International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Comparative Performance Analysis of Lung Cancer Detection using Naïve Bayes, Support Vector Machine, K-Nearest Neighbor and Decision Tree

Mohtar Yunianto[1*], Suparmi Suparmi[2], Cari Cari[3], Tonang Dwi Ardyanto[4]

Submitted: 10/11/2022 **Accepted:** 11/02/2023

Abstract: This study aims to determine the best classification technique for lung cancer detection. Four different machine learning algorithms are implemented, which are Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor Classifier (KNN), and Decision Tree (DT). The classification was carried out on 140 CT Scan data from the Lung Image Database Consortium image collection (LIDC-IDRI) dataset. Furthermore, the preposition started with a variety of filtering methods, The segmentation used was Otsu thresholding, which was textured with extraction using 11 features. The best results were obtained using DT, Low pass filter, and GLCM segmentation angle of 45^0 with performance results of 99.00% accuracy, 100.00% sensitivity, and 98.04% specificity for training data, as well as 96.25% accuracy, 95.12% sensitivity and 97.44 % specificity for test data.

Keywords: *Comparative, Decision tree, K-nearest neighbor, Lung cancer, Naïve Bayes, Support vector machine*

1. Introduction

Cancer occurs due to cell changes, which cause uncontrolled cell growth and division. It can also cause tumours, impairment of the immune system, and abnormalities that prevent the body from functioning properly [1]. Furthermore, lung cancer is the leading cause of mortality among other types globally, accounting for 18.4% of all death cases [2, 3]. It is generally detected between the age of 55 and 70 years because early diagnosis is quite difficult. The presence of small nodules in the organ is an early sign of abnormalities and they have the potential to become cancerous [4].

The condition can be detected through X-ray examination, which provides different images of a normal lung or cancerous lung [5]. As technology advances in the health sector, medical imaging is often used in the clinical diagnosis and treatment of various diseases [6]. Computed Tomography (CT) has been widely used to detect lung disorders, including pneumoconiosis, pneumonia, lung disease, edema, and cancer [7]. The image results obtained from the scan are often blurry or lack contrast, which makes the diagnosis and prognosis very difficult, especially in the early stages with small cancer cells [3]. This is due to the presence of noise that can reduce the quality. Therefore, this defect must be removed or reduced to make the information in the image clearer [8].

Technological developments in CT increased the detail of the anatomical slices of the body with smaller sizes and also provided better image quality. This then causes an increase in the number of datasets generated, where one scanning process can produce 500 slices [9]. A radiologist observing one slide often requires 2 – 3.5 minutes [10] with 68% accuracy in examining images and can be detected up to 82% more accurately than two radiologists [11]. The most challenging task performed on data generated by CT scans is the analysis and evaluation, hence, the assistance of a Computer-Aided Detection (CAD) is needed [12]. Image processing techniques with CAD have gained attention in all sectors, especially health, in terms of lung cancer detection [13].

Digital image processing consists of several stages, where the first step is image preprocessing to reduce noise and improve quality. It then segments to distinguish the region of interest (ROI) from other structures, followed by feature extraction to extract different features from the image. The final stage is classification, which is performed to evaluate and diagnose ROI based on the extracted features [14,15]. Various studies showed that the use of these methods can outperform radiologists, and has a 38-100% accuracy in detecting nodules as well as their location [16].

[1] *Physics Department, Universitas Sebelas Maret, Surakarta, Indonesia ORCID ID : 0000-0003-3715-4989*
[2] *Physics Department, Universitas Sebelas Maret, Surakarta, Indonesia ORCID ID : 0000-0003-3358-6469*
[3] *Physics Department, Universitas Sebelas Maret, Surakarta, Indonesia ORCID ID : 0000-0002-5717-1156*
[4] *Clinical Pathology Department, Universitas Sebelas Maret, Surakarta, Indonesia, ORCID ID : 0000-0003-4267-7282*
* *Corresponding Author Email: mohtaryunianto@staff.uns.ac.id*

2. Related Works

The first CAD used to detect cancer nodules was carried out in the late 1980s with the limited ability of computers to perform image analysis [17]. Technological developments and the discovery of computers with high performance in computing processes and artificial neural networks have increased the development of studies in this field.

Kulkarni et al. classified lung cancer stages starting with the pre-processing using a Gabor filter, followed by segmentation with a marker-based watershed. Subsequently, feature extraction was performed using geometric features, followed by SVM for the classification process [18]. Aggarwal et al. [19] classified 90 images using a median filter, thresholding and morphological closing operation. The feature extraction used eight geometric features, while the classification with LDA obtained an 84% accuracy and 53.33% specificity.

Magdy et al. compared four types of classification, namely NN, SVM, NB, and linear classifier, for 83 CT image data from TCIA [12]. The preposition stage using a wiener filter was followed by AM-FM modelling Partial Least Squares Regression. The KNN obtained an Accuracy of 64%, 55% Sensitivity, and 72% Specificity, the SVM had a 90% accuracy, 85% sensitivity, and 97% Specificity, while Naïve Bayes had 82% accuracy, 82% sensitivity, and 82% specificity. The linear classifier obtained an accuracy of 95% with 94% sensitivity and 97% specificity.

Punithavathy et al. [20] recorded an accuracy of 92.67% for detecting lung nodules using the wiener filter with CLAHE technique, ROI extraction, feature extraction using 3 GLCM features, and classification with FCM Clustering. Meanwhile, [21] compared three classification methods, namely MLP, KNN, and SVM, for 60 CT LIDC images using median filters, histogram equalization, region growing, thresholding in the preposition, segmentation processes as well as feature extraction using five morphological features. The results showed that MLP obtained an accuracy of 90.41%, 73.55% sensitivity, and 94.68% specificity, while the KNN had an accuracy of 91.20 %, 81.76% sensitivity, and 93.59% specificity. The SVM had an accuracy of 90.60 %, 73.44% sensitivity, and 94.94% specificity.

Taher et al., with 100 image data from the Tokyo Center for lung cancer, compared two classification methods, namely ANN and SVM, with the Bayesian framework in terms of their preposition stage, mean shift technique for segmentation, and features Nucleus to Cytoplasm (NC) ratio, perimeter, density, curvature, circularity and Eigen ratio for feature extraction. Based on the results, ANN had a sensitivity of 94%, 83% specificity, and 90% accuracy of 90%, while SVM had 97% sensitivity, 96% specificity, and 97% accuracy [22]. The values obtained by Riti et al. for the three parameters was 85% [23], using the Otsu thresholding for the segmentation process. Furthermore, the convexity, solidity, circularity, and compactness for feature extraction and classification were carried out with MLP. Hasnely et al. [24] also used Otsu thresholding with MLP to classify cancer nodules with ROI and feature extraction variations using six statistical methods (Histogram) and 5 GLCM features. For the histogram, an accuracy of 80% was obtained, along with 88% sensitivity and 72% specificity. For GLCM, an accuracy of 96% was recorded with a sensitivity of 96%, and specificity of 96%. The combination of both methods produced an accuracy of 98%, 96% sensitivity, and 96% specificity.

El-regaily et al. [25] used 400 LIDC CT scan images and using Thresholding, thorax, lung extractions, and reconstruction for the preposition and segmentation processes. Structure extraction, tabular structure elimination, and Rule-based classifier obtained an accuracy of 70.53%, 77.77% sensitivity, and 69.5% specificity. A total of 4,682 image data from TCIA were used by Kalaivani et al. with Histogram Equalization, Binarization in the preposition process, Region props function for feature extraction, and backpropagation classification with an accuracy of 78% [26]. Lavanya et al. [27] used data from LIDC with prepositions using a winer filter, a segmentation process using the FLICM algorithm, feature extraction with three geometric features, and classification using BPN. An accuracy of 85.9% was recorded, along with 90.87% sensitivity and 84.77% specificity.

Bao et al. compared four classification methods: SVM, BPNN, PNN, and k-means Clustering. The proposition stage was carried out with High boost filtering, followed by a segmentation process using FCM Clustering Algorithm, while statistical methods were used for feature extraction. The accuracy values for SVM, PNN, BPNN, and K-means clustering were 85%, 82%, 86%, and 81%, respectively [28]. Variations of feature extraction in the form of LTCoP and LBP were performed by Bruntha et al. [29] for 50 images from LIDC. The preposition stage used median filtering, followed by intensity thresholding segmentation, while the classification was performed with SVM. The LTCop and LBP had 91.5% and 89.2%, respectively. Furthermore, Karthiga et al [30] compared the FMSVM, MLP, KNN, SVM and I-NBC classifications, which had accuracies of 98%, 52%, 46%, 96%, and 98%, respectively. The stages of Minimum Mean Square Error preposition, ROI segmentation, and feature extraction using 10 parameters were also compared.

Image data from LIDC was tested by Narayanan et al. using the preposition median filtering stage for segmentation with thresholding. The feature extraction stage used seven geometric features, while the classification used ANN with an accuracy of 92.2% [31]. Perumal et al. [32] used the

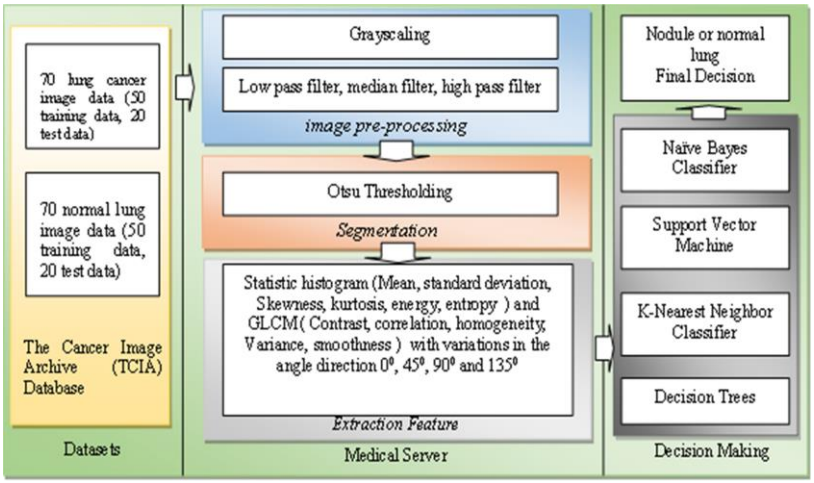


Fig 1. Block diagram of the proposed system

Enhanced Artificial Bee Colony (EABC) Optimization classification and obtained a sensitivity of 92.4%. The prepositions were carried out with CLAHE, followed by segmentation using Sobel, while feature extraction used five levels of haar wavelet transform.

Listyalina et al. [33] designed software with artificial neural networks from X-Ray images. The median filter served as the filtering method, while the discrete cosine transformation and adaptive histogram equalization were used for image feature extraction. The level of accuracy obtained was 72.97% in detecting lung cancer from X-ray results. Firdaus et al. [34] discussed the development of an image-based lung cancer detection system with feature detection in the form of GLCM and SVM as the classification method. In determining lung cancer diagnosis as benign or malignant, this system had an accuracy of 83.33%.

Furthermore, Bhatt & Soni [35] proposed an approach for classifying the condition with a classification technique in the form of CBIR-based BOF and grouping pattern categories with K-means clustering. This method provided a training accuracy of 99% and a testing accuracy of 98.56%. A total of 120 CT Scan image data were tested by Yunianto et al. [36], using a median filter in the preposition process, and Otsu Thresholding in the segmentation. It was then extracted using the GLCM feature with variations in angle direction, and the classification process used Naïve Bayes with an accuracy of 88.33%.

3. Material and Method

Based on a review of relevant studies in the related works, the proposed method of classifying images was based on the determined procedure. Data acquisition was carried out for .png format images. They were then processed in the pre-processing stage to improve the quality and eliminate noise by choosing between low pass, median, and high pass filtering. Segmentation was performed using Otsu thresholding, while feature extraction was used to determine the characteristics possessed by the image using the statistical method, namely Histogram and GLCM with variations in the angle direction of 0^0, 45^0, 90^0, and 135^0. Furthermore, the database obtained from the extraction process was then classified using the NB, SVM, KNN, and DT. These four classification methods were included in the Top 10 algorithms in data mining [37]. Fig. 1 shows a block diagram of the proposed system.

3.1. Datasets

The research data was obtained from The Lung Image Database Consortium image collection (LIDC-IDRI) Dataset, which was downloaded from https://nbia.cancerimagingarchive.net/nbia-search/ using the NBIA Data Retriever [38] in DICOM format for further analysis. Furthermore, the 3D Slicer 4.11[39] was used to retrieve the slice number in the source database, marked by four radiologists, and then converted into .png format. A total of 140 images were used, and they consist of 100 training and 40 testing data.

3.2. Image pre-processing

Image quality improvement was done before processing to increase the quality and correct errors [40]. The Grayscaling stage [41] and variations of the filtering process in the form of low pass, median, and high pass filters [42-45] were used for this process. The low pass filter process passes low-frequency and attenuates signals with higher frequencies than the cutoff frequency [46]. Median filtering is an effective technique used to reduce noise without removing the edges [47-50]. The high pass filter maintains high frequencies, which makes the image clearer with sharper edges [51].

3.3. Segmentation

Image segmentation involves separating objects from other

items or the backgrounds from an image. The thresholding method has a threshold value used to convert a grayscale image into the binary form [52]. This study used the Otsu Thresholding method [53, 54].

3.4. Feature Extraction

The feature extraction or separation process aims to find some texture features of the image, thereby facilitating accurate classification and segmentation. The classification depends on the quality of the features produced. In this study, six parameters of the first order Histogram-based Matrix were used: mean, standard deviation, Skewness, Kurtosis, Energy, and Entropy [55, 56]. Furthermore, five parameters of second-order GLCM were used to evaluate image features related to order histogram statistics, namely contrast, correlation, homogeneity, variance, and smoothness [57-59].

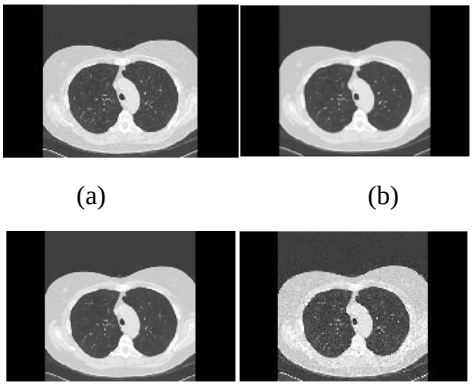
3.5. Classification

The classification used was Naïve Bayes [60] with a probability in the maximum posterior, which only took the largest class. The posterior probability was only calculated based on the main form [61]. The support vector machine (SVM), which is a supervised learning algorithm model used in data analysis for classification and regression, needs to be developed at AT&T Bell Laboratory [62-66]. Furthermore, the K-Nearest Neighbor (KNN) is a method for classifying objects based on the closest learning data. A decision tree was combined with the supervised machine learning method [67-69].

4. Results and Discussion

4.1. Preprocessing

The inputted .png format image sample was converted to grayscale and then processed into low pass, median, and high pass filters to determine the form with the best performance.

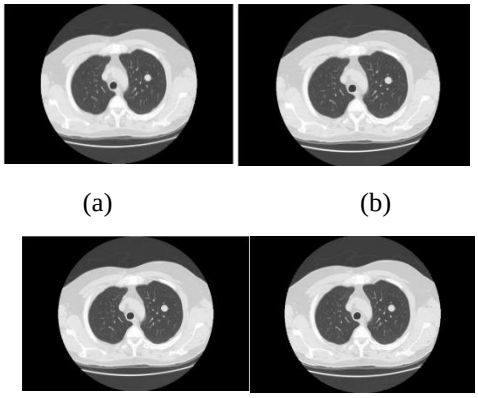


(a) (b)

(c) (d)

Fig 2. Image of normal lung (a) input image, (b) low pass filtering (c) median filtering, (d) high pass filtering.

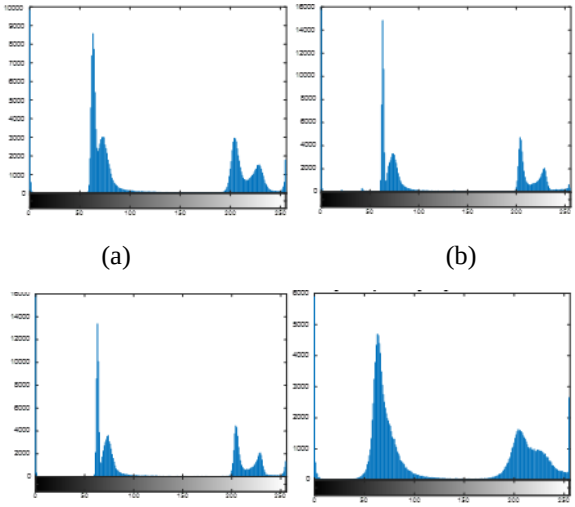
The changes that occurred cannot be seen visually, as shown in Fig. 2 and 3, but the histogram in the program has different values. This was indicated by the histogram results produced in Fig. 4 and 5.



(a) (b)

(c) (d)

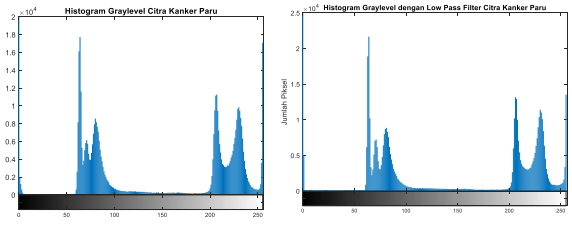
Fig 3. Image of lung cancer nodules (a) input image, (b) low pass filtering, (c) median filtering, (d) high pass filtering.



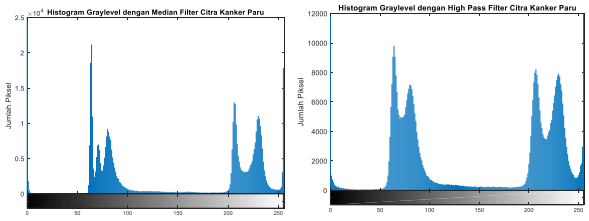
(a) (b)

(c) (d)

Fig 4. Histogram of normal lung image (a) input image, (b) low pass filtering, (c) median filtering, (d) high pass filtering.



(a) (b)



(c) (d)

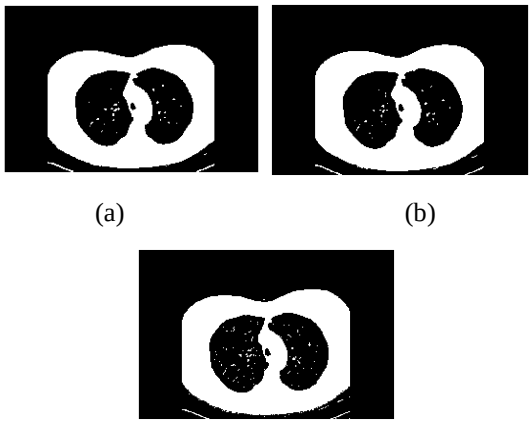
Fig 5. Histogram of lung cancer nodules image (a) input image, (b) low pass filtering, (c) median filtering, (d) high pass filtering

Histograms from the low pass and medium filtering have almost the same intensity and number of pixels. They can both reduce noise by making the intensity more even and smoothing the image. The grayscale segmentation and low pass filter obtained less noise and a higher average accuracy value [70]. In MRI, the median type processing can also remove noise [71].

The image produced in the high pass filter process is sharper than the original form. Furthermore, the edge pixels are shown to be brighter, while the non-edge pixels are darker. During improvement with the high pass filter method, the image produced was too sharp compared to the original [72].

4.2. Segmentation

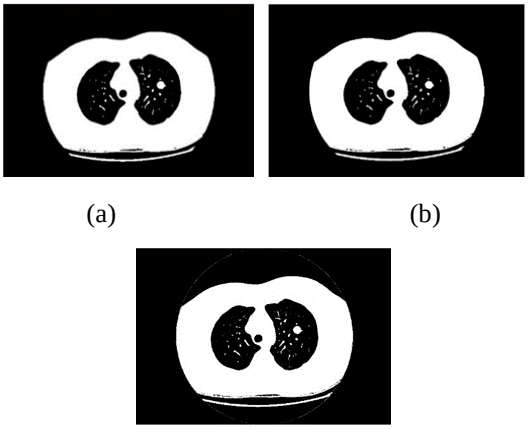
The segmentation was carried out with used the Otsu Thresholding method, where the threshold value is automatically calculated based on the input image. This method was performed with a discriminant analysis approach [73]. Therefore, the segmentation can now separate the lung object from its background. Each image has a different intensity of gray level and a threshold value. The Otsu thresholding segmentation for the normal lung image and lung cancer nodule image is presented in Fig. 6 and 7, respectively.



(a) (b)

(c)

Fig 6. (a) Otsu thresholding using a low pass filter for normal lung images, (b) Otsu thresholding using a median normal lung image filter, (c) Otsu thresholding using a high pass filter for normal lung images



(a) (b)

(c)

Fig. 7 (a) Otsu thresholding using a low pass filter for lung cancer nodules, (b) Otsu thresholding using a median filter for lung cancer nodules, (c) Otsu thresholding using a high pass filter for lung cancer nodules

Fig. 6 and 7 show that the edges of the image were detected after using the Otsu thresholding method. The lung object and the image background can be seen, hence, the foreground and background can be distinguished with this process. A CT scan of the organ consists of high-intensity pixels in the body and low-intensity pixels in the lungs and surrounding cavities, which can be separated with a thresholding process by automatically calculating the optimal threshold value [74]. In Fig. 6, the image results were in the form of a dark lung, with white spots as the remaining noise. Meanwhile, in Fig. 7, there are nodules on the CT scan image of the cancerous organ, where the size looks larger than the surrounding noise. The nodule image has more white spots than the normal form. These results are similar to Otsu thresholding in brain MRI image study using the threshold method and GLCM with the KNN algorithm. Threshold white patches are part of the tumor [75].

Table 1. The feature extraction on the lung cancer nodules image segmented using a low pass filter.

Angle direction	*contrast*	*correlation*	*Energy*	*Homogeneity*	*Mean*	*St. Deviation*	*Entropy*	*Variance*	*Smoothness*	*Kurtosis*	*Skewness*
0^0	0.005	0.987	0.613	0.997	0.257	0.437	0.822	0.146	1.000	2.235	1.111
45^0	0.008	0.978	0.609	0.996	0.257	0.437	0.822	0.146	1.000	2.235	1.111
90^0	0.006	0.983	0.611	0.997	0.257	0.437	0.822	0.146	1.000	2.235	1.111
135^0	0.009	0.977	0.609	0.996	0.257	0.437	0.822	0.146	1.000	2.235	1.111

Table 2. The feature extraction on normal lung images segmented using a low pass filter

Angle direction	*contrast*	*correlation*	*Energy*	*Homogeneity*	*Mean*	*St. Deviation*	*Entropy*	*Variance*	*Smoothness*	*Kurtosis*	*Skewness*
0^0	0.007	0.977	0.663	0.996	0.208	0.406	0.737	0.132	1.000	3.076	1.441
45^0	0.013	0.956	0.657	0.993	0.208	0.406	0.737	0.132	1.000	3.076	1.441
90^0	0.010	0.969	0.660	0.995	0.208	0.406	0.737	0.132	1.000	3.076	1.441
135^0	0.013	0.960	0.657	0.993	0.208	0.406	0.737	0.132	1.000	3.076	1.441

Table 3. The feature extraction on the lung cancer nodules image segmented using a median filter

Angle direction	*contrast*	*correlation*	*Energy*	*Homogeneity*	*Mean*	*St. Deviation*	*Entropy*	*Variance*	*Smoothness*	*Kurtosis*	*Skewness*
0^0	0.005	0.987	0.613	0.997	0.256	0.437	0.821	0.146	1.000	2.245	1.116
45^0	0.008	0.978	0.610	0.996	0.256	0.437	0.821	0.146	1.000	2.245	1.116
90^0	0.007	0.983	0.612	0.997	0.256	0.437	0.821	0.146	1.000	2.245	1.116
135^0	0.009	0.977	0.609	0.996	0.256	0.437	0.821	0.146	1.000	2.245	1.116

Table 4. The feature extraction on normal lung images segmented using a median filter

Angle direction	*contrast*	*correlation*	*Energy*	*Homogeneity*	*Mean*	*St. Deviation*	*Entropy*	*Variance*	*Smoothness*	*Kurtosis*	*Skewness*
0^0	0.008	0.977	0.666	0.996	0.205	0.404	0.732	0.131	1.000	3.135	1.461
45^0	0.013	0.959	0.660	0.993	0.205	0.404	0.732	0.131	1.000	3.135	1.461
90^0	0.010	0.968	0.663	0.995	0.205	0.404	0.732	0.131	1.000	3.135	1.461
135^0	0.013	0.956	0.660	0.993	0.205	0.404	0.732	0.131	1.000	3.135	1.461

Table 5. The feature extraction on the lung cancer nodules image segmented using a high pass filter

Angle direction	*contrast*	*correlation*	*Energy*	*Homogeneity*	*Mean*	*St. Deviation*	*Entropy*	*Variance*	*Smoothness*	*Kurtosis*	*Skewness*
0^0	0.008	0.978	0.609	0.996	0.257	0.437	0.823	0.147	1.000	2.231	1.110
45^0	0.011	0.972	0.607	0.995	0.257	0.437	0.823	0.147	1.000	2.231	1.110
90^0	0.010	0.975	0.608	0.995	0.257	0.437	0.823	0.147	1.000	2.231	1.110
135^0	0.011	0.971	0.606	0.994	0.257	0.437	0.823	0.147	1.000	2.231	1.110

Table 6. The feature extraction on normal lung images segmented using high pass filter

Angle direction	*contrast*	*correlation*	*Energy*	*Homogeneity*	*Mean*	*St. Deviation*	*Entropy*	*Variance*	*Smoothness*	*Kurtosis*	*Skewness*
0^0	0.016	0.950	0.654	0.992	0.208	0.506	0.738	0.133	1.000	3.068	1.438
45^0	0.020	0.939	0.650	0.990	0.208	0.506	0.738	0.133	1.000	3.068	1.438
90^0	0.019	0.942	0.651	0.990	0.208	0.506	0.738	0.133	1.000	3.068	1.438
135^0	0.020	0.939	0.650	0.990	0.208	0.506	0.738	0.133	1.000	3.068	1.438

4.3. Feature Extraction

Feature extraction used six parameters of first-order Histogram-based Matrix and five parameters of second-order GLCM with a distance of 1 pixel as well as angles of 0°, 45°, 90°, and 135°. Furthermore, Tables 1 – 6 show the results of the average feature extraction values for lung cancer nodules and normal lung images with various filters.

The correlation value revealed the relationship between 1 pixel and its neighboring variant, the energy and homogeneity values are a measure of the similarity and concentration of intensity pairs in the image. The smoothness value indicates the smoothness of the intensity. Meanwhile, the kurtosis and skewness indicate the sharpness and relative slope of the histogram curve.

Based on the statistical value obtained, the normal lung images have a higher contrast than the cancerous nodule. Furthermore, the correlation, energy, homogeneity, smoothness, kurtosis, and skewness of the lung cancer nodule image was higher than the normal.

The irregularity of the gray intensity and the average value in the image as indicated by the entropy and mean values. The standard deviation shows the histogram variance obtained from the image. Furthermore, the normal lung has a higher mean, standard deviation, and entropy compared to the feature extracted lung cancer nodule. The variance is a variation of the co-occurrence matrix element, which shows the image with a small gray degree transition. Hence, the value recorded for both of them was small.

The feature extraction was compared with the input value, where 0 indicated the image of lung cancer nodules, and 1 indicated that of the normal lung. It was then used as a database for the classification process.

Table 7. The performance of the training data classification process using Naïve Bayes

Angle direction	Low pass filter			Median filter			High pass filter		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
0^0	97.00%	96.08%	97.96%	97.00%	96.08%	97.96%	97.00%	97.96%	96.08%
45^0	97.00%	96.08%	97.96%	97.00%	96.08%	97.96%	95.00%	94.12%	95.92%
90^0	98.00%	98.00%	98.00%	97.00%	96.08%	97.96%	95.00%	94.12%	95.92%
135^0	97.00%	96.08%	97.96%	95.00%	95.92%	94.12%	95.00%	94.12%	95.92%

Table 8. Performance of the training data classification process using SVM

Angle direction	Low pass filter			Median filter			High pass filter		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
0^0	57.00%	70.59%	54.22%	56.00%	68.75%	53.57%	56.00%	68.75%	53.57%
45^0	56.00%	66.67%	53.66%	58.00%	72.22%	54.88%	59.00%	80.00%	55.29%
90^0	55.00%	66.67%	52.94%	54.00%	64.29%	52.33%	56.00%	68.75%	53.57%
135^0	58.00%	70.00%	55.00%	55.00%	66.67%	52.94%	56.00%	68.75%	53.57%

Table 9. The performance of the training data classification process using KNN

Angle direction	Low pass filter			Median filter			High pass filter		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
0^0	58.00%	59.52%	56.90%	54.00%	54.76%	53.45%	65.00%	65.96%	64.15%
45^0	64.00%	64.00%	64.00%	57.00%	57.78%	56.36%	73.00%	74.47%	71.70%
90^0	60.00%	60.00%	60.00%	54.00%	54.55%	53.57%	63.00%	63.83%	62.26%
135^0	67.00%	67.35%	66.67%	59.00%	60.47%	57.89%	71.00%	73.33%	69.09%

Table 10. The performance of the training data classification process using Decision Tree

Angle direction	Low pass filter			Median filter			High pass filter		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
0^0	97.00%	97.96%	96.08%	97.00%	97.96%	96.08%	95.00%	95.92%	94.12%
45^0	99.00%	100.00%	98.04%	95.00%	97.87%	92.45%	94.00%	95.83%	92.31%
90^0	98.00%	98.00%	98.00%	97.00%	97.96%	96.08%	95.00%	95.92%	94.12%
135^0	98.00%	98.00%	98.00%	97.00%	97.96%	96.08%	93.00%	95.74%	90.57%

4.4. Classification

The classification process used the NB, SVM, KNN, and DT methods. The database from the feature extraction was entered as a dataset. These methods can calculate every opportunity for existing features. Hence, new values obtained are used to calculate data accuracy. A confusion matrix was combined with the actual and predictive values, which were useful for measuring the performance level of the program test results. The True Positive (TP) parameter was used when the normal image was read normally, while the False Positive (FP) was used when the normal image was read as a lung cancer nodule. The performance results were calculated based on the equation [76].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}\ x100\% \quad (1)$$

$$Sensitivity = \frac{TP}{TP+FN}\ x100\% \quad (2)$$

$$Specificity = \frac{TN}{TN+FP}\ x100\% \quad (3)$$

By using equations 1-3, the accuracy, sensitivity, and specificity values were obtained from the applied classification process, as shown in Tables 7-10.

Based on the performance data for the four types of classification methods in Table 7-10, the type with the best value is presented in Table 10, namely the DT classification with a GLCM 45^0 angle direction and low pass filter, which has a 99.00% accuracy, 100.00% sensitivity, and 98.04% specificity. In the training program, the variation was used as input for the classification process in the test data. In the testing process using the test data program, a total of 40 images were entered. Using equations 1-3, the accuracy value was 96.25%, while the sensitivity and specificity were 95.12% and 97.44%, respectively. Comparing these results with previous studies in the related works section [17-36] has the best accuracy value.

5. Conclusion

Image classification of lung cancer and normal lung was done using the NB, SVM, KNN, and DT methods. Based on the method used, the best performance value was obtained from the training process using variations of low pass filter, otsu thresholding segmentation, GLCM feature extraction

with 45 angle direction, and classification using DT, which had 99.00% accuracy, 100.00% sensitivity, and 98.04% specificity. The performance for the test data obtained 96.25% accuracy, 95.12% sensitivity, and 97.44% specificity. Therefore, the method is good for predicting and classifying lung cancer and normal lung CT Scan image results.

Author contributions

Mohtar Yunianto[1*]: Dataset Preparation, Methodology, Writing-Reviewing and Editing. **Suparmi Suparmi**[2]: Conceptualization, Writing-Original draft preparation**. Cari Cari**[3]: Field study, Conceptualization, Validation. **Tonang Dwi Ardiyanto**[4]: Dataset Preparation, Field study, Writing-Reviewing and Editing.

Conflict of interest

The authors declare that the present study has no conflict of interest

Acknowledgements

The authors would like to thank LPPM UNS for providing the fund through the Doctoral Dissertation Research Grant with the contract number: 260/UN27.22/HK.07.00/2021. The authors would also like to express gratitude to Ms Haya Alvinesha, Ms Armilya, Ms Meilina, Ms Umi Salamah, Mr Nuryani and Mr Fuad Anwar for the discussion and assistance in this research.

Data Availability Statement

Data available in a publicly accessible repository (The Lung Image Database Consortium image collection (LIDC-IDRI) Dataset, https://nbia.cancerimagingarchive.net/nbia-search/) that does not issue DOIs.

Ethical approval (and / or if people are involved) Ethical approval

The study did not require ethical approval.

Ethical confirmation

This article does not contain any studies with animals by human participants or any authors.

References

[1]. Nasser, I. M., & Abu-Naser, S. S. Lung Cancer Detection Using Artificial Neural Network. International Journal of Engineering and Information Systems (IJEAIS), 3(3), 17-23, 2019.

[2]. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 68(6), 394-424, 2018, doi: 10.3322/caac.21492

[3]. Moreno, S., Bonfante, M., Zurek, E., & Juan, H. S. Study of Medical Image Processing Techniques A lied to Lung Cancer. Proceedings of the 14th Iberian Conference on Information Systems and Technologies (CISTI), 1-6, 2019, doi: 10.23919/cisti.2019.8760888

[4]. Pawar, V. J., Kharat, K. D., Pardeshi, S. R., & Pathak, P. D. Lung Cancer Detection System Using Image Processing and Machine Learning Techniques. International Journal of Advanced Trends in Computer Science and Engineering, 9(4), 5956-5963, 2020, doi: 10.30534/ijatcse/2020/260942020

[5]. Listyalina, L. Peningkatan Kualitas Citra Foto Rontgen Sebagai Media Deteksi Kanker Paru. Respati Jurnal Ilmiah Teknologi Informasi, 12(4). 110-118, 2017, doi: 10.35842/jtir.v12i34.1

[6]. Gao, J., Wang, B., Wang, Z., Wang, Y., & Kong, F. A Wavelet Transform-Based Image Segmentation Method. Optik, 208, 164123, 2020, doi: 10.1016/j.ijleo.2019.164123

[7]. Hollings, N., & Shaw, P. Diagnostic imaging of lung cancer. European Respiratory Journal, 19(4), 722–742, 2002, doi: 10.1183/09031936.02.00280002

[8]. Trisnawati, L., & Hakim, L. Segmentasi Citra Ct Scan Lung Menggunakan Deteksi Tepi Sobel Dan Metode Distance Regularized Level Set Evolution (DRLSE). Explore IT: Jurnal Keilmuan dan Aplikasi Teknik Informatika, 10(1), 1-13, 2018, doi: 10.35891/explorit.v10i1.1670

[9]. Demir, Ö., & Çamurcu, A.Y. Computer-aided detection of lung nodules using outer surface features. Bio-Medical Materials and Engineering, 26(1), S1213–S1222, 2015, doi: 10.3233/BME-151418

[10]. Bogoni, L., Ko, J. P., Alpert, J., Anand, V., Fantauzzi, J., Florin, C. H., Koo, C.W., Mason, D., Rom, W., Shiau, M., Salganicoff, M., & Naidich, D, P. Impact of a computer-aided detection (CAD) system integrated into a picture archiving and communication system (PACS) on reader sensitivity and efficiency for the detection of lung nodules in thoracic CT exams. Journal of Digital Imaging, 25, 771–781, 2012, doi: 10.1007/s10278-012-9496-0

[11]. Al Mohammad, B., Brennan, P. C., & Mello-Thoms, C. A review of lung cancer screening and the role of computer-aided detection. Clinical Radiology, 72(6), 433–442, 2017, doi: 10.1016/j.crad.2017.01.002

[12]. Magdy, E., Zayed, N., & Fakhr, M. Automatic classification of normal and cancer lung CT images using multiscale AM-FM features. Journal of Biomedical Imaging, 11, 2015, doi: 10.1155/2015/230830

[13]. Bhattacharya, S., Maddikunta, P. K. R., Pham, Q. V., Gadekallu, T. R., Chowdhary, C. L., Alazab, M., & Piran, M. J. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A

survey. Sustainable Cities and Society, 65, 102589, 2-18, 2021, doi: 10.1016/j.scs.2020.102589

[14]. Wayne Niblack. An introduction to digital image processing. Strandberg Publishing Company, DNK, 1985, https://dl.acm.org/doi/abs/10.5555/4901

[15]. Lee, H., Matin, T., Gleeson, F., & Grau, V. Medical image computing and computer assisted intervention 2017., in Miccai 10433, 108–115, 2017.

[16]. Dou, Q., Chen, H., Yu L., Qin, J., & Heng, P. A. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. IEEE Transactions on Biomedical Engineering, 64(7), 1558-1567, 2017, doi: 10.1109/tbme.2016.2613502

[17]. Nasrullah, N., Sang, J., Alam, M. S., Mateen, M., Cai, B., & Hu, H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors, 19(17), 3722, 2019, doi: 10.3390/s19173722

[18]. Kulkarni, A., & Panditrao, A. Classification of lung cancer stages on CT scan images using image processing. Proceedings of the 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, 1384-1388, 2014, doi: 10.1109/icaccct.2014.7019327

[19]. Aggarwal, T., Furqan, A., & Kalra, K. Feature extraction and LDA based classification of lung nodules in chest CT scan images. Proceedings of the 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 1189-1193, 2015, doi: 10.1109/icacci.2015.7275773

[20]. Punithavathy, K., Ramya, M. M., & Poobal, S. Analysis of statistical texture features for automatic lung cancer detection in PET/CT images. Proceedings of the 2015 International Conference on Robotics, Automation, Control and Embedded Systems (RACE), 1-5, 2015, doi: 10.1109/race.2015.7097244

[21]. Farahani, F. V., Ahmadi, A., & Zarandi, M. F. Lung nodule diagnosis from CT images based on ensemble learning. Proceedings of the 2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 1-7, 2015, doi: 10.1109/cibcb.2015.7300281

[22]. Taher, F., Werghi, N., & Al-Ahmad, H. Computer aided diagnosis system for early lung cancer detection. Algorithms, 8(4), 1088-1110, 2015, doi: 10.3390/a8041088

[23]. Riti, Y.F., Nugroho, H.A., Wibirama, S., Windarta, B. & Choridah, L. Feature extraction for lesion margin characteristic classification from CT Scan lungs image. Proceedings of the 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 54-58, 2016, doi: 10.1109/icitisee.2016.7803047

[24]. Hasnely, H., Nugroho, H.A., Wibirama, S., Windarta, B. & Choridah, L. Texture feature extraction for the lung lesion density classification on computed tomography scan image. Communications in Science and Technology, 1(1), 2016, doi: 10.21924/cst.1.1.2016.14

[25]. El-Regaily, S. A., Salem, M. A. M., Aziz, M. H. A., & Roushdy, M. I. Lung nodule segmentation and detection in computed tomography. Proceedings of the Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), 72-78, 2017, doi: 10.1109/intelcis.2017.8260029

[26]. Kalaivani, S., Chatterjee, P., Juyal, S., & Gupta, R. Lung cancer detection using digital image processing and artificial neural networks. Proceedings of the International conference of Electronics, Communication and Aerospace Technology (ICECA), 2, 100-103, 2017, doi: 10.1109/iceca.2017.8212773

[27]. Lavanya, M., and Kannan, P. M. Lung lesion detection in CT scan images using the fuzzy local information cluster means (FLICM) automatic segmentation algorithm and back propagation network classification. Asian Pacific Journal of Cancer prevention, 18(12), 3395-3399, 2017, doi: 10.22034/fapjcp.2017.18.12.3395

[28]. Bao, J., Walliander, M., Kovacs, F., Hemmes, A., Sarhadi, V., Lundin, J., Hovarth P., & Verschuren, E. Spa-RQ: an image analysis tool to visualise and quantify spatial phenotypes a lied to nonsmall cell lung cancer. science report, 9. 17613, 2019, doi: 10.1038/s41598-019-54038-9

[29]. Bruntha, P. M., Pandian, S. I. A., Anitha, J., Mohan, P., & Dhanasekar, S. Local Ternary Co-occurrence Patterns based Lung Nodules Detection. Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS),489-492,2020, doi: 10.1109/icaccs48705.2020.9074411

[30]. Karthiga, B., & Rekha, M. Feature extraction and I-NB classification of CT images for early lung cancer detection. Materials Today: Proceedings, 33(7), 3334-3341, 2020, doi: 10.1016/j.matpr.2020.04.896

[31]. Narayanan L. A., & Jeeva, J. B. (2015). A Computer Aided Diagnosis for detection and classification of lung nodules. Proceedings of the 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), 1-5. doi: 10.1109/isco.2015.7282242

[32]. Perumal, S., & Velmurugan, T. Lung cancer detection and classification on CT scan images using enhanced artificial bee colony optimization. International Journal of Engineering & Technology, 7(2.26), 74-79, 2018, http://dx.doi.org/10.14419/ijet.v7i2.26.12538

[33]. Listyalina, L., Utari, E. L., & Puspaningtyas, D. E. Penentuan Penyakit Paru Dengan Menggunakan Jaringan Saraf Tiruan. Simetris: Jurnal Teknik Mesin,

Elektro dan Ilmu Komputer, 11(1), 233-240, 2020, doi: 10.24176/simet.v11i1.3667

[34]. Firdaus, Q., Sigit, R., Harsono, T., & Anwar, A. Lung Cancer Detection Based On CT-Scan Images with Detection Features Using Gray Level Co-Occurrence Matrix (GLCM) and Support Vector Machine (SVM) Methods. Proceedings of the 2020 International Electronics Symposium (IES), 643-648, 2020, doi: 10.1109/IES50839.2020.9231663

[35]. Bhatt, S. D., & Soni, H. B. Image Retrieval using Bag-of-Features for Lung Cancer Classification. Proceedings of the 6th International Conference on Inventive Computation Technologies (ICICT), 531-536, 2021, doi: 10.1109/icict50816.2021.9358499

[36]. Yunianto, M., Soeparmi, S., Cari, C., Anwar, F., Septianingsih, D. N., Ardyanto, T. D., Pradana, R. F. Klasifikasi Kanker Paru Paru menggunakan Naïve Bayes dengan Variasi Filter dan Ekstraksi Ciri GLCM. Indonesian Journal of Applied Physics, 11(2), 256 - 268, 2021, doi: 10.13057/ijap.v11i2.53213

[37]. Wu, X., Kumar, V., Quinlan, R. J., Ghosh, J., Yang, Q., Motoda, H., McLachlan G.J., Ng. A., Liu. B., Yu, P. S., Zhou, Z., Steinbach, M., Hand, D. J., & Steinberg, D. Top 10 algorithms in data mining. Knowledge and information systems, 14(1), 1-37, 2008, doi: 10.1007/s10115-007-0114-2

[38]. NBIA Data Retriever https://wiki.cancerimagingarchive.net/display/NBIA/Version+4.2

[39]. 3D Slicer 4.11 https://download.slicer.org/

[40]. Malik, B., Singh, J. P., Singh, V. B. P., & Naresh, P. Lung Cancer Detection at Initial Stage by Using Image Processing and Classification Techniques. International Research Journal of Engineering and Technology (IRJET), 3(11), 781-786, 2016.

[41]. Kalaivani, S., Chatterjee, P., Juyal, S. and Gupta, R. Lung cancer detection using digital image processing and artificial neural networks. Proceedings of the 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 100-103, 2017, doi: 10.1109/iceca.2017.8212773

[42]. Yadav, A. K., Roy, R., Rajkumar., Vaishali., & Somwanshi, D. Thresholding and Morphological Based Segmentation Techniques for Medical Images. Proceedings of the 2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 1-5, 2016, doi: 10.1109/icraie.2016.7939573

[43]. Kumar, A., & Raheja, S. Edge detection using guided image filtering and enhanced ant Colony optimization. Procedia Computer Science, 173(1), 8-17, 2020, doi: 10.1016/j.procs.2020.06.003

[44]. Wu, J., Tong, T., Chen, Y., Kang, X., & Sun, W. An adversarial learning framework with cross-domain loss for median filtered image restoration and anti-forensics. Computers & Security, 112(1), 1-16, 2021, doi: 10.1016/j.cose.2021.102497

[45]. Khorsheed O, K, Produce Low-Pass and High-Pass Image Filter In Java. International Journal of Advances in Engineering & Technology, 7(3), 712-722, 2014.

[46]. Makandar, A. and Halalli, B. Image enhancement techniques using highpass and lowpass filters. International Journal of Computer Applications, 109(14), 21-27, 2015, doi: 10.5120/19256-0999

[47]. Sevani, A., Modi, H., Patel, S. & Patel, H. Implementation of image processing techniques for identifying different stages of lung cancer. International Journal of Applied Engineering Research, 13(8), 6493-6499, 2018.

[48]. Listiyani, E. Implementasi Adaptive Median Filter Sebagai Reduksi Noise Pada Citra Digital, Thesis (Undergraduate), 2013, https://repository.dinamika.ac.id/id/eprint/79/ .

[49]. Pratap, G.P., & Chauhan, R.P. Detection of Lung cancer cells using image processing techniques. Proceedings of the 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), 1-6, 2016, doi: 10.1109/icpeices.2016.7853347

[50]. Hussain, A. & Khunteta, A. Semantic Segmentation of Brain Tumor from MRI Images and SVM Classification using GLCM Features. Proceedings of the 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 38-43, 2020, doi: 10.1109/icirca48905.2020.9183385

[51]. Singh, S., Sharma, A., Mittal, M., Monika. Performance Evaluation of High Pass, Low Pass and Median filter on Webcam Pictures. Proceedings 4th International Conference on "Computing for Sustainable Global Development, 6025-6029. 2017.

[52]. Ganesh, P. S., Kumar, T. S., Kumar, M., & Kumar, M.S.R. Brain Tumor Detection and Classification Using Image Processing Techniques. Brain Tumor Detection and Classification Using Image Processing Techniques, 4(3). 32-38, 2021, DOI:10.48175/ijarsct-V4-I3-005

[53]. Katherine, K., Rulaningtyas, R. & Ain, K. CT scan image segmentation based on hounsfield unit values using Otsu thresholding method. Journal of Physics: Conference Series. 1816(1), 012080). 1-7, 2021, doi:10.1088/1742-6596/1816/1/012080

[54]. Arifin, T. Analisa Perbandingan Metode Segmentasi Citra Pada Citra Mammogram. Jurnal

Informatika, 3(2). 156-163, 2016, doi: 10.31294/ji.v3i2.1169

[55]. Radi, R, M., & Purnomo, M. H. Combination of first and second order statistical features of bulk grain image for quality grade estimation of green coffee bean. ARPN Journal of Engineering and Applied Sciences, 10(18), 8165-8174, 2015.

[56]. Jiao, Y., Ijurra, O. M., Zhang, L., Shen, D., & Wang, Q. Curadiomics: A GPU-based radiomics feature extraction toolkit. Lecture Notes in Computer Science, vol 11991. Springer, Cham., 44-52, 2020, doi: 10.1007/978-3-030-40124-5_5

[57]. Zotin, A., Hamad, Y., Simonov, K., & Kurako, M. Lung Boundary Detection for Chest X-Ray Images Classification Based on GLCM and Probabilistic Neural Networks. Procedia Computer Science, 159, 1439-1448, 2019. doi: 10.1016/j.procs.2019.09.314

[58]. Ankita, R., Kumari, C. U., Mehdi, M. J., Tejashwini, N., & Pavani, T. Lung Cancer Image-Feature Extraction and Classification using GLCM and SVM Classifier. International Journal of Innovative Technology and Exploring Engineering, 8(11), 2211-2215, 2019, doi: 10.35940/ijitee.K2044.0981119

[59]. Haralick, R. M., Shanmugam, K., & Dinstein, I. H. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, (6), 610-621, 1973, doi: 10.1109/TSMC.1973.4309314.

[60]. Zaw, H. T., Maneerat, N., & Win, K. Y. Brain Tumor Detection Based on Naïve Bayes Classification. Proceeedings of the 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST), 1-4, 2019, doi: 10.1109/ICEAST.2019.8802562

[61]. Karthick, G., & Harikumar, R. Comparative Performance Analysis of Naive bayes and SVM classifier for Oral X-ray images. Proceeding 4th International Conference on Electronics and Communication Systems (ICECS), 88-92, 2017, doi: 10.1109/ECS.2017.8067843

[62]. Vapnik, V., & Lerner, A. Pattern recognition using generalized portrait method, Automation and remote control, 24, 774-780, 1963.

[63]. Boser, B. E., Guyon, I. M., & Vapnik, V. N. A Training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory. 144-152, 1992, doi: 10.1145/130385.130401

[64]. Cortes. C., & Vapnik, V. Support-vector networks, Mach Learn 20, 273–297, 1995, doi: 10.1007/BF00994018

[65]. Li. W., Dai. D., Tan, M., Xu, D., & Gool L. V. Fast Algorithms for Linear and Kernel SVM+. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2258-2266, 2016, doi: 10.1109/CVPR.2016.248

[66]. Amancio, D. R., Comin, C. H., Casanova, D., Travieso G., Bruno, O. M., Rodrigues, F, A., & Costa L, F. A systematic comparison of supervised classifiers, PLoS One, 9(4): e94137, 2014, doi: 10.1371/journal.pone.0094137

[67]. Günaydin, Ö., Günay, M., & Şengel, Ö. Comparison of lung cancer detection algorithms. Proceeding of 2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019, 2019, doi: 10.1109/ebbt.2019.8741826

[68]. Larose, D.T. & Larose, C. D. Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc. USA, 2005.

[69]. Radhika, P. R., Nair, R. A. S., & Veena, G. A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms. Proceedings of 2019 3rd IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 1–4, 2019, doi: 10.1109/icecct.2019.8869001

[70]. Wijaya, R. S. D., Adiwijaya., Suksmono, A. B., & Mengko, T. L. Segmentasi Citra Kanker Serviks Menggunakan Markov Random Field dan Algoritma K-Means. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 5(1), 139-147, 2021, doi: 10.29207/resti.v5i1.2816

[71]. Hussain, A., & Khunteta, A. Semantic Segmentation of Brain Tumor from MRI Images and SVM Classification using GLCM Features. Proceeding of the Second International Conference on Inventive Research in Computing Applications (ICIRCA), 38-43, 2020, doi: 10.1109/ICIRCA48905.2020.9183385

[72]. Khairunnisa., & Sianturi, C. F. Perbaikan Kualitas Citra Menggunakan Metode High Pass Filter dan Sharpening. Jurnal Multimedia dan Teknologi Informasi (Jatilima), 2(2), 57-65, 2020.

[73]. Pratama, A. R., Juwita, A. R., & Al Mudzakir, T. Klasifikasi Daging Sapi Berdasarkan Ciri Warna Dengan Metode Otsu dan Euclidean Distance. Techno Xplore: Jurnal Ilmu Komputer dan Teknologi Informasi, 5(1). 26-32, 2020, doi: 10.36805/technoxplore.v5i1.1011

[74]. Kanagaraj, G., & Kumar, P. S. Pulmonary Tumor Detection by Virtue of GLCM. Journal of Scientific an Industrial Research (JSIR), 79(02), 132-134, 2020.

[75]. Sofian, J., & Laluma, R. H. Klasifikasi Hasil Citra Mri Otak untuk Memprediksi Jenis Tumor Otak dengan Metode Image Threshold dan GLCM Menggunakan Algoritma K-NN (Nearest Neighbor) Classifier Berbasis Web. Infotronik: Jurnal Teknologi Informasi dan Elektronika, 4(2), 51-56, 2019, doi: 10.32897/infotronik.2019.4.2.258

[76]. Shaukat, F., Raja, G., Ashraf, R., Khalid, S., Ahmad, M., & Ali, A. Artificial Neural Network Based Classification of Lung Nodules in CT Images Using Intensity, Shape and Texture Features. Journal of Ambient Intelligence and Humanized Computing, 10(10), 4135-4149, 2019, doi: 10.1007/s12652-019-01173-w