Lung Cancer Nodules Detection Using Ideal Features Extraction Technique in CT Images

Vikul J. Pawar(✉), P. Premchand, and I. Govardhanrao

Computer Science and Engineering Department, University College of Engineering Osmania University, Hyderabad, India
vikul.pawar@gmail.com

Abstract. In the present time, worldwide the number of patients related to lung cancer disease getting increase exponentially, accordingly the application of Computer Aided Diagnosis (CAD) system building association with medical science to deliver pertinent solution using Image Processing and Machine Learning Techniques. This paper presenting a model for Lung Cancer nodules detection in CT image by employing proposed work in the progressive phases, the first step is image pre-processing which uses standard LIDC-IDRI images as input images, the pre-processing approach employ the denoising technique to remove the speckle noises from images, then by applying adaptive contrast enhancement (CLAhe) technique the quality of input CT image is improved. The second step works on to segment the Region of Interest (ROI) using LevelSet segmentation algorithm, the third step employed the learnable Feature Extraction technique from suspected (ROI) in CT image, the feature to be extracted from CT images are Texture Features such as Grey-Level Co-occurrence Matrix (GLCM), Grey-Level Run-Length Matrix (GLRM), Local Binary Pattern (LBP), Shape-Based Features such as Perimeter, Area, Irregularity Index, Solidity, Equivalent Diameter, Convex Area, and Statistical Features such as Mean, Median, Mode, Entropy, Moment, Skewness, and Kurtosis. The optimized measurable features are offered as input to the Hybrid-Layer Convolutional Neural Network, Hybrid-CNN applied Enhanced Cat-Swarm Optimization algorithm for optimal weight selection. The convolutional neural network trained based on feature values by varying the training percentage of dataset (in the range of 50%, 60%, 70%, 80%, and 90%) and the proposed model attains the elevated accuracy of 92.93%. The performance evaluation metrics are used such as Sensitivity, Specificity, Accuracy and F-Measure to evaluate the robustness of proposed Hybrid-CNN classification model.

Keywords: Lung Cancer Nodules · CT Images · CAD · Pre-processing · Segmentation · Feature Extraction · Classification

1 Introduction

The Lung-cancer disease is becoming prominent source of deaths for human beings in worldwide, research statistics shows that the 2.3 million patients diagnosed with cancer out of which around 1.8 million patents are died due the lung cancer only in the

R. Manza et al. (Eds.): ACVAIT 2022, AISR 176, pp. 39–57, 2023.
https://doi.org/10.2991/978-94-6463-196-8_5

year 2020 [1–3]. Detecting Lung-Cancer in early stage of it can raise the prospects of patients to survive up to five years after getting treated, the well-known standard lung cancer diagnosis methods are using X-rays, CT images, Blood sample and Biopsy are used by medical professional. The most impactful scanning method is used in diagnosing the disease is low dose CT scan, the LDCT images are less pertaining to clear the lung nodules in CT images through normal human visual perception and conducting scanning of High Dose CT scan techniques are more hazardous on human body [4, 5], the substantial number of research work is conducted for lung cancer detection in its early stage of it using CAD systems [6]. In such cases computer-based technology is coming with enormous solution, similarly computer assisted diagnosis (CAD) system are becoming more effective for offering the solution to assist the medical system [7, 8]. The CAD system is substitutional system which avoid the situation of indecisiveness in lung cancer disease diagnosis from input CT image of LIDC dataset, consequently the CAD systems assisting medical professionals for precise diagnosis by reducing the erroneous falsification [9–11]. The revolutionary methods of Machine-Learning and Image-Processing techniques maintains detection of nodules and nodules-classification based on their type, shape, dimension, and additional features in CT images [12, 13]. Correspondingly the manual CT scan examination is a time overwhelming and complex task to work on massive amount of dataset for every patient. As the size and structural patterns fluctuates in each slice of CT image becomes difficult for radiologist to categorize the nodules. The basic steps involved in this proposed work for Lung Cancer detection are Image Pre-processing using Adaptive Median Filtering Technique and contrast enhancement using histogram equalization technique, secondly Image Segmentation using improvise LevelSet segmentation method. In segmented ROI possesses the nodules which has certain properties of grey-level intensity, shape, and structure in third stage extracting and optimizing ideal features from CT images, and finally in classification step where the detected nodules are classified whether it is benign or malignant.

2 Related Work

In Ayman E. et al. [14], authors projected model for predicting lung cancer nodules from LIDC-IDRI dataset of CT images, the convolution neural network classification technique is employed for lung cancer nodules detection. The provided input images are converted into stack encoder (SAE) for pro-cessing the input image, then by extracting significant features from input image the proposed model established Convolution Neural Network (CNN) and deep learning neural network classification for detecting the nodules whether it is benign or malignant. The proposed system achieves the accuracy of 84.32%.

B Halalli et al. [15], the authors of paper have stated that computer-aided-diagnosis (CAD) system for detecting cancer in Brest and to assist radiologist by applying artificial intelligence-based algorithm in Image-Processing techniques. The proposed CAD system accomplished the work in several phases such as image-pre-processing, Segmentation-of-images, Extraction-of -Features, classification. This proposed model is more applicable for detecting of accumulation in mammography based on feature extraction, further the accumulated mass is classified whether it in normal or abnormal using classification technique. The present research strengthened the medical image analysis step by developing robust model, the performance evaluation metrics are improving sensitivity rate by reducing the computing cost of the system. The quality of input image is improved using Weiner Filtering and CLAHE histogram equalization technique, then by applying contour-based segmentation method the ROI is segmented. In feature extraction step the measurable features such as Wavelet dB4, GLCM, SFTA features are isolated, those relative features-values offered to classifier (SVM) for classifying the mass type belong cancerous or non-cancerous.

J Wason, et al. [16], this paper comprises of discussion of application of image processing techniques for lung cancer detection using CT images, the use CAD system for lung cancer detection can classify the tumours are belongs to benign or malignant. The steps involved in CAD system are Image pre-processing, Segmentation, Feature Extraction and Classification. Compellingly the authors of this article suggested a path to enrich the superior accuracy by engaging the productive stages of CAD system.

M. Keshani et al. [17], the author has presented feature extraction techniques in CT image, this work has employed the segmentation method known as adaptive fuzzy thresholding to separate the nodules in CT images, then consecutive step acquires hole-free lung-mask by applying widow size 5×5 and 23×23, then in next step to segment large and small size nodules are the masked window is rotated with 45degree by modified dimension of window size 50×50 and 25×25. The fourth step of segmentation allows to apply standard threshold algorithm to separate the region beyond thorax portion of bones province of chest region which are present in inner portion image. Conclusively the established mask returns to the Active-contour model as input. This projected work has successfully extracted the region of interest using adaptive fuzzy thresholding technique with measurable features learning purpose in classification step.

R. Shojaii et al. [18] introduced the segmentation technique to acquires the lung region using watershed segmentation algorithm. The fundamental application of this algorithm is to classify the internal and external boundary of lung region, by marking gradient region of lung image is separated. This work has applied smoothening function using Rolling bar filtering technique, this method has complementary function to filling up the unfilled portion of original boundaries of image.

E. M. van Rikxoort et al. [19], The authors have applied hybrid segmentation model for segmenting lung region, profoundly this method has accomplished in four steps, first step applied the region growing algorithm and morphological methods to automatically segment the lung region. Concurrently, this step found some of the errors in segmentation, but those are promptly removed by tis proposed algorithm. The proposed algorithm is speedy and little intricate to accomplish the task, but the lung region is segmented using multi-atlas technique. Decisively proposed work is evaluated based on error detection method to counteract the error.

J. Lee et al. [20], this work has introduced CAD model which computes the multiple geometric factors to categorize the alleged nodules in CT image based on its structure, elongation, dimension, conjecture, intensity, and some of the more pixel-based features. The further classification step classifies the nodules belongs to benign or malignant. Gurcan et al. [21] the author of this paper has presented the applicability of some of the statistical features to be extracted in 3D object for the purpose of classification nodules using measuring components such as volume, superficial area, grey values, standard-deviation, kurtosis, and skewness of the grey values of each pixel.

3 Methods and materials

The proposed computer assisted diagnosis system for lung cancer nodule detection using ideal feature extraction techniques and Hybrid-CNN classification model is demonstrated in Fig. 1, the intense analysis is conducted to implement the proposed CAD system [22], the methods implicated for accomplishment of proposed work has stated in subsequent phases of it is as follows.

3.1 Dataset

The proposed model uses the LIDC-IDRI dataset in DICOM image format, the size of images is in 512 $\times$ 512 pixel, the nodules size is categorized with Nodules $\leq$ 3 mm, Nodules $\geq$ 3 mm, Non_Nodules $\geq$ 3 mm. This dataset comprises of patients having Benign or Non-Malignant cases, Malignant cases, Metastatic Lesion, which are confirmed Malignant cases [23].

3.2 Image Preprocessing

The medical images may consist of different types of noises in it, the adaptive median filtering technique is more pertinent approach to eliminate such type noises from CT images. The proposed work applying the Adaptive Median filtering preserve the useful information in CT image and overcome on the disadvantages of standard Median filtering techniques.

3.2.1 Contrast Limited Adaptive Histogram Equalization (CLAhe)

The histogram equalization technique is employed to improve the contrast of input image, the Contrast-Limited Adaptive Histogram Equalization (CLAhe) is preferably used in this work to enhance the specific region of CT image. The CLAhe method works in several steps, first step is split the image into several similar size of region which should not be overlap, in second step histogram is calculated for each divided region, the third step calculates the Cumulative Distribution (CDF) function in Eq. (1), using Clip-point limit (β), the fourth step performs the Bilinear Interpolation operation to reduce the artificial relics and finally the output image is formed with improved contrast quality.

$$CDF = \sum_{i=0}^{l} f_j(j_i) \tag{1}$$

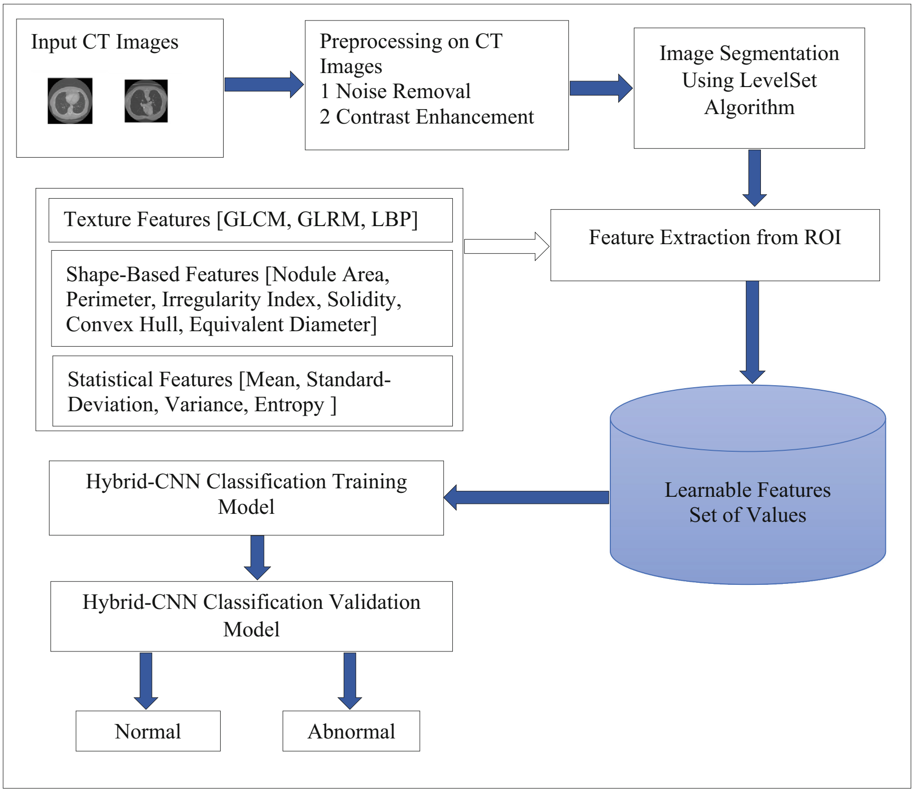

Fig. 1. Block diagram of Computer Assisted Diagnosis System for Lung Cancer Detection Using Hybrid-CNN Classification Model

3.3 Segmentation

The one more valuable step in CAD system is segmentation, which segment the suspicious region from CT image, basically the segmentation is applied on the basis of discontinuity and similarity property type of pattern in any image. The segmentation for discontinuity is applied when there is disordering pixel intensities are reflected, and segmentation for similarity is performed when similar types of pixels patterns are found. The proposed CAD model is applying narrative enhanced LevelSet algorithm for segmentation of region of interest (ROI) in CT images.

The enhanced LevelSet segmentation approach is employed based on the standard Level Set Segmentation method, for separating the exterior region the speed function of standard algorithm is uses moving curve with curving-based velocities. The curve distribution is generated from levelset function, $\phi = (x, y, z) = 0$, initial distribution is set by LevelSet segmentation algorithm is zero. Though, the levelset function, $\phi = (x, y, z)$ is assigned to SDF function, similarly the Signed Distance Function is defined in Eq. (2). For moving curve velocity to develop boundary of exterior region with respect to SDF function the speed function Φ_t is defined in Eq. (3).

$$SDF = (x, y, z) \tag{2}$$

$$\Phi_t = F|\nabla_t| = 0 \tag{3}$$

The LevelSet algorithm is prominently separating the suspicious ROI from CT images by estimating the expectant values of FC-Means and K-means algorithm, the experimental enhancement in proposed work use disaco function isolate the boundary of suspicious ROI.

3.4 Feature Extraction

Major contribution of proposed model is on feature extraction method, feature extraction is a significant phase in nodule detection. The feature extraction endorsed the certain features, those measurable features are used to train the Convolutional Neural Network (CNN), and convincingly based on the parametric feature value nodules are categorized in normal and abnormal class in given input CT image. The preceding step in CAD system extract the suspicious ROI, from the segmented ROI the several features such Texture features, Shape-Based Features, and Statistical Features are estimated the parametric values, these extracted features values are useful for analysis of disease [24, 25].

3.4.1 Texture Features

The texture measures the spatial variation in pixel intensity, the texture features compute the transformation-based model for given object in input image. The texture features compute the Grey-Level Co-occurrence Matrix (GLCM), Grey-Level Run Length Matrix (GLRM), and Local Binary Pattern (LBP) [26] [27].

3.4.1.1 Glcm

The GLCM measures the second order textural features such as Energy, Correlation, Contrast, Homogeneity, and Entropy.

Energy

The energy feature measures the intimacy between the elements which are distributed in spatial pattern of image. The energy feature is measured in Eq. (4).

$$Energy(En) = \sum_{i=1} \sum_{j=1} (px(i, j))^2 \tag{4}$$

Correlation

The correlation measures the probability of pairs of pixels which are connected to each other, this measuring parameter recognizes correlation with neighbouring pixel values, the correlation features estimation equation is defined in Eq. (5).

$$Correlation(Cr) = \sum_{i,j=0}^{G-1} P(i, j)(i - \mu_i)\left(j - \mu_j\right)/\sigma_i\sigma_j \tag{5}$$

Contrast

The contrast measures the pixel intensity between neighboring pixels, under contrast estimation equation evaluate the inertia and variance of given input image, the contrast feature values calculation is represented in Eq. (6).

$$Contrast(Co) = \sum\nolimits_{i,j=0}^{N-1} P(i-j)^2 \tag{6}$$

Homogeneity

The Homogeneity reveals the uniformity in texture of image and organize the local variation in texture of image. The lack of intra-regional variation is observed when the high-level values is indicated by homogeneity in the distribution of texture. The equation for Homogeneity is defined in equation no. (7).

$$Homogeneity(Ho) = \sum\nolimits_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2} \tag{7}$$

Entropy

The entropy measures the uncertainty in pixel distribution in texture of image, entropy calculated the probability of random pixel values at each position of *(i,j)*, entropy calculation shown in Eq. (8).

$$Entropy(En) = -\sum\nolimits_{i=1} \sum\nolimits_{j=1} (pl(i,j))log(pl(i,j)) \tag{8}$$

3.4.1.2 GLRM

The Grey-Level Run-Length Matrix (GLRM) feature counts the total number of pixels present with similar intensity values in specific direction [28]. The 2-D matrix of GLRM consist of pixel intensity value i and j, and angle ϴ indicates the element present in specific direction. The GLRM is extracting and calculating total fourteen features in all (which are known as SRE (Short-Run-Emphasis), LRE(Long-Run-Emphasis), GLN (Gray_level nonuniformity), GLNN, RLN, RLNN, RP(Run Percentage), LGLRE (Low _Gray_Level Run-Emphsis), HGLRE (High_Gray_Level Run-Emphasis), SRLGLE, SRLHGLE, LRLGLE, LRHGLE, and GLV) [29]. The grey-level run length are group of elements of sequential and colinear pixel points having analogous grey level run-length values, the GLRM is calculate the Run-Length (RNL) in Eq. (9).

$$RNL(\theta) = (g(i,j)|\theta),\ 0 \le i \le GY_{maxe},\ 0 \le j \le RNL_{maxe} \tag{9}$$

Here, GY_{maxe} indicates the maximum grey-level values and RNL_{maxe} represents the maximum run-length values.

3.4.1.2 LBP

The LBP feature extraction method calculates the LBP value and validating that the limit of designated value is compared with adjacent pixel and replacing the conclusive binary value. Furthermore, LBP converts the subsequent positive values encode as 1 and negative value is set 0. The binary values are arranged from left in clockwise direction the generated code is accepted as LBP code. The LBP is extracting 100 features.

3.4.2 Shape-Based Features

The Shape-Based features analyzes the attributes at each pixel level, the shape-based features have properties of instinctiveness and optical in nature [30]. This research work is centering to compute the shape-based features such as Nodule Perimeter, Nodule Area, Nodule Irregularity Index, Convex Hull, Solidity, Equivalent Diameter [31].

Nodules Perimeter

The Perimeter is a property is used to calculate the suspicious edges of nodules, if identifies the distance amongst adjacent pixels over the ROI and computing the periphery of nodules shown in Eq. (10), perimeter will return the unknown values when there is incoherence in pixel values [32].

$$Nodule\ Perimeter(PN) = \sum_{i=0}^{L} \sum_{j=0}^{W} p(i, j) \tag{10}$$

Here, L indicates length, W indicates width, and $p(i, j)$ is the pixel values in region.

Nodules Area

The Area of nodules defines the abnormal growth of tissues in CT image, which is calculated using equation no (11), equation evaluates the region in image having binary value as 1.

$$Nodule\ Area(AN) = \sum_{m=0}^{L} \sum_{n=0}^{W} j(m, n) \tag{11}$$

Here, $j(m, n)$ are the pixel values of suspicious region, L indicates length, W indicates width.

Nodule Irregularity Index

The edge of nodule is characterized by computing irregularity index in, the Nodule Irregularity Index is calculated using Eq. (12).

$$Irregularity\ Index(IN) = \frac{4\Pi \times AN}{(PN)^2} \tag{12}$$

Here, AN is Area, and PN is Perimeter of nodule, the nodule identification through Irregularity Index (IN) is evaluating the value of roundedness index = 1 only for circle, and for any other shape it should be less than 1. The high probability of object is to be nodule is presumed based on high ranking of circularity of object.

Convex Hull

The convex hull is collection of P points. Which is the tiniest region in convex polygon with the same beginning and end point. The points are collected in counter-clockwise direction around the polygon. Convex area of nodule is ordered the sum of pixel in convex image. The established leaping frame is considered as convex image, and leaping frame is convex hull.

Nodule Solidity

Nodule solidity measure the area wrapped under pixel intensity, though the elements present in nodule regions are highly concentrated in natural surroundings. The nodule solidity is calculated in Eq. (13).

$$Nodule\ Solidity(SN) = \frac{AN}{Convex\ Hull} \tag{13}$$

Here, AN represents the Nodule Area.

Nodule Equivalent Diameter

Equivalent Diameter is typical accumulation method to calculate the diameter of circle having equivalent aggregation of partitioned area, it is calculated using Eq. (14).

$$Nodule\, Solidity(SN) = \frac{AN}{Convex\, Hull} \tag{14}$$

3.4.3 Statistical Features

The statistical features are extracted to evaluate the nodules mathematical parametric value, the statistical property of nodule can be applied for the stage of classification to classify the nodules category. The statistical features calculated through Mean, Standard-Deviation, Variance, Entropy, Moment, Skewness, and Kurtosis. The size of statistical features is 7, which are combinedly applied together as parametric value to the next classification model to classify the nodule [33].

3.5 Classification Using Hybrid-CNN

Convolutional neural-networks are the standard network which applies the mathematical formulation with nonlinear activation functions on the input CT images which are compiled from defined sources, conclusively these neural networks are producing final results as detected nodules are benign or malignant [34, 35]. This research is optimized weight distribution hybrid CNN architecture set with convolutional filters or kernels, the extracted features from input dataset are provided to Hybrid CNN architecture and accomplished feature mapping using transformation function. In fully-connected layer all layers of CNN layer are connected every neuron with all other neuron in impending interlinked layers. The well optimized CNN layers in hybrid layer are in three dimensional, the neurons in CNN are designed in the form of height, width, and depth. In hybrid-CNN layers selects a tiny ROI of image and analyzes the features parametric values of selected portion and subsequently the decision each portion of image will merge together to get the complete result. ReLU layers are used as activation function to stimulates the neuron and Max-pooling layers is applied to optimize the parametric values in hybrid CNN model. Moreover, q^{th} feature mapping is accomplished by the s^{th} layers, the position of feature value (p, r) is defined as $F^{s}{}_{prq}$ is exhibited in Eq. (15).

$$F^{s}_{p,r,q} = W_k^{s^T} J^{s}_{p,r} + B^{s}_k \tag{15}$$

where, q^{th} filters value is given in s^{th} layer, W_k is the elective weight in the given equation and B_k is the bias weight. Sequentially, the optimize weight evaluation technique is implemented using Enhanced Cat-Swarm Optimization (ECSO) Algorithm to enhance the accuracy of proposed Hybrid-CNN approach. Crucial layer of CNN model is the Sigmoid layer function, this act as activation function to give conclusive result for detected nodule is normal or abnormal. The sigmoid function is derived in Eq. (16), the sigmoid function is specified as $f(y_k)$.

$$f(y_k) = \frac{1}{1 + e^{y_k}} \tag{16}$$

Training the Hybrid-CNN using ECSO weight optimization model classifies the nodule in conclusive activation known as sigmoid function, it classifies the detected nodule based on the evaluation value 0 and 1. This works has distributed the training percentage of dataset in range of 50, 60, 70, 80, and 90, the subsequent section discusses the performance analysis of proposed work.

4 Result and Discussion

The CT images are most substantial way to diagnose the lungs related diseases, equally in detecting lung nodules LIDC-IDRI dataset images are exclusively used as source of input in CAD computer-based lung cancer diagnosis system. This research work has conducted in distinct stages of it, which are discussed independently in prevalent sections of this research paper, this section is overseeing the performance analysis of proposed work. This work accomplished with minimum software and hardware requirements, Software: Python- Version: 3.7, IDE: pycharm: Version: 2019.2.4, OS-Windows 10. Hardware: Intel Core-i5, RAM-8GB, ROM-More than 100 GB, GPU-Yes,CPU-1.7 Ghz. This research work has used CT images of 793 malignant images and 1323 is benign images. The Fig. 2 shows the transformation of input CT images in (a) and (d), the pre-processed CT images (b), (e) are improved the contrast of images, then by segmenting ROI the measurable features are extracted shown in (c) for benign case and (f) for malignant case. The learnable features are used for training the Hybrid-CNN model with optimizing weight selection method using Enhanced-CSO algorithm, by varying the training percentage with 50, 60, 70, 80, and 90 percentage the experiments is supervised in 50 epochs.

The Hybrid-CNN classification model is trained using ideal features value of Shape-Based Features and Statistical Features, the pattern of extracted feature parameter values are used to analyze the category Benign and Malignant Cases, the benign and malignant cases are discussed in Table 1, the nodules Area and Perimetric values in malignant case are higher as compared to benign cases. The smaller values of irregularity index and nodule Solidity indicates malignant nodules, the diameter of malignant nodules is comparably greater than the benign case. Moreover, Table 2 the statistical parameters such as Mean, Standard Deviation and Entropy are relatively high-ranking values in malignant nodules.

The intent of measuring performance analysis is to investigate the robustness of proposed classification model using ideal feature extraction technique is measured through Accuracy, Specificity, Sensitivity and F-Measure. However, to examine sturdiness of proposed model it is compared with aforementioned standard classification techniques such as DBN [36], SVM [37], and CNN [38]. Table 3, represents the Sensitivity measurement by varying the training percentage of between 50, 60, 70, 80 and 90, the objective of sensitivity is measure the truly classified positive cases amongst all actual positive cases in used dataset. The lower sensitivity ratio for Hybrid-CNN model is 90.37% at 50^{th} training percentage and higher sensitivity ratio is 92.9% at 90^{th} training percentage. Moreover, at 90^{th} training percentage the sensitivity ration for Hybrid-CNN is 4.2% better than CNN, 4.5% is better than SVM and 4.72% higher than DBN.

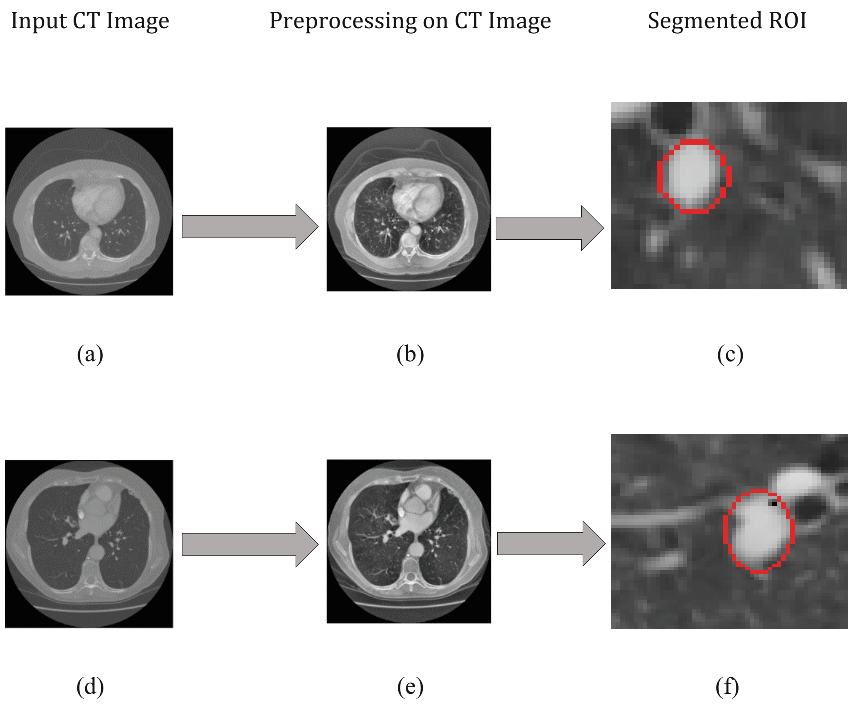

Fig. 2. Feature Extraction: (a) and (d) are the input CT images, (b) and (e) are the pre-processed CT Images, (c) is Feature Extracted in Benign case and (f) Feature Extracted in Malignant case.

Table 1. Shape-based Features Parameter Analysis for categorizing Benign and Malignant Cases

Shape-Based Features		
Shape-Based Parameter	Benign Cases	Malignant Cases
Nodule-Perimeter	56.341	182.362
Nodule-Area	381	699
Nodules-Irregularity index	0.8932	0.2103
Convex Hull	91	689
Nodule-Solidity	0.951	0.556
Nodules-Equivalent Diameter	7.4567	19.9099

The Fig. 3 represents the sensitivity ratio of Hybrid-CNN through graphical analysis in comparison with DBN, SVM and CNN, at distinct training percentage proposed model delivering superior sensitivity ratio which can be clearly visualize in graph.

The Specificity measures the ratio of detecting negative cases or the normal cases amongst all actual normal cases in the given dataset, Table 4 is pertaining the Specificity of Hybrid-CNN model and compared with standard classification model such DBN [36],

Table 2. Statistical Features Parameter Analysis for categorizing Benign and Malignant Cases

Statistical Features		
Statistical Parameter	Benign Cases	Malignant Cases
Mean	0.0218	0.0879
Variance	5.12E + 06	2.19E + 07
Standard deviation	3.06E + 03	5.16E + 03
Entropy	0.0021	0.0039

Table 3. Sensitivity ratio at distinct training percentage

Sensitivity					
Training (%)	50	60	70	80	90
DBN [36]	78.16	78.61	81.77	86.62	88.18
SVM [37]	78.39	79.61	83.53	86.93	88.4
CNN [38]	84.16	87.3	87.31	88.07	88.7
Proposed Hybrid-CNN	90.37	91.13	92.3	92.62	92.9

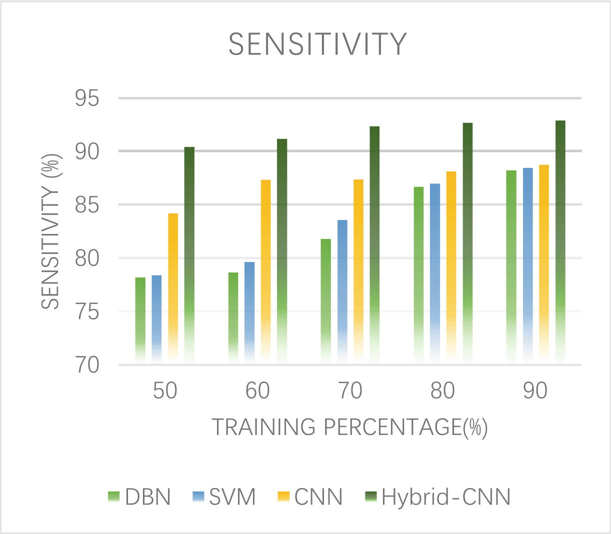

Fig. 3. Sensitivity analysis of Hybrid-CNN in comparison with DBN, SVM and CNN

SVM [37], and CNN [38]. The proposed classification technique is dispensing excellent results for detecting benign cases from all actual benign cases from dataset. The Specificity ratio of Hybrid-CNN attains 10.31%, 10.43%, and 11.18 superior values at 90th training (%) than other customary models such as CNN, SVM, and DBN respectively.

The Fig. 4, explicitly demonstrate the graphical representation of Specificity values for Hybrid-CNN and likening the other conventional classification models such as DBN, SVM, and CNN. The irrefutable analysis reveals the critical study of classification models that exhilarating the companionship within it.

Correspondingly, proposed Hybrid-CNN classification model is highly improved approach, which relatively well distinguishable in comparison with prevalent classification model, the Table 5, includes the results of Accuracy measurement of Hybrid-CNN and contrasted with DBN [36], SVM [37], and CNN [38]. The proposed Hybrid-CNN achieves the 4.4%, 5.02%, 5.05% upper values compare to classical classification model such as CNN, SVM and DBN at 90th percentage.

Table 4. Specificity ratio at distinct training percentage

Specificity					
Training (%)	50	60	70	80	90
DBN [36]	63.47	74.57	83.03	84.62	85.82
SVM [37]	79.76	80.87	84.62	85.64	86.57
CNN [38]	82.77	83.08	85.15	86.28	86.69
Proposed Hybrid-CNN	94.05	95.09	96.13	96.62	97

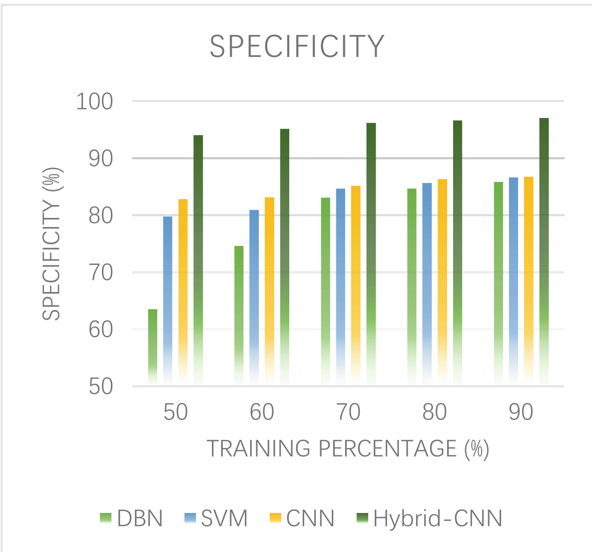

Fig. 4. Specificity analysis of Hybrid-CNN in comparison with DBN, SVM and CNN

Table 5. Accuracy ratio at distinct training percentage

Accuracy					
Training Percentage	50	60	70	80	90
DBN [36]	79.87	79.87	82.42	85.26	87.88
SVM [37]	79.87	79.87	83.43	87.71	87.91
CNN [38]	82.31	85.27	86.79	88.29	88.53
Proposed Hybrid-CNN	90.07	91.08	91.51	92.28	92.93

The influence of vigorous Hybrid-CNN classification model can be evidently validated through the graphical analytical study in Fig. 5, the Accuracy value at all training percentage perceives the improved results, at 90th percentage archiving the best accuracy is 92.93% which exceptionally marginal in comparison with other classification model.

The F-measure calculating the accuracy of classification model based on the Recall and Precision ratio of the classification analysis system, F-measure the single score value for positive nodule prediction amongst all positive cases in dataset. The Table 6 comprises of F-Measure ratio for Hybrid-CNN and differed with DBN [36], SVM [37], and CNN [38] classification model.

The Fig. 6 shows the graphical analysis F-Measure ratio for Hybrid-CNN model and compared with other classification model.

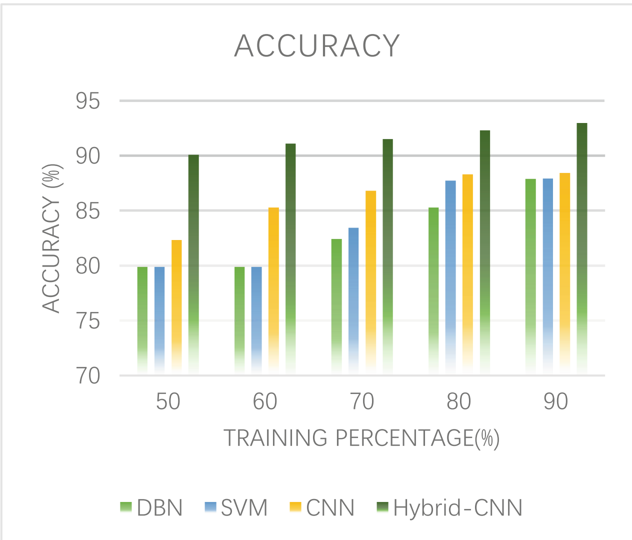

Fig. 5. Accuracy analysis of Hybrid-CNN in comparison with DBN, SVM and CNN

Table 6. F-Measure ratio at distinct training percentage

F-Measure					
Training Percentage	50	60	70	80	90
DBN [36]	84.71	85.92	86.62	86.88	89.67
SVM [37]	84.94	85.94	87.74	87.93	89.71
CNN [38]	88.13	88.15	88.73	89.74	89.84
Proposed Hybrid-CNN	93.76	93.84	93.92	93.94	93.96

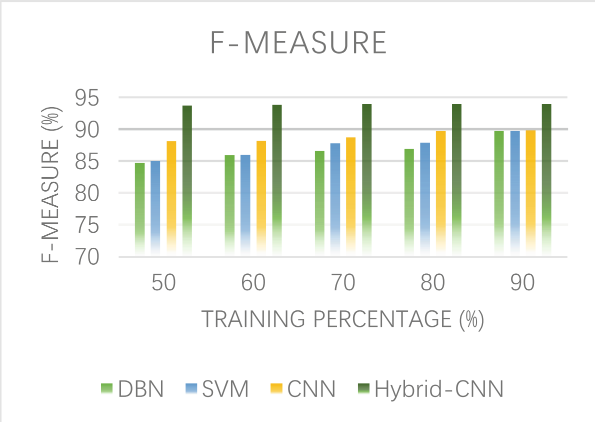

Fig. 6. F-Measure analysis of Hybrid-CNN in comparison with DBN, SVM and CNN

5 Conclusion

This research work has implemented proposed CAD system using Hybrid-CNN architecture for lung cancer detection, the crucial steps involved in proposed models are Preprocessing, Segmentation-(ROI), Extraction of Features, and ending with Classification. Specifically, the first step applies the preprocessing technique using Adaptive-Median Filtering method to eliminate the speckle noises from CT images and applying Contrast-Limited-Adaptive Histogram-Equalization method (CLAhe) to contrast enhancement in input CT image. The enhanced CT images are positively affecting to improve the results of proposed model. Later in the second stage of work is segmenting the Region of Interest (ROI), which is proportionally leaning with suspicious nodules in CT image. The third step of Feature Extraction employs segmented ROI as input to obtain the learnable features-value in ROI, this research model has contributed to extract the Texture Features, Shape-Based Features, and Statistical Features from CT images. The measurable Shape-Based Features are Nodule-Perimeter, Nodule-Area, Nodules-Irregularity index, Convex Hull, Nodule-Solidity, Nodules-Equivalent Diameter, and Statistical Features are Mean, Variance, Standard deviation, and Entropy. The most significant final step is classification using Hybrid-CNN classification model employed

Enhanced CSO algorithm for optimally selecting the weight to improve the performance of proposed system. Analysing critical performance evaluation of proposed model this paper has determined Sensitivity, Specificity, Accuracy and F-Measure ratio. The results proposed model are compared with standard classification model such DBN, SVM and CNN, likewise the Hybrid-CNN model attains the Sensitivity-92.9%, Specificity-97% and Accuracy-92.93% which is remarkably improving the results of proposed model. The crucial F-Measure for evaluating positivity ratio for proposed Hybrid-CNN model is delivering 93.96% ratio. Established performance analytical study exhibits that the proposed model of Lung Cancer detection is a robust model for the diagnosis of disease.

Acknowledgments. The authors would like to thank Head CSE Prof. P. Shyamala, UCE, Osmania University, Hyderabad. And Prof. Suresh Lokhande, Dean BOS in CSE department, UCE Osmania University, Hyderabad, India for their continuous motivation, help and always being there to support for motivation.

Authors' Contributions. The proposed work has major contributions of segmentation of ROI using enhanced Levelset segmentation method, which is consisting of suspicious region in input CT images, secondly is extracting prominent features from region of interest, and thirdly implementing classification Hybrid-CNN model using improved weight optimization algorithm.

References

1. World Health Organization(WHO), Global Health Estimates 2020-Deaths by Cause, Age, Sex, by Country and by Region, 2000–2019. WHO; 2020. Accessed December 11, 2020.
2. Bray F, Laversanne M, Weiderpass E, Soerjomataram,",The ever-increasing importance of cancer as a leading cause of premature death worldwide.", Cancer. In-press in Web of Science.
3. S, Hyuna, Jacques Ferlay, L.Rebecca, Siegel, L Mathieu, I Soerjomataram, Ahmedin DMV, Jemal B Freddie, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries", A Cancer Journal for Clinician, Pages 209–249, Volume:71, Issue:3, May-June-2021.
4. B. N. Narayanan, R C Hardie, Kebede, T.M, "Performance analysis of a computer-aided detection system for lung nodules in CT at different slice thicknesses." Journal of Medicine-Imaging 2018, 5. 5–10.
5. M. Nishio, M. Nishizawa, O. Sugiyama, R. Kojima , M, Yakami, T. Kuroda,K. Togashi, "Computer-aided diagnosis of lung nodule using gradient tree boosting and Bayesian optimization." PLoS ONE 2018.
6. L. Li,Y Wu, Y Yang, L. Li, and B. Wu. "A New Strategy to Detect Lung Cancer on CT Images" InJune-2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC) (pp. 716–722). IEEE.
7. G, Zhang, Z Yang, L Gong, S. Jiang, L. Wang, X, Cao, L. Wei, H. Zhang, Z. Liu, "An Appraisal of Nodule Diagnosis for Lung Cancer in CT Images." Journal of Medicine System 2019, 43-:181.
8. S. A. Regaily, M. A. Salem,M H Aziz Abdel, M. I. Roushdy, "Survey of Computer Aided Detection Systems for Lung Cancer in Computed Tomography," Current-. Medicin Image. Revi.-2018, :14. 3to18.

9. Y. Chunran, W. Yuanvuan, and GX Yi, " .Automatic Detection and Segmentation of Lung Nodule on CT Images," :2018 -11 International-Congress on Imag and Signal Processing.-BioMedical-Engineering and Informaticss (.CISP-BMEI), 1to 6.
10. S.M.B.Netto, A.C Silva, R.ANunes, M. Gattass, "Automatic segmentation of lung nodules with growing neural gas and support vector machine," Computer. Biolo. Medi.: 2012,-42, page-1110–1121.
11. S.L.A. Lee, A.Z Kouzani, E.J Hu, "Automated identification of lung nodules," .In Proceedings-of the 10th Workshop on Multimedia Signal Processing (MMSP), Cairns, Australia, 8,9,10-October 2008; pp:497–502.
12. D. Cascio, R, Magro F, Fauci, M, Iacomi, G. Raso, "Automatic detection of lung nodules in CT datasets based on stable 3D mass–spring models," Compute. Biom. Medic. :2012,-42, pp:1098–1109.
13. S. Akram, M.Y Javed, M. U. Akram, U. Qamar, A. Hassan, "Pulmonary Nodules Detection and Classification Using Hybrid Features from Computerized Tomographic Images," Jou. Of Medic. Imaging Health-Informa. :2016, 6, pp:252–259.
14. Ayman El-Baz, Garth M. Beache, Georgy Gimel'farb, et al. Computer-aided diagnosis systems for lung cancer: challenges and methodologies, International Journal Biomed. Imaging 2013 (2013) 46, https://doi.org/10.1155/2013/942353.
15. Halalli Bhagirathi & Makandar Aziz, "Computer Aided Diagnosis -Medical Image Analysis Techniques", Published on 20December 2017, Intech-Open Book Series, DOI: https://doi.org/10.5772/intechopen.69792.
16. Vasanth Jeyaprakash Wason, Nagarajan Ayyappan, "Image processing techniques for analyzing CT scan images towards the early detection of lung cancer," Bioinformation.:2019; Volume15(Issue-8): pp:-596–599. doi: https://doi.org/10.6026/97320630015596.
17. Keshani M, Azimifar Z, Boostani R, Shakibafar A., "Lung nodule segmentation using active contour modeling" published in 6th Iranian conference on machine vision and image processing, IEEE (2010, October), pp. 1–6.
18. Shojaii R, Alirezaie, P. Babyn "Automatic lung segmentation in CT images using watershed transform" Published in IEEE international conference on image processing-2005, volu. 2, IEEE, September 2005, pp. II:-1270.
19. Van Rikxoort E. M., de Hoop B., Viergever M.A., Prokop M., van Ginneken B. "Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection", Medicne of Phy, 36(7) (2009), pp. 2934-2947.
20. Lee I J., Gamsu G, Czum J, Wu N., Johnson R., Chakrapani S. "Lung nodule detection on chest CT: evaluation of a computer-aided detection (CAD) system", Korean Jou Radiolog, Volume:6 (Issue-2), 2005 pp. 89–93.
21. N. Metin, Gurcan, B Sahiner, L. Hadjiiski, N. Petrick, C.Heang-Ping, A. Ella, Kazerooni, Philip N. Cascade, "Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system ", Medici. Phy. Volu:29 Issue-11, November 2002.
22. V Pawar, Kailash Kharat, Suraj Pardeshi, Prashant Pathak, "Lung Cancer Detection System Using Image Processing and Machine Learning Techniques" in International Journal of Advanced Trends in Computer Science and Engineering, Scopus Indexed [ISSN 0974–2034] Volume 9, Issue No.4, July – August 2020. Available Online at https://doi.org/10.30534/ijatcse/2020/260942020.
23. A. G.Samuel, G. McLennan, L. Bidaut, M. F. McNitt, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. L. Henschke, E A Hoffman, "Data From LIDC-IDRI", 2015. Available online: https: //wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI.
24. Priyanka, and Dharmender Kumar, "Feature Extraction and Selection of kidney Ultrasound Images Using GLCM and PCA", Procedi ComputerScience, volu.: 167, 16 April 2020 pp. 1722–1731.

25. K D. Kharat, V J. Pawar, Suraj R. Pardeshi, "Feature Extraction and selection from MRI Images for the brain tumor classification," Presented in IEEE International conference on Communication and Electronics Systems (ICCES) 2016 Coimbatore India.
26. Ojala, T; Pietikainen, M; & Maenpaa, T; "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns" In IEEE Transactions on Pattern Analysis and Machine Intelligence, volume:.24, Issue-no:-7, July 2002, pp. 971–987.
27. Radhakrishnan Manavalan & Kuttiannan Thangavel, "Comparative Analysis of Feature Extraction Methods for the Classification of Prostate Cancer from TRUS Medical Images", IJCSI International Journal of Computer Science Issues, Volume:09, Issue-no.1, Jan-2012.
28. J. K. Kim, M, P. Jeong, S. Koun Song & H P Wook, "Texture Analysis and Artificial Neural Network for Detection of Clustered Microcalcifications on Mammograms," Published in IEEE, 1997;pp.199 – 206,
29. A. J. Masino, A. Kaitlin, Folweiler, "Unsupervised learning with GLRM feature selection reveals novel traumatic brain injury phenotypes", Machine Learning for Health (ML4H) Workshop at Neur IPS-2018.
30. Guo, Z;, Li, Y; Wang, Y; Liu, S; Lei, T; & Fan, Y; "A method of effective text extraction for complex video scene", Math. Problems in Engineering, volume: 2016,Article ID 2187647, 11 pages, 2016.
31. Aarthy, K. P; & Ragupathy, U. S. ; "Detection of lung nodule using multiscale wavelets and support vector machine", International Journal of Soft Computing and Engineering (IJSCE), volume:02, issue-no.03, July 2012.
32. Gozalez R, Woods R. E., "Digital Image Processing Using Matlab," second-edition, Gatesmark, USA, 2002, chat.-12, pp:642–654.
33. https://en.wikipedia.org/wiki/Statistic
34. Wu Jianxin, "Introduction to Convolutional Neural Networks," National Key Lab for Novel Software Technology Nanjing University, China, 01 May 2017.
35. Zeiler, M. D; Fergus, R.; "Visualizing and understanding convolutional networks", in-Proceedings of the European Conference on Computer Vision (ECCV), 2014, pp. 818–833.
36. Shanthi V., Sridevi G., Charanya R. & Josphin, J; "Mary, "Deep Belief Network (DBN) Classification For Lung Cancer Prediction Using KNN Classifier", European Journal of Molecular & Clinical Medicine, ISSN 2515–8260 Volume 07, Issue 09, 2020.
37. Makaju S, Prasad P.W.C., Alsadoon Abeer, Singh A. K., Elchouemi A.,"Lung Cancer Detection using CT Scan Images", sixth International Conference on Smart Computing and Communications, ICSCC 2017, 7–8- December-2017, Kurukshetra, India, 1877–0509 © 2018 The Authors. Published by Elsevier B.V.
38. Jin Xin-Yu, Zhang Yu-Chen, Jin Qi-Liang, "Pulmonary nodule detection based on CT images using Convolution neural network," 9th International Symposium on Computational Intelligence and Design 2016, 2473-3547/16 2016, DOI https://doi.org/10.1109/ISCID.2016.52.