Hybrid Sparrow-Fly Swarm Optimized Support Vector Machine Approach for Automated Lung Cancer Identification from CT Imaging

Dr. Sumitra Padmanabhan[1], Guntaj J[2], Dr. Arvind Kumar Pandey[3], Kalyan Acharjya[4], Sunil MP[5], Deepak Minhas[6]

[1]Associate Professor, Department of uGDX, ATLAS SkillTech University, Mumbai, Maharashta, India, Email Id- sumitra.padmanabhan@atlasuniversity.edu.in, Orcid Id- 0000 0003 4846 080X
[2]Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India. guntaj.j.orp@chitkara.edu.in https://orcid.org/0009-0006-2409-4648
[3]Associate Professor, Department of Computer Science & IT, ARKA JAIN University, Jamshedpur, Jharkhand, India, Email Id- dr.arvind@arkajainuniversity.ac.in, Orcid Id- 0000-0001-5294-0190
[4]Assistant Professor, Maharishi School of Engineering & Technology, Maharishi University of Information Technology, Uttar Pradesh, India, Email Id- kalyan.acharjya@gmail.com, Orcid Id- 0000-0001-7802-7332
[5]Assistant Professor - 3, Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Karnataka - 562112, India, Email Id- mp.sunil@jainuniversity.ac.in, Orcid Id- 0000-0002-7737-4145
[6]Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh- 174103 India. deepak.minhas.orp@chitkara.edu.in https://orcid.org/0009-0002-0585-6046

Unregulated cell growth makes lung cancer (LC) more effective than other respiratory disorders. One of the main and obvious strategies to increase patient survival rates is early detection of LC. Using a hybrid Sparrow-Fly Swarm Optimized Support Vector Machine (HS-FSO-SVM), lung tumors in CT (Computed Tomography) scans were automatically identified in this study. The suggested technique provides accurate and efficient production by integrating the processes of data collecting, data pre-processing, feature extraction and classification. At this stage, we are collecting high-resolution CT images, which are a vital tool for diagnosing LC. Next, a Savitzky-Golay filter is used for data preparation to improve picture quality by cutting down on noise and increasing signal-to-noise ratios generally. Linear discriminant analysis (LDA) is used for feature extraction. To reduce the complexity of the data while maintaining crucial information, LDA is used to find the most distinguishing characteristics in the pre processed CT images. This improves the categorization process's efficacy and efficiency. An HS-FSO-SVM is used in our approach's classification phases. MATLAB software was utilized in this study to generate findings. Our methodology was evaluated by means of a comparison with existing approaches by utilizing metrics such as precision, accuracy, recall and F1-measure. The corresponding values were determined to be 96.7%, 97%, 93.5% and 90.9%. To fine-tune the SVM parameters and enhance the classifier's performance, a combination of the Sparrow and Firefly optimization approaches has been developed. These are analyzed using an immediately trained neural network with deep learning (DL) for predicting LC. Finally, the result demonstrates that our method has superior classification accuracy than the competing systems.

Keywords: Lung Cancer, CT Images, Identification, Patients and Prediction.

1. Introduction

The illness caused by abnormal tissue development in the human body is known as cancer. One of the most significant illnesses that present a risk to human health today is cancer [1]. One of the well-known medical imaging modalities is used to gather pictures for films in the CT scan. While X-rays are used to see a CT scan, radio waves are used to notice the difference in MRI (magnetic resonance imaging) pictures. Most clinicians prefer CT scans over MRIs for their patients because of the higher processing cost of CT data compared to MRI data. The primary benefit of a CT scan is that it can be completed quickly and shows the thin structure, tissues and organs of the brain [2]. Physicians and radiologists utilize CT scans to diagnose and confirm the existence of illnesses, which depict the morphologic levels of diseases directly, characterize the trends and severity of diseases, track the progression of diseases clinically and assess how well the patients respond to treatment. Spiral scans, made possible by the volumetric CT technology, eliminate artifacts from partial volume impacts, heart motion and uneven breathing cycles while cutting down on scan duration in thoracic imaging [3]. In the realm of medical image processing and detection technologies, the proposed method for automatic LC identification from CT imaging is a novel and cutting-edge method. Early identification is essential to improve patient results in LC, a deadly illness with a high death rate. Tobacco use is the primary contributor to LC and the prevalence of the disease has either peaked or increases in a number of nations. This indicates that the incidence of LC will continue to rise for at least the next several decades [4]. The two primary kinds of LC are non-small-cell LC (NSCLC) and small-cell LC (SCLC). Smoking, toxic airborne particles, sex, genes, age and other variables are generally the causes of LC [5]. The main goal of this research is to create an automated system that can identify instances of LC from CT scans using a hybrid strategy that combines sparrow-fly swarm optimization approaches.

2. Related works

According to the author of, [6] offered an extensive structure for identifying LC in chest by low-dose computed tomography (LDCT) scans. Their methodology involves training a cancer forecaster on the Kaggle data sample after a nodule identifier on the LIDC-IDRI data sample. The most deadly cancer in the world was LC. Research indicated that early diagnosis by LDCT scans can decrease the number of fatalities resulting from the condition. Study [7] suggested a novel approach to LC prediction by utilizing machine learning (ML) and optimized manipulation of images was presented. Images from CT scan datasets of non-small cell LC are gathered to identify the disease. The obtained pictures were analyzed using the multilayer brightness-preserving technique, which carefully inspected each pixel, removed noise, and improved the lung image's quality. To describe a DL [8] approach for automated LC diagnosis in the paper. The proposed framework uses procedures to identify and evaluate the nodule's malignancy based on 3D lung CT scan pictures. Based on the 3D Deep Convolutional Neural Network, a novel computer-aided decision support system for lung nodule identification was developed (3DDCNN) [9]. When radiologists were making decisions about diagnosing LC, the decision support technology gave them a second viewpoint. They utilized a multi-region Proposal Network (mRPN) and median intensity projection to automatically choose possible regions of interest using the 3-dimensional information from CT images. Author [10] described an automated method for identifying LC using CT scan images. The suggested approach for the identification of LC employs techniques such as pre-processing images using median filtering, followed by segmenting the lung region of interest by mathematical morphological procedures. The accuracy and other performance metrics of the current designs of LC detection and classification systems were constrained since they rely on hand-engineered methods. Motivated by the remarkable DL performance in several tasks linked to identification, an ALCDC [11] system employing DL was presented, utilizing CT scan images. The suggested an automatic 3-D lung segmentation method for computed tomography (CT) [12] images was characterized using an active contour model. Using the suggested segmentation model, the active contour model (ACM) and the local image bias field formulation were integrated. According to the author of, [13] created a replicable ML module for the identification of LC and to assess as well as contrast the methodologies along with the outcomes of the most efficient algorithms created during the Data Science Bowl on Kaggle. The majority of solutions utilized separate modules for segmentation, classification and image preprocessing. The described the development of an intelligent system for diagnosing lung tumors utilizing a variety of image-processing techniques. The supported vector machine (SVM) [14] kernel was used in the simulated phases of image improvement, post-processing, segmentation, feature extraction and classification. To described fully automated lung segmentation and identification system for 3D pulmonary X-ray CT [15] images was provided. The approach has the benefit of separating the nodules that are connected to the lung wall, which were eliminated by conventional methods for segmenting the lungs. The lungs' CT images were acquired from the Cancer Imaging Archive (CIA), and the emphasis was on reducing misclassification to improve lung image quality and LC detection [16] dataset. The weighted mean equalization of the histograms technique was used to erase noise from the images, improving their quality. The enhanced profuse clustering method (IPCT) was used to segment the afflicted zone. The affected area offers a range of spectral characteristics.

2.1 Problem statement

LC requires extra attention in the various illness assessment procedures since it affects both men and women, increasing the fatality rate. Traditional LC prediction algorithms failed to achieve the necessary accuracy due to poor image quality that interfered with the segmentation procedure. The previous few decades have seen a significant problem with the identification of early phase LC using chest CT scans, since accurate early identification of malignant lung nodules can extend the patient's life. Methods for automated LC diagnosis are crucial. Due to the striking structural similarities among malignant and benign lung tumors, it is difficult to correctly identify LC in its premature stage. Another difficult issue is identifying the LC tumor in its early stages. A significant number of lives might be saved by early tumor detection. To overcome the above mentioned issues, we proposed a HS-FSO-SVM method for the automatic detection of LC from CT images.

3. Methodology

In this section, we will discuss the materials and methods for the proposed system. We collect the LIDC data samples. The study used the Savitzky- Golay filter for preprocessing the collected data. The preprocessed data were extracted by utilizing linear discriminant analysis (LDA). The samples were optimized by using the hybrid sparrow-fly swarm optimized support vector machine (HS-FSO-SVM). The architecture of the recommended system is depicted in Figure 1.

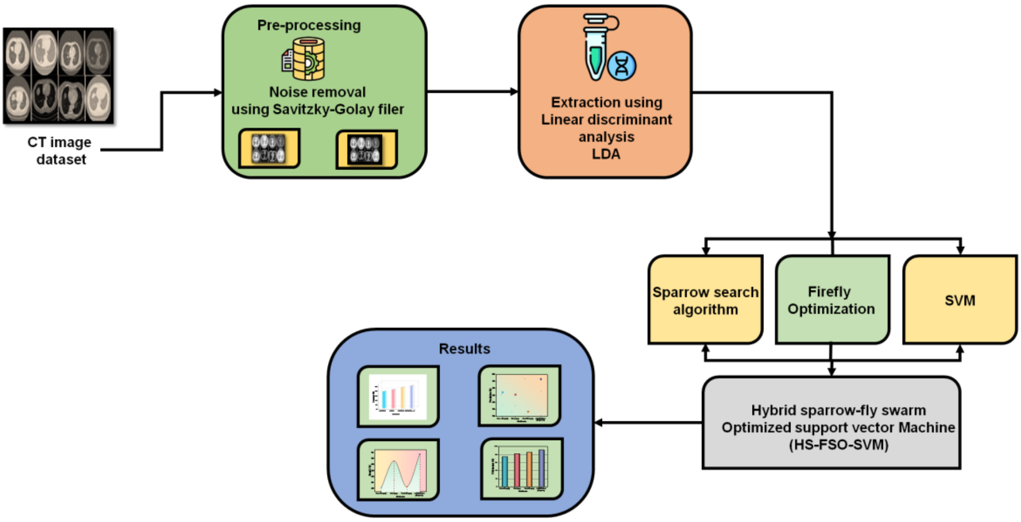

Figure 1: Framework for Proposed HS-FSO-SVM [Source: Author]

3.1 Dataset

The Lung Image Database Consortium (LIDC) is launched by National Cancer Institute (NCI) to enhance investigation and development efforts. Three kinds of objects were added to the LIDC database, which is 4 radiologists will label: nodules with a diameter of three millimeters or more, with assumed histology, nodules fewer than 3 millimeters in diameter, with an unclear nature and benign non-nodules under 3 mm in size. There are 1008 records of patients in the database. Figure 2 shows the example CT scan images from the LIDC dataset [20].

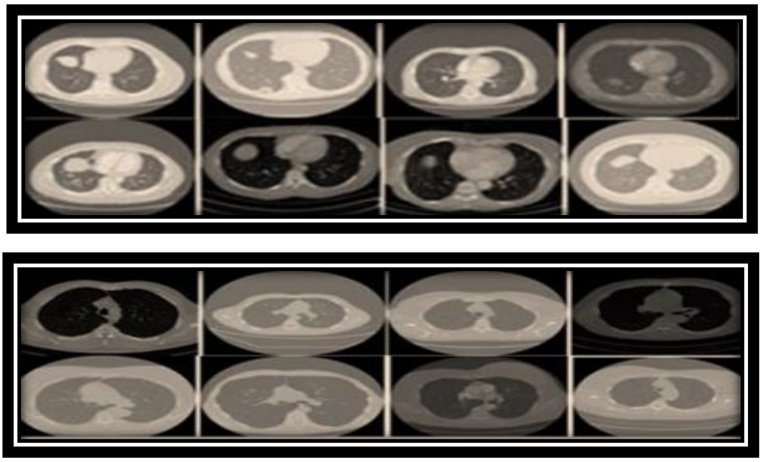

Figure 2: CT images from the LIDC dataset [20]

3.2 Data preprocessing

Preparing a raw data for analysis or ML involves cleaning and formatting. This process is known as data preparation.

3.2.1 Savitzky –Golay filter

A less complicated least squares-fit convolution was presented by Savitzky and Golay for smearing and computing the variance of a spectrum, or a collection of consecutive numbers. The weight coefficients, which are called coefficients in the following, using quadratic least squares fits in the filter window when a signal is present. This polynomial is meant to minimize the bias generated by the filter while preserving the higher moments in the data. As long as the points are graphed continuously as well as smoothly and as long as the points are distributed along the chosen border at a predefined along with a consistent spacing, any series of successive data can be filtered using this method. Time series unquestionably assures these circumstances. Equation (1) is for smoothing time series using least-squares convolution.

$$A_h^* = \frac{\sum_{z=-b}^{z=b} x_z A_{h+z}}{B} \tag{1}$$

Where B is the quantity of convoluting integers, A represents initial time series value, A^* is the subsequent time series value, x_z is the coefficient for the z-th time series value of the filter and smoothing window size $(2z + 1)$ is represented by B. The original coordinate data table's running index is represented by the index h. The smoothing array's dimensions are $2m + 1$, where the smoothing window's half-width is b. The factors of a Savitzky-Golay filter (X_z)can be acquired using a modified version of Savitzky as well as Golay's method, or by applying the formulas.

3.3 Feature extraction

Feature extraction aims to highlight the salient features and patterns in the data so that ML algorithms can analyze the data more quickly and provide accurate predictions or classifications.

3.3.1 Linear discriminant analysis (LDA)

One common use of LDA is to identify the linear characteristics that optimize among-class separation of data while reducing in-class dispersion. Consider a training data set containing b samples$\{t^1 \dots . t^B\}$, where t^h represents a column vector of distance e for each example. The training examples are instances of the R classes. Let b_e = $|X_e|$be the collection of the instances of class r and let b_e be the total number of examples in class $e = 1 \dots E$. In LDA, the scatter matrices within and among classes are computed as follows in Equation (2):

$$F_y = \frac{\sum_e \sum_{h \in x_e} (t^h - b_e)(t^h - b_e)^S}{N}, F_o = \frac{\sum_e b_e (b_e - b)(b_e - b)^S}{N} \quad (2)$$

Where $t^h$is the average of the eth class, $n_e = \frac{1}{n_e} \sum_{h \epsilon C_e} t^h$, and $b = \frac{1}{N} \sum_h t^h$is the average of the data sample. To look for a straightforward adjustment $t \rightarrow Y^S t$ that improves the virtual among-class conflict to the inside-group conflict, where Y is a $r \times r' matrix$ and r is the desired size. Demonstrating that $F_o y = \lambda F_y y$ is the result of the correspondence between the generalized eigenvectors and the r largest eigenvalues is represented by the columns of the ideal Y. This conclusion leads to the simultaneous diagonalization by Y of the scatter matrices $Y^S F_o Y$and $Y^S F_y Y$. As an alternative, LDA rectifies data within and across classes.

3.4 Proposed method for the study objective

The HS-FSO-SVM technique combines the robustness of SVM with optimization inspired by nature to assist in the automated diagnosis of LC from CT images. For detecting malignant spots in CT images, our hybrid technique performs well in improving feature selection and fine-tuning SVM parameters.

3.4.1 Support vector machine (SVM)

The SVM is capable of generalization and adheres to the structural risk minimization concept. The SVM offers notable benefits for managing tiny samples and nonlinear data due to its adaptive learning and nonlinear approximation capabilities. $\boldsymbol{\phi(z)}$ is the nonlinear mapping function that maps the $\boldsymbol{u}^{\text{th}}$ input sample $\boldsymbol{z_u}$ for a given dataset from a low-dimensional vector space to a higher-dimensional vector space. In high dimensional space, the linear regression function can be simplified to the following in Equation (3):

$$\boldsymbol{d(z) = q . \phi(z) + vz \in E^s, v \in E} \quad (3)$$

In the sample space, $\boldsymbol{z}$ is the input vector, $\boldsymbol{E}$ represents the field of real numbers, sample space's dimension is denoted by $\boldsymbol{s}$,$\boldsymbol{\omega}$represents weight, $\boldsymbol{v}$ is the bias parameter and $\boldsymbol{d(z)}$ signals the predicted values. By using the structural risk reduction concept, the regression of the support vectors issue is changed to a quantitative enhancement issue with restrictions. Two slack parameters, $\boldsymbol{\xi_u}$and $\boldsymbol{\xi_u^*}$, are introduced and the fitting mistake is taken into account. The following are the restrictions and optimization function in Equations (4-5):

$$\boldsymbol{min \frac{1}{2} \|\omega\|^2 + X \sum_{h=1}^{b} (\xi_u + \xi_u^*)} \quad (4)$$

$$\begin{cases} t_u - \omega . z_u - v \le \varepsilon + \xi_u \\ \omega . z_u - v - t_u \le \varepsilon + \xi_u^* \\ \quad \xi_u \ge 0 \\ \xi_u^* \ge 0 \; u = 1, 2, .. b \end{cases} \tag{5}$$

The number of samples is denoted by $u = 1, 2, \ldots, b$. ε Denote the loss function and X is the punishment factor, $t_u = d(z_u)$. When the Lagrange multiplier is included, the equation's solution can be converted into the Lagrange equation's saddle point result. The final optimization issue is given as follows after using the duality theorem and computing the partial derivative of the Equations (6-7):

$$max\left\{-\frac{1}{2}\sum_{u,h=1}^{b}(\alpha_u - \alpha_u^*)(\alpha_h - \alpha_h^*)J(z_u - z_h) + \sum_{u=1}^{b}(\alpha_u - \alpha_u^*)t_u - \varepsilon\sum_{u=1}^{b}(\alpha_u + \alpha_u^*)\right\}$$

$$s.t. \sum_{u=1}^{b}(\alpha_h - \alpha_h^*) = 0 \; \alpha_h, \alpha_h^* \in [0, X] \tag{6}$$

Where α_u, α_u^* are the Lagrange multiplier and $\alpha_h > 0, \alpha_h^* > 0$, $J(z_u, z_h) = \varphi(z_u). \varphi(z_h)$indicates function for kernel, $u, h = 1, 2, \ldots, b$.

Additionally, the nonlinear function can be written:

$$d(z) = \sum_{u=1}^{b}(\alpha_h - \alpha_h^*)J(z_u, z_h) + v \tag{7}$$

The radial basis function's larger convergence areas and superior resolution make it regarded as the optimal classification function. The kernel function in this study is the radial basis function, which has the following definition in Equation (8):

$$J(z_u, z_h) = \exp\left(-\frac{\|z - z_u\|}{2\delta^2}\right) \tag{8}$$

Where δ is the radial basis kernel function's width. The connection among the original input and output spaces can be seen by applying the radial basis function. The parameter d and the penalty component C in the conventional SVM are selected empirically.

3.4.2 Sparrow search algorithm

The three stages of the SSA are discoverer, follower and investigator. The discoverer follows through its name by locating food, pursuing it and directing others in the proper direction. Because 20% of the population is made up of this group, the discoverer looks for a variety of foods. The formula for updating the discoverer's location is in Equation (9).

$$Z_{u,h}^{r+1} = \begin{cases} Z_{u,h}^{r} . exp\left(\frac{-g}{\alpha . N}\right), if \; E_2 < AR. \\ Z_{u,h}^{r} + P.Kif \; E_2 \ge AR \end{cases} \tag{9}$$

In Equation (9), g stands for the number of iterations that are currently in progress; N is the maximum number of iterations, $Z_{u,h}^{r}$ is uth sparrow's current location in the hth dimension, $\alpha\varepsilon[0, 1]$ is a random number, E_2 and AR are the security and warning values, respectively. P Is a normal distribution of random numbers and K indicates that there is one with all of the elements of $1 \times D$. When $E_2 < [0, 1], AR < [0.5, 1], P$ it means that the discoverer can use a broad search method and that the community environment is secure at this moment, with no predators around the sparrows. $\boldsymbol{E_2 \geq AR}$ means that someone in the group has found the predat'or and alerted others, everyone in the group will behave in an anti-predatory manner and the discoverer will guide the follower to a safe place.

Following the discoverer, followers look for food and the area surrounding the discoverer's location. The location of followers modifies the formula in this Equation (10):

$$\boldsymbol{Z_{u,h}^{r+1}} = \begin{cases} \boldsymbol{P.exp\left(\frac{Z_{worst}^{r}-Z_{u,h}^{r}}{u^2}\right) if u > \frac{b}{2}} \\ \boldsymbol{Z_o^{r+1} + \left|Z_{u,h}^{r} - Z_O^{r+1}\right|.L^{+}.K, otherwise} \end{cases} \tag{10}$$

In Equation (10), $\boldsymbol{P}$ is a matrix of size $\boldsymbol{1 \times d}$, where each element has a value of $\boldsymbol{1}$ or $\boldsymbol{-1}$, and $\boldsymbol{L^{+} = L^{R}(LL^{R})^{-1}}$. $\boldsymbol{Z_o}$is the ideal position held by the discoverer, while $\boldsymbol{Z_{worst}}$ indicates the worst position currently. At this stage, when $\boldsymbol{u > b/2}$, the population of sparrows will launch a counterattack when it perceives a threat.

3.4.3 Firefly optimization

The Firefly optimization (FFO) is a stochastic global optimization method influenced by biology or nature. Every firefly in the population indicates a potential solution in the study space and FFO is a population-based meta-heuristic. Using light flashes, the FFO mimics the mating and information-sharing processes of fireflies. This section explains the firefly's primary behavior. Approximately two thousand different species of fireflies exist and the majority of them create brief, melodic flash illumination. Three primary purposes for firefly flash illumination are as follows:

- Attracting potential victims.
- Attraction of potential mates.
- Offers a warning system.

3.4.4 Hybrid Sparrow-Fly Swarm Optimization

It is possible to achieve successful automated LC detection from CT scans by combining the Sparrow Search Algorithm with Firefly Optimization. This hybrid technique improves feature selection and classification accuracy by utilizing the search powers of both algorithms. The pseudo code for hybrid sparrow-fly swarm optimization is represented in Algorithm 1.

Algorithm 1: Hybrid sparrow-fly swarm optimization

```
Objective function $\boldsymbol{d(z)}$,    $\boldsymbol{z = (z_1, \dots . z_s)^R}$
Generate initial population of sparrow fireflies $\boldsymbol{z_u (u = 1, 2, \dots, b)}$
Light intensity $\boldsymbol{U_u}$ at $\boldsymbol{z_u}$is determined by $\boldsymbol{d(z_u)}$
Define light absorption coefficient $\boldsymbol{\gamma}$
while($\boldsymbol{r < MaxGeneration}$)
for$\boldsymbol{u = 1 : b}$ $\boldsymbol{all\ b}$sparrow fireflies
for$\boldsymbol{h = 1 : b}$  $\boldsymbol{all\ b}$ sparrow fireflies
```

```
if($(\boldsymbol{U_h} > \boldsymbol{U_u})$, Move sparrow firefly $\boldsymbol{u}$ towards $\boldsymbol{h}$ in $\boldsymbol{s}$-dimension; end if
prettinessvary with distance $\boldsymbol{e}$throughexp$[-\boldsymbol{\gamma e}]$
    Examine new solutions and modernize light intensity
end for$\boldsymbol{h}$
end for$\boldsymbol{u}$
Rank the sparrow fireflies and find the current best
end while
Post process outcomes and visualization
```

4. Result

MATLAB software is used in this research to stimulate results. To evaluate the efficiency of our proposed method by comparing it with traditional methods such as Conventional neural network with recurrent neural network (CNN- RNN) [17], Optimal deep neural network (ODNN) [18] and Gray wolf optimization with Recurrent neural network (GWO-RNN) [19] with regard to f1-score, accuracy, precision and recall. As a result, the outcomes of our suggested technique are higher than other existing approaches.

4.1 Accuracy

Accuracy is one performance statistic that's applied to classification activities. One way to assess the measurement's efficiency is to compare the input data's actual labels with the anticipated labels. When the predicted and actual labels match, the prediction is considered correct. Accuracy measures from CT Imaging's ability to diagnose lung cancer. Figure 3 and Table 1 demonstrate the comparison of accuracy between the existing and proposed methods. HS-FSO-SVM has accuracy value of 97%, this proves that our method performs superior than other conventional methods in Equation (11).

$$\boldsymbol{Accuracy = \frac{Amount of perfectly categorized data instances}{Total data instances} \times 100} \qquad (11)$$

Table 1: Comparison of accuracy

Methods	Accuracy (%)
CNN - RNN [17]	72.2
ODNN [18]	80.2
GWO-RNN [19]	89.8
HS-FSO-SVM [Proposed]	97

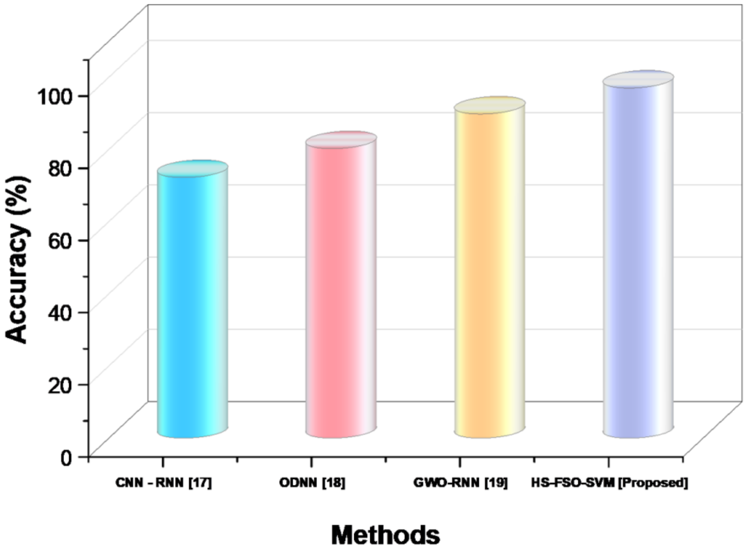

Figure 3: Comparison of accuracy between proposed and traditional methods

4.2 Precision

One typical performance metric for classification tasks is precision. A model determines the percentage of genuine positive predictions that is, accurate positive predictions that are based on the quantity of accurate forecasts it produced. Precision would be the percentage of correct cancer predictions among the forecasts with a positive cancer label using CT scans. By dividing the number of trustworthy and untrustworthy forecasts by the total number of accurate projections, one can calculate the accuracy. A comparison of precision among conventional and recommended methods is depicted in Figure 4 and Table 2. As a result, the value for HSFSO-SVM is higher by 96.7 %. This value proved that our proposed method performs better than other methods as shown in Equation (12).

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \times \mathbf{100} \tag{12}$$

Table 2: Comparison of precision

Methods	Precision (%)
CNN - RNN [17]	87.3
ODNN [18]	85.6
GWO-RNN [19]	73.5
HS-FSO-SVM [Proposed]	96.7

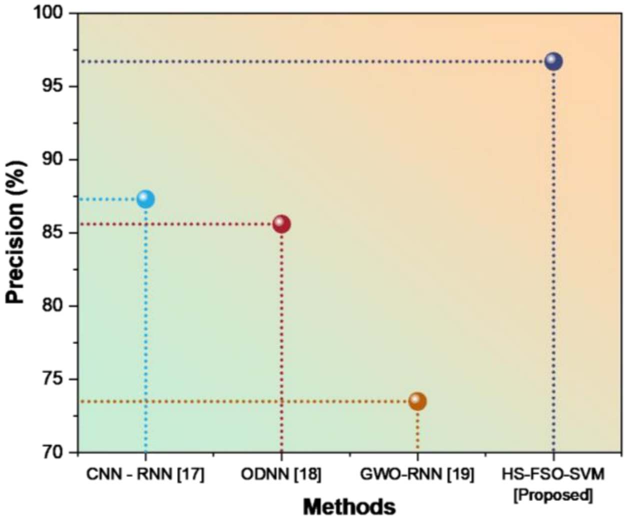

Figure 4: Comparison of precision between suggested and existing methods

4.3 Recall

Recall is a crucial performance evaluation. In other words, recall can be defined as sensitivity and true positive rate. The whole amount of projected positive outcomes and expected favorable outcomes is divided by the total amount of correctly anticipated positive results and wrongly estimated unpleasant consequences to get the recall score. Recall is the key to identify the true positives and minimizing false negatives in LC detection using CT imaging and precise cancer diagnosis. The contrast of recall among traditional and suggested method is shown in Figure 5 and Table 3. Our proposed approach has a recall value of 93.5%, it proved that the recommended model performs superior than other methods as shown in Equation (13).

$$Recall = \frac{TP}{TP+False\ Negative\ (FN)} \times 100 \qquad (13)$$

Table 3: Comparison of recall

Methods	Recall (%)
CNN - RNN [17]	82.3
ODNN [18]	91.3
GWO-RNN [19]	82.7
HS-FSO-SVM [Proposed]	93.5

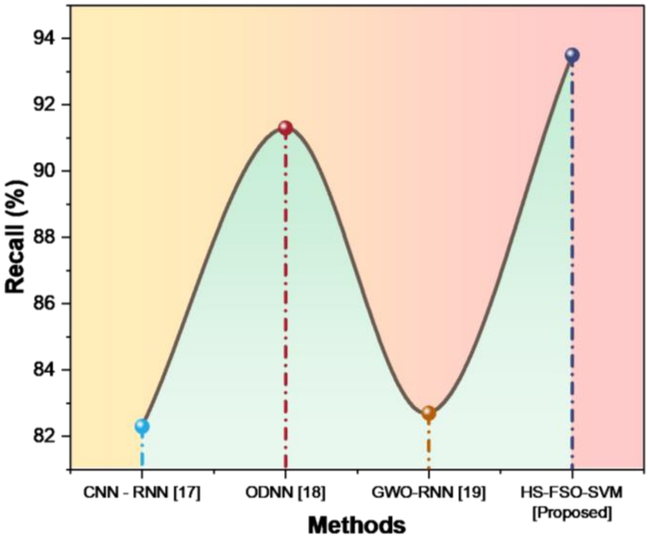

Figure 5: Comparison of recall among suggested and existing methods

4.4 F1- measure

The F1-measure is quite useful for finishing classification tasks. A higher F1 value suggests that the model fairly balances accuracy and recall. The accuracy and total recall components of the F1-score are measured on a scale from 0 to 1, with 1 representing the best outcome. An increased F1-score suggests that CT Imaging offers a reliable and well-rounded approach to lung cancer diagnosis. A comparison of the f1-score between existing and proposed method is shown in Figure 6. As a result, the value for HSFSO-SVM is higher by 90.9 % when compared to other conventional methods, it proved that our proposed method performs superior than other techniques. Figure 6 and Table 4 represent the outcome of the F1-measure.

Table 4: Comparison of F1-measure

Methods	F1-Measure (%)
CNN - RNN [17]	74.2
ODNN [18]	81.9
GWO-RNN [19]	85.7
HS-FSO-SVM [Proposed]	90.9

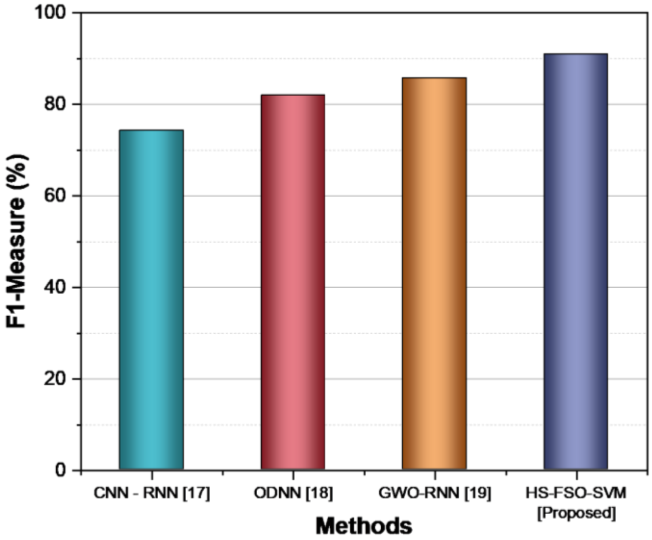

Figure 6: Comparison of recall among suggested and existing methods

5. Discussion

Cancer is a disease that develops in humans as a consequence of abnormal growth of tissue in the body. Cancer is one of the most dangerous diseases that can affect a person's health and poses a threat to society. Since benign and malignant lung tumors share a surprising amount of anatomical similarities, it is very challenging to accurately diagnose LC at an early stage is the major drawback in existing methods such as GWO-RNN [19], ODNN [18] and CNN-RNN [17]. Existing method does not provide images in better quality. The principal cause of LC is the use of tobacco products and the incidence of the disease has either reached its highest point or it is still growing in a number of different nations. In this study, our suggested model is compared with other traditional approaches with regard to accuracy, precision, recall and F1-measure. The values that our methods produce for accuracy, precision, recall and F1-measure are 97%, 96.7%, 93.5% and 90.9% respectively. The values presented above demonstrate that the approach that is recommended performs better than other approaches. Our methodology is very well suited for the automated prediction of LC using CT scans.

6. Conclusion

The disease known as LC, which is described by unchecked cell division, is more curative than other respiratory disorders. One important and obvious way to increase patient survival rates is to detect LC in its early stages. This study employed a HS-FSO-SVM to autonomously identify lung tumors from CT data. By combining the steps of data collection, preprocessing, feature extraction and classification, the proposed method generates results that are accurate and effective. A hybrid optimization strategy combining the Firefly and Sparrow techniques has been created to optimize the SVM parameters and improve the efficiency of the classifier. In summary, our approach outperforms the rival systems in terms of classification accuracy.

The use of intricate optimization techniques can make it challenging to understand the model's conclusions, which could limit its comprehension in clinical settings where patients' and doctors' acceptance along with trust are earned. In the future, the method might be modified for near-real-time or real-time diagnosis, providing a useful tool for early identification of LC and improving patient results through more prompt therapies.

References

1. Toğaçar, Mesut, Burhan Ergen, and Zafer Cömert. "Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks." Biocybernetics and Biomedical Engineering 40, no. 1 (2020): 23-39.https://doi.org/10.1016/j.bbe.2019.11.004
2. Khan, M. Attique, Sadia Rubab, Asifa Kashif, Muhammad Imran Sharif, Nazeer Muhammad, Jamal Hussain Shah, Yu-Dong Zhang, and Suresh Chandra Satapathy. "Lungs cancer classification from CT images: An integrated design of contrast based classical features fusion and selection." Pattern Recognition Letters 129 (2020): 77-85.https://doi.org/10.1016/j.patrec.2019.11.014
3. Riquelme, Diego, and Moulay A. Akhloufi. "Deep learning for lung cancer nodules detection and classification in CT scans." Ai 1, no. 1 (2020): 28-67. https://doi.org/10.1007/s11042-019-08394-3
4. Thanoon, Mohammad A., Mohd Asyraf Zulkifley, Muhammad Ammirrul Atiqi Mohd Zainuri, and Siti Raihanah Abdani. "A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images." Diagnostics 13, no. 16 (2023): 2617.https://doi.org/10.3390/diagnostics13162617
5. Chen, Bihong T., Zikuan Chen, Ningrong Ye, Isa Mambetsariev, Jeremy Fricke, Ebenezer Daniel, George Wang et al. "Differentiating peripherally-located small cell lung cancer from non-small cell lung cancer using a CT radiomic approach." Frontiers in Oncology 10 (2020): 593.https://doi.org/10.3389/fonc.2020.00593
6. Perez, Gustavo, and Pablo Arbelaez. "Automated lung cancer diagnosis using three-dimensional convolutional neural networks." Medical & Biological Engineering & Computing 58 (2020): 1803-1815.https://doi.org/10.1007/s11517-020-02197-7
7. Shakeel, P. Mohamed, M. A. Burhanuddin, and Mohammad Ishak Desa. "Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier." Neural Computing and Applications (2022): 1-14.https://doi.org/10.1007/s00521-020-04842-6
8. Sang, Jun, Mohammad S. Alam, and Hong Xiang. "Automated detection and classification for early stage lung cancer on CT images using deep learning." In Pattern recognition and tracking XXX, vol. 10995, pp. 200-207. SPIE, 2019.https://doi.org/10.1117/12.2520333
9. Masood, Anum, Po Yang, Bin Sheng, Huating Li, Ping Li, Jing Qin, Vitaveska Lanfranchi, Jinman Kim, and David Dagan Feng. "Cloud-based automated clinical decision support system for detection and diagnosis of lung cancer in chest CT." IEEE journal of translational engineering in health and medicine 8 (2019): 1-13.https://doi.org/10.1109/JTEHM.2019.2955458
10. Nadkarni, Nidhi S., and Sangam Borkar. "Detection of lung cancer in CT images using image processing." In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 863-866. IEEE, 2019.https://doi.org/10.1109/ICOEI.2019.8862577
11. Abdul, Wadood. "An automatic lung cancer detection and classification (ALCDC) system using convolutional neural network." In 2020 13th International Conference on Developments in eSystems Engineering (DeSE), pp. 443-446. IEEE,

2020.https://doi.org/10.1109/DeSE51703.2020.9450778

12. Kasinathan, Gopi, Selvakumar Jayakumar, Amir H. Gandomi, Manikandan Ramachandran, Simon James Fong, and Rizwan Patan. "Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier." Expert Systems with Applications 134 (2019): 112-119.https://doi.org/10.1016/j.eswa.2019.05.041
13. Yu, Kun-Hsing, Tsung-Lu Michael Lee, Ming-Hsuan Yen, S. C. Kou, Bruce Rosen, Jung-Hsien Chiang, and Isaac S. Kohane. "Reproducible machine learning methods for lung cancer detection using computed tomography images: Algorithm development and validation." Journal of Medical Internet research 22, no. 8 (2020): e16709.https://doi.org/10.2196/16709
14. Vijh, Surbhi, Deepak Gaur, and Sushil Kumar. "An intelligent lung tumor diagnosis system using whale optimization algorithm and support vector machine." International Journal of System Assurance Engineering and Management 11 (2020): 374-384.https://doi.org/10.1007/s13198-019-00866-x
15. Shariaty, F., S. Hosseinlou, and V. Yu Rud’. "Automatic lung segmentation method in computed tomography scans." In Journal of Physics: Conference Series, vol. 1236, no. 1, p. 012028. IOP Publishing, 2019.https://doi.org/10.1088/1742-6596/1236/1/012028
16. Shakeel, P. Mohamed, Mohd Aboobaider Burhanuddin, and Mohamad Ishak Desa. "Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks." Measurement 145 (2019): 702-712.https://doi.org/10.1016/j.measurement.2019.05.027
17. Moitra, Dipanjan, and Rakesh Kr Mandal. "Automated AJCC staging of non-small cell lung cancer (NSCLC) using deep convolutional neural network (CNN) and recurrent neural network (RNN)." Health information science and systems 7 (2019): 1-12.https://doi.org/10.1007%2Fs13755-019-0077-1
18. Lakshmanaprabu, S. K., Sachi Nandan Mohanty, K. Shankar, N. Arunkumar, and Gustavo Ramirez. "Optimal deep learning model for classification of lung cancer on CT images." Future Generation Computer Systems 92 (2019): 374-382.https://doi.org/10.1016/j.future.2018.10.009
19. Gunjan, Vinit Kumar, Ninni Singh, Fahimudin Shaik, and Sudipta Roy. "Detection of lung cancer in CT scans using grey wolf optimization algorithm and recurrent neural network." Health and Technology 12, no. 6 (2022): 1197-1210.https://doi.org/10.1007/s12553-022-00700-8
20. Sajja, Tulasi, Retz Devarapalli, and Hemanth Kalluri. "Lung Cancer Detection Based on CT Scan Images by Using Deep Transfer Learning." Traitement du Signal 36, no. 4 (2019): 339-344.https://doi.org/10.18280/ts.360406