IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Hybrid the Support Vector Machine and Firefly Algorithm to Detect Lung Cancer

[1]Neha Sharma, [2]Mandeep Kaur

[1]Research Scholar, [2]Assistant Professor

[1]Electronics and Communication Engineering,

[1]Punjabi University, Patiala, India.

Abstract: In the last few years, the death rate in men and women increased due to lung cancer. Thus, early detection of lung cancer increases the chances of survival of patients. In this paper, we have designed a method by hybrid the support vector machine and firefly algorithm to detect and classify lung cancer. The proposed method contains three steps, namely, pre-processing, appropriate parameter value selection, and classification. We have used the median filter in the pre-processing step to remove noise. After that, the SVM algorithm is applied to detect and classify lung cancer. In the last, the firefly algorithm is used to select appropriate parameter value selection for the SVM algorithm to enhance accuracy. The experimental results were performed on the standard dataset. The performance analysis of the proposed method was done using various parameters. The results show that the proposed method is superior as compared to the existing methods such as SVM, KNN, RF, and GNB.

***Index Terms* - Firefly Algorithm, Lung Cancer, Median Filter, Support Vector Machine.**

I. Introduction

Lung cancer is also called "Lung Carcinoma", is cancer affected by uncontrolled cell growth. The growth of this cell can spread to other parts of the body through the transfer process [1]. Generally, there are two types of lung cancer, one is "small cell lung cancer (SCLC)" and the other is "no-small cell lung cancer (NSCLC)". The main symptoms of lung cancer are "cough, weight loss, shortness of breath and chest pain". One of the main causes of lung cancer is smoking. In addition to passive smoke, air pollution and genetic factors may also cause lung cancer. In addition to other risk factors, the prevention of smoking plays an important role in the prevention of lung cancer [2]. Lung cancer is one of the most dangerous and common cancers in the world. Early diagnosis of "lung cancer" can prolong the survival time of patients. For doctors, CT scans can be difficult to determine the stage of cancer. In the technological era, computers can help to calculate the phase of lung cancer more precisely [3].

Various image processing and machine learning techniques have been used in the literature to detect lung cancer. Various authors have developed models for the use of statistical parametric methods for physicians to detect structural stages of lung cancer based on "Gray Matter-Level Matrix " (GLCM) analysis [4]. Some techniques include image acquisition, pre-processing, feature unpacking, and eventual sorting. Two methods were used to analyze the characteristics: structured image analysis based on the "Gray Level Co-Occurrence Matrix " (GLCM) and statistical parameter approach. In the traditional method, the statistical parameter approach and the SVM classifier provide better performance [5 - 6]. Because SVM performance is sensitive to ker and C control parameters, these parameters should be carefully selected to increase classification accuracy.

The main contribution of this paper is to design a method that accurately detects and classifies lung cancer. To achieve this goal, the firefly algorithm and support vector machine (SVM) algorithm are hybrid. In the proposed method, the SVM parameters, namely, kernel function parameter (γ) and the regularization parameter (C) values optimally selected using the firefly algorithm. Besides that, the pre-processing of the lung cancer images is done using a median filter to remove the noise. The experimental results show that the improved SVM algorithm provides better accuracy, precision, recall, and F1-score value as compared to the existing algorithms such as SVM, KNN, RF, and GNB [6].

The rest of the paper is as follows. Section 2 gives an overview of the median filter, support vector machine, and firefly algorithm. Section 3 presents the proposed method. Section 4 shows the experimental results were performed to validate the proposed method. In last, the conclusion and future work is drawn in Section 5.

II. Related Work

In this section, to understand the proposed method, an overview of median filter, support vector machine, and firefly algorithm is given.

2.1 Median Filter

The median filter is used to record the size of all the vectors in the mask and to differentiate by size. This is based on moving the window in the image (as in the curved position) and assuming that the output pixels are the average illumination inside the input window. Because a simple median filter is higher than the average filter, we get the median of the data, not the average of the image. The median size of the pixel is then used to replace the pixel being studied. The median of the image is stronger when exposed to noise [7].

2.2 Support Vector Machine

The SVM classification is based on obtaining an upper bound between two classes [8]. Linear isolation can take place directly, but in non-linear cases, the data are placed in characteristic spaces and isolation is performed in extreme spaces. A kernel is a string used to map data from an input space that defines a kernel function. Three types of kernel functions are available. They are called Linear, Polynomial, and RBF. Polynomials and RBF kernels are used to capture nonlinear data over space. The SVM classifier can be written in the same way as Equation (1), and the RBF kernel function is shown in Equation (2).

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i k(x_i, x) + b \quad (1)$$

$$k(x, x') = exp\left(\frac{||x - x\prime||^2}{2\sigma^2}\right) \quad (2)$$

$$\gamma = \frac{1}{2\sigma^2} \quad (3)$$

Equation (3) shows γ, the width of the nucleus, which is a positive number that determines the coefficient of the scale of the nucleus is used to define the shape of the "peak" and is wider or sharper. An SVB classifier with an RBF core is given by Equation (4).

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i exp\left(\frac{||x - x\prime||^2}{2\sigma^2}\right) + b \quad (4)$$

$$f(x) = C \sum_{i=1}^{N} \alpha_i y_i exp\left(\frac{||x - x\prime||^2}{2\sigma^2}\right) + b \quad (5)$$

An SVM classifier containing an RBF core has two parameters; core size (γ) and box limit (C). Box constraints are control parameters that increase the marginal level and control the gap between training data errors. Equation (5) shows SVM (C). To achieve the optimal parameters, the RBF kernel variable was scaled on the SVM.

2.3 Firefly Algorithm

The Firefly algorithm is a "swarm based metaheuristic algorithm " was presented by Young [9]. The algorithm mimics the interaction of glare with flashing light. The algorithm considers all fireflies to be unisex, which means that any firefly can be attracted to other fireflies. The charm of the hearth is directly proportional to its brightness and depends on the purpose and function. The firefly pulls a brighter fire cake. Furthermore, the clarity decreases according to the inverse quadratic law from a distance, as shown in the equation. (6)

$$I \propto \frac{1}{r^2} \quad (6)$$

If light passes through an environment with a light absorption coefficient γ, the light intensity can be given by the distance r from the source, for example, by the equation. (7)

$$I = I_0 e^{-\gamma r^2} \quad (7)$$

where I_0 is light intensity at the source. Similarly, the brightness, β, can be given as in Eq. (8).

$$\beta = \beta_0 e^{-\gamma r^2} \quad (8)$$

The general illumination function of Bright ≥ 1 is equal to (9) [5]. In fact, a uniform decrease function can be used.

$$\beta = \beta_0 e^{-\gamma r^{\omega}} \quad (9)$$

In the algorithm, randomly generated solutions, such as fire flies, are distributed by light energy based on the performance of a target function. This intensity is assumed to be directly proportional to the light of the flame. In the event of a reduction problem, the solution with the lowest functional value shall be assigned the maximum light intensity. Once the intensity or clarity of the solution has been determined, all igniters better illuminate the flame. For the brightest fire, it moves randomly to its neighbor and does a local search. Therefore, for two fires, if j firefly is clearer than i, then i is the equation. (10)

$$x_i \coloneqq x_i + \beta_0 e^{-\gamma r_{ij}^2}(x_j - x_i) + \propto (\in () - 0.5) \quad (10)$$

β0 is an attractive property of xj at rj = 0 [5] In order for the author to implement implementation0 = 1, γ is an algorithmic parameter that determines how much the renewal process depends on the distance between the two fires, α. is a parameter of the random motion step length algorithm, and ε () is a uniformly distributed random vector with a value between 0 and 1. it is omitted in accordance with Equation (11)

$$x_b \coloneqq x_b + \propto (\in () - 0.5) \quad (11)$$

These fireflies location updates will continue until final conditions are met. The completion criterion may be the maximum number of iterations, the tolerance of the optimal value if known, or the absence of consecutive iteration improvements.

III. PROPOSED METHOD

This section shows the proposed method is designed to classify the lung cancer images. The flowchart of it is shown in Figure 1. Initially, the lung cancer standard dataset is taken from Kaggle. After that, the median filter is applied to remove the noise present in the lung cancer images. Next, the filtered image is processed and its statistical features, namely, mean, variance, and standard deviation are extracted. After the feature was extracted, we have trained the SVM classifier for classification purposes and calculated performance parameters for it. In the last, to enhance the accuracy of the SVM classifier, we have employed a firefly optimization algorithm that optimally selects the classifier parameters such as kernel scale and box constraint.

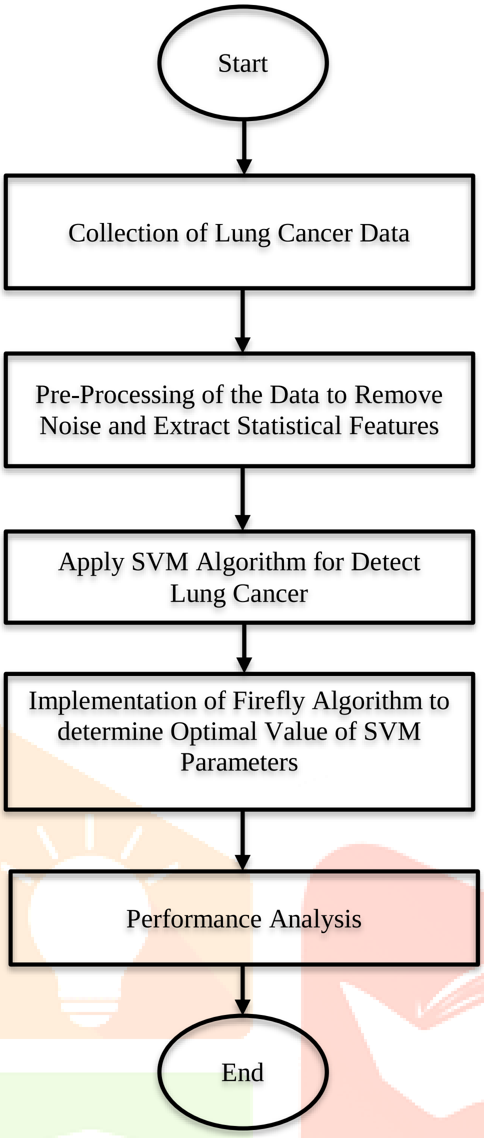


Figure 1 Flowchart of the Proposed Method

IV. Experimental Results

This section shows the experimental results were performed for the proposed method to validate it over the existing methods. The standard dataset was downloaded from cancerimagingarchive.net and Kaggle [10]. The dataset contains a large amount of dataset and lung cancer images of dimension 512 × 512. Next, we have measured the various performance parameters for it, as explained below [11].

- Accuracy: In terms of ranking, accuracy is the record common rating indicator. Precision can be defined as the sum of the positive real and negative real in relation to the total number of cases. Calculate the equation (12)

“Accuracy=(TP+TN)/(TP+FP+TN+FN)” (12)

- Precision: It determines the correct actual level of positive recognition. Simple “precision “can be expressed by dividing the number of true positive numbers by the number of true positive numbers and the number of false positive numbers. Calculate the equation (13)

“Precision =(TP)/(TP+FP) “ (13)

- Recall: Explain the precise ratio of real positive recognition. A added positive recall can be stated as a real positive number divided by a real positive number and a false negative number. Calculate the equation (14)

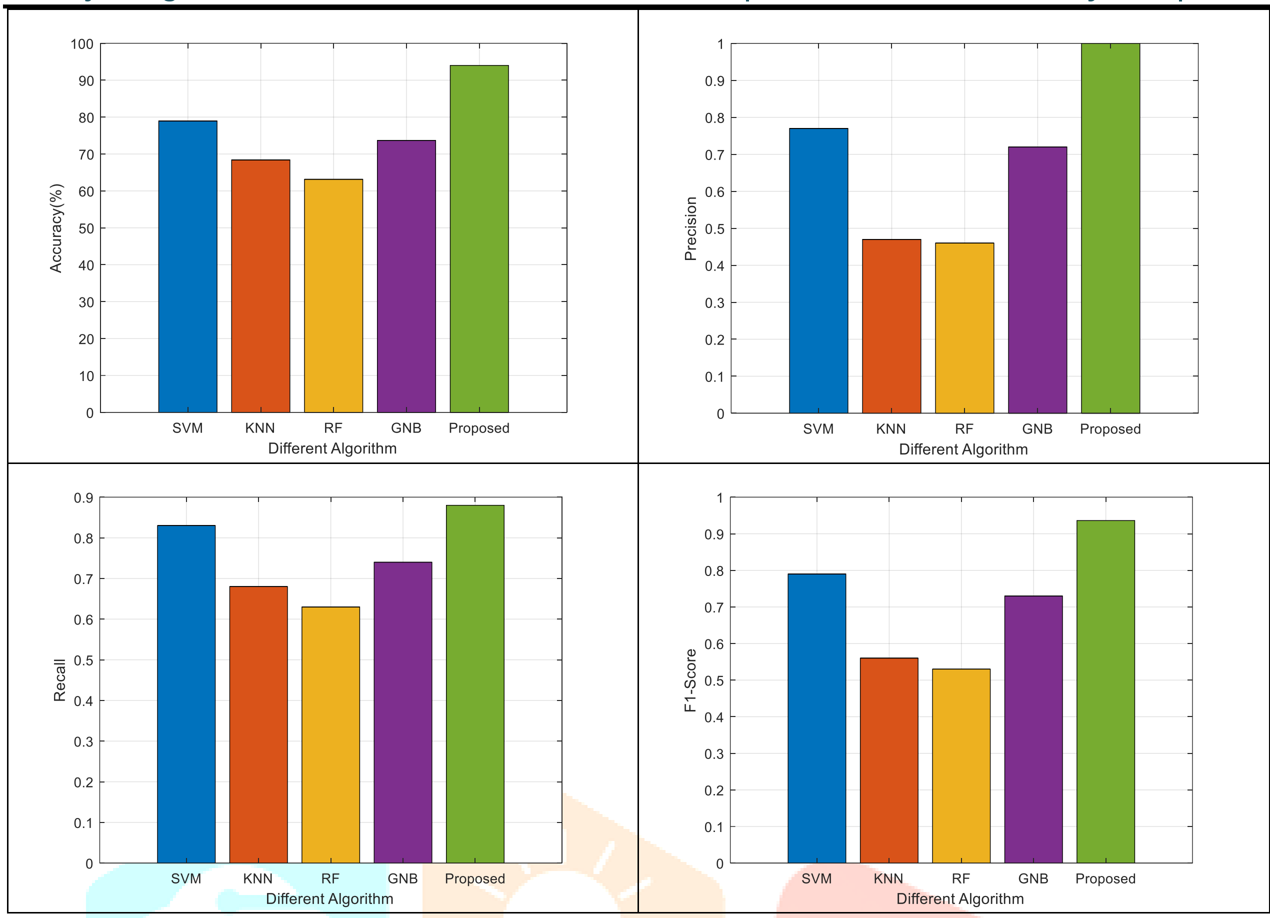
“Recall =(TP)/(TP+FN) “ (14)

- F1 Score: The F1 Score is the 2 x ((precision x recall)/ (precision+recall)). It is as well-known as the F Score or the F Measure. Calculate the equation (15)

“F1 Score= 2x ((precision x recall)/ (precision + recall))” (15)

Next, the comparative analysis of the proposed method is done with the existing methods in Table 1. The results show that the proposed method achieves better accuracy, precision, recall, and F1 score due to selecting optimal values of SVM parameters using the firefly algorithm.

Table 1 Comparative Analysis with the Existing Methods [6] in terms of Accuracy, Precision, Recall, and F1 score.



V. Conclusion and Future Scope

In this paper, we have presented a method that accurately detects lung cancer as compared to the existing methods. The reason behind this is that the appropriate values of the support vector machine parameters are calculated using the firefly algorithm. In the future, the other artificial intelligence algorithm and multi-objective-based fitness function will be explored to improve the accuracy.

References

[1] Lovly, C. M., & Horn, L. (2012). Strategies for overcoming EGFR resistance in the treatment of advanced-stage NSCLC. *Current treatment options in oncology*, *13*(4), 516-526.

[2] Okuyama, A., Shibata, A., & Nishimoto, H. (2018). Critical points for interpreting patients' survival rate using cancer registries: A literature review. *Journal of epidemiology*, *28*(2), 61-66.

[3] Kulkarni, A., & Panditrao, A. (2014, May). Classification of lung cancer stages on CT scan images using image processing. In *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies* (pp. 1384-1388). IEEE.

[4] Mudathir, O., Kamil, A. E., Salah, S., Gamar, M., & Nouraldaem, Z. (2021). Detection of lung Cancer on CT Scan Using Image Processing Techniques. *FES Journal of Engineering Sciences*, *10*(1), 1-5.

[5] Herman, G. T. (2009). *Fundamentals of computerized tomography: image reconstruction from projections*. Springer Science & Business Media.

[6] Islam, M., & Rab, R. (2019, October). Analysis of CT Scan Images to Predict Lung Cancer Stages Using Image Processing Techniques. In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 0961-0967). IEEE.

[7] Yasmeen, D., Nisha, S. S., & Sathik, M. M. (2019). Analytical Study of Various Filters in Lung CT Images.

[8] Zainuddin, A., Mansor, W., Lee, K. Y., & Mahmoodin, Z. (2018). Performance of support vector machine in classifying EEG signal of dyslexic children using RBF kernel. *Indonesian Journal of Electrical Engineering and Computer Science*, *9*(2), 403-409.

[9] Khan, W. A., Hamadneh, N. N., Tilahun, S. L., & Ngnotchouye, J. M. (2016). A review and comparative study of firefly algorithm and its modified versions. *Optimization Algorithms-Methods and Applications*, 281-313.

[10] https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70224216

[11] Ahn, H. K., Lee, H., Kim, S. G., & Hyun, S. H. (2019). Pre-treatment 18F-FDG PET-based radiomics predict survival in resected non-small cell lung cancer. *Clinical radiology*, *74*(6), 467-473.