CHAPTER 5

CLASSIFICATION OF NODULES EMPLOYING
SUPPORT VECTOR MACHINE (SVM)

5.1 INTRODUCTION

Cancer forms in tissues of the lung, usually in the cells lining air passages. Methods of cancer detection include new imaging technologies, tumor markers, and biopsy procedures.

The complexity of the structures in the chest including the ribs, mediastinum and pulmonary vessels can make it difficult to identify separate pulmonary nodules that may represent an early lung cancer from normal anatomy. Computer-aided detection is a method that can be used to assist the radiologist in the search for lung cancer. The software highlights abnormalities that may be overlooked by the radiologist on an initial search.

Scientists (Disabled-world.com, 2011) have identified over 480 molecules whose concentration in the blood changes when a person develops lung cancer. These molecules are present in the blood cells either in increased or decreased quantities. The molecules are nucleic acids which form in the body when certain genes are transcribed.

The changes in the blood also occur if the tumor is still in a very early stage. In lung cancer, there are four different stages. The prognosis for patients in stage 3 and 4 is still very poor even today; even with the most modern therapies, the point of death can only be postponed. Lung cancer in
stage 1, on the other hand, can be treated surgically and it can even be cured in many cases. Today, however, a tumor is seldom detected so early, namely in only about 15% of all cases. If a simple screening blood test would increase this percentage, a large proportion of lung cancer patients could survive. By contrast, to date, over 80% of all lung cancer patients die within two years after diagnosis, since the tumor is already far too advanced.

In the future, a lung cancer screening test may become part of routine practice: The doctor takes a blood sample from his/her patient, and within 24 hours, he knows with a high degree of certainty whether the patient has lung cancer or not – even though the patient does not yet have any symptoms.

The researchers are presently planning an analogous but much larger study with ten times as many patients, in order to confirm the results. If the present results prove to be true in such a study, there would no longer be anything standing in the way of developing the blood test to the point of being ready to be put on the market.

Qionghua Weng et al (2009) discussed computer-aided diagnosis system using support vector machine for nodule detection in chest radiographs. In this system, multi-scale difference approach was implemented to extract initial candidate nodules, features are extracted and SVM is applied and a sensitivity of 0.86 was achieved with 5.21 FPs/image.

Support Vector Machines are extensively used in the classification of candidate nodules (Paola Campadelli et al 2005) where cost-sensitive SVMs trained with very unbalanced datasets achieved promising results in terms of sensitivity and specificity.
This chapter introduces a proposed Computer Aided Diagnosis (CAD) system for detection of lung nodules using the Support Vector Machine. The lung cancer detection system comprises of a number of steps in which, initially different image processing techniques such as Bit-Plane Slicing, Erosion, Median Filter, Dilation, Outlining, Lung Border Extraction and Flood-Fill algorithms are applied for extraction of lung region. Then for segmentation, Modified Fuzzy Possibilistic C-Means algorithm is used and for learning and classification Support Vector Machine is used.

5.2 PROPOSED METHODOLOGY

The first stage of the proposed technique is lung region extraction using several image processing techniques. The second stage is segmentation of extracted lung region using Modified Fuzzy Possibilistic C-Means (MFPCM) algorithm. Finally, Support Vector Machine (SVM) is applied in order to classify the cancer nodules.

The five phases included in the proposed computer aided diagnosis system for lung cancer detection are as follows:

- Extraction of lung region from chest computer tomography images
- Segmentation of Lung region using Modified Fuzzy Possibilistic C-Means
- Feature extraction from the segmented region
- Formation of diagnosis rules from the extracted features
- Classification of benign and malignant nodules using SVM.
Phase 1: Extraction of Lung Region from Chest Computer Tomography Images

Extraction of the lung region is the first and foremost phase in the proposed scheme. Here, the basic image processing algorithms such as Bit-Plane Slicing, Erosion, Median Filter, Dilation, Outlining, Lung Border Extraction and Flood-Fill algorithms are utilized. Bit-plane slicing algorithm acts as data compression technique, which helps in minimizing the number of slices to be analyzed for the detection of cancerous nodules. This, in turn reduces the time taken for analyzing the slices. This technique was then followed by other techniques to delete the details that are not relevant, from the image.

Phase 2: Segmentation of Lung Region Using Modified Fuzzy Possibilistic C-Means

After extracting the lung region, it is followed by the second phase where segmentation of lung region is done which would identify the Region of Interest (ROI). Modified Fuzzy Possibilistic C-Means (MFPCM) is used in the proposed approach for segmentation, as explained in section 4.2.1 of this thesis as it is better when compared to FPCM.

Phase 3: Feature Extraction from the Segmented Region

In the following phase, features were obtained from the segmented image. These features are used to eliminate false positive candidates and identify true positives.
Phase 4: Formation of Diagnosis Rules from the Extracted Features

After the necessary features are extracted, diagnosis rules are formed as in section 3.4.3. These rules can be passed on to classifier in order to detect the cancer nodules.

Phase 5: Classification of Benign and Malignant Nodules

The final phase in the proposed CAD system is the classification of benign and malignant nodules. The classifier used in this approach is Support Vector Machine.

The feature vectors extracted from the images are fed as input to the SVM classifier.

5.3 SUPPORT VECTOR MACHINE (SVM)

SVM introduced by Cortes is generally used for classification purpose. SVMs are efficient learning approaches for training classifiers based on several functions like polynomial functions, radial basis functions, neural networks etc. It is considered as a supervised learning approach that produces input-output mapping functions from a labeled training dataset. SVM has significant learning ability and hence is broadly applied in pattern recognition.

SVMs are universal approximators which depend on the statistical and optimizing theory. The SVM is particularly striking the biological analysis due to its capability to handle noise, large dataset and large input spaces.

The fundamental idea of SVM can be described as follows (Figure 5.1):

- Initially, the inputs are formulated as feature vectors.
• Then, by using the kernel function, these feature vectors are mapped into a feature space.

• Finally, a division is computed in the feature space to separate the classes of training vectors.

Usually, global hyper plane is sought by the SVM in order to separate both the classes of examples in training set and avoid over fitting. This phenomenon of SVM is more superior in comparison to other machine learning techniques which are based on artificial intelligence.

![Input Space Feature Space](image)

**Figure 5.1 Principle of SVM**

The mapping of the input-output functions from a set of labeled training data set is generated by the supervised learning method called SVM. In a high dimensional feature space, SVM uses a hypothesis space of linear functions which are trained with a learning technique from optimization theory that employs a learning bias derived from statistical learning theory.

In Support Vector machines, the classifier is created using a hyper-linear separating plane as shown in Figure 5.2. The SVM provides the ideal solution for problems which are not linearly separated in the input space. The original input space is non-linearly transformed into a high dimensional feature space, where an optimal separating hyper plane is found and the problem is solved. A maximal margin classifier with respect to the training data is obtained when the separating planes are optimal.
The support vectors are the points which are at the margin and the solution is based only on these data points. This is the unique feature of this technique. When a feature space uses a set of nonlinear basis function, the linear SVM can be extended to nonlinear SVM. The data points can be separated linearly in the feature space which are having higher dimension.

![Input Space $\mathbb{R}^d$](image)

**Figure 5.2** A separating hyper plane in the feature space corresponding to a non-linear boundary in the input space.

A significant feature of the SVM is that this transformation need not be implemented to determine the separating hyper plane in the possibly very-high dimensional feature space, instead a kernel representation can be used for the purpose of determining the separating hyper plane, where the solution evaluated at the support vectors is written as a weighted sum of the values of certain kernel functions.

The mapping of the data from a lower-dimensional input space to a higher-dimensional feature space is possible when SVM acts as a binary classifier. The data is linearly separable into two classes by this mapping. The application of the binary classifiers to various multi category problems is made possible by using one versus-all approach, for which $c$ (number of classes) binary classifiers should be built for SVM to differentiate one class from all the other classes. Similarly, when the one-versus-one comparison
approach is used, \([c(c-1)/2]\) binary classifiers should be built for SVM in such a way that it must be able to distinguish between every two class combination. Thus, the overall complexity of the classifier increases, when the number of classes ‘c’ increases.

For binary classification SVM determines an Optimal Separating Hyperplane (OSH) which produces a maximum margin between two categories of data. To create an OSH, SVM maps data into a higher dimensional feature space and carries out this nonlinear mapping with the help of a kernel function. Then, SVM builds a linear OSH between two classes of data in the higher feature space. Data vectors that are closer to the OSH in the higher feature space are known as Support Vectors (SVs) and include all data necessary for classification. The theory of SVM is described below.

The training set \(D = \{(x_i, y_i)\}_{i=1}^{l}\) with every input \(x \in \mathbb{R}^n\) and an associated output \(y_i \in \{-1, +1\}\) are considered. Every input \(x\) is initially mapped into a higher dimension feature space \(F\), through a nonlinear mapping. If the data are linearly non-separable in \(F\), then a vector \(w \in F\) and a scalar ‘b’ will exist which describe the separating hyper plane as:

\[
y_i(w \cdot x_i + b) \geq 1 - \xi_i, \forall i
\]

(5.1)

where \(\xi (\geq 0)\) are known as slack variable. The hyper plane that optimally splits the data in ‘\(F\)’ is one that

\[
\text{minimise } \frac{1}{2} w \cdot w + C.
\]

subject to \(y_i(w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall i\)

(5.2)
where $C$ is known as regularization parameter that finds the tradeoff between maximum margin and minimum classification error. By creating a Lagrangian, the optimal hyper plane based on the previous equation, may be provided as the solution of

$$\text{maximize } W(\alpha) = \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i=1}^{L} \sum_{j=1}^{L} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

subject to $\sum_{i=1}^{L} y_i \alpha_i = 0, \, 0 \leq \alpha_i \leq C, \forall i$ \hspace{1cm} (5.3)

where $\alpha_1, \ldots, \alpha_L$ represents the nonnegative Lagrangian multipliers. The data points $x_i$ that corresponding to $\alpha_i > 0$ are considered as Support Vectors. The weight vector $w$ is described as

$$w = \sum_{i=1}^{L} y_i \alpha_i x_i$$ \hspace{1cm} (5.4)

For any test vector $x \in \mathbb{R}^n$, the classification result is then described as

$$y = \text{sign}(w \cdot z + b) = \text{sign}(\sum_{i=1}^{L} \alpha_i y_i K(x_i, x) + b)$$ \hspace{1cm} (5.5)

Equation 5.1 to 5.5 explains the application of SVM to classification of nodules. A kernel function and the parameters should be selected for constructing the support vector machine classifier. Here, three kernel functions are used to construct SVM classifiers:

- Linear kernel function
- Polynomial kernel function
- Radial basis function
**SVM kernel functions**

The classification ability of feature combinations in gait applications is obtained with first attempt work of SVM kernel function. The three main kernel functions are used for our study here.

**Radial Basis Function Kernel:** The kernel function is defined as

\[
K(x,z) = \exp\left\{-\frac{|x-z|^2}{2\sigma^2}\right\}, \quad \sigma \text{ is the width of the function.}
\]

**Linear Kernel:** The Linear kernel is the simplest kernel function. It is specified by the inner product \(K(x, z) = \langle x, z \rangle\) additionally with an optional constant \(c\). Kernel algorithms using a linear kernel are usually equivalent to their non-kernel counterparts.

\[
K(x, z) = x^Tz + c
\]

**Polynomial Kernel:** The Polynomial kernel is a non-stationary kernel. Polynomial kernels are suitable when the all the training data is normalized.

\[
k(x, z) = (\alpha x^Tz + c)^d
\]

Modifiable parameters are the slope \(\alpha\), the constant term \(c\) and the polynomial degree \(d\).

After the learning process is completed by providing several conditions, the proposed technique would be able to detect the presence of cancer in the lung region automatically.

The most used kernel function for SVM is Radial Basis Function (RBF) because of their localized and finite responses across the entire range.
of real x-axis. The classification accuracy of RBF kernel was high; also, the bias value and the error rate of RBF kernel were small when compared to other kernels.

CROSS-FOLD VALIDATION

Cross-validation is a method for analyzing how the results of a statistical analysis will generalize to an independent data set. It is used in situations, where the goal is prediction, and to estimate how accurately a predictive model will perform in practice. One round of cross-validation includes dividing or partitioning a sample of data into complementary subsets, performing the analysis on one subset (training set), and validating the analysis on the other subset (validation set or testing set). Multiple rounds of cross-validation are performed using different partitions to reduce variability, and the validation results are averaged over the rounds.

In k-fold cross-validation, the sample data is randomly partitioned into k subsamples. Among the k subsamples, a single subsample is kept as the validation data for testing the model, and the remaining k–1 subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation data. The k results from the folds then can be averaged or combined to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. 10-fold cross-validation is used in this study. The 300 samples collected for this study are split into 10 subsets and each subset is validated against the remaining subsets and this process is repeated 10 times and the results are averaged.

Confusion matrix was used to calculate the performance of the classifier. Figure 5.3 shows the confusion matrix. It is a specific table that
helps to visualize the performance of a learning algorithm. Each column of the matrix represents the predicted class, and each row represents the actual class.

<table>
<thead>
<tr>
<th>True positives</th>
<th>False positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>False negatives</td>
<td>True negatives</td>
</tr>
</tbody>
</table>

Figure 5.3 Confusion Matrix

The elements on the diagonal represent the correctly predicted elements and off the diagonal represent the misclassified elements.

5.4 PERFORMANCE EVALUATION

The performance of the proposed scheme was evaluated using the following parameters:

- Sensitivity
- Specificity
- Accuracy
- Classification time

Sensitivity

Sensitivity is also known as recall rate and is defined as the proportion of actual positives which are correctly classified. (i.e.) the percentage of patients with malignant nodules identified as cancer patients.

<table>
<thead>
<tr>
<th>Table 5.1 Sensitivity of the SVM Kernels</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kernel</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Linear (Dot)</td>
</tr>
<tr>
<td>Polynomial</td>
</tr>
<tr>
<td>Radial Basis Function (RBF)</td>
</tr>
</tbody>
</table>
Table 5.1 shows the sensitivity obtained by applying the proposed CAD system to the CT images. The CAD system used MFPCM for segmentation and SVM for classification. The system is tested with three different kernels dot, polynomial and Radial Basis Function. The default kernel is linear kernel or dot product. In the polynomial kernel, 3 have been chosen as the polynomial order. Finally, RBF Kernel was used and the sensitivity obtained by three kernels was tabulated. Table 5.1 shows that the RBF kernel in SVM performs better when compared to other two kernels.

The true positives (malignant nodules) as per training and testing phase were 112 and 106 which means that the false negatives (missed nodules) were 11 and 17 in training and testing phase respectively, in the case of RBF. It is clear from the table that some of the nodules identified as true positives by the classifier have a size greater than or equal to 2mm, as the number of nodules equal to 2mm in size was 28 which indicates that MFPCM in combination with SVM enables the early detection of cancerous nodules in Computer Tomography images. The number of true negatives (benign nodules) reported by the classifier was 159 and 150 in training and testing phase which means that the false positives (misclassified nodules) were 18 and 27 in the training and testing phase respectively.

**Specificity**

Specificity is defined as the proportion of actual negatives which are correctly classified. (i.e.) the percentage of people with benign nodules identified as healthy people.
Table 5.2  Specificity of the SVM Kernels

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Specificity (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Standard Deviation</td>
<td>Testing</td>
</tr>
<tr>
<td>Linear (Dot)</td>
<td>84.18</td>
<td>0.68</td>
<td>81.92</td>
</tr>
<tr>
<td>Polynomial</td>
<td>86.44</td>
<td>0.61</td>
<td>83.61</td>
</tr>
<tr>
<td>Radial Basis Function (RBF)</td>
<td>89.83</td>
<td>0.51</td>
<td>84.74</td>
</tr>
</tbody>
</table>

Table 5.2 shows the specificity obtained by the three kernels. The true negatives (benign nodules) identified as per the RBF Kernel was 159 and 150 nodules in training and testing phase. Sensitivity obtained by RBF Kernel was better when compared to other kernels. The number of false positives (misclassified nodules) was 18 and 27 in training and testing phase respectively in the case of RBF Kernel.

Accuracy

Table 5.3 Accuracy of SVM Kernels

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Accuracy (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Standard Deviation</td>
<td>Testing</td>
</tr>
<tr>
<td>Linear (Dot)</td>
<td>84.67</td>
<td>0.62</td>
<td>82</td>
</tr>
<tr>
<td>Polynomial</td>
<td>87.33</td>
<td>0.56</td>
<td>84</td>
</tr>
<tr>
<td>Radial Basis Function (RBF)</td>
<td>90.33</td>
<td>0.48</td>
<td>85.33</td>
</tr>
</tbody>
</table>

The accuracy resulted of the proposed CAD system with different kernels of SVM for the training phase and testing phase is shown in Table 5.3.
Table 5.3 and Figure 5.4 show the performance measures of Linear, Polynomial and RBF Kernels of SVM and it is clear that RBF Kernel performs better.

**Classification Time**

Classification Time is the time taken by the classifier in 10-cross fold validations to identify the nodules using different kernels of SVM in training and testing phase.

**Table 5.4 Classification Time of CAD System with SVM Kernels**

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Classification Time (secs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
</tr>
<tr>
<td>Linear (Dot)</td>
<td>1.19</td>
</tr>
<tr>
<td>Polynomial</td>
<td>1.12</td>
</tr>
<tr>
<td>Radial Basis Function (RBF)</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Table 5.4 shows the average classification time of the proposed approach with different kernels of SVM. From the table it is observed that the
average classification time taken by the training and testing samples in SVM approach using RBF is just around 0.75 and 0.61 seconds.

![Classification Time of the Proposed CAD system with SVM](image)

**Figure 5.5 Classification Time of the Proposed CAD system with SVM**

Figure 5.5 shows the graphical representation of the classification time of the SVM.

### 5.5 SUMMARY

This approach provides a computer aided diagnosis system for detection of lung cancer. In the first phase of the proposed technique, the lung region is extracted from the chest tomography image. The different basic image processing techniques are used for this purpose. In the second phase, extracted lung is segmented with the help of Modified Fuzzy Possibilistic C-Means algorithm and the features are extracted. The classification is performed using Linear (dot), Polynomial and RBF Kernels of SVM to detect the occurrence of cancer nodules and the different performance measures are compared.

The next chapter deals with the next proposed approach called “Classification of Nodules extending Extreme Learning Machine”.