Texture Filter Optimization Using Particle Swarm Optimization for Efficient Lung Image Classification

S.Vatchala¹, Sridevi.R², A. Sathish³, S. Ravimaran⁴, M. Nallusamy⁵
¹Department of School of Computer Science and Engineering (SCOPE), VIT – Chennai Campus
²K Ramakrishnan College of Engineering Trichy, India
³,⁵Department of Computer Science and Engineering, Roever Engineering College, Perambalur
⁴Department of Artificial Intelligence and Data Science, Sarananthan College of Engineering
*Corresponding author: S.Vatchala, Department of School of Computer Science and Engineering (SCOPE), VIT – Chennai Campus, Email: vatchala.s@vit.ac.in

Submitted: 24 April 2023; Accepted: 13 May 2023; Published: 07 June 2023

ABSTRACT

There are several millions of images that have been generated by the hospitals and healthcare centres on a daily basis. Imaging was employed as the preferred tool of diagnostics by many more medical procedures. Detecting lung cancer at an early stage can to a significant extent enhance survival chances but can also be challenging to detect the various stages of lung cancer with a lesser number of symptoms. Medical image processing is favoured by Gabor filter and for the application of signal processing, it is a sharp cut-off. A step in pre-processing is the next step in machine learning and this is very effective in the reduction of dimensionality and removal of irrelevant data. It also helps in increasing accuracy and in learning accuracy improvement aside from the comprehensibility of results. Particle Swarm Optimization (PSO) has been used widely for solving problems in optimization which is a problem of feature selection. For the PSO, every solution has been looked at as a particle and the algorithm looks for an ideal solution taking into consideration all particles.

Keywords: Medical Images, Gabor filter, Feature selection, Particle Swarm Optimization (PSO), Naïve Bayes and Random Tree

INTRODUCTION

The human lungs are organs that are cone-shaped resembling a sponge [1]. The lungs are a part of an apparatus that is complex and this expands and also relaxes several thousands of times each day pull in oxygen and expend carbon dioxide. The disorders of the lung can result in problems and affect airways. For curing such diseases, it is crucial that they are identified correctly. Owing to the new development in technology and different imaging modalities like the Magnetic Resonance Imaging (MRI) and the Computed Tomography (CT), the diagnosis of lung diseases is made easier. The challenges faced are how to analyze a great volume of information and automate the process for the diagnosis and treatment of diseases.

To automate the classification of MRI images, it is required to classify the texture and identify the region of lung affected. This type of texture classification is the primary domain found in the analysis of texture in image processing. This analysis will be crucial in different applications that are used in the analysis of computer images.
made for either the classification or the segmentation of images that were based on both local, as well as spatial variations of either intensity or colour. There can be a segmentation or classification that is successful in needing an efficient image texture description. Another important area of application in such texture classification can be the biomedical surface and industrial inspection. These may be the identification of disease and defects and the segmentation of the aerial or satellite imagery that is content-based access to the databases of images [2, 3]. An image classification is for analyzing the properties of various features for organizing them into different classes. The classifiers employ two phases: training where the characteristic properties belonging to image features isolated based on the class. And in the subsequent testing phase, feature-space partitions are employed for the classification of image features.

The feature can be a piece of information that is quite significant and is extracted from the image providing a detailed understanding of an image. These features will be based on intensity and geometric in nature. The measurements of shape will be the measures of physical dimension to characterize the object’s appearance. These are the only features that are taken to be extracted. The Gabor filters are widely used in various applications of computer vision that includes analysis of texture, classification, edge detection, and segmentation. Furthermore, these features will be extracted using the Gabor Filters (or Gabor Features). These are intertwined using signals that result in the Gabor space that has the advantage of a Gabor function and this will be a good fit to the weight functions and their receptive weight. For the applications of image processing, the Gabor filter can be useful in edge detection. To identify a particular image, it is suitable for spatial locations in the distinction between objects in the image. The main activations are extraction creating a sparse object representation [4, 5].

To address the high dimensionality of the feature set obtained from images, feature selection techniques are applied. This refers to a technique that chooses certain minimized features that are applicable to improve the accuracy of classification. These methods are: the Filter, the Wrapper and the Hybrid method. Swarm Intelligence (SI) is used for tackling feature selection as it is a proven technique to solve computational problems that are NP-hard to find a feature subset that is optimal. In recent years, PSO is applied successfully to plenty of areas of research. There has been a demonstration that the PSO can get better results that is faster and cheaper compared to the other methods. One more reason for the PSO to be attractive is the lesser number of parameters that need adjustment. There is yet another version that has some variations and is useful in varied applications. Particle Swarm Optimization is an approach that is very useful in being used across various applications which are in addition to the regular ones.

In this paper, the technique of optimization used for the classification of various lung diseases has been presented. The features of the lung images has been chosen using the PSOalgorithm proposed along with the Gabor filter, and these classifiers were used for the classification of datasets into the relevant ones. In section 2, the literature related to the work is explained. The techniques used are explained in Section 3, results are elaborated in Section 4 and the conclusion is made in Section 5.

LITERATURE SURVEY

Tun and Soe [6] employed a new median filter for the pre-processing of images. For purposes of segmentation, Otsu’s method of thresholding was used. In the case of feature extraction, all physical measures with Gray-Level Co-occurrence Matrix (GLCM) based method were used. The Artificial Neural Network (ANN) had been applied to classifying the stages of the disease. Computed Tomography (CT) scan image was found suitable to the diagnosis of lung cancer. The paper has been presented for implementing feature classification after extraction of the lung cancer nodule. For implementing this MATLAB software was employed. The technique enables doctors and radiologists to understand the condition of the diseases early to avoid serious consequences.

Alakwaa et al [5] had demonstrated an new Computer-Aided Diagnosis (CAD) system that was employed for classifying lung cancer of the CT scans that have unmarked nodules which is a dataset in the Kaggle Data Science Bowl, 2017. The CAD system had three phases (segmentation, detection of nodule candidates, and classification of malignancy), that permits
certain efficient training or detection with plenty of generalizability to all other types of cancers. Thresholding had been employed as an approach to initial segmentation for segmenting out the lung tissue from the remaining CT scan. The next best segmentation of lungs is produced by thresholding. An initial approach had been directly fed to the segmented CT scans within the 3D Convolutional Neural Networks (CNNs) used for classification which was inadequate. As opposed to the modified U-Net trained on a LUNA16 data employed to identify the nodules. There had also been a U-Net nodule recognition that had generated several false positives, in order to ensure the CT regions that had segmented lungs were the nodule candidates. It was decided by U-Net outputs that were fed inside the 3D CNNs for an ultimate classification of the CT scan either as positive or negative. Test set accuracy which was about 86.6% which was produced by the 3D CNNs. The CAD system had outperformed the current ones in literature with various training and testing phases needing plenty of labelled data.

Kaur and Gupta [6] had proposed an automated approach to classification along with feature selection of lung diseases with CT images. ACT image of the lung is used as the input. There are filters that were used for choosing features once the MAD technique is used after extraction. There is a new technique of feature selection which is the hybridization of a PSO for choosing features after the MAD technique. As soon as the features are chosen, they will be classified with the MLP-NN classifier. For the purpose of this paper, the techniques of optimization were employed for the process of feature selection.

Kumar et al [7] had explored the expedient image segmentation algorithm employed for medical images in order to curtail the interpretation of physicians of CT images. These modern modalities of medical imaging were able to generate larger images that were very grim for analysing them manually. These segmentation algorithms were dependent on the convergence time and exactitude. There was also a compelling need for exploring and also for implementing the evolutionary algorithms for solving the problems that were connected with the segmentation of medical images. Other frequently diagnosed types of cancer all over the world among men are lung cancer.

METHODOLOGY
In this study, there are a total of five algorithms taken for study which are the K-Median Clustering, the Particle Swarm Optimization, the Guaranteed Convergence Particle Swarm Optimization (GCPSO), the K-Means Clustering and the Inertia-Weighted Particle Swarm Optimization.

Gabor Filter
The actual performance of the median, the adaptive median, and the average filters in the stage of pre-processing was compared and its adaptive median filter was found to be suitable for the CT images. Also, there was a contrast that was observed in an image that was enhanced by means of using adaptive histogram equalization. A pre-processed image that had an improved quality was subject to four algorithms. All practical results were verified for 20 sample images with the MATLAB, and the GCPSO had a high accuracy of 95.89%. The Gabor filters have proven to be very good band-pass filters applied to one-dimensional signals like speech signals. Gabor filters were also employed in image representation, image coding, image retrieval, edge detection, and also in texture segmentation [8-10].

The Gabor filter has an application related to the computer with its inspiration from biological findings based on the similarity of the 2D Gabor filters along with their neuron fields found in their respective visual cortex. In connection to spatial domains or frequency domains, a localization of joints that is optimal will be found. For edge detection purposes, vehicle detection, image retrieval, image coding and texture analysis, their applications are found for different images or applications. A 2D Gabor filter functionality can be represented as a Gaussian function modulated using complex sinusoidal signals. The formulation of the 2-D Gabor filter \( g(x, y) \) is in (1 and 2):(1)

\[
\begin{align*}
\sigma_x & = x \cos \theta + y \sin \theta \\
\sigma_y & = -x \sin \theta + y \cos \theta
\end{align*}
\]

Wherein, \( \sigma_x \) and \( \sigma_y \) are the scaling parameters in filter that determine the pixel size and also its neighbourhood in which a weighted summation \( \theta (\theta \in [0, \pi]) \) will specify the Gabor filter orientation. \( W \) refers to the radial

\[
\begin{align*}
\sigma_x & = x \cos \theta + y \sin \theta \\
\sigma_y & = -x \sin \theta + y \cos \theta
\end{align*}
\]
frequency along with the sinusoid. This filter further responds to the bar or the edge with a normal parallel to the orientation which is $\theta$ for the sinusoid. Gabor function in the form of Fourier transform is shown as per (1) to (3):

$$\sigma_u = \frac{1}{2\pi \sigma_x, \sigma_y} = \frac{1}{2\pi \sigma_x, \sigma_y}$$

Wherein, $\sigma_u$ and $\sigma_v$ and the Fourier domain representation is as in (3) that indicates the extent of modification for each frequency component of the input image:

$$G(u,v) = \exp \left[ -\frac{1}{2} \left( \frac{(u-W^2)}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right) \right]$$

(3)

In which the $\sigma_u = \frac{1}{2\pi \sigma_x, \sigma_y} = \frac{1}{2\pi \sigma_x, \sigma_y}$ and the representation of the Fourier domain as per (3) will specify the amount in which the filter will modify every component of frequency in the input image.

**Correlation based Feature Selection (CFS)**

The rank of CFS has been attributed in accordance with a correlation-based heuristic evaluation function [10]. This function assesses the subsets that are made of several attribute correlated to the class label and independent of each other. This CFS technique will assume that all inappropriate features will have a low level of correlation and overlooked. At the same time, the excess features need to be examined and they tend to be reasonably well-correlated with more than one attribute. A CFS is found to be a heuristic as it can assess the utility or the value of a feature subset according to (4). The usefulness of individual features was depicted here to predict a new class label and an inter-correlation level between them.

$$\text{Merits} = \frac{k^2 f}{\sqrt{k + k(k-1)f}}$$

(4)

In this, merit refers to a heuristic “merit” for S a feature subset with k features. $F_{cf}$ is a class correlation existing between the average feature and $F_{cf}$ is the average inter-correlation that is feature to feature.

**Proposed Particle Swarm Optimization (PSO) Optimization with Gabor Filter**

A metaheuristic technique that is inspired by the swarm and its social behaviour is the Particle swarm optimization (PSO). Here, a technique that is based on population is perceived among birds and fish to look out for the best path. As suggested by the name, the process of optimization of the PSO was done on a swarm of particles. The PSO contains a warm in which every particle has a certain position within its search space moving around with some velocity. This particle will choose the path that is the best for every iteration using memory and further by learning the path which is effective followed by the swarm earlier. A new position will be selected based on earlier knowledge by the position that is self-best and best with very few assumptions regarding the optimization of the problem when searching large spaces with candidate solutions [11-13].

Every particle performs with “closeness” with a global minimum that is measured in accordance with a fitness function that is predefined. In case the search space is found to be D-dimensional with m particles, every particle will be located at a position $X_i = [x_{i1}, x_{i2}, ..., x_{id}]$ with a velocity $V_i = [v_{i1}, v_{i2}, ..., v_{id}]$, wherein i=1, 2, ..., m. For the PSO algorithm, every particle will move to its best position (pbest) that is denoted as $P_{best,v} = [p_{best,1}, p_{best,2}, ..., p_{best,m}]$ and further the best position of the entire swarm (gbest) shown as $G_{best} = [g_{best,1}, g_{best,2}, ..., g_{best,m}]$. Each particle will change its position in accordance with its velocity and this is generated randomly towards its pbest and gbest positions. For every particle i and dimension s, a new velocity $v_{is}$ with position $x_{is}$ is computed as per equation (5):

$$x_{is} = \left[ x_{is} - \frac{w}{t} \right]$$

Wherein t refers to the actual iteration number. Inertial weight w will be used for controlling velocity and to balance both exploration and exploitation abilities of the algorithm. There is a large w value that keeps all particles under a high velocity thus preventing them from getting trapped inside the local optima. There can also be a small value of w that maintains the particles at a velocity that c1 and c2 refer to the coefficients of acceleration which can determine if the particles want to be closer to both pbest and gbest positions. Now, $b_1$ and $b_2$ are the independent random numbers that get distributed uniformly.
between 0 and 1. The criterion of termination of the PSO will further include the actual number of generations and their designated \( p_{\text{best}} \) value without any noticeable improvement to the \( p_{\text{best}} \).

A Gabor filter is now used to extract textural features of an image connected with a certain band of frequency. The texture of the image will be a quasi-periodic signal. The concentration of energy will be within the frequency range. For this, a response band is crucial and this is in case the range of frequency is corresponding to the Gabor filters. If it is not the case, there can be suppression of its output. Thus, for a relevant spectrum, it can be captured using a well-designed set of such Gabor filters with different width and direction. More often, there can be a huge bank of filters that is employed for tasks of texture analysis. This may be quite expensive owing to its inverse Fourier transform but is also quite stable in connection with sequence textures. This is used for the optimization of the Gabor Filter bank requiring to be applied. Thus, the time taken for execution is reduced and suitable parameters of filter are used to detect and further estimate motion.

Image classification performance with the PSO is dependent on the Gabor filter quality. The lesser the convolution operation performed, the more is the rate of recognition of the algorithm and this means the Gabor filters are better.

### Classifiers

The classification will have two different phases, the first being the training phase where huge datasets will be supplied along with an analysis wherein rules are created. For executing the second phase, an evaluation of datasets with archives of the accuracy of patterns of classification. This section describes the Naïve Bayes and the Random Tree methods.

#### Naïve Bayes Classifier

Naïve Bayes is a Statistical Bayesian Classifier that was quite simple. It is known as Naïve since it makes an assumption that all the variables will provide to classification and will be mutually correlated. Such assumptions are referred to as class conditional independence. This is also known as the Independence Bayes, Simple Bayes or the Idiot’s Bayes. They further predict probabilities of their class membership like the probability of the data item given to a certain class label. This Bayesian classification was ideally based on a Bayesian Theorem mentioned below [14-16]: if the \( X \) is a new data sample that has an unknown class label, \( H \) is a hypothesis so that the data sample \( X \) can be part of a particular class \( C \). The Bayes theorem is employed to compute posterior probability \( P(C|X) \), which is from \( P(C) \), \( P(X) \), and \( P(X|C) \) as per (6):

\[
P(C|X) = \frac{P(X|C)P(C)}{P(X)}
\]

(6)

Wherein, \( P(C|X) \) refers to the target class and its posterior probability. \( P(C) \) is the prior probability of a class. \( P(X|C) \) indicates the chances of a predictor probability of a class, and \( P(X) \) will be the prior probability for the predictor of class:

The advantages are that it will need only a short time for computation and training, it removes unwanted information and has good performance.

#### Random Tree

The Random Tree is known as a supervised Classifier with an ensemble algorithm of learning that can generate plenty of individual learners. There is a bagging idea that is employed for the construction of a random dataset for a decision tree. For any type of standard tree, each node will
be split by making use of the best ones among the predictor subsets chosen for the node. This algorithm is capable of dealing with several types of problems of classification as well as regression. The random trees will be a new group (or ensemble) of predictors and this is known as the forest. The mechanism for classification is as follows: a random tree classifier will get an input feature vector to classify with each tree found in the forest and will output the class label received for many such “votes” [17-19].

Firstly, training data will be sampled with a replacement for every single tree as in the case of Bagging. Next, during the time a tree is grown, as opposed to computing the best split for every node only one random subset for the attributes that are taken into consideration for a node. The best split is then computed for the subset. These trees are for classifying random forests and model trees. The random model trees further combine the model trees with random forests. These random trees will use the produce for the purpose of choosing the split and will induce a reasonable number of such balanced trees in which a single global setting for a ridge value will work across the leaves and this will further simplify the process of optimization [20-21].

RESULTS AND DISCUSSION

In this work the Gabor filter parameters are optimized using PSO for efficient feature extraction. Three types of lung images are classified after feature extraction namely Normal (300 CT images), Bronchiectasis (200 CT images) and Pleural Effusion (100 CT images). Table 1 and figure 2 to 4 shows the results of classification accuracy, sensitivity and specificity.

<table>
<thead>
<tr>
<th>TABLE 1: Summary of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Classification accuracy</td>
</tr>
<tr>
<td>Sensitivity for Normal</td>
</tr>
<tr>
<td>Sensitivity for Bronchiectasis</td>
</tr>
<tr>
<td>Sensitivity for Pleural Effusion</td>
</tr>
<tr>
<td>Specificity for Normal</td>
</tr>
<tr>
<td>Specificity for Bronchiectasis</td>
</tr>
<tr>
<td>Specificity for Pleural Effusion</td>
</tr>
</tbody>
</table>

![Classification Accuracy Graph](image)

FIGURE 2: Classification accuracy for PSO-Gabor-Random Tree
Table 1 and figure 2 shows that the classification accuracy of PSO-Gabor-Random Tree performs better by 9.13%, by 5.12% and by 2.96% than Gabor filter-Naïve Bayes, Gabor filter-Random Tree and PSO-Gabor filter- Naïve Bayes respectively.

![Figure 3: Sensitivity for PSO-Gabor-Random Tree](image-url)

Table 1 and figure 3 shows that the sensitivity of PSO-Gabor-Random Tree performs better by 7.93%, by 4.52% and by 3.7% than Gabor filter- Naïve Bayes, Gabor filter-Random Tree and PSO-Gabor filter- Naïve Bayes respectively for normal. The sensitivity of PSO-Gabor-Random Tree performs better by 11.7%, by 6.02% and by 1.97% than Gabor filter- Naïve Bayes, Gabor filter-Random Tree and PSO-Gabor filter- Naïve Bayes respectively for Bronchiectasis. The sensitivity of PSO-Gabor-Random Tree performs better by 8%, by 5.26% and by 2.59% than Gabor filter- Naïve Bayes, Gabor filter-Random Tree and PSO-Gabor filter- Naïve Bayes respectively for Pleural Effusion.

![Figure 4: Specificity for PSO-Gabor-Random Tree](image-url)
Table 1 and figure 4 shows that the Specificity of PSO-Gabor-Random Tree performs better by 3.68%, by 2.24% and by 0.63% than Gabor filter-Naïve Bayes, Gabor filter-Random Tree and PSO-Gabor filter-Naïve Bayes respectively for normal. The Specificity of PSO-Gabor-Random Tree performs better by 4.87%, by 2.46% and by 1.75% than Gabor filter-Naïve Bayes, Gabor filter-Random Tree and PSO-Gabor filter-Naïve Bayes respectively for Bronchiectasis. The Specificity of PSO-Gabor-Random Tree performs better by 6.25%, by 3.6% and by 2.15% than Gabor filter-Naïve Bayes, Gabor filter-Random Tree and PSO-Gabor filter-Naïve Bayes respectively for Pleural Effusion.

CONCLUSION

The technique and the process employed for creation of these images of the human body for different clinical purposes and further for analyzing and diagnosing in medical science is referred to as medical imaging. Lung diseases are the disorders that are capable of affecting the lungs which are the organs that help us breathe. This is a found to be a very common medical condition all the world over and more so in India. This technique of Particle Swarm Optimization (PSO) is considered to be a novel one of global optimization and the method is based on population and is also stochastic. It is a very simple algorithm and is very effective in a varied range of functions.

REFERENCES

8. R.Sridevi, S.Srimathi,(2020),” Populating Indian GST Details Into Java Apache Derby Database Powered By Glass Fish Server” Injventive Communication and Computational Technologies, springer series, Page(s): 409-421.
Texture Filter Optimization Using Particle Swarm Optimization For Efficient Lung Image Classification


