Lung Cancer Detection using Bio-Inspired Algorithm in CT Scans and Secure Data Transmission through IoT Cloud

C.Venkatesh¹
Research Scholar, Department of ECE
Koneru Lakshmaiah Education Foundation
Green Fields, Vaddeswaram, Guntur-522502
Andhra Pradesh, India

Polaiah Bojjia²
Professor, Department of ECE
Koneru Lakshmaiah Education Foundation
Green Fields, Vaddeswaram, Guntur-522502
Andhra Pradesh, India

Abstract—Primary recognition of pulmonary cancer nodules eloquently increases the odds of survival, also leads to solid problem to resolve, as it often relies on a tomography scan filmic examination. By increasing the possibility of effective treatment, earlier tumor diagnosis decreases lung cancer mortality. Radiologists usually diagnose lung cancer on medical images by a systematic analysis that consumes more time and is unreliable often, because of the substantial improvement in the transmission of data in the healthcare sector, the protection and integrity of medical data has been a huge problem for healthcare applications. This study utilizes computational intelligence techniques. For detection and data transmission, a novel Hybrid model is therefore proposed in this paper. Two steps are involved in the proposed method where diverse image processing procedures are used to detect cancer in the first step using MATLAB and data transfer to authorized persons via the IoT cloud in the second stage. The simulated steps include pre-processing, segmentation by Otsu thresholding along with swarm intelligence algorithm, extraction of features by local binary pattern and classification using the support vector machine (SVM). This work demonstrates the dominance of swarm-intelligent framework over the conventional algorithms in terms of performance metrics like sensitivity, accuracy and specificity as well as training time. The tests carried out show that the model built can achieve up to 92.96 percent sensitivity, 93.53 percent accuracy and 98.52 percent specificity.

Keywords—Pulmonary; mortality; carcinogenic; swarm intelligence; IoT

I. INTRODUCTION

A malicious tumor characterized by uninhibited cell evolution in lung tissues is lung cancer. Carcinomas are the majority of cancers that originate in the lungs. Most of the patients are diagnosed at an advanced stage due to no apparent early cancer symptoms [1], which typically results in high costs and a worse prognosis. In medical diagnosis and treatment, medical imagery has become important. These images playa extensive part in clinical applications since medical professionals expose attention in exploring the interior structure [2]. Several procedures have been established based on cross-sectional images, such as magnetic resonance imaging (MRI) or computed tomography (CT) or other topographic modes [3,4,5]. The application of medical image processing has played an important role in both technological and clinical aspects in helping to identify and examine anomalies by making it easier for medical practitioners to work with more scientific and sophisticated approaches to solve the problem [6]. A CT Scan obtains images of an organ that cannot be seen on a regular x-ray that results in earlier diagnosis [7]. The biggest issue with lung cancer is that these cases of cancer are later diagnosed, making treatments more complicated and decreasing the probability of survival subsequently [8]. It is therefore important to recognize a modern, robust method for diagnosing lung cancer at an earlier stage [9]. For cancer diagnosis, CT scan images are being used; they are analyzed by radiologists to recognize and identify nodules into malignant and benign nodules [10]. These techniques, require highly trained radiologists who are not in particular, accessible to people in remote regions. In addition, in manual testing, there seems to be a significant chance of human error, and therefore optimization-based systems are required that can assist radiologists in diagnosing and help minimize the incidence of false results [11]. To detect the nodules, their form, scale, and other characteristics from CT scans, digital image processing techniques can be used. In order to design specialist support systems for the diagnosis of various diseases such as lung cancer identification, medical image processing has been widely and rapidly implemented. In addition, the existence of nodules that define a patient's destiny is also very complex, as their shape and size differ from slice to slice. They are often connected, such as arteries or bronchioles, to other pulmonary structures [12]. It can also vary the color in which they appear on CT scans. These variables contribute to the difficulty of defining them.

In this work an efficient framework is proposed to decipher the lung cancer at an early stage and also data transmission to medical practitioners. Detection stage involves pre-processing, separation of nodules with optimization, feature extraction and classification. Transmission stage involves transmission of statistical parameters through IoT as well as MATLAB IoT cloud Thing speak. As direct data transmission is not possible, thing speak module has been considered for effective transmission. The structure of this paper contains Section II: related work, discusses about the
previous works, Section III proposed methodology represents the methods, block diagram and corresponding algorithms, Section IV shows segmentation with optimization concepts, Section V is the Extraction by LBP method, Section VI is Classification by SVM and Section VII presents the Simulation results, provides output images, statistical values and corresponding thingspeak plots.

II. RELATED WORK

Malayil Shanid et al. [13] in 2020 presented a pulmonary cancer detection system with SE (slap elephant) optimization and deep learning techniques. By this work authors gained 96 percent accuracy.

Noor Khehrab et al. [14] in 2020 presented a pulmonary nodule detection system with thresholding and statistical features techniques. By this work authors gained 93.75 percent sensitivity.

Shankar et al. [15] in 2019 documented an Alzheimer's identification technique that uses the gray-level run-length matrix and scale-invariant conversion to extract different features. By this framework 96.23 percent accuracy is gained.

K.Senthil Kumar et al. [16] in 2019 recognized a lung cancer detection scheme by GCPSO. By this model 95 percent meticulousness is acquired.

C.Venkatesh et al. [17] in 2019 projected a detection scheme by genetic approach. By this approach 90 percent precession.

Vijh et al. [18] in 2019 proposed a detection procedure using whale optimization algorithm and SVM. By this work authors gained 95 percent accuracy.

Preethijo et al. [19] in 2019 projected a respiratory cancer recognition strategy with the SVM classifier using fuzzy c & k-mean partition methodologies. By this model less than 93 percent accuracy is gained.

S.Perumal et al. [20] in 2018 documented an enhanced ABC optimization for cancer detection and classification. 92 percent proficiency is accomplished by using this procedure.

Uc-ar et al. [21] in 2019 recommended a detection model by Laplacian and Gaussian filter model with CNN architecture. In this method 72.97 percent precision is attained.

In all the above conventional (existing) techniques, the accuracy is lower. In this paper, therefore, an assorted approach is projected where PSO has been used for segmentation to obtain greater accuracy along with SVM classifier and LBP for feature extraction.

III. PROPOSED METHODOLOGY

Fig. 1 shows a detailed view of the proposed system where it involves two phases. In first phase lung cancer is identified from CT images using the optimization method of swarm intelligence. In second phase the data transferred through thingspeak and IoT. Initially, the CT input images of lung cancer are read from private and public databases. The attained CT images typically encompass a noise [22]. In pre-processing step by the use of median filter the noise is condensed. Then, output image of filter is segmented by swarm optimization with the Otsu thresholding technique. The partitioned image then endures an extraction process by LBP to excerpt textural topographies. Then, the extracted topographies fed to classification stage to detect whether image is normal or abnormal. If the image is anomalous the attributes are determined and transferred to the medical persons via Thingspeak or IoT.

A. Pre-processing

Optimal reliability inspection is improved by image pre-processing. All images probably contain noise, so the image has to be pre-processed by median filtering to suppress the noise [23]. It improves the aesthetic value and accuracy of the image.

B. Median Filtering

This filter reduces the noise of salt and pepper and also retains the image edges. The random bit error in a communication channel generates salt and pepper noise. The median filter is a basic regional sliding kernel that swaps the kernel's centre point with the kernel's average of all the pixels [24].

Fig. 1. Proposed Framework.
IV. SEGMENTATION WITH OPTIMIZATION

A. Segmentation

Segmentation tends to slice the image pixels into sections that are directly connected to the objects of the image. Typically, it is the basic step for all the computer vision systems [25]. Usually, segmentation algorithm relies on pixel intensities. All the algorithms entail certain threshold parameter to be set. The appropriate threshold results in greater segmentation. The threshold value is set bestowing to the intensity values [26]. To achieve best threshold value in this work otsu thresholding techniques is used.

B. Otsu Thresholding

Otsu thresholding is based on the basic idea of identifying a threshold that mitigates the weighted variance in the class, which is the same as optimizing the variance among classes [27]. It works directly on the gray-level bimodal histograms. Also no other entity structure description and regional continuity is required. It has set numbers, but can be modified to suit locally [28].

C. Algorithm

1. Estimate Histogram and likelihood of every intensity value
2. Initialize xi(0) and μi(0)
3. Move across all feasible points t=1 Upgrade wi and μ
4. Determine σb2(t) and considered it as preferred threshold
5. Measure two maxima’s of σ12 is the higher limit and the value limit is greater or equal to σ22
6. optimal threshold = (Th1+Th2)/2

D. Particle Swarm Optimization

It is a metaheuristic process used effectively in the study of medical images [29]. It mimics the communal movements of food-seeking birds [30]. Because of simplicity and generality, this algorithm has been effectively used for cancer detection. PSO falls quickly, however, into the local optimal solution. The argument and alliance of information is a good basic principle of PSO. In this process every particle has a preliminary position and velocity [31]. Each particle's position signifies a probable solution and has a fitness value calculated by its fitness function. The position and speed are altered based on the fitness value and gets updated. After updating with a group of random particles, the procedure pursuits for optima. The equations to update position and speed are as follows [32]:

\[ S(t + 1) = S(t) + a_1 c_1 \left[ P_{\text{best}}(t) - m(t) \right] + a_2 c_2 \left[ G_{\text{best}}(t) - m(t) \right] \]

\[ m(t+1)=m(t)+V(t+1) \]

Where \( c_1 \), \( c_2 \) are random values and \( a_1 \), \( a_2 \) are acceleration values

The optimistic concert relies on the fitness function. The equation of fitness function is given below:

\[ f = \sum_{k=1}^{n} \frac{\text{interband distance}}{\text{intraband distance}} \]

Where \( k \) is the number of bands

The weight of inertia enables global searching which enhances the rate of convergence and also reduces the iterations, while a small weight of inertia enables local searching [1].

\[ S(t + 1) = W \ast S(t) + a_1 c_1 \left[ P_{\text{best}}(t) - m(t) \right] + a_2 c_2 \left[ G_{\text{best}}(t) - m(t) \right] \]

Where \( w \) is weight of inertia the values of constant and random inertia are 0.7 & 0.5+rad()/2 respectively.

Algorithm

1. Initialize with some random position and velocity vectors for the particles.
2. For all particles in the group calculate fitness value.
3. When fitness(p) is better compared to fitness(Pbest) and P_best = P.
4. Assign G_best as the best particle value.
5. Determine each particle speed velocity is measured for each particle.
6. Update speed and position of particles.

V. FEATURE EXTRACTION BY LOCAL BINARY PATTERN (LBP)

In diverse fields, the LBP method has been used. It is a texture description operator based on symbols of variances between central and adjacent pixels [33]. In this technique a binary cypber of every pixel is gained by thresholding its surrounding pixel with the centre pixel. If the value of adjacent pixel is superior or equal to threshold value it assigns 1 otherwise 0. First, to evaluate the frequency values of binary patterns, a histogram is constructed [34]. The likelihood of a binary pattern contained in the image is represented by texture characteristics. The equation of LBP is as follows:

\[ LBP(m_c,n_c) = \sum_{k=0}^{k-1} 2^k f(g_k - g(m_c,n_c)) \]

Where \( g(m_c,n_c) \) is the grey values of center and surrounding pixels, \( f(x) \) is the function whose value is 0 if \( g<0 \) and 1 if \( g\geq0 \). Finally, the LBP value is the center pixel \((m_c,n_c)\).

VI. CLASSIFICATION BY SVM

SVM was formerly used to categorize linearly detached data for binary classification. The preliminary purpose is to discover an optimum hyper plane [35,36]. The Hyper plane is a two-class frontier. It not separates two classes but also enhances the boundary between two classes. The boundary is the major distance in each class among the hyper plane and the nearest data [37]. The ideal boundary is attained by maximizing the distance between the support vector and the hyper plane. Let \( m=(x_1,x_2) \) and \( W(m,-1) \) then for each class hyper plane can be expressed as \( y(t) \) and the equations can be written as follows:

\[ y(t) = mx + n \]
\[ m x_1 - x_2 + n = 0 \]  \hfill (6)
\[ w. x + n = 0 \]  \hfill (7)

If hyperplane is defined then based on assumptions the hypothesis function can be written as follows:

\[ h_f(x_i) = \begin{cases} +1 & \text{if } w. x + n \geq 0 \\ -1 & \text{if } w. x + n < 0 \end{cases} \]  \hfill (8)

From the above equation if the point is above the plane then it is categorized as +1 otherwise -1. The data set used in the proposed method is as shown in Fig. 2.

VII. SIMULATION RESULTS

In this work, firstly by using MATLAB, the lung cancer is detected and secondly, the attributes attained are transferred to doctors using Thingspeak. The CT images are collected from private hospital in Ananthapuramu. In this approach the problem of thresholding is considered as an optimization issue and can be resolved by the particle swarm principle. In this work algorithms were implemented using MATLAB (R2017b) on an Intel Core i5 PC at 1.80Ghz with a total physical memory of 8GB RAM.

A. Detection Phase using MATLAB

In Fig. 3 and 4, the input and median filter output of CT lung cancer images are shown. The CT image commonly has noise with less falsification. The input image is processed into a median filter to abolish noise and falsification in the image.
To discern mutant in CT image, together with optimization technique, the filter output image is partitioned with Otsu thresholding. At first, the CT image is partitioned through simple Otsu thresholding in which the segmented classes, "all category variance," are greatly increased. By refining by particle swarm optimization, the results acquired from the thresholding method should be optimized.

In PSO, by following the maximum particles present, the potential solutions, called particles, pass through the space of the problem. In Fig. 5 and 6, the segmented output images are shown. Fig. 7 shows the classifier output. The features of image are extracted by feature extraction with LBP after classification.

1) Statistical Results of Existing Method: The traits attained from the proposed model method are shown in Table I. In the abovementioned table, the proposed method is proved as best to obtained the less MSE at 0.186 also high PSNR at 42.729 and high accuracy at 96.550% as compared to conventional systems.

B. Data Transmission

Finally, the obtained result is plotted as graph in the ThingSpeak and is shared to the authorized personnel. Fig. 8 to 13 shows the ThingSpeak plots which are shared to medical professionals.
Fig. 9. Specificity Data Plot of ThingSpeak.

Fig. 10. Mean Square Error Data Plot of ThingSpeak.

Fig. 11. PSNR Data Plot of ThingSpeak.

Fig. 12. Entropy Data Plot of ThingSpeak.

Fig. 13. Correlation Data Plot of ThingSpeak.

VIII. CONCLUSION

In this paper, a new strategy to early detection, prediction and diagnosis has been introduced in order to improve patient safety and mitigate the risks also the data is transferred to medical professionals through MATLAB IoT cloud called Thingspeak. The image pre-processing and segmentation procedures are used for partition the lung nodule along with particle swarm algorithm. The several features are extracted by LBP to the study of statistical information that assists in the decision-making process by SVM whether the tumour is malicious or non-malicious. The proposed approach outperforms by providing an accuracy of 96.5 percent. The interpretation of the obtained results are limited with accuracy and specificity parameters due to convergence of local optima in the algorithm. Further, these results can be analyzed more effectively using deep learning techniques and advanced hardware processors in near future.
REFERENCES


379 | Page