CLASSIFIER FOR THE CLASSIFICATION OF POTENTIAL MICROCALCIFICATION

The microcalcifications are considered to be important signs of breast cancer. It has been reported in literature that 30% - 50% of breast cancer detected radiographically show microcalcifications on mammograms. Histologic examinations report 62 to 79% of breast carcinomas reveals microcalcifications. Correlation between the presence of microcalcifications and the presence of the breast cancer suggests that the accurate detection of microcalcifications will improve the efficacy of mammography as a diagnostic procedure.

5.1 Introduction

Microcalcifications detection task for the diagnosis of breast cancer is a difficult one. In most of the cases simple oversight by radiologists and dense breasts may lead to the failure of microcalcifications detection. Figure 5.1 shows microcalcification clusters.

![Figure 5.1 Microcalcification clusters](image)

Following factors contribute major problems in analyzing and detecting microcalcifications in mammograms.

- Microcalcifications are very tiny objects and can be described as granular, linear, or irregular. The sizes of microcalcifications vary from 0.1 mm to 1.0 mm, with an average diameter of about 0.3 mm. Due to superimposition of microcalcification on the breast parenchymal textures and noise, tiny microcalcification can hardly be seen on image.
Classifier for the Classification of Potential Microcalcification

- In mammogram, microcalcifications often appear in an inhomogeneous and some region of the background may be brighter than the microcalcifications. In some case microcalcifications have low contrast to the background and the intensity and size of the microcalcifications can be very close to noise.

In order to preserve all true MCs and to reduce the number of false positives in this stage, thresholds must be carefully determined. This chapter focuses on the design of new CAD techniques for the detection and classification of the potential microcalcification. The proposed methodology consists of segmentation, feature extraction and classification of potential microcalcification. The method does not treat the detection of microcalcification as simple blob detection. Section 5.2 explains the detection of potential microcalcification using a new approach. Section 5.3 explains the feature selection. Techniques used for the classification of potential microcalcifications are explained in section 5.4. Results and discussions are drawn in section 5.5. Conclusion is drawn in section 5.6.

5.2 Segmentation

Segmentation technique uses an entropy thresholding algorithm and is the modified algorithm of [104]. Primary step of algorithm filter mammogram with a morphological white top hat for detecting the small bright points. Top hat filter has two steps. First step is a morphological opening which involves erosion followed by dilation. Second step enhances the visibility and detectability of microcalcifications. This step involves subtraction of the opened image from the original image. The new segmentation scheme accepts the filtered image and uses the ranked thresholds which are explained in next section to detect all the maximum number of points inside the mammogram. These points are later transformed to the potential microcalcification based on some selection methods.

5.2.1 Detection of potential microcalcification

This phase removes the background without reducing the microcalcifications by a top hat filter, fixes optimal thresholds for segmentation of filtered image and detects the potential microcalcification. The new algorithm for point detection is explained in figure 5.2.
Let $i(x,y)$ be the digitized mammogram image $IM$ of size $N \times N$ with $L$ grey levels. Perform the following steps.

1. Filter the image $IM$ by a top hat filter.
2. For all the grey levels of $g$ from 1 to $L$ of $IM$ compute the image entropy $[104]$. Image entropy and the corresponding $g$ are stored in the arrays $IE$ and $G$ respectively.
3. Rank the array $IE$ (i.e., obtain the descending order of $IE$) and then rank the corresponding $G$ array. Array $G$ stores the optimal thresholds (grey levels) in the descending order.
4. To detect greater possible quantity of points the top 9 optimal thresholds are used to segment the filtered image. When each of these thresholds is applied the newly obtained points are added and the repeated points are deleted.

**Figure 5.2 Algorithm for point detection**

Two selection procedures are used to transform a point to potential microcalcification. Area is used as the first selection method to transform the detected points to potential microcalcification. Second selection method uses grey gradient to transform the detected points to potential microcalcification. After the transformation potential microcalcification are classified as microcalcification and non-microcalcification with the available truth information and with the support of radiologists. Only 5% of potential microcalcification was microcalcifications and others were not. Percentage indicates a class imbalance problem and this issue is addressed by using classifiers that introduces balanced learning for the accurate classification of potential microcalcifications.

**5.3 Feature Extraction**

In order to accurately classify the potential microcalcification into microcalcification a set of features are extracted. The following features are extracted. Average grey level, standard deviation of grey level, edge strength, background grey level, foreground background ratio and difference, compactness, elongation, shape and invariant moment, second order histogram related features, five features and three features related to contrast. These features are passed through a feature selection process and features that present high correlation with other features are removed.
Classifier for the Classification of Potential Microcalcification

Feature selection use CfsSubsetEval (WEKA Machine learning Tool). It prefers subsets of features that are highly correlated with the class while having low intercorrelation. After the feature selection procedure the following features are selected. Absolute contrast, standard deviation of grey levels, difference ratio, area, compactness, entropy, angular second moment, correlation, sum entropy.

5.4 Classification of Potential Microcalcification

Most of the traditional classifiers generally perform well if number of instances for both classes is perfectly balanced. If training aims overall accuracy many traditional methods perform well only during the testing time when classes (i.e., Non-Microcalcification and Microcalcification classes) in data set are relatively balanced. Most of traditional learning methods predict only the majority class (Non-Microcalcification) correctly but not the minority class (Microcalcification). This section explains the array of techniques that are used to test the potential microcalcification classification performance for the imbalanced data sets. Methods discussed in this chapter for the classification of potential microcalcification handles the imbalanced class problem effectively.

Method 1. CCW-kNN

When dealing with highly imbalanced data, kNN algorithms lead to suboptimal classification performance on the minority class (Microcalcification). This issue is addressed by a CCW (Class Confidence Weights) [105] that uses the probability of attribute values given class labels to weight prototypes in kNN. CCW is able to correct the inherent bias to the microcalcification class in kNN algorithms on any distance measurement.

Let \((a_i, b_i)\) (for \(i = 1, n\)) represent training data, \(a_i\) are feature vectors, \(b_i\) belongs to \{Non-Microcalcification, Microcalcification\}. kNN algorithm finds group of prototypes from the training set that are closest to a test instance by a certain distance measure. When the \(k\) neighbours vary widely in their distances and nearest neighbours are reliable, the neighbours are weighted by the multiplicative-inverse (MI) or the additive-inverse (AI) of their distances.

**MI:**
\[
b_i' = \arg\max_c \left\{ \sum_{a_i \in \phi(a_i)} \sum_{b_i = c} T(b_i = c) \cdot (1/dist(a_i, a_)) \right\}
\]

**AI:**
\[
b_i' = \arg\max_c \left\{ \sum_{a_i \in \phi(a_i)} \sum_{b_i = c} T(b_i = c) \cdot (1 - dist(a_i, a_i) / \text{dist maximum}) \right\}
\]
where $b't$ is a predicted label, $T(\cdot)$ is an indicator function that returns 1 if its condition is true and 0 otherwise, $\phi(a_t)$ denotes the set of $k$ training instances closest to $a_t$, $dist(a_t, a_i)$ represents the distance between the test point $a_t$ and a prototype $a_i$, and $dist_{maximum}$ is the maximum possible distance between two training instances in the feature space which normalizes $dist(a_t, a_i)/dist_{maximum}$ to the range of [0,1]. CCW-kNN capture the probability of attributes values given a class label. CCW on a training instance $i$ is defined as follows:

$$w^{CCW}_i = p(a_i|b_i)$$

where $a_i$ and $b_i$ represent the attribute vector and the class label of instances $i$. Now the classification rule integrated with CCW is:

$$b't = \arg\max_{c \in \{Non-Microcalcification, Microcalcification\}} \sum_{a_i \in \phi(a_t)} T(b_i = c) \cdot w^{CCW}_i$$

CCW changes the bases of $k$NN rule from using priors to posteriors and handle effectively the class imbalance problem.

### Method2. CCP-C4.5 and CCP-CART

Performance of C4.5 and CART depends on the assumption that there is an equal amount of information for both classes contained in the training data. C4.5 and CART will have poor performance if the training data set tends to have an imbalanced class distribution. CCP-based decision trees [106] (CCP-C4.5 & CCP-CART) are used which uses CCP-embedded entropy/Gini index as splitting criteria, and Fisher’s Exact Test (FET) as the determinant of when to prune the branches. CCP is the measure of splitting attributes during decision tree construction. The approach of replacing a conventional splitting measure like entropy or Gini index by CCP is a generic mechanism for C4.5 and CART decision trees that perform better for balanced data sets. Approach can be applied to decision tree algorithm such as C4.5 and CART. For imbalanced data set, high confidence rules do not necessarily imply high significance in imbalanced data, and some significant rules may not yield high confidence so that the splitting criteria in C4.5 is suitable for balanced but not imbalanced data sets. Decision tree based on CART will too suffer from the imbalanced class problem. The weakness of the support-confidence framework and the factor that results in the poor performance of entropy and Gini index are well addressed by a new measure called Class Confidence Proportion (CCP). Suppose a
training data set which consists of n records, and the antecedents (denoted by A and Neg A) and class (b and Neg b) distributions are in the form of Table 5.1.

Table 5.1 Notations for class based on association analysis

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>Neg A</th>
<th>Total instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>p</td>
<td>q</td>
<td>p+q</td>
</tr>
<tr>
<td>Neg b</td>
<td>r</td>
<td>s</td>
<td>r+s</td>
</tr>
<tr>
<td>Total attributes</td>
<td>p+r</td>
<td>q+s</td>
<td>n</td>
</tr>
</tbody>
</table>

During the calculation of CCP measurement the method will focus on each class (i.e., Non-Microcalcification and Microcalcification) and find the most significant antecedents associated with that class. Even though the high frequency with which the Non-Microcalcification class present together with antecedent does not mean antecedent "implies" the Non-Microcalcification class as this particular class is a majority class. Class Confidence (CC) is defined as $\text{CC}(A \rightarrow b) = \text{Supp}(A \cup b)/\text{Supp}(b)$. It finds the most interesting antecedents from all the classes. In the notation of Table 5.1 CC can be expressed as $\text{CC} (A \rightarrow b) = TP/(TP+FN)$ and $\text{CC}(A \rightarrow \text{Neg } b) = FP/(FP+TN)$, where denominator represents the actual positive and negative instances respectively. CC is focusing on the recall than the precision. It can be noted that if CC is high then the rules will be the significant ones. But high CC rules obtained are not sufficient for solving classification problems. It should be ensured that the classes implied by those rules are not only of high confidence, but more interesting than their corresponding alternative classes. So the proportion of one Class Confidence (CC) over that of all classes is the Class Confidence Proportion (CCP). The CCP of rule $A \rightarrow b$ is defined as

$$\text{CCP}(A \rightarrow b) = \text{CC}(A \rightarrow b) / (\text{CC}(A \rightarrow b)+\text{CC}(A \rightarrow \text{Neg } b))$$

A rule with high CCP means that, compared with its alternative class, the class this rule implies has higher CC, and consequently is more likely to occur together with this rule’s antecedents regardless of the proportion of classes in the data set.

Method 3. SMOTE + C4.5

As explained earlier in previous section C4.5 alone cannot perform well in the classification of imbalanced data sets. One positive solution to the imbalanced dataset
problem is a pre-processing step in order to balance the class distribution. So the imbalanced data sets are preprocessed using SMOTE before C4.5 is trained.

**Synthetic Minority Over-sampling Technique (SMOTE)**

SMOTE [107] is an over-sampling technique. SMOTE over sampling approach attempts to add information to the training set by introducing new non-replicated minority class examples. This technique generates synthetic instances rather than replicating minority class instances. The algorithm is described in figure 5.3.

| A is the original data set and B is the set of positive instances (minority class instances) |
| For each instance m in B |
| Find the k-nearest neighbors (minority class instances) to m in B |
| Obtain n by randomizing one from k instances |
| difference = m – n |
| gen = random number between 0 and 1 |
| x = m + difference * gen |
| Add x to O |
| End for |

Figure 5.3. The Synthetic Minority Over-sampling Technique (SMOTE)

**Method 4. Safe-Level SMOTE + C4.5**

This method preprocesses the imbalanced data sets using Safe-Level SMOTE algorithm before C4.5 is trained. Safe-Level-Synthetic Minority Oversampling Technique [108] assigns each positive instance its *safe level* before generating synthetic instances. Each synthetic instance is positioned closer to the largest safe level so all synthetic instances are generated only in safe regions. The instance is nearly noise if the safe level of an instance is close to 0. Safe-Level-SMOTE algorithm is showed in figure 5.4.

**Description of variables used in algorithm**

p is an instance in the set of all original positive instances A.

n is a selected nearest neighbours of p.

s is a synthetic instance.
bln is safe level of p
xy is safe level of n
sl_ratio is safe level ratio.
NUMAT is the number of attributes.
DIFFERENCE is the difference between the values of n and p at the same attribute id.
gap is a random fraction of DIFFERENCE.
|A| is the number of all positive instances in A
A’ is a set of all synthetic instances returned when the algorithm terminates

|Input: A set of all original positive instances A |
|Output: A set of all synthetic positive instances A’ |
1. A’ = ∅ 
2. for each positive instance p in A { 
3. compute k nearest neighbours for p in A and randomly select one from the k nearest neighbours, call it n 
4. bln = the number of positive stances in k nearest neighbours for p in A 
5. xy = the number of positive stances in k nearest neighbours for n in A 
6. if (xy ≠ 0) { ; sl is safe level. 
7. sl_ratio = bln / xy ; sl_ratio is safe level ratio. 
8. } 
9. else { 
10. sl_ratio = ∞ 
11. } 
12. if (sl_ratio = ∞ AND bln = 0) { ; the 1st case 
13. does not generate positive synthetic instance 
14. } 
15. else { 
16. for (atti = 1 to NUMAT) { ; NUMAT is the number of attributes. 
17. if (sl_ratio = ∞ AND bln ≠ 0) { ; the 2nd case 
18. gap = 0 
19. } 
20. else if (sl_ratio = 1) { ; the 3rd case 
21. generate a random number between 0 and 1, call it gap 
22. } 
23. else if (sl_ratio > 1) { ; the 4th case 
24. generate a random number between 0 and 1/sl_ratio, call it gap 
25. } 
26. else if (sl_ratio < 1) { ; the 5th case 
27. generate a random number between 1-sl_ratio and 1, call it gap 
28. } 
29. DIFFERENCE = n[atti] - p[atti] 
30. s[atti] = p[atti] + gap * DIFFERENCE 
31. } 
32. A’ = A’ ∪ {s} 
33. } 
34. }
Method 5. Borderline-SMOTE + C4.5

This method preprocesses the imbalanced data sets using Borderline SMOTE algorithm [109] before C4.5 is trained. Borderline-SMOTE are different from many over-sampling methods in which all the minority examples or a random subset of the minority class are over-sampled. It is based on SMOTE (Synthetic Minority Over-sampling Technique). k Nearest Neighbours of the same class are calculated for every minority example and random selection of some examples are performed according to the over-sampling rate. Then along the line new synthetic examples are generated between its selected nearest neighbours and the minority example.

C4.5 is considered as a base classifier along with the above sampling techniques since in imbalanced domains it has been widely used. Also it has been considered as one of the top ten data mining algorithm.

5.5 Results and Discussions

In this section, results of potential microcalcification detection and experimental results from CCW-kNN, CCP-C4.5, CCP-CART, SMOTE + C4.5, Safe-Level SMOTE + C4.5 and Borderline-SMOTE + C4.5 classifiers are presented. Quantitative evaluations are used to validate the effectiveness of proposed techniques. Proposed techniques described in this chapter use 22 out of 250 mammogram images used in the breast contour extraction technique discussed in the last chapter. These 22 mammogram images contain microcalcification clusters. The following sections explain the experimentation and results of every stage of the study.

5.5.1 Results of potential microcalcification detection

First step in the segmentation phase detected the potential microcalcification using top hat morphological filtering that remove the mammogram background and attenuates the potential microcalcification. It enhances the image contrast. Structuring element as square kernels of size 21x21 has been used to get the best results. In second step, from all the grey levels from 1 to L image entropy has been computed, stored computed values in the array and ranked them (descending order). As discussed earlier
in the section (detection of potential microcalcification) maximum available points are detected using top 9 thresholds and subsequently using two selection methods these points are transformed to potential microcalcification. Application of top 9 thresholds accumulates all available points and contributes in the detection of maximum available potential microcalcifications. This procedure also enhances the detection and classification of high percentage of microcalcification.

Figure 5.5 shows the applications of top 9 thresholds, top 7 thresholds, and top 6 thresholds to detect the points in mammogram. It is observed that increase in number of top thresholds proportionally increase the number of detected points in image. One interesting factor observed in the experimental analysis is the application of top ranked threshold for point detection. Application of this threshold results in points detection and it is observed that 92% of points were microcalcifications. But application of this threshold results in some false positive detection.

![Images](a) (b) (c)

Figure 5.5 Applications of thresholds to detect the points from mammogram (a) top 9 thresholds (b) top 7 thresholds (c) top 6 thresholds

5.5.2 Unbalanced samples of dataset

Points are passed through the selection methods as discussed earlier for selecting potential microcalcification. The additional information included with the MIAS data set includes regions in the mammograms where microcalcifications are located. It is supposed that potential microcalcification within these regions are mainly microcalcifications. With this information and the support of radiologists, all the potential microcalcification located in these 22 mammograms was preclassified into microcalcification, and non-microcalcifications. But percentage of microcalcifications was very less. This indicates class imbalance problem in the data set. After the microcalcification features were extracted from each potential microcalcification, the
Classifier for the Classification of Potential Microcalcification

feature selection processes extract only the relevant features. Aimed to improving the classification performance several methods are used in proposed work to deal the unbalanced data sets. Array of methods like CCW-kNN, CCP-C4.5, CCP-CART, SMOTE + C4.5, Safe-Level SMOTE + C4.5 and Borderline-SMOTE + C4.5 has been used to address the issue of this imbalanced distribution of elements in each class. Most of these methods reduced false positive detection.

5.5.3 Parameter selection

Each classifier model described in the previous section is associated with few model parameters. This parameters need to be fine-tuned for best performance. For each classifier model the parameter specification is set as shown below.

Configuration parameter for CCW-kNN
- Number of neighbours (k) = 5

Configuration parameter for CCP based decision trees
- Confidence level of Fisher’s Exact Test (FET) for pruning = 0.01

Parameter specification for C4.5 + SMOTE

Parameter settings for C4.5
- Prune = True
- Confidence level = 0.25
- Minimum number of item-sets per leaf = 2
- Confidence = Laplace smoothing

Parameter settings for SMOTE
- Number of neighbours (k) = 5
- Quantity = Balance
- Distance = Heterogeneous value difference
- Percentage = 100 metric

Configuration parameters for Safe Level SMOTE+C4.5

Parameter settings for C4.5
- Prune = True
- Confidence level = 0.25
- Minimum number of item-sets per leaf = 2
- Confidence = Laplace smoothing

Safe Level SMOTE
Classifier for the Classification of Potential Microcalcification

- Number of neighbours (k) = 5

**Configuration parameters for Borderline SMOTE+C4.5**

Parameter settings for C4.5
- Prune = True
- Confidence level = 0.25
- Minimum number of item-sets per leaf = 2
- Confidence = Laplace smoothing

Borderline SMOTE
- Number of neighbours (k) = 5

### 5.5.4 ROC analysis

Quantitative evaluations are used to validate the effectiveness of proposed methods. In a two-class problem containing positive and negative samples it has been mentioned the true positive and true negative as correctly classified positive and negative samples, false positive and false negative for incorrectly classified positive and negative samples. For quantitative evaluations following metrics are determined as follows.

\[
\text{Recall (TPrate)} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad \text{and} \quad \text{FPrate} = \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}}
\]

For a given classifier an ROC curve is a plot of the classification sensitivity as the ordinate versus the specificity as the abscissa.

### 5.5.5 Performance by different classifier models

Before starting with the analysis, the results for all the proposed methods in the experimental study are summarized in Table 5.2. For more meaningful interpretation the classification results are summarized in the figure 5.6. Figure 5.7 to 5.12 show the ROC Curve for all classifier models. Ten-fold cross-validation was performed and means and standard deviations of the metrics are reported.

---

Table 5.2 Mean and standard deviation of the measures of proposed methods
### Results of Classifier Models

<table>
<thead>
<tr>
<th>Classifier Models</th>
<th>Mean AUC</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCW-kNN</td>
<td>.909</td>
<td>.0316</td>
</tr>
<tr>
<td>CCP-CART</td>
<td>.771</td>
<td>.0461</td>
</tr>
<tr>
<td>CCP-C4.5</td>
<td>.9201</td>
<td>.0582</td>
</tr>
<tr>
<td>SMOTE+ C4.5</td>
<td>.7097</td>
<td>.0231</td>
</tr>
<tr>
<td>Safe-Level-SMOTE+ C4.5</td>
<td>.8554</td>
<td>.0264</td>
</tr>
<tr>
<td>Borderline-SMOTE+ C4.5</td>
<td>.8706</td>
<td>.0267</td>
</tr>
</tbody>
</table>

#### Figure 5.6

Results obtained with different classifier models.
Figure 5.7 ROC Curve for CCW-kNN

Figure 5.8 ROC Curve for CCP-C4.5

Figure 5.9 ROC Curve for CCP-CART
Classifier for the Classification of Potential Microcalcification

Figure 5.10 ROC Curve for SMOTE+C4.5

Figure 5.11 ROC Curve for SAFELEVEL-SMOTE+C4.5

Figure 5.12 ROC Curve for BORDERLINE-SMOTE+C4.5
Several two-tailed Student's t-tests at a level of significance of .05 were performed in order to compare the mean measure of each method with the mean of the other methods. In Table 5.3 the results of the statistical comparison between mean measures of each method versus the mean of the other methods has been shown. A statistical comparison between each method yields a two-tailed value for rejecting / accepting the null hypothesis that their corresponding ROC curves have the same area under them.

Table 5.3 Results of the statistical comparison between the mean measures of each method versus the mean of the other methods

<table>
<thead>
<tr>
<th></th>
<th>CCW-kNN</th>
<th>CCP-CART</th>
<th>CCP-C4.5</th>
<th>SMOTE+C4.5</th>
<th>Safe-Level SMOTE+C4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCP-CART</td>
<td>REJECT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCP-C4.5</td>
<td>ACCEPT</td>
<td>REJECT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMOTE+C4.5</td>
<td>REJECT</td>
<td>REJECT</td>
<td>REJECT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safe-Level SMOTE+C4.5</td>
<td>REJECT</td>
<td>REJECT</td>
<td>REJECT</td>
<td>REJECT</td>
<td></td>
</tr>
<tr>
<td>Borderline SMOTE + C4.5</td>
<td>REJECT</td>
<td>REJECT</td>
<td>ACCEPT</td>
<td>REJECT</td>
<td>ACCEPT</td>
</tr>
</tbody>
</table>

Statistical comparison between CCW-kNN and CCP-CART yields a two-tailed p-value of 1.17 E-05 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between CCW-kNN and CCP-C4.5 yields a two-tailed p-value of 0.6732 accepting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between CCW-kNN and SMOTE+C4.5 yields a two-tailed p-value of 1.01E-08 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between CCW-kNN and Safe-Level SMOTE+ C4.5 yields a two-tailed p-value of 0.0033 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.
Classifier for the Classification of Potential Microcalcification

Statistical comparison between CCW-kNN and Borderline SMOTE + C4.5 yields a two-tailed p-value of 0.0232 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between CCP-CART and CCP-C4.5 yields a two-tailed p-value of 7.68E-05 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between CCP-CART and SMOTE+C4.5 yields a two-tailed p-value of 0.007 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between CCP-CART and Safe-Level SMOTE+ C4.5 yields a two-tailed p-value of 0.00093 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between CCP-CART and Borderline SMOTE + C4.5 yields a two-tailed p-value of .0002 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between CCP-C4.5 and SMOTE+C4.5 yields a two-tailed p-value of 5.53E-06 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between CCP-C4.5 and Safe-Level SMOTE+ C4.5 yields a two-tailed p-value of 0.0168 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between CCP-C4.5 and Borderline SMOTE + C4.5 yields a two-tailed p-value of 0.0537 accepting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between SMOTE+C4.5 and Safe-Level SMOTE+ C4.5 yields a two-tailed p-value of 1.27E-08 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between SMOTE+C4.5 and Borderline SMOTE + C4.5 yields a two-tailed p-value of 3.8E-09 rejecting the null hypothesis that their corresponding ROC curves have the same area under them.

Statistical comparison between Safe-Level SMOTE+C4.5 and Borderline SMOTE + C4.5 yields a two-tailed p-value of 0.271 accepting the null hypothesis that their corresponding ROC curves have the same area under them.
The following observations have been made. Results demonstrate that CCW-kNN and CCP-C4.5 are similar in performance in terms of Az. It is noted that in sampling methods, Safe-Level SMOTE+C4.5 and Borderline SMOTE + C4.5 are similar in performance in terms of Az since the statistical comparison between the mean measures of both methods yields a p-value of 0.271 where as SMOTE+C4.5 yields a notably lesser ROC curve than other sampling methods. CCP-C4.5 and Borderline SMOTE + C4.5 are similar in performance in terms of Az since the statistical comparison between the mean measures of both methods yields a p-value of 0.0537. Among the CCP decision tree methods, CCP-C4.5 performs better than CCP-CART. CCP-C4.5 and CCW-kNN yields a notably high ROC curve than other classifier models.

Several algorithms have been developed for detection of microcalcification. To evaluate proposed approach comparisons with many state of the art methods of microcalcification detection discussed in the literature review have been carried out and many of proposed method outperform all other methods. The comparison carried out with other methods is discussed in the next section.

5.5.6 Comparison with other methods

In this section experimental results are compared with current state-of-the-art approaches for microcalcification detection. Each approach by researchers’ uses a different set of images from different databases and the number of images used for the experimental study is also varies. So comparisons have been done only in a qualitative way. Table 5.4 shows comparison of the experimental results obtained with state-of-the-art approaches for microcalcification detection. Even though some researchers used more than one method for microcalcification detection Table 5.4 reports only the highest performance result. The results show that proposed method has the highest Az value of ROC curve among other classification methods.

Table 5.4 Comparison of the experimental results with state-of-the-art approaches for micro-calcification detection

<table>
<thead>
<tr>
<th>Method proposed by</th>
<th>Result(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chang et.al.[57]</td>
<td>Az=.90</td>
</tr>
<tr>
<td>Nunes et.al.[58]</td>
<td>Az=.92</td>
</tr>
</tbody>
</table>
### 5.6 Conclusions

The main focus of this chapter is on improving the potential microcalcification detection and classification performance. It has been explained in the beginning of this chapter the major difficulties in the detection of microcalcification. Having identified the real cause of problem in detection and classification a new technique has been proposed for the potential microcalcification detection and classification. The effective performance of proposed new technique in the potential microcalcification detection has been reported. It has been highlighted why traditional classifiers are sensitive to the imbalanced data classification. Having identified the cause of problem with traditional classifiers some techniques have been proposed that can effectively handle the class imbalance problem for the classification of potential microcalcification. Proposed method includes CCW-kNN, CCP-C4.5, CCP-CART, SMOTE + C4.5, Safe-Level SMOTE + C4.5 and Borderline-SMOTE + C4.5. Quantitative evaluations are used to validate the effectiveness of proposed methods by ROC analysis. Two-tailed Student's t-tests at a level of significance of .05 were performed in order to compare the mean measure of each method with the means of the other methods. CCP-C4.5 outperforms the other methods and it has been recommended for the classification of potential microcalcification. To evaluate proposed techniques, comparisons are carried out with many state of the art methods of microcalcification detection discussed in the literature review and some of the proposed method outperforms other methods.