Article

Lung Cancer Prediction Using Robust Machine Learning and Image Enhancement Methods on Extracted Gray-Level Co-Occurrence Matrix Features

Lal Hussain [1,2,*], Hadeel Alsolai [3], Siwar Ben Haj Hassine [4], Mohamed K. Nour [5], Mesfer Al Duhayyim [6], Anwer Mustafa Hilal [7,*], Ahmed S. Salama [8], Abdelwahed Motwakel [7], Ishfaq Yaseen [7] and Mohammed Rizwanullah [7]

[1] Department of Computer Science and Information Technology, King Abdullah Campus Chatter Kalas, University of Azad Jammu and Kashmir, Muzaffarabad 13100, Pakistan

[2] Department of Computer Science and Information Technology, Neelum Campus, University of Azad Jammu and Kashmir, Athmuqam 13230, Pakistan

[3] Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; hadeel@pnau.edu.sa

[4] Department of Computer Science, College of Science and Arts at Muhayel, King Khalid University, Abha 62529, Saudi Arabia; siwar@kku.edu.sa

[5] Department of Computer Science, College of Computing and Information System, Umm Al-Qura University, Makkah 21955, Saudi Arabia; nour@uqu.edu.sa

[6] Department of Computer Science, College of Sciences and Humanities-Aflaj, Prince Sattam Bin Abdulaziz University, Alfaj 16828, Saudi Arabia; mesfer@psau.edu.sa

[7] Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia; abdelwahed@psau.edu.sa (A.M.); ishfaq@psau.edu.sa (I.Y.); rizwanullah@psau.edu.sa (M.R.)

[8] Electrical Engineering, Faculty of Engineering & Technology, Future University in Egypt, New Cairo 11845, Egypt; salama@fue.edu.eg

* Correspondence: lall_hussain2008@live.com (L.H.); a.hilal@psau.edu.sa (A.M.H.)

Abstract: In the present era, cancer is the leading cause of demise in both men and women worldwide, with low survival rates due to inefficient diagnostic techniques. Recently, researchers have been devising methods to improve prediction performance. In medical image processing, image enhancement can further improve prediction performance. This study aimed to improve lung cancer image quality by utilizing and employing various image enhancement methods, such as image adjustment, gamma correction, contrast stretching, thresholding, and histogram equalization methods. We extracted the gray-level co-occurrence matrix (GLCM) features on enhancement images, and applied and optimized vigorous machine learning classification algorithms, such as the decision tree (DT), naïve Bayes, support vector machine (SVM) with Gaussian, radial base function (RBF), and polynomial. Without the image enhancement method, the highest performance was obtained using SVM, polynomial, and RBF, with accuracy of (99.89%). The image enhancement methods, such as image adjustment, contrast stretching at threshold (0.02, 0.98), and gamma correction at gamma value of 0.9, improved the prediction performance of our analysis on 945 images provided by the Lung Cancer Alliance MRI dataset, which yielded 100% accuracy and 1.00 of AUC using SVM, RBF, and polynomial kernels. The results revealed that the proposed methodology can be very helpful to improve the lung cancer prediction for further diagnosis and prognosis by expert radiologists to decrease the mortality rate.

Keywords: GLCM features extraction; image enhancement; machine learning; neural network; image adjustment

1. Introduction

Every year, the American Cancer Society estimate the new cancer cases in the United States regarding the new number of cancer cases and deaths on population-based cancer. In the year 2021, about 1,898,160 new cancer cases were encountered, and 608,570 deaths were projected in the United States [1], and 85% will be non-small cell lung cancer (NSCLC) [2,3], among others. NSCLC can be detected with radiofrequency (RF) excision and stereotactic body radiotherapy (SBRT) methods. Lung cancer has two main types, i.e., NSCLC and small cell lung carcinoma (SCLC). Both of these types of cancer spread in different ways and require different treatment accordingly. NSCLC spreads very slowly and is different than SCLC, which is related to smoking, and propagates very rapidly, forms tumors, and spreads in whole body. The deaths based on SCLC are proportional to the number of cigarettes smoked [4].

This study is specifically designed to propose automated tools which improve lung cancer prediction. The tumor type and stage can further improve the early detection of NSCLC. Usually, due to the late detection of cancer cases, there is a 16% survival rate of NSCLC for five years. Additionally, chemotherapy is a standard therapy for SCLC, and the response for SCLC patients is 60%. Thus, the recurrence rate is a few months, resulting in an overall survival of 6% for SCLC. In the past few decades, the survival rates for these two types of cancer have been changed.

The diagnosis of lung cancer is mainly based on findings from traditional chest radiography [5], bronchoscopy, computed tomography (CT) [6], magnetic resonance (MR) [7], positron emission tomography (PET), and biopsy. Due to late diagnosis, lung cancer prognosis remains poor, and advanced metastasis is usually present at the time of presentation. The 5-year persistence ratio for lung cancer confined to the lungs is approximately 54%, but only 4% for inoperable, advanced lung cancer. Treatment modalities depend on the extent and type of the cancer, and include surgery, radiotherapy, and chemotherapy. Computed tomography (CT) is a very sensitive tool for diagnosing, assessing the extent of tumor growth, and monitoring disease progression.

The main objective of this study is to utilize different image enhancement methods to improve lung cancer prediction by improving the image quality, and to further investigate which enhancement method is more robust. To the best of our knowledge, these methods have not been utilized on this dataset. We utilized image enhancement methods, such as image adjustment, gamma correction, thresholding, contrast stretching for improving input image quality and noise removal, etc., and compared results without image enhancement methods. The image enhancement methods improved the prediction performance on the Lung Cancer Alliance dataset. After utilizing image enhancement methods, the texture features based on GLCM were computed. Previously, researchers utilized different image enhancement methods, such as histogram equalization (HE), to improve the image appearance. For low-contrast foreground and background images, HE increases contrast and decreases intensity [8]. The image adjustment method was used by [9] to improve the resolution in a 3-dimensional glasses viewing system. Moreover, contrast adjustment was utilized by [10] to enhance MRI images of visual attenuation. Similarly, different authors [11] utilized enhanced methods to remove gamma correction, haze, external noise, etc. Section 1 is the introduction, Section 2 explains the material and methods utilized in our study, Section 3 describes the results and discussions, and Section 4 details the conclusion along with limitations and future recommendations.

2. Materials and Methods

2.1. Dataset

The public dataset [12] was utilized as provided by the Lung Cancer Alliance (LCA). The LCA is a national nonprofit organization which provides patient advocacy and support. Moreover, this web-based database repository facilitates the researchers. A Digital Imaging and Communications in Medicine (DICOM) format is used for database images. There were 76 patients, with a total of 945 images, including 377 from NSCLC and 568 from SCLC.

The same dataset is utilized and detailed in [13]. We utilized 10-fold cross validation, which minimizes the chances of overfitting [14].

2.2. Data Augmentation

To avoid overfitting, the researchers used different image data augmentation methods for increasing data comprising geometric transformation, and photometric shifting and primitive data augmentation methods. These methods include rotation, flipping, cropping, shearing, and translation in the geometric transformation.

Flipping

Vyas et al. 2018 [1] proposed the flipping method, which reflects an image around its horizontal or vertical axis or both. This method is helpful for maximizing the image number in the dataset without requiring other artificial processing techniques.

Rotation

Sifre and Mallat 2013 [2] proposed another geometric data augmentation method known as rotation, which rotates the images around the axis in the right or left direction by an angle between 1 to 359.

Shearing

Vyas et al. 2018 [1] proposed the shearing method, which changes the original image along the x and y directions. In this case, the existing object shape is changed to an image.

Cropping

Image cropping was utilized by Sifre and Mallat 2013 [2], which is also known as scaling or zooming. It magnifies the image using cropping.

Translation

Using the translation method [1], the object is moved from one position to another position in the image. During this process, the image's black or white part is left after translation, which preserves the image data, or it includes the Gaussian noise or is randomized. The translation can also take place in X, Y, or both directions.

2.3. Image Enhancement

For machine learning, feature extraction is utilized to extract the most relevant information from images, which are then fed as input to different machine learning algorithms. However, image enhancement methods before computing the features can further improve the prediction performance. Good-quality images and the improved visual effects of images can be further helpful for diagnostic systems. In this study, we utilized different image enhancement methods based on various factors to improve the image quality before computing features, and fed them to machine learning algorithms.

Figure 1 shows the workflow in predicting lung cancer. First, we applied image enhancement methods, such as gamma correction at various thresholds, image adjustment, threshold, contrast stretching, and histogram equalization. We then computed the texture and GLCM features, and fed these features into machine learning classification algorithms.

2.3.1. Image Adjustment

With the image adjustment, the grayscale images are converted into new image. The resultant images are adjusted with high and low intensities of the input image. The grayscale image contrast is enhanced in this way, used for further diagnostic procedures [15].

$$k = \left(\frac{l-n}{h-n} \right)^{gamma} * (d-o) + c \tag{1}$$

Here, the lower range is denoted by '*l*', and the higher range is represented by '*h*'; '*n*' denotes the input, and '*o*' is the output, where *c* and *d* denote prevailing lower and higher pixel values, respectively.

The quality of the image is determined by gamma; the brighter or higher image value is determined by gamma, where gamma > 1 reduces the output image and the

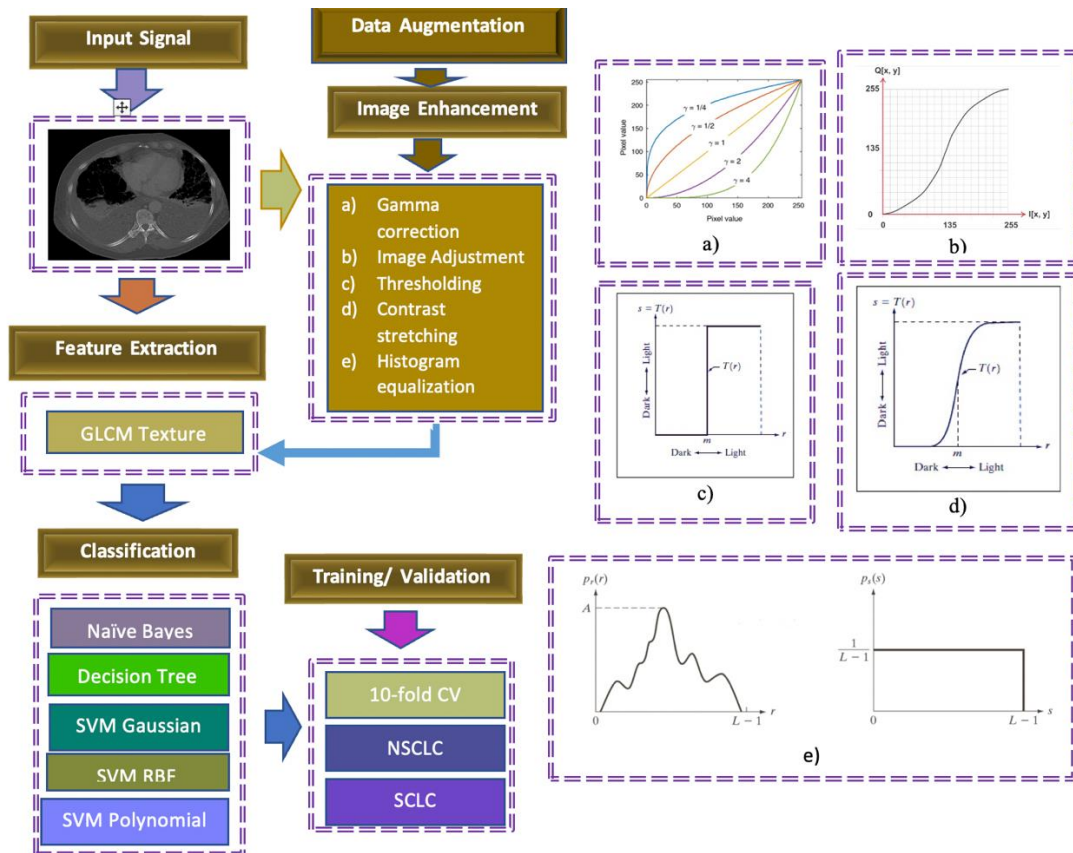image becomes darker [16]. The medical diagnostic system was designed by [17] using contrast enhancement.

Figure 1. Schematic diagram for lung cancer detection based on image enhancement methods on GLCM texture features by utilizing machine learning techniques.

2.3.2. Gamma Correction

This method was utilized as a nonlinear method for brightness decoding and encoding to adopt with human visual insight. The transformation method is mathematically represented as follows:

$$T(l) = l_{max} \left(\frac{l}{l_{max}} \right)^{\gamma} \tag{2}$$

Here, l denotes the intensity level, and l_{max} denotes the maximum intensity level of the image [18]. Usually, most of the image-taking devices do not take exact luminance and familiarize nonlinearity, thus gamma correction is required to improve the luminance. The power function in Equation (1) compensates for the nonlinearity introduced to properly reproduce the actual luminance. The gamma transformation curves for different values of gamma are plotted in Figure 1a. The value of $\gamma < 1$ raises the brightness, where a value of $\gamma > 1$ increases the darkness of the image. The value of $\gamma = 1$ shows that the gamma correction curve reduces towards the identity, and, hence, does not alter the identity. The luminance nonlinear dynamics for lower gray-levels greatly increase, whereas the dynamics for moderate gray-levels slightly increase, and the dynamic ranges for higher gray-levels are compressed. In this study, we applied the gamma values as $\gamma = [0.03, 0.4, 0.5, 0.1, 1, 2, 4]$ according to the scientific requirements [19–21].

2.3.3. Contrast Stretching

In the image enhancement, contrast stretching is also recognized as normalization. By stretching the intensity value ranges, the image quality is enhanced. To enhance the

images, we specified the upper- and lower-pixel value limits as (0.02–0.98) and (0.05–0.95) by considering the highest and lowest pixel values in the images. Consider a as lower limit and b as upper limit, where existing lowest and highest pixel values are denoted by c and d, respectively. Then, each pixel, p, is scaled according to the following equation:

$$P_{out} = (P_{in} - c)\left(\frac{b-a}{d-c}\right) + a \tag{3}$$

The outlier having very low and high value can affect severely the value of c or d. To avoid this, first, a histogram for the images is taken, and then, c and d as 5th and 95th percentiles are selected in the histogram.

2.3.4. Thresholding

In the first step, Otsu's Thresholding Technique [22] is used to remove the noisy part of the image. In 1975, Otsu proposed Otsu's Method to determine the optimum threshold. The Otsu value depends on the discrimination analysis, which maximizes the separability level of classes in gray-level images [23]. Ostu proposed a method between the summation of sigma function and class variance according to the below equations.

$$f(t) = \sigma_0 + \sigma_1 \tag{4}$$

$$\sigma_0 = \omega_0 \left(\mu_0 - \mu_T\right)^2, \; \sigma_1 = \omega_1 \left(\mu_1 - \mu_T\right)^2 \tag{5}$$

The input image mean intensity is denoted by μ_T. For bi-level thresholding, the mean level (μ_i) of two classes is obtained using the below equations [24]:

$$\mu_0 = \sum_{i=0}^{t-1} \frac{ip_i}{\omega_0} \; , \;\; \mu_1 = \sum_{i=1}^{L-1} \frac{ip_i}{\omega_1} \tag{6}$$

The optimal threshold value by maximizing the between-class variance function can be computed using the following:

$$t^* = \text{argmax}(f(t)) \tag{7}$$

2.4. Feature Extraction

In medical image diagnostic systems, the next important step is to compute the most relevant features. Researchers are paying much attention to compute the most relevant features [17]. In the field of medical imaging problems, the medical data are collected without sacrificing the result quality [25–31]. To capture the most relevant properties, we computed GLCM features.

Gray-Level Co-Occurrence Matrix (GLCM)

In the machine learning (ML) techniques, the most important step is to compute most relevant features to capture the maximal hidden information present in the data of interest. The GLCM features were applied from an input image by applying transition on two pixels with gray-level. GLCM features were originally proposed by [32] in 1973, which characterized texture utilizing diverse quantities acquired from second-order image statistics. Two steps are required to compute the GLCM features. Firstly, the pair-wise spatial co-occurrence of image pixels are separated by distance, d, and direction angle, θ. A spatial relationship between two pixels is created for neighboring and reference pixels. In the second step, GLCM features are computed with scaler quantities, which utilize the representation of numerous aspects of an image. This process produces the gray-level co-occurrence matrix that contains several gray-level pixel combinations in an image of interest or the specific portion of an image [32]. The resultant GLCM matrix comprises MxM, where M denotes the gray-level numbers in an image. To compute the GLCM, we utilized the distance, d = {1, 2, 3, 4}, and angle, θ = {0°, 45°, 90° and 135°}, for directions.

Consider the pixel probability, $P(i, j, d, \theta)$, indicates the two pixels' probability separated by a particular distance having gray-levels i and j [33–35]. The GLCM-based texture features were composed of contrast, sum of square variance, cluster shade [36], correlation [37], and two values of homogeneity [37–39]. GLCM features have been successfully utilized in the classification of breast tissues [40], and many other medical imaging problems [36,41–43] detailed in [44–46].

2.5. Classification

After extracting the features, the next important step is to train and choose the classification algorithm. We applied and optimized the SVM with polynomial, radial base function (RBF) and Gaussian kernels, naïve Bayes, and decision tree. Vladimir Vapnik, in 1979, proposed SVM for classification problems. SVM has recently gained much popularity, being widely used in large margin classification problems, including medical diagnosis areas [47,48], machine learning [49], and pattern recognition [50,51]. SVM has also been successfully used in many other applications, such as signature and text recognition, face expression recognition, speech recognition, biometrics, emotion recognition, and several content-based applications, as detailed in [52–54]. Naïve Bayes belong to the family of probabilistic networks based on Bayes' theorem. Based on its performance, naïve Bayes is used in many applications, detailed in [55–58]. Moreover, the decision tree algorithms have also been used in many applications of medical, economic, and scientific applications [59].

2.5.1. Support Vector Machine (SVM)

In classification problems, the SVM is utilized as a supervised ML algorithm. It is used in many applications, including medical systems [47,48], pattern recognition [60], and machine learning [61]. Recently, SVM has been utilized in speech recognition, text recognition, biometrics, and image retrieval problems. To separate the nonlinear problems, SVM using the largest margin constructed a hyperplane. The good margin produced the lowest generalization error.

2.5.2. Decision Tree (DTs)

Breiman, in 1984, proposed the DTs algorithm [62], which serves as a learning algorithm, decision support tool, or predictive model to handle large input data, and predict the class label or target on numerous input variables. The DTs classifiers check and compare similarities in the dataset, and rank them to distinct classes. The DTs algorithm was utilized by [63] to classify data based on a choice of attributes to fix and maximize the data division.

2.5.3. Naïve Bayes (NB)

The NB algorithm was introduced by Wallace and Masteller in 1963, based on the family of the probabilistic classifier. In 1960, this algorithm was used for clustering problems. Due to the large computational errors, the NB methods are biased. This problem can be minimized by reducing the probability valuation errors. However, this may not guarantee the reduction of errors. The poor performance is obtained due to the bias-variance decomposition among probability computation performance and classification errors [64]. Recently, NB has been used in many applications, detailed in [55–58], due to its good performance [65].

2.6. Training/Testing Data Formulation

The jack-knife 10-fold cross-validation method was utilized for training/testing data validation. This is one of the most widely used methods used for validation when there is a small dataset, to avoid the overfitting. During this process, our data are simultaneously utilized for training and testing. The data are initially divided into 10 folds, where, in training, 9 folds participate, and the remaining folds are used for predictions based on training data performance. The entire process is repeated 10 times to predict each class sample. The unseen sample-predicted labels are finally utilized to determine the classifi-

cation accuracy. To avoid overfitting, k-fold cross-validation is used where the dataset is small, as one of the standard approaches used by the researchers [14]. In our case, there were a total of 945 images from both classes. For a larger dataset, the holdout method is usually preferred. For model tuning, the dataset is split into multiple train-test bins. In the standard iterative process, the k-1 folds are involved in the training of the model, and the rest of folds are used for model testing. The general k-fold cross-validation method is reflected in Figure 2 below:
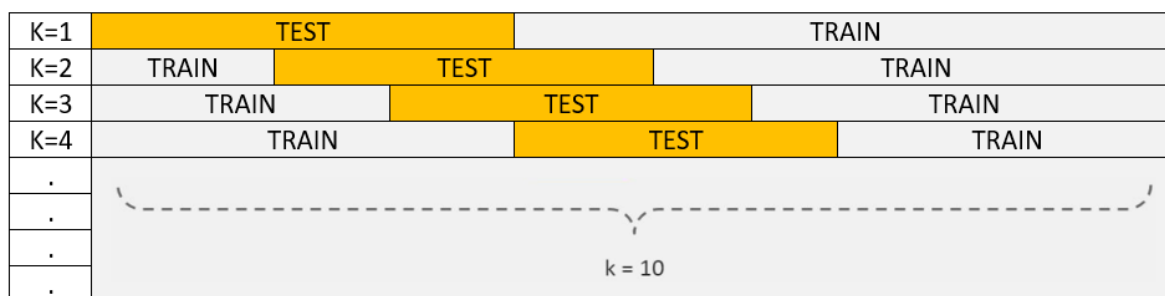

K=1	TEST				TRAIN			
K=2	TRAIN	TEST			TRAIN			
K=3	TRAIN		TEST		TRAIN			
K=4	TRAIN			TEST		TRAIN		
.								
.				k = 10				
.								
.								

Figure 2. K-fold cross-validation.

2.7. Performance Evaluation

We computed the ML algorithms' performance based on specificity, sensitivity, negative predictive value (NPV), positive predictive value (PPV), area under the receiver operating curve (AUC), and accuracy using 10-fold cross-validation, as detailed in [66–68].

3. Results and Discussions

This study was specifically conducted to improve lung cancer prediction performance by optimizing feature extraction and machine learning methods. We first extracted the 22 gray-level co-occurrences (GLCM) features, and applied vigorous ML algorithms to classify the NSCLC from SCLC. We then used an image enhancement method, i.e., image adjustment, before extracting the GLCM feature, and then applied machine learning algorithms. The results reveal that the prediction performance based on image enhancement methods was improved with all the classification algorithms.

Figure 3a reflects the texture features extracted from NSCLC and SCLC MRI images. We utilized the supervised machine learning algorithms. The SVM Gaussian and RBF provided the highest performance, with sensitivity (99.82%), specificity and PPV (100%), and accuracy (99.89%), followed by SVM polynomial, with sensitivity and NPV (100%), specificity (99.72%), and accuracy (99.89%). The decision tree yielded an accuracy of 99.35%, whereas the naïve Bayes yielded an accuracy of 88.57%. Figure 3b reflects the lung cancer performance detection by first applying the image adjustment method, and then computing the texture features from NSCLC and SCLC MRI images. The highest performance was obtained using SVM Gaussian and RBF, with 100% performance metrics; followed by SVM polynomial, with accuracy (99.89%) and AUC (0.9999); decision tree, with accuracy (98.37%) and AUC (0.9649); naïve Bayes, with accuracy (84.87%) and AUC (0.9649). Figure 3c reflects the lung cancer performance detection by first applying the gray-level thresholding method, and then computing the texture features from NSCLC and SCLC MRI images. The highest performance was obtained using SVM polynomial, yielding an accuracy (99.46%) and AUC (0.9971); followed by SVM polynomial, with accuracy (97.93%) and AUC (0.9977); SVM Gaussian, with accuracy (99.71%) and AUC (0.9731); decision tree, with accuracy (96.52%) and AUC (0.9731); and naïve Bayes, with accuracy (0.9652) and AUC (0.9731). Figure 3d reflects the lung cancer performance detection by first applying the histogram equalization method, and then computing the texture features from NSCLC and SCLC MRI images. SVM polynomial provided accuracy (97.82%) and AUC (0.9971); followed by SVM RBF, with accuracy (96.84%) and AUC (0.9912); SVM Gaussian, with accuracy (96.30%) and

AUC (0.9868); decision tree, with accuracy (98.80%) and AUC (0.9821); and naïve Bayes, with accuracy (90.53%) and AUC (0.9821).
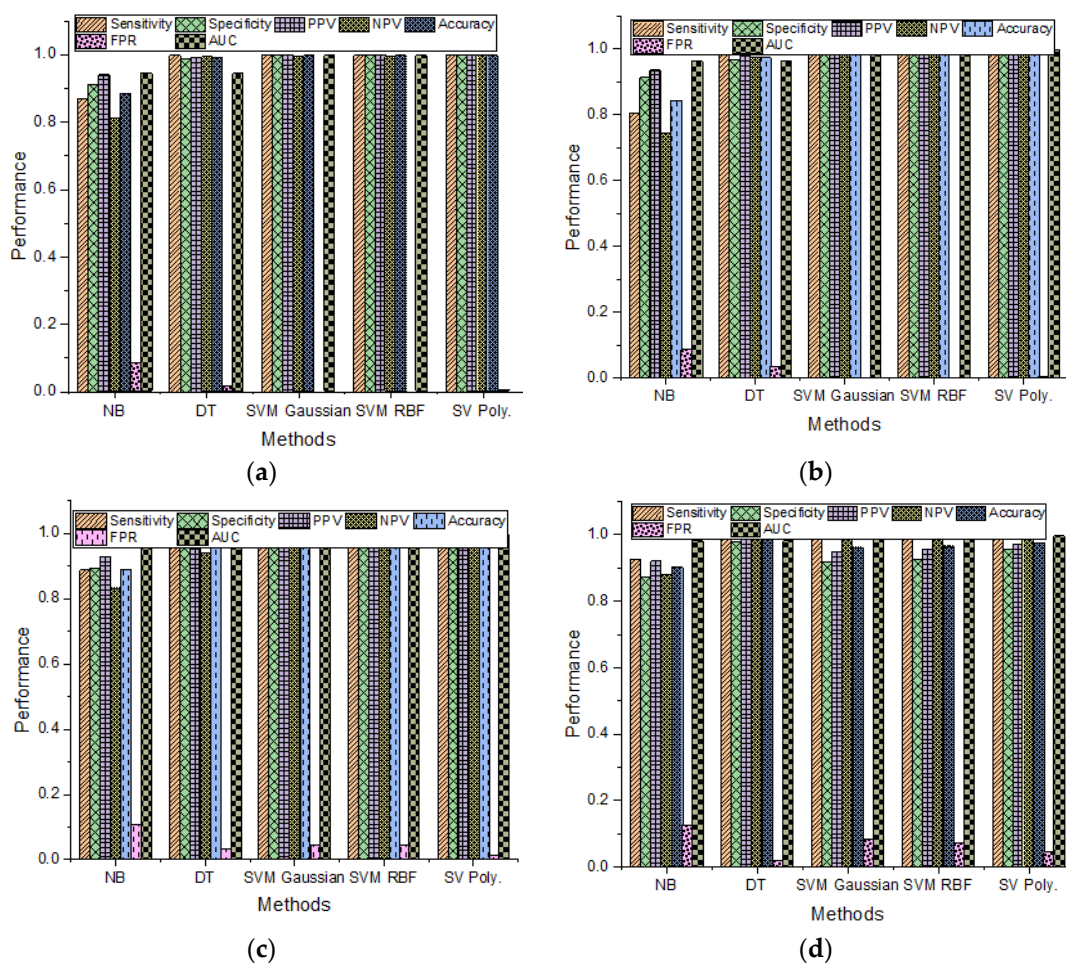

Figure 3. Detecting lung cancer based on different image enhancement methods: (**a**) texture features only, (**b**) image adjustment, (**c**) gray-level thresholding, (**d**) histogram equalization.

Figure 4 reflects the AUC to discriminate the NSCLC from SCLC by computing texture features and applying the image adjustment method and (b) the gamma correction method with a gamma value of 0.04. We then applied the supervised machine learning algorithms. The highest separation using image adjustment (Figure 4a) was yielded using SVM Gaussian and RBF, with AUC (1.00); followed by SVM polynomial, with AUC (0.999); and decision tree and naïve Bayes, with AUC (0.9649). The highest separation by extracting gamma correction with a gamma value of 0.04 (Figure 4b) was obtained using SVM RBF, with AUC (0.9956); followed by SVM polynomial, with AUC (0.9930); SVM Gaussian, with AUC (0.9941); and decision tree and naïve Bayes, with AUC (0.9821).

Figure 5 indicates the lung cancer detection based on the image enhancement method, e.g., contrast stretching at different selected thresholds. Figure 5a utilized a threshold of (0.05, 0.95). The highest detection performance was yielded using SVM Gaussian and SVM RBF, with 100% performance; followed by SVM polynomial, with accuracy (99.67%), AUC (0.9916); decision tree, with accuracy (98.37%), AUC (0.9518); naïve Bayes, with accuracy (81.37%), AUC (0.9518). Figure 5b utilized a threshold of (0.01, 0.90). The SVM Gaussian and SVM polynomial provided the highest accuracy (99.89%), with AUC (0.9999); followed by SVM RBF, with accuracy (99.56%), AUC (0.9999); SVM Gaussian, with accuracy (99.46%), AUC (0.9999); decision tree, with accuracy (99.24%), AUC (0.9613); and naïve Bayes, with accuracy (80.63%), AUC (0.9613). Figure 5c utilized a threshold of (0.02, 0.98). The maximum detection performance was yielded using SVM Gaussian and SVM RBF,

with accuracy (100%), AUC (1.00); followed by SVM polynomial, with accuracy (99.89%), AUC (0.9994); decision tree, with accuracy (98.48%), AUC (0.9465); and naïve Bayes, with accuracy (84.11%), AUC (0.9465).
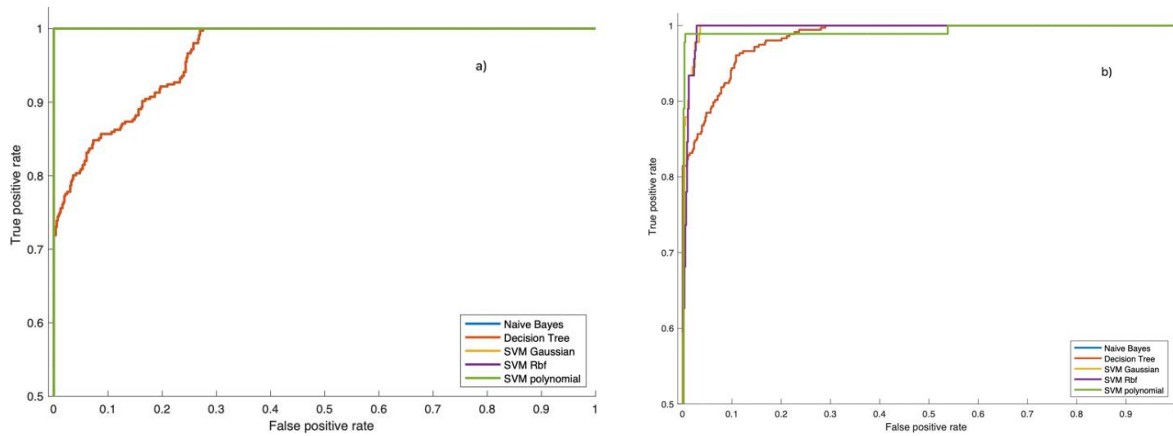

Figure 4. Area under the receiver operating characteristic (AUC) curve: (**a**) texture with image adjustment, (**b**) gamma correction at gamma value of 0.04 to distinguish the NSCLC from SCLC.
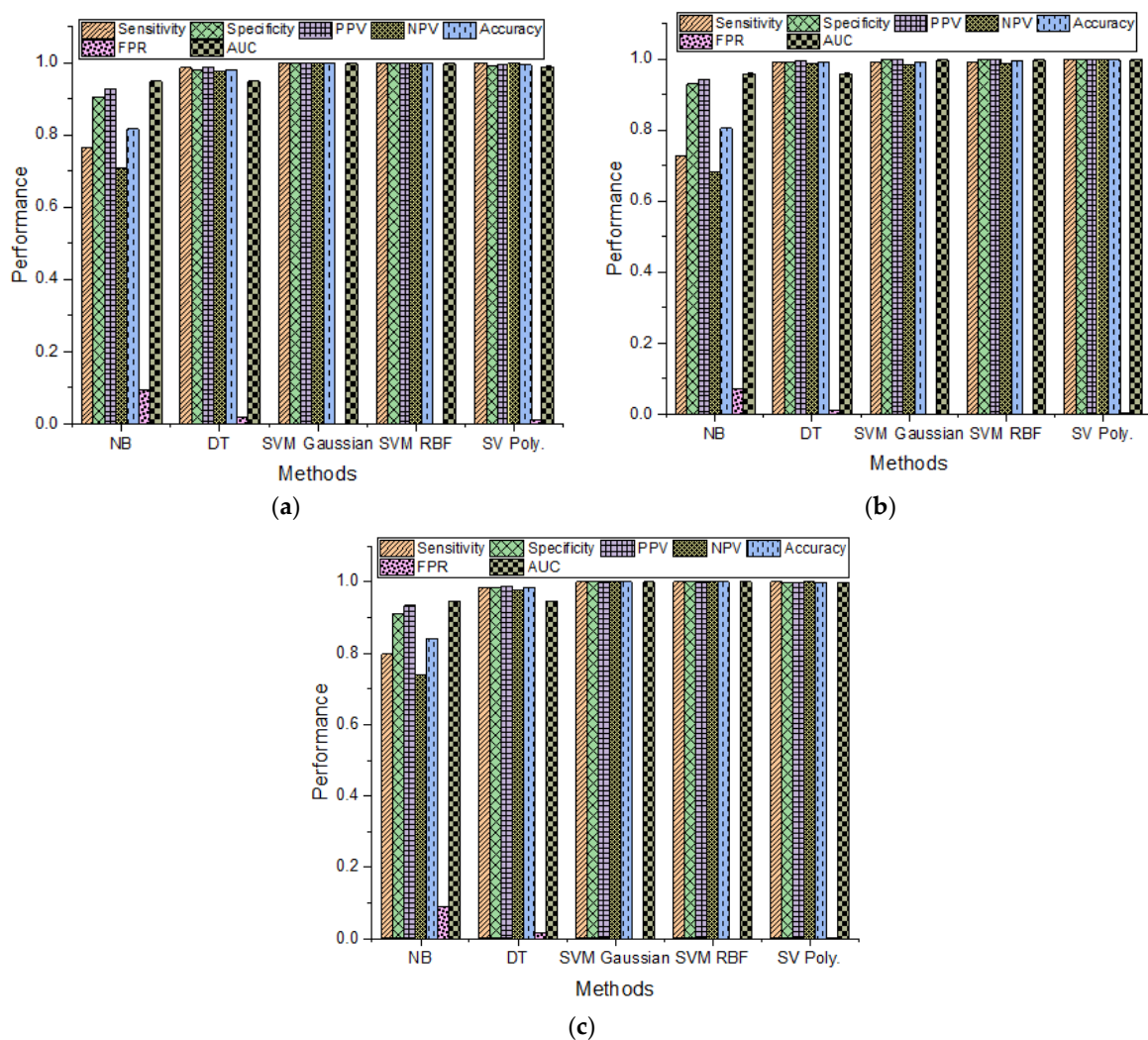

(**a**)

(**b**)

(**c**)

Figure 5. Lung cancer detection based on contrast stretching at different selected threshold ranges: (**a**) (0.05–0.95), (**b**) (0.01–0.90), (**c**) (0.02–0.98).

Figure 6 indicates lung cancer detection based on the gamma correction method at different selected gamma values. Figure 6a utilized a gamma value of 0.04. The DT produced the highest detection performance, with accuracy (98.15%), AUC (0.9821); followed by naïve Bayes, with accuracy (92.49%), AUC (0.9821); SVM polynomial, with accuracy (70.84), AUC (0.993); SVM RBF, with accuracy (69.21%), AUC (0.7472); and SVM Gaussian, with accuracy (0.6910), AUC (0.9941). Figure 6b utilized a gamma value of 0.3. The highest detection performance was yielded using SVM polynomial, with accuracy (98.69%), AUC (1.00); followed by decision tree, with accuracy (98.59%), AUC (0.9856); SVM Gaussian, with accuracy (96.84%), AUC (0.9989); SVM RBF, with accuracy (96.52), AUC (0.9989); SVM RBF, with accuracy (69.21%), AUC (0.7472); SVM Gaussian, with accuracy (0.6910), AUC (0.9941); and naïve Bayes, with accuracy (94.56%), AUC (0.9856). Figure 6c utilized a gamma value of 0.5. The SVM RBF yielded a higher performance, with accuracy (100%), AUC (1.00); followed by SVM Gaussian and polynomial, with accuracy (99.89%), AUC (0.9986); decision tree, with accuracy (99.134%), AUC (0.9469); and naïve Bayes, with accuracy (82.59%), AUC (0.9469). Figure 6d utilized a gamma value of 0.7. The SVM Gaussian produced the highest detection performance, with accuracy (100%), AUC (1.00); followed by SVM RBF and polynomial, with accuracy (99.89%), AUC (0.9987); decision tree, with accuracy (98.91%), AUC (0.9664); and naïve Bayes, with accuracy (85.53%), AUC (0.9664). Figure 6e utilized a gamma value of 0.7. The SVM Gaussian and RBF provided accuracy (100%), AUC (1.00); followed by SVM polynomial, with accuracy (99.89%), AUC (0.9991); decision tree, with accuracy (99.24%), AUC (0.9795); and naïve Bayes, with accuracy (88.14%), AUC (0.9795). Figure 6f utilized a gamma value of 4.0. The highest detection performance was yielded using SVM polynomial, with accuracy (%), AUC (1.00); followed by SVM polynomial, with accuracy (99.89%), AUC (0.9991); decision tree, with accuracy (99.24%), AUC (0.9795); and naïve Bayes, with accuracy (88.14%), AUC (0.9795).

Figure 7 reflects the AUC separation at different folds from 1 to 10 for each of the classification algorithms. The maximum AUC was covered at all the folds for SVM Gaussian, RBF, polynomial, and decision tree, whereas the naïve Bayes yielded a slight lower AUC for folds 1 to 10 using gamma correction at a gamma value of 0.04.

Figure 8 reflects the lung cancer prediction performance using the decision tree model by extracting GLCM-based texture features and applying image enhancement gamma correction methods with gamma = 0.04 and 0.5 using 10-fold cross-validation techniques. The corresponding prediction plots with *mean* $\pm$ *standard deviation* are reflected in Figure 9. Using gamma correction at gamma = 0.04, the NSCLC TP predicted are 318, and 38 were FP. Likewise, for SCLC, 552 were predicted as TN, and 11 were predicted as FN. Using gamma correction at gamma = 0.5, NSCLC yielded 349 as TP, and 7 as FP, whereas SCLC yielded 559 as TN, and 4 as FN. In Figure 8, the blue color represents the NSCLC, and the red color represents the SCLC. The solid line represents correct predictions, and the bold line with crosses represents incorrect predictions. As the confusion matrix indicates, using gamma = 0.04, there were 38 misclassification examples represented as FP, i.e., NSCLC misclassified to SCLC, and 11 examples were misclassified as FN, i.e., SCLC misclassified to NSCLC. In the case of gamma = 0.5, for NSCLC, only seven examples were misclassified to SCLC, and for SCLC, only four examples were misclassified to NSCLC, as reflected in Figure 9a,b. Figure 9c presents the predictions of NSCLC and SCLC plotted within the 1*std*. The misclassification examples are highlighted with a bold black line and cross symbols at different extracted features.
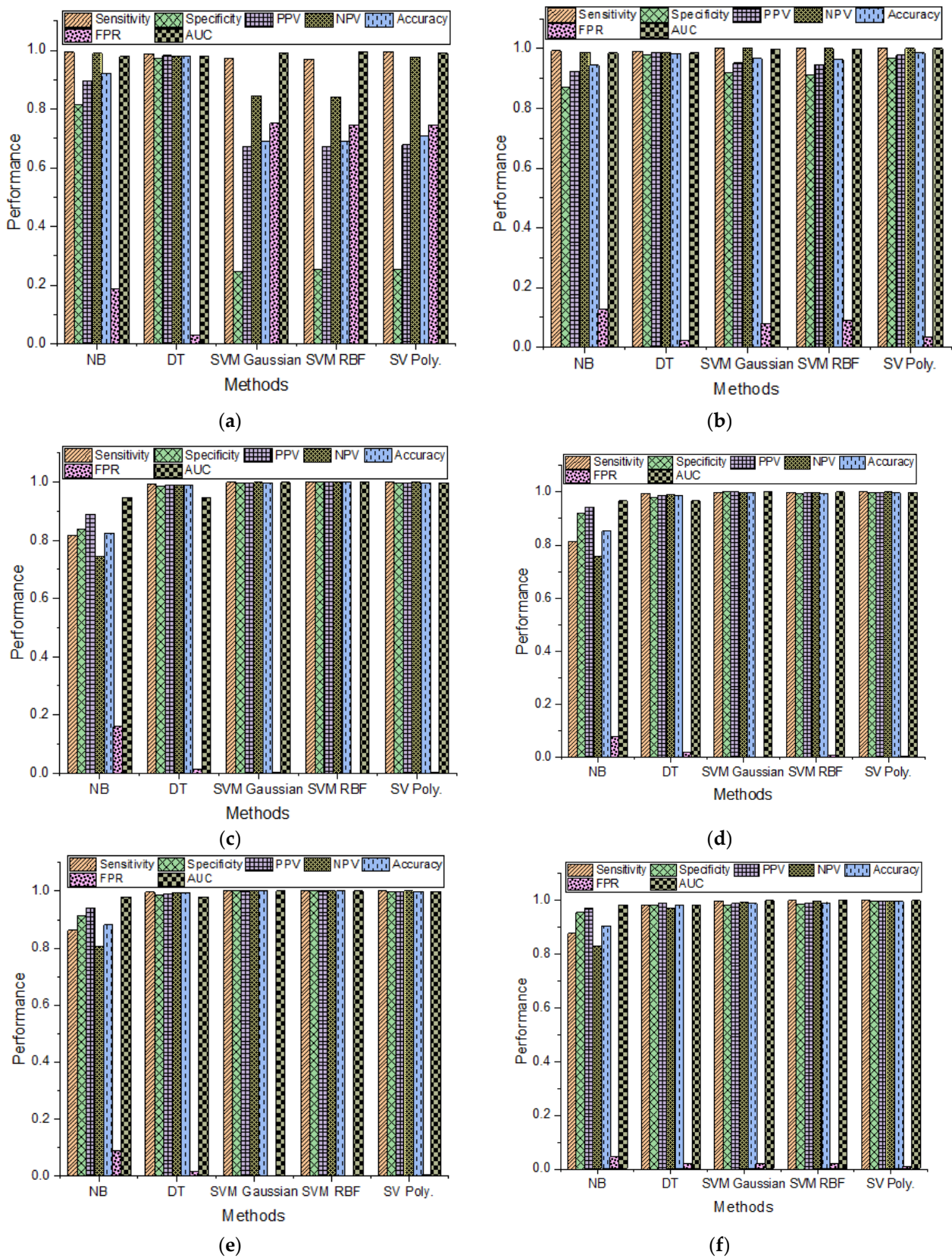
(**a**)

(**b**)

(**c**)

(**d**)

(**e**)

(**f**)

Figure 6. Lung cancer detection based on gamma correction method at different selected gamma values: (**a**) 0.04, (**b**) 0.3, (**c**) 0.5, (**d**) 0.7, (**e**) 0.9, (**f**) 4.0.
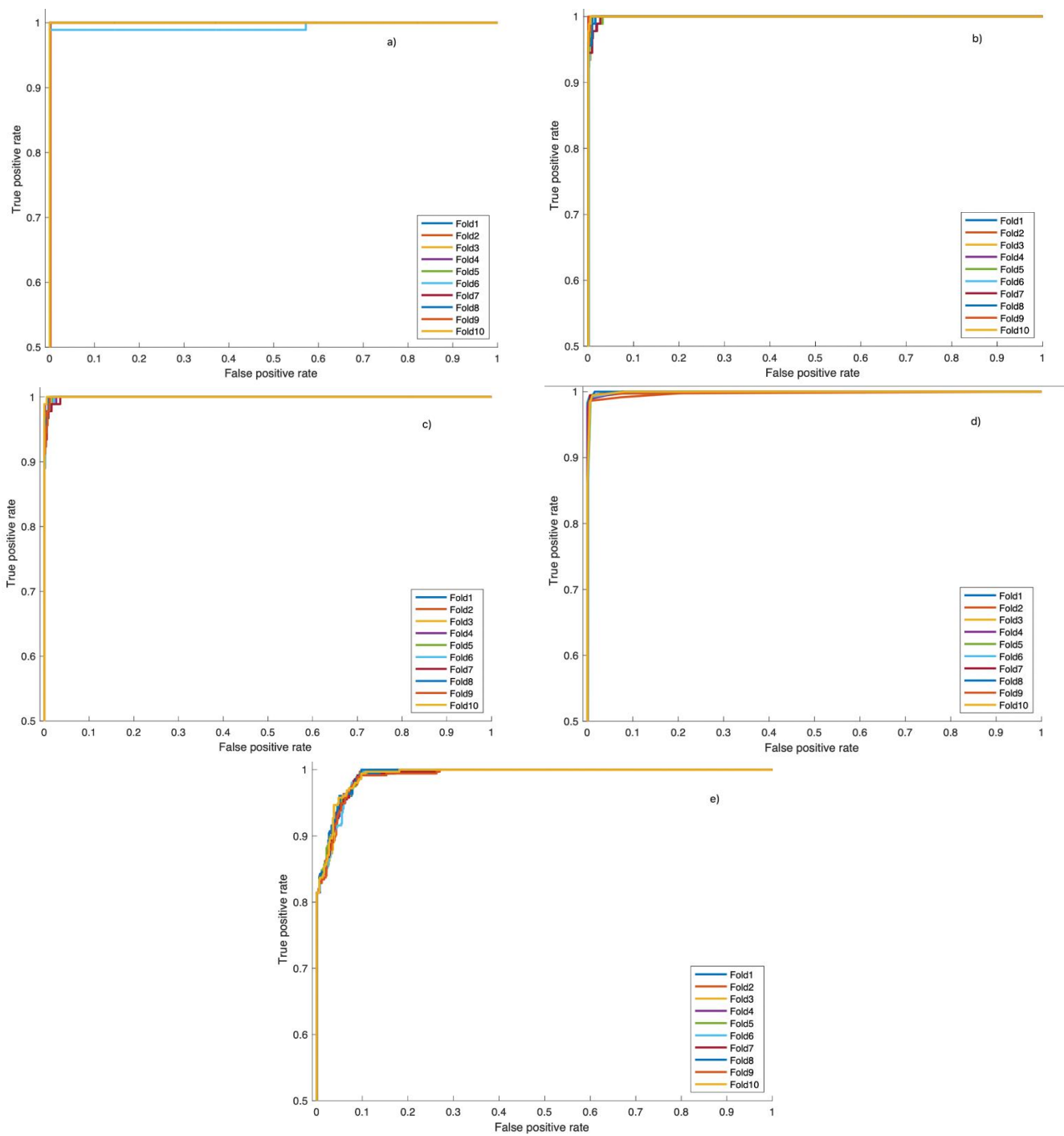
Figure 7. GLCM-based texture features by applying an image enhancement method to lung cancer detection, and applying ML techniques using 1 to 10 cross-fold validation: (**a**) SVM polynomial, (**b**) SVM RBF, (**c**) SVM Gaussian, (**d**) decision tree, (**e**) naïve Bayes.
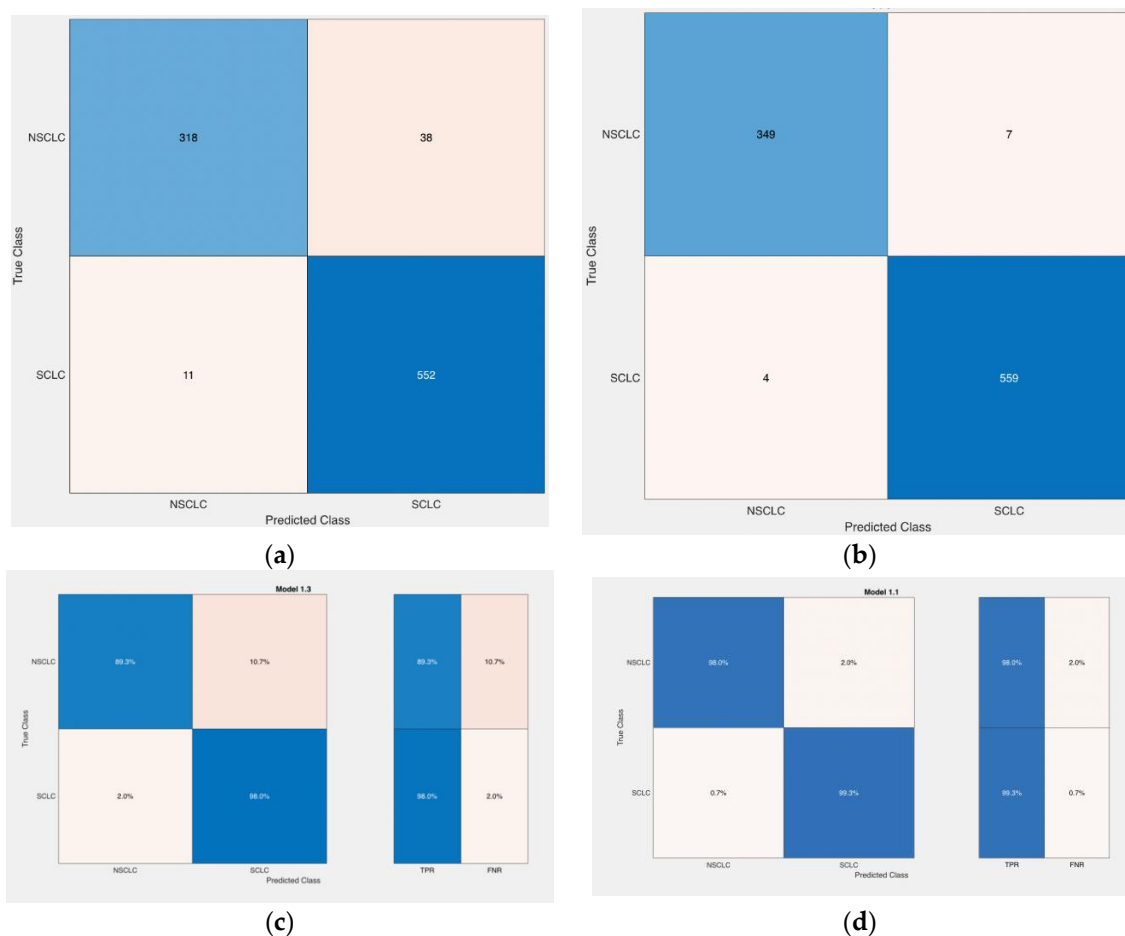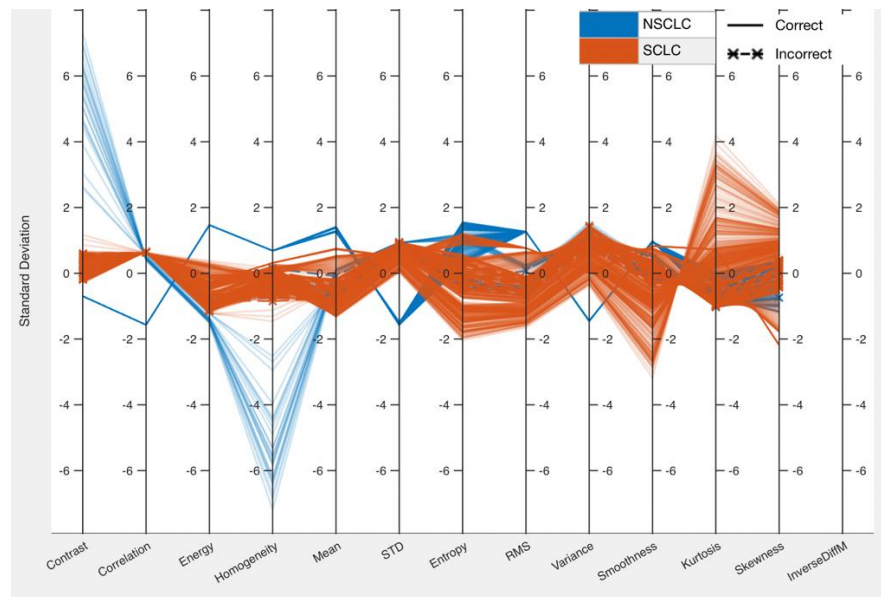
Figure 8. Lung cancer prediction performance based on gamma correction methods using DT course: (**a**) confusion matrix with gamma = 0.04, (**b**) confusion matrix with gamma = 0.5, (**c**) TPR (sensitivity) with gamma = 0.04, (**d**) TPR (sensitivity) with gamma = 0.5.

Table 1 reflects the results based on data augmentation methods to avoid overfitting. We utilized the augmentation methods using 'RandRotation'= $(-20, 20)$, 'RandXReflection' = 1, 'RandYReflection' = 1, 'RandXTranslation' = $(-3, 3)$, and 'RandYTranslation' = $(-3, 3)$ on GLCM features extracted from lung cancer SCLC and NSCLC, and then applied the image enhancement method gamma correction with gamma = 0.5. Using data augmentation, we obtained mostly similar results as those obtained without data augmentation. The highest accuracy was yielded using SVM Gaussian, with 100% accuracy and 1.0 of AUC; followed by SVM RBF, with accuracy (99.89%), AUC (1.0); SVM polynomial, with accuracy (99.89%), AUC (0.9999); decision tree, with accuracy (98.91%), AUC (0.98); and naïve Bayes, with accuracy (98.91%) and AUC (0.98).

Table 2 reflects the lung cancer detection results using different feature extraction and classification methods. In the past, researchers utilized different automated approaches to detect lung cancer. Sousa et al. [69] applied different features of extraction methods, such as gradient, histogram, and spatial methods, and obtained an overall accuracy of 95%. Dandil et al. [6] also extracted different features, such as GLCM, shape-based features, and statistical and energy features, and achieved an accuracy of 95%; however, a sensitivity of 97% was obtained, and a specificity of 96%. Nasrulla et al. [70] computed the statistical features and obtained a sensitivity of 94%, specificity of 90%, and AUC of 0.990. Han et al. [71] used machine learning techniques to distinguish the SCLC types, and achieved an accuracy of 84.10%. Grossman et al. [72] applied EfficientNet to deep learning, and obtained a highest accuracy of 90%. Hussain et al. [13] computed different entropic-based features and computed the nonlinear dynamics to distinguish the SCLC from NSCLC with the highest

significant results (*p*-value < 0.000000). In this study, we first applied image enhancement methods, such as gamma correction at different gamma values, contrast stretching at different thresholds, image adjustment, and histogram processing methods, and then computed the GLCM texture features. We obtained the improved detection results. The researchers in the existing studies did not utilize image enhancement methods and data augmentation techniques. Image enhancement methods on acquired data further improves the image quality, thereby improving the detection results. The proposed methods improved the classification results which can best be utilized by concerned health practitioners to further improve diagnostic capabilities.
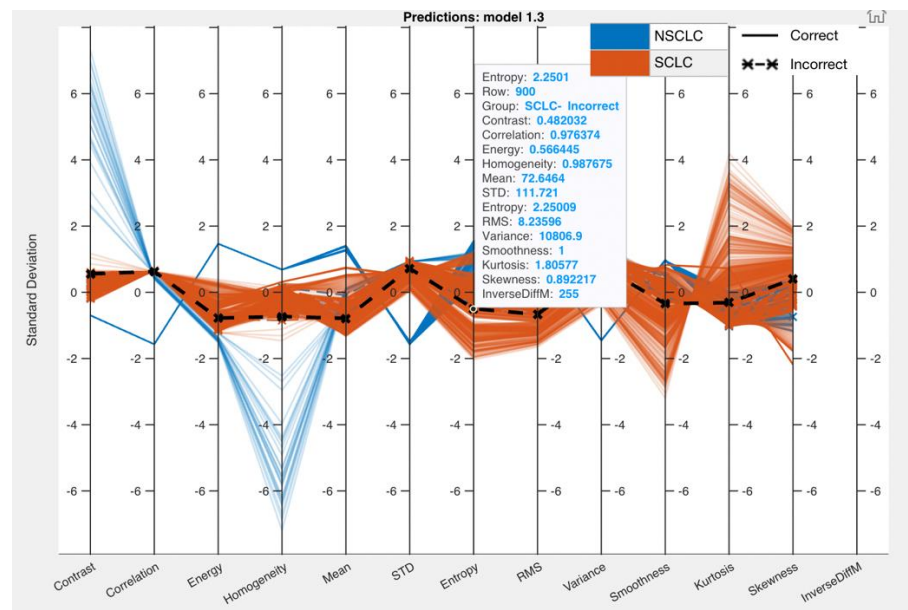

(**a**)
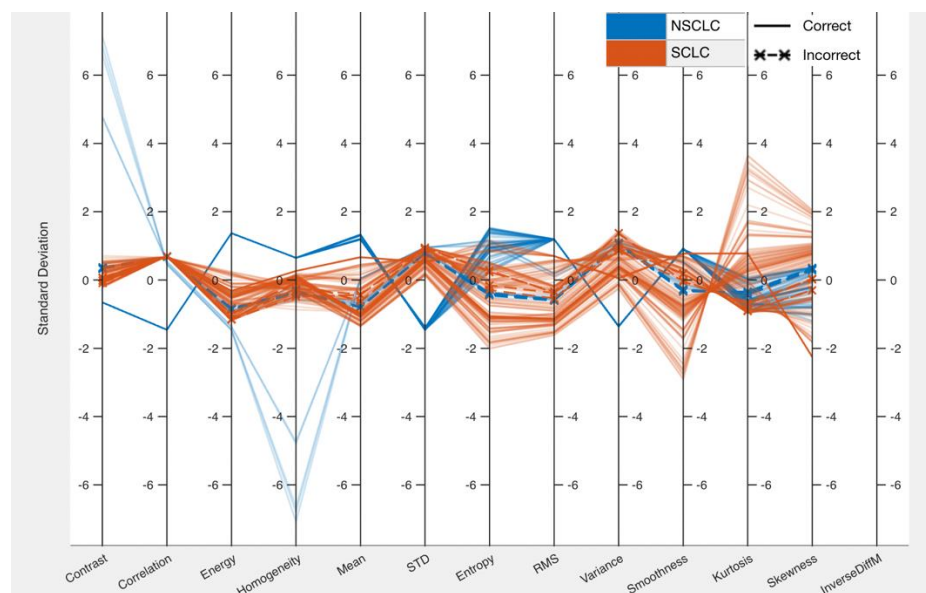
(**b**)

Figure 9. *Cont.*

(**c**)

Figure 9. Parallel prediction plots (**a**) using gamma correction at gamma = 0.04 on predictions, (**b**) gamma correction with gamma = 0.5 on predictions, (**c**) gamma correction with gamma = 0.04 data using 1*std*.

Table 1. Lung cancer detection performance based on data augmentation with gamma correction (gamma = 0.5), and applying the ML methods.

Method	Sensitivity	Specificity	PPV	NPV	Accuracy	FPR	AUC
Naïve Bayes	0.8686	0.9101	0.9386	0.8141	0.8847	0.08989	0.98
Decision Tree	0.9911	0.986	0.9911	0.986	0.9891	0.01404	0.98
SVM Gaussian	1	1	1	1	1	0	1
SVM RBF	0.9982	1	1	0.9972	0.9989	0	1
SVM Poly.	1	0.9972	0.9982	1	0.9989	0.002809	0.9999

Table 2. Comparison of findings with previous studies.

Author	Features Used	Performance
Sousa et al. [69]	1. Gradient 2. Histogram 3. Spatial	Sensitivity = 84%, Specificity = 96% Accuracy = 95%
Dandil et al. [6]	1. GLCM 2. Shape 3. Statistical 4. Energy	Sensitivity = 97%, Specificity = 94% Accuracy = 95%
Nasrulla et al. [70]	1. Statistical	Sensitivity = 94%, Specificity = 90% AUC = 0.990
Han et al. [71]	Machine learning and deep learning methods to distinguish SCLC types	Accuracy =84.10%
Grossman et al. [72]	EfficientNet using transfer learning to distinguish NSCLC from SCLC	Accuracy = 90%
Gao et al. [73]	Machine learning to classify subtypes of NSCLC	AUC = 0.972

Table 2. *Cont.*

Author	Features Used	Performance
Hussain et al. [13]	Texture features using MFE with standard deviation, Morphological features using RCMFE with mean EFDs features using MFE	*p*-value (1.95×10^{-50}) *p*-value (3.01×10^{14}) *p*-value (1.04×10^{-13})
This study	Texture features using SVM polynomial Image Adjustment using SVM RBF and Polynomial Contrast stretching at threshold of (0.02,0.98) using SVM RBF and Polynomial Gamma Correction at gamma value 0.9	Sensitivity = 100%, Specificity = 99.72% Accuracy = 99.89 Sensitivity = 100%, Specificity. = 100% Accuracy = 100% Sensitivity = 100%, Specificity = 100% Accuracy = 100% Sensitivity = 100%, Specificity = 100% Accuracy = 100%

For contrast stretching, we utilized the different threshold ranges, i.e., (0.1, 0.90; 0.02, 0.98; 0.05, 0.95). The interval ranges (0.02, 0.98) and (0.05, 0.95) yielded the highest improved detection performance, which indicates that these ranges are more appropriate to further enhance the image quality for this lung cancer dataset. The threshold range (0.02, 0.98) yielded further improved performance for the selected classifiers. Likewise, the image adjustment enhancement method also yielded higher detection performance for all the classifiers, similar to the contrast stretching range (0.02, 0.98). For gamma correction, we set different values of gamma, such as 0.04, 0.4, 0.5, 0.7, 0.9, and 4.0. The detection performance at lower and higher gamma values reduced the detection performance; however, the mid gamma values yielded a higher detection performance. The gamma value of 0.9 yielded the highest detection performance for all the classifiers, and the decision tree and naïve Bayes algorithms yielded a higher detection performance than other contrast stretching and image adjustment methods. The gray-level thresholding and histogram equalization methods do not enhance the detection performance much.

4. Conclusions

Lung cancer is the most threatening cancer type in the world. It is the most common and leading cause of deaths internationally. The incidences of cancer-related deaths have multiplied unexpectedly, and lung cancer has come to be the most prevalent cancer in the majority of countries. This study was conducted to distinguish between the groups of NSCLC and SCLC by first extracting hand-crafted texture features and employing supervised machine learning algorithms, such as naïve Bayes, and decision tree and SVM with RBF, Gaussian, and polynomial kernels. We then applied different image enhancement methods, such as image adjustment contrast stretching, thresholding, gamma correction, etc., before computing texture features, and fed them into machine learning algorithms. The image enhancement methods further improved the detection performance to detect lung cancer. In order to ovoid overfitting, we also applied data augmentation methods. The results revealed that the proposed methods are very robust in improving the further diagnosis of lung cancer prognosis by expert radiologists.

Limitations and Future Recommendations: The present study was carried out on a small lung cancer dataset provided by the Lung Cancer Alliance on lung cancer types, i.e., NSCLC and SCLC. In the future, we will apply the proposed methods based on image enhancement, feature extraction, and ranking and machine learning methods on larger datasets with more clinical details, disease severity levels, and more types, and larger datasets acquired on different imaging modalities.

Abbreviations

(GLCM)	Gray-level co-occurrence matrix
(SVM)	Support Vector Machine
(RBF)	Radial Base Function
(DT)	Decision Tree
(NSCLC)	Non-Small Cell Lung Cancer
(SCLC)	Small Cell Lung Cancer
(SBRT)	Stereotactic body radiotherapy
(CT)	Computed Tomography
(MR)	Magnetic Resonance
(HE)	Histogram equalization
(LCA)	Lung Cancer Alliance
(DICOM)	Digital Imaging and Communications in Medicine

References

1. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2021. *CA. Cancer J. Clin.* **2021**, *71*, 7–33. [CrossRef] [PubMed]
2. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2018. *CA. Cancer J. Clin.* **2018**, *68*, 7–30. [CrossRef]
3. Oser, M.G.; Niederst, M.J.; Sequist, L.V.; Engelman, J.A. Transformation from non-small-cell lung cancer to small-cell lung cancer: Molecular drivers and cells of origin. *Lancet Oncol.* **2015**, *16*, e165–e172. [CrossRef]
4. Krishnaiah, V.; Narsimha, G.; Chandra, D.N.S. Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques. *Int. J. Comput. Sci. Inf. Technol.* **2013**, *4*, 39–45.
5. Giger, M.L.; Chan, H.-P.P.; Boone, J. Anniversary Paper: History and status of CAD and quantitative image analysis: The role of Medical Physics and AAPM. *Med. Phys.* **2008**, *35*, 5799–5820. [CrossRef] [PubMed]
6. Dandıl, E. A Computer-Aided Pipeline for Automatic Lung Cancer Classification on Computed Tomography Scans. *J. Healthc. Eng.* **2018**, *2018*, 1–12. [CrossRef] [PubMed]
7. Biederer, J.; Ohno, Y.; Hatabu, H.; Schiebler, M.L.; van Beek, E.J.R.; Vogel-Claussen, J.; Kauczor, H.-U. Screening for lung cancer: Does MRI have a role? *Eur. J. Radiol.* **2017**, *86*, 353–360. [CrossRef]
8. Patil, N.K.; Vasudha, S.; Boregowda, L.R. A Novel Method for Illumination Normalization for Performance Improvement of Face Recognition System. In Proceedings of the 2013 International Symposium on Electronic System Design, Singapore, 10–12 December 2013; Volume 3, pp. 148–152.
9. Nishihara, I.; Nakata, T. Dynamic Image Adjustment Method and Evaluation for Glassless 3D Viewing Systems. In Proceedings of the 2015 International Conference on Cyberworlds (CW), Visby, Sweden, 7–9 October 2015; 5, pp. 121–124. [CrossRef]
10. Zhu, R.; Li, X.; Zhang, X.; Xu, X. MRI enhancement based on visual-attention by adaptive contrast adjustment and image fusion. *Multimed. Tools Appl.* **2021**, *80*, 12991–13017. [CrossRef]
11. Ngo, D.; Lee, S.; Nguyen, Q.H.; Ngo, T.M.; Lee, G.D.; Kang, B. Single image haze removal from image enhancement perspective for real-time vision-based systems. *Sensors* **2020**, *20*, 5170. [CrossRef]
12. Lung Cancer Alliance Dataset. Available online: http://www.giveascan.org (accessed on 1 January 2021).

13. Hussain, L.; Aziz, W.; Alshdadi, A.A.A.; Ahmed Nadeem, M.S.; Khan, I.R.; Chaudhry, Q.-U.-A. Analyzing the Dynamics of Lung Cancer Imaging Data Using Refined Fuzzy Entropy Methods by Extracting Different Features. *IEEE Access* **2019**, *7*, 64704–64721. [CrossRef]

14. Chen, W.; Cockrell, C.; Ward, K.R.; Najarian, K. Intracranial pressure level prediction in traumatic brain injury by extracting features from multiple sources and using machine learning methods. In Proceedings of the 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Hong-Kong, China, 18–21 December 2010; pp. 510–515.

15. Soliman, S.R.; Zayed, H.H.; Selim, M.M.; Kasban, H.; Mongy, T. Image quality enhancement in Neutron Computerized Tomography based on projection exposure time adjustment. *Appl. Radiat. Isot.* **2019**, *154*, 108862. [CrossRef] [PubMed]

16. Paul, E.M.; Perumal, B.; Rajasekaran, M.P. Filters Used in X-Ray Chest Images for Initial Stage Tuberculosis Detection. In Proceedings of the 2018 International Conference on Inventive Research in Computing Applications (ICIRCA 2018), Coimbatore, India, 11–12 July 2018; pp. 235–239.

17. Hongjuan, Y.; Decai, M.; Yunchu, Z.; Jianrong, C. Preprocessing of automobile engine connecting rod based on shadow removal and image enhancement. In Proceedings of the 2021 International Conference on Communications, Information System and Computer Engineering (CISCE), Shenzhen, China, 14–16 May 2021; pp. 428–432.

18. Tiwari, M.; Gupta, B. Brightness preserving contrast enhancement of medical images using adaptive gamma correction and homomorphic filtering. In Proceedings of the 2016 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 5–6 March 2016; pp. 1–4.

19. Farid, H. Blind inverse gamma correction. *IEEE Trans. Image Process.* **2001**, *10*, 1428–1433. [CrossRef] [PubMed]

20. Bhandari, A.K.; Kumar, A.; Singh, G.K.; Soni, V. Dark satellite image enhancement using knee transfer function and gamma correction based on DWT–SVD. *Multidimens. Syst. Signal Process.* **2016**, *27*, 453–476. [CrossRef]

21. Ngo, D.; Kang, B. Taylor-Series-Based Reconfigurability of Gamma Correction in Hardware Designs. *Electronics* **2021**, *10*, 1959. [CrossRef]

22. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *PAMI-8*, 679–698. [CrossRef]

23. Kurban, T.; Civicioglu, P.; Kurban, R.; Besdok, E. Comparison of evolutionary and swarm based computational techniques for multilevel color image thresholding. *Appl. Soft Comput.* **2014**, *23*, 128–143. [CrossRef]

24. Bhandari, A.K.; Kumar, A.; Singh, G.K. Modified artificial bee colony based computationally efficient multilevel thresholding for satellite image segmentation using Kapur's, Otsu and Tsallis functions. *Expert Syst. Appl.* **2015**, *42*, 1573–1601. [CrossRef]

25. Hussain, L.; Ali, A.; Rathore, S.; Saeed, S.; Idris, A.; Usman, M.U.; Iftikhar, M.A.; Suh, D.Y. Applying Bayesian Network Approach to Determine the Association Between Morphological Features Extracted from Prostate Cancer Images. *IEEE Access* **2018**, *7*, 1586–1601. [CrossRef]

26. Hussain, L.; Saeed, S.; Awan, I.A.; Idris, A.; Nadeem, M.S.A.; Chaudhary, Q.-A. Detecting Brain Tumor Using Machine Learning Techniques Based on Different Features Extracting Strategies. *Curr. Med. Imaging Rev.* **2019**, *15*, 595–606. [CrossRef]

27. Hussain, L.; Aziz, W.; Saeed, S.; Shah, S.A.; Nadeem, M.S.A.; Awan, A.; Abbas, A.; Majid, A.; Zaki, S.; Kazmi, H. Complexity analysis of EEG motor movement with eye open and close subjects using multiscale permutation entropy (MPE) technique. *Biomed. Res.* **2017**, *28*, 7104–7111.

28. Hussain, L.; Ahmed, A.; Saeed, S.; Rathore, S.; Awan, I.A.; Shah, S.A.; Majid, A.; Idris, A.; Awan, A.A. Prostate cancer detection using machine learning techniques by employing combination of features extracting strategies. *Cancer Biomark.* **2018**, *21*, 393–413. [CrossRef] [PubMed]

29. Hussain, L.; Aziz, W.; Saeed, S.; Rathore, S.; Rafique, M. Automated Breast Cancer Detection Using Machine Learning Techniques by Extracting Different Feature Extracting Strategies. In Proceedings of the 2018 17th IEEE International Conference on Trust, Security and Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, USA, 1–3 August 2018; pp. 327–331.

30. Hussain, L. Detecting epileptic seizure with different feature extracting strategies using robust machine learning classification techniques by applying advance parameter optimization approach. *Cogn. Neurodyn.* **2018**, *12*, 271–294. [CrossRef]

31. Hussain, L.; Rathore, S.; Abbasi, A.A.; Saeed, S. Automated lung cancer detection based on multimodal features extracting strategy using machine learning techniques. In *Proceedings of the Medical Imaging 2019: Physics of Medical Imaging*; Bosmans, H., Chen, G.-H., Gilat Schmidt, T., Eds.; SPIE: Bellingham, WA, USA, 2019; Volume 10948, p. 134.

32. Harlick, R.M.; Shanmugam, K. ITS'Hak Dinstein. Textural feature for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *6*, 610–621. [CrossRef]

33. Khuzi, A.M.; Besar, R.; Zaki, W.M.D.W. Texture features selection for masses detection in digital mammogram. *IFMBE Proc.* **2008**, *21*, 629–632. [CrossRef]

34. Nguyen, V.D.; Nguyen, D.T.; Nguyen, T.D.; Pham, V.T. An Automated Method to Segment and Classify Masses in Mammograms. *Eng. Technol.* **2009**, *3*, 942–947.

35. Nithya, R.; Santhi, B. Classification of Normal and Abnormal Patterns in Digital Mammograms for Diagnosis of Breast Cancer. *Int. J. Comput. Appl.* **2011**, *28*, 975–8887. [CrossRef]

36. Nithya, R. Comparative study on feature extraction. *J. Theor. Appl. Infromat. Technol.* **2011**, *33*, 7.

37. Manjunath, S.; Guru, D.S. Texture Features and KNN in Classification of Flower Images. *IJCA* **2010**, *1*, 21–29.

38. Soh, L.; Tsatsoulis, C.; Member, S. Texture Analysis of SAR Sea Ice Imagery. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 780–795. [CrossRef]

39. Berbar, M.A. Hybrid methods for feature extraction for breast masses classification. *Egypt. Informatics J.* **2017**, 1–11. [CrossRef]
40. Beura, S.; Majhi, B.; Dash, R. Neurocomputing Mammogram classi fi cation using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing* **2015**, *154*, 1–14. [CrossRef]
41. Parvez, A. Feature Computation using CUDA Platform. In Proceedings of the 2017 International Conference on Trends in Electronics and Informatics (ICEI), Tirunelveli, India, 11–12 May 2017; pp. 296–300.
42. Rathore, S.; Hussain, M.; Khan, A. Automated colon cancer detection using hybrid of novel geometric features and some traditional features. *Comput. Biol. Med.* **2015**, *65*, 279–296. [CrossRef] [PubMed]
43. Amrit, G.; Singh, P. Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. *Neural Comput. Appl.* **2018**, *3456789*, 6863–6877.
44. Qureshi, S.A.; Raza, S.E.A.; Hussain, L.; Malibari, A.A.; Nour, M.K.; ul Rehman, A.; Al-Wesabi, F.N.; Hilal, A.M. Intelligent Ultra-Light Deep Learning Model for Multi-Class Brain Tumor Detection. *Appl. Sci.* **2022**, *12*, 3715. [CrossRef]
45. Hammad, B.T.; Ahmed, I.T.; Jamil, N. A Steganalysis Classification Algorithm Based on Distinctive Texture Features. *Symmetry* **2022**, *14*, 236. [CrossRef]
46. Patel, D.; Thakker, H.; Kiran, M.B.; Vakharia, V. Surface roughness prediction of machined components using gray level co-occurrence matrix and Bagging Tree. *FME Trans.* **2020**, *48*, 468–475. [CrossRef]
47. Dobrowolski, A.P.; Wierzbowski, M.; Tomczykiewicz, K. Multiresolution MUAPs decomposition and SVM-based analysis in the classification of neuromuscular disorders. *Comput. Methods Programs Biomed.* **2012**, *107*, 393–403. [CrossRef]
48. Subasi, A. Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. *Comput. Biol. Med.* **2013**, *43*, 576–586. [CrossRef]
49. Papadopoulos, H.; Vovk, V.; Gammerman, A. Guest editors' preface to the special issue on conformal prediction and its applications. *Ann. Math. Artif. Intell.* **2015**, *74*, 1–7. [CrossRef]
50. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef] [PubMed]
51. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [CrossRef] [PubMed]
52. Rajnoha, M.; Burget, R.; Dutta, M.K. Offline handwritten text recognition using support vector machines. In Proceedings of the 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2–3 February 2017; pp. 132–136.
53. Phadikar, S.; Sinha, N.; Ghosh, R.; Ghaderpour, E. Automatic Muscle Artifacts Identification and Removal from Single-Channel EEG Using Wavelet Transform with Meta-Heuristically Optimized Non-Local Means Filter. *Sensors* **2022**, *22*, 2948. [CrossRef] [PubMed]
54. Ahmed, M.Z.I.; Sinha, N.; Phadikar, S.; Ghaderpour, E. Automated Feature Extraction on AsMap for Emotion Classification Using EEG. *Sensors* **2022**, *22*, 2346. [CrossRef]
55. Zaidi, N.A.; Du, Y.; Webb, G.I. On the Effectiveness of Discretizing Quantitative Attributes in Linear Classifiers. *IEEE Access* **2020**, *8*, 198856–198871. [CrossRef]
56. Zhang, J.; Chen, C.; Xiang, Y.; Zhou, W.; Xiang, Y. Internet Traf fi c Classi fi cation by Aggregating Correlated Naive Bayes Predictions. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 5–15. [CrossRef]
57. Chen, C.; Zhang, G.; Yang, J.; Milton, J.C.; Alcántara, A.D. An explanatory analysis of driver injury severity in rear-end crashes using a decision table/Naïve Bayes (DTNB) hybrid classifier. *Accid. Anal. Prev.* **2016**, *90*, 95–107. [CrossRef]
58. Bermejo, P.; Gámez, J.A.; Puerta, J.M. Speeding up incremental wrapper feature subset selection with Naive Bayes classifier. *Knowl. Based Syst.* **2014**, *55*, 140–147. [CrossRef]
59. Rissanen, J.J. Fisher information and stochastic complexity. *IEEE Trans. Inf. Theory* **1996**, *42*, 40–47. [CrossRef]
60. Bousquet, O.; Boucheron, S.; Lugosi, G. Introduction to statistical learning theory. In *Summer School on Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 169–207.
61. Gammerman, A.; Luo, Z.; Vega, J.; Vovk, V. (Eds.) *Conformal and Probabilistic Prediction with Applications*; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2016; Volume 9653, ISBN 978-3-319-33394-6.
62. Ariza-Lopez, F.J.; Rodriguez-Avi, J.; Alba-Fernandez, M.V. Complete Control of an Observed Confusion Matrix. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018), Valencia, Spain, 22–27 July 2018; Volume 2018, pp. 1222–1225.
63. Wang, L.-M.; Li, X.-L.; Cao, C.-H.; Yuan, S.-M. Combining decision tree and Naive Bayes for classification. *Knowl. Based Syst.* **2006**, *19*, 511–515. [CrossRef]
64. Fang, X. Inference-based naive bayes: Turning naive bayes cost-sensitive. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 2302–2314. [CrossRef]
65. Yuan, G.-X.; Ho, C.-H.; Lin, C. Recent Advances of Large-Scale Linear Classification. *Proc. IEEE* **2012**, *100*, 2584–2603. [CrossRef]
66. Rathore, S.; Hussain, M.; Ali, A.; Khan, A. A recent survey on colon cancer detection techniques. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2013**, *10*, 545–563. [CrossRef]
67. Fergus, P.; Hussain, A.; Hignett, D.; Al-Jumeily, D.; Abdel-Aziz, K.; Hamdan, H. A machine learning system for automated whole-brain seizure detection. *Appl. Comput. Informat.* **2016**, *12*, 70–89. [CrossRef]
68. Asim, Y.; Raza, B.; Malik, A.K.; Rathore, S.; Hussain, L.; Iftikhar, M.A. A multi-modal, multi-atlas-based approach for Alzheimer detection via machine learning. *Int. J. Imaging Syst. Technol.* **2018**, *28*, 113–123. [CrossRef]

69. da Silva Sousa, J.R.F.; Silva, A.C.; de Paiva, A.C.; Nunes, R.A. Methodology for automatic detection of lung nodules in computerized tomography images. *Comput. Methods Programs Biomed.* **2010**, *98*, 1–14. [CrossRef] [PubMed]

70. Nasrullah, N.; Sang, J.; Alam, M.S.; Xiang, H. Automated detection and classification for early stage lung cancer on CT images using deep learning. In Proceedings of the Pattern Recognition and Tracking XXX, Baltimore, MD, USA, 15–16 April 2019; p. 27.

71. Han, Y.; Ma, Y.; Wu, Z.; Zhang, F.; Zheng, D.; Liu, X.; Tao, L.; Liang, Z.; Yang, Z.; Li, X.; et al. Histologic subtype classification of non-small cell lung cancer using PET/CT images. *Eur. J. Nucl. Med. Mol. Imaging* **2021**, *48*, 350–360. [CrossRef] [PubMed]

72. Grossman, R.; Haim, O.; Abramov, S.; Shofty, B.; Artzi, M. Differentiating Small-Cell Lung Cancer from Non-Small-Cell Lung Cancer Brain Metastases Based on MRI Using Efficientnet and Transfer Learning Approach. *Technol. Cancer Res. Treat.* **2021**, *20*, 153303382110049. [CrossRef]

73. Gao, Y.; Song, F.; Zhang, P.; Liu, J.; Cui, J.; Ma, Y.; Zhang, G.; Luo, J. Improving the Subtype Classification of Non-small Cell Lung Cancer by Elastic Deformation Based Machine Learning. *J. Digit. Imaging* **2021**, *34*, 605–617. [CrossRef]