



A deep-XAI method for histopathological image classification: Utilizing transformer based FNet architecture and LIME method

Histopatolojik görüntü sınıflandırması için açıklanabilir derin öğrenme yöntemi: Transformer tabanlı FNet mimarisi ve LIME metodunun kullanımı

Delal ŞEKER^{1*}, Aşlı AKHAN², Abdulnasır YILDIZ¹

¹Department of Electrical and Electronics Engineering, Dicle University, Diyarbakır, Türkiye.
delalkabak93@gmail.com, abnayil@dicle.edu.tr

²Selahaddin Eyyubi State Hospital, Diyarbakır, Türkiye.
dr.asli.unlu@gmail.com

Received/Geliş Tarihi: 29.05.2024
Accepted/Kabul Tarihi: 12.03.2025

Revision/Düzeltilme Tarihi: 24.02.2025

doi: 10.5505/pajes.2025.98572
Research Article/Araştırma Makalesi

Abstract

The purpose of the study is to improve the cancer detection in medical images using the Fourier Net (FNet) architecture and the Local Interpretable Model-agnostic Explanations (LIME) method. The FNet architecture excels in extracting features from high-dimensional images and anatomical representations. LIME, on the other hand, is an algorithm to make the model's decisions interpretable. After applying the FNet architecture to the existing data, the LIME explainability method has been applied to determine whether the model outputs meaningful results from the image. Using deep learning techniques, the proposed algorithm represents cancer types with distinctive and robust features. An additional assessment by an expert pathologist was carried out to prove the results obtained after the LIME interpretation. Thus, medical professionals and researchers will be able to evaluate whether this method developed using FNet and LIME can provide a more interpretable and effective approach to cancer diagnosis. The proposed study lays the foundation for developing effective systems that assist doctors and pathologists in evaluating histopathological tissue images. Additionally, this study aims to enhance the reliability of machine learning methods.

Keywords: FNet, explainable artificial intelligence, lung cancer, colon cancer, pathological assessment

Öz

Çalışmanın amacı, Fourier Net (FNet) mimarisi ve Yerel Yorumlanabilir Model-Agnostik Açıklamalar (LIME) yöntemi kullanılarak tıbbi görüntülerde kanser tespiti geliştirmektir. FNet mimarisi, yüksek boyutlu görüntülerden ve anatomik temsillerden öne çıkan özellikleri başarıyla çıkarma yeteneğine sahiptir. Öte yandan, LIME, modelin kararlarını yorumlanabilir hale getirmek için kullanılan bir algoritmadır. Mevcut verilere FNet mimarisi uygulandıktan sonra, modelin görüntülerden anlamlı sonuçlar üretip üretmediğini değerlendirmek amacıyla LIME açıklanabilirlik yöntemi uygulanmıştır. Önerilen algoritma, derin öğrenme teknikleri kullanılarak kanser türlerini belirgin ve güçlü özelliklerle temsil etmektedir. Ayrıca, LIME yorumlamasından elde edilen sonuçların doğruluğunu kanıtlamak amacıyla uzman bir patoloj tarafından ek bir değerlendirme gerçekleştirilmiştir. Bu sayede, tıp uzmanları ve araştırmacılar, FNet ve LIME kullanılarak geliştirilen bu yöntemin kanser teşhisinde daha yorumlanabilir ve etkili bir yaklaşım sağlayıp sağlamadığını değerlendirebileceklerdir. Önerilen çalışma, doktorlar ve patolojların histopatolojik doku görüntülerini değerlendirmesine yardımcı olacak etkili sistemlerin geliştirilmesi için bir temel oluşturmaktadır. Ek olarak, bu çalışma makine öğrenimi yöntemlerinin güvenilirliğini artırmayı amaçlamaktadır.

Anahtar kelimeler: FNet, açıklanabilir yapay zeka, akciğer kanseri, kolon kanseri, patolojik değerlendirme

1 Introduction

Globally, cancer is the second leading cause of mortality, following cardiovascular diseases [1]. According to the 2022 WHO reports, approximately 10 million people died from cancer. Lung cancer is the leading cause of cancer-related mortality, accounting for 1.817.469 fatalities. Colon cancer ranks fifth in mortality, with 538.167 deaths reported. In the male population, colon and lung cancers have a higher prevalence [2]. The organism consists of countless cells that proliferate and reproduce through division. Certain cell populations may become defective or are replaced by normal cells after a certain age. However, errors occurring during cell division process or genetic mutations can lead to uncontrolled cell proliferation. This uncontrolled proliferation disrupts the body's normal functioning, leading to development of a disease known as cancer. Unlike healthy cells, cancer cells can proliferate without responding to growth signals, fail to

undergo apoptosis when needed, and thus contribute to tumor formation[3]. Colon & lung cancer exhibit distinct symptoms and causes. These cancers are among the most prevalent malignancies and can sometimes develop simultaneously [4]. In a study spanning 76 months, the medical care and outcomes of inmates diagnosed with simultaneous colon & lung cancers were investigated. Among 3102 patients diagnosed with lung cancer, 17 individuals received a diagnosis of colon cancer within one month. Significantly, chronic smoking and overweight were not identified as particular aspects among the inmates. While the simultaneous occurrence of both cancers is rare, especially in incidentally diagnosed cases, it is crucial to appropriately manage such patients appropriately [5]. Cancer patients may not show symptoms at the initial phase of the disease or exhibit only a few symptoms, but when symptoms fully manifest, the lesion has progressed significantly, indicating some delays in treatment [6]. Molecular imaging technologies, including MRI, PET and X-ray are capable of early

*Corresponding author/Yazışılan Yazar

diagnosis and staging [7]. However, chest X-rays haven't been studied as a screening test for lung cancer in high risk individuals [8]. Furthermore, diagnosing cancer cases can be time-consuming and subjective, particularly in the early stages, among doctors. Early detection and diagnosis methods can help overcome these challenges [7]. Another approach used in monitoring to especially lung nodule development is based on the analysis of multiple images [9]. Artificial intelligence techniques play a crucial role in the early diagnosis of biomedical images, predicting disorders, and identifying serious health issues [10].

When considering the findings in the literature, it is evident that there are various approaches for the evaluation and classification of histopathological tissue images as follows: in their study, Mangal et al. (2020) utilized a convolutional-based diagnostic system to evaluate colon and lung cancer data. This shallow CNN model classified the 3-classes lung data as 97.2% and the 2-classes colon data as 96% [11]. Masud et al. (2021), aimed to automate cancer diagnosis to save time and costs by employing a hybrid approach based on deep learning. In this convolution-based model, they achieved a rate of 96.33% accuracy [12]. Adu et al. (2021) introduced a capsule-based method for the automatic detection and classification of colon & lung cancers. This research developed a novel hybrid feature technique utilizing encoders and a new horizontal compression function to suppress the feature set. The recommended methodology yielded an accuracy of 99.23% [13]. On the other hand, Ali & Ali (2021) developed a hybrid model combining traditional layers, separable convolution layers, and capsule networks to discriminate colon and lung cancers. In the pre-processing stage, they applied gamma correction, color balancing, image sharpening, and multiscale merging. The current model utilizes two inputs: raw images processed through standard convolutional layers in the first input block and preprocessed images through separable convolution layers in the second block, both fed into a capsule network for classification. Experimental outcomes yielded that the proposed model accomplished meaningful success across all classes within a 99.58% overall accuracy for colon and lung cancer [14]. Ibrahim & Talaat (2022) reported a deep learning-based approach utilizing EfficientNetB7 for feature extraction and Modified Neural Networks (MNNs) for the multi-way classification of colon and lung cancers. Their results indicated that the system achieved an accuracy of 99.5% [15]. Meanwhile, Talukder et al. (2022), introduced a hybrid method for identifying colon and lung cancers. This approach used MobileNet to extract features and an ensemble soft voting classifier. The simulation results demonstrated that this technique achieved a 99.3% accuracy in identifying colon and lung cancers [16]. In their study, Raju & Rau (2022) utilized visualization techniques, GradCam and SmoothGrad, to better classify histopathological images and enhance their quality. They classified five different colon and lung tissue histopathological images using MobileNetV2 and InceptionResnetV2 architecture. This suggested framework accurately identified cancer tissues up to 99.95%. Their aim was to help healthcare providers in developing a computerized and robust system to detect many types of colon & lung cancers [17]. Kumar et al. (2022) compare handcrafted and deep learning-based feature extraction methods for lung and colon cancer classification. Handcrafted features and deep features from seven transfer learning models were evaluated using Gradient Boosting (GB), SVM-RBF, Multi-Layer Perceptron (MLP), and Random Forest (RF) classifiers. Deep features

significantly outperformed handcrafted features, with DenseNet-121 and RF achieving 98.60% accuracy, 98.63% precision, 0.985 F1-score, and 1.0 ROC-AUC [18]. In their study, Wahid et al. (2023) compared their proposed simple customized CNN model with three different pre-trained CNN models, namely GoogLeNet, ResNet18 and ShuffleNet V2. The evaluation metrics provided the highest accuracy with ResNet18 for classifying lung cancer and with ShuffleNet V2 for evaluating colon data, achieving 99.87% accuracy. Their proposed customized CNN model achieved accuracy of 93.02% for lung data and 88.26% for colon data. Moreover, they demonstrated that their proposed CNN model had a shorter training time compared to GoogLeNet and ResNet18 [19]. Tummala et al. (2023) achieved 99.97% an accuracy on the test set for 5-classes classification based on colon & lung data using 5-fold cross-validated and tested EfficientNetV2 models. They also strengthened their study using Saliency maps [20]. Mehmood et al. (2022) proposed a transfer learning-driven approach to automatically categorize colon and lung cancer pathological images. In this technique, they applied the contrast enhancement method not to all images but only to the class with low performance to reduce computational cost and increase accuracy (increased from 89% to 98.4%)[21]. Fan et al. (2021) proposed a transfer learning architecture incorporating softmax and SVM for categorizing colon and lung cancers. To improve classifying accuracy, the fully connected layer of the softmax classifier was input into the SVM classifier. Their findings showed that the proposed model achieved an accuracy of 99.4% on the LC25000 data [22]. Aitazaz et al. (2023) evaluated the diagnostic performance of Vision Transformer (ViT) vs. pre-trained CNN architectures for colon and lung images. The study's findings revealed that the ViT architecture outperformed the pre-trained CNNs. Moreover, the results demonstrated that the ViT-L32 model, which has more layers, achieved higher accuracy compared to the ViT-B32 model [23]. Ahmet et al. (2023) present an interpretable lung cancer diagnostic system utilizing various ML models including Random Forest, Decision Tree, Logistic Regression, and Naive Bayes classifier. The research is conducted using the same dataset of lung tissue images. Random Forest and Logistic Regression classifiers emerged with the highest accuracy, reaching 97% in predicting lung cancer among the diverse ML models explored. Additionally, the study employs SHapley Additive exPlanations (SHAP) and LIME to demonstrate the explainability of the employed ML structures [24].

Considering the aforementioned studies, most of the research is built on convolutional based deep architectures and conventional ML algorithms for histopathological image assessment and classification. A few studies combine their results with explainable methods. In this study, the transformer-based Fourier Net (FNet) architecture is utilized. FNet is a transformer architecture that consists of a Fourier sub-layer followed by a feedforward sub-layer in each layer, without attention layers [25]. With this transformer based deep learning model, it is observed that a large number of images can be trained with few parameters, resulting in a fast and effective outcome. An additional ViT model is utilized considering same hyper-parameters with FNet so that we will validate our results with different transformer based network. To strengthen the study, the effective classification of the FNet model outputs are explained using the LIME model. According to our knowledge, none of the previous studies have implemented Fourier Transforms instead of the attention mechanism in Transformer block to carry out an effective classification task for current

dataset Additive explainable methods (i.e. LIME) are gaining popularity to elaborate the black-box behavior of deep layered architectures. Additionally, images classified as false negatives and false positives, obtained through the confusion matrix, are re-interpreted with the assistance of an expert pathologist. The FNet and ViT architectures will classify 3 classes of lung tissue images, 2 classes of colon histopathological images and 2 classes of microscopic pathological breast cancer dataset. LIME will test a pathology image, which is desired to be analyzed by the trained machine learning algorithm, and will provide insight into which parts of the image the algorithm extracts the most features from. Following this process, the validation aspect of the relevant algorithm will be observed by comparing the regions drawn with LIME with the regions evaluated by the expert pathologist.

The rest of the paper is organized as follows: a brief data description will be given in the Materials section. The working mechanism of the FNet deep learning network and LIME explainer method will be explained with details in Method section. Moreover, required pre-processing steps, and calculation of performance metrics will be illustrated. Result and discussion section will put forward the crucial outcomes of the current study by comparison of literature studies. Finally, the current paper will be concluded with a brief conclusion section

2 Material and methods

In the current study, histopathological tissue images of the lung with three different classes and colon including 2 different classes are labeled by reading them in a computer environment. The images are resized to smaller dimensions to improve the performance of the processing machine (Step 1). Subsequently, the labeled images are split into training and test sets before the classification stage (Step 2). The labeled data belonging to relevant classes were passed through the classification stage using the proposed deep learning architecture with an interpretable approach (Step 3), and the created algorithm was

tested with performance metrics (Step 4). Finally, an expert physician interpretation for regions sketched by LIME has been performed (Step 5). All the mentioned steps are provided in a flowchart diagram in Figure 1.

2.1 Materials

In this research, a dataset named LC25000, containing histopathological images of colon and lung cancer from the Kaggle website, is employed to assess the proposed method [26]. Current dataset, created by Andrew Borkowski and his team at James A. Haley Veterans' Hospital in Tampa, Florida, contains 25,000 images, including two types of colon cancer and three types of lung cancer. The original image dimensions are (1024,768) pixels, but the images have been resized by the data owners to (768,768) pixels. The dataset contains a total of 750 lung and 500 colon histopathological images. Through data augmentation methods (i.e. right and left rotations, horizontal and vertical flips), a total of 25000 images have been generated. The total space occupied by the entire dataset is 1.85 GB. To provide a balanced dataset, the images are evenly distributed among the five types, with 5000 images allocated to each type. The focus of this study's analysis will be on 3 classes for lung cancer lung benign tissue, lung adenocarcinoma, and lung squamous cell carcinoma; 2 classes for colon cancer benign tissue and colon adenocarcinoma. This distribution allows for comprehensive examination and comparison across different cancer types, facilitating a more nuanced understanding of diagnostic patterns and potential treatment strategies. Lung adenocarcinoma arises in glandular cells in the lungs and constitutes over 40% of lung cancers. Lung squamous cell carcinoma, on the other hand, originates in the bronchi of the lungs and accounts for over 30% of lung cancers, making it the second most common type [27]. The residual category is benign, remaining localized without metastasizing to other bodily regions [28].

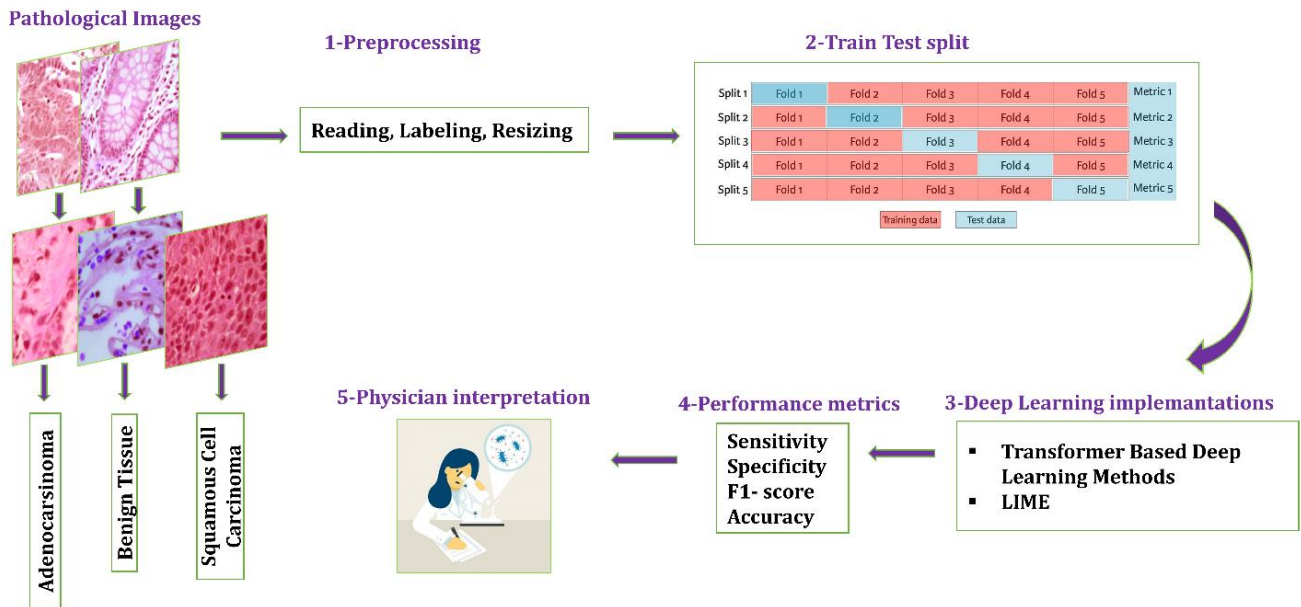


Figure 1. The steps followed in the proposed study

Colon adenocarcinoma is a type of cancer that begins in the cells lining the inner surface of the colon. It is the most common kind of colon cancer, explaining most instances. Adenocarcinomas are typically developed from adenomatous polyps, which are precancerous growths in the colon [29]. Colon benign tissue refers to non-cancerous or non-malignant tissue in the colon. This tissue does not exhibit abnormal growth or cancerous characteristics. In this study, a breast cancer dataset is also additionally used to evaluate the accuracy of the proposed model. The BreakHis dataset, developed by Spanhol et al., consists of all slides obtained from a total of 82 patients. This dataset is not augmented and artificially generated. All images are unique. This dataset has been divided into patches, resulting in a two-classes dataset containing 2,479 benign and 5,304 malignant samples. The proposed model is also applied to this breast cancer dataset, and the results are compared [30].

2.2 A modern multilayer perceptron for image classification: Fourier Net

Transformer-based attention mechanism is a deep learning model that has achieved significant success, particularly in Natural Language Processing (NLP). It was initially introduced in the study titled "Attention is All You Need" by Vaswani et al. (2017) [31]. Transformers have emerged as a prominent structure to overcome many limitations in previous competitive models. This model has been effectively utilized in tasks of language understanding and many other sequential datasets. The attention mechanism is a crucial structure that forms the foundation of the transformer-based model. This mechanism allows measuring and weighting the relationship of each element in an input sequence with all other elements. In other words, each input element is associated with a weight that determines how much attention should be paid to other elements. Thus, the contribution of each element to the information can be measured based on the content of other elements.

The attention mechanism provides more effective information processing, especially when working with long and interconnected sequences. Additionally, this mechanism enables parallel processing due to its ability to focus on different regions of the input, speeding up both training and usage of the model [32]. The transformer architecture, utilizing this attention mechanism, provides a structure capable of performing language understanding tasks more effectively and in parallel compared to previous models. In their study, Lee-Thorp et al. (2021) investigated whether specific token mixing mechanisms could be used to simplify the relatively complex self-attention layers in the transformer encoder architecture [25]. Initially, they substituted the attention sub-layer with two parameterized matrix multiplications: one for mixing the row dimension and the other for mixing the hidden dimension. Encouraged by the promising results of this straightforward linear mixing scheme, they explored the effectiveness of faster linear transformations. Surprisingly, the Fourier Transform, despite having no parameters, scaled exceptionally well for long inputs, especially on GPUs due to the Fast Fourier Transform (FFT). This method achieved performance nearly equivalent to more complex linear mixing techniques.

The FNet model is based on the non-parameterized Fourier Transform. The FNet model uses a block very similar to the Transformer block. However, FNet changes the self-attention layer in the Transformer block with a non-parameterized two-dimensional Fourier transform layer. One dimensional Fourier Transform is applied across patches, and another one-

dimensional Fourier Transform is applied across channels. The Fourier Transform is a mathematical operation that breaks down a function into its individual frequency components [33]. When applied to a sequence $\{x_n\}$ where $n \in [0, N-1]$, DFT is formally defined as in Equation 1:

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{-2\pi}{N}nk}, 0 \leq k \leq N-1 \quad (1)$$

This method involves directly applying the Discrete Fourier Transform (DFT) matrix to the input sequence. The DFT matrix, denoted as W , which is a Vandermonde matrix constructed for the roots of unity with an additional normalization factor as given in Equation 2:

$$W_{nk} = (e^{\frac{-2\pi}{N}nk} / \sqrt{N}) \quad (2)$$

, in each $n, k = 0, \dots, N-1$. This matrix product operation has a computational complexity of $O(N^2)$, which is higher than the FFT. However, it has been observed to be faster for relatively shorter sequences on Tensor Processing Units (TPUs).

In current study, the hyper-parameters for the FNet architecture are selected based on experimental trials and preliminary testing, with the goal of achieving an optimal balance between model performance and training efficiency. The selected hyper-parameters are utilized based on initial experiments to ensure robust training while mitigating the risk of overfitting. The details of the hyper-parameters are as follows:

Weight Decay of 0.0001: This value is chosen to prevent overfitting by applying L2 regularization to the model's weights.

Batch Size: A batch size of 128 is selected considering memory constraints and the ability of achieving stable gradient updates during training.

Number of Epochs: A total of 50 epochs is used, providing sufficient time for the model to converge based on the learning curves observed during initial trials.

Dropout Rate: A dropout rate of 0.2 is applied to prevent overfitting, and this value demonstrated effective performance for the task at hand.

Image Size: Input images are resized to 64x64 pixels to strike a balance between computational efficiency and the retention of relevant features for classification.

Patch Size: The patch size is set to 8x8 to capture finer details in the image while maintaining a reasonable model size.

Number of Patches with 64: The number of patches is determined based on the image size and patch size to ensure sufficient representation of the spatial information in the image.

Embedding Dimension: An embedding dimension of 256 is selected based on empirical testing to provide adequate capacity for feature representation while maintaining computational efficiency.

Number of Block: The number of blocks is determined through experimentation, where four blocks provided optimal trade-off between performance and computational cost. This value determines how many FNet layer (Fourier transform-based processing blocks) would be used in the FNet architecture. This is a hyper-parameter affecting the depth of the model.

The selection of these hyper-parameters is guided by monitoring the model's performance on the validation set and observing the learning curves generated from the initial trials. This approach ensured that the most convenient combination of hyper-parameters is chosen for robust model training and enhanced generalization.

As highlighted in the FNet study, the model's performance can be further enhanced by increasing the embedding dimensions, increasing the number of FNet blocks, and training the model within more epochs. Additionally, experimenting with larger input image sizes and varying patch sizes may also lead to improved results. The FNet architecture is known to scale efficiently with long inputs, operate significantly faster than attention-based Transformer models, and deliver competitive accuracy results, making it suitable for large-scale applications [25].

The illustration of a FNet architecture with an N encoder block can be seen in Figure 2. In essence, the self-attention sublayer in each Transformer encoder layer is replaced with a Fourier sublayer. This Fourier sublayer implements a two-dimensional Discrete Fourier Transform (DFT) on the embedding input, which is characterized by its dimensions, mainly sequence length and hidden dimension. Specifically, it performs one dimensional DFT along the sequence dimension denoted in Equation 3:

$$y = R(F_{seq}(F_h(x))) \quad (3)$$

,where F_{seq} and another one dimensional DFT along the hidden dimension are represented as F_h .

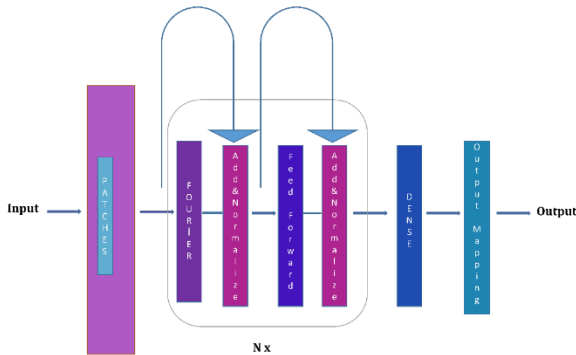


Figure 2. The internal blocks of FNet architecture

2.3 Local interpretable model-agnostic explanations (LIME)

Deep learning (DL) models often attract attention due to their effective performance on high dimensional data despite their complex structures. However, these models are often referred to as "black box" models [34]. This term indicates the difficulty in understanding the internal workings and decision processes of DL models. This situation limits the use of DL in critical applications. Explainable models are employed to understand and make transparent the internal workings of DL algorithms. In this study, following the implementation of the FNet model, one of the explainable models, the LIME method, is used. The LIME method interprets decisions based on a specific set of examples. It adopts a model agnostic approach and is compatible with any machine learning model, which makes it commonly used as a basic explainer [35] as illustrated in Equation (4),

$$Explanation(x) = \underset{g \in G}{\operatorname{argmin}} L(f, g, \pi_x) + \Omega(g) \quad (4)$$

, where x represents the example to be explained, g is linear regression or decision tree as a surrogate model, f is primary model making predictions like the random forest, π_x is proximity metric for the example x , L is function minimizing the loss function such as the least mean square error and finally $\Omega(g)$ represents the function used to keep the feature count of the surrogate model low. However, in practice, the feature count is often determined by the expert.

2.4 Preprocessing steps for data preparation

The images in the current dataset have been resized to (128, 128, 3), and the pixel values, originally ranging from 0 to 255, have been normalized to the range [0, 1]. The dataset is split into 5 folds using the train-test split approach. For each fold, 20% of the remaining training set was selected as the validation set. During training, the training and validation sets are used to observe the loss and accuracy values across epochs. The splitting is performed in a way that ensures an equal number of samples from each class using the 'Stratified' function from the 'scikit-learn' library.

2.5 Performance metrics for diagnostic metrics

Performance metrics including precision, recall, F1-score, and accuracy values are measured. Precision quantifies the fraction of correctly predicted positive cases among all cases predicted as positive. The precision is calculated by dividing the number of true positives (TP) by the sum of true positives and false positives (FP). Recall, also known as sensitivity or true positive rate (TPR), assesses the proportion of correctly predicted positive cases out of all actual positive cases. It is computed by dividing the number of true positives (TP) by the sum of TPs and false negatives (FN). The F1-score represents the harmonic mean of precision and recall, providing a single metric that balances both precision and recall. Accuracy measures the overall correctness of the model's predictions, calculated as the ratio of correct predictions to total predictions. The summary of performance metrics calculation can be found in Table 1.

Table 1. Evaluation of performance metrics

Precision		Recall	
TP		TP	
TP + FP		TP + FN	
F1-score		Accuracy	
2x Precision x Recall		TP + TN	
Precision + Recall		TP + FP + FN + TN	

A confusion matrix is a tool utilized to assess the effectiveness of a classification model. It compares the actual classes with the classes predicted by the model, showing correct and incorrect predictions. It typically includes four main terms: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The metrics for benign tissue (BT, BNG), adenocarcinoma (ADC), and squamous cell carcinoma (SCC) for the three-classes lung cancer data have been calculated as shown in Table 2.

Additionally, the metrics for benign tissue (BT, BNG) and adenocarcinoma (ADC) for the two-classes colon cancer data have been written according to Table 3.

Table 2. Evaluation of lung cancer performance metrics

	Predicted				
	BT	ADC	SCC	FN	TN
Observed	TP for BT:			BT-ADC+	ADC-ADC+
	BT	BT-BT	BT-ADC	BT-SCC	ADC-SCC+
				BT-SCC	SCC-ADC+
					SCC-SCC
	TP for ADC:			ADC-BT+	BT-BT+
	ADC	ADC-BT	ADC-ADC	ADC-SCC	BT-SCC+
				ADC-SCC	SCC-BT+
					SCC-SCC
	TP for SCC:			SCC-BT+	BT-ADC+
	SCC	SCC-ADC	SCC-ADC	SCC-SCC	ADC-BT+
					ADC-ADC
	FP	ADC-BT+	BT-ADC+	BT-SCC+	
		SCC-BT	SCC-ADC	ADC+SCC	

Table 3. Evaluation of colon cancer performance metrics

	Predicted		
	ADC	TP	FP
Observed	BT	FN	TN

Training accuracy represents the ratio of correct predictions on the training dataset by a model. Validation accuracy, on the other hand, measures the performance of the model on the validation dataset. Both training and validation accuracies are important metrics for assessing the overall classification ability of a model. Training loss is a metric that measures the

performance of a model on the training dataset. Validation loss, on the other hand, evaluates the performance of the model on the validation dataset. Loss quantifies how far the predictions of the model are from the actual values. Lower loss values may indicate better model performance. Loss is typically calculated using the following formula:

$$Loss = \frac{1}{N} \sum_{i=1}^N Error(y_i, y'_i) \quad (5)$$

, where N represents the number of samples, y_i denotes the actual value, and y'_i refers to the predicted value. In a well-trained machine learning architecture, the accuracy value of the training-validation set should consistently increase over epochs, while the loss values of the training-validation set should consistently decrease. In the current study, the FNet architecture was trained along 50 epochs.

3 Results and discussion

The FNet architecture has been trained over 50 epochs to classify histopathological images of colon (adenocarcinoma-benign) and lung (adenocarcinoma-squamous cell carcinoma-benign). Moreover, we additionally included breast (benign-malignant) tissue cell images to validate the robustness of proposed architecture. We put the results of this additional experiment as an ablation study. During the training phase of the proposed FNet architecture, accuracy and loss values are calculated after each epoch. Finally, confusion matrices for the classes are computed for test set. The relevant representations are observed in Figure 3 for colon, lung and breast tissue images.

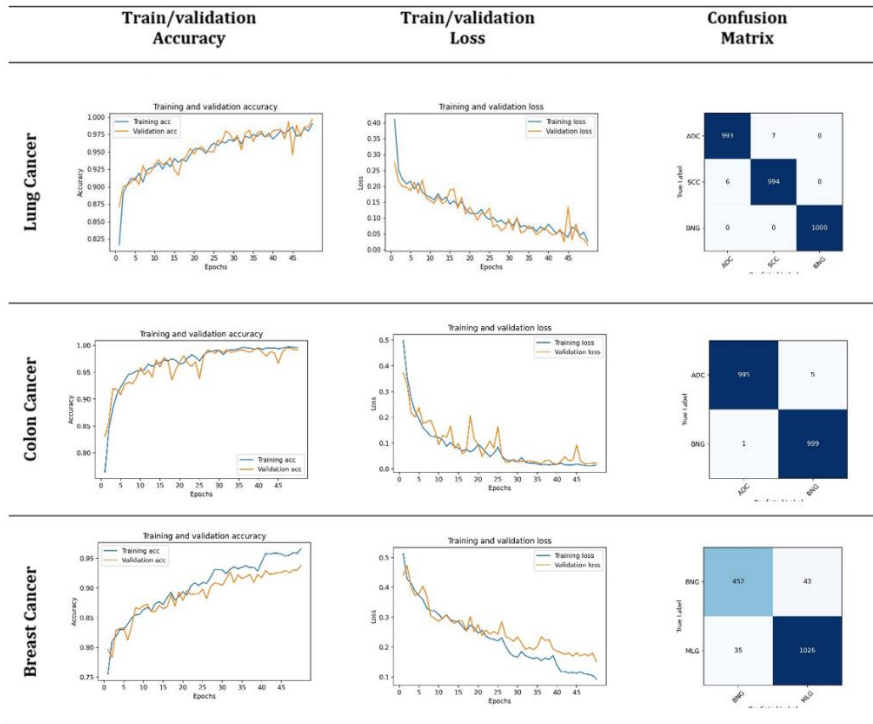


Figure 3. Accuracy-loss plots during the epochs of the training phase and confusion matrix during the testing phase for colon, lung and breast histopathological images (ADC: Adenocarcinoma, SCC: Squamous cell carcinoma, BNG: Benign tissue, MLG: Malignant tissue)

In terms of lung tissue images, all adenocarcinoma images of lung tissues are correctly identified except 7 of them are classified as squamous cell carcinoma. 6 of squamous cell carcinoma are incorrectly classified as adenocarcinoma and the rest of them are correctly discriminated. All benign tissues are correctly identified. In terms of colon tissues, 995 out of 1000 images of adenocarcinoma are classified as benign tissue images. Only single benign image is classified as adenocarcinoma. Finally, while observing the confusion matrix of breast tissue classification, 43 out of 495 benign are incorrectly classified as malignant. Moreover, only 35 malignant images are incorrectly identified as benign among 1061 images. The number of images in each class for breast tissue dataset is not balanced and proposed FNet still yields fair performance metrics as it can be seen in Table 4. Besides utilizing FNet architecture, classification results of ViT are also provided as an ablation study. Both deep learning algorithms cover 3 different datasets. We can clearly observe the superiority of FNet on different datasets and its comparison with another transformer-based algorithm (i.e. ViT). When it comes to computational complexity, FNet serves faster, accurate and less complex training stage. In FNet, the number of trainable parameters is 580,611, meanwhile, in ViT algorithm, around 22M parameter is observed. FNet yields more accurate results with less computational expenses.

Most of the colon and lung cancer multi-classification approaches given in the introduction section rely on deep learning methods and show promising levels of accuracy. In the literature, studies on breast histopathology, as well as colon and lung datasets, show that LIME analyses applied to test data after machine learning-based classification consistently demonstrate their reliability across diverse medical histopathology datasets [36]. In the medical field, interpretability is just as crucial as accuracy, as errors can have life-threatening consequences. While deep learning has indeed been successful in many disciplines, its fundamental 'black box' nature hinders its reliability in the medical field. For reliable disease diagnosis, the interpretability and transparency are

necessary in addition to accurately image classification. Meaningful results in this context require large amounts of high-quality data, significant computational burden, and efficient model architecture. Consequently, this approach can provide competitive accuracy along with being more interpretable and understandable. Alongside performance metrics, a medical image classification approach also should prioritize interpretability and transparency, fostering user trust and adoption of the method. Therefore, adding interpretability to deep learning models is one of the popular research topics. In the proposed study, applying the LIME interpretability method to train the FNet architecture and highlight the learned regions in the images has significantly enhanced the research value. This approach has made the internal learning mechanisms of the FNet architecture more understandable and interpretable, clarifying which features and regions the model emphasizes while learning. The use of the LIME method has increased the interpretability of deep learning models, which are often referred to as 'black boxes,' particularly in image-based learning processes, by reducing the model's complexity. Thus, highlighting the learned regions has helped us understand the model's decisions and learning processes more transparently, enhancing the overall impact of the proposed method.

3.1 Evaluations of tissue images by experienced pathologist

Mutations lead to the formation of masses composed of cells that proliferate uncontrollably, independent of normal physiology, and behave differently from the tissue in which they originate, known as neoplasms or tumors. Neoplasms are categorized as benign and malignant based on their behavior in the organism they develop in. Malignant neoplasms tend to cause damage at the site of development (destruction) and show a tendency to spread to distant tissues (metastasis). They are named according to the cell they originate from and classified as carcinoma (of epithelial cell origin) and sarcoma (of stromal cell origin) [37].

Table 4. Performance metrics obtained from proposed deep learning architecture (ADC: Adenocarcinoma, SCC: Squamous cell carcinoma, BNG: Benign tissue, MLG: Malignant tissue)

Transformer Models	Tissue Datasets	Performance Metrics	Best Fold
FNet Number of Trainable Parameters: 580,611	Lung	Precision	ADC:0.99; SCC:0.99; BNG:1
		Recall	ADC:0.99; SCC:0.99; BNG:1
		F1-score	ADC:0.99; SCC:0.99; BNG:1
		Accuracy	100%
	Colon	Precision	ADC:1; BNG:1
		Recall	ADC:0.99; BNG:1
		F1-score	ADC:1; BNG:1
		Accuracy	100%
	Breast	Precision	BNG:0.93; MLG:0.96
		Recall	BNG:0.91; MLG:0.97
		F1-score	BNG:0.92; MLG:0.96
		Accuracy	0.95%
ViT Number of Trainable Parameters: 21,659,588	Lung	Precision	ADC:0.98; SCC:0.99; BNG:1
		Recall	ADC:0.99; SCC:0.98; BNG:1
		F1-score	ADC:0.99; SCC:0.99; BNG:1
		Accuracy	0.99%
	Colon	Precision	ADC:1; BNG:0.99
		Recall	ADC:0.99; BNG:1
		F1-score	ADC:0.99; BNG:0.99
		Accuracy	0.99%
	Breast	Precision	BNG:0.90; MLG:0.96
		Recall	BNG:0.93; MLG:0.95
		F1-score	BNG:0.91; MLG:0.96
		Accuracy	0.94%

Lung adenocarcinoma and squamous cell carcinoma are originated from the alveolar and bronchial epithelium of the lung; colon adenocarcinoma is arisen from the crypt epithelium and both of these carcinomas evolved from precursor lesions. In pathology routine, to diagnose invasive carcinoma, besides observing abnormal cellular appearance, it is necessary to identify the abnormal behavior of these cells structurally. Cellular atypia is mostly evaluated by looking at the nucleus characteristics of the cell (such as enlargement, irregular borders, hyperchromasia, etc.). The invasion of atypical cells beyond the basal membrane, which should confine them, into the stroma (infiltration), along with stromal alterations (desmoplasia), supports the diagnosis of invasive carcinoma

[38]. Pathologists evaluate tissues under the microscope using different magnifications of objectives to assess the described histological changes in wide areas. The areas observed in the photo frames in this study are insufficient to observe these histological changes. Therefore, based on the diagnosis in the database accepted as correct by the pathologist, interpretations have been made using a few representative images for classification in Table 5-6. In addition, the term "benign" used in the classification in the study represents normal tissue. In squares where the texture is correctly classified, artificial intelligence has not interpreted the areas selected as having a normal appearance throughout the area.

Table 5. Evaluations of colon tissue images by experienced pathologist within LIME explanations

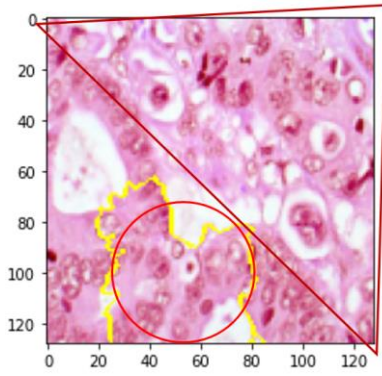
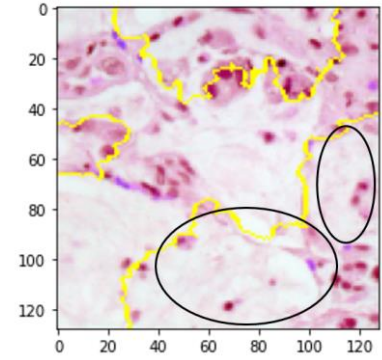
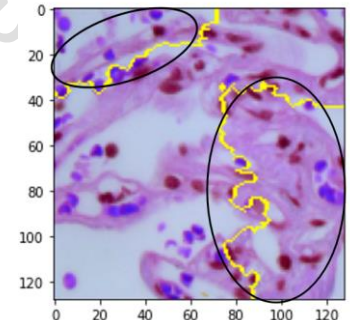
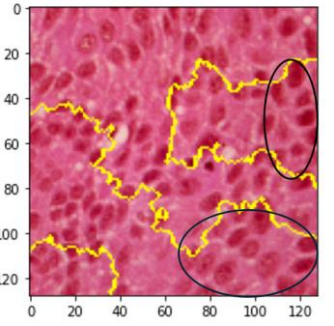
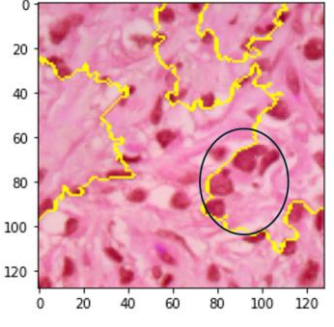
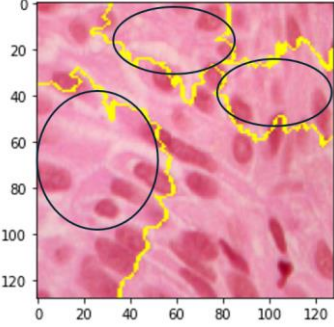
Identified Tissue Image	Explanations
	Correctly classified as adenocarcinoma In this image, the kernel sizes selected by artificial intelligence are similar and exhibit a monotonous appearance in the area. Unlike normal epithelial cells, clustering, vesicular appearance, and prominent nucleoli stand out. By considering these features as atypical, one could have made the correct classification. In fact, the nuclear characteristics within the red triangle in the photograph are more supportive for the diagnosis of carcinoma.
	It is adenocarcinoma, but misclassified as benign The artificial intelligence may have considered this square as normal tissue due to the low cellularity of the selected areas (marked by as circle). Additionally, the existing nuclei in the selected area appear small and monotonous.

Table 6. Evaluations of lung tissue images by experienced pathologist within LIME explanations

Identified tissue image	Explanations
	Correctly classified as adenocarcinoma In this photo frame, cells in the area identified diagnostically by artificial intelligence exhibit atypical cytological features such as enlargement, shape and size variation, and hyperchromasia. Although no obvious invasion is observed, accurate classification may have been made based on this atypical appearance.

	<p>Correctly classified as squamous cell carcinoma In the cells marked in this square with oval, nuclear enlargement, hyperchromasia, and irregular shape are prominent. Although obvious invasion is not observed, nuclei are more densely seen in the areas marked by artificial intelligence.</p>
	<p>It is benign but misclassified as squamous cell carcinoma In this image, enlargement, hyperchromasia, shape, and size variations are observed in the nuclei marked with circular symbols within the area identified diagnostically by artificial intelligence. These features may be considered as atypical findings.</p>
	<p>It is squamous cell carcinoma, but misclassified as benign In this image, areas recognized as peripheral by artificial intelligence typically exhibit cytoplasm, with fewer instances of nuclear structures. These nuclei are similar in size and shape, presenting a monotonous appearance. Moreover, due to the denser presence of cytoplasm in enclosed areas, it may be deemed as normal tissue by artificial intelligence.</p>

4 Conclusion

In this study, we have demonstrated the efficacy of integrating the FNet architecture with the LIME method for the detection of colon & lung cancer in histopathological images. The FNet architecture, which utilizes a transformer-based approach with a Fourier transform layer, proved to be effective in extracting meaningful features from pathological medical images. The application of the LIME method further enhanced the interpretability of the model's decisions, allowing for clearer insights into the decision-making process of the deep learning model.

Our experimental results indicate that the proposed deep-XAI method achieves high accuracy in classifying different types of colon & lung cancers. The integration of expert pathologist evaluations with LIME outputs validated the reliability of the model's predictions. This methodology not only enhances cancer detection accuracy but also offers a transparency and interpretable framework, which is vital for clinical contexts where comprehending the reasoning behind a model's prediction.

Despite the promising results, the proposed study has some limitations. One significant drawback is the computational

complexity associated with the Fourier-Net architecture, which may pose challenges in terms of processing time and resource requirements, particularly with very large datasets. Additionally, while the LIME method enhances interpretability, it is inherently local and may not fully capture the global behavior of the model, potentially leading to incomplete explanations in some cases. Furthermore, the reliance on a single dataset (LC25000) limits the generalizability of the findings; diverse datasets with varying conditions and image

qualities are needed to validate the robustness of the proposed method.

Overall, the integration of F-Net and LIME offers a robust and interpretable solution for histopathological image analysis, contributing to more effective and reliable cancer diagnosis systems. This study lays the groundwork for future research in developing advanced, interpretable AI systems to support medical professionals in making informed decisions in cancer diagnosis.

The impact of this study on existing literature is substantial. It advances the field of medical image analysis by integrating an effective transformer-based architecture with explainable AI techniques, addressing the critical need for transparency in deep learning models used in clinical settings. This approach

bridges the gap between high-performance AI models and their interpretability, enhancing trust among medical professionals and promoting the integration of AI into routine diagnostic practices.

Future researches should focus on leveraging vision-based foundation models for generating feature embeddings and utilizing the extracted embeddings with novel classifiers to improve classification performance. Additionally, integrating various explainable AI techniques, such as SHAP, Grad-CAM, and Rollout, could further enhance model interpretability and provide deeper insights into the decision-making process. These advancements hold the potential to create robust, interpretable, and effective AI systems for cancer classification.

5 Author contribution statement

In the current study, Author1 has contributed to Conceptualization, Methodology, Software, Writing - Original Draft, and Visualization; Author2 has contributed to Conceptualization, Writing - Original Draft, Visualization, and Author3 has contributed to Supervision, Writing - Review & Editing.

6 Ethics committee approval and conflict of interest statement

There is no need to obtain permission from the ethics committee for the article prepared. There is no conflict of interest with any person/institution in the article prepared.

7 Acknowledgement

This study was supported by Dicle University Department of Scientific Research Projects (Project Number: MÜHENDİSLİK.17.019). A sincere appropriation to Julius Bamwenda for his diligent proofreading of the manuscript.

8 References

- [1] D. Collaborators, "Global , regional , and national life expectancy , all-cause mortality , and cause-specific mortality for 249 causes of death , 1980 – 2015 : a systematic analysis for the Global Burden of Disease Study 2015," pp. 1980–2015, 2017, doi: 10.1016/S0140-6736(16)31012-1.
- [2] B. F. Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, "WHO." [Online]. Available: <https://gco.iarc.who.int>
- [3] D. Hanahan and R. A. Weinberg, "The Hallmarks of Cancer," *Cell*, vol. 100, no. 1, pp. 57–70, 2000, doi: [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
- [4] P. L. Nunez and R. Srinivasan, "A theoretical basis for standing and traveling brain waves measured with human EEG with implications for an integrated consciousness," *Clin. Neurophysiol.*, vol. 117, no. 11, pp. 2424–2435, 2006, doi: 10.1016/j.clinph.2006.06.754.
- [5] K. Kurishima *et al.*, "Lung cancer patients with synchronous colon cancer," *Mol. Clin. Oncol.*, pp. 137–140, 2017, doi: 10.3892/mco.2017.1471.
- [6] M. del Re *et al.*, "Implications of KRAS mutations in acquired resistance to treatment in NSCLC," *Oncotarget*, vol. 9, no. 5, pp. 6630–6643, 2018, doi: 10.18632/oncotarget.23553.
- [7] D. Crosby *et al.*, "Early detection of cancer," *Science* (80-.), vol. 375, no. 6586, p. eaay9040, 2022, doi: 10.1126/science.aay9040.
- [8] C. Bladder *et al.*, "Bladder Cancer Early Detection , Diagnosis , and Staging Can Bladder Cancer Be Found Early," *Am. Cancer Soc.*, no. cancer.org, pp. 1–24, 2023, [Online]. Available: <https://www.cancer.org/content/dam/CRC/PDF/Public/8661.00.pdf>
- [9] E. Sümer, M. Engin, M. Ağildere, and H. Oğul, "Monitoring Nodule Progression in Chest X-ray Images," *Pamukkale Univ. J. Eng. Sci.*, vol. 24, no. 5, pp. 934–941, 2018, doi: 10.5505/pajes.2018.89166.
- [10] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Med. Image Anal.*, vol. 79, p. 102470, 2022, doi: 10.1016/j.media.2022.102470.
- [11] S. Mangal, A. Chaurasia, and A. Khajanchi, "Convolution Neural Networks for diagnosing colon and lung cancer histopathological images." 2020.
- [12] M. Masud, N. Sikder, A.-A. Nahid, A. K. Bairagi, and M. A. AlZain, "A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using a Deep Learning-Based Classification Framework," *Sensors*, vol. 21, no. 3, 2021, doi: 10.3390/s21030748.
- [13] K. Adu, Y. Yu, J. Cai, K. Owusu-Agyemang, B. A. Twumasi, and X. Wang, "DHS-CapsNet: Dual horizontal squash capsule networks for lung and colon cancer classification from whole slide histopathological images," *Int. J. Imaging Syst. Technol.*, vol. 31, no. 4, pp. 2075–2092, 2021, doi: <https://doi.org/10.1002/ima.22569>.
- [14] M. Ali and R. Ali, "Multi-Input Dual-Stream Capsule Network for Improved Lung and Colon Cancer Classification," *Diagnostics*, vol. 11, no. 8, 2021, doi: 10.3390/diagnostics11081485.
- [15] N. yahia Ibrahim and A. S. Talaat, "An Enhancement Technique to Diagnose Colon and Lung Cancer by using Double CLAHE and Deep Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 8, 2022, doi: <https://doi.org/10.14569/IJACSA.2022.0130833>.
- [16] M. A. Talukder, M. M. Islam, M. A. Uddin, A. Akhter, K. F. Hasan, and M. A. Moni, "Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning," *Expert Syst. Appl.*, vol. 205, p. 117695, 2022, doi: <https://doi.org/10.1016/j.eswa.2022.117695>.
- [17] M. S. N. Raju and B. S. Rao, "Classification of Colon and Lung Cancer through Analysis of Histopathology Images Using Deep Learning Models," *Ing. des Syst. d'Information*, vol. 27, no. 6, pp. 967–971, 2022, doi: 10.18280/isi.270613.
- [18] N. Kumar, M. Sharma, V. P. Singh, C. Madan, and S. Mehandia, "An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images," *Biomed. Signal Process. Control*, vol. 75, p. 103596, 2022, doi: <https://doi.org/10.1016/j.bspc.2022.103596>.
- [19] R. R. Wahid, C. Nisa', R. P. Amaliyah, and E. Y.

- Puspaningrum, "Lung and colon cancer detection with convolutional neural networks on histopathological images," *AIP Conf. Proc.*, vol. 2654, no. 1, p. 20020, Feb. 2023, doi: 10.1063/5.0114327.
- [20] S. Tummala, S. Kadry, A. Nadeem, H. T. Rauf, and N. Gul, "An Explainable Classification Method Based on Complex Scaling in Histopathology Images for Lung and Colon Cancer," *Diagnostics*, vol. 13, no. 9, 2023, doi: 10.3390/diagnostics13091594.
- [21] S. Mehmood *et al.*, "Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning With Class Selective Image Processing," *IEEE Access*, vol. 10, pp. 25657–25668, 2022, doi: 10.1109/ACCESS.2022.3150924.
- [22] J. Fan, J. Lee, and Y. Lee, "A Transfer Learning Architecture Based on a Support Vector Machine for Histopathology Image Classification," *Appl. Sci.*, vol. 11, no. 14, 2021, doi: 10.3390/app11146380.
- [23] T. Aitazaz, A. Tubaishat, F. Al-Obeidat, B. Shah, T. Zia, and A. Tariq, "Transfer learning for histopathology images: an empirical study," *Neural Comput. Appl.*, vol. 35, no. 11, pp. 7963–7974, 2023, doi: 10.1007/s00521-022-07516-7.
- [24] M. S. Ahmed, K. N. Iqbal, and M. G. R. Alam, "Interpretable Lung Cancer Detection using Explainable AI Methods," in *2023 International Conference for Advancement in Technology (ICONAT)*, 2023, pp. 1–6, doi: 10.1109/ICONAT57137.2023.10080480.
- [25] J. Lee-Thorp, J. Ainslie, I. Eckstein, and S. Ontañón, "FNet: Mixing Tokens with Fourier Transforms," *NAACL 2022 - 2022 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Proc. Conf.*, pp. 4296–4313, 2022, doi: 10.18653/v1/2022.naacl-main.319.
- [26] 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)," pp. 1–2, 2019.
- [27] H. Lemjabbar-Alaoui, O. U. I. Hassan, Y.-W. Yang, and P. Buchanan, "Lung cancer: Biology and treatment options," *Biochim. Biophys. Acta - Rev. Cancer*, vol. 1856, no. 2, pp. 189–210, 2015, doi: <https://doi.org/10.1016/j.bbcan.2015.08.002>.
- [28] J. H. Schiller *et al.*, "Comparison of Four Chemotherapy Regimens for Advanced Non-Small-Cell Lung Cancer," *N. Engl. J. Med.*, vol. 346, no. 2, pp. 92–98, 2002, doi: 10.1056/NEJMoa011954.
- [29] A. Banerjee, S. Pathak, V. D. Subramaniam, D. G., R. Murugesan, and R. S. Verma, "Strategies for targeted drug delivery in treatment of colon cancer: current trends and future perspectives," *Drug Discov. Today*, vol. 22, no. 8, pp. 1224–1232, 2017, doi: <https://doi.org/10.1016/j.drudis.2017.05.006>.
- [30] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, 2016, doi: 10.1109/TBME.2015.2496264.
- [31] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is All You Need," in *31st Conference on Neural Information Processing Systems (NIPS)*, Long Beach, 2015.
- [32] E. Yazan and M. F. Talu, "Integration of attention mechanisms into segmentation architectures and their application on breast lymph node images," *Pamukkale Univ. J. Eng. Sci.*, vol. 29, no. 3, pp. 248–255, 2023, doi: 10.5505/pajes.2022.07838.
- [33] G. Çelik and M. F. Talu, "Generating the image viewed from EEG signals," *Pamukkale Univ. J. Eng. Sci.*, vol. 27, no. 2, pp. 129–138, 2021, doi: 10.5505/pajes.2020.76399.
- [34] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Medical Image Analysis*, vol. 79, Elsevier B.V., Jul. 01, 2022, doi: 10.1016/j.media.2022.102470.
- [35] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," *NAACL-HLT 2016 - 2016 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Proc. Demonstr. Sess.*, pp. 97–101, 2016, doi: 10.18653/v1/n16-3020.
- [36] D. Kaplun, A. Krasichkov, P. Chetyrbok, N. Oleinikov, A. Garg, and H. S. Pannu, "Cancer Cell Profiling Using Image Moments and Neural Networks with Model Agnostic Explainability: A Case Study of Breast Cancer Histopathological (BreakHis) Database," *Mathematics*, vol. 9, no. 20, 2021, doi: 10.3390/math9202616.
- [37] J. C. Kumar, Vinay Abbas, Abul K Aster, "Robbins Basic Pathology, Ninth Edition," in *Robbins Basic Pathology*. [Online]. Available: https://books.google.com.tr/books?id=aLV9nU_7X6UC&printsec=copyright&redir_esc=y#v=onepage&q&f=false
- [38] J. M. JR Goldblum, LW Lamps, *Rosai and Ackerman's Surgical Pathology*. 2017.