CHAPTER 4

ANALYSIS OF SVM AND KNN AS CLASSIFIERS FOR EIT IMAGES

The details of classifiers implemented in this investigation are presented in this chapter. Section 4.3 gives the framework of the image classification. Section 4.4 deals with the Support Vector Machine classifier and section 4.5 explain the k-Nearest Neighbour classifier.

4.1 INTRODUCTION

Humans have the abilities to classify the images without exertion than computers. Automatic image classification is a way for assigning a medical image to a class between a numbers of images groups. Classification is a decision-making theoretic approaches to the identification of images, based on their similarities. All classification techniques are based on the assumption that the image in demand describes one or more features and that each of these features belong to one of several dissimilar and exclusive classes. The main three classification techniques are supervised, unsupervised and semi supervised. In supervised technique, priori information or labeled datas are used for classification. In unsupervised technique, a large number of unknown pixels are examined and divides into a number of classes based on clustering. A semi-supervised (Zhu & Xiaojin 2005) technique is used unlabeled datas in order to avoid the expert.
4.2 LITERATURE SURVEY ON IMAGE CLASSIFIER

Niranjan & Ashish (2014) was Survey on Various Classification Techniques for Medical Image. Data Classification and Regression Trees (CART) algorithm was established by Leo Breiman et al. (1984). CART creates trees which have binary split on nominal or interval inputs for a nominal, ordinal or interval target. The CART algorithm supports the towing splitting criteria which can be used for multiclass problem. It uses the minimal cost complexity trimming to remove irrelevant features from the classifier.

The Cat Swarm Optimization decision tree algorithm was developed by Quinlan (1994) which can build either a derision tree or a rule set. It uses the information gain ratio to estimate split at each interval node of tree by slitting the data according to the field. Chi-squared Automated Interaction Detection (CHAID) algorithm was suggested by Gordon (1980) in which, chi-square statistics method is used to identify optimal split to build a decision tree. CHAID uses a Bonferroni adjustment for the number of categorical values of the input variables, thereby mitigating the bias towards inputs with many values.

Quick, Unbiased, Efficient, Statistical Tree (QUEST) algorithm was developed by Lim et al. (2000). QUEST is a binary classification method for building decision trees. In order to reduce the processing time required for CART, this method uses the quadratic disseminate analyses on selected input variables to get the best split, thereby improving the speed over. CART use to determine the optimal split.

Discriminate analysis was introduced by Fisher (1936) in which a multivariate statistical technique used to build a predictive or descriptive model of group discriminate based on observed predictor variables and
classify each observation into one of the groups. Several conventional classifiers are available but now a day’s most of the works mainly depend on Artificial Intelligence (AI) techniques which yield highly accurate results than the conventional classifiers. Michael et al. (2000) developed an interactive tool to classify the healthy and the tumorous MR brain images. But the drawback of this system is very low accuracy compared to the AI techniques. However this method offers a faster convergence rate. So this method mostly focused to improve the convergence rate. Ronald et al. (2000) have evidently mentioned the usage of Artificial Neural Networks (ANN) to increase the accuracy of the classifiers. A comparative analysis was performed along with the Linear Discriminant Classifier based on head and neck carcinoma detection.

The application of various ANN for image classification is studied by Egmont-Pertersen et al. (2002); Jude Hemanth et al. (2009); Jude Hemanth (2013); Niranjan & Ashish 2014). The poor convergence rate of the conventional neural networks is described in these reports. So these reports explain the importance of modified neural networks with greater convergence rate for image classification applications. Lukas et al. (2004) have classified four different types of tumor using LDA technique. But the classification accuracy is relatively very low. This work also recommended the various reasons for misclassifications.

Guo Zheng et al. (2006) have performed Support Vector Machine (SVM) based classification for MR glioma images. This method was performed better than rule based systems but the accuracy is low. Another drawback of this method is the lack of generalizing capability and used with only glioma images. Arjun et al. (2004) have used the Kohonen neural networks for image classification. In this work, some modifications have been done to improve the performance.
Sandeep et al. (2006) have used a combination of wavelets and Support Vector Machine (SVM) for classifying the abnormal and the normal images. This report shows that the hybrid SVM is performed better than the Kohonen neural networks. But the main drawback of this system is the small size of the dataset used for implementation. The classification accuracy results may reduce when the size of the dataset is increased.

A Least Square SVM (LS-SVM) for brain tumor recognition is proposed by Luts et al. (2007). Both bi-level classification and multiclass classification are performed in this work to show the superior nature of the proposed method over the conventional classifiers. This report also specified an important note that the differences between various algorithms increase when the number of classes increase. Thus, this work suggested the necessity for multiclass classification techniques than bi-level classification techniques.

Another version of LS-SVM is proposed by Selvaraj et al. (2007). A wide-range comparative analysis is performed between the SVM, neural classifier and the statistical classifiers. Results show that the SVM yields better classification accuracy. But only bi-level classification is performed in this work which is insufficient for judging the performance of the system.

Georgiadis et al. (2008) have developed the modified Probabilistic Neural Network for tumor image classification. The Least square feature transformation based Probabilistic Neural Network (PNN) was used for differentiating the abnormal images such as metastasis, glioma and meningioma. A comparative analysis is also performed with SVM. This work concluded that the transform based PNN is superior in terms of classification accuracy than the SVM.

Yamashita et al. (2008) have implemented the artificial neural networks to classify the different grades of abnormal images. This report
recommended a practical method for selection of database. The training of ANN is mainly dependent on input data and hence a wide range of pattern is required for high accuracy. This report also emphasized the difficulty in collecting a large dataset of different uncommon patterns.

A time efficient neural network such as PNN is used by Ibrahim (2008) for pattern classification problems. This work gives the importance for convergence than the classification accuracy. The result shows that the PNN is better than the conventional neural networks in terms of training time period. Georgiadis et al. (2008) have developed a computer aided system for discriminating the primary and secondary tumors in which Probabilistic Neural classifier is used. The report registers high classification accuracy but the size of the dataset is comparatively very small.

Felix & Liuis (2008) have implemented the Statistical classifiers for classifying the different types of tumor. This classification is performed on proton Magnetic Resonance Spectroscopy images. A comparative analysis with neural classifier is also performed in this work. This report concluded that a combined statistical and neural classifier increased the accuracy to higher extent.

Palaniappan (2009) has used an enhanced Adaptive Resonance Theory (ART) neural network for classification applications. This employed the Genetic Algorithm (GA) approach to select the order of training patterns to enhance the classification performance. This work is conducted on various datasets. But the main drawback is that the classification accuracy results are different for different datasets. Merinsky (2009) has implemented a self-organizing neural network based automated system for glioma detection. The main disadvantages of this system are the low classification accuracy and the lack of multiclass analysis.
Chandra et al. (2009) have used the RBF kernel based SVM for brain tumor detection. The results of SVM are compared with AdaBoost, a machine learning algorithm. Experimental results illustrated the superior nature of SVM over the other classifiers. Hamilton-Wright et al. (2007) has demonstrated image classification based on fuzzy approach using the pattern discovery algorithm. Experiments are conducted on various real-world datasets and the results concluded that the proposed algorithm yield good results when compared with the other classifiers.

Lin et al. (2006) have used the combination of SVM and fuzzy rules for pattern classification. The result shows that this method is accurate, fast and robust. Zhang et al. (2011) have suggested a new hybrid method to classify the images. Forward Neural Network (FNN) was used to design a new hybrid method. Mehdi (2011) has implemented the Back Propagation Network based Pattern recognition for classifying the brain tumor images. The drawback of this approach is the usage of PCA has enlarged the computational difficulty.

El-Sayed et al. (2010) have used the Back Propagation Network based bilevel classification technique for tumor images. (Maitra et al. 2006; Zhang et al. 2010) have implemented Feed forward neural networks based binary classification. But, the major disadvantages of these works are lesser accuracy and high convergence rate.

Zhang et al. (2011) proposed the semi supervised based maximum margin analysis for interactive image classification. A variety of relevance feedback (RF) schemes have been developed as a powerful tool to bridge the semantic gap between low-level visual features and high-level semantic concepts, and thus to improve the performance of the systems.
4.3 FRAMEWORK OF THE IMAGE CLASSIFICATION

In this work, three different types of classifiers are used to classify the EIT Lung images for abnormality detection as shown in Figure 4.1. The main focus in this chapter is only on SVM and kNN classifiers but ELM classifier is employed in the next chapter to support the comparative analysis. These classifiers are executed for normal and abnormal EIT image classification and the classifiers are analyzed in terms of classification accuracy.

![Flow diagram for the image classification](image)

Figure 4.1 Flow diagram for the image classification

4.4 SUPPORT VECTOR MACHINE

The EIT problem is an ill posed and the traditional neural network methods have grieved difficulties with generalization, creating models that can over fit the data. This is an importance of the optimization algorithms
used for parameter selection and the arithmetical procedures used to select the best model. Boser et al. (1992) have developed the Support Vector Machines (SVM) which is mostly accepted due to elegant features and encouraging performance. Support Vector Machines (SVM) is a statistical supervised machine learning technique which is used both for classification and for regression. The original SVM proposal was aimed at both the binary classification problem (Ruiz-Gonzalez et al. 2014), considering only two possible classification classes, and the multiclass classification problem, which considers more than two classification classes.

Binary linear SVM classification implements the calculation of the optimal hyperplane decision boundary, separating one class from the other, on the basis of a training dataset. Optimality can be assumed, depending on whether perfect classification of the training dataset is feasible and desired, in two separate ways (Ruiz-Gonzalez et al. 2014):

i) If perfect separability of training dataset classes can be achieved, Hard Margin optimality can be used. In this case, the hyperplane decision boundary is chosen to maximize the distance from the hyperplane to the nearest training data point.

ii) If perfect classification is not desired or if it is impossible, Soft Margin optimality is used. In this case, the hyperplane selection is a customizable tradeoff between minimizing the misclassification rate and maximizing the distance to the nearest properly classified training point.

The decision boundary hyperplane in SVM classification is calculated by employing the training dataset. This decision boundary is completely determined by Support Vectors, a subset of training input vectors
which by themselves alone lead to the same decision boundary. After this hyperplane is determined, the SVM classifier is ready to be used with a different dataset from the one used in the training stage. The assigned class, labeled either +1 or −1, depends on the side of the decision boundary on which the input vector falls. Figure 4.2 shows a linear SVM-based classification, both in the case of linearly separable classes and non-linearly separable classes.

Figure 4.2  SVM classifier corresponding to (a) a linearly separable pattern, (b) a non-linearly separable pattern

In mathematically, the general SVM linear binary classification problem can be formulated as follows:

For a given training dataset, \( \{x_i, d_i\}_{i=1}^N \), the main goal is to calculate the optimal weight vector \( w \), such that satisfy the following constraints:

\[
d_i(w^T x_i + b) \geq 1 - \xi_i \forall i = 1, 2, \ldots, N
\]

\[
\xi_i \geq 0, \quad \forall i = 1, 2, \ldots, N
\]

and such that the following cost function is minimized:
\[ \Phi(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \] (4.2)

where, \( x_i \in \mathbb{R}^{m_0} \) indicates the \( i \)th input vector, \( d_i \in \{-1, 1\} \) represents the class corresponding to the \( i \)th input vector, \( \xi = \{\xi_i\}_{i=1}^{N} \) represents the slack variables, and the constant \( C \) is a user-specified parameter that determines the tradeoff between misclassification and maximum inter class margin.

In practice, most classification problems cannot be solved by using a simple hyperplane as the decision boundary. In such cases a more complex and elaborate decision boundary is required. SVM attains this goal by increasing the dimensionality of the input space, of dimension \( m_0 \), by applying a nonlinear transformation, denoted by \( \varphi(\cdot) \), into a feature space of dimension \( m_f > m_0 \). This transformation aids to reduce the misclassification probability in the transformed feature space. The most commonly used transformation functions are radial basis functions, higher-order polynomials, and sigmoids. Figure 4.3 represents a graphical example of an SVM nonlinear classification.

\[ \text{Figure 4.3 SVM classifier with a nonlinear kernel (a) the input space (b) the feature space} \]
The boundary in the nonlinear classification problem is still a hyperplane, not in the original input space but in the feature space, and can be expressed as the point’s \( \phi(x) \) that satisfies that:

\[
    w^T \phi(x) + b = 0
\]

(4.3)

where, \( x \in \mathbb{R}^m \) and \( \phi(x) \in \mathbb{R}^{m_f} \).

Following the application of the Lagrange multipliers method, it has been shown that the optimal weight vector can be expressed as (Haykin 1998):

\[
    w = \sum_{i=1}^{N} \alpha_i d_i \phi(x_i)
\]

(4.4)

where, \( \alpha_i \) stands for the Lagrange multiplier coefficients. Therefore, the optimal decision boundary can be rewritten as:

\[
    \sum_{i=1}^{N} \alpha_i d_i \phi(x_i)^T \phi(x) + b = 0
\]

(4.5)

Renaming \( u_i = \alpha_i d_i \) and \( K(x, x_i) = \phi(x_i)^T \phi(x) = \phi(x)^T \phi(x) = K(x, x_i) \), the decision function, \( y \), can be expressed as:

\[
    y = \sum_{i=1}^{N} u_i K(x, x_i) + b
\]

(4.6)

In case of linear classifiers, \( K(x, x_i) \) is the conventional Euclidean inner product of the input vector \( x \) with the Support Vector \( x_i \). In case of nonlinear classifiers, \( K(x, x_i) \) is the conventional Euclidean inner product of the nonlinear transformation \( \phi(x) \) of the input vector \( x \) with the nonlinear transformation \( \phi(x_i) \) of the Support Vector \( x_i \). The decision function in
equation (4.6) results in the architecture depicted in Figure 4.4, once the proper weights and Support Vectors have been computed in the training stage. Only the Support Vectors have to be considered, as they are the only vectors that generate non-zero $a_i$ coefficients.

Classification is executed by identifying the sign of the output value, $y$, in equation (4.6). If $\text{sign}(y) = +1$, then this input is labeled as class +1 and if otherwise as class −1.

Figure 4.4 Architecture of a SVM classifier

The most well-known and widely used nonlinear kernels are Radial Basis Function (RBF), sigmoids and polynomials.

i) The RBF kernel can be expressed as $K(x, x_i) = \exp(-\gamma \|x-x_i\|^2), \gamma > 0$ where $\gamma$ is a user-defined parameter;

ii) the sigmoidal kernel can be expressed as $K(x, x_i) = \tanh(\gamma x^T x_i)$, where $\gamma > 0$ and $c_0 < 0$ are user-defined parameters;
iii) The d-order polynomial kernel can be expressed as $K(x, x_i) = \tanh (\gamma x^T x_i + c_0)^d$, where $\gamma$ and $c_0$ are user-defined parameters and where $d$ denotes the polynomial degree.

4.4.1 Training Algorithm of SVM

Step 1 The inputs are created as set of features.

Step 2 These set of features are mapped into a feature space using the kernel function.

Step 3 A segmentation is calculated in the feature space to isolate the classes of training vectors

4.5 K-NEAREST NEIGHBOUR

K-Nearest Neighbours algorithm is a non-parametric method used for classification. kNN is an instance-based learning method in which the function is estimated locally and all calculation is put off until classification is over. The $kNN$ algorithm is very simple machine learning algorithm. Each sample in data set has $n$ attributes which was combine to form an $n$-dimensional vector:

$$X = (x_1, x_2, \cdots, x_n)$$

(4.7)

These $n$ attributes are the independent variables. Each sample has the dependent variable, denoted by $y$, whose value mainly depends on $n$ attributes $x$. Assume that $y = f(x)$ is a categoric variable and $f$ is a scalar function. Suppose set of $T$ vectors are given together with their corresponding classes:

$$X^i, y^i \text{ for } i = 1, 2, 3, \ldots, T$$

(4.8)
This set of T is referred as the training set. Suppose a given new sample is $X = u$ and need to find the class of this sample belongs. If the function $f$ is known value, then $v = f(u)$ can be easily calculated to know the class this new sample, but here $f$ is unknown. The main idea in kNN methods is to recognize k samples in the training set whose independent variables $x$ are similar to $u$, and to use these k samples to classify this new sample into a class, $v$. Assume that $f$ is a smooth function which is used to look the nearer variable from the samples in training data and then to calculate the values of $v$ from the values of $y$ for these samples.

A neighbour is a distance measure which can be calculated between samples based on the independent variables. For this study, the Euclidean distance is used to find the distance between the samples. The Euclidean distance between the points $x$ and $u$ is

$$d(x, u) = \sqrt{\sum_{i=1}^{n} (x_i - u_i)^2}$$ (4.9)

In case of $k=1$, find the sample in the training set that is closest to $u$ and fix as $v = y$. Here $y$ is the class of the nearest neighbouring sample. In case of large number of samples in training set, this single nearest neighbour can be very powerful tool to classify samples.

For kNN case, the idea of 1-NN can be extended as follows. First find the nearest k neighbours of $u$ and then use a majority decision rule to classify the sample. When noise present in the training data, the higher values of $k$ offers smoothing which decreases the possibility of over-fitting. When $k = n$, the number of samples in the training data set are simply guessing the class that has the majority in the training data for all samples irrespective of $u$. 
This leads to over-smoothing because there is no information about the dependent variable in the independent variables.

4.6 RESULTS AND DISCUSSION

The experiments are performed on the Intel Core i3 processor with speed of 2.53 GHz and 4 GB RAM. The software used for these work is MATLAB version 8.0. There are two classifiers are discussed in this chapter which are trained and tested independently with the datasets.

The SVM which is implemented in this work is binary classification. In general, the RBF kernel is a reasonable first choice. This kernel non-linearly maps samples into a higher dimensional space so it can handle the case when the relation between class labels and features is non-linear. In addition, the sigmoid kernel behaves like RBF for certain parameters (Lin & Lin 2003). Another reason is the number of hyperparameters which influences the complexity of model selection. The polynomial kernel has more hyperparameters than the RBF kernel. So, RBF is used as the parameter for kernel type. The gamma value is calculated based on the number of training features and rho value is set as default. Two classes of data are used. The labels for these classes are using ‘0’ and ‘1’ for ‘normal’ and ‘abnormal’ respectively. The number of normal images used for training set is 25 whereas for abnormal images is 21. For the testing, normal and abnormal images are used is 10 and 9 respectively. Numerous testing have been done with EIT images.

Figure 4.5 shows the classification results of GA-SVM classifier for RBF kernel. In this case, the number of features in the input data is ten. So the calculated value of gamma for these features is 0.1 and the constant C is defined as 20 which control the tradeoff between misclassification and maximum inter class margin. As a result, one normal image is classified as
abnormal. The classification accuracy of GA-SVM classifier with RBF kernel for testing data is 95% under the gamma value of 0.1.

Figure 4.5 Outputs of GA-SVM classifier with RBF kernel

Figure 4.6 shows the classification results of Inf-SVM classifier with RBF kernel. In this classifier, one normal image is classified as abnormal images. The classification accuracy of Inf-SVM classifier with RBF kernel for testing data is 95% under gamma value of 0.1. In Figure 4.5 and 4.6, the combination of blue circle and red squares are representing the normal images and other colors are representing the abnormal images.

To assess the accuracy of an image classification, it is common practice to create a confusion matrix. A confusion matrix is a table with two rows and two columns that report the number of False Positive (FP), False Negative (FN), True Positive (TP) and True Negative (TN). In a confusion matrix, the classification results are compared to additional ground truth information.
The strength of a confusion matrix is that it identifies the nature of the classification errors, as well as their quantities. The results of confusion matrix highly depend on the selection of ground truth / test set pixels.

In machine learning, a confusion matrix is a specific table layout that allows visualization of the performance of an algorithm. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The results of 10 different trails of confusion matrix for the selected features using the both classifiers are averaged separately for the performance analysis. Based on the confusion matrix values, Sensitivity, specificity, Accuracy and False discovery rate (FDR) were calculated. Tables 4.1 and 4.2 shows the confusion matrix of GA-SVM classifier and Inf-FS-SVM classifier for EIT images respectively.

Figure 4.6 Outputs of Inf-FS-SVM classifier with RBF kernel
Table 4.1  Confusion Matrix of GA-SVM Classifier for EIT Images to detect the abnormality

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>accuracy</th>
<th>FDR</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>20</td>
<td>20</td>
<td>1</td>
<td>5</td>
<td>0.800</td>
<td>0.952</td>
<td>0.870</td>
<td>0.130</td>
<td>0.200</td>
</tr>
<tr>
<td>Testing</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0.900</td>
<td>0.889</td>
<td>0.895</td>
<td>0.105</td>
<td>0.100</td>
</tr>
<tr>
<td>Overall</td>
<td>29</td>
<td>28</td>
<td>2</td>
<td>6</td>
<td>0.829</td>
<td>0.933</td>
<td>0.877</td>
<td>0.123</td>
<td>0.171</td>
</tr>
</tbody>
</table>

Table 4.2  Confusion Matrix of Inf-FS-SVM Classifier for EIT Images to detect the abnormality

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>accuracy</th>
<th>FDR</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>20</td>
<td>20</td>
<td>1</td>
<td>5</td>
<td>0.800</td>
<td>0.952</td>
<td>0.870</td>
<td>0.130</td>
<td>0.200</td>
</tr>
<tr>
<td>Testing</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0.900</td>
<td>0.889</td>
<td>0.895</td>
<td>0.105</td>
<td>0.100</td>
</tr>
<tr>
<td>Overall</td>
<td>29</td>
<td>28</td>
<td>2</td>
<td>6</td>
<td>0.829</td>
<td>0.933</td>
<td>0.877</td>
<td>0.123</td>
<td>0.171</td>
</tr>
</tbody>
</table>

Table 4.3 shows the classification results of GA-kNN classifier. The number of nearest neighbors used in the GA-kNN classifier is 1 (k=1). The Euclidean distance is used to find the distance between the samples. The overall classification accuracy for GA-kNN classifier is 86.2%. In this case, two abnormal images are classified as normal images and seven normal images are classified as abnormal images. Totally 56 EIT images are correctly classified. Table 4.4 shows the classification results of Inf-FS-kNN classifier. The number of nearest neighbors used in the Inf-FS-kNN classifier is also 1 (k=1). The classification result for the class 0 is 87.0% and for the class 1 is 94.7%. The overall classification accuracy for Inf-FS-kNN classifier is
89.2%. In this case, two abnormal images are classified as normal images and one normal image is classified as abnormal image.

Table 4.3 Confusion Matrix of GA-kNN Classifier for EIT Images to detect the abnormality

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>accuracy</th>
<th>FDR</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>20</td>
<td>19</td>
<td>1</td>
<td>6</td>
<td>0.769</td>
<td>0.950</td>
<td>0.848</td>
<td>0.048</td>
<td>0.231</td>
</tr>
<tr>
<td>Testing</td>
<td>8</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0.889</td>
<td>0.900</td>
<td>0.895</td>
<td>0.111</td>
<td>0.111</td>
</tr>
<tr>
<td>Overall</td>
<td>28</td>
<td>28</td>
<td>2</td>
<td>7</td>
<td>0.800</td>
<td>0.933</td>
<td>0.862</td>
<td>0.067</td>
<td>0.200</td>
</tr>
</tbody>
</table>

Table 4.4 Confusion Matrix of Inf-FS-kNN Classifier for EIT Images to detect the abnormality

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>accuracy</th>
<th>FDR</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>21</td>
<td>19</td>
<td>2</td>
<td>4</td>
<td>0.840</td>
<td>0.905</td>
<td>0.870</td>
<td>0.130</td>
<td>0.160</td>
</tr>
<tr>
<td>Testing</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0.900</td>
<td>1.000</td>
<td>0.947</td>
<td>0.053</td>
<td>0.100</td>
</tr>
<tr>
<td>Overall</td>
<td>30</td>
<td>28</td>
<td>2</td>
<td>5</td>
<td>0.857</td>
<td>0.933</td>
<td>0.892</td>
<td>0.108</td>
<td>0.143</td>
</tr>
</tbody>
</table>

Receiver Operating Characteristic (ROC) curves is generally used in binary classification to evaluate the output of a classifier. A ROC curve consists of true positive rate on the Y axis, and false positive rate on the X axis which means that a larger area under the curve is usually better. The steepness of ROC curves is very important, since it is model to maximize the true positive rate while minimizing the false positive rate. The figure 4.7 shows the ROC curve for the GA-kNN, Inf-kNN, GA-SVM and Inf-SVM.
classifiers for trail 1. The area under the curve (AUC) values for GA-kNN, Inf-kNN, GA-SVM and Inf-SVM classifiers are 93.5%, 90%, 94% and 95% respectively for trail 1.

![ROC curve](image)

**Figure 4.7 ROC curve**

### 4.7 CONCLUSION

In this chapter, two classifiers are proposed for EIT image classification. Confusion matrices are used to evaluate the results of classification. Therefore, the classification error rate can be easily implicit from a confusion matrix. The results of these classifiers are compared for the two different feature selection techniques. The Feature selection based classifiers are found to hold the ability of precisely classifying the EIT images. The performance analysis can be done based on a confusion matrix which is described in the chapter 6.