6

Faiz Mohammad Karobari\* and Hosahally Narayangowda Suresh

# Histopathological Image Segmentation Using Modified Kernel-Based Fuzzy C-Means and Edge Bridge and Fill Technique

https://doi.org/10.1515/jisys-2018-0316 Received July 27, 2018; previously published online March 20, 2019.

**Abstract:** Histopathological lung cancer segmentation using region of interest is one of the emerging research area in the field of health monitoring system. In this paper, the histopathological images were collected from the database Stanford Tissue Microarray Database (TMAD). After image collection, pre-processing was performed using a normalization technique, which enhances the quality of the histopathological image by eliminating unwanted noise. After pre-processing, segmentation was carried out using the modified kernel-based fuzzy c-means clustering (KFCM) approach along with the edge bridge and fill technique (EBFT). It was a flexible high-level machine learning technique to localize the object in a complex template. The experimental result shows that the proposed approach segments the normal and abnormal cancer regions by means of precision, recall, specificity, accuracy, and Jaccard coefficient. The proposed methodology improved the classification accuracy in lung cancer segmentation up to 2.5–5% compared to the existing methods deep convolutional neural network (DCNN) and diffusion-weighted approach.

**Keywords:** Deep convolutional neural network; kernel-based fuzzy c-means; lung cancer segmentation; normalization.

Mathematics Subject Classification: 68U10; 92C50.

# 1 Introduction

Lung cancer is one of the predominant cancer types, which causes more than 1.4 million deaths annually [19]. Generally, there are two types of lung cancers: the non-small cell lung cancer (NSCLC) with 80%–85% of lung cancer patients and small cell lung cancer (SCLC) with 10%–15% of lung cancer patients [17]. The evaluation of microscopic histopathology slides by experienced pathologists is indispensable for establishing the diagnosis and to define the subtypes and types of lung cancers, which includes two major types of NSCLC: adenocarcinoma and squamous cell carcinoma [5, 10, 23, 26]. The distinction of squamous cell carcinoma from adenocarcinoma is essential for chemotherapeutic selection because certain antineoplastic agents are contraindicated for squamous cell carcinoma patients, due to decreased efficacy or increased toxicity [11, 12]. However, the qualitative evaluation of well-known histopathology pattern (cancer grade classification) is insufficient to predict the survival rate of patients with lung adenocarcinoma, and only the best-characterized histopathology features achieve the modest agreements among experienced pathologists [20, 24]. Poor cancer differentiation and slide quality are related with diagnostic agreement [6]. Recently, several researches were carried out to describe the additional visual features for prognostic detection of patients with lung adenocarcinoma [3, 15]. Still, there is considerable improvement needed for the inter-rater agreements of these features [8]. Erroneous or subjective evaluation of the histopathology image leads to poor therapeutic choice

<sup>\*</sup>Corresponding author: Faiz Mohammad Karobari, Department of Electronics and Communication Engineering, KNS Institute of Technology, Kogilu Main Road, Yelahanka Hobli, Tirumenahalli, RK Hegde Nagar, Bengaluru, Karnataka 560064, India, e-mail: faizkarobari@gmail.com

Hosahally Narayangowda Suresh: Department of Electronics and Instrumentation Engineering, Bangalore Institute of Technology, Bangalore, India; and Research Guide, Visvesvaraya Technological University, Belagavi, Karnataka, India

that results in decreased survival and loss of life quality in several patients [16]. Unsupervised image processing technology shows high accuracy, consistency, and efficiency in histopathology evaluations and also provides decision support for ensuring the diagnostic consistency. Automated histopathological analysis is valuable in prognostic determinations of various malignancies [9, 22].

In this research, histopathology lung images were collected from the dataset TMAD. After the collection of histopathology images, a new segmentation methodology was implemented along with the edge bridge and fill technique (EBFT) for histopathological lung cancer segmentation. Currently, most of the segmentation algorithms are working based on Euclidean distance, which helps to identify the similarity between the data objects. Euclidean distance is computationally inexpensive and simple, but it is sensitive to outliers and perturbations. Currently, with the popular usage of support vector machine, a new direction appears to use kernel functions. The kernel function projects the data into high-dimensional space, where data can be easily separated. To perform this operation, kernel trick is adopted for transforming the linear algorithm into nonlinear algorithm using a dot operator. The kernel-based FCM works well only if the clusters are round shaped, and it works badly with non-convex shapes, due to the hard assignments of points to the clusters. To address this concern, a new clustering methodology was proposed (modified kernel-based fuzzy c-means clustering (KFCM)) for histopathological lung cancer segmentation. In this experimental research, the Euclidean distance measure was replaced with correlation distance measure because it includes numerous advantages. The distribution of correlation between the random vectors becomes narrowly focused around zero as the dimensionality increases, so the significance of small correlation grows with increasing dimensionality. Also, it is very effective in capturing the similarity between the patterns of feature vectors. Finally, the segmented output was compared with the existing methodologies by means of precision, recall, specificity, accuracy, and Jaccard coefficient.

This research paper is organized as follows. Section 2 surveys several recent papers on histopathological lung cancer detection. In Section 3, an effective approach (modified KFCM with EBFT) for histopathological lung cancer segmentation is presented. Section 4 shows the quantitative and comparative analysis of proposed and existing segmentation approaches. The conclusion is made in Section 5.

### 2 Literature Review

Researchers suggested several techniques for the detection of histopathological lung cancer. A brief evaluation of some important contributions to the existing literatures is presented in this section.

Xing and Yang [21] presented a new segmentation methodology, which combines an effective local repulsive balloon snake deformable approach (bottom-up) and a robust selection-based sparse shape approach (top-down) to tackle the segmentation concerns. The developed approach extensively tested on 62 cases with 6000 tumor cells. The experimental outcome confirms that the developed methodology delivers better segmentation performance compared to the existing methods. While performing with semi-supervised methodologies, the semantic gap was maximized between the feature values, which leads to poor segmentation rate.

Coudray et al. [4] developed an effective classification methodology: deep learning convolutional neural network for classifying the whole-slide histopathology images into squamous cell carcinoma and adenocarcinoma. The advantage of developed methodology was it was fully automated; there was no user intervention for segmenting the normal and cancer tissues, so the time consumption of detecting the cancer cell was very less. Extensive experiments were carried out on real-time histopathological images for demonstrating the robustness of the proposed scheme. The developed technique was less suitable for recognizing the inclined histopathology images because it took a high response time and identification rate.

Zhang et al. [25] developed an effective segmentation methodology: Gaussian-based hierarchical voting and repulsive balloon approach for classifying the lung cancers as adenocarcinoma and squamous carcinoma. The developed methodology delivers an effective histopathological lung cancer segmentation in the spatial localization and structural composition. The experiment was carried out on publicly available database, Cancer Genome Atlas (TCGA), in order to validate the developed method accuracy, robustness, and segmentation speed. In some cases, the training data were supervised evaluation or manual adjustment, which is needed to be automated.

Khosravi, et al. [13] presented an effective computational approach based on the convolution neural network for classifying dissimilar histopathology images across dissimilar types of cancer. In this literature, the developed machine learning approach splits the cancer cells from normal tissues and also handles the cell appearance variations. The developed algorithm was tested on different histopathology images with different variations in appearance and shapes. The experimental outcome shows that the developed histopathology cancer cell detection methodology was very effective compared to the existing methodologies. The developed methodology would fail, when the growth of the cancer cell boundaries was constrained.

Vu et al. [18] developed a new feature discovery methodology: discriminative feature-oriented dictionary learning (DFDL) approach for disease grading and classification in histopathology. In this literature, the developed dictionary learning approach was tested on three challenging image datasets: animal diagnostics lab dataset, cancer genome atlas dataset, and intraductal breast lesions histopathological images. The experimental performance suggested that the developed approach was robust and effective in comparison with existing approaches. In a large dataset, DFDL failed to achieve better segmentation by means of accuracy.

An unsupervised algorithm (modified KFCM with EBFT) is implemented for enhancing the performance of histopathological lung cancer segmentation and to overcome the above-mentioned drawbacks.

# 3 Proposed Methodology

The proposed methodology for segmenting the normal and abnormal regions from lung histopathological image is divided into three major steps: image collection, image pre-processing, and feature extraction. The workflow of the proposed histopathological lung cancer detection system is represented in Figure 1. The brief description about the proposed technique is described below.

### 3.1 Image Collection

In the initial stage of histopathological lung cancer segmentation, histopathological images are taken from the standard benchmark dataset: TMAD. This dataset contains 205,161 images, which archives 349 distinct probes on 1488 tissue micro-array slides. In that, 31,306 histopathological images for 68 probes on 125 slides are released to the public. The TMAD combines the NCI thesaurus ontology for probing tissues

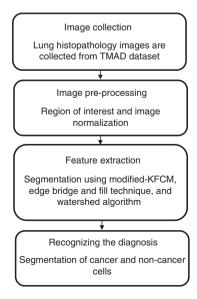


Figure 1: Workflow of the proposed system.

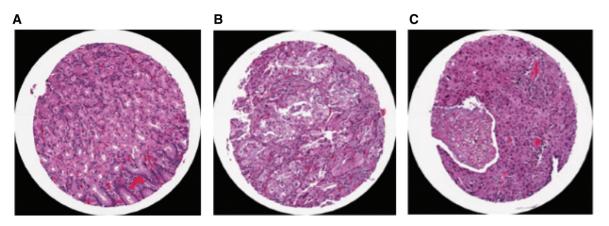
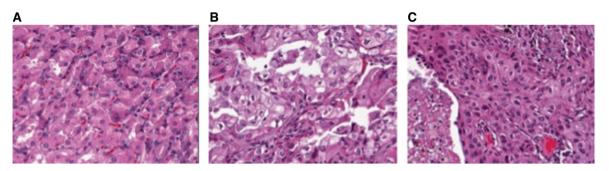


Figure 2: Sample images collected from the TMAD database.
(A) Normal image, (B) adenocarcinoma image, (C) squamous image.



**Figure 3:** ROI applied images.

(A) ROI-applied normal image, (B) ROI-applied adenocarcinoma image, (C) ROI-applied squamous image.

in the cancer domain. For lung cancer, the TMAD database contains one normal image, 220 adenocarcinoma images, and 68 squamous histopathological images. The sample collected histopathological images of normal, adenocarcinoma, and squamous are represented in Figure 2A–C.

After obtaining the histopathological images, an important step: region of interest (ROI) is carried out on the collected images. The ROI is defined as a subset of histopathological image or a database identified for a specific purpose. Therefore, the size of the selected ROI is 256  $\times$  256, which is one-fourth of the original image 1024  $\times$  1024. The sample ROI applied histopathological images of normal, adenocarcinoma, and squamous are denoted in the Figure 3A–C.

## 3.2 Pre-processing of Histopathological Image

The ROI-applied histopathological images are used for pre-processing. In this research, normalization is performed on the histopathological image for enhancing or de-noising the images. Most of the histopathological images are collected or captured from the equipment's, whereas the collected histopathological images majorly consist of two noises: impulse noise and machinery noise (electrical and mechanical noises) [1]. The normalization methodology is very effective to remove impulse and machinery noise and also helps to enhance the image quality significantly.

Normalization modifies the pixel intensity range values to improve the quality of images by reducing the noise from histopathology images. Then, the deformation and alternations that occurred in the histopathological image by inaccurate image capture were evaluated. Usually, image normalization contains pixel intensity variations. The entire histopathological images are converted into pre-defined values because it is a pixel-wise procedure. The general formula of image normalization approach is denoted in equation (1).

$$I_n = (I - Min) + \frac{newMax - newMin}{Max - Min} + newMin$$
 (1)

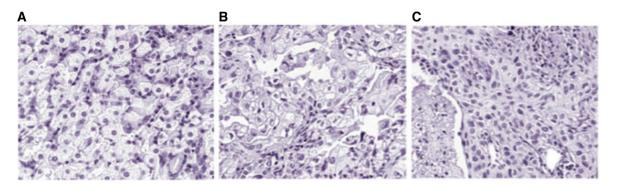


Figure 4: Preprocessed image. (A) Normalized normal image, (B) normalized adenocarcinoma image, (C) normalized squamous image.

where I is represented as the original histopathology image,  $I_n$  is stated as the new image, (Min = 0, Max =255) is specified as the intensity range of the original image, and (newMin, newMax) is stated as the intensity range of the new histopathology image. The sample normalized histopathology images of normal, adenocarcinoma, and squamous are represented in Figure 4A-C.

### 3.3 Histopathological Image Segmentation

The normalized histopathological images are used for segmentation; an effective methodology modified KFCM is undertaken for segmenting the normal and abnormal regions of histopathological images. Generally, image segmentation is the procedure of sub-dividing an image into different regions that are homogeneous with respect to some image features. The aim of image segmentation is to extract and detect the particular region from an image [14]. In this scenario, consider *I* as an input histopathological image that consists of a set of  $p_i$  color images at pixel  $i(i=1,2,\ldots,N)$  and  $P=\{p_1,p_2,p_3\ldots p\}\subset R^k$ , respectively, in the kdimensional area. The cluster centers in the histopathological images are denoted as  $Q = \{q_1, q_2, q_3 \dots q_c\}$ , where *c* is said to be a positive integer (2 <  $c \ll N$ ), and  $u_{ij}$  is the membership value for each pixel *i* in the *j*-th cluster  $(j = 1, 2, \dots c)$ . The clusters formed in the image space are combined by assigning a separate membership value to all pixels in the KFCM algorithm. The objective function or general equation of KFCM is written in equation (2).

$$J_{KFCM} = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^{m} \| p_{i} - q_{j} \|^{2}, \quad 1 \leqslant m < \infty$$
 (2)

where *m* is an exponent of regularization to the degree of fuzziness, m > 1, and  $||p_i - q_i||^2$  is the grayscale Euclidean distance between i and  $q_i$ , which is stated in equation (3).

$$\sum_{j=1}^{c} u_{ij} = 1, \quad u_{ij} \in [0, 1], \quad 0 \leqslant \sum_{i=1}^{N} u_{ij} \leqslant N$$
 (3)

Utilizing the membership function from the alternate optimization, the cluster centers are updated iteratively using equations (4) and (5).

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} (\|p_i - q_j\|^2 / \|p_i - q_k\|^2)^{1/(m-1)}}$$
(4)

$$q_{j} = \frac{\sum_{i=1}^{N} u_{ij}^{m} p_{i}}{\sum_{i=1}^{N} u_{ij}^{m}}$$
 (5)

The presence of noise is decreased by adding the spatial information of neighboring pixels that is denoted in equation (6).

$$J_{KFCM-S} = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^{m} \|p_{i} - q_{j}\|^{2} + \frac{\alpha}{N_{R}} \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^{m} \left( \sum_{r \in N_{i}} \|p_{r} - q_{j}\|^{2} \right)$$
(6)

where  $\alpha$  denotes a spatial information,  $N_i$  and  $N_r$  are defined as the set of pixel and cardinality of the pixels employed in the system. To avoid the neighborhood function, the term  $\frac{1}{N_R}\sum_{r\in N_i}\left\|p_i-q_j\right\|^2$  is replaced with  $\left\|\dot{p}_i-q_j\right\|^2$ , where,  $\dot{p}$  is a color scale-filtered image, and the Euclidean distance is replaced by the correlation distance measure.

The updated equation is represented in equation (7):

$$J_{KFCM-S(1,2)} = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^{m} \|p_{i} - q_{j}\|^{2} + \alpha \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^{m} (\|\hat{p}_{i} - q_{j}\|^{2})$$

$$(7)$$

In this research, a modified KFCM is proposed, which calculates the parameter  $\eta_j$  at every step of the iterations to replace  $\alpha$  for every cluster [7]. The correlation function is used to calculate the parameter value, which is represented in equation (8).

$$\eta_{j} = \frac{\min_{j' \neq j} (1 - C(q'_{j}, q_{j}))}{\max_{k} (1 - C(q_{k}, p))}$$
(8)

Here, C is a correlation distance measure or correlation function. The general identification of C requires a large number of patterns, and also, many cluster centers are required to find the optimal value for  $\eta_i$ . To overcome this problem, the combination of spatial context and scale information are made using a fuzzy factor. The fuzzy factor  $F_{ij}$  is included in the objective function of the KFCM, which is stated in equation (9).

$$J_{C-KFCM} = \sum_{i=1}^{N} \sum_{i=1}^{c} \left[ u | i j^{m} | \left\| p_{i} - q_{j} \right\|^{2} + F_{ij} \right]$$
(9)

Then, the altered fuzzy factor  $F'_{ij}$  is derived using equation (10).

$$F'_{ij} = \sum_{C \in N_i, i \neq k} w_{ik} (1 - u_{ij})^m \tag{10}$$

This altered fuzzy controls the local neighbor relationship and replaces the distance with a correlation function, where  $w_{ik}$  denotes the fuzzy factor i, and  $1 - C(p_i, q_j)$  denotes correlation metric function. The segmented histopathological images of normal, adenocarcinoma, and squamous are given in Figure 5A–C.

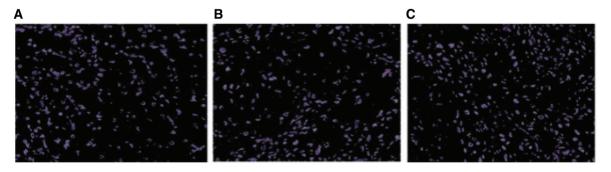


Figure 5: Segmented image.

(A) Segmented normal image, (B) segmented adenocarcinoma image, (C) segmented squamous image.

### 3.3.1 Pseudo Code of Modified KFCM

In this section, the pseudo code for the modified KFCM is presented. In the modified KFCM, a few changes are carried out in step 5, compared to the conventional KFCM. Finally, the iteration stops and returns after calculating the missing value using equation (10).

- Fix the number of clusters c, and m > 1 for some positive constant.
- Set  $p_i = 0$ , if  $p_i$  is a missing feature.
- Update all membership function  $u_{ii}$  using equation (4).
- Update all prototypes  $q_i$  using equation (5).
- Replace the Euclidean distance with the correlation function *C*.
- Calculate the missing value using equation (10).
- End

### 3.4 Edge Bridge and Fill Technique

After histopathology cell segmentation, the EBFT is utilized to bridge the gaps in the edges of the histopathology lung image for separating the overlapped nuclei and non-nuclei cells. The step-by-step procedure of the EBFT is described below.

- Initially, the extreme boundary line of the overlapped cell is kept, and all the inner portions were removed. The overlapped cell region is connected to the background black region because it is necessary to distinguish the overlapped cell region from the background. The overlapped cell edge needs to be closed, so that the cell region can discriminate from the background.
- To close the open-edge portion of the cell region, an iterative thinning (skeletenization) and thickening (dilation) morphology is utilized in this research work. Dilation is carried out with a square structural element or disk of size four. The dilation of histopathology image A by the structuring element B is described using equation (11).

$$AxorB = b \in BA_b \tag{11}$$

- In this research work, the size of the structural element is 12. All overlapped cells in the histopathology image segmented are shown in Figure 6. To make the segmentation of the overlapped cell more reliable and robust for any size of cell, it is better to use the small-size structural element rather than use the large structural element. In this experimental research, the disk structural element with radius 4 is utilized followed by skeletenization iteratively three times. This procedure depends on the number of iterations that happens after the closing of the overlapped cell edge; hence, the maximum limit of iterations does not affect the reliability of the overlapped cell segmentation process.

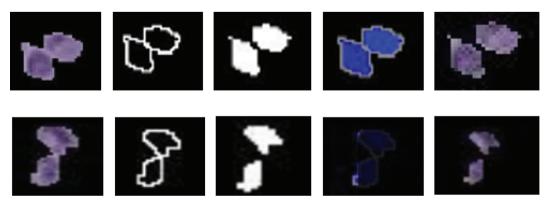


Figure 6: Sample image of the EBFT.

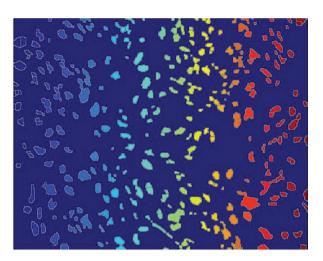


Figure 7: Sample-labeled image.

- The dilated thick line from equation (11) is skeletonized by eliminating the pixels from each side with respect to the central pixels of the line. This procedure makes the boundary lines more sharp with less gap. Then, the sharp boundary line is added to the overlapped cell without any gap.
- The holes in the overlapped cells are filled as shown in Figure 6. In the filled cell image, the overlapped nuclei and non-nuclei cells are effectively distinguished.

The labeling is utilized to better distinguish the overlapped nuclei cells and non-nuclei cells. The sample-labeled image is denoted in Figure 7.

# 4 Experimental Analysis

For experimental simulation, MATLAB (version 2017a) was employed on a PC with 3.2 GHz with i5 processor [2]. In order to estimate the efficiency of the proposed algorithm, the performance of the proposed method was compared with the diffusion-weighted approach [22] and DCNN [13] on a database: TMAD. The performance of the proposed methodology was compared by means of precision, recall, specificity, accuracy, and Jaccard coefficient.

### 4.1 Performance Measure

Performance measure is defined as the regular measurement of the outcomes and results that develops a reliable information about the effectiveness and efficiency of the proposed system. Also, it is the procedure of reporting, collecting, and analyzing information about the performance of a group or individual. The general formula for calculating the precision, recall, specificity, and accuracy of lung cancer detection is mathematically represented in equations (12)–(15).

$$Precision = \frac{TP}{(FP + TP)} \times 100$$
 (12)

$$Recall = \frac{TP}{(FN + TP)} \times 100 \tag{13}$$

Specificity = 
$$\frac{TN}{FP + TN} \times 100$$
 (14)

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP} \times 100$$
 (15)

Additionally, for segmentation validation, the Jaccard coefficient is expressed in terms of TP, TN, FP, and FN counts, which is obtained by matching the segmented result to the ground truth image. The general formula utilized to calculate the Jaccard coefficient is represented in equation (16).

$$Jaccard coefficient = \frac{TP}{(FP + TP + FN)} \times 100$$
 (16)

where FP is represented as false positive, TP is denoted as true positive, FN is indicated as false negative, and TN is specified as true negative.

### 4.2 Quantitative Analysis on TMAD Dataset

In this experimental analysis, the TMAD dataset is used for comparing the performance evaluation of the proposed approach and the existing segmentation methods. In Table 1, performance evaluation of the proposed method (modified KFCM) and the segmentation methods (FCM and KFCM) are validated by means of precision, recall, and Jaccard coefficient. The TMAD dataset contains three classes of histopathology lung images: normal, adenocarcinoma, and squamous. Here, the performance evaluation is validated for one sample image in each class with two random multi-region ROIs. The validation result shows that the proposed method outperformed the existing methodologies by means of precision, recall, and Jaccard coefficient.

For the collected histopathological image, the mean or average precision of the proposed technique is: for the modified KFCM, 68.32%, and the existing methodologies FCM and KFCM delivers 56.71% and 58.4% of the average precision. The average recall of the proposed technique is 99.22%, and the existing methodologies delivered 96.38% and 97.33% of the average recall. Similarly, the average Jaccard coefficient of the proposed technique delivers 67.96%, and the existing methodologies attained 55.55% and 57.82% of the average Jaccard coefficient. The graphical comparison of the average precision, recall, and Jaccard coefficient is denoted in Figures 8 and 9. Similarly, the standard deviation of the proposed technique precision is 5.2561, and the existing techniques (FCM and KFCM) delivered 2.837 and 3.0480. The standard deviation of the proposed technique recall is 0.5178, and the existing techniques achieved 1.0200, and 0.1892. Finally, the standard deviation of

Table 1: Precision, recall, and Jaccard coefficient comparison of the proposed and existing methods.

Images	Clustering technique	TP	TN	FP	FN	Precision (%)	Recall (%)	Jaccard coefficient (%)
Class: normal	FCM	3480	59,178	2520	130	58	96.3989	56.77
	KFCM	3611	60,181	2380	100	60.2737	97.3053	59.2842
	Modified KFCM	4801	72,681	1947	35	71.14	99.28	70.70
Class: adenocarcinoma	FCM	3821	65,646	3326	186	53.463	95.3581	52.1069
	KFCM	4018	67,706	3256	122	55.2378	97.0531	54.3267
	Modified KFCM	4724	77,734	2853	63	62.3466	98.6839	61.83
Class: squamous	FCM	3436	50,637	2419	92	58.6849	97.3923	57.777
	KFCM	3546	51,689	2293	85	60.7296	97.659	59.8582
	Modified KFCM	4556	59,711	1816	13	71.5003	99.7155	71.35

Figure 8: Graphical comparison of precision and recall.

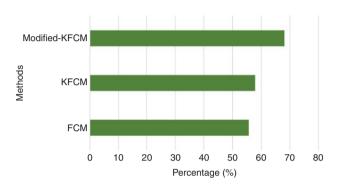


Figure 9: Graphical comparison of the Jaccard coefficient.

**Table 2:** Accuracy and specificity comparison of the existing and proposed method.

Images	Clustering technique	TN	TP	FP	FN	Specificity (%)	Accuracy
Class: normal	FCM KFCM Modified KFCM	49,923 51,957 59,000	3045 3128 4149	2401 2209 2008	166 146 81	95.4113 95.9218 99.8629	95.3777 95.9001 96.79
Class: adenocarcinoma	FCM KFCM Modified KFCM	50,946 51,440 64,948	3132 3179 3713	2584 2580 2181	109 104 12	95.1728 95.224 99.9815	95.2564 95.3161 96.90
Class: squamous	FCM KFCM Modified KFCM	60,831 64,194 74,338	4223 4401 4678	2648 2482 2315	74 56 47	95.8285 96.2775 99.9368	95.9838 96.432 97.09

the proposed technique Jaccard coefficient is 5.3154 and the existing techniques (FCM and KFCM) delivered 3.0264 and 3.0415.

In Table 2, performance evaluation of the existing and proposed method is validated in terms of accuracy and specificity. In this section, performance evaluation is validated for one sample image in each class

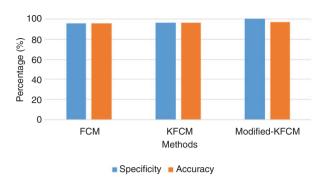


Figure 10: Graphical comparison of accuracy and specificity.

of TMAD dataset. For the collected image, the average accuracy of the proposed technique is: for the modified KFCM, 96.92%, and the existing methodologies FCM and KFCM delivered 95.53% and 95.88% of the average accuracy. In addition, the average specificity of the proposed technique is 99,92%, and the existing methodologies delivered 95.47% and 95.80% of the average specificity. Correspondingly, the standard deviation of the proposed technique's accuracy is 0.1517, and the existing techniques (FCM and KFCM) delivered 0.3914 and 0.5602. Similarly, the standard deviation of the proposed technique specificity is 0.0598, and the existing techniques (FCM and KFCM) achieved 0.3286 and 0.5346.

Tables 1 and 2 confirmed that the proposed approach performs effectively compared to the existing segmentation method on the TMAD dataset. The modified KFCM encodes both the local and shape features of the wavelet transformed histopathology image for improving the segmentation efficiency of lung cancer. The proposed modified KFCM algorithm fuses different pixel information, which represents a different correlation space to produce a new correlation value. The graphical comparison of the average accuracy and specificity is denoted in Figure 10.

### 4.3 Comparative Analysis

Tables 3 and 4 represent the comparative study of the existing work and the proposed work performance. Yin et al. [22] developed an image analysis chain that was utilized to study the correlation between tumor cellularity from serial histopathological slides of a resected NSCLC tumor and diffusion coefficient (D value) calculated from the diffusion-weighted magnetic resonance imaging (DWI). On digitized histological image, color deconvolution along with cell nuclei segmentation was used in determining the cell type and local two-dimensional densities. Then, the DWT sequence information was over-laid with resected histology using

Table 3: Performance comparison of the proposed and existing approach by means of accuracy and specificity.

Method	Cancer stages	Accuracy (%)	Specificity (%)
Diffusion weighted [22]	Combined stages	$\textbf{94.7} \pm \textbf{0.016}$	$94.3 \pm 0.029$
Modified KFCM	Normal	$\textbf{97.53} \pm \textbf{0.007}$	$99.95 \pm 0.0121$
	Adenocarcinoma	$\textbf{97.02} \pm \textbf{0.023}$	$\textbf{99.94} \pm \textbf{0.074}$
	Squamous	$\textbf{97.31} \pm \textbf{0.012}$	$\textbf{99.97} \pm \textbf{0.0225}$
	Combined stages	$\textbf{97.28} \pm \textbf{0.014}$	$99.95 \pm 0.0075$

Table 4: Performance comparison of the proposed and existing approach by means of precision, recall, and Jaccard coefficient.

Methodology	Cancer stages	Precision (%)	Recal1 (%)	Jaccard coefficient (%)
DCNN [13]	Combined stages	92	92	92
Modified KFCM	Normal	99.21	71.81	91.40
	Adenocarcinoma	98.90	64.58	94.12
	Squamous	99.54	70.04	95.81
	Combined stages	99.21	68.81	93.77

non-invasive imaging modality data and prominent anatomical hallmarks of histology tissue blocks. Finally, spatial tumor cell density and cell number information were determined on the basis of the DWT data. For sample 30 histopathological images, the average accuracy of the proposed technique is: for the modified KFCM, 97.28  $\pm$  0.014% and for the existing methodology: diffusion weighted [22] delivers 94.7  $\pm$ 0.016% of the segmentation accuracy. Similarly, the specificity of the proposed technique delivers 99.95  $\pm$  0.0075%, and the existing methodology delivers 94.3  $\pm$  0.029% of the specificity. Compared to the existing method, the proposed methodology shows 3–5% of improvement in specificity and accuracy. The graphical comparison of accuracy and specificity is denoted in Figure 11.

In Table 4, performance evaluation of the proposed method is: for the modified KFCM and the existing methodology: DCNN [13] is validated by means of precision, recall, and Jaccard coefficient. Khosravi et al. [13] developed numerous computational methodologies on the basis of CNN and also built a stand-alone pipeline for classifying the histopathology images across dissimilar types of cancer. In this research paper, stand-alone pipeline demonstrates the discriminate between two sub-types of lung cancer, five bio-markers of breast cancer, and four bio-markers of bladder cancer. The classification phase includes the ensemble of two algorithms (ResNe and inception), a basic CNN architecture, and Google's Inceptions with three training approaches. The average precision of the proposed technique is: for the modified KFCM, 99.21% is delivered and the existing methodology: DCNN attains 92% of segmentation precision. Similarly, the Jaccard coefficient of the proposed technique delivers 93.77%, and the existing methodology achieves 92% of the Jaccard coefficient. Compared to the existing method, the proposed methodology shows better segmentation results by means of precision and Jaccard coefficient. Additionally, the recall of the proposed technique delivers 68.81%, and the existing methodology delivers 92% of the recall value. The proposed approach shows less recall value because while performing with multiple ROI on input histopathology image, the size of the pixel is decreased that reduces the performance of the proposed approach. The graphical representation of precision, recall, and Jaccard coefficient is denoted in Figure 12.

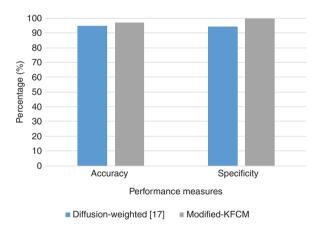


Figure 11: Graphical comparison of the proposed and existing method by means of accuracy and specificity.

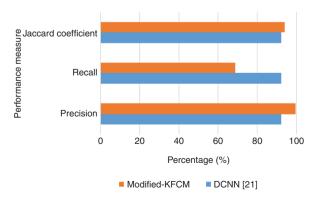


Figure 12: Graphical comparison of the proposed and existing method by means of precision, recall, and Jaccard coefficient.

### 4.4 Discussion about Proposed Methodology

Histopathology image segmentation plays a major role in cancer diagnosis, which delivers the necessary information for separating the non-cancer region from the cancer region. In this experimental research, histopathological image segmentation is carried out for the disease: lung cancer, which is one of the growing diseases in the medical field. Previously, several methodologies were developed to distinguish the cancer cells from non-cancer cells, but existing researches do not concentrate on the overlapped cancer and noncancer cells. It is essential to distinguish the overlapped cancer and non-cancer cells for best recognition or segmentation. In this research, a new unsupervised machine learning approach was developed for segmenting the overlapped cancer cells from non-cancer cells. The effectiveness of the proposed methodology is shown in Tables 3 and 4. The performance analysis is verified by determining the performance metrics like precision, recall, specificity, accuracy, and Jaccard coefficient. Under such circumstance, the accuracy of the proposed methodology is 2.5% better than the existing approach (diffusion weighted). Additionally, precision and the Jaccard coefficient of the proposed methodology shows 7% and 1.5% improvement compared with the existing approach (DCNN). The proposed segmentation methodology includes numerous advantages: assists the doctors during surgery, cost efficient related to other existing machine-learning approaches, and earlier detection of lung diseases.

# 5 Conclusion

In this research paper, a new texture-based histopathological segmentation methodology is proposed, which is based on the modified KFCM with ROI. The modified KFCM is the most effective methodology in histopathological lung cancer segmentation. In this experimental research, the modified KFCM, along with the EBFT, is utilized for segmenting the cancer and non-cancer regions in histopathological images. The proposed methodology effectively combines the advantage of the EBFT and the modified KFCM methodology. The experimental investigation was verified on a publicly available database (TMAD dataset), which shows a superiority of the proposed methodology. The modified KFCM scheme delivered an effective segmentation performance, compared to the other obtainable approaches in histopathological lung cancer detection. The proposed methodology showed 2.5-5% of the enhancement in segmentation compared with the existing methods in terms of segmentation accuracy. In future work, an appropriate classification methodology will be used in classifying the cancer and non-cancer regions in the segmented histopathology image.

# **Bibliography**

- [1] L. Azevedo, A. M. Faustino and J. M. R. Tavares, Segmentation and 3D reconstruction of animal tissues in histological images, in: Computational and Experimental Biomedical Sciences: Methods and Applications, pp. 193-207, Springer, Cham., 2015.
- [2] S. Beagum, N. Dey, A. S. Ashour, D. Sifaki-Pistolla and V. E. Balas, Nonparametric de-noising filter optimization using structure-based microscopic image classification, Microsc. Res. Tech. 80 (2017), 419-429.
- [3] R. H. J. Breuer, P. E. Postmus and E. F. Smit, Molecular pathology of non-small-cell lung cancer, Respiration 72 (2005), 313-330.
- [4] N. Coudray, A. L. Moreira, T. Sakellaropoulos, D. Fenyo, N. Razavian and A. Tsirigos, Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning, bioRxiv (2017), 1559.
- [5] N. Dey, A. S. Ashour, A. S. Ashour and A. Singh, Digital analysis of microscopic images in medicine. J. Adv. Microsc. Res. 10 (2015), 1-13.
- [6] M. Dietel, L. Bubendorf, A. M. C. Dingemans, C. Dooms, G. Elmberger, R. C. García, K. M. Kerr, E. Lim, F. López-Ríos, E. Thunnissen and P. E. Van Schil, Diagnostic procedures for non-small-cell lung cancer (NSCLC): recommendations of the European Expert Group, Thorax 71 (2015), 177-184.
- [7] Y. Ding and X. Fu, Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm, Neurocomputing 188 (2016), 233-238.
- [8] R. Dorantes-Heredia, J. M. Ruiz-Morales and F. Cano-García, Histopathological transformation to small-cell lung carcinoma in non-small cell lung carcinoma tumors, Transl. Lung Cancer Res. 5 (2016), 401-412.

- [9] L. He, L. R. Long, S. Antani and G. Thoma, Computer assisted diagnosis in histopathology, Sequence and Genome Analysis: Methods and Applications 3 (2010), 271-287.
- [10] S. Hore, S. Chakroborty, A. S. Ashour, N. Dey, A. S. Ashour, D. Sifaki-Pistolla, T. Bhattacharya and S. R. Chaudhuri, Finding contours of hippocampus brain cell using microscopic image analysis. J. Adv. Microsc. Res. 10 (2015), 93-103.
- [11] H. D. Hosgood, C. Farah, C. C. Black, M. Schwenn and J. M. Hock, Spatial and temporal distributions of lung cancer histopathology in the state of Maine, Lung Cancer 82 (2013), 55-62.
- [12] J. R. F. Junior, M. Koenigkam-Santos, F. E. G. Cipriano, A. T. Fabro and P. M. de Azevedo-Marques, Radiomics-based features for pattern recognition of lung cancer histopathology and metastases, Comput. Methods Programs Biomed. 159 (2018), 23-30.
- [13] P. Khosravi, E. Kazemi, M. Imielinski, O. Elemento and I. Hajirasouliha, Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images, EBioMedicine 27 (2018), 317-328.
- [14] Z. Ma, J. M. R. Tavares and R. N. Jorge, A review on the current segmentation algorithms for medical images, in: Proceedings of the 1st International Conference on Imaging Theory and Applications (IMAGAPP), 2009.
- [15] A. L. Moreira and M. Mino-Kenudson, Update on histologic classification of non-small cell lung cancer, Diagn. Histopathol. 20 (2014), 385-391.
- [16] M. G. Oser, M. J. Niederst, L. V. Sequist and J. A. Engelman, Transformation from non-small-cell lung cancer to small-cell lung cancer: molecular drivers and cells of origin, Lancet Oncol. 16 (2015), e165-e172.
- [17] J. C. Sieren, J. Weydert, A. Bell, B. De Young, A. R. Smith, J. Thiesse, E. Namati and G. McLennan, An automated segmentation approach for highlighting the histological complexity of human lung cancer, Ann. Biomed. Eng. 38 (2010), 3581-3591.
- [18] T. H. Vu, H. S. Mousavi, V. Monga, G. Rao and U. A. Rao, Histopathological image classification using discriminative feature-oriented dictionary learning, IEEE Trans. Med. Imaging 35 (2016), 738-751.
- [19] C. W. Wang, Robust automated tumour segmentation on histological and immunohistochemical tissue images, PloS One 6 (2011), 15818.
- [20] M. Wang, F. Tang, X. Pan, L. Yao, X. Wang, Y. Jing, J. Ma, G. Wang and L. Mi, Rapid diagnosis and intraoperative margin assessment of human lung cancer with fluorescence lifetime imaging microscopy, BBA Clin. 8 (2017), 7-13.
- [21] F. Xing and L. Yang, Robust selection-based sparse shape model for lung cancer image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 404-412, Springer, Berlin, Heidelberg, 2013.
- [22] Y. Yin, O. Sedlaczek, B. Müller, A. Warth, M. González-Vallinas, B. Lahrmann, N. Grabe, H. U. Kauczor, K. Breuhahn, I. E. Vignon-Clementel and D. Drasdo, Tumor cell load and heterogeneity estimation from diffusion-weighted MRI calibrated with histological data: an example from lung cancer, IEEE Trans, Med. Imaging 37 (2018), 35-46.
- [23] K. H. Yu, C. Zhang, G. J. Berry, R. B. Altman, C. Ré, D. L. Rubin and M. Snyder, Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features, Nat. Commun. 7 (2016), 12474.
- [24] K. H. Yu, G. J. Berry, D. L. Rubin, C. Ré, R. B. Altman and M. Snyder, Association of omics features with histopathology patterns in lung adenocarcinoma, Cell Syst. 5 (2017), 620-627.
- [25] X. Zhang, H. Su, L. Yang and S. Zhang, Fine-grained histopathological image analysis via robust segmentation and largescale retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5361-5368, 2015.
- [26] X. Zhang, F. Xing, H. Su, L. Yang and S. Zhang, High-throughput histopathological image analysis via robust cell segmentation and hashing, Med. Image Anal. 26 (2015), 306-315.