Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset

Hamdalla F.Kareem[1], Muayed S. AL-Husieny[2], Furat Y. Mohsen[3], Enam A. Khalil[4], Zainab S. Hassan[5]
[1,2]Department of Electrical Engineering, College of Engineering, Wasit University, Wasit, Iraq
[3,4,5]The Oncology Teaching Hospital, The Medical City, Baghdad, Iraq

Article Info	ABSTRACT
	This paper concerns the development/analysis of the IQ-OTH/NCCD lung cancer dataset. This CT-scan dataset includes more than 1100 images of diagnosed healthy and tumorous chest scans collected in two Iraqi hospitals. A computer system is proposed for detecting lung cancer in the dataset by using image-processing/computer-vision techniques. This includes three preprocessing stages: image enhancement, image segmentation, and feature extraction techniques. Then, support vector machine (SVM) is used at the final stage as a classification technique for identifying the cases on the slides as one of three classes: normal, benign, or malignant. Different SVM kernels and feature extraction techniques are evaluated. The best accuracy achieved by applying this procedure on the new dataset was 89.8876%.

Corresponding Author:

Hamdalla Fadil Kareem Al-Yasriy
Department of Electrical Engineering, College of Engineering
Wasit University, Al-Rabie Dist., Al-Kut, Wasit, Iraq
Email: hamdallak@uowasit.edu.iq

1. INTRODUCTION

According to the statistics of the American cancer society, lung cancer is the primary cancer killer in the United States [1]. The overall estimated number of the new cases of all types of cancer in 2013 was 1660290 (854790 for men and 805500 for women), of which the number of lung cancer accounted for 13.7% of the incidences. Whereas the total number of estimated death cases of cancer were 580350 cases (306920 for men and 273430 for women), 27.5% of these were for cases diagnosed with lung cancer with close shares for both genders [2]. According to official records of the the Iraqi ministry of health, lung cancer cases were the second among the most cancer types in 2016, where there were 2123 people of the two genders diagnosed with lung cancer. This represents about 8.31% with a small rise from the ratio of 2015. The distribution among genders was slightly different. It represented 13.27% of all types of cancers in men, while accounting for 12.7% of the total cancer types in women. Cancer is the fourth cause of mortality in the eastern Mediterranean region and is the third cause of death in Iraq, and this rate is generally rising continuously [3, 4]. The primary and most important cause of this increase is thought to be smoking. Other factors include pollution, unhealthy diet, continuous exposure to industrial and agricultural carcinogens, and reduced physical activity. The total count of cancer deaths in Iraq in 2014 was (8211), about (4525) in males and (3959) in females. Of which lung cancer deaths scored about 16.31%

Lung cancer is recognized as one of the most critical and most lethal kinds of cancer. In addition to the truth, the survival percentage after late diagnosis is shallow, and sometimes it is almost non-existent. This deadly disease occurs because of the uncontrolled growth of malignant cells within one or both lungs, and its

danger lies through the ability of these cells to spread to the rest of the human body. To eliminate this risk, early detection is the best solution. It saves the lives of people and keeps a lot of time on doctors to determine the appropriate treatment for the case. The increasing rates and the vicious nature of lung cancer worldwide puts added pressure on health society to find better, faster, and more accurate methods for the early diagnosis of this type of diseases. Computer based methods are known for their success in various industrial but more importantly medical fields. An essential factor in the quest for effective battling of a disease is the early and accurate detection and diagnosis of the medical condition. Computers resources can be utilized in a fast and scalable fashion.

Many studies employed machine learning and computer vision techniques for detecting lung cancer. For examples, the authors in [5] combined artificial neural networks and fuzzy clustering methods. Their method was applied to a dataset of about 1000 images and obtained accuracy about 97%. Artificial neural networks were employed again in another study in [6]. This one was limited to some symptoms of lung cancer, such as: Yellow fingers, Wheezing, Chest pain, Fatigue, Allergy, and etc. They achieved overall accuracy about 96.67%. The researchers in [7] developed a computer-aided system based on artificial neural networks, their accuracy was around 90.63%. Genetic algorithm was applied by the authors in [8] for feature selection. They employed SVM and ANN classification with reported accuracy at 95.87% and 93.66% respectively. Another study in [9] proposed a CAD system where the authors claim to estimate survival rate in addition to the detection and prediction of lung cancer. They extracted features from a dataset containing 909 image and got an accuracy of about 96%. Artificial intelligence was used in almost all fields of biomedical engineering and informatics such as screening and diagnosis of breast cancer [10, 11], diagnosis of Heart diseases [12-14], and also in diagnosing and classification of diabetes [15].

The main setback for computer vision applications is the large dimensionality of the problem. In order to achieve credible results with such problems, datasets ought to be very large. The vast majority of success stories in computer based medical diagnosis rely on two pillars: an effective algorithm, and a good, large dataset. This research introduces a new large marked dataset of CT scan images collected at two specialist hospitals in Iraq. This dataset is then tested and benchmarked by using computer vision methods to create a baseline for later investigation.

2.   THE IQ-OTH/NCCD LUNG CANCER DATASET

The Iraq-Oncology teaching hospital/national center for cancer diseases (IQ-OTH/NCCD) lung cancer dataset was collected in the above-mentioned specialist hospitals over a period of three months in fall 2019. It includes CT scans of patients diagnosed with lung cancer in different stages, as well as healthy subjects. IQ-OTH/NCCD slides were marked by oncologists and radiologists in these two centers. The dataset contains a total of 1190 images representing CT scan slices of 110 cases as shown in Figure 1. These cases are grouped into three classes: normal, benign, and malignant. of these, 40 cases are diagnosed as malignant; 15 cases diagnosed with benign; and 55 cases classified as normal cases. The CT scans were originally collected in DICOM format. The scanner used is SOMATOM from Siemens. CT protocol includes: 120 kV, slice thickness of 1 mm, with window width ranging from 350 to 1200 HU and window center from 50 to 600 were used for reading, with breath hold at full inspiration. All images were de-identified before performing analysis. Written consent was waived by the institutional oversight and review board of the participating medical centers which approved the study. Each scan contains several slices. The number of these slices range from 80 to 200 slices, each of them represents an image of the human chest with different sides and angles. The 110 cases vary in gender, age, educational attainment, area of residence and living status. Some of them are employees of the Iraqi ministries of Transport and Oil, others are farmers and gainers. Most of them come from places in the middle region of Iraq, particularly, the provinces of Baghdad, Wasit, Diyala, Salahuddin, and Babylon. The dataset can be accessed online on Kaggle at [16, 17].
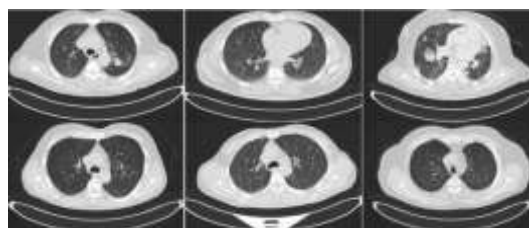

Figure 1. Example CT scan images from the IQ-OTH/NCCD dataset, some of them are for diagnosed lung cancer and benign tumor patients and others are for healthy

3.   METHODS AND IMPLEMENTATION

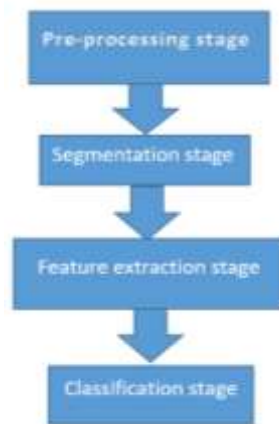In the proposed method, four steps were applied to analyze the input images as shown in Figure 2.

Figure 2. Main steps of the proposed automatic diagnosis method

3.1. Preprocessing stage

Original Images are initially subjected to preprocessing. which includes enhancing image quality that might have deteriorated due to poor illumination, and eliminating of noise [18]. This is achieved by bit plane slicing which furthermore reduces their original sizes and focuses them [19]. Figure 4 shows an example input image alongside the result after applying bit plane slicing. Next, Gaussian filtering is applied [20, 21]. This image processing technique is applied for enhancing the image and eliminating noise. Then, the erosion technique, which is a morphological operation is used to decrease the sizes of objects. This helps in removing small anomalies by subtracting objects with a radius smaller than the structuring element [18]. The result of applying Gaussian filter and erosion are as depicted in Figures 5 and 6 respectively. Then, the outlining operation [22] utilized for forming the outline of the entire image containing lungs, ribcage, etc., is performed by making all darker pixels in the original image black and all the lighter pixels in the original image white, also the darker pixels at the edges of images become white. This is done to extract the lung borders as shown in the Figure 7 At the end of the enhancement stage, the lung regions are readied for the process segmentation of the nodules from the lung CT image as depicted in Figure 8.

Figure 3. Original image

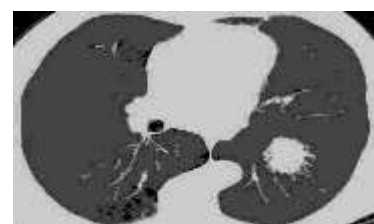Figure 4. After bit-plane slicing

Figure 5. After Gaussian filter

Figure 6. After erosion operation

Figure 7. Extracting the lung borders

Figure 8. Defining the lung areas

3.2. Segmentation stage

The second stage employed here is the segmentation stage. The aim of this stage is to extract the nodules in the lungs. The extracted nodules will be used to derive the features fed to the classifier. Otsu's thresholding technique [23, 24] is applied in this research to extract the nodules from the extracted region of the patients' lungs. The technique works by finding a threshold value which reduces the variances of the segmented regions. This techniques seems capable to obtain excellent outputs when the input image histogram has distinct foreground-background peaks [24]. The output from the segmentation method of one of the input images is shown in Figure 9.

Figure 9. Segmentation of cancer nodules

3.3. Feature extraction stage

Two types of features are extracted and used in this research, Gabor filter [25] and GLCM [26]: Gabor filter catches the total frequency spectrum, the pair amplitude, and phase. For the operation of Gabor feature extraction, the image has been convoluted with each of the Gabor filter group at all pixels [27]. The GLCM refer to Gray Level Co-occurrence Matrix [28]. The technique which consists of two steps is widely used in the applications of the image analysis, especially in the biomedical field. At the first step, GLCM matrix is computed, then the texture features based on the previously computed GLCM matrix are determined in the second step of this feature extraction method [29]. GLCM explains how often each grey level occurs at a pixel located at a fixed geometric position related to each other pixel, as a function of the grey level. The features extracted are as given below:

Contrast, refers to the measure of intensity or the grey-level variations in the image between the reference pixel and its neighborhood [30]:

$$Contrast = \sum_{i}^{Ng} \sum_{j}^{Ng} (i, j)^2 P(i, j) \tag{1}$$

Homogeneity: calculates how tight the distribution of the elements in the GLCM matrix to its diagonal [30]:

$$Homogenity = \sum_{i} \sum_{j} \frac{1}{1 - (i - j)^2} P(i, j) \tag{2}$$

Entropy: measures the degree of disorder in the image [29]:

$$Entropy = -\sum_{i} \sum_{j} Pd(i, j) \ln Pd(i, j) \tag{3}$$

Energy gives the sum of square elements in the GLCM matrix. It derived from the Angular Second Moment (ASM). which estimates the local uniformity of the grey levels [29, 30]. The value of the ASM will be large if pixels are very similar, and this value will be small when this pixel is different.

$$Energy = \sqrt{\sum_{i} \sum_{j} Pd^2(i, j)} \tag{4}$$

Correlation: gives the linear dependency of the grey level values in the co-occurrence matrix [29],

$$Correlation = \sum_{i}^{N} \sum_{j}^{N} \frac{(i - \mu x)(j - \mu y) p(i, j)}{\partial x \partial y} \tag{5}$$

Where Ng is the number of gray levels, Pd is the normalized symmetric GLCM of dimension $Ng \times Ng$, and (i, j) is the (i, j) the element of the normalized GLCM.

3.4. The classification stage

For classification, this paper employs SVM [31, 32] which is a supervised learning algorithm [33] as a learning model. It is a powerful machine learning techniques for classification and regression [34]. This algorithm achieves the classification of its input by computing the optimal separating hyperplane. This hyperplane is characterized by having the most possible distance to the classes' nearest points (support vectors). The above-mentioned approach ensures that with larger margins, the classifier yields a lower generalization error as depicted by Figure 7a [35] in which H1 does not divide the two classes; H2 divides them but with a small separating margin while H3 separates the classes with the best margin. Separation hyperplanes as shown in Figure 10.
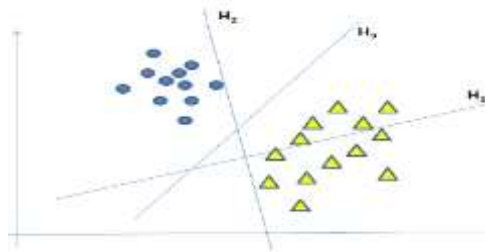

Figure 10. Separation hyperplanes

The essential ideas behind the Support Vector Machine algorithm can be explained by understanding only four fundamental thoughts: the separating hyperplane, the maximum-margin hyperplane, the soft margin and the kernel function. The above concepts will be described with the aid of the example proposed by Golub et al [36]:

A fast look at Figure 11(a) illustrates that green-colored data that represent a specified class, and the red colored data represents another class. A simple rule applied to separate the data by forming an edge, as presented in Figure 11(b). In order to describe the concept of a separating hyperplane, if the data contains only single class so the 'space' is represented by a line with a one-dimension, this line possible to be divided in half applying a particular point, this showed in Figure 11(c). In the case of two-dimensional presented in Figure 11(b), a direct line can divide the area in half, but in three-dimensional, there is a necessity for a plane for sharing the area by using a mathematical trick to extrapolating the data to a higher dimensions; this showed in Figure 11(d).

The difficility showned in Figure 11(e) gives a big question, which one of the classifiers can be providing the most desirable classifier? By special consideration can getting straightforward thought of choosing the plane or line that is the one in the midpoint, another opinion might go with selecting the plan which divides the two groups but leaves the maximum range than each item of the different classes; this is shown in Figure 11(f), when explaining the range that separates hyperplane and the most next class vector as "the maximum-margin hyperplane", various existing data sets not possible to be separating as clear; rather, they seem similar to the example in Figure 11(g), when the data set includes an "error'" the surrounded object. we need a technique to be effective in dealing with the mistakes in the data by enabling some objects to drop at the 'reverse area' of the separating hyperplane. For handling states similar to those, this algorithm should be altered by appending the "soft margin" by enabling several data objects to force their move over the separating hyperplane margin without influencing the last decision as shown in Figure 11(h).

Let's assume having only a single class as shown in Figure 11(i), in such state, the maximum-margin that separates hyperplane is a particular point. The difficulty here is there is none particular point that able to separating the categories and offering a "soft margin", so the kernel function adds to the data an additional dimension as shown in Figure 11(j) by squaring the original values and a simple direct line will sepreates the data. But the data in Figure 11(k), which isn't separable by such a direct plan. Still, a simplistic kernel function that outlines the data of the two-dimensions area to area with four-dimensions, the result of this operation is given as the bowed line in Figure 11(k). The data on the Figure 11(l) is the equivalent as Figure 11(k), though the extended hyperplane originates from an SVM which applies a high-dimensions kernel function.Then the decision is that the border separating the groups, which is so particular to the patterns in the training dataset.

In this case, The algorithm is assumed to become overfitted. And hence, it will not be generalizing well while performed with different examples [37].
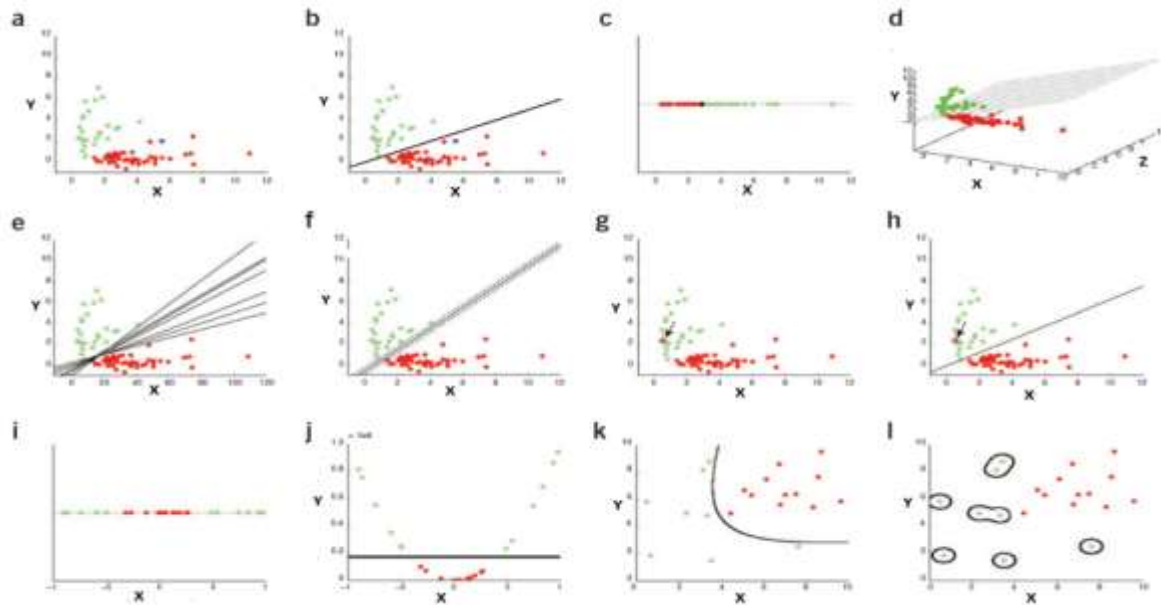

Figure 11. The essentials of the Support Vector Machine algorithm

4.   RESULTS AND DISCUSSION

In this paper, three types of features are employed for training the model. First, the Gabor filter only, second GLCM matrix only, and third a combination of the two mentioned features. SVM was applied using three kernels, linear kernel, RBF kernel and polynomial kernel. The dataset was split into two parts, 70% training set and 30% testing set.

Referring to Table 1, results show that generally in all types of feature extraction, the polynomial kernel gives the best accuracy followed by the RBF kernel and the lowest accuracy was for linear kernel. Regarding the extracted features, combining the Gabor filter with GLCM give the best accuracy as described above and the loss of the model:

Table 1. Accuracy table

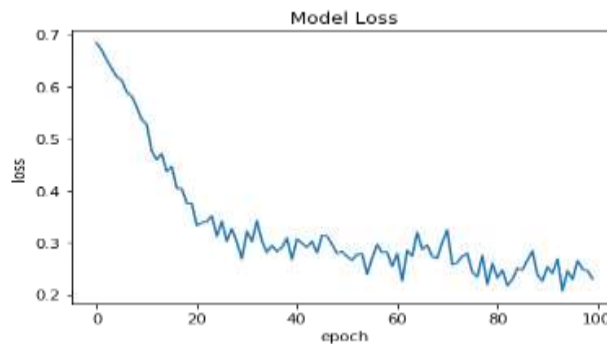| Feature extraction techniques	Linear kernel	RBF kernel	Polynomial kernel
Gabor technique	78.3351%	55.5556%	80.0127%
GLCM technique	81.0126%	79.2222%	85.3933%
Combined Gabor & GLCM techniques	82.0225%	80%	89.8876%

Figure 12. SVM Model Loss

The confusion matrix of the whole 110 cases with the values of true positives, false positives, true negatives and false negatives are shown in the Table 2. The performance metrics of the proposed model based on the above confusion matrix such as sensitivity (Recall) that represents the proportion of true positives that are correctly identified by the test [38]. Specificity that is the proportion of true negatives that are correctly identified by the test [38]. Precision or the positive predictive value that refers to the portion of related cases among the total cases and F score that Represents a combination between precision and recall [39]. These metrics were calculated for the results of the methods employed in this paper, the outcomes are listed in Table 3 below, sensitivity and specificity with their high values show that the algorithm produces robust discrimination between healthy and cancerous images which implies that this method can be reliably used for the diagnosis of lung cancer by using CT scan images. Below there is a comparison between above used technique and some related works:

Table 2. The confusion matrix

Actual class	Predicted class	
	Non-malignant (positive)	Malignant (negative)
Non-malignant	68 (TP)	2 (FN)
Malignant	1 (FP)	39 (TN)

Table 3. Performance metrics of the proposed model

Performance metrics	Sensitivity	Specificity	Precision	F Score
Value	97.143%	97.5%	98.551%	97.842%

Table 4. Comparison with related works

Author	Techniques	performance
Taher et al [5]	FCM & HNN	Sensitivity=83%, specificity=99%, and accuracy=98%
Dandıl et al [7]	ANN	Sensitivity=92.3%, specificity=89.47%, and accuracy=90.63%
Diaz et al [8]	GA, SVM, and ANN	Accuracy of 95.87% and 93.66 % respectively
Lee et al [40]	CNN	94.61% Accuracy

5.   CONCLUSIONS

Lung cancer is One of the most dangerous and widespread cancers. This danger depends on the stage at which the disease discovered. For this purpose, we proposed in this study a new dataset of lung cancer CT scans, the IQ-OTH/NCCD dataset. This dataset was analyzed to develop an automatic technique for diagnosis of lung cancer. In this approach image processing which includes enhancement, segmentation, and feature extraction was first applied, followed by the selection of features and finally the classification of images by using SVM. Three different kernels were used for the training of SVM, linear, RBF and Polynomial. The best accuracy of 89.88% was achieved when using SVM with polynomial kernel and a combination of two extracted features, Gabor filter and GLCM. The associated performance metrics with their scores at or above 90% such as sensitivity and specificity suggest that the used method is reliable in separating healthy and malignant chest CT scans. This method is intended as a baseline for further testing and analysis of the new IQ-OTH/NCCD dataset.

REFERENCES

[1]   "American Lung Association, URL www.lung.org."
[2]   Siegel, Rebecca, Naishadham, Deepa, Jemal, and Ahmedin., "Cancer statistics, 2013," *A Cancer Journal For Clinicians,* vol. 63, no. 1, pp. 11-30, 2013.
[3]   Republic of Iraq, Ministry of Health\Environment, and I. C. Board., "Annual Report Iraqi Cancer Registry 2016.
[4]   Republic of Iraq, Ministry of Health\Environment, and I. C. Board., "Annual Report Iraqi Cancer Registry 2015,"
[5]   Taher, Fatma Sammouda, and Rachid., "Lung cancer detection by using artificial neural network and fuzzy clustering methods," in *2011 IEEE GCC Conference and Exhibition (GCC)*, pp. 295-298: IEEE, 2011.
[6]   Nasser, Ibrahim M Abu-Naser, and S. S., "Lung Cancer Detection Using Artificial Neural Network," *International Journal of Engineering Information Systems,* vol. 3, no. 3, pp. 17-23, 2019.
[7]   Dandıl, M. E. Emre Çakiroğlu, Ziya Özkan, Murat Kurt, Özlem Kar Canan, and Arzu., "Artificial neural network-based classification system for lung nodules on computed tomography scans," in *2014 6th International conference of soft computing and pattern recognition (SoCPaR)*, pp. 382-386: IEEE, 2014.
[8]   Diaz, Joey Mark Pinon, Raymond Christopher Solano, and Geoffrey., "Lung cancer classification using genetic algorithm to optimize prediction models," in *IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications*, pp. 1-6, IEEE, 2014.

[9]   Ada¹ and R. Kaur, "Early detection and prediction of lung cancer survival using neural network classifier," ed: IJAIEM, 2013.

[10]   Ahmad, L Gh Eshlaghy, AT Poorebrahimi, A Ebrahimi, and H. M. I. M Razavi, "Using three machine learning techniques for predicting breast cancer recurrence," *Health Med Inform,* vol. 4, no. 124, p. 3, 2013.

[11]   Asri, Hiba Mousannif, Hajar Al Moatassime, Hassan Noel, and Thomas., "Using machine learning algorithms for breast cancer risk prediction and diagnosis," *Procedia Computer Science,* vol. 83, pp. 1064-1069, 2016.

[12]   Fei and Sheng-wei., "Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine," *Expert Systems with Applications,* vol. 37, no. 10, pp. 6748-6752, 2010.

[13]   Acharya *et al.*, "Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals," *Applied Intelligence,* vol. 49, no. 1, pp. 16-27, 2019.

[14]   Muayed S. AL-Huseiny, Noor K. Abbas, and A. S. Sajit., "Diagnosis of Arrhythmia Based on ECG Analysis Using CNN," *Bulletin of Electrical Engineering and Informatics (BEEI),* vol. 9, no. 3, 2020.

[15]   Kumari and V Anuja Chitra, "Classification of diabetes disease using support vector machine," *International Journal of Engineering Research Applications,* vol. 3, no. 2, pp. 1797-1801, 2013.

[16]   H. F. Al-Yasriy, "The IQ-OTH/NCCD lung cancer dataset URL https://www.kaggle.com/hamdallak/the-iqothnccd-lung-cancer-dataset," *kaggle.com.*

[17]   H. F. Al-Yasriy, "The IQ-OTHNCCD lung cancer dataset URL https://data.mendeley.com/datasets/bhmdr45bh2/1," *Mendeley Data,* 2020.

[18]   E. Emirzade, "A Computer Aided Diagnosis System for Lung Cancer Detection using SVM," Near East University, 2016.

[19]   Alshayeji, Mohammad H Al-Rousan, Mohammad Hassoun, and D. T., "Detection method for counterfeit currency based on bit-plane slicing technique," *International Journal of Multimedia Ubiquitous Engineering,* vol. 10, no. 11, pp. 225-242, 2015.

[20]   Ito, Kazufumi Xiong, and Kaiqi, "Gaussian filters for nonlinear filtering problems," *IEEE transactions on automatic control,* vol. 45, no. 5, pp. 910-927, 2000.

[21]   Deng, Guang Cahill, and LW, "An adaptive Gaussian filter for noise reduction and edge detection," in *1993 IEEE conference record nuclear science symposium and medical imaging conference*, pp. 1615-1619: IEEE, 1993.

[22]   H. Huang, "Biomedical image processing," *Critical reviews in bioengineering,* vol. 5, no. 3, pp. 185-271, 1981.

[23]   H. J. Vala and A. Baxi, "A review on Otsu image segmentation algorithm," *International Journal of Advanced Research in Computer Engineering Technology,* vol. 2, no. 2, pp. 387-389, 2013.

[24]   N. Senthilkumaran and S. Vaithegi, "Image segmentation by using thresholding techniques for medical images," *Computer Science Engineering: An International Journal,* vol. 6, no. 1, pp. 1-13, 2016.

[25]   D. Zheng, Y. Zhao, and J. Wang, "Features extraction using a gabor filter family," in *Proceedings of the sixth Lasted International conference, Signal and Image processing, Hawaii*, 2004.

[26]   P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using GLCM approach," *International journal of scientific research publications,* vol. 3, no. 5, pp. 1-5, 2013.

[27]   X. Wang, X. Ding, and C. Liu, "Gabor filters-based feature extraction for character recognition," *Pattern recognition,* vol. 38, no. 3, pp. 369-379, 2005.

[28]   R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, cybernetics,* no. 6, pp. 610-621, 1973.

[29]   N. Zayed and H. A. Elnemr, "Statistical analysis of Haralick texture features to discriminate lung abnormalities," *Journal of Biomedical Imaging,* vol. 2015, p. 12, 2015.

[30]   Er. Kanchan Sharma, Er. Priyanka , Er. Aditi Kalsh, and E. K. Saini, "GLCM and its Features," *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE),* vol. Volume 4, no. Issue 8, August 2015.

[31]   C. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining knowledge discovery,* vol. 2, no. 2, pp. 121-167, 1998.

[32]   C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, no. 3, pp. 273-297, 1995.

[33]   B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144-152: ACM, 1992.

[34]   S. Ganesan, T. Subashini, and K. Jayalakshmi, "Classification of X-rays using statistical moments and SVM," in *2014 International Conference on Communication and Signal Processing*, pp. 1109-1112: IEEE, 2014.

[35]   A. C. G. Thome, "SVM classifiers – concepts and applications to character recognition, advances in character recognition," 2012.

[36]   T. R. Golub *et al.*, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science,* vol. 286, no. 5439, pp. 531-537, 1999.

[37]   W. S. Noble, "What is a support vector machine?," *Nature biotechnology,* vol. 24, no. 12, p. 1565, 2006.

[38]   D. G. Altman and J. M. Bland, "Diagnostic tests. 1: Sensitivity and specificity," *British Medical Journal,* vol. 308, no. 6943, p. 1552, 1994.

[39]   L. Torgo and R. Ribeiro, "Precision and recall for regression," in *International Conference on Discovery Science*, pp. 332-346: Springer, 2009.

[40]   J. Lee, "Lung Cancer Detection Method using Convolution Neural Network," *University of Kyung Hee.*