Turkish Journal of Computer and Mathematics Education Vol.12 No.12 (2021), 3376-3389

Research Article

Prediction of Lung Cancer Using FSSO Optimization and Deep Learning based CNN Algorithm

[1]Dr.L.Malathi, [2]M.Priyanka, [3]Dr.N.Mohanapriya* and [4]P. Brindha

[1]Associate Professor, Department of Computer Science and Engineering, Vivekananda College of Engineering for Women, Elayampalayam, Tiruchengode - 637 205
[2]PG Scholar, Department of Computer Science and Engineering, Vivekananda College of Engineering for Women, Elayampalayam, Tiruchengode - 637 205.
[3]Assistant Professor, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Elayampalayam, Tiruchengode - 637 205.
[4]Assistant Professor, Department of Computer Science and Engineering,Vivekanandha College of Engineering for Women, Elayampalayam, Tiruchengode - 637 205.

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 23 May 2021

Abstract

Lung cancer is the uncontrollable growth of cells within the lungs. Commonly cancer occurs in both male and female. This causes a significant breathing problem in both inhaling and exhales a part of the chest. Earlier identification of cancer is the best way for reducing mortality rate. And also identification of lung cancer is the challenging task for every researchers and doctors. In this study, we proposed the deep learning based convolutional neural network (CNN) for classification and prediction of lung cancer in earlier stage. By improving the classification accuracy, we using different significant methods as averaging histogram equalization (AVHEQ) for pre-processing to enhance and remove the noise from the dataset image , social spider optimization algorithm (FSSO) for segment the cancer affected region and extraction technique for extracted the affected region from entire image. Finally the deep learning classifier model to predict the lung cancer in earlier stage. In this model, LIDC-IDRI dataset is used to analyse the results. And also we compare the different machine learning classifier model with deep learning method. The performance of the proposed model is calculated by using different parametric measures. Finally the deep learning model achieve the better classification accuracy than other compared machine learning classifier.

Keywords: social spider optimization algorithm, LIDC-IDRI dataset, convolutional neural network, Lung cancer and optimization.

1 Introduction

Cancer is a disease characterized by abnormal cell proliferation with the ability to assault and spread to other sections of the human body. Lung cancer (malignant) and non-malignant (benign) are minor growths of lung cells. The early diagnosis of malignant nodules in the lungs is important in the complex pathology [1]. Early-stage lung nodules resemble benign nodules and require a diagnosis based on minor morphological changes, location, and clinical biomarkers [2]. The challenge is to calculate the likelihood of malignancy in early lung cancers. With needles [3], however, health professionals mainly use invasive techniques such as biopsy or surgery to distinguish between malignant and benign nodules in the lungs. With such delicate and delicate organs, invasive techniques are risky and increase patient discomfort

It is a life-threatening condition, and precise prognosis is critical for good clinical analysis and therapy advice. Even for professionals with extensive expertise, detecting it correctly remains a difficulty. It is difficult to make a trustworthy early diagnosis, which is done by physicians based on the diagnosis of cancer symptoms. Among the various cancer, LC is the foremost reason for mortality in both mankinds in the present time, with an inspiring figure of around five million deaths every year [4 and 5]. LC is kinds of healthcare application in DNN. Computer pictorial representation (CT) will deliver valuable info once designation respiratory organ sicknesses. The chief goal of this work is to spot cancer nodules within the lungs from a given input image of the lungs and to arrange LC and its harshness.

The CADx system is designed to classify nodules discovered into benign nodules [6 and 8]. Because the likelihood of malignancy is highly tied to geometric size, form, and appearance, CADx can differentiate between benign and malignant pulmonary nodules using useful criteria such as texture, form, and growth rate. Thus, the success of a certain CADx system may be judged in terms of diagnosis accuracy, speed, and automation level [9 and 10].

In recent years, neural networks, rebranded as “deep learning,” have begun to outperform classical AI in every crucial task: speech recognition, picture characterization, and creating meaningful, readable sentences. Deep learning not only speeds up the important work, but it also increases the computer's precision and the performance of CT image recognition and classification.

2 Literature Review

Salsabil Amin et.al [11] introduced a technique for detecting LC using computed tomography. Pretreatment was carried out using stretching and reverse growth. An arrangement of morphological operations, threshold and area culture was used to separate the lung image. A rule-based classifier was used that condensed

the number of false positives. The scheme provided an accuracy of 70.53%. In the upcoming, CNN or SVM will be proposed for classification.

P.B.Sangamithraa et.al [12] offered a scheme for classifying CT images of lung cancer as malignant or non-malignant. The image noise was removed using a median filter and a Wiener filter. Fuzzy k-means were performed for segmentation and the outcome was refined by grouping the EK means. The functions were requested and transmitted to the Back Propagation Network (BPN). The scheme provided an accuracy of 90.87%. SVM is recommended for greater accuracy.

Manasee Kurkure et.al [13] has projected the method of Nave Bays and a genetic algorithm (GA) for identifying lung cancer nodes. The histogram was originally applied to the image. Strong edges have been identified by the Smart Edge sensor. Naval bases were used for taxonomy and the taxonomy was improved using a GA. An accuracy of 80%. The downside was that GA is slow and needs difficult calculations.

Emre Dandil et.al [14] has created a CAD ideal to classify lung cancer types. The similarity of the histogram was used to increase the contrast by changing the intensity of the image. The lobes of the lobes were removed by morphological surgery and the remaining portions on the edge were detached using a double cutoff. A self-organized map was used to divide the lungs. The GLCM was used for feature extraction and was used for pivotal study to further reduce the feature. Artificial neural networks (ANNs) are implemented for classification. The CAD scheme showed 90.63% accuracy.

3 Proposed methodology

In this section we briefly describe about the proposed methodology, which is showed in figure 1. The proposed method has several phases such as pre-processing, image segmentation, feature extraction and lung cancer classifcation. The pre-processing phase is performed by using image averaging histogram equalization approach (AVHEQ) for improving quality of image such as contrast enhancement and brightness preservations. The image segmentation phase is performed by using FSSO, spectral features extraction are used to extract the feature from image, and CNN algorithm for used to detecting the lung cancer predictions. we use LIDC-IBRI dataset. Finally, the system is tested for system performance using MATLAB-based simulation results.

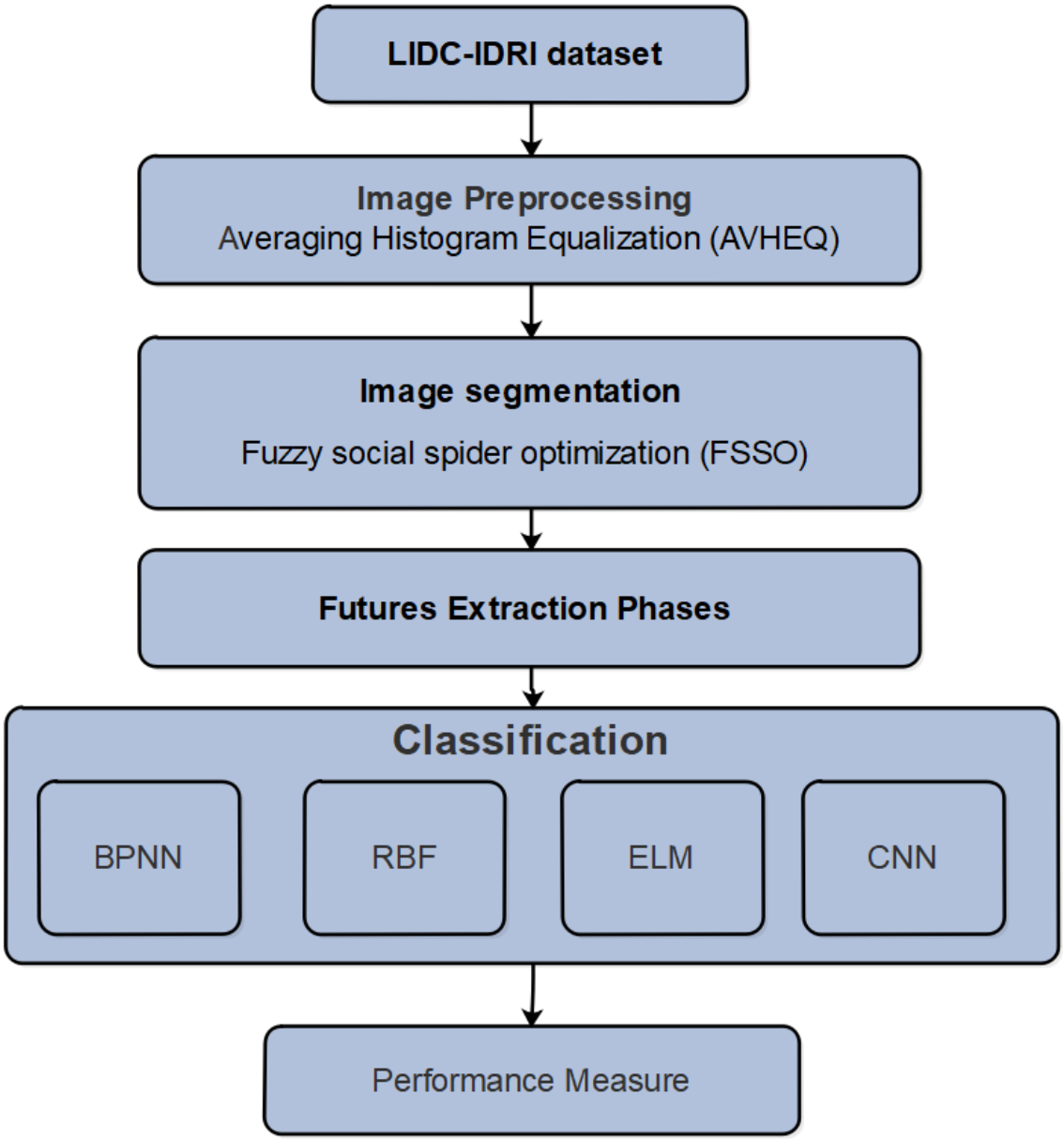

Figure 1: Work flow Diagram of Proposed Methodology

3.1 LIDC-IBRI dataset collection

The Lung Imaging Database Consortium (LIDC-IDRI) image collection includes chest register tomography (CT) demonstrations and screen tests for lung lesions with extended specification lesions. This website is a global resource that enables personal computers to identify, diagnose, and demonstrate (CAD) lung lesions. To collect this information, 10 scientists focused on eight restored imaging institutes to collect 1,018 cases. Each subject records the outcomes of a two-part image representation technique executed by four knowledgeable thoracic radiologists, combining images from four clinical breast CTs and the associated XML archive. In the proposed system, 70 images were trained and 30 images were tested on the LIDC-IBRI data set. Some images of the sample data set are defined in Figure 2.

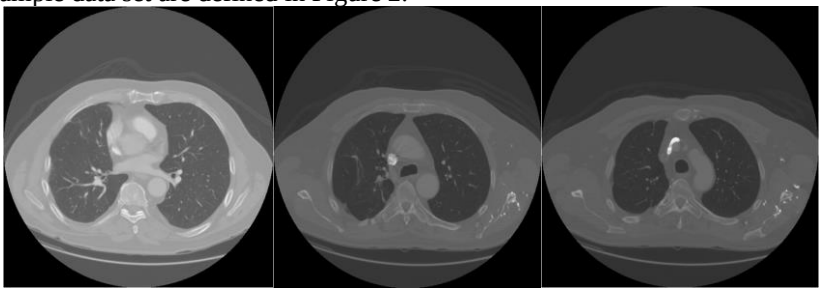

Figure 2: LIDC-IBRI dataset Sample images.

3.2 Pre-processing phase

3.2.1 Conventional histogram equalization

Consider an input color image in (RGB) format of sized $U \times V$ in width-by-height is given. The magnitude of each color channel is bounded within $[0 \quad L-1]$ for typical 8-bit storage, where $L = 2^8 = 256$. The image can be described by

$$\tau(u,v) = \{R(u,v) \quad G(u,v) \quad B(u,v)\} \tag{3}$$

Where (u,v) is the image's pixel coordinate. Most ways convert a color image to a gray one, for example, by extracting and improving the color saturation (HSI) intensity and the intensity of the I-channel. That is

$$\{HSI\} \quad \leftarrow \tau \quad (\{RGB\}) \tag{4}$$

Where $\tau(.)$ is the RGB to HSI transformation. A histogram methods is first constructed for the input image by

$$h(i) = \{n(i)\} \tag{5}$$

Where i is the intensity index and n (i) is the number of pixels having the i^{th} intensity. A cumulative density function is further obtained from

$$c(i) = \frac{1}{N}\sum_{j=0}^{i} n(j), \quad c(L-1) = 1, \quad N = U \times V. \tag{6}$$

The increased gray picture comes from the equalization procedure

$$I_{enh} = I_{\min} + (I_{\max} - I_{\min}) \times c(i), \tag{7}$$

Where $I_{\min}$ and $I_{\max}$ are the minimum and maximum intensities, and usually $I_{\min} = 0$ and $I_{\max} = L-1.$

3.2.2 Histogram separation

Since the image output is a uniform distribution, its medium intensity is increased using standard histogram equalization

$$I_m = \sum_{i=0}^{L-1} (i \times p_o(i)), \quad i = 0,..., \quad L-1, \tag{8}$$

Where $P_0(i) = h_0(i)/N = 1/L$ and $h_0 = (i)$ represents the histogram of the image $I_{enh.}$ it is deviated from the input image and may be regards as undesirable. Efforts were made in the development of bi-histogram based approaches to remedy this drawback by citing the brightness preserving concept.

The input image is separated into two sub-images, and conventional histogram equalization is then performed independently on each sub-image. The distinction is that, whereas the latter employed the median value, the first work used medium light as the separation level. As pixels are determined by the picture content under and above the threshold, the mean output brightness is not exactly the same as the average input brightness.

After the bi-histogram equalization work, methods were developed which restrict the peaks of the histogram with the mean and median values of each subsection. The image is divided into the low and high regions, for example

$$I_{lo} = \{I(i) \mid 0 < i \le I_m\}, \tag{9}$$

$$I_{hi} = \{I(j) \mid I_{m+1} < j \le L-1\}. \tag{10}$$

The output image is obtained from the aggregation of two enhanced sub-images as

$$I_{enh} = I_{enh,lo} \cup I_{enh,hi,} \tag{11}$$

Where the enhanced sub-images $I_{enh,lo}$ and $I_{enh,hi}$ are separately generated by performing the conventional histogram equalization discussed in the last section on the two sub-images I_{lo} and I_{hi}. The target distributions of these enhanced sub-images are the clipped sub-histograms.

3.2.3 Histogram clipping

Conventional historograms have over-improved effects, which often produce odd objects of sight. Artifacts have been noticed when intensity changes during equalization are too extreme. It is usually observed at levels of intensity where the input picture histogram peaks. The picture input is separated into four quadrants using the medium to divide the average values of each of the four sub-images into two sub-images. There has

been a further change in division of the input picture with the mean intensity. In contrast to the other, a minimum histogram element and mean and median value for the sub-images were used to select the clipping limit. That's what it is.

$$P_{lo}(i) = \min\{h(i), \quad \mu(i), \quad m(i)\}, \quad 0 < i \le I_m, \tag{12}$$

$$P_{hi}(j) = \min\{h(j), \quad \mu(j), \quad m(j)\}, \quad I_{m+1} < j \le L-1, \tag{13}$$

Where i denotes the sub image range $o < i \le I_m$ and $\mu(j)$ and $m(j)$ are defined accordingly for $I_{m+1} < j \le L-1$. The target histograms to be used in the equalization then become

$$h_{lo}(i) = \begin{cases} h_{lo}(i), & h_{lo}(i) \le p_{lo}(i) \\ p_{lo}(i), & otherwise. \end{cases} \tag{14}$$

$$h_{hi}(j) = \begin{cases} h_{hi}(j), & h_{hi}(j) \le p_{hi}(j) \\ p_{hi}(j), & otherwise. \end{cases} \tag{15}$$

Procedures are now available in numerous problem domains as modifications and variants originate from the bi-histogram concept. However, an accurate correspondence of the mean brightness of the input and output image and at the same time the picture contrast improves and the viewing of objects does not occur.

3.2.4 Histogram smoothing

Alternatives from a different viewpoint are still feasible, however ways based on the separation of the sub-image had become popular in improving contrast and ensuring that brightness is maintained. The greatest histogram can be created by smoothering the picture histogram in order to decrease artifacts created by equalization to the uniform histogram, as usually practiced in sub-image operations.

The image histogram is calculated by calculating the average values for each intensity level in the histogram in this technique. The average is performed by the intensity of the divisor. The process is describable as

$$h(i) \leftarrow \frac{h(i)}{i \pm w(i)}, \tag{16}$$

Where the range of division is a function of the intensity and is determined from

$$w(i) = \begin{cases} 2(i+1)-1, & 0 < i \le L/2 \\ 2(L-i+1)-1, & L/2 < i \le L-1. \end{cases} \tag{17}$$

In the mid intensity region, the resulting smooth histogram flatters and replicates the input histogram at both ends of the intensity range. In the equalization procedure, the adjusted histogram is subsequently used as the target distribution. On the other hand, because the histogram is only adjusted moderately at the two ends, the result image contains a little amount of artifacts, where the highest intensities are visible. Figure 3 shows the improved and original photographs.

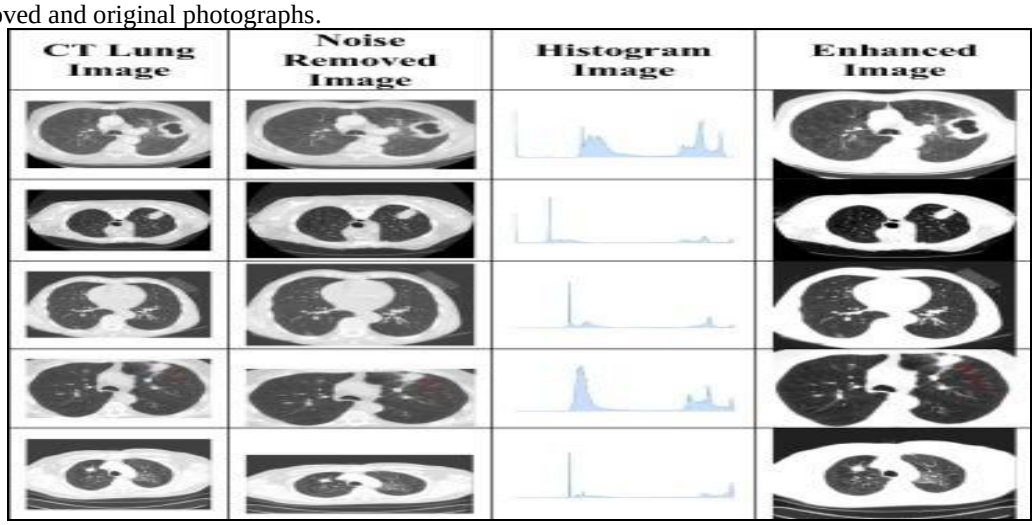

Figure 3: Sample pre-processed images

4 Segmentation

For the medical picture segmentation process, a number of technologies were created. Thresholding, regional growth and edge detection are the primary fragmentation approaches. The advanced techniques are model-based, graph-based and partial differential equations (PDE). A gray medical picture usually offers information about intensity and gradient. Segmentation technology can therefore be classified into intensity and gradient methods. The thresholding, region growth and graphics algorithms are intensity-based. Clustering, modeling, active contour, PDE, and graph-based approaches are gradient based approaches.

The threshold and regional development are a fundamental component to the intensity-based approaches. Thresholding is a basic, but highly helpful method for obtaining pixel intensity within the desired range. Regional development is also a small and effective method for achieving pixel intensity comparable to the initial seed locations. For intensity-based segmentation, the graph based technique can be utilized. The conventional graph or grab is a well-known approach for the segmentation of images based on information about intensity. The most significant way of distinguishing the two pixel intensity model is through the classic graph cut which is utilized as a minimum cutting tool. The minimum is cut to object and background in the graph partitions of the image. Several strategies for the use of intensity information in medical image segmentation have been proposed. Clustering methods are pixel information statistical approaches used to determine whether or not a pixel pertained to a cluster. Examples of clustering approaches are the k-means, FCM and Gussian Mixture Models (GMM). In the segmentation of medical gray photos the FCM is commonly used.

Beforehand, intensity-based approaches and gradient-based methods can be classified. The two forms of information, such as intensity and gradient information, are often lacking in precise segmentation results using these methods. Intensity-based detection may not identify areas if an imaging has a slim difference in intensity across regions, but if the inhomogeneity of the local pixels in the regions is excessively great the intensity-based sensing frequently reduces the regions. The dividing scheme that is done with the graph-screening method and divided in Figure 4.

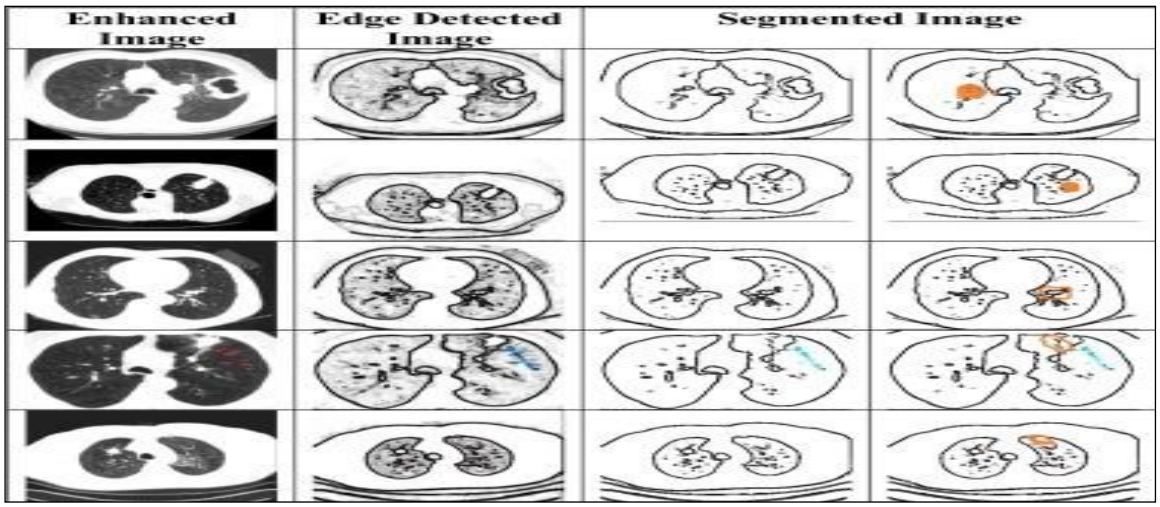

Figure 4: Sample segmented images

If a picture has a little difference of gradient between areas, the detection based on the gradients may shrink down or detect nothing, however if the difference of gradient is too large with other structures, the result of the segmentation can be stuck. Several strategies were used to resolve efficiently the segmentation problem. Hence, clustering based segmentation approach is used in this work for extract the affected region of the lung cancer image.

4.1 Social Spider Optimization (SSO)

The population-based strategies to solve the problem of optimization and the co-operative social spider behaviour. The SSO algorithm assumes the whole of the shared web search. The fitness value of specific solutions is determined by every social spider. Both men and women are the two types of agents to find the optimal answers. Every SSO person is led by a number of evolutionary operators, followed by various cooperative behaviours. In the colony, the cooperative conduct of SSO, depending on gender, is often expected. The social spider's good characteristic is that it is very much based on women. The SSO begins by specifying the amount of women and men in the search area that are divided as individuals..

The female spiders $'N_f'$ are choosing randomly that range between 65-90% from the total population of $'N'$. So, $'N_f'$ is compute as follows,

$$N_f = floor[(0.9 - rand.0.25).N] \tag{18}$$

From the above equation, the $rand$ is a random number that ranges between $[0,1]$, the total population is denoted by $'N'$. The floor $'(.)'$ denotes maps a real value to an integer value. The male spiders $'N_m'$ is compute as the balance between $'N_f'$ and $'N'$. It is considered as follows,

$$N_m = N - N_f \tag{19}$$

4.1.1 Fuzzy SSO for Fuzzy Clustering

The standard SSO clustering algorithms are known as fuzzy SSO algorithms in combination with a theory termed clustering problems. The following steps are developed for fluid clustering issues as an acceptable algorithm.

Step 1: Initialization: In this step, each individual spider is randomly selected from the given datasets.

$$S = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nc} \end{bmatrix} \tag{20}$$

Where $'\mu'$ is a membership value of the $'S'$. The spider swarm is well-defined as follows,

$$S_{P\times T} = [S_{M\times T} \bigcup S_{F\times T}] \tag{21}$$

Where, $'P'$ is denoted the size of population. The $'M'$ and $'F'$ are the male and female spiders respectively. The spider is taking dimensionality $T = [D \times K]$.

Step 2: Objective function evaluation each observation of the datasets

Objective function associated with each spider is computed as follows

$$J_m(S_i) = \sum_{j=1}^{K}\sum_{i=1}^{n} \mu_{ij}^m d_{ij} \tag{22}$$

Step 3: Weight assignment to each individual spider

Weight in each spider that is related with it, based on the following fitness functions, is estimated,

$$W(i) = \frac{worst_f - J_m(S_i)}{worst_f - best_f} \tag{23}$$

Where, $J_m(S_i)$ is the fitness value which is twisted by calculation of spider positions $'S_i'$ with consider to the objective functions $J_m(.)$.

The lowest values is considered as the best spider that denoted by $best_f$ as follows,

$$best_f = \min_{\{k\in 1,2,....,N\}} (J(S_k) \tag{24}$$

The highest fitness values is considered as the worst spider that denoted by $worst_f$,

$$best_f = \min_{\{k\in 1,2,....,N\}} (J(S_k) \tag{25}$$

Step 4: Male population fragmentations

The two sorts of male population are the dominant and non-dominant. For separation of the male spiders, the following formula.

$$S_M(i) = \begin{cases} S_{DM}(i) & if W(i) \geq median(W) \\ S_{NM}(i) & otherwise \end{cases} \tag{26}$$

$'S_{DM}'$ - denotes the dominant male spiders which have larger fitness value over the non-dominant that undertake a process of mating. $'S_{NM}'$ - is denote non-dominant male spiders.

Step 5: Positions updates

(i) Female Positions updates: On the community web, the spider movement of women depends on the internal position that is affected by the two elements, including the cycle of reproduction and curiosity. The position of the female spider is updated,

$$S_F^{k+1}(i)=\begin{cases} S_F^k(i)+\alpha\vartheta_{i,l}(S^l-S_F^k(i))+\beta\vartheta_{i,g}(S^g-S_F^k(i))+.....+\oplus\ (rand-0.5)\ \ if\ rand<PF \\ S_F^k(i)-\alpha\vartheta_{i,l}(S^l-S_F^k(i))-\beta\vartheta_{i,g}(S^g-S_F^k(i))+.....+\oplus\ (rand-0.5) \quad otherwise \end{cases} \tag{27}$$

From the above equation, $'\alpha','\beta',rand$ and $'\oplus'$ randomly selected from the uniform distribution between the spiders $\left[0,\frac{P}{N_F+N_M}\right]$ Here, $\vartheta_{i,l}=W_l e^{-\partial(S(i),S^l)^2}$ are the vibrations established from the neighboring or limited spider $'S^l'$. $'S^l'$ must having maximum weight then $'S(i)'$, $\vartheta_{i,g}=W_g e^{-\partial(S(i),S^g)^2}$ are the signals established by the i[th] spider from the spider having supreme weight or global S^g among all. The process of vibration is important for spiders to share information.

(ii) Updating the positions of Dominant Male:

The goal of the dominant male is to move to the female spiders in order to make reproduction process. The following formula is denoted positions updates,

$$S_{DM}^{k+1}(i)=S_{DM}^k(i)+\alpha\vartheta_{i,F}(S_F-S_{DM}^k(i))+\theta(rand-0.5) \tag{28}$$

Here, $\vartheta_{i,l}=W_F e^{-\partial(S(i),S_F)^2}$ is the nearby female spider received the vibration. The vibration process can integrate population diversity.

(iii) Updating the positions of Non-Dominant Male:

The main objective of non-dominant male is collected themselves in the center (radius) of the dominant male spider to take the due benefit of the left over.

$$S_{NM}^{k+1}(i)=S_{NM}^k(i)+\alpha(\overrightarrow{X_w}-S_{NM}^k(i)) \tag{29}$$

$'\overrightarrow{X_w}'$ is the weighted mean attained from the male population weight.

Step 6: Violation check on the communal web

After the job updates are completed, each spider should be placed on the common web. Each spider shall be checked and de-limited if required.

Step 7: Mating of the spiders

Radius of each dominant male determined as follows,

$$Radius=\frac{R_{\max}-R_{\min}}{2D} \tag{30}$$

Where, $'D'$ is a dimensionality of the datasets, $'R_{\min}'$ and $'R_{\max}'$ are the minimum and maximum values of the datasets respectively. This ensures that the data sets provide a fresh solution. In the radius, a number of females are generally identified. When women are not found, the matching process is invalid. If two dominant men are present within the radius, the male with the highest weight can generate the latest bran. The newest offspring consists of a probability approach for the roulette wheel. The spider with the highest fitness value and a weak solution will be exchanged by an innovative baby spider. The next new group will be the new Spiders.

Step 8: Results

As an optimal cluster, the least value of the goal function. Convergence of algorithms can either be accomplished by an iteration which reaches zero for a certain number of generations.

5 Feature extraction phase

The segmented picture is translated into the characterization phase, which produces the different spectral properties such as mean, third skewness, standard differential and fourth movement kurtosis because the lung cancer-related features are efficiently detected. Table 1 shows the predictable spectral characteristics.

Table 1: Details of features

Features	Formula
Mean	$\mu=1/N\sum_{i=1}^{N}S_i$
Standard deviations	$\sigma=\sqrt{1/N\sum_{i=1}^{N}(S_i-\mu)^2}$
Third movement skewness	$sk=\sqrt{1/N\times\sigma^3\times\sum_{i=1}^{N}(S_i-\mu^3)^{1/3}}$

Fourth moment kurtosis	$ku = \sqrt{1/N \times \sigma^4 \times \sum_{i=1}^{N} (S_i - \mu^4)^{1/4}}$

6 Classification

The final phase of this work is the detection of lung cancer by using well-known deep neural network algorithm such as CNN is applied to solve any real-world problem with high efficient manner.

6.1 Convolutional Neural Networks (CNNs)

A CNN is the DNN class and architecture that is most frequently used. Their common weight and invariance characteristics also mean shift invariant or invariant space artificial neural networks (SIANN). They have image and video recognition applications, recommendation systems, picture classification, medical image analysis, natural language processing, brain computer interfaces and time series.

Multilayer perceptrons are regularized versions of CNNs. The meaning of multilayer perceptrons is usually full networked, which means every neuron in one layer is connected to the next layer of all neurons. These networks' "full connectivity" allows them to overfit data. Typical regularization methods include adding some kind of weight measurement to the loss function. CNNs have a different regularization approach: they exploit the hierarchical pattern of data and use smaller and simpler patterns to build more complex patterns. CNNs are thus on the lower extreme in terms of connectivity and complexity.

Biological processes inspired CNN because the pattern of neuronal connectivity was similar to the visual cortes of animals. In the narrow region of the visual field known as the receptive area, individual cortical neurons respond to stimuli only. The receptive fields of the various neurons partially overlap to cover the entire field of vision. In comparison to other image classification algorithms, CNNs use relatively little pre-processing. This means that the network learns that filters have been manufactured by hand in traditional algorithms. This independence from prior knowledge and human effort is a major advantage in functional design.

6.2 Architecture of CNN

An input layer, cached layers, and output layer are part of a CNN. In each FFNN, any middle layer is called overshadowed, because the activation function and final convolution mask their inputs and outputs. In a CNN, the hidden layers contain convolution layers. This typically includes a multiplication or other dot product layer and is usually activated by ReLU. Followed by other convolution layers like layers in pooling, fully connected layers and layers of normalization. The following subsection discussed about the layers of CNN.

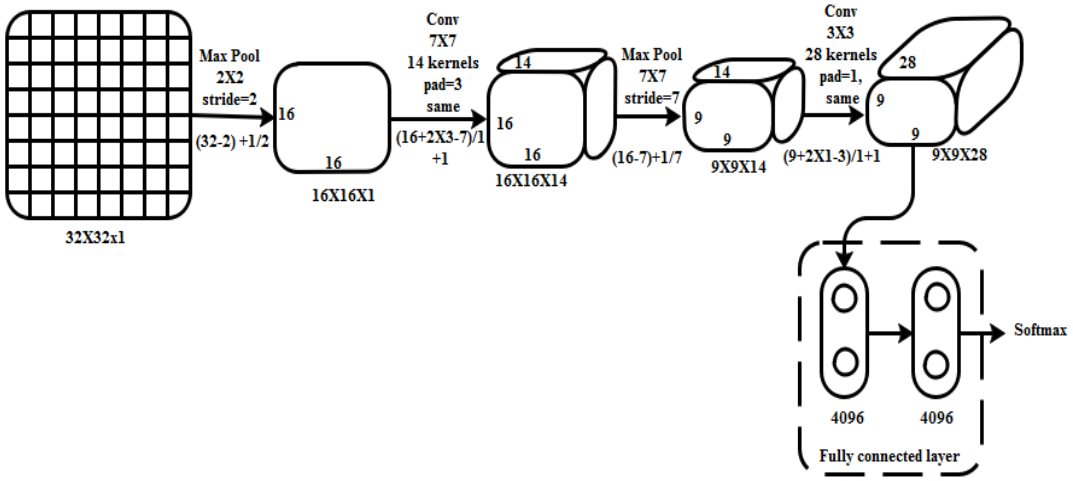

Figure 5: Design of deep CNN networks.

6.2.1 Convolutional layers

In a CNN, the input is a tensor any format. A convolutional layer within a neural network should have the following attributes:

6.2.2 Pooling layers

In order to optimize underlying computations the revolutionary networks may have local or global pooling layers. Power layers minimize the data size by joining neuron cluster outputs in one layer in the next layer to one neuron.

6.2.3 Fully connected layers

Each neuron in one layer is connected to each neuron in a different layer in fully connected layers. It is like a typical neural network of multi-layer perceptrons (MLP). A fully linked layer leads to the flattened matrix to identify pictures.

6.2.4 Receptive field

Each neuron receives information from some places in the preceding layer in neural networks. Each neuron receives information from each neuron of the previous layer in a completely linked layer. Only a restricted portion of the last layer termed the receptive field of the neuron receives input in a convolutionary layer.

6.2.5 Weights

The result value is calculated by using a special function for input values from the receptive area of the preceding layer by each neuron in a neural network. A weight and distortion vector determines the function applied to the input data (typically real numbers). Learning involves adapting these weights and biases iteratively.

7 Results and discussion

The performance of proposed CNN prediction algorithm is discussed and analyzed by using well-known dataset such as lung cancer dataset. Also, strength of proposed prediction algorithm is compared with some well-known benchmark algorithm such as BPNN, RBF and ELM. The performance is analyzed by using various performance measures such as accuracy, specificity, precision and recall.

7.1 Performance measures

The performance comparison is very significant task in any machine learning algorithms. Hence, in this research work, the performance is analyzed by using various performance measures such as accuracy, specificity, precision and recall.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100 \tag{31}$$

$$Specificity = \frac{TN}{TN + FN} * 100 \tag{32}$$

$$\Pr ecisions = \frac{TP}{TP + FP} \tag{33}$$

$$\mathrm{Re}\, call = \frac{TP}{TP + FN} \tag{34}$$

Table 2: Comparative analysis of Accuracy

Methods	Image1	Image 2	Image 3	Image 4	Image 5
BPNN	89.34	90.36	91.83	90.37	91.58
RBF	93.29	93.08	93.53	92.82	94.83
ELM	95.82	95.84	97.36	95.72	97.72
CNN	97.92	96.83	98.83	97.91	98.35

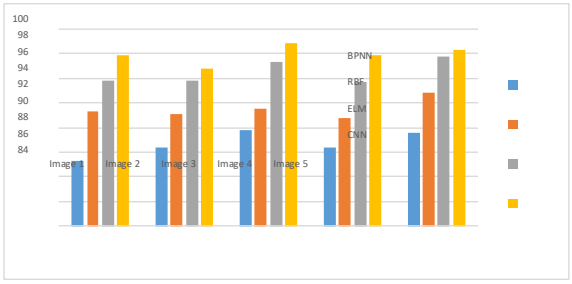

Figure 6: Comparative analysis of accuracy

Table 3: Comparative analysis of specificity

Methods	Image1	Image 2	Image 3	Image 4	Image 5
BPNN	88.93	89.26	90.46	91.37	89.03
RBF	91.32	92.52	93.82	93.86	92.62
ELM	93.03	95.62	94.64	96.37	94.83
CNN	98.27	97.93	96.48	99.04	97.65

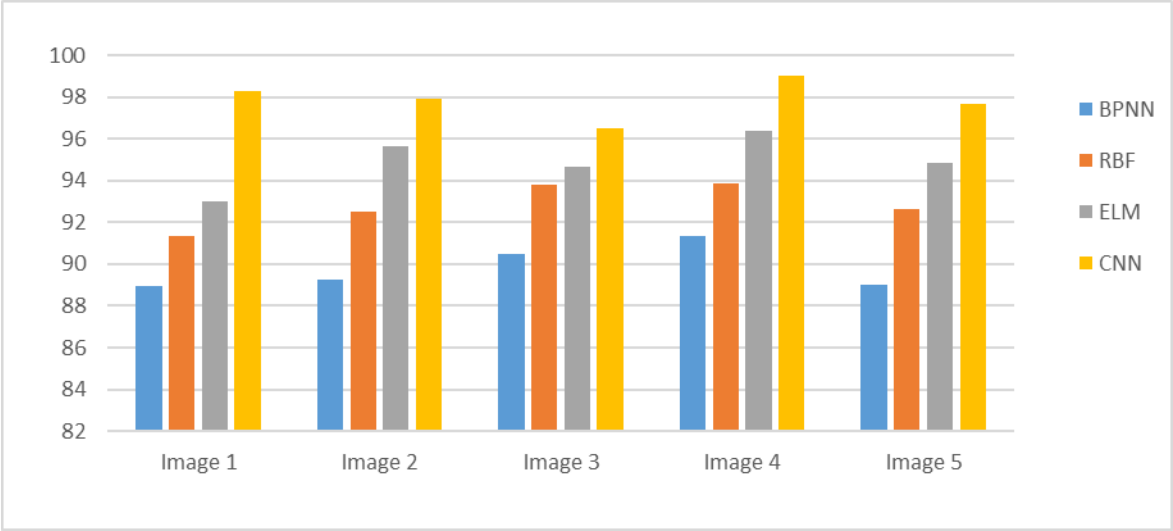

Figure 7: Comparative analysis of Specificity

The proposed work is concentrated lung cancer prediction for analyzing disease. This work consists of four phases such as pre-processing, image segmentation, feature extractions, and disease prediction. The pre-processing step is accomplished by using AVHEQ method for enhancing the quality of the lung images. The spectral features are obtained in features extraction step which is accomplished by using segmented images. Finally, we used CNN algorithm for detecting the lung disease which is affected in lung images.

Table 2 shows the performance comparisons results on accuracy and Figure 6 shows graphical representation. Table 3 shows performance comparisons on specificity and Figure 7 shows the graphical representation of performance comparisons. Table 4 shows the performance comparison on precision and its graphical representation is shows in Figure 8. Whereas, recall value is shows in Table 5 and its graphical

representation is show in figure 9. From the experimental results, the proposed CNN prediction algorithm produced higher detection accuracy when compared with conventional BPNN, RBF and ELM.

Table 4: Comparative analysis of recall

Methods	Image1	Image 2	Image 3	Image 4	Image 5
BPNN	87.46	88.92	87.64	90.31	89.02
RBF	89.36	93.42	90.42	92.04	91.64
ELM	93.72	96.26	92.28	95.27	94.82
CNN	96.81	97.92	96.80	96.83	95.59

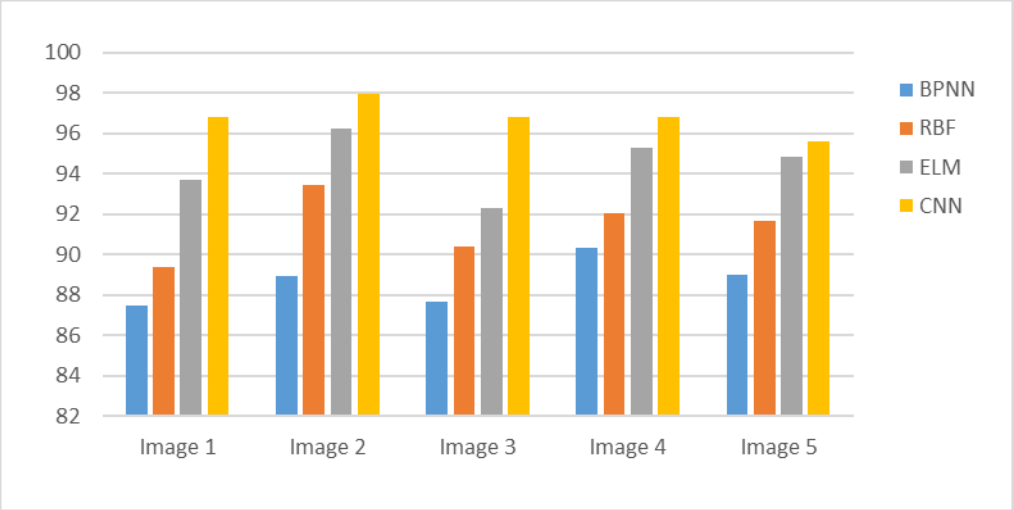

Figure 8: Comparative analysis of recall

Table 5: Comparative analysis of precisions

Methods	Image1	Image 2	Image 3	Image 4	Image 5
BPNN	87.64	88.25	88.93	86.03	89.71
RBF	89.82	89.72	91.04	89.54	92.08
ELM	92.92	94.51	93.82	92.29	95.59
CNN	95.92	96.83	94.27	96.19	98.35

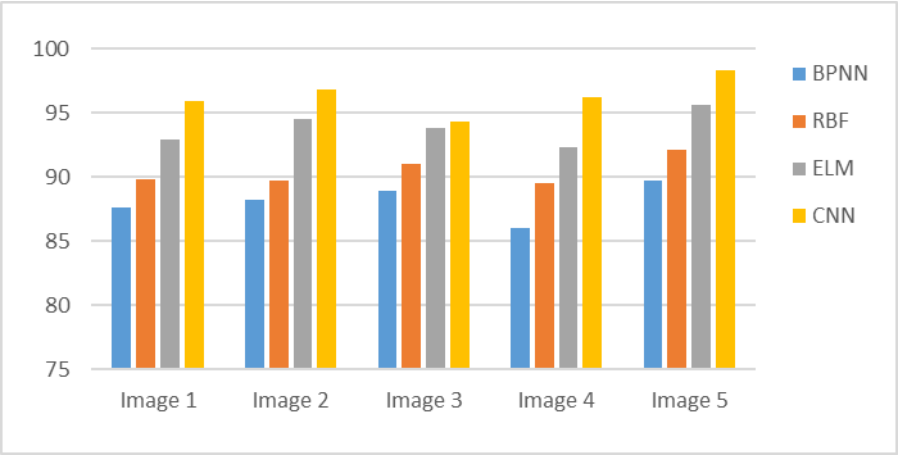

Figure 9: Comparative analysis of precision

8 Conclusions

Lung cancer has been found to be a widespread and most common cause of mortality. Lung cancer is an extremely deadly disease. The purpose of this work is to forecast early detection using classifiers with optimal characteristics. In this research, the CNN prediction algorithm is proposed to predict the lung cancer disease. The lung CT images are obtained from CIA which contains 100 images that is separated into 70 images for training and 30 images for testing. The qualities of the images are enhanced by AVHEQ algorithm by calculating average value of histogram. After improving quality of images, the FSSO segmentation algorithm is used to segment the affected part of lung images. Then, the spectral related features are extracted from the images. The extracted features are predicted either image is affected by disease or not. The proposed CNN algorithm is applied successfully which is obtained higher performance when compared with some conventional algorithms.

Reference

[1].Nair, M.; Sandhu, S.S.; Sharma, A.K. Cancer molecular markers: A guide to cancer detection and management. Semin. Cancer Biol. 2018, 52, 39–55.

[2].Silvestri, G.A.; Tanner, N.T.; Kearney, P.; Vachani, A.; Massion, P.P.; Porter, A.; Springmeyer, S.C.; Fang, K.C.; Midthun, D.; Mazzone, P.J. Assessment of plasma proteomics biomarker's ability to distinguish benign from malignant lung nodules: Results of the PANOPTIC (Pulmonary Nodule Plasma Proteomic Classifier) trial. Chest 2018, 154, 491–500.

[3].Shi, Z.; Zhao, J.; Han, X.; Pei, B.; Ji, G.; Qiang, Y. A new method of detecting pulmonary nodules with PET/CT based on an improved watershed algorithm.

[4].Lee, K.S.; Mayo, J.R.; Mehta, A.C.; Powell, C.A.; Rubin, G.D.; Prokop, C.M.S.; Travis, W.D. Incidental Pulmonary Nodules Detected on CT Images: Fleischner 2017. Radiology 2017, 284, 228–243.

[5].Bjerager, M.; Palshof, T.; Dahl, R.; Vedsted, P.; Olesen, F. Delay in diagnosis of lung cancer in general practice. Br. J. Gen. Pract. 2006, 56, 863–868.

[6].Bogoni, L.; Ko, J.P.; Alpert, J.; Anand, V.; Fantauzzi, J.; Florin, C.H.; Koo, C.W.; Mason, D.; Rom, W.; Shiau, M.; et al. Impact of a computer-aided detection (CAD) system integrated into a picture archiving and communication system (PACS) on reader sensitivity and efficiency for the detection of lung nodules in thoracic CT exams. J. Digit. Imaging 2012, 25, 771–781.

[7].Al Mohammad, B.; Brennan, P.C.; Mello-Thoms, C. A review of lung cancer screening and the role of computer-aided detection. Clin. Radiol. 2017, 72, 433–442.

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

[9].Diederich, S.; Heindel, W.; Beyer, F.; Ludwig, K.; Wormanns, D. Detection of pulmonary nodules at multirow-detector CT: Effectiveness of double reading to improve sensitivity at standard-dose and low-dose chest CT. Eur. Radiol. 2004, 15, 14–22.

[10]. Demir, Ö.; Çamurcu, A.Y. Computer-aided detection of lung nodules using outer surface features. Bio-Med. Mater. Eng. 2015, 26, S1213–S1222.

[11]. Yu, L.; Dou, Q.; Chen, H.; Heng, P.-A.; Qin, J. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. IEEE Trans. Biomed. Eng. 2016, 64, 1558–1567.

[12]. Salsabil Amin El-Regaily, Mohammed Adbel Megeed Salem, Mohamed Hassan Abdel Aziz, Mohamed Ismail Roushdy," Lung nodule segmentation and detection in computed tomography", Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), 2017, ISBN: 978-1-5386-0821-0.

[13]. P.B.Sangamithraa, S.Govindaraju," Lung tumour detection and classification using EK-mean clustering", International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016, ISBN: 978-1-4673-9338-6.

[14]. Manasee Kurkure, Anuradha Thakare, "Lung cancer detection using genetic approach", International Conference on Computing Communication Control and automation (ICCUBEA), 2016, ISBN: 978- 1-5090-3291-4,

[15]. Emre Dandil, Murat Cakrolu, Ziya Eksi, Murat Ozkan, Ozlem Kar Kurt, Arzu Canan," Artificial neural network-based classification system for lung nodules on computed tomography scans", 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), 2014, ISBN: 978-1-4799-5934-1.