

# Classification of Normal and Abnormal Lung CT-scan Images Using Cellular Learning Automata

Nooshin Hadavi, Md Jan Nordin, Ali Shojaeipour and Mohammad Faidzul Nasrudin

Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia

## Article history

Received: 04-04-2019

Revised: 12-10-2019

Accepted: 14-10-2019

## Corresponding Authors

Nooshin Hadavi

Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia

Email: Nooshin.Hadavi@gmail.com

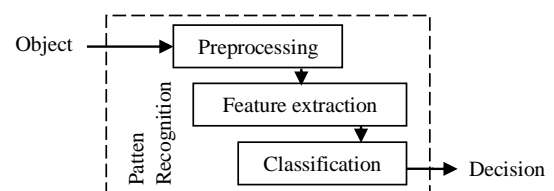
**Abstract:** This paper proposes a medical pattern recognition system based on the Cellular automata (CA). CA or cellular machine is a dynamic mathematical model that consists of several similar and simple units organized by considerably simple local rules. Each cell acts as a simple computer automaton. This can lead to the implementation of the complex computations through uncomplicated methods. However, the CA model needs to determine certain rules for specific use and this model is regarded as suitable for modelling certain systems. To overcome this problem, a method is needed through which the favorable rules are extracted. Cellular Learning Automata (CLA) model is obtained from developing CA by appending a Learning Automaton (LA) to each cell. Many applications of CA are known today, especially in the field of pattern recognition. Therefore in this study, we use the CLA to design an automatic system to diagnosis the images which contain cancer tissue. Hence in this study, after applying the required approaches on lung Computed Tomography (CT) images, images are classified through the CLA model and the proposed methods are evaluated in terms of sensitivity, specificity and accuracy. The proposed system promises a flexible and low complexity model. The method has been tested on 22 slices of CT scan images from a real-world dataset and has yielded satisfactory results. The model with a low error rate (0.09), yielded a favorable accuracy (95.4%).

**Keywords:** Computer-Aided Diagnosis, Image Processing, Pattern Recognition, Classification, Cellular Learning Automata, Lung Cancer

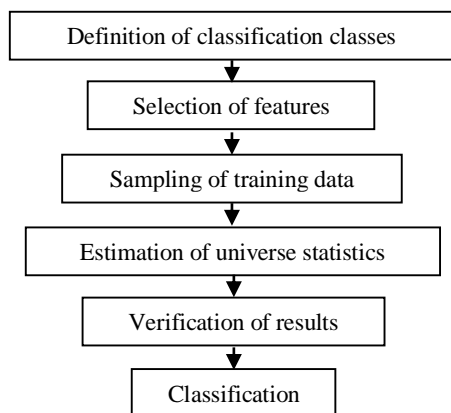
## Introduction

Pattern recognition is the study that identifies an object through a signal or image, based on its features which are obtained from a set of related data. In the pattern recognition approach, the extracted information from each image is contrasted into an explicit structural model of the desired object and recognition is achieved from the analysis of the images. The image or video will be searched to find a model based on the extracted information. Recently, the pattern recognition approach has become a topic of interest in the study of medical diagnosis and employing the technique has added significant contributions to the advancement of this field. Hence, several medical pattern recognition systems have been suggested so far.

The operation of pattern recognition is similar to the classification and the final aim of pattern recognition is to extract the patterns and classify them so that optimum results can be obtained. A pattern recognition system is formed based on basic steps including the pre-processing, feature extraction and selection, as well as classifier design and optimization as illustrated in Fig. 1.



**Fig. 1:** Overview of a pattern recognition system



**Fig. 2:** Procedures of classification

The pattern recognition (PR) system needs an efficient, precise and highly reliable classifier. Therefore, if the methods are wisely selected, it can increase the detection and classification ability of the system. Assigning each instance to a class is called classification which is a problem for the PR approach (Jamshidnezhad and Nordin, 2012). The pattern recognition system gives a class label as an output, for example, “1” or “0” are classes label in a quality control test of a product (Liu *et al.*, 2006). An image includes the quantitative spectral information which is used by the image classifier in order to discriminate the images from each other. The interpretation technique of visual images is based on size, shape, pattern, texture, shadow, tone and association (Kahaki *et al.*, 2017). Classification involves the steps illustrated in Fig. 2.

In general, digital image classification techniques can be divided into two main categories: Supervised and unsupervised classification. In order to perform image classification, the images need to be analyzed and suitable decision rules should be extracted. In the supervised classification technique, the classes are determined and representative samples of the pixels are identified by the computer. When there is less information about certain object in some cases (images), in order to perform classification, the unsupervised classification techniques are utilized (Parashar and Harish Kundra, 2014). Most of the classifiers are supervised and need to be trained with labeled data. Hence to design a supervised classifier, it is imperative to prepare the training data and for this purpose we should obtain some objects such as: Enhancing the images, segment the region of interest, select the aim part according to the extracted features, prepare it for training the classifier and finally, training the classifier (Ahmed and Nordin, 2011).

## Background and Related Works

Pattern Recognition (PR) is the implementation of computer algorithms in order to discover the regularities

in a dataset. The regularities are then utilized to perform some proceeding, such as classifying the data into separate categories (Bishop, 2006). Basically, in the PR approach, a sample should be recognized as the criteria of measurements. There are several PR techniques that are widely used and each method has its relative benefits in various operational situations (e.g., Neural Networks (Raj *et al.*, 2016), PCA (Segreto *et al.*, 2014), Fuzzy sets (Mitra and Pal, 2005).

In order to carry out the PR in the medical field, special medical knowledge and experiences are required. The type of required information is often not fixed and constant and some have relatively low-level features such as texture, shape and other pixels based on the statics extracted from the images and utilized for recognition (Homma, 2009). The PR approach based on medical images is generally conducted by extracting the diseased tissue from the subject and then analyzing it. The step that follows is the detection of some regulation that is used to classify the images into the correct class which will be used to diagnosis. Automatic systems that are designed for medical diagnosis through the integration of computers and medical science are called Computer Aided Diagnosis (CAD).

Cellular Automata (CA) or cellular machine is a dynamic mathematical model, the model concepts were initially introduced by Neumann in 1966 (Wongthanavasu and Ponkaew, 2013). By following the cellular automata studied by Ulam, Conway (introduced the Game of life) and Wolfram (studied and presented the Wolfram rules) (Sarkar, 2000; Shiraishi *et al.*, 2011). The main purpose of this model is to design a self-reproductive system which is computationally complete (Rennard, 2000). This model consists of cell lattices where each cell has a set of stats and work based on the local rules. The CA model works in discrete time and the cells states are updated according to local rules. In time  $t+1$ , the state of a cell is determined based on the previous state of the cell and its neighbors in time ‘ $t$ ’ (Abu Dalhoum and Al-Dhamari, 2010). Three types of cellular automata have been introduced so far; one, two and three-dimensional cellular automata (Hadavi *et al.*, 2014; Liu *et al.*, 2006). For this study, we used the two dimensional (2D) CA. The 2D-CA model can be shown by a 4-tuple,  $(A, S, N, F)$  where  $A$  is a set of cells,  $S$  shows a set of possible states for the cells,  $N$  represents the neighborhood position and  $F$  is a function that obtains the new state for the cell (Fathy Navid and Bagheri Aghababa, 2013):

$$F : S | N | \rightarrow S \quad (1)$$

Although each cell of the CA model works using simple rules, by local interaction, the cells lattice can

perform complex computations in less complex approaches, which is one of CA's advantages. It should be noted that as a problem, the CA model needs a certain form of rules and these rules should be carefully determined (Rosin, 2006). As a solution to this issue, an approach which is able to automatically generate the rules is required. Making the CA intelligent and improving them by adding the learning capability is one of these methods. In CA, the state of a cell is defined by a rule, according to the dependence of the neighborhood of the cell. The activities scope of the cells is defined by the neighborhood. Therefore, the type and radius of the neighborhood can impress the simulation results. Figure 3 illustrates some of the most common neighborhoods for 2D-CA (Janssen, 2010).

For calculating the new state of a cell, it is required to know how can be referenced each cell's neighbors. For one-dimensional CA, it is simple as the cell that is indexed with (*i*) has two neighbors, which are (*i*-1) and (*i*+1). In the 2D-CA, the cells do not have a single index; they indexed by (*x*, *y*), according to the rows and

columns (Shiffman, 2012). The neighbors of a cell (*x*, *y*) are shown in Fig. 4.

As previously mentioned, the next state (*t*+1) depends on the state (*t*) of the cell and its neighbors. More neighborhood is given by:

$$S_{i,j}(t+1) = f(S_{i-1,j-1}(t) \oplus S_{i-1,j}(t) \oplus S_{i-1,j+1}(t) \oplus S_{i,j-1}(t) \oplus S_{i,j}(t) \oplus S_{i,j+1}(t) \oplus S_{i+1,j-1}(t) \oplus S_{i+1,j}(t) \oplus S_{i+1,j+1}(t)) \quad (2)$$

Logical Exclusive-OR operation is represented by  $\oplus$  in Equation (2).

The operation of the Cellular Automata (CA) on the medical images was surveyed by Wongthanavas (2011). The CA techniques appeared as a natural tool for image processing. CA model has several advantages; it has a local nature and the computing implementation is parallel and straightforward. The author expressed in his report stated that CA has resulted in favorable medical image processing. In this study, several uniform CA algorithms have been applied to various mammography images.

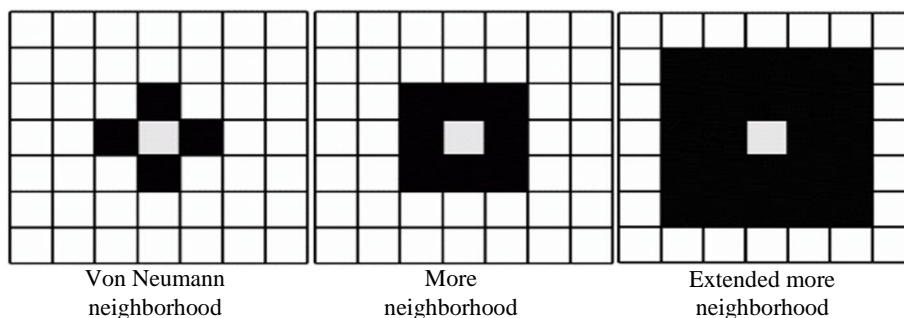


Fig. 3: Some cellular automata neighborhood samples

	<i>x</i> -1	<i>x</i>	<i>x</i> +1
<i>y</i> -1	<i>x</i> -1, <i>y</i> -1	<i>x</i> , <i>y</i> -1	<i>x</i> +1, <i>y</i> -1
<i>y</i>	<i>x</i> -1, <i>y</i>	( <i>x</i> , <i>y</i> )	<i>x</i> +1, <i>y</i>
<i>y</i> +1	<i>x</i> -1, <i>y</i> +1	<i>x</i> , <i>y</i> +1	<i>x</i> +1, <i>y</i> +1

Fig. 4: The neighborhoods of a cell (*x*, *y*)

The use of a group of classifiers can increase the classification accuracy in comparison to using only single classifier. Mostly, these groups of classifiers are constructed using various methods, where a specific issue would be assigned by the user. To overcome this problem, a group of classifiers which are constructed based on the CA is proposed using a self-organizing system (Kokol *et al.*, 2004). A purpose of their study was to create a combination of several classifiers in a group which would be represented by the CA. They used the self-organizing ability of CA to combine the classifiers using an ensemble approach and various classifiers are assigned to each cell. The simple Classifier Cellular Automata (CCA) is compared with various decision tree approaches (greedy, genetically and boosting) and the results indicated the CCA gives the best results in the case of medical data in comparison with other classifiers. Povalej *et al.* (2005) also used the CA to solve a classification problem in another study by combining the various classifiers. In this system, the successful information in the classifiers is stored in the corresponding cells. The energy of each cell is determined based on the transition rules. Based on these transition rules, the neighborhood cells will engage in some interaction to execute the learning processing.

The CA model is a simple distributed system that uses parallel architecture advantage, where each cell is considered a simple computer machine. Abu Dalhoum and Al-Dhamari (2010) in their study proposed a CA model for use in the neuroimaging field, particularly to classify functional Magnetic Resonance Imaging (fMRI) of the brain. In their study, two classifiers which are the Support Vector Machine (SVM) and CA have been applied on the data and the results were compared. The results demonstrated that the CA classifier obtained better accuracy and sensitivity compared to the SVM classifier.

By using the classification approach that is based on the relation between attributes, some suitable rules are extracted to classify the data. As an important point in machine learning, it is worthy to note that the process of learning leads to the improvement of the model that enables it to learn the training problem efficiently and provide a solution.

Therefore, a new classification method based on cellular learning automata as a data classifier which involved three steps was presented (Esmailpour *et al.*, 2012). The first step utilizes a complete graph where the pattern was extracted. In the next step of the CLA model, a strong pattern (the most repeated pattern) was selected. In the final step, the rate of compliance was calculated according to the value of the reward and penalty to test the model. After applying the suggested method on some of the UCI Machine

Learning datasets, it was compared with previously performed methods such as Naïve Bayes (NB), Decision tree (C4.5) and Multilayer Perception (MLPN) and observations were made regarding the time of training and classification accuracy. In terms of training time, the CLA managed to obtain the second rank. The use of the programming language and techniques are very effective in the overall time of training, while accuracy is influenced by the suggested algorithm. The evaluation results showed that the suggested model obtained a favorable yield in both accuracy and training time. It is also simple to implement and has low computational complexity.

A computer-aided diagnosis system was proposed by Chen *et al.* (2008) to distinguish between cancerous cells and normal cells based on the cellular automata and learning methods. In their study, the differences between the cancerous cells and normal cells were detected by utilizing the pattern recognition techniques and cellular automata with evolutionary learning. Making a system that is able to detect features independently was the main aim of the mentioned study. The study conducted a computer analysis on the microscopic images of the cell and consisted of three steps. The first step was the recognition of the cells borders along with the segmentation. The next step involved extracting the features. The last step was the classification. Initially, for detection of the cell boundaries, the images noises were removed by applying *Photo Impact* (an image editing tool). After the pre-processing step, the format of the saved images was bitmapped and then the images were saved in the ASCII format. Initially, according to the obtained cancerous cells patterns, some Feature Detectors (FDs) were created. The generated FDs were compared with the obtained cell patterns to establish whether the pattern of a cell matched any system FDs; this cell was labelled as a cancerous cell and if no pattern was established, it would be labeled as a normal cell. An FD will be known as a favorable FD when it conforms to a minimum of one cancerous pattern and does not have conformity to any normal cell. This will allow the CA model to improve the performance of lesser-performing FDs. The achieved results indicated that the suggested system had a high ability to distinguish between the normal and cancerous cells.

Cellular Automata model only can accept the reinforcement signal as input and is not able to receive another input such as the state of the environment. To solve this drawback, Ahangaran *et al.* (2017) proposed a new model of CLA, in which an extra input is entered to each cell, in addition to the reinforcement signal. This extra input is the information from the environment. It can improve the computational power and flexibility of the model. It evaluated the proposed model for classification,

clustering and image segmentation. Their model provided acceptable result on various data. In terms of the performance of the model in the classification task, the CLA model managed to achieve the second level after SVM by an average accuracy of 84%.

## Materials and Methods

Automatic Lung cancer detection is considered as a pattern recognition problem, in which the final step of the data processing is data classification. The data are the lung CT scan images of humans. As mentioned earlier, a PR system involves three basic steps: Pre-processing, feature selection and extraction and classifier design. Since the current study investigates a PR problem, the structure of the proposed system is based on three main parts. These three parts are referred to as the pre-processing, feature extraction and cancerous nodule detection and training and recognition. The overview of the proposed model is illustrated in Fig. 5.

### Data

This study was implemented to images from the Parsian medical imaging center lung CT database. Two sets of lung CT scan images were needed. One of them were used as the training set and another one was required for validation. This database contained 73 CT images with the .dcm format in which 70% of the dataset (51 images) were used for training and 30% images (22 images) were used for testing the proposed technique.

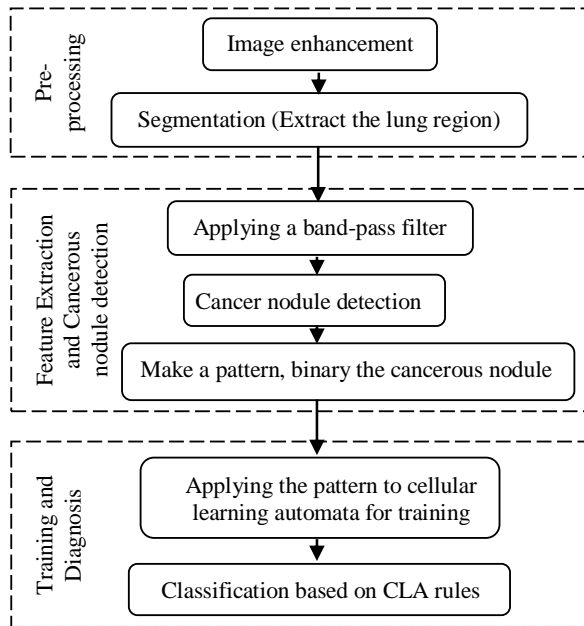


Fig. 5: Overview of the suggested system

### Pre-Processing

The pre-processing results included a compact representation of the pattern. In the current study, the pre-processing step was divided into two parts: Image enhancement and image segmentation. Most of the images include some undesirable details such as blurring and noise, which should be removed, so the image enhancement methods were applied to the images to enhance the important features. Compared to various enhancing methods, the Gabor filter had a good effect on the data images. Gabor filter is a band-pass filter containing both orientation-selective and frequency-selective properties and comprises optimal joint resolution in both frequency and special domain (Wang *et al.*, 2005). Hence, Gabor was appropriate for removing the noise and enhancing the medical images. A Gabor function is shown as the following:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{\tilde{x}^2/\tilde{\sigma}^2 + \tilde{y}^2/\tilde{\sigma}^2}{2}\right] \cos(2\pi\omega\tilde{x}) \quad (3)$$

Where:

$$\begin{aligned} \tilde{x} &= x \cos\theta + y \sin\theta \\ \tilde{y} &= -x \sin\theta + y \cos\theta \end{aligned} \quad (4)$$

where the orientation of Gabor filters is depicted by  $\theta(\theta \in [0, \pi])$  and the variance of Gaussian that envelop along the  $x$  and  $y$  axes are  $\sigma_x$  and  $\sigma_y$  respectively, the sine radian frequency is shown by  $\omega$ . The Gaussian envelop in the following:

$$G(u, v) = \exp\left[-\left(\frac{(u - \omega)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right)/2\right] \quad (5)$$

Where:

$$\sigma_u = 1/2\pi\sigma_x, \sigma_v = 1/2\pi\sigma_y \quad (6)$$

Wang and Sun (2010).

In the next part of the pre-processing step, to focus on the Region of Interest (ROI), the segmentation method was performed. The main goal of segmentation is to change the representation of the image, which makes it easy to analyze (Sharma and Jindal, 2011). Thresholding is one of the famous segmentation methods, which is quick, easy and has low space keeping. As its name implies, this method works based on one or several assigned thresholds.

Another notable method is the region growing. The region growing method consists of simple concepts, multi-criteria that can be selected at once time and the border obtained by this method is very accurate and thin and it is able to distinguish the regions with similar attributes (Kamdi and Krishna, 2012; Saliba and Dipanda, 2013). The

region growing segmentation approach is based on the threshold, which starts from the determined seed point and checks all the neighboring pixels according to the similarity with seed point; the region will frequently grow if the similarity range of the neighbor is sufficiently high. If the difference of the newly discovered region is more than the threshold (maximum intensity distance = 0.92), the process would be stopped. The basic formulation of this method is:

- (a)  $\bigcup_{i=1}^n R_i = R$
- (b)  $R_i$  is a connected region,  $i = 1, 2, \dots, n$
- (c)  $R_i \cap R_j = \emptyset$  for all  $i = 1, 2, \dots, n$
- (d)  $P(R_i) = \text{True}$  for  $i = 1, 2, \dots, n$ .
- (e)  $P(R_i \cap R_j) = \text{False}$  for any adjacent region  $R_i$  and  $R_j$

Wherein:

- (a) During the segmentation process, every pixel must be in a region.
- (b) Desired points should be connected.
- (c) The regions should be separated.
- (d) Shows the requirement for optimum coordination attributes of pixels in a segmented region. If all  $R_i$  pixels in have the same gray level  $P(R_i) = \text{True}$ .
- (e) Shows in the sense of predicate P, the region  $R_i$  and  $R_j$  are different

### Features Extraction

There are many possible features that have been investigated and used as characters. Two major categories can be assumed for pattern recognition features which are structural and statistical features. Several appearances based on statistical features were suggested by the researchers, but that required prior modeling assumptions that would enable the system to overcome the inaccuracy and inability. This is the main disadvantage of statistical pattern recognition approaches (Gaikwad *et al.*, 2010). Structural pattern recognition is more flexible than statistical pattern recognition (Spillmann *et al.*, 2006). Structural pattern recognition has been used and was successful in various applications such as digit recognition (Tuba *et al.*, 2016), shape classification (Chen *et al.*, 2009; Ramesh *et al.*, 2015) and bioinformatics (Marchiori, 2013). Most feature extraction methods are supervised; this means that they need to have prior knowledge about the pattern and predefined training samples. Feature selection is the process of selecting the best subset of the input space in a way that concludes optimum results (Kahaki *et al.*, 2016). After selecting the optimum subset, a classifier should be designed.

Lung cancer diagnosis is one of the many applications of the PR approach. As previously mentioned, PR works as a classification process. The goal of PR applications is extracting patterns optimally and separating the classes. For this purpose, accurate and

principle features need to be extracted. There are several methods for the feature extraction and classification stage that help researchers achieve better results. In this phase, a frequent domain filter is applied to images based on texture features. Since the image consists of three main layers (down or background, up or surface and median or content) using the high-pass filter, low-pass filter and the concurrent implementation of both of them and subtracting the yielded images from the original images, a band-pass filter is specifically designed for the detection of a cancer nodule. Therefore, by utilizing a band-pass filter in the frequent domain, the cancer nodule is detected and then it is imported to a cellular lattice as a pattern instead of some statistical features (size, contrast). The ideal band-pass filter is:

$$H(u, v) = \begin{cases} 1 & \text{if } D_L \leq D(u, v) \leq D_H \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where,  $D(u, v)$  represents the distance between the frequent center and point  $(u, v)$  and the low frequency bound and high bound are shown by  $D_L$  and  $D_H$  respectively ( $D_L = 60, D_H = 120$ ).

### Training and Diagnosis

After the images passed all of the phases above, the cancer nodules would be identified. According to previous entries, we intended to utilize the achieved images from prior phases as a pattern for training the classifier. Therefore, for generating a pattern of cancer nodules, the gained images were converted into a binary format. To train the CA, we used the CA model which has been improved by the learning automat, so- also known as the *learning automata (CLA)*

To solve the training problem of the CA model, a new approach was proposed, which involved allocating a learning automaton to each cell. This approach led to the development of the CA to become a powerful and effective model for many problems, which are decentralized (Esmailpour *et al.*, 2012). Adjusting the state of the transition probability for the CAs is the basic concept behind the cellular learning automata model. The CLA model is designed for systems where their components worked based on communication with other components and use each other's experiences to improve their behavior. The CLA model can be defined as a 6-tuple (S, V, N, R, L, and C), where S represents a finite set of possible states for a cell and the assigned value cell shown by V in time t:

$$s = \{s_1, s_2, \dots, s_n\} \quad v = \{v_1, v_2, \dots, v_k\} \quad (8)$$

Value of  $s_i$  in time t is shown by  $v'(s_i)$ .

N is the cell neighborhood set; R describes the supervisor's rules which control the assigning

rewards-penalty the cells. Cellular learning is shown by L and this model needs a function that is displayed by C for the purpose of dividing the reward to the lattices cells. Therefore:

$$\text{If } W(s_i) = N(s_i) \cup \{s_i\}, r = r(W(s_i)) r \in R$$

$s_i$ : Selected action in step  $k$ :

$$\text{sum} = \sum \forall w(s_i) \exists s_i \in w(s_i) cw((s_i); s_i) \quad (9)$$

$$r(s_i) = \begin{cases} \text{reward} & \text{sum} \geq \theta \\ \text{penalty} & \text{sum} < \theta \end{cases}$$

$r$ : Learning parameter.

After assigning the reward and penalty to the cells, their action's probability vector is updated as follows:

$$p_j(k+1) = \begin{cases} p_j(k) + b \times [1 - p_j(k)] & \text{if } i = j \\ p_j(k) - b \times p_j(k) & \text{if } i \neq j \end{cases} \quad (10)$$

$b \in (0, 1)$ ,  $p(k)$ : Probability of the cells action.

In Fig. 6, the pseudo code for CLA training algorithm is illustrated.

The Cellular Automata model is selected for the classification stage in this study. By utilizing the training set, some patterns have been identified and the related rules have been generated. Therefore, according to the generated rules, the pixels of the image are compared in terms of compliance with a pattern. Hence, if the discovered region matches the rules, the system diagnoses the image as an image of lungs with cancer. Otherwise, the image is classified in the non-cancer lungs class.

#### CLA training algorithm

```

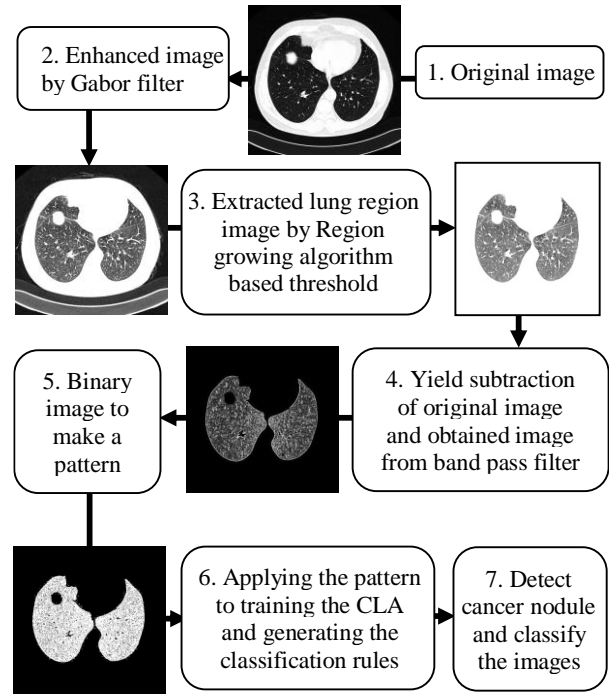
Input: binary pattern of cancer nodule.
Result: training cellular automata.
for:  $i = 1$  to all training CT image of the
database;
    apply the obtained pattern to cellular
learning automata;
    each automaton gives the initial state
based on the pattern ( $\alpha$ );
    for each pixel of pattern  $[x],[y]$ ;
        check the value of neighbors in the  $3 \times 3$ 
neighborhood ( $\beta$ );
        if the neighbor rate > assigned rate
            Reward the pattern
        else
            Penalty the pattern
    end;
    convert the found pattern to rule;
    update the probability action set ( $P$ );
end;
end;
    
```

**Fig. 6:** Pseudocode for CLA training algorithm

## Research Methodology

In order to evaluate the proposed approach, we used a lung CT scan image dataset, which was obtained from Parsian medical imaging center. A set of 22 images were used for testing the model, which consisted of 11 lung cancer images and 11 non-cancer images. The initial images that were obtained by a digital CT scanner were saved by the Digital Imaging and Communications in Medicine (DICM) file extension. Because the DICOM images include many details, they take a considerable space so, for working on images in this investigation, the DICOM images were converted to images with the JPEG format. Each image has a size of approximately 513 KB, with dimensions of  $512 \times 512$  pixels. Figure 7 illustrates all the steps of the suggested method.

According to this figure, the image was enhanced using the Gabor filter in the first step. In the next step, we applied the region growing algorithm based on the threshold to segment the lung area from the bones. Then the image was passed from a band-pass filter. To obtain the difference, the obtained image from the band-pass filter was subtracted from the enhanced image. In the next stage, the image was converted to a binary image. The binary image entered to CLA model as an input to locate the effective rules which described the cancerous tissue. Through these stored rules in the database, the test images were classified as either cancerous or non-cancerous cells.



**Fig. 7:** Steps of the suggested lung cancer recognition system (1. Original image 2. Enhancement 3. RoI, 4. Apply band-pass filter 5. Image binarization, 6. Training 7. Classification)

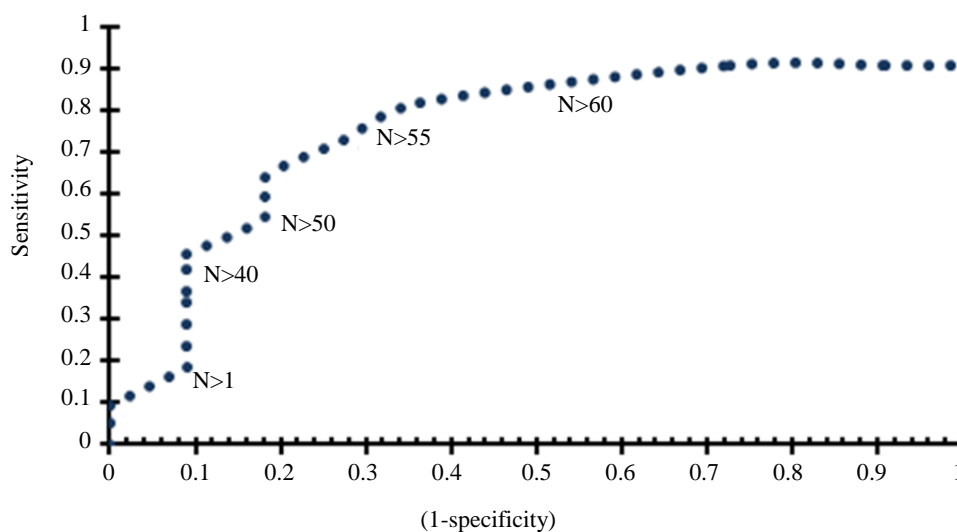
As previously mentioned, various paradigms have been proposed and used for medical images processing. Convolutional neural network (CNN) is one of the well-known models which have been employed for many tasks. Therefore, Anthimopoulos *et al.* (2016) designed a CNN based system to classify the lung images. This model provides promising results on 120 CT scan images, but it is a time-consuming procedure and the training process becomes slower due to a large number of parameters. This paradigm also needs a large number of training images. Support Vector Machine (SVM) is another popular method utilized by Kaucha *et al.* (2017) to classify the lung CT scan images. The model is designed based on a combination of image processing and data mining. The result of experiments in this study shows that the system is able to classify the images with 95.16% accuracy. This model was applied and evaluated on the LIDC dataset which contained 1018 cases. But the auteurs expressed that, this model demanded more data and it required considerable computations.

## Results and Discussion

There are various methods for perusing the result of the classification. The most traditional methods are calculating the Accuracy, True Positive Rate (TPR) and True Negative Rate (TNR) and drawing the Receiver Operating Characteristics (ROC) curve:

$$\text{Sensitivity percentage or TPR} = 100 \times \frac{TP}{TP + FN} \quad (11)$$

$$\text{Specificity percentage or TNR} = 100 \times \frac{TN}{FP + FN} \quad (12)$$



**Fig. 8:** The ROC curve of sensitivity corresponding to the specificity

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (13)$$

In Equations (11), (12) and (13), TP defines the True Positive, TN represents the True Negative, FP shows the False Positive and FN is the False Negative. The achieved values are shown in Table 1 and the evaluation is performed to illustrate the performance of the suggested system.

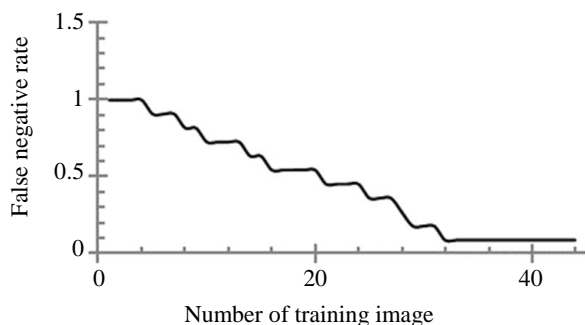
According to Table 1, the results of the proposed system show that the suggested system successfully diagnosed 10 cancer images and 11 non-cancer images and the sensitivity of the system is 90.9%. The specificity was evaluated (100%), which indicates that the proposed algorithm has just one misdiagnosis. The accuracy of the measuring system is 95.4%, which proves that it has a suitable level of reliability.

The receiver operating characteristic (ROC) curve is a graphical plot and is the best way to compare diagnostic tests. Determining the diagnostic performance of a test or evaluating the capability of a diagnosis system to distinguish the diseased cases from the normal cases, it is developed by utilizing the ROC curve. Each point on the ROC curve represents the relation between sensitivity (TP/ (TP+FN)) and specificity (TN/ (TN+FP)) with a spatial decision threshold. Figure 8 shows the sensitivity of the system based on the ROC curve. Whatever the area under the curve is greater, the performance of the system is higher. N is the radius of the neighborhood.

**Table 1:** The performance rates of the system

TP	TN	FP	FN	TPR	TNR	ACC
10	11	0	1	90.9%	100%	95.4%





**Fig. 9:** Error rate during training

According to Fig. 8, it is clearly observed that, by choosing a low neighborhood (N) radius, the proposed algorithm fail to achieve favorable results. When the number of neighborhood radius reaches a range between 58 and 60 cells, suitable results are obtained by the algorithm. However, after this range with an increasing neighborhood radius, only the computation cost is increased, whereas the results have not improved. The neighborhood radius of approximately 60 cells obtains the best result which corresponds with a 0.909 sensitivity value. If it exceeds this amount, the sensitivity of the system will not increase anymore. Hence, choosing a favorable neighborhood radius is one of the more critical characteristics, which affects the diagnosis system's performance. As shown in Fig. 9, with each training epoch, the capability of the system in terms of grouping will be increased and the possibility of error is reduced.

## Recommendation

Based on the results, the error rate decreased over the training time shows that if we have appropriate and sufficient training data, the efficient rules can be found and the system becomes more robust. After reaching the radius of 60 neighbor cells, the system provides the best results in terms of accuracy. Hence, choosing an appropriate neighborhood radius is one of the more important characteristics, which affects the diagnosis system performance. By utilizing a low complex method, the obtained system does not need complex computing operation and prohibitive hardware.

## Limitation

In this study, the proposed model is applied to the CT lung images with 512\*512 resolution. This size can vary in terms of the images captured by different devices and can affect the results. In this procedure, we just used gray level images. The neighborhood radius of approximately 60 cells provides the best result in this work, but it can be different for various tasks.

## Conclusion

According to the last image processing techniques which have been used for the lung CT- scan images processing, we have selected a few methods which have better performance to apply on the target database. In the previously suggested CAD systems, various classifiers were used such as neuronal networks, genetic algorithm (Daliri, 2012) and Decision tree (Tartar *et al.*, 2013), which were applied on different databases using diverse image preprocessing techniques in each enhancement, segmentation and feather extraction steps. The number of existing images in each used database, the percentage of selected images to training and testing are also dissimilar. While each one of these conditions can affect the yield results, the comparison of the systems under different conditions is not completely correct. Regarding the achieved accuracy, the proposed system is classified in a relatively good category in terms of performance.

In this study, a medical pattern recognition system was designed based on a proposed algorithm using cellular learning automata. The suggested model works based on the communication between its components. Through these relations, the required rules for classification are extracted. With generated rules, the classifier is able to classify images into the correct classes. As previously mentioned, the CLA model communicates with the environment. Therefore, the selection and performance actions receive an answer from the environment. Through this procedure, the CLA is able to learn from an unknown environment then make a wise decision for further. The result of the system evaluation indicates that the proposed model is able to obtain a favorable accuracy value. The implemented system is simple and its complexity and computation costs are low and if training is done properly, the model is able to decrease the error rate, which improves the system's reliability.

## Future Works

In this study, we aimed to investigate some image processing techniques and used the pattern recognition methods to design an assistant system to distinguish the normal from the abnormal lung CT scan images. With regard to the goal of this paper to create a roadmap for the study of the CAD systems, their demands are considered. Previously accomplished works are investigated to find a possible solution. As a result, some techniques for solving the problems have been suggested and evaluated on the dataset and the end results and findings were discussed. In this regard as future works; the system accuracy could be improved by applying a detection technique based on a combination of the appearance and static features of the images. Apart from that, only the gray level images were used in the

experiment to achieve the recognition results, so additional experiments involving color images would be worthwhile to investigate

## Acknowledgement

The researchers wish to thank Universiti Kebangsaan Malaysia (UKM) for supporting this work by research grant DIP-2016-018.

## Author's Contributions

**Nooshin Hadavi:** Review the literature, conducting the development, experiments, and validation of the results; manuscript writing.

**Md Jan Nordin:** Principal supervisor and advisor of work; data analysis review; manuscript revision and language corrections.

**Ali shojaeipour:** Assistance in practice related to experiments; implementation and validation of the proposed system and writing the manuscript.

**Mohammad Faidzul Nasrudin:** Advisor of writing the manuscript.

## Ethics

The corresponding author confirms that there are no ethical issues related to this research.

## References

- Abu Dalhoum, A.L. and I. Al-Dhamari, 2010. FMRI brain data classification using cellular automata. Proceedings of the 10th WSEAS International Conference on Applied Informatics and Communications and 3rd WSEAS International Conference on Biomedical Electronics and Biomedical Informatics, Aug. 20-22, World Scientific and Engineering Academy and Society (WSEAS), Taiwan, pp: 348-352.
- Ahangaran, M., N. Taghizadeh and H. Beigy, 2017. Associative cellular learning automata and its applications. *Applied Soft Comput.*, 53: 1-18. DOI: 10.1016/j.asoc.2016.12.006
- Ahmed, H.S.S. and M.J. Nordin, 2011. Improving diagnostic viewing of medical images using enhancement algorithms. *J. Comput. Sci.*, 7: 1831-1838. DOI: 10.3844/jcssp.2011.1831.1838
- Anthimopoulos, M., S. Christodoulidis, L. Ebner, A. Christe and S. Mougiakakou, 2016. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans. Medical Imag.*, 35: 1207-1216. DOI: 10.1109/TMI.2016.2535865
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. 1st Edn., Springer, New York, ISBN-10: 0387310738, pp: 738.
- Chen, J.C., C.M. Yeh and J.E. Tzeng, 2008. Pattern differentiation of glandular cancerous cells and normal cells with cellular automata and evolutionary learning. *Expert Syst. Applic.*, 34: 337-346. DOI: 10.1016/j.eswa.2006.09.035
- Chen, L., J.J. McAuley, R.S. Feris, T.S. Caetano and M. Turk, 2009. Shape classification through structured learning of matching measures. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Jun. 20-25, IEEE Xplore Press, Miami, FL, USA. DOI: 10.1109/CVPR.2009.5206792
- Daliri, M., 2012. A hybrid automatic system for the diagnosis of lung cancer based on genetic algorithm and fuzzy extreme learning machines. *J. Med. Syst.*, 36: 1001-1005. DOI: 10.1007/s10916-011-9806-y
- Esmailpour, M., V. Naderifar and Z. Shukur, 2012. Cellular learning automata approach for data classification. *Int. J. Innovative Comput. Inform. Control*, 8: 8063-8076.
- Fathy Navid, A.H. and A. Bagheri Aghababa, 2013. Cellular Learning Automata and its Applications. In: *Emerging Applications of Cellular Automata*, Salcido, A. (Ed.), InTech, pp: 85-111.
- Gaikwad, S.K., B.W. Gawali and P. Yannawar, 2010. A review on speech recognition technique. *Int. J. Comput. Applic.*, 10: 16-24. DOI: 10.5120/1462-1976
- Hadavi, N., M. Nordin and A. Shojaeipour, 2014. Lung cancer diagnosis using CT-scan images based on cellular learning automata. Proceedings of the International Conference on Computer and Information Sciences, Jun. 3-5, IEEE Xplore Press, Kuala Lumpur, Malaysia. DOI: 10.1109/ICCOINS.2014.6868370
- Homma, N., 2009. Pattern Recognition in Medical Image Diagnosis. In: *Pattern Recognition*, Yin, P.Y. (Ed.), InTech, Japan.
- Jamshidnezhad, A. and J.M. Nordin, 2012. A heuristic model for optimizing fuzzy knowledge base in a pattern recognition system. *J. Scientific Industrial Res.*, 71: 341-34.
- Janssen, M.A., 2010. Cellular Automata. *Games and Gossip*.
- Kahaki, S.M.M., M.J. Nordin, A.H. Ashtari and S.J. Zahra, 2016. Deformation invariant image matching based on dissimilarity of spatial features. *Neurocomputing*, 175: 1009-1018. DOI: 10.1016/j.neucom.2015.09.106
- Kahaki, S.M.M., M.J. Nordin, W. Ismail, S.J. Zahra and R. Hassan, 2017. Blood cancer cell classification based on geometric mean transform and dissimilarity metrics. *Pertanika J. Sci. Technol.*, 25: 223-234.
- Kamdi, S. and R. Krishna, 2012. Image segmentation and region growing algorithm. *Int. J. Comput. Technol. Electron. Eng.*, 2: 103-107.

- Kaucha, D.P., P. Prasad, A. Alsadoon, A. Elchouemi and S. Sreedharan, 2017. Early detection of lung cancer using SVM classifier in biomedical image processing. Proceedings of the IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, Sept. 21-22, IEEE Xplore Press, Chennai, India.  
DOI: 10.1109/ICPCSI.2017.8392305
- Kokol, P., P. Povalej, M. Lenic and G. Štiglic, 2004. Building Classifier Cellular Automata. In: Cellular Automata, Sloot, P.A., B. Chopard and A. Hoekstra (Eds.), Springer, Berlin Heidelberg, pp: 823-830.
- Liu, J., J. Sun and S. Wang, 2006. Pattern recognition: An overview. *Int. J. Comput. Sci. Network Security*, 6: 57-61.
- Marchiori, E., 2013. Class dependent feature weighting and k-nearest neighbor classification. Proceedings of the IAPR International Conference on Pattern Recognition in Bioinformatics, Jun. 17-20, Nice, Springer, France, pp: 69-78.  
DOI: 10.1007/978-3-642-39159-0\_7
- Mitra, S. and S.K. Pal, 2005. Fuzzy sets in pattern recognition and machine intelligence. *Fuzzy Sets Syst.*, 156: 381-386. DOI: 10.1016/j.fss.2005.05.035
- Parashar, P. and E. Harish Kundra, 2014. Comparison of various image classification methods. *Int. J. Adv. Sci. Technol.*, 2: 20-23.
- Povalej, P., P. Kokol, T.W. Družovec and B. Stiglic, 2005. Machine-Learning with cellular automata. Proceedings of the 6th international conference on Advances in Intelligent Data Analysis, Sept. 08-10, Springer, Madrid, Spain, pp: 305-315.  
DOI: 10.1007/11552253\_28
- Raj, V., S. Magg and S. Wermter, 2016. Towards effective classification of imbalanced data with convolutional neural networks. Proceedings of the 7th IAPR Workshop on Artificial Neural Networks in Pattern Recognition, Sept. 28-30, Ulm, Germany, pp: 150-162. DOI: 10.1007/978-3-319-46182-3\_13
- Ramesh, B., C. Xiang and T.H. Lee, 2015. Shape classification using invariant features and contextual information in the bag-of-words model. *Patt. Recogn.*, 48: 894-906.  
DOI: 10.1016/j.patcog.2014.09.019
- Rennard, J.P., 2000. Introduction to cellular automata. PhD Thesis, Universitat Trier.
- Rosin, P.L., 2006. Training cellular automata for image processing. *IEEE Trans. Image Process.* 15: 2076-2087. DOI: 10.1109/TIP.2006.877040
- Saliba, E. and A. Dipanda, 2013. An overview of pattern recognition.
- Sarkar, P., 2000. A brief history of cellular automata. *ACM Comput. Surveys*, 32: 80-107.  
DOI: 10.1145/349194.349202
- Segreto, T., A. Simeone and R. Teti, 2014. Principal component analysis for feature extraction and NN pattern recognition in sensor monitoring of chip form during turning. *CIRP J. Manufact. Sci. Technol.*, 7: 202-209.  
DOI: 10.1016/j.cirpj.2014.04.005
- Sharma, D. and G. Jindal, 2011. Identifying lung cancer using image processing techniques. Proceedings of the International Conference on Computational Techniques and Artificial Intelligence, (TAI' 11).
- Shiffman, D., 2012. Cellular Automata. In: *The Nature of Code: Simulating Natural Systems with Processing*, Fry, S. (Ed.), the Nature of Code. New York.
- Shiraishi, J., Q. Li, D. Appelbaum and K. Doi, 2011. Computer-aided diagnosis and artificial intelligence in clinical imaging. *Seminars Nuclear Med.*, 41: 449-462. DOI: 10.1053/j.semnuclmed.2011.06.004
- Spillmann, B., M. Neuhaus, H. Bunke, E. Pękalska and R.P. Duin, 2006. Transforming Strings to Vector Spaces using Prototype Selection. In: *Structural, Syntactic and Statistical Pattern Recognition*, Yeung, D.Y. (Ed.), Springer, Berlin, pp. 287-296.
- Tartar, A., N. Kılıç and A. Akan, 2013. A new method for pulmonary nodule detection using decision trees. Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Jul. 3-7, IEEE Xplore Press, Osaka, Japan. DOI: 10.1109/EMBC.2013.6611257
- Tuba, E., M. Tuba and D. Simian, 2016. Handwritten digit recognition by support vector machine optimized by bat algorithm. Proceedings of the 24th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, May 30-Jun. 3, University of West Bohemia, Plzen, Czech, pp: 369-376.
- Wang, J. and X. Sun, 2010. Fingerprint image enhancement using a fast Gabor filter. Proceedings of the 8th World Congress on Intelligent Control and Automation, Jul. 7-9, IEEE Xplore Press, Jinan, China. DOI: 10.1109/WCICA.2010.5554350
- Wang, X., X. Ding and C. Liu, 2005. Gabor filters-based feature extraction for character recognition. *Patt. Recogn.*, 38: 369-379.  
DOI: 10.1016/j.patcog.2004.08.004
- Wongthanavasus, S. and J. Ponkaew, 2013. Cellular Automata for Pattern Recognition. In: *Emerging Applications of Cellular Automata*, Salcido, A. (Ed.), InTech, pp: 53-68.
- Wongthanavasus, S., 2011. Cellular Automata for Medical Image Processing. In: *Cellular Automata-Innovative Modelling for Science and Engineering*, Salcido, A. (Ed.), InTechOpen, pp: 395-410.