Engineering and Technology Journal
Journal homepage: https://etj.uotechnology.edu.iq

Traditional machine learning techniques for explainable AI in lung and colon cancer classification

Mahmood Th. Amer*, Zainab N. Sultani

Computer Science Dept., College of Science, Al-Nahrain University, Baghdad, Iraq.
*Corresponding author Email: mo.alnuimi1593@gmail.com

HIGHLIGHTS

- An image-based method was developed to classify lung and colon cancer using CLAHE-enhanced histopathology images.
- Color histograms and LBP features were combined to improve classification across five cancer-related classes.
- XGBoost with RFE achieved 99.80% accuracy by selecting the most relevant handcrafted features.
- CLAHE preprocessing enhanced feature clarity, improving model accuracy and interpretability.
- SHAP and LIME tools were used to explain model decisions, supporting transparent AI-driven cancer diagnosis.

Keywords:
Lung cancer
LC25000 dataset
Colon cancer
Machine learning
Explainable AI.

ABSTRACT

Lung and colon cancers are two of the most common and deadly tumors around the world, creating significant public health concerns. Artificial intelligence (AI) and machine learning (ML) have heavily improved cancer research, particularly in early detection, histopathological analysis, and personalized therapy planning. However, despite their remarkable accuracy, ML models sometimes lack transparency, making explainability crucial in medical applications. Although various machine learning (ML)-based classifications for cancer models exist, their interpretation is not understood. The current research overcomes the diagnostic gap by developing a highly accurate system that uses XAI (Explainable Artificial Intelligence) methods to clarify its predictions. We used Kaggle's LC25000 dataset, which included histology images for lung and colon tumors in humans. To determine the best cancer classification strategy, we tested various machine learning algorithms, including Random Forest, Decision Tree, Support Vector Machine (SVM), and Extreme Gradient Boosting. Furthermore, XAI approaches such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (Shapley Additive Explanations) were utilized to evaluate model predictions and identify important information affecting classification outcomes. XGBoost confirmed that it was useful in identifying colon and lung cancer by achieving the highest accuracy of 99.80% among the models used. Also, XAI techniques offered useful information on the most significant features. SHAP analysis highlighted LBP and color histogram features as key for distinguishing lung and colon tissues, while LIME confirmed their importance by identifying critical image regions influencing predictions.

1. Introduction

Cancer is the leading cause of mortality worldwide, second only to cardiovascular disease [1]. In 2018, they were responsible for 2.9 million new instances and more than 2.5 million fatalities in the United States alone [2]. Lung cancer is responsible for 18.4% of cancer fatalities, whereas 9.2% of cases are of colon cancer [3-4]. Approximately 17% of lung and colon cancer cases occur simultaneously. This is unlikely, yet in the absence of early exploration, cancer cell transmission across these two organs is rather common [5]. It is difficult to detect cancer without a comprehensive diagnostic test such as a computed tomography (CT) scan, magnetic resonance imaging (MRI), positron emission tomography (PET) scan, ultrasound, or biopsy. In many situations, individuals show little to no symptoms in the early stages, and when symptoms do appear, the diagnosis is usually too late [1]. Currently, adequate therapy as well as quick detection are just ways to minimize deaths from cancer [6]. This study only uses histological images to diagnose colon and lung cancers automatically. Health professionals commonly utilize histopathological images for diagnosis, and they are critical in estimating patients' chances of survival [1]. Traditionally, health experts had to go through a lengthy procedure to diagnose cancer by reviewing histopathological images; however, with the technological tools accessible today, this process can now be completed with less time and effort [5].

With the growth of AI and machine learning for medical imaging, image processing methods such as image vision, classification, and segmentation assist radiologists in diagnosing illnesses as early as feasible compared to manual examinations [7]. The advancement of machine learning has permitted the creation of decision support systems that display predictive

http://doi.org/10.30684/etj.2025.158895.1935
Received 06 April 2025; Received in revised form 05 June 2025; Accepted 15 June 2025; Available online 20 July 2025
2412-0758/University of Technology-Iraq, Baghdad, Iraq

accuracy, outperforming the skills of humans in some contexts [8]. However, the model's complexity level has increased due to this advancement, leading to models of black boxes that conceal their underlying reasoning from users [8]. This is especially noticeable when users need to interpret the output of AI systems [9]. Because technology is used in medical applications such as cancer prediction, the solutions must be dependable and efficient, as they put human lives at risk, where errors can have serious implications [10]. Because black-box characteristics are classified across multiple fields, people and experts find these ML models' applications unreliable [11].

To solve this problem, explainable artificial intelligence (XAI) was created, explaining how machine learning (ML) makes decisions. Explainable machine learning aims to make "black box" modeling decisions more transparent and trustworthy [12]. To put it simply, XAI is the action of making AI systems simpler for people to understand [13-15]. Shapley Additive ExPlanation (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are two major XAI approaches [16-17]. The paper's primary achievements include establishing an accurate and interpretable diagnostic approach for lung and colon cancers. Color histograms and Local Binary Patterns (LBP) were used to determine important image features. After extracting features, Recursive Feature Elimination (RFE) was applied to streamline the feature set. These features were then used as input into classification models. The system delivers remarkable diagnostic performance using machine learning models such as Decision Tree, Random Forest, SVM, and XGBoost with XAI techniques. Integrating Explainable AI (XAI) approaches such as SHAP and LIME promotes transparency by exposing the characteristics that influence predictions. This method builds confidence, facilitates clinical decision-making, and reduces the gap between useful applications for healthcare and potent AI technologies, resulting in earlier identification and better patient outcomes. The following sections comprise the paper's structure: The related work is covered in the second section, the material and methods are covered in the third, and the results and the discussion are covered in the fourth. Additionally, the conclusion is covered in the fifth section.

Developments in machine learning (ML) and deep learning (DL) techniques have improved the classification of lung and colon tumors. Researchers have examined various methods and procedures to improve diagnosis accuracy and model interpretability, two factors crucial in clinical applications. The LC25000 dataset has been utilized in recent research to classify histological images shows in Table 1.

Chehade et al. [18], concentrate on classifying photos of lung and colon cancer histopathology using machine learning techniques. Histopathological image downsizing to 200×200 pixels can be found in the LC25000 collection. During preprocessing, contrast is enhanced using the Unsharp Masking (UM) approach. The three feature extraction techniques are first-order statistics, the gray-level co-occurrence matrix (GLCM), and Hu invariant moments. The Recursive Feature Elimination (RFE) method is used for feature selection. Many machine learning models, such as XGBoost, Support Vector Machine (SVM), Random Forest (RF), Linear Discriminant Analysis (LDA), and Multilayer Perceptron (MLP), are used for categorization. When training on the entire LC25000 dataset (25,000 images) with 70% training and 30% testing, XGBoost obtained the greatest accuracy of 99%. When Explainable AI (XAI) techniques are implemented, feature contributions are understood using SHAP (Shapley Additive exPlanations). The study demonstrates that machine learning algorithms can accurately diagnose lung and colon cancer and generate easily comprehensible results. Al-Jabbar et al. [19], classified lung and colon cancers. The averaging filter and CLAHE are employed to increase contrast and lower noise. While DWT, LBP, FCH, and GLCM extract handcrafted features, GoogLe Net and VGG-19 extract deep features. The overlap in selecting features has been removed via PCA. By attaining 99.64% accuracy with VGG-19 and handcrafted features and 99.22% accuracy with GoogLeNet and handcrafted features, the main classifier, ANN, demonstrated the effectiveness of feature fusion in cancer diagnosis.

Table 1: Comparing the obtained outcomes with different methods utilizing the same lung and colon cancer dataset

Focus	Machine Learning Model	Results	Ref.
Classification: Lung & Colon Cancer	XGBoost, SVM, and other ML models trained on the LC25000 dataset with feature engineering techniques	XGBoost Accuracy 99%	[18]
Early diagnosis of lung and colon cancer	GoogLeNet, VGG-19, ANN	achieving 99.64% accuracy with VGG-19, 99.22% accuracy with GoogLeNet	[19]
Classification of lung and colon cancer using an ensemble deep learning approach with advanced hyperparameter tuning.	Ensemble of Xception and MobileNet architectures	Achieved exceptional precision and recall for malignant tissues, with a 99.44% classification accuracy.	[20]
Lung and colon cancer classifier using Explainable AI (XAI)	Lightweight Multi-Scale CNN (LW-MS-CCN) with Grad-CAM & SHAP for explainability	Achieved 99.20% accuracy on LC25000 dataset	[21]

Vanitha et al. [20], offered a deep-learning ensemble method for detecting lung and colon cancer using histopathology images. It uses the LC25,000 dataset, which comprises 25,000 histopathology images divided into five groups. The images were consistently downsized to 224×224 pixels and evaluated using two deep learning architectures, Xception and MobileNet. These algorithms were adapted for classifying cancer after being pre-trained on the ImageNet dataset. An ensemble technique was used to improve classification accuracy, pooling the output of both models. Explainable AI (XAI) was developed using Gradient-weighted Class Activation Mapping (Grad-CAM), visually depicting the model's decision-making procedure. The suggested methodology demonstrated a remarkable 99.44% accuracy rate in categorizing histological images of colon and lung cancer.

A lightweight multi-scale convolutional neural network (LW-MS-CCN) model was suggested by Hasan et al. [21], for classifying lung as well as colon cancers using histopathology images. The dataset used in the current study is LC25000. The

dataset was prepared by downsizing to 180×180 pixels and using data augmentation techniques like flips and rotations to improve the model's resilience. By selecting and extracting features using multi-scale convolutional layers, a machine could identify patterns in histopathological images at the global and local levels. The model's 99.20% accuracy in cancer classification shows how well it worked. Furthermore, Explainable AI (XAI) tools such as Grad-CAM and SHAP were included in the model to visually understand its judgments, improving the algorithm's transparency and dependability for medical professionals.

2. Materials and methods

Figure 1 shows the overall layout of the recommended method for classifying lung and colon cancer using histopathological images. The LC25000 dataset contains 25,000 images, of which 5,000 have been reduced to 250 by 250 pixels for every class. Prior to collecting each image's attributes, the contrast of each image was increased using the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique. The most effective features are selected using a feature selection technique called Recursive Feature Elimination. 80% of the dataset was utilized for training (chosen at random), while 20% of it was used for testing. The training set's image properties are used for training a machine learning method. The efficacy of the model is assessed using the testing set image features. Lastly, SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) are employed to improve interpretation and better understand the predictions.

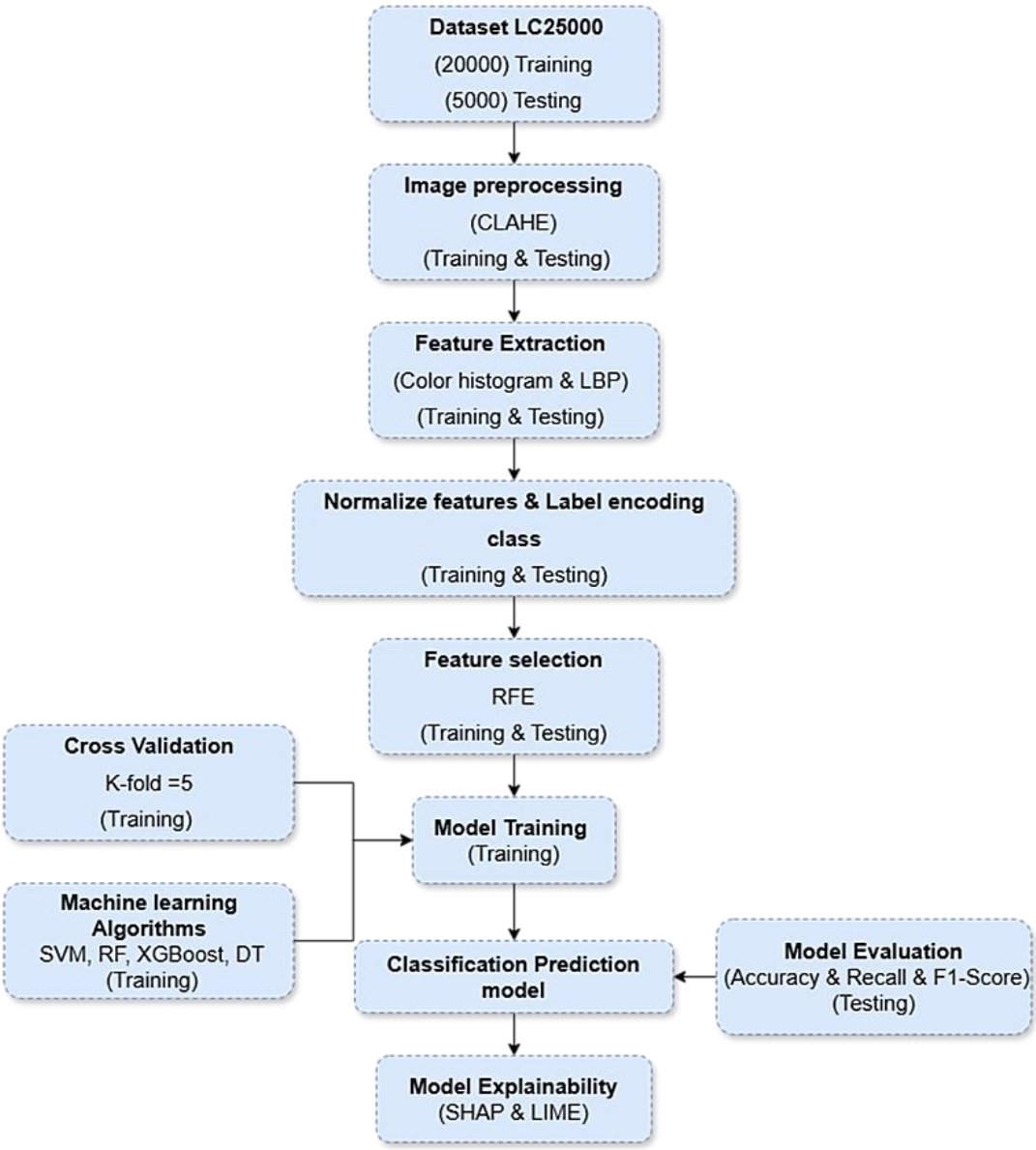

Figure 1: Block diagram of the approach used to classify cancer types by histopathological images

2.1 Lung and colon cancers datasets

This paper made use of A. Borkowski and his associates' histopathology images (LC25000) dataset that was released [22]. There are 25,000 images of lung and colon tissues in this collection, arranged within five categories, with 15,000 depicting lung cancer and 10,000 depicting colon cancer [23]. The LC25000 dataset is divided into five files, each representing one class/category: lung aca (adenocarcinoma), lung scc (squamous carcinoma), lung n (benign), colon aca (adenocarcinoma), and colon n (benign). Each folder contains 5,000 tissue images [24]. All images on LC25000 have the same resolution, 768x768

pixels. The original developers also enhanced the images (random rotation, horizontal and vertical flip) to increase the data in the dataset [25]. Figure 2 (a - e) depicts images from the dataset that demonstrate significant tissue types such as colon adenocarcinoma, lung squamous cell carcinoma, benign colon and lung tissue, and lung adenocarcinoma, illustrating the variety of the dataset. Table 2 shows 25,000 histopathology images from the LC25000 dataset split into tissues of the colon (10,000 samples) and the lung (15,000 samples). Each group contains both normal and malignant samples.

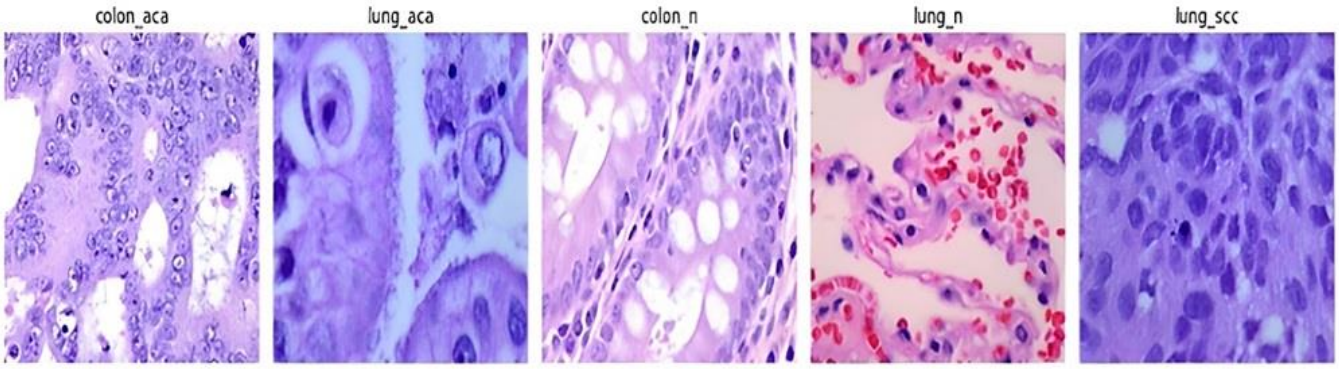

Figure 2: Several histopathology image samples from the LC25000 dataset (a) colon adenocarcinoma (b) lung adenocarcinoma (c) normal colon (d) normal lung (e) lung squamous cell carcinoma

Table 2: Summary of the LC25000 dataset

Cancer Type	Samples	Colon and Lung
Colon adenocarcinoma	5000	**10000**
Colon benign tissue	5000	
Lung adenocarcinoma	5000	**15000**
Lung benign tissue	5000	
Lung squamous cell carcinoma	5000	
Total	**25000**	

2.2 Data preprocessing

The input data was preprocessed before training the model. First, the images were scaled to 250*250 pixels using bicubic interpolation. Bicubic interpolation produces higher-quality images with smoother transitions. It is generally the better choice for medical images due to its ability to preserve important image details, making it better suited for applications that require high accuracy in visualizing structures. CLAHE filtering was applied to enhance image quality for better suitability in the learning process.

2.2.1 Contrast limited adaptive histogram equalization (CLAHE) filter

CLAHE (Contrast Limited Adaptive Histogram Equalization) is an adaptable histogram equalization technique [26, 27] that limits the excessive improvement of image contrast. CLAHE is commonly used to improve image contrast in computer vision and pattern recognition applications [28-29]. CLAHE was added to the image's L component to boost visibility before being converted back to RGB [30]. The LAB color space separates color and luminance information, making it easier to use contrast enhancement techniques [31-32]. Figure 3 shows the procedures for applying a contrast algorithm to the image's luminance channel. Figure 4 depicts the application of CLAHE to a histopathology image. This preprocessing stage is critical in medical imaging because it improves fine details, resulting in more accurate feature extraction and analysis.

2.3 Feature extraction

Feature extraction (FE) is critical in image retrieval, processing, data mining, and computer vision. FE is the process of extracting useful information from unstructured data. It is used to extract the most distinguishing features from a dataset (image, text, voice), which are then utilized to represent and describe the data. This paper employs color histogram and local binary pattern (LBP) feature extraction.

2.3.1 color histogram

The color histogram (CH) is the most extensively used technique for obtaining an image's color feature [33,34]. Proposed as a global color descriptor that examines all statistical color frequencies in an image [35]. An HSV color space was used because the hue and saturation of the human visual system are extremely similar [36]. Color histograms based on the HSV color space are more precise than other popular color representations, such as color layout descriptors, color moments, and color structure descriptors based upon retrieval in the HSV color space. Figure 5 (a - d) shows a histopathology image's HSV (hue, saturation, value) color space decomposition. A 512-bin color histogram uses 8 bins per channel (H: 0-256, S: 0-256, V: 0-256) to represent the color distribution of an image in HSV space. Figure 6 presents the extraction of color features from the Histopathology image for analysis using the "Color Histogram". The HSV (hue, saturation, value) color distribution is represented by red, green, and blue curves, respectively.

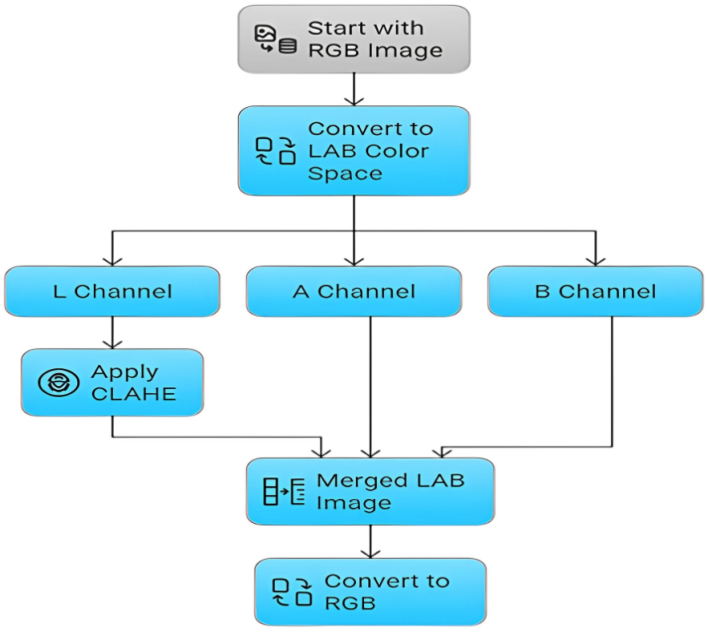

Figure 3: Enhance image by CLAHE

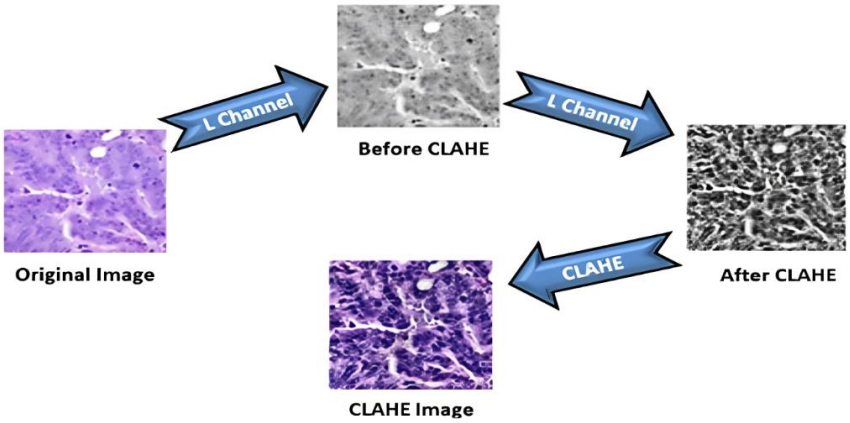

Figure 4: CLAHE technique on the LC25000 Dataset

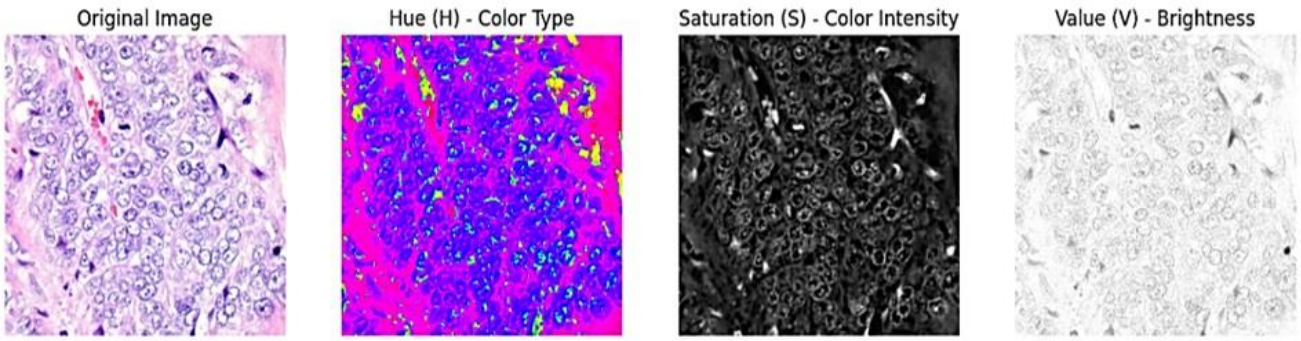

Figure 5: HSV color space decomposition for histopathology image (a) Original (b) Hue (c) Saturaction (d) Brightness

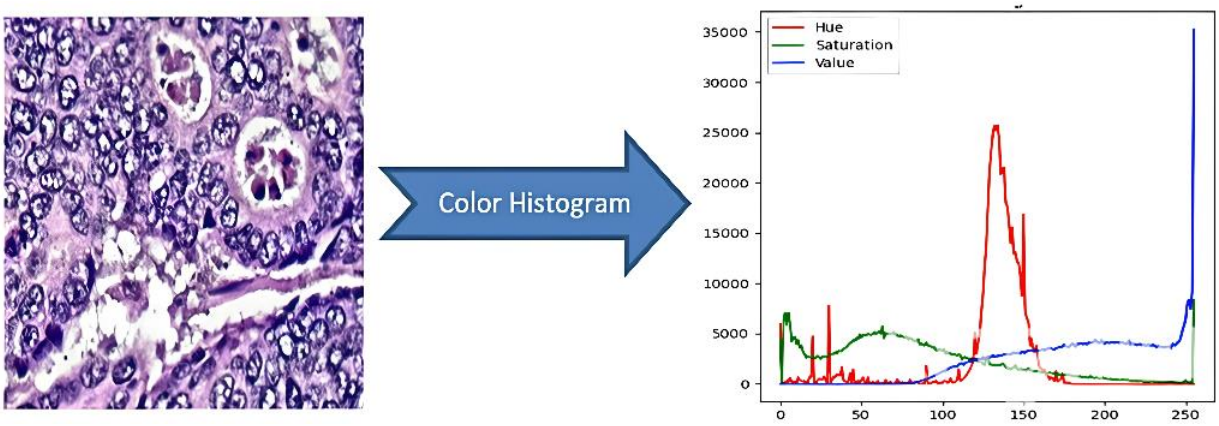

Figure 6: Color feature extraction using HSV histogram in histopathology images

2.3.2 local binary pattern (LBP)

One of the most often used texture descriptors in computer vision is LBP. Most applications utilize LBP histograms as texture features, creating a high-dimensional feature space, especially for tasks involving categorizing color textures. The LBP operator is commonly used to produce texture features for gray-level image classification because of its inherent simplicity and resilience [37]. This procedure applies throughout the full image to build an LBP feature map, capturing texture patterns for machine learning-based image classification and recognition tasks. Figure 7 shows LBP, where a grayscale image's pixels are compared to the center (e.g., 87), forming an 8-bit binary (e.g., 11100001 → 225), creating a texture-based feature map.

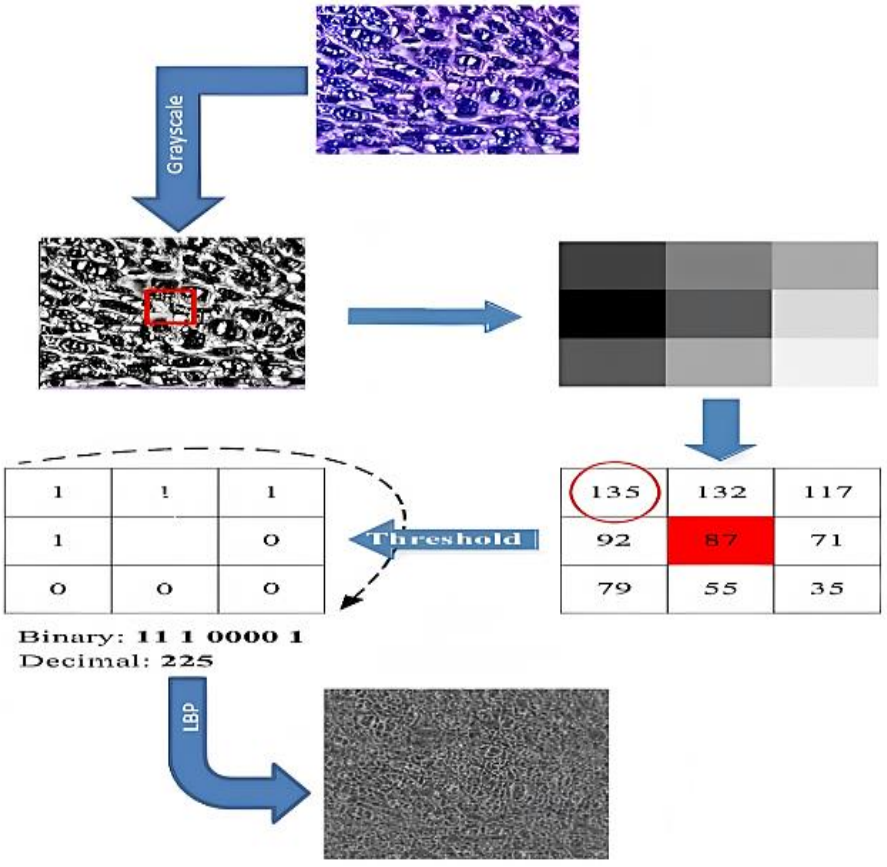

Figure 7: LBP feature extraction

Table 3 classifies distinct patterns of LBP into two distinct groups: uniform and non-uniform. Uniform patterns have no more than two-bit transitions, such as (00000000, 00000001). These patterns depict basic textures like smooth areas, edges, and corners. Non-uniform patterns, like 10101010, feature more than two-bit transitions and represent complex textures. Based on this classification, 10 LBP features were extracted.

Figure 8 illustrates the feature extraction process for image analysis. It has 512 color histogram features and 10 additional Local Binary Pattern (LBP) features, for a total of 522. These properties are useful for image categorization and texture analysis.

Table 3: LBP binary patterns and texture classification

Bin LBP pattern	Binary pattern	Uniform/Non-Uniform
0 (Smooth Regions)	00000000	Uniform
1 (Edges)	00000001	Uniform
2 (Edges)	00000011	Uniform
3 (Edges)	00000111	Uniform
4 (Edges)	00001111	Uniform
5 (Corners)	11100011	Uniform
6 (Corners)	10001111	Uniform
7 (Corners)	11110001	Uniform
8 (Corners)	11110011	Uniform
9 (Complex Textures)	10101010	Non-Uniform

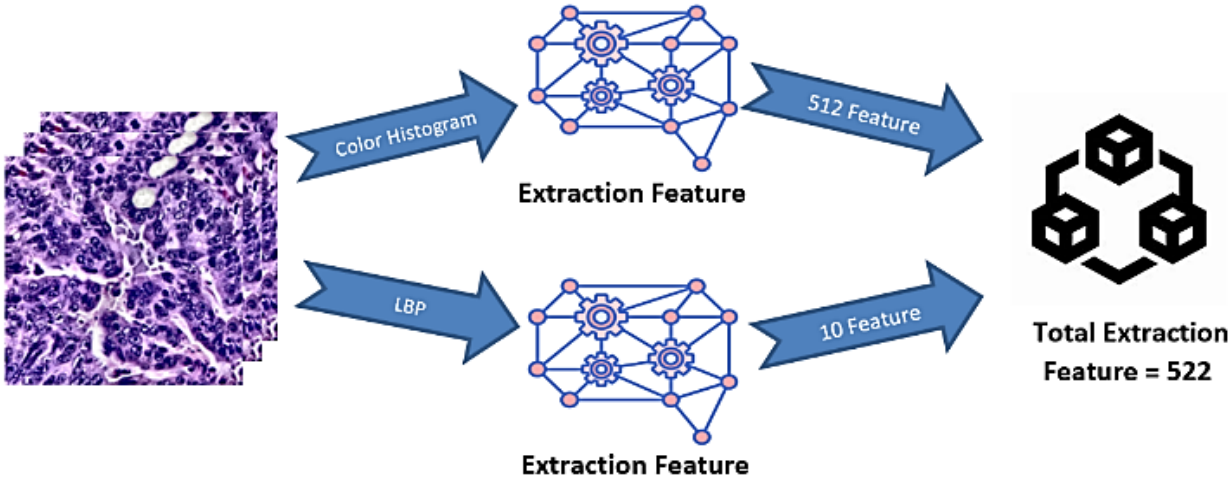

Figure 8: Color histogram and LBP feature extraction

2.4 Feature selection - recursive feature elimination (RFE)

Recursive Feature Elimination (RFE) is an approach of generality that may be used with most models [38]. RFE is a common pruning approach for selecting features that enhance classification performance [39,40]. RFE is the process of gradually removing the least important features from a model by recursively training it and measuring feature relevance until the required number of features is achieved [41]. It assists in identifying the most relevant subset of characteristics for the best model performance.

2.5 Machine learning models

Over the past ten years, machine learning techniques have become a significant threat to traditional statistical models in terms of prediction [42]. Machine learning (ML) technology greatly aided the development of prediction systems, producing better and more affordable solutions. We experimented with various machine learning classifiers, such as decision Trees, support vector machines (SVM), random forests, and XGBoost. Decision trees are the most popular and efficient tool for categorizing and predicting difficulties [43]. It is hierarchically structured, including root, branches, and internal and leaf nodes. It has a tree-like structure and is used to create a model that predicts the target value using judgments from the training dataset [44]. Support Vector Machines (SVMs) have been developed as an effective classifier and regression tool. [45]. Random forests are a set of tree predictors in which each tree is determined by the values of a random vector selected separately and uniformly across the forest [46]. It is a quick and computationally effective method for dealing with huge datasets [47]. XGBoost is a powerful gradient tree-boost software that can handle large-scale machine learning workloads [48]. Because of its outstanding performance and low time and memory requirements have been widely used in various research fields since its introduction, including cancer diagnosis and medical record analysis [49]. The benefit of regularization is that it helps prevent overfitting, allowing the model to generalize better to unseen data [50].

The model's hyperparameters were set to maintain computational efficiency while enhancing the model's ability to generalize well using grid search. Tables 4, 5, 6, and 7 display the DT, SVM, RF, and XGBoost hyperparameters, respectively.

Figure 9 illustrates that Grid search is a systematic method for selecting the best hyperparameters for machine learning models by exhaustively evaluating predefined parameter combinations. The process begins by defining a parameter grid, which specifies the range of values for each hyperparameter (e.g., learning rate, tree depth, or regularization strength). The algorithm then trains and evaluates the model for each possible combination of these values, typically using cross-validation to ensure robust performance estimation. In this case, the model's performance is measured using a predefined metric, such as accuracy. After testing all combinations, the hyperparameter set that yields the highest performance is selected as the optimal configuration. The key benefit of grid search is that it guarantees finding the best combination of hyperparameters within the specified search space. While it can be computationally expensive, grid search provides a reliable and straightforward method for hyperparameter optimization, improving the model's performance and generalization ability.

Table 4: Hyperparameters of the decision tree model

hyperparameters	value
Maximum depth of each tree	10
Minimum number of leaf nodes	5
Minimum number to split	5
criterion	entropy
Random state	42
Minimum impurity decrease	0.01

Table 5: Hyperparameters of the SVM model

hyperparameters	value
kernel	linear
gamma	scale
C	0.1
probability	True
Random State	42

Table 6: Hyperparameters of the random forest model

hyperparameters	value
bootstrap sampling	True
Maximum depth of each tree	10
Minimum number of leaf nodes.	4
Minimum number to split	10
Number of trees	150
Maximum Features	sqrt
Random State	42

Table 7: Hyperparameters of the XGBoost model

hyperparameters	value
Column subsampling by tree	0.7
Learning rate	0.05
Maximum depth of each tree	4
Number of trees	200
Subsample ratio of instances	0.7
Evaluation metric	Mlogloss
Minimum child weight	10
Gamma	1.0
Regularization alpha	2.0
Regularization lambda	1.0
Random state	42
Use a label encoder	False

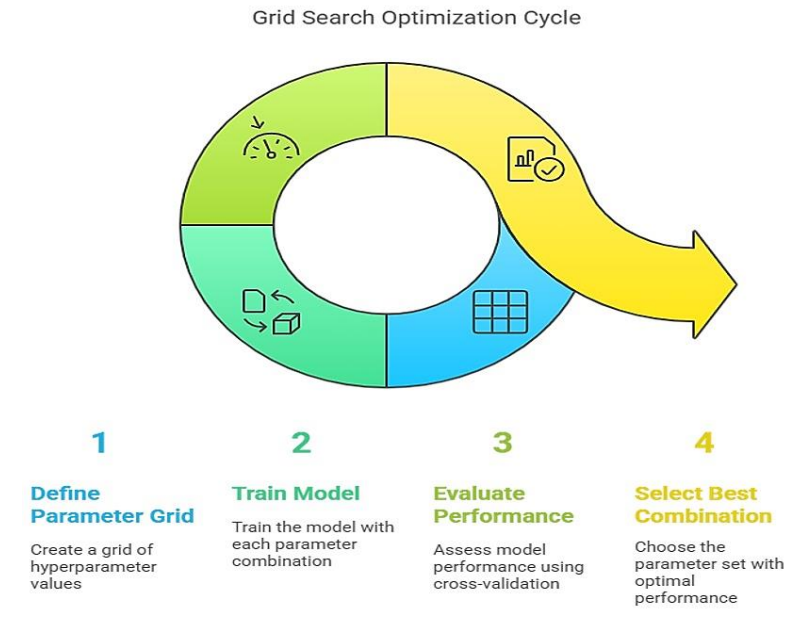

Figure 9: Grid search optimization for hyperparameter tuning

2.6 Model explainability

SHAP and LIME are techniques for explaining machine learning predictions. SHAP assigns feature importance using game theory, while LIME approximates model behavior locally. Both improve model interpretability.

SHAP uses values to demonstrate the transparency of the machine learning models. It calculates how much each attribute contributes to the forecast by calculating Shapley values throughout the dataset. Proposes explaining model prediction for a given input by computing each feature's contribution [51]. Equation 1 calculates the SHAP value (ϕ_i) for each feature by considering its contribution across all possible feature subsets, ensuring a fair distribution of feature importance.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} \left(f(S \cup \{i\}) - f(S)\right) \quad (1)$$

N collection of all features, S subset of N excluding feature i, f (S) models prediction utilizing only the features in S, f (S∪{i}) models prediction that involve feature i, ϕ_i is the SHAP value, representing the contribution of feature i to the final prediction.

LIME is another interpretable model used to show machine learning models' interpretability. Unlike SHAP, it supports local model interpretation. XAI approach for successfully explaining classifier or regressor predictions by approximating them locally with an interpretable model [52]. LIME is model-agnostic, which means it may be used with any machine-learning model [52]. To maintain both the interpretability and fidelity of the surrogate model to the original model, we optimize the objective function presented in Equation 2.

$$arg \min_{g \in G} \; L(f, g, \pi_x) + \Omega(g) \quad (2)$$

G is the set of possible interpretable models, L (f, g, π_x) is the loss function measuring the variation from the initial model f and the surrogate model g, weighted by π_x , which assigns higher importance to points closer to x, Ω (g) is a complexity penalty to ensure interpretability.

3. Results and discussion

The Lung and Colon Cancer Histopathological Image Dataset includes 25000 images separated into five groups (colon_aca, lung_aca, colon_n, lung_n, and lung_scc), each with 5000 images. The images are divided into two folders. The training folder has 20000 images, whereas the testing folder contains 5000 images. All images across the five classes were resized to 250 × 250 pixels. After that, feature extraction was performed using two texture analysis techniques: color histogram and LBP. Table 7 highlights the evaluation of model performance using these individual techniques and their combinations to determine the most effective feature extraction strategy with and without using the CLAHE filter. Following that, Table 8 highlights that the feature set was refined and optimized using Recursive Feature Elimination (RFE) without a CLAHE filter. Table 9 shows the results of the CLAHE filter to improve image quality and learning performance.

Table 8 compares the classification accuracy of various machine learning models using three feature extraction strategies: Color Histogram, Local Binary Patterns (LBP), and their combination. The results indicate a clear trend in feature effectiveness, which can be explained based on the features' nature and complementary properties.

color histogram alone

- This feature captures the global color distribution in an image, which is particularly effective when color is a strong discriminative cue for classification.
- It consistently achieves high accuracy across all models (above 95%), suggesting that color information alone is highly informative for the task.

LBP alone

- LBP is a texture descriptor that captures local spatial patterns, which is valuable in images where surface texture contributes to class distinction.
- However, its performance is notably lower than Color Histogram in all models (dropping as low as ~83% with SVM).
- This suggests that texture information alone may be insufficient for accurate classification in this dataset, possibly due to low texture variation or the dominance of color-based cues.

color histogram and LBP

- The combined feature set consistently outperforms the individual features, achieving the highest accuracy (e.g., 99.56% with XGBoost).
- This improvement highlights the complementary nature of the features: while color histogram captures global color distribution, LBP adds local texture information that can resolve ambiguities when color alone is not decisive.
- The synergy between these features likely allows the model to better generalize across varied image characteristics.
- This analysis suggests that feature diversity enhances model performance by providing complementary information. XGBoost benefits the most from the combined features, due to its ability to handle complex feature interactions and perform fine-grained splits in high-dimensional space.

Table 8: Comparing the accuracy outcomes using the extraction of features for the various models without the CLAHE filter

Feature extraction	RF	XGBoost	SVM	DT
Color histogram	97.56%	98.98%	95.96%	97.32%
LBP	86.94%	86.12%	83.52%	84.26%
Color histogram + LBP	98.02%	99.56%	97.08%	98.26%

Table 9 compares the performance of four models on image classification using different features selected by Recursive Feature Elimination (RFE). Decision Tree and SVM performance remain consistent, with only minor gains across different feature subsets, demonstrating that they are less susceptible to the selected number of features. In contrast, XGBoost and Random Forest gain considerably from additional features, obtaining the maximum accuracies (99.58%) with 100 features.

RFE improves the performance of XGBoost and Random Forest by identifying the most important features, allowing these models to capture complicated patterns and increase accuracy. Random Forest works particularly well by aggregating forecasts from several decision trees using a diversified feature set. Overall, RFE contributes the most to XGBoost and Random Forest, resulting in significant gains in classification performance.

Table 9: Accuracy of different models with varying numbers of features selected via (RFE) without the use of CLAHE filter

Model	NF= 50	NF= 100	NF= 200	NF= 250
Random Forest	98.78%	98.70%	98.52%	98.18%
XGBoost	99.32%	99.58%	99.50%	99.46%
SVM	96.46	97.14	97.8	97.10
Decision Tree	98.28%	98.32%	98.22%	98.26

Note: NF =number of feature selection

Table 10 compares the accuracy results of several feature extraction models with the CLAHE filter. Without RFE, XGBoost had the greatest classification accuracy of 99.72%, followed closely by Random Forest at 98.70%. SVM and Decision Tree also performed well, scoring 97.92% and 98.52%, respectively. After adding RFE, all models improved somewhat, with Random Forest climbing to 98.78% and XGBoost to 99.80%. SVM and Decision Tree also improved performance, achieving 97.94% and 98.60%, respectively.

Table 10: Comparing the accuracy outcomes using the extraction of features and RFE for the various models using the CLAHE filter

Accuracy with CLAHE	RF	XGBoost	SVM	DT
Accuracy without RFE	98.70%	99.72%	97.92%	98.52%
Accuracy with RFE =250	98.78%	99.80%	97.94%	98.60%

Table 11 shows average cross-validation accuracy across all four models for machine learning. Cross-validation entails training the model on many data sets and comparing its accuracy on unidentified data to evaluate a model's performance and generalizability. With an impressive 98.44% accuracy rate, the Random Forest model demonstrates its strong effectiveness and capacity for generalization over several data splits. With an accuracy of 99.46%, XGBoost showcases its exceptional capacity to manage complex data and consistently deliver strong results during cross-validation. At 98.25% and 97.93%, respectively, Decision Trees and SVMs have somewhat lower average accuracies. Despite their continued high performance, these models fall short of Random Forest and XGBoost in terms of accuracy and consistency across a wide range of data sets. The results generally show that Random Forest and XGBoost perform better than other ML models in cross-validation, providing higher reliability and accuracy. These models' performance may be explained by using cross-validation (5-fold) to determine the perfect hyperparameters.

Table 11: Average cross-validation accuracy with the use of the CLAHE filter

Model	Result
Random forest	98.44%
XGBoost	99.46%
SVM	97.93%
Decision Tree	98.25%

Table 12 displays the model performance based on multiple classification measurements, including F1 Score, Precision, Recall, and Accuracy. XGBoost is the top performer, with flawless 98.80% Precision, Recall, and F1 scores, demonstrating its extraordinary ability to categorize positive and negative cases properly. Overall, Random Forest and XGBoost excel at classification, particularly in balancing precision and recall, whereas Decision Tree and SVM are dependable but somewhat less successful.

Table 12: Performance metrics of machine learning classification models with RFE = 250 Features and CLAHE filter

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	98.78%	98.78%	98.78%	98.78%
XGBoost	99.80%	99.80%	99.80%	99.80%
SVM	97.94%	97.94%	97.94%	97.94%
Decision tree	98.60%	98.60%	98.60%	98.60%

The LC25000 dataset consists of five distinct tissue classes: Colon Adenocarcinoma (colon_aca), a malignant tumor from glandular cells in the colon, showing irregular and disorganized tissue structure. Colon Benign (colon_n) normal colon tissue with well-organized gland structures and no signs of cancer. Lung adenocarcinoma (lung_aca) is a common lung cancer starting in mucus-producing glands and is usually found in the outer lung areas. Lung Benign (lung_n) healthy lung tissue with normal cell patterns and no cancerous changes. Lung Squamous Cell Carcinoma (lung_scc) lung cancer type arising from squamous cells in the airways, often linked to smoking. Table 13 evaluates classification accuracy before and after feature selection. The performance metrics are expressed in percentages, clearly comparing how feature selection impacts each model's effectiveness across different tissue types.

colon adenocarcinoma (colon_aca)

XGBoost: 99.90%, RF: 98.30%, DT: 99.10%, SVM: 99.90%

This class continues to demonstrate exceptionally high classification accuracy across all models. The consistently strong performance indicates that colon_aca has distinct and easily learnable features. The near-perfect scores suggest that even with different modeling strategies, the class is robustly separable from others.

normal colon (colon_n)

XGBoost: 99.90%, RF: 99.80%, DT: 98.90%, SVM: 100%

This class is also very easily classifiable across all models. The features characterizing normal colon tissue appear well-defined and do not overlap with pathological classes, resulting in uniformly high accuracy post-feature selection.

lung adenocarcinoma (lung_aca)

XGBoost: 99.90%, RF: 98.80%, DT: 97.40%, SVM: 95.50%

This class shows strong performance, particularly with XGBoost and RF. However, SVM lags, which may indicate that this model is less effective at exploiting the refined feature set. The class seems to benefit from sophisticated ensemble methods that can better capture nuanced patterns.

normal lung (lung_n)

XGBoost: 100%, RF: 100%, DT: 99.60%, SVM: 100%

This is the best-performing class overall. The features representing normal lung tissue are highly distinctive, leading to near-perfect or perfect classification across all models. Feature selection likely has minimal impact because this class's inherent separability is already very high.

lung squamous cell carcinoma (lung_scc)

XGBoost: 99.30%, RF: 97.00%, DT: 98.00%, SVM: 94.30%

Among all classes, lung_scc exhibits the most variation in accuracy across models, and the lowest performance with SVM. While XGBoost maintains high accuracy, RF and especially SVM show reduced performance compared to other tissue types. This suggests that lung_scc may be more challenging to represent, potentially due to feature overlap with other lung cancer types or greater intra-class variability, making it more sensitive to the choice of features and model complexity.

Colon tissue types (both normal and cancerous) are consistently well-classified with very high accuracy, indicating that their features are inherently discriminative and stable across different models. Normal lung tissue is also trivially separable, reflecting a strong signal in the feature space. Lung cancer subtypes, particularly lung_scc, benefit most from thoughtful feature representation and model selection, revealing a greater dependence on feature quality and model capacity to capture complex patterns.

Table 13: Model accuracy comparison across All classes on the LC25000 dataset

Class name	XGBoost	RF	DT	SVM
colon_aca	99.90%	98.30%	99.10%	99.90%
colon_n	99.90%	99.80%	98.90%	100%
lung_aca	99.90%	98.80%	97.40%	95.50%
lung_n	100%	100%	99.60%	100%
lung_scc	99.30%	97.00%	98.00%	94.30%

Along with classification accuracy, Figure 10 (a - d) shows the confusion matrix for each model after feature selection in the LC25000 dataset, which provides a detailed view of class-wise prediction performance, misclassifications, and the distribution of true versus predicted labels. The authors of the most recently published research publications employed deep learning to categorize histological images of colon and lung tumors. Table 12 compares the results obtained using the 25,000 images of the LC25000 dataset with those of publications across the literature. The results of our research show that feature engineering accomplishes results comparable to deep learning techniques. Our XGBoost model obtained 99.80% accuracy in classifying colon and lung cancer. This paper aims to show an interest in feature engineering techniques rather than competing with current approaches and illustrate how machine learning models might produce better results. Machine learning techniques

for healthcare image classification are evaluated in Table 14. For example, Chehade et al. [18] used SHAP with XGBoost and other ML models, attaining a 99% accuracy in XGBoost with SHAP explanation. Using Grad-CAM, Vanitha et al. [20] evaluated explanation in an ensemble of Xception and MobileNet models with 99.44% accuracy. Hasan et al. [21] developed (LW-MS-CCN) with 99.20% accuracy using Grad-CAM and SHAP.

Table 14: Accuracy and explanation comparison of models for histopathological Image classification

Machine Learning Model	Results	Explainable artificial intelligence (XAI)	Ref.
XGBoost, SVM, and other ML models trained on the LC25000 dataset	XGBoost Accuracy 99%	SHAP for explainability	[27]
Ensemble of Xception and MobileNet architectures.	Achieved accuracy of 99.44%	Grad-CAM for explainability	[29]
Lightweight Multi-Scale CNN (LW-MS-CCN)	Achieved accuracy of 99.20%	Grad-CAM & SHAP for explainability	[30]
Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), Decision Tree	XGBoost Accuracy 99.80%	LIME & SHAP for explainability	Proposed model

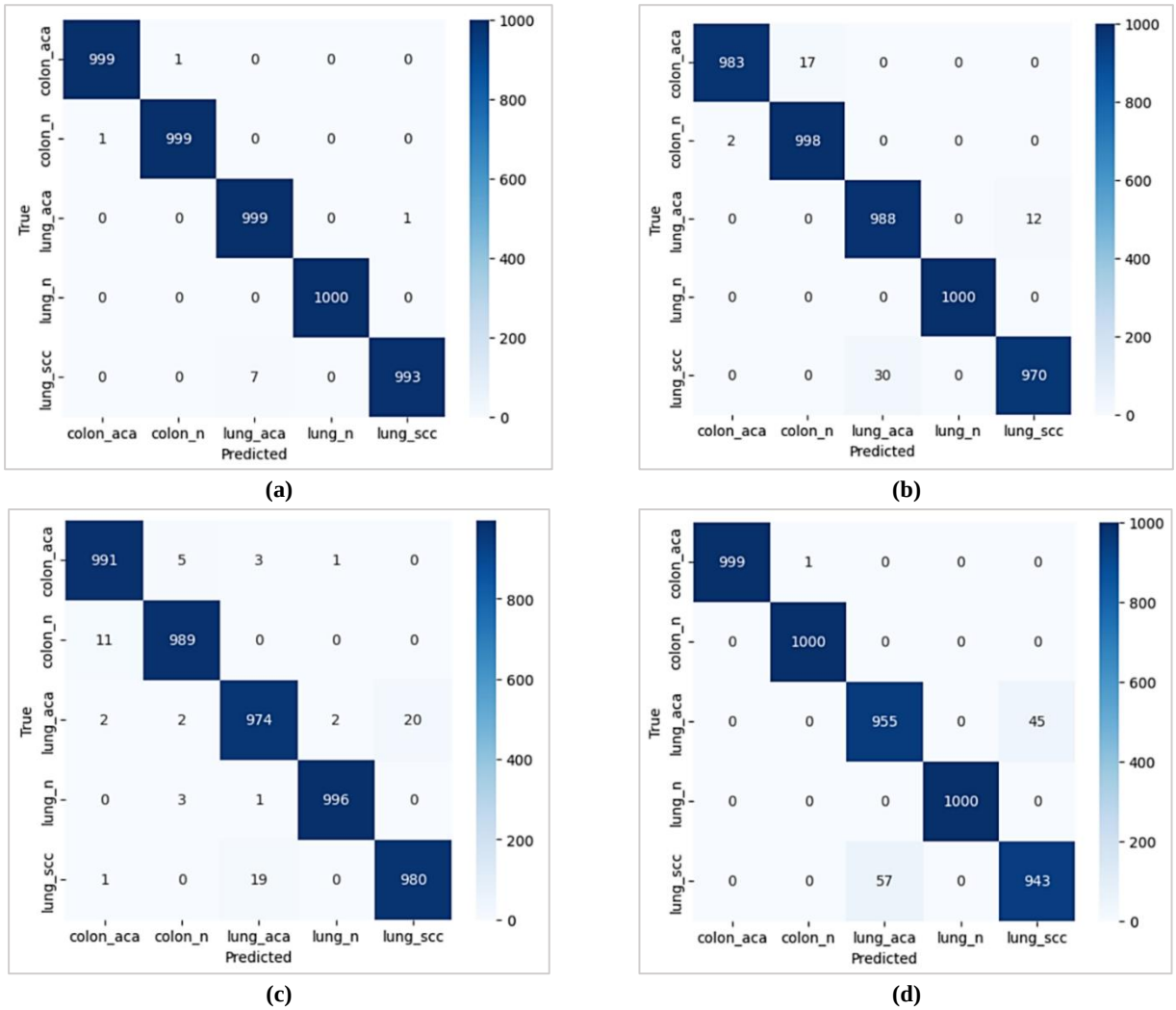

Figure 10: Accuracy of models on LC25000 dataset using confusion matrix (after feature selection) (a) xgboost (b) RF (c) DT (d) SVM

Using XGBoost, the proposed model achieves 99.80% accuracy, surpassing all prior models. Unlike much of the previous research, which has focused primarily on accuracy, this combination sheds light on the importance of both local (LIME) and global (SHAP) factors, resulting in an extremely accurate and interpretable model. This is especially useful for medical applications, as reliability and trust rely on understanding model judgments. Medical practitioners may evaluate forecasts and enhance decision-making processes by promoting openness through combining LIME and SHAP.

SHAP for feature importance enables better model diagnostics, optimized feature sets, and improved clinical interpretability in cancer image classification tasks. Each bar represents a feature (identified by an HSV index). The x-axis shows the mean SHAP value, indicating how much each feature contributes, on average, to the model's final decision. The mean SHAP value for

a feature is calculated by averaging the absolute SHAP values of that feature across all instances in the dataset. SHAP assigns an importance value to each feature for a particular prediction, indicating how much that feature contributes to the final model output. To evaluate overall importance, we compute the average of the absolute values of a feature's SHAP contributions across all samples. This ensures that both positive and negative impacts are considered equally, as we are interested in the magnitude of the contribution, not its direction. Figure 11 shows SHAP bar plots for the dataset LC25000. In the LC25000 dataset, the SHAP bar plot reveals that color features derived from specific HSV bins, such as HSV-475, HSV-460, and HSV-171, dominate the XGBoost model's decision-making. These high-ranking features likely capture important color and tissue colorations that differ across cancer types. Overall, the model relies on a specific subset of HSV features to effectively classify images, making these color-based features highly valuable for tissue discrimination in LC25000. The prominence of HSV features in the dataset validates using color histogram features in the HSV color space as a strong descriptor for histopathology image classification tasks.

The SHAP beeswarm plots provided are a visual explanation tool used to interpret the predictions of the XGBoost classifier for medical image classification tasks. These plots are particularly useful for understanding how individual features (like HSV color components and local binary pattern texture descriptors) contribute to a model's prediction for each class in a dataset. The datasets analyzed are LC25000, which includes histological images of lung and colon tissue. Each plot corresponds to a specific class (e.g., lung_aca, colon_n) and shows how various features impact the model's output for that class.

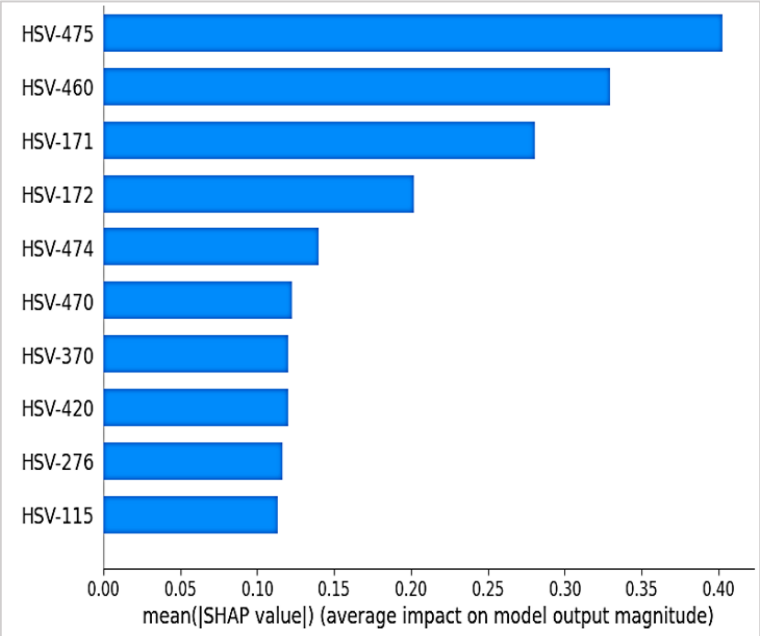

Figure 11: SHAP feature importance bar plots for the LC25000 dataset based on the XGBoost classifier

In each SHAP beeswarm plot, the x-axis represents SHAP values, which quantify the impact of each feature on the prediction. A positive SHAP value means the feature increases the likelihood of the model predicting the given class, while a negative SHAP value indicates that the feature decreases that likelihood. The y-axis lists the feature names, such as HSV-460 or LBP_4. Each dot represents a single sample, with its horizontal position showing the strength and direction of the feature's contribution for that sample. The color of the dots represents the feature value: red dots indicate high feature values and blue dots indicate low feature values

This color coding helps you see whether high or low values of a feature are associated with a stronger impact on the prediction. For instance, if red dots are clustered to the right (positive SHAP values), that feature's high values strongly contribute to the class prediction. Conversely, if blue dots are on the right, low values push the model toward the class prediction. The length and spread of the horizontal lines (the "beeswarm") for each feature show the variability and importance of that feature across all samples:

Longer lines mean the feature has a large and varied impact across the dataset. This indicates that the feature is generally important and influences predictions differently across different samples. Shorter lines mean the feature has a more consistent or smaller impact, possibly playing a minor or more stable role in decision-making. Figure 12 (a – e) shows the SHAP Beeswarm Plot per Class for the LC25000 dataset based on the XGBoost classifier.

In the LC25000 plots:

a) colon_aca class, features like HSV460, HSV017, and HSV440 have the highest SHAP values, indicating they strongly predict adenocarcinoma. High values in these HSV channels positively influence the model's output for this class.
b) colon_n relies on features like HSV360, HSV276, and HSV447, which help the model distinguish normal colon tissue, although the SHAP impact is more mixed, suggesting overlap with other classes.
c) lung_aca class, features such as HSV370, HSV174, and HSV172 significantly contribute to classification, with high values strongly pushing predictions toward adenocarcinoma .

d) lung_n, HSV475, HSV267, and HSV577 dominate, where high feature values help the model identify normal lung tissue accurately.
e) lung_scc class depends on fewer but more distinct features like HSV171 and HSV172, which show a clear influence, suggesting that squamous cell carcinoma has unique visual traits easily picked up by the model.

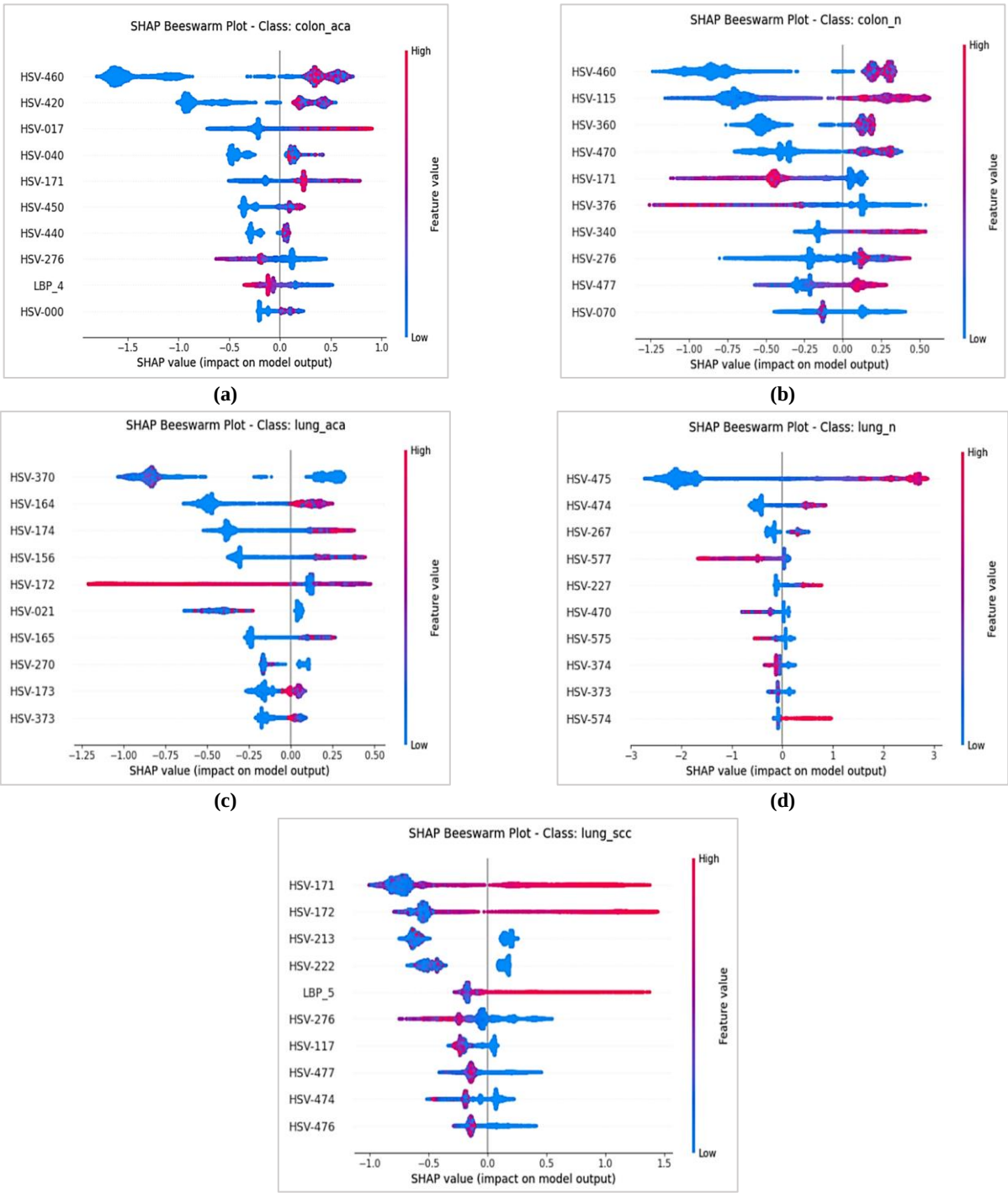

Figure 12: SHAP beeswarm plot per class for the LC25000 dataset based on the XGBoost classifier a) colon adenocarcinoma b)normal colon c) lung adenocarcinoma d) normal lung e) lung squamous cell carcinoma

LIME for tabular data explains a single prediction by showing which features had the most impact. It does this by slightly changing the input and seeing how the model's output changes, making complex models easier to understand. Figure 13 illustrates a LIME (Local Interpretable Model-Agnostic Explanations) explanation that shows how the machine learning model confidently classified the input histopathological image as colon adenocarcinoma (colon_aca) with a predicted probability of 1.00. This type of explanation is particularly valuable in medical diagnostics, as it allows clinicians to interpret the model's decision on a per-sample basis. The visualization highlights the most influential features that contributed to the model's decision, breaking them into those supporting the "colon_aca" prediction and those opposing other potential classifications such as "colon_n" (normal colon), "lung_aca" (lung adenocarcinoma), "lung_n" (normal lung), and "lung_scc" (lung squamous cell carcinoma). The features supporting the "colon_aca" prediction included color histogram (HSV) and texture (LBP) features, such as HSV-171, HSV-017, and LBP_2. Specifically, the condition $-0.35 < \text{HSV-171} \leq 1.04$ was associated with a contribution

weight of 0.15, indicating its significant influence in favoring the colon_aca class. Similarly, HSV-017 > –0.11 and LBP_2 > 0.68 contributed weights of 0.13 and 0.09, respectively.

These features likely capture specific chromatic and structural characteristics of malignant colon tissue, distinct from other classes' features. On the other hand, features such as LBP_6 ≤ –0.67 had a lower contribution weight (0.05), suggesting a mild opposition to the colon_aca classification, potentially because such texture values are less common in adenocarcinoma cases. The right-hand side of the figure presents a tabular view of the actual feature values extracted from the image. For example, HSV-171 had a value of –0.04, HSV-017 measured 0.41, and LBP_2 was significantly high at 1.87, all of which fall within the influential thresholds that supported the colon_aca prediction. This transparency confirms the model's decision and allows medical experts to cross-reference these values with known pathological patterns in cancer diagnosis. Compared to global explanation methods like SHAP, which offer generalized insights across the entire dataset, LIME's local explanation approach focuses on the current image. It yields actionable, interpretable results for that individual prediction. This can be particularly critical in clinical settings, where understanding the "why" behind a model's classification helps build trust and supports more informed decision-making.

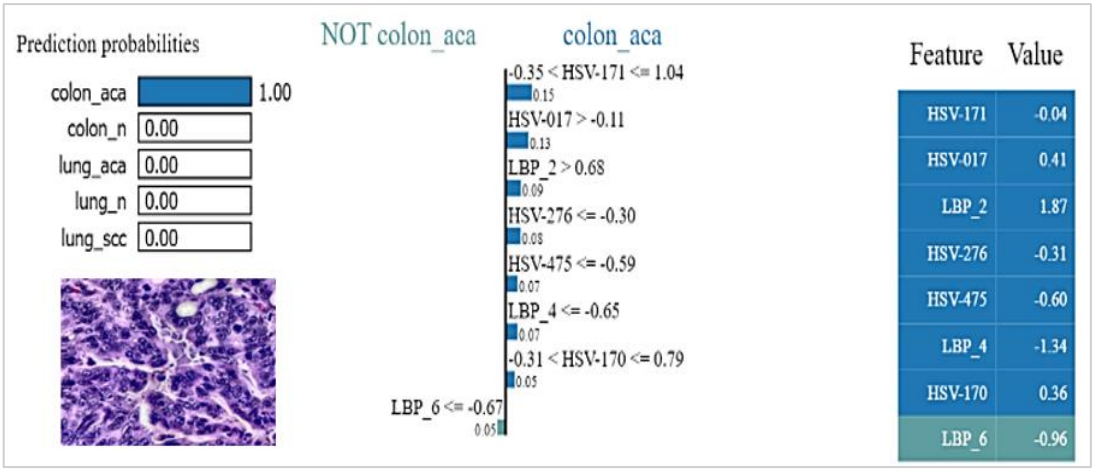

Figure 13: XGBoost model LIME explainability for a single instance (colon_aca)

Figure 14 shows a step-by-step procedure for enhancing, segmenting, and explaining a histopathology image utilizing Contrast Limited Adaptive Histogram Equalization (CLAHE), Superpixel Segmentation, and Local Interpretable Model-Agnostic Explanations. The top-left image displays the original histopathology image without any alterations. Applying CLAHE (top-right) enhances the contrast, making fine histological details more visible. Its superpixel segment (bottom-right) connects similar regions, depicted in yellow, to aid in targeted feature extraction. Lastly, the LIME explanation (bottom-left) highlights the most important points that influence the model's conclusion while excluding the less important ones. This method greatly enhances the understanding and trustworthiness of AI-based histopathology classified images. Superpixel divides guarantee organized feature extraction, the CLAHE stage improves transparency, and LIME offers an unambiguous presentation of model decisions. This method enhances model transparency and helps with crucial diagnostic choices by enabling medical personnel to assess AI predictions. The process guarantees that AI models learn useful patterns rather than worthless noise by concentrating on medically significant fields, improving the reliability of these models for healthcare applications.

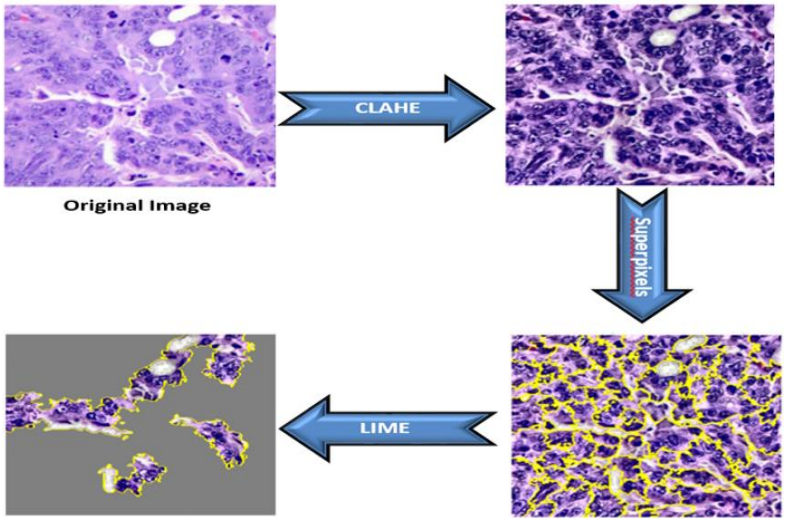

Figure 14: LIME Explanation for Histopathological Image Classification

LIME is a valuable interpretability technique, especially effective in misclassification cases or uncertain predictions. When a model outputs similar probabilities for two or more classes, it indicates ambiguity in decision-making, precisely the scenario

where LIME is most insightful. By providing a localized explanation around a single prediction, LIME allows researchers and practitioners to understand which specific features contributed most to the classification and in which direction. This granular insight is particularly useful in medical image classification tasks, such as distinguishing between histopathological images of lung and colon cancers.

Figure 15 presents a LIME explanation for a lung_aca prediction, offering a visual breakdown of how individual features influenced the model's decision. On the vertical axis, we see the top features expressed as conditional statements involving HSV color channel values, such as "-0.61 < HSV-156 ≤ 0.51"—representing specific pixel value ranges extracted from the image using color-based feature engineering. These features are ordered from top to bottom by their impact magnitude on the model's output. The horizontal axis quantifies each feature's contribution to the prediction, with values ranging from -0.015 to +0.015. Positive values (green bars) indicate features that support the predicted class (lung_aca), whereas negative values (red bars) represent features that push against this class, nudging the model toward lung_scc in this case.

The bar chart shows that "-0.61 < HSV-156 ≤ 0.51" had the strongest positive influence, increasing the confidence in lung_aca prediction by about +0.0145. Other supporting features include "HSV-245 > 0.14" and "HSV-272 > 0.33", which made smaller yet meaningful contributions. On the other hand, "HSV-271 > 0.65", "HSV-376 ≤ -0.42", and "HSV-152 > 0.29" had the most negative impact, with values close to -0.013, pulling the prediction away from lung_aca, most likely toward lung_scc. This tension between supportive and opposing features visually confirms that the model's prediction was made under considerable uncertainty, with small shifts in feature values potentially tipping the outcome toward another class. Thus, the LIME chart reveals what influenced the model and how strongly it influenced it, enabling a transparent interpretation of near-boundary or misclassified cases.

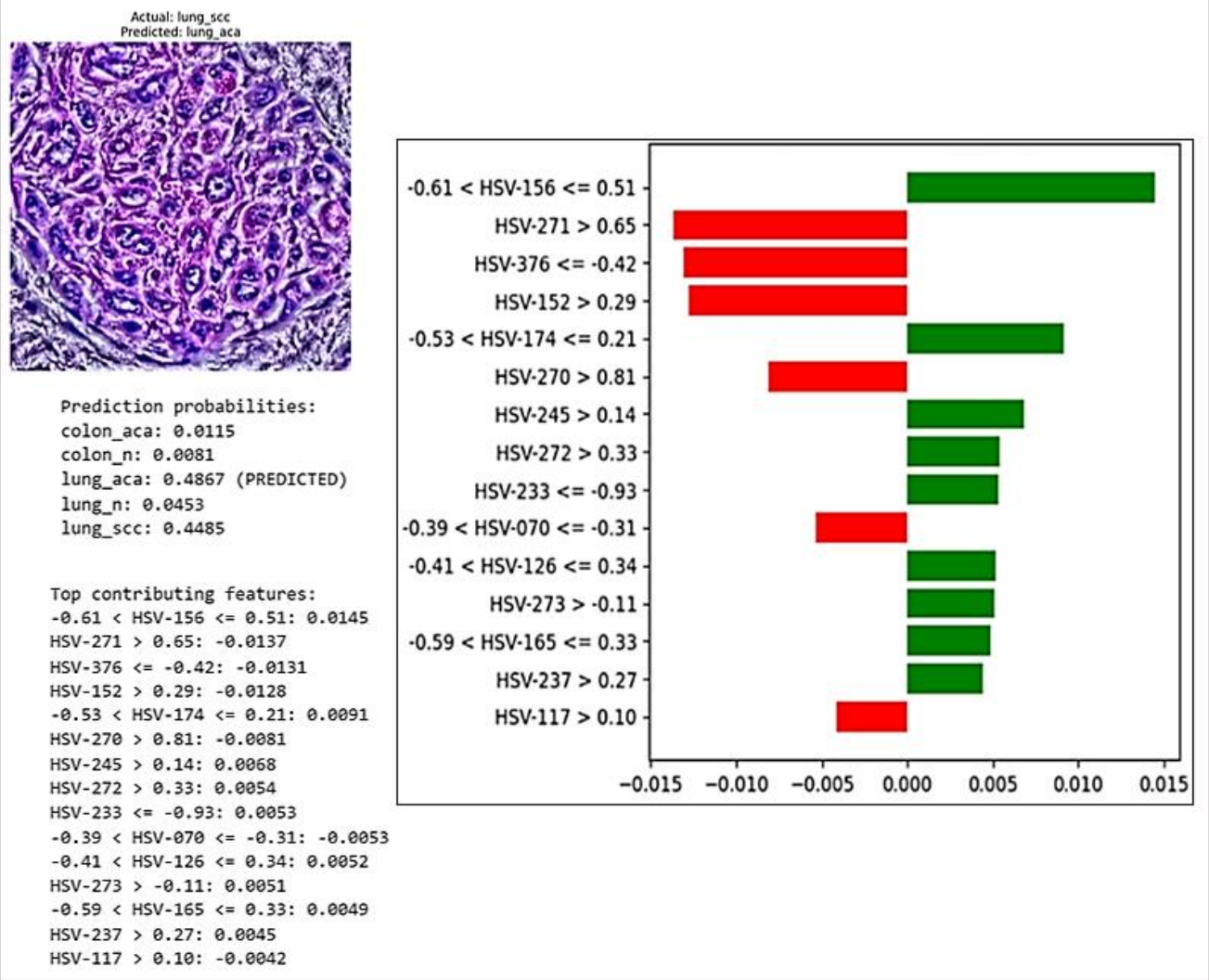

Figure 15: LIME-based feature contributions toward Lung_aca prediction

4. Conclusion

This study presents research that introduces a robust and interpretable machine learning pipeline for classifying lung and colon cancer histopathological images. We achieved highly discriminative feature sets through comprehensive feature engineering using color histograms and Local Binary Patterns (LBP) and refinement via Recursive Feature Elimination (RFE). Among the classifiers evaluated, the XGBoost model consistently outperformed others, reaching an exceptional accuracy of 99.80%. Contrast Limited Adaptive Histogram Equalization (CLAHE) preprocessing significantly improved model performance

across all classifiers. As demonstrated in Table 10, including CLAHE led to improved classification accuracies—for instance, XGBoost accuracy increased from 99.56% to 99.80%, and Random Forest from 98.02% to 98.78% after applying RFE. Table 11 further confirms the benefit with higher cross-validation accuracy for models trained on CLAHE-enhanced images, particularly for XGBoost (99.46%) and Random Forest (98.44%). These improvements indicate that CLAHE enhances image contrast, leading to more discriminative feature extraction, ultimately boosting classification accuracy and reliability.

Explainable AI (XAI) tools such as SHAP and LIME were employed to ensure model transparency and interpretability. SHAP provided global and local insights into feature contributions, enabling us to identify that color and texture features played critical roles in distinguishing benign and malignant tissues. LIME further enhanced interpretability by visually highlighting the specific regions of histopathological images influencing model decisions, thereby offering meaningful support for clinical decision-making. This study validates that integrating advanced feature extraction, RFE-based feature selection, CLAHE preprocessing, and XAI methods significantly enhances AI models' predictive performance and trustworthiness in medical imaging applications.

This study demonstrates that feature selection significantly enhances classification performance for certain tissue types in the LC25000 dataset, particularly for challenging classes like Lung Squamous Cell Carcinoma (lung_scc). While classes such as Colon Adenocarcinoma, Normal Colon, and Normal Lung are inherently easier to classify—with models achieving near-perfect accuracy—classes like lung_aca and especially lung_scc show more sensitivity to feature representation, benefiting from more informative and well-selected features. Confusion matrices were generated for each model to assess classification effectiveness thoroughly, offering insights into model strengths and class-specific weaknesses. These matrices revealed that most misclassifications were concentrated in the lung cancer subclasses, reinforcing the importance of tailored feature selection strategies for such categories.

Additionally, to better understand model decisions and improve interpretability, we employed LIME (Local Interpretable Model-agnostic Explanations). LIME is particularly valuable for analysing misclassified instances, as it identifies and displays the most influential features contributing to a prediction. By visualizing these feature weights, LIME helps uncover potential reasons behind classification errors, providing both model transparency and guidance for further refinement of feature engineering or model tuning. In future research, we plan to expand and enhance the proposed pipeline in several directions:

1) Explore advanced feature extraction techniques such as Haralick texture features, Gabor filters, Zernike moments, and Histogram of Oriented Gradients (HOG), which may offer additional discriminatory power for capturing morphological and textural patterns specific to cancer subtypes.
2) Apply deep learning-based feature extraction through convolutional neural networks (CNNs) to learn hierarchical and spatially rich representations directly from histopathological images. When combined with handcrafted features, this could potentially improve classification performance.
3) Test generalizability by evaluating the proposed framework on various histopathological datasets from different medical centers, scanners, and staining protocols to ensure robustness and real-world applicability.
4) Integrate multi-modal data, such as patient metadata or genomic profiles, alongside histopathological images to develop more holistic and precise cancer diagnostic models.

These future directions aim to improve the clinical readiness of AI-assisted diagnostic tools, ensuring that they not only achieve high accuracy but also earn the trust of medical professionals through transparency, adaptability, and consistent performance.

Author contributions

Conceptualization, **M. Amer** and **Z. Sultani**; data curation, **M. Amer**; formal analysis, **M. Amer**; investigation, **M. Amer**; methodology, **M. Amer**; project administration, **Z. Sultani**; resources, **M. Amer**; software, **M. Amer**; supervision, **Z. Sultani**; validation, **Z. Sultani** and **M. Amer**; visualization, **M. Amer**; writing—original draft preparation, **M. Amer**; writing—review and editing, **Z. Sultani**. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author.

Conflicts of interest

The authors declare that there is no conflict of interest.

References

[1] M. Masud, N. Sikder, A. A. Nahid, A. K. Bairagi, M. A. AlZain**,** A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework, Sensors, 21 (2021) 748. https://doi.org/10.3390/s21030748

[2] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R.L. Siegel, I. Soerjomataram, A. Jemal, Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin., 74 (2024) 229-263. https://doi.org/10.3322/caac.21834

[3] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin., 68 (2018) 394–424. https://doi.org/10.3322/caac.21492

[4] A. Bermúdez, I. Arranz-Salas, S. Mercado, J. A. López-Villodres, V. González, F. Ríus, M. V. Ortega, C. Alba, I. Hierro, D. Bermúdez, Her2-Positive and Microsatellite Instability Status in Gastric Cancer Clinicopathological Implications, Diagnostics, 11 (2021) 944. https://doi.org/10.3390/diagnostics11060944

[5] M. Togaçar, Disease type detection in lung and colon cancer images using the complement approach of inefficient sets, Comput. Biol. Med., 137 (2021) 104827. https://doi.org/10.1016/j.compbiomed.2021.104827

[6] L. F. Sánchez-Peralta, L. Bote-Curiel, A. Picón, F. M. Sánchez-Margallo, J. B. Pagador, Deep learning to find colorectal polyps in colonoscopy: A systematic literature review, Artif. Intell. Med., 108 (2020) 101923. https://doi.org/10.1016/j.artmed.2020.101923.

[7] N. Wijethilake, D. Meedeniya, C. Chitraranjan, I. Perera, M. Islam, H. Ren, Glioma survival analysis empowered with data engineering A survey, IEEE Access, 9 (2021) 43168–43191. https://doi.org/10.1109/ACCESS.2021.3065965.

[8] N. Alangari, M. El Bachir Menai, H. Mathkour, I. Almosallam, Exploring evaluation methods for interpretable machine learning: A survey, Information, 14 (2023) 469. https://doi.org/10.3390/info14080469.

[9] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, C. Mooney, Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review, Appl. Sci., 11 (2021) 5088. https://doi.org/10.3390/app11115088.

[10] J. Xu, P. Yang, S. Xue, B. Sharma, M. Sanchez-Martin, F. Wang, B. Parikh, Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives, Hum. Genet., 138 (2019) 109–124. https://doi.org/10.1007/s00439-019-01970-5

[11] O. Loyola-Gonzalez, Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view, IEEE Access, 7 (2019) 154096–154113. https://doi.org/10.1109/ACCESS.2019.2949286

[12] S. K. Ghosh, A. H. Khandoker, Investigation on explainable machine learning models to predict chronic kidney diseases, Sci. Rep., 14 (2024) 3687. https://doi.org/10.1038/s41598-024-54375-4

[13] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion, 58 (2020) 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[14] A. Adadi, M. Berrada, Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI), IEEE Access, 6 (2018) 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

[15] Z. C. Lipton, The mythos of model interpretability, Queue, 16 (2018) 31–57. https://doi.org/10.1145/3236386.3241340

[16] C. Hu, et al., Application of interpretable machine learning for early prediction of prognosis in acute kidney injury, Comput. Struct. Biotechnol. J., 20 (2022) 2861–2870. https://doi.org/10.1016/j.csbj.2022.06.003

[17] S. M. Lundberg, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell., 2 (2020) 56–67. https://doi.org/10.1038/s42256-019-0138-9

[18] A. Hage Chehade, N. Abdallah, J. M. Marion, M. Oueidat, P. Chauvet, Lung and colon cancer classification using medical imaging: A feature engineering approach, Phys. Eng. Sci. Med., 45 (2022) 729–746. https://doi.org/10.1007/s13246-022-01139-x

[19] M. Al-Jabbar, M. Alshahrani, E. M. Senan, I. A. Ahmed, Histopathological analysis for detecting lung and colon cancer malignancies using hybrid systems with fused features, Bioengineering, 10 (2023) 383. https://doi.org/10.3390/bioengineering10030383

[20] K. Vanitha, S. S. Sree, S. Guluwadi, Deep learning ensemble approach with explainable AI for lung and colon cancer classification using advanced hyperparameter tuning, BMC Med. Inform. Decis. Mak., 24 (2024) 222. https://doi.org/10.1186/s12911-024-02628-7

[21] M. A. Hasan, F. Haque, S. R. Sabuj, H. Sarker, M. O. F. Goni, F. Rahman, M. M. Rashid, An end-to-end lightweight multi-scale CNN for the classification of lung and colon cancer with XAI integration, Technologies, 12 (2024) 56. https://doi.org/10.3390/technologies12040056

[22] A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, S. M. Mastorides, Lung and Colon Cancer Histopathological Image Dataset (LC25000), arXiv, 1912 (2019) 12142. https://arxiv.org/abs/1912.12142

[23] S. A. El-Ghany, M. Azad, M. Elmogy, Robustness fine-tuning deep learning model for cancer diagnosis based on histopathology image analysis, Diagnostics, 13 (2023) 699. https://doi.org/10.3390/diagnostics13040699

[24] S. Tummala, S. Kadry, A. Nadeem, H. T. Rauf, N. Gul, An explainable classification method based on complex scaling in histopathology images for lung and colon cancer, Diagnostics, 13 (2023) 1594. https://doi.org/10.3390/diagnostics13091594

[25] M. Sakli, C. Essid, B. B. Salah, H. Sakli, Flexible framework for lung and colon cancer automated analysis across multiple diagnosis scenarios, Int. J. Adv. Comput. Sci. Appl., 16 (2025). https://dx.doi.org/10.14569/IJACSA.2025.0160258

[26] J. C. M. Dos Santos, G. A. Carrijo, C. F. d. S. De Cardoso, J. C. Ferreira, P. M. Sousa, A. C. Patrocínio, Fundus image quality enhancement for blood vessel detection via a neural network using CLAHE and Wiener filter, Res. Biomed. Eng., 36 (2020) 107–119. https://doi.org/10.1007/s42600-020-00046-y

[27] A. M. Reza, Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement, J. VLSI Signal Process, Syst. Signal Image Video Technol., 38 (2004) 35–44. https://doi.org/10.1023/B:VLSI.0000028532.53893.82

[28] T. Ayyavoo, J. J. Suseela, Illumination pre-processing method for face recognition using 2D DWT and CLAHE, IET Biometrics, 7 (2017) 380–390. https://doi.org/10.1049/iet-bmt.2016.0092

[29] S. Sahu, A. K. Singh, S. P. Ghrera, M. Elhoseny, An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE, Optics & Laser Technol., 110 (2019) 87–98. https://doi.org/10.1016/j.optlastec.2018.06.061

[30] W. N. J. H. W. Yussof, M. Man, R. Umar, A. N. Zulkeflee, E. A. Awalludin, N. Ahmad, Enhancing moon crescent visibility using contrast-limited adaptive histogram equalization and bilateral filtering techniques, J. Telecommun. Inf. Technol., 1 (2022) 3–13. http://dx.doi.org/10.26636/jtit.2022.155721

[31] J. Ahmad, M. Batool, K. Kim, Sustainable wearable system: Human behavior modeling for life-logging activities using K-Ary tree hashing classifier, Sustainability, 12 (2020) 10324. https://doi.org/10.3390/su122410324

[32] I. Nosheen, A. Naseer, A. Jalal, Efficient vehicle detection and tracking using blob detection and Kernelized filter, 2024 5th International Conference on Advancements in Computational Sciences, 2024, 1–8. https://doi.org/10.1109/ICACS60934.2024.10473292

[33] Q. Zhao, J. Yang, H. Liu, Stone images retrieval based on color histogram, IEEE Int. Image Anal. Signal Process., 2009 (2009) 157–161. https://doi.org/10.1109/IASP.2009.5054590

[34] J. Hafner, H. Sawhney, W. Equitz, M. Flickner, W. Niblack, Efficient color histogram indexing for quadratic form distance functions, IEEE Trans. Pattern Anal. Mach. Intell., 17 (1995) 729–736. https://doi.org/10.1109/34.391417

[35] M. J. Swain, D. H. Ballard, Color indexing, Int. J. Comput. Vis., 7 (1991) 11–29. https://doi.org/10.1007/BF00130487

[36] S. Sural, G. Qian, and S. Pramanik, Segmentation and histogram generation using the HSV color space for image retrieval, Proceedings of IEEE International Conference on Image Processing, 2 (2002) 589–592. https://doi.org/10.1109/ICIP.2002.1040019

[37] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognit., 29 (1996) 51–59. https://doi.org/10.1016/0031-3203(95)00067-4

[38] V. E. Staartjes, L. Regli, C. Serra, Machine learning in clinical neuroscience: foundations and applications, Conference proceedings Machine Learning in Clinical Neuroscience Foundations and Applications, 2017. http://dx.doi.org/10.1007/978-3-030-85292-4

[39] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn., 46 (2002) 389–422. https://doi.org/10.1023/A:1012487302797

[40] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res., 3 (2003) 1157. http://dx.doi.org/10.1162/153244303322753616

[41] C. A. Ramezan, Transferability of recursive feature elimination (RFE)-derived feature sets for support vector machine land cover classification, Remote Sens., 14 (2022) 6218. https://doi.org/10.3390/rs14246218

[42] N. K. Ahmed, A. F. Atiya, N. E. Gayar, H. El-Shishiny, An empirical comparison of machine learning models for time series forecasting, Econometric Rev., 29 (2010) 594–621. https://doi.org/10.1080/07474938.2010.481556

[43] L. U. Ying, Decision tree methods: applications for classification and prediction, Shanghai Arch. Psychiatry, 27 (2015) 130. https://doi.org/10.11919/j.issn.1002-0829.215044

[44] B. Charbuty, A. Abdulazeez, Classification based on decision tree algorithm for machine learning, J. Appl. Sci. Technol. Trends, 2 (2021) 20–28. https://doi.org/10.38094/jastt20165

[45] C. Gold, P. Sollich, Model selection for support vector machine classification, Neurocomputing, 55 (2003) 221–249. https://doi.org/10.1016/S0925-2312(03)00375-8

[46] L. Breiman, Random forests, Mach. Learn., 45 (2001) 5–32. https://doi.org/10.1023/A:1010933404324

[47] T. M. Oshiro, P. S. Perez, J. A. Baranauskas, How many trees in a random forest?, Mach. Learn. Data Min. Pattern Recognit., 8 (2012) 154–168. https://doi.org/10.1007/978-3-642-31537-4_13

[48] M. Shepovalov, V. Akella, FPGA and GPU-based acceleration of ML workloads on Amazon cloud – A case study using gradient boosted decision tree library, Integration, 70 (2020) 1–9. https://doi.org/10.1016/j.vlsi.2019.09.007

[49] C. W. Wang, Y. C. Lee, E. Calista, F. Zhou, H. Zhu, R. Suzuki, D. Komura, S. Ishikawa, S.-P. Cheng, A benchmark for comparing precision medicine methods in thyroid cancer diagnosis using tissue microarrays, Bioinformatics, 34 (2017) 1767–1773. https://doi.org/10.1093/bioinformatics/btx838

[50] N. Nusrat, S. B. Jang, A comparison of regularization techniques in deep neural networks, Symmetry, 10 (2018) 648. https://doi.org/10.3390/sym10110648

[51] S. Das, M. Sultana, S. Bhattacharya, D. Sengupta, D. De, XAI–reduct: accuracy preservation despite dimensionality reduction for heart disease classification using explainable AI, J. Supercomput., 79 (2023) 18167–18197. https://doi.org/10.1007/s11227-023-05356-3

[52] J. An, Y. Zhang, I. Joe, Specific-input LIME explanations for tabular data based on deep learning models, Appl. Sci., 13 (2023) 8782. https://doi.org/10.3390/app13158782