ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

Τμήμα Ιατρικής

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

VRIJE UNIVERSITEIT BRUSSEL

Department of Electronics and Informatics

Διανεπιστημιακό Πρόγραμμα Μεταπτυχιακών Σπουδών στη Βιοϊατρική Τεχνολογία

Μεταπτυχιακή Εργασία:

Αυτόματη ανίχνευση ύποπτων μικροαποτιτανώσεων σε υψηλής ανάλυσης, τρισδιάστατη απεικόνιση μαστού

της

Ευγενίας Παπαβασιλείου

Διπλωματούχου Ηλεκτρολόγου Μηχανικού και Τεχνολογίας Υπολογιστών

Πάτρα, Ιούλιος 2014

UNIVERSITY OF PATRAS

Department of Medicine

NATIONAL TECHNICAL UNIVERSITY OF ATHENS

Department of Electrical and Computer Engineering

VRIJE UNIVERSITEIT BRUSSEL

Department of Electronics and Informatics

Interuniversity Postgraduate Programme on Biomedical Engineering

Master Thesis:

Automatic Detection of Suspicious Microcalcifications in high Resolution 3-D Breast Imaging

of

Evgenia Papavasileiou

The Advisory and Examining Committee

Prof. Nicolas Pallikarakis - University of Patras, Greece

Prof. Bart Jansen - Vrije Universiteit Brussel, Belgium

Prof. Eleni Kostaridou - University of Patras, Greece

Prof. Dimitrios Koutsouris – National Technical University of Athens, Greece

Acknowledgements

I would like to express my sincere appreciation to a lot of people for their assistance during the elaboration of this Master Thesis. In particular, I wish to express my sincere gratitude to my two supervisors: Professor Nicolas Pallikarakis and Professor Bart Jansen for trusting me with the assignment of this thesis and their countless time to guide my work and help me to develop research skills and critical thinking. Moreover, I would like to thank Prof. Pallikarakis, director of this master program, for accepting me as a student in this interuniversity program and Prof. Jansen for accepting me as an exchange student in the Vrije Universiteit Brussel. Also, I thank all the members of my committee for their time and their precious recommendations. In addition, I am thankful to Dr. Frederik Temmermans for his time, his support and his useful comments. Finally, I thank Mrs Eleni Panoutsopoulou for her administrative support.

Last but not least, I thank my parents, Archontoula and Niko, my brother Vasili, my grandmother Evgenia and my friend Panagioti for their support and encouragement. Mom, honestly, thank you for everything.

To my mother Archontoula,

Εκτεταμένη ελληνική περίληψη

Ο καρκίνος του στήθους είναι ο κακοήθης όγκος ο οποίος σχηματίζεται από τον ταχύ και ανεξέλεγκτο πολλαπλασιασμό κυττάρων του αδένα του μαστού τα οποία έχουν εκτραπεί από τη φυσιολογική λειτουργία τους και αναπτύσσονται σε βάρος του φυσιολογικού ιστού (Φύσσας, 2006).

Ο καρκίνος του μαστού θεωρείται ένας από τους πιο συχνούς τύπους καρκίνου και μία από τις σημαντικότερες αιτίες θανάτου παγκοσμίως (Stewart et al, 2003). Στις ΗΠΑ τα κρούσματα ξεπερνούν τα 200,000 ετησίως ενώ στην Ελλάδα διαγιγνώσκονται περίπου 4,500 γυναίκες με καρκίνο του μαστού κάθε χρόνο (Eucan, 2014).

Είναι αξιοσημείωτο το γεγονός ότι στην Ευρώπη το 60% των κρουσμάτων καρκίνου διαγιγνώσκεται σε πρώιμο στάδιο ενώ στην Ελλάδα το ποσοστό μετά βίας αγγίζει το 5% (Karkinos, 2014). Τα στοιχεία αυτά καταδεικνύουν πόσο ελλιπής είναι η σχετική ενημέρωση μεταξύ των Ελληνίδων, γεγονός εξαιρετικά λυπηρό, αν λάβουμε υπόψη τις δυνατότητες πλήρους ίασης που παρέχει μία έγκαιρη διάγνωση.

Το πενταετές ποσοστό επιβίωσης σε περιπτώσεις διάγνωσης σε πρώιμο στάδιο φθάνει ως και το 95%, γεγονός που υποδηλώνει πως ο καρκίνος του μαστού μπορεί να αντιμετωπιστεί επιτυχώς για την πλειονότητα των γυναικών που φροντίζουν να τον εντοπίσουν έγκαιρα, μέσω συχνών προληπτικών ελέγχων και εξετάσεων.

Για το σκοπό αυτό έχουν αναπτυχθεί πολλαπλές απεικονιστικοί μέθοδοι μεταξύ των οποίων η μαστογραφία, ο γενετικός έλεγχος, ο υπέρηχος και η μαγνητική τομογραφία. Η μαστογραφία είναι η κύρια μέθοδος απεικόνισης του στήθους. Πρόκειται για ακτινολογική μέθοδο δυο διαστάσεων η οποία επιτρέπει την απεικόνιση της μορφολογίας και της δομής των ανατομικών στοιχείων του μαστού. Ανάμεσα στα ευρήματα των μαστογραφιών οι μικροασβεστώσεις είναι ιδιαίτερα σημαντικές. Οι ασβεστώσεις είναι μικρά αποθέματα ασβεστίου στον μαστικό αδένα. Διακρίνονται σε μικροασβεστώσεις και μακροασβεστώσεις με βάση το εάν η διάμετρος τους είναι μικρότερη ή μεγαλύτερη από 1 mm (Karahaliou et al., 2012).

Οι μικροασβεστώσεις είναι ιδιαίτερα σημαντικές διότι επιτρέπουν την έγκαιρη διάγνωση. Έχει καταγραφεί ότι περίπου 30-40% των καρκίνων του στήθους ανιχνεύονται αποκλειστικά από την παρουσία τους. Επίσης, το 70-80% της

ιστοπαθολογικής εξέτασης μαστικών βιοψιών απέδειξε την παρουσία μικροασβεστώσεων (Wilde 1999).

Ωστόσο, δεν υποδεικνύουν όλες οι μικροασβεστώσεις κακοήθεια: είναι παρούσες και στις υγιείς γυναίκες. Για τον χαρακτηρισμό τους ως καλοήθεις ή κακοήθεις υπάρχουν πολλοί παράγοντες οι οποίοι πρέπει να λαμβάνονται υπόψη: το σχήμα, το αν σχηματίζουν συμπλέγματα, η θέση στο μαστό, η συχνότητα και η κατανομή τους (Andreadis et al., 2011; Karahaliou et al., 2012; Sickles, 1984; Willekens et al., 2014).

Το σχήμα είναι ένας από τους σημαντικότερους παράγοντες ο οποίος βοηθάει στη διάκριση ανάμεσα σε καλοήθεις και κακοήθεις μικροασβεστώσεις (Sickles, 1984), δηλαδή ασβεστώσεις οι οποίες είναι παρούσες σε καλοήθεις και κακοήθεις αλλοιώσεις αντίστοιχα. Λεπτά, γραμμικά, καμπυλόγραμμα σχήματα καθώς και σχήματα με διακλαδώσεις μπορούν να υποδεικνύουν κακοήθεια. Αντίθετα, στρογγυλά, οβάλ και καλά καθορισμένα σχήματα μπορούν να υποδεικνύουν καλοήθεια.

Ωστόσο, το σχήμα των μικροασβεστώσεων δεν μπορεί να απεικονιστεί σωστά στην μαστογραφία καθώς η ανάλυση ειναι πολύ χαμηλή και λόγω του ότι η μαστογραφία αποτελεί την δισδιάστατη προβολή ενός τρισδιάστατου αντικειμένου. Ως εκ τούτου, η υπέρθεση ιστού ιστών συχνά δημιουργεί μοτίβα τα οποία φαίνονται ως κακοήθη ή αλλάζουν την μορφή των πραγματικών (Elter et al., 2009).

Σε περίπτωση υποψίας κακοήθειας, διεξάγεται βιοψία. Η βιοψία περιλαμβάνει την παρακέντηση του μαστού με σκοπό την απομάκρυνση ιστού και κυττάρων από μια ύποπτη περιοχή. Εν συνεχεία, οι ιστοί που εξήχθησαν εξετάζονται ανατομοπαθολογικά, δηλαδή τοποθεντούνται κάτω από το μικροσκόπιο για περαιτέρω ανάλυση από τον παθολόγο. Αυτός είναι ο μοναδικός ασφαλής τρόπος για σαφή και αλάνθαστη διάγνωση καρκίνου. Αξίζει να τονιστεί ότι το 65-80% των βιοψιών αποδεικνύονται ότι είναι καλοήθεις όγκοι και άλλα ευρήματα άσχετα με τον καρκίνο. Ο μεγάλος αριθμός των μη απαραίτητων βιοψιών προκαλεί μεγάλο αίσθημα δυσφορίας τόσο ψυχολογικό όσο και σωματικό για τις ασθενείς καθώς και επιπλέον μη απαραίτητα έξοδα για την ίδια την ασθενή και για το σύστημα υγείας. Επίσης, οι μικροασβεστώσεις αυτές καθ’ εαυτές δεν αναλύονται. Ως εκ τούτου, έχει τεθεί το ερώτημα εάν οι μη απαραίτητες βιοψίες μπορούν να αποφευχθούν εάν το σχήμα των μικροασβεστώσεων μπορούσε να μελετηθεί πιο λεπτομερώς.

Για το σκοπό αυτό έχουν πραγματοποιηθεί έρευνες (Temmermans et al., 2013; Temmermans 2014; Willekens et al., 2014) για την ανάλυση του σχήματος των μικροασβεστώσεων με χρήση μικροτομογραφίας (μCT), μιας μη επεμβατικής απεικονιστική μεθόδου υψηλής ανάλυσης με την οποία το τρισδιάστατο σχήμα των αντικειμένων ανακατασκευάζεται. Το σύνολο των δεδομένων που χρησιμοποίηθηκε σε αυτή την εργασία προέρχεται από αυτές τις πηγές (Temmermans et al., 2013; Temmermans, 2014; Willekens et al., 2014). Συγκεκριμένα, σε γυναίκες των οποίων η μαστογραφία έδειχνε υποψία κακοήθειας πραγματοποίηθηκε βιοψία και μετά από ανατομοπαθολογική εξέταση προέκυψαν 50 καλοήθη και 50 κακοήθη δείγματα. Από τα κακοήθη δείγματα εξήχθησαν 2034 μικροασβεστώσεις και από τα καλοήθη 1651 (Temmermans, 2014). Η ταξινόμηση αυτή των μικροασβεστώσεων σε καλοήθεις και κακοήθεις γίνεται αποκλειστικά με βάση το δείγμα στο οποίο βρέθηκαν. Το γεγονός ότι καλοήθεις μικροασβεστώσεις μπορούν να υπάρχουν σε κακοήθη δείγματα και το αντίστροφο δεν λαμβάνεται υπόψη. Ως εκ τούτου, δεν υπάρχει μια βάση αναφοράς (ground truth) για τις ίδιες τις μικροασβεστώσεις, γεγονός που προκαλεί προβλήματα κατά την εισαγωγή των μικροασβεστώσεων σε έναν ταξινομητή.

Για την καλύτερη κατανόηση του συγκεκριμένου προβλήματος παρατίθεται το ακόλουθο παράδειγμα. Δυο μικροασβεστώσεις δίνονται ως παραδείγματα εκπαίδευσης σε ένα ταξινομητή. Η μια απο αυτές τις μικροασβεστώσεις είναι καλοήθης αλλά βρέθηκε μέσα σε κακοήθες δείγμα και άρα εσφαλμένα χαρακτηρίστηκε ως κακοήθης και η άλλη είναι όντως κακοήθης. Και οι δύο μικροασβεστώσεις χαρακτηρίζονται από συγκεκριμένα χαρακτηριστικά σχήματος (shape features) που υποδεικνύουν αντιστοίχως καλοήθεια και κακοήθεια. Κατά την εισαγωγή των δυο μικροασβεστώσεων στην διαδικασία εκπαίδευσης του ταξινομητή, αυτός θα «μάθει», ότι δυο μικροασβεστώσεις με διαφορετικά χαρακτηριστικά σχήματος αντιστοιχούν στην ίδια τάξη, την «κακοήθεια». Ως εκ τούτου, θα υπάρχει ασάφεια όταν ο εκπαιδευμένος πια ταξινομητής θα χρειαστεί να ταξινομήσει μια νέα άγνωστη μικροασβέστωση (Papavasileiou et al., 2015) .

Τα προβλήματα αυτά θα μπορούσαν να εάν οι μικροασβεστώσεις ταξινομηθούν πρώτα με βάση το σχήμα τους. Για το σκοπό αυτό, πριν την είσοδο των μικροασβεστώσεων στον ταξινομητή, εισάγεται ένα βήμα ομαδοποίησης (clustering): μια κατηγορία αλγορίθμων μη επιβλεπόμενης μηχανικής μάθησης (Papavasileiou et al.,

2015). Με αυτό τον τρόπο ο ταξινομητής εκπαιδεύεται να ταξινομεί μικροασβεστώσεις στο σωστό σύμπλεγμα (cluster) χωρίς να χρειάζεται να γνωρίζει την «ετικέτα» που είχε ανατεθεί στις μικροασβεστώσεις σύμφωνα με την ανατομοπαθολογική εξέταση.

Ως clustering ορίζεται η διαδικασία ομαδοποίησης όμοιων αντικειμένων (Hartigan, 1975). Κάθε ομάδα, η οποία ονομάζεται σύμπλεγμα, αποτελείται από αντικείμενα τα όποια είναι όμοια μεταξύ τους με βάση κάποιο μέτρο, που είναι η συνάρτηση απόστασης και ανόμοια με τα αντικείμενα άλλων συμπλεγμάτων. Υπάρχουν πολλοί διαθέσιμοι αλγόριθμοι ομαδοποίησης. Στην παρούσα εργασία χρησιμοποίθηκε ο αλγόριθμος K-means για δύο διαφορετικές αποστάσεις (τετραγωνική ευκλίδεια και cityblock) καθώς και ο αλγόριθμος Minkowski Weighted K-means (MWK-means).

Για την περιγραφή των μικροασβεστώσεων χρησιμοποιήθηκαν 20 χαρακτηριστικά σχήματος. Ανάμεσα σε αυτά υπάρχουν 8 παραδοσιακά χρησιμοποιούμενα χαρακτηριστικά σχήματος τα οποία περιλαμβάνουν: τον όγκο, την επιφάνεια, τον λόγο επιφάνεια ανα όγκο, την μέγιστη και την ελάχιστη διάμετρο του περικλειόμενου ελλειψοειδούς ελαχίστου όγκου, δύο διαφορετικούς ορισμούς της συμπαγότητας (compactness) και την επιμήκυνση (elongation). Επιπλέον, χρησιμοποιήθηκαν 12 νέα χαρακτηριστικά τα οποία περιγράφουν τις ιδαιτερότητες της συνοριακής ζώνης των μικροασβεστώσεων με το φόντο και ονομάζονται συνοριακά χαρακτηριστικά (boundary zone features).

Ένα από τα σημαντικότερα προβλήματα στους αλγόριθμους ομαδοποίησης, είναι ο εκ των προτέρων ορισμός του αριθμού των συμπλεγμάτων τα οποία θα δημιουργηθούν. Στην παρούσα εργασία, ο αριθμός των συμπλεγμάτων μεταβάλλεται από 2 σε 10 και τελικά ως «ιδανικός αριθμός» συμπλεγμάτων επιλέγεται αυτός για τον οποίο η μέση «εντροπία» παίρνει πολύ χαμηλές τιμές. Ως «εντροπία» ορίστηκε μια συνάρτηση η οποία δείχνει την κατανομή καλοηθών και κακοηθών μικροασβεστώσεων μέσα στο σύμπλεγμα. Χαμηλές τιμές εντροπίας αντιστοιχούν σε συμπλέγματα στα οποία ο χαρακτηρισμός ενός συμπλέγματος ως καλοήθες ή κακοήθες είναι πιο καθαρός. Κάθε σύμπλεγμα χαρακτηρίζεται ως καλοήθες ή κακοήθες με μια συγκεκριμενη πιθανότητα η οποία βασίζεται σε αυτή την «εντροπία». Η αντίστοιχη πιθανότητα ανατίθεται σε κάθε μικροασβέστωση που βρίσκεται μέσα στο σύμπλεγμα.

Στη συνέχεια οι μικροασβεστώσεις δίνονται ως είσοδοι σε έναν ταξινομητή. Ως ταξινομητής επιλέγεται ένα τεχνητό νευρωνικό δίκτυο (ANN) και μια μηχανή διανυσμάτων υποστήριξης (SVM). Ο ταξινομητής εκπαιδεύεται να ταξινομεί κάθε μικροασβέστωση στο σωστό σύμπλεγμα δηλαδή σε κάποιο συγκεκριμένο σχήμα. Αφού κάθε σύμπλεγμα χαρακτηρίζεται ως καλοήθες ή κακοήθες με μια συγκεκριμένη πιθανότητα, κάθε μικροασβέστωση ταξινομημένη σε αυτό χαρακτηρίζεται αντίστοιχα ως καλοήθης ή κακοήθης με αντίστοιχη πιθανότητα. Αυτή η ταξινόμηση αποτελεί ένα ενδιάμεσο βήμα ταξινόμησης στο επίπεδο των μικροασβεστώσεων το οποίο θα χρησιμοποιηθεί εν συνεχεία για την ταξινόμηση του δειγματος. Επιπλέον, αυτή η προσέγγιση ταξινόμησης αποτελεί εξέλιξη της μέχρι στιγμής δυαδικής ταξινόμησης της μικροασβέστωσης ως καλοήθης ή κακοήθης. Τέλος, το δείγμα χαρακτηρίζεται ως καλοήθες ή κακοήθες ορίζοντας ένα όριο στον αριθμό των κακοηθών μικροασβεστώσεων οι οποίες απαιτούνται για τον χαρακτηρισμό του δείγματος. Για το σκοπό αυτό, υιοθετούνται τα αποτελέσματα της ταξινόμησης των μικροασβεστώσεων του προηγούμενου βήματος και ακολουθούνται δυο προσεγγίσεις μια σταθμισμένη και μια μη σταθμισμένη. Στην πρώτη περίπτωση οι πιθανότητες των μικροασβεστώσεων που υπολογίστηκαν στο προηγούμενο βήμα χρησιμοποιούνται για τον χαρακτηρισμό του δειγματος. Το τελικό αποτέλεσμα ταξινόμησης του δείγματος συγκρίνεται με το αποτέλεσμα της ανατομοπαθολογικής εξέτασης και έτσι υπολογίζονται η ακρίβεια (accuracy), η ευαισθησία(sensitivity) και η ειδικότητα (specificity).

Τα πειράματα χωρίστηκαν σε τρεις κατηγορίες ανάλογα με τα χαρακτηριστικά τα οποία χρησιμοποιούνται για την ομαδοποίηση και την ταξινόμηση των μικροασβεστώσεων: μονο τα 8 παραδοσιακά χρησιμοποιούμενα χαρακτηριστικά, μόνο τα 12 συνοριακά χαρακτηριστικά και το σύνολο των δύο μαζί (20 χαρακτηριστικά).

Από τα αποτελέσματα των πειραμάτων για την ταξινόμηση στο επίπεδο των μικροασβεστώσεων οι δύο ταξινομητές (ANNs & SVMs) παρουσίασαν πολύ κοινή απόδοση παρόλο που τα ANNs είναι κατά πολύ πιο απαιτητικά όσον αφορά τους επεξεργαστικούς πόρους και χρόνους. Επίσης, τα παραδοσιακά χρησιμοποιούμενα χαρακτηριστικά σχήματος παρουσιάζουν καλύτερη απόδοση από τα χαρακτηριστικά συνοριακής ζώνης και από το σύνολο των δύο χαρακτηριστικών μαζί. Αυτά τα αποτελέσματα ωστόσο πρέπει να ερμηνεύονται προσεκτικά λόγω των προβλημάτων

έλλειψης βάσης αναφοράς. Επίσης, η ταξινόμηση στο επίπεδο των μικροασβεστώσεων αποτελεί ένα ενδιάμεσο βήμα για την τελική ταξινόμηση του δείγματος.

Στο επιπεδο ταξινόμησης του δείγματος, ο κύριος στόχος είναι η μεγιστοποίηση της ευαισθησίας, καθώς είναι σημαντικότερο να μπορούν να ανιχνεύονται σωστά όλες οι κακοήθεις περιπτώσεις σε σχέση με την ακρίβεια, με ταυτόχρονα υψηλά ποσοστά ειδικότητας και ακρίβειας. Τόσο στην σταθμισμένη όσο και στη μη σταθμισμένη περίπτωση τα οχτώ παραδοσιακά χαρακτηριστικά σχήματος παρουσιάζουν καλύτερη απόδοση σε σχέση με τις δυο αλλες κατηγορίες. Το καλύτερο αποτέλεσμα επιτυγχάνεται, στην σταθμισμένη περίπτωση με χρήση του K-means αλγορίθμου, τα 8 χαρακτηριστικά σχήματος, την απόσταση cityblock και 9 συμπλέγματα. Με αυτή την προσέγγιση επιτυγχάνεται ευαισθησία 100%, ειδικότητα 42.6% και ακρίβεια 72.7% και αποτελούν βελτίωση αντίστοιχα 2%, 2.6% και 2.7% σε σχέση με την βιβλιογραφία (Temmermans et al., 2013; Temmermans, 2014; Willekens et al., 2014).

Η επίτευξη 100% ευαισθησίας επιτρέπει να αποφευγονται οι περιπτώσεις παράλειψης κάποιων κακοηθών δειγμάτων. Αυτή η βελτίωση είναι αποτέλεσμα της αποφυγής των προβλημάτων που προκαλεί η έλλειψη βάσης αναφοράς. Ο αλγόριθμος που προτείνει η παρούσα διπλωματική εργασία μπορεί να χρησιμοποιηθεί για να αποφεύγονται ορισμένες μη απαραίτητες ανατομοπαθολογικές εξετάσεις. Επιπλέον, η τρισδιάτατη απεικόνιση υψηλής ανάλυσης μπορεί στο μέλλον να χρησιμοποιηθεί απευθείας στην ασθενή (in vivo) αυτό θα σήμαινε την αποφυγή σημαντικού αριθμού μη απαραίτητων βιοψιών.

Κλείνοντας, η παρούσα εργασία εισάγει δύο καινοτομίες: αποφεύγονται τα προβλήματα που δημιουργεί η έλλειψη βάσης αναφοράς και υπάρχει μια μετατόπιση από δυαδική ταξιμόμηση σε μια σταθμισμένη με πιθανότητες ταξινόμηση. Η μεγάλη πληθώρα διαθέσιμων αλγορίθμων ομαδοποίησης και ταξινομητών καθώς και η χρήση εναλλακτικών προσεγγίσεων για την ενσωμάτωση των πιθανότητων στην διαδικασία χαρακτηρισμού του δείγματος αποτελούν ένα προσοδοφόρο χώρο για την περαιτέρω βελτίωση των αποτελέσματων.

Summary

This Master Thesis presents a novel classification approach for microcalcifications (MCs) extracted from core biopsy tissue samples digitized using micro-CT, a high-resolution 3D imaging modality. MCs are tiny spots of calcium that may occur in the female breast. Although they are common in healthy woman, they are often an early sign of breast cancer. In case of suspiciousness, a biopsy is conducted and the extracted tissue is pathologically analysed for the presence of cancer cells. However, the MCs themselves are mostly not analysed. As a result, there is a ground truth for the tissue samples but not for the individual MCs.

This ground truth problem can be bypassed if the MCs are first grouped explicitly based on their shape. By clustering MCs according to their shape-features, similar shapes are grouped together in clusters. This way, the classifier is trained to classify MCs in the correct cluster, i.e. it learns to distinguish among different shapes of MCs, independently of their original class labels. This thesis investigates whether the use of a clustering method as a preprocessing step before training the classifier would improve the obtained classification results.

In cluster analysis one of the biggest difficulties is the a priori definition of the number of clusters. For this reason, the number of generated clusters is varied from 2 to 10 and as the ideal number of clusters is selected the one for which the mean entropy is lowest, because the lower the entropy of a cluster, the clearer the designation of a cluster as benign or malignant. Each of the clusters is characterized as benign or malignant with a certain probability based on this entropy and thus each object of this cluster is assigned a corresponding label with the same probability. Next, the objects are introduced to a classifier that learns to classify them to the correct cluster, i.e. to the correct shape. In the final step, the sample is classified as benign or malignant by setting a threshold on the percentage of MCs classified as malignant and the result is compared with the outcome of the anatomopathological examination. In addition, the MC's cluster probability can now be used as a weight when determining the sample's class.

At the sample level, this approach delivers a sensitivity of 100%, a specificity of 42,6% and an accuracy of 72,7% which is respectively an improvement of 2%, 2,6% and 2,7% percent compared to the state of the art.

This approach introduces two novelties; the ground truth issues concerning each MC's class are avoided and there is a shift from binary classification to a probability-weighted classification. The vast amount of clustering algorithms and classifiers as well as the use of various approaches for incorporating the cluster probabilities in the classification of the sample hold the potential to further improve these results.

The results of this master thesis have been accepted to be presented as an oral paper in the 6th European Conference of the International Federation for Medical and Biological Engineering (MBEC), in Dubrovnik, Croatia; 'Papavasileiou, E., Temmermans, F., Jansen, B., Willekens, I., Van de Casteele, E., De Mey, J., Deklerck R. & Hostens, J. (2015, January). Shape-Based Clustering and Classification of Breast Microcalcifications in Micro-CT Images. In *6th European Conference of the International Federation for Medical and Biological Engineering* (pp. 160-163), Dubrovnik, Croatia'

List of figures

List of Tables

Contents

1. Chapter: Introduction

1.1 Introduction

Breast cancer is considered to be one of the most frequent types of cancer and one of the most significant causes of death worldwide (Stewart et al., 2003). In the United States, more than 200,000 cases of breast cancer occur every year, while in Greece 4,500 women are diagnosed with breast cancer annually (Eucan, 2014).

Breast cancer's early detection has a key role in survival. Many breast imaging modalities exist that assist in detecting suspicious lesions that otherwise would have been noticeable in no less than five years, leading to reduction in mortality from breast cancer. It is very remarkable that in Europe 60% of the breast cancer cases are diagnosed in early stage, while in Greece this percentage barely reaches 5% (Eucan, 2014). These facts reveal how insufficient the briefing of Greek women about breast cancer is, especially if we take into consideration the high chances of full recovery for the early diagnosed cases.

1.2 Problem Statement

Breast imaging has a key role in the early detection of breast cancer. Mammography is the standard of care in breast screening. An important early indicator of potential presence of breast cancer is the appearance of microcalcifications (MCs). MCs are tiny spots of calcium deposit that may occur in the breast. Although MCs are associated with breast cancer, they are common in healthy woman as well. Among others, the shape of the MCs is an important factor used to discriminate between benign and malignant abnormalities. However, their characterization as benign or malignant based on their appearance in mammograms is a difficult task even for expert radiologists. Due to the fact that a mammogram is a 2D image of a 3D object, superposition of breast tissue often produces patterns that appear like suspicious masses to a radiologist or alters the appearance of real mammographic lesions, thus underestimating the importance of the findings (Elter et al., 2009; Karahaliou et al., 2012; Temmermans et al., 2013; Willekens et al., 2014).

In case the radiologist considers a region suspicious, a biopsy is conducted. The extracted tissue is then anotomopathologically investigated for the presence of cancer cells. The MCs themselves however are mostly not analyzed. Many biopsies are conducted that turn out to be negative. As such, the question has been raised whether some biopsies might have been avoided if the shape of the MCs could be analyzed in more detail (Papavasileiou et al., 2015).

Therefore, studies have been presented to analyze the shape of calcifications using high-resolution 3D imaging (Temmermans et al., 2013; Willekens et al., 2014). Based upon extracted shape features, automated classification techniques have been presented (Temmermans et al., 2013). Typically, these classifiers learn to assign a benign or malignant label to individual calcifications based upon the nature of the sample they originate from. The fact that benign calcifications may occur in malignant samples and vice versa is not taken into account. As a consequence, a bias in the training process is introduced. This thesis presents a methodology that aims to avoid this predicament.

1.3 Thesis Purpose

The dataset consists of 3685 microcalcifications (MCs) extracted from 50 benign and 50 malignant samples. It is not possible to assign the label "benign" or "malignant" for each MC. Instead, the label is assigned to each MC depending on whether it was found in a benign or a malignant sample, after anatomopathological examination. This means that no ground truth exists for the actual MCs. As a consequence, benign MCs may exist in malignant samples, and in this case they may be incorrectly classified as malignant. Similarly, malignant MCs may exist in benign samples.

This problem can be overcome if the MCs are first grouped explicitly based on their shape. By clustering MCs according to their shape-features, similar shapes are grouped together in clusters. In this way, the classifier is trained to classify MCs in the correct cluster, i.e. it learns to distinguish among different shapes of MCs, independently of their original class labels. So, the classifier classifies objects exclusively based on their shape features and without any knowledge of the class label that was assigned during the histopathological examination. In general, the question that

this thesis is trying to answer is the following; are the shape features an objective characterisation of the shape of microcalcifications so that benign and malignant shapes could be distinguished?

1.4 Thesis Structure

The presented thesis consists of 9 chapters, including the current introduction chapter. In Chapter 2, the basic aspects of breast anatomy, breast cancer, breast imaging techniques and biopsy are discussed. Chapter 3 provides information on CADe and CADx systems and on the categories of features that are used in CAD systems. Chapter 4 is focused on the classification methods and Chapter 5 on cluster analysis. Chapter 6 provides the features that were adopted in this thesis. Chapter 7 presents the approach that is implemented in the current thesis and Chapter 8 presents the obtained results. This thesis closes with Chapter 9, where the main conclusions are discussed.

1.5 Publications

Papavasileiou, E., Temmermans, F., Jansen, B., Willekens, I., Van de Casteele, E., De Mey, J., Deklerck R. & Hostens, J. (2015, January). Shape-Based Clustering and Classification of Breast Microcalcifications in Micro-CT Images. In *6th European Conference of the International Federation for Medical and Biological Engineering* (pp. 160-163). Springer International Publishing

2. Chapter: Breast Cancer

2.1 Breast Anatomy

Each breast lies over the pecoral muscle of the chest wall and extends from the 2nd to the 7th rib and across the sternum to the anterior axillary line (Δημητρόπουλος, 2000). (Figure 2.1). The mammary gland has a discoid shape except for a projection that is formed in the upper outer quadrant, the so- called Spence's tail. The mammary gland is surrounded by fat and skin and consists of 15-20 lobes. Each of these lobes ends with its main duct in the nipple. The main ducts branch and eventually end in lobules. The lobules are again a tree like structure where the branches are called ductules and the leaves aveoli or Terminal Ductal Lobulo- alveolar Units (TDLUs) (Figure 2.2). The alveoli are responsible for breast milk production. They are surrounded by tiny muscles that squeeze them to push milk out into the ductules (Δημητρόπουλος, 2000).

The breasts of women in their 20s differ hormonally and biologically from the breasts of women in their 50s who gradually enter into menopause. During the reproductive age the high levels of oestrogens in conjunction with progesterone maintain dense breasts, while the breasts of older women generally contain a larger proportion of fat. Using radiographic imaging, abnormalities are more difficult to detect in dense breast than in fat breasts.

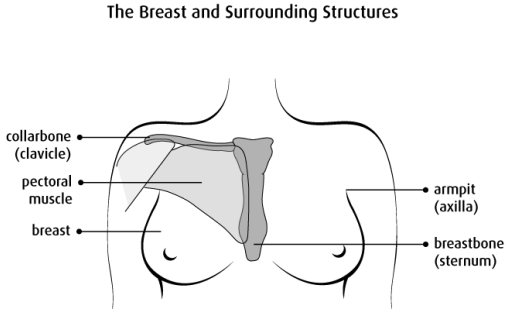


Figure 2.1: The breast and the surrounding tissues

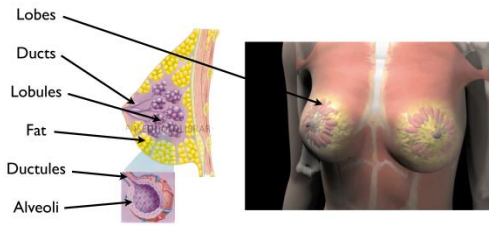


Figure 2.2: Breast anatomy

2.2 Breast Cancer

Breast cancer is the leading type of cancer in women and the second most important cause of death, after colon cancer. It is a malignant tumour caused by the uncontrolled proliferation of abnormal cells of the mammary gland which have diverted from their normal function and grow at the expense of normal tissue. It originates from the breast tissue, most commonly from the inner lining of milk ducts or the lobules that supply the ducts with milk and can be invasive or not (Φύσσας, 2006). Cancers originating from ducts are known as ductal carcinomas, while those originating from lobules are known as lobular carcinomas. If the cancer infilters adjoining parts of the breast it is said to be invasive. The most common types of breast cancer are:

- Ductal Carcinoma in Situ (DCIS)
- Lobular Carcinoma in Situ (LCIS)
- Invasive Ductal Carcinoma (IDC)
- Invasive Lobular Carcinoma (ILC)

The frequency of the cases that are diagnosed with these types of breast cancer are 2.5%, 2.5%, 85%, and 10% respectively (Breast Cancer, 2014).

Intraductal or ductal carcinoma in situ is the proliferation of malignant epithelial cells confined to ducts, with no evidence of invasion through the basement membrane, as illustrated in Figure 2.3.

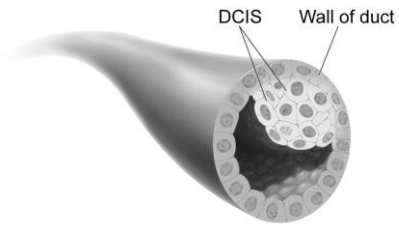


Figure 2.3: Ductal Carcinoma in Situ (DCIS)

2.3 Breast Cancer's risk factors

The causes of breast cancer are unknown. However, there are some risk factors that are considered to influence the possibility of developing a malignant mass in the breast (Φύσσας, 2006).

Inheritance: It is estimated that only 5-10 % of the breast cancer's cases are associated with heredity factors. The majority of the cases result from impairment to breast cells' genetic material that are caused by various factors during the life of a person.

Age: The risk of developing a breast cancer is increasing with age. A woman in her 60s is 100 times more likely to develop breast cancer than a woman in her early 20s. More specifically, for women at the age of 15-39 years old the percentage of breast cancer cases is 0.5%, for women at the age of 40-59 years old the percentage is increased to 4% and for women over 60 years old, the percentage is calculated at 7%.

Disorders of menstruation and pregnancy: There are several facts that suggest that women with early start of menstruation or late menopause have an increased risk of breast cancer. Also, the lifetime exposure to oestrogen plays a fundamental role in the development of breast cancer. A similar correlation has also emerged for women who remained childless or did not have complete pregnancies.

Weight Gain: Studies indicate that weight gain after the age of 20 may increase the risk of breast cancer. In particular, the risk triplicates if the body mass index is at peak levels after the age of 50 years old.

Alcohol and smoking: Recent studies have shown that alcohol and smoking can increase the risk of breast cancer up to 25% and 50-60%, respectively.

2.4 Breast Cancer's Diagnosis Systems

Breast screening is the medical screening of asymptomatic women for breast cancer in an attempt to achieve an early diagnosis. This is based on the assumption that early detection of breast cancer can improve the cancer's outcomes and increase the survival rates. A number of screening modalities have been developed over the years in order to assist in finding a lesion at a very primitive form, including clinical and self breast examinations, mammography, genetic screening, ultrasound and magnetic resonance imaging.

2.4.1 Clinical and Self Breast Examination

Clinical and self-breast examination are considered the easiest and most painless screening methods of breast examination that involve feeling of the breasts for lumps, distortions and swelling. Clinical breast examination is performed by physicians in annual base while breast self-examination (Figure 2.4) is done by the woman herself once in a month. However, after the age of 40 it is an accompanying method and cannot replace mammography.

Figure 2.4: Self-Breast Examination

2.4.2 Mammography

Mammography is a low-energy 2-D X-ray breast imaging modality which allows visualisation of the morphology and structure of the anatomical data of the human breast and of pathological lesions. The value of mammography against breast cancer is invaluable as it can detect growing tumours that otherwise would become perceptible and palpable by the patient or the doctor, not earlier than two years later (Cady et al., 2004).

Several types of mammography exist depending on the technology used (analogue, digital, magnetic mammography) and the reason for which the mammography is conducted (screening or confirmation of previous diagnosis).

2.4.2.1 The process of mammography

The analogue and digital mammographies are performed with the use of a special radiodiagnostic machine, known as the mammography unit, which may be analogue or digital, respectively (Κανδαράκης, 2004). In order to obtain a mammogram, each breast is placed between two plates that are known as compression paddle (Figure 2.5). The compression paddle pushes the breast so as to achieve the same thickness throughout the whole breast as well as maximum resolution.

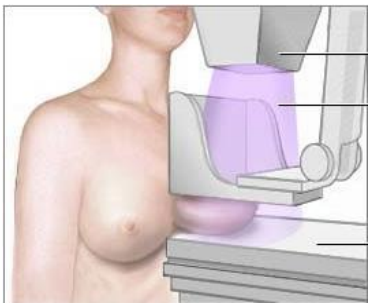


Figure 2.5: compression paddle

Then, the X ray tube emits ionizing radiation of low intensity, which penetrates the breast and weakens unevenly as it passes from different types of tissue. This heterogeneous attenuation is recorded by the image formation system lying underneath the breast.

The image formation system may be analogue or digital. In the first case an analogue film is placed underneath the compression paddle together with reinforcing plates, while in the second case a Flat Panel Active Matrix is used, where the image is formed digitally using pixels (Charge-Coupled Device- CCD) (Hyoung-Koo).

The picture that is formed by these two methods consists of different scales of grey as shown in Figure 2.6. Each shade of grey represents a certain amount of X ray radiation that falls on the analogue film or on the digital image detector. This quantity depends on the attenuation of the initial radiation during its passage through the human breast. The black colour and the shades of gray indicate very little attenuation of the initial radiation. On the other hand, white and open shades of gray correspond to total or very high attenuation of the initial radiation (Κανδαράκης, 2004).

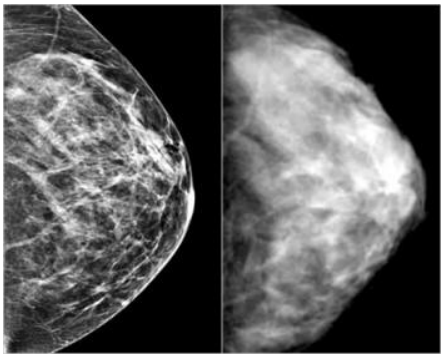


Figure 2.6: Left: mammogram. Right: analogue mammogram

During the procedure of a mammography two images are taken from each breast from two different angles. The one irradiation of the breast is performed with the compression paddle being in a parallel position in relation to the floor (Craniocaudal-CC). The other irradiation is performed at 4 degrees in relation to the floor (Mediolateral oblique-MLO).

2.4.2.2 Findings on mammograms

Distinguishing breast cancer on a mammogram is not always easy. The American College of Radiology (ACR, 2014) defines the main signs of breast cancer on a mammogram as:

- Masses
- Clusters of microcalcifications (MCs)
- Deterioration of the structure of adjacent tissues (architectural distortions)

2.4.2.2.1 Nodular shadows (Masses)

The nodular shadows can occur with distinct boundaries (circumscribed lesions), sharp limits with spikes (spiculated masses) and with less distinct boundaries (ill-defined masses). In general, radiation impermeable nodular shadows are basically malignant, while radiation permeable masses are usually benign.

In general, when examining mammograms contour, density, shape, orientation and size of the shadings should be taken into consideration for discriminating between benign and malignant.

2.4.2.2.2 Calcifications

In many cases it is possible that a lesion may be displayed in a mammogram in the form of microcalcifications with or without shading. Calcifications are small deposits of calcium in the mammary gland. They appear in a mammogram as bright structures of high Signal to Noise Ratio (SNR) due to the large attenuation coefficient of calcium (Figure 2.7). If the calcifications have a diameter less than 1mm they are characterized as microcalcifications (MCs), whereas if their diameter is more than 1 mm, they are characterised as macrocalcifications (Karahaliou et al., 2012). A group of calcifications is considered as a cluster if more than 5 calcifications appear in an area of 1 cm^2.

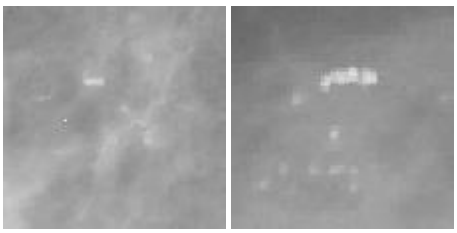


Figure 2.7: Microcalcifications. (A) Malignant and (B) Benign.

Since Salomon's radiographs of mastectomy specimens in 1913 (Salomon, 1913) it has been reported that MCs are associated with breast cancer (Gershon – Cohen et al., 1962) were the first to report that the irregular, clustered appearance of calcifications was associated with breast cancer in 1962.

MCs allow for early diagnosis. It has been reported that approximately 30-40% of breast cancers are detected exclusively by the presence of MCs. E.g. 85% of DCIS cases are discovered exclusively by the appearance of MCs. Also, 70-80% of breast cancers' histopathological examination revealed the presence of microcalcifications (De Wilde, 1999).

However, MCs do not always indicate malignancy; they are also common in healthy women. There are several factors that have to be taken into consideration such as shape, size, clustering, location, frequency and distribution (Andreadis et al., 2011; Karahaliou et al., 2012; Sickles, 1984; Willekens et al., 2014).

Shape is an important factor that aids to distinguish between benign and malignant MCs (Sickles, 1984), i.e. calcifications present in a benign or a malignant lesion respectively. In (Lanyi, 1983) Lanyi classified 5641 MCs according to their shape, categorising them as punctate, bean-shaped, linear or branching. In 95% of cases, the MCs had more than two configurations within individual clusters (polymorphy). Thin, linear, curvilinear, branching shapes may suggest malignancy while round, oval and well delineated shapes may indicate benign lesions (Sickles, 1984).

Malignant calcifications have a wide variation in shape. In general, three basic forms can be distinguished (Tabár, 2011); casting type, granular type and powderish calcifications.

Casting type calcifications have linear, fragmented, occasionally branching calcifications with irregular contours. These calcifications are typically formed within ducts affected by ductal carcinoma.

Granular type calcifications are individually discernible particles that resemble granulated sugar or crushed stone.

Powederish calcifications are very small particles of calcium. On a mammogram they are only visible if they are grouped together.

In practice, the resolution of a mammogram is too low for clear shape differentiation. Therefore, another important discriminator is the distribution of the calcifications. According to the BI-RADS atlas (Tabár, 2011) the distributions of calcifications are defined as (Figure 2.8):

Diffuse or Scattered: diffuse calcifications may be scattered calcifications or multiple similar appearing clusters of calcifications throughout the whole breast.

Regional: scattered in a larger volume (> 2 cc) of breast tissue and not in the expected ductal distribution.

Clustered: at least 5 calcifications occupy a small volume of tissue

Linear: calcifications arrayed in a line, which suggests deposits in a duct.

Segmental: calcium deposits in ducts and branches of a segment or lobe

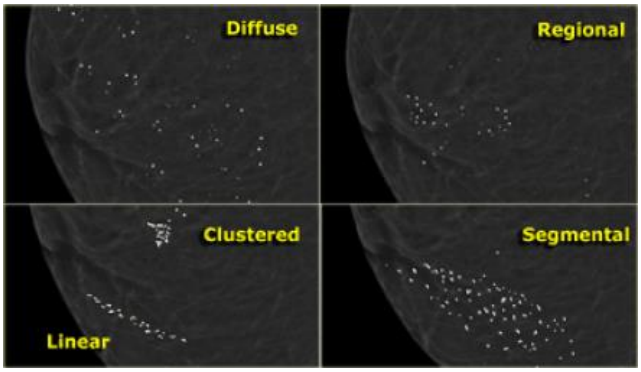


Figure 2.8: Distribution of Microcalcifications

2.4.2.2.3 Architectural Distortions

Architectural distortion is defined as a rupture of the normal architecture with no visible mass (Karahaliou, 2009). Parenchymal asymmetry between the two breasts may indicate the presence of breast cancer.

2.5 Biopsy

A biopsy is the examination that rejects or confirms the diagnosis of breast cancer on a mammogram or on other examination. The breast biopsy involves the aspiration of the breast in order to remove tissue/cells from a suspicious region. Subsequently, the tissues are anatomopathologically investigated, i.e. they are placed under a microscope for observation and further analysis by a pathologist. Several types of biopsy procedures exist. The most common types include Fine Needle Aspiration Biopsy, Core Needle Biopsy, Stereotactic Core Needle Biopsy, Vacuum Assisted Core Biopsy and Surgical Biopsy (ACS, 2014; NHS, 2009; Λάππας, 2014. The type of biopsy used for each case depends on many factors, such as lesion's suspiciousness, size and location, as well as patient's other medical problems and personal preferences.

2.5.1 Fine Needle Aspiration Biopsy

In fine needle aspiration biopsy (FNAB), a very thin needle (thinner than the needles used for blood tests) is used in order to aspirate a small amount of tissue from the suspicious area. The use of local anaesthetic is optional, as it has been noticed that the administration of the numbing medicine may be more painful than the biopsy itself.

The area to be biopsied can be located by various means. If it can be felt, the doctor locates the lump and guides the needle there. If the suspicious area cannot be felt then two methods for guiding the needle can be used; ultrasound guided biopsy or stereotactic needle biopsy. In the first technique the doctor uses ultrasound imaging to watch the needle on the screen as he moves it towards the mass. In the second one, the exact position of the mass is mapped with the aid of computers that use mammograms taken from two angles.

Once the needle is in place, fluid or tissue is drawn out. If clear fluid is withdrawn, the lump is more likely a benign cyst. Bloody or cloudy fluid can mean either a benign cyst or, less often, cancer. If the lump is solid, small pieces of tissue are drawn out.

Nevertheless, FNAB may miss cancer if the needle does not get a tissue sample from the cancer cells (Lieske et al., 2006).

2.5.2 Core Needle biopsy

Core needle biopsy (CNB) can provide increased sensitivity and specificity compared to FNAC (Perry et al., 2008). It is a similar procedure compared to FNAB with the exceptions that it uses a slightly larger, hollow needle and that local anaesthesia is usually required**.**

During CNB the doctor guides the needle to the suspicious area either by feeling a palpable lump or with the aid of imaging (ultrasound or X- ray). Then the needle is used to withdraw small cylinders (or cores) from the abnormal area. Core biopsy is preferred for lesions of architectural distortion and MCs. If MCs are extracted during the core biopsy, it is essential that specimen radiography is applied in order to verify the presence of calcifications.

2.5.2.1 Stereotactic Core Needle Biopsy

A stereotactic core needle biopsy uses x-ray equipment and a computer to map the area where the needle will be located. It is a type of biopsy recommended for suspicious cases where MCs are found and no mass can be felt or seen on ultrasound (ACS, 2014; Λάππας, 2014).

2.5.2.2 Vacuum-assisted core biopsy

Vacuum-assisted biopsies can be conducted with systems like the Mammotome®. For this type of biopsy the skin is numbed and a small cut (less than 0.6 cm) is made (ACS, 2014; Λάππας, 2014). A hollow probe is put in through the cut and guided into the abnormal area of breast tissue using x-rays, ultrasound, or MRI. A cylinder of tissue is then pulled into the probe through a hole in its side, and a rotating knife inside the probe cuts the tissue sample from the rest of the breast.

The advantages of this method is that it allows multiple tissue samples to be removed through one small opening, with very little scarring and without the need of stitches.

2.5.3 Surgical Biopsy

When needle biopsy is insufficient to conclude about the malignancy of a suspicious area, surgical biopsy may be recommended (ACS, 2014). A surgical biopsy, also known as open biopsy, is done in an operating room. During a surgical biopsy a cut is made in the breast to remove all or part of the lump for further examination. Apart from local anaesthesia, sedation through a line in the patient's arm may also be used to help the patient relax during the procedure. In some cases, general anaesthesia may also be used. Surgical biopsies usually last an hour and the recovery period is less than two hours.

An open biopsy that removes only part of a lump of suspicious tissue is called an incisional biopsy, while one that removes the entire lump is called excisional biopsy. An incisional biopsy is usually done when the lump is quite large. However, if the tissue is proven to be malignant the remaining portion of the lump will be removed surgically, during a second surgery.

When a breast mass or an area of calcifications is not palpable, but it looks suspicious on the mammogram, the surgeon may use a procedure called "wire localisation" to help identify the tissue for later surgical biopsy. For this procedure, the breast is numbed with local anaesthetic and a thin, hollow needle is inserted into the breast guided by ultrasound's or mammography's support. Next, a very thin wire with a small hook at the end is placed to pinpoint the suspicious area. The needle is then removed leaving the wire in order to help the surgeon find the part of the breast tissue that will be removed.

Following a surgical breast biopsy, stitches are often needed and a short scar in the shape of a line is left. In addition, bleeding, soreness and swelling are also expected for a few days. Finally, depending on the size of the extracted tissue, this procedure can affect the shape of the breast.

2.6 Anatomical pathological investigation

The tissue samples extracted during biopsy are placed in formol tubes, from which they are taken out later and placed on a glass plate in order to be radiographically scanned. During this procedure, it is validated if the target MCs were correctly extracted from the suspicious region. Following this scan the tissues are divided into two blocks and are conserved in paraffin. One block includes the samples with calcifications, and the other block comprises the samples without calcifications. During the anatomo - pathological investigation, the anatomical pathologist cuts slices from these blocks which are investigated under a microscope. Typically, one third of the blocks is analysed and the pathologist will then make a diagnosis based upon the identified cells.

2.7 Diagnostic classification

According to European guidelines for quality assurance in breast cancer screening and diagnosis (Perry et al, 2008) a simple five-point classification system should be used as described below;

Radiology

- R1 Normal/benign
- R2 A lesion having benign characteristics
- R3 An abnormality present of indeterminate significance
- R4 Features suspicious of malignancy
- R5 Malignant features

While this system is sufficient for most working purposes, if desired, the ACR BIRADS system can be used which is more complex but more precise classification in terms of percentage likelihood (ACR, 2003).

Fine Needle Aspiration Cytology

- C1 Inadequate for diagnosis
- C2 Benign epithelial cells
- C3 Atypia probably benign
- C4 Suspicious of malignancy
- C5 Malignant

Core Biopsy/Histology

- B1 Unsatisfactory/normal breast tissue
- B2 Benign
- B3 Benign but of uncertain malignant potential
- B4 Suspicious of malignancy
- B5 Malignant

2.8 Targets of a Breast Assessment Unit

According to European guidelines for quality assurance in breast cancer screening and diagnosis (Perry et al, 2008) the following percentages are standardized:

Table 2-1: Percentage targets for quality assurance in breast diagnosis

Percentage	Minimum Standard	Expected
of image guided FNAC procedures with an insufficient result (C1)	<25%	<15%
of image guided FNAC procedures from lesions subsequently proven to be malignant having an insufficient result (C1)	< 10%	< 5%
of women with breast cancer having a non-operative diagnosis of malignancy (FNAC/CB reported as definitely malignant)	> 70%	> 90%

2.9 Cytology/histology quality assurance

According to European guidelines the following percentages are standardize for quality assurance in breast cancer cytology/histology.

Absolute sensitivity considers only the definitely malignant results (C5 or B5)

Complete sensitivity considers all abnormal results (above and equal to B3, C3)

Table 2-2: Suggested Thresholds for FNAC performance

	Minimum	Preferred
Absolute Sensitivity	>60%	>70%
Complete Sensitivity	>80%	>90%
Specificity	>55%	>65%
Positive Predictive Value (C5)	>98%	>99%
False Negative Rate	<6%	<4%
False Positive Rate	<1%	<0.5%

Table 2-3: Suggested Thresholds for Core Biopsy performance

	Minimum	Preferred
Absolute Sensitivity	>70%	>80%
Complete Sensitivity	>80%	>90%
Specificity	>75%	>85%
Positive Predictive Value (C5)	>99%	>99.5%
False Negative Rate	<0.5%	<0.1%
False Positive Rate	<15%	<10%

3. Chapter: Computer Aided Diagnosis Systems (CAD)

3.1 Introduction

Early detection of breast cancer has a key role in addressing it. For this reason, many breast imaging modalities have been developed to assist in finding a lesion at a very primitive form. In conjunction with increased public awareness prompting for monthly self-breast examination and annual examination by physician, it yields a reduction in mortality from breast cancer (Karahaliou et al., 2012).

Even though mammography is the current standard for breast screening, the characterization of lesions as benign or malignant based on their appearance in mammograms is a difficult task even for expert radiologists. Due to the fact that a mammogram is a 2D image of a 3D breast, superposition of breast tissue often produces patterns that appear like suspicious masses to a radiologist or alters the appearance of real mammographic lesions (Karahaliou et al., 2012; Elter et al., 2009).

In general there are two categories of typical errors when examining mammograms; false positive and false negative results. False positive results occur when the radiologist recognises a lesion as malignant when it is actually benign. False negative errors occur when a malignant region is not recognised by the radiologist.

The aforementioned errors have resulted in a percentage of omitted cancers of 10-30%, necessitating the conduction of biopsies. However, it is reported that less than 30% of all breast biopsies actually show a malignant pathology. The high number of unnecessary breast biopsies causes major mental and physical discomfort for the patients as well as unnecessary expenses (Elter et al., 2009; Kopans, 1992).

Computer aided Detection (CADe) and Diagnosis Systems (CADx) systems have been proposed in the past years and they have a key role in detection and diagnosis of breast lesions.

3.2 Computer Aided Detection and Diagnosis Systems

CADe and CADx systems aim to support radiologists in the discrimination of benign and malignant mammographic lesions and to increase the positive predictive value of mammographic interpretation. The term "CADe/x" refers to formulating the clinical detection or diagnosis problem into the context of quantitative image feature extraction and pattern classification with the goal of solving it automatically (Duncan et al., 2000).

CADe systems have been developed to improve the ability of radiologists in detecting lesions through identification of suspicious regions of masses and MC clusters in an image. The input of the system is a mammographic image and the output is the location and the boundaries of a suspicious lesion, known as Region of Interest (ROI).

CADx systems assist radiologists in the process of diagnosis and classification of lesions as malignant or benign, thus affecting the subsequent patient management (follow-up or biopsy). The input of a CADx system is the area containing the abnormality (ROI) and the output is a probability that the lesion is malignant

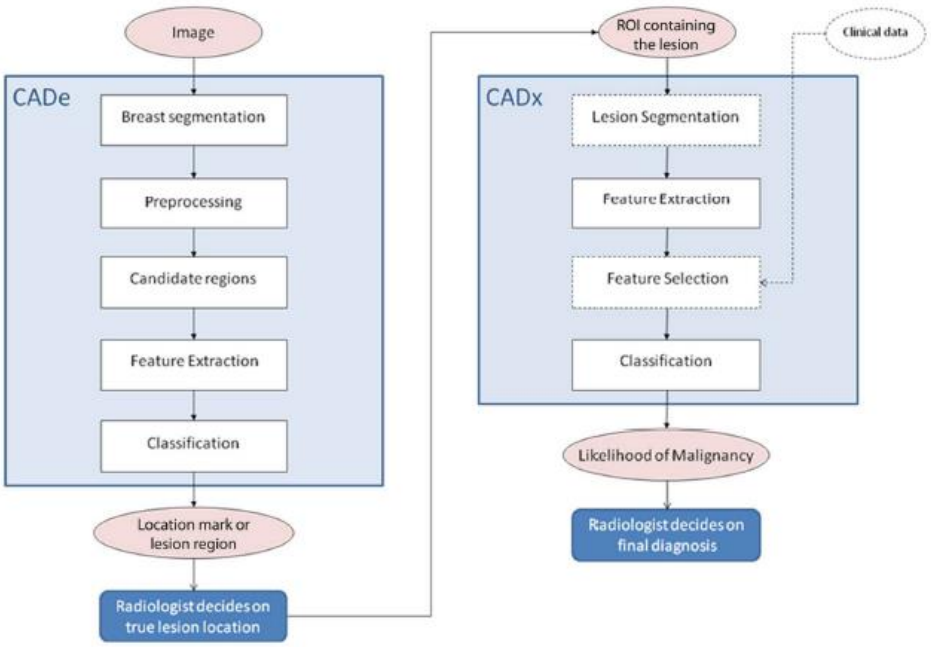


Figure 3.1: CADe and CADx systems

Typical phases of a CAD system

Most CADe and CADx systems include similar processing steps, such as segmentation, feature extraction, feature selection and classification.

Starting from a digital or digitized mammogram the first operations are the processing ones. Here the breast is segmented and some filtering or normalisation is accomplished in order to improve the quality of the image and reduce the noise. Then a signal extraction step is performed. In this phase objects similar to the lesions are isolated by means of different techniques (Suri et al., 2006).

After that a set of features is calculated on the extracted signal. A feature is an individual measurable heuristic property of a phenomenon being observed. The extracted features represent a mathematical description of characteristics that are helpful for isolating the lesion or for distinguishing malignant and benign lesions. It is probably the most important step for a characterisation of a Region of Interest (ROI) containing MCs. The set of features of a given data instance is often grouped into a feature vector in order to be treated mathematically. Basically, researchers have investigated two types of features; those traditionally used by radiologists: intensity based and geometric features and high order features that may be not as intuitive to radiologists: texture features.

If the number of features is large compared to the number of training data, the generalisation ability of the conventional classifiers used in the following step may not be good. Therefore, to improve the generalisation ability, a small set of features is retained in an extra step known as Feature Selection step. It is difficult to predict which of the features' combination will have more accurate results. The only way to ensure an optimal feature vector is through exhaustive search of all possible subsets of features. However, the search space to be explored is too big.

Finally, a classification step is performed where the extracted features are provided as inputs to the classifier. This phase is known as false positive reductions. It is necessary to set up a classifier that hopefully maintains all the true detected signals and at the same time rejects almost all the false positive signal.

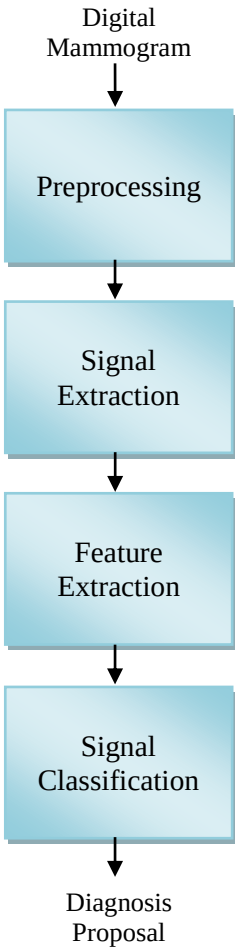


Figure 3.2: Flowchart of CADx system

3.3 CADx systems for clusters of microcalcifications

3.3.1 Morphology based CADx schemes

In the clinical practice the diagnosis of MC clusters is based on their morphological properties (shape, size, intensity) and the distribution of individual particles within a cluster. In this context, shape and intensity features are mainly exploited for the characterisation of a suspicious lesion. More specifically, the features that are used are; area, perimeter, elongation, circularity compactness, eccentricity, moment ratio, axis ratio, concavity index, effective thickness, volume and mean intensity, background intensity, contrast and edge strength (Karahaliou et al., 2012).

CADx schemes that classify MC clusters are based on two categories of cluster features (Karahaliou et al., 2012):

Category I: Cluster features based on descriptive statistics (e.g. average, standard deviation, coefficient of variation, maximum, median, range) of morphological properties of individual particles.

Category II: Cluster features describing cluster morphology considering the cluster as an entire object (cluster area, diameter, perimeter, circularity, eccentricity, elongation, solidity and cluster background intensity). In this category the spatial distribution of individual particles within a cluster is also considered (number of MCs, structural index, proximity to the nearest MC, cluster density, as well as distance to pectoral and breast edge).

3.3.2 Texture-based CADx schemes

A malignancy usually changes the texture of the surrounding tissue. Based on this hypothesis, CADx schemes can also exploit textural features. The systems based on texture analysis seem to provide better results than those based on morphology (Karahaliou et al., 2007). CAD systems based on morphology quantify properties that are visible to the eye. In this way they imitate the radiologists' assessments, but in an objective manner. Schemes based on texture analysis not only imitate the radiologists perception when visible lesions in the texture exist, but also enhance radiologists' capabilities by extracting and quantifying properties that associate with the diagnosis but are not visible to the naked eye.

3.4 Extracted Features

In literature a vast variety of features for the analysis of a cluster containing MCs has been exploited. These features can be categorised into six main groups (Andreadis et al., 2011).

1) Shape features of individual MCs

According to BIRADS system, the morphology of MCs is an important factor for their discrimination. In general, round and oval particles with lucent centres are considered benign, while thin, linear and small particles are considered an important sign of malignancy. The shape features that are usually extracted are: area, perimeter, compactness, circularity, elongation, eccentricity, spread, F3_F1 metric. In addition,

statistics of each feature like mean value, median and standard deviation are calculated for the whole cluster.

2) Cluster shape features

Except for the morphology of individual particles the morphology of the whole cluster is also considered by radiologists. Features such as the number of MCs in the cluster, the calcification coverage, area, circularity, eccentricity and the perimeter of the cluster can be calculated

3) Distribution features

The characteristics of this category describe the distribution of particles of MCs within the cluster. According to the BIRADS system, the MCs that appear in clusters are considered less suspicious of malignancy as opposed to linear, partial and scattered ones. The distances among individual MCs and the distances to the centre of the cluster are examples of features belonging to this category.

4) Optical density of individual MCs

These features are related to the grey values of each individual particle. For each MC its brightness is measured and its contrast to the surrounding tissue is estimated. The features of this category are highly dependent on the segmentation phase of the CAD pipeline.

5) Textural features

First order statistics (FOS) and Grey Level Co-occurrence Matrixes (GLCMs) are the most popular features of this category.

GLCMs: these features, proposed by Haralick et al (Haralick et al., 1973), are used for the characterisation of texture patterns and describe how often different pixel values occur in an image. They provide information concerning image texture heterogeneity and coarseness, which is not necessarily visually perceived.

FOS: depend only on single pixel values and represent properties of the intensity histogram of the region of interest (i.e. energy, entropy and square root of the coefficients norm) extracted from wavelet or multiwavelet analysis. Statistics about the grey values of the ROI, kurtosis, skewness and statistics related to the distribution of the histogram all belong to this category.

6) BIRADS descriptors

Features of this category involve the subtlety of an image, the radiologist's assessment, the description of the shape of MCs within the cluster and the age of the patient. These features are coded into numerical values following a rank ordering system proposed by Lo et al (Lo et al., 2003).

4. Chapter: Classification

4.1 Introduction

In general, the recognition or the classification of an object (pattern) can be described as the process by which signals that correspond to an object are classified to one of a finite set of categories, known as classes (Δερματάς, 1997). These categories concern groups of objects with similar characteristics or properties. Pattern recognition can be defined as the classification of data based on knowledge already gained or on statistical information extracted from patterns and/or their representations (Murty et al., 2011). Each object is described by properties known as features. The features constitute a description of all the known characteristics of the instance. In a classification problem every instance should be represented by the same set of features that are given in the form of a vector.

In general, pattern recognition is categorised according to the type of the learning procedure that is used to generate the output value. In a supervised learning problem the classifier is able to predict a value for a given valid input, after its training with a set of training examples (Δερματάς, 1997). The training set consists of a set of instances that have been properly labelled with the correct output. Then, a learning procedure generates a model that attempts to perform as well as possible on the training data and generalise as well as possible to new data (Zhang, 2000). On the other hand, unsupervised learning assumes unlabeled training data and attempts to find inherent patterns in the data that can be used to determine the correct output value for the new data instances. Finally, in reinforcement learning the examples of desired outputs are not given to the algorithm, as in supervised learning, but instead they have to be discovered by a process of trial and error (Barto, 1998).

Classification comprises the fourth step of a Computer Aided Diagnosis (CAD) procedure. In a CAD system designed for the classification of suspicious lesions, the characterisation of a lesion as benign or malignant is a supervised learning problem that has only two discrete output values; benignity and malignancy (Elter et al., 2009). For this reason, it is regarded as a two-class classification problem. There is a large number of techniques that have been developed for supervised classification based on Artificial Intelligence (Logical/Symbolic techniques), Perceptron-based techniques and Statistics

(Bayesian Networks, Instance-based techniques) (Kotsiantis, 2007). The most frequently used classifiers in CAD systems are K-Nearest Neighbours (KNN), Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and Linear Discriminant Analysis (LDA). Examples of classifiers that are less frequently applied in CADx systems include Bayes classification, generalized dynamic fuzzy neural networks, and rule-based expert systems (Elter et al., 2009).

In general, KNN is one of the simplest and most popular classifiers and one of the most common classifiers used to categorize lesions and microcalcifications (Nakayama et al., 2007). The classifier can distinguish unknown patterns based on the similarity with known examples. In order to do that, it calculates the distances of the unknown pattern to every existing pattern and finds the K nearest patterns as a basis for classification. The unknown pattern is then classified to the class to which the majority of its k nearest neighbours belong.

Artificial Neural Networks (ANNs) are inspired from the way the biological neural systems operate. ANNs simulate the human brain in the way by which knowledge is acquired and in the way by which the neurons are connected to each other with the aid of synapses bearing specific weights (Haykin, 1994).

A Support Vector Machine (SVM) is a binary classifier which abstracts a decision boundary in a multi-dimensional space using an appropriate subset of the training set of vectors. These vectors are called support vectors and geometrically are those training patterns that are closest to the decision boundary (Murty et al., 2011).

4.2 Artificial Neural Networks

An Artificial Neural Network (ANN) is a computational model inspired by the way the biological neural systems operate. It is a mathematical model which consists of a large number of independent computing elements, called neurons which are interconnected and organised in layers (Haykin, 1994).

4.2.1 Characteristics of ANNs

Artificial neural networks have the ability to learn from their environments and to improve their performance through learning. Learning is achieved through training;

an iterative process of gradual adjustments of the network's parameters, usually of the weights and of the bias levels (Haykin, 1994). In order to give a more formal definition of learning, the definition adapted from Mendel and McClaren is cited next; "Learning is a process by which the free parameters of a neural network are adapted through a process of stimulation by the environment in which the network is embedded. The type of learning is determined by the manner in which the parameter changes take place" (Mendel et al, 1970).

Training the neural network with pairs of known input patterns and their corresponding outputs, i.e. their corresponding classes, makes it possible to become specialised in this set of patterns. In this way, the outputs of unknown patterns of similar nature can be predicted next. In addition, ANNs are nonlinear models, which makes them flexible in modelling real world complex relationships. They have the ability to extract information from complex data, and learn from their environments in order to improve their performance. Also, they have the ability to organise themselves during the procedure of learning and they are able to operate in real time because of their parallel connection. Finally, ANNs are able to adapt their synaptic weights to changes in the surrounding environment which assures that a neural network that is trained to operate in a specific environment can be easily retrained to deal with minor changes in the operating environment conditions (Haykin, 1994).

4.2.2 The model of the artificial neuron

A neural network consists of interconnected computational units called neurons. The neuron is the basic information processing unit that transforms the input vector into a single output which is then connected to the inputs of the other neurons (Διαμαντάρας, 2007). In analogy with a biological neuron, an artificial neuron consists of a set of connections called synapses. Each neuron consists of multiple inputs and one single output. Each synapse is characterised by a weight that gives different levels of importance to the inputs (figure 4.1). In this way, an input signal x_i in the i^{th} input of the neuron is multiplied with the synaptic weight w_i. The weight can be positive or negative depending on whether the load that is released from the synapse stimulates the neuron to produce pulses with greater frequency or it suppresses it.

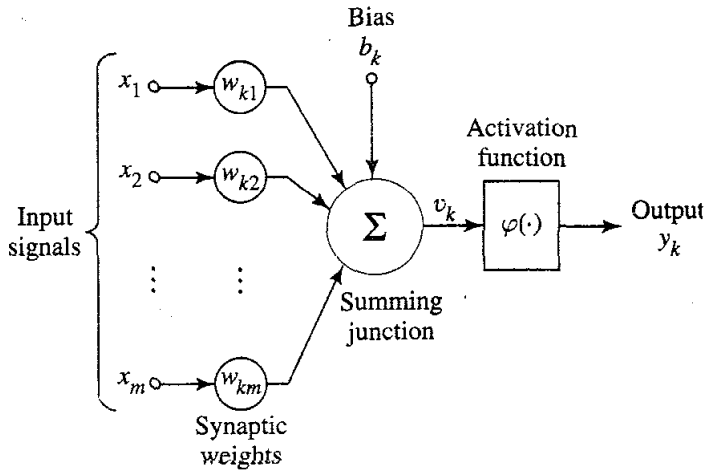


Figure 4.1: Non linear model of a neuron (Haykin, 1994)

The body of the neuron consists of an adder for summing the weighted inputs and an activation function $(\varphi(.))$ for reducing the amplitude of the neuron's output. The model also includes an externally applied bias b_k which aims to reduce the influence of the input to the activation function.

Using mathematics a k-neuron can be described with the following equations:

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \qquad (4.1)$$

$$y_k = \varphi(u_k - b_k) \qquad (4.2)$$

The signal u_k is also known as activation signal.

The aforementioned equations can be altered in order to include the external bias as an additional synapse with input $x_0 = -1$ and synaptic weight $w_{k0} = b_k$.

$$v_k = \sum_{j=0}^{m} w_{kj} x_j \qquad (4.3)$$

$$y_k = \varphi(v_k) \qquad (4.4)$$

Then the model of the neuron k is reformulated as in figure 4.2. Although the models in figures 4.1 and 4.2 differ in appearance, mathematically they are equivalent.

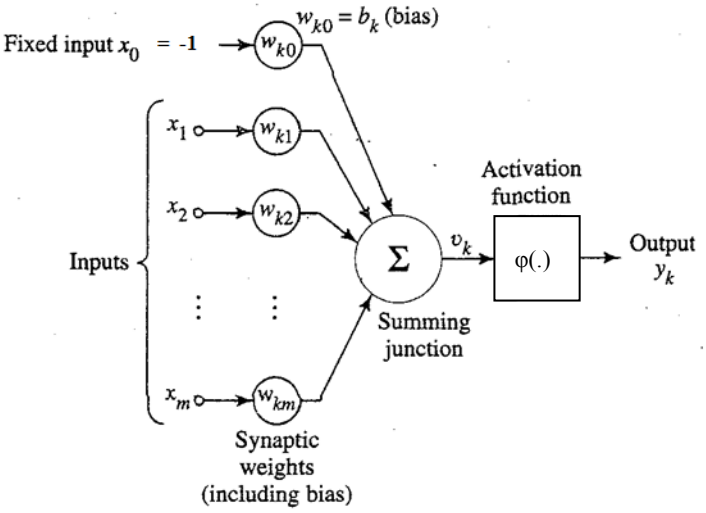


Figure 4.2: linear model of a neuron (another form) (Haykin, 1994)

Every neuron is in an internal state that is called 'activation level' and it constitutes the exit of the neuron. The neuron's activation depends on its inputs and on its activation function. So, the output of the neuron is defined by the activation function in terms of the induced input signal. The most frequently used activation functions are the following:

Threshold function

$$\varphi(v) = \begin{cases} 1, v \geq 0 \\ 0, v < 0 \end{cases} \qquad (4.5)$$

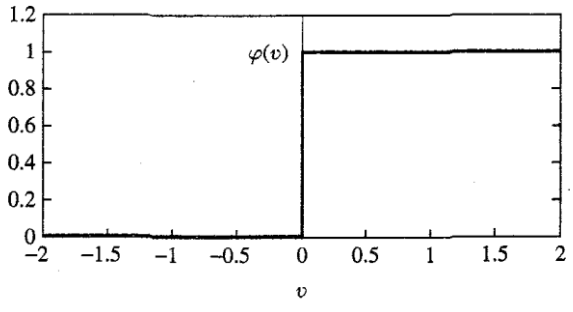


Figure 4.3: Threshold function

Piecewise-Linear Function

$$\varphi(v) = \begin{cases} 1, & v \geq \frac{1}{2} \\ v, -\frac{1}{2} \leq x \leq \frac{1}{2} \\ 0, & v \leq -\frac{1}{2} \end{cases} \quad (4.6)$$

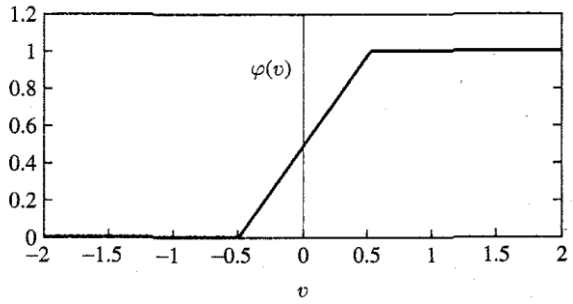


Figure 4.4: Piecewise-Linear Function

Sigmoid function

The sigmoid function is by far the most common form of activation function that is used in building artificial neural networks. An example of a sigmoid function is the logisti function defined by:

$$\varphi(v) = \frac{1}{1 + e^{-av}} \quad (4.7)$$

Parameter "α" is the slope parameter of the sigmoid function. By varying the parameter α, sigmoid functions of different slopes are acquired as illustrated in figure 4.5.

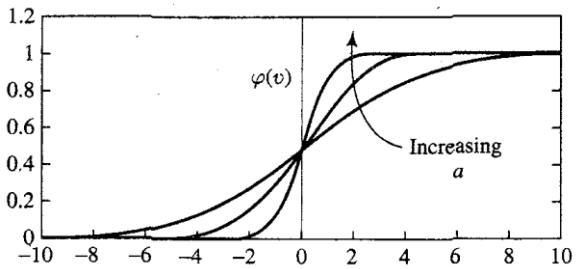


Figure 4.5: Sigmoid function

4.2.3 Neural Network Architectures

The way by which the neurons of the neural network are structured is closely linked to the learning algorithm that is used to train the network. As far as the architecture of the neural networks is concerned, two types of neural networks can be distinguished (Haykin, 1994):

- Feedforward or acyclic neural networks, where there are not any repetitive loops
- Recurrent neural networks, where there are repetitive loops through feedback connections

4.2.3.1 Feedforward networks

Two types of feedforward networks can be distinguished; single layer and multilayer networks. In a layered neural network the neurons are organised in layers. A simple single layer neural network consists of one input layer of source nodes and one output layer of neurons. The input layer does not count as a layer, because no computation is performed in that one.

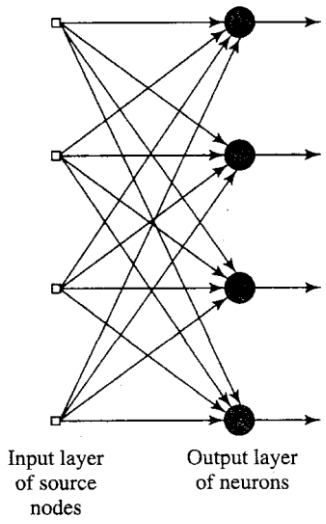


Figure 4.6: Single layer feedforward network (Haykin, 1994)

On the other hand, in a multilayer network there may be one or more intermediate layers of neurons, between the input nodes and the output layer, called hidden layers. By adding more hidden layers, the network is enabled to extract higher -

order statistics, a property particularly useful when the size of the input layer is large (Haykin, 1994). In feedforward networks, the information from the input signals flows in one direction, from the input layer to the output layer. Typically, the inputs of one hidden layer are the outputs of the preceding hidden layer.

In the following figure (figure 4.7), a multilayer feedforward network is illustrated. This network is said to be fully connected, in the sense that every node in each layer is connected to every other node in the adjacent forward network. However, there are networks in which the nodes of a layer are not connected to all the neurons of the following layer. This type of network is called partially connected, since some of the synaptic connections are missing.

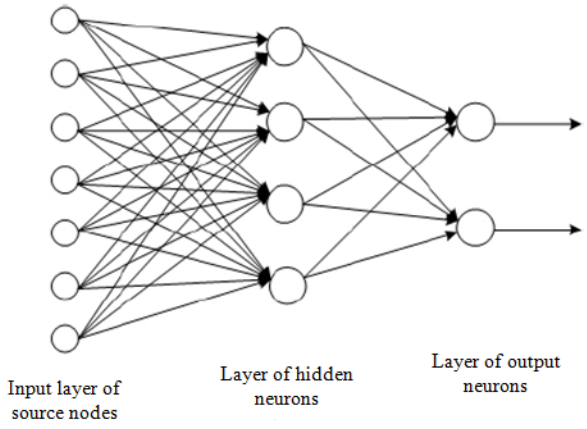


Figure 4.7: Fully connected feedforward multilayer network

4.2.3.2 Recurrent networks

Recurrent networks are different from feedforward networks in the sense that they have at least one feedback loop. In this type of networks, the output signals of one or more neurons from the output and/or the hidden layers, feed the inputs of other neurons. Also, it is possible that the signal of a neuron feeds its own input. This type of loops are called self-feedback loops. In the following figures (figure 4.8 and 4.9), two examples of recurrent networks are depicted. The network of figure 4.8 is a single layer network, with each neuron feeding its output signal back to the inputs of all the other neurons. Figure 4.9 illustrates another type of a recurrent network with hidden neurons, where the feedback connections originate both from hidden and output neurons.

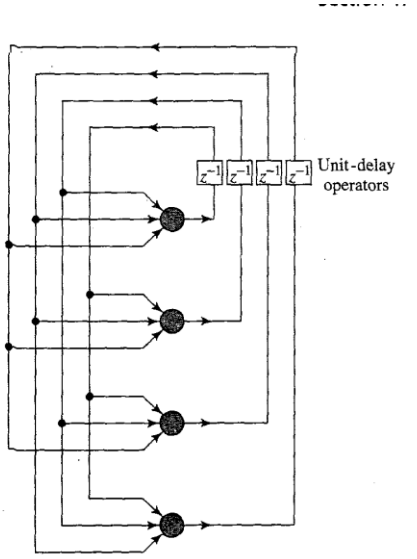


Figure 4.8: Recurrent network with no self-feedback loops and no hidden layers of neurons (Haykin, 1994)

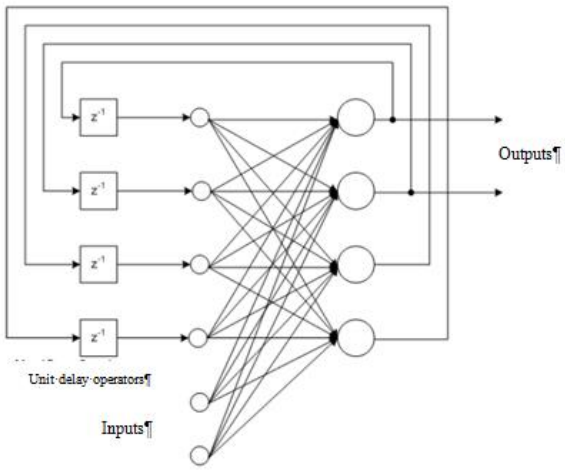


Figure 4.9: Recurrent network with feedbacks originating from hidden and output neurons (Haykin, 1994)

4.2.4 Training Neural Networks

One of the most important properties of a neural network is its ability to learn from its environment and improve its performance through learning. A neural network learns about its environment through an interactive process of adjusting its synaptic weights and bias levels that takes place ideally after each iteration of the learning process (Haykin, 1994). Two types of neural networks' learning algorithms (also known as training algorithms) can be distinguished;

- Supervised learning

- Unsupervised learning
- Reinforcement learning

4.2.4.1 Supervised Learning

In general neural networks are not aware of their surrounding environment. However, in supervised learning, there is an external "teacher" who has knowledge of the environment that is represented in form of a set of input-output examples. When a training vector is given as input to the neural network, the teacher provides the network with the desired output for that input which represents the network's optimum response. Then, the network's parameters (weights and bias levels) are adjusted under the influence of the training vector and the error signal. The error signal results from the difference between the network's actual and desired output. This adjustment is repeated iteratively in a step by step manner with the purpose of eventually making the neural network to emulate the teacher in the best possible way. In this way, the knowledge of the environment, known to the teacher, is transfered to the neural network through the training procedure as fully as possible. Next, the teacher is removed and the network is able to deal with the environment by itself.

This type of supervised learning is the error correction learning described in the next subsection (*4.2.4.1.1)*. It is a closed loop feedback system, where the environment is not included in the loop. The mean square error, or the sum of squared errors over the training samples can be considered as the performance measures of the system.

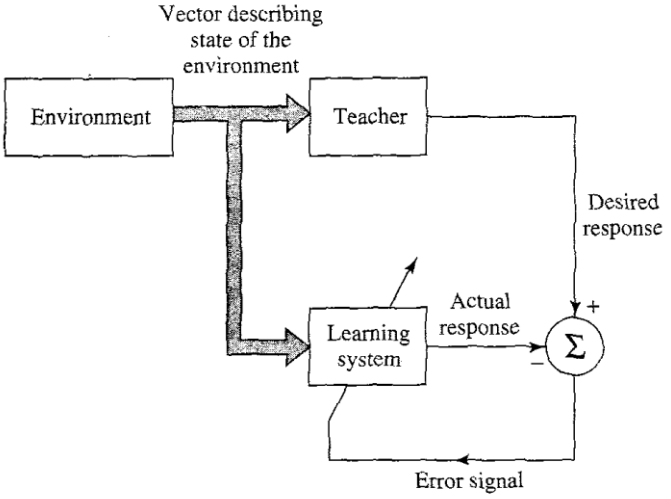


Figure 4.10: Block diagram of supervised learning (Haykin, 1994)

Examples of this type of training algorithms are the least-mean-square (LMS) algorithm of Widrow and Hoff (Widrow et al., 1960) and its generalisation known as back propagation (BP) (Werbos P.J. 1974). The LMS algorithm consists of only one neuron, while BP consists of a number of connected neurons in layers.

4.2.4.1.1 Error-Correction Learning

To illustrate this type of learning, a multilayer feedforward network is considered which consists of one input layer, one or more layers of hidden neurons and one output layer consisting of only one neuron.

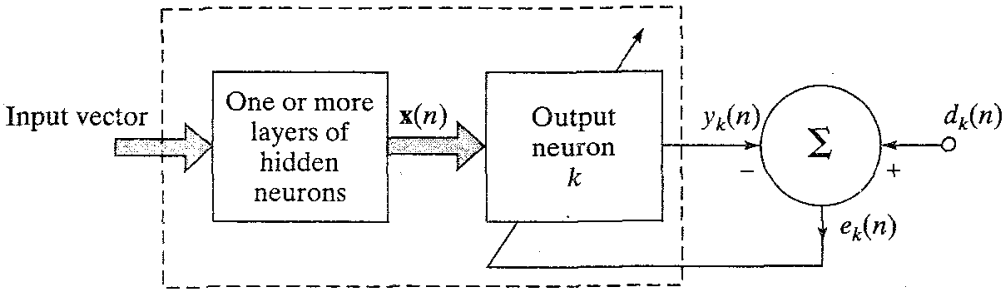


Figure 4.11: Multilayer feedforward network with one or more layers of hidden neurons and one neuron at the output layer (Haykin, 1994)

Neuron k is driven by a signal vector $x(n)$ produced by the hidden neurons which are driven by an input vector applied to the input layer of the neural network. The output signal of the neuron k is denoted by $y_k(n)$. This output signal, which is the only output of the neural network, is then compared to a desired response (target output) $d_k(n)$ and thus an error signal $e_k(n)$ is produced.

$$e_k(n) = d_k(n) - y_k(n) \tag{4.8}$$

This error signal motivates a control mechanism whose purpose is to apply a series of corrective adjustments to the synaptic weights of neuron k in order to make the output signal $y_k(n)$ come closer to the desired response $d_k(n)$ in a step-by-step iterative procedure until the synaptic weights are essentially stabilized. This objective is achieved by minimizing a cost function $\mathcal{E}(n)$,that is the instantaneous value of the error energy:

$$\mathcal{E}(n) = \frac{1}{2}\sum_{k\in C} e_k^2(n) \tag{4.9}$$

Where the set C includes all the neurons of the output layer. Let N be the total number of patterns in the training set. Then the average squared error energy is defined as follows:

$$\mathcal{E}_{av} = \frac{1}{N}\sum_{n=1}^{N} \mathcal{E}(n) \tag{4.10}$$

The instaneous error energy $\mathcal{E}(n)$ and the average error energy $\mathcal{E}_{av}$ are functions of the synaptic weights and the bias levels of the ANN. For a given training set, the fucntion $\mathcal{E}_{av}$ represents the cost function, as the measure of the efficiency of the learning process. The purpose of the learning process is the adjustment of the free parameters of the ANN (synaptic weights and bias levels) in order to minimize the average error energy $\mathcal{E}_{av}$. The weights are updated on a pattern-by-pattern basis, i.e. the adjustments of the weights are done according to the errors that are calculated for each pattern presented to the network.

Assuming a neuron j, and a set of function signals $y_i(n)$ produced by a layer of neurons to its input. Then the induced local field produced at the input of the activation function is as follows:

$$v_j(n) = \sum_{i=0}^{m} w_{ji}\,(\mathrm{n}) y_i(\mathrm{n}) \tag{4.11}$$

Then the function signal at the output of the j^{th} neuron at iteration n is

$$y_j(n) = \varphi(v_j(n)) \tag{4.12}$$

The backpropagation algorithm applies a correction $\Delta w_{ji}(n)$ to the synaptic weight $w_{ji}(n)$ which is proportional to the partial derivative $\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)}$. It can be shown that the applied correction is given by:

$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} \tag{4.13}$$

Where η is called the learning rate parameter.

4.2.4.2 Learning without a teacher

Learning without a teacher does not include a "teacher" supervising the learning process. This means that there are not any labeled examples to be learned by the network. This type of learning includes two subtypes;

- Reinforcement learning
- Unsupervised learning

4.2.4.2.1 Reinforcement learning

In reinforcement learning, the learning of an input-output mapping is performed through continued interaction with the environment in order to minimize a scalar index of performance (Haykin, 1994). In figure 4.12, the block diagram of one form of reinforcement learning system is shown. The system is built around a critic that converts a primary reinforcement signal received from the environment into a reinforcement signal of higher quality called the heuristic reinforcement signal, both of which are scalar inputs (Barto, 1998). The system is designed to learn under delayed reinforcement, which means that the system observes a temporal sequence of stimuli also received from the environment, which eventually results in the generation of the heuristic reinforcement signal. The goal of learning is to minimize a cost-to-go function defined as the expectation of the aggregated cost of actions taken over a sequence of steps instead of simply the immedicate cost (Haykin, 1994). The function of the learning machine, which constitutes the second component of the system is to discover these actions and to feed them back to the environment.

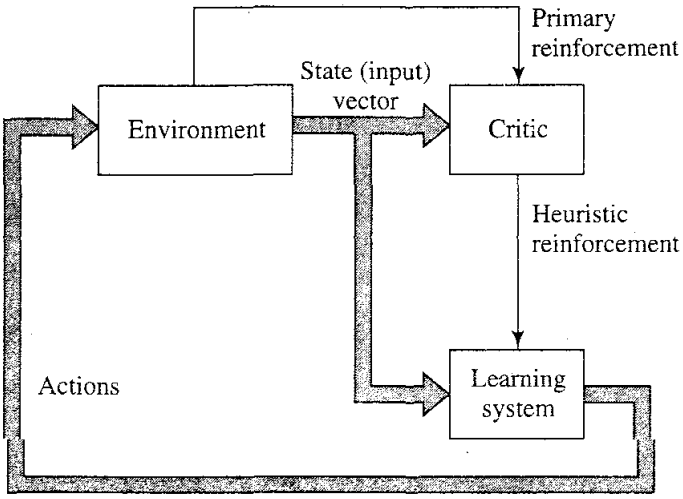


Figure 4.12: Block diagram of reinforcement learning (Haykin, 1994)

4.2.4.2.2 Unsupervised learning

In unsupervised learning also known as self-organized learning there is no external teacher or critic to oversee the learning process as shown in figure 4.13. Rather, a task-independent measure of the quality of representation that the network is required to learn is provided and the network's parameters are optimized with respect to it. Once the network becomes tuned to the statistical regularities of the input data, it develops the ability to form internal representations for encoding features of the input and thereby to automatically create new classes (Becker, 1991).

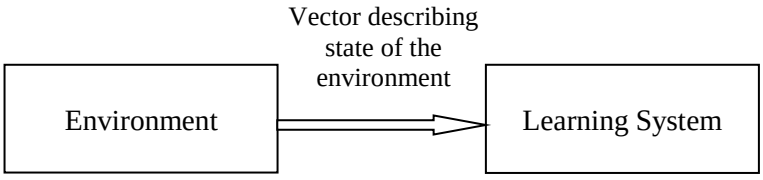


Figure 4.13: Block diagram of unsupervised learning (Haykin, 1994)

4.3 Support Vector Machines

4.3.1 Introduction

Support vector machines (SVMs) are a supervised learning-based method, primarily used for binary and multi-class classification (Vapnik, 2000). Given a set of training examples, each marked as belonging to one of two classes, an SVM training

algorithm builds a model in order to assign new examples into one of these classes. An SVM builds a hyperplane (or a set of hyperplanes) in a high or infinite dimensional space that can be used for classification and regression. SVMs revolve around the notion of "margin", which is the distance of a hyperplane to the nearest training data point of any class (Nakayama et al., 2007). A good separation is achieved by the hyperplane that has the largest margin, because the larger the margin, the better the separation of the instances and thus the lower the generalization error of the classifier (Nakayama et al., 2007).

Two types of Support Vector Machines can be distinguished; hard and soft partitioning SVMs, depending on whether the training data in the input space are linearly separable or not (Abe, 2010).

4.3.2 Hard-Margin Support Vector Machines

A set of patterns is said to be linearly separable if a hyperplane can be found so that patterns of different classes fall on different sides of the hyperplane. In the following figure (Figure 4.14), for simplicity reasons, two-dimensional data is considered, where each pattern is represented by a point in the two-dimensional space. Assuming five patterns that belong to class labelled as X and four patterns of class O, a line e.g. $x_1 = x_2$ can be found so that all the X patterns fall on the left side of the line and all the O patterns on the right side (Murty et al., 2011).

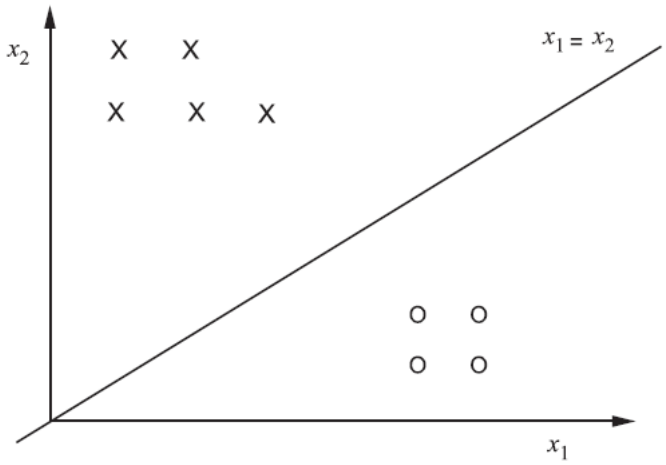


Figure 4.14: Linearly seperable 2D data (Murty et al., 2011)

Another way of separating the patterns of the two classes is the following:

- all the patterns of class X satisfy the property that $x_1 < x_2$or $x_1 - x_2 < 0$ and equivalently
- all the patterns of class O satisfy the property $x_1 > x_2$ or $x_1 - x_2 > 0$.

There are infinite ways of realising the decision boundary. All the lines can be written in the following form:

$$f(x) = w_1 x_1 + w_2 x_2 + b = 0 \quad (4.14)$$

In the multidimensional space, the linear decision boundary becomes a hyperplane and it can be represented by the following equation:

$$f(x) = w^T x + b = 0 \quad (4.15)$$

where w and x are m-dimensional vectors (Murty et al., 2011).

Let M m-dimensional training inputs x_i $(i = 1, \dots, M)$ belong to Class 1 or Class 2 with the associated labels be $y_i = 1$ for Class 1 and $y_i = -1$ for Class 2. For the linearly separable data, the decision function for the classification of an unknown pattern is defined as follows:

$$D(x) = sgn(w^T x + b) \quad (4.16)$$

where w is an $m-$dimensional vector, b is a bias term (the term-b is called threshold)

Equivalently the decision function can be written as shown in equations 4.13 and 4.14:

$$D(x) = w^T x + b \quad (4.17)$$

and

$$w^T x_i + b \begin{cases} > 1 & for\ y_i = 1 \\ < -1 & for\ y_i = -1 \end{cases} \text{ for } i = 1, \dots, M \quad (4.18)$$

Here, 1 and -1 on the right-hand sides of the inequalities of equation 4.14 can also be constants $a > 0$ and $-a$, respectively (Abe, 2010). Equation 4.14 is equivalent to

$$y_i(w^T x_i + b) \geq 1 \text{ for } i = 1, \dots, M \tag{4.19}$$

The hyperplane

$$D(x) = w^T x + b = c \text{ for } -1 < c < 1 \tag{4.20}$$

forms a separating hyperplane that separates x_i $(i = 1, \dots, M)$. When c = 0, the separating hyperplane is in the middle of the two hyperplanes with $c = 1$ and $c = -1$. The distance between the separating hyperplane and the training data sample nearest to the hyperplane is called the margin. Assuming that the hyperplanes $D(x) = 1$ and $D(x) = -1$ include at least one training data sample, the hyperplane $D(x) = 0$ has the maximal margin for $-1 < c < 1$. The region $\{x| -1 \leq D(x) \leq 1\}$ is the generalization region for the decision function (Abe, 2010).

The optimal separating hyperplane can be found by minimizing the squared norm of the separating hyperplane. The minimization can be set up as a convex quadratic programming problem (Kotsiantis, 2007):

$$Minimize_{w,b} \Phi(\boldsymbol{w}) = \frac{1}{2} \|\boldsymbol{w}^2\| \tag{4.21}$$

Subject to $y_i(w^T x_i + b) \geq 1$ for $i = 1, \dots, l$

In general, there is an infinite number of decision functions which are separating hyperplanes and satisfy equation 4.15. Figure 4.15 shows two decision functions that satisfy equation 4.15. The generalization ability depends on the location of the separating hyperplane, and the hyperplane with the maximal margin is called the optimal separating hyperplane.

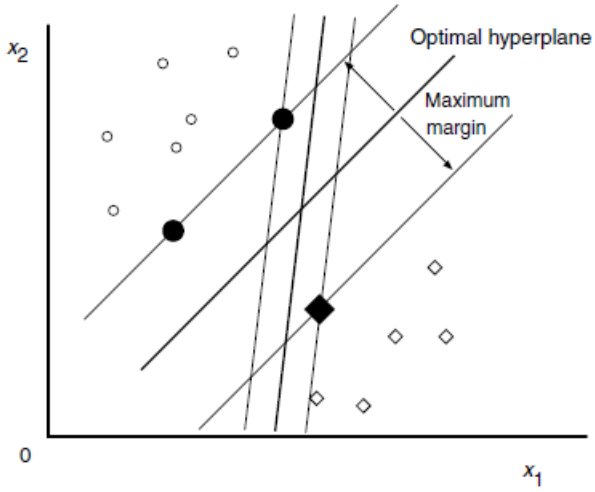


Figure 4.15: Optimal separating hyperplane in a two-dimensional space (Abe, 2010)

Once the optimal separating hyperplane is found, data points that lie on its margin are known as support vector points and the plane can be represented as a linear combination of only these points. Other data points are ignored. Therefore, the model complexity of an SVM is unaffected by the number of features encountered in the training data (the number of support vectors selected by the SVM learning algorithm is usually small). For this reason, SVMS are well suited to deal with learning tasks where the number of features is large with respect to the number of training instances (Kotsiantis, 2007).

4.3.3 Soft-Margin Support Vector Machines

When the training data is not linearly separable, the SVM may not be able to find any separating hyperplane. The problem can be addressed by using a soft margin that accepts some misclassification of the training instances (Veropoulos et al., 1999). This can be done by introducing non negative slack variables $\xi_i,\ i = 1, \dots, M$ into equation.4.15 which then becomes:

$$wx_i - b \geq +1 - \xi, \text{ for } y_i = +1$$
$$wx_i - b \leq -1 + \xi \text{ for } y_i = -1 \qquad (4.22)$$
$$\xi \geq 0$$

In most real-world problems, the training data is non-separable, thus no hyperplane fullfilling all conditions of equation 4.15 exists and thus the generalization ability is degraded. One solution to the inseparability problem is to map the data onto a higher dimensional space and define a separating hyperplane there. This higher-dimensional space is called the feature space (Kotsiantis, 2007). With an appropriately chosen transformed feature space of sufficient dimensionality, any consistent training set can be made separable. A linear separation in transformed feature space corresponds to a non-linear separation in the original input space (Abe, 2010).

Using the nonlinear vector function $\varphi(x) = (\varphi_1(x), \dots, \varphi_l(x))$ that maps the m-dimensional input vector x into the l -dimensional feature space, the linear decision function in the feature space is given by

$$D(x) = w^T \varphi(x) + b \tag{4.23}$$

where w is an l -dimensional vector and b is a bias term (Abe, 2010).

In order to avoid mappings, a kernel function can be defined in order to allow inner products to be calculated directly in feature space without performing mappings. Once a hyperplane has been created, the kernel function is used to map new points into the feature space for classification. The selection of an appropriate kernel function is important, since the kernel function defines the transformed feature space in which the training set instances will be classified (Kotsiantis, 2007). Some popular kernels are the following:

$$K(x, y) = (x \cdot y + 1)^p \tag{4.24}$$

$$K(x, y) = e^{\frac{-\|x-y\|^2}{2\sigma^2}} \tag{4.25}$$

$$K(x, y) = tanh(kx \cdot y - \delta)^p \tag{4.26}$$

5. Chapter: Cluster analysis

5.1 Introduction

Clustering or cluster analysis is the process of grouping similar objects (Hartigan, 1975; Duda, et al., 1973). Each group, called cluster, consists of objects (patterns) that are similar between themselves and dissimilar to objects of other groups. This brings up the notion of similarity. Similarity is quantified using a distance measure in a way so that the distance between objects of the same cluster (intra-cluster distance) is low and the distance between objects of different clusters (inter-cluster distance) is high (Murty et al, 2011). In the following figure (figure 5.1) an example of 3 clusters is provided to assist in the illustration of this notion. For simplicity reasons two dimensional data are considered, where each pattern is represented by a point in the two dimensional space.

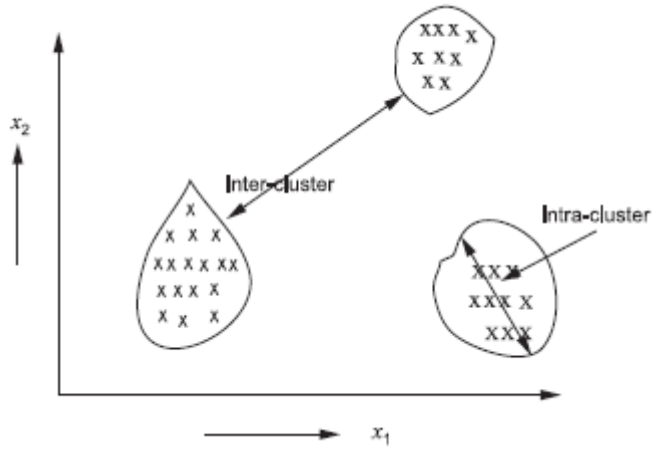


Figure 5.1: Example of 3 clusters in the 2D space (Murty et al, 2011)

In contrast to the supervised learning problem, where the labels of the data are known, clustering is an unsupervised learning problem. This means that it does not require the objects to be labeled in order to assign a new object to a cluster. Clustering has applications in many areas; biometrics, bioinformatics, document analysis and recognition, information retrieval, remote data analysis, target recognition and data mining (Murty et al, 2011).

5.2 Basic Steps in Clustering

The basic steps of the clustering procedure are illustrated in the Figure 5.2:

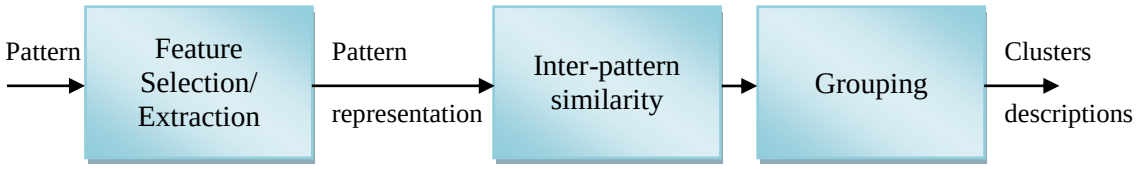


Figure 5.2: The basic stages in clustering

1. Feature Extraction/Feature Selection: As in supervised learning an object is represented by a set of features. A pattern is the representation of an object by the values taken by the features. Features must be properly selected in order to encode as much information as possible depending on the classification problem.

2. Definition of an appropriate similarity measure: A distance measure is the metric used to compute the similarity between pattern representations. Patterns that are more similar are closer to each other and thus they have smaller distance. Some of the most commonly used distance measures are presented in section 5.3.1.

3 Selection of the clustering algorithm and using it in order to generate a partition of the clusters. The most common clustering algorithms are described in section 5.3

5.3 Clustering algorithms

A clustering algorithm is a learning procedure that tries to identify the specific characteristics of the clusters underlying the data set (Theodoridis et al, 2009). There is a wide variety of clustering algorithms. A common distinction between clustering algorithms is based on whether the partitions that are generated are overlapping or not. In this context, there is the distinction between hard and soft clustering paradigms. Examples of hard clustering algorithms are the "hierarchical clustering algorithms", where a nested sequence of partitions is generated, and the "partitional clustering algorithms" where a partition of the given data is generated. Soft clustering algorithms are based on fuzzy sets, rough sets, artificial neural networks, or evolutionary

algorithms, specifically genetic algorithms (Murty et al, 2011). In soft clustering, each object belongs to a cluster with some probability, whereas in hard clustering an object either belongs to a cluster or not.

5.3.1 Distance measures

The results of the clustering algorithms may depend on the distance measure that is used. A distance measure should have the following properties (Murty et al, 2011):

- Positive reflexivity $d(x, x) = 0$
- Symmetry $d(x, y) = d(y, x)$
- Triangular inequality $d(x, y) \leq d(x, z) + d(z, y)$

The most common distance measure that is used in clustering algorithms is the squared Euclidean distance. Assuming two M-dimensional points $x = (x_k)$ and $y = (y_k)$, the squared Euclidean distance can be defined as follows:

$$d(x, y) = \sum_{k=1}^{M} |x_k - y_k|^2 \qquad (5.8)$$

Another common distance measure, the Euclidean distance is defined by the equation 5.2:

$$d^2(x, y) = \sqrt{\sum_{k=1}^{M} |x_k - y_k|^2} \qquad (5.9)$$

The Cityblock distance, also known as the Manhattan distance, is commonly used as well and is defined by the equation 5.3:

$$d^1(x, y) = \sum_{k=1}^{M} |x_k - y_k| \qquad (5.10)$$

The Minkowski p-metric is a generalisation of the aforementioned distances and it is defined as in the equation 5.4 (De Amorim and Mirkin., 2012):

$$d^p(x,y) = \left(\sum_{k=1}^{M} |x_k - y_k|^p\right)^{\frac{1}{p}} \quad (5.11)$$

From this equation (Eq. 5.4), it is obvious that the Minkowski p-metric is equivalent to the Cityblock distance, if p=1 and to the Euclidean distance if p=2. For p→∞;the distance metric is called Chebyshev or Maximum metric and it takes the following form (Eq. 5.5):

$$d^{\infty}(x,y) = max_{k=1,\dots M}|x_k - y_k| \quad (5.12)$$

5.3.2 Hierarchical Clustering

Hierarchical Clustering algorithms build a hierarchy of data partitions that is represented using a tree structure, called dendrogram. These tree structures allow a root cluster having branch clusters each of which may later become a root cluster and generate more branches. Hierarchical algorithms are further subdivided into agglomerative and divisive (Murty et al, 2011; Theodoridis et al, 2009).

Agglomerative algorithms: these algorithms produce a sequence of clusterings of decreasing number of clusters at each step using a bottom-up approach. They start with having n singleton clusters, each one containing one of the n patterns of the input data set. At each step, the most similar pair of clusters is merged to reduce the size of the partition by one. An important property of agglomerative algorithms is that once two patterns are placed in the same cluster at a level, they remain in the same cluster at all subsequent levels.

Divisive algorithms: they use a top-down strategy for generating the partitions of the data. They start with a single cluster having all the n patterns and at each successive step, a cluster is split. This process continues until there is only one pattern in a cluster or a collection of singleton clusters. Similarly with the agglomerative algorithms, once

two patterns are placed in two different clusters at a level, they remain in different clusters at all subsequent levels.

Decisions concerning which clusters should be combined or whether a cluster should be split require a measure of similarity between patterns. This is achieved by using an appropriate distance measure and a linkage criterion. The former is used in order to specify the similarity between two patterns and so creating the clusters. The later specifies the similarity between pairs of patterns in order to merge or split clusters in the second place, i.e. it constitutes a distance measure between clusters.

The most commonly used linkage criteria between two sets of observations A and B are the following (De Amorim, 2011):

Single linkage clustering: it defines the distance between two clusters as the minimum possible. This method finds the two patterns, one of each cluster, that are closest to each other. The following equation gives the distance between clusters X and Y that is given by the distance between the two closest entities $x \in X, y \in Y$ in the two clusters.

$$D(X,Y) = \min_{x \in X, y \in Y} d(x,y) \qquad (5.13)$$

Maximum or full linkage clustering: This method finds two entities $x \in X, y \in Y$ that are the farthest from each other.

$$D(X,Y) = \max_{x \in X, y \in Y} d(x,y) \qquad (5.14)$$

5.3.3 Partitional Clustering

Partitional clustering algorithms are based on the optimisation of an objective function J. This is usually done in an iterative way until a local optimum of J is determined. A major advantage of partitional clustering is the fact that iterative optimisation may gradually improve clusters (Berkhin, 2006).

This category includes probabilistic algorithms, like the Expectation–maximization algorithm as well as centroid-based algorithms like the k-means algorithm and all its variations, e.g. k-medoids, Fuzzy c-means and k-means++.

5.4 K-means Clustering

K-means originated independently in the works of MacQueen (MacQueen,. 1967) and Ball and Hall (Ball et al, 1967) and it is the most popular clustering algorithm that produces non-overlapping clusters. It is said to be more efficient than the hierarchical algorithms (Manning et al., 2008).

K-means partitions a dataset of N objects $Y = \{y_1, y_2, \dots, y_N\}$, each represented by M features, into K non-empty and non-overlapping clusters $S = \{S_1, S_2, \dots, S_K\}$. Each cluster is represented by its gravity centre; an M-dimensional vector called centroid; $c_k \in C = \{c_1, c_2, \dots, c_K\}$.

The algorithm iteratively minimizes the following criterion, which represents the sum of the within-cluster distances to centroids.

$$W(S,C) = \sum_{k=1}^{K} \sum_{i \in S_k} d(y_i, c_k) \qquad (5.15)$$

where $d(y_i, c_k)$ is the dissimilarity metric between y_i and its respective centroid c_k.

This criterion can also be written as follows:

$$W(S,C) = \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{M} s_{ik}\, d(y_{ij}, c_{kj}) \qquad (5.16)$$

where y_{ij} is the value of the j^{th} feature at entity i and s_{ik} is a variable representing the binary cluster membership such that $s_{ik} = 1$ if $i \in s_{ik}$ and $s_{ik} = 0$ otherwise.

In most cases the dissimilarity measure is chosen to be the Euclidean distance. Then the aforementioned criterion can be rewritten as follows:

$$W(S,C) = \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{M} s_{ik} \, (y_{ij} - c_{kj})^2 \qquad (5.17)$$

In other words, the goal is to find values for the $\{s_{ik}\}$ and the $\{c_{kj}\}$ so as to minimise the cost function W. This can be done through an iterative procedure in which each iteration involves two successive steps corresponding to successive optimisations with respect to the $\{s_{ik}\}$ and the $\{c_{kj}\}$. At first, some initial values for the centroids $\{c_1, c_2, \dots, c_K\}$ are chosen. Then in the first step, the cost function W is minimised with respect to the $\{s_{ik}\}$, keeping the centroids $\{c_1, c_2, \dots, c_K\}$ fixed. In the second step, W is minimised with respect to the centroids $\{c_1, c_2, \dots, c_K\}$ keeping $\{s_{ik}\}$, fixed. This two-stage optimisation is repeated until convergence (Bishop, 2006).

The K-means algorithm can be defined more formally in four steps (Mirkin, 2005):

1. Initialization: Define the value of the number of clusters K and initialize randomly the centroids $\{c_1, c_2, \dots, c_K\}$ also referred as seeds.

2. Cluster update: Given the K centroids, assign the N entities to their respective centroid using the minimum distance rule.

3. Stop condition: Check whether there is a change in the formation of the clusters. If there are changes then go to step 4, else, the clustering task is assumed to be finished and the generated partitions of clusters $S = \{S_1, S_2, \dots, S_K\}$ are final. Other possible stop conditions are a limitation in the number of iterations or a threshold for the objective function.

4. Centroids update: Update the centroids of each cluster to the centre of gravity of the cluster. If the squared Euclidean distance is used as distance metric, the centroids are set equal to the means of their clusters.

5.5 Finding the initial centroids

Finding the initial centroids and the number of clusters K automatically are among the most desirable properties of clustering algorithms (Callan, 2003). The problem of finding the number of clusters K and the centroids has been widely addressed in literature, with the "intelligent K-means" (iK-means) method (Mirkin, 2005; Callan, 2003) being one of the most popular.

Intelligent K-Means utilizes the so-called anomalous clusters that are found before running the K-Means itself. Anomalous clusters are extracted one-by-one until no unclustered patterns remain. At the end, the centroids of the largest anomalous clusters are used to initialize K-Means. Each of the anomalous clusters is built by taking, as its initial centroid, the entity that is farthest away from a pre-specified ''reference point''. Then the anomalous cluster is filled in by the entities that are nearer to it than to the reference point, which is iteratively updated by updating the centre of the anomalous cluster in the manner of K-Means itself, until the cluster stops changing. The resulting anomalous cluster is removed from the dataset, and the procedure is repeated with the reference point unmoved. When all entities have been clustered in this way, the centroids of the non-singleton anomalous clusters are used to both set K and initialize K-Means. If there is no information regarding the reference point then it is set to be the central point of the dataset that minimizes the summary distance to all the data entities.

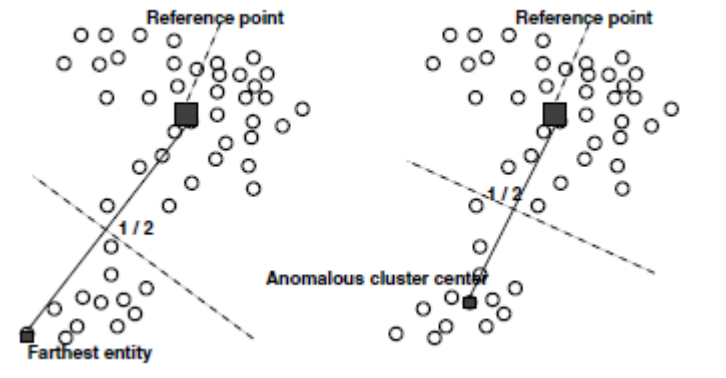


Figure 5.3: Anomalous pattern clustering (Mirkin, 2005)

Formally the anomalous pattern algorithm can be defined in the following stages (Mirkin, 2005):

1. Pre-processing: Specify a reference point a. If there is no information regarding the reference point, it is set to be the central point of the dataset that minimizes the summary distance to all the data entities.

2. Initial Setting: Take as initial centroid c, the entity that is farthest away from the pre-specified reference point.

3. Cluster update: Create a cluster S, of entities that are closer to c than to a. An entity y_i should be assigned to S if $d(y_i, c) < d(y_i, a)$.

4. Centroid update: Calculate the central point of the cluster S, c' and check whether it differs from the previous centroid c. If c' and c differ update the centroid by assigning $c \leftarrow c'$and return to step 3.

5. Output: Return the list S of entities and centroid s.

Like K-Means itself, the Anomalous pattern alternately minimizes a criterion,

$$W(S, c) = \sum_{i \in S} d(y_i, c) + \sum_{i \notin S} d(y_i, a) \qquad (5.18)$$

The intelligent K-means algorithm uses an anomalous pattern iteratively to get one cluster at a time and can be formally defined in the following steps (Mirkin, 2005):

1. Setting: Specify the cluster discarding threshold used to remove all anomalous clusters whose size is less than the threshold, and standardize the dataset.

2. Anomalous pattern: Apply the anomalous pattern algorithm using the centre of gravity of the whole dataset as a reference point. The position of this centre is not changed during any part of the algorithm.

3. Control: Stop if all entities are clustered, otherwise remove the anomalous cluster found in the previous step from the dataset, and return to step 2. Other possible stopping conditions exist, like the reach of a pre-specified number of

clusters.

4. Discard of small clusters: Remove these clusters that are smaller than the pre-specified threshold, defined in the first step. If there are clusters with only one element (singletons), there may be errors in the data, possibly data entry mistakes.

5. K-Means: Run K-Means using the found centroids, and the full original dataset.

5.6 Feature Weighting in K-means

One of the basic drawbacks of the K-means algorithm is that it treats all features equally and thus it may have difficulties in clustering data with irrelevant, or noise features. This is not desirable in many applications, such as data mining, where data often contains a large number of diverse variables. In this section some feature weighting algorithms based on K-means are described.

5.6.1 Weighted K-means

Chan, Huang and their colleagues (Chan et al., 2004; Huang et al, 2005; Huang et al, 2008) developed the Weighted K-means algorithm (WK-means) by modifying Eq. 5.9 in order to assign weights to features based on their importance in clustering.

The criterion of the K-means algorithm is altered as follows:

$$W(S,C,w) = \sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{M} s_{ik}\, w_j^{\beta}(y_{ij} - c_{kj})^2 \qquad (5.19)$$

The weights w_j are non-negative and they sum to unity, such as $\sum_{j=1}^{M} w_j = 1$. The exponent β is a parameter defined by the user and expresses the degree of the weights' contribution to the distance. Thus, the real values of the weights depend on the

parameter β which can take values such that $\beta < 0$ or $\beta > 1$ (Chan et al., 2004; Huang et al, 2005; Huang et al, 2008).

WK-means adds an extra step to the standard K-means algorithm in order to update the feature weights based on the current partition of the data. The weights are updated according to the following equation:

$$w_j = \frac{1}{\sum_{u=1}^{h} \left[\frac{D_j}{D_u}\right]^{1/(\beta-1)}} \qquad (5.20)$$

where

$$D_j = \sum_{k=1}^{K} \sum_{i=1}^{N} s_{ik}(y_{ij} - c_{kj})^2 \qquad (5.21)$$

is the sum of the within-cluster distances of feature j and h is the number of features for which $D_j \neq 0$.

Lower weights are assigned to features that have larger D_j, thus reducing the impact of noisy or insignificant features in the clustering process.

The distance measure is altered in order to take into consideration the weights introduced in this algorithm. The distance between the j^{th} feature of an entity y_{ij} and the centroid c_k is given by;

$$d(y_i, c_k) = \sum_{j=1}^{M} w_j^{\beta} \left|y_{ij} - c_{kj}\right|^2 \qquad (5.22)$$

The equation 5.13 is not applicable in two cases:

i) When $D_u = 0$ for $u \in [1, M]$. To solve the issue of dividing by zero it is suggested adding a non-negative constant to each item in the definition of D_u. Such a

constant might be the average dispersion of all features in the dataset (Chan et al., 2004; Huang et al, 2005; Huang et al, 2008).

ii) When $\beta = 1$. In this case the WK-means criterion of equation 5.12 is minimised by setting $w_{j^*} = 1$ where j^* represents the feature with the smallest sum of the within cluster variance. All other feature weights are set equal to 0.

Huang et al. (Chan et al., 2004; Huang et al, 2005; Huang et al, 2008) noticed that a feature may have greater relevance within one cluster, than within another one. For this reason they extended the criterion of Eq. 5.12 so that the feature weights become cluster-specific. For this purpose, some small modifications should take place to w_j and D_j These are changed to w_{kj} and D_{kj}, such as

$$w_{kj} = \frac{1}{\sum_{u=1}^{h} \left[\frac{D_{kj}}{D_{ku}}\right]^{1/(\beta-1)}} \qquad (5.23)$$

and

$$D_{kj} = \sum_{i=1}^{N} s_{ik}(y_{ij} - c_{kj})^2 \qquad (5.24)$$

where *k=1,...K*

WK-means algorithm inherits some of the insufficiencies of the K-means algorithm. These include the random initialisation of the centroids that do not guarantee an optimal solution as well as the random selection of the initial weights, that may not represent well the true significance of the features.

5.6.2 Intelligent Weighted K-means

One of the first approaches in extending WK-means was the integration of the anomalous cluster initialisation approach (Mirkin, 2005). De Amorim suggested that the

initialization step can be further extended to find initial weights for the features (De Amorim, 2011). The new algorithm is called intelligent Weighted K-means (iWK-means). The differences in the initialisation part of the iWK-means, compared to the IK-means algorithm, are:

- in the initialisation step the feature weights are set equal to $1/M$, ensuring that they sum up to unity.
- the distance measure that is used is the weighted distance metric (eq. 5.15)
- in the cluster update step there is an extra step of weight updating using eq. 5.13

After the initial clusters, centroids and weights have been found, the WK-means algorithm is ran.

5.6.3 Minkowski Weighted K-means

The WK-means algorithm was further extended by De Amorim and Mirkin in (De Amorim and Mirkin., 2012) by changing the distance measure from the squared Euclidean distance to the Minkowski β-metric.

$$d_\beta(y_i, c_k) = \sum_{j=1}^{M} w_j^\beta \left| y_{ij} - c_{kj} \right|^\beta \qquad (5.25)$$

In this way, the criterion of Equation 5.12 is modified to the Minkowski Weighted K-means (MWK-means) criterion:

$$\begin{aligned} W_\beta(S, C, w) &= \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{M} s_{ik}\, w_j^\beta \left| y_{ij} - c_{kj} \right|^\beta \\ &= \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{M} s_{ik} \left| w_j y_{ij} - w_j c_{kj} \right|^\beta \end{aligned} \qquad (5.26)$$

The difference between the criterions (5.12) and (5.19) is the change of the distance exponent from 2 to β, thus referring to the β-power of the Minkowski β-metric between the rescaled points $w_j y_{ij}$ and rescaled centroids $w_j c_{kj}$.

The differences of MWK-means algorithm compared to the WK-means algorithm are:

- in the initialisation step the value of the Minkowski exponent β is also selected, such as $\beta \geq 1$
- the distance measure used is the β-power of the Minkowski β-metric (eq. 5.18)
- the centroid that is calculated as the average of the cluster, it is called "the Minkowski centre". The new centroid is found by finding a real value c that minimizes the β-power of the Minkowski distance to the cluster's entities according to the equation:

$$d(c) = \sum_{i \in S_k} w_j^{\beta} \left| y_{ij} - c \right|^{\beta} \qquad (5.27)$$

  This problem is solved with the use of a steepest descent algorithm and it is described below.

- in the weight update step the feature weights are computed according to the equation:

$$w_j = \frac{1}{\sum_{u=1}^{h} \left[\frac{D_{j\beta}}{D_{u\beta}} \right]^{1/(\beta-1)}} \qquad (5.28)$$

where

$$D_{j\beta} = \sum_{k=1}^{K} \sum_{i=1}^{N} s_{ik} \left| y_{ij} - c_{kj} \right|^{\beta} \qquad (5.29)$$

and h the number of features where $D_j \neq 0$.

The algorithm proposed by De Amorim (De Amorim, 2011) in computing the Minkowski centre is illustrated in the following paragraphs. First, it is noted that each feature is a set of reals y_i, $i = 1,2, \dots, N_k$, where N_k is the number of entities in the cluster k. Also, the Minkowski centre is a convex function of c in the space of $\beta \geq 1$ and thus it can be minimised using a steepest-descent procedure. For $\beta = 1$, the minimum of equation 5.20 is known to be the median of y_i, and thus the minimisation method only needs to be applied in cases where $\beta > 1$.

The steepest-descent algorithm used to find the Minkowski's centre of a set $\{y_i\}$ of reals so that $y_1 \leq y_2 \dots \leq y_{N_k}$.

1. Initialisation: Initialise with $c_0 = y_{i^*}$, the minimiser of $d(c)$ on the set $\{y_i\}$ and a positive learning rate λ that can be taken, as approximately 10% of the difference $y_{N_\kappa} - y_1$.

2. Update c_1: Compute $c_0 - \lambda d'(c_0)$ and take it as c_1 if it falls within the interval $[y_{i'}, y_{i''}]$. Otherwise, decrease λ by approximately 10%, and repeat the step.

3. Stop condition: Test whether c_1 and c_0 coincide with a pre-specified threshold. If they do, halt the process and output c_1 as the optimal value of c. If not, move on to the next step.

4. Update c_0: Test whether $d(c_1) \leq d(c_0)$. If it is, set $c_0 = c_1$ and $d(c_0) = d(c_1)$ and go to step 2. If not, decrease λ a by approximately 10%, and go to step 2 without changing c_0.

Finally, the MWK-means can be further extended so that features are cluster specific as in (Chan et al., 2004; Huang et al, 2005; Huang et al, 2008). Similarly to the WK-means, w_j and $D_{j\beta}$ are changed to w_{kj} (Eq. 5.16) and $D_{kj\beta}$,

where

$$D_{kj\beta} = \sum_{i=1}^{N} s_{ik} \left| y_{ij} - c_{kj} \right|^{\beta} \qquad (5.30)$$

and $k=1,...K$

5.6.4 Intelligent Minkowski Weighted K-means

Intelligent Minkowski Weighted K-means (iMWK-means) is an algorithm developed by De Amorim and Mirkin. (De Amorim and Mirkin., 2012) resulting by the use of the iK-means in the Minkowski space. The iMWK-means algorithm uses the Intelligent K-means to find the anomalous clusters and the Minkowski distance to find the initial centroids and weights (De Amorim and Komisarczuk, 2012). The difference, compared to the iK-means, in the initialisation step is that the origin is now the Minkowski centre of the dataset.

6. Chapter: Dataset and Feature Extraction

6.1 Dataset

The dataset used in this thesis comes from the research of (Temmermans et al., 2013; Temmermans, 2014; Willekens et al., 2013) and it consists of 50 benign and 50 malignant biopsy samples. 2034 calcifications were extracted from the malignant samples and 1651 calcifications from the benign samples (Temmermans, 2014).

For the extraction of biopsy samples minimally invasive vacuum-assisted stereotactic breast biopsies were performed in the Radiology department of Brussels' university hospital (UZ Brussel) with the Mammotome Biopsy Stem (Ethicon Endo-Surgery (EES), Inc., Johnson & Johnson, Langhorne PA, Pennsylvania, USA) (Temmermans et al., 2013; Temmermans, 2014; Willekens et al., 2013; Papavasileiou et al., 2015).

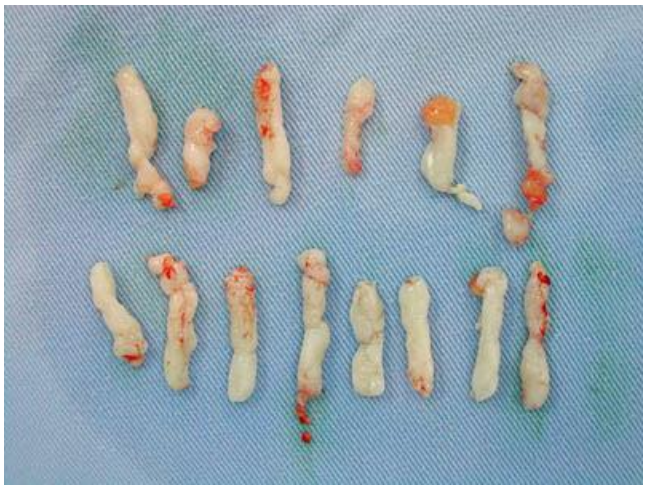


Figure 6.1: Example of extracted tissue samples from a mammotome core biopsy (HCMA,2014)

After the biopsy, the samples were radiographically scanned in order to confirm the extraction of some of the target microcalcifications (MCs) (Liberman et al., 1994). Subsequently, the calcifications were stored in blocks of paraffin and they were anatomopathologically investigated. During the anatomopathological investigation, the pathologist cuts slices from these blocks and makes a diagnosis based on the nature of the identified cells. Then the blocks were scanned using X-ray micro-computed tomography (micro-CT) (Temmermans, 2014). Micro-CT is a non-invasive high-resolution imaging method, which has been primarily used for bone

imaging studies and material analysis.

The 3D images that were generated (Temmermans et al., 2013; Temmermans, 2014; Willekens et al., 2013) had a voxel-width of 8.66 μm. Some examples of the created 3D renderings are shown in the figures below (6.2-6.4):

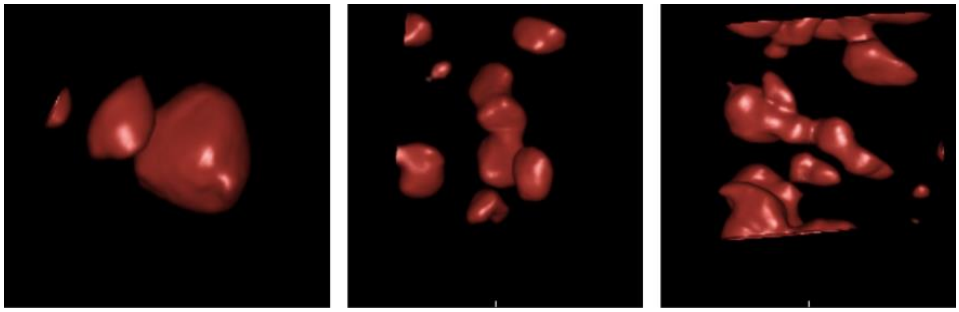


Figure 6.2: Typical sphere-like benign calcifications (Temmermans, 2014)

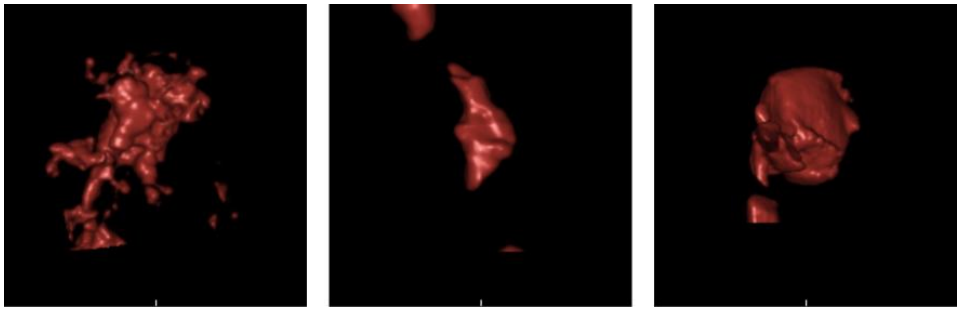


Figure 6.3: Typical roughly shaped malignant calcifications (Temmermans, 2014)

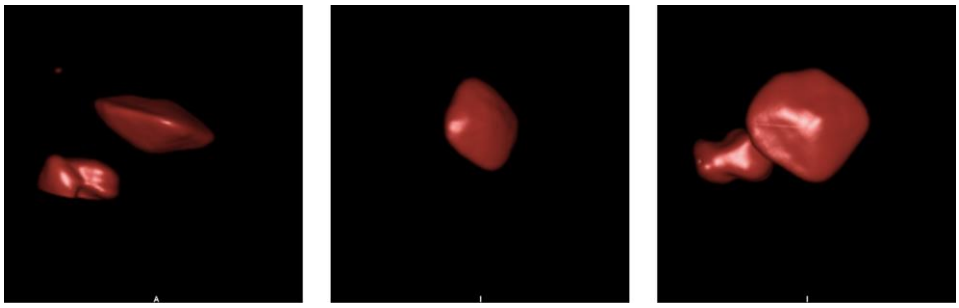


Figure 6.4: Pyramidal crystal shape calcifications (Temmermans, 2014)

6.2 Features

The features used in this thesis come from the work of (Temmermans et al., 2013; Temmermans, 2014; Willekens et al., 2013). The first step of feature extraction is the extraction of MCs from the background tissue. MCs can be extracted by using a thresholding operation since they are very distinguishable because of their higher attenuation compared to the background and the paraffin in which they are stored. Thresholding is followed by a connected component extraction using a 3D voxel connectivity of 6. Only the connected components with a size larger than 10 voxels are retained. In addition, an upper threshold is set corresponding with the size of a sphere with a diameter of 1mm which is the maximum size of a MC (Temmermans, 2014; Papavasileiou et al., 2015).

Assuming that the set of MCs is the following; $\{cc_1, cc_2, \dots cc_N\}$. Then, for each MC cc_i a volume of interest Ωcc_i is determined as the bounding cuboid with an additional margin of 30 voxels in each direction. Subsequently, the binary mask M_i and the gray-value sub image C_i are formulated as follows (Temmermans, 2014; Papavasileiou et al., 2015):

$$C_i(x) = I(x) \text{ with } x \in \Omega_{cc_i} \tag{6.1}$$

$$M_i(x) = \begin{cases} 1 \; if \; x \in \Omega_{cc_i} \\ 0 \; if \; x \in \Omega_{cc_i}/cc_i \end{cases} \tag{6.2}$$

The next step is the calculation of shape features on the MCs. Many of these features are based on the associated minimum volume-enclosing ellipsoid which is calculated using the Kachiyan algorithm (Khachiyan, 1996) and it is uniquely defined by its three principal axes$\{d_i^{(1)}, d_i^{(2)}, d_i^{(3)}\}$.

The feature set (Temmermans, 2014; Papavasileiou et al., 2015) incorporates the traditionally used shape features including volume, minimum and maximum diameter, compactness and elongation.

Volume

Volume of the calcification is defined as the total amount of voxels of the calcification:

$$F_i^{(1)} = \sum_{x \in \Omega_{cc_i}} \mathrm{M_i(x)} \tag{6.3}$$

Surface

Surface is defined as the number of perimeter voxels:

$$F_i^{(2)} = \sum_{x \in \Omega_{cc_i}} \mathrm{S_i(x)} \tag{6.4}$$

where S_i is the binary mask that contains the perimeter pixel of M_i with a 3D connectivity of six.

Surface over volume

$$F_i^{(3)} = \frac{F_i^{(2)}}{F_i^{(1)}} \tag{6.5}$$

Maximum Diameter

The maximum diameter of the minimum volume enclosing ellipsoid is calculated using the Kachiyan algorithm (Khachiyan, 1996):

$$F_i^{(4)} = max[d_i^{(1)}, d_i^{(2)}, d_i^{(3)}] \tag{6.6}$$

Minimum Diameter

The minimum diameter of the minimum volume enclosing ellipsoid calculated using the Kachiyan algorithm (Khachiyan, 1996):

$$F_i^{(5)} = min[d_i^{(1)}, d_i^{(2)}, d_i^{(3)}] \tag{6.7}$$

Compactness

This feature calculates how compact a microcalcification is. The definition by Murphy et al is adopted (Murphy et al., 2009).

$$F_i^{(6)} = \frac{F_i^{(1)}}{\left(F_i^{(4)}\right)^3} \tag{6.8}$$

Compactness

The feature compactness as defined by Kyongtae Bae et al (Bae et al., 2005):

$$F_i^{(7)} = \frac{F_i^{(1)}}{\prod_{j=1}^{3} d_i^{(j)}} \tag{6.9}$$

Elongation

The feature elongation as defined by Kyongtae Bae et al (Bae et al., 2005):

$$F_i^{(8)} = \frac{F_i^{(4)}}{F_i^{(5)}} \tag{6.10}$$

In addition to these 8 shape features, 12 novel features were introduced by (Temmermans et al., 2013; Temmermans, 2014; Willekens et al., 2013) that capture the specifics of the boundary zone of the calcification. Benign calcifications are expected to be clearly delineated, while malignant calcifications are expected to have more spiculated shapes. When a calcification is clearly delineated, there will be a clear border between calcification and background. However, when the calcification is not clearly delineated, there will be a more smooth transition from calcification to background. Therefore, some new features were proposed by (Temmermans et al., 2013; Temmermans, 2014; Willekens et al., 2013) that help to discriminate based upon this aspect. The basic idea is the comparison of the gray levels in the kernel region with the grey levels in the surrounding tissue.

For this purpose, for each MC cc_i a distance image D_i is created using the volume of interest Ω_{cc_i} as follows:

$$D_{i(x)} = \min_{y \in cc_i} [d(x,y)] \tag{6.11}$$

where $d(x,y)$ represents the Euclidean distance between the voxels x and y. Hence, the following boundary zones are obtained:

$$B_i^{(j)} = \{x | max[(j-1)*T, 1] \leq D_i(x) < j*T\} \tag{6.12}$$

where T represents the width of the boundary zone, and $j \in \{1,2,\dots,N\}$ is the index of the boundary zone. Both T and N were determined heuristically and set to 5 and 3 respectively. For generalization purposes, the zone corresponding to the

segmented kernel of MC cc_i shall be denoted by $B_i^{(0)}$. For each zone, $B_i^{(0)}$to $B_i^{(3)}$, the mean and standard deviation of the grey values were included as a feature. Both the mean and the standard deviation were expected to be higher in borders that contain spikes.

$$F_i^{(9)} = \frac{\sum_{x \in B_i^{(0)}} C_i(x)}{\sum_{x \in B_i^{(0)}} 1} = \mu_i^{(0)} \tag{6.13}$$

$$F_i^{(10)} = \sqrt{\frac{\sum_{x \in B_i^{(0)}} \left(C_i(x) - F_i^{(9)} \right)^2}{\sum_{x \in B_i^{(0)}} 1}} = \sigma_i^{(0)} \tag{6.14}$$

$$F_i^{(11)} = \frac{\sum_{x \in B_i^{(1)}} C_i(x)}{\sum_{x \in B_i^{(1)}} 1} = \mu_i^{(1)} \tag{6.15}$$

$$F_i^{(12)} = \sqrt{\frac{\sum_{x \in B_i^{(1)}} \left(C_i(x) - F_i^{(9)} \right)^2}{\sum_{x \in B_i^{(1)}} 1}} = \sigma_i^{(1)} \tag{6.16}$$

$$F_i^{(13)} = \frac{\sum_{x \in B_i^{(2)}} C_i(x)}{\sum_{x \in B_i^{(2)}} 1} = \mu_i^{(2)} \tag{6.17}$$

$$F_i^{(14)} = \sqrt{\frac{\sum_{x \in B_i^{(2)}} \left(C_i(x) - F_i^{(9)} \right)^2}{\sum_{x \in B_i^{(2)}} 1}} = \sigma_i^{(2)} \tag{6.18}$$

$$F_i^{(15)} = \frac{\sum_{x \in B_i^{(3)}} C_i(x)}{\sum_{x \in B_i^{(3)}} 1} = \mu_i^{(3)} \tag{6.18}$$

$$F_i^{(16)} = \sqrt{\frac{\sum_{x \in B_i^{(3)}} \left(C_i(x) - F_i^{(9)} \right)^2}{\sum_{x \in B_i^{(3)}} 1}} = \sigma_i^{(3)} \tag{6.19}$$

$$F_i^{(17)} = \frac{\mu_i^{(0)}}{\mu_i^{(1)}} = \frac{\frac{\sum_{x \in B_i^{(0)}} C_i(x)}{\sum_{x \in B_i^{(0)}} 1}}{\frac{\sum_{x \in B_i^{(1)}} C_i(x)}{\sum_{x \in B_i^{(1)}} 1}} \quad (6.20)$$

$$F_i^{(18)} = \frac{\mu_i^{(1)}}{\mu_i^{(2)}} = \frac{\frac{\sum_{x \in B_i^{(1)}} C_i(x)}{\sum_{x \in B_i^{(1)}} 1}}{\frac{\sum_{x \in B_i^{(2)}} C_i(x)}{\sum_{x \in B_i^{(2)}} 1}} \quad (6.21)$$

$$F_i^{(19)} = \frac{\mu_i^{(2)}}{\mu_i^{(3)}} = \frac{\frac{\sum_{x \in B_i^{(2)}} C_i(x)}{\sum_{x \in B_i^{(2)}} 1}}{\frac{\sum_{x \in B_i^{(3)}} C_i(x)}{\sum_{x \in B_i^{(3)}} 1}} \quad (6.22)$$

$$F_i^{(20)} = \frac{\mu_i^{(0)}}{\mu_i^{(3)}} = \frac{\frac{\sum_{x \in B_i^{(0)}} C_i(x)}{\sum_{x \in B_i^{(0)}} 1}}{\frac{\sum_{x \in B_i^{(3)}} C_i(x)}{\sum_{x \in B_i^{(3)}} 1}} \quad (6.23)$$

7. Chapter: Experimental Procedure

7.1 Ground Truth Issues

The label "benign" or "malignant" cannot be assigned to an individual microcalcification (MC). Instead, the label is assigned to a MC depending on whether it was found in a benign or a malignant sample, after anatomopathological examination (Temmermans, 2014). This means that no ground truth exists for the actual microcalcifications (MCs). As a consequence, benign MCs may exist in malignant samples, and incorrectly labelled as malignant. Since the blocks are not analysed completely, there is a low chance that malignant cells exist in the unanalysed part and therefore there is a low chance of false-negatives (Grimes et al., 2001).

In order to understand the ground truth issues that motivated in the devising of the algorithm described in this thesis, the following problem is illustrated. Two MCs are given as training examples to a classifier. One of these MCs is benign and was incorrectly classified as malignant and the other one is malignant. The MCs are characterized by shape features indicating benignity and malignancy, respectively. By introducing these training examples to a classifier, the latter learns that two different sets of shape features, with benign and malignant characteristics, correspond to the same class. As a result, there is degree of ambiguity when the classifier is asked to label a new instance (Papavasileiou et al., 2015).

Figures 7.1 and 7.2 show examples of benign MCs with malignant characteristics and malignant MCs with benign characteristics, respectively.

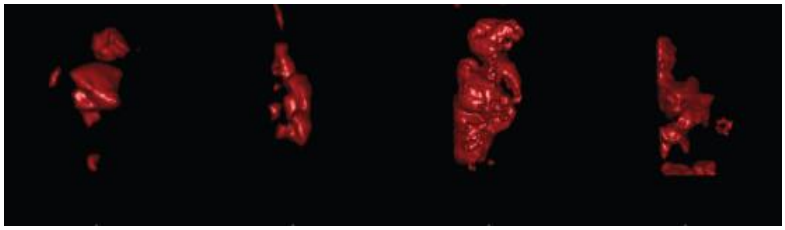


Figure 7.1: Benign MCs incorrectly classified as malignant (Temmermans, 2014)

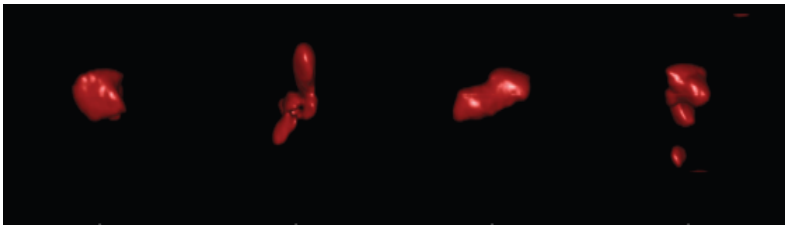


Figure 7.2: Malignant MCs incorrectly classified as benign (Temmermans, 2014)

7.2 Proposed Approach

The ground truth problem can be bypassed if the MCs are firstly grouped explicitly based on their shape. By clustering MCs according to their shape-features, similar shapes are grouped together in clusters. This way, the classifier is trained to classify MCs in the correct cluster, i.e. it learns to distinguish among different shapes of MCs, independently of their original class labels (Papavasileiou et al., 2015). So, the classifier classifies objects exclusively based on their shape features and without any knowledge of the class label that was assigned during the histopathological examination.

In order to do this an intermediate step of Clustering is introduced in the pipeline of the CAD system between the Feature Extraction and Classification steps. The CAD system's first steps of pre-processing, signal extraction and feature extraction remain the same as described in chapter 3.

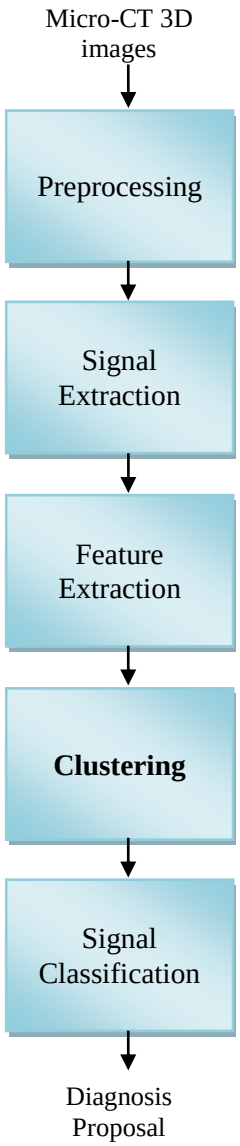


Figure 7.3: Proposed approach of the CAD system

7.3 Experiments

The dataset consists of 3685 MCs, each represented by the 20 features described in the 6th chapter. The performance of the proposed approach is investigated on three sets of features:

- the 8 shape features, $F_i^{(1)} - F_i^{(8)}$
- the 12 boundary zone features, $F_i^{(9)} - F_i^{(20)}$ and
- the combination of both groups of features, $F_i^{(1)} - F_i^{(20)}$.

For each of the aforementioned cases, the first step is to apply 10-fold cross

validation to the data. In general, in *k*-fold cross-validation, the original sample is randomly partitioned into *k* subsamples. One of these subsamples is used 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, known as *k* folds, where each of the *k* subsamples are used once as data for testing. So, in the case of 10 fold cross validation, 10 partitions of data are generated. Each partition consists of one set on which training is going to be applied (training set) and one set on which the model is going to be tested (testing set).

Next, the training set of each of the 10 partitions is grouped in clusters. The following algorithms are used as clustering algorithms:

- K-means
- Weighted K-means
- Intelligent Weighted K-means
- Minkowski Weighted K-means
- Intelligent Minkowski Weighted K-means

In cluster analysis one of the biggest difficulties is the definition of the number of clusters a priori. This cannot be done in an effective way since knowing the number of clusters before performing the clustering itself would require knowing the structure of the data, which is actually the goal of the cluster analysis. From the algorithms described in the 5th Chapter only the IWK-means and IMWK-means algorithms define a way to find the ideal number of clusters. For the other clustering algorithms, the number of generated clusters is varied from 2 to 10, thus resulting in 9 different cases of clustered data. Then, in order to decide which of these sets of clusters should be kept, a new measure is introduced that was named "entropy".

This "entropy" illustrates the distribution of the number of objects belonging to class "Benignity" and class "Malignancy" in each cluster. Let "a" be the number of objects belonging to class "Benignity" and "b" the number of objects belonging to class "Malignancy" in a cluster. Then the entropy is defined as:

$$Entropy = \frac{min(a,b)}{max(a,b)} \qquad (7.1)$$

Entropy can take values between 0 and 1. The maximum value of entropy ($max_{entropy}$=1) is achieved when a=b, i.e. when the numbers of objects belonging to

these two classes are equal. The minimum value of entropy ($min_{entropy}$=0) is achieved when a=0 or b=0, i.e. when all the objects in this cluster belong to class “M” or class “B” respectively. The following figure illustrates an approximation of the way the entropy is distributed.

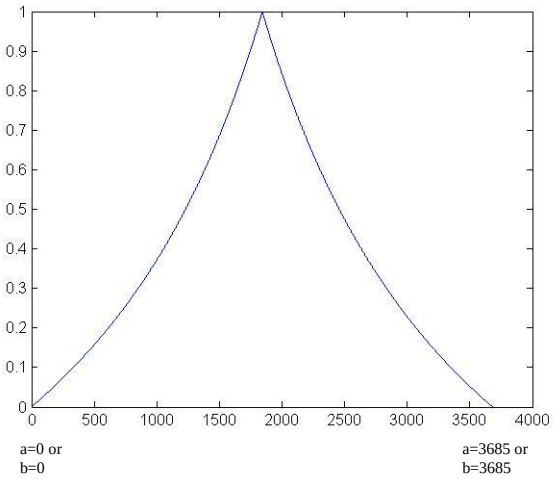


Figure 7.4: Entropy's distribution

Subsequently, for each of the 9 different cases of clustered data the mean entropy over its 10 folds is calculated. Finally, as ideal number of clusters is selected the clustering partition whose mean entropy is the lowest, because the lower the entropy of a cluster, the clearer the designation of a cluster as benign or malignant.

Each of the clusters is characterized as benign or malignant with a certain probability based on this entropy and thus each object of a cluster is assigned a corresponding label with the same probability.

Following in the classification step, the clustered objects are introduced to a classifier, either to an Artificial Neural Network or a Support Vector Machine. The classifier is trained based on the clustered training set to classify the objects to the correct cluster, i.e. to the correct shape. Next, the classifier is asked to predict the output cluster for each of the 10 testing sets.

After classification, an intermediate step of binary classification is performed. Each of the 3685 MCs have been classified to a cluster for which the label benign or malignant is known. Thus each MC is assigned a corresponding label. Then, this label is compared to the original ground truth and the percentage of "correctly" classified objects is calculated.

The final aim is the characterization of a sample as benign or malignant. For this purpose, the binary classification results for individual MCs are adopted and a threshold on the amount of objects classified as malignant is defined. Therefore, a sample is classified as malignant if the number of MCs that have been characterized as malignant exceeds a threshold. Otherwise, the sample is considered benign. By varying the threshold, a higher sensitivity can be achieved in return of a lower specificity and vice versa (Temmermans, 2014).

Another approach for the characterization of the sample is by weighting each MC with a weight linear to the probability of the cluster to which the MC belongs. Since each MC is assigned a cluster label and each cluster has a probability of being malignant, the MC's cluster probability can now be used as a weight when determining the sample's class (Papavasileiou et al., 2015). In this way, the MCs that belong to a cluster with a high probability of being benign or malignant (low entropy) will be assigned a higher weight compared to the MCs of clusters with lower probability. Thus, the MCs that belong to clusters which are clearer characterized as benign or malignant will contribute more in the characterization of the sample. Then, the threshold technique is used as described previously.

Finally, the results of the sample's classification are compared to the outcome of the anatomopathological examination. Thus, the accuracy, sensitivity and specificity of the correctly classified samples are calculated.

Accuracy is defined as the percentage of correctly classified objects and it indicates the ability of the system to make correct predictions on unknown data. The Accuracy is defined as follows:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP}$$

where TP: True Positives, FP: False Positives, TN: True Negatives, FN False Negatives.

Sensitivity or true positive rate is defined as the percentage of correctly classified malignant samples and it indicates the ability of the system to make correct diagnosis. Sensitivity is defined as follows:

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity is defined as the percentage of correctly classified benign samples and it indicates the ability of the system to correctly identify the benign cases. It is defined as follows:

$$Specificity \quad = \frac{TN}{TN \; + \; FP}$$

The approach described in this section introduces two novelties. Firstly, the issues that have emerged concerning each MC's ground truth are avoided. Secondly, there is a shift from binary classification to a probability-based classification.

.

8. Chapter: Results and Discussion

In this chapter the results of the experiments described in Chapter 7 are presented. At first, the binary classification results are presented and then the results of sample's classification. The results concern the experiments that were conducted with the application of the following combinations as far as the clustering algorithms and the classifiers are concerned:

Clustering Algorithm	Classifier
K-means (Cityblock distance)	Artificial Neural Network
K-means (Euclidean distance)	Artificial Neural Network
K-means (Cityblock distance)	Support Vector Machines
K-means (Euclidean distance)	Support Vector Machines
Minkowski Weighted K-means	Artificial Neural Networks
Minkowski Weighted K-means	Support Vector Machines

As discussed in Chapter 7, the performance of the new approach is investigated on 3 groups of features; the 8 shape features, the 12 boundary zone features and on the combination of both (20 features).

Section 8.1 presents the classification results at the microcalcification level, while section 8.2 presents the final results of the sample's classification.

8.1 Individual Classification Results

At the microcalcification level the label assigned to an individual MC is compared to the original ground truth and thus the accuracy of the binary classification is obtained. In the following tables (8.1-8.3), the mean and the standard deviation of the accuracy for different cases of "ideal" number of clusters are illustrated. It is reminded that the number of clusters is obtained according to the criterion of minimum entropy as this described in Chapter 7.

In Tables 8.1 and 8.2 the results of accuracy for clustering with K-means with two different distance measures; squared Euclidean distance and Cityblock distance, respectively, are presented.

Table 8-1: Results of accuracy for squared Euclidean distance using K-means as clustering algorithm, for features 1-8 (shape features), 9-20 (boundary zone features) and for the ensemble of all features.

	7		8		9		10	
	ANN	SVM	ANN	SVM	ANN	SVM	ANN	SVM
1-8	64.04 3.10	63.80 2.85	65.83 3.47	65.37 3.22	64.88 3.16	64.69 3.43	65.81 2.42	65.86 2.93
9-20	59.73 2.22	59.38 1.55	60.32 4.01	60.41 4.02	60.60 1.48	59.62 1.68		
ALL	64.02 2.30	62.82 2.47	64.64 2.27	62.47 2.11	64.15 2.77	63.07 2.59	65.00 2.88	63.58 1.89

Table 8-2: Mean and standard deviation results of accuracy using K-means algorithm and Cityblock distance, for features 1-8 (shape features), 9-20 (boundary zone features) and for the ensemble of all features

	7		8		9		10	
	ANN	SVM	ANN	SVM	ANN	SVM	ANN	SVM
1-8			66.21 2.03	65.86 2.05	67.14 3.11	67.25 2.62	67.38 2.34	67.41 2.39
9-20	58.62 3.30	58.72 3.51	59.35 2.40	59.78 3.14	59.73 2.78	60.24 3.06	60.65 2.56	60.19 3.09
ALL	64.10 1.60	62.52 2.06	63.80 2.01	62.52 1.75	64.53 2.20	63.20 1.94		

In Tables 8.3-8.5, the results of accuracy for clustering with MWK-means and optimal values of β are presented. In order to find the optimal value of β in the MWK-means, the values of β are varied such that $\beta \in [1.0,5]$ with steps of 0.1. Only the cases which resulted in higher accuracy are shown.

Table 8-3: Mean and standard deviation results of accuracy using MWK-means algorithm on the first set of shape features (Features 1-8)

	7		8		9		10	
β	ANN	SVM	ANN	SVM	ANN	SVM	ANN	SVM
1,7	67,19 2,08	67,30 2,53	66,51 5,00	66,43 4,14	67,90 2,26	68,79 2,51		
1,9			67,35 2,24	67,87 2,09	68,03 2,17	68,01 2,41	68,03 1,96	67,98 1,97
2,3	67,46 2,62	67,79 2,18	67,35 2,60	67,11 2,05	67,65 2,28	65,65 3,62		
2,7	67,08 2,26	67,11 2,58	67,55 3,28	67,44 3,03	67,49 2,63	67,63 2,19	65,81 4,59	67,03 3,59
2,8	67,52 1,58	67,60 1,30	67,49 1,88	67,35 2,41	67,19 1,99	67,49 2,09	64,83 5,88	66,78 1,86
3,1	66,38 2,93	66,13 2,49	66,32 3,95	67,08 3,34	66,14 1,83	66,43 2,19	67,38 2,21	66,83 2,23

Table 8-4: Mean and standard deviation results of accuracy using MWK-means algorithm on the second set of boundary zone features (Features 9-20)

	7		8		9		10	
β	ANN	SVM	ANN	SVM	ANN	SVM	ANN	SVM
1,9			58,7 2,59	58,64 2,92	58,56 2,61	57,99 3,27	58,91 1,79	59,32 2,47
2	57,70 3,71	59,97 4,20	58,86 2,72	57,88 3,40	58,86 2,58	58,84 2,74	58,37 2,79	59,03 3,07
4,7	59,59 1,91	59,97 2,03	60,93 3,89	59,38 3,58	61,06 2,69	60,98 2,24	59,92 2,84	59,81 2,69
4,8	58,57 3,06	57,85 2,95	60,19 2,98	60,24 3,15	61,14 2,50	61,41 1,90		
4,9	60,32 2,92	60,08 3,03	59,62 2,10	59,65 1,47	60,35 2,46	60,70 2,42	61,19 2,59	61,30 3,73

Table 8-5: Mean and standard deviation results of accuracy using MWK-means algorithm on both sets of features (Features 1-20)

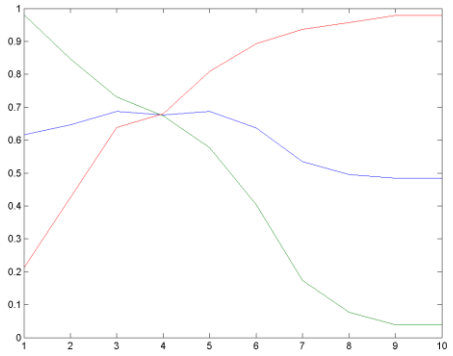
	7		8		9		10	
β	ANN	SVM	ANN	SVM	ANN	SVM	ANN	SVM
1,7	66,05 2,73	64,99 3,14	66,38 2,99	65,37 3,36	66,76 2,74	66,16 3,49	64,19 4,12	64,64 4,57
1,9			66,47 2,05	64,80 1,55	65,54 1,50	64,50 2,11	66,87 2,09	65,35 1,57
2,1	64,48 2,48	63,53 1,95	64,78 2,21	63,15 2,64	66,32 2,57	65,08 2,47	65,51 2,51	63,20 2,70

8.2 Sample's Classification Results

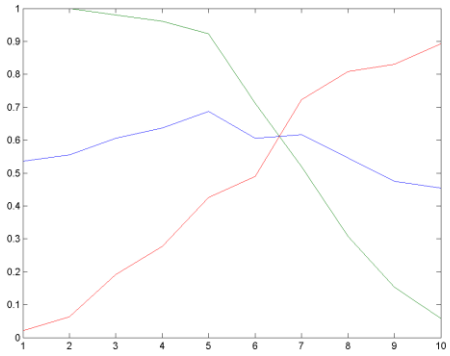
In this section the best results of the sample's classification are presented. As discussed in chapter 7, two approaches were applied on the sample's classification; the one where the MC's cluster probability is used as a weight and the other where this weight is not taken into account.

Table 8-6: Results of sample's classification accuracy without taking into account the cluster's probability, for features 1-8 (shape features), 9-20 (boundary zone features) and for the ensemble of all features

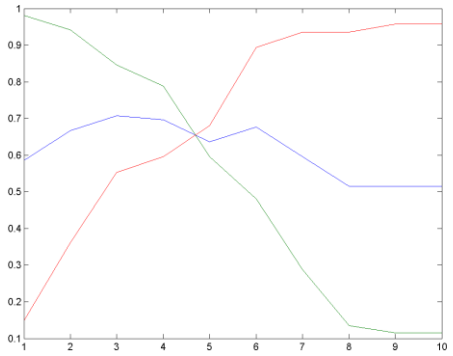
	Kmeans-sqEuclidean			Kmeans- Cityblock			MWKmeans		
	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.
1-8	61.6	98.1	21.3	71.7	80.8	61.7	62.6	100	21.3
9-20	68.7	92.3	42.6	60.6	94.2	23.4	64.6	94.2	31.9
ALL	70.7	84.6	55.3	60.6	76.9	42.6	65.7	90.4	38.3



(a)

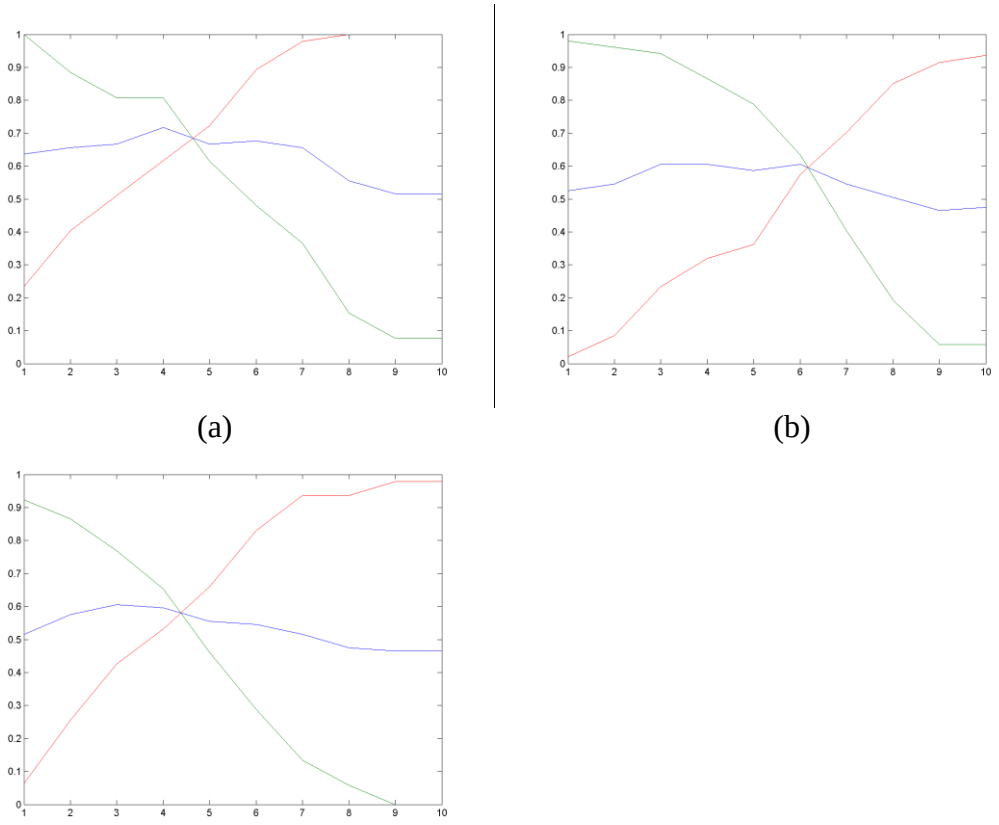


(b)



(c)

Figure 8.1: Accuracy (blue line), sensitivity (green line) and specificity (red line) for an increasing threshold (from 10 to 100 percent), for K-means with squared Euclidean distance and (a) shape features, (b) boundary zone features, (c) the ensemble of all features (non-weighted sample classification)



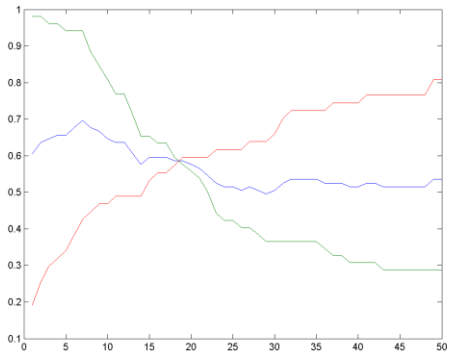
(a) (b)

(c)

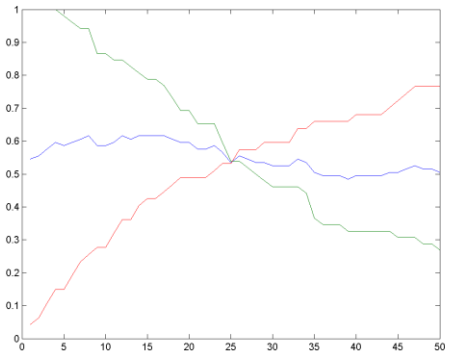
Figure 8.2: Accuracy (blue line), sensitivity (green line) and specificity (red line) for an increasing threshold (from 10 to 100 percent), for K-means with Cityblock distance and (a) shape features, (b) boundary zone features, (c) the ensemble of all features (non-weighted sample classification)

Table 8-7: Results of sample's classification accuracy taking into account the cluster's probability, for features 1-8 (shape features), 9-20 (boundary zone features) and for the ensemble of all features

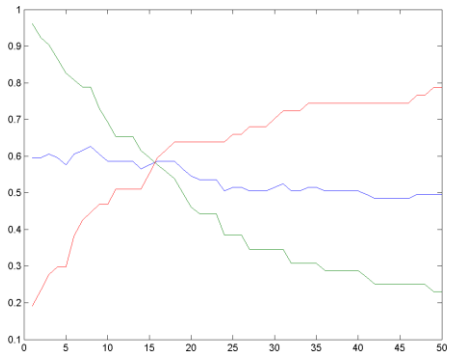
	Kmeans-sqEuclidean			Kmeans- Cityblock			MWKmeans		
	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.
1-8	64.6	98.1	27.7	72.7	100	42.6	64.6	100	25.5
9-20	61.6	94.2	25.5	65.7	80.8	48.9	66.7	96.2	34.0
ALL	69.7	94.2	42.6	63.6	82.7	42.6	63.6	96.2	27.7



(a)

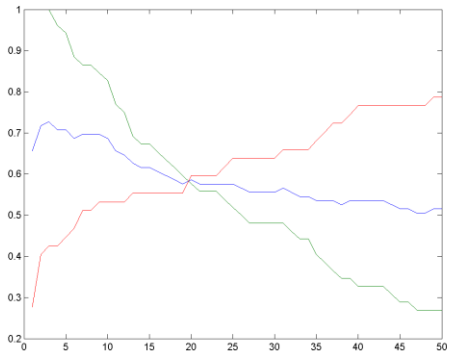


(b)

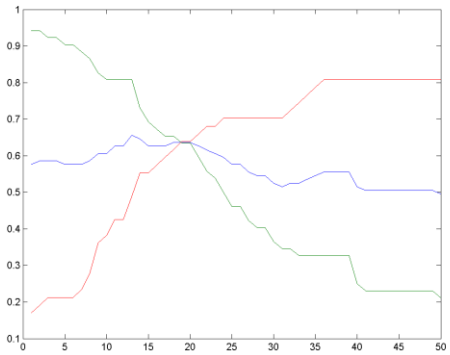


(c)

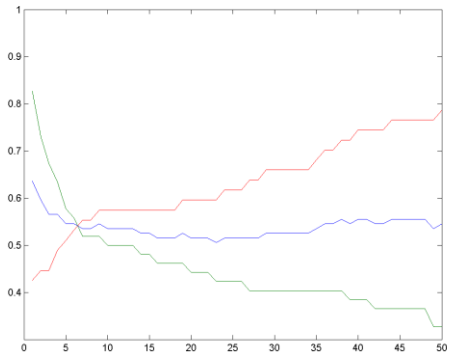
Figure 8.3: Accuracy (blue line), sensitivity (green line) and specificity (red line) for an increasing threshold (from 5 to 100 percent), for K-means with squared Euclidean distance and (a) shape features, (b) boundary zone features, (c) the ensemble of all features (weighted sample classification)



(a)



(b)

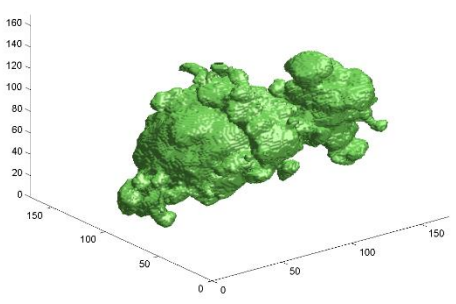


(c)

Figure 8.4: Accuracy (blue line), sensitivity (green line) and specificity (red line) for an increasing threshold (from 5 to 100 percent), for K-means with Cityblock distance and (a) shape features, (b) boundary zone features, (c) the ensemble of all features (weighted sample classification)

8.3 Examples of clustered microcalcifications

In this section, some examples of clustered MCs in the generated clusters are illustrated.



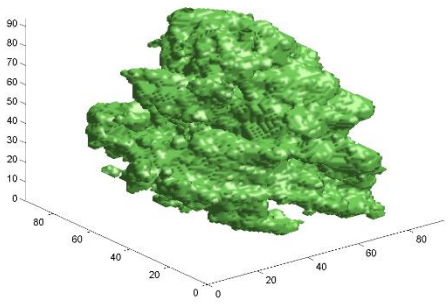


Figure 8.5: Examples of clustered MCs, characterized as malignant and benign during histopathological examination

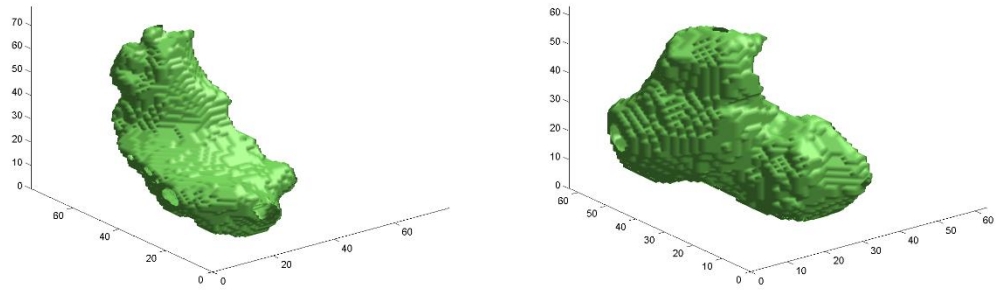


Figure 8.6: Examples of clustered microcalcifications, both characterized as malignant during histopathological examination

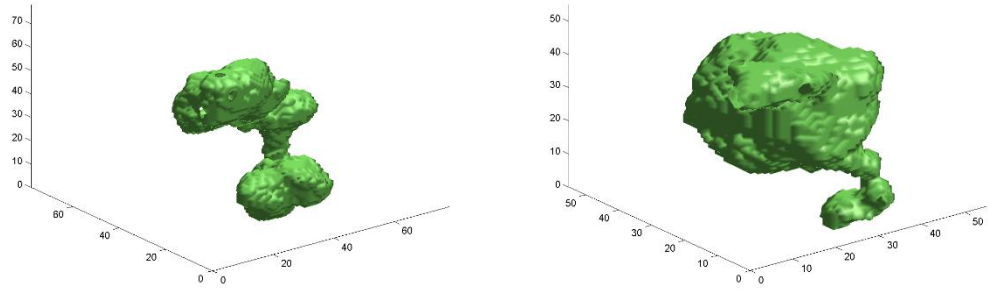


Figure 8.7: Examples of clustered microcalcifications, both characterized as benign during histopathological examination

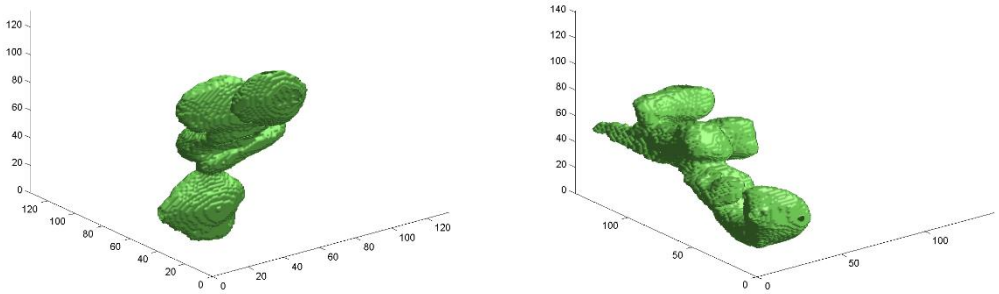


Figure 8.8: Examples of clustered microcalcifications, characterized as benign and malignant during histopathological examination

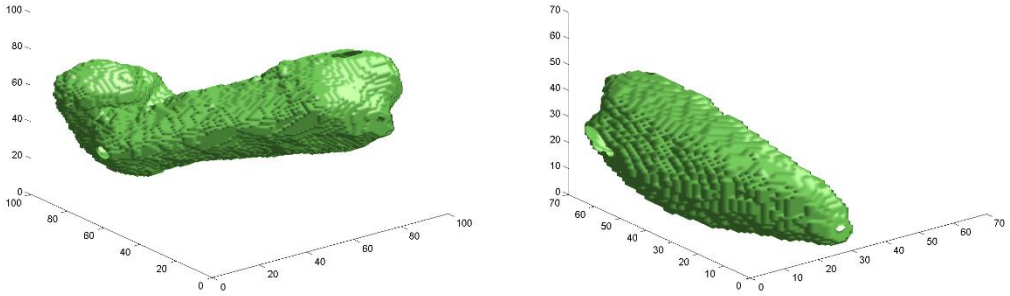


Figure 8.9: Examples of clustered microcalcifications both characterized as malignant during histopathological examination

8.4 Discussion

From Tables 8.1-8.5, it becomes obvious that ANNs and SVMs show very similar performance (Papavasileiou et al., 2015), despite the fact that ANNs need more processing time and more processing resources than SVMs. In addition, the legacy features (shape features 1-8) outperform the boundary zone features with accuracies ranging from 63.80% to 68.79% and from 57.70% to 61.41%, respectively. The shape features show similar performance to the ensemble of all features with accuracies ranging from 62.47% to 66.87%. These are opposed to (Temmermans et al, 2013; Temmermans, 2014) where the same problem of classifying MCs was explored but without the pre-clustering step. In that approach the accuracy results for the shape features, the boundary zone features and the ensemble of all features were: from 64.3% to 68.7%, from 60.5% to 73.2% and from 65.4% to 73.8% respectively. The latter reveals that the previous approach (Temmermans et al, 2013; Temmermans, 2014) without the pre-clustering phase showed slightly higher overall accuracy at the calcification level (Papavasileiou et al., 2015).

However, it should be mentioned that these results should be interpreted carefully because of the ground truth issues. Moreover, the binary classification consists an intermediate step of calculating the final sample classification results in this thesis approach.

At the sample level (Tables 8.6-8.7), the ultimate goal is the optimization of sensitivity, with respect to high levels of specificity and accuracy. This is because the percentage of correctly classified malignant samples is more important than the

accuracy itself. Both in the weighted and the non-weighted approach the shape features outperform the boundary zone features and the ensemble of features. In more detail, the resulted percentages for the non-weighted approach are; from 80.8% to 100%, from 92.3% to 94.2% and from 76.9% to 90.4%. For the weighted approach the corresponding percentages are; from 98.1% to 100%, from 80.8% to 96.2% and from 82.7% to 96.2%. From the latter it is apparent that the shape features outperform the other feature categories.

The best result can be reached with the weighted approach, with K-means using the 8 traditionally used shape features, the Cityblock distance and 9 clusters. This approach delivers a sensitivity of 100%, a specificity of 42,6% and an accuracy of 72,7% for a threshold of 15% malignant MCs per sample. This is respectively an improvement of 2%, 2,6% and 2,7% compared to the state of the art (Temmermans et al, 2013; Temmermans, 2014).

From the figures 8.5-8.9 it becomes obvious that by clustering MCs according to their shape features, similar shapes are grouped together in clusters. Especially, from the figures 8.1 and 8.8 the ground truth issues are illustrated because even though these MCs share common shape features, they were not classified as the same (as malignant or as benign) during the histopathological examination.

9. Chapter: Conclusions

Breast cancer is one of the most common types of cancers and a leading cause of death worldwide. Its diagnosis in an early stage may contribute to its effective treatment. For this purpose, many imaging modalities have been developed in order to find suspicious lesions and microcalcifications (MCs), i.e. tiny spots of calcium that may indicate malignancy. Mammography is the current standard procedure for imaging the breast tissue. However, the characterization of MCs as benign or malignant based on their appearance is a difficult task because of the superposition of breast tissue and thus the alteration of their real appearance. In case of suspiciousness, a biopsy is conducted and the extracted tissue is pathologically analysed for the presence of cancer cells. However, the MCs themselves are mostly not analysed and therefore ground truth exists for the tissue samples but not for the individual MCs.

The current thesis presented a novel classification approach for MCs extracted from core biopsy tissue samples digitized using micro-CT, a high-resolution 3D imaging modality. The aim of this master thesis was the investigation of whether the introduction of a clustering step before classification could improve the sample's classification results.

By clustering MCs according to their shape-features, similar shapes are grouped together in clusters. This way, the classifier is trained to classify MCs in the correct cluster, i.e. it learns to distinguish among different shapes of MCs, independently of their original class labels.

K-means and Minkowski Weighted K-means (MWK-means) were selected as clustering algorithms while Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) as classification algorithms. In this framework, MCs that share common shape features are grouped together in clusters. Thus, clusters of MCs with common shape characteristics are formed. In addition, each cluster is characterized as benign or malignant with a certain probability. After clustering and classification, an intermediate step of binary classification is performed which will then be used for the classification of the sample, by varying a threshold of the number of malignant MCs that are required in order to characterize the sample as malignant.

The results at the classification level revealed that the two clustering

algorithms and the two classification algorithms statistically show similar performance. As far as the processing time and resources are concerned though, ANNs are far more demanding. In addition, the traditionally used shape features outperform the boundary zone features and show similar performance to the ensemble of all features. It should be mentioned that these results should be interpreted carefully because of the ground truth issues at the calcification level.

However, the binary classification consists an intermediate step of calculating the final sample classification results. At the sample level the ultimate goal is the optimization of sensitivity. For the characterization of the sample two approaches were proposed; a weighted and a non-weighted. Both in the weighted and the non-weighted approach the shape features outperform the boundary zone features and the ensemble of all features. The best result is reached with the weighted approach which delivers a sensitivity of 100%, a specificity of 42,6% and an accuracy of 72,7% for a threshold of 15% malignant MCs per sample. This is an improvement of 2%, 2,6% and 2,7% compared to the state of the art.

The achievement of 100% of sensitivity allows to avoid the case of missing any malignant samples. This improvement is the result of avoiding the bias introduced by the fact that the ground truth is distorted, i.e. the fact that malignant samples may contain unsuspicious calcifications which are considered malignant during the learning process. With this approach some unnecessary anatomopathological investigations might be avoided. Furthermore, if the future 3D high-resolution imaging would be applicable in vivo, the number of unnecessary biopsies could be decreased, thus reducing any unnecessary expenses and physical and mental discomfort for the patient.

The approach of this thesis introduces two novelties; the ground truth issues concerning each MC's class are bypassed and there is a shift from binary classification to a probability-weighted classification. The vast amount of clustering algorithms and classifiers as well as the use of various approaches for incorporating the cluster probabilities in the classification of the sample hold the potential to further improve these results.

References

Abe, S. (2010). *Support vector machines for pattern classification.* Springer.

ACR, American College of Radiology (2003). Breast Imaging Reporting and Data System Atlas (BI-RADS®Atlas) Reston, VA: American College of Radiology

ACR, American College of Radiology (2014). Available Online; http://www.acr.org/

ACS. American Cancer Society (2014). Available Online; *www.cancer.org/treatment/understandingyourdiagnosis/examsandtestdescriptions/forwomenfacingabreastbiopsy/breast-biopsy-biopsy-types*

Andreadis, I. I., Spyrou, G. M., & Nikita, K. S. (2011). A comparative study of image features for classification of breast microcalcifications. *Measurement Science and Technology*, *22*(11), 114005.

Bae K.T., Kim J.S., Na Y.H., Kim J.H., and Jin-Hwan K. (2005) Pulmonary nodules: Automated detection on ct images with morphologic matching algorithm - preliminary results1. Radiology, 236(1):286–293

Ball, G. and Hall, D. (1967) A clustering technique for summarizing multivariate data, *Behav. Sci*., vol. 12, pp. 153–155.

Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT press.

Becker, S. (1991). Unsupervised learning procedures for neural networks. *International Journal of Neural Systems*, 2(01n02), 17-33

Berkhin, P. (2006) *A survey of clustering datamining techniques*. In Jacob Kogan, Charles Nicholas, and Marc Teboulle (eds.), Grouping Multidimensional Data: Recent Advances in Clustering, pp. 25–71. Springer. 372, 520.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 1, p. 740). New York: springer.

Breast Cancer (2014). Available Online: http://www.borstkanker.nl

Cady, B., & Michaelson, J. S. (2001). The life-sparing potential of mammographic screening. *Cancer*, *91*(9), 1699-1703.

Callan, R. (2003) *Artificial Intelligence*. Palgrave Macmillan, NY, USA

Chan, E. Y., Ching, W. K., Ng, M. K., & Huang, J. Z. (2004). An optimization algorithm for clustering using weighted dissimilarity measures. *Pattern recognition*, *37*(5), 943-952.

Chiang, M. M. T., & Mirkin, B. (2010). Intelligent choice of the number of clusters in K-Means clustering: an experimental study with different cluster spreads. *Journal of classification*, *27*(1), 3-40.

De Amorim, R. C, & Mirkin, B. (2012). Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering. *Pattern Recognition*, 45(3), 1061-1075.

De Amorim, R. C. (2011). Learning feature weights for K-Means clustering using the Minkowski metric, Doctoral Dissertation, Birkbeck, University of London

De Amorim, R. C., & Komisarczuk, P. (2012). On initializations for the minkowski weighted k-means. In *Advances in Intelligent Data Analysis XI* (pp. 45-55). Springer Berlin Heidelberg

De Wilde V. (1999). Mammaire microcalcificaties: significant van den aantal criteria aan de hand van een patientenstudie. PhD thesis, Vrije Universiteit Brussel

Duda, R. O. and Hart, P. E. (1973) *Pattern Classification and Scene Analysis*, New York: J. Wiley & Sons

Duncan, J. S., & Ayache, N. (2000). Medical image analysis: Progress over two decades and the challenges ahead. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22*(1), 85-106.

Elter, M., & Horsch, A. (2009). CADx of mammographic masses and clustered microcalcifications: a review. *Medical physics*, *36*(6), 2052-2068.

Eucan (2014), Available Online: eco.iarc.fr/EUCAN/.

Gershon-Cohen J, Yiu LS, Berger SM (1962): The diagnostic importance of calcareous patterns in roentgenography of breast cancer. AJR 1962, 88:1117–1125.]

Grimes M.M., Karageorge L.S., Hogge J.P. (2001). Does exhaustive search for microcalcifications improve diagnostic yield in stereotactic core needle breast biopsies? *Modern pathology*, 14(4), 350-353.

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6), 610-621.

Hartigan, J. A. (1975) Clustering Algorithms, John Wiley and Sons, Inc., New York, NY. USA

Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall PTR.

HCMA. The Hong Cong Medical Association (2014) Available Online: www.hkma.org

Huang, J. Z., Ng, M. K., Rong, H., & Li, Z. (2005). Automated variable weighting in k-means type clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *27*(5), 657-668

Huang, J. Z., Xu, J., Ng, M. And Ye, Y. (2008) Weighting Method for Feature Selection in K-Means. In: H. Liu and H. Motoda (Ed.) *Computational Methods of Feature Selection*, Chapman & Hall/CRC, pp 193-209.

Hyoung-Koo Lee: '*Technology of digital radiography*', Department of Biomentical Enginnering, College of Medicine, The Catholic University of Korea.

Karahaliou Anna (2009), *Texture Analysis of Mammographic Images for Computer-Aided Breast Cancer Diagnosis*, PhD dissertation, University of Patras.

Karahaliou, A. N., Arikidis, N. S., Skiadopoulos, S. G., Panayiotakis, G. S., & Costaridou, L. I (2012). Computerized Image Analysis of Mammographic Microcalcifications: Diagnosis and Prognosis

Karahaliou, A., Skiadopoulos, S., Boniatis, I., Sakellaropoulos, P., Likaki, E., Panayiotakis, G., & Costaridou, L. (2007). Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis. *The British Journal of Radiology*, Vol. 80, No. 956, pp. 648-656, ISSN 0007-1285

Karkinos (2014) Available Online: www.karkinos24.gr/index.php/karkinostoumastou.

Khachiyan L.G.(1996). Rounding of polytopes in the real number model of computation Mathematics of Operations Research, 21(2):307–320

Kopans, D. B. (1992). The positive predictive value of mammography. *AJR. American journal of roentgenology*, *158*(3), 521-526.

Kotsiantis, S. B. (2007). Supervised machine learning: a review of classification techniques. *Informatica (03505596), 31(3).*

Lanyi, M. (1983). "Pattern analysis of 5641 microcalcifications in 100 mammary duct carcinomas: polymorphism." *RoFo: Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin* 139.3: 240-248.

Liberman, L., Evans 3rd, W. P., Dershaw, D. D., Hann, L. E., Deutch, B. M., Abramson, A. F., & Rosen, P. P. (1994). Radiography of microcalcifications in stereotaxic mammary core biopsy specimens. *Radiology*, *190*(1), 223-225.

Lieske, B., Ravichandran, D., & Wright, D. (2006). Role of fine-needle aspiration cytology and core biopsy in the preoperative diagnosis of screen-detected breast carcinoma. *British journal of cancer*, *95*(1), 62-66.

Lo, J. Y., Gavrielides, M. A., Markey, M. K., & Jesneck, J. L. (2003, May). Computer-aided classification of breast microcalcification clusters: Merging of features from image processing and radiologists. In *Medical Imaging 2003* (pp. 882-889). International Society for Optics and Photonics.

MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (Vol. 1, p. 6). Cambridge: Cambridge university press.

Mendel, J. M. & Fu, K. S., (1970), *Adaptive, Learning and Pattern Recognition Systems: Theory and Applications*, Academic Press, New York.

Mirkin, B. (2005) *Clustering for Data Mining: A Data Discovery Approach*, Boca Raton Fl., Chapman and Hall/CRC

Murphy K., B. van Ginneken, A.M.R. Schilham, B.J. de Hoop, H.A. Gietema, and M. Prokop (2009). A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest neighbour classification. Medical Image Analysis, 13(5):757{770, 2009. Includes Special Section on the 12th International Conference on Medical Imaging and Computer Assisted Intervention.

Murty, M. N. and Devi, V. S. (2011). *Pattern Recognition, An Algorithmic Approach*, Springer, London DOI 10.1007/978-0-85729-495-1

Nakayama, R., Watanabe, R., Namba, K., Takeda, K., Yamamoto, K., Katsuragawa, S., & Doi, K. (2007). An Improved Computer-aided Diagnosis Scheme Using the Nearest Neighbor Criterion for Determining Histological Classification of Clustered Microcalcifications. *Methods Inf Med*, *46*, 716-722.

NHS Cancer Screening Programmes, (2001) "Guidelines for non-operative diagnostic procedures and reporting in breast cancer screening, Prepress Projects, England. Available Online: www.cancerscreening.nhs.uk/breastscreen/ publications/nhsbsp50.pdf

Papavasileiou, E., Temmermans, F., Jansen, B., Willekens, I., Van de Casteele, E., De Mey, J., Deklerck R. & Hostens, J. (2015, January). Shape-Based Clustering and Classification of Breast Microcalcifications in Micro-CT Images. In *6th European Conference of the International Federation for Medical and Biological Engineering* (pp. 160-163). Springer International Publishing.

Perry, N., Broeders, M., De Wolf, C., Törnberg, S., Holland, R., & Von Karsa, L. (2008). European guidelines for quality assurance in breast cancer screening and diagnosis. — . *Annals of Oncology*, *19*(4), 614-622.

Salomon, Albert. "Beitrage zur pathologie und klinik der mammacarcinome."*Arch Klin Chir* 101 (1913): 573-668.

Sickles, E. A. (1984). Mammographic features of" early" breast cancer.*American Journal of Roentgenology*, *143*(3), 461-464.

Stewart, Bernard W et al (2003). *World cancer report*. Iarc,.

Suri J.S, Rangayyan R.M (2006), Recent Advances in Breast Imaging, *Mammography, and Computer-Aided Diagnosis of Breast Cancer*. SPIE Publications

Tabár, L., & Dean, P. B. (2011). *Teaching atlas of mammography*. Thieme.

Temmermans Frederik, *Visual search in mobile and medical applications: Feature extraction and classification, interoperable image search and human-machine interaction*, PhD dissertation, 2014.

Temmermans, F., Jansen, B., Willekens, I., Van de Casteele, E., Deklerck, R., Schelkens, P., & De Mey, J. (2013, September). Classification of microcalcifications using micro-CT. In SPIE *Optical Engineering+ Applications* (pp. 88561B-88561B). International Society for Optics and Photonics

Theodoridis, S, Koutroumbas, K. (2009) *Pattern Recognition*, Academic Press, Elsevier

Vapnik, V. (2000). *The nature of statistical learning theory*. Springer.

Veropoulos, K., Campbell, C., & Cristianini, N. (1999, May). Controlling the sensitivity of support vector machines. In *Proceedings of the international joint conference on artificial intelligence* (Vol. 1999, pp. 55-60).

Werbos P.J. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, Doctoral Dissertation, Applied Mathematics, Harvard University, Boston, MA,

Widrow B & and Hoff (1960). *Adaptive switching circuits*, IRE WESCON Convention Record, pages 96-104, IREm

Willekens, I., Van de Casteele, E., Buls, N., Temmermans, F., Jansen, B., Deklerck, R., & de Mey, J. (2014). High-resolution 3D micro-CT imaging of breast microcalcifications: a preliminary analysis. *BMC cancer*, *14*(1), 9

Zhang, G. P. (2000). Neural networks for classification: a survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 30(4), 451-462.

Δερματάς, Β. (1997). *Αναγνώριση Προτύπων Ι*, Πανεπιστημιακές Εκδόσεις, Πανεπιστήμιο Πατρών—Dermatas, V. (1997) Pattern Recognition I, University Press, University of Patras

Δημητρόπουλος Ν., Κεραμόπουλος Α., *«Η μαστογραφία στην κλινική πράξη» (2000)*, Εκδόσεις Παρισιάνος

Διαμαντάρας, Κ.(2007), *Τεχνητά Νευρωνικά Δίκτυα*, Εκδόσεις Κλειδάριθμος,. Diamantaras K (2007), *Artificial Neural Networks*, Kleidarithmos press.

Κανδαράκης Ι.: «*Φυσικές και τεχνολογικές αρχές ακτινοδιαγνωστικής*» (2004), Εκδόσεις ΕΛΛΗΝ, Τρίτη έκδοση

Λάππας Σπ. Ηλίας, Μη Ψηλαφητές Αλλοιώσεις του μαστού: Χειρουργικές επιλογές - Επικουρική Θεραπεία. Online. Available: http://www.ilias-lappas.gr/article004/index.html Last accessed March 2014.

Φύσσας. Γ. Π., *«Ο μαστός και οι παθήσεις του»* (2006) Αθήνα, εκδοτικός οίκος Α.Α. Λιβάνη