JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

AUTOMATED CLASSIFICATION OF LUNG CANCER IMAGES USING KERNEL PRINCIPLE COMPONENT ANALYSIS BASED SUPPORT VECTOR MACHINE CLASSIFIER

M. Shenbagapriya[1], Dr. M. Vanitha[2]

[1]Research Scholar, Department of Electronics & Communication Engineering, Sri Satya Sai University of Technology and Medical Sciences, India
[2]Research Guide, Department of Electronics & Communication Engineering, Sri Satya Sai University of Technology and Medical Sciences, India

Email: [1]shenbagapriya4@gmail.com, [2]vanijsk79@gmail.com

Received: 14 March 2020 Revised and Accepted: 8 July 2020

ABSTRACT: In this paper, we propose an Automated Classification Model (ACM) to classify large sets of lung cancer images. In this model, the images are pre-processed for noise removal and the features are extracted using Kernel Principle Component Analysis (KPCA). The images are classified finally using Support Vector Machine (SVM) classifier. The ACM is a classical model for lung cancer images that has improved the feature extraction and the classification. It is trained to classify the exact lesion into the large collection of images of lung cancer and is then tested in real-time images. In terms of accurately, the experimental results are compared with existing methods. The experimental findings show that ACM is better classified in precision, specificity and sensitivity than conventional classifiers.

KEYWORDS: Kernel Principle component Analysis, Support Vector Machine, Lung Cancer Images, Automated Classification model

I. INTRODUCTION

The most frequently diagnosed cancer is lung cancer that leads to over one-quarter the world's deaths from carcinogenic disease. Roughly 85% of all primary lung malignancies are non-small cells of the lung, while approximately 15% are small cells of the lung. There is therefore the first and key step in individual therapy and systematic treatment decisions is a correct classification of lung cancer. Both methods are invasive diagnostic methods that are not only limited by the risk of surgery and biopsy, they also require time. In summary, developing non-invasive ways to classify lung cancer subtypes can help doctors make better treatment decisions. In addition, they can provide other approaches in patients with not enough tissue to investigate histopathology to find the result [1] [12] [13].
In the leading death diseases, this disease is important because it is difficult to recognize lung cancer as compared to other diseases. This is mainly due to the small size of the lesion, which is called a nodule. Tumor cell sizes in the benign stage are so small but over time they increase tumor sizes and the tumor is malignant. Radiologist easily acknowledged this illness at that time, but patients were too late. The initial level of the disease needs to be monitored. If it is diagnosed early, the survival rate can be increased. The latest computer vision researchers have introduced computer systems [3] – [8] that automatically detect and classify healthy and tumor regions [2].
The present study is designed to classify the large data sets related to pulmonary cancer. The Automated Classification Model (ACM) This framework is trained in classifying the exact injury across the large database of images of lung cancer and is then tested on images in real time.
The outline of the paper is given below: Section 2 gives the detailed discussion of the proposed classification method. Section 3 provides the performance measures used to evaluate the proposed method. Section 4 evaluates the proposed method and provide discussions over it. Finally, section 5 concludes the entire work with future direction.

JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

II. METHODS

The architecture of proposed lung cancer images diagnostic method is given in figure 1. The images are pre-processed for noise removal and then the features are extracted using KPCA. The images are finally classified using SVM classifier.

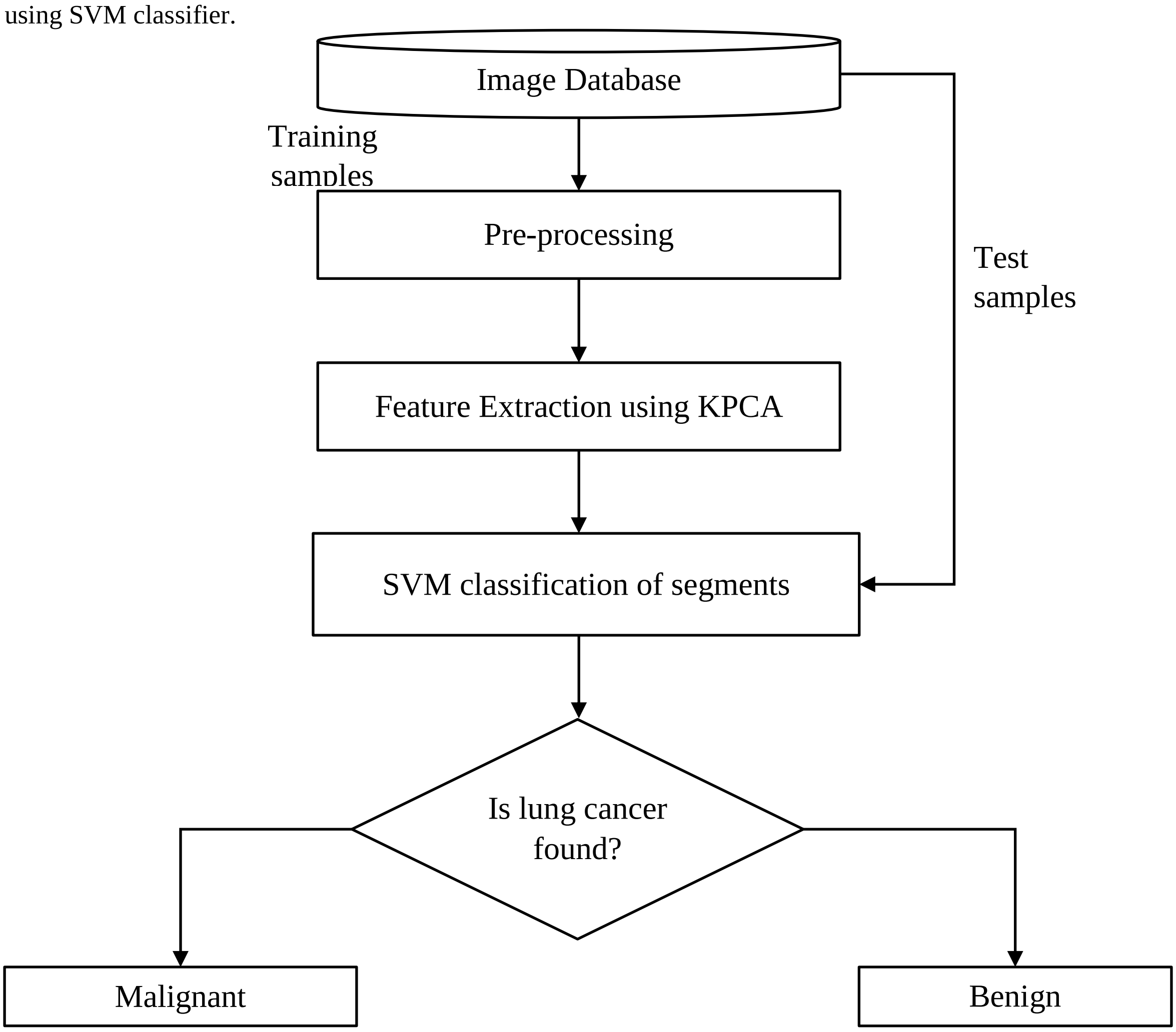


Figure 1. Automated Classification Model

Pre-Processing

The preprocessing phase is here used to improve the final result of the classification process. In the first stage, a medium sift filtering model was offered to remove artifacts from the image. Input image. The thresholds will be converted to a binary format for the image extracted. A contour-drawn step is then made to retrieve the contours of the images against the objects. In addition, the highest contour mask has been used to maintain a large object in the production of a mask. The noise in the mask must then be removed and the contrast improvement performed using the CLAHE [9] model. Finally, the contrast in the preprocessed image was improved.

Feature Extraction

The PCA is a typical strategy linked to reduction of dimensionality and extraction emphasis. The PCA technique can only extract the linear dataset structure data, however, this non-linear structure information cannot be extracted. KPCA [10] is a PCA enhanced that removes the main components through a non-linear strategy of the kernel. A major understanding behind KPCA is that input data (information) should be converted into a high dimension element F, where PCA is completed and the component Vector understood in F is not to be registered expressly when used even though the internal result of $2F$ vectors and the kernel function is simply processed.

Assume $x_1, x_2. . . x_n \in R_d$, which are considered to be preparing number of tests n for training the KPCA. The i^{th} pixel using KPCA for the element t_i is obtained by the following expression.

$$t_i = \frac{1}{\sqrt{\lambda_i}} \gamma_i^T \left[k(x_1, x_{new}), k(x_2, x_{new}), ..., k(x_n, x_{new}) \right]^T \quad (1)$$

The component vector of KPCA is thus acquired for feature extraction.

JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

SVM classifier

The dual problem of looking for the Lagrange function threshold point, which can be reduced to the quadratic programming, is resolves in order to find a hyperplane divider in the SVM [11] algorithm. In general, even with default settings, the SVM classifications offer a high quality data classification. To increase the accuracy of the SVM classification, the location of the wrongly classified objects are analysed. False objects are often located near the separate hyperplane. Therefore, the additional tools must be employed to improve the grading quality of objects in lesion separation.

The SVMs gradually become known as an intense device for classification of information and the estimation of function. The SVM unfolds a direct classification of a quadratic programming problem and the SVM can be displayed with every preparation focus. The central idea of SVM is presented quickly as follows:

Assume the test set $\{x_i, y_i\}$, i=1,2,...,N where x_i is regarded as the i^{th} input vector of a d-dimensional image and y_i is regarded as the class label of x_i obtained from KPCA and that is either +1 or −1.

$$Y(x) = sign[w^T \phi + b] \tag{2}$$

where

w is considered as a weight vector,

b is considered as an inclination term and

$\varphi(x)$ is considered as a nonlinear function that maps pixel in higher resolution images.

The inclination term, b and the weight vector, w needs to be resolved that accounts for acquiring both the b and w. The enhancement to obtain the b and w is given below, which is subjected to the uniform distribution.

$$\min\ j = (w,b,e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{n} e_i^2 \tag{3}$$

The uniform distribution for the high dimensional image is given below

$$y_i = [w^T \phi(x_i) + b] = 1 - e_{i,} \quad \text{i=1, 2...N}$$

Here,

γ is regarded as the regularization parameter,

e_i is regarded as the order variable and

J is regarded as the cost function that limits the error during classification. Hence the Lagrangian function is then defined as follows:

$$L(w,b,e,\alpha) = J(w,b,e) - \sum_{i=1}^{N} \alpha_i \{y_i [w^T \phi(x_i) + b] - 1 + e_i\} \tag{4}$$

where

α_i is considered as the mean Lagrange multipliers (i=1,2,...,N). The Eq.(4) is obtained by segregating the Lagrangian function that takes into account the parameters w, b, e_i, α_i and the conditions that limits to 0. The pixel by pixel deduction is now accessible and then the classification using SVM based on Eq. (4) is given by Eq.(5):

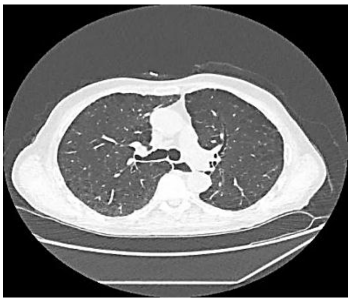
$$y(x) = sign\left(\sum_{i=1}^{N} y_i \alpha_i K(x, x_i) + b) \right) \tag{5}$$

The radial basis function (RBF) is used as a kernel function, which is given below:

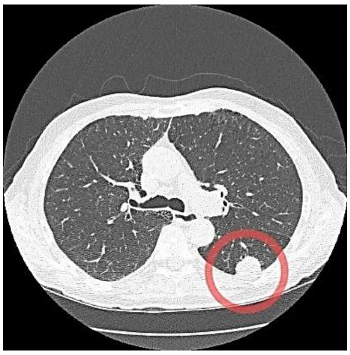
$$K(x, x_i) = \phi(x)^T \phi(x_i) = \exp\left(\frac{-(\|x - x_i\|)^2}{2\sigma^2} \right) \tag{6}$$

1. Performance measures

The ACM's performance is compared to five different classifiers: The performance is compared with various metrics including precision, sensitivity and specificity. Classification performance is measured with 500 images of lung cancer, and tested on 100 images.



(a) Normal Lungs Image



(b) Abnormal Lung image
Figure 3. MRI lungs image

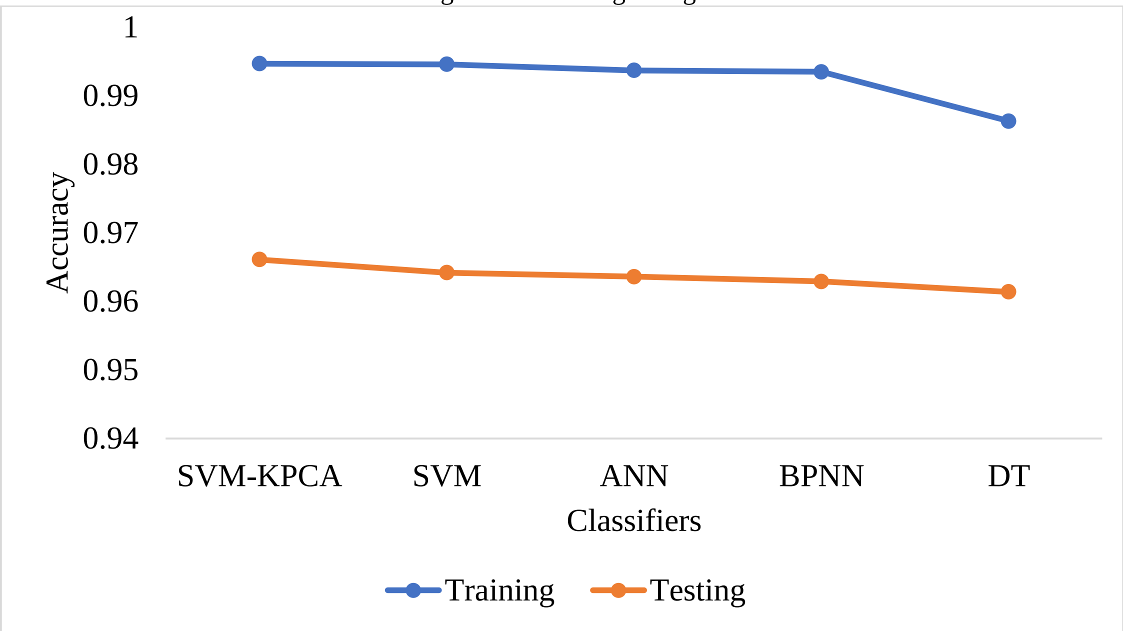


Figure 4: Accuracy of entire classifiers for training and testing

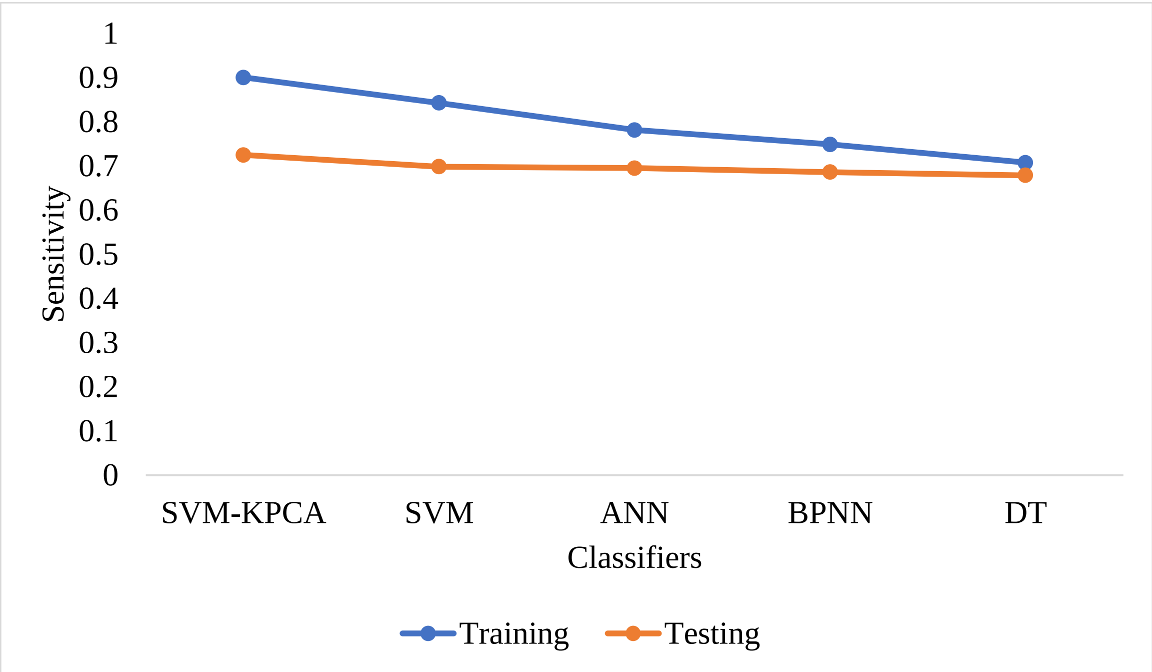


Figure 5: Sensitivity of entire classifiers with 100 and 200 test images

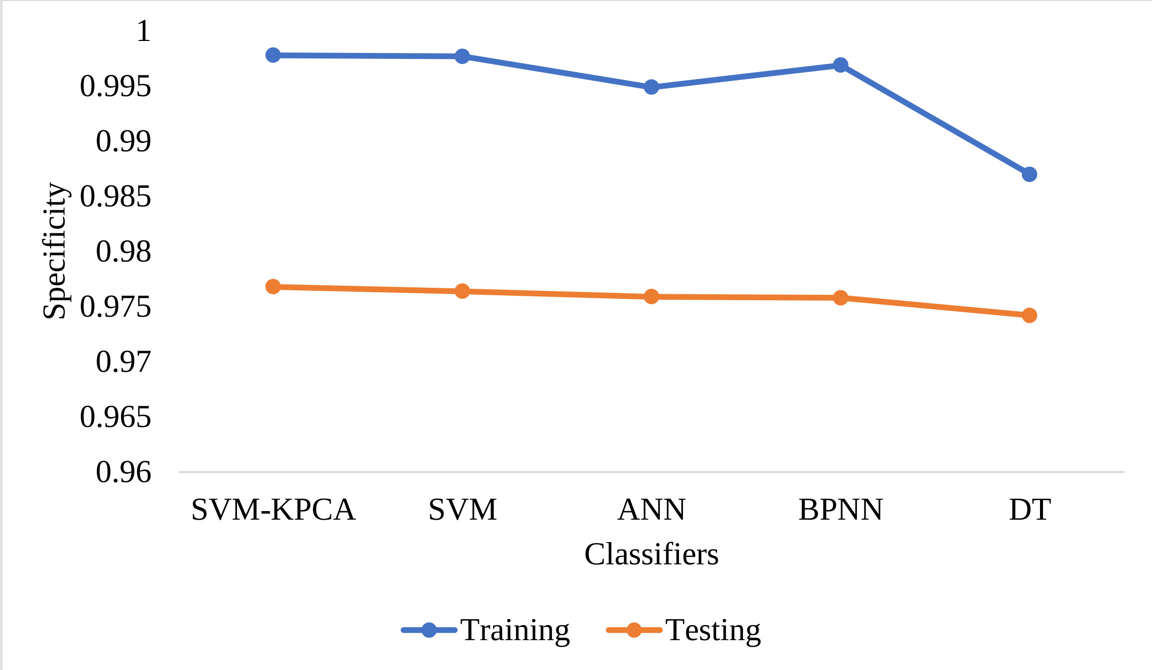


Figure 6: Specificity of entire classifiers with 100 and 200 test images

The results demonstrate that the proposed SVM with ACM is higher than the other methods in the classification of lung cancer. The simulation checks that the system is accurate with low computational images and makes it difficult when it is supplied with higher dimensional images, to achieve greater accuracy.

II. CONCLUSIONS

An ACM that classifies the large datasets of lung cancer images is suggested in this paper. The pre-processing of the image to remove the noises in the samples are carried out well. The extraction of features using KPCA enables the SVM to classify the images into benign or malignant. This is designed to classify the exact lesion into the large lung image collection and tested on test datasets, which are not used for training. The results of the simulation show that the ACM improvement in the classification accuracy of the lung cancer images and has improved its performance compared to existing methods.

III. REFERENCES

[1] Guo, Y., Song, Q., Jiang, M., Guo, Y., Xu, P., Zhang, Y., ... & Yao, X. (2020). Histological Subtypes Classification of Lung Cancers on CT Images Using 3D Deep Learning and Radiomics. *Academic Radiology*.

[2] Khan, M. A., Rubab, S., Kashif, A., Sharif, M. I., Muhammad, N., Shah, J. H., ... & Satapathy, S. C. (2020). Lungs cancer classification from CT images: An integrated design of contrast based classical features fusion and selection. *Pattern Recognition Letters*, *129*, 77-85.

[3] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems* (pp. 3146-3154).

[4] Robertson, S., Azizpour, H., Smith, K., & Hartman, J. (2018). Digital image analysis in breast pathology—from image processing techniques to artificial intelligence. *Translational Research*, *194*, 19-35.

[5] Saritas, M. M., & Yasar, A. (2019). Performance analysis of ANN and Naive Bayes classification algorithm for data classification. *International Journal of Intelligent Systems and Applications in Engineering*, *7*(2), 88-91.

[6] Tandel, G. S., Biswas, M., Kakde, O. G., Tiwari, A., Suri, H. S., Turk, M., ... & Madhusudhan, B. K. (2019). A review on a deep learning perspective in brain cancer classification. *Cancers*, *11*(1), 111.

[7] Tang, A., Hallouch, O., Chernyak, V., Kamaya, A., & Sirlin, C. B. (2018). Epidemiology of hepatocellular carcinoma: target population for surveillance and diagnosis. *Abdominal Radiology*, *43*(1), 13-25.

[8] Weitschek, E., Di Lauro, S., Cappelli, E., Bertolazzi, P., & Felici, G. (2018). CamurWeb: A classification software and a large knowledge base for gene expression data of cancer. *BMC bioinformatics*, *19*(10), 354.

[9] Sahu, S., Singh, A. K., Ghrera, S. P., & Elhoseny, M. (2019). An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE. *Optics & Laser Technology*, *110*, 87-98.

[10] Mansouri, M., Nounou, M., Nounou, H., & Karim, N. (2016). Kernel PCA-based GLRT for nonlinear fault detection of chemical processes. *Journal of Loss Prevention in the Process Industries*, *40*, 334-347.
[11] Deng, W., Yao, R., Zhao, H., Yang, X., & Li, G. (2019). A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm. *Soft Computing*, *23*(7), 2445-2462.
[12] Babu.M., Dr.G.Nanthakumar (2019) Diagnosis of Liver Images Lesions in MR Images using Improved Segmentation and Classification Task, IJRTE- International Journal of Recent Technology and Engineering, ISSN No: 2277-3878, Volume-8, Issue-1S4, Pages 748-752.
[13] Babu.M., Dr.G.Nanthakumar "(2019) , Liver Lesions Diagnosis Using GWO – Anfis Framework, IJEAT- International Journal of Engineering and Advanced Technology, ISSN No: 2249-8958, Volume-8, Issue-6, August 2019, Pages 3692-3696.