STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **14**, December 2025, pp 3447–3467.
Published online in International Academic Press (www.IAPress.org)

An Explainable Vision Transformer-Based Web Application for Medical Decision-Making: Case of Colon Cancer

Mohamed Abderraouf Ferradji [1,2,*], Asma Merabet [3], Faycal Zetoutou [1], Samir Balbal [1,2]

[1] *Department of computer science, Faculty of Sciences, Ferhat Abbas University Setif-1, Setif, Algeria*
[2] *Artificial Intelligence Laboratory, Department of computer science, Faculty of Sciences, Ferhat Abbas University Setif-1, Setif, Algeria*
[3] *DeLIAOA Laboratory, Department of Computer science, University of Oum El Bouaghi, Oum El Bouaghi, Algeria*

Abstract Despite the impressive research related to the application of artificial intelligence in the medical field, its adoption in real clinical settings, especially in medical decision-making, remains very limited. Therefore, our objective in this work is to develop a deep learning-based web application that supports medical decision-making. In addition to enabling efficient interaction and knowledge sharing among medical professionals, our web application also provides an accurate prediction system for colon cancer. This system is based on a Vision Transformer (ViT) deep learning model, which is characterized by its attention mechanism that ensures rich contextual representations and captures long-distance dependencies within images. To promote physicians' confidence in the intelligent system, our approach provides clear visual explanations of the ViT predictions using the XAI method LIME. The validation of our model was conducted on a merged dataset of LC25000 and DigestPath images, with an additional external evaluation on the EBHI-Seg dataset. The experimental results demonstrate the competitive performance of the proposed ViT-based approach, which achieved perfect accuracy on the LC25000 dataset, 94.90% on the challenging merged dataset, and a robust accuracy of 92.17% on the unseen EBHI-Seg dataset. This remarkable performance makes the model suitable for real-world clinical applications.

Keywords Vision Transformer, Explainable AI, Web Application, Smart System, Medical Decision Support, Colon Cancer

DOI: 10.19139/soic-2310-5070-3035

1. Introduction

With the increasing need of efficient healthcare services around the world due to modern lifestyles that cause serious diseases, the healthcare domain has become more challenging with increasingly complex clinical cases[1, 2, 3]. This requires healthcare professional to continually update their medical knowledge, to connect to each other, and to adopt the latest practices and technologies to effectively address the needs of their patients [4, 5, 6].
Making efficient medical decisions is considered the most important process that can improve healthcare quality. Therefore, errors in this critical process must be avoided [7]. The lack of physician experience in generating relevant medical decisions or validating probabilities is the main obstacles that affects healthcare quality [8]. The lack of accessible patient information and poor communication between medical staff also present serious challenges [9]. Therefore, trying to integrate and profit from what the information technologies (IT) field has reached can greatly enhance medical healthcare. Nowadays, many powerful tools have been developed by IT engineers that can enhance communication between healthcare professionals and patients and make it more user-friendly [10, 11]. Additionally, they also present modern artificial intelligence (AI) approaches that identify patterns within large amounts of any data types. These approaches are very promising for detecting abnormalities in medical imaging, which can be challenging for human analysis [12].

*Correspondence to: Mohamed Abderraouf Ferradji (Email: mohamed.ferradji@univ-setif.dz)

ISSN 2310-5070 (online) ISSN 2311-004X (print)

Machine learning (ML) and deep learning (DL) are subfields of artificial intelligence that have seen significant growth in various domains, especially in the medical field. A large number of research studies aim to adopt new ML and DL technologies to support medical decision-making. The detection of cancerous cells in medical imaging is one of the major research areas in AI that has attracted considerable interest in recent years [13].

With over 200 different types of cancer identified, colorectal cancer, which is the focus of our study, is the third most common type of cancer and the second leading cause of death due to cancer. The continuous increase of this cancer type is largely related to the modern lifestyle, which is characterized by reduced physical effort and unhealthy diets based on fast food and processed products. In addition, smoking and alcohol consumption are major risk factors contributing to the spread and severity of this cancer type. Since colorectal cancer often shows no symptoms in its initial stages, early diagnosis is considered one of the most important factors in reducing its mortality rate [14, 15, 16, 17].

Histopathology is one of the most common diagnostic methods for colorectal cancer. The preparation process involves cutting tissue samples into extremely thin layers, and then staining them with specific chemicals that highlight cellular components (nuclei in blue and cytoplasm in pink). These samples are then examined under a light or electron microscope and can be stored in numerical image form [18, 19]. These images can be used to train machine learning and deep learning algorithms for later automated detection of tumor cells and reduce workload of specialized doctors [20].

In the AI field, the focus on proposing increasingly efficient techniques has led to the development of complex machine learning and deep learning models that are generally incomprehensible, even to AI experts. These models are often considered black boxes that are difficult to interpret for the majority of users. This lack of transparency leads to grow-ing concerns about their deployment in critical fields such as medicine, where doctors need to clearly understand the decision-making processes of these algorithms [21]. Consequently, despite significant efforts by researchers to create promising AI algorithms for enhancing medical decisions, these advances remain theoretical proposals in scientific literature, with very limited real-world application in clinical settings. To address this challenge, the first necessary step is the integration of these AI techniques into physicians' actual clinical routines [22, 23]. These AI-integrated solutions must be as transparent as possible by taking advantage of advancements in the field of Explainable AI (XAI) to generate human-understandable explanations for the decision-making processes of ML and DL models [24, 25].

Therefore, our objective in this work is to integrate modern AI techniques into the clinical setting through a deep learning web-based application for medical decision-making with a focus on colon cancer detection. We turned to web technology for its simple accessibility from any internet-connected device (smartphones and computers), which provides an immediately deployable solution for real clinical practice. This web application serves a dual purpose. Primarily, it enables efficient instantaneous interaction between medical professionals for discussing clinical cases encountered in their clinical setting, in order to reach the most relevant medical decision. Furthermore, this application also provides an effective deep learning-based tool for colon cancer detection. This smart tool adopts a Vision Transformer deep learning model (ViT), which is the transformer model version adapted for computer vision [26].

Transformer models were first proposed in the field of natural language processing (NLP), where they have achieved great success thanks to their high performance based on the attention mechanism [26, 27]. Compared to recurrent neural networks (RNNs), which process data sequentially and convolutional neural networks (CNNs), which employ local filters, the attention mechanism enables the model to simultaneously analyze the relationships between all elements in a sequence [27, 28, 29]. This strategy enhances the ability to capture long-distance dependencies and produces rich contextual representations [30]. Models such as BERT and OpenAI's GPT-2 and GPT-3, which rely on this approach, have revolutionized the field of natural language processing [31].

To ensure high usability of the proposed application in clinical settings, an Explain-able AI (XAI) solution was developed for our ViT model to provide clear and hu-man-understandable explanations of its medical decision-making. The XAI approach adopted in this work is Local Interpretable Model-Agnostic Explanations (LIME), which offers a visual explanation of the complex model's predictions by highlighting regions that strongly influence the prediction results. This is achieved by masking regions of the image and observing the corresponding impact on the model's output to identify the most critical areas for the decision [32].

The ViT model integrated into our application was rigorously evaluated to demonstrate its robustness. The experimental results demonstrate that our model is characterized by its high predictive accuracy, even on external unseen data. Furthermore, a comparative analysis with recently proposed deep learning models highlights the competitive performance of our model. These results confirm its strong generalization capability and high reliability to support medical decision-making in clinical settings. The primary contributions of this study are as follows :

- Design and development of an interactive web application that enables physicians to share clinical cases and conduct teamwork discussion-based medical decision-making.
- Integrate an efficient Vision Transformer (ViT) model for colon cancer detection.
- Incorporate the XAI method LIME in order to provide physicians with visual and interpretable explanations of ViT model predictions.
- Conduct rigorous validation of the model based on a merged dataset (LC25000 + DigestPath) and an unseen external dataset to promote the reliability of predictions in real-world clinical settings.
- Present an innovative combination enabling physicians' expertise sharing and explainable ViT-based predictions to enhance medical decision-making.

The rest of this paper is organized as follows: section 2 is devoted to related work. Section 3 presents the proposed architecture of the intelligent web application. Sections 4 and 5 explain in detail the adopted ViT deep learning model and the XAI LIME method, respectively. Section 6 describes the developed application. Section 7 presents and dis-cusses the prediction performance of the ViT model. Section 8 provides the conclusion of this work.

2. Related Work

In this section, we highlight recent work interested in the development of intelligent web applications for medical decision-making. We focus on the types of machine learning and deep learning models integrated into these applications, as well as their prediction performance and reliability in clinical settings. Furthermore, a summary comparison of these works is presented in Table 1, providing a concise overview of their key characteristics.

In [33], the authors propose a medical web application with mobile compatibility. This application supports medical functions such as appointment scheduling and an intelligent deep learning system for brain cancer and diabetic retinopathy detection. The work presented in this paper focuses on the application's functionalities and does not provide detailed experimental evaluation results of the deep learning models, which affects the reliability of adopting this application in real-world clinical settings.

[34] presents an improved version of the same application, with enhancements especially in the user interface intuitiveness and the number of diseases taken into account by the intelligent system. In this version, keratoconus, breast cancer, and pneumonia have been introduced for deep learning-based detection.

The web application presented in [35] offers a chatbot tool for medical diagnosis. This tool integrates two models for medical image detection: pneumonia via chest X-rays and ocular pathologies via optical coherence tomography scans. Additionally, the application includes a model capable of extracting physiological parameters for diabetes prediction through the analysis of text-based conversations between the chatbot and the user. Despite the cloud services (OpenAI, Teachable Machine) adopted by this application, which avoid expensive infrastructure, there are concerns regarding the confidentiality of patient data analyzed through these services.

In [36], a comparison between various machine learning models is conducted to determine their performance in diabetes prediction on multiple datasets. The results of this study demonstrate that the Support Vector Machine (SVM) algorithm achieved the highest accuracy, particularly on the Pima Indians Diabetes Dataset (PIDD). For this reason, SVM was deployed in the Flask-based web application that takes clinical parameters as input and returns an instant prediction (positive/negative) for diabetes diagnosis.

In [37], the authors have conducted a comparative study between several machine learning models to determine the best performing one for stroke prediction. Compared to K Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machine (SVM), Naive Bayes (NB), and eXtreme Gradient Boosting (XGB), this study found that Random Forest is the most efficient model, with an accuracy of 90.36%. Therefore, Random Forest was selected for integration into the web application for stroke prediction. This application has also employed Explainable AI (XAI) techniques, such as SHAP and LIME, which provide doctors with a clear explanation of the prediction process.

The smart web application proposed in [38] is intended for prostate cancer detection using the EfficientNet-B1 deep learning model. Based on input biopsy images, the output result of this model presents the severity grade of this type of cancer, where Grade 1 corresponds to a less aggressive cancer and Grade 5 to a highly dangerous form. This study focused more on the usability evaluation of the application rather than the efficiency of the deep learning model, which is important for the reliability of this application in real clinical settings.

[39] presents a web application for the prediction of COVID-19. Based on the COVID-19 symptoms and presence dataset, different machine learning models were evaluated to determine the most relevant one to be integrated into the application. The dataset used includes 20 binary symptom inputs (0 = symptom absent, 1 = symptom pre-sent) and 1 binary result output (COVID-19 = presence/absence). This study showed that RF, SVM, KNN, and ANN achieved the best prediction results with an equal accuracy of 98.84%. The source code of the application, available in the GitHub repository, demonstrates that SVM is the model integrated into the web application.

The web application presented in [40] is based on a sequential convolutional neural network (Seq-CNN) proposed for skin cancer detection. The model of this application takes dermatoscopic images as input and returns one of the seven types of skin lesions de-fined in the HAM10000 dataset. The experimental study demonstrated that this application achieves a detection accuracy of 96.25%. Furthermore, the application was evaluated through a user satisfaction survey conducted with 150 participants. The results showed that 76% rated their overall satisfaction as 5 (Very Satisfied).

In [41], the authors proposed a simple intelligent web application for heart disease prediction. To obtain an effective web application in a real clinical setting, a comparison between several machine learning models was conducted to determine the most suitable one for heart disease prediction. This study was performed on two numerical datasets (CHSLB and Cleveland), and KNN was identified as the most effective model, achieving an accuracy of 100% on the CHSLB dataset and 97.83% on the Cleveland dataset. These results demonstrate that KNN is the most appropriate model for this smart web application.

Another web application was developed for the prediction of heart disease in [42]. An analysis of several datasets in this context was carried out to identify the 13 most significant features for heart disease prediction from 75 initial features. Compared to other intelligent web applications that conduct comparative studies of several ML algorithms to determine the most appropriate one, this application integrates multiple ML models and simultaneously presents the prediction results of all the considered models.

The application presented in [43] focuses on the diagnosis of Kawasaki disease. To achieve this objective, a dataset of 3,650 samples was analyzed, with the goal of reducing the number of features from 43 to 10 by selecting the most important ones using Shapley Additive Explanations (SHAP). A comparative study among 10 machine learning algorithms was carried out to determine the most relevant one to be integrated into the pro-posed application. This analysis revealed that XGBoost is the best-performing model, with an accuracy of 93.81%.

In [44], a smart web application based on Azure Machine Learning Studio was developed for the early prediction of diabetes. Among several tested machine learning models (logistic regression, neural networks, SVM, etc.),

Decision Forest (DF), which achieved an accuracy of 81.2%, was selected for its excellent performance to be integrated into the smart web tool. The performance evaluation of the considered machine learning (ML) models is based on the Pima Indians Diabetes Dataset (PIDD), which includes 768 samples with 8 inputs and a binary output result, allowing prediction of whether the patient is diabetic or not.

Table 1. Summary of intelligent web applications for medical decision-making

Reference	Disease(s)	Model(s)	XAI Method	Model accuracy
(Mustapha et al., 2023) [33]	Brain Cancer, Diabetic Retinopathy	CNN	None	Not mentioned
(Driss et al., 2024) [34]	Keratoconus, Breast Cancer, Pneumonia	CNN	None	Keratoconus: 98.74%, Breast Cancer: 86.25%, Pneumonia: 96.02%
(Pires, 2024) [35]	Pneumonia, Ocular Diseases, Diabetes	CNN + OpenAI LLMs (GPT) (for routing & NLP); Shallow NN + OpenAI LLMs (GPT) (for routing & NLP)	None	Not mentioned
(Ahmed et al., 2021) [36]	Diabetes	SVM	None	78.125%
(Mridha et al., 2023) [37]	Stroke	Random Forest	SHAP, LIME	90.36%
(Singh et al., 2023) [38]	Prostate Cancer	CNN (EfficientNet-B1)	None	Accuracy not mentioned; Quadratic weighted kappa: 0.862 (95% CI: 0.840–0.884)
(Villavicencio et al., 2022) [39]	COVID-19	SVM	None	98.84%
(Siddique et al., 2024) [40]	Skin Cancer	Sequential CNN (Seq-CNN)	None	96.25%
(Absar et al., 2022) [41]	Coronary Heart Disease (CHD)	KNN	None	CHSLB dataset: 100%; Cleveland dataset: 97.83%
(Rahman, 2022) [42]	Heart Disease	KNN, XgBoost, LR, SVM, AdaBoost, DT, Naive Bayes, RF	None	DT: 99%, RF: 99%, XgBoost: 95%, KNN: 89%, SVM: 85%, LR: 85%, AdaBoost: 83%, Naive Bayes: 82%
(Duan et al., 2025) [43]	Kawasaki Disease (KD)	XGBoost	SHAP	93.81%
(Raheem & Al-Qurabat, 2022) [44]	Diabetes	Decision Forest	None	81.2%

3. Proposed Architecture of the Explainable Vision Transformer-Based Web Application

As presented in Figure 1, the architecture of the proposed system aims to enhance medical decision-making by allowing doctors to support their diagnostic decisions either by sharing medical cases with other experienced

doctors or by using the deep learning system for medical image-based diagnosis. Therefore, the proposed web application can be divided into two main phases: patient data collection phase and medical decision-making phase.

3.1. Patient Data Collection Phase

The first phase is devoted to patient data gathering, where the doctor can add patients and generate their personal healthcare records. This data includes the patient's personal information, the reason for the clinical consultation as described by the doctor, as well as complementary data. The collected data can later be used to create new medical datasets or to enrich existing ones.

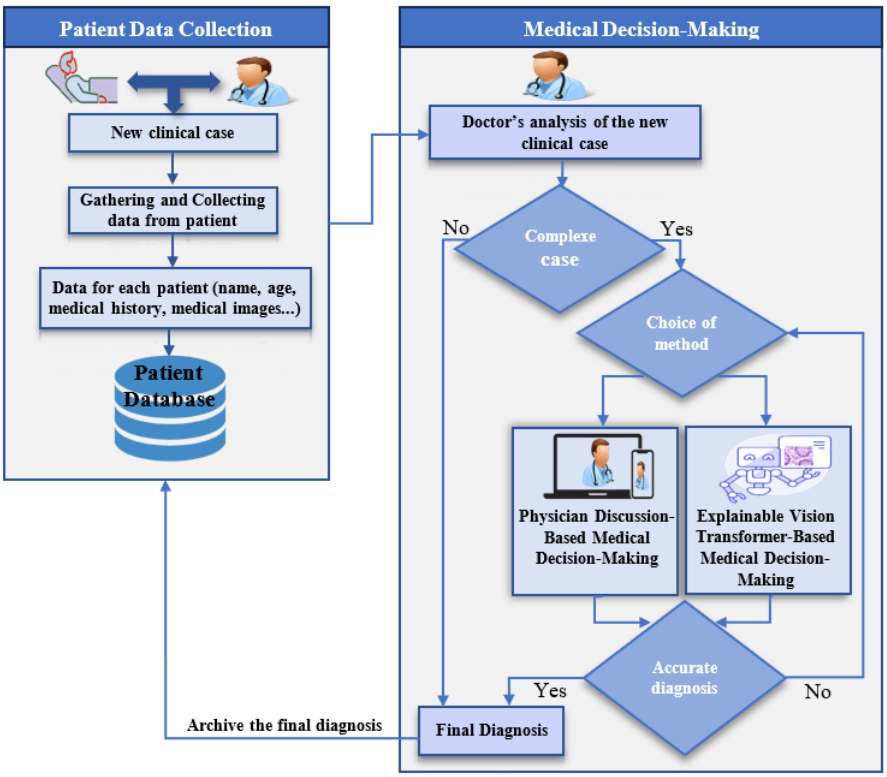

Figure 1. The proposed architecture of the web application

3.2. Medical Decision-Making Phase

The second phase is intended for medical decision-making. In simple cases, the doc-tor can directly provide the final decision. For more complex situations, they may employ one or both of the following methods to reach an effective decision.

3.2.1. Explainable Vision Transformer-Based Medical Decision-Making: This medical decision-making method involves the use of the proposed explainable vision transformer system, which is focused on colon cancer detection. This system employs a ViT deep learning model for colon cancer detection, along with a LIME explainable model that aims to promote the doctor's confidence in the system's results by providing clear explanations of how it achieves its predictions. A detailed description of the ViT deep learning model and the LIME XAI method, which covers their architectures and components, is provided in the next sections

3.2.2. Physicians Discussion-Based Medical Decision-Making: The physician who encounters a complex or ambiguous clinical case can share it with selected colleagues. The physician initiating the case is considered the team leader, responsible for coordinating the discussion and guiding the team's work. Each invited physician has access to the patient's medical records and can provide observations, comments, and diagnostic ideas.
During the group discussion, physicians are strongly encouraged to engage in a structured exchange of ideas and interpretations. In this collaborative process, based on the available information and their medical expertise, the participating physicians collectively propose a list of possible diagnoses. This step is considered a brainstorming session, allowing the team to generate a limited set of potential solutions that will later be analyzed [45]. The early formulation of this list helps structure the clinical problem by focusing on a restricted number of plausible hypotheses. This strategy makes medical decision-making more efficient and reduces physicians' memory overload [46]. Once all possible diagnoses have been generated, the team carefully discusses and evaluates each of them. Through this process, the less likely hypotheses are progressively removed, leading the team toward the most relevant final diagnosis.
This structured, discussion-based approach is recognized as a best practice for effective teamwork and collective medical decision-making. By promoting knowledge sharing and consensus building among physicians, this method enhances diagnostic accuracy, supports interprofessional collaboration, and reduces the likelihood of individual bias or diagnostic error. Such an approach is widely acknowledged in modern healthcare environments as a key factor for improving patient safety and decision-making quality [47].
When there is disagreement between the AI system and the physicians' decision, the application does not override human expertise. Otherwise, the XAI method provides clear explanations of the model's predictions, which helps the medical team understand the model's thinking and decide whether to revise or confirm their diagnosis. In this way, the AI serves as a helpful tool to support, not replace, physicians' decision-making.

4. Fundamentals of Vision Transformers

Vision transformers are a type of deep learning model designed for the computer vision process. They were inspired by the success of transformer models used in NLP by applying a self-attention mechanism to understand the relationships between different parts of an image. Compared to CNNs, which hierarchically analyze the image to extract low-level features such as edges and textures, and then high-level structures such as objects, ViTs can directly capture the global context by establishing semantic links between all regions of the image from the beginning. As presented in figure 2, vision transformers split the image into small square regions, called patches, which correspond to the words of a sentence in NLP transformers. These 2D patches are flattened into linear embedding vectors that are processed by the transformer encoder to analyze the global dependencies within the image. The output of the transformer encoder is passed into small multi-layer perceptron for the final prediction or classification [26, 30].
Formally, each input image $\chi \in \mathbb{R}^{H\times W\times C}$ is divided into N patches $x_p \in \mathbb{R}^{(P\times P\times C)}$
Where :
H, *W*, and *C* represent the height, width, and number of channels of the input image, respectively.

The patch size is $p \times p$, and the number of patches is: $N = \frac{H\times W}{P^2}$. These patches are then flattened into vectors of size $P^2 \cdot C$, and then passed through a linear embedding layer which transforms them into new vectors of size D.

A special classification token [CLS] is added to the beginning of the sequence of patches, and will be used for the final prediction. Positional embeddings are added to the vectors to preserve the position information of the patches in the image. Therefore, the input sequence of the transformer encoder is: $z_0 \in \mathbb{R}^{(N+1)\times D}$ and will be processed by a series of L layers.

Each transformer layer begins with a normalization operation:

$$z'_{l-1} = \text{Norm}(z_{l-1}) \tag{1}$$

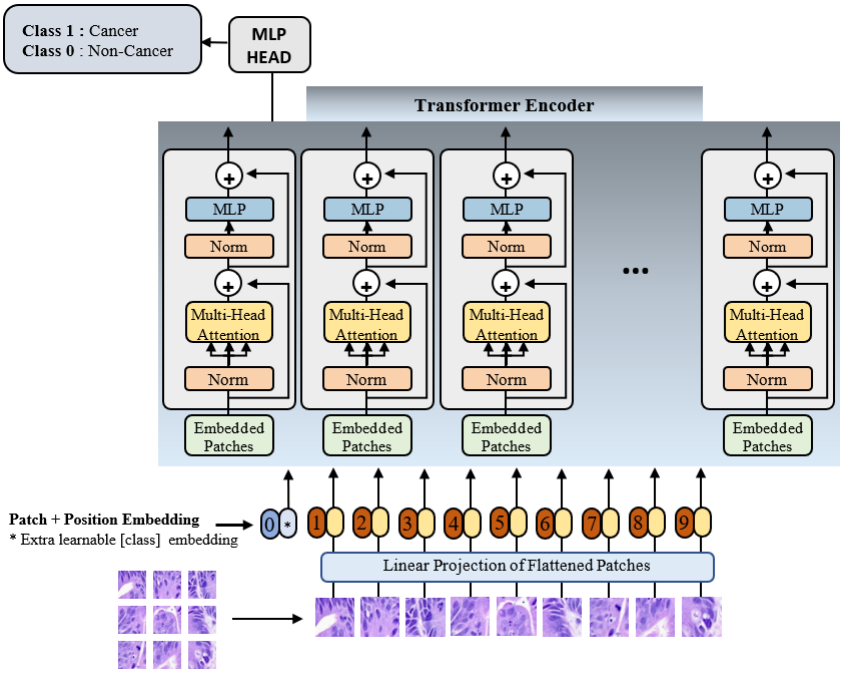

Figure 2. The architecture of vision transformer based on [26]

This normalization is followed by a Multi-Head Self-Attention (MSA) mechanism that allows each token to interact with all other tokens in the sequence. For each attention head i among the h heads:

$$\text{head}_i = \text{Attention}(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \tag{2}$$

where:
$Q_i = z'_{l-1} W_i^Q$, $K_i = z'_{l-1} W_i^K$, $V_i = z'_{l-1} W_i^V$
$W_i^Q \in \mathbb{R}^{D \times d_k}$, $W_i^K \in \mathbb{R}^{D \times d_k}$, $W_i^V \in \mathbb{R}^{D \times d_v}$
$Q_i \in \mathbb{R}^{N \times d_k}$, $K_i \in \mathbb{R}^{N \times d_k}$, $V_i \in \mathbb{R}^{N \times d_v}$
$d_k = d_v = \frac{D}{h}$
The output of the MSA is obtained by concatenating the h heads:

$$\text{MSA}(z'_{l-1}) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h) W^O \tag{3}$$

where:
$W^O \in \mathbb{R}^{h d_v \times D}$
This output is combined with the original via a residual connection:

$$z''_l = \text{MSA}(z'_{l-1}) + z_{l-1} \tag{4}$$

The vector z''_l is then passed through another normalization operation, the result of which serves as input to the Multi-Layer Perceptron (MLP) block. This block is composed of two fully connected layers with a GELU activation function. Similar to the MSA block, the MLP block also employs a residual connection, which leads to the following expression:

$$z_l = \text{MLP}(\text{Norm}(z''_l)) + z''_l \tag{5}$$

At the output of the last block, we obtain the final representation: $Z_L \in \mathbb{R}^{(N+1)\times D}$ The [CLS] token, located at the first position z_L^0, is then used for the final prediction:

$$y = \text{MLP}_{\text{Predict}}(z_L^0) \tag{6}$$

5. The Local Interpretable Model-Agnostic Explanations (LIME) Method

XAI methods can be categorized according to four main criteria. The first depends on the agnosticism of the model, which allows the distinction between *model-agnostic methods* applicable to all DL and ML models, and *model-specific methods* designed for a particular type of model. The second criterion corresponds to the scope, which allows the distinction between *local scope methods* that provide an explanation for a particular prediction, and *global scope methods* that explain the general behavior of the model. The third criterion concerns the type of data, where some approaches are adapted to images, others to textual, tabular, or even graphical data. The last criterion depends on the type of explanation: some methods provide a visual representation of the relationships between dataset variables, others are based on the identification of the most influential features in the output results, and still others rely on simplified models (called surrogates) that provide a local approximation of complex models [21].
Local Interpretable Model-Agnostic Explanations (LIME), as the name suggests, is a method for explaining the predictions of any machine learning (ML) or deep learning (DL) model, regardless of the data type (images, text, tabular, graphs). It relies on generating a surrogate model that approximates the complex model around a particular prediction in order to produce a human-understandable explanation of the results. The main idea behind this approach is to generate perturbed input data around a specific instance. These generated data are then used to train a simple local surrogate model to provide an interpretable explanation of the prediction process [32, 48].
Formally, Lime can be considered as a mathematical optimization problem. The objective is to find a simple model $g \in G$ that maximizes the fidelity to the complex model f in the neighborhood while keeping the interpretability as simple as possible.

$$\varepsilon(x) = \arg\min_{g \in G} \; L(f, g, \pi_x) + \Omega(g)$$

where:
$x \in \mathbb{R}^{dim}$ is the predicted instance that we want to explain,
$f : \mathbb{R}^{dim} \to \mathbb{R}$ is the complex model being explained,
g is the surrogate model used to approximate f in the vicinity of x,
G is the family of possible interpretable models,
π_x is the local neighborhood of x,
$\Omega(g)$ is the complexity measure of the surrogate model g.

To explain the local behavior of the complex model f, the LIME method generates a set of samples by randomly modifying the features of the instance to be explained x. These perturbations allow to create a new dataset Z of samples reflecting the neighborhood of x. These new samples are then weighted by the function $\pi_x(z_i),\ z_i \in Z$, which favors the samples closest to x:

$$\pi_x(z) = \exp\left(-\frac{D(x,z)^2}{\sigma^2}\right) \tag{8}$$

Then, LIME selects a surrogate model g from the class G by minimizing the following weighted linear regression objective:

$$L(f, g, \pi_x) = \sum_{z, z' \in Z} \pi_x(z)\big(f(z) - g(z')\big)^2 \tag{9}$$

where:
z is a perturbed version of x, used to explore the behavior of model f

z' is a simplified version (binary representation) of z, intended for model g
$D(x, z)$ is a distance function
σ is a parameter that controls the size of the neighborhood around x.

6. The Developed Explainable Vision Transformer-Based Web Application

The proposed application was developed using modern web technologies. The front-end interface employs ReactJS, which is considered a widely utilized open-source JavaScript library for building user interfaces [49]. The back-end development is based on the Django framework (Python), which follows a Model-View-Template (MVT) architecture [50]. Django's strengths are mainly due to its Object-Relational Mapper (ORM), which enables efficient manipulation of relational databases [51]. In addition, the Django REST Framework (DRF) provides powerful tools for building robust and scalable APIs that facilitate data exchange between front-end and back-end systems [52].
The Vision Transformer (ViT) model was developed and trained separately using deep learning libraries such as TensorFlow and Keras. After training, the model was serialized as a .pkl file and integrated into the web application. The developed LIME explainer function then uses this pre-trained model to generate human-understandable explanations for ViT model predictions. On the server side, this function produces a visual explanation by identifying the pixel regions that most significantly contribute to the prediction.

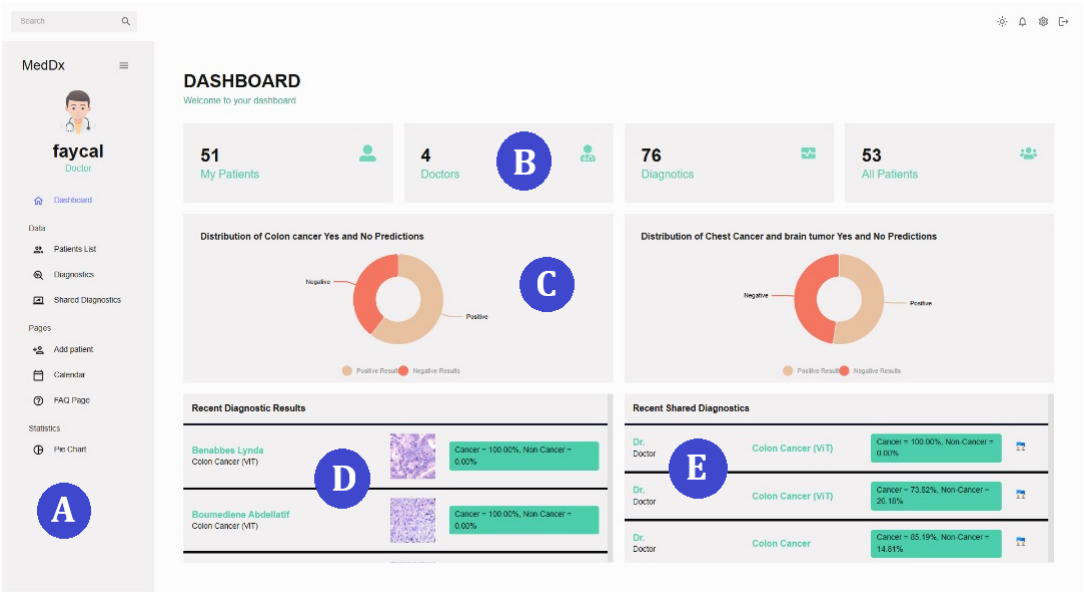

Figure 3. The doctor's personal dashboard

As shown in Figure 3, the physician's personal dashboard is accessible via standard username and password authentication. This interface provides an overview of the application's main data and features. The sidebar (Figure 3-A) offers navigation links to key application components. Statistical cards (Figure 3-B) display general metrics, including the number of physicians, patients, and clinical cases handled by the concerned doctor. A pie chart (Figure 3-C) visually represents the distribution of diagnostic predictions, including colon cancer and other diseases under study, such as brain cancer. Finally, the dashboard lists the most recent clinical cases (Figure 3-D) and the latest cases shared with other physicians (Figure 3-E).

To initiate the decision-making process for a new clinical case, the doctor must first create the medical record of the concerned patient. This record includes the patient's personal information, medical history, and complementary data. Such data may include medical images (e.g., a histopathological image of colon tissue) or textual PDF

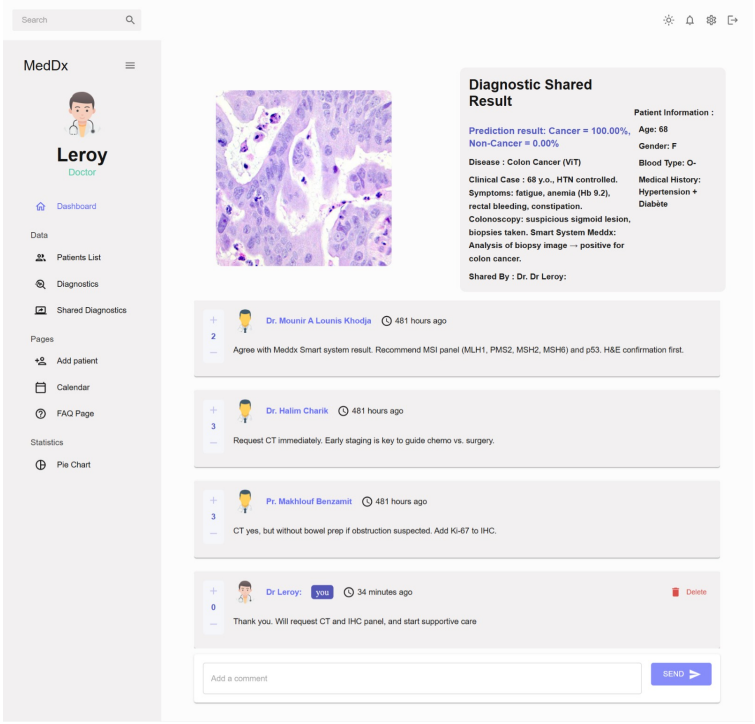

Figure 4. Shared clinical case for group decision-making

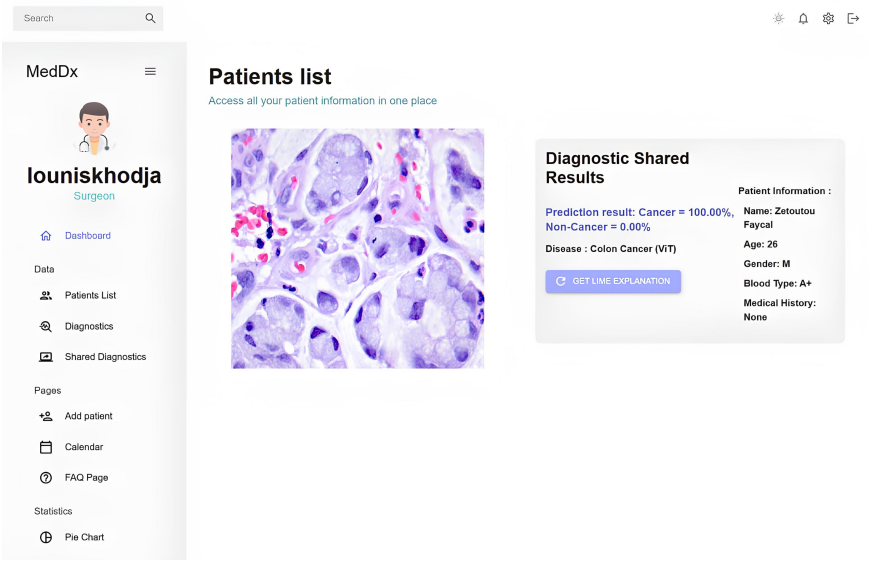

Figure 5. Smart tool for colon cancer case prediction

documents. Once the new patient is added to the patient list, the doctor can share the clinical case (Figure 4) with other selected colleagues to achieve effective collective decision-making. The involved doctors can then discuss their different opinions and comments on the clinical case. A group voting tool enables them to collectively determine the most appropriate decision based on shared knowledge and complementary expertise.

The doctor can also use the smart medical decision tool for colon cancer prediction. In this option, the doctor must

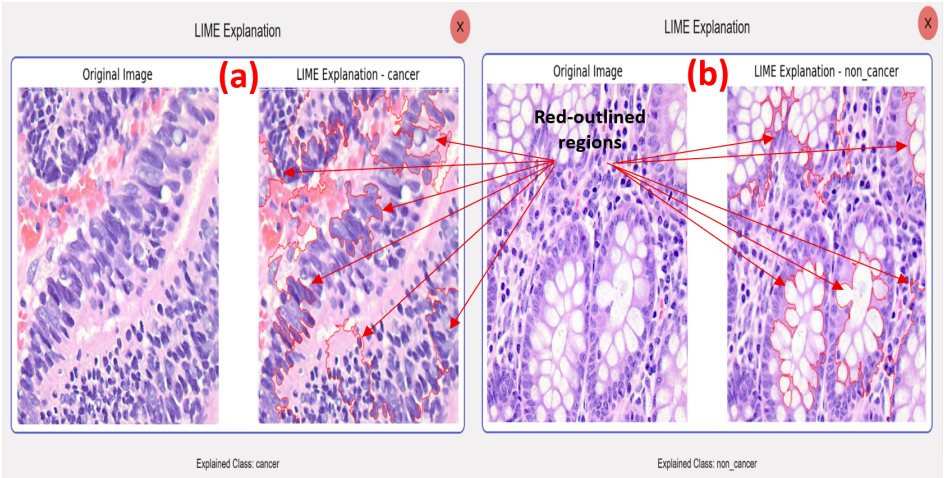

Figure 6. Prediction explanation using the integrated LIME method for (a) colon cancer and (b) non-cancerous tissue

upload a histopathological image of colon tissue, which will be processed by the integrated ViT model. The tool then displays the prediction result (Figure 5) and offers the doctor the possibility of obtaining an explanation of the prediction based on the integrated LIME method. In this case (Figure 6), the original histopathological image is displayed along with an annotated image with red-outlined regions highlighting the pixel areas that significantly contribute to the prediction. This solution provides the doctor with transparency regarding the prediction process of the vision transformer (ViT) model. Such visibility strengthens doctors' confidence in the medical decisions proposed by our smart system and promotes its integration into routine clinical practice.

7. Experimental Results

This section presents a comprehensive evaluation of our web application, covering both user acceptance and technical performance. The evaluation begins with a technology acceptance analysis based on the TAM model, followed by rigorous validation of the ViT model's performance and a quantitative analysis of LIME explanations.

7.1. Technology Acceptance and Usability Assessment:

To assess the potential adoption and success of the proposed environment, this study was based on the Technology Acceptance Model (TAM) framework, originally developed by Davis in 1989 [53]. A survey instrument was adapted from this model to measure user acceptance. Our target population consisted of 50 students from the faculty of medicine. This group was selected as it represents future medical experts who are primary candidates to benefit effectively from smart Information and Communication Technology (ICT) applications in the medical field.
The questionnaire was designed with 20 items, categorized into four key TAM constructs:

- Perceived Ease of Use (PEOU): The degree to which a person believes that using the system would be free of effort.
- Perceived Usefulness (PU): The degree to which a person believes that using the system would enhance their job performance.
- Attitude Toward Using (ATU): The individual's positive or negative feelings about using the system.
- Behavioral Intention to Use (BI): The degree to which a person has formulated conscious plans to use or not use the system.

All items were rated on a 5-point Likert scale, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Example items include: "Overall, I find this app easy to use" (PEOU), "I find this application useful in daily clinical practice" (PU), "Using this application seems beneficial to me for practice" (ATU), and "I plan to use this app regularly for medical diagnosis in the future" (BI). The internal consistency and reliability of the items for each subscale were assessed using Cronbach's alpha. The analysis demonstrated strong reliability, with scores exceeding the minimum acceptable threshold of 0.7 [54, 55]: PEOU ($\alpha = 0.766$), PU ($\alpha = 0.836$), ATU ($\alpha = 0.851$), and BI ($\alpha = 0.788$). This confirms the questionnaire's robustness as a measurement tool. The comprehensive results of the user perception evaluation are detailed in Table 2.

Table 2. Perceptions of the smart web application (5-point Likert scale: 1 represented "strongly disagree" and 5 represented "strongly agree") (n = 50).

Construct	Avg	Std	Cronbach's alpha
Perceived ease of use	4.2067	0.6774	0.766
Perceived usefulness	4.1600	0.3672	0.836
Attitude toward using	3.9240	0.5052	0.851
Behavioral intention to use	4.3000	0.4881	0.788

As shown in Table 2, participants demonstrated a positive acceptance of the smart web application. Analysis of the four TAM subscales revealed high scores on the Likert scale, particularly for Behavioral Intention (Avg = 4.30) and Perceived Ease of Use (Avg = 4.2067). The internal consistency of the subscales was confirmed by Cronbach's alpha coefficients, which ranged from 0.77 to 0.85. All values exceeded the recommended threshold of 0.70, thus validating the questionnaire's reliability. The high score for Behavioral Intention suggests a high probability of the smart application's adoption. The strong scores for Perceived Ease of Use (Avg = 4.21) and Perceived Usefulness (Avg = 4.16) demonstrate that the application meets essential criteria for usability. The moderate score for Attitude (Avg = 3.92) may reflect an adaptation period required for its adoption. Although the present evaluation focused on assessing user acceptance through the Technology Acceptance Model (TAM), a future usability study is planned to complement these findings. This study will specifically involve practicing physicians to analyze how they interact with both the AI diagnostic tool and the collaborative discussion module. The objective will be to evaluate the impact of the system on diagnostic workflow, team communication efficiency, and decision-making time. Quantitative indicators such as task completion time, number of discussion iterations, and consensus rate will be combined with qualitative feedback collected through standardized usability instruments such as the System Usability Scale (SUS) [56] and the NASA Task Load Index (NASA-TLX) [57]. The outcomes of this evaluation will guide future optimization of the the smart web application's design and improve its integration into real clinical workflows.

7.2. ViT Model Performance Evaluation

This study adopts the LC25000 dataset [58], which consists of 25,000 lung tissue and colon tissue images with a resolution of 768 × 768 pixels, presented in the RGB color space. These images provide an equal distribution across five classes, serving as indicators of the presence or absence of colon and lung cancer, with 5000 images per class. There are three classes concerning lung cancer : Lung adenocarcinomas, Lung squamous cell carcinomas, and Benign lung tissues and two concerning colon cancer: Colon adenocarcinomas, Benign colonic tissues. Our work focuses on 10 000 colon tissue images, which are predivided into 9000 images for training and validation and 1000 images for testing.

Despite the large use of the LC25000 dataset in recent research on lung and colon cancer, it's important to note that it is primarily composed of a small base set of only 1,250 images. These images are equitably distributed across the five classes, with 250 original images for each one. The final size of this dataset (25,000 images) is achieved through extensive data augmentation using rotations and flips. This fundamental characteristic can affect the ability of models trained on this dataset to generalize their performance in real-world clinical settings. This limitation is particularly critical for heavier and more complex architectures, such as the vision transformer (ViT), which are

at high risk of overfitting when trained on the LC25000 colon cancer images. To overcome this challenge, we expanded the dataset by adding 10,000 images from the DigestPath dataset [59], with 5,000 additional images for each class (colon adenocarcinomas, benign colonic tissues). To ensure an in-depth evaluation of the generalization capability of the ViT model, a further test on unseen data was conducted using a subset of 1,200 images (600 per class) from a third independent dataset, EBHI-Seg [60]. Therefore, the data distribution across the training, validation, and test sets is detailed in Table 3.

Table 3. Distribution of images across training, validation, and test sets

Dataset	Total images used	Training		Validation		Test	
		Size	Percentage	Size	Percentage	Size	Percentage
LC25000	10000	7200	72%	1800	18%	1000	10%
DigestPath	10000	8000	80%	1000	10%	1000	10%
Total (Development Set)	20000	15200	76%	2800	14%	2000	10%
EBHI-Seg (External Test)	1200	/	/	/	/	1200	100%

The ViT model performance is evaluated based on the most common metrics in deep learning (DL) and machine learning (ML). Therefore, we use accuracy, which evaluates the model's overall performance in correctly predicting the class of all images. Precision defines the proportion of colon cancer cases correctly identified among all cases predicted as colon cancer. Recall defines the proportion of colon cancer cases correctly identified among all true cases of colon cancer. The F1-score provides a balance between precision and recall by calculating their harmonic mean. The AUC-ROC (Area Under the ROC Curve) evaluates the model's ability to distinguish between colon cancer cases and non-colon cancer cases.

Figure 7 shows the evolution of the accuracy curves during the 20 training epochs. We observe that the training accuracy and the validation accuracy increase consistently to very high values, close to 1.0. This demonstrates that the model has an efficient learning capacity. The small and stable gap between these curves, without observable divergence, indicates the absence of overfitting. This performance demonstrates that our model has a strong generalization capacity on unseen data.

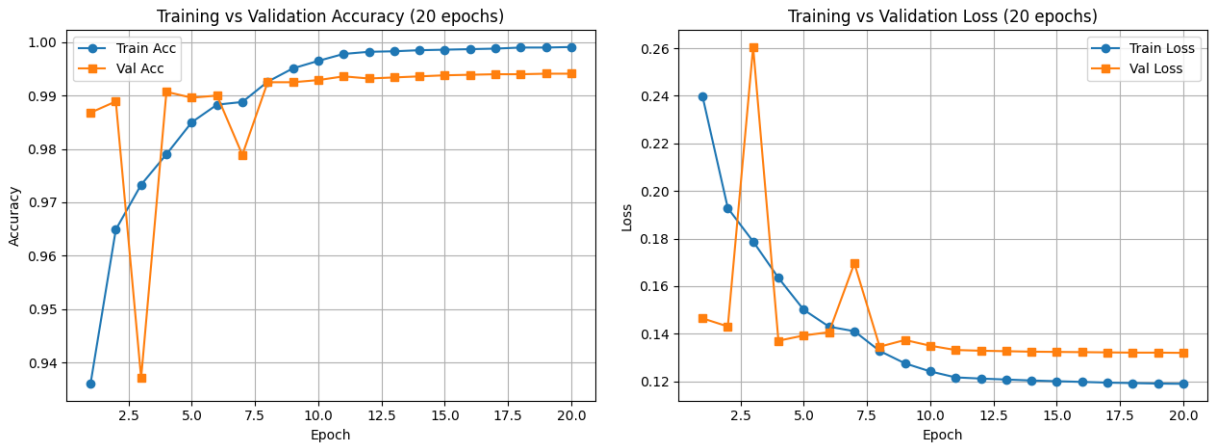

Figure 7. Model performance: accuracy and loss curves

Moreover, Figure 7 illustrates the evolution of the training and validation loss curves. These curves demonstrate a rapid reduction in the loss function value, which stabilizes at approximately 0.12 after only a few epochs. The parallel convergence of the two curves towards very low values confirms the effectiveness of the training process and the ability of the model to make predictions with a high level of confidence and low error.

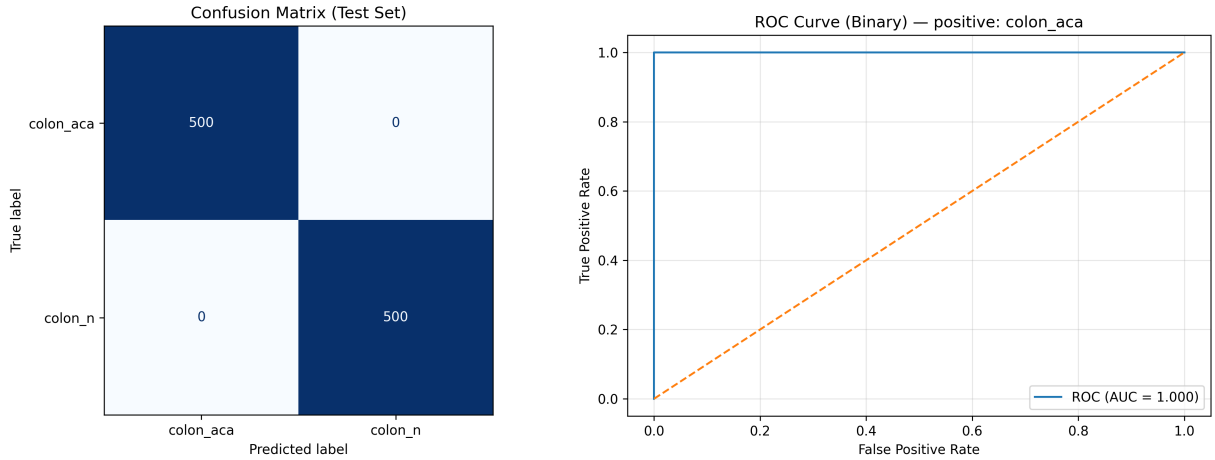

Figure 8. Confusion matrix and ROC curve for the LC25000 test set

On the LC25000 dataset, the confusion matrix (Figure 8) shows that the model correctly classified all 500 cases of colon adenocarcinoma and all 500 cases of normal tissue. The ROC curve (Figure 8) passes through the upper left corner, with an area under the curve (AUC) of 1.00, indicating perfect performance and ideal separability between the two classes. This means the model achieved a perfect true positive rate and true negative rate, with no false positives or false negatives on this specific dataset.
On the merged LC25000 and DigesPath dataset (Figure 9), the model maintains a high AUC of 0.988. The corresponding confusion matrix (Figure 9) shows high performance with 942 true positives and 956 true negatives from 2000 test images, reflecting excellent prediction accuracy.

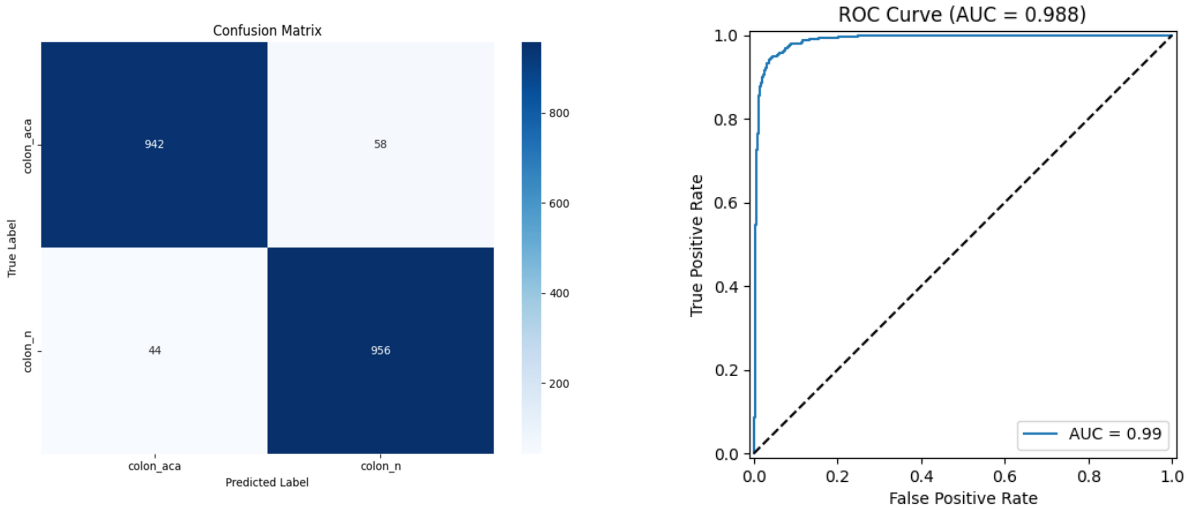

Figure 9. Confusion matrix and ROC curve for the merged LC25000 and DigestPath test set

The most critical evaluation for the ViT model generalization is performed on external dataset. As presented in (Figure 10), the test on the EBHI-Seg dataset achieved an AUC of 0.976. The confusion matrix for this test (Figure 10) shows 552 true positives and 554 true negatives from the 1200 images used, which demonstrates strong performance on this unseen external dataset.

Based on the employed metrics, Table 4 demonstrates that the ViT model achieved perfect performance across accuracy, precision, recall, and F1-score on the LC25000 dataset. On the more challenging merged dataset, the model also demonstrated excellent and balanced performance, achieving 94.90% across all metrics, which indicates reliable prediction without any preference toward a specific class. Furthermore, the EBHI-Seg dataset test results confirm the strong generalization capability of the ViT model, with an accuracy of 92.17%. This represents a key indicator of the model's reliability for real-world clinical settings.

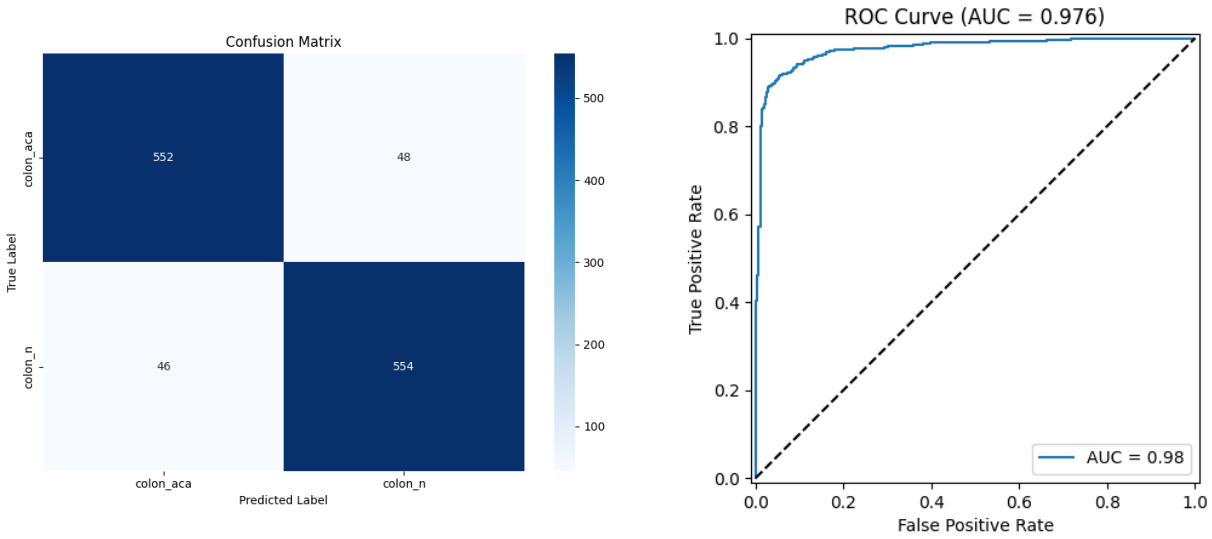

Figure 10. Confusion matrix and ROC curve for the external EBHI-Seg test set

Table 4. Performance metrics of the ViT model on the LC25000, merged, and external test datasets

Dataset	Accuracy	Precision	Recall	F1-Score
LC25000	100%	100%	100%	100%
LC25000 + DigestPath	94.90%	94.91%	94.90%	94.90%
EBHI-Seg (External Test)	92.17%	92.17%	92.17%	92.17%

The comparative analysis presented in Table 5 demonstrates that although the proposed ViT model has a larger architectural size (327.3 MB), it achieves an effective balance of speed, accuracy, and generalization that is essential for clinical practice. Its inference time of 0.016 seconds per image remains highly competitive, being only a little slower than SWIN transformer (0.011 s) and comparable to CNN-based models (0.013 s), which enables rapid diagnostic outcomes. This practical efficiency is further supported by its superior predictive robustness, as confirmed by McNemar's test. While all models, including our ViT, achieved perfect accuracy on the LC25000 dataset with no significant differences ($p = 1.0$), their performance varied considerably on the unseen EBHI-Seg dataset. In this challenging scenario, our model achieved a high accuracy of 92.17%, clearly outperforming SWIN transformer (83.30%), CNN (50.30%), and MobileNetV2 (53%). McNemar's test verified that these performance differences are statistically significant ($p = 0.000$), confirming the model's strong potential for effective integration into real-world clinical workflows.

Table 5. Performance comparison of ViT versus SWIN, CNN, and MobileNetV2 on LC25000 and EBHI-Seg datasets using McNemar's test ($p < 0.05$).

Model	Size (MB)	Inference time (sec/img)	LC25000			EBHI-Seg		
			Accuracy	P value	Significance ($p < 0.05$)	Accuracy	P value	Significance ($p < 0.05$)
Proposed ViT	327.3	0.016	100%	-	-	92.17%	-	-
SWIN	105	0.011	99.7%	1.0	No	83.30%	0.000	Yes
CNN	8.5	0.013	100%	1.0	No	50.30%	0.000	Yes
MobilNetV2	8.5	0.013	100%	1.0	No	53%	0.000	Yes

To quantitatively evaluate the reliability of the LIME explanations, we performed a feature removal test on 35 randomly selected histopathological images. On average, the highlighted regions covered 31.56% of the image area, while the mean confidence drop reached 53.89% after these regions were masked. This substantial decline in model confidence demonstrates that the areas identified by LIME correspond to features that the Vision Transformer heavily relies on for classification. Moreover, the moderate proportion of the highlighted area indicates that the explanations remain spatially concentrated rather than dispersed, providing interpretable and localized visual insights. Overall, these findings confirm the relevance and consistency of LIME explanations in identifying diagnostically meaningful regions in colon cancer histopathology images.

In this study, we adopted LIME as the primary explainability method due to its model-agnostic design and clinician-friendly visualization through superpixel perturbations. Nonetheless, several transformer-oriented XAI approaches have recently emerged. For example, gradient-based extensions of Grad-CAM have been adapted for Vision Transformers [61], enabling class-specific attribution maps by computing gradients with respect to attention blocks or heads. Similarly, Attention Rollout proposed by Abnar and Zuidema [62] quantifies the cumulative attention flow across layers, providing a global visualization of how input patches influence the [CLS] token representation. While these methods may produce more accurate representations of the model's internal reasoning, they typically require architectural modifications and are less intuitive for clinical users. We prioritized LIME for its ease of interpretation and integration, though combining it with Grad-CAM or Attention Rollout presents a promising direction for future enhancements.

Table 6 presents the results of the comparative analysis we conducted to position our work within recent research focused on colon cancer detection. This table uses the LC25000 dataset as a common reference point to enable an equitable comparison.

Table 6. Performance metrics of deep learning models on the LC25000 dataset for colon cancer detection

Reference	Model	XAI	Accuracy	Precision	Recall	F1-Score
(Mangal et al., 2020) [63]	CNN	None	96.61%	-	-	-
(Sakr et al., 2022) [64]	CNN	None	99.50%	99%	100%	99.49%
(Bukhari et al., 2020) [65]	RESNET-50	None	93.91%	95.74%	96.77%	96.26%
(Mohamed et al., 2024) [66]	FMO-CNN	None	97.65%	93.89%	94.87%	96.76%
(Shahadat et al., 2024) [67]	CNN	None	100%	100%	100%	100%
(Opee et al., 2025) [68]	CNN	Grad-CAM, LIME	99.40%	99.00%	99.00%	99.00%
(Giammarco et al., 2024) [69]	CNN (MobileNet)	Grad-CAM	99.9%	99.9%	99.9%	99.9%
Proposed model	**ViT**	**LIME**	**100%**	**100%**	**100%**	**100%**

Compared to the CNN models presented in this study, our ViT model demonstrates that the transformer architecture, even if conceptually different from CNNs, is not only competitive but can also outperform them in colon cancer prediction. This superior performance reflects the ability of attention mechanisms, which characterize the ViT model, to capture the global context and long-distance dependencies that may be missed by traditional CNN approaches.

While the majority of models presented in these studies show good prediction performance, ranging from 96.61% to 100%, their evaluation based only on the LC25000 dataset limits the generalizability of the results and affects their reliability in clinical settings. On the other hand, our study provides a more comprehensive evaluation that

confirms strong reliability by including results on the merged dataset and, most significantly, an external test on unseen data from the EBHI-Seg dataset.
Furthermore, the models presented in [63, 64, 65, 66, 67] are considered black boxes. As indicated previously, confidence in and understanding of the decision-making process are critically important in the medical field, and the lack of transparency in these models represents a major obstacle to their clinical adoption.
The works presented in [68] and [69] use the XAI Grad-CAM method, which applies only to CNN architectures. This approach is efficient and visual but remains limited to the convolutional layers of the CNN. The integration of LIME in [68] reflects a desire to provide a second interpretable approach that offers a different form of explanation based on superpixels. Our model, on the other hand, uses only the LIME method. This choice is strategic and particularly judicious for a ViT model. In contrast to Grad-CAM, LIME is a model-agnostic method that works with any type of architecture (CNN, ViT, etc.). This flexibility is crucial for a non-convolutional model like ViT.
It should be noted that the ViT model is generally heavier and more complex than lightweight CNN architectures. However, integrating it into a web application overcomes this problem. In this case, the doctor uses the web interface of our application without taking hardware constraints into account, and the computing power is managed server-side.

7.3. Limitations and Future Directions

Despite the promising performance of our Vision Transformer model, this study has limitations related to training data diversity. The publicly available datasets used lack comprehensive demographic and technical metadata, potentially restricting the model's generalizability to broader clinical populations and settings. Additionally, the ViT architecture's computational complexity may challenge deployment on resource-constrained devices. While our web-based application effectively avoids this limitation by handling computation server-side, future work should explore lightweight transformer variants. Regarding the explainability component, while LIME successfully identifies regions that influence the model's predictions, this technical validation requires complementary evaluation by pathologists to assess the clinical relevance of the highlighted areas.
Future work will focus on three main directions: validating the model on multi-center datasets specifically collected for this purpose with standardized patient information; investigating optimized and computationally efficient architectures, including lightweight transformer variants such as MobileViT, to ensure reliable performance across diverse clinical environments; and conducting a user study with pathologists to evaluate the clinical relevance, interpretability, and reliability of LIME explanations in real diagnostic contexts.

8. Conclusion

This work presents an intelligent web application dedicated to medical decision-making with a focus on colon cancer. Unlike the majority of studies, which are limited to theoretical proposals of complex black-box models, our approach provides not only a powerful vision transformer (ViT) model but also an operational solution centered on the physician as the end user. This solution bridges the gap between advances in artificial intelligence research for the medical field and their effective adoption in real clinical settings.
Rigorous validation of our ViT model on multiple datasets demonstrates its exceptional performance not only on the LC25000 dataset but also on a merged dataset (LC25000 + DigestPath) and on external unseen data from the EBHI-Seg dataset. These results confirm its strong predictive reliability in real clinical settings. Furthermore, the comparative analyses highlight the remarkable competitive position of our approach among recent CNN architectures for colon cancer detection.
To enhance the interpretability and clinical reliability of the vision transformer model's predictions, our approach provides clear and intuitive visual explanations using the LIME (XAI) method. A quantitative evaluation of the LIME explanations confirmed that the highlighted regions are critically important for the model's predictions, as masking these areas led to a substantial decrease in model confidence. This reinforces the clinical relevance and credibility of the explanations provided.
In addition, a technology acceptance assessment based on the TAM framework revealed positive feedback from

users, with high scores in perceived ease of use and a strong behavioral intention to use the system. This confirms that the application is not only accurate and interpretable but also user-friendly and likely to be adopted in real clinical environments.

Furthermore, the deployment of our XAI-based deep learning model in a web application makes it immediately usable and removes hardware constraints for clinicians by centralizing computing power on the server side. Therefore, the strength of this research lies in three major advantages: the ease of sharing clinical cases and exchanging medical expertise, the reliable predictive performance provided by the vision transformer model, and the transparency of decisions ensured by the use of the model-agnostic explanatory method LIME.

REFERENCES

1. R. Nason, *Challenges of implementing complexity in healthcare*, Healthcare Management Forum, vol. 36, no. 6, pp. 368–372, 2023. https://doi.org/10.1177/08404704231191956
2. T. Wilson, T. Holt, and T. Greenhalgh, *Complexity and clinical care*, BMJ, vol. 323, no. 7314, pp. 685–688, 2001. https://doi.org/10.1136/bmj.323.7314.685
3. P. E. Plsek, and T. Greenhalgh, *The challenge of complexity in health care*, BMJ, vol. 323, no. 7313, pp. 625–628, 2001. https://doi.org/10.1136/bmj.323.7313.625
4. L. M. Allen, D. Balmer, and L. Varpio, *Physicians' lifelong learning journeys: A narrative analysis of continuing professional development struggles*, Medical Education, vol. 58, no. 9, pp. 1086–1096, 2024. https://doi.org/10.1111/medu.15375
5. F. Kobrai-Abkenar, S. Salimi, and P. Pourghane, *"Interprofessional Collaboration" among Pharmacists, Physicians, and Nurses: A Hybrid Concept Analysis*, Iranian Journal of Nursing and Midwifery Research, vol. 29, no. 2, pp. 238–244, 2024. https://doi.org/10.4103/ijnmr.ijnmr_336_22
6. M. Bekbolatova, J. Mayer, C. W. Ong, and M. Toma, *Transformative potential of AI in healthcare: definitions, applications, and navigating the ethical landscape and public perspectives*, Healthcare, vol. 12, no. 2, pp. 125, 2024. https://doi.org/10.3390/healthcare12020125
7. F. Atallah, R. F. Hamm, C. M. Davidson, C. A. Combs, and Patient Safety and Quality Committee, *Society for Maternal-Fetal Medicine special statement: Cognitive bias and medical error in obstetrics—Challenges and opportunities*, American Journal of Obstetrics and Gynecology, vol. 227, no. 2, pp. B2–B10, 2022. https://doi.org/10.1016/j.ajog.2022.04.033
8. A. S. Elstein, and A. Schwarz, *Clinical problem solving and diagnostic decision making: selective review of the cognitive literature*, BMJ, vol. 324, no. 7339, pp. 729–732, 2002. https://doi.org/10.1136/bmj.324.7339.729
9. M. Leonard, S. Graham, and D. Bonacum, *The human factor: the critical importance of effective teamwork and communication in providing safe care*, BMJ Quality & Safety, vol. 13, no. suppl 1, pp. i85–i90, 2004. https://doi.org/10.1136/qshc.2004.010033
10. G. Fan, Z. Deng, and B. Wang, *Social benefits of online peer information exchange among doctors: an empirical study on an online health community in China*, Behaviour & Information Technology, vol. 44, no. 11, pp. 2596–2619, 2024. https://doi.org/10.1080/0144929X.2024.2401869
11. A. I. Stoumpos, F. Kitsios, and M. A. Talias, *Digital transformation in healthcare: technology acceptance and its applications*, International Journal of Environmental Research and Public Health, vol. 20, no. 4, pp. 3407, 2023. https://doi.org/10.3390/ijerph20043407
12. S. Zeb, F. N. U. Nizamullah, N. Abbasi, and M. Fahad, *AI in healthcare: revolutionizing diagnosis and therapy*, International Journal of Multidisciplinary Sciences and Arts, vol. 3, no. 3, pp. 118–128, 2024. https://doi.org/10.47709/ijmdsa.v3i3.4546
13. C. Kaur, and U. Garg, *Artificial intelligence techniques for cancer detection in medical image processing: A review*, Materials Today: Proceedings, vol. 81, pp. 806–809, 2023. https://doi.org/10.1016/j.matpr.2021.04.241
14. D. S. Chan, M. Cariolou, G. Markozannes, K. Balducci, R. Vieira, S. Kiss, et al., *Post-diagnosis dietary factors, supplement use and colorectal cancer prognosis: a Global Cancer Update Programme (CUP Global) systematic literature review and meta-analysis*, International Journal of Cancer, vol. 155, no. 3, pp. 445–470, 2024. https://doi.org/10.1002/ijc.34906
15. H. Zahed, X. Feng, M. Sheikh, F. Bray, J. Ferlay, O. Ginsburg, et al., *Age at diagnosis for lung, colon, breast and prostate cancers: An international comparative study*, International Journal of Cancer, vol. 154, no. 1, pp. 28–40, 2024. https://doi.org/10.1002/ijc.34671
16. O. Attallah, M. F. Aslan, and K. Sabanci, *A framework for lung and colon cancer diagnosis via lightweight deep learning models and transformation methods*, Diagnostics, vol. 12, no. 12, pp. 2926, 2022. https://doi.org/10.3390/diagnostics12122926
17. A. Woźniacki, W. Książek, and P. Mrowczyk, *A novel approach for predicting the survival of colorectal cancer patients using machine learning techniques and advanced parameter optimization methods*, Cancers, vol. 16, no. 18, pp. 3205, 2024. https://doi.org/10.3390/cancers16183205
18. A. A. Khan, M. Arslan, A. Tanzil, R. A. Bhatty, M. A. U. Khalid, and A. H. Khan, *Classification of colon cancer using deep learning techniques on histopathological images*, Migration Letters, vol. 21, no. S11, pp. 449–463, 2024.
19. Y. Xu, L. Jiao, S. Wang, J. Wei, Y. Fan, M. Lai, and E. I. C. Chang, *Multi-label classification for colon cancer using histopathological images*, Microscopy Research and Technique, vol. 76, no. 12, pp. 1266–1277, 2013. https://doi.org/10.1002/jemt.22294
20. A. B. Hamida, M. Devanne, J. Weber, C. Truntzer, V. Derangère, F. Ghiringhelli, et al., *Deep learning for colon cancer histopathological images analysis*, Computers in Biology and Medicine, vol. 136, pp. 104730, 2021. https://doi.org/10.

1016/j.compbiomed.2021.104730
21. P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, *Explainable AI: A review of machine learning interpretability methods*, Entropy, vol. 23, no. 1, pp. 18, 2020. https://doi.org/10.3390/e23010018
22. C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, *Key challenges for delivering clinical impact with artificial intelligence*, BMC Medicine, vol. 17, no. 1, pp. 195, 2019. https://doi.org/10.1186/s12916-019-1426-2
23. J. Yin, K. Y. Ngiam, and H. H. Teo, *Role of artificial intelligence applications in real-life clinical practice: systematic review*, Journal of Medical Internet Research, vol. 23, no. 4, pp. e25759, 2021. https://doi.org/10.2196/25759
24. A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, et al., *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*, Information Fusion, vol. 58, pp. 82–115, 2020. https://doi.org/10.1016/j.inffus.2019.12.012
25. R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, *A survey of methods for explaining black box models*, ACM Computing Surveys (CSUR), vol. 51, no. 5, pp. 1–42, 2018. https://doi.org/10.1145/3236009
26. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., *An image is worth 16x16 words: Transformers for image recognition at scale*, arXiv preprint arXiv:2010.11929, 2020.
27. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., *Attention is all you need*, Advances in Neural Information Processing Systems, pp. 5998–6008, 2017.
28. S. Chang, Y. Zhang, W. Han, M. Yu, X. Guo, W. Tan, et al., *Dilated recurrent neural networks*, Advances in Neural Information Processing Systems, vol. 30, 2017.
29. Z. Zuo, B. Shuai, G. Wang, X. Liu, X. Wang, B. Wang, and Y. Chen, *Convolutional recurrent neural networks: Learning spatial dependencies for image representation*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 18–26, 2015.
30. S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, *Transformers in vision: A survey*, ACM Computing Surveys (CSUR), vol. 54, no. 10s, pp. 1–41, 2022. https://doi.org/10.1145/3505244
31. D. Rothman, *Transformers for Natural Language Processing*, Packt Publishing Ltd, 2022.
32. M. T. Ribeiro, S. Singh, and C. Guestrin, *"Why should I trust you?" Explaining the predictions of any classifier*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144, 2016. https://doi.org/10.1145/2939672.2939778
33. A. Mustapha, K. Abdellah, L. Mohamed, L. Khalid, H. Hamid, and K. Ali, *A mobile and web application for diseases classification using Deep Learning*, SoftwareX, vol. 23, pp. 101488, 2023. https://doi.org/10.1016/j.softx.2023.101488
34. L. M. Driss, A. Mustapha, L. Mohamed, H. Hamid, O. Najia, A. Roukéya, and A. E. Mohamed, *DLDiagnosis: A mobile and web application for diseases classification using Deep Learning*, SoftwareX, vol. 26, pp. 101745, 2024. https://doi.org/10.1016/j.softx.2024.101745
35. J. G. Pires, *A conversational artificial intelligence based web application for medical conversations: a prototype for a chatbot*, medRxiv, pp. 2023–12, 2024. https://doi.org/10.1101/2023.12.31.23300681
36. N. Ahmed, R. Ahammed, M. M. Islam, M. A. Uddin, A. Akhter, M. A. Talukder, and B. K. Paul, *Machine learning based diabetes prediction and development of smart web application*, International Journal of Cognitive Computing in Engineering, vol. 2, pp. 229–241, 2021. https://doi.org/10.1016/j.ijcce.2021.12.001
37. K. Mridha, S. Ghimire, J. Shin, A. Aran, M. M. Uddin, and M. F. Mridha, *Automated stroke prediction using machine learning: an explainable and exploratory study with a web application for early intervention*, IEEE Access, vol. 11, pp. 52288–52308, 2023. https://doi.org/10.1109/ACCESS.2023.3278273
38. A. Singh, S. Randive, A. Breggia, B. Ahmad, R. Christman, and S. Amal, *Enhancing prostate cancer diagnosis with a novel artificial intelligence-based web application*, Cancers, vol. 15, no. 23, pp. 5659, 2023. https://doi.org/10.3390/cancers15235659
39. C. N. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, *Development of a machine learning based web application for early diagnosis of COVID-19 based on symptoms*, Diagnostics, vol. 12, no. 4, pp. 821, 2022. https://doi.org/10.3390/diagnostics12040821
40. A. Siddique, K. Shaukat, and T. Jan, *An intelligent mechanism to detect multi-factor skin cancer*, Diagnostics, vol. 14, no. 13, pp. 1359, 2024. https://doi.org/10.3390/diagnostics14131359
41. N. Absar, E. K. Das, S. N. Shoma, M. U. Khandaker, M. H. Miraz, M. R. I. Faruque, et al., *The efficacy of machine-learning-supported smart system for heart disease prediction*, Healthcare, vol. 10, no. 6, pp. 1137, 2022. https://doi.org/10.3390/healthcare10061137
42. M. M. Rahman, *A web-based heart disease prediction system using machine learning algorithms*, Network Biology, vol. 12, no. 2, pp. 64, 2022.
43. M. Duan, Z. Geng, L. Gao, Y. Zhao, Z. Li, L. Chen, et al., *An interpretable machine learning-assisted diagnostic model for Kawasaki disease in children*, Scientific Reports, vol. 15, no. 1, pp. 7927, 2025. https://doi.org/10.1038/s41598-025-92277-1
44. R. Raheem, and A. Al-Qurabat, *Developing a predictive health care system for diabetes diagnosis as a machine learning-based web service*, Journal of University of Babylon for Pure and Applied Sciences, vol. 30, pp. 1–32, 2022.
45. L. Wijnia, S. M. Loyens, and R. M. Rikers, *The Problem-Based Learning Process: An Overview of Different Models*, The Wiley Handbook of Problem-Based Learning, pp. 273–295, 2019. https://doi.org/10.1002/9781119173243.ch12
46. A. S. Elstein, *What goes around comes around: return of the hypothetico-deductive strategy*, Teaching and Learning in Medicine: An International Journal, vol. 6, no. 2, pp. 121–123, 1994. http://dx.doi.org/10.1080/10401339409539658
47. M. A. Ferradji and A. Zidani, *Collaborative environment for remote clinical reasoning learning*, International Journal of E-Health and Medical Communications (IJEHMC), vol. 7, no. 4, pp. 62–81, 2016. https://doi.org/10.4018/IJEHMC.2016100104
48. C. Molnar, *Interpretable machine learning*, 2020. https://christophm.github.io/interpretable-ml-book/
49. S. Chen, U. R. Thaduri, and V. K. R. Ballamudi, *Front-end development in React: an overview*, Engineering International, vol. 7, no. 2, pp. 117–126, 2019. https://doi.org/10.18034/ei.v7i2.662

50. A. Holovaty, and J. Kaplan-Moss, *The definitive guide to Django: Web development done right*, Apress, Berkeley, CA, 2009. https://doi.org/10.1007/978-1-4302-1937-8_4
51. A. Szumowska, M. Burzańska, P. Wiśniewski, and K. Stencel, *Efficient implementation of recursive queries in major object relational mapping systems*, International Conference on Future Generation Information Technology, Springer, Berlin, Heidelberg, pp. 78–89, 2011. https://doi.org/10.1007/978-3-642-27142-7_10
52. M. Kumar, and R. Nandal, *Role of Python in Rapid Web Application Development Using Django*, Proceedings of the International Conference on Innovative Computing and Communication, New Delhi, India, pp. 1–13, 2024. https://dx.doi.org/10.2139/ssrn.4751833
53. F. D. Davis, *Perceived usefulness, perceived ease of use, and user acceptance of information technology*, MIS Quarterly, vol. 13, no. 3, pp. 319–340, 1989. https://doi.org/10.2307/249008
54. L. J. Cronbach, *Coefficient alpha and the internal structure of tests*, Psychometrika, vol. 16, no. 3, pp. 297–334, 1951. https://doi.org/10.1007/BF02310555
55. M. Tavakol and R. Dennick, *Making sense of Cronbach's alpha*, International Journal of Medical Education, vol. 2, pp. 53–55, 2011. https://doi.org/10.5116/ijme.4dfb.8dfd
56. J. Brooke, *SUS – A quick and dirty usability scale*, Usability Evaluation in Industry, vol. 189, no. 194, pp. 4–7, 1996.
57. S. G. Hart and L. E. Staveland, *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research*, in *Advances in Psychology*, vol. 52, pp. 139–183, 1988. https://doi.org/10.1016/S0166-4115(08)62386-9
58. A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides, *Lung and Colon Cancer Histopathological Image Dataset (LC25000)*, arXiv:1912.12142, 2019. https://doi.org/10.48550/arXiv.1912.12142
59. S. Gupta, and H. S. Khanuja, *DigestPath Dataset*, Kaggle, 2021. https://www.kaggle.com/datasets/mittalswathi/digestpath-dataset
60. L. Shi, X. Li, W. Hu, H. Chen, J. Chen, Z. Fan, et al., *EBHI-Seg: A Novel Enteroscope Biopsy Histopathological Haematoxylin and Eosin Image Dataset for Image Segmentation Tasks*, arXiv:2212.00532, 2022. https://doi.org/10.3389/fmed.2023.1114673
61. J. Leem and S. Seo, *Attention Guided CAM: Visual Explanations of Vision Transformer Guided by Self-Attention*, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 4, pp. 2956–2964, 2024. https://doi.org/10.1609/aaai.v38i4.28077
62. S. Abnar and W. Zuidema, *Quantifying Attention Flow in Transformers*, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 4190–4197, 2020. https://doi.org/10.18653/v1/2020.acl-main.385
63. S. Mangal, A. Chaurasia, and A. Khajanchi, *Convolution neural networks for diagnosing colon and lung cancer histopathological images*, arXiv:2009.03878, 2020. https://doi.org/10.48550/arXiv.2009.03878
64. A. S. Sakr, N. F. Soliman, M. S. Al-Gaashani, P. Pławiak, A. A. Ateya, and M. Hammad, *An efficient deep learning approach for colon cancer detection*, Applied Sciences, vol. 12, no. 17, pp. 8450, 2022. https://doi.org/10.3390/app12178450
65. S. U. K. Bukhari, A. Syed, S. K. A. Bokhari, S. S. Hussain, S. U. Armaghan, and S. S. H. Shah, *The histological diagnosis of colonic adenocarcinoma by applying partial self supervised learning*, medRxiv, 2020. https://doi.org/10.1101/2020.08.15.20175760
66. A. A. A. Mohamed, A. Hançerlioğullari, J. Rahebi, R. Rezaeizadeh, and J. M. Lopez-Guede, *Colon cancer disease diagnosis based on convolutional neural network and fishier mantis optimizer*, Diagnostics, vol. 14, no. 13, pp. 1417, 2024. https://doi.org/10.3390/diagnostics14131417
67. N. Shahadat, R. Lama, and A. Nguyen, *Lung and colon cancer detection using a deep AI model*, Cancers, vol. 16, no. 22, pp. 3879, 2024. https://doi.org/10.3390/cancers16223879
68. S. A. Opee, A. A. Eva, A. T. Noor, S. M. Hasan, and M. F. Mridha, *ELW-CNN: An extremely lightweight convolutional neural network for enhancing interoperability in colon and lung cancer identification using explainable AI*, Healthcare Technology Letters, vol. 12, no. 1, pp. e12122, 2025. https://doi.org/10.1049/htl2.12122
69. M. D. Giammarco, F. Martinelli, A. Santone, M. Cesarelli, and F. Mercaldo, *Colon cancer diagnosis by means of explainable deep learning*, Scientific Reports, vol. 14, no. 1, pp. 15334, 2024. https://doi.org/10.1038/s41598-024-63659-8