DEEP NEURAL NETWORK BASED CLASSIFICATION MODEL FOR FEATURE EXTRACTION

[1]Dinesh Valluru [2]I.Jasmine Selva kumari Jeya

[1]Research scholar, Anna university, Chennai-600025, Tamil Nadu, India
[2]Associate Professor, Department of Computer Science and Engineering, Hindusthan College of Engineering and Technology, Coimbatore - 641 032, Tamil Nadu, India
Corresponding Author : Dinesh Valluru , dinesh.phd2018@gmail.com

Abstract

Background:

Computer aided diagnosis model developed for the detection of lung cancer can leads to higher survival rate. In recent days, Deep Learning (DL) models finds successful for various applications in medical domain.

Materials and Methods:

In this paper, an effective feature extraction based lung cancer classification model has been presented. The proposed model makes use of Hybrid Feature Extraction (HFE) with Deep Neural Network (DNN) based classification models. The proposed HFE-DNN model undergoes pre-processing to enhance the image quality. Then, feature extraction process is carried out by the use of HFE technique followed by DNN based data classification. The outcome of the HFE-DNN technique can assist doctors in the diagnosis process of lung cancer.

Results and Conclusion:

An extensive simulation results takes place on the benchmark HFE-DNN model and the results confirmed its superior characteristics over the existing methods. The simulation outcome pointed out that the proposed HFE-DNN model has offered effective classification outcome with the maximum accuracy of 85.79%, sensitivity of 79.55% and specificity of 89.03%.

Keywords: Lung cancer; Feature extraction; Deep learning; Classification; Computer aided diagnosis model.

1. Introduction

Computer vision techniques are composed of analytical power for treating non-invasive as well as medical investigation. The obtained clinical images like X-rays, CT, and MRI are the commonly applied objectives for the disease diagnosis [1]. Among all other images, CT image plays a vital role in medical application that captures the images in film types. The contrast of MRI images can be pointed under the application of radio waves while CT scan is monitored by exploiting X-rays. Since MRI images are highly expensive interms of processing cost, most of the doctors prefer to use CT scans for the treatment. The major benefit of CT scan is, it is accompanied in minimum duration of time and offers the illustration of brain, tissues, and thin structure. Basically, lungs cancer is the most deadly disease and many people have lost their life due to this cancer. Based on the IARC survey, around 2.09 million patients were being positive for lungs cancer and 1.76 million were fatal. Typically, an elderly person gets affected by lung cancer when compared to other age individuals [2]. It is the leading cancer that causes death in short span of time as well as it is very complex to detect the presence of lung cancer.

This complexity is due to the tiny size of lesion which is named as a nodule. At the benign stage, tumor cell size would be very small; but when it is not detected at the primary stage the tumor size would be improved and grows into malignant. At this point, radiologist analyzes the disease very easily; however it is hard to save the person's life. In order to save the people's life, primary detection is more essential [3]. Recently, the computer vision tends to evolve programmed systems that is capable of detecting the cancer in an automated fashion and categorize the normal and abnormal (tumor) regions.

As mentioned before, lung cancer is the dominative disease for male and females. According to the report, 1.2 million people are analysed with this cancer, and around 1.1 million people died out of lung cancer [4]. The lifetime of a person can be improved by early prediction of disease. But, the primary detection is a tedious operation. Only some patients can be diagnosed accurately when it is initial or middle stage. As an inclusion Computer-aided diagnosis (CAD) model was more applicable for radiologist to predict and diagnose the anomalies at previous stage. This CAD system is an alternate option for radiologists in prior to recommend a biopsy test. The current research work, it is monitored that strategies of Neural Networks (NN) has been extensively applied for detecting lung cancer using the clinical images. To classify the lung cancer, it is employed with few models on the basis of NN as reported in this study.

Abdulla et al. [5] applied a CAD according to Artificial Neural Networks (ANN) to classify the lung cancer. Hence, the features applied for classifying lung cancer are area, perimeter as well as shape. Under this classification, higher accuracy could be obtained. Camarlinghi et al. [6] projected a CAD system is applied for automated prediction of a nodule. Here, the sensitivity would be partial with 3 FP/scan. Al-Kadi et al. [7] utilized a classifier according to fractal texture features which is accomplished with moderate accuracy. van Ginneken et al. [8] related and integrated 6 CAD techniques to predict the pulmonary nodules. The concatenations of 6 models are capable of detecting every nodules at 2 false positive predictions for each scan and minimum nodules with lower false positives. Cascio et al. [9] implied a CAD approach to select the lung nodules exist in Computer Tomography (CT) images.

In recent times, various traditional approaches are executed to classify the medical image which applies 4 levels from pre-processing to classification. Initially, pre-processing is one of the important step in medical imaging if the classical schemes are used as the actual CT images are composed with some noisy factors that produce negative features. Therefore, the negative features alleviate the classification accuracy hence minimization of irregular feature becomes a challenging task. The nodules are very smaller in size and very hard to identify the usual and tumor images by feature-centric modules. Sample CT images of benign as well as malignant are depicted in Fig. 1.
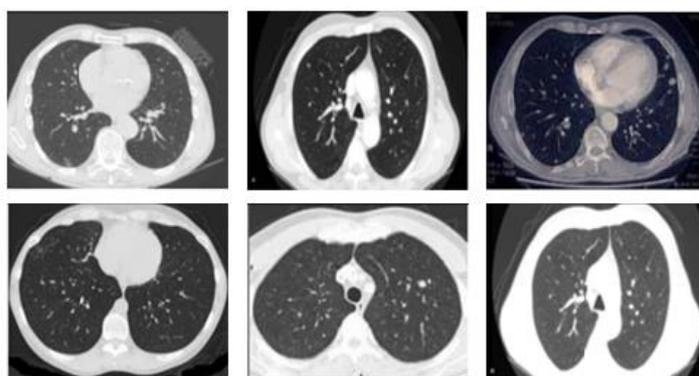

Fig. 1. Sample CT lung images

Globally, lungs cancer is one of the fatal diseases [10]. Massive numbers of CAD modules were established in this work by many developers, also there are several other problems of maximum false rate and minimum identification of lungs lesions. Usually, a CAD system depends upon best features extracting methods like classical as well as DL [11]. Here, the

major attention is on the conventional feature-centric approaches as it is comprised with lower images. Jinsa et al. [12] proposed a CAD system to classify the lungs cancer from CT images. It helps to obtain features from classified lung images and applied some well-known metrics like mean, SD, skewness and so on. Such metrics are again used by feedforward NN (FF-NN) to classify the lung images and attained with a higher accuracy. Kumar et al. [13] projected an auto encoder (AE) oriented DL approach to classify the lung cancer either as cancerous or non-cancerous. Also, it is employed with LIDC dataset to validate the system and reached maximum accuracy. As a result, the newly proposed technique helps the radiologist for early detection of lungs cancer.

Lakshmanaprabu et al. [14] employed a best DL approach to classify the lung cancer from CT images. The obtained feature from a deep model undergoes optimization under the application of Modified Gravitational Search Algorithm (MGSA). Followed by, it reduces the unwanted features by applying LDA and reaches better classification accuracy. Nasrullah et al. [15] established an automatic method to process the lungs cancer classification from CT images. There are 2 deep networks such as U-Net and faster RCNN are interconnected with combined link network to process the feature extraction. Such features undergo classification by applying Gradient Boosting Machine (GBM) and validate the 1200 CT images to accomplish better sensitivity function. Zhang et al. [16] employed hybrid features centric approach to classify the nodules from CT images. The combination of CNN, LBP, and HOG features are capable of obtaining optimal characterization of lungs nodules. These combined features are consequently classified by GBM classification model and attained higher accuracy.

In recent days, several works based on the classification of lung CT images are developed. [17] presented a secured indexing of lung CT image (SILI), a secured way to index the lung CT images with the patient's detail. Authentication is achieved by the use of sender's logo information and the secret key is utilized to embed the watermark into the host image. The simulation values indicated the presented method is robust to unauthorized access, noise, blurring, and intensity based attacks. [18] concentrated on the tuning of weights and bias of Radial Basis Function Neural Network (RBFNN) classifier by the use of presented Real Coded Genetic Algorithm (RCGA). The operators present d in RCGA allows the method to determine the weights and bias value so that minimum Mean Square Error (MSE) is attained. It is tested against Lung Image Database Consortium (LIDC) database and Real time database. The attained results exhibited superior results over the compared methods. [19]

presented lung cancer detection using Optimal Support vector machine, and the same was stored in Cloud for better accessibility. The main reason behind selecting optimal features was fine-tuning the SVM for better prediction. The feature selection in the SVM classification was controlled by the Hybridized algorithm namely modified grey wolf optimization algorithm combined with genetic algorithm (GWO-GA). The result shows better feature selection and classification[20].

In this paper, an efficient feature extraction relied lung cancer classification model has been proposed. The projected model applies Hybrid Feature Extraction (HFE) with Deep Neural Network (DNN) based classification models. The proposed HFE-DNN model undergoes pre-processing to improve the image quality. Later, feature extraction is carried out by the application of HFE technique followed by DNN based data classification. The result of HFE-DNN technique can assist doctors in diagnosing the presence of lung cancer. An extensive simulation results takes place on the benchmark HFE-DNN model and the results confirmed its superior characteristics over the existing methods.IoT has been rapidly growing area, which can be applied in all the areas of Engineering. Recently IoT is mostly used in Healthcare services for assisting doctors to simplify their services. In future, this research work can be extended to implement using IoT based technologies.

2. Proposed model

The projected model is applied for classifying CT images of human lung which that has the levels as given in the following:

- Pre-processing
- Feature extraction
- Reduction
- Classification

In the beginning, CT images are utilized to enhance the superiority of images after using feature extraction strategy for extracting the features like histogram, Texture, and wavelet of images oriented procedures. At the classification stage, CT lung images are labelled into their respective classes. During training process, the proposed model will be trained using the chosen feing along with the selected features of training data. At the time of a testing phase, the simulation outcome of DNN classifier implies that the images are comprised with lung cancer or normal. Fig. 2 shows the architecture of HFE-DNN model.
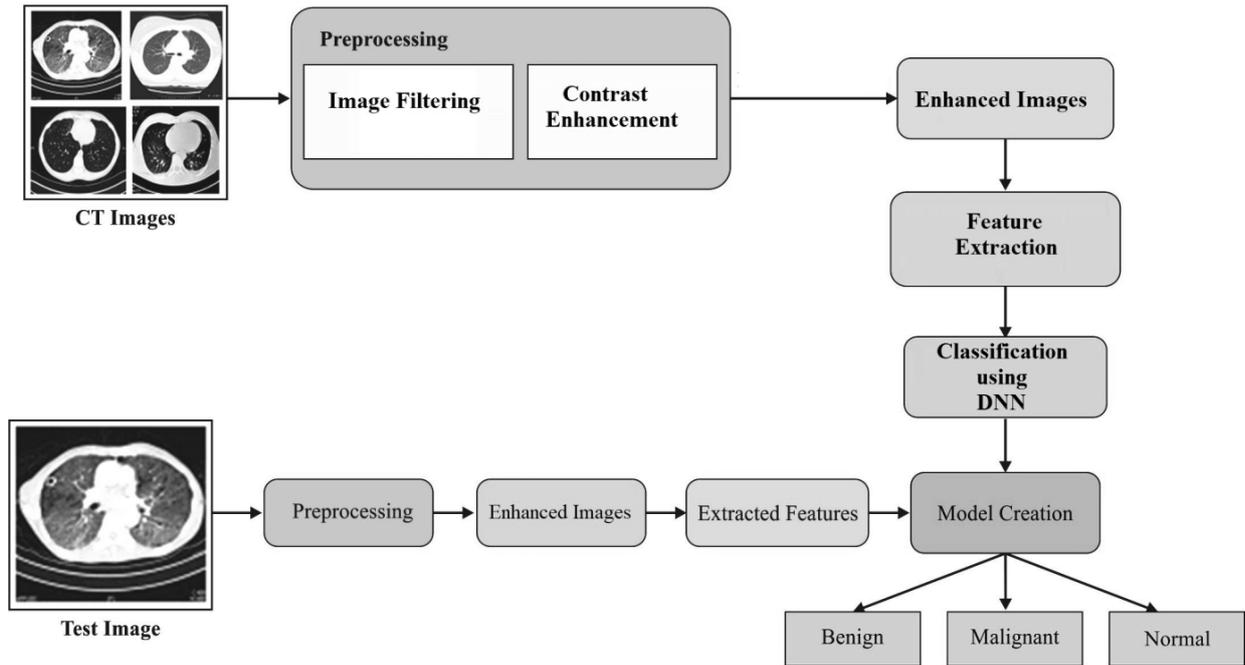
Fig. 2. Architectureof HFE-DNNalgorithm

2.1. Pre-processing

The gathered medical database images are developed with some noisy factors. When an image has noise and desired or nearby pixel exist closer to 0 or 255, it will be replaced with its intermediate value. Once the noisy content has been eliminated from the database, it is assumed to contrast improvement as adaptive histogram equalization.

$$Contrast(i,j) = rank * max\_intensity\,(i,j) \;\therefore\; Initiallyrank = 0$$

$$rank = 0 + 1 \tag{1}$$

The histogram in main location of every line has been obtained by applying primary positions of final row, through reducing the behind columns such as novel leading rows. The difficulty of a CT images are improved and fixed with the limit, this it analyzes the gray level of an image and change dispersion of 2 nearby grayscale levels in a novel histogram.

2.2. Feature extraction

This process makes the image in the exclusive form of matrix vector. It decreases the dimension in image processing on the basis of image classification. It encloses the diminishing input data as decreased representative collection of features. The features are

applied to the classification model. Here, histogram, texture and wavelet features are filtered from the available CT images.

2.2.1. Histogram features

Here, the images can be shown with respect to pixels. It shows the number of pixels from an image at every power. From the applied images, the grayscale value of the images is estimated using a histogram approach. In this model, it has 256 gray levels from the range of 0 to 255. There are also some features like Variance, mean, skewness and kurtosis, SD.

Variance: This metric provides the values of grayscale variations from the average grayscale value. The statistical distributions, such as difference in line length for a specific rage is applied to differentiate minimum contrast level from textures.

Mean: It provides the maximum gray level of all regions and applicable in hostile area of power.

Standard Deviation: It can be computed from the square root of a variance that represents contrast level of the applied images. It can be calculated using the maximum and minimum variance measures. It depicts that a maximum contrast image is composed with greater variance while a minimum contrast image has less variance.

Skewness: It has been estimated on the basis of tail of a histogram and is classified into 2 sets, namely, positive and negative

Kurtosis: It is defined as value of feasible distribution of a real-time arbitrary variables and it represents abnormal image. Kurtoses as well as Skewness are employed in statistical investigation to attain the structure of distribution.

2.2.2. Texture features

It is filtered from the applied images once the histogram feature is determined. Since the abnormal region is generally distributed in the images, texture orientation of all classes becomes important for attaining optimal classifier outcome. Gray level co-occurrence matrix (GLCM) shows a statistics model to review the surface which considers the spatial association of pixels. The GLCM operates the texture of images through calculating reappearance of identical pair of pixels. In general, the features are estimated by applying GLCM probability values with the range of 22 features and is defined in Eq. (2).

$$G_{P_{ij}} = F_{ij} \Big/ \sum_{i,j=0}^{L-1} F_{ij} \qquad (2)$$

From the above function, F_{ij} is a number of occurrences among 2 grayscale levels, L shows the count of quantized grayscale levels, 'i' and 'j' indicates the displacement vectors for particular window sizes.

Energy: It ensures that the higher constant values in grayscale level distribution would shape the higher strength of plane.

Entropy: It is defined as the amount of data from an image that is needed for compression. The image which has minimum entropy shows lower contrast and maximum execution of image pixels from allocated values.

Homogeneity: It is typically named as contrast minute that estimates homogeneous nature of images by considering the prevalent values for insignificant grayscale modification in pair elements. Similarly, the homogeneity is estimation that defines prevalent rates for minor contrast images.

Contrast: It is used in calculating the spatial recurrence of an image and diverse moments of GLCM. It shows the difference among maximum and base values of nearby pixel values.

Correlations: It is applied to determine the linear dependency of grayscale values of nearby pixel values. The observation of electronic image association represents an optic process which misuse tracking as well as register images, which are used in values of image variation.

2.2.3. Wavelet-based features

These features are able to provide an image handling data due to the advantageous features. The DWT tends to linear transformation that is a function on data vectors which has the related energy. Here, the features are extracted in terms of 2 phases as given here. Initially, the sub-band of original image can be deployed and subbands are estimated using diverse resolutions. Wavelet is additional mathematical approach used for extraction and isolates the wavelet coefficients from the image. The mean detection of DWT coefficients are labelled by assuming the actual coarse coefficient.

$$Coeff[a_t] = \delta_{at} \qquad (3)$$

where δ_{at} implies the mean value and is induced in low pass channels that analyzes the minimum recurrence image inside a cut off repetition.

2.3. DNN based classification

Typically, it can be operated as FFNN system and it belongs to unsupervised pretraining model with greedy layer wise training. In this model, data flow from input to output layers with no looping function. The major benefit of DNN classification is, the classification probabilities of missed values is lower. The DNN model implements 1-layer from unsupervised pre-training stage. It assigns a classifier value $f(x)$ at the time of prediction stage. All input data sample $x = [x_1, \ldots . x_N]$ is a forward pass. Then, f is a function that has a sequence of layers to calculate values as shown in Eq. (4):

$$Z_{ij=}x_i w_{ij}; Z_j = \sum_i Z_{ij} + b_j; X_j = g(Z_j), \qquad (4)$$

where input of a layer is given as x_i, output layer is x_j, and w_{ij} implies the modelling parameters and $g(Z_j)$ analyze the mapping function. Layer-wise relevance propagation degrades the classifier output $f(x)$ with respect to relevance's r_i attributing to all input component x_i that contributes in making classification decision as given in Eq. (5)

$$f(x) = \sum_i r_i, \qquad (5)$$

where $r_i > 0$ denotes a positive evidence that supports classification decision and $r_i < 0$ signifies a negative evidence; else, it is named as neutral evidence, though relevant attributesr_i is estimated by applying Eq. (6). The common structure of DNN is shown in Fig. 3.

$$r_i = \sum_j \frac{zij}{\sum_i zij}. \qquad (6)$$

The DNN is capable to predict the irregular feature coherence of input. The DNN offers a hierarchical feature learning system. Finally, the main aim of DNN is to manage the complex function which illustrates the higher level abstraction.
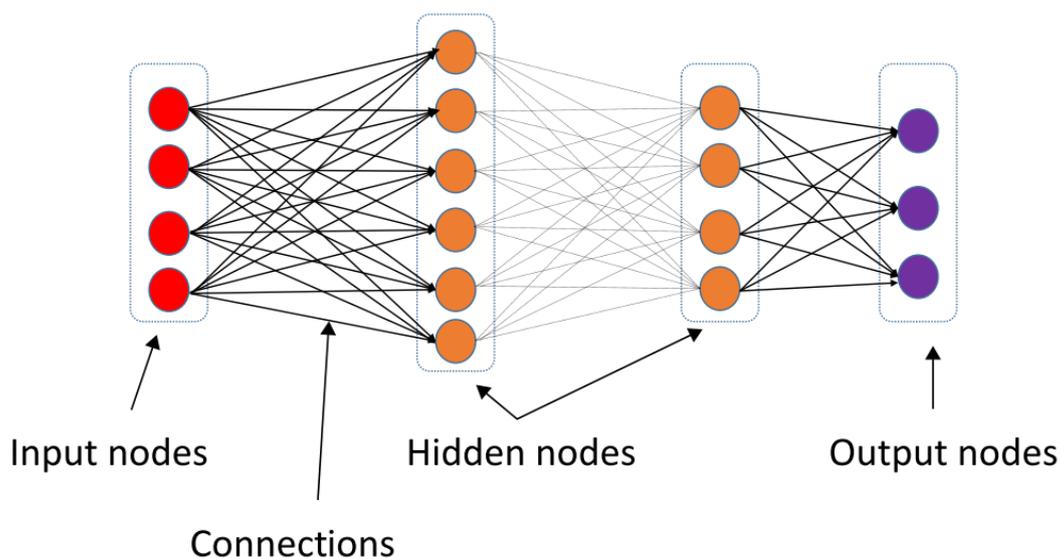
Fig. 3. General architectures of DNN

3. Performance Validation

In this section, the extensive experimental analysis of the proposed HFE-DNN model is carried out on the benchmark image database and the results are examined under diverse aspects.

3.1. Dataset used

The proposed HFEM+DNN technique is executed in MATLAB 2014b. To apply the results of the HFEM+DNN model, a standard image database with maximum low-dosages and saved lung CT images are considered [21]. The images are basically in 1.25mm slice thickness and are generated in single breath in and breath out. The location of a nodule is found by the expert and also provided in a database. Here, a total of 35 images present under normal class and 33 images exist under malignant category. Likewise, a collection of 32 images exists under the benign class.

3.2. Results analysis

Table 1gives a confusion matrix for 3 categories on the applied images by newly developed model. The table values shows that the presented model correctly classifies a set of 6 images as normal, 9 images as malignant and 9 images as benign type. Table 2 and Fig. 4offers the estimation of confusion matrix present in Table 3 by means of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) correspondingly.

Table 1 Confusion Matrix of Proposed HFEM+DNN

Category	Normal	Malignant	Benign	Total
Normal	6	1	1	8
Malignant	0	9	2	11
Benign	2	0	9	11
Total	**8**	**10**	**12**	**30**

Table 2 Manipulations from Confusion Matrix for Proposed HFEM+DNN

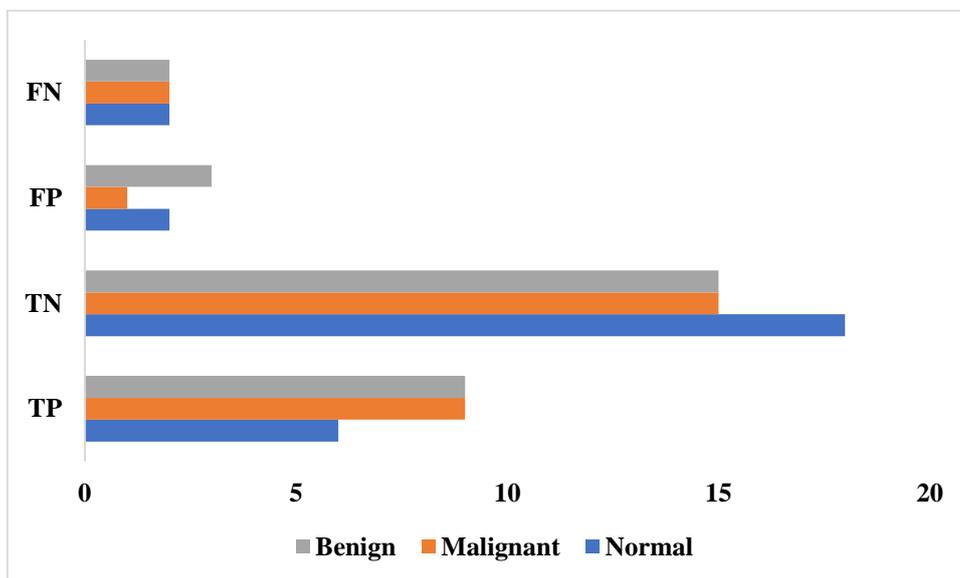| Category	Normal	Malignant	Benign
TP	6	9	9
TN	18	15	15
FP	2	1	3
FN	2	2	2

Fig. 4. Confusion matrix for different classes

Table 3 and Fig. 5offers the classifier outcomes of the applied method with respect to sensitivity, specificity and accuracy under 3 classes of lung CT images. By seeking the table, it can be seen that the normal images undergo classification with the sensitivity of 75%,

specificity of 90% and accuracy of 85.71% respectively. In line with this, it is clear that the malignant images are classified with the sensitivity of 81.82%, specificity of 93.75% and accuracy of 88.89% correspondingly. Similarly, the benign image undergoes classificationwith sensitivity of 81.82%, specificity of 83.33% and accuracy of 82.76% respectively.

Table 3 Results of Proposed HFEM+DNN

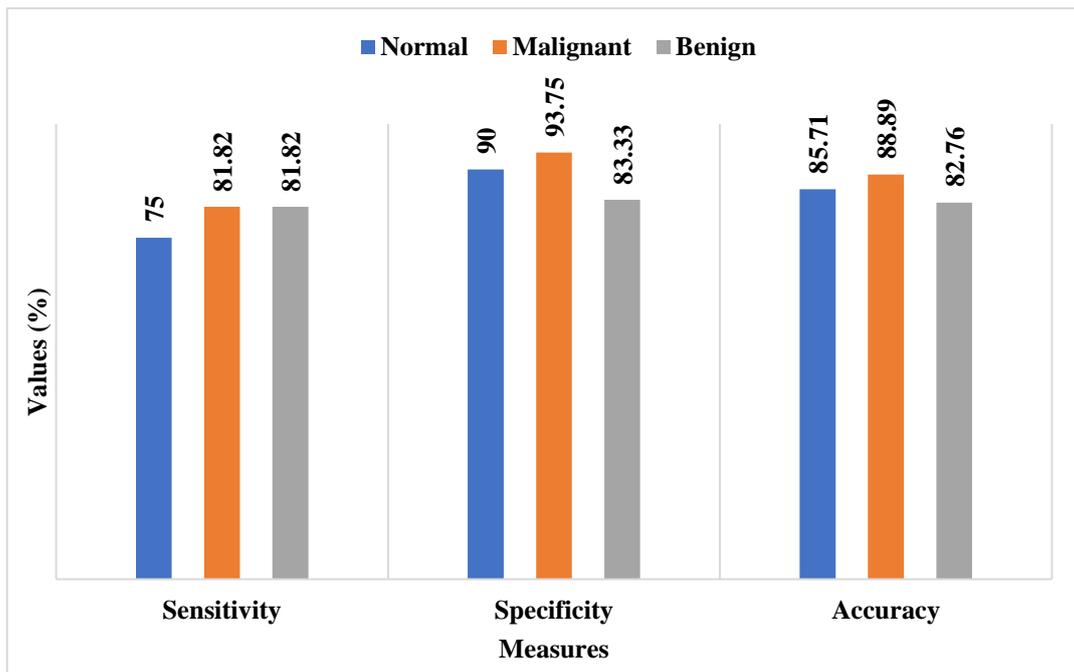| Category	Sensitivity	Specificity	Accuracy
Normal	75.00	90.00	85.71
Malignant	81.82	93.75	88.89
Benign	81.82	83.33	82.76

Fig. 5. Results of Proposed HFEM+DNN

Table 4 and Fig. 6 provide the comparative analysis of proposed and traditional models by means of distinct measures.During the performance assessment interms of accuracy, it can be seen that the CoKNN model has leads to worse outcome by offering a minimal accuracy of 74.60%. At the same time, it is shown that the WKNN model has provided slightly better

results with the accuracy of 76%. Equally, the QSVM model has offered better outcomes with an accuracy of 76%. Also, the linear model has outperformed the earlier model by achieving an accuracy of 77%. But, the MLP model has tried to manage well and ended up with an accuracy of 82%. Along with that, the RBF model has provided acceptable performance and leads to a competitive accuracy of 84% whereas the proposed model has offered supreme outcome with the highest accuracy of 85.79%.

During the performance validation with respect to sensitivity, it is shown that the QSVM model has provides pooroutcome by a minimal sensitivity of 72%. Simultaneously, it is shown that the CoKNN model has offered slightly better results with the sensitivity of 73%. Likewise, the WKNN model has provided better results with the sensitivity of 74%. In addition, the MLP model has outperformed the previous methods by reaching a sensitivity of 77%. However, the HFEM+DNN model has attempted to manage well and ended up with a sensitivity of 79.55%. On the same way, the RBF model has given reasonable function and leads to a competitive sensitivity of 86% while the linear model has offered supreme outcome with the highest sensitivity of 89%.

Table 4 Comparative analysis of different models

Methods	Accuracy	Sensitivity	Specificity
HFEM+DNN	85.79	79.55	89.03
MLP	82.00	77.00	72.00
RBF	84.00	86.00	54.00
Linear	77.00	89.00	36.00
QSVM	76.00	72.00	89.00
CoKNN	74.60	73.00	80.00
WKNN	75.80	74.00	84.00

While assessing the results with respect to specificity, it is implemented that the linear model tends to worse outcome by offering a lower specificity of 36%. Meanwhile, it is exhibited that the MLP model has given slightly better results with the specificity of 54%. Additionally, the MLP approach performs better when compared with traditional approaches by reaching a

specificity of 72%. Hence, the CoKNN techniques has tried to balance well and concluded with a specificity of 80%.
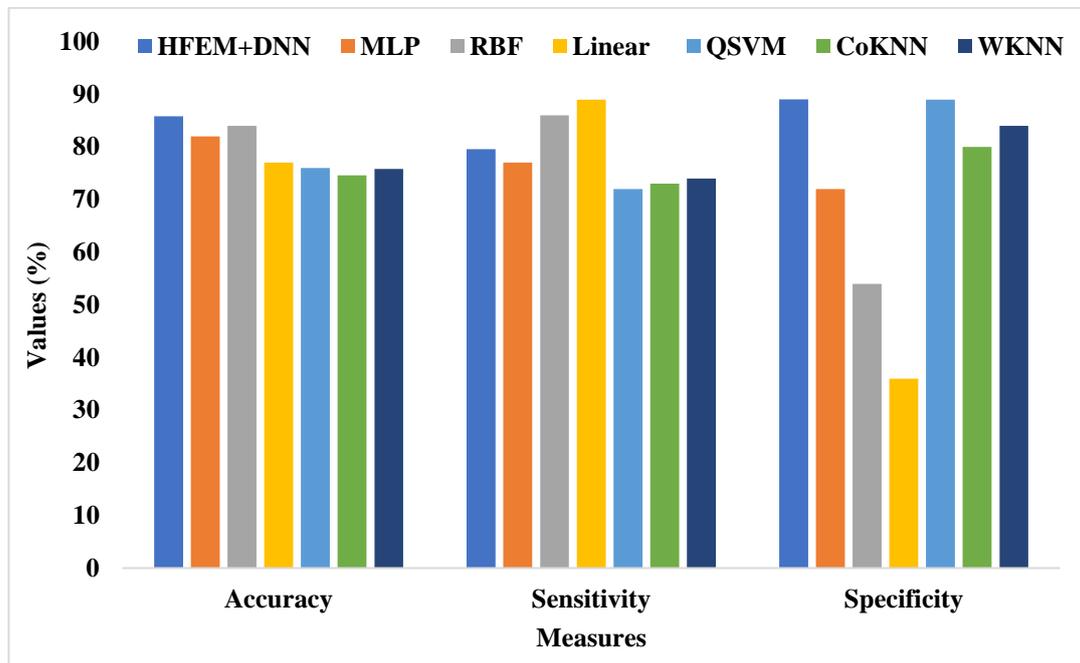

Fig. 6. Comparison of HFE-DNN with existing models

At the same time, the WKNN model has given manageable performance and provides a competing specificity of 84%. In line with this, the QSVM model slightly moderate result with specificity of 89% while the proposed system has attained optimized result with the maximum specificity of 89.03%.

4. Conclusion

This paper has introduced an effective feature extraction based lung cancer classification model called HFE-DNN model. The proposed model operates on three major processes namely preprocessing, feature extraction and classification. Initially, the images are preprocessed and the image quality gets improvised. Then, feature extraction process takes place using HFE technique followed by DNN based data classification. The outcome of the HFE-DNN technique can help doctors in the diagnostic process of lung cancer. An extensive simulation results takes place on the benchmark HFE-DNN model and the results confirmed its superior characteristics over the existing methods. The simulation outcome pointed out that the proposed HFE-DNN model has offered effective classification outcome with the maximum accuracy of 85.79%, sensitivity of 79.55% and specificity of 89.03%. As a part of

future scope, the performance of the HFE-DNN model can be enhanced by the use of segmentation techniques.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1]   J. Kuruvilla, K. Gunavathi, Lung cancer classification using neural networks for CT images, Comput. Methods Programs Biomed. 113 (2014) 202–209.

[2]   R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2019, CA Cancer J. Clin. 69 (2019) 7–34.

[3]   S. Lakshmanaprabu, S.N. Mohanty, K. Shankar, N. Arunkumar, G. Ramirez, Optimal deep learning model for classification of lung cancer on ct images, Fut. Gen. Comput. Syst. 92 (2019) 374–382.

[4]   D.M. Parkin, Global cancer statistics in the year 2000, Lancet Oncology 2 (2001) 533–543.

[5]   A.A. Abdulla, S.M. Shaharum, Lung cancer cell classification method using artificial neural network, Information Engineering Letters 2 (March) (2012) 50–58.

[6]   N. Camarlinghi, et al., Combination of computer-aided detection algorithms for automatic lung nodule identification, International Journal of Computer Assisted Radiology and Surgery 7 (2012) 455–464.

[7]   O.S. Al-Kadi, D. Watson, Texture analysis of aggressive and non aggressive lung tumor CE CT images, IEEE Transactions on Biomedical Engineering 55 (2008) 1822–1830.

[8]   B. van Ginneken, et al., Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study, Medical Image Analysis 14 (2010) 707–722.

[9]   R. Bellotti, D. Cascio, et al., A CAD system for nodule detection in low-dose lung CTs based on region growing and a new active contour model, International Journal of Medical Physics Research and Practice 34 (2007) 4901–4911.

[10]   S. Naqi, M. Sharif, M. Yasmin, S.L. Fernandes, Lung nodule detection using polygon approximation and hybrid features from CT images, Curr. Med. Imaging Rev. 14 (2018) 108–117.

[11]   C.-H. Chen, C.-K. Chang, C.-Y. Tu, W.-C. Liao, B.-R. Wu, K.-T. Chou, et al., Radiomic features analysis in computed tomography images of lung nodule classification, PLoS One 13 (2018) e0192002.

[12]   Han, Y., Ma, Y., Wu, Z. et al. Histologic subtype classification of non-small cell lung cancer using PET/CT images. Eur J Nucl Med Mol Imaging (2020). https://doi.org/10.1007/s00259-020-04771-5

[13]   D. Kumar, A. Wong, D.A. Clausi, Lung nodule classification using deep features in CT images, in: 2015 12th Conference on Computer and Robot Vision, 2015, pp. 133–138.

[14]   Revathi, M., Jeya, I.J.S. & Deepa, S.N. Deep learning-based soft computing model for image classification application. Soft Comput (2020).

[15]   N. Nasrullah, J. Sang, M.S. Alam, H. Xiang, Automated detection and classifi- cation for early stage lung cancer on CT images using deep learning, Pattern Recognition and Tracking XXX, 2019.

[16]   G. Zhang, Z. Yang, L. Gong, S. Jiang, L. Wang, Classification of benign and malignant lung nodules from CT images based on hybrid features, Phys. Med. Biol. (2019).

[17]   I. Jasmine SelvakumariJeya, and Suganthi, J., 2015. RONI based secured and authenticated indexing of lung CT images. *Computational and Mathematical Methods in Medicine*, *2015*.

[18]   I. Jasmine SelvakumariJeya, and S.N, Deepa, 2016. Lung cancer classification employing proposed real coded genetic algorithm based radial basis function neural network classifier. *Computational and mathematical methods in medicine*, *2016*.

[19]   Valluru, D., Jeya, I.J.S. IoT with cloud based lung cancer diagnosis model using optimal support vector machine. Health Care Management Science (2019). https://doi.org/10.1007/s10729-019-09489-x

[20]   Revathi, M., Jeya, I.J.S. & Deepa, S.N. Deep learning-based soft computing model for image classification application. Soft Comput 24, 18411–18430 (2020).

[21]   http://www.via.cornell.edu/lungdb.html