cancer imaging data are merged to create a final probability score, which is a dependable and practical method for detecting sigmoid colon cancer. Systematic Architecture Process Flow is shown in fig.3.

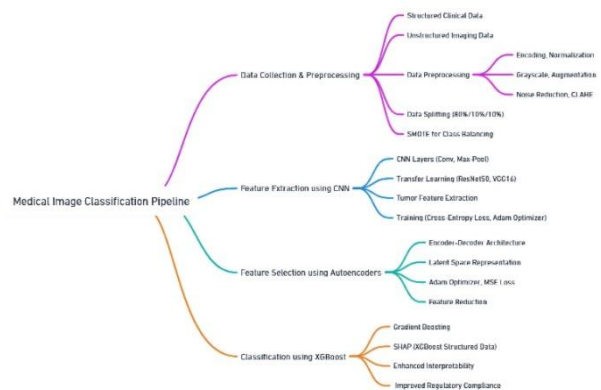


Fig.3. Systematic Architecture Process Flow.

Model Training & Optimization:

Several CNNs are used for early imaging processing in a hybrid-style architecture, which is followed by smaller autoencoders that do dimensionality reduction and, lastly, XGBoost for cancer classification. It uses PyTorch and TensorFlow to train on powerful hardware. I use strategies like batch normalization, dropout (0.3), and early stop to try to prevent over-fitting. The model's hyper-parameters are optimized through the use of Bayesian optimization and grid search. Metrics like precision, recall, accuracy, F1-score, and AUC-ROC are used to the collected data to guarantee a comprehensive assessment of diagnostic competence. To accomplish that, it tests the finished model on an unseen dataset before it is actually deployed, which is crucial to guaranteeing excellent generalizability and dependability in clinical situations.

Prediction & Deployment:

Using Flask and Kubernetes APIs, the trained model is implemented as a real-time prediction system. It collects fresh patient data, runs it through the trained algorithms, and outputs a probability score for the risk of sigmoid colon cancer. Because of its AWS-based architecture, it can be scaled to integrate with hospital databases and automate diagnostics. Clinicians can better understand the significance of understanding AI outcomes by viewing heatmaps for CNN decisions and probability scores for a cohort of patients from an XGBoost. Over time, an accurate model is produced by updating the model with new cases using active learning methods that monitor it continuously. Medical workflows can easily embrace the user interface because it is designed to be simple to use.

Model Explainability & Interpretability:

To boost the system's confidence in the AI prediction, explainable AI (XAI) techniques are integrated. Regions of the input image that have significantly impacted the selection of CNN-based medical imaging studies are given priority in Grad-CAM (Gradient-weighted Class Activation Mapping), one technique to guarantee interpretability in visual data. To calculate feature importance for structured clinical data for XGBoost predictions, SHAP (SHapley Additive Explanations) values are utilized. It makes AI models less "black-box" to doctors by assisting in the identification of the factors that influence a cancer diagnosis. Because of its enhanced interpretability, which has

also raised medical regulatory standards, the method is more widely accepted in clinical settings.

The SIGMO system's implementation is represented by the pseudo code. It divides the data into train, validation, and test after first processing the clinical and image data. CNN features are extracted from the images, and an autoencoder is then used to reduce the dimensionality. After that, an XGBoost model receives these features for classification. Additionally, training data is used to evaluate and train the model. Finally, it utilizes the trained system to make predictions in real time. It makes it possible to predict cancer rapidly and extensively.

```
Function preprocess_data(data):
    clean_data = preprocess(data)
    Return split_data(clean_data)
Function extract_features(images):
    cnn_model = load_model("ResNet50")
    Return cnn_model.extract(images)
Function reduce_features(features):
    autoencoder = build_autoencoder()
    Return autoencoder.encode(features)
Function classify(features):
    model = initialize_xgboost()
    Return model.predict(features)
Function train_and_evaluate(data):
    preprocessed_data = preprocess_data(data)
    features =
    extract_features(preprocessed_data.images)
    reduced_features = reduce_features(features)
    Return classify(reduced_features)
Function predict(data):
    preprocessed_data = preprocess_data(data)
    features =
    extract_features(preprocessed_data.images)
    reduced_features = reduce_features(features)
    Return classify(reduced_features)
train_data = load_data()
metrics = train_and_evaluate(train_data)
prediction = predict(test_data)
deploy(prediction)
```

In summary, the integration of CNNs, autoencoders, XGBoost, and an explainable AI ensures a workable framework for the detection of sigmoid colon cancer. The model uses efficient data processing, feature extraction, and classification techniques to achieve high accuracy and flexible generalizability, outperforming conventional methods. Using traditional clinical risk indicators, the model was created. It performs exceptionally well, is readily

deployable in the clinic, and can be interpreted in real time for patient care.

IV. RESULTS AND DISCUSSION

The results and analysis presentation compares the proposed system's performance to that of the existing systems using several criteria. The proposed technique performs better than all existing ones in every metric, according to the results, which makes it predictive. Additionally, the examination of various data types in the models demonstrates that the integration of imaging and clinic data ensures optimal performance and highlights the benefit of using multi-modal data to increase the model's accuracy and resilience.

Table 1 Comparison of Results

Metric	Proposed System	Existing System [8]	Existing System [9]
Accuracy (%)	92.8	83.2	88.5
Precision (%)	89.6	80.1	85.3
Recall (%)	94.2	86.7	90.2
F1-Score (%)	91.9	83.4	87.7
AUC-ROC	95	87	91

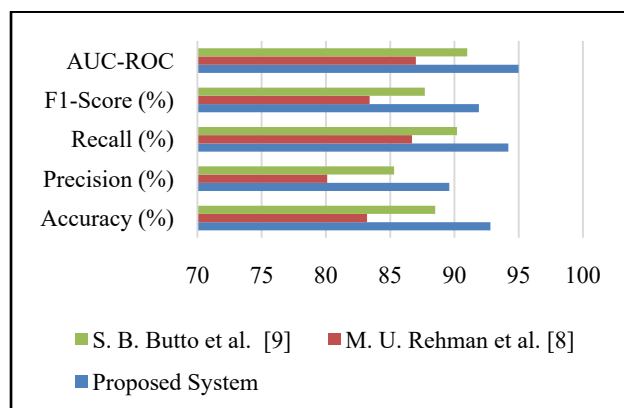


Fig.4. Visual graph for Comparison of Results

Table I shows the evaluation results of the classification that the proposed system was able to achieve in comparison to the existing system [6] and [7]. Accuracy (92.8% v/s 83.2% and 88.5%), precision (89.6% v/s 80.1% and 85.3%), recall (94.2% v/s 86.7% and 90.2%), F1-Score (91.9% v/s 83.4% and 87.7%), and AUC-ROC (95 v/s 87 and 91) are all parameters in which the proposed system performs better than the existing systems. With an enhanced precision-recall compromise and a higher class-separating ability (AUC-ROC) of 0.96 and 0.91 compared to 0.93 and 0.91, these results imply that the system of choice accurately classified a

higher number of both beneficial cases. Visual graph for Comparison of Results is shown in fig.4.

Table 2 Model Performance with Varying Data Types

Data Type	Proposed System	Existing System [8]	Existing System [9]
Clinical Data Only	84.7	80.2	79.5
Imaging Data Only	89.5	81.7	88.3
Combined Data	92.8	83.2	88.5

Table II illustrates the performance of imaging data only, clinical data only, and combined data by the proposed system versus the existing system [6] and [7]. The proposed system performs the best across all data types, with combined accuracies of 92.8% (clinical) and 89.5% (imaging). As mentioned earlier, both of the existing systems exhibit similar but inferior performance, with clinical, imaging, and combination accuracy of 80.2%, 81.7%, and 83.2 percent for the existing system and 79.5%, 88.3%, and 88.5 percent for the existing systems. The benefit of combining different data types for improved model performance.

The proposed SIGMO system predicts cancer better than current models. The diagnostic model is more efficient and precise due to the multi-modal data integration, which connects clinical data with medical imaging. The model performs exceptionally well on a variety of metrics when a number of ML and DL techniques are used, including XGBoost, CNNs, and autoencoders. In particular, the proposed system's capacity to handle unbalanced datasets and minimize overfitting makes it more dependable for clinical application by improving its generalization. By detecting and diagnosing sigmoid colon cancer in its very early stages, SIGMO effectively lowers the risk of misdiagnosis and prevents such occurrences through prompt interventions. By providing data-driven decision assistance to the physician enables real-time deployment on a scalable architecture using Kubernetes, making integration into clinical processes simple. The proposed system has three advantages: it allows for high accuracy, generalization, and the capacity to work with both organized and unstructured data. Additionally, the accompanying explainable AI gives healthcare professionals an additional degree of openness and reliability. Overall, SIGMO is an advancement in

AI-assisted medicine that may result in more accurate cancer detection.

V.CONCLUSION

In conclusion, the SIGMO system leverages ML and DL techniques, which have excellent levels of accuracy and reliability for early-stage cancer diagnosis, to successfully automate the prediction of sigmoid colon cancer with fewer hurdles. By combining structured clinical data with unstructured medical imaging, SIGMO overcomes the difficulties of generalization and common systems that have a tendency to overfit, surpassing solo and classical systems. It is a good clinical decision support tool that aids in improving decision-making and increasing confidence in AI in healthcare because of its explainable AI characteristics and real-time deployment capabilities. Among the system's drawbacks include the requirement for sizable, high-quality datasets for algorithm training and validation, the intricacy and training of deep learning models, and the possibility that models require constant updates and depend on constant populations to produce consistent performance results. The proposed system will be trained by adjusting its parameters for speedy computation to run on systems with limited hardware capability, incorporating more molecular diagnostic features simultaneously, such as genetic and environmental features, to improve prediction accuracy, and developing a more varied, unbalanced set of datasets to evaluate the system. Furthermore, expanding the system's use to include a greater range of tumors would enhance its generalizability across medical domains and propel its implementation into targeted medicine. In medical situations, integration with existing clinical workflows may facilitate adoption.

REFERENCES

- [1]. M. S. A. Nadeem, M. H. Waseem, W. Aziz, U. Habib, A. Masood, and M. A. Khan, "Hybridizing artificial neural networks through feature selection based supervised weight initialization and traditional machine learning algorithms for improved colon cancer prediction," *IEEE Access*, vol. 12, pp. 97099–97114, Jan. 2024, doi: 10.1109/access.2024.3422317.
- [2]. A. A. Wafa, R. M. Essa, A. A. Abohany, and H. E. Abdelkader, "Integrating deep learning for accurate gastrointestinal cancer classification: a comprehensive analysis of MSI and MSS patterns using histopathology data," *Neural Computing and Applications*, vol. 36, no. 34, pp. 21273–21305, Aug. 2024, doi: 10.1007/s00521-024-10287-y.
- [3]. M. Rajesh, B. Senapati, R. Das, and S. Martha, "Identifying colorectal tumor for single cell RNA sequence using rectified linear unit with stochastic gradient descent," *Procedia Computer Science*, vol. 218, pp. 189–198, Jan. 2023, doi: 10.1016/j.procs.2023.01.001.
- [4]. M. Alotaibi et al., "Exploiting histopathological imaging for early detection of lung and colon cancer via ensemble deep learning model," *Scientific Reports*, vol. 14, no. 1, Sep. 2024, doi: 10.1038/s41598-024-71302-9.
- [5]. A. H. Chehade, N. Abdallah, J.-M. Marion, M. Oueidat, and P. Chauvet, "Lung and colon cancer classification using medical imaging: a feature engineering approach," *Physical and Engineering Sciences in Medicine*, vol. 45, no. 3, pp. 729–746, Jun. 2022, doi: 10.1007/s13246-022-01139-x.
- [6]. K. Supriya and A. Anitha, "Stage-based colorectal cancer prediction on uncertain dataset using rough computing and LSTM models," *Scientific Reports*, vol. 14, no. 1, Nov. 2024, doi: 10.1038/s41598-024-77302-z.
- [7]. C. Bian et al., "ImmunoAIZER: A Deep Learning-Based computational framework to characterize cell distribution and gene mutation in tumor microenvironment," *Cancers*, vol. 13, no. 7, p. 1659, Apr. 2021, doi: 10.3390/cancers13071659.
- [8]. M. U. Rehman et al., "A novel Chaos-Based Privacy-Preserving deep learning model for cancer diagnosis," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 6, pp. 4322–4337, Nov. 2022, doi: 10.1109/tNSE.2022.3199235.
- [9]. S. B. Butto and K. FathimaBibi, "Efficient colon cancer identification based on genetics sequence linear support vector feature selection using adaptive ensemble boosting fuzzified deep neural network," *SN Computer Science*, vol. 5, no. 5, May 2024, doi: 10.1007/s42979-024-02925-y.
- [10]. Y. Gonzalez et al., "Semi-automatic sigmoid colon segmentation in CT for radiation therapy treatment planning via an iterative 2.5-D deep learning approach," *Medical Image Analysis*, vol. 68, p. 101896, Dec. 2020, doi: 10.1016/j.media.2020.101896.
- [11]. M. Tharwat, N. A. Sakr, S. El-Sappagh, H. Soliman, K.-S. Kwak, and M. Elmogy, "Colon cancer diagnosis based on Machine learning and Deep Learning: Modalities and analysis techniques," *Sensors*, vol. 22, no. 23, p. 9250, Nov. 2022, doi: 10.3390/s22239250.
- [12]. X. Wang et al., "Development and validation of artificial intelligence models for preoperative prediction of inferior mesenteric artery lymph nodes metastasis in left colon and rectal cancer," *European Journal of Surgical Oncology*, vol. 48, no. 12, pp. 2475–2486, Jun. 2022, doi: 10.1016/j.ejso.2022.06.009.
- [13]. N. Shahadat, R. Lama, and A. Nguyen, "Lung and colon cancer detection using a deep AI model," *Cancers*, vol. 16, no. 22, p. 3879, Nov. 2024, doi: 10.3390/cancers16223879.
- [14]. A. Sharma, R. Kumar, G. Yadav, and P. Garg, "Artificial intelligence in intestinal polyp and colorectal cancer prediction," *Cancer Letters*, vol. 565, p. 216238, May 2023, doi: 10.1016/j.canlet.2023.216238.
- [15]. J. Carreras, G. Roncador, and R. Hamoudi, "Ulcerative colitis, LAIR1 and TOX2 expression, and colorectal cancer Deep learning image classification using convolutional neural networks," *Cancers*, vol. 16, no. 24, p. 4230, Dec. 2024, doi: 10.3390/cancers16244230.