Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακό Πρόγραμμα Σπουδών

Εφαρμοσμένη Πληροφορική της Υγείας και Τηλεϊατρική

Μεταπτυχιακή Διατριβή

Αυτόματη Επισημείωση Ιατρικής Εικόνας

Αναστάσιος Γιαννουλάκης

Επιβλέπων Καθηγητής

Ζήνωνας Θεοδοσίου

Δεκέμβριος 2019

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακό Πρόγραμμα Σπουδών

Εφαρμοσμένη Πληροφορική της Υγείας και Τηλεϊατρική

Μεταπτυχιακή Διατριβή

Αυτόματη Επισημείωση Ιατρικής Εικόνας

Αναστάσιος Γιαννουλάκης

Επιβλέπων Καθηγητής

Ζήνωνας Θεοδοσίου

Η παρούσα πτυχιακή εργασία υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων για απόκτηση μεταπτυχιακού τίτλου σπουδών στην Εφαρμοσμένη Πληροφορική της Υγείας και Τηλεϊατρική από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών του Ανοικτού Πανεπιστημίου Κύπρου.

Δεκέμβριος 2019

Περίληψη

Η ιατρική απεικόνιση αποτελεί βασικό στοιχείο της σύγχρονης υγειονομικής περίθαλψης. Λόγω του μεγάλου αριθμού ιατρικών εικόνων, υπάρχει επείγουσα ανάγκη για έναν αποδοτικό μηχανισμό που μπορεί να ταξινομεί και να αναζητά ιατρικές εικόνες σε υψηλό σημασιολογικό επίπεδο. Η αυτόματη επισημείωση των ιατρικών εικόνων αποτελεί βασική προϋπόθεση για τη δημιουργία σημασιολογικών αρχείων που μπορούν να χρησιμοποιηθούν για την ενίσχυση της τεκμηριωμένης διάγνωσης, της εκπαίδευσης των ιατρών και της βιοϊατρικής έρευνας. Θεωρείται, επίσης, βασική προϋπόθεση για την ανάπτυξη μηχανών αποτελεσματικής αναζήτησης και ανάκτησης ιατρικών εικόνων, εκθέσεων και συναφών δημοσιεύσεων. Το αντικείμενο της παρούσας μελέτης είναι η βιβλιογραφική επισκόπηση και η αναλυτική μελέτη των μεθόδων αυτόματης επισημείωσης ιατρικών εικόνων. Η έρευνα της βιβλιογραφίας περιστρέφεται γύρω από δύο άξονες. Ο πρώτος αφορά την επισκόπηση αλγορίθμων και μεθόδων αυτόματης επισημείωσης εικόνας. Παρουσιάζοντας τις βασικές αρχές και καθιερωμένες μεθόδους αυτόματης επισημείωσης εικόνας, επιχειρεί να δώσει μια επισκόπηση αυτού του σημαντικού ερευνητικού πεδίου, αποσκοπώντας στη διάδοση της γνώσης και στην προσέλκυση μεγαλύτερου ενδιαφέροντος από διάφορες ερευνητικές κοινότητες για την ταχεία προώθηση της έρευνας στον τομέα αυτό. Ο δεύτερος άξονας αφορά μία εκτενή ανασκόπηση εβδομήντα εννέα επιστημονικών άρθρων που μελετούν συστήματα αυτόματης επισημείωσης ιατρικών εικόνων. Κατά την ανάλυση και αξιολόγησή τους παρουσιάζεται η συνεισφορά τους στην έρευνα της αυτόματης επισημείωσης ιατρικής εικόνας, τα προβλήματα που επέλυσαν αλλά και οι υφιστάμενοι περιορισμοί. Τα συμπεράσματα που εξάγονται, μπορούν να συμβάλουν στην ανάπτυξη ενός αποδοτικού συστήματος αυτόματης επισημείωσης ιατρικής εικόνας.

Summary

Medical imaging is an essential element of modern healthcare. Due to the large number of medical images, there is an urgent need for an efficient mechanism that can sort and search for medical images at a high semantic level. Automatic annotation of medical images is a key prerequisite for creating semantic records that can be used to enhance computer-aided diagnosis, physician education, and biomedical research. It is also considered to be a key prerequisite for the development of effective search and retrieval engines for medical images, reports and associated publications. The purpose of the present study is to review the literature and to analyze in detail the methods of automatic annotation of medical images. The research of the literature revolves around two axes. The first concerns the overview of algorithms and methods of automatic image annotation. Introducing the basic principles and established methods of automatic image annotation, it attempts to give an overview of this important research field, with the aim of disseminating knowledge and attracting greater interest from various research communities to rapidly promote research in this field. The second axis concerns an extensive review of seventy nine scientific articles studying medical image annotation systems. Their analysis and evaluation reveal their contribution to the research of automatic medical image annotation, the problems that have been solved and the existing limitations. The findings can contribute to the development of an efficient automatic medical image annotation system.

Ευχαριστίες

Η συγγραφή της διπλωματικής διατριβής αποτελεί την ολοκλήρωση μιας πορείας δύο ετών. Θα ήθελα να εκφράσω τις ευχαριστίες μου στην τριμελή επιτροπή. Ξεκινώντας από τον Δρ. Χαράλαμπο Μπαλή και τον Δρ. Μάριο Νεοφύτου τους οποίους είχα τη χαρά να έχω και καθηγητές μου στο μεταπτυχιακό πρόγραμμα. Οι συμβουλές τους και η καθοδήγησή τους υπήρξαν πολύτιμος οδηγός για μένα. Ιδιαιτέρως, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου Δρ. Ζήνων Θεοδοσίου για την άριστη επικοινωνία, την κατανόηση, την αστείρευτη υπομονή του και την πολύτιμη αρωγή που μου προσέφερε κατά τη διάρκεια της εκπόνησης της παρούσας εργασίας. Όλοι συνέβαλαν στην επίτευξη του τελικού αποτελέσματος.

Περιεχόμενα

Εικόνες

Πίνακες

Κεφάλαιο 1
Εισαγωγή

Στο πρώτο κεφάλαιο παρουσιάζεται η ταυτότητα της έρευνας. Αναφέρεται ο σκοπός της εργασίας καθώς και η συνεισφορά της στην έρευνα του επιστημονικού πεδίου της αυτόματης επισημείωσης της ιατρικής εικόνας. Περιγράφεται η μεθοδολογία που ακολουθήθηκε στην ανασκόπηση της βιβλιογραφίας η οποία έχει δύο στόχους: ο πρώτος, να τεθεί με ακρίβεια το πρόβλημα της αυτόματης επισημείωσης και να περιγραφούν οι βασικές πτυχές του και ο δεύτερος, η ανάλυση και η αξιολόγηση μεθόδων και συστημάτων αυτόματης επισημείωσης ιατρικών εικόνων.

1.1 Ο σκοπός της εργασίας

Η ιατρική απεικόνιση αποτελεί βασικό στοιχείο της σύγχρονης υγειονομικής περίθαλψης καθώς διαδραματίζει σημαντικό ρόλο στη διάγνωση των ασθενειών, τον προγραμματισμό της θεραπείας και την αξιολόγηση της απόκρισης του ασθενή στη θεραπευτική αγωγή. Ο αριθμός των ψηφιακών ιατρικών εικόνων που παράγεται καθημερινά κατά την κλινική πράξη, είναι πολύ μεγάλος και διαρκώς αυξάνεται την τελευταία δεκαετία.

Αυτές οι ιατρικές εικόνες αποθηκεύονται σε βάσεις δεδομένων μεγάλης κλίμακας και μπορούν να διευκολύνουν τους ιατρούς, τους επαγγελματίες της υγείας, τους ερευνητές και τους φοιτητές στο κλινικό τους έργο για παροχή υγειονομικής περίθαλψης υψηλού επιπέδου στον ασθενή και μπορούν να παράσχουν πολύτιμες πληροφορίες που θα προωθήσουν και θα υποστηρίξουν την ιατρική έρευνα. Επιπλέον, η δημιουργία ψηφιακών βιβλιοθηκών ιατρικών εικόνων κρίνεται απαραίτητη για την εκπαίδευση του ιατρικού προσωπικού καθώς επίσης και για την αξιολόγηση μεθόδων αυτόματης επεξεργασίας και ανάλυσης ιατρικής εικόνας από τους ερευνητές της επιστήμης υπολογιστών.

Λόγω της αυξανόμενης χρήσης ψηφιακών ιατρικών εικόνων υπάρχει ανάγκη να αναπτυχθούν προηγμένες τεχνικές ανάκτησης πληροφοριών οι οποίες μπορούν να βελτιώσουν την αποτελεσματικότητα της περιήγησης και αναζήτησης μεγάλων βάσεων δεδομένων ιατρικών εικόνων. Μεταξύ των διαφόρων προηγμένων τεχνικών ανάκτησης

πληροφοριών, η επισημείωση εικόνας θεωρείται ως προϋπόθεση για τη διαχείριση μιας βάσης δεδομένων ψηφιακών εικόνων.

Εάν οι εικόνες είναι επισημειωμένες με λεκτική περιγραφή, η αναζήτηση με βάση λέξεις-κλειδιά μπορεί να χρησιμοποιηθεί για την ανάκτηση των εικόνων. Ωστόσο, η τακτικής της χειροκίνητης επισημείωσης, ειδικά σε μεγάλες βάσεις δεδομένων εικόνων, έχει σημαντικούς περιορισμούς. Οι μη αυτόματες επισημειώσεις απαιτούν πολύ χρόνο και είναι δαπανηρές για την υλοποίησή τους. Καθώς ο αριθμός των αρχείων σε μια βάση δεδομένων αυξάνεται, είναι ανέφικτο να εντοπιστούν και να χαρακτηριστούν χειροκίνητα όλες οι ιδιότητες του περιεχομένου μίας εικόνας. Οι επισημειώσεις που εξαρτώνται από τον ανθρώπινο παράγοντα, δεν καταφέρνουν να αντιμετωπίσουν την απόκλιση των υποκειμενικών αντιλήψεων. Όταν διαφορετικά άτομα πραγματοποιούν επισημείωση μιας εικόνας, παρέχουν συνήθως διαφορετική περιγραφή ανάλογη με τις υποκειμενικές αντιλήψεις τους. Επιπλέον είναι δύσκολο να αποδίδεται συγκεκριμένη περιγραφή για ορισμένα περιεχόμενα της εικόνας. Το σχήμα των οργάνων στις ιατρικές εικόνες, για παράδειγμα, είναι πολύ περίπλοκο για να περιγραφεί.

Το ερευνητικό ενδιαφέρον έχει εστιάσει στην κατασκευή αυτομάτων μεθόδων επισημείωσης ιατρικών εικόνων (automatic medical image annotation) που είναι άμεσα συνδεδεμένες με την αυτόματη ταξινόμηση εικόνας. Η αυτόματη επισημείωση εικόνας είναι μια μέθοδος που εκχωρεί αυτόματα ένα σύνολο γλωσσικών όρων στις εικόνες προκειμένου να τις κατηγοριοποιήσει εννοιολογικά και να παράσχει τα μέσα για την αποτελεσματική πρόσβαση σε εικόνες από βάσεις δεδομένων.

Στόχος της παρούσας μελέτης είναι η βιβλιογραφική επισκόπηση και η αναλυτική μελέτη των μεθόδων αυτόματης επισημείωσης ιατρικών εικόνων. Παρουσιάζοντας τις βασικές αρχές και καθιερωμένες μεθόδους αυτόματης επισημείωσης εικόνας, επιχειρεί να δώσει μια επισκόπηση αυτού του σημαντικού ερευνητικού πεδίου. Επιπλέον, αποσκοπεί στη διάδοση της γνώσης των διαφορετικών προσεγγίσεων αυτόματης επισημείωσης στις εφαρμογές διαχείρισης ιατρικών εικόνων και στην προσέλκυση μεγαλύτερου ενδιαφέροντος από διάφορες ερευνητικές κοινότητες για την ταχεία προώθηση της έρευνας στον τομέα αυτό.

Η μεταπτυχιακή εργασία διερευνεί τους λόγους που υπαγορεύουν την ανάπτυξη αποδοτικών μεθόδων αυτόματης επισημείωσης ιατρικής εικόνας, καθώς επίσης παρουσιάζει και αξιολογεί τις πιο αναγνωρισμένες μεθόδους που έχουν αναπτυχθεί. Παράλληλα, εξετάζονται περιορισμοί και προβλήματα που ανέκυπταν τις τελευταίες δύο δεκαετίες στη διάρκεια των οποίων η αυτόματη επισημείωση εικόνας σημείωσε σημαντικά βήματα προόδου. Τέλος, καταγράφονται οι αναδυόμενες κατευθύνσεις και οι σύγχρονες τάσεις στον τομέα της αυτόματης επισημείωσης ιατρικής εικόνας.

1.2 Η συνεισφορά της εργασίας

Οι ολοένα αυξανόμενες συλλογές εικόνων διαφόρων ειδών οδήγησαν στην ανάγκη ανάπτυξης μεθόδων για αποτελεσματική ταξινόμηση, αναζήτηση και ανάκτησή τους. Αυτό ισχύει ιδιαίτερα για τις συλλογές ιατρικών εικόνων όπου το μέγεθος των εικόνων που συλλέγονται μέσω των ημερήσιων κλινικών διαδικασιών, μπορεί να είναι τεράστιο. Λόγω του μεγάλου αριθμού ιατρικών εικόνων, υπάρχει επείγουσα ανάγκη για έναν αποδοτικό μηχανισμό που μπορεί να ταξινομεί και να αναζητά ιατρικές εικόνες σε υψηλό σημασιολογικό επίπεδο.

Η αυτόματη επισημείωση των ιατρικών εικόνων αποτελεί βασική προϋπόθεση για τη δημιουργία σημασιολογικών αρχείων που μπορούν να χρησιμοποιηθούν για την ενίσχυση της τεκμηριωμένης διάγνωσης, της εκπαίδευσης των ιατρών και της βιοϊατρικής έρευνας. Επίσης θεωρείται βασική προϋπόθεση για την ανάπτυξη μηχανών αποτελεσματικής αναζήτησης και ανάκτησης ιατρικών εικόνων.

Στις δύο τελευταίες δύο δεκαετίες, έχουν καταβληθεί σημαντικές προσπάθειες για την ανάπτυξη διαφόρων μεθόδων αυτόματης επισημείωσης. Αναπτύχθηκαν μοντέλα επισημείωσης που βασίζονται στη μηχανική μάθηση και σε μεθόδους ταξινόμησης, τεχνικές επισημείωσης βασισμένες στην ανάκτηση εικόνας, αλλά και μέθοδοι βαθιάς μάθησης. Οι προτεινόμενες προσεγγίσεις επισημείωσης εικόνας μπορούν να κατηγοριοποιηθούν με πολλούς τρόπους. Στην ελληνική βιβλιογραφία απουσιάζει μέχρι και σήμερα, μια γενική ταξινόμηση και σε βάθος ανασκόπηση των μεθόδων αυτόματης επισημείωσης εικόνας. Για το λόγο αυτό θεωρούμε ότι η μελέτη μας η οποία δεν περιορίζεται απλά στην παρουσίαση διαφόρων μεθόδων αλλά ακολουθεί μια σαφώς καθορισμένη ταξινόμηση στην προσπάθεια

εξήγησής τους, είναι απαραίτητη γιατί καλύπτει όλες τις πτυχές της αυτόματης επισημείωσης εικόνας.

1.3 Μεθοδολογία της έρευνας και περιορισμοί

Η μελέτη μας επιχειρεί μια βιβλιογραφική επισκόπηση του τομέα της αυτόματης επισημείωσης εικόνας. Η έρευνα της βιβλιογραφίας περιστρέφεται γύρω από δύο άξονες. Ο πρώτος αφορά την επισκόπηση αλγορίθμων και μεθόδων αυτόματης επισημείωσης εικόνας. Στο πλαίσιο αυτό μελετήθηκε ένα σύνολο δημοσιευμένων άρθρων σε έγκυρα επιστημονικά περιοδικά και πρακτικά επιστημονικών συνεδρίων με σκοπό να παρουσιαστούν τα βασικά βήματα που ακολουθούνται στην αυτόματη επισημείωση εικόνας συμπεριλαμβάνοντας, αυτά της εξαγωγής οπτικών χαρακτηριστικών και αναπαράστασης της εικόνας καθώς και της ταξινόμησης. Οι μέθοδοι εξαγωγής οπτικών χαρακτηριστικών αναλύονται διεξοδικά καθώς αποτελούν το πρώτο βήμα κάθε μεθόδου επισημείωσης. Παρουσιάζονται οι κύριοι κλάδοι της αυτόματης επισημείωσης εικόνας ταξινομημένοι με βάση τέσσερα κριτήρια: την μέθοδο μάθησης, το σύνολο δεδομένων εκπαίδευσης, το παραγόμενο μοντέλο και το μήκος της παραγόμενης επισημείωσης. Επιχειρώντας να καλυφθεί κάθε πτυχή της αυτόματης επισημείωσης εικόνας καταγράφονται τα μέτρα αξιολόγησης που χρησιμοποιούνται για τη σύγκριση της απόδοσης των διαφόρων μεθόδων και συστημάτων καθώς και διαθέσιμες βάσεις δεδομένων για την ανάπτυξη και δοκιμή τους.

Ο δεύτερος άξονας αφορά μία εκτενή ανασκόπηση εβδομήντα εννέα επιστημονικών άρθρων που μελετούν συστήματα αυτόματης επισημείωσης ιατρικών εικόνων. Όλες αυτές οι μελέτες παρουσιάζονται συγκριτικά ως προς τις μεθόδους που ακολουθούν για την αναπαράσταση της εικόνας και την ταξινόμηση. Κατά την ανάλυση και αξιολόγησή τους παρουσιάζεται η συνεισφορά τους στην έρευνα της αυτόματης επισημείωσης ιατρικής εικόνας, τα προβλήματα που επέλυσαν αλλά και οι υφιστάμενοι περιορισμοί που υποδεικνύουν τις μελλοντικές κατευθύνσεις της έρευνας στην περιοχή της αυτόματης επισημείωσης εικόνας. Τα συμπεράσματα που εξάγονται, μπορούν να συμβάλουν στην ανάπτυξη ενός αποδοτικού συστήματος αυτόματης επισημείωσης ιατρικής εικόνας.

1.4 Διάρθρωση της εργασίας

Η εργασία αναπτύσσεται σε επτά κεφάλαια. Στο **πρώτο** εισαγωγικό κεφάλαιο παρουσιάζεται ο σκοπός της εργασίας, τονίζεται η συνεισφορά της και αναλύεται η μεθοδολογία που

ακολουθήθηκε. Στο **δεύτερο** κεφάλαιο, γίνεται μία σύντομη εισαγωγή στην έννοια της αυτόματης επισημείωσης. Μετά από μία σύντομη αναφορά των λόγων που καθιστούν αναγκαία την επισημείωση εικόνας κατά την ανάλυση της ιατρικής εικόνας, παρουσιάζεται η δομή ενός τυπικού συστήματος αυτόματης επισημείωσης. Η εισαγωγή στην αυτόματη επισημείωση ολοκληρώνεται με μία σύντομη βιβλιογραφική ανασκόπηση της εξέλιξης του επιστημονικού πεδίου της αυτόματης επισημείωσης στη διάρκεια των τριών τελευταίων δεκαετιών. Στο **τρίτο** κεφάλαιο, εξετάζονται διεξοδικά τα χαρακτηριστικά που χρησιμοποιούνται για την αναπαράσταση της εικόνας. Η εξαγωγή χαρακτηριστικών αποτελεί το πρώτο βήμα της σημασιολογικής κατανόησης της εικόνας. Στο **τέταρτο** κεφάλαιο, γίνεται μια εκτενής παρουσίαση των μεθόδων αυτόματης επισημείωσης εικόνας. Οι πολυάριθμες τεχνικές που συναντώνται στη βιβλιογραφία, ταξινομούνται ως προς τη μέθοδο μάθησης, το σύνολο των δεδομένων εκπαίδευσης, το παραγόμενο μοντέλο και το μήκος της παραγόμενης επισημείωσης. Ιδιαίτερη αναφορά γίνεται στις μεθόδους επισημείωσης που στηρίζονται στη βαθιά μάθηση καθώς αποτελούν τεχνολογίες αιχμής. Στο **πέμπτο** κεφάλαιο, συνοψίζουμε καθιερωμένες μετρικές και ανοικτές βάσεις δεδομένων αναφοράς για την αξιολόγηση μεθόδων αυτόματης επισημείωσης εικόνας. Στο **έκτο** κεφάλαιο, παρουσιάζουμε συγκριτικά μια σειρά από μελέτες μεθόδων αυτόματης επισημείωσης ιατρικών εικόνων. Η ανάλυση και η αξιολόγησή τους αποκαλύπτει τα προβλήματα τα οποία επέλυσαν αλλά και τα ανοικτά ζητήματα που αποτελούν αντικείμενο έρευνας στον τομέα της αυτόματης επισημείωσης ιατρικής εικόνας. Τέλος, στο **έβδομο** κεφάλαιο, συνοψίζουμε τα γενικότερα συμπεράσματα που προκύπτουν από την ανάλυση και αξιολόγηση των συστημάτων αυτόματης επισημείωσης ιατρικής εικόνας και μπορούν να αποτελέσουν αντικείμενο περαιτέρω διερεύνησης στο μέλλον.

Κεφάλαιο 2
Αυτόματη Επισημείωση Εικόνας

Η αυτόματη επισημείωση ιατρικής εικόνας αποτελεί ένα ενεργό διεπιστημονικό πεδίο. Στο κεφάλαιο αυτό περιγράφουμε συνοπτικά τους περιορισμούς των συστημάτων ανάκτησης εικόνων με βάση το περιεχόμενο που ανέδειξαν την ανάγκη για την αυτόματη επισημείωση των εικόνων. Μετά από μία αναφορά στη βασική ιδέα της αυτόματης επισημείωσης εικόνας, παρουσιάζεται μια σύντομη ιστορική αναδρομή των εξελίξεων που σημειώθηκαν στον τομέα της αυτόματης επισημείωσης εικόνας τις τελευταίες τρεις δεκαετίες.

2.1. Συστήματα ανάκτησης ιατρικής εικόνας με βάση το περιεχόμενο

Η εικόνα είναι ένα από τα σημαντικότερα εργαλεία στην ιατρική δεδομένου ότι παρέχει μια μέθοδο για τη διάγνωση, την παρακολούθηση των αντιδράσεων στη φαρμακευτική αγωγή και τη θεραπεία ασθενών έχοντας το πλεονέκτημα ότι είναι μια πολύ γρήγορη μη επεμβατική διαδικασία με ελάχιστες παρενέργειες και με μια εξαιρετική σχέση κόστους-αποτελέσματος (Amaral et al., 2010). Κατά τα τελευταία είκοσι πέντε χρόνια το πεδίο της διαγνωστικής ιατρικής απεικόνισης αναπτύχθηκε ταχέως και άλλαξε λόγω της εισαγωγής νέων μορφών απεικόνισης και της ενίσχυσης των υφιστάμενων τεχνικών (Nalini and Malleswari, 2017). Καθώς νέες, πιο αποτελεσματικές συσκευές λήψης εικόνων αναπτύσσονται συνεχώς ώστε να παράγονται ακριβέστερες πληροφορίες και η χωρητικότητα αποθήκευσης δεδομένων αυξάνεται, παρατηρείται μια σταθερή αύξηση του αριθμού των ιατρικών εικόνων που παράγονται. Ένα χαρακτηριστικό παράδειγμα αυτής της τάσης αποτελεί το Τμήμα Ακτινολογίας του Πανεπιστημιακού Νοσοκομείου της Γενεύης, όπου οι ιατρικές εικόνες που παρήγαγε ημερησίως ανήλθαν από 12.000 το 2002 σε 50.000 το 2007 (Amaral et al., 2010). Με μια τέτοια εκθετική αύξηση των ιατρικών δεδομένων σε ψηφιακές βιβλιοθήκες, γίνεται όλο και πιο δύσκολο να εκτελεστούν αναλύσεις που στηρίζονται στην αναζήτηση και ανάκτηση πληροφοριών.

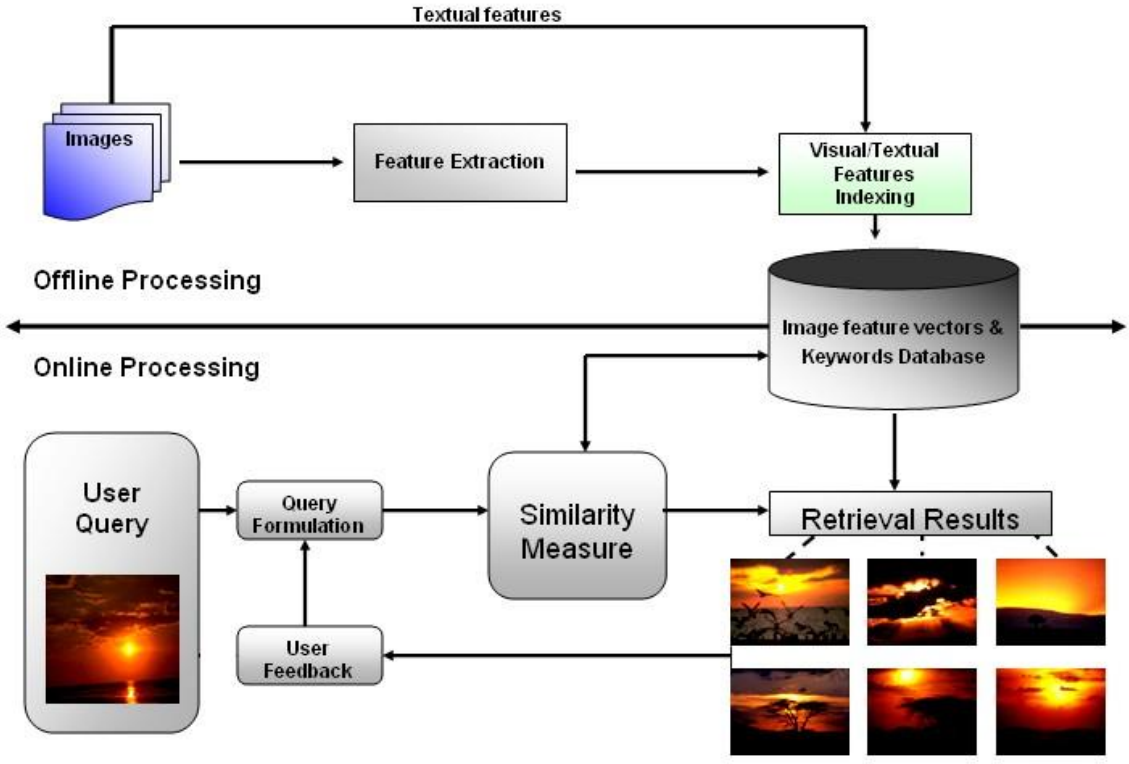
Ένα νοσοκομείο μπορεί να παράγει καθημερινά δεκάδες χιλιάδες ψηφιακές ιατρικές εικόνες. Τέτοιες ψηφιακές ιατρικές εικόνες περιλαμβάνουν ακτινογραφίες (X-ray), υπολογιστικές

τομογραφίες (computed tomography - CT), απεικονίσεις μαγνητικού συντονισμού (magnetic resonance imaging - MRI), απεικονίσεις λειτουργικής μαγνητικής τομογραφίας (functional magnetic resonance imaging - fMRI), φασματοσκοπίες μαγνητικού συντονισμού (magnetic resonance spectroscopy - MRS), απεικονίσεις μαγνητικής πηγής (magnetic source imaging - MSI), ψηφιακές αφαιρετικές αγγειογραφίες (DSA), τομογραφίες εκπομπής ποζιτρονίων (PET), υπερηχογραφήματα (US), ενδοσκοπήσεις, κ.ά. Λόγω της αυξανόμενης χρήσης ψηφιακών ιατρικών εικόνων, υπάρχει ανάγκη να αναπτυχθούν προηγμένες τεχνικές ανάκτησης πληροφοριών, οι οποίες μπορούν να βελτιώσουν την αποτελεσματικότητα της περιήγησης και αναζήτησης σε μεγάλες βάσεις δεδομένων ιατρικών εικόνων (Wei and Chen, 2012).

Οι χρήσεις που μπορεί να έχει ένα τέτοιο σύστημα ανάκτησης, μπορούν να κατηγοριοποιηθούν σε τρεις μεγάλες κατηγορίες: (α) διδασκαλία, καθώς οι καθηγητές και οι εκπαιδευόμενοι μπορούν να ανακτήσουν παρόμοιες οπτικά περιπτώσεις (πιθανώς με διαφορετική διάγνωση) χωρίς τη χρήση προσωπικών πληροφοριών για τους ασθενείς, (β) έρευνα, καθώς τα πρότυπα που μπορούν να οδηγήσουν σε νέες γνώσεις, μπορεί να βρεθούν ευκολότερα, και (γ) διάγνωση, καθώς οι γιατροί μπορούν να συγκρίνουν εικόνες που παρουσιάζουν παθολογικές περιπτώσεις με τις εικόνες των υγιών οργάνων ή να αναζητήσουν παρόμοιες οπτικά περιπτώσεις για να αναλύσουν τη θεραπεία που εφαρμόστηκε. Επιπλέον, θα μπορούσε να χρησιμοποιηθεί στο πεδίο της συλλογιστικής βάσει υποθέσεων ή της τεκμηριωμένης ιατρικής, που υπάρχει ανάγκη εύρεσης παρόμοιων ιατρικών περιπτώσεων (Ko et al., 2012).

Επειδή η ανάκτηση πληροφορίας κειμένου είναι ήδη ένα ώριμο ερευνητικό πεδίο, η χρήση κειμένου για την περιγραφή του περιεχομένου και του πλαισίου μιας εικόνας διαδραματίζει σημαντικό ρόλο στην ανάπτυξη μιας ιατρικής βάσης δεδομένων (Ko et al., 2012, Amaral et al., 2010). Οι πρώτες μηχανές αναζήτησης που αναπτύχθηκαν, και είναι ευρέως διαδεδομένες, έχουν υιοθετήσει προσεγγίσεις ανάκτησης εικόνων βάσει κειμένου (text-based). Αυτές οι λύσεις παρουσιάζουν ωστόσο σημαντικούς περιορισμούς επειδή οι ψηφιακές εικόνες προς αναζήτηση είτε δεν έχουν επισημειωθεί είτε έχουν επισημειωθεί χρησιμοποιώντας ανακριβείς λέξεις-κλειδιά (Maher, 2017). Μια μελέτη των Güld et al. (2002) αναφέρει ότι περίπου το 15% των επισημειώσεων στις κεφαλίδες DICOM (Digital Imaging and Communications in Medicine) είναι εσφαλμένες. Επίσης, ολόκληρη η κεφαλίδα DICOM χάνεται πολύ συχνά ως συνέπεια της συμπίεσης εικόνας.

Η ανάκτηση εικόνων βάσει περιεχομένου (Content Based Image Retrieval - CBIR) εμφανίστηκε ως ένα πολλά υποσχόμενο υποκατάστατο προκειμένου να ξεπεραστούν οι προκλήσεις που αντιμετωπίζουν οι τεχνικές ανάκτησης εικόνας που βασίζονται σε κείμενο. Στην πραγματικότητα, οι ψηφιακές εικόνες, οι οποίες ανακτώνται χρησιμοποιώντας το σύστημα CBIR, αντιπροσωπεύονται από ένα σύνολο οπτικών χαρακτηριστικών. Όπως απεικονίζεται στην εικόνα 1, ένα τυπικό σύστημα CBIR αποτελείται από ένα offline στάδιο που στοχεύει στην εξαγωγή και την αποθήκευση των διανυσμάτων οπτικών χαρακτηριστικών από τις εικόνες της βάσης δεδομένων. Από την άλλη πλευρά, το online κομμάτι της εφαρμογής επιτρέπει στο χρήστη να ξεκινήσει την εργασία ανάκτησης παρέχοντας την εικόνα του ερωτήματος. Τέλος, ένα τυπικό σύστημα CBIR επιστρέφει ένα σύνολο εικόνων οπτικά «όμοιων» με το ερώτημα του χρήστη. Ωστόσο, το κύριο μειονέκτημα ενός τέτοιου συστήματος συνίσταται στην υπόθεση ότι η οπτική ομοιότητα αντανακλά τη σημασιολογική ομοιότητα. Αυτή η υπόθεση ωστόσο δεν ικανοποιείται λόγω του σημασιολογικού χάσματος μεταξύ της σημασίας (υψηλότερου επιπέδου) και των οπτικών χαρακτηριστικών (χαμηλού επιπέδου) (Maher, 2017). Επιπρόσθετα, το γεγονός ότι ο χρήστης θα πρέπει πάντα να έχει διαθέσιμη την εικόνα που θα δοθεί στο ερώτημα, δημιουργεί αρκετούς περιορισμούς στη χρήση ενός τέτοιου συστήματος.



Εικόνα 1. Επισκόπηση ενός τυπικού συστήματος CBIR (Maher, 2007).

Το κύριο πρόβλημα που σχετίζεται με την ποιότητα των αποτελεσμάτων είναι ο τρόπος κατασκευής μιας ακριβούς αναπαράστασης του περιεχομένου της εικόνας, ώστε να μπορούν να ανακτηθούν όλες οι αντιληπτικά παρόμοιες εικόνες. Λόγω του σημασιολογικού χάσματος, ακόμη και οι σύγχρονες μέθοδοι βασίζονται σε χαρακτηριστικά χαμηλού επιπέδου, όπως το χρώμα, η υφή ή το σχήμα.

Το πρόβλημα αυτό είναι ακόμα πιο δύσκολο σε ιατρικό πλαίσιο, επειδή οι ιατρικές εικόνες, όπως οι ακτινογραφίες και οι μαγνητικές τομογραφίες, έχουν συγκεκριμένα χαρακτηριστικά που πρέπει να λαμβάνει υπόψη το σύστημα CBIR. Οι ιατρικές εικόνες είναι συνήθως χαμηλής ανάλυσης εικόνες με υψηλό θόρυβο. Είναι εικόνες μόνο έντασης και περιέχουν λιγότερες πληροφορίες σχετικά με το χρώμα καθώς το μεγαλύτερο μέρος του αφορά γεωμετρικές πληροφορίες. Οι ιατρικές εικόνες που προκύπτουν από διαφορετικές μεθόδους απεικόνισης, όπως για παράδειγμα, ακτινογραφική προβολή (π.χ. ακτίνες Χ, πυρηνική ιατρική) και τομογραφία (π.χ. CT, MRI, υπερηχογράφημα), επιβάλλουν μοναδικούς, η καθεμία, περιορισμούς που εξαρτώνται από την εικόνα σχετικά με τη φύση των χαρακτηριστικών που είναι διαθέσιμα για εξαγωγή. Τα χαρακτηριστικά της υφής και του σχήματος είναι πιο ισχυρά χαρακτηριστικά στην ανάλυση των ιατρικών εικόνων (Nalini and Malleswari, 2017).

Οι πληροφορίες που μεταφέρονται από ιατρικές εικόνες διαφέρουν από τις γενικές εικόνες στα εξής: (1) Οι ιατρικές γνώσεις που συλλέγονται από ιατρικές εικόνες συνήθως δεν είναι ακριβείς, (2) Τα χωρικά δεδομένα της πληροφορίας που μεταφέρει μια ιατρική εικόνα μπορεί να μην εκφράζονται σε κατάλληλη (συμβατική) γλώσσα, (3) Ένα μεγάλο μέρος των πληροφοριών της εικόνας είναι γεωμετρικού χαρακτήρα (Nalini and Malleswari, 2017).

Σημαντικός αριθμός ερευνών αναδεικνύει μια σειρά από νέες προκλήσεις, οι οποίες έχουν ιδιαίτερη σημασία για τα συστήματα CBIR. Μία από αυτές είναι η αυτόματη επισημείωση ιατρικής εικόνας (Computer-aided Medical Image Annotation) (Maher, 2017).

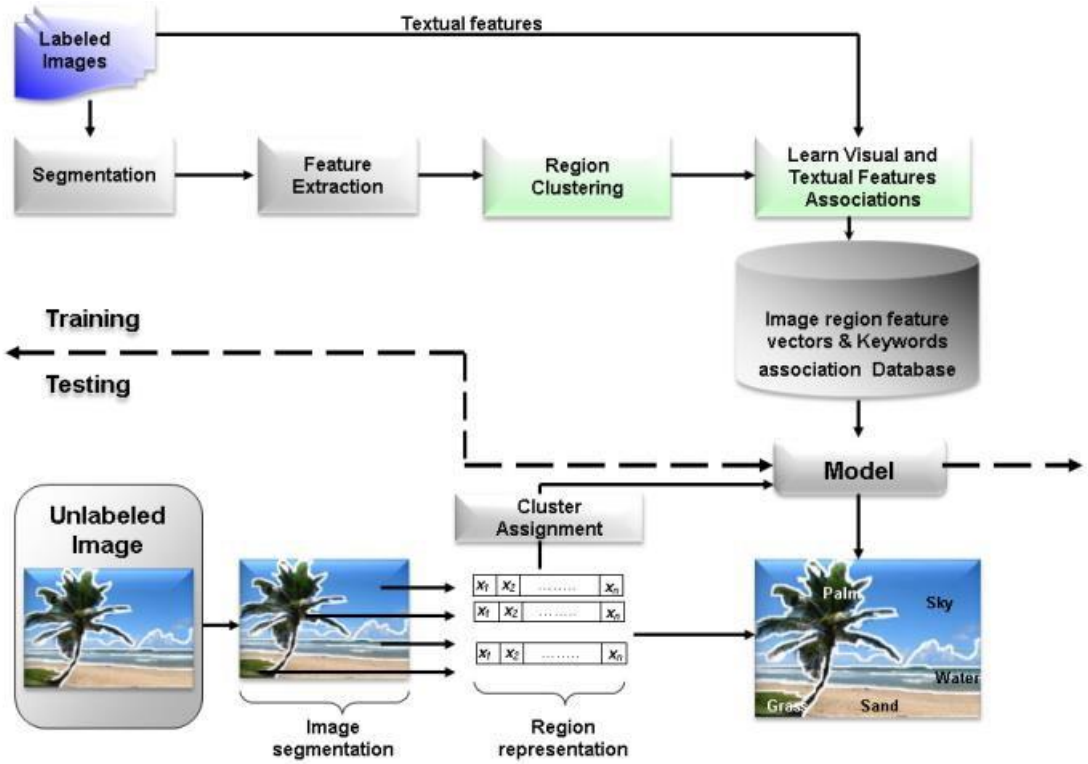
2.2. Αυτόματη επισημείωση ιατρικής εικόνας

Η επισημείωση εικόνας θεωρείται βασική προϋπόθεση για σημασιολογικές ιατρικές μηχανές αναζήτησης που επιτρέπουν στο ιατρικό προσωπικό να βρίσκει αποτελεσματικότερα ιατρικές εικόνες, εκθέσεις και συναφείς δημοσιεύσεις. Η αυτόματη σημασιολογική επισημείωση κρίνεται απαραίτητη επειδή είναι δύσκολο, χρονοβόρο και ακριβό να επισημειωθεί

χειροκίνητα το πλούσιο περιεχόμενο των ιατρικών εικόνων (Kumar et al., 2016). Η περίπτωση του έργου caBIG (cancer Biomedical Informatics Grid), ενός κυβερνητικού προγράμματος των ΗΠΑ για την ανάπτυξη ενός ανοιχτού κώδικα, ανοιχτής πρόσβασης δικτύου πληροφοριών για ασφαλή ανταλλαγή δεδομένων σχετικά με την έρευνα για τον καρκίνο (Channin et al., 2009), που προέβλεπε μια βιβλιοθήκη λογισμικού που θα μπορούσε να χρησιμοποιηθεί για την επισημείωση μεγάλων συλλογών εικόνων, αποτελεί χαρακτηριστικό παράδειγμα των δυσκολιών που συνεπάγεται η διαδικασία της μη αυτόματης επισημείωσης. Οι Wennerberg et al. (2011) βελτίωσαν την αποτελεσματικότητα αυτής της χειροκίνητης διαδικασίας επισημείωσης χρησιμοποιώντας ένα εργαλείο μοντελοποίησης οντολογιών που αναγνωρίζει και ταξινομεί θραύσματα μιας οντολογίας που σχετίζονται με την εργασία επισημείωσης. Ωστόσο, αυτές οι προσεγγίσεις των χειροκίνητων επισημειώσεων απαιτούν από τους ιατρούς να καθορίσουν υποκειμενικά τις ετικέτες που σχετίζονται με μια συγκεκριμένη εικόνα με βάση τη γνώση και την προηγούμενη εμπειρία τους (Kumar et al., 2016). Αντίθετα, η αυτόματη επισημείωση εικόνας πραγματοποιείται βάσει ποσοτικοποιήσιμων χαρακτηριστικών εικόνας. Ο συνδυασμός χαρακτηριστικών που υπάρχουν σε κάθε εικόνα, όπως το χρώμα (color), η υφή (texture) και το σχήμα (shape), οδηγεί στην επιλογή των σχετικών επισημειώσεων (Kumar et al., 2016).

Η ανάγκη δημιουργίας μεγάλων βάσεων ιατρικών εικόνων για βελτιωμένη ανάλυση δεδομένων προϋποθέτει την τυποποίηση ιατρικών επισημειώσεων γεγονός που οδήγησε σε μια μεγάλη επένδυση για την ανάπτυξη δομημένων μεθοδολογιών αναφοράς μέσω της χρήσης κοινών λεξικών όπως το RadLex (Marvasti et al., 2017). Η προσπάθεια αυτή συντέλεσε στην ανάπτυξη μεθόδων επισημείωσης ιατρικής εικόνας που χρησιμοποιούν αυτά τα λεξικά. Η αυτόματη επισημείωση ιατρικής εικόνας (Computer-aided Medical Image Annotation – CMIA) μπορεί να περιγραφεί ως το έργο εκχώρησης ετικετών υψηλού επιπέδου (για παράδειγμα όροι RadLex) σε έννοιες (σημασιολογικά χαρακτηριστικά) χρησιμοποιώντας χαρακτηριστικά εικόνας χαμηλού επιπέδου, που μπορούν να χρησιμοποιηθούν ως ετικέτες της εικόνας σε ιατρικά συστήματα αναζήτησης/ανάκτησης εικόνων/εγγράφων και στη δημιουργία δομημένων ακτινολογικών εκθέσεων, οι οποίες αποτελούν το κύριο εμπόδιο για την πλήρη αξιοποίηση των εργαλείων ανάλυσης δεδομένων στην ιατρική πληροφορική (Marvasti et al., 2017).

Ένα σύστημα αυτόματης επισημείωσης ιατρικής εικόνας (CMIA), χρησιμοποιεί ένα σύνολο επισημειωμένων εικόνων για εκπαίδευση. Αρχικά, κάθε εικόνα τμηματοποιείται σε περιοχές ενδιαφέροντος και τοπικά χαρακτηριστικά εξάγονται και χρησιμοποιούνται για να περιγράψουν κάθε περιοχή. Η κατάτμηση της εικόνας μπορεί να πραγματοποιηθεί είτε με μία προσέγγιση πλέγματος όπου η εικόνα διαιρείται σε ένα σύνολο σταθερού μεγέθους τμημάτων (μπλοκ), είτε με μία προσέγγιση βάσει περιοχής, όπου η εικόνα διαχωρίζεται σε ομοιογενείς περιοχές, δηλαδή περιοχές που μοιράζονται κοινά χαρακτηριστικά. Στην ιδανική περίπτωση, κάθε περιοχή αντιστοιχεί σε ένα διαφορετικό αντικείμενο στην εικόνα. Μετά την τμηματοποίηση, κάθε περιοχή αντιπροσωπεύεται από ένα διάνυσμα (φορέα) χαρακτηριστικών (Maher, 2017). Η εικόνα 2 δείχνει τη γενική αρχιτεκτονική ενός τυπικού συστήματος επισημείωσης εικόνας.



Εικόνα 2. Επισκόπηση ενός τυπικού συστήματος αυτόματης επισημείωσης εικόνας (Maher, 2007).

Μετά την κατάτμηση όλων των εικόνων εκπαίδευσης και την εξαγωγή οπτικών χαρακτηριστικών από τις περιοχές τους, χρησιμοποιείται ένας αλγόριθμος μηχανικής μάθησης για να μάθει το σύστημα συσχετίσεις ή κοινές κατανομές πιθανότητας μεταξύ αυτών των χαρακτηριστικών και των λέξεων-κλειδιών που χρησιμοποιούνται για την επισημείωση των εικόνων. Το τμήμα ελέγχου του συστήματος παίρνει, ως είσοδο, μια μη επισημειωμένη εικόνα, την χωρίζει σε ομοιογενείς περιοχές, εξάγει και κωδικοποιεί το οπτικό

περιεχόμενο κάθε περιοχής σε διανύσματα χαρακτηριστικών. Στη συνέχεια, χρησιμοποιεί τις αποκτηθείσες συσχετίσεις ή τις κοινές κατανομές πιθανοτήτων για να συναγάγει το σύνολο των λέξεων-κλειδιών που περιγράφουν καλύτερα τα οπτικά χαρακτηριστικά. Αυτές οι λέξεις-κλειδιά χρησιμοποιούνται στη συνέχεια για την επισημείωση της εικόνας (Maher, 2017).

2.3. Ιστορικό της επισημείωσης εικόνας: μια επισκόπηση

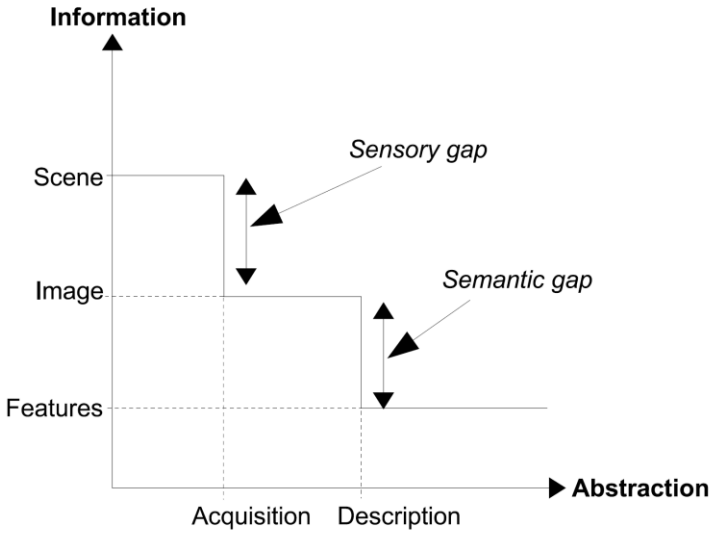
Στην ενότητα αυτή παραθέτουμε μία σύντομη επισκόπηση της εξέλιξης της αυτόματης επισημείωσης εικόνας κατά τη διάρκεια των τριών τελευταίων δεκαετιών.

2.3.1. Πρώτη δεκαετία (1990 -2000)

Οι απαρχές της επισημείωσης εικόνων τοποθετούνται τη δεκαετία του ΄90. Τα έτη 1994-2000 μπορούν να θεωρηθούν ως η αρχική φάση της έρευνας και της ανάπτυξης της ανάκτησης εικόνων με βάση το περιεχόμενο (CBIR) και κατ' επέκταση της επισημείωσης εικόνας (Datta et al., 2008). Ιδιαίτερα, μετά το 1996 παρατηρείται πολύ μεγάλη αύξηση στον αριθμό των άρθρων που δημοσιεύονται σχετικά με την ανάκτηση και επισημείωση εικόνας.

Η πρόοδος που σημειώνεται κατά τη διάρκεια αυτής της φάσης, συνοψίζεται από τους Smeulders et al. (2000), σε μία μελέτη που είχε σαφή επιρροή στην πρόοδο που σημειώθηκε κατά την επόμενη δεκαετία. Κρίνεται, σκόπιμο να παρουσιάσουμε μια σύντομη αναφορά των ιδεών, των επιρροών και των τάσεων των πρώτων αυτών χρόνων καθώς τα προβλήματα που τέθηκαν, καθόρισαν και παρακίνησαν τις περισσότερες από τις μελλοντικές ερευνητικές προσπάθειες.

Τα δύο βασικά προβλήματα που ανέδειξε η έρευνα σε αυτή την πρώιμη φάση, ήταν το αισθητηριακό (sensory gap) και το σημασιολογικό κενό (semantic gap). Το αισθητηριακό κενό είναι το κενό μεταξύ της σκηνής στον πραγματικό κόσμο και των πληροφοριών σε μια (υπολογιστική) περιγραφή που προέρχεται από μια καταγραφή αυτής της σκηνής. Το σημασιολογικό κενό είναι η έλλειψη σύμπτωσης μεταξύ των πληροφοριών που μπορούν να εξαχθούν από τα οπτικά δεδομένα και την ερμηνεία που έχουν τα ίδια δεδομένα για έναν χρήστη σε μια δεδομένη κατάσταση (Datta et al., 2008). Ενώ το πρώτο αναγνωρίζει τις προκλήσεις από το περιεχόμενο εικόνας λόγω των περιορισμών στην καταγραφή, το δεύτερο θέτει το ζήτημα της ερμηνείας των εικόνων από τον χρήστη καθώς είναι εγγενώς δύσκολο για το οπτικό περιεχόμενο να τη συλλάβει.



Εικόνα 3. Sensory gap: η διαφορά μεταξύ μιας σκηνής του πραγματικού κόσμου και της αναπαράστασής της σε μια εικόνα, Semantic gap: η διαφορά μεταξύ των χαρακτηριστικών χαμηλού επιπέδου και του πραγματικού περιεχομένου της εικόνας (Clouard et al., 2010).

Η αντιμετώπιση του αισθητηριακού χάσματος σημαίνει την επιλογή των κατάλληλων χαρακτηριστικών που δίνουν στον χρήστη τις πληροφορίες που χρειάζεται από μια εικόνα, λαμβάνοντας υπόψη ότι ορισμένες πληροφορίες χάνονται κατά τη δημιουργία της εικόνας. Τα πρώτα χρόνια, τα χαρακτηριστικά χαμηλού επιπέδου (υφή, χρώμα, σχήμα κ.λπ.) εξάγονται με χειροκίνητες τεχνικές εξαγωγής χαρακτηριστικών και δεν υπάρχει συσχέτιση μεταξύ αυτών των χαρακτηριστικών χαμηλού επιπέδου και των χαρακτηριστικών κειμένου.

Πολλές μελέτες συνεισέφεραν στην κατεύθυνση της εξαγωγής χαρακτηριστικών χρώματος, της υφής και του σχήματος από τις εικόνες. Μεταξύ της πρώτης χρήσης των ιστογραμμάτων χρώματος για την ευρετηρίαση εικόνων ήταν αυτή των Swain and Ballard (1991). Η εξαγωγή χαρακτηριστικών σε συστήματα όπως το QBIC (Flickner et al., 1995), το Pictoseek (Gevers and Smeulders, 2000) και το VisualSEEK (Smith and Chang, 1997) είναι επίσης αξιοσημείωτες.

Οι Huang et al. (1999) προτείνουν τα διαγράμματα συσχέτισης χρώματος (color correlograms) ως βελτιώσεις στα ιστογράμματα, τα οποία λαμβάνουν υπόψη και τη χωρική κατανομή των χρωμάτων. Τα φίλτρα Gabor χρησιμοποιήθηκαν με επιτυχία για τοπική εξαγωγή σχήματος με σκοπό την αντιστοίχιση και την ανάκτηση στους Manjunath and Ma (1996). Οι μετασχηματισμοί wavelet χρησιμοποιήθηκαν για τη βελτίωση της εξαγωγής χαρακτηριστικών χρώματος στο σύστημα WBIIS (Wang et al., 1998). Τα αμετάβλητα τοπικά χαρακτηριστικά για την ανάκτηση εικόνας (Schmid and Mohr, 1997) έλαβαν σημαντική προσοχή ως μέσο

γεφύρωσης του αισθητηριακού κενού. Εργασίες εξαγωγής τοπικών χαρακτηριστικών που βασίζονται σε προεξέχοντα (salient) σημεία περιοχών της εικόνας, όπως των Tuytelaars and Van Gool (1999), βρήκαν εφαρμογή σε τομείς όπως η ανάκτηση εικόνων.

Η αναγνώριση αντικειμένων σε εικόνες, ήταν επίσης ένα πολύ δύσκολο πρόβλημα το οποίο τέθηκε κατά την περίοδο αυτή. Οι Smeulders et al. (2000) στην έρευνά τους κατηγοριοποίησαν την αναγνώριση αντικειμένων σε εικόνες, ως ισχυρή (strong) / αδύναμη (weak) τμηματοποίηση (ομαδοποίηση με βάση δεδομένα), διαμέριση (ομαδοποίηση ανεξάρτητη από δεδομένα, για παράδειγμα σταθερά μπλοκ εικόνας) και εντοπισμό προτύπου (ομαδοποίηση με βάση ένα καθορισμένο πρότυπο). Αναλύονται επίσης τα πλεονεκτήματα και οι περιορισμοί της τμηματοποίησης της εικόνας και παρουσιάζονται προσεγγίσεις που μπορούν να αποφύγουν την ισχυρή τμηματοποίηση ενώ εξακολουθούν να χαρακτηρίζουν τη δομή της εικόνας επαρκώς για την ανάκτηση και την επισημείωση.

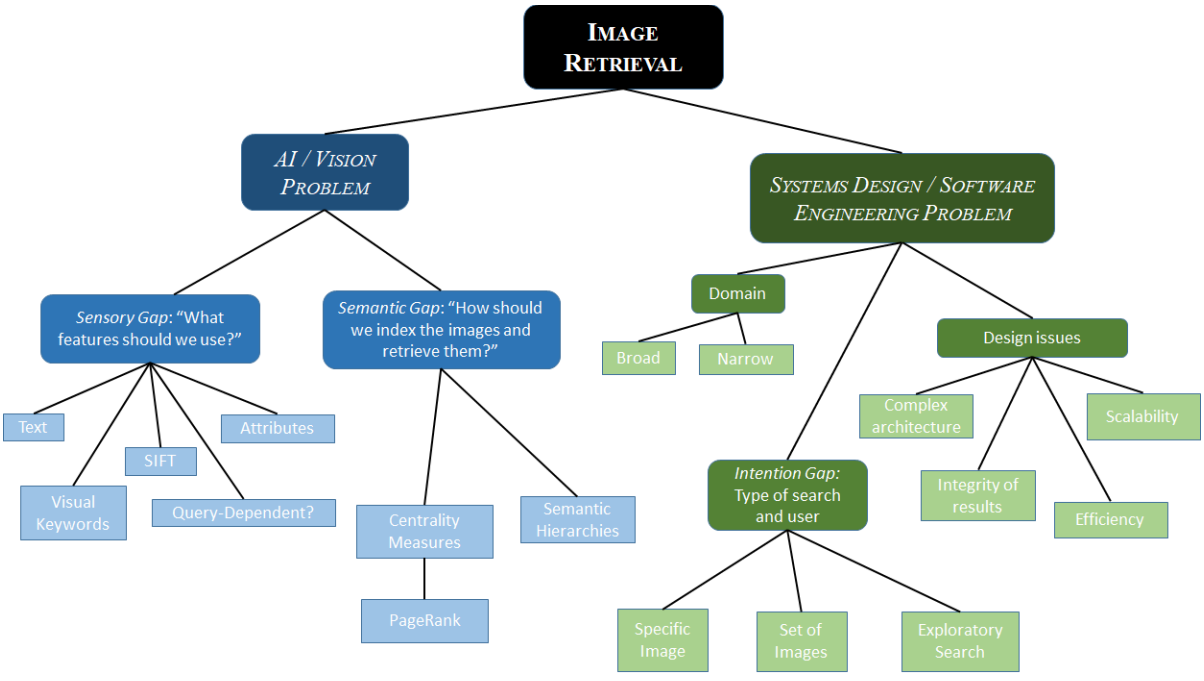
Στον τομέα της τμηματοποίησης της εικόνας σημειώθηκε σημαντική πρόοδος. Στους Zhu and Yuille (1996), οι ιδέες της τμηματοποίησης με ανάπτυξη περιοχών (snake and region growing) συνδυάστηκαν σε ένα πλαίσιο με αρχές ενώ οι Shi and Malik (2000) εφάρμοσαν διαμέριση γράφων για το σκοπό αυτό. Η χρήση γράφων για την αναπαράσταση των χωρικών σχέσεων μεταξύ αντικειμένων, ειδικά προσανατολισμένων προς την ιατρική απεικόνιση, διερευνήθηκε από τους Petrakis and Faloutsos (1997). Στους Smith and Chang (1997), δισδιάστατες συμβολοσειρές χρησιμοποιήθηκαν για τον χαρακτηρισμό των χωρικών σχέσεων μεταξύ περιοχών. Μια μέθοδος για την αυτόματη επιλογή χαρακτηριστικών προτάθηκε από τους Swets and Weng (1996).

Μετά την εξαγωγή των οπτικών χαρακτηριστικών της εικόνας, η ερώτηση παρέμεινε ως προς τον τρόπο με τον οποίο θα μπορούσαν να αναπροσαρμοστούν και να συνδυαστούν μεταξύ τους για την αποτελεσματική ανάκτηση και επισημείωση της εικόνας. Οι μέθοδοι που προτάθηκαν, αποσκοπούσαν ουσιαστικά στη μείωση όσο το δυνατόν περισσότερο του σημασιολογικού κενού. Στους Smeulders et al. (2000), παρουσιάζονται διεξοδικά τα μέτρα ομοιότητας που χρησιμοποιούνται στις εφαρμογές CBIR της περιόδου αυτής. Στενά συνδεδεμένος με τα μέτρα ομοιότητας είναι ο τρόπος με τον οποίο τα οπτικά χαρακτηριστικά συναντούν τις ανάγκες των χρηστών και, πιο πρακτικά, πώς μπορούν να τροποποιηθούν σταδιακά με την ανατροφοδότηση από τον χρήστη. Από την άποψη αυτή, μια σημαντική

πρόοδος στην τεχνολογία αλληλεπίδρασης του χρήστη για την ανάκτηση εικόνων ήταν η ανάδραση σχετικότητας (relevance feedback - RF). Μια σημαντική πρώιμη εργασία που εισήγαγε την RF στον τομέα ανάκτησης εικόνας ήταν αυτή των Rui et al. (1998), η οποία εφαρμόστηκε στο σύστημα MARS το οποίο ανέπτυξαν.

Συστήματα ανάκτησης εικόνων με βάση το περιεχόμενο (CBIR) που ξεχώρισαν σε αυτή την πρώιμη εποχή ήταν τα IBM QBIC (Query By Image Content) (Flickner et al., 1995), VIRAGE (Gupta and Jain, 1997) και NEC AMORE (Mukherjea et al., 1999) στον εμπορικό τομέα, και τα MIT Photobook (Pentland et al., 1996), Columbia VisualSEEK, WebSEEK (Smith and Chang, 1997), UCSB NeTra (Ma and Manjunath, 1997) και Stanford WBIIS (Wang et al., 1998) στον ακαδημαϊκό τομέα (Datta et al., 2008).

Οι περισσότερες από τις προτεινόμενες μεθόδους και συστήματα που αναπτύχθηκαν κατά την πρώτη περίοδο, χρησιμοποίησαν τα χειροκίνητα εξαχθέντα χαρακτηριστικά και κλασικούς ταξινομητές για να επισημειώσουν μια εικόνα. Ο κύριος προβληματισμός όλων των μεθόδων είναι η αντιμετώπιση του σημασιολογικού κενού. Όλες οι προτεινόμενες μέθοδοι είναι επιβλεπόμενης μάθησης και η ανάκτηση βασίζεται σε CBIR (Bhagat and Choudhary, 2018).



Εικόνα 4. Τα κυριότερα προβλήματα που σχετίζονται με την ανάκτηση εικόνας, όπως αυτά διατυπώθηκαν και σχηματοποιήθηκαν στο τέλος της πρώτης δεκαετίας (Chi and Cristante, 2015).

Στη μελέτη των Smeulders et al. (2000), εκτός από την παρουσίαση των βασικών μεθόδων επισημείωσης και της ανάκτησης που ξεχώρισαν κατά την πρώτη περίοδο, θίγονται πρακτικά ζητήματα όπως η υλοποίηση του συστήματος και η αρχιτεκτονική, οι περιορισμοί τους και ο τρόπος αντιμετώπισής τους, ο ρόλος του χρήστη, η οπτικοποίηση των αποτελεσμάτων και τέλος η αξιολόγηση του συστήματος. Στα προβλήματα που διατυπώθηκαν, προτείνουν λύσεις, θέτουν τους στόχους της μελλοντικής έρευνας. Τα περισσότερα από τα άρθρα της δεύτερης δεκαετίας ακολούθησαν τις κατευθύνσεις που υπαγορεύονται στην έρευνα των Smeulders et al. (2000).

2.3.2. Δεύτερη δεκαετία (2000-2010)

Στην πρώτη περίοδο, οι ερευνητές συνειδητοποίησαν το πρόβλημα της σημασιολογικής απόκλισης και στη δεύτερη, ενδιάμεση, περίοδο διερεύνησαν διάφορες μεθόδους για να το αντιμετωπίσουν. Από το 2000 έως το 2008, διεξήχθη εκτεταμένη έρευνα για να αντιμετωπιστεί το πρόβλημα του σημασιολογικού κενού. Μια λεπτομερής μελέτη αυτών των τεχνικών παρουσιάζεται στη μελέτη των Datta et al. (2008).

Στις περισσότερες δημοσιευμένες μελέτες χρησιμοποιούνται χειροκίνητα εξαγόμενα χαρακτηριστικά, τα οποία όμως δίνουν καλύτερα αποτελέσματα και αντιμετωπίζουν αποτελεσματικά το σημασιολογικό κενό (Wang et al.,2008). Πολλές state-of-the-art τεχνικές αλλά και βασικές προσεγγίσεις αναπτύσσονται κατά τη διάρκεια αυτής της περιόδου. Το ερευνητικό ενδιαφέρον επικεντρώνεται κυρίως στην ανίχνευση της συσχέτισης μεταξύ οπτικών και λεκτικών χαρακτηριστικών.

Το χρονικό διάστημα μεταξύ 2002 και 2008 χαρακτηρίζεται ως η ενδιάμεση περίοδος της έρευνας και ανάπτυξης σχετικά με τις τεχνικές επισημείωσης και ανάκτησης εικόνων. Σε αυτή την ενδιάμεση περίοδο, οι τεχνικές μηχανικής μάθησης χρησιμοποιούνται εκτεταμένα. Ως εκ τούτου, αυτή η περίοδος είναι αφιερωμένη στη χρήση τεχνικών μηχανικής μάθησης για την επισημείωση και ανάκτηση εικόνας (Bhagat and Choudhary, 2018). Οι Datta et al. (2008) εξέτασαν μερικές από τις πιο ελπιδοφόρες σχετικές εργασίες που πραγματοποιήθηκαν κατά την ενδιάμεση περίοδο και έθεσαν τις μελλοντικές κατευθύνσεις της έρευνας στο πεδίο της επισημείωσης και ανάκτησης εικόνων. Κατά την πρώτη και δεύτερη δεκαετία, σχεδόν όλες οι μέθοδοι επισημείωσης εικόνων βασίζονταν στην εποπτευόμενη μάθηση, όπου το σύνολο των δεδομένων εκπαίδευσης παρέχεται πλήρως επισημειωμένο με ένα σύνολο ετικετών. Οι

περισσότερες από τις προτεινόμενες μεθόδους στην ενδιάμεση περίοδο ακολούθησαν την τεχνική ανάκτησης εικόνων με βάση το περιεχόμενο (Bhagat and Choudhary, 2018).

2.3.3. Τρίτη δεκαετία

Μετά το 2010, τεχνικές βαθιάς μηχανικής μάθησης (deep learning) χρησιμοποιούνται εκτενώς για τη διαδικασία επισημείωσης και ανάκτησης. Η χρήση χαρακτηριστικών βασισμένων σε συνελικτικά νευρωνικά δίκτυα (CNN) και τα χαρακτηριστικά που εξάγονται από προ-εκπαιδευμένα δίκτυα όπως τα AlexNet (Krizhevsky et al., 2012) και VGGNet που βασίζονται στα CNN, χρησιμοποιούνται στην αυτόματη επισημείωση εικόνας.

Ο περιορισμός της ανάκτησης εικόνας βάσει περιεχομένου (CBIR) οδήγησε την έρευνα στην ανάκτηση εικόνων βάσει κειμένου (TBIR). Στην ανάκτηση εικόνων βάσει κειμένου (TBIR) χρησιμοποιούνται σημασιολογικές λέξεις-κλειδιά για την ανάκτηση των εικόνων. Η TBIR απαιτεί να έχει προηγηθεί επισημείωση της εικόνας με σημασιολογικές λέξεις-κλειδιά. Επίσης, εξαιτίας του περιορισμού της εποπτευόμενης μάθησης η οποία απαιτεί μεγάλο αριθμό επισημειωμένων δεδομένων εκπαίδευσης, οι ερευνητές διερευνούν πλέον μεθόδους μηχανικής μάθησης με μερική επίβλεψη (SSL) και χωρίς επίβλεψη. Μετά το 2010, το ενδιαφέρον των ερευνητών μετατοπίστηκε προς την κατεύθυνση ημι-επιβλεπόμενων μεθόδων επισημείωσης εικόνας, όπου το μοντέλο εκπαιδεύεται χρησιμοποιώντας ελλιπείς ετικέτες σε δεδομένα εκπαίδευσης για επισημείωση εικόνας μεγάλης κλίμακας. Σε μεθόδους επισημείωσης μεγάλης κλίμακας με βάση την ημι-επιβλεπόμενη μάθηση (semi-supervised learning - SSL), η πραγματική πρόκληση είναι να αντιμετωπιστεί το σύνολο δεδομένων το οποίο περιέχει θόρυβο και στο οποίο ο αριθμός των εικόνων αυξάνεται διαρκώς. Πρόσφατα, οι ερευνητές έχουν αρχίσει να αναζητούν τεχνικές επισημείωσης εικόνων χωρίς επίβλεψη, όπου το σύνολο δεδομένων εκπαίδευσης δεν έχει επισημανθεί καθόλου και μόνο μεταδεδομένα (URL, κείμενα που πλαισιώνουν τις εικόνες κ.λπ.) συνοδεύουν το σύνολο των δεδομένων εκπαίδευσης. Οι μέθοδοι επισημείωσης που βασίζονται στη μάθηση χωρίς επίβλεψη, βρίσκονται σε πρώιμο στάδιο εξέλιξης (Bhagat and Choudhary, 2018).

2.3.4. Σύνοψη

Το πρόβλημα του σημασιολογικού κενού που εντοπίστηκε ήδη από τα πρώτα χρόνια, περιορίστηκε ή σχεδόν επιλύθηκε κατά τη δεύτερη δεκαετία σε ένα πλαίσιο επιβλεπόμενης μάθησης. Με την πάροδο των ετών, παρατηρείται μια μετατόπιση από τις τεχνικές

χειροκίνητης εξαγωγής χαρακτηριστικών στη συσχέτιση μεταξύ των λεκτικών και οπτικών χαρακτηριστικών και στην εξαγωγή χαρακτηριστικών με βάση μεθόδους βαθιάς μάθησης. Παράλληλα, σημειώνεται μία μετατόπιση από τη CBIR στη TBIR. Το πρόβλημα ωστόσο, που καλούνται να αντιμετωπίσουν οι ερευνητές είναι αυτό της μεγάλης κλίμακας των βάσεων δεδομένων εικόνων, το οποίο συνεχίζει να αυξάνεται. Σήμερα, οι ερευνητές προσπαθούν να διερευνήσουν τη δυνατότητα του SSL να αντιμετωπίσει το αυξανόμενο μέγεθος του συνόλου δεδομένων.

Η πρόσφατη πρόοδος στη βαθιά μάθηση και η τεράστια επιτυχία που σημείωσαν τα συνελικτικά νευρωνικά δίκτυα (ConvNets) στο διαγωνισμό ImageNet το 2012, έχει επιφέρει επανάσταση στη μηχανική όραση. Τα ConvNets αποτελούν πλέον την κυρίαρχη προσέγγιση για όλες σχεδόν τις εργασίες αναγνώρισης και ανίχνευσης προσεγγίζοντας τις ανθρώπινες επιδόσεις. Ενώ η εκπαίδευση τέτοιων μεγάλων δικτύων θα μπορούσε να διαρκέσει εβδομάδες μόνο πριν από λίγα χρόνια, η πρόοδος στο υλικό των υπολογιστικών συστημάτων, το λογισμικό και η παραλληλοποίηση του αλγορίθμου εκπαίδευσης έχουν μειώσει τους χρόνους κατάρτισης σε λίγες μόνο ώρες (LeCun et al., 2015). Οι εξελίξεις αυτές δίνουν τη δυνατότητα κατάργησης της χειροποίητης εξαγωγής χαρακτηριστικών και της αξιοποίησης χαρακτηριστικών που εξάγονται από τα συνελικτικά δίκτυα.

Μία συνοπτική παρουσίαση των μεθόδων επισημείωσης εικόνων τις τελευταίες τρεις δεκαετίες παρουσιάζεται στον Πίνακα 1.

	Πρώιμη περίοδος	Ενδιάμεση περίοδος	Τελευταία δεκαετία
Εξαγωγή χαρακτηριστικών	Χειροκίνητα χαρακτηριστικά	Συσχέτιση οπτικών και λεκτικών χαρακτηριστικών	Συσχέτιση οπτικών και λεκτικών χαρακτηριστικών, χαρακτηριστικά βαθιάς μάθησης
Μάθηση	Επιβλεπόμενη Μάθηση	Επιβλεπόμενη Μάθηση[1]	Ημι-επιβλεπόμενη Μάθηση[1], Μάθηση χωρίς επίβλεψη[2]
Ανάκτηση	CBIR	CBIR	TBIR
Πρόβλημα που αντιμετωπίστηκε	Σημασιολογικό κενό	Αναζήτηση βάσει εικόνας	Μάθηση από ανοργάνωτα και θορυβώδη δεδομένα κειμένου

[1] Οι περισσότερες από τις μεθόδους ακολούθησαν μόνο αυτή την προσέγγιση.

[2] Λίγες από τις μεθόδους ξεκίνησαν να ακολουθούν αυτή την προσέγγιση.

Πίνακας 1. Μερικά από τα κυριότερα χαρακτηριστικά των μεθόδων επισημείωσης εικόνων τις τελευταίες τρεις δεκαετίες (Bhagat and Choudhary, 2018).

Κεφάλαιο 3
Εξαγωγή χαρακτηριστικών και αναπαράσταση εικόνων

Το πρώτο βήμα για την αυτόματη επισημείωση της εικόνας είναι η αναπαράσταση του περιεχομένου της με βάση οπτικά χαρακτηριστικά που εξάγονται από αυτήν. Στο κεφάλαιο αυτό γίνεται μία ανασκόπηση των μεθόδων εξαγωγής οπτικών χαρακτηριστικών και αναπαράστασης της εικόνας που ακολουθούνται στα συστήματα αυτόματης επισημείωσης και ανάκτησης εικόνας.

3.1. Χαρακτηριστικά αναπαράστασης εικόνας

Η Αυτόματη επισημείωση εικόνας είναι μια τεχνική που εκχωρεί αυτόματα ένα σύνολο γλωσσικών όρων στις εικόνες προκειμένου να τις κατηγοριοποιήσει σημασιολογικά παρέχοντας το μέσο για την αποτελεσματική πρόσβαση σε βάσεις δεδομένων – αποθετήρια – εικόνων (Deselaers et al., 2007).

Στα τυπικά συστήματα ανάκτησης πληροφοριών βάσει περιεχομένου (CBIR), είναι πάντα σημαντικό να επιλεγεί μια κατάλληλη αναπαράσταση των αρχείων. Η ποιότητα των αποτελεσμάτων της ανάκτησης εξαρτάται από την ποιότητα της εσωτερικής αναπαράστασης του περιεχομένου τους (Martinet and Elsayad, 2012). Τα κλασικά μοντέλα ανάκτησης πληροφοριών θεωρούν συνήθως ότι ένα αρχείο περιγράφεται από ένα σύνολο χαρακτηριστικών (features) ή περιγραφέων (descriptors). Στην ανάκτηση αρχείων κειμένου για παράδειγμα, οι περιγραφείς παίρνουν τη μορφή αντιπροσωπευτικών όρων ευρετηρίου, που είναι λέξεις-κλειδιά που εξάγονται από τη συλλογή κειμένων. Στην ανάκτηση εικόνων επειδή μία εικόνα είναι μια μη δομημένη σειρά εικονοστοιχείων, το πρώτο βήμα της σημασιολογικής κατανόησης είναι να σχεδιαστεί ένας εξαγωγέας χαρακτηριστικών που θα μεταμορφώνει τα ακατέργαστα δεδομένα (όπως οι τιμές των εικονοστοιχείων της εικόνας) σε μια κατάλληλη εσωτερική αναπαράσταση ή ένα διάνυσμα χαρακτηριστικών, από το οποίο το υποσύστημα μάθησης, συχνά ένας ταξινομητής, θα μπορούσε να ανιχνεύσει ή να ταξινομήσει πρότυπα στα δεδομένα εισόδου (LeCun et al., 2015). Η κατάλληλη

αναπαράσταση της εικόνας βάσει οπτικών χαρακτηριστικών βελτιώνει σημαντικά την απόδοση των τεχνικών σημασιολογικής μάθησης.

Τα οπτικά χαρακτηριστικά, που ονομάζονται επίσης χαρακτηριστικά χαμηλού επιπέδου, προέρχονται αντικειμενικά από τις εικόνες και όχι από εξωτερική σημασιολογία. Δεδομένου ότι τα χαρακτηριστικά που εξάγονται από τις εικόνες, πρέπει να έχουν νόημα για το άτομο που αναζητά μία εικόνα, τα οπτικά χαρακτηριστικά που χρησιμοποιούνται στα συστήματα ανάκτησης και επισημείωσης εικόνων χωρίζονται κυρίως σε τρεις ομάδες: χρώμα, σχήμα και υφή (Tsai and Hung, 2008).

Κατά την εξέταση εικόνων προκύπτει το πρόβλημα της σημασιολογικής απόκλισης (Smeulders et al., 2000). Στην περίπτωση των εικόνων, εξαιτίας της απόστασης μεταξύ του ακατέργαστου σήματος, δηλαδή, του πίνακα των εικονοστοιχείων της εικόνας και της ερμηνείας τους, είναι δύσκολο να εξάγεται αυτόματα μια ακριβής σημασιολογική αναπαράσταση του περιεχομένου τους. Οι παραδοσιακές τεχνικές αυτόματης επισημείωσης προσπαθούν να συσχετίσουν τα χαρακτηριστικά χαμηλού επιπέδου (ή περιγραφείς[1] χαμηλού επιπέδου) με σημασιολογικές πληροφορίες (ή περιγραφείς υψηλού επιπέδου).

3.1.1. Χαρακτηριστικά χαμηλού επιπέδου

Ένας περιγραφέας χαμηλού επιπέδου είναι μια συνεχής ή διακριτή αριθμητική ή συμβολική μέτρηση που υπολογίζεται απευθείας από το σήμα (εικονοστοιχεία εικόνας), σε ολόκληρη ή μέρος μίας εικόνας. Οι περιγραφείς χαμηλού επιπέδου περιλαμβάνουν συνήθως το χρώμα, την υφή και το σχήμα. Υποστηρίζουν μετρήσεις που εκτελούνται απευθείας από το σήμα, με ευθύ τρόπο, χωρίς εξωτερική γνώση - διαδικασία μάθησης - ούτε συνολική στατιστική ανάλυση άλλων αρχείων. Αντιπροσωπευτικά παραδείγματα αποτελούν, ο δυαδικός τελεστής τοπικού προτύπου (Local Binary Pattern - LBP) (Ojala et al., 2002) και ο αμετάβλητος στην κλιμάκωση μετασχηματισμός χαρακτηριστικών (Scale-Invariant Feature Transform - SIFT) (Lowe, 2004).

Ένας περιγραφέας χαμηλού επιπέδου ορίζεται για τη λήψη μιας συγκεκριμένης οπτικής ιδιότητας μιας εικόνας, είτε συνολικά, για να καταγράψει τα γενικά χαρακτηριστικά της

[1] Σημειώνεται ότι και οι δύο όροι (περιγραφείς και χαρακτηριστικά) είναι παρόμοιοι και μπορούν να χρησιμοποιηθούν εναλλακτικά.

εικόνας, είτε τοπικά για μια συγκεκριμένη ομάδα εικονοστοιχείων που αντιστοιχούν σε μία περιοχή ενδιαφέροντος ή σε κάποιο αντικείμενο της εικόνας. Τα πιο συχνά χρησιμοποιούμενα χαρακτηριστικά περιλαμβάνουν αυτά που αντικατοπτρίζουν το χρώμα, την υφή, το σχήμα και τα σημεία ενδιαφέροντος (salient / interest points) σε μια εικόνα.

3.1.2. Χαρακτηριστικά υψηλού επιπέδου

Ένας περιγραφέας υψηλού επιπέδου είναι ένα κομμάτι ανθρώπινα ερμηνευόμενων σημασιολογικών πληροφοριών που περιγράφουν μία εικόνα (ή μέρος μίας εικόνας). Οι περιγραφείς υψηλού επιπέδου αντιπροσωπεύουν τη σημασιολογία της εικόνας, όπως ένα σύνολο λέξεων-κλειδιών ή μια περιγραφή κειμένου. Αντιπροσωπεύουν τον τελικό στόχο της επισημείωσης, της ευρετηρίασης, της ανίχνευσης ιδεών υψηλού επιπέδου ή γενικότερα της αυτόματης δημιουργίας σημασιολογικών περιγραφέων.

Οι περισσότερες από τις μεθόδους αυτόματης επισημείωσης σκοπεύουν να κάνουν το σύστημα να μάθει ένα μοντέλο αντιστοιχίας μεταξύ χαρακτηριστικών χαμηλού και υψηλού επιπέδου. Μόλις μάθει αυτό το μοντέλο αντιστοιχίας, το σύστημα είναι σε θέση να παράγει χαρακτηριστικά υψηλού επιπέδου από ένα δεδομένο σύνολο χαρακτηριστικών χαμηλού επιπέδου, δηλαδή το σύστημα είναι σε θέση να εξαγάγει τη σημασιολογία από τα χαρακτηριστικά που εξάγονται απευθείας από το σήμα. Για παράδειγμα, το αποτέλεσμα ενός ταξινομητή εικόνας που κατηγοριοποιεί εικόνες ως εσωτερικές (indoor) και εξωτερικές (outdoor), θεωρείται ένας περιγραφέας υψηλού επιπέδου.

3.2. Οπτικά χαρακτηριστικά εικόνας

Το οπτικό περιεχόμενο μιας εικόνας μπορεί να εκπροσωπείται από συνολικά (global) ή τοπικά (local) χαρακτηριστικά. Τα συνολικά χαρακτηριστικά λαμβάνουν υπόψη όλα τα εικονοστοιχεία μιας εικόνας. Τα ιστογράμματα χρωμάτων, για παράδειγμα, μπορούν να εξαχθούν για να αναπαραστήσουν ή να περιγράψουν το συνολικό χρωματικό περιεχόμενο των εικόνων. Ωστόσο, καθώς τα συνολικά χαρακτηριστικά λαμβάνουν υπόψη τα οπτικά χαρακτηριστικά ολόκληρης της εικόνας, δεν μπορούν να περιγράψουν διαφορετικά μέρη μιας εικόνας. Από την άλλη πλευρά, η τμηματοποίηση της εικόνας σε τοπικό περιεχόμενο, σε διαφορετικές περιοχές ή τμήματα, είναι σε θέση να παρέχει πιο λεπτομερείς πληροφορίες για τις εικόνες (Tsai and Hung, 2008).

Αν και στις υπάρχουσες τεχνικές επισημείωσης εικόνας χρησιμοποιούνται τόσο χαρακτηριστικά που αντιπροσωπεύουν συνολικά την εικόνα όσο και χαρακτηριστικά που αντιπροσωπεύουν περιοχές της, η τάση είναι να χρησιμοποιούνται χαρακτηριστικά που βασίζονται σε περιοχές (Zhang et al., 2012).

Η ενότητα που ακολουθεί παρέχει μια σύντομη ανασκόπηση των αλγορίθμων τμηματοποίησης που χρησιμοποιούνται συνήθως στην αυτόματη επισημείωση εικόνας.

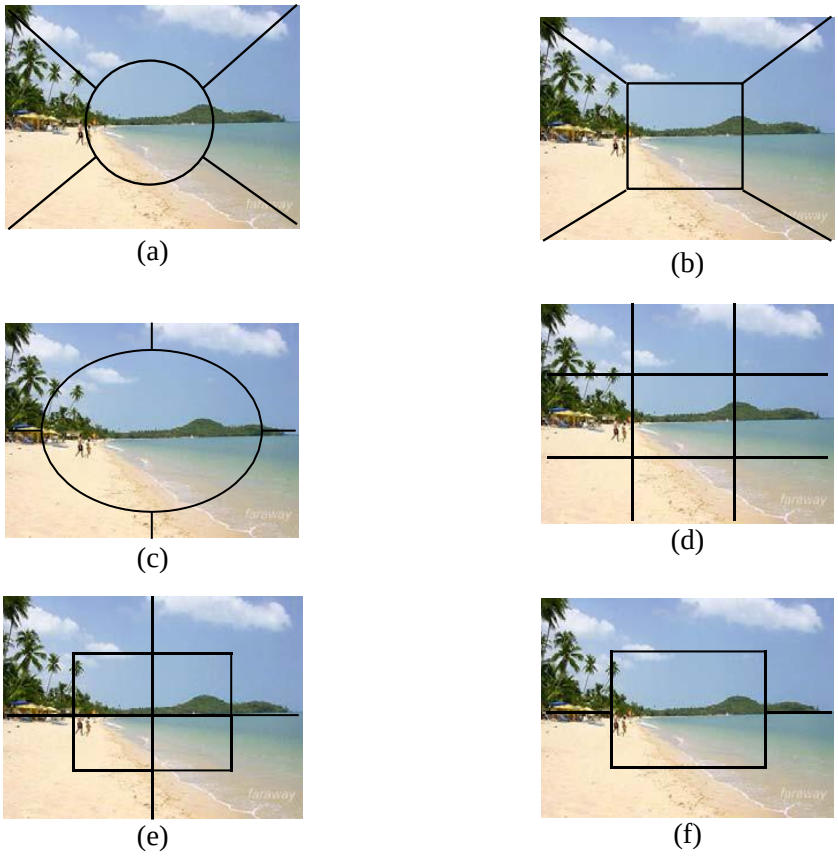
3.2.1. Τμηματοποίηση εικόνας

Η τμηματοποίηση της εικόνας είναι συνήθως το πρώτο βήμα για την ανάλυση της εικόνας. Ένας αλγόριθμος τμηματοποίησης διαχωρίζει την εικόνα σε διαφορετικά τμήματα με βάση την ομοιογένεια χαρακτηριστικών. Υπάρχουν πολλές προσεγγίσεις τμηματοποίησης στη βιβλιογραφία, όπως μέθοδοι που βασίζονται στο πλέγμα (grid based), στις ομάδες (clustering based), στα περιγράμματα (contour based), σε στατιστικά μοντέλα (model based), στους γράφους (graph based) και στην σταδιακή επέκταση περιοχών (region growing based) (Zhang et al., 2012).

3.2.1.1. Τμηματοποίηση σε μπλοκ

Επειδή η αυτόματη τμηματοποίηση της εικόνας είναι μια δύσκολη εργασία, πολλές τεχνικές απλοποιούν αυτή την εργασία χρησιμοποιώντας την προσέγγιση με βάση το πλέγμα (grid-based ή block-based) για την κατά προσέγγιση διαίρεση των εικόνων σε ένα σύνολο τμημάτων – μπλοκ (blocks-tiles) σταθερού μεγέθους όπως φαίνεται στην εικόνα 5.

Μετά την ολοκλήρωση της τμηματοποίησης σε μπλοκ, τα οπτικά χαρακτηριστικά μπορούν να εξαχθούν από τις περιοχές για αναπαράσταση τοπικών χαρακτηριστικών. Η προσέγγιση με βάση μπλοκ έχει το πλεονέκτημα ότι έχει μικρό υπολογιστικό κόστος. Ωστόσο, αυτή η απλή προσέγγιση αδυνατεί να αναπαραστήσει με ακρίβεια τις σημασιολογικές έννοιες (αντικείμενα) σε μία εικόνα. Ένα ενιαίο μπλοκ αποτελείται συχνά από τμήματα οπτικά διαφορετικών αντικειμένων. Επιπλέον, είναι δύσκολο να καθοριστεί το μέγεθος των μπλοκ για την αναπαράσταση της εικόνας. Επομένως, τα χαρακτηριστικά που εξάγονται από μια περιοχή συνήθως δεν είναι ακριβή. Με κατάλληλη εφαρμογή, μπορεί ωστόσο να χρησιμοποιηθεί σε συγκεκριμένες εφαρμογές όπως για παράδειγμα στην ταξινόμηση ιατρικής εικόνας.



Εικόνα 5. Παραδείγματα τμηματοποίησης εικόνας βάσει μπλοκ (Tsai and Hung, 2008).

3.2.1.2. Τμηματοποίηση εικόνας με ομαδοποίηση

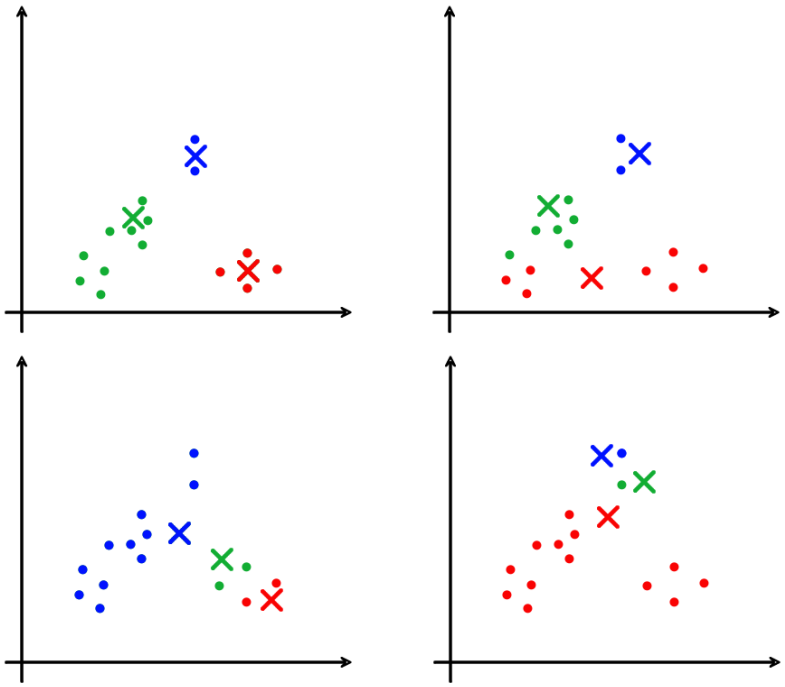
Αλγόριθμοι ομαδοποίησης (clustering-based), όπως ο k-means, χρησιμοποιούνται για την ομαδοποίηση εικονοστοιχείων σε διαφορετικές ομάδες, όπου κάθε ομάδα ταυτίζεται με μια περιοχή. Στις περισσότερες περιπτώσεις, μια εικόνα χωρίζεται πρώτα σε μπλοκ μεγέθους 4×4 pixels. Χαρακτηριστικά χρώματος και / ή υφής εξάγονται για κάθε μπλοκ. Στη συνέχεια, εφαρμόζεται ο αλγόριθμος k-means για την ομαδοποίηση των διανυσμάτων χαρακτηριστικών (Zhang et al., 2012).

Ο αλγόριθμος k-means ξεκινάει με k τυχαία σημεία, που ονομάζονται κεντροειδή της ομάδας και δηλώνουν το κέντρο βάρους της ομάδας. Το k δηλώνει το πλήθος των ομάδων που θα δημιουργήσει ο αλγόριθμος. Ο αλγόριθμος εκτελεί επαναληπτικά δύο βήματα. Το πρώτο αφορά την ανάθεση κάθε διανύσματος σε κάποια ομάδα. Με χρήση κάποιου μέτρου απόστασης, αναθέτει το εξεταζόμενο διάνυσμα στην ομάδα, της οποίας το κεντροειδές απέχει λιγότερο από το συγκεκριμένο διάνυσμα. Το δεύτερο βήμα αφορά τον επαναπροσδιορισμό και τη μετατόπιση του κεντροειδούς κάθε ομάδας. Με βάση τον μέσο όρο των διανυσμάτων κάθε ομάδας, επανυπολογίζονται τα κεντροειδή της κάθε ομάδας

ώστε το κεντροειδές να είναι πιο αντιπροσωπευτικό στην πρόσφατα διαμορφωμένη ομάδα. Ο αλγόριθμος εκτελεί επαναληπτικά αυτά τα δύο βήματα μέχρις ότου τα κεντροειδή των ομάδων να μετατοπίζονται ελάχιστα και σε απόσταση μικρότερη από κάποια δοθείσα τιμή κατωφλίου. Ως εναλλακτικό κριτήριο τερματισμού του αλγορίθμου μπορεί να χρησιμοποιηθεί και ο αριθμός επαναλήψεων του αλγορίθμου (Βερύκιος κ.ά., 2015).

Αρχικοποίησε τυχαία τα k κεντροειδή των ομάδων $\mu_1, \mu_2, \dots, \mu_k$.
Επανάλαβε
{
▪ Εξέτασε κάθε διάνυσμα και ανέθεσε το στην ομάδα με το πλησιέστερο κεντροειδές ($\boldsymbol{min}\left\|\boldsymbol{x}^{(i)} - \boldsymbol{\mu_k}\right\|^2$)
▪ Επανυπολόγισε τα κεντροειδή υπολογίζοντας το μέσο όρο των δειγμάτων της ομάδας
}

Πίνακας 2. Ο αλγόριθμος k-means (Βερύκιος κ.ά., 2015).



Εικόνα 6. Ο αλγόριθμος k-means. Τυχαία αρχικοποίηση κεντροειδών. Πάνω αριστερά έχουμε την καλύτερη περίπτωση. Ακολουθεί μια λιγότερο ποιοτικά καλή ομαδοποίηση (πάνω δεξιά). Στις δυο τελευταίες περιπτώσεις είναι προφανές ότι η αρχικοποίηση επηρεάζει αρνητικά τη διαδικασία ομαδοποίησης, καθώς οι δύο ομάδες περιέχουν πολύ λίγα δείγματα, ενώ η μία περιέχει όλα τα υπόλοιπα δείγματα (Βερύκιος κ.ά., 2015).

Μια περιοχή σχηματίζεται με τα εικονοστοιχεία που ανήκουν σε μπλοκ της ίδιας ομάδας. Το βασικό μειονεκτήματα του αλγορίθμου $k - means$ είναι ότι πρέπει να προκαθοριστεί ο

αριθμός των ομάδων, δηλαδή του αριθμού k, ευριστικά. Μια ακατάλληλη επιλογή του k μπορεί να έχει ως αποτέλεσμα κακή απόδοση του αλγορίθμου. Ένα άλλο μειονέκτημα είναι ότι ο αλγόριθμος υποθέτει ότι τα δεδομένα είναι σε σφαιρικές ομάδες, έτσι ώστε οι μέσες τιμές να βρίσκονται κοντά στα κέντρα των ομάδων. Αυτή η υπόθεση, ωστόσο, συνήθως δεν ισχύει (Zhang et al., 2012).

3.2.1.3. Τμηματοποίηση βάσει περιγράμματος

Η κύρια ιδέα της τμηματοποίησης βάσει περιγράμματος (contour based segmentation) είναι να αναπτυχθεί μια καμπύλη γύρω από ένα αντικείμενο. Η ανάπτυξη σταματά όταν η καμπύλη συμπέσει με το περίγραμμα ενός αντικειμένου (Zhu and Yuille, 1996, Chan and Vese, 2001). Σε αντίθεση με την ομαδοποίηση, οι αλγόριθμοι τμηματοποίησης που βασίζονται σε περιγράμματα, δεν χρειάζονται τον προκαθορισμό του αριθμού των ομάδων. Το βασικό πρόβλημα αυτής της προσέγγισης είναι η εξάρτησή της από την ακριβή ανίχνευση των ακμών που υπόκεινται σε θόρυβο. Ως εκ τούτου, χρειάζεται συχνά ο ανθρώπινος παράγοντας για τον ακριβή καθορισμό του περιγράμματος καθιστώντας την προσέγγιση εφαρμόσιμη μόνο σε συγκεκριμένους τομείς, όπως για παράδειγμα εργαλεία επεξεργασίας εικόνας (Zhang et al., 2012).

3.2.1.4. Τμηματοποίηση βάσει στατιστικών μοντέλων

Αλγόριθμοι τμηματοποίησης που βασίζονται σε στατιστικά μοντέλα έχουν επίσης προταθεί. Μεταξύ αυτών, ο αλγόριθμος Blobworld (Carson et al., 2002) που χρησιμοποιείται ευρέως. Στον αλγόριθμο αυτό, κάθε εικονοστοιχείο αντιπροσωπεύεται από ένα 8-διάστατο διάνυσμα χαρακτηριστικών χρώματος, υφής και θέσης. Ένα εικονοστοιχείο της εικόνας μοντελοποιείται ως τυχαία μεταβλητή που ακολουθεί την κανονική (Gaussian) κατανομή. Οι παράμετροι της κανονικής κατανομής υπολογίζονται χρησιμοποιώντας τον αλγόριθμο μεγιστοποίησης προσδοκίας (Expectation Maximization - EM). Μόλις εντοπιστούν οι παράμετροι του μοντέλου, υπολογίζεται η σχέση εικονοστοιχείου – περιοχής χρησιμοποιώντας τις εκ των υστέρων (a posteriori) πιθανότητες. Η σχέση εικονοστοιχείου – περιοχής χρησιμοποιείται για τμηματοποίηση της εικόνας. Ένα από τα βασικά ζητήματα της προσέγγισης αυτής είναι το υπολογιστικό κόστος καθώς ο EM είναι αλγόριθμος βελτιστοποίησης.

3.2.1.5. Τμηματοποίηση βάσει γράφων

Οι Shi and Malik (2000) προτείνουν έναν αλγόριθμο τμηματοποίησης βασισμένο σε τομές γράφων γνωστό ως αλγόριθμο κανονικοποιημένης τομής (NCut). Η μέθοδος NCut αναπαριστά μια εικόνα ως γράφο όπου οι κορυφές είναι εικονοστοιχεία της εικόνας και τα βάρη των ακμών αντιπροσωπεύουν τις ομοιότητες χαρακτηριστικών μεταξύ εικονοστοιχείων. Η τμηματοποίηση των εικόνων γίνεται, στη συνέχεια, ένα πρόβλημα διαμέρισης γράφων. Η ιδέα είναι να χωριστούν οι κορυφές του γραφήματος σε διαφορετικά σύνολα έτσι ώστε να ελαχιστοποιηθεί η συνολική ομοιότητα μεταξύ των διαφορετικών συνόλων. Κάθε σύνολο θεωρείται περιοχή. Δεδομένου ότι ο αριθμός των εικονοστοιχείων μιας εικόνας είναι μεγάλος, ο αριθμός των πιθανών διαμερίσεων του γράφου είναι εκθετικός. Ως αποτέλεσμα, είναι δαπανηρό να υπολογιστεί η βέλτιστη διαμέριση. Οι Tao et al. (2007) βελτιώνουν τον NCut με προ-τμηματοποίηση εικόνων. Αντί της χρήσης εικονοστοιχείων, οι περιοχές της αρχικής τμηματοποίησης χρησιμοποιούνται ως κορυφές στον αλγόριθμο NCut. Ως εκ τούτου, το υπολογιστικό κόστος μειώνεται και η απόδοση αυξάνεται. Ο βασικός αλγόριθμος NCut βασίζεται μόνο σε χαρακτηριστικά χρώματος. Οι Malik et al. (2001) τον επεκτείνουν ώστε να ενσωματώνει χαρακτηριστικά υφής.

3.2.1.6. Τμηματοποίηση με επέκταση περιοχής

Ο ευρέως χρησιμοποιούμενος αλγόριθμος JSEG (Deng and Manjunath, 2001) είναι μια μέθοδος σταδιακής επέκτασης περιοχών (region growing). Ομαδοποιεί εικονοστοιχεία ή μικρότερες περιοχές σε μεγαλύτερες περιοχές. Αρχικά, τα χρώματα των εικονοστοιχείων της εικόνας ποσοτικοποιούνται σε έναν αριθμό κλάσεων και τα εικονοστοιχεία της εικόνας αντικαθίστανται με τις ετικέτες κλάσης του χρώματος. Καταρτίζεται ένας πίνακας των κλάσεων και ακολουθείται η ανάπτυξη της περιοχής στον χάρτη κλάσεων. Τα εικονοστοιχεία με πιο ομογενείς γείτονες υποτίθεται ότι είναι εικονοστοιχεία εσωτερικού χώρου πιθανών περιοχών. Αυτά τα εικονοστοιχεία επιλέγονται ως υποψήφια σημεία σπόροι και οι περιοχές αναπτύσσονται γύρω από αυτές τις περιοχές σπόρους. Επειδή αυτή η μέθοδος αναζητά ομοιογένεια χρώματος και υφής, οι κατά τμήματα περιοχές έχουν πολύ ομοιογενή χαρακτηριστικά. Η προσέγγιση αυτή έχει χρησιμοποιηθεί ευρέως στην ανάκτηση εικόνων.

3.2.1.7. Σύνοψη

Η ακρίβεια του αλγορίθμου τμηματοποίησης διαδραματίζει κρίσιμο ρόλο στη σημασιολογική επισημείωση της εικόνας. Η σημασιολογική τμηματοποίηση απαιτεί το κάθε εικονοστοιχείο

να πάρει ετικέτα από μια κλάση αντικειμένων, και μπορεί να εφαρμοστεί με επιβλεπόμενο τρόπο ή με μία μέθοδο αδύναμης επίβλεψης ή με μη-επιβλεπόμενο τρόπο. Πρόσφατες μελέτες αναφέρουν τη χρήση βαθιών συνελικτικών δικτύων για την σημασιολογική τμηματοποίηση αντικειμένων (Bhagat and Choudhary, 2018).

Έχουν αναφερθεί διάφορες μέθοδοι επισημείωσης εικόνας που χρησιμοποιούν τεχνικές τμηματοποίησης για την αναγνώριση αντικειμένων. Η ισχυρή τμηματοποίηση είναι απαραίτητη για να επισημειωθεί σημασιολογικά η εικόνα, όμως δύσκολα μπορεί να επιτευχθεί. Ως εκ τούτου, αρκετοί ερευνητές προσπάθησαν να εφαρμόσουν αδύναμη τμηματοποίηση. Εναλλακτικά, η τμηματοποίηση μπορεί να παραλειφθεί και άλλοι τύποι χαρακτηριστικών μπορούν να χρησιμοποιηθούν για επισημείωση εικόνας (Bhagat and Choudhary, 2018).

3.2.2. Χαρακτηριστικά χρώματος

Από τα διάφορα οπτικά χαρακτηριστικά μιας εικόνας, το χρώμα είναι το πιο απλό χαρακτηριστικό. Το χρώμα αποτελεί ένα από τα πιο σημαντικά χαρακτηριστικά των εικόνων, καθώς το ανθρώπινο μάτι είναι ευαίσθητο στα χρώματα. Το χρώμα, είναι το πλέον χρησιμοποιούμενο οπτικό χαρακτηριστικό για την ανάκτηση εικόνων βάσει περιεχομένου, λόγω της υπολογιστικής αποδοτικότητας της εξαγωγής του που απλοποιεί την αναγνώριση αντικειμένων (Gonzalez and Woods, 2002). Το χρώμα θεωρείται ισχυρός περιγραφέας καθώς τα χαρακτηριστικά που εξάγονται από αυτό, είναι αμετάβλητα στην περιστροφή και τον μετασχηματισμό. Εάν χρησιμοποιείται κανονικοποίηση, τότε τα χαρακτηριστικά χρώματος είναι επίσης αμετάβλητα στην κλιμάκωση (Bhagat and Choudhary, 2018).

Τα χαρακτηριστικά χρώματος ορίζονται με βάση ένα συγκεκριμένο χρωματικό χώρο ή μοντέλο. Όλα τα χρώματα μπορούν να αναπαρασταθούν με μεταβλητούς συνδυασμούς των τριών επονομαζόμενων βασικών χρωμάτων: κόκκινο (R), πράσινο (G) και μπλε (B). Υπάρχουν μερικοί άλλοι χρωματικοί χώροι για την απεικόνιση του χαρακτηριστικού του χρώματος όπως οι HSV, L*u*v*, YIQ, HMMD, κ.λπ. (Gonzalez and Woods, 2002).

Μόλις καθοριστεί ο χρωματικός χώρος, το χρώμα εξάγεται από εικόνες ή περιοχές. Στην βιβλιογραφία έχουν προταθεί διάφοροι περιγραφικοί δείκτες χρώματος, όπως τα ιστογράμματα χρώματος, οι χρωματικές ροπές (Yu et al., 2002), το διάγραμμα χρωματικής

συσχέτισης (Huang et al., 1997), το διάνυσμα χρωματικής συνοχής (Color Coherence Vector) (Pass and Zabith, 1996) κ.ά. Το πρότυπο MPEG-7 (Manjunath et al., 2002) τυποποιεί επίσης ορισμένα χαρακτηριστικά χρώματος, όπως τον κυρίαρχο περιγραφέα χρώματος (DCD), τον περιγραφέα δομής χρώματος (CSD) και τον κλιμακωτό περιγραφέα χρωμάτος (SCD).

Το ιστόγραμμα χρώματος το οποίο αντιπροσωπεύει την κατανομή του αριθμού των εικονοστοιχείων σε κάθε κβαντισμένη τιμή χρώματος (bin), είναι μια αποτελεσματική αναπαράσταση του περιεχομένου χρώματος μιας εικόνας που χρησιμοποιείται συχνά για ευρετηρίαση και ανάκτηση εικόνας (Tsai and Hung, 2008). Το ιστόγραμμα χωρίζει το χρωματικό χώρο της εικόνας, το σύνολο όλων των πιθανών χρωμάτων, σε διαφορετικές περιοχές τιμών που καλύπτουν το χρωματικό χώρο και μετράει τη συχνότητα εμφάνισης της κάθε περιοχής τιμών, δηλαδή των αριθμό των εικονοστοιχείων που το χρώμα τους ανήκει στη συγκεκριμένη περιοχή τιμών (bin) (Zhang et al., 2012).

Το χρωματικό ιστόγραμμα είναι ένα ισχυρό χαρακτηριστικό καθώς δεν επηρεάζεται από αλλαγές όπως η περιστροφή και η μετατόπιση γύρω από τον άξονα προβολής και μπορεί να χαρακτηρίσει τη συνολική και τοπική κατανομή των χρωμάτων σε μια εικόνα (Wei and Chen, 2012). Ωστόσο, ένα χρωματικό ιστόγραμμα δεν περιέχει χωρικές πληροφορίες των εικονοστοιχείων. Συνεπώς, οπτικά διαφορετικές εικόνες μπορούν να έχουν παρόμοια χρωματικά ιστογράμματα. Επιπλέον, η διάσταση ενός ιστογράμματος είναι συνήθως πολύ μεγάλη (Zhang et al., 2012).

Οι χρωματικές ροπές (Color Moments) είναι ένα από τα απλούστερα χαρακτηριστικά χρώματος. Χρησιμοποιούνται σε πολλά συστήματα ανάκτησης. Η βάση των χρωματικών ροπών στηρίζεται στην υπόθεση ότι η κατανομή του χρώματος σε μια εικόνα μπορεί να ερμηνευτεί ως κατανομή πιθανοτήτων. Οι κατανομές πιθανοτήτων χαρακτηρίζονται από έναν αριθμό μοναδικών μέτρων (π.χ. μέτρα θέσης, όπως η μέση τιμή και μέτρα μεταβλητότητας, όπως η διασπορά). Συνεπώς, εάν το χρώμα μιας εικόνας ακολουθεί μια ορισμένη κατανομή πιθανοτήτων, τα μεγέθη (ροπές) αυτής της κατανομής μπορούν στη συνέχεια να χρησιμοποιηθούν ως χαρακτηριστικά για την αναγνώριση αυτής της εικόνας με βάση το χρώμα. Οι πιο κοινές ροπές είναι η μέση τιμή (Mean), η τυπική απόκλιση (Standard deviation) και η λοξότητα ή ασυμμετρία (Skewness). Συνήθως υπολογίζονται ξεχωριστά για κάθε χρωματικό κανάλι. Επομένως, εννέα χαρακτηριστικά αποτελούν το διάνυσμα

χαρακτηριστικών (Stricker and Orengo, 1995). Αυτές οι λειτουργίες είναι χρήσιμες όταν υπολογίζονται για μία περιοχή ή ένα αντικείμενο της εικόνας. Η μέθοδος των χρωματικών ροπών έχει το χαμηλότερης διάστασης διάνυσμα χαρακτηριστικών και τη χαμηλότερη υπολογιστική πολυπλοκότητα. Ως εκ τούτου, είναι η πιο κατάλληλη για ανάκτηση εικόνων. Ωστόσο, οι χρωματικές στιγμές δεν αρκούν για να αντιπροσωπεύσουν όλες τις πληροφορίες χρώματος μιας εικόνας (Zhang et al., 2012).

Μέση τιμή (Mean)	*Τυπική Απόκλιση (Standard Deviation)*	*Ασυμμετρία ή Λοξότητα (Skewness)*
$\boldsymbol{E_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij}}$	$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^2}$	$s_i = \sqrt[3]{\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^3}$
Η μέση τιμή μπορεί να γίνει κατανοητή ως η μέση χρωματική τιμή στην εικόνα	Η τυπική απόκλιση είναι η τετραγωνική ρίζα της διακύμανσης της κατανομής.	Πόσο και προς ποια κατεύθυνση αποκλίνει η κατανομή από την πλήρη συμμετρία (skewness=0)

όπου $\boldsymbol{p_{ij}}$ η τιμή του του j-οστού εικονοστοιχείου της εικόνας στο i-οστό χρωματικό κανάλι.

Πίνακας 3. Οι τρεις πιο κοινές χρωματικές ροπές (Stricker and Orengo, 1995).

Το διάνυσμα χρωματικής συνοχής (Color Coherence Vector - CCV) ενσωματώνει χωρικές πληροφορίες στο βασικό χρωματικό ιστόγραμμα. Διαχωρίζει κάθε (περιοχή τιμών) χρωματικό εύρος του ιστογράμματος σε δύο τμήματα: συνεκτικά και μη συνεκτικά μέρη. Το συνεκτικό τμήμα περιλαμβάνει τα εικονοστοιχεία που είναι χωρικά συνδεδεμένα. Το μη συνεκτικό τμήμα περιλαμβάνει τα απομονωμένα εικονοστοιχεία (Pass and Zabith,1996). Καθώς το CCV συλλαμβάνει χωρικές πληροφορίες, συνήθως αποδίδει καλύτερα από ένα χρωματικό ιστόγραμμα. Ωστόσο, η διάσταση ενός CCV είναι διπλάσια από ένα συμβατικό ιστόγραμμα (Zhang et al., 2012).

Ένα διάγραμμα χρωματικής συσχέτισης (Color correlogram) είναι η έγχρωμη εκδοχή της μήτρας συν-εμφάνισης σε επίπεδο γκρίζου (grey-level co-occurrence matrix). Χαρακτηρίζει την κατανομή των ζευγών χρωμάτων σε μια εικόνα. Ένα έγχρωμο διάγραμμα συσχέτισης μπορεί να αντιμετωπιστεί ως τρισδιάστατο ιστόγραμμα όπου οι πρώτες δύο διαστάσεις αντιπροσωπεύουν τα χρώματα κάθε ζεύγους εικονοστοιχείων και η τρίτη διάσταση είναι η χωρική απόστασή τους (Huang et al., 1997). Έτσι, σε ένα διάγραμμα συσχέτισης, κάθε περιοχή τιμών (i, j, k) αντιπροσωπεύει τον αριθμό ζεύγους χρωμάτων (i, j) σε απόσταση k.

Το χρωματικό διάγραμμα συσχέτισης υπολογίζεται για οριζόντια απόσταση *k = 1*. Τα διαγράμματα συσχέτισης μπορούν επίσης να υπολογιστούν για άλλες αποστάσεις. Η απόδοση του χρωματικού διαγράμματος συσχέτισης είναι καλύτερη τόσο από το ιστόγραμμα όσο και από το CCV, επειδή συμπεριλαμβάνει τόσο τα επίπεδα έντασης όσο και τα χωρικά πρότυπα σε μια εικόνα. Ωστόσο, είναι πολύ πιο περίπλοκο λόγω της μεγάλης διαστασιμότητας και των πολλαπλών επεξεργασιών των πινάκων (Zhang et al., 2012).

Μεταξύ των περιγραφέων χρώματος του προτύπου MPEG-7, ο κλιμακωτός περιγραφέας χρώματος (Scalable Color Descriptor – SCD) είναι ένας περιγραφέας που βασίζεται στο ιστόγραμμα. Ο SCD είναι βασικά ένα ιστόγραμμα στο χρωματικό χώρο HSV. Διαφέρει από το συμβατικό ιστόγραμμα λόγω της δυνατότητας κλιμάκωσης. Η κλιμάκωση επιτυγχάνεται με δύο τρόπους: τη μείωση του αριθμού των περιοχών τιμής χρώματος με μετασχηματισμό Haar και την αφαίρεση μερικών λιγότερο σημαντικών δυαδικών ψηφίων από τις ποσοτικοποιημένες (ακέραιες) αναπαραστάσεις των τιμών των περιοχών. Ωστόσο, τα πειραματικά αποτελέσματα δείχνουν ότι μια τέτοια μείωση της κλίμακας επηρεάζει σημαντικά την απόδοση ανάκτησης (Manjunath et al., 2002). Επιπλέον, ο περιγραφέας SCD δεν περιλαμβάνει καμία χωρική πληροφορία. Επομένως, παρουσιάζει παρόμοια προβλήματα με το συμβατικό ιστόγραμμα.

Ο περιγραφέας δομής χρώματος (Color Structure Descriptor – CSD) είναι επίσης ένας περιγραφέας βάσει ιστογράμματος (Manjunath et al., 2002). Το ιστόγραμμα CSD δημιουργείται μετακινώντας ένα δομικό στοιχείο (π.χ. τετραγωνικό παράθυρο) σε όλη την εικόνα. Η περιοχή τιμών i του CSD ιστογράμματος δείχνει πόσες φορές το παράθυρο περιέχει τουλάχιστον ένα εικονοστοιχείο με χρώμα i. Εάν το παράθυρο έχει μέγεθος 1 pixel, το CSD είναι ένα συνηθισμένο ιστόγραμμα. Η απόδοση του CSD εξαρτάται από το μέγεθος και τη δομή του παραθύρου, τα οποία είναι δύσκολο να προσδιοριστούν. Επιπλέον, είναι υπολογιστικά πιο «δαπανηρό» από το SCD (Zhang et al., 2012).

Ο κυρίαρχος περιγραφέας χρώματος (Dominant Color Descriptor – DCD) είναι επίσης μια παραλλαγή του ιστογράμματος. Ο DCD επιλέγει ένα μικρό αριθμό χρωμάτων από τις υψηλότερες περιοχές τιμών ενός ιστογράμματος. Ο αριθμός των περιοχών τιμής χρώματος που επιλέγεται ως DCD, εξαρτάται από το όριο του ύψους της περιοχής τιμών. Το MPEG-7 προτείνει ότι ένα έως οκτώ χρώματα είναι επαρκή για να αντιπροσωπεύσουν μια περιοχή

(Manjunath et al., 2002). Σε αντίθεση με το παραδοσιακό ιστόγραμμα, τα επιλεγμένα χρώματα στο DCD προσαρμόζονται στην περιοχή αντί να τοποθετούνται στο χρωματικό χώρο. Έτσι, η αναπαράσταση χρώματος με DCD είναι πιο ακριβής και συμπαγής από ότι στο συμβατικό ιστόγραμμα. Έχει αποδειχθεί ότι ο DCD είναι επαρκής για να αντιπροσωπεύει τις πληροφορίες χρώματος μιας περιοχής. Επιπλέον, η διάσταση χαρακτηριστικών του DCD είναι χαμηλή και ο υπολογισμός του είναι σχετικά «φθηνός» (Zhang et al., 2012).

Ο Πίνακας 4 παρέχει μια σύνοψη των διαφορετικών χαρακτηριστικών (περιγραφέων) χρώματος που εξετάστηκαν.

Περιγραφέας χρώματος	Πλεονεκτήματα	Μειονεκτήματα
Ιστόγραμμα	Απλό στον υπολογισμό, διαισθητικό	Υψηλή διάσταση, καμία χωρική πληροφορία, ευαίσθητο στον θόρυβο
CM – χρωματικές ροπές	Συμπαγείς, ανθεκτικές	Δεν είναι αρκετές για να περιγράψουν όλα τα χρώματα, καμία χωρική πληροφορία
CCV	Χωρική πληροφορία	Υψηλή διάσταση, υψηλό κόστος υπολογισμού
Correlogram	Χωρική πληροφορία	Πολύ υψηλό κόστος υπολογισμού, ευαίσθητο στον θόρυβο, την περιστροφή και την κλιμάκωση
DCD	Συμπαγής, ισχυρός, αντιληπτή ερμηνεία	Χρειάζεται μετα-επεξεργασία για χωρική πληροφορία
CSD	Χωρική πληροφορία	Ευαίσθητος στον θόρυβο, την περιστροφή και την κλιμάκωση
SCD	Συμπαγής αν χρειαστεί, επεκτασιμότητα	Καμία χωρική πληροφορία, λιγότερο ακριβής εάν είναι συμπαγής

Πίνακας 4. Παρουσίαση διαφορετικών περιγραφέων χρώματος (Zhang et al., 2012).

Οι έγχρωμες ιατρικές εικόνες παράγονται συνήθως σε διαφορετικά τμήματα του ανθρώπινου σώματος και διαφορετικές απεικονιστικές μεθόδους. Το χρώμα στις ιατρικές εικόνες συχνά αποκαλύπτει πολλά χαρακτηριστικά αλλοιώσεων διαδραματίζοντας επίσης σημαντικό ρόλο στη μορφολογική διάγνωση (Wei and Chen, 2012). Τα προβλήματα στις έγχρωμες ιατρικές εικόνες αφορούν στην ανακριβή αναπαραγωγή χρώματος, στις ακατέργαστες διαβαθμίσεις χρώματος και στην ανεπαρκή πυκνότητα των εικονοστοιχείων. Επομένως, η αποτελεσματική χρήση των διαφόρων πληροφοριών χρώματος στις ιατρικές εικόνες περιλαμβάνει τις απόλυτες τιμές χρώματος, τις αναλογίες καθενός από τα τρία βασικά χρώματα, τις διαφορές χρώματος σε γειτονικές περιοχές και τα εκτιμώμενα δεδομένα φωτισμού.

Επιπλέον, οι περισσότερες από τις τεχνικές ιατρικής απεικόνισης, όπως οι XRAY, CT και MRI, παράγουν ιατρικές εικόνες που είναι συνήθως σε αποχρώσεις του γκρι. Για τέτοιου είδους grayscale εικόνες, η ανάκτηση εικόνας βάσει περιεχομένου μπορεί να θεωρήσει το χρώμα ως δευτερεύον χαρακτηριστικό, επειδή τα επίπεδα γκρίζου παρέχουν περιορισμένες πληροφορίες σχετικά με το περιεχόμενο μιας εικόνας. Για ειδικούς σκοπούς, σε ορισμένες grayscale εικόνες έχουν προστεθεί ψευδοχρώματα για την ενίσχυση συγκεκριμένων περιοχών. Καθώς τα ψευδοχρώματα προστίθενται τεχνητά για να διευκολύνουν την ανθρώπινη παρατήρηση, μπορεί να μην ερμηνεύονται με τον ίδιο τρόπο από διαφορετικούς χρήστες. Αυτή η επεξεργασία ψευδοχρωματισμού αυξάνει τις δυσκολίες στην επισημείωση και την ανάκτηση εικόνων (Wei and Chen, 2012).

3.2.3. Χαρακτηριστικά υφής

Η υφή (texture) είναι ένα άλλο σημαντικό χαρακτηριστικό της εικόνας. Ενώ το χρώμα αποτελεί συνήθως μια ιδιότητα ενός εικονοστοιχείου, η υφή μπορεί να μετρηθεί μόνο από μια ομάδα εικονοστοιχείων. Λόγω της ισχυρής διακριτικής ικανότητάς του, το χαρακτηριστικό γνώρισμα της υφής χρησιμοποιείται ευρέως σε τεχνικές ανάκτησης εικόνας και σημασιολογικής μάθησης (Zhang et al., 2012). Η υφή παίζει σημαντικό ρόλο στο ανθρώπινο οπτικό σύστημα αντίληψης. Ελλείψει οποιουδήποτε ορισμού της υφής, τα επιφανειακά χαρακτηριστικά και η εμφάνιση ενός αντικειμένου μπορούν να θεωρηθούν ως η υφή αυτού του αντικειμένου. Περιγράφοντας την υφή, οι ερευνητές της μηχανικής όρασης την περιγράφουν ως το περιεχόμενο μιας εικόνας μετά την εξαγωγή των χρωμάτων και των τοπικών σχημάτων (Bhagat and Choudhary, 2018).

Η υφή μπορεί να είναι κανονική ή τυχαία. Οι περισσότερες φυσικές υφές είναι τυχαίες. Οι κανονικές υφές αποτελούνται από υφές που έχουν κανονική ή σχεδόν τακτική διάταξη πανομοιότυπων ή τουλάχιστον παρόμοιων συνιστωσών. Οι ακανόνιστες υφές αποτελούνται από ακανόνιστες και τυχαίες διατάξεις στοιχείων που σχετίζονται με ορισμένες στατιστικές ιδιότητες (Tsai and Hung, 2008).

Η αναπαράσταση της υφής στην ανάκτηση εικόνων μπορεί να χρησιμοποιηθεί για δύο τουλάχιστον σκοπούς. Πρώτον, μια εικόνα μπορεί να θεωρηθεί ως ψηφιδωτό που αποτελείται από διαφορετικές περιοχές υφής. Αυτές οι περιοχές μπορούν να

χρησιμοποιηθούν ως παραδείγματα για την αναζήτηση και ανάκτηση παρόμοιων περιοχών. Δεύτερον, η υφή μπορεί να χρησιμοποιηθεί για αυτόματη επισημείωση του περιεχομένου μιας εικόνας. Για παράδειγμα, η υφή μίας μολυσμένης περιοχής δέρματος μπορεί να χρησιμοποιηθεί για την επισημείωση περιοχών με την ίδια μόλυνση. Αυτό που πρέπει να τονιστεί είναι ότι ιατρικές εικόνες που βασίζονται σε διαφορετικές τεχνικές ιατρικής απεικόνισης έχουν διαφορετικά χαρακτηριστικά. Οι ιστολογικές εικόνες μικροσκοπίου διαθέτουν μοναδικές υπογραφές χρώματος και κυτταρικές υφές. Οι εικόνες υπερήχων των μεγάλων οργάνων φαίνεται να κυριαρχούνται από τις υφές, συνεπώς δίνουν έμφαση στην εξαγωγή μιας ολικής ιδιότητας και όχι τοπικών χαρακτηριστικών. Οι ακτινογραφίες θώρακα είναι προβολές πολλών επικαλυπτόμενων δομών. Οι διαδικασίες ταξινόμησης ακτινογραφιών αξιοποιούν χαρακτηριστικά με βάση την υφή και όχι το σχήμα. Οι τομογραφικές εικόνες επιτρέπουν εύκολα την αναγνώριση μη επικαλυπτόμενων γεωμετρικά οριοθετημένων οργάνων και ιστών ως συλλογή μεμονωμένων χαρακτηριστικών. Η υφή περιέχει σημαντικές πληροφορίες σχετικά με τις δομικές διατάξεις των επιφανειών τους και τις σχέσεις με τους περιβάλλοντες ιστούς (Wei and Chen, 2012).

Η υφή έχει μελετηθεί αρκετά στην περιοχή επεξεργασίας εικόνας και στον τομέα της μηχανικής όρασης. Έχουν προταθεί διάφορες τεχνικές για την εξαγωγή χαρακτηριστικών υφής. Οι μέθοδοι εξαγωγής χαρακτηριστικών υφής μπορούν να ταξινομηθούν, με βάση το πεδίο από το οποίο εξάγεται το χαρακτηριστικό της υφής, γενικά σε χωρικές (spatial) και φασματικές (spectral).

3.2.3.1. Χωρικές μέθοδοι εξόρυξης χαρακτηριστικών υφής

Στη χωρική προσέγγιση, τα χαρακτηριστικά υφής εξάγονται με υπολογισμό των στατιστικών των εικονοστοιχείων ή ανίχνευση των τοπικών δομών εικονοστοιχείων στο αρχικό πεδίο εικόνας. Οι χωρικές τεχνικές εξαγωγής χαρακτηριστικών υφής μπορούν να ταξινομηθούν περαιτέρω ως δομικές (Structural), στατιστικές (statistical) και βασισμένες σε μοντέλα (Model Based) (Tsai and Hung, 2008).

Οι δομικές τεχνικές περιγράφουν τις υφές χρησιμοποιώντας ένα σύνολο θεμελιακών στοιχείων υφής (textons ή texture elements) και τους κανόνες τοποθέτησής τους. Τα textons είναι οργανωμένα σε έναν γραμμικό περιγραφέα και τεχνικές συντακτικής αναγνώρισης προτύπων χρησιμοποιούνται για να εντοπίσουν την ομοιότητα δύο περιγραφών.

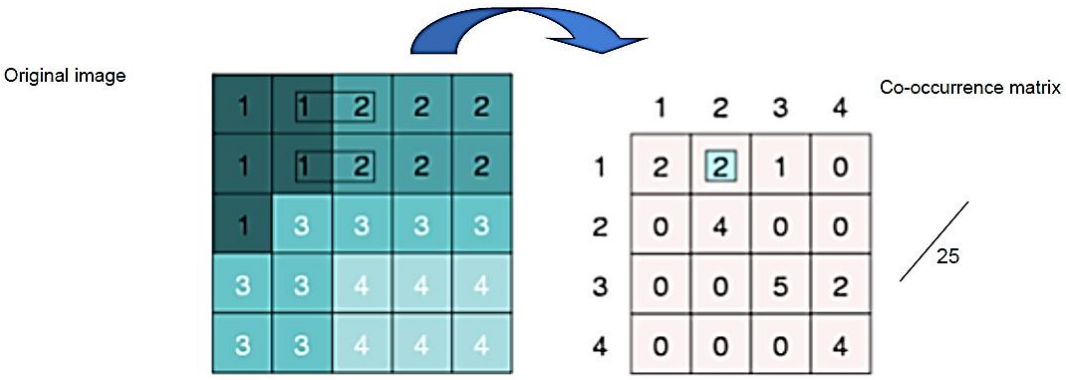
Ένα στατιστικό χαρακτηριστικό υφής περιγράφει την υφή ως μέτρο χαμηλού επιπέδου στατιστικών στοιχείων μιας grayscale εικόνας. Τα πιο κοινά στατιστικά χαρακτηριστικά στο πεδίο του χώρου είναι οι ροπές υφής, τα χαρακτηριστικά υφής Tamura και ο πίνακας συν-εμφάνισης σε επίπεδο γκρίζου (Grey Level Co-occurance Matrix).

Η μέθοδος GLCM είναι μία από τις στατιστικές τεχνικές στην ανάλυση της υφής για τον υπολογισμό των ιδιοτήτων της υφής που σχετίζονται με τα στατιστικά στοιχεία δεύτερης τάξης της εικόνας. Ουσιαστικά, ο πίνακας GLCM είναι μια εκτίμηση της κοινής συνάρτησης πυκνότητας πιθανότητας των τιμών ενός ζεύγους εικονοστοιχείων σε μια εικόνα. Ο πίνακας GLCM θα μπορούσε να εκφραστεί σύμφωνα με την ακόλουθη έκφραση:

$$Pd, \theta(i,j), \qquad (i,j = 0,1,\dots,N-1)$$

όπου i και j δείχνουν το επίπεδο γκρίζου των δύο εικονοστοιχείων που απέχουν μεταξύ απόσταση d κατά γωνία θ και N είναι ο αριθμός των επιπέδων γκρίζου στην εικόνα.

Τα χαρακτηριστικά υφής που εξάγονται από τους πίνακες συν-εμφάνισης των επιπέδων του γκρίζου (GLCM) περιλαμβάνουν τη συσχέτιση, την αντίθεση, την ομοιογένεια κ.ά. (Fesharaki and Pourghassem, 2013)



Εικόνα 7. Ο πίνακας συνεμφάνισης επιπέδου γκρι (Fesharaki and Pourghassem, 2013).

Τα χαρακτηριστικά Tamura είναι μία ομάδα από έξι χαρακτηριστικά της υφής, τα οποία είναι η τραχύτητα (Coarseness), η αντίθεση (Contrast), η κατευθυντικότητα (Directionality), η γραμμικότητα (Line-likeness), η κανονικότητα (Regularity) και η ανωμαλία (Roughness) της επιφάνειας. Αναπαριστώνται από ένα διάνυσμα 18 διαστάσεων (Tamura et al., 1978). Τα στατιστικά χαρακτηριστικά είναι συμπαγή και ισχυρά, επειδή έχουν μεγάλη θεωρητική

θεμελίωση. Ωστόσο, δεν επαρκούν για να περιγράψουν τη μεγάλη ποικιλία υφών σε μία εικόνα (Zhang et al., 2012).

Στις τεχνικές που βασίζονται σε μοντέλα, η υφή ερμηνεύεται χρησιμοποιώντας στοχαστικά (τυχαία) ή παραγωγικά (generative) μοντέλα. Οι παράμετροι του μοντέλου χαρακτηρίζουν την υποκείμενη ιδιότητα υφής της εικόνας. Δημοφιλή μοντέλα υφής που αναφέρουν οι Zhang et al. (2012) είναι το τυχαίο μοντέλο Markov (Markov Random field - MRF), το μοντέλο ταυτόχρονης αυτοπαλινδρόμησης (Simultaneous Auto-Regressive - SAR) και η διάσταση των φράκταλς (Fractal Dimension - FD). Καθώς αυτά τα μοντέλα περιλαμβάνουν βελτιστοποίηση, είναι συνήθως υπολογιστικά δαπανηρά.

Οι χωρικές μέθοδοι εξαγωγής χαρακτηριστικών υφής είναι εύκολο να κατανοηθούν και πολλές από αυτές έχουν ακόμη και σημασιολογική ερμηνεία. Δεν απαιτούν το σχήμα της περιοχής να είναι κανονικό και μπορούν να εφαρμοστούν απρόσκοπτα και στις ακανόνιστες περιοχές. Ωστόσο, αυτά τα χαρακτηριστικά είναι συνήθως ευαίσθητα στο θόρυβο και τις στρεβλώσεις. Επιπλέον, πολλές από αυτές τις μεθόδους περιλαμβάνουν πολύπλοκες διαδικασίες αναζήτησης και βελτιστοποίησης που δεν έχουν γενικές λύσεις.

Ο Πίνακας 5 συνοψίζει τις διαφορετικές χωρικές μεθόδους εξαγωγής χαρακτηριστικών υφής (Zhang et al., 2012).

Μέθοδος χαρακτηριστικών υφής	Πλεονεκτήματα	Μειονεκτήματα
Texton	Διαισθητική	Ευαίσθητη στο θόρυβο, την περιστροφή και την κλιμάκωση, δύσκολο να καθοριστούν τα textons
GLCM	Διαισθητική, συμπαγής, ισχυρή	Υψηλό κόστος υπολογισμού, δεν επαρκεί για να περιγράψει όλες τις υφές
Tamura	Αντιληπτικά σημαντική / σημασιολογική ερμηνεία	Μικρός αριθμός χαρακτηριστικών
SAR	Συμπαγής, ανθεκτική, αναλλοίωτη στην περιστροφή	Υψηλό κόστος υπολογισμού, δύσκολο να καθοριστεί το μέγεθος του προτύπου
FD	Συμπαγής, Αντιληπτικά σημαντική	Υψηλό κόστος υπολογισμού, ευαίσθητη στην κλιμάκωση

Πίνακας 5. Αντιπαραβολή διαφορετικών χωρικών μεθόδων εξαγωγής χαρακτηριστικών υφής (Zhang et al., 2012).

3.2.3.2. Φασματικές τεχνικές εξαγωγής χαρακτηριστικών υφής

Στις φασματικές τεχνικές εξαγωγής χαρακτηριστικών υφής, μια εικόνα μετασχηματίζεται στον τομέα της συχνότητας και στη συνέχεια το χαρακτηριστικό υπολογίζεται από τη μετασχηματισμένη εικόνα. Οι πιο κοινές φασματικές τεχνικές περιλαμβάνουν το μετασχηματισμό Fourier (Fourier Transform - FT), τον διακριτό μετασχηματισμό συνημίτονου (Discrete Cosine Transform - DCT), το μετασχηματισμό Wavelet και τα φίλτρα Gabor.

Οι FT και DCT υπολογίζονται πολύ γρήγορα αλλά δεν είναι αμετάβλητοι στην κλιμάκωση και την περιστροφή. Ο μετασχηματισμός Wavelet είναι αποτελεσματικός και ισχυρός, αλλά καταγράφει μόνο οριζόντια και κάθετα χαρακτηριστικά. Ανάμεσά τους, τα χαρακτηριστικά του Gabor είναι τα πιο εύρωστα, επειδή καταγράφουν χαρακτηριστικά εικόνας σε πολλαπλές κατευθύνσεις και σε πολλαπλές κλίμακες. Τα χαρακτηριστικά που στηρίζονται στο μετασχηματισμό curvelet (Sumana et al., 2008), έχουν αξιοσημείωτα πλεονεκτήματα έναντι των χαρακτηριστικών του Gabor και των χαρακτηριστικών wavelet, επειδή είναι πιο αποτελεσματικά στη λήψη καμπυλόγραμμων ιδιοτήτων, όπως οι γραμμές και ακμές.

Το πρόβλημα με αυτές τις φασματικές μεθόδους είναι ότι μπορούν να εφαρμοστούν μόνο σε τετραγωνικές περιοχές λόγω της χρήσης του αλγορίθμου FFT (Fast Fourier Transform). Οι περισσότερες από τις υπάρχουσες τεχνικές που βασίζονται σε περιοχές ορίζουν μια περιοχή ως σύνολο μικρών τετραγώνων μεγέθους 4×4 pixel και εφαρμόζουν φασματικό μετασχηματισμό σε αυτά τα μπλοκ, επειδή τα μικρά μπλοκ είναι πιθανώς ομοιογενή. Τα χαρακτηριστικά μιας περιοχής υπολογίζονται τότε ως τα μέσα χαρακτηριστικά αυτών των μπλοκ. Αυτή η μέθοδος έχει το μειονέκτημα ότι τα μπλοκ είναι πολύ μικρά για να συλλαμβάνουν επαρκείς πληροφορίες ακμής. Για την επίλυση αυτού του προβλήματος, οι Islam et al. (2009) προτείνουν μια μέθοδο πλήρωσης (padding) για να μετασχηματίσουν μια ακανόνιστη περιοχή υφής σε μια τετράγωνη περιοχή υφής. Αυτή η μέθοδος δέχεται επίσης μεγάλες περιοχές για να εξάγει σημαντικά χαρακτηριστικά υφής.

Στον πίνακα 6 συνοψίζονται τα πλεονεκτήματα και μειονεκτήματα διαφορετικών μεθόδων εξαγωγής χαρακτηριστικών υφής στο πεδίο του φάσματος της εικόνας (Zhang et al., 2012).

Μέθοδος χαρακτηριστικών υφής	Πλεονεκτήματα	Μειονεκτήματα
FT/DCT	Γρήγορος υπολογισμός	Ευαίσθητη στην κλιμάκωση και την περιστροφή
Wavelet	Γρήγορος υπολογισμός, πολλαπλής ανάλυσης	Ευαίσθητη στην περιστροφή, περιορισμένοι προσανατολισμοί
Gabor	Πολλαπλής κλίμακας, Πολλαπλών-προσανατολισμών, ισχυρή	Χρειάζεται κανονικοποίηση περιστροφής, απώλεια φασματικών πληροφοριών λόγω ατελούς κάλυψης του επιπέδου
Curvelet	Πολλαπλής ανάλυσης, πολλαπλών προσανατολισμών, ανθεκτική	Χρειάζεται κανονικοποίηση περιστροφής

Πίνακας 6. Αντιπαραβολή διαφορετικών φασματικών μεθόδων εξαγωγής χαρακτηριστικών υφής (Zhang et al., 2012).

3.2.4. Χαρακτηριστικά σχήματος

Το σχήμα είναι ένα από τα πιο σημαντικά χαρακτηριστικά για την περιγραφή του περιεχομένου ή των αντικειμένων μιας εικόνας. Οι άνθρωποι μπορούν εύκολα να αναγνωρίσουν διαφορετικές εικόνες και να τις ταξινομήσουν σε διαφορετικές κατηγορίες μόνο από το περίγραμμα ενός αντικειμένου σε μια δεδομένη εικόνα. Καθώς το σχήμα μπορεί να μεταφέρει κάποιο είδος σημασιολογικής πληροφορίας που έχει νόημα στην ανθρώπινη αναγνώριση, χρησιμοποιείται ως χαρακτηριστικό γνώρισμα για την απεικόνιση ενός αντικειμένου εικόνας (Tsai and Hung, 2008, Wei and Chen, 2012).

Σε ορισμένες ιατρικές εικόνες, το σχήμα είναι το πιο σημαντικό χαρακτηριστικό για την περιγραφή των παθολογιών στις ιατρικές εικόνες. Για παράδειγμα, οι εικόνες ακτίνων Χ της σπονδυλικής στήλης παρουσιάζονται με γκρι κλίμακα και παρέχουν λίγες πληροφορίες όσον αφορά την υφή για την ανατομία που μας ενδιαφέρει. Το σχήμα των σπονδύλων είναι το πιο σημαντικό χαρακτηριστικό που περιγράφει διάφορες παθολογίες στις εικόνες ακτίνων Χ της σπονδυλικής στήλης. Μία από τις προκλήσεις είναι ότι, ενώ οι διαφορές μεταξύ φυσιολογικών και παθολογικών συνθηκών είναι λεπτές, τα σχήματα του ίδιου τύπου παθολογίας παρουσιάζουν μεγαλύτερες διακυμάνσεις. Ως εκ τούτου, είναι δύσκολο να επιλεγεί ένας κατάλληλος περιγραφέας σχήματος που δεν θα αντιπροσωπεύει μόνο τις ανατομικές δομές, αλλά θα διατηρεί επίσης αρκετές πληροφορίες για μέτρηση ομοιότητας (Wei and Chen, 2012).

Οι τεχνικές εξαγωγής χαρακτηριστικών σχήματος ταξινομούν ευρέως σε δύο κύριες ομάδες: μεθόδους που βασίζονται σε περιγράμματα (ή ανίχνευση ορίων) (contour based) και σε περιοχές (region based) (Tsai and Hung, 2008). Οι μέθοδοι που βασίζονται σε περιγράμματα υπολογίζουν τα χαρακτηριστικά του σχήματος μόνο από το περίγραμμα του σχήματος, ενώ οι μέθοδοι με βάση τις περιοχές εξάγουν χαρακτηριστικά από ολόκληρη την περιοχή. Επειδή οι τεχνικές που βασίζονται σε περιγράμματα χρησιμοποιούν μόνο ένα τμήμα της περιοχής, είναι πιο ευαίσθητες στον θόρυβο από τις τεχνικές που βασίζονται στην περιοχή, καθώς μικρές αλλαγές στο σχήμα επηρεάζουν σημαντικά το περίγραμμα του σχήματος. Επομένως, η ανάκτηση έγχρωμων εικόνων συνήθως χρησιμοποιεί λειτουργίες σχήματος με βάση την περιοχή (Zhang et al., 2012).

Ένας αριθμός απλών χαρακτηριστικών σχήματος που βασίζονται στο περίγραμμα μιας περιοχής, χρησιμοποιείται συνήθως στην ανάκτηση έγχρωμων εικόνων, συμπεριλαμβανομένων των ροπών σχήματος, της περιοχής (area) (Duygulu et al., 2002, Jeon et al., 2004), της κυκλικότητας (circularity) και της εκκεντρότητας (eccentricity) κ.ά. Η κυκλικότητα μετρά την αναλογία της περιοχής σε σχέση με το περίγραμμα. Η εκκεντρότητα ή επιμήκυνση (elongation) είναι η αναλογία του μήκους του κύριου άξονα με εκείνο του δευτερεύοντος άξονα. Μεμονωμένοι απλοί περιγραφείς σχήματος δεν είναι συνήθως εύρωστοι και συνεπώς, συνδυάζονται για να δημιουργήσουν έναν πιο αποτελεσματικό περιγραφέα σχήματος (Zhang et al., 2012).

Πιο περίπλοκα χαρακτηριστικά σχήματος χρησιμοποιούνται συνήθως σε συγκεκριμένες εφαρμογές όπως η ανάκτηση εμπορικών σημάτων και η ταξινόμηση αντικειμένων, όπου το σχήμα είναι το πιο σημαντικό χαρακτηριστικό. Οι Park et al. (2007) ανέπτυξαν ένα σύστημα το οποίο εξάγει αυτόματα έννοιες υψηλού επιπέδου από τις εικόνες χρησιμοποιώντας οντολογίες και κανόνες εξαγωγής συμπερασμάτων. Στη μέθοδό τους χρησιμοποιούν τους οπτικούς περιγραφείς του προτύπου MPEG-7 για την εξαγωγή των οπτικών χαρακτηριστικών της εικόνας. Ως περιγραφέα σχήματος χρησιμοποιούν τον περιγραφέα περιγράμματος σχήματος (Contour Shape descriptor) του MPEG-7.

Η προσέγγιση του περιγραφέα Fourier είναι μία από τις σημαντικές προσεγγίσεις που βασίζονται σε περιγράμματα. Βασική ιδέα του είναι να χρησιμοποιήσει τα εικονοστοιχεία του ορίου του αντικειμένου για να υπολογίσει το περίγραμμά του. Ο μετασχηματισμός

Fourier χρησιμοποιείται για το περίγραμμα του σχήματος για να υπολογιστούν οι συντελεστές Fourier (Zhang et al., 2008). Στη συνέχεια, οι συντελεστές Fourier που είναι αμετάβλητοι στη μετατόπιση, στην κλιμάκωση, στην περιστροφή και στην αλλαγή του αρχικού σημείου, χρησιμοποιούνται ως χαρακτηριστικά σχήματος. Η προσέγγιση του περιγραφέα Fourier παρουσιάζει σημαντικά πλεονεκτήματα συγκριτικά με άλλους περιγραφείς σχήματος που βασίζονται στα περιγράμματα. Το κυριότερο είναι ότι η επίδραση του θορύβου καθώς και μικρών αλλαγών στο σχήμα που επηρεάζουν το περίγραμμα του σχήματος, μειώνεται αποτελεσματικά με την ανάλυση του σχήματος στον τομέα των συχνοτήτων. Επίσης τα χαρακτηριστικά από την προσέγγιση του περιγραφέα Fourier αποτελούν μια συμπαγή περιγραφή και είναι εύκολο να κανονικοποηθούν. Το υπολογιστικό κόστος του περιγραφέα Fourier είναι χαμηλό ενώ έχει καλύτερη απόδοση σε σύστημα ανάκτησης σε σύγκριση με διαφορετικές προσεγγίσεις εξαγωγής χαρακτηριστικών σχήματος. Έχουν προταθεί τροποποιημένες εκδόσεις για τη βελτίωση της απόδοσης της προσέγγισης περιγραφέα Fourier έχουν προταθεί. Οι Zhang and Lu (2004) μετασχηματίζουν μια εικόνα από το καρτεσιανό σύστημα συντεταγμένων στο σύστημα πολικών συντεταγμένων. Ο μετασχηματισμός Fourier χρησιμοποιείται στην μετασχηματισμένη εικόνα για να βελτιώσει την απόδοση του περιγραφέα Fourier.

Σε σύγκριση με τα χαρακτηριστικά χρώματος και υφής, τα χαρακτηριστικά σχήματος περιγράφονται συνήθως μετά την τμηματοποίηση των εικόνων σε περιοχές ή αντικείμενα. Συνεπώς, η αποτελεσματική εξαγωγή χαρακτηριστικών σχήματος εξαρτάται από τις μεθόδους τμηματοποίησης (Tsai and Hung, 2008).

3.2.5. Χωρική σχέση

Η χωρική σχέση (Spatial Relationship) περιγράφει τη θέση του αντικειμένου μέσα σε μια εικόνα ή τις σχέσεις μεταξύ αντικειμένων. Τα αντικείμενα και οι χωρικές σχέσεις (όπως αριστερά, μέσα και πάνω) μεταξύ αντικειμένων σε μια εικόνα χρησιμοποιούνται για να αντιπροσωπεύσουν το περιεχόμενο της εικόνας. Μια εικόνα μπορεί να χωριστεί σε έναν αριθμό μπλοκ και τα χρώματα, η υφή και / ή τα χαρακτηριστικά σχήματος εξάγονται από κάθε ένα από τα μπλοκ. Στη συνέχεια, μπορούμε να τα προβάλουμε κατά μήκος των αξόνων x και y, όπως οι σχέσεις «αριστερά / δεξιά», «κάτω / πάνω» μεταξύ τους (Tsai and Hung,2008). Για κάθε προβολή, η σχέση μεταξύ αντικειμένων αντιπροσωπεύεται από μια σειρά (διάνυσμα) συμβόλων. Τα σύμβολα προέρχονται από δύο σύνολα: το σύνολο των

συμβόλων αντικειμένων και το σύνολο συμβόλων σχέσεων, όπως «αριστερά / δεξιά», «κάτω / πάνω».

Έχουν προταθεί διάφορες παραλλαγές αυτής της μεθόδου. Αυτές οι προσεγγίσεις διαφέρουν ως προς τον αριθμό των σχεσιακών τελεστών (σύμβολα) και τον τρόπο με τον οποίο ορίζουν αυτές τις σχέσεις. Η εικόνα 8 παρουσιάζει το παράδειγμα της αναπαράστασης μιας δισδιάστατης συμβολοσειράς. Η εικόνα αποσυντίθεται σε περιοχές (μπλοκ). Για απλότητα, τα αναγνωριστικά των μπλοκ χρησιμοποιούνται ως σύμβολα αντικειμένων. Στην περίπτωση αυτή χρησιμοποιούνται δύο σύμβολα σχέσης $'<'$ και $'='$. Σε οριζόντιες και κάθετες κατευθύνσεις, το σύμβολο $'<'$ υποδηλώνει σχέσεις αριστερά-δεξιά και κάτω-πάνω αντίστοιχα. Το σύμβολο $'='$ σημαίνει τη χωρική σχέση «στην ίδια χωρική θέση με». Μια $2D$ συμβολοσειρά παίρνει τη μορφή (u, v), όπου u και v είναι οι σχέσεις αντικειμένων στην οριζόντια και κάθετη κατεύθυνση, αντίστοιχα. Η εικόνα 1 (d) δείχνει την $2D$ συμβολοσειρά για την εικόνα 1 (α) (Zhang et al., 2012).

(a)

d		
	b	c
a	a	

(b)

Σύμβολα Αντικειμένων	a, b, c, d

(c)

Σχεσιακοί Τελεστές	***Σημασία***
<	*Αριστερά-δεξιά ή πάνω-κάτω*
=	*Στην ίδια χωρική θέση με*

(d)

(a = d < a = b < c, a = a < b = c < d)

Εικόνα 8. Παρουσίαση μιας δισδιάστατης συμβολοσειράς: (a) μια εικόνα διαχωρισμένη σε μπλοκ, (b) σύμβολα αντικειμένων ως ονόματα μπλοκ, (c) ορισμοί σχεσιακών συμβόλων, και (d) μία δισδιάστατη συμβολοσειρά για την εικόνα (a) (σχήμα από τους (Zhang et al., 2012).

Η $2D$ συμβολοσειρά και οι παραλλαγές της μπορούν να χρησιμοποιηθούν ως συνολικά χαρακτηριστικά για αναπαράσταση με βάση την περιοχή, με την προϋπόθεση ότι τα αντικείμενα ορίζονται καλά από τις τμηματοποιημένες περιοχές. Καθώς οι αλγόριθμοι τμηματοποίησης συχνά διαιρούν ένα αντικείμενο σε διαφορετικά θραύσματα, η $2D$ συμβολοσειρά συνήθως δεν δίνει ακριβή αναπαράσταση. Στην πράξη, συνήθως χρησιμοποιείται η σχετική θέση των περιοχών. Οι Hsu et al. (1996) ανιχνεύουν αντικείμενα όπως οστά, όγκους του εγκεφάλου και περιγράμματα στήθους σε ιατρικές εικόνες ακτίνων Χ προτείνοντας ένα μοντέλο βασισμένο στη γνώση. Το μοντέλο αντιπροσωπεύει επιλεγμένα χαρακτηριστικά και χωρικές σχέσεις μεταξύ τους με τη μορφή ιεραρχίας τύπων.

3.2.6. Χαρακτηριστικά σημείων ενδιαφέροντος εικόνας

Στην ανάκτηση εικόνων χρησιμοποιούνται συνολικά χαρακτηριστικά χρώματος ή υφής για την περιγραφή του περιεχομένου της εικόνας. Το πρόβλημα με αυτή την προσέγγιση είναι ότι αυτά τα συνολικά χαρακτηριστικά δεν μπορούν να συλλάβουν όλα τα μέρη της εικόνας που έχουν διαφορετικά χαρακτηριστικά. Επομένως, είναι απαραίτητος ο τοπικός υπολογισμός πληροφοριών της εικόνας. Με τη χρήση σημαντικών σημείων που αντιπροσωπεύουν τοπικές πληροφορίες, μπορούν να υπολογιστούν περισσότερα διακριτικά χαρακτηριστικά (Sebe et al., 2003).

Προεξέχοντα σημεία που υπάρχουν στην εικόνα, τα οποία συνήθως προσδιορίζονται από το χρώμα, την υφή ή τα τοπικά σχήματα, χρησιμοποιούνται για την παραγωγή τοπικών χαρακτηριστικών. Παρόλο που τα χαρακτηριστικά χρώματος και υφής χρησιμοποιούνται συνήθως για να αντιπροσωπεύουν τα περιεχόμενα της εικόνας, τα προεξέχοντα σημεία (salient points) παρέχουν πιο διακριτικά χαρακτηριστικά. Ελλείψει τμηματοποίησης σε επίπεδο αντικειμένων, τα σημαντικότερα σημεία δρουν ως αδύναμη τμηματοποίηση και μπορούν να αναπαραστήσουν μια εικόνα. Προεξέχοντα σημεία μπορεί να υπάρχουν σε διαφορετικές θέσεις της εικόνας και δεν είναι κατ' ανάγκη γωνίες, δηλαδή, μπορεί να είναι επίσης ομαλές γραμμές.

Οι Sebe et al. (2003) σύγκριναν τα κυριότερα σημεία που εξήχθησαν, χρησιμοποιώντας έναν αλγόριθμο εξαγωγής σημαντικών σημείων βασισμένο σε μετασχηματισμό κυματιδίων wavelet με έναν αλγόριθμο ανίχνευσης γωνιών (Harris corner detector). Η μελέτη τους έδειξε ότι η εξαγωγή πληροφοριών χρώματος και υφής στις τοποθεσίες που δίνουν τα σημαντικότερα σημεία, μας προσφέρει σημαντικά βελτιωμένα αποτελέσματα όσον αφορά την ακρίβεια ανάκτησης, την υπολογιστική πολυπλοκότητα και το χώρο αποθήκευσης των διανυσμάτων χαρακτηριστικών σε σύγκριση με τις προσεγγίσεις συνολικών χαρακτηριστικών. Αν και τα προεξέχοντα σημεία υποκαθιστούν την τμηματοποίηση (επειδή η τμηματοποίηση είναι μια εύθραυστη εργασία), αν χρησιμοποιούνται σημαντικά σημεία μαζί με τμηματοποίηση, δίνουν πολύ περισσότερα διακριτικά χαρακτηριστικά (Bhagat and Choudhary, 2018).

Πρόσφατα, τα χαρακτηριστικά βασισμένα στο μετασχηματισμό χαρακτηριστικών αμετάβλητων στην κλιμάκωση (SIFT) (Lowe, 2004) έχουν γίνει πολύ πιο δημοφιλή. Ο SIFT

είναι ένας περιγραφέας χαρακτηριστικών αμετάβλητος στην κλιμάκωση και την περιστροφή που βασίζεται σε ιστόγραμμα προσανατολισμένο στις ακμές, το οποίο εξάγει βασικά σημεία (σημεία ενδιαφέροντος) και τους περιγραφείς του. Μια παρόμοια μέθοδος που υπολογίζει ένα ιστόγραμμα της κατεύθυνσης των κλίσεων σε ένα εντοπισμένο τμήμα μιας εικόνας που ονομάζεται ιστόγραμμα προσανατολισμένων κλίσεων (Histogram of Oriented Gradients - HOG) είναι ένας ισχυρός περιγραφέας χαρακτηριστικών. Επίσης, μια ταχύτερη έκδοση του αλγορίθμου SIFT που ονομάζεται επιταχυνόμενη εξαγωγή εύρωστων χαρακτηριστικών (Speeded-Up Robust Features - SURF), προτάθηκε από τους Bay et al. (2006). Για εφαρμογές σε πραγματικό χρόνο, προτάθηκε από τους Rosten and Drummond (2006) ένας αλγόριθμος μηχανικής μάθησης ανίχνευσης γωνίας που ονομάζεται χαρακτηριστικά από επιταχυνόμενο δοκιμαστικό τμήματος (Features from Accelerated Segment Test - FAST).

Οι περιγραφείς SIFT και SURF μετατρέπονται συνήθως σε δυαδικές συμβολοσειρές για να επιταχυνθεί η διαδικασία αντιστοίχισης. Ωστόσο, τα δυαδικά ισχυρά ανεξάρτητα στοιχειώδη χαρακτηριστικά (Binary Robust Independent ELementary Features - BRIEF) παρέχουν μια πιο σύντομη διαδικασία για την ανίχνευση δυαδικών συμβολοσειρών απευθείας χωρίς να υπολογίζονται οι περιγραφείς. Αξίζει να σημειωθεί ότι ο BRIEF είναι ένας περιγραφέας χαρακτηριστικών, όχι ένας ανιχνευτής χαρακτηριστικών (Calonder et al., 2010). Ο SIFT χρησιμοποιεί έναν περιγραφέα 128 διαστάσεων και ο SURF έχει έναν περιγραφέα 64 διαστάσεων, συνεπώς και οι δύο έχουν μεγάλη υπολογιστική πολυπλοκότητα. Μια εναλλακτική λύση του SIFT και του SURF που ονομάζεται «προσανατολισμένος FAST και περιστρεφόμενος» BRIEF (Oriented FAST and Rotated BRIEF - ORB) προτάθηκε από τους Rublee et al. (2011).

Η εξαγωγή χαρακτηριστικών είναι ένα από τα βασικά στάδια ενός συστήματος επισημείωσης εικόνας. Τα οπτικά χαρακτηριστικά έχουν σημαντικό ρόλο στην ταυτοποίηση και την αναγνώριση αντικειμένων και την αναπαράσταση του περιεχομένου της εικόνας. Διαφορετικοί τύποι χαρακτηριστικών (υφή, χρώμα, SIFT κ.λπ.) έχουν διαφορετικά χαρακτηριστικά. Τα τοπικά χαρακτηριστικά (SIFT, SURF, σχήμα, κ.λπ.) περιγράφουν περιοχές εικόνας ενώ τα συνολικά χαρακτηριστικά αντιπροσωπεύουν μια εικόνα ως σύνολο. Έτσι, τα τοπικά χαρακτηριστικά προσδιορίζουν μια εικόνα και χρησιμοποιούνται για την αναγνώριση αντικειμένων, τα συνολικά χαρακτηριστικά αποτελούν μια γενίκευση της εικόνας και είναι κατάλληλα για ανίχνευση αντικειμένων. Οι πρόσφατες μέθοδοι εξαγωγής χαρακτηριστικών

(SIFT, SURF, κ.λπ.) έχουν πλέον καθιερωθεί και αποτελούν βασικές μεθόδους εξαγωγής χαρακτηριστικών. Ωστόσο, συνολικά χαρακτηριστικά όπως η υφή και το χρώμα έχουν τη σημασία τους και θα συνεχίσουν να χρησιμοποιούνται στην πράξη (Bhagat and Choudhary, 2018).

Ένα αντικείμενο συνήθως περιέχει πολλά χαρακτηριστικά σε μια εικόνα, οπότε η ενσωμάτωση χαρακτηριστικών έχει λάβει όλο και περισσότερη προσοχή τα τελευταία χρόνια. Βασική ιδέα του είναι να υπολογίσει περισσότερα διακριτικά χαρακτηριστικά ενσωματώνοντας διάφορα χαρακτηριστικά. Οι προσεγγίσεις ενσωμάτωσης χαρακτηριστικών μπορούν να ταξινομηθούν σε προσεγγίσεις σειριακής ολοκλήρωσης και σε παράλληλης ενσωμάτωσης ανάλογα με το εάν τα χαρακτηριστικά χρησιμοποιούνται κατά σειρά ή όχι (Zhang et al., 2012).

Κεφάλαιο 4
Ταξινόμηση τεχνικών επισημείωσης εικόνας

Ο άνθρωπος διαθέτει ένα απίστευτο οπτικό σύστημα ερμηνείας. Το ανθρώπινο οπτικό σύστημα ερμηνεύει την εικόνα ενώ παράλληλα συνδέει το θέμα με τα αντικείμενα που υπάρχουν στην εικόνα. Ο άνθρωπος μπορεί να περιγράψει αβίαστα τα αντικείμενα και τα σχετικά χαρακτηριστικά τους σε μία εικόνα. Οι ερευνητές της μηχανικής όρασης επιχείρησαν εντατικά τις τελευταίες δεκαετίες να καταστήσουν το σύστημα ηλεκτρονικών υπολογιστών ικανό να μιμηθεί αυτή την ικανότητα του ανθρώπου. Η αυτόματη επισημείωση εικόνας (AIA) αποτελεί ένα βήμα προς αυτήν την κατεύθυνση, όπου ο στόχος είναι να ανιχνεύσουμε κάθε αντικείμενο που υπάρχει στην εικόνα και να αναθέσουμε τις αντίστοιχες ετικέτες για να περιγράψουμε τα περιεχόμενα της εικόνας.

Εμπνευσμένοι από το μοντέλο συν-εμφάνισης λέξης που προτάθηκε από τους Mori et al. το 1999, όλο και περισσότεροι μελετητές στρέφονται στη διεξαγωγή μελετών σχετικά με την επισημείωση εικόνων αναπτύσσοντας πιθανοτικούς και μη πιθανοτικούς αλγορίθμους, μεθόδους βασισμένες στη μηχανική μάθηση και μεθόδους βασισμένες στην ανάκτηση εικόνας, μεθόδους επιβλεπόμενης, ημι-επιβλεπόμενης και μη επιβλεπόμενης μάθησης. Αυτά τα επιτεύγματα έχουν ενισχύσει την ανάπτυξη της AIA σε μεγάλο βαθμό κατά τη διάρκεια των τελευταίων δύο δεκαετιών.

Στη βιβλιογραφία απουσιάζει μέχρι και σήμερα, μια γενική ταξινόμηση και βαθιά ανασκόπηση των μεθόδων AIA. Η πρώτη εμπεριστατωμένη επισκόπηση των μεθόδων επισημείωσης επιχειρήθηκε από τους Zhang et al. (2012). Η έρευνά τους που καλύπτει τις εξελίξεις στον τομέα της AIA τις δύο πρώτες δεκαετίες, αναλύει βασικά ζητήματα όπως η εξαγωγή χαρακτηριστικών και οι μέθοδοι μηχανικής μάθησης. Μία πρόσφατη μελέτη των Cheng et al. (2018) παρουσιάζει μια πολύ γενικευμένη κατηγοριοποίηση των μεθόδων AIA. Το άρθρο τους επικεντρώνεται στην παρουσίαση και επεξήγηση των βασικών μεθόδων AIA. Όμως η AIA αποτελεί ένα τεράστιο διεπιστημονικό ερευνητικό πεδίο που ενσωματώνει τα επιτεύγματα από την εξόρυξη δεδομένων (data mining), τη σημασιολογική ανάλυση

(semantic analysis), την επεξεργασία φυσικής γλώσσας (Natural Language Processing), την αυτόματη βαθιά κατανόηση κειμένου (Automatic Deep Understanding - ADU) την ανάλυση και αναγνώριση κειμένου, τα συστήματα πολυμέσων, τη μηχανική μάθηση, τη μηχανική όραση και ακόμη και τη βιολογία και τη στατιστική (Cheng et al., 2018).

Συνεπώς, κάθε προσπάθεια ερμηνείας αυτών των μεθόδων απαιτεί μια σαφώς καθορισμένη κατηγοριοποίηση για να καλύψει όλες τις πτυχές των μεθόδων της AIA. Σε αντίθεση με τη μελέτη των Cheng et al. (2018), οι Bhagat and Choudhary (2018) υιοθετούν μια διαφορετική προσέγγιση για την ταξινόμηση των μεθόδων της AIA, η οποία είναι πιο εξειδικευμένη και αναλυτική επιχειρώντας να καλύψουν κάθε πτυχή της AIA. Στα πλαίσια της παρούσας εργασίας κρίνουμε σκόπιμο να παρουσιάσουμε τους κύριους κλάδους της AIA ταξινομημένους με βάση τέσσερις βασικούς άξονες: τη μέθοδο μάθησης, το σύνολο δεδομένων εκπαίδευσης, το παραγόμενο μοντέλο και το μήκος της παραγόμενης επισημείωσης. Οι καινοτόμες τεχνικές που βασίζονται στη βαθιά μάθηση και ειδικότερα τα βαθιά συνελικτικά νευρωνικά δίκτυα εξετάζονται ξεχωριστά.

4.1. Μέθοδοι επισημείωσης βασισμένες στη μάθηση

Τα οπτικά χαρακτηριστικά δεν επαρκούν για την επισημείωση όλων των αντικειμένων που υπάρχουν στην εικόνα. Επιπλέον, τα οπτικά στοιχεία δεν αντιπροσωπεύουν το περιεχόμενο της εικόνας στο σύνολό της. Για να επισημάνει με ακρίβεια τα περιεχόμενα της εικόνας, το μοντέλο θα πρέπει να μάθει χαρακτηριστικά από διάφορες διαθέσιμες πηγές. Η συσχέτιση μεταξύ των ετικετών είναι μια τέτοια πηγή χαρακτηριστικών. Η αξιοποίηση της συσχέτισης ετικετών βοηθά επίσης στην αντιμετώπιση του προβλήματος ελλιπών ετικετών και ετικετών με θόρυβο. Τα βασισμένα στη μάθηση μοντέλα επικεντρώνονται κυρίως στην εκπαίδευση του μοντέλου από πολλαπλές πηγές του συνόλου εκπαίδευσης. Τα περισσότερα από τα μοντέλα μάθησης εκμεταλλεύονται το χώρο των χαρακτηριστικών εισόδου και έχουν την ικανότητα να αντιμετωπίσουν το πρόβλημα των ελλιπών ετικετών. Η μάθηση από διάφορες πηγές (οπτικά χαρακτηριστικά, χαρακτηριστικά κειμένου) βελτιώνει τη δυνατότητα γενίκευσης του μοντέλου (Bhagat and Choudhary, 2018).

Όταν ένα μοντέλο διερευνά πολλαπλές αναπαραστάσεις των χαρακτηριστικών και αποθηκεύει τις ιδιότητες κάθε όψης ξεχωριστά σε ένα διάνυσμα χαρακτηριστικών, αυτά τα διανύσματα χαρακτηριστικών πρέπει να συνδυάζονται έτσι ώστε το ενιαίο διάνυσμα

χαρακτηριστικών που θα προκύψει, να έχει φυσική σημασία και να έχει μεγαλύτερη διακριτική ισχύ από τα χαρακτηριστικά από μία μόνο αναπαράσταση.

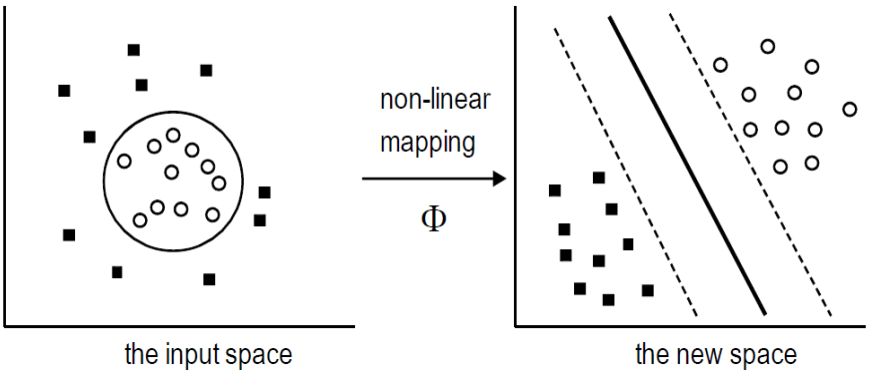
Τα μοντέλα επισημείωσης με βάση τη μάθηση μπορεί να ταξινομηθούν στις ακόλουθες κατηγορίες: παραδοσιακή μάθηση μονής ετικέτας με δυαδική ταξινόμηση (single-label), μάθηση πολλαπλών ετικετών (multi-label), μάθηση πολλαπλών στιγμιότυπων πολλαπλών ετικετών (multi-instance multi-label), μάθηση πολλαπλών όψεων (multi-view) και μάθηση μετρικής απόστασης (distance metric).

4.1.1. Επισημείωση μονής ετικέτας με δυαδική ταξινόμηση

Στην προσέγγιση αυτή, τα χαρακτηριστικά χαμηλού επιπέδου εξάγονται από το περιεχόμενο της εικόνας και τα χαρακτηριστικά τροφοδοτούνται απευθείας σε ένα συμβατικό δυαδικό ταξινομητή που δίνει μια ψήφο «ναι» ή «όχι». Η έξοδος του ταξινομητή είναι η σημασιολογική έννοια που χρησιμοποιείται για την επισημείωση της εικόνας. Η ιδέα της επισημείωσης μονής ετικέτας είναι ισοδύναμη με τη συλλογική επισημείωση, δηλαδή αντί να επισημειώνονται οι εικόνες μεμονωμένα, οι εικόνες πρώτα ομαδοποιούνται και στη συνέχεια επισημειώνονται συλλογικά. Τα πιο διαδεδομένα εργαλεία μηχανικής μάθησης περιλαμβάνουν μηχανές διανυσμάτων υποστήριξης (SVM), τεχνητά νευρωνικά δίκτυα (ANN) και δέντρα αποφάσεων (DT). Στη συνέχεια, εξετάζουμε συνοπτικά κάθε μία από αυτές τις κατηγορίες.

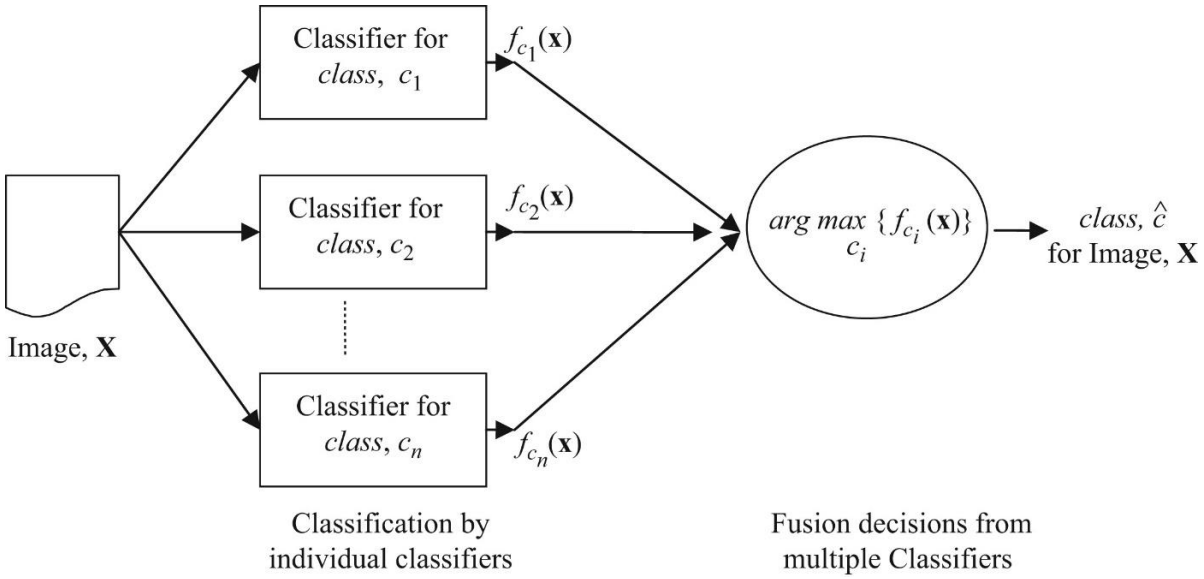
4.1.1.1. Επισημείωση εικόνας με Μηχανές Διανυσμάτων Υποστήριξης

Μια μηχανή διανυσμάτων υποστήριξης (SVM) είναι ένας επιβλεπόμενος ταξινομητής. Έχει αποδειχθεί ότι έχει μεγάλη αποτελεσματικότητα στην ταξινόμηση δεδομένων μεγάλης διάστασης, ειδικά όταν το σύνολο των δεδομένων εκπαίδευσης είναι μικρό. Ένας SVM μπορεί να ταξινομήσει τόσο γραμμικά όσο και μη γραμμικά δεδομένα λόγω της χρήσης των συναρτήσεων απεικόνισης πυρήνα. Το πλεονέκτημα των SVM έναντι άλλων ταξινομητών είναι ότι επιτυγχάνει να εντοπίσει τα βέλτιστα όρια των κλάσεων, βρίσκοντας τη μέγιστη απόσταση μεταξύ των κλάσεων (Cortes and Vapnik,1995), όπως φαίνεται στο σχήμα 9. Οι SVM έχουν εφαρμοστεί με επιτυχία σε μια σειρά από προβλήματα ταξινόμησης, όπως ταξινόμηση κειμένου, αναγνώριση αντικειμένου και επισημείωση εικόνας.



Εικόνα 9. Η λειτουργία του μοντέλου SVM (Tsai and Hung, 2008)

Ένας ταξινομητής SVM λειτουργεί με την εύρεση υπερεπεπιπέδου από ένα σύνολο δειγμάτων εκπαίδευσης για να τα διαχωρίσει. Κάθε δείγμα εκπαίδευσης αντιπροσωπεύεται με ένα διάνυσμα χαρακτηριστικών και μια ετικέτα κλάσης. Το υπερεπίπεδο υπολογίζεται με τέτοιο τρόπο ώστε να μπορεί να διαχωρίσει το μεγαλύτερο μέρος των δειγμάτων της ίδιας κλάσης από όλα τα άλλα δείγματα. Ένας SVM είναι βασικά ένας δυαδικός ταξινομητής. Ωστόσο, η αυτόματη ταξινόμηση και επισημείωση εικόνας απαιτεί ένα ταξινομητή πολλαπλών κλάσεων.



Εικόνα 10. Ταξινομητής πολλαπλών κλάσεων που χρησιμοποιεί πολλούς δυαδικούς SVM ταξινομητές (Zhang, et al., 2012).

Η πιο κοινή προσέγγιση είναι η εκπαίδευση ενός ξεχωριστού SVM για κάθε έννοια, με κάθε SVM να παράγει μια τιμή πιθανότητας. Κατά τη διάρκεια της φάσης δοκιμών, οι αποφάσεις όλων των ταξινομητών συγχωνεύονται για να προκύψει η τελική ετικέτα κλάσης μιας δοκιμαστικής εικόνας. Το σχήμα 10 δείχνει αυτή τη διαδικασία. Ο ταξινομητής είναι μια

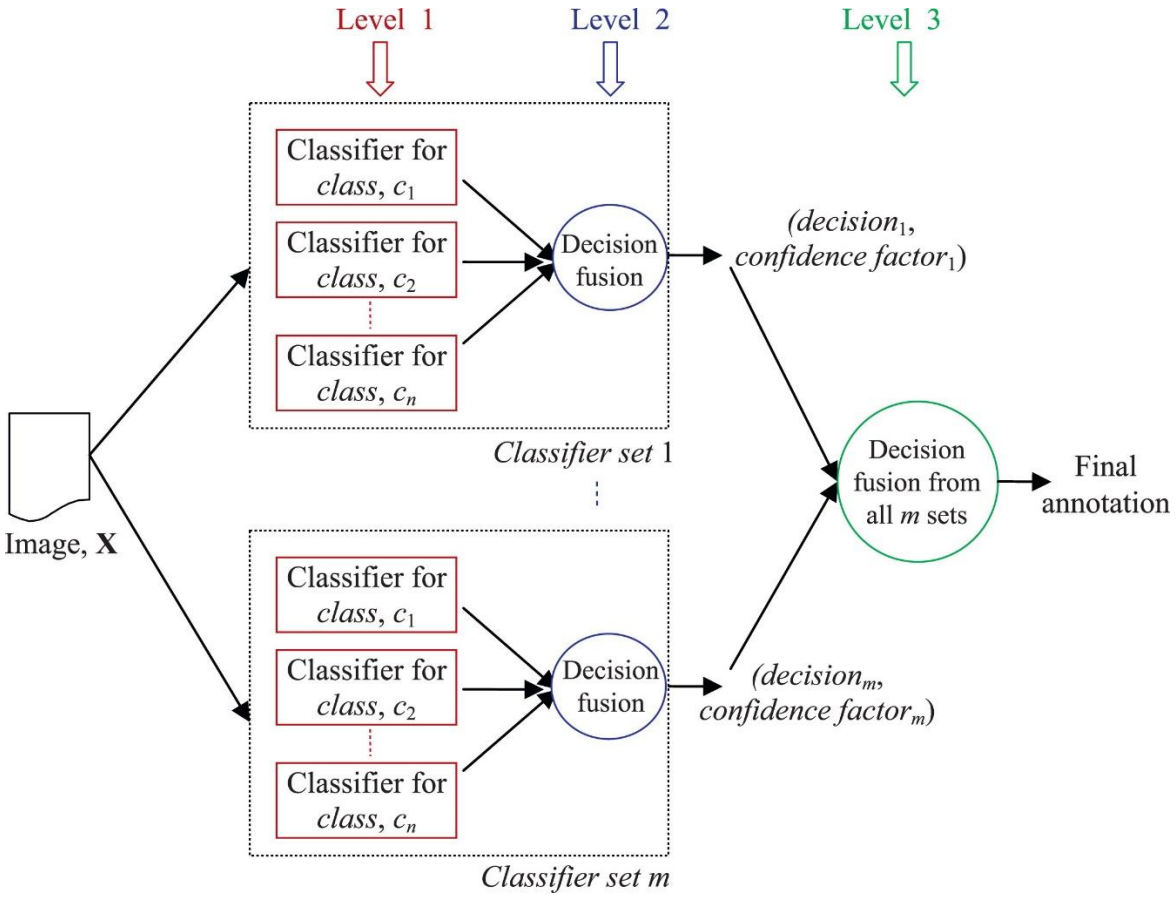
διαδικασία δύο επιπέδων. Το βασικό επίπεδο αποτελείται από πολλούς δυαδικούς ταξινομητές και το δεύτερο επίπεδο συγχωνεύει τις αποφάσεις από τους ταξινομητές του βασικού επιπέδου (Zhang, et al., 2012).

Οι Chapelle et al. (1999) χρησιμοποιούν αυτό το βασικό πλαίσιο για να εκπαιδεύσουν 14 ταξινομητές SVM για 14 κλάσεις. Οι εικόνες αντιπροσωπεύονται με ιστογράμματα HSV διάστασης $n = 4096$. Για την εκπαίδευση ενός SVM για μια συγκεκριμένη έννοια, οι εικόνες του συνόλου εκπαίδευσης που ανήκουν σε αυτή την έννοια θεωρούνται θετικά δείγματα ενώ τα άλλα θεωρούνται ως αρνητικά δείγματα. Επομένως, κάθε εκπαιδευόμενος ταξινομητής μπορεί να θεωρηθεί ως «ένας εναντίον όλων» (one-vs-all) ταξινομητής. Κατά τη διάρκεια της δοκιμής, κάθε ταξινομητής δημιουργεί μια πιθανοτική απόφαση. Η κλάση με τη μέγιστη πιθανότητα επιλέγεται ως έννοια της εικόνας δοκιμής.

Η προσέγγιση αυτή λειτουργεί καλά για μικρό αριθμό εννοιών. Η ποιότητα της ταξινόμησης υποβαθμίζεται με την αύξηση του αριθμού των εννοιών λόγω της αύξησης του θορύβου και της ανισορροπίας των κλάσεων στα δεδομένα εκπαίδευσης. Για να είναι πιο αποδοτικές, μερικές προσεγγίσεις χρησιμοποιούν πολλαπλές ομάδες ταξινομητών SVM, όπως φαίνεται στο σχήμα 11 (Zhang et al., 2012). Κάθε ομάδα SVM είναι παρόμοια με έναν ταξινομητή πολλαπλών κλάσεων, που παρουσιάζεται στο σχήμα 10. Κάθε ομάδα ταξινομητών ταξινομεί ανεξάρτητα μια εικόνα εισόδου. Η τελική απόφαση αποτελεί τη συγχώνευση όλων των αποφάσεων όλων των συνόλων. Στη βιβλιογραφία αναφέρονται πολλές ερευνητικές εργασίες που χρησιμοποιούν αυτό το πολυεπίπεδο πλαίσιο και διαφέρουν ανάλογα με τον τρόπο που εκπαιδεύονται οι επιμέρους ταξινομητές στο πρώτο επίπεδο και τον τρόπο που οι αποφάσεις συγχωνεύονται σε επόμενα επίπεδα.

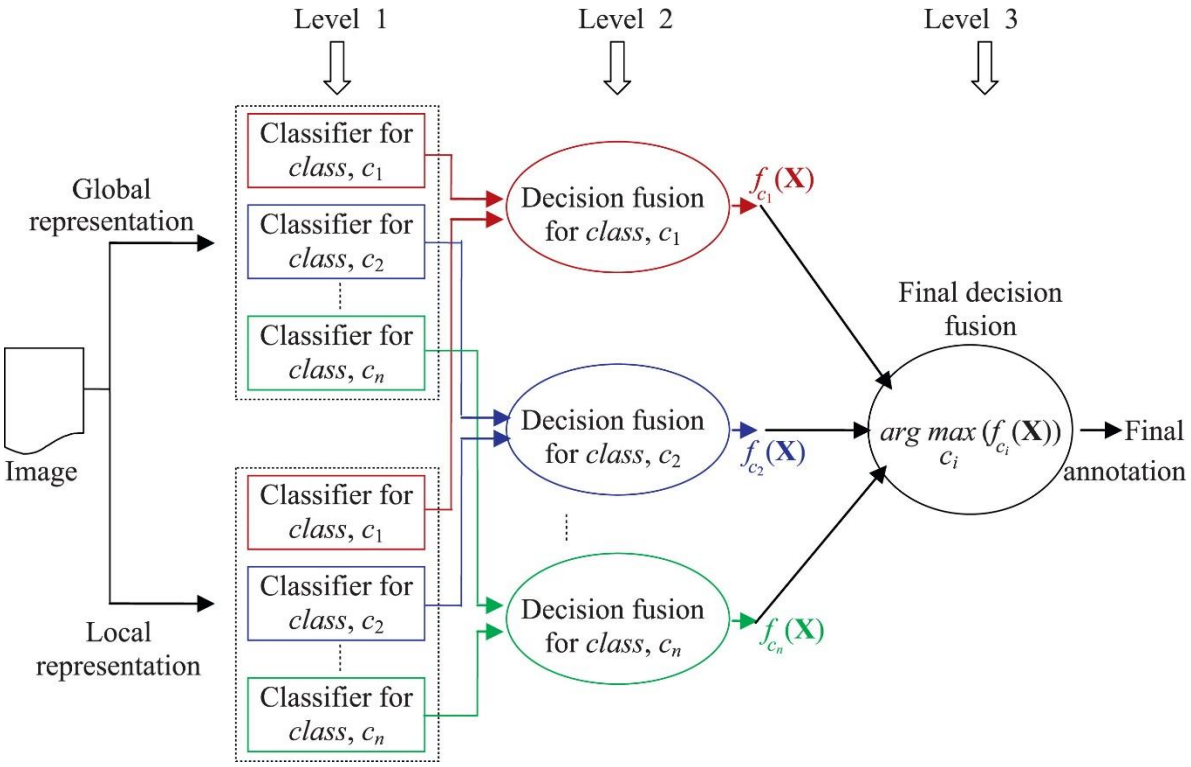
Οι Goh et αl. (2005) χρησιμοποιούν την παραπάνω προσέγγιση τριών επιπέδων, που παρουσιάζεται στο σχήμα 11, για να ταξινομήσουν τις εικόνες σε μία από 116 έννοιες κλάσεων. Κατά τη διάρκεια της εκπαίδευσης, κάθε ομάδα ταξινομητών εκπαιδεύεται χρησιμοποιώντας διαφορετικό υποσύνολο δειγμάτων εκπαίδευσης. Κατά τη στιγμή της επισημείωσης, κανονικοποιούν τις πιθανοτικές εξόδους των ταξινομητών του πρώτου επιπέδου πριν τις χρησιμοποιήσουν στη διαδικασία συγχώνευσης του δεύτερου επιπέδου. Κατά τη διάρκεια της διαδικασίας συγχώνευσης, υπολογίζουν επίσης έναν παράγοντα εμπιστοσύνης (confidence factor), επιπλέον της απόφασης με τη μεγαλύτερη πιθανότητα. Ο

παράγοντας εμπιστοσύνης είναι συνάρτηση τόσο της υψηλότερης πιθανοτικής τιμής όσο και της διαφοράς μεταξύ των δύο πιο πιθανών αποφάσεων. Στο τρίτο επίπεδο, οι παράγοντες εμπιστοσύνης της ίδιας έννοιας προστίθενται μαζί. Η έννοια με τη μέγιστη αθροιστικά εμπιστοσύνη είναι η τελική απόφαση. Με την προσέγγιση αυτή η θορυβώδης έξοδος από ένα σύνολο ταξινομητών αντισταθμίζεται χρησιμοποιώντας τις αποφάσεις άλλων ομάδων ταξινομητών.



Εικόνα 11. Επισημείωση εικόνας με πολλαπλές ομάδες SVM ταξινομητών (Zhang et al., 2012).

Οι Qi and Han (2007) χρησιμοποιούν ένα παρόμοιο πλαίσιο με τους Goh et al. (2005) αλλά συγχωνεύουν τις αποφάσεις με διαφορετικό τρόπο όπως φαίνεται στο σχήμα 12. Χρησιμοποιούν τόσο συνολικά όσο και τοπικά οπτικά χαρακτηριστικά σε δύο διαφορετικές ομάδες SVM. Οι ίδιες εικόνες χρησιμοποιούνται για την εκπαίδευση και των δύο κατηγοριών ταξινομητών. Το σύνολο SVM με αναπαράσταση εικόνων βάσει συνολικών χαρακτηριστικών λειτουργεί με τον ίδιο τρόπο που φαίνεται στο σχήμα 10. Για το άλλο σύνολο SVM, κάθε εικόνα αντιπροσωπεύεται με τα χαρακτηριστικά της περιοχής ενδιαφέροντος. Αντί να συγχωνεύουν τις αποφάσεις σύνολο προς σύνολο, τις συνδυάζουν ταξινομητή προς ταξινομητή (ανά κλάση). Επομένως, αυτή η προσέγγιση μπορεί να αντισταθμίσει τους περιορισμούς ενός τύπου οπτικών χαρακτηριστικών από τον άλλο.



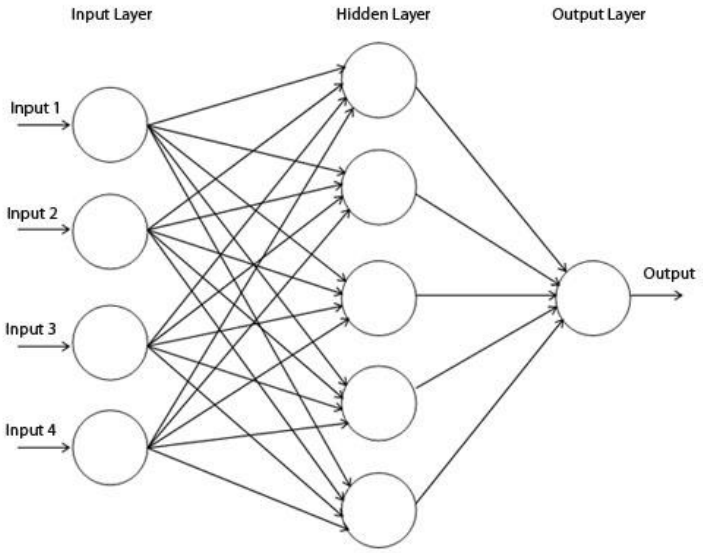
Εικόνα 12. Οι SVM ταξινομητές πολλαπλών κλάσεων τριών επιπέδων που χρησιμοποιούν οι Qi και Han (Zhang et al., 2012).

Οι ταξινομητές SVM έχουν παρουσιάσει αξιοσημείωτες επιδόσεις στον τομέα της επισημείωσης εικόνων. Το πλεονέκτημα είναι ότι ένας SVM μπορεί να μάθει από ένα μικρό σύνολο δειγμάτων επειδή χρειάζεται μόνο τα δείγματα (γνωστά ως «διανύσματα υποστήριξης») κοντά στο διαχωριστικό υπερεπίπεδο. Ωστόσο, οι SVM παρουσιάζουν πρόβλημα ανισορροπίας κλάσης, πράγμα που σημαίνει ότι έχουν κακή απόδοση σε δεδομένα που δεν έχουν υπολογιστεί ισόρροπα. Δυστυχώς, η ανισορροπία κλάσεων είναι ένα κοινό φαινόμενο στα δεδομένα εικόνας. Για παράδειγμα, για ένα συγκεκριμένο θέμα, ο αριθμός των αρνητικών δειγμάτων είναι συχνά πολύ μεγαλύτερος από τον αριθμό των θετικών δειγμάτων. Επιπλέον, τα θετικά δείγματα είναι σχετικά μεταξύ τους, αλλά κάθε αρνητικό δείγμα συχνά ανήκει σε ξεχωριστές σημασιολογικές ομάδες. Αυτά τα προβλήματα υποβαθμίζουν την ποιότητα των ταξινομητών.

4.1.1.2. Επισημείωση εικόνας με τεχνητά νευρωνικά δίκτυα

Ένα τεχνητό νευρωνικό δίκτυο (Artificial neural network – ANN) είναι μία μέθοδος μάθησης που μπορεί να μάθει από παραδείγματα και μπορεί να αποφασίσει για ένα νέο δείγμα. Ένα ANN αποτελείται από πολλαπλά επίπεδα διασυνδεδεμένων κόμβων, τα οποία είναι επίσης γνωστά ως νευρώνες ή perceptrons. Επομένως, ένα ANN ονομάζεται επίσης πολλαπλών επιπέδων perceptron (multilayer perceptron – MLP). Το πρώτο επίπεδο είναι το επίπεδο

εισόδου που έχει αριθμό νευρώνων ίσο με τη διάσταση του δείγματος εισόδου. Ο αριθμός των νευρώνων στο επίπεδο εξόδου είναι ίσος με τον αριθμό των κλάσεων. Αυτό σημαίνει ότι ένα ANN μπορεί να μάθει πολλαπλές κλάσεις ταυτόχρονα, αν και υπάρχουν ANN μιας μόνο κλάσης (για ταξινόμηση σε μία μόνο κλάση).



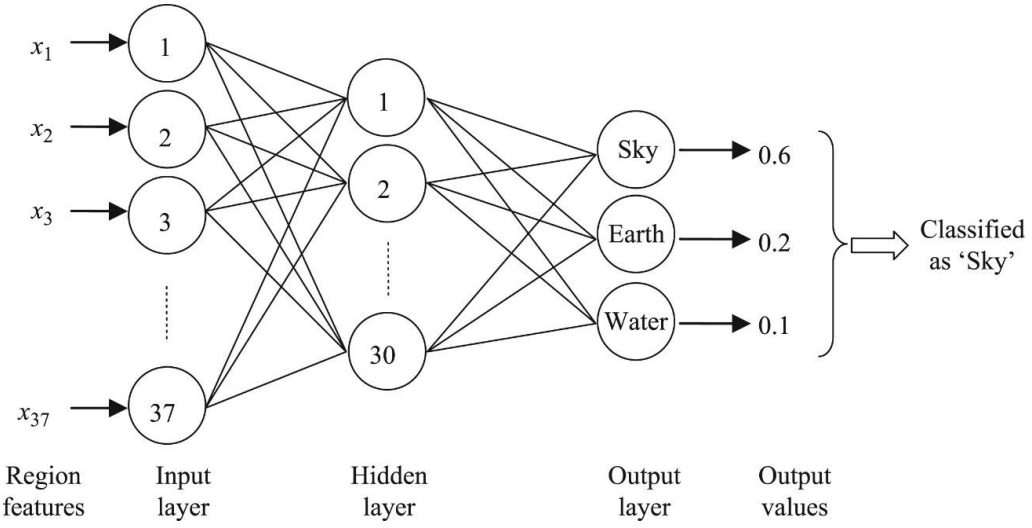
Εικόνα 13. Νευρωνικό δίκτυο τριών στρωμάτων (Tsai and Hung, 2008)

Ο αριθμός των κρυφών επιπέδων και ο αριθμός των νευρώνων σε κάθε κρυφό επίπεδο αποτελούν ανοικτά ζητήματα στα νευρωνικά δίκτυα και η επιλογή τους γίνεται συνήθως εμπειρικά. Οι συνδέσεις μεταξύ νευρώνων (συνάψεις) διαφορετικών επιπέδων συνοδεύονται από βάρη. Κάθε νευρώνας λειτουργεί ως στοιχείο επεξεργασίας και διέπεται από μια συνάρτηση ενεργοποίησης η οποία παράγει εξόδους βάσει των βαρών των συνδέσεων και των εξόδων των νευρώνων στα προηγούμενα επίπεδα. Κατά τη διάρκεια της εκπαίδευσης, το ANN μαθαίνει τα βάρη των συνδέσεων έτσι ώστε να ελαχιστοποιείται το συνολικό μαθησιακό σφάλμα. Κατά την ταξινόμηση ενός νέου δείγματος, κάθε νευρώνας εξόδου παράγει ένα μέτρο εμπιστοσύνης και η κλάση που αντιστοιχεί στο μέγιστο μέτρο δηλώνει την απόφαση για το δείγμα (Tsai and Hung, 2008, Zhang et al., 2012).

Ένα ANN μπορεί να χρησιμοποιηθεί τόσο για ρητή ταξινόμηση εικόνων, περιοχών ή εικονοστοιχείων, όσο και για έμμεση αντιστοίχιση ασαφών αποφάσεων σε εικόνες. Οι Frate et al. (2007) χρησιμοποιούν ένα ANN τεσσάρων επιπέδων για την ταξινόμηση εικονοστοιχείων δορυφορικών εικόνων σε μία από τις τέσσερις κατηγορίες: βλάστηση, άσφαλτο, κτίρια και γυμνό έδαφος. Με βάση τη βέλτιστη πειραματική απόδοση,

χρησιμοποιούν ένα δίκτυο από δύο κρυμμένα επίπεδα, το καθένα από τα οποία αποτελείται από 20 νευρώνες.

Οι Kuroda and Hagiwara (2002) χρησιμοποιούν τέσσερα διαφορετικά ANN τριών επιπέδων για ιεραρχική ταξινόμηση περιοχών εικόνας. Οι αριθμοί των νευρώνων που χρησιμοποιούνται στα κρυμμένα στρώματα των τεσσάρων δικτύων είναι 30, 10, 20 και 20, αντίστοιχα. Το σχήμα 14 δείχνει πώς το πρώτο δίκτυο ταξινομεί μια περιοχή εικόνας σε μία από τις τρεις ευρείες κατηγορίες όπως ο ουρανός, το νερό και η γη. Τα διανύσματα, διάστασης 37, των χαρακτηριστικών περιοχής τροφοδοτούνται στο στρώμα εισόδου. Κάθε κόμβος του στρώματος εξόδου αντιστοιχεί σε μία από τις κλάσεις, για παράδειγμα, ουρανό, νερό και γη και παράγει μια τιμή πιθανότητας. Η κλάση της περιοχής εισόδου καθορίζεται από την μέγιστη πιθανότητα. Οι περιοχές του ουρανού και της γης ταξινομούνται περαιτέρω σε πιο συγκεκριμένες κατηγορίες χρησιμοποιώντας τα άλλα δύο ANN. Το τέταρτο ANN δεν ταξινομεί καμία περιοχή. Αντ' αυτού, συνδυάζει μια εικόνα με ένα διάνυσμα 18 διαστάσεων που περιέχει αμφίβολες λέξεις εντυπώσεων και συγκεκριμένα ουσιαστικά των ταξινομημένων περιοχών εικόνας, όπως "βουνό" από την κατηγορία της γης ή "συννεφιασμένος ουρανός" από την κατηγορία του ουρανού (Tsai and Hung, 2008).



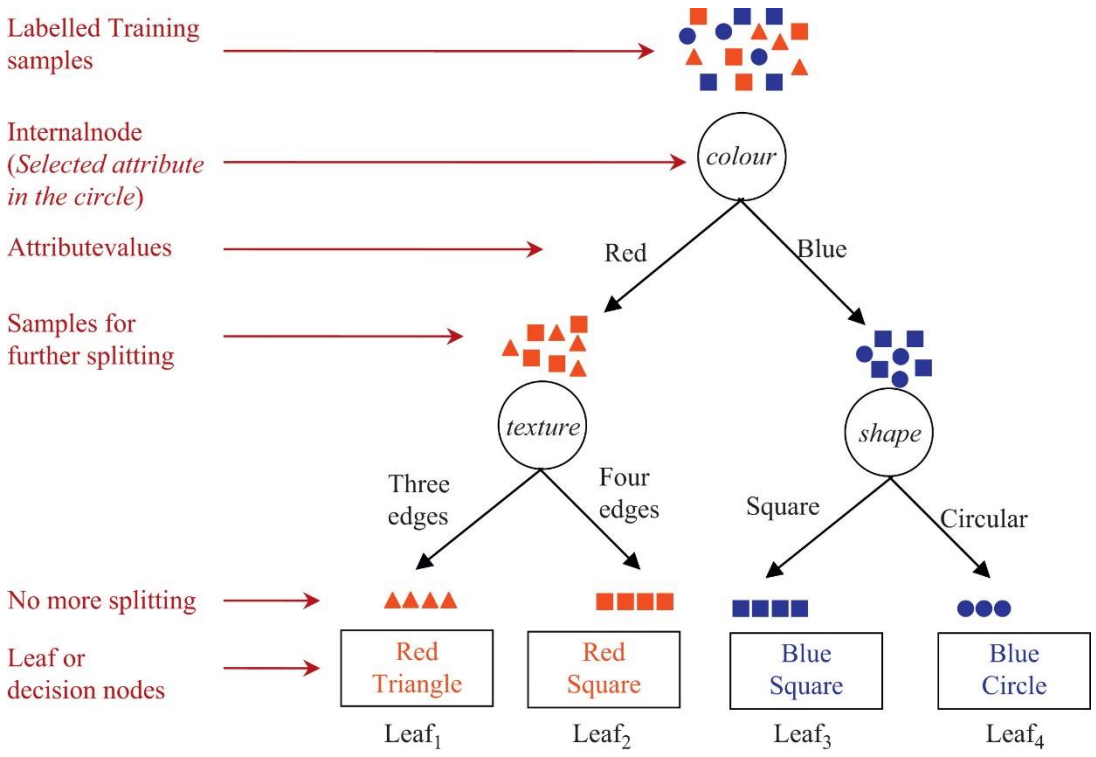
Εικόνα 14. Ταξινόμηση μίας περιοχής με χρήση ενός ΤΝΔ (Kuroda and Hagiwara, 2002).

Το νευρωνικό δίκτυο έχει το πλεονέκτημα ότι οι έξοδοι των νευρώνων του στρώματος εξόδου καθορίζονται από τα προηγούμενα στρώματα και τις συνδέσεις. Δεν χρειάζεται άλλη ρύθμιση παραμέτρων ή οποιαδήποτε παραδοχή σχετικά με την κατανομή των χαρακτηριστικών. Ωστόσο, υπάρχουν πολλά ουσιαστικά ζητήματα με τα ANN. Πρώτον, η ακρίβεια της ταξινόμησης εξαρτάται από τον αριθμό των κρυμμένων στρωμάτων και των νευρώνων. Δεύτερον, στις περισσότερες ερευνητικές εργασίες με ANN, οι αριθμοί κρυφών στρωμάτων

και νευρώνων δεν δικαιολογούνται. Τρίτον, η επιλογή κατάλληλων συναρτήσεων ενεργοποίησης για τους νευρώνες είναι επίσης ένα ζήτημα. Τέταρτον, η εκπαίδευση (η εύρεση των βέλτιστων βαρών συνδέσεων) διαρκεί πολύ και μπορεί να εγκλωβιστεί σε τοπικά μέγιστα. Τέλος, ένα ANN λειτουργεί σαν ένα «μαύρο κουτί» που σημαίνει ότι η ακριβής σχέση μεταξύ εισόδου και εξόδου δεν είναι διαφανής και είναι δύσκολο να ερμηνευτεί (Zhang et al., 2012).

4.1.1.3. Επισημείωση εικόνας με δέντρα απόφασης

Ένα δέντρο απόφασης (Decision Tree - DT) είναι ένα εργαλείο λήψης αποφάσεων ή ταξινόμησης πολλαπλών σταδίων. Ανάλογα με τον αριθμό των αποφάσεων που λαμβάνονται σε κάθε εσωτερικό κόμβο του δέντρου, ένα DT μπορεί να ονομαστεί δυαδικό ή n-αδικό δέντρο. Σε αντίθεση με τα άλλα μοντέλα ταξινόμησης των οποίων οι σχέσεις εισόδου-εξόδου είναι δύσκολο να περιγραφούν, η σχέση εισόδου-εξόδου σε ένα DT μπορεί να εκφραστεί χρησιμοποιώντας ανθρώπινα κατανοητούς κανόνες, *if-then* ($AN - TOTE$).



Εικόνα 15. Μάθηση με δέντρο απόφασης (Zhang et al., 2012).

Ένα DT εκπαιδεύεται χρησιμοποιώντας ένα σύνολο επισημειωμένων δειγμάτων εκπαίδευσης. Τα δείγματα αντιπροσωπεύονται από έναν αριθμό χαρακτηριστικών. Κατά τη διάρκεια της εκπαίδευσης, ένα DT κατασκευάζεται με την αναλογική κατανομή των

δειγμάτων εκπαίδευσης σε ομάδες που δεν αλληλεπικαλύπτονται και κάθε φορά που τα δείγματα διαιρούνται, το χαρακτηριστικό που χρησιμοποιείται για τη διαίρεση απορρίπτεται. Η διαδικασία συνεχίζεται μέχρις ότου όλα τα δείγματα μιας ομάδας που ανήκουν στην ίδια κλάση ή το δέντρο φτάσει στο μέγιστο βάθος του, όταν δεν υπάρχει κανένα χαρακτηριστικό για να το διαχωρίσει. Η διαδικασία αποτυπώνεται στην εικόνα 15. Το δέντρο έχει δύο τύπους κόμβων: εσωτερικούς και κόμβους φύλλων. Κάθε εσωτερικός κόμβος συνδέεται με μια απόφαση που διέπεται από ένα συγκεκριμένο χαρακτηριστικό, το οποίο διαιρεί τα δείγματα της εκπαίδευσης με τον πιο αποτελεσματικό τρόπο (Zhang et al., 2012). Κάθε κόμβος φύλλων αντιπροσωπεύει την έκβαση (κλάση) των περισσότερων δειγμάτων που ακολουθούν τη διαδρομή από τη ρίζα του δέντρου στο αντίστοιχο φύλλο. Οι κόμβοι των φύλλων μπορούν να εκφραστούν με μοναδικούς κανόνες $if - then - else$.

Για την επισημείωση ενός νέου δείγματος, το δέντρο διασχίζεται από τον κόμβο ρίζας σε έναν κόμβο φύλλων χρησιμοποιώντας την τιμή χαρακτηριστικού του νέου δείγματος. Η απόφαση για ένα δείγμα είναι το αποτέλεσμα του κόμβου των φύλλων όπου φθάνει το δείγμα.

Οι Huang et al. (1998) κατασκευάζουν ένα δέντρο ταξινόμησης για ιεραρχική ταξινόμηση έντεκα κατηγοριών. Αναφέρουν ότι η χρήση του χρωματικού διαγράμματος συσχέτισης (color correlogram) αποδίδει καλύτερα αποτελέσματα από το γενικό χρωματικό ιστόγραμμα και το δέντρο ταξινόμησης υπερβαίνει τον παραδοσιακό ταξινομητή k - πλησιέστερων γειτόνων.

Εκτός από την ταξινόμηση με βάση ένα μόνο δέντρο, οι Máree et al. (2005) χρησιμοποιούν σύνολο πολλαπλών DTs για επισημείωση και ταξινόμηση εικόνων. Κάθε DT είναι παρόμοιο με το κλασσικό DT που φαίνεται στο σχήμα 15, όμως, καθένα από αυτά εκπαιδεύεται διαφορετικά. Κατά τη διάρκεια του διαχωρισμού ενός εσωτερικού κόμβου, ένα χαρακτηριστικό επιλέγεται τυχαία αντί για την επιλογή του καλύτερου. Το δέντρο που χτίζεται με αυτό τον τρόπο, ονομάζεται τυχαίο δέντρο (random tree). Δημιουργείται ένα σύνολο τυχαίων δέντρων με αυτόν τον τρόπο. Κατά την ταξινόμηση, ένα δοκιμαστικό δείγμα διαβιβάζεται μέσω όλων των τυχαίων δέντρων και χρησιμοποιείται μια πλειοψηφική τιμή (majority vote) για τον προσδιορισμό της ετικέτας κλάσης της εικόνας.

Σε σύγκριση με άλλες μεθόδους μάθησης, ένα DT είναι απλό στην ερμηνεία και κατανόηση του και μπορεί να μάθει με μικρό αριθμό δειγμάτων. Είναι επίσης ισχυρό για ατελή και θορυβώδη δεδομένα. Ένα DT συνήθως απαιτεί διακριτές τιμές χαρακτηριστικών ως εισόδους. Στη βιβλιογραφία χρησιμοποιούνται διάφοροι αλγόριθμοι δημιουργίας δέντρων απόφασης, συμπεριλαμβανομένων των ID3, C4.5, και CART. Αυτοί οι αλγόριθμοι DT διαφέρουν ανάλογα με τον τύπο των χαρακτηριστικών, τα κριτήρια επιλογής χαρακτηριστικών, το αποτέλεσμα κ.λπ. Αν και οι αλγόριθμοι C4.5 και CART μπορούν να λειτουργήσουν με συνεχή χαρακτηριστικά, αποδίδουν φτωχότερα σε σύγκριση με διακριτά χαρακτηριστικά (Zhang et al., 2012).

Ένα άλλο ζήτημα με τα DT είναι ότι οι αλγόριθμοι C4.5 και CART έχουν σχεδιαστεί για χαρακτηριστικά μονής τιμής. Δεν λειτουργούν για υψηλής διάστασης διανύσματα χαρακτηριστικών. Οι Liu et al. (2008) πρότειναν μια μέθοδο διακριτοποίησης χαρακτηριστικών βασισμένη σε ένα σημασιολογικό πρότυπο για δεδομένα εικόνας για την αντιμετώπιση του προβλήματος. Ωστόσο, αυτή η τεχνική είναι χρήσιμη μόνο όταν τα διανύσματα χαρακτηριστικών έχουν την ίδια διάσταση. Επομένως, είναι απαραίτητη μια πιο στιβαρή τεχνική διακριτοποίησης χαρακτηριστικών, η οποία μπορεί να συμβάλει στη διακριτοποίηση διανυσμάτων χαρακτηριστικών μεταβλητής διάστασης όπως ο περιγραφέας DCD του προτύπου MPEG-7.

Σε προσεγγίσεις επισημείωσης μονής ετικέτας (single labelling annotation) οι εικόνες ταξινομούνται σε διαφορετικές κλάσεις και κάθε κλάση επισημειώνεται με μια ετικέτα (που περιγράφει μία έννοια). Το μειονέκτημα αυτού του τύπου προσέγγισης είναι ότι δεν λαμβάνει υπόψη το γεγονός ότι πολλές εικόνες ανήκουν σε πολλές κλάσεις. Ένας τρόπος για να λυθεί αυτό το πρόβλημα είναι να επισημειωθεί κάθε κλάση με πολλές λέξεις-κλειδιά (πολλαπλές ετικέτες) που αντικατοπτρίζουν διαφορετικά θέματα – έννοιες – μέσα στην κλάση. Ένα άλλο ζήτημα με την επισημείωση μόνης ετικέτας είναι ότι οι εικόνες σε κάθε κλάση δεν κατατάσσονται, με συνέπεια τη μειωμένη ακρίβεια ανάκτησης (Bhagat and Choudhary, 2018).

4.1.2. Μάθηση πολλαπλών ετικετών - Multi-label learning (MLL)

Οι περισσότερες εικόνες περιέχουν περισσότερα από ένα αντικείμενα, επομένως μία ενιαία ετικέτα δεν περιγράφει σωστά το περιεχόμενο της εικόνας. Η μάθηση πολλαπλών ετικετών

(MLL) σημαίνει την εκχώρηση πολλαπλών ετικετών σε ένα στιγμιότυπο. Το μοντέλο που βασίζεται στη μάθηση πολλαπλών ετικετών χειρίζεται πολλαπλές κλάσεις ετικετών και αποδίδει πολλαπλές ετικέτες σε μια εικόνα βάσει του περιεχόμενού της. Η μάθηση πολλαπλών κλάσεων είναι μια διαφορετική μορφή μάθησης πολλαπλών ετικετών καθώς στην πρώτη, από τις πολλές κλάσεις μόνο μια κλάση έχει εκχωρηθεί σε ένα στιγμιότυπο ενώ στην τελευταία, μέσα από ένα σύνολο κλάσεων μπορούν να εκχωρηθούν πολλαπλές κλάσεις σε μια περίπτωση (στιγμιότυπο). Η MLL συνήθως υποθέτει ότι το σύνολο δεδομένων εκπαίδευσης είναι πλήρως επισημειωμένο, αν και η πρόσφατη πρόοδος στην κατεύθυνση της μάθησης πολλαπλών ετικετών με μερική επίβλεψη είναι αξιοσημείωτη. Έχουν προταθεί διάφορες μέθοδοι MLL βασισμένες σε ημι-επιβλεπόμενη μάθηση όπου το μοντέλο εκπαιδεύεται από ελλιπείς ετικέτες (Bhagat and Choudhary, 2018).

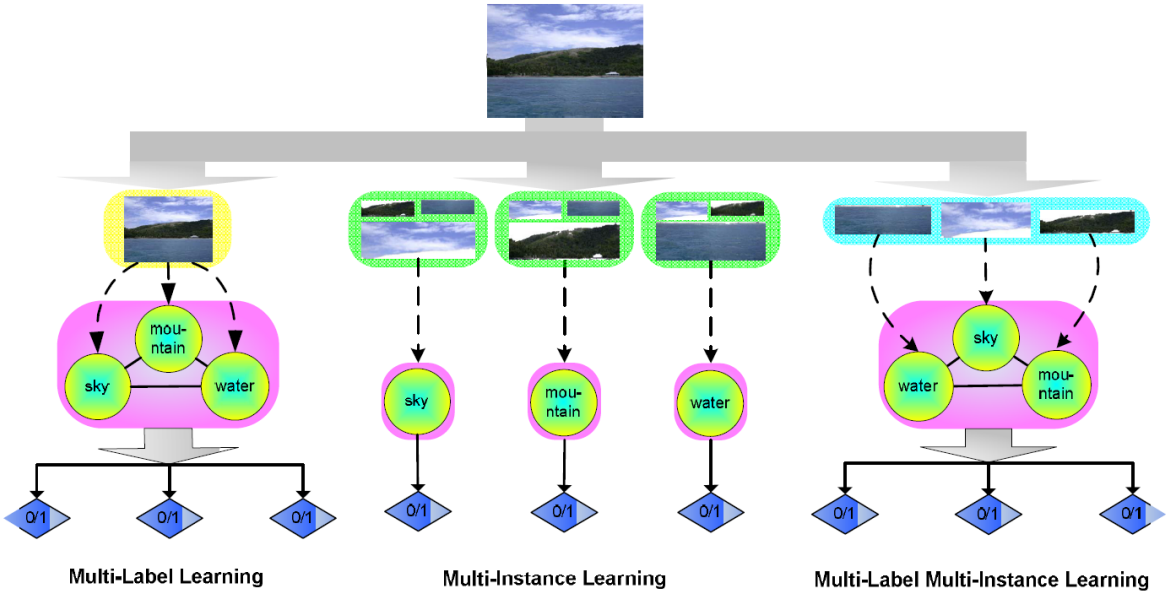
Οι τεχνικές MLL μπορούν να υλοποιηθούν με τη χρήση διακριτικών μοντέλων, μοντέλων βασισμένων στον πλησιέστερο γείτονα, μοντέλων βασισμένων σε γράφους και διάφορες άλλες τεχνικές που θα αναλυθούν στη συνέχεια. Μια λεπτομερής ανασκόπηση των μεθόδων MLL καταγράφεται στην εργασία των Zhang and Zhou (2014).

4.1.3. Μάθηση πολλαπλών στιγμιοτύπων – πολλαπλών ετικετών

Η μάθηση πολλαπλών στιγμιοτύπων (Multi-instance learning, MIL) είναι ένας τρόπος χειρισμού των αντικειμένων που περιγράφονται από πολλαπλά στιγμιότυπα (σχήμα 16). Στη προσέγγιση της MIL, μια εικόνα αναπαρίσταται με ένα «σάκο», που περιέχει μια σειρά στιγμιοτύπων που αντιστοιχούν στις περιοχές (αντικείμενα) της εικόνας. Εάν οποιαδήποτε από αυτά τα στιγμιότυπα σχετίζεται με μια ετικέτα, η εικόνα θα συσχετιστεί με την ετικέτα. Η προσέγγιση αυτή είναι γνωστή ως «σάκος χαρακτηριστικών» (bag of features) ή «σάκο περιοχών» (bag of regions) ή πολλαπλά στιγμιότυπα (multiple instances). Σε σύνθετες εφαρμογές, όπου μία εικόνα έχει διάφορα στιγμιότυπα (αντικείμενα), επομένως πολλαπλά διανύσματα χαρακτηριστικών και μόνο ένα από αυτά τα διανύσματα αναπαριστά αυτό το αντικείμενο, τότε η μάθηση ονομάζεται MIL (Bhagat and Choudhary, 2018).

Το σύνολο εκπαίδευσης στην προσέγγιση MIL έχει συνήθως ελλιπείς ετικέτες. Η ασάφεια που συνδέεται με το σύνολο δεδομένων εκπαίδευσης, όπου κάθε εικόνα έχει πολλά στιγμιότυπα, μπορεί να επιλυθεί χρησιμοποιώντας το MIL. Η παρουσία τουλάχιστον ενός στιγμιότυπου στο σάκο σημαίνει ότι ο σάκος είναι θετικός, διαφορετικά ο σάκος χαρακτηρίζεται αρνητικός.

Από την άλλη πλευρά, το MLL ασχολείται με την περίπτωση όπου ένα παράδειγμα ανήκει ταυτόχρονα σε περισσότερες από μία κλάσεις εξόδου. Στο MLL, ένα στιγμιότυπο μπορεί να ταξινομηθεί σε περισσότερες από μία κλάσεις, πράγμα που σημαίνει ότι οι κλάσεις δεν είναι αμοιβαία αποκλειόμενες (Bhagat and Choudhary, 2018).



Εικόνα 16. Σύγκριση τριών προσεγγίσεων βασισμένων στη μάθηση.
Από τα αριστερά προς τα δεξιά, η μάθηση πολλαπλών ετικετών (MLL), η μάθηση πολλαπλών στιγμιοτύπων (MIL) και το κοινό πλαίσιο μάθησης πολλαπλών στιγμιοτύπων πολλαπλών ετικετών (MIML). Το MLL καταγράφει τις συσχετίσεις των ετικετών, ενώ το MIL μοντελοποιεί τη σύνδεση μεταξύ ετικετών και περιοχών. Το πλαίσιο MIML μοντελοποιεί και οι δύο σχέσεις ταυτόχρονα (Zha et al., 2008).

Η μοντελοποίηση των σχέσεων μεταξύ ετικετών και περιοχών (αντί για ολόκληρη την εικόνα) μειώνει το θόρυβο στον αντίστοιχο χώρο χαρακτηριστικών και συνεπώς τα εκπαιδευόμενα μοντέλα είναι πιο ακριβή. Ωστόσο, οι μέθοδοι που ακολουθούν την προσέγγιση MIL, επικεντρώνονται κατά κύριο λόγο στο σενάριο της μεμονωμένης ετικέτας και τα προβλήματα πολλαπλών ετικετών πρέπει να προσεγγίζονται ετικέτα προς ετικέτα. Δηλαδή, οι συσχετίσεις των ετικετών δεν λαμβάνονται υπόψη σε αυτές τις μεθόδους ταξινόμησης που βασίζονται σε MIL (Zha et al., 2008).

Η μάθηση πολλαπλών στιγμιοτύπων πολλαπλών ετικετών (MIML) είναι ένας συνδυασμός MIL και MLL. Το MIML είναι ουσιαστικά ένα πρόβλημα μάθησης με επίβλεψη με ασαφές σύνολο εκπαίδευσης όπου κάθε αντικείμενο στο σετ εκπαίδευσης έχει πολλαπλά στιγμιότυπα και ανήκει σε πολλές κλάσεις. Η εικόνα επισημειώνεται με μια ετικέτα έννοιας εάν κάποια από τις περιοχές / στιγμιότυπα του σάκου συνδέεται με την ετικέτα. Ως αποτέλεσμα, μια εικόνα επισημειώνεται με πολλαπλές ετικέτες (Zhang et al., 2012).

Ένα από τα πρώιμα έργα που προτείνονται στο πλαίσιο του MIML, είναι οι μέθοδοι MIMLBoost και MIMLSVM (Zhou and Zhang, 2007) για την ταξινόμηση σκηνών. Η παραδοσιακή εποπτευόμενη μάθηση (μονού-στιγμιοτύπου, μονής-ετικέτας) είναι προφανώς μια εκφυλισμένη έκδοση της εκμάθησης πολλαπλών στιγμιοτύπων καθώς και μια εκφυλισμένη έκδοση της μάθησης πολλαπλών ετικετών, ενώ η παραδοσιακή εποπτευόμενη μάθηση, η εκμάθηση πολλαπλών στιγμιοτύπων και η εκμάθηση πολλαπλών ετικετών είναι όλες εκφυλισμένες εκδόσεις του MIML. Σε αυτή τη διαδικασία εκφυλισμού, μερικές από τις σημαντικές πληροφορίες στο σύνολο εκπαίδευσης μπορεί να χαθούν. Για να αντιμετωπιστεί αυτό το πρόβλημα, προτείνεται μια εκτεταμένη έκδοση, που ονομάζεται άμεση MIMLSVM (D-MIMLSVM), η οποία μπορεί να χειριστεί και την ανισορροπία κλάσεων, χρησιμοποιώντας είτε το MIL είτε το MLL ως γέφυρα (Zhou and Zhang, 2007).

4.1.4. Μάθηση πολλαπλών αναπαραστάσεων (Multi-view Learning)

Ένα αντικείμενο έχει διάφορα χαρακτηριστικά και μπορεί να αναπαρασταθεί χρησιμοποιώντας πολλαπλά σύνολα χαρακτηριστικών. Αυτά τα σύνολα χαρακτηριστικών λαμβάνονται από διαφορετικές αναπαραστάσεις (με βάση την υφή, το σχήμα, το χρώμα κ.λπ.). Αυτά τα διαφορετικά σύνολα χαρακτηριστικών είναι συνήθως ανεξάρτητα αλλά έχουν συμπληρωματική φύση και μόνο μία αναπαράσταση δεν είναι επαρκής για την ταξινόμηση ενός αντικειμένου. Οι συνδυασμοί όλων των συνόλων χαρακτηριστικών δίνουν πολύ περισσότερη διακριτική ισχύ για τον χαρακτηρισμό ενός αντικειμένου. Όταν συνδυάζουμε όλα τα σύνολα χαρακτηριστικών απευθείας σε μια διανυσματική μορφή (χωρίς να ακολουθεί κανένας συστηματικός κανόνας αλληλουχίας), η συσχέτιση δεν έχει νόημα, καθώς κάθε σύνολο χαρακτηριστικών έχει συγκεκριμένη στατιστική ιδιότητα. Η μάθηση πολλαπλών αναπαραστάσεων ασχολείται με το πρόβλημα της μηχανικής μάθησης, το οποίο παρέχει συστηματικό συνδυασμό χαρακτηριστικών από διαφορετικές αναπαραστάσεις (πολλαπλά σύνολα διακριτών χαρακτηριστικών) με τέτοιο τρόπο ώστε το διάνυσμα χαρακτηριστικών που δημιουργείται, να έχει φυσική σημασία. Στην εκμάθηση πολλαπλών αναπαραστάσεων, τα ετερογενή χαρακτηριστικά από διαφορετικές αναπαραστάσεις ενσωματώνονται για να εκμεταλλευτούν τη συμπληρωματική φύση των διαφορετικών αναπαραστάσεων στο σύνολο εκπαίδευσης. Η συνένωση χαρακτηριστικών από δύο ή περισσότερες αναπαραστάσεις μπορεί να υλοποιηθεί χρησιμοποιώντας διάφορες μεθόδους (Bhagat and Choudhary, 2018).

Οι Xia et al. (2010) διερευνώντας τη συμπληρωματική ιδιότητα των διαφορετικών αναπαραστάσεων χρησιμοποιούν ενσωμάτωση χαμηλής διάστασης, όπου η ενσωμάτωση χαρακτηριστικών από διαφορετικές αναπαραστάσεις έχει σχεδόν ομαλή κατανομή σε όλες τις αναπαραστάσεις. Η πιθανοτική μέθοδος multi-view μάθησης (m-SNE) που ανέπτυξαν, χρησιμοποιεί την απόσταση των διαφορετικών αναπαραστάσεων ανά ζεύγη για να υπολογίσει την κατανομή πιθανότητας με την εκμάθηση του βέλτιστου συντελεστή συνδυασμού σε όλες τις αναπαραστάσεις. Στη μελέτη των Yu et al. (2014), αυτή η απόσταση ζεύγους αντικαθίσταται με απόσταση υψηλής τάξης χρησιμοποιώντας υπεργράφο όπου κάθε δείγμα δεδομένων αντιπροσωπεύεται με μια κορυφή σε υπεργράφο και ένα κεντροειδές και ο αριθμός των k πλησιέστερων γειτόνων του χρησιμοποιείται για να συνδεθεί μια κορυφή με άλλες κορυφές χρησιμοποιώντας υπερακμές. Η προτεινόμενη μέθοδος πιθανοτικού πλαισίου (HD-MSL) των Yu et al. (2014) σημείωσε εξαιρετική ακρίβεια ταξινόμησης. Μια λεπτομερής ανασκόπηση των διαφορετικών προσεγγίσεων μάθησης πολλαπλών αναπαραστάσεων περιλαμβάνεται στη μελέτη των Xu et al. (2013) στην οποία αναλύονται λεπτομερώς διάφορα ζητήματα της μάθησης πολλαπλών αναπαραστάσεων.

4.1.5. Μάθηση μετρικής απόστασης

Κατά την επισημείωση εικόνας με βάση οπτικά χαρακτηριστικά, ανακύπτει το πρόβλημα του σημασιολογικού κενού. Μια άλλη προσέγγιση για την επισημείωση είναι η επισημείωση που βασίζεται στην ανάκτηση (retrieval based annotation). Στην επισημείωση που βασίζεται στην ανάκτηση, γίνεται ανάκτηση ενός συνόλου παρόμοιων εικόνων και μια άγνωστη εικόνα επισημαίνεται με βάση τις ετικέτες των ανακτημένων εικόνων. Για να εντοπιστούν παρόμοιες εικόνες σε ένα σύνολο εκπαίδευσης, πρέπει να μετρηθεί η απόσταση μεταξύ των εικόνων. Η εκμάθηση αυτών των αποστάσεων μεταξύ των εικόνων ονομάζεται μάθηση μετρικής απόστασης - distance metric learning (DML). Η ακρίβεια και η αποτελεσματικότητα της μετρικής απόστασης παίζουν σημαντικό ρόλο στη επιλογή των γειτόνων ή στην επιλογή παρόμοιων εικόνων (Bhagat and Choudhary, 2018).

Η απόσταση μεταξύ των εικόνων αποκτάται με βάση την οπτική ομοιότητα και την ομοιότητα μεταξύ ετικετών. Μια γραμμική μετρική απόστασης συχνά δεν είναι σε θέση να καταγράψει με ακρίβεια την πολυπλοκότητα του προβλήματος (πολυδιάστατα δεδομένα, μη γραμμικά όρια κλάσεων). Αντίθετα, μια μη γραμμική μετρική μπορεί να αντιπροσωπεύει σύνθετα

πολυδιάστατα δεδομένα και μπορεί να συλλάβει μη γραμμικές σχέσεις μεταξύ των στιγμιότυπων δεδομένων με των συναφών πληροφοριών.

Θα επιχειρήσουμε να παρουσιάσουμε τη βασική ιδέα της DML με βάση την απόσταση Mahalanobis. Ας θεωρήσουμε ένα σύνολο δεδομένων που περιέχει n εικόνες και κάθε εικόνα αντιπροσωπεύεται σε ένα χώρο d-διαστάσεων, δηλαδή, $x_i \in R^d$. Η απόσταση Mahalanobis μεταξύ της εικόνας x_i και της εικόνας x_j ορίζεται ως:

$$d_M(x_i, x_j) = \|x_i - x_j\|_M^2 = (x_i - x_j)^T M(x_i - x_j) \quad (4.1)$$

όπου, M είναι ένας προκαθορισμένος πίνακας που ικανοποιεί την ιδιότητα μιας έγκυρης μετρικής και ο στόχος της DML είναι να μάθει μια βέλτιστη μέτρηση Mahalanobis M από τα δεδομένα εκπαίδευσης.

Οι παράπλευρες πληροφορίες που προέρχονται από τις ετικέτες και το πλούσιο περιεχόμενο των εικόνων, που αναφέρονται ως αβέβαιες παράπλευρες πληροφορίες (uncertain side information), οδηγούν σε νέα πρόκληση για τη DML. Συμβατικά, οι μέθοδοι DML απαιτούν η εργασία της μάθησης για ρητές παράπλευρες πληροφορίες να δίνεται με τη μορφή είτε ετικετών κλάσης για ταξινόμηση εικόνας, είτε ζευγών περιορισμών για ομαδοποίηση και ανάκτηση. Εδώ οι περιορισμοί ανά ζεύγος (pairwise constraints) χρησιμοποιούνται για να μετρήσουμε εάν δύο εικόνες είναι παρόμοιες (“must-link”) ή ανόμοιες (“cannot-similar”). Επειδή οι περισσότερες εικόνες στην AIA επισημειώνονται με έναν αριθμό ετικετών, είναι δύσκολο να προσδιοριστεί αν δύο εικόνες σχηματίζουν τους περιορισμούς “must-link” (πρέπει να συνδέονται).

Η ενσωμάτωση της DML στην AIA μπορεί να συμβάλει στην αναζήτηση των πλησιέστερων γειτόνων και έτσι να βελτιώσει την ακρίβεια επισημείωσης. Χαρακτηριστικά παραδείγματα μοντέλων που ενσωματώνουν την DML είναι το μοντέλο UDML και το μοντέλο πιθανοτικής μέτρησης απόστασης (Probabilistic DML).

Το μοντέλο UDML (Wu et al., 2011) χρησιμοποιεί τόσο ετικέτες κειμένου όσο και οπτικό περιεχόμενο για μάθηση μετρικής και συνδυάζει επαγωγική (inductive) και μεταγωγική (transductive) μάθηση σε ένα συστηματικό πλαίσιο. Σε αντίθεση με τη κλασική DML, το μοντέλο UDML στοχεύει να μάθει αποτελεσματικές μετρικές από τις έμμεσες παράπλευρες πληροφορίες. Η εξαγωγή των παράπλευρων πληροφοριών μπορεί να επιτευχθεί σε μια

μορφή «τριάδας», δηλ., (x, x^+, x^-), στην οποία η εικόνα x και η εικόνα x^+ είναι παρόμοιες, ενώ η εικόνα x και η εικόνα x^- είναι ανόμοιες. Η επαγωγική (inductive) διατύπωση μάθησης για τη βελτιστοποίηση της μέτρησης (μετρικής) απόστασης από παράπλευρες πληροφορίες παρουσιάζεται στις εξισώσεις 4.2 και 4.3:

$$\min_{M>0} J_1(M) \triangleq \frac{1}{N_p} \sum_{i=1}^{N_Q} \sum_{\forall \left(x_{qi}, x_{k_i}^+, x_{k_i}^-\right) \in p_i} l\left(M; \left(x_{qi}, x_{k_i}^+, x_{k_i}^-\right)\right) \quad (4.2)$$

$$l\left(M; \left(x_{qi}, x_{k_i}^+, x_{k_i}^-\right)\right) = max\left\{0, 1 - \left[d_M\left(x_{qi}, x_{k_i}^-\right) - d_M\left(x_{qi}, x_{k_i}^+\right)\right]\right\} \quad (4.3)$$

όπου το N_p υποδηλώνει τον συνολικό αριθμό των τριάδων και η συνάρτηση απώλειας (loss function) βελτιστοποιεί τη μετρική τιμωρώντας τη μεγάλη απόσταση μεταξύ δύο παρόμοιων εικόνων και τη μικρή απόσταση μεταξύ δύο ανόμοιων εικόνων.

Επιπλέον, οι Wu et al. (2011) ανέπτυξαν επίσης μια μεταγωγική προσέγγιση για την ενσωμάτωση των ετικετών κειμένου και του οπτικού περιεχομένου των εικόνων των μέσων κοινωνικής δικτύωσης ως εξής:

$$\min_{M>0} J_2(M) \triangleq \sum_{i,j} w_{ij} \left\| x_i - x_j \right\|_M^2 \quad (4.4)$$

όπου ο παράγοντας w_{ij} αντιπροσωπεύει την ομοιότητα συνημίτονου μεταξύ των δύο διανυσμάτων ετικετών των δύο εικόνων.

Η εξίσωση υποδεικνύει ότι εάν δύο εικόνες μοιράζονται παρόμοιες ετικέτες κειμένου, η «οπτική» τους απόσταση αναμένεται να είναι μικρή. Τέλος, η διαμόρφωση μιας ενιαίας μάθησης μετρικής απόστασης επιτυγχάνεται με τη σύνθεση του επαγωγικού και του μεταγωγικού τύπου:

$$\min_{M>0} J(M) \triangleq \frac{1}{2} tr(M^T M) + cJ_1(M) + \lambda J_2(M) \quad (4.5)$$

Το μοντέλο PDML (Wu et al., 2009) μπορεί να εξάγει πιθανοτικές παράπλευρες πληροφορίες από τα δεδομένα χρησιμοποιώντας ένα μοντέλο γράφων και στη συνέχεια ένας πιθανοτικός αλγόριθμος, ο RCA (Relevance Component Analysis) χρησιμοποιείται για την εύρεση μιας βέλτιστης μετρικής από τις πιθανές παράπλευρες πληροφορίες.

4.2. Επισημείωση εικόνας με βάση το πλήθος των ετικετών

Κατά την AIA, εκχωρούνται σε μία εικόνα ετικέτες με βάση τα περιεχόμενά της. Ο αριθμός ετικετών που σχετίζονται με τις εικόνες, μπορεί να είναι σταθερός ή μεταβλητός. Ο σταθερός αριθμός ετικετών αναφέρεται στην εκχώρηση ενός αριθμού n ετικετών στις εικόνες. Μόλις οριστεί ο αριθμός n, παραμένει σταθερός για όλες τις εικόνες. Στην προσέγγιση αυτή, ο αριθμός των αντικειμένων που υπάρχουν στην εικόνα, δεν έχει πολύ μεγάλη σημασία. Ωστόσο, τα περιεχόμενα της εικόνας μπορεί να χρησιμοποιηθούν για την επισημείωση. Αυτός ο τύπος επισημείωσης μπορεί να οδηγήσει σε υψηλότερο ψευδώς θετικό (False Positive) αριθμό στα αποτελέσματα, ειδικά όταν ο αριθμός αντικειμένων που υπάρχουν στην εικόνα είναι μικρότερος από το n (Bhagat and Choudhary, 2018).

Στην πραγματικότητα, μια εικόνα μπορεί να περιέχει πολλά αντικείμενα, οπότε κατά την επισημείωση της εικόνας, οι ετικέτες πρέπει να αντιστοιχιστούν στα σχετικά αντικείμενα. Η επισημείωση όλων των σχετικών αντικειμένων μπορεί να έχει ως αποτέλεσμα έναν μεταβλητό αριθμό ετικετών για τις διαφορετικές εικόνες. Ενώ είναι σχετικά εύκολο να εφαρμοστεί επισημείωση εικόνας για έναν σταθερό αριθμό ετικετών, οι μεταβλητού πλήθους ετικέτες αντιπροσωπεύουν τα πραγματικά περιεχόμενα μιας εικόνας. Και οι δύο προσεγγίσεις, σταθερού πλήθους ετικέτες και ετικέτες μεταβλητού πλήθους αποτελούν μέρος της επισημείωσης πολλαπλών ετικετών εφόσον είναι $n > 1$.

4.2.1. Σταθερού πλήθους ετικέτες

Η συμβατική προσέγγιση επισημείωσης ακολουθεί τον ορισμό σταθερού πλήθους ετικετών, όπου ο αριθμός των ετικετών έχει οριστεί για όλες τις εικόνες δοκιμής ανεξαρτήτως του περιεχομένου τους. Στην προσέγγιση επισημείωσης καθορισμένου πλήθους ετικετών, λαμβάνεται ένα σύνολο υποψήφιων ετικετών και ο τελικός κατάλογος των ετικετών ορίζεται από τις υποψήφιες ετικέτες. Όταν ληφθούν οι υποψήφιες ετικέτες, η επιλογή της τελικής λίστας ετικετών επισημείωσης είναι ουσιαστικά ένα πρόβλημα μεταφοράς ετικετών (label transfer problem). Τα συστήματα μεταφοράς ετικετών ασχολούνται με τη διαδικασία λήψης αποφάσεων, όπου μια λίστα με λέξεις-κλειδιά είναι επιλεγμένες ως τελικές ετικέτες για οποιαδήποτε εικόνα ερωτήματος. Μια τιμή πιθανότητας αποδίδεται σε κάθε υποψήφια ετικέτα και όλες οι υποψήφιες ετικέτες κατατάσσονται σύμφωνα με την πιθανότητά τους.

Αυτή η πιθανότητα υποδεικνύει τη βαθμολογία εμπιστοσύνης για την υποψήφια ετικέτα επισημείωσης.

Οι προσεγγίσεις επισημείωσης με βάση την ταξινόμηση (κατάταξη) περιστρέφονται γύρω από τις ακόλουθες μεθόδους: (i) Οι υποψήφιες επισημειώσεις κατατάσσονται σύμφωνα με τις πιθανότητές τους και επιλέγονται οι κορυφαίες n εικόνες με την υψηλότερη πιθανότητα ως τελική επισημείωση, (ii) Μπορεί να ακολουθηθεί μια προσέγγιση βασισμένη σε τιμή κατωφλίου, όπου όλες οι υποψήφιες ετικέτες των οποίων η πιθανότητα είναι μεγαλύτερη από το όριο (κατώφλι), επιλέγονται ως τελικές ετικέτες.

Όταν οι κορυφαίες n ετικέτες επιλέγονται ως τελική επισημείωση, η προσέγγιση επισημείωσης είναι εύκολο να αξιολογηθεί. Η τιμή του n είναι μεταβλητή, αλλά μόλις αυτή οριστεί, επιλέγει τον ίδιο αριθμό ετικετών για όλες τις εικόνες. Η επιλογή της κατάλληλης τιμής του n είναι σημείο προβληματισμού. Αν το n είναι πολύ μικρό, η δυνατότητα ανάκλησης (recall) της επισημείωσης θα υποβαθμιστεί. Εάν το n είναι πολύ μεγάλο, η ακρίβεια της μεθόδου επισημείωσης θα μειωθεί. Το μέγεθος του n συνήθως ορίζεται στο 5, ή μερικές φορές μπορεί να είναι μεταβλητό.

Η δημοτικότητα των προσεγγίσεων επισημείωσης με σταθερού πλήθους ετικέτες είναι αποτέλεσμα εύκολων και απλών κριτηρίων αξιολόγησης. Επίσης, οι διάφορες state-of-art μέθοδοι για επισημείωση με σταθερό πλήθος και η έλλειψη βασικών και σύγχρονων μεθόδων επισημείωσης αυθαίρετου πλήθους ετικετών είναι ο λόγος πίσω από την κλίση προς την προσέγγιση επισημείωσης σταθερού πλήθους ετικετών. Οι state-of-art μέθοδοι βοηθούν στη συγκριτική αξιολόγηση των προτεινόμενων μεθόδων (Bhagati and Choudhary, 2018).

4.2.2. Μεταβλητού πλήθους ετικέτες

Όλα τα σχετικά αντικείμενα που υπάρχουν στην εικόνα θα πρέπει να φέρουν μια ετικέτα. Έτσι, το πλήθος των ετικετών καθορίζεται από το περιεχόμενο της εικόνας. Η επισημείωση μιας εικόνας με όλες τις σχετικές ετικέτες είναι ο ρεαλιστικός τρόπος επισημείωσης της εικόνας. Η πρόβλεψη του σωστού πλήθους των ετικετών είναι πρωταρχική εργασία οποιασδήποτε προσέγγισης επισημείωσης μεταβλητού πλήθους ετικετών, έτσι ώστε να μπορούν να εκχωρηθούν οι κατάλληλες ετικέτες από μια λίστα ετικετών. Οι μεταβλητού

πλήθους ετικέτες μπορούν να προβλεφθούν χρησιμοποιώντας μια ελαφρώς τροποποιημένη προσέγγιση μεταφοράς ετικετών. Όταν έχει οριστεί μια τιμή κατωφλίου, όλες οι ετικέτες που έχουν βαθμό εμπιστοσύνης μεγαλύτερο από το κατώφλι επιλέγονται ως τελικές ετικέτες. Ο αριθμός των ετικετών που έχουν βαθμολογία εμπιστοσύνης μεγαλύτερο από το όριο, μπορεί να διαφέρει από εικόνα σε εικόνα, με αποτέλεσμα μεταβλητού πλήθους ετικέτες. Όμως, η εύρεση του βέλτιστου κατωφλίου είναι σημείο προβληματισμού (Bhagat and Choudhary, 2018).

Πρόσφατα, τεχνικές βαθιάς μάθησης χρησιμοποιούνται για την πρόβλεψη ετικετών μεταβλητού πλήθους. Η βασική ιδέα πίσω από τη μέθοδο επισημείωσης ετικετών αυθαίρετου πλήθους με βαθιά μάθηση είναι η χρήση ενός βαθιού νευρωνικού δικτύου που μπορεί να δημιουργήσει αυτόματα μια λίστα ετικετών.

4.3. Επισημείωση εικόνας με βάση το σύνολο δεδομένων εκπαίδευσης

Για να σχεδιάσουμε ένα μοντέλο επισημείωσης, το πρώτο πράγμα που πρέπει να εξετάσουμε είναι το σύνολο των δεδομένων εκπαίδευσης. Ο τύπος του συνόλου δεδομένων εκπαίδευσης έχει σημαντική επίπτωση στο τελικό μοντέλο επισημείωσης. Το σύνολο δεδομένων μπορεί να κατηγοριοποιηθεί σε τρεις κατηγορίες (Bhagat and Choudhary, 2018):

(i) Το σύνολο εκπαίδευσης είναι πλήρως επισημειωμένο. Αυτό σημαίνει ότι όλες οι εικόνες στο σύνολο εκπαίδευσης είναι επισημειωμένες χειροκίνητα και περιέχουν ένα πλήρες σύνολο ετικετών. Η εκπαίδευση με αυτούς τους τύπους δεδομένων είναι γνωστή ως εποπτευόμενη μάθηση. Η απόκτηση ενός πλήρως επισημειωμένου συνόλου δεδομένων είναι δαπανηρή και χρονοβόρα, αλλά εάν είναι διαθέσιμο, μπορεί να σχεδιαστεί ένα πολύ αποτελεσματικό μοντέλο βασισμένο σε μάθηση με επίβλεψη.

(ii) Το σύνολο εκπαίδευσης έχει ελλιπείς ετικέτες. Αυτού του τύπου σύνολα δεδομένων είτε είναι επισημειωμένα με το χέρι είτε η επισημείωση εκτελείται από τους χρήστες μέσω ιστότοπων κοινωνικής δικτύωσης, αλλά οι εικόνες δεν είναι πλήρως επισημειωμένες. Καθώς οι εικόνες επισημειώνονται από τους χρήστες, οι εικόνες εκπαίδευσης ενδέχεται να περιέχουν θορυβώδεις ετικέτες ή να έχουν ένα ελλιπές σύνολο ετικετών. Η εκπαίδευση με αυτό το είδος δεδομένων είναι γνωστή ως ημι-εποπτευόμενη (Semi-Supervised Learning - SSL) ή ασθενώς εποπτευόμενη ή ενεργή μάθηση. Τα μοντέλα επισημείωσης που βασίζονται

σε SSL, έχουν μεγάλες δυνατότητες, καθώς μπορούν να προσαρμοστούν σε πολύ μεγάλης κλίμακας και επεκτάσιμο σύνολο δεδομένων.
(iii) Το σύνολο δεδομένων εκπαίδευσης δεν είναι επισημειωμένο. Οι εικόνες στο σύνολο δεδομένων εκπαίδευσης δεν φέρουν ετικέτες, αλλά κάθε εικόνα είναι εφοδιασμένη με μερικά μεταδεδομένα (metadata), όπως για παράδειγμα διεύθυνση URL ή η επικεφαλίδα (header) ενός αρχείου DICOM στην περίπτωση μιας ιατρικής εικόνας. Σε τέτοιες περιπτώσεις, οι υποψήφιες ετικέτες για επισημείωση πρέπει να εξάγονται από τα μεταδεδομένα που σχετίζονται με τις εικόνες. Η εκπαίδευση με αυτό το είδος δεδομένων είναι γνωστή ως μη εποπτευόμενη μάθηση. Είναι δύσκολο να σχεδιάσουμε ένα μοντέλο επισημείωσης βασισμένο στη μη εποπτευόμενη μάθηση με υψηλή αποτελεσματικότητα και ακρίβεια.

Η αντιπροσωπευτική ακρίβεια του συνόλου δεδομένων εκπαίδευσης έχει άμεση επίπτωση στην απόδοση οποιουδήποτε μοντέλου μάθησης. Εάν το σύνολο δεδομένων περιέχει οποιοδήποτε θόρυβο, τότε πρέπει να αφαιρεθεί πριν από την εκπαίδευση ενός μοντέλου. Η αποκατάσταση θορυβώδους ή ελλιπούς συνόλου δεδομένων οδηγεί σε μία προσέγγιση βασισμένη στην ενεργή μάθηση (SSL).

4.3.1. Μάθηση με επίβλεψη

Όταν το σύνολο δεδομένων εκπαίδευσης συνοδεύεται από αντίστοιχες ετικέτες εξόδου, ο σχεδιασμός του μοντέλου είναι κάπως ευθύς (straightforward). Το σύνολο εποπτευόμενων δεδομένων εκπαίδευσης δίνεται ως $\{(X_1, Y_1), (X_2, Y_2), \ldots\ldots, (X_m, Y_m)\}$ όπου το X_i είναι το σύνολο εισόδου και το Y_i είναι το αντίστοιχο σύνολο εξόδου. Στην περίπτωση πολλαπλών κλάσεων η i-οστη ετικέτα εξόδου $Y_i = (y_i^1, y_i^2, \ldots\ldots, y_i^P)$ όπου P είναι ο αριθμός των κλάσεων εξόδου για τα i-οστα δεδομένα εισόδου. Τα δεδομένα εισόδου μπορούν επίσης να έχουν πολλαπλά χαρακτηριστικά που αντιπροσωπεύονται ως $X_i = (x_i^1, x_i^2, \ldots\ldots, x_i^N)$ όπου N είναι ο αριθμός των χαρακτηριστικών για τα i-οστα δεδομένα εισόδου. Κατά την επισημείωση, μια εικόνα μπορεί να επισημειώνεται με πολλαπλές λέξεις-κλειδιά και ως εκ τούτου ένα σύνολο δεδομένων εκπαίδευσης δίνεται ως $(x_i^1, x_i^2, \ldots\ldots, x_i^N,\ y_i^1, y_i^2, \ldots\ldots, y_i^P)$ όπου x_i^j είναι ένα στοιχείο του συνόλου εισόδου X_i των χαρακτηριστικών εικόνας και το y_i^j είναι ένα στοιχείο του συνόλου εξόδου Y_i των λέξεων-κλειδιών για την i-οστη εικόνα εκπαίδευσης. Τα χαρακτηριστικά εισόδου μπορούν να είναι είτε χαρακτηριστικά που εξάγονται από τις εικόνες εισόδου, είτε συσχετισμένα χαρακτηριστικά που εξάγονται από την εικόνα και τις ετικέτες συλλογικά.

Το μέγεθος του συνόλου δεδομένων εκπαίδευσης μπορεί επίσης να επηρεάσει τη συνολική απόδοση του μοντέλου, καθώς περισσότερες εικόνες εκπαίδευσης σημαίνουν μικρότερο σφάλμα γενίκευσης. Το πλεονέκτημα της εποπτευόμενης μάθησης είναι ότι το σύνολο δεδομένων εκπαίδευσης επιτρέπει στο μοντέλο να μάθει τα χαρακτηριστικά των εννοιών και τον κανόνα ταξινόμησης. Έτσι, ένα μεγάλο σύνολο εκπαίδευσης θα ενισχύσει την ικανότητα του εκπαιδευμένου μοντέλου. Η βασική ιδέα πίσω από όλα τα μοντέλα επισημείωσης εικόνας που βασίζονται στην εποπτευόμενη μάθηση, είναι η αποτελεσματική εκμετάλλευση των επισημειωμένων δεδομένων στο σύνολο εκπαίδευσης.

Τα διακριτικά μοντέλα επισημείωσης επιβλεπόμενης μάθησης εκπαιδεύουν έναν ταξινομητή εκμεταλλευόμενα τα οπτικά και λεκτικά χαρακτηριστικά των επισημειωμένων εικόνων. Η ανισορροπία κλάσης αποτελεί επίσης ένα ζήτημα που αντιμετωπίζει το πλήρως επισημειωμένο σύνολο δεδομένων. Κάθε ετικέτα θα πρέπει να έχει έναν επαρκή και σχεδόν παρόμοιο αριθμό εικόνων του συνόλου εκπαίδευσης για να αποφευχθεί η υποπροσαρμογή του ορίου λήψης αποφάσεων. Έτσι, δημιουργεί το πρόβλημα της κλιμάκωσης (scalability problem).

Η εργασία επισημείωσης πολλαπλών ετικετών όπου ο αριθμός των ετικετών είναι μεταβλητός για διαφορετικές εικόνες περιπλέκει περαιτέρω το σχεδιασμό και την πολυπλοκότητα του εποπτευόμενου μοντέλου. Εάν υπάρχει ένας μεγάλος αριθμός ετικετών εξόδου, τότε ο χρόνος εκπαίδευσης του μοντέλου είναι πολύ υψηλός. Στην εποπτευόμενη μάθηση, μόλις εκπαιδευτεί το μοντέλο, το σύνολο δεδομένων εκπαίδευσης καθίσταται άχρηστο. Έτσι, ο χρόνος εκπαίδευσης του μοντέλου θεωρείται συνήθως ως η υπολογιστική πολυπλοκότητα του μοντέλου (Bhagat and Choudhary, 2018).

4.3.2. Ημι-εποπτευόμενη μάθηση

Το κύριο μειονέκτημα της εποπτευόμενης μάθησης είναι ότι απαιτεί μεγάλο αριθμό επισημειωμένων εικόνων εκπαίδευσης, που είναι πολύ δύσκολο να αποκτηθούν. Επίσης, για ένα σύνολο δεδομένων μεγάλης κλίμακας, ο χρόνος εκπαίδευσης των εποπτευόμενων μοντέλων είναι συνήθως πολύ υψηλός. Για να αντιμετωπιστούν αυτές οι επιπλοκές, προτείνεται μια νέα μέθοδος για την εκπαίδευση του μοντέλου που ονομάζεται SSL ή μάθηση με χαλαρή επίβλεψη ή ενεργός μάθηση. Η προσέγγιση SSL χρησιμοποιεί μόνο ένα μικρό

αριθμό επισημειωμένων δεδομένων και χρησιμοποιεί μη επισημειωμένα δεδομένα για την εκπαίδευση ενός μοντέλου. Η SSL μπορεί να αντιμετωπίσει το θορυβώδες, ελλιπές και μη ισορροπημένο σύνολο δεδομένων εκπαίδευσης. Το βασικό κίνητρο πίσω από ένα μοντέλο που βασίζεται σε SSL, είναι να μειώσει το μέγεθος του επισημειωμένου συνόλου εκπαίδευσης. Όταν το σύνολο δεδομένων εικόνων επισημειώνεται από τον χρήστη, όπως συμβαίνει στην περίπτωση των βάσεων δεδομένων εικόνων ImageNet και NUS-WIDE, οι επισημειωμένες εικόνες είναι συνήθως θορυβώδεις, καθώς οι ετικέτες που έχουν ανατεθεί, ενδέχεται να μην αντιπροσωπεύουν με ακρίβεια το περιεχόμενο των εικόνων και είναι είτε ελλιπείς είτε υπερβολικά πολλές (over-tagged). Η υπερ-επισημείωση (over-tagging) είναι ένα είδος θορύβου στις ετικέτες που πρέπει να αφαιρεθεί. Οι θορυβώδεις ετικέτες πρέπει να αντικαθίστανται από σχετικές ετικέτες που αντικατοπτρίζουν με ακρίβεια τις έννοιες της εικόνας. Διάφορες μέθοδοι για τη μείωση των ετικετών έχουν προταθεί στη βιβλιογραφία. Για την αντιμετώπιση ελλιπών ετικετών, έχουν προταθεί επίσης διάφορες τεχνικές, στις οποίες αναφερόμαστε στη συνέχεια.

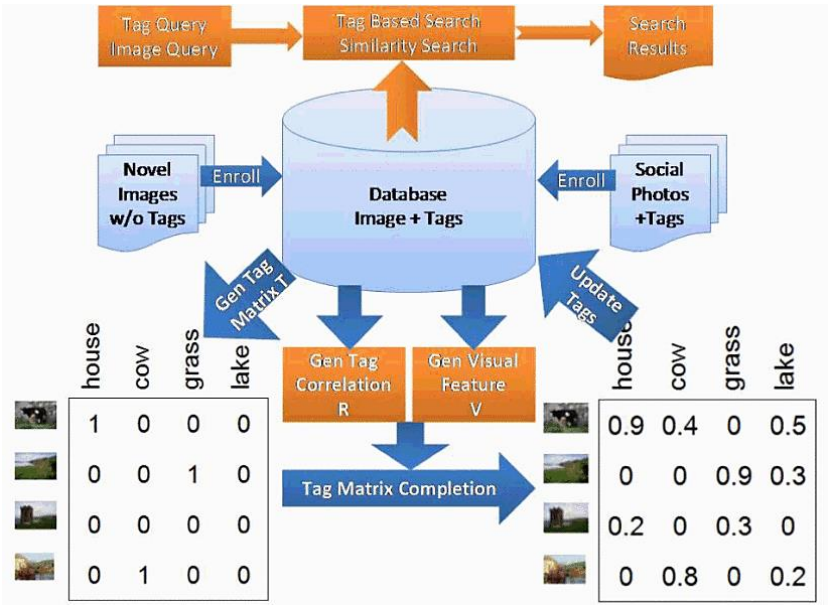
4.3.2.1. Μέθοδοι AIA με βάση την συμπλήρωση των ετικετών

Οι μέθοδοι AIA με βάση την συμπλήρωση ετικετών είναι αρκετά διαφορετικές από τις άλλες μεθόδους επισημείωσης εικόνων και έχουν συγκεντρώσει το ερευνητικό ενδιαφέρον τα τελευταία χρόνια. Οι μέθοδοι AIA συχνά υποθέτουν ότι οι εικόνες στο σύνολο δεδομένων εκπαίδευσης είναι πλήρως επισημειωμένες με κατάλληλες ετικέτες. Ωστόσο, πρόσφατες μελέτες έχουν δείξει ότι οι χειροκίνητες ετικέτες είναι συχνά αναξιόπιστες και ασυμβίβαστες. Η καινοτομία των μεθόδων AIA με βάση την συμπλήρωση των ετικετών είναι ότι οι ετικέτες που λείπουν, μπορούν να συμπληρωθούν αυτόματα χωρίς διαδικασία εκπαίδευσης και ότι οι θορυβώδεις ετικέτες για τις δεδομένες εικόνες μπορούν να διορθωθούν (Cheng et al., 2018). Παρόλο που έχουν αναπτυχθεί διάφορες μέθοδοι συμπλήρωσης ετικετών πάνω σε διαφορετικά πλαίσια, όλες εστιάζουν στη συνεκτικότητα (συνοχή) του περιεχομένου και τη σχέση των ετικετών. Ολόκληρο το σύνολο των δεδομένων εκπαίδευσης μπορεί να αναπαρασταθεί ως ένας αρχικός πίνακας ετικετών με κάθε σειρά να αντιστοιχεί σε μια εικόνα και κάθε στήλη σε μια ετικέτα. Η συμπλήρωση ετικετών λειτουργεί σε επίπεδο πίνακα, ανακτώντας τον αρχικό πίνακα, προσδιορίζοντας σωστές συσχετίσεις μεταξύ εικόνων και ετικετών.

Οι μέθοδοι AIA με βάση την συμπλήρωση των ετικετών μπορούν να χωριστούν περαιτέρω σε μεθόδους που βασίζονται στη συμπλήρωση πίνακα (matrix completion-based methods), μεθόδους βασισμένες σε γραμμική ανακατασκευή χώρου (linear space reconstruction-based methods), μεθόδους που βασίζονται σε ομαδοποίηση υποχώρων (subspace cluster-based methods) και μεθόδους βασισμένες σε χαμηλής τάξης παραγοντοποίηση πίνακα (low-rank matrix factorization-based methods).

4.3.2.1.1. Μέθοδοι βασισμένες στη συμπλήρωση πίνακα

Το μοντέλο TMC που έχουν προτείνει οι Wu et al. (2013) για την επισημείωση και ανάκτηση εικόνων μοντελοποιεί τη διαδικασία συμπλήρωσης των ετικετών σε ένα πρόβλημα συμπλήρωσης πίνακα. Η σχέση μεταξύ των ετικετών και των εικόνων περιγράφεται από ένα πίνακα ετικετών, όπου κάθε καταχώρηση στον πίνακα ετικετών αντιπροσωπεύει τη συνάφεια μιας ετικέτας με μια εικόνα.



Εικόνα 17. Το πλαίσιο για την συμπλήρωση του πίνακα ετικετών και την εφαρμογή του στην αναζήτηση εικόνων.

Δεδομένης μιας βάσης δεδομένων εικόνων επισημειωμένων με ορισμένες ετικέτες, ο προτεινόμενος αλγόριθμος δημιουργεί αρχικά έναν πίνακα ετικετών που υποδηλώνει τη σχέση μεταξύ των εικόνων και των αρχικά καθορισμένων ετικετών. Στη συνέχεια συμπληρώνει αυτόματα τον πίνακα ετικετών, ενημερώνοντας την βαθμολογία σχετικότητας των ετικετών σε όλες τις εικόνες. Ο συμπληρωμένος πίνακας ετικετών θα χρησιμοποιηθεί για αναζήτηση εικόνων με βάση ετικέτες ή αναζήτηση ομοιότητας εικόνας (Wu et al., 2013).

Έστω n και m ο αριθμός των εικόνων και οι διαθέσιμες ετικέτες, αντίστοιχα. Έστω ότι $\hat{T} \in \mathbb{R}^{n \times m}$ υποδηλώνει τον πίνακα ετικετών που προέρχεται από το μη αυτόματη επισημείωση,

όπου $\widehat{T_{ij}} = 1$ υποδεικνύει ότι η εικόνα i φέρει την ετικέτα j ενώ $\widehat{T_{ij}} = 0$ ότι η εικόνα i δεν φέρει ετικέτα j. Στη συνέχεια, τα οπτικά χαρακτηριστικά της εικόνας απεικονίζονται από τον πίνακα $\hat{V} \in \mathbb{R}^{n \times d}$, όπου κάθε εικόνα μπορεί να περιγραφεί με d είδη χαρακτηριστικών. Επιπλέον, η συσχέτιση μεταξύ των ετικετών $R \in \mathbb{R}^{m \times m}$ λαμβάνεται υπόψη σε αυτό το μοντέλο και το R_{ij} αντιπροσωπεύει τη συσχέτιση μεταξύ της ετικέτας i και της ετικέτας j. Τέλος, ο πίνακας $T \in \mathbb{R}^{n \times m}$ υποδηλώνει τον πλήρη πίνακα ετικετών που πρέπει να υπολογιστεί.

Έτσι, το μοντέλο TMC στοχεύει στη βελτιστοποίηση του πίνακα ετικετών ελαχιστοποιώντας τη διαφορά μεταξύ της ομοιότητας που βασίζεται σε ετικέτες και της ομοιότητας που βασίζεται στο οπτικό περιεχόμενο. Η ακόλουθη εξίσωση βελτιστοποίησης χρησιμοποιείται για τον υπολογισμό του πλήρους πίνακα ετικετών:

$$\min_{T \in \mathbb{R}^{n \times m}} \|T \cdot T^T - V \cdot V^T\|_F^2 + \lambda \|T^T \cdot T - R\|_F^2 + \eta \left\|T - \hat{T}\right\|_F^2 \quad (4.6)$$

όπου $\lambda > 0$ και $\eta > 0$ είναι παράμετροι που καθορίζονται από cross validations.

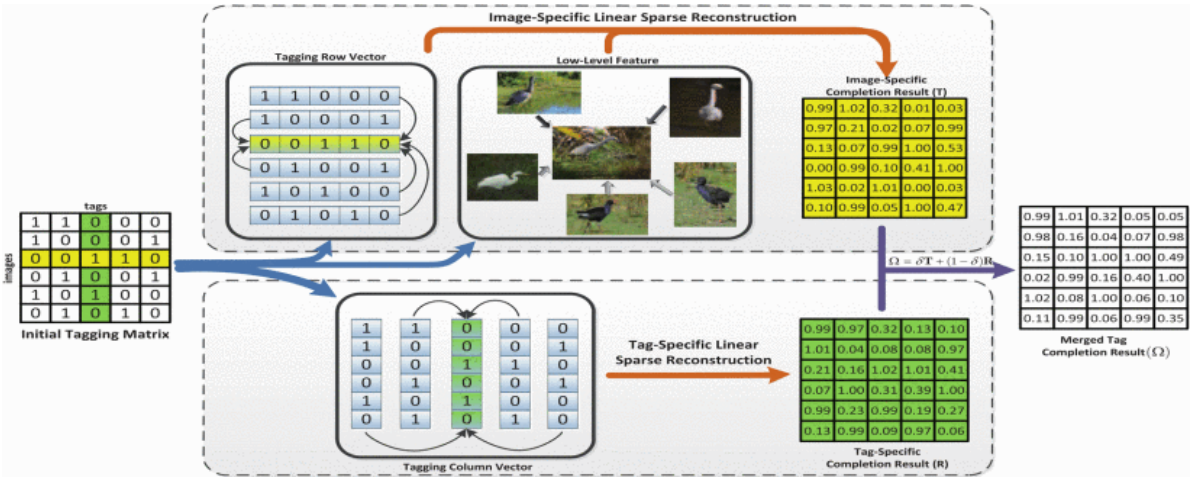
Η προτεινόμενη μέθοδος εμπίπτει στην κατηγορία της ημι-εποπτευόμενης μάθησης στο ότι εκμεταλλεύεται τόσο τις επισημειωμένες εικόνες όσο και τις εικόνες χωρίς ετικέτα για να βρει τον βέλτιστο πίνακα ετικετών.

4.3.2.1.2. Μέθοδοι που βασίζονται στην γραμμική ανακατασκευή χώρου

Οι Lin et al. (2013) παρουσίασαν ένα σχήμα για την συμπλήρωση των ετικετών μιας εικόνας μέσω γραμμικών αραιών ανακατασκευών (Linear Sparse Reconstructions - LSR) που σχετίζονται με την εικόνα και την ετικέτα. Το μοντέλο LSR διατυπώνει τις ανακατασκευές που σχετίζονται με την εικόνα και την ετικέτα ως κυρτό πρόβλημα βελτιστοποίησης κάτω από περιορισμένους για την αραιότητα. Η συγκεκριμένη ανακατασκευή της εικόνας χρησιμοποιεί τις οπτικές και σημασιολογικές ομοιότητες μεταξύ των εικόνων, ενώ η ειδική ανακατασκευή της ετικέτας ανιχνεύει τη συνάφεια μεταξύ των ετικετών. Τέλος, το LSR ομαλοποιεί και συγχωνεύει τα αποτελέσματα συμπλήρωσης των ετικετών που προκύπτουν από τις δύο γραμμικές ανακατασκευές υιοθετώντας ένα σταθμισμένο γραμμικό συνδυασμό στην εξίσωση:

$$\Omega = \delta T + (1 - \delta)R \quad (4.7)$$

όπου Ω είναι το αναμενόμενο τελικό αποτέλεσμα, τα T και R είναι τα κανονικοποιημένα αποτελέσματα συμπλήρωσης από ανακατασκευές ειδικά για την εικόνα και την ετικέτα αντιστοίχως και δ είναι μια παράμετρος βαρύτητας που κυμαίνεται από 0 έως 1.



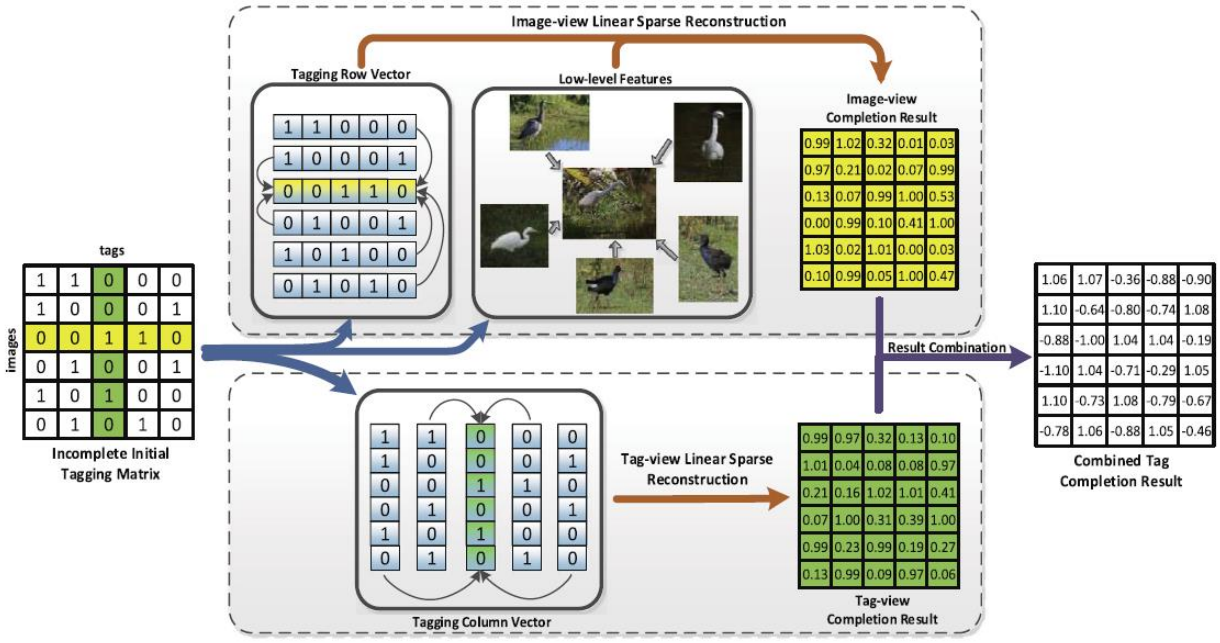
Εικόνα 18. Το πλαίσιο του LSR, απεικονίζεται με εικονικά δεδομένα.

Δεδομένου ότι δεν έχει συμπληρωθεί ο αρχικός πίνακας ετικετών, το LSR εκτελεί ξεχωριστά την συμπλήρωση ετικετών και από τις δύο όψεις, της εικόνας (άνω τετράγωνο με διακεκομμένη γραμμή) και της ετικέτας (κάτω τετράγωνο με διακεκομμένη γραμμή) και στη συνέχεια κανονικοποιεί και συγχωνεύει τα αντίστοιχα αποτελέσματα (Lin et al., 2013).

Το μοντέλο LSR βελτιώνει την απόδοση συμπλήρωσης ετικετών για κάθε ετικέτα και κάθε εικόνα, αλλά πιθανώς με το κόστος της υψηλής υπολογιστικής πολυπλοκότητας σε περίπτωση εικόνων ή ετικετών μεγάλης διάστασης.

Οι Lin et al. (2014) προχώρησαν στην περαιτέρω επέκταση και βελτίωση του LSR, με το μοντέλο γραμμικής αραιής ανακατασκευής διπλής όψης (Dual-view Linear Sparse Reconstruction – DLSR). Το DSLR πραγματοποιεί την συμπλήρωση των ετικετών μέσω της ανακατασκευής κάθε εικόνας και κάθε ετικέτας, αντίστοιχα. Ο στόχος του DSLR είναι να καταστήσει τις μεθόδους ανακατασκευής πιο αποτελεσματικές και πρακτικές, καθώς η μέθοδος LSR είναι υπολογιστικά δαπανηρή. Η μελέτη επικεντρώνεται κυρίως στην αξιοποίηση των ίδιων βαρών ανακατασκευής των χαρακτηριστικών και των αρχικών ετικετών αντί των διαφορετικών βαρών και στην διερεύνηση μιας καλύτερης στρατηγικής για το συνδυασμό των ανακατασκευασμένων διανυσμάτων επισημείωσης από τη σκοπιά της εικόνας και της ετικέτας. Για το συνδυασμό των ανακατασκευασμένων διανυσμάτων

επισημείωσης, η μελέτη αντιμετωπίζει το ανακατασκευασμένο διάνυσμα επισημείωσης από τη σκοπιά της εικόνας $t1$ και το ανακατασκευασμένο διάνυσμα επισημείωσης από τη σκοπιά της ετικέτας $t2$ ως αποτέλεσμα της ανάκτησης σχετικών ετικετών για μια δεδομένη προς συμπλήρωση εικόνα και τις αρχικά επισημειωμένες ετικέτες της από δύο διακριτές "μηχανές αναζήτησης".



Εικόνα 19. Το προτεινόμενο πλαίσιο του DLSR, με εικονικά δεδομένα.
Δοθείσης ενός ελλιπούς πίνακα ετικετών, το DLSR εκτελεί ξεχωριστά την συμπλήρωση ετικετών από την όψη της εικόνας (άνω τετράγωνο με διακεκομμένη γραμμή) και την όψη της ετικέτας (κάτω τετράγωνο με διακεκομμένη γραμμή) και στη συνέχεια συνδυάζει τα αντίστοιχα αποτελέσματα για καλύτερη πρόβλεψη των σχετικών ελλιπών ετικετών (Lin et al., 2014).

4.3.2.1.3. Μέθοδοι που βασίζονται σε ομαδοποίηση υποχώρων

Το μοντέλο ομαδοποίησης υποχώρων και συμπλήρωσης πίνακα (Subspace Clustering and Matrix Completion – SCMC) που προτείνεται από τους Hou and Lin (2015), εκτελεί διαδοχικά τη συμπλήρωση και τη βελτίωση των ετικετών. Αρχικά αντιμετωπίζει το πρόβλημα της συμπλήρωσης ετικετών σε ένα πλαίσιο ομαδοποίησης υποχώρων, υποθέτοντας ότι οι εικόνες έχουν δειγματοληφθεί από μια ένωση πολλαπλών γραμμικών υποχώρων και ότι οι αντίστοιχες ετικέτες τους σχηματίζουν ένα συμβατό υπο-πίνακα. Το μοντέλο στη συνέχεια βελτιώνει τον πίνακα ετικετών χρησιμοποιώντας ένα μοντέλο συμπλήρωσης πίνακα για να περιορίσει το σημασιολογικό χάσμα καθώς και την ακεραιότητα του πίνακα ετικετών.

Το μοντέλο SCMC χρησιμοποιεί τον αλγόριθμο LRR για να ομαδοποιήσει τα διανύσματα οπτικών χαρακτηριστικών σε διαφορετικούς υποχώρους. Ο αλγόριθμος LRR δίνει ως έξοδο

ένα μπλοκ-διαγώνιο πίνακα συνάφειας, στον οποίο κάθε υπο-πίνακας αντιστοιχεί σε ένα υποχώρο (ομάδα). Η εικόνα μπορεί να ομαδοποιηθεί σύμφωνα με τον πίνακα συνάφειας και η συμπλήρωση των ετικετών μπορεί να πραγματοποιηθεί με έναν αλγόριθμο μεταφοράς ετικετών για την συμπλήρωση ετικετών σε κάθε ομάδα χρησιμοποιώντας τη συχνότητα ετικετών, τη συνεμφάνιση ετικετών και την τοπική συχνότητα.

Η βελτίωση των ετικετών έχει ως στόχο να διορθώσει τις «θορυβώδεις» ετικέτες. Το πρόβλημα βελτίωσης των ετικετών μπορεί να αντιμετωπιστεί ως πρόβλημα συμπλήρωσης πίνακα, όπου η πρόκληση είναι να διαγράψει τις ετικέτες που δεν είναι αξιόπιστες στον πίνακα προτιμήσεων του στοιχείου-χρήστη και να «συμπληρώσει» τις ετικέτες που λείπουν δοθέντος ενός δείγματος παρατηρούμενων προτιμήσεων.

Οι μέθοδοι που βασίζονται σε ομαδοποίηση υποχώρων είναι ανώτερες από τις παραδοσιακές μεθόδους ομαδοποίησης επειδή αφενός δεν χρειάζεται να μετρήσουν την ομοιότητα μεταξύ χαρακτηριστικών και αφετέρου μπορούν να μοντελοποιήσουν με ακρίβεια τις κατανομές των χαρακτηριστικών της εικόνας (Cheng et al., 2018).

4.3.2.1.4. Μέθοδοι που βασίζονται σε χαμηλής τάξης παραγοντοποίηση πινάκων

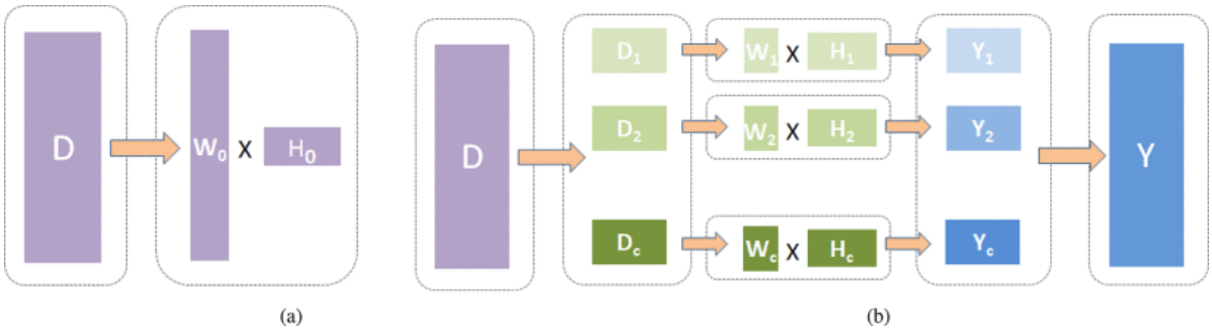
Ο αλγόριθμος συμπλήρωσης των ετικετών που προτείνεται από τους Li et al. (2014), έχει σχεδιαστεί με τα ακόλουθα χαρακτηριστικά: 1) Χαμηλή τάξη και αραιότητα σφάλματος (Low-rank and error sparsity): ο ατελής αρχικός πίνακας επισημείωσης D αποσυντίθεται στον πλήρη πίνακα επισημείωσης A και ένα αραιό πίνακα σφάλματος E, δηλαδή, $\mathrm{D} = \mathrm{U} \cdot \mathrm{V} + \mathrm{E}$. Αυτή η χαμηλής τάξης διατύπωση που εμπεριέχει αραιή κωδικοποίηση επιτρέπει στον αλγόριθμό των Li et al. (2014) να ανακτήσει λανθάνουσες δομές από θορυβώδη αρχικά δεδομένα αποφεύγοντας ωστόσο την υπερβολική μείωση του θορύβου, 2) Συνέπεια δομής τοπικής ανακατασκευής (Local reconstruction structure consistency). Για να κατευθύνουν την συμπλήρωση του D, οι τοπικές δομές γραμμικής ανακατασκευής στο χώρο των χαρακτηριστικών και στον χώρο των ετικετών αποκτώνται και διατηρούνται στον πίνακα U και V αντίστοιχα. Ένα τέτοιο σχήμα θα μπορούσε να μετριάσει το αρνητικό αποτέλεσμα των αποστάσεων που μετρήθηκαν από τα χαρακτηριστικά χαμηλού επιπέδου και τις ελλιπείς ετικέτες. Έτσι, ο αλγόριθμος επιδιώκει την αξιοποίηση όσο το δυνατόν περισσότερων πληροφοριών χωρίς ωστόσο τον κίνδυνο να οδηγηθεί σε υποβέλτιστη επίδοση.

Το τοπικά ευαίσθητο μοντέλο χαμηλής τάξης (LSLR) των Li et al. (2016) εκτελεί την συμπλήρωση ετικετών εικόνας εκτιμώντας ένα συνολικό μη γραμμικό μοντέλο με μια συλλογή από τοπικά γραμμικά μοντέλα. Σε σύγκριση με τις μεθόδους που βασίζονται στη γραμμική δομή, τα μη γραμμικά μοντέλα μπορούν να διερευνήσουν αποτελεσματικά την περίπλοκη συσχέτιση μεταξύ εικόνων και ετικετών. Λαμβάνοντας υπόψη τον ελλιπή αρχικό πίνακα ετικετών $D_{m\times n}$ (όπου m και n υποδηλώνει τον αριθμό των εικόνων και των ετικετών, αντίστοιχα) και τον πίνακα οπτικών χαρακτηριστικών, έστω $X_{n\times d}$ (όπου d η διάσταση του χώρου των οπτικών χαρακτηριστικών), στοχεύει στην ανάκτηση του πλήρους πίνακα ετικετών Y.

Η προεπεξεργασία στοχεύει στο να μάθει το μοντέλο την κατάλληλη αναπαράσταση για τη διαμέριση των δεδομένων. Όλες οι εικόνες στο σύνολο δεδομένων διαχωρίζονται σε πολλές ομάδες σύμφωνα με το σημασιολογικό περιεχόμενο. Στη συνέχεια, ένα τοπικό μοντέλο υπολογίζεται μέσω της παραγοντοποίησης του πλήρους πίνακα Y_i σε ένα βασικό πίνακα W_i και ενός αραιού πίνακα συντελεστών H_i, όπως παρουσιάζεται στην εξίσωση:

$$Y_i = W_i \cdot H_i, \ \forall\, i \in 1, 2, \ldots, c \qquad (4.8)$$
$$W_i \in R_{n_i\times k} \ και \ H_i \in R_{k\times m}$$

όπου n_i είναι ο αριθμός των δειγμάτων στην $i-οστή$ ομάδα. Ο τελικός πλήρης πίνακας ετικετών Y λαμβάνεται με την ενσωμάτωση όλων των υποπινάκων Y_i.



Εικόνα 20. Πλαίσιο του προτεινόμενου μοντέλου LSLR.
(α) το τμήμα προεπεξεργασίας, το οποίο μαθαίνει μια αναπαράσταση χαμηλού επιπέδου της εικόνας (W0) κατάλληλη για διαμέριση. (b) το τοπικά ευαίσθητο πλαίσιο, όπου ο αρχικός πίνακας ετικετών D χωρίζεται σε c ομάδες, στη συνέχεια, ένα τοπικό γραμμικό μοντέλο εκπαιδεύεται για κάθε ομάδα μέσω παραγοντοποίησης πίνακα. Ο τελικός ολοκληρωμένος πίνακας επιτυγχάνεται με την ενσωμάτωση των αποτελεσμάτων Yi (Li et al., 2016).

4.3.3. Μη εποπτευόμενη μάθηση

Οι μέθοδοι που βασίζονται στη μάθηση χωρίς επίβλεψη, είναι μία από τις πιο ελκυστικές μεθόδους επισημείωσης καθώς είναι απόλυτα κατάλληλες για το μεγάλο αριθμό εικόνων που είναι διαθέσιμες στις μέρες μας. Το ισχυρότερο πλεονέκτημα αυτών των μεθόδων είναι ότι δεν απαιτούν επισημειωμένες (πλήρως ή μερικώς) εικόνες εκπαίδευσης. Παρόλα αυτά οι μέθοδοι επισημείωσης που βασίζονται στη μάθηση χωρίς επίβλεψη, χρειάζονται δεδομένα κειμένου για την επισημείωση οποιασδήποτε εικόνας χωρίς ετικέτα. Οι υποψήφιες ετικέτες όμως στην περίπτωση αυτή, εξορύσσονται από τα μεταδεδομένα. Όπως γνωρίζουμε, κάθε εικόνα στον Παγκόσμιο Ιστό έχει μια διεύθυνση URL, κάποιο κείμενο που περιβάλλει την εικόνα και κάποιες άλλες πληροφορίες που σχετίζονται με αυτή. Αυτές οι πληροφορίες και το κείμενο που σχετίζονται με την εικόνα ονομάζονται μεταδεδομένα. Μια μη επιτηρούμενη μέθοδος επισημείωσης εξορύσσει τις ετικέτες από αυτά τα μεταδεδομένα και τις εκχωρεί στην εικόνα. Δεδομένου ότι οι υποψήφιες ετικέτες προέρχονται από τα μεταδεδομένα εικόνας, οι μη εποπτευόμενες μέθοδοι μάθησης μπορούν να επισημειώσουν την εικόνα χωρίς να εκπαιδεύσουν ένα μοντέλο. Τα μεταδεδομένα συνήθως παρέχουν μια σημαντική ένδειξη σχετικά με τις έννοιες της εικόνας. Παρόλο που όλο το κείμενο που υπάρχει στα μεταδεδομένα δεν έχει σημασία για την επισημείωση, τα μεταδεδομένα περιέχουν σχεδόν όλες τις υποψήφιες ετικέτες που μπορούν να περιγράψουν τέλεια τα περιεχόμενα και τις έννοιες της εικόνας. Η εξόρυξη των υποψήφιων ετικετών από τα μεταδεδομένα και η συσχέτιση μεταξύ των υποψήφιων ετικετών για την παραγωγή των τελικών ετικετών για την εικόνα είναι μια δύσκολη εργασία. Η εξόρυξη των ετικετών από τα μεταδεδομένα επιτρέπει επίσης την εμφάνιση ετικετών μεταβλητού μήκους. Τα μεταδεδομένα των εικόνων μπορεί να είναι θορυβώδη και αδόμητα, επομένως η μέθοδος ανίχνευσης της υποψήφιας ετικέτας πρέπει να είναι αρκετά ισχυρή (Bhagat and Choudhary, 2018).

4.4. Προσέγγιση επισημείωσης βάσει μοντέλου

Ένας τρόπος προσέγγισης του προβλήματος της επισημείωσης είναι η εκπαίδευση ενός μοντέλου για την επισημείωση των άγνωστων εικόνων. Ο πρωταρχικός στόχος αυτής της προσέγγισης (Model-based) είναι να εκπαιδεύσει ένα μοντέλο επισημείωσης από τα δεδομένα εκπαίδευσης μέσω του συσχετισμού οπτικής πληροφορίας και κειμένου έτσι ώστε το εκπαιδευμένο μοντέλο να μπορεί με ακρίβεια να επισημειώσει μία άγνωστη εικόνα. Οι πολυάριθμες μέθοδοι επισημείωσης μπορούν να ταξινομηθούν με βάση το παραγόμενο μοντέλο στις ακόλουθες κατηγορίες: μέθοδοι AIA βασισμένες στο παραγωγικό μοντέλο

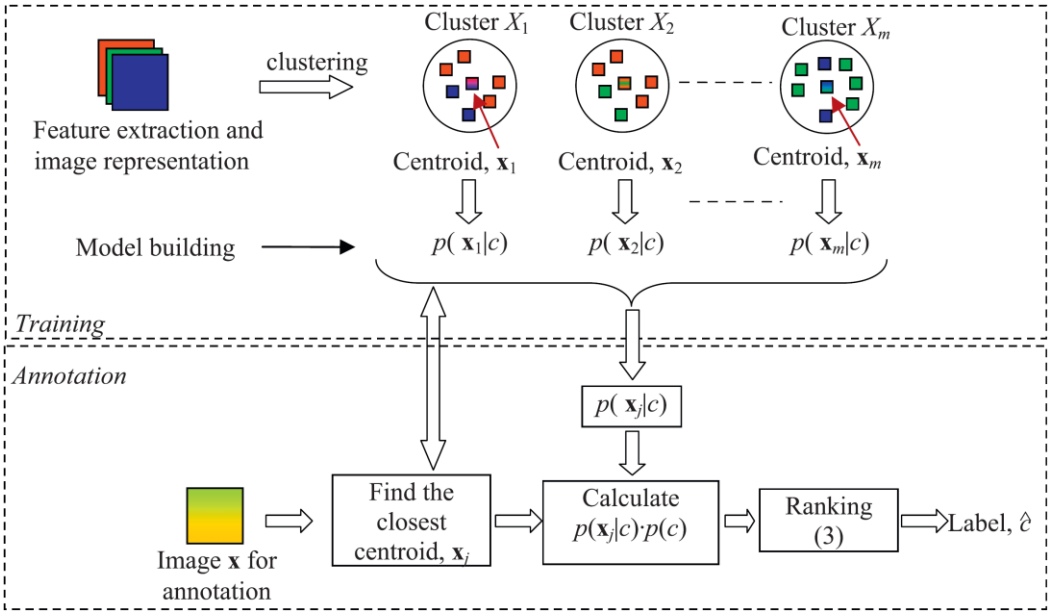
(generative model), οι οποίες είναι αφιερωμένες στη μοντελοποίηση της κοινής κατανομής των οπτικών χαρακτηριστικών και ετικετών εικόνας, μέθοδοι AIA με βάση το διακριτικό μοντέλο (discriminative model), οι οποίες θεωρούν την εργασία επισημείωσης ως πρόβλημα ταξινόμησης πολλαπλών ετικετών, μέθοδοι AIA που βασίζονται στο μοντέλο πλησιέστερων γειτόνων (nearest neighbor based model), οι οποίες υποθέτουν ότι εικόνες με παρόμοια οπτικά χαρακτηριστικά έχουν μεγάλη πιθανότητα να μοιράζονται παρόμοιες ετικέτες, και μέθοδοι AIA βασισμένες στη βαθιά μάθηση (deep learning based model), οι οποίες χρησιμοποιούν αλγορίθμους βαθιάς μάθησης για την εξαγωγή ισχυρών οπτικών χαρακτηριστικών ή εξαντλητικών παράπλευρων πληροφοριών για την AIA, ειδικά για AIA μεγάλης κλίμακας.

Οι προαναφερόμενες τέσσερις κατηγορίες μεθόδων AIA μπορούν να ταξινομηθούν περαιτέρω σε διάφορες υποκατηγορίες ανάλογα με τις υποκείμενες ιδέες τους.

4.4.1. Μέθοδοι AIA βασισμένες στο παραγωγικό μοντέλο

Οι μέθοδοι AIA που βασίζονται στο μοντέλο είναι αρκετά δημοφιλείς και σημείωσαν σημαντική πρόοδο κατά τη δεύτερη δεκαετία (2000 – 2010). Το παραγωγικό στοχεύει στην εκμάθηση μιας κοινής κατανομής πάνω σε οπτικά και λεκτικά χαρακτηριστικά, έτσι ώστε το εκπαιδευμένο μοντέλο να μπορεί να προβλέψει την υπό όρους πιθανότητα ετικετών που έχουν τα χαρακτηριστικά εικόνας. Τα παραγωγικά μοντέλα βασίζονται συνήθως σε μοντέλα συνάφειας (relevance models), μοντέλα μίγματος (mixture models) και σε μοντέλα θέματος (topic models).

Οι μέθοδοι AIA που βασίζονται στο παραγωγικό μοντέλο, παράγουν την πιθανότητα μιας ετικέτας εικόνας, υπολογίζοντας ένα κοινό πιθανοτικό μοντέλο χαρακτηριστικών εικόνας και λέξεων από ένα σύνολο δεδομένων εκπαίδευσης. Ένα τέτοιο πιθανοτικό (probabilistic) εργαλείο παρέχουν οι Bayesian μέθοδοι επισημείωσης. Οι Bayesian μέθοδοι λειτουργούν βρίσκοντας την εκ των υστέρων (posterior) πιθανότητα ότι μια εικόνα ανήκει σε οποιαδήποτε συγκεκριμένη έννοια, δεδομένης της παρατήρησης ορισμένων χαρακτηριστικών από την εικόνα ή την περιοχή. Αυτό επιτρέπει την αντιστοίχιση μιας εικόνας σε πολλαπλές έννοιες και την ταξινόμηση εικόνων με την ίδια έννοια σύμφωνα με τις πιθανότητες.



Εικόνα 21. Το γενικό Bayesian μοντέλο επισημείωσης (Zhang et al., 2012).

Δεδομένου ενός συνόλου εικόνων $\{I_1, I_2, \dots, I_N\}$ από ένα σύνολο δεδομένων σημασιολογικών κλάσεων $\{c_1, c_2, \dots, c_N\}$, το Bayesian μοντέλο υπολογίζει την εκ των υστέρων (posterior) πιθανότητα μια άγνωστη εικόνα I να ανήκει στην κλάση c_i από τις υπό συνθήκη πιθανότητες και τις προηγούμενες (prior). Ας υποθέσουμε ότι μια εικόνα I αντιπροσωπεύεται από το διάνυσμα χαρακτηριστικών x. Με δεδομένες τις πιθανότητες $p(c_i)$ και τις υπό συνθήκη πυκνότητες πιθανότητας $p(x|c_i)$, η ζητούμενη πιθανότητα καθορίζεται από τη σχέση:

$$p(c_i|x) = \frac{p(x|c_i) \cdot p(c_i)}{p(x)} \qquad (4.9)$$

Από την εξίσωση (4.9), μπορεί να γίνει αντιληπτό ότι ένα Bayesian πλαίσιο συνίσταται σε τέσσερα μέρη: ένα στοιχείο εξόδου $p(c_i|x)$ και τρία στοιχεία εισόδου: $p(x|c_i), p(c_i)$, and $p(x)$. Επειδή η κατανομή $p(x)$ είναι συνήθως ομοιόμορφη για όλες τις κλάσεις, η κλάση της εικόνας I μπορεί να αποφασιστεί χρησιμοποιώντας το κριτήριο μεγιστοποίησης της εκ των υστέρων πιθανότητας (maximising a posterior criterion - MAP):

$$\hat{c} = arg \max_{c_i}\{p(c_i|x)\} \approx arg \max_{c_i}\{p(x|c_i) \cdot p(c_i)\} \qquad (4.10)$$

Το κρίσιμο μέρος της Bayesian επισημείωσης είναι να μοντελοποιήσουμε τις υπό συνθήκη πιθανότητες, επειδή οι προηγούμενες πιθανότητες $p(c_i)$ μπορούν να βρεθούν από τη

συχνότητα των δειγμάτων που ανήκουν στην έννοια c_i. Τα διάφορα Bayesian μοντέλα ποικίλλουν από τον τρόπο με τον οποίο μοντελοποιούν τις υπό συνθήκη πιθανότητες $p(x|c_i)$.

Υπάρχουν γενικά δύο τύποι προσεγγίσεων για τη μοντελοποίηση των υπό συνθήκη πιθανοτήτων, η μη-παραμετρική προσέγγιση που ακολουθείται στα μοντέλα συνάφειας και η παραμετρική προσέγγιση που ακολουθείται στα μοντέλα μίγματος (Zhang et al., 2012).

4.4.1.1. Το μοντέλο συνάφειας

Οι μέθοδοι AIA που βασίζονται στο μοντέλο συνάφειας υλοποιούνται γενικά σε τρία βήματα: (i) Προσδιορίζουν τις κοινές κατανομές των οπτικών χαρακτηριστικών της εικόνας και των ετικετών, (ii) Υπολογίζουν την εκ των υστέρων (posterior) πιθανότητα κάθε ετικέτας για τις μη επισημειωμένες εικόνες, (iii) Για την επισημείωση μιας νέας εικόνας επιλέγουν την ετικέτα με την υψηλότερη πιθανότητα.

Τα μοντέλα συνάφειας ακολουθούν τη μη-παραμετρική προσέγγιση σύμφωνα με την οποία, οι υπό όρους πιθανότητες υπολογίζονται χωρίς προηγούμενη παραδοχή για την κατανομή των χαρακτηριστικών της εικόνας. Αντίθετα, η πραγματική κατανομή χαρακτηριστικών αποκτάται από τα χαρακτηριστικά των δειγμάτων κατάρτισης χρησιμοποιώντας ορισμένα στατιστικά στοιχεία. Στην πράξη, τα χαρακτηριστικά εικόνας πρώτα ποσοτικοποιούνται σε ομάδες χρησιμοποιώντας έναν συγκεκριμένο αλγόριθμο ομαδοποίησης. Στη συνέχεια, τα συνεχή χαρακτηριστικά αντικαθίστανται από τα κεντροειδή της ομάδας. Αυτή η διαδικασία διακριτοποιεί το χώρο των χαρακτηριστικών εικόνας. Οι υπό συνθήκη πιθανότητες για κάθε κλάση υπολογίζονται με την εύρεση της συχνότητας των δειγμάτων που ανήκουν στην κλάση αυτή. Για παράδειγμα, αν το πλησιέστερο κεντροειδές του διανύσματος χαρακτηριστικών x είναι το x_j, η πιθανότητα $p(x|c)$ στην εξίσωση (4.9) μπορεί να υπολογιστεί ως:

$$p(x|c) \approx p(x_j|c) = \frac{\text{No. of samples in } x_j \text{ which are from concept, c}}{\text{Total no. of samples from concept, c}} \qquad (4.11)$$

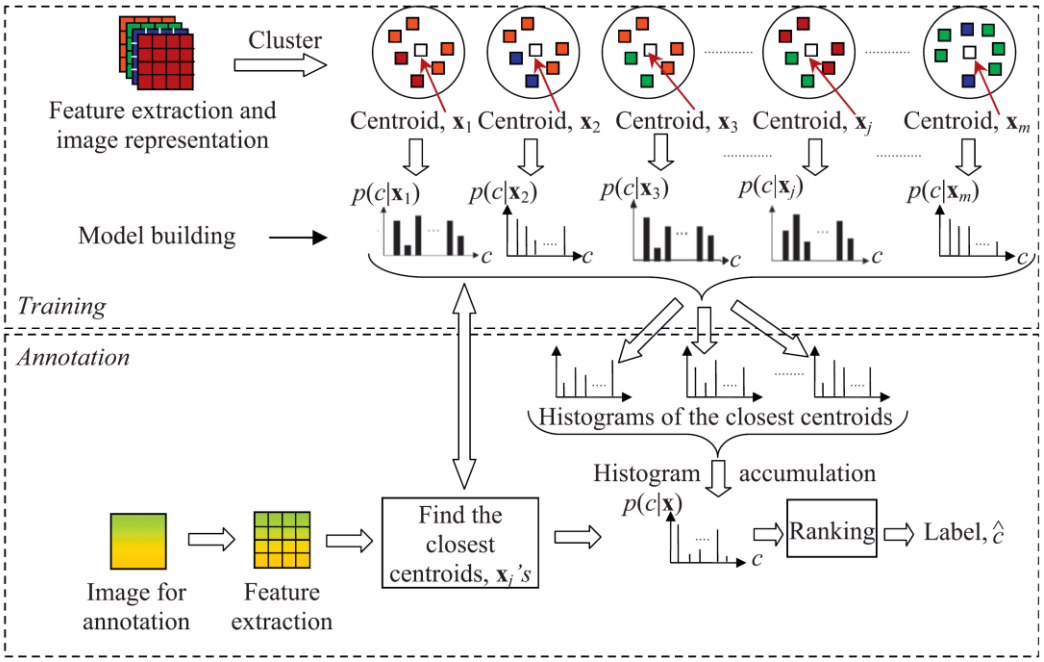
Η πλήρης διαδικασία επισημείωσης με βάση αυτή την προσέγγιση παρουσιάζεται στο σχήμα 21. Λαμβάνοντας μια νέα εικόνα, τα χαρακτηριστικά της εξάγονται και συγκρίνονται με τα κέντρα των ομάδων. Επιλέγονται τα κοντινότερα κέντρα ομάδας. Τα μοντέλα των υπό συνθήκη πιθανοτήτων που αντιστοιχούν στα επιλεγμένα κέντρα ομάδας, χρησιμοποιούνται

στη συνέχεια για τον υπολογισμό των εκ των υστέρων πιθανοτήτων. Το κριτήριο MAP (εξίσωση 4.10) χρησιμοποιείται στη συνέχεια για να επισημειώσει το μοντέλο τη νέα εικόνα.

Στο έργο των Mori et al. (1999), οι εικόνες εκπαίδευσης χωρίζονται σε μπλοκ και τα μπλοκ ομαδοποιούνται. Κάθε μπλοκ κληρονομεί όλες τις επισημειώσεις των γονικών εικόνων και κάθε ομάδα είναι μια συλλογή από έννοιες από όλα τα μπλοκ σε αυτήν. Η εκ των υστέρων πιθανότητα $p(c|x_j)$ μοντελοποιείται ως συν-εμφάνιση της λέξης c μέσα στην ομάδα X_j,

$$p(c|x_j) = \frac{\text{Total no. of annotation c inherited into cluster } X_j}{\text{Total no. of all annotation words in } X_j} \quad (4.12)$$

Η εκ των υστέρων πιθανότητα κάθε έννοιας c_i υπολογίζεται για την ομάδα X_j. Ως αποτέλεσμα, δημιουργείται ένα ιστόγραμμα εννοιών (concept histogram) για κάθε κεντροειδές ομάδας x_j. Κατά τη διάρκεια της επισημείωσης, μια άγνωστη εικόνα χωρίζεται σε μπλοκ. Για κάθε μπλοκ στην άγνωστη εικόνα, επιλέγονται τα κεντροειδή των ομάδων x_j που είναι πλησιέστερα στο μπλοκ. Τα ιστογράμματα των επιλεγμένων κεντροειδών αθροίζονται. Οι έννοιες με μεγάλες τιμές πιθανότητας (οι υψηλότεροι ιστοί στο αθροιστικό ιστόγραμμα) χρησιμοποιούνται ως επισημειώσεις της εικόνας δοκιμής. Ο αλγόριθμος των Mori et al. (1999) απεικονίζεται στο σχήμα 22.



Εικόνα 22. Το μοντέλο συν-εμφάνισης λέξης των Mori et al.(1999) (Zhang et al., 2012).

Το μοντέλο μετάφρασης (Translation Model) δημιουργεί μια «ένα προς ένα» αντιστοίχιση ανάμεσα σε ένα blob και μια λέξη (Duygulu et al., 2002). Σε αυτό το μοντέλο, οι περιοχές (blob) από όλες τις εικόνες του συνόλου εκπαίδευσης αρχικά ομαδοποιούνται και αντιπροσωπεύονται από το δείκτη του πλησιέστερου κεντροειδούς της ομάδας (blob). Στη συνέχεια, κάθε blob συνδέεται με μια λέξη στο λεξιλόγιο, όμοια με τη διαδικασία εκμάθησης ενός λεξικού. Η συσχέτιση κάθε blob με μια λέξη στο λεξιλόγιο, επιτυγχάνεται μεγιστοποιώντας την από κοινού πιθανότητα υπολογίζοντας πρώτα την πιθανότητα κάθε blob να σχετίζεται με μία λέξη σε κάθε μία από τις εικόνες. Διατυπώνουν ένα πρόβλημα βελτιστοποίησης για να μάθουν την πιθανότητα μιας λέξης c δεδομένου ενός blob X_j, δηλαδή, $p(c|X_j)$. Η διαδικασία μεγιστοποίησης γίνεται μέσω του αλγόριθμου expectation – maximization (EM), ο οποίος είναι υπολογιστικά δαπανηρός και χρονοβόρος. Κατά τη διάρκεια της επισημείωσης, οι περιοχές μιας εικόνας δοκιμής εκπροσωπούνται από τα πλησιέστερα κεντροειδή (blobs) και συνεπώς, η επισημείωση κάθε περιοχής προσδιορίζεται χρησιμοποιώντας τις πιθανότητες μετάφρασης.

Σε αντίθεση με το μοντέλο μετάφρασης, οι Jeon et al. (2004) αντιστοιχίζουν λέξεις σε ολόκληρες εικόνες αντί για συγκεκριμένα blobs. Μοντελοποιούν την εκ των υστέρων πιθανότητα $p(c|I)$ ως την από κοινού πιθανότητα λέξεων και blobs, και ονομάζουν το μοντέλο τους Cross Media Relevance Model (CMRM). Στη μέθοδο CMRM, μια εικόνα I αντιπροσωπεύεται από ένα σύνολο από blobs $\{b_1, b_2, \ldots, b_n\}$ και η υπό συνθήκη πιθανότητα η εικόνα να ανήκει σε μια κλάση w προσεγγίζεται ως:

$$p(w|I) = p(w|b_1, b_2, \ldots, b_n) \quad (4.13)$$

Το σύνολο εκπαίδευσης που προέρχεται από τις επισημειωμένες εικόνες χρησιμοποιείται για να εκτιμηθεί η από κοινού κοινή πιθανότητα για τη λέξη w και τα blobs $\{b_1, b_2, \ldots, b_n\}$. Η κοινή κατανομή πιθανότητας μπορεί να υπολογιστεί πάνω στην εικόνα j στο σύνολο εκπαίδευσης T ως:

$$p(w, b_1, b_2, \ldots, b_m) = \sum_{j \in T} p(j)\, p(w, b_1, b_2, \ldots, b_m|j) \quad (4.14)$$

Καθώς η εικόνα j είναι γνωστή, η πιθανότητα $p(j)$ είναι σταθερή για ολόκληρο το σύνολο εκπαίδευσης. Υποθέτοντας ότι οι λέξεις w και τα blobs $\{b_1,\ b_2\ ,\ldots,\ b_n\}$ είναι ανεξάρτητα, ένα μοντέλο λέξης και ένα μοντέλο blob κατασκευάζονται για κάθε ξεχωριστή εικόνα εκπαίδευσης j. Έτσι, η εξίσωση (4.14) μπορεί να διορθωθεί ως εξής:

$$p(w, b_1,\ b_2\ ,\ldots,\ b_m) = \sum_{j \in T} p(j)\, p(w|\, j) \prod_{i=1}^{n} p(b_i|\, j) \quad (4.15)$$

$$p(w|\, j) = \left(1 - \alpha_j\right) \frac{\#(\mathrm{w}, \mathrm{j})}{|j|} + \alpha_j \frac{\#(\mathrm{w}, \mathrm{T})}{|T|} \quad (4.16)$$

$$p(b_i|\, j) = \left(1 - \beta_j\right) \frac{\#(b_i, \mathrm{j})}{|j|} + \beta_j \frac{\#(b_i, \mathrm{T})}{|T|} \quad (4.17)$$

όπου $\#(w, j)$ υποδηλώνει τη συχνότητα που εμφανίζεται η λέξη w στις λέξεις-κλειδιά της εικόνας j, και $\#(w, T)$ δηλώνει τη συχνότητα που εμφανίζεται σε όλες τις λέξεις-κλειδιά στο σύνολο εκπαίδευσης T. Η έννοια του $\#(b_i, j)$ και $\#(b_i, T)$ είναι παρόμοια με $\#(w, j)$ και $\#(w, T)$. Εδώ $|j|$ σημαίνει την καταμέτρηση όλων των λέξεων και των blobs που εμφανίζονται στην εικόνα j, και $|T|$ αντιπροσωπεύει το συνολικό μέγεθος του εκπαιδευτικού συνόλου. Ειδικά, α_j και β_j είναι ρυθμιζόμενες παράμετροι.

Το προτεινόμενο μοντέλο μπορεί να χρησιμοποιηθεί για την ταξινόμηση των εικόνων καθώς και για τη δημιουργία μίας καθορισμένου πλήθους ετικετών επισημείωσης από τις ταξινομημένες εικόνες. Το μοντέλο CMRM έχει καλύτερη ακρίβεια επισημείωσης από το μοντέλο μετάφρασης. Ωστόσο, η απόδοση εξαρτάται από την επιλογή των κατάλληλων παραμέτρων α_j και β_j.

Στις μεθόδους TM και CMRM απαιτείται η διακριτοποίηση των συνεχών διανυσμάτων χαρακτηριστικών. Το μοντέλο CRM (Lavrenko et al., 2004) χρησιμοποιεί διανύσματα χαρακτηριστικών συνεχών τιμών για να περιγράψει την εικόνα, δεδομένου ότι η ποσοτικοποίηση συνεχών διανυσμάτων χαρακτηριστικών σε ένα διακριτό λεξιλόγιο θα απωλέσει κάποιες απαραίτητες πληροφορίες της εικόνας (Duygulu et al., 2002). Επιπλέον, το CRM χρησιμοποιεί περιοχές αντί για blobs για να περιγράψει τη δοσμένη εικόνα, καθώς η ικανότητα επισημείωσης του μοντέλου CMRM είναι ευαίσθητη σε σφάλματα λόγω ομαδοποίησης.

Υποτίθεται ότι κάθε εικόνα περιέχει πολλές διαφορετικές περιοχές $\{r_1, r_2, \ldots, r_n\}$, και κάθε περιοχή είναι ένα στοιχείο του R και περιέχει τα εικονοστοιχεία ορισμένων αντικειμένων που ξεχωρίζουν στην εικόνα. Μια συνάρτηση φ μοντελοποιείται για την απεικόνιση της περιοχής εικόνας $r \in R$ σε διανύσματα πραγματικών λέξεων $g \in \mathbb{R}^k$ και η τιμή $\varphi(r)$ αντιπροσωπεύει ένα σύνολο χαρακτηριστικών μιας περιοχής της εικόνας. Στη συνέχεια, η κοινή κατανομή των περιοχών εικόνας $\{r_1, r_2, \ldots, r_n\}$ και ενός συνόλου λέξεων $\{w_1, w_2, \ldots, w\}$ υπολογίζεται από την εξίσωση:

$$p(r_A, w_B) = \sum_{j \in T} p_T(j) \prod_{b=1}^{n_B} p_v(w_b | j) \int_{\mathbb{R}^k} p_R(r_a | g_a)\, p_\varphi(g_a | j)\, dg_a \quad (4.18)$$

όπου $p_R(r_a | g_a)$ αντιπροσωπεύει μια συνολική κατανομή πιθανότητας υπεύθυνη για τη αντιστοίχιση των διανυσμάτων $g \in \mathbb{R}^k$ σε περιοχές εικόνας $r \in R$. Υποτίθεται ότι το διάνυσμα χαρακτηριστικών $\varphi(r)$ όλων των περιοχών σε κάθε εικόνα ακολουθεί τη Gaussian (κανονική) κατανομή. Το $p_v(w_b | j)$ υπολογίζεται με τη χρήση της πολυωνυμικής κατανομής, από την ακόλουθη εξίσωση:

$$p(r|g) = \begin{cases} \frac{1}{N_g}, & \alpha\nu\ \ \varphi(r) = g \\ 0, & \alpha\lambda\lambda\iota\acute{\omega}\varsigma \end{cases} \quad (4.19)$$

$$p_\varphi(g | j) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sqrt{2^k \cdot \pi^k \cdot |\Sigma|}} e^{(g - \varphi(r_i))^T \Sigma^{-1} (g - \varphi(r_i))} \quad (4.20)$$

$$p_v(v | j) = \frac{\mu \cdot p_v + N_{v,i}}{\mu + \Sigma_{v'} N_{v',i}} \quad (4.21)$$

Το μοντέλο MBRM των Feng et al. (2004) χρησιμοποιεί πολλαπλά μοντέλα Bernoulli (κατανομή Bernoulli αντί της πολυωνυμικής κατανομής) για να υπολογίσει τις πιθανότητες λέξεων, όπως φαίνεται από την εξίσωση (4.22). Η υπόθεση είναι ότι πρέπει να δοθεί περισσότερη προσοχή στην ίδια τη λέξη και όχι στη συχνότητά της. Με άλλα λόγια, η παρουσία ή η απουσία μιας λέξης αφορά την επισημείωση της εικόνας και όχι η συχνότητα της ίδιας της λέξης που χρησιμοποιείται.

$$p_v(v | j) = \frac{\mu \cdot \delta_v + N_v}{\mu + N} \quad (4.22)$$

Αργότερα οι Liu et al. (2007a), επέκτειναν το CMRM για να αναλύσουν επίσης τη σχέση μεταξύ λέξεων. Αυτό το ολοκληρωμένο μοντέλο CMRM στο οποίο συνδυάζονται συσχετίσεις λέξεων, ανάκτηση εικόνων και τεχνικές αναζήτησης ιστού για να λύσουν το πρόβλημα επισημείωσης ονομάζεται dual-CMRM.

4.4.1.2. Το μοντέλο μίγματος

Το μοντέλο του μίγματος βασίζεται στο παραμετρικό μοντέλο. Στην παραμετρική προσέγγιση, ο χώρος χαρακτηριστικών θεωρείται ότι ακολουθεί έναν ορισμένο τύπο μιας γνωστής συνεχούς κατανομής. Επομένως, η υπό συνθήκη πιθανότητα $p\,(x \mid c)$ διαμορφώνεται χρησιμοποιώντας αυτή την κατανομή χαρακτηριστικών. Η γενική διαδικασία είναι παρόμοια με αυτή που παρουσιάζεται στο σχήμα 21. Χαρακτηριστικά ή περιοχές ομαδοποιούνται πρώτα και ποσοτικοποιούνται και το μοντέλο της υπό συνθήκη πιθανότητας κατασκευάζεται για κάθε ομάδα (ή blob).

Η υπό συνθήκη πιθανότητα $p\,(x \mid c)$ διαμορφώνεται συνήθως ως πολυπαραγοντική Gaussian κατανομή, όπως φαίνεται στην εξίσωση (4.23):

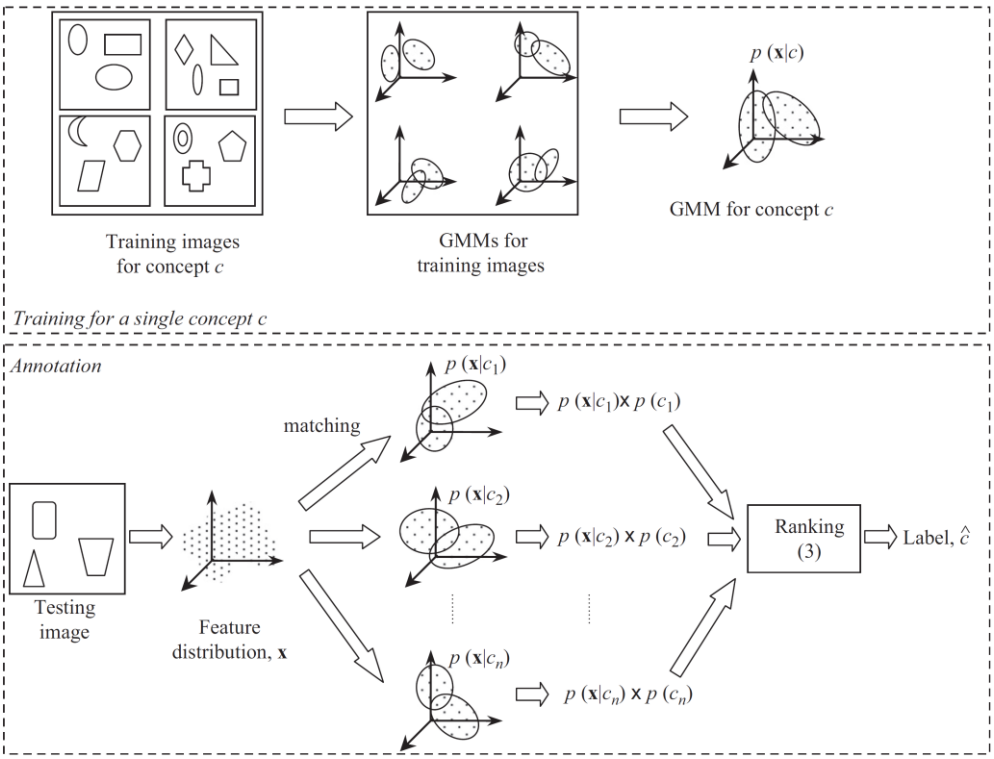
$$p(x|c) = \frac{1}{\sqrt{2^d \cdot \pi^d \cdot |\Sigma|}} e^{-(x-\bar{x})^T \Sigma^{-1} (x-\bar{x})} \quad (4.23)$$

όπου d είναι η διάσταση του διανύσματος των χαρακτηριστικών, $\bar{x}$ και Σ είναι η μέση τιμή και ο πίνακας συνδιακύμανσης υπολογιζόμενος από τα διανύσματα χαρακτηριστικών των εικόνων εκπαίδευσης που ανήκουν στην έννοια c.

Οι Carneiro et al. (2007), ακολουθούν την προσέγγιση αυτή, και μαθαίνουν τα υπό συνθήκη μοντέλα πιθανοτήτων έννοια προς έννοια και στη συνέχεια χρησιμοποιούν τα μοντέλα για να επισημειώσουν άγνωστες εικόνες. Δεν χωρίζουν τις εικόνες σε περιοχές, αλλά αντίθετα, υποθέτουν ότι τα χαρακτηριστικά της εικόνας ακολουθούν συγκεκριμένες Gaussian κατανομές και απευθείας εκπαιδεύουν ένα GMM για κάθε εικόνα εκπαίδευσης μέσα σε μια έννοια χρησιμοποιώντας τον αλγόριθμο μεγιστοποίησης προσδοκίας (EM). Αυτό ισοδυναμεί με μια ταυτόχρονη διαδικασία τμηματοποίησης και μηχανικής μάθησης. Στη συνέχεια, αναπτύσσουν το GMM μοντέλο της έννοιας από τον μέσο όρο των μεμονωμένων GMM εντός της έννοιας. Στο στάδιο της επισημείωσης, ένα GMM μαθαίνεται για την άγνωστη εικόνα και

αυτό το GMM αντιστοιχίζεται στη συνέχεια στο GMM μοντέλο κάθε έννοιας. Οι έννοιες με την καλύτερη αντιστοίχιση επιλέγονται ως οι επισημειώσεις για την άγνωστη εικόνα. Το σχήμα 23 αποτυπώνει αυτή τη διαδικασία. Το μειονέκτημα είναι ότι η εκτίμηση των μοντέλων GMM είναι περίπλοκη λόγω της χρήσης της μεθόδου βελτιστοποίησης ΕΜ.

Οι Wang et al. (2009), χρησιμοποιούν ένα Gaussian Mixture Model (GMM) για την εξαγωγή χαρακτηριστικών και προτείνουν το αραιό πλαίσιο κωδικοποίησης για την επισημείωση εικόνας. Το προτεινόμενο GMM είναι εμπνευσμένο από χαρακτηριστικά διακριτού μετασχηματισμού συνημιτόνου (DCT) και χρησιμοποιούν μάθηση υποπεριοχών για να αξιοποιήσουν αποδοτικά τις πληροφορίες πολλαπλών ετικετών για την εξαγωγή χαρακτηριστικών.



Εικόνα 23. Μοντελοποίηση των υπό συνθήκη πιθανοτήτων και επισημείωση εικόνας με χρήση ιεραρχικών GMMs από τους Carneiro et al. (2007) (Zhang et al., 2012).

4.4.1.3. Το μοντέλο θέματος

Το μοντέλο θέματος είναι ένας άλλος τύπος ευρέως χρησιμοποιούμενων παραγωγικών μοντέλων για την AIA. Οι μέθοδοι AIA που βασίζονται στο μοντέλο θέματος θεωρούν τις επισημειωμένες εικόνες ως δείγματα από ένα συγκεκριμένο μείγμα θεμάτων, όπου κάθε θέμα (topic) είναι μια κατανομή πιθανότητας πάνω σε χαρακτηριστικά εικόνας και λέξεις

επισημείωσης. Ένα μοντέλο θέματος είναι ένα ισχυρό μη-επιβλεπόμενο εργαλείο για την ανάλυση εγγράφων κειμένου. Ορισμένες καθιερωμένες προσεγγίσεις ανάλυσης εγγράφων, όπως η pLSA (probabilistic latent semantic analysis - pLSA) και η LDA (latent Dirichlet Allocation - LDA), έχουν προσαρμοστεί επιτυχώς για να χειριστούν την κοινή μοντελοποίηση των οπτικών και περιεχομένων πληροφοριών, χρησιμοποιώντας την έννοια του θέματος.

Το μοντέλο pLSA υποθέτει ότι μια ομάδα συν-εμφανιζόμενων λέξεων συνδέεται με ένα λανθάνον θέμα. Γενικά, ένα θέμα είναι μια εννοιολογική ιδέα και χαρακτηρίζεται από μια σειρά σχετικών λέξεων. Για παράδειγμα, εάν η “Microsoft” θεωρείται θέμα, τότε οι λέξεις “Bill Gates” και “Microsoft Windows” πιθανότατα εμφανίζονται συχνά σε αυτό το θέμα. Δεδομένου ότι το μοντέλο pLSA υποθέτει την ύπαρξη ενός λανθάνοντος θέματος z στη παραγωγική διαδικασία για κάθε στοιχείο x_j σε ένα συγκεκριμένο έγγραφο d_i, η κοινή πιθανότητα (joint probability) του στοιχείου x και του εγγράφου d υπολογίζεται ως:

$$p(x_j, d_i) = p(d_i) \sum_k p(z_k|d_i) \cdot p(x_j|z_k) \quad (4.24)$$

Το μοντέλο pLSA έχει βελτιωθεί με πολλούς τρόπους. Για παράδειγμα, οι Lienhart και Romberg (2009) πρότειναν ένα πολυεπίπεδο μοντέλο pLSA. Η πολυεπίπεδη δομή διευκολύνει την επέκταση των κανόνων μάθησης και συμπερασμάτων σε περισσότερα επίπεδα και τρόπους.

Η βασική ιδέα του μοντέλου LDA είναι ότι τα έγγραφα περιγράφονται ως τυχαία μείγματα πάνω σε λανθάνοντα θέματα, όπου κάθε θέμα χαρακτηρίζεται από μία κατανομή πάνω σε λέξεις. Οι Blei and Jordan (2003) πρότειναν το Correlation LDA, το οποίο είναι μια επέκταση του LDA, για να συνδέσουν λέξεις και εικόνες. Αυτό το μοντέλο υποθέτει ότι μπορεί να χρησιμοποιηθεί μια κατανομή Dirichlet για τη δημιουργία του μίγματος λανθάνοντων παραγόντων. Αυτό το μίγμα λανθάνοντων παραγόντων στη συνέχεια χρησιμοποιείται για τη δημιουργία λέξεων και περιοχών. Ο αλγόριθμος EM χρησιμοποιείται για την εκτίμηση των παραμέτρων αυτού του μοντέλου. Τα αποτελέσματα των πειραμάτων που παρουσιάζονται στους Blei and Jordan (2003), υποδεικνύουν ότι η LDA έχει χαμηλότερη πολυπλοκότητα στην αναπαράσταση εγγράφων κειμένου απ’ ό,τι μια τυπική bag-of-word προσέγγιση

(χαμηλότερη πολυπλοκότητα υποδεικνύει καλύτερη απόδοση γενίκευσης) και επίσης μεγαλύτερη ακρίβεια στις εργασίες ταξινόμησης εγγράφων.

Το μοντέλο tr-mmLDA (topic-regression multi-modal Latent Dirichlet Allocation) παρουσιάζει μια διαφοροποιημένη μέθοδο με σκοπό την εκμάθηση της κοινής κατανομής κειμένου και χαρακτηριστικών εικόνας. Το μοντέλο παρέχει μια εναλλακτική μέθοδο για την εκμάθηση δύο ομάδων κρυφών θεμάτων και ενσωματώνει μια μονάδα γραμμικής παλινδρόμησης για την καταγραφή στατιστικών συσχετισμών μεταξύ εικόνων και κειμένου. Αυτό το μοντέλο tr-mmLDA είναι αρκετά διαφορετικό από τα προηγούμενα μοντέλα θέματος που μοιράζονται μόνο ένα σύνολο λανθανόντων θεμάτων μεταξύ δύο τύπων δεδομένων. Επιπλέον, το tr-mmLDA μπορεί να χειριστεί τις διαφορές στον αριθμό των θεμάτων στις δύο μεθόδους δεδομένων, ένα πλεονέκτημα σε σχέση με τα προηγούμενα μοντέλα στα οποία το άθροισμα των λανθανόντων θεμάτων πρέπει να αποφασιστεί χειροκίνητα (Putthividhy et al., 2010).

Τα παραγωγικά μοντέλα έχουν συμβάλει σημαντικά στην ανάπτυξη της AIA και πολλές μέθοδοι AIA εμπνέονται από παραγωγικά μοντέλα. Ωστόσο, υπάρχουν τρεις κύριες ελλείψεις στις μεθόδους AIA που βασίζονται σε παραγωγικά μοντέλα. Το πρώτο είναι ότι τα παραγωγικά μοντέλα εκτιμούν την κοινή πιθανότητα των χαρακτηριστικών της εικόνας και των επισημειώσεων, αλλά δεν μπορούν να εγγυηθούν τη βελτιστοποίηση της πρόβλεψης της ετικέτας. Το δεύτερο είναι ότι τα παραγωγικά μοντέλα μπορεί να μην είναι σε θέση να καταγράψουν την περίπλοκη σχέση μεταξύ χαρακτηριστικών εικόνας και ετικετών. Η τρίτη είναι το υψηλό υπολογιστικό κόστος που οφείλεται από στη χρήση περίπλοκων αλγορίθμων, όπως ο αλγόριθμος βελτιστοποίησης EM και στις πολυάριθμες ρυθμίσεις παραμέτρων (Cheng et al., 2018).

4.4.2. Μέθοδοι AIA βασισμένες στο διακριτικό μοντέλο

Οι μέθοδοι AIA βασισμένες σε διακριτικά μοντέλα θεωρούν την επισημείωση εικόνας ως πρόβλημα ταξινόμησης πολλαπλών ετικετών. Το πρόβλημα αυτό επιλύεται με την εκμάθηση ενός ανεξάρτητου δυαδικού ταξινομητή για κάθε ετικέτα και στη συνέχεια με τη χρήση των δυαδικών ταξινομητών για την πρόβλεψη των ετικετών για τις μη επισημειωμένες εικόνες.

Τα περισσότερα διακριτικά μοντέλα βασίζονται στη μηχανή διανυσμάτων υποστήριξης (SVM) ή στις παραλλαγές της. Τα διακριτικά μοντέλα χρησιμοποιούνται εκτεταμένα για την

επισημείωση ιατρικών εικόνων όπου ο SVM χρησιμοποιείται ως ταξινομητής. Διακριτικά μοντέλα βασισμένα στα τεχνητά νευρωνικά δίκτυα (ANN) έχουν επίσης εφαρμοστεί για επισημείωση εικόνας.

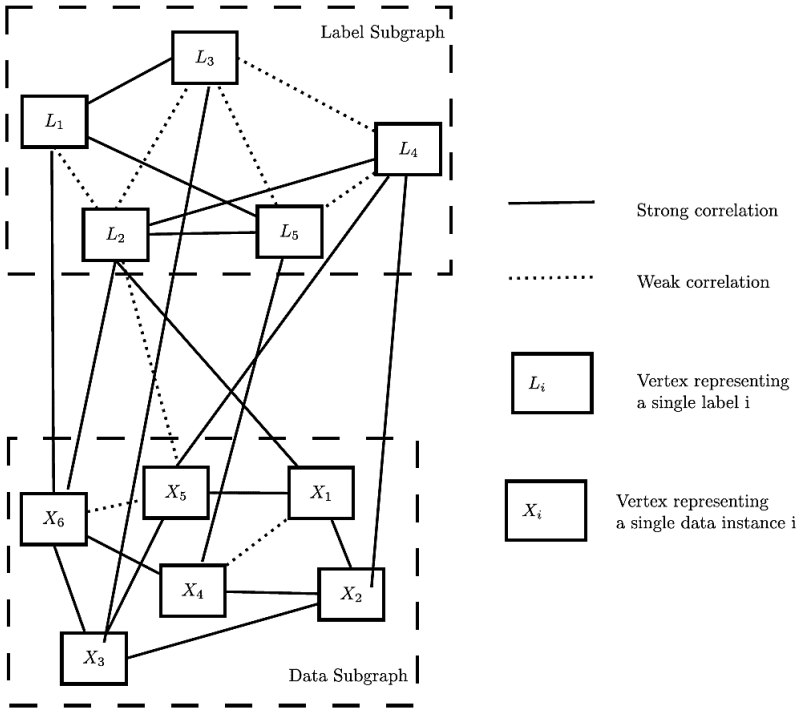
Την τελευταία δεκαετία, το ενδιαφέρον των ερευνητών έχει στραφεί σε μεθόδους μάθησης με μερική επίβλεψη (Semi-Supervised Learning - SSL), λόγω της έλλειψης επαρκούς αριθμού δεδομένων εκπαίδευσης με πλήρη επισημείωση και του σχετικού κόστους. Τα τελευταία χρόνια, οι ερευνητικές προσπάθειες για ημι-εποπτευόμενη μάθηση έχουν αφιερωθεί κυρίως στις μεθόδους μάθησης με βάση γράφους, οι οποίες μοντελοποιούν ολόκληρα τα δεδομένα ως ένα γράφο. Η συσχέτιση των ετικετών μπορεί εύκολα να ενσωματωθεί στο γράφο σε αυτή τη μέθοδο που βασίζεται στη διάδοση. Υπάρχουν κυρίως δύο τρόποι χρήσης των συσχετίσεων των ετικέτων στη μέθοδο μάθησης με βάση το γράφο. Η συσχέτιση των ετικετών χρησιμοποιείται ως μέρος των βαρών του γράφου ή ως πρόσθετος περιορισμός.

4.4.2.1. Μοντέλο με βάση γράφο

Η βασική ιδέα πίσω από το μοντέλο που βασίζεται σε γράφους είναι να σχεδιαστεί ένας γράφος από τα οπτικά χαρακτηριστικά και τα χαρακτηριστικά κειμένου με τέτοιο τρόπο ώστε η μεταξύ τους συσχέτιση να μπορεί να αναπαρίσταται με τη μορφή κορυφών και ακμών και η εξάρτησή τους να μπορεί να εξηγηθεί. Τα σημεία δεδομένων (οπτικά χαρακτηριστικά των εικόνων) και οι ετικέτες μπορούν να αναπαριστώνται ως χωριστοί υπογράφοι και οι ακμές αντιπροσωπεύουν τη συσχέτιση μεταξύ των υπογράφων (Wang et al., 2011). Αυτοί οι υπογράφοι συνδέονται με ένα πρόσθετο διμερές γράφο που συνεπάγονται οι εκχωρήσεις των ετικετών. Ο αλγόριθμος τυχαίου περιπάτου εκτελείται τόσο στη σχέση της κλάσης με την εικόνα όσο και στη σχέση μεταξύ κλάσεων, εξετάζοντας κάθε κλάση και τις ετικέτες της ως σημασιολογική ομάδα. Ειδικά, η σχέση μεταξύ κλάσεων περιγράφεται με ασύμμετρο τρόπο για να μιμηθεί τη σημασιολογική σχέση στον πραγματικό κόσμο και η σχέση μεταξύ κλάσης και εικόνας χρησιμοποιείται για την ευθεία πρόβλεψη ετικετών για μη επισημειωμένες εικόνες.

Η σημασιολογική συσχέτιση μεταξύ ετικετών μπορεί να αναπαρασταθεί χρησιμοποιώντας διασυνδεδεμένους κόμβους που βοηθούν στην επισημείωση εικόνας πολλαπλών ετικετών. Το μοντέλο γράφων μπορεί επίσης να χρησιμοποιηθεί για να εντοπίσει τη συσχέτιση μεταξύ των ετικετών. Σε μια τέτοια περίπτωση, οι κορυφές αντιπροσωπεύουν ετικέτες και οι ακμές

αντιπροσωπεύουν συσχέτιση μεταξύ ετικετών. Το σχήμα 24 δίνει ένα παράδειγμα συσχέτισης μεταξύ οπτικών χαρακτηριστικών και χαρακτηριστικών κειμένου χρησιμοποιώντας ένα γράφο.



Εικόνα 24. Παράδειγμα μοντέλου γράφου (Bhagat and Choudhary, 2018).

Το μοντέλο γράφων μπορεί να χρησιμοποιηθεί τόσο σε επιβλεπόμενο όσο και σε το ημι-επιβλεπόμενο πλαίσιο για να μοντελοποιήσει την εγγενή δομή των δεδομένων εικόνας και από επισημειωμένες και αλλά και από μη επισημειωμένες εικόνες (Bhagat and Choudhary, 2018).

Το τυχαίο πεδίο Markov (Markov Random Field – MRF) είναι ένα πιθανοτικό μοντέλο μη κατευθυνόμενου γράφου. Έχει χρησιμοποιηθεί για να διερευνήσει σημασιολογικές σχέσεις ανάμεσα σε έννοιες και χαμηλού επιπέδου χαρακτηριστικά στην AIA. Τα διακριτικά μοντέλα που βασίζονται στο MRF, αποδείχθηκαν πιο αποτελεσματικά από το παραγωγικό μοντέλο που πάσχει από την αδύναμη ικανότητα μάθησης λόγω της έλλειψης κατάλληλης στρατηγικής μάθησης για τον χαρακτηρισμό του σημασιολογικού πλαισίου (Cheng et al., 2018).

Οι Xiang et al. (2009) και οι Llorente and Manmatha (2010) παρουσίασαν προσεγγίσεις βασισμένες στο MRF για να αξιοποιήσουν τις σημασιολογικές εξαρτήσεις των εικόνων. Οι Xiang et al. (2009) υιοθέτησαν το MRF για να μοντελοποιήσουν το πλαίσιο των σχέσεων μεταξύ των σημασιολογικών εννοιών με τους υπογράφους των λέξεων-κλειδιών που παράγονται από δείγματα εκπαίδευσης για κάθε λέξη-κλειδί. Μία συνάρτηση δυνητικών θέσεων και μια συνάρτηση δυνητικών ακμών ορίζονται ως πρότυπο της κοινής πιθανότητας ενός χαρακτηριστικού εικόνας και μιας λέξης.

Οι Llorente and Manmatha (2010) δημιούργησαν ένα μη κατευθυνόμενο γράφο όπου ένας κόμβος θα μπορούσε να είναι μια εικόνα των συνόλων δεδομένων δοκιμής ή ενός ερωτήματος. Η μελέτη επικεντρώθηκε στη διερεύνηση των εξαρτήσεων μεταξύ χαρακτηριστικών εικόνας και λέξεων, των εξαρτήσεων μεταξύ δύο λέξεων και των εξαρτήσεων μεταξύ των χαρακτηριστικών εικόνας και λέξεων. Η καινοτομία της μεθόδου έγκειται στη χρήση διαφορετικών πυρήνων (όπως ο πυρήνας "τετραγωνικής ρίζας" ή ο Laplacian πυρήνας) στη μη-παραμετρική εκτίμηση της πυκνότητας, καθώς και η χρήση ρυθμίσεων για να εξερευνήσουν σημασιολογικές σχέσεις μεταξύ εννοιών. Έτσι, είναι εύκολο να συγκρίνουν και να αναλύσουν τις επιδόσεις σε πολλές διαφορετικές ρυθμίσεις.

Οι μέθοδοι επισημείωσης εικόνας βασισμένες σε γράφους είναι μεταγωγικές και μπορούν να προβλέψουν μόνο ετικέτες για συγκεκριμένα μη επισημειωμένα δείγματα. Με άλλα λόγια, για να επισημειωθεί μια νέα εικόνα δοκιμής, η εικόνα δοκιμής πρέπει να προστεθεί πρώτα στο μη επισημειωμένο σύνολο και στη συνέχεια η φάση εκπαίδευσης θα επαναληφθεί. Ωστόσο, αυτό δεν είναι πρακτικό για εργασίες μαζικής επισημείωσης εικόνων στις μέρες μας.

4.4.3. Μοντέλα που βασίζονται στον πλησιέστερο γείτονα

Τα μοντέλα που βασίζονται στον πλησιέστερο γείτονα (Nearest Neighbor) επικεντρώνονται κατά κύριο λόγο στην επιλογή παρόμοιων γειτόνων και μετά στη διάδοση των ετικετών στην εικόνα δοκιμής, καθώς στηρίζονται στην υπόθεση ότι οπτικά παρόμοιες εικόνες είναι πιο πιθανό να μοιράζονται κοινές ετικέτες. Οι όμοιοι γείτονες μπορούν να προσδιοριστούν από την ομοιότητα μεταξύ δύο εικόνων (οπτική ομοιότητα) ή την ομοιότητα μεταξύ εικόνας και ετικέτας ή και στα δύο. Χρησιμοποιείται ένα μέτρο (μια μετρική) απόστασης για την επιλογή παρόμοιων γειτόνων. Η αποτελεσματικότητα ενός μέτρου της απόστασης διαδραματίζει

ζωτικό ρόλο στην επιλογή των σχετικών και κατάλληλων γειτόνων, και συνεπώς στη συνολική απόδοση της μεθόδου. Ο πλησιέστερος γείτονας είναι ένας μη παραμετρικός ταξινομητής που μπορεί να επεξεργαστεί άμεσα τα δεδομένα χωρίς παραμέτρους μάθησης. Διάφορες τεχνικές επισημείωσης εικόνων που χρησιμοποιούν μοντέλα βασισμένα σε πλησιέστερο γείτονα, παρουσιάζονται στη βιβλιογραφία (Bhagat and Choudhary, 2018, Cheng et al., 2018).

Το μοντέλο Joint Equal Contribution (JEC) των Makadia et al. (2008) είναι ένα από τα πιο κλασικά μοντέλα πλησιέστερων γειτόνων. Δημιουργεί μια οικογένεια πολύ απλών και διαισθητικών βασικών μεθόδων για επισημείωση εικόνας. Το μοντέλο JEC χρησιμοποιεί συνολικά χαρακτηριστικά χρώματος και υφής και έναν απλό συνδυασμό βασικών μέτρων απόστασης για να εντοπίσει πλησιέστερους γείτονες μιας δεδομένης εικόνας. Στη συνέχεια, οι λέξεις-κλειδιά εκχωρούνται στην εικόνα δοκιμής χρησιμοποιώντας έναν άπληστο αλγόριθμο μεταφοράς ετικετών (label transfer), ο οποίος επιλέγει τις λέξεις-κλειδιά από τον πλησιέστερο γείτονα ή τους γείτονες αφού προηγουμένως τις ταξινομήσει με βάση τη συχνότητά τους.

Το κύριο πρόβλημα με το μοντέλο πλησιέστερου γείτονα είναι ότι απαιτεί εντελώς χειροκίνητα επισημειωμένο σύνολο εκπαίδευσης. Επίσης, στο σύνολο εκπαίδευσης κάθε ετικέτα πρέπει να έχει επαρκή αριθμό εικόνων και επιπλέον ο αριθμός εικόνων ανά ετικέτα πρέπει να είναι σχεδόν ο ίδιος. Το πρόβλημα της ανισορροπίας κλάσεων (class-imbalance) είναι αρκετά κοινό όταν το μέγεθος του λεξιλογίου ετικετών είναι μεγάλο. Σημαίνει ότι υπάρχει μεγάλη διακύμανση μεταξύ του αριθμού των εικόνων που αντιστοιχούν σε διαφορετικές ετικέτες. Στις περισσότερες περιπτώσεις, η ανισορροπία των κλάσεων οδηγεί σε κακή επισημείωση καθώς οι μέθοδοι AIA που βασίζονται στους πλησιέστερους γείτονες, επισημειώνουν τις εικόνες με τη βοήθεια παρακείμενων (γειτονικών) εικόνων. Όσο πιο συχνά χρησιμοποιείται μια υποψήφια ετικέτα για να περιγράψει τους γείτονές της, τόσο μεγαλύτερη είναι η πιθανότητα αυτή η ετικέτα να χρησιμοποιηθεί για την επισημείωση μιας γειτονικής μη επισημειωμένης εικόνας. Με άλλα λόγια, εάν μια ετικέτα αντιπροσωπεύεται μόνο από λίγες περιπτώσεις (στιγμιότυπα), η πιθανότητα να χρησιμοποιηθεί η ίδια ετικέτα για μία εικόνα χωρίς ετικέτα θα είναι πολύ μικρή.

Ένα άλλο πρόβλημα, η αδύναμη επισημείωση (weak-labeling), οφείλεται σε περιορισμούς της χειροκίνητης επισημείωσης σε κάποιο βαθμό. Από τη μια πλευρά, αυτό σημαίνει ότι ένας

σημαντικός αριθμός διαθέσιμων εικόνων δεν επισημειώνεται με όλες τις σχετικές ετικέτες. Από την άλλη πλευρά, υποδεικνύει ότι ορισμένες εικόνες ενδέχεται να επισημειώνονται με μη σχετικές ετικέτες. Αναμφισβήτητα, η αδύναμη επισημείωση θα προκαλέσει επίσης κακή επισημείωση.

Το μοντέλο διάδοσης ετικέτας (tag propagation) TagProp (Guillaumin et al., 2009) ενσωματώνει μια μέθοδο σταθμισμένου πλησιέστερου γείτονα και τις δυνατότητες μετρικής μάθησης σε ένα διακριτικό πλαίσιο. Μεταφέρει ετικέτες λαμβάνοντας ένα σταθμισμένο συνδυασμό της παρουσίας και της απουσίας ετικέτας των γειτόνων. Τα βάρη των γειτόνων βασίζονται στην κατάταξη των γειτόνων ή την απόσταση και ορίζονται αυτόματα μεγιστοποιώντας την πιθανότητα των επισημειώσεων σε ένα σύνολο εικόνων εκπαίδευσης. Με βάση την κατάταξη των βαρών, ο k-οστός γείτονας λαμβάνει πάντοτε ένα σταθερό βάρος, ενώ τα βάρη βάσει απόστασης μειώνονται εκθετικά με την απόσταση. Το μοντέλο TagProp επιτρέπει την ενσωμάτωση της μετρικής μάθησης. Αυτό της επιτρέπει να βελτιστοποιήσει ένα συνδυασμό διαφόρων μέτρων απόστασης μεταξύ χαρακτηριστικών εικόνας για να καθορίσει τα βάρη των γειτόνων, για την πρόβλεψη της ετικέτας. Επιπλέον, εισάγει μοντέλα λέξης διακριτικής φύσης, τα οποία αυξάνουν την πιθανότητα για σπάνιες ετικέτες και μειώνουν την πιθανότητα των συχνών ετικετών ταυτόχρονα για υπέρβαση του προβλήματος της ανισορροπίας τάξης. Το μοντέλο πρόβλεψης ετικετών TagProp είναι εννοιολογικά απλό, αλλά ξεπερνά πολλές state-of-the-art μεθόδους.

Το μοντέλο 2PKNN (two pass KNN) (Verma and Jawahar, 2012), μια παραλλαγή δύο βημάτων του κλασικού αλγόριθμου k-πλησιέστερων γειτόνων, αντιπροσωπεύει μια κλασική λύση για την επίλυση προβλημάτων που σχετίζονται με την ανισορροπία κλάσεων και την αδύναμη επισημείωση. Το 2PKNN χρησιμοποιεί τους δύο τύπους ομοιότητας σε δύο περάσματα. Στο πρώτο πέρασμα, χρησιμοποιείται ομοιότητα εικόνας προς ετικέτα και στο δεύτερο πέρασμα χρησιμοποιείται ομοιότητα εικόνας προς εικόνα. Αναγνωρίζει όλους τους σχετικούς σημασιολογικούς γείτονες για κάθε ετικέτα επιλέγοντας k παρόμοιες εικόνες στο λεξιλόγιο. Έτσι, μπορεί να διασφαλιστεί ότι κάθε ετικέτα εμφανίζεται τουλάχιστον k φορές στο σύνολο εκπαίδευσης.

Εμπνευσμένοι από το μοντέλο 2PKNN, οι Bakliwal and Jawahar (2015) στην προσπάθειά τους να χειριστούν το πρόβλημα της ανισορροπίας των κλάσεων με τρόπο ώστε κάθε ετικέτα να

εμφανίζεται τουλάχιστον k φορές στα δεδομένα εκπαίδευσης, προτείνουν έναν αλγόριθμο σταθμισμένου πλησιέστερου γείτονα ο οποίος αποδίδει σημασία στις ετικέτες με βάση την ομοιότητα εικόνας και στη συνέχεια υπολογίζει τη βαθμολογία για κάθε ετικέτα για μια νέα εικόνα. Επιπλέον, η ακρίβεια επισημείωσης βελτιώνεται καθώς αναθέτουν μεταβλητό αριθμό ετικετών στις εικόνες, σε αντίθεση με τις προηγούμενες μελέτες που αναθέτουν σταθερό αριθμό ετικετών στις εικόνες χωρίς ετικέτα.

Εκτός από τις προσπάθειες που καταβάλλονται για ανοικτά ζητήματα όπως η DML, η ανισορροπία των κλάσεων και η αδύναμη επισημείωση, μελέτες σχετικές με άλλες πτυχές όπως η συνάφεια των ετικετών (Tian and Shen, 2014) και ο μεταβλητός αριθμός των πλησιέστερων γειτόνων (Lin et al., 2012), αποσκοπούν επίσης να βελτιώσουν τις επιδόσεις των μεθόδων AIA με βάση μοντέλα πλησιέστερων γειτόνων. Οι Tian and Shen (2014) ανέπτυξαν ένα μοντέλο, το LSLabel, που στοχεύει στη μάθηση της συνάφειας των ετικετών. Αντί να εκτιμάται η συνάφεια της ετικέτας για μια εικόνα από την συχνότητα των επισημειώσεων που προκύπτει από τους κοντινότερους γείτονές της, οι Tian and Shen (2014) εκτιμούν τη συνάφεια των ετικετών, υπολογίζοντας τη συνάφεια μεταξύ «συνόλου ετικετών» και εικόνας και τη συνάφεια της ετικέτας με τις άλλες ετικέτες σε ένα κοινό πλαίσιο. Οι Lin et al. (2012) χρησιμοποίησαν ένα περιορισμένο εύρος παρά έναν ταυτόσημο και σταθερό αριθμό οπτικών γειτόνων για την πρόβλεψη της ετικέτας. Το μοντέλο τους, το TagSearcher, θεωρείται αρκετά ανώτερο από πολλές προηγούμενες μεθόδους βασισμένες σε οπτικά πλησιέστερους γείτονες που είναι ευαίσθητες στον αριθμό των οπτικά παρόμοιων γειτόνων.

4.4.4. Μέθοδοι AIA βασισμένες στη βαθιά μάθηση

Την τελευταία δεκαετία έχουμε γίνει μάρτυρες της σημαντικής ανάπτυξης τεχνικών βαθιάς μάθησης, οι οποίες επιτρέπουν τη αναπαράσταση οπτικών χαρακτηριστικών για την υποβοήθηση του έργου της AIA. Οι τελευταίες εξελίξεις στην βαθιά μάθηση επιτρέπουν μια ποικιλία βαθιών μοντέλων για μεγάλης κλίμακας επισημείωση εικόνων (Cheng et al., 2018).

Η AIA που βασίζεται σε βαθιά μάθηση μπορεί να συνοψιστεί σε δύο πτυχές. Πρώτον, στη δημιουργία ισχυρών οπτικών χαρακτηριστικών με χρήση συνελικτικών νευρωνικών δικτύων (ConvNets), για επισημείωση εικόνας. Δεύτερον, στην εξαγωγή παράπλευρων πληροφοριών (όπως οι σημασιολογικές σχέσεις μεταξύ ετικετών) για την AIA. Η AIA βασισμένη στη βαθιά μάθηση είναι μια αρκετά νέα αλλά ελπιδοφόρα κατεύθυνση για την AIA (Cheng et al., 2018).

4.4.4.1. Ισχυρά οπτικά χαρακτηριστικά

Η εξαγωγή οπτικών χαρακτηριστικών είναι ο πιο καθοριστικός παράγοντας στην επισημείωση εικόνας. Τα παραδοσιακά οπτικά χαρακτηριστικά που εξάγονται με άλλες μεθόδους από την εικόνα είναι στοχαστικά και μη ικανοποιητικά. Εμπνευσμένοι από την επιτυχία των ConvNets στην μηχανική όραση, οι ερευνητές τείνουν να χρησιμοποιούν το ConvNet για να παράγουν ισχυρά οπτικά χαρακτηριστικά για την AIA.

Το μοντέλο CNN+WARP (Weighted Approximate Ranking - Ζυγισμένη κατά προσέγγιση κατάταξη) των Gong et al. (2013) χρησιμοποιεί την κατάταξη για την κατάρτιση συνελικτικών νευρωνικών δικτύων για προβλήματα επισημείωσης εικόνας πολλαπλών ετικετών. Η αρχιτεκτονική του ConvNet που χρησιμοποιούν, διαθέτει πέντε συνελικτικά στρώματα και τρία πυκνά συνδεδεμένα στρώματα. Για ένα σύνολο εικόνων x, το συνελικτικό δίκτυο υποδηλώνεται με $f(.)$ όπου τα συνελικτικά στρώματα και τα πυκνά συνδεδεμένα στρώματα φιλτράρουν τις εικόνες. Η έξοδος του $f(.)$ είναι μια συνάρτηση βαθμολόγησης του σημείου δεδομένων x που περιέχει ένα διάνυσμα ενεργοποιήσεων. Υποτίθεται ότι υπάρχουν n εικόνες και c ετικέτες για την εκπαίδευση. Η συνάρτηση απώλειας WARP ελαχιστοποιεί την ακόλουθη ποσότητα:

$$J = \sum_{i=1}^{r} \sum_{j=1}^{c_+} \sum_{k=1}^{c_-} L(r_j) max\left(0, 1 - f_j(x_i) + f_k(x_i)\right) \quad (4.25)$$

όπου $L(.)$ είναι μια συνάρτηση στάθμισης για διαφορετικές τάξεις, και r_j είναι η τάξη για την κλάση j για την εικόνα i. Η συνάρτηση στάθμισης $L(.)$ ορίζεται ως:

$$L(.) = \sum_{j=1}^{r} a_j \quad (4.26)$$

όπου, a_j ορίζεται ως $1 / j$, ενώ τα βάρη που ορίζονται από το $L(.)$ Ελέγχουν το top-k της βελτιστοποίησης. Συγκεκριμένα, αν μια θετική ετικέτα καταταχθεί στην κορυφή της λίστας ετικετών, τότε η $L(.)$ θα αποδώσει ένα μικρό βάρος στην απώλεια και δεν θα κοστίσει υπερβολικά την απώλεια. Ωστόσο, εάν μια θετική ετικέτα δεν κατατάσσεται στην κορυφή, το

$L(.)$ θα αποδώσει πολύ μεγαλύτερο βάρος στην απώλεια, που ωθεί τη θετική ετικέτα στην κορυφή. Επιπλέον, η τάξη r_j εκτιμάται από τη σχέση:

$$r_j = \left| \frac{c-1}{s} \right| \quad (4.27)$$

για τις c κλάσεις και s δοκιμές δειγματοληψίας.

Στη συνέχεια υπολογίζεται η υποκλίση για αυτό το στρώμα κατά τη διάρκεια της βελτιστοποίησης. Για συγκρίσεις, οι Gong et al. (2013) χρησιμοποίησαν ένα σύνολο 9 διαφορετικών οπτικών χαρακτηριστικών (GIST, D-SIFT, D-CSIFT, D-RGBSIFT, H-SIFT, H-CSIFT, H- RGBSIFT, HOG και χαρακτηριστικά χρώματος). Με βάση αυτά τα χαρακτηριστικά, δύο απλοί αλλά ισχυροί ταξινομητές (kNN και SVM) χρησιμοποιήθηκαν για επισημείωση εικόνας. Μέσω της σύγκρισης των πλαισίων βασισμένων σε χαρακτηριστικά του ConvNet με τους ταξινομητές που βασίζονται στα βασικά οπτικά χαρακτηριστικά, τα πειραματικά αποτελέσματα έδειξαν ότι το ConvNet είχε καλύτερη απόδοση από τις υπάρχουσες μεθόδους με οπτικά χαρακτηριστικά στην επισημείωση εικόνας.

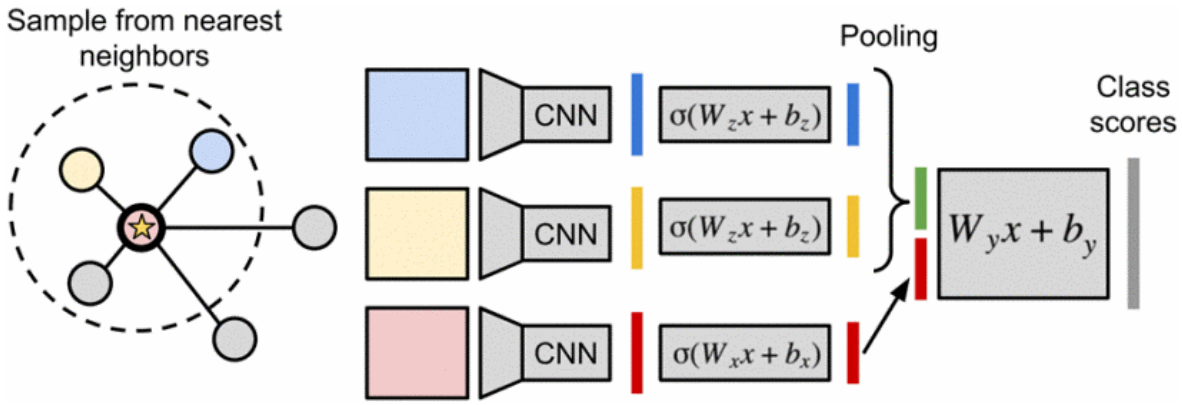
Οι Mayhew et al. (2016), ακολουθώντας την ίδια μεθοδολογία, εκπαίδευσαν δύο διαφορετικούς αλγόριθμους επισημείωσης εικόνας (TagProp και 2PKNN) με χαρακτηριστικά που προέρχονται από δύο αρχιτεκτονικές ConvNet (AlexNet και VGG-16). Τα πειραματικά αποτελέσματα αποκάλυψαν ότι καλύτερη, ή τουλάχιστον παρόμοια, απόδοση επισημείωσης επιτεύχθηκε με τη χρήση χαρακτηριστικών που προέρχονται από ένα βαθύ συνελικτικό νευρωνικό δίκτυο παρά με τη χρήση περισσότερων χειροποίητων χαρακτηριστικών. Επιπλέον, η μελέτη τους απέδειξε την ιδέα ότι συμπληρωματικές πληροφορίες τόσο στα βαθιά όσο και στα χειροποίητα χαρακτηριστικά θα μπορούσαν να χρησιμοποιηθούν από κοινού για τη βελτίωση της απόδοσης.

Το μοντέλο CCA-KNN (Murthy et al., 2015) βασίζεται στο πλαίσιο Canonical Correlation Analysis (CCA), το οποίο βοηθά στη μοντελοποίηση τόσο των οπτικών χαρακτηριστικών (χαρακτηριστικά ConvNet) όσο και των χαρακτηριστικών κειμένου (διανύσματα ενσωμάτωσης λέξεων) των δεδομένων. Αποδείχθηκε ότι τα χαρακτηριστικά του ConvNet πλεονεκτούν έναντι 15 χειροποίητων χαρακτηριστικών σε υπάρχοντα μοντέλα,

συμπεριλαμβανομένων των JEC και 2PKNN. Επιπλέον, η μελέτη τους έδειξε ότι τα διανύσματα ενσωμάτωσης λέξεων αποδίδουν καλύτερα από τα δυαδικά διανύσματα ως αναπαράσταση των ετικετών που σχετίζονται με μια εικόνα.

Δεδομένου ότι η ποιότητα της αρχικής επισημείωσης του συνόλου δεδομένων έχει μεγάλη επίδραση στην απόδοση του AIA, το μοντέλο Multitask Voting automatic image annotation CNN (MVAIACNN) των Wang et al. (2017) υιοθετεί τη μέθοδο πολλαπλής ψηφοφορίας μέσω ενός μηχανισμού μάθησης πολλαπλών στόχων για την επιλογή των συνόλων δεδομένων εκπαίδευσης και δοκιμής. Συνδυάζοντας τη μέθοδο πολλαπλής μάθησης με το Bayesian μοντέλο πιθανότητας, η μέθοδος MV επιτυγχάνει τη σωστή ετικέτα. Τέλος, προτάθηκε το μοντέλο AIACNN, το οποίο περιέχει πέντε συνελικτικά στρώματα για την εξαγωγή χαρακτηριστικών ιεραρχικά και τέσσερα στρώματα συγκέντρωσης (pooling layers), ακολουθούμενα από δύο πλήρως συνδεδεμένα στρώματα και το επίπεδο εξόδου softmax που υποδεικνύει την ταυτότητα των κλάσεων. Το MVAIACNN έχει ρηχά στρώματα και θεωρεί κάθε κλάση απευθείας ως ετικέτα, χρησιμοποιώντας τις ακατέργαστες εικόνες ως εισόδους για επισημείωση εικόνας μεγάλης κλίμακας. Σε κάποιο βαθμό, λιγότερα στρώματα μειώνουν τα ελαττώματα στην απόδοση που προκαλούνται από περισσότερα στρώματα.

Οι Johnson et al. (2015) πρότειναν ένα μοντέλο για τη δημιουργία «γειτονιών» σχετικών εικόνων με παρόμοια μεταδεδομένα κοινωνικού δικτύου. Τα μεταδεδομένα που μεταφέρονται από τις περισσότερες εικόνες στον ιστό, όπως οι δημιουργούμενες από τον χρήστη ετικέτες και οι ομάδες που έχουν επιλεγεί από την κοινότητα, μπορούν να είναι εξαιρετικά κατατοπιστικές όσον αφορά τα σημασιολογικά περιεχόμενα των εικόνων. Οι τύποι μεταδεδομένων εικόνων που εξετάζονται από τους Johnson et al. (2015) περιλαμβάνουν ετικέτες χρηστών, σύνολα φωτογραφιών και ομάδες εικόνων. Τα φωτογραφικά σύνολα είναι εικόνες που συλλέγονται συνήθως από τον ίδιο χρήστη. Για παράδειγμα, φωτογραφίες από μια αθλητική συνάντηση φορτώνονται από τον ίδιο χρήστη του κοινωνικού δικτύου. Οι ομάδες εικόνων είναι ένα σύνολο εικόνων που ανήκουν στην ίδια περίσταση, έννοια και γεγονός στην τοποθεσία του κοινωνικού δικτύου, π.χ. ένα σύνολο εικόνων που περιέχουν όλες πορτοκάλια στο κοινωνικό δίκτυο.



Εικόνα 25. Σχέδιο του μοντέλου των Johnson et al. (2015).

Για να κάνουν προβλέψεις για μια εικόνα, δοκιμάζουν αρκετούς από τους πλησιέστερους γείτονές της για να σχηματίσουν μια «γειτονιά» και χρησιμοποιούν ConvNet για να εξαγάγουν οπτικά χαρακτηριστικά. Υπολογίζουν τις αναπαραστάσεις κρυφής κατάστασης για την εικόνα και τους γείτονές της, στη συνέχεια εργάζονται πάνω στη συνένωση αυτών των δύο αναπαραστάσεων για να υπολογίσουν τις βαθμολογίες της κλάσης.

Το μοντέλο αυτόματου κωδικοποιητή πολλαπλών προβολών (multi-view stacked auto-encoder – MVSAE) των Yang et al. (2015), δημιουργεί ένα νέο πλαίσιο SAE για επισημείωση εικόνας. Τα χαρακτηριστικά εικόνας χρησιμοποιούνται συνήθως ως είσοδος του μοντέλου και οι λέξεις-κλειδιά χρησιμοποιούνται ως αντικείμενα του μοντέλου στα πιο πολλά μοντέλα AIA βασισμένα στα βαθιά νευρωνικά δίκτυα. Επιπλέον, αρκετά κρυφά επίπεδα χρησιμοποιούνται για τη μοντελοποίηση της περίπλοκης σχέσης μεταξύ οπτικών χαρακτηριστικών και ετικετών. Δεδομένου ότι η απόδοση του βαθιού νευρωνικού δικτύου εξαρτάται σε μεγάλο βαθμό από τις αρχικές παραμέτρους, το μοντέλο MVSAE υιοθετεί τις προ-εκπαιδευμένες παραμέτρους για τη βελτιστοποίηση του μοντέλου. Συγκεκριμένα, αρχικά, το οπτικό χαρακτηριστικό I ως είσοδος του μοντέλου x χρησιμοποιείται για την εκπαίδευση του SAE για τη δημιουργία της αρχικής κατανομής πιθανότητας λέξεων-κλειδιών $D1$. Στη συνέχεια, οι I και $D1$ χρησιμοποιούνται ως νέες είσοδοι μοντέλου x για την επανεκπαίδευση του μοντέλου SAE για τη δημιουργία της τελικής κατανομή πιθανότητας της λέξης-κλειδιού $D2$. Τέλος, οι λέξεις-κλειδιά της εικόνας $\hat{T}$ λαμβάνονται από την $D2$.

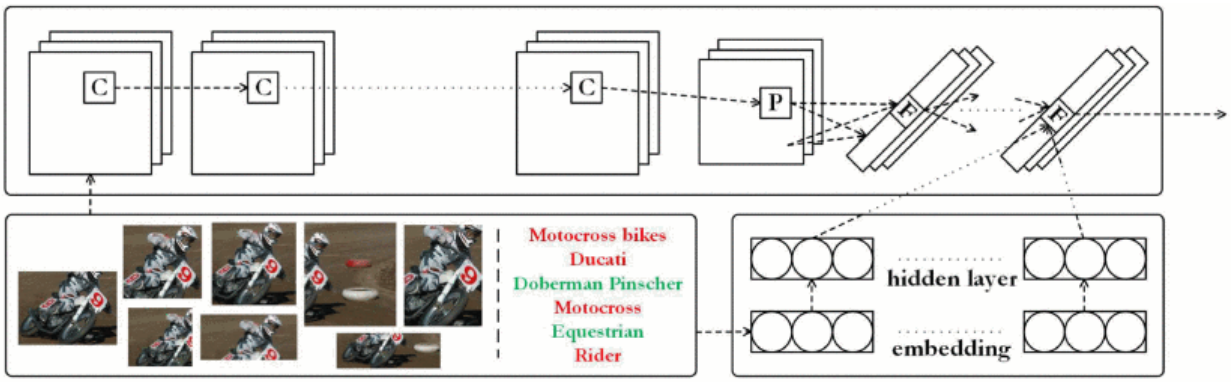
4.4.4.2. Παράπλευρες πληροφορίες

Το πλαίσιο CNN-RNN των Wang et al. (2016) χρησιμοποιεί αναδρομικά νευρωνικά δίκτυα (Recurrent Neural Network – RNN) για να συλλάβει σχέσεις ετικετών υψηλού επιπέδου σε ένα μέτριο επίπεδο υπολογιστικής πολυπλοκότητας. Σε αυτό το πλαίσιο, τα CNN και RNN χρησιμοποιούνται από κοινού για την εξαγωγή της αναπαράστασης εικόνας και της

συσχέτισης μεταξύ των γειτονικών ετικετών, βάσει των οποίων υπολογίζονται οι τελικές έξοδοι, όπως η πιθανότητα της ετικέτας. Είναι σημαντικό να ταξινομηθούν οι ετικέτες για την εκπαίδευση μοντέλων πολλαπλών ετικετών CNN-RNN. Στο πλαίσιο του CNN-RNN, οι θέσεις (orders) των ετικετών στην κατάταξη καθορίζονται σύμφωνα με τις συχνότητες εμφάνισής τους στα δεδομένα εκπαίδευσης.

Το μοντέλο RIA των Jin and Nakayama (2016), χρησιμοποιεί επίσης το πλαίσιο CNN-RNN για επισημείωση εικόνας. Εμπνευσμένο από την πρόσφατη επιτυχία του RNN στην περιγραφή εικόνας με λεζάντες (Vinyals et al., 2015), το RIA χρησιμοποιεί το CNN για να εξαγάγει οπτικά χαρακτηριστικά εικόνας και το RNN για να δημιουργήσει την ακολουθία ετικετών από τα οπτικά χαρακτηριστικά ένα προς ένα. Τα πλεονεκτήματα του RNN στην AIA εντοπίζονται σε δύο πτυχές. Από τη μία πλευρά, το RNN μπορεί να παράγει αποτελέσματα με διαφορετικό μήκος (λέξεις-ετικέτες διαφορετικού μήκους). Από την άλλη πλευρά, το RNN είναι σε θέση να αναφερθεί σε προηγούμενες εισόδους στην πρόβλεψη της εξόδου στο τρέχον χρονικά βήμα ροής. Η περιγραφή των εικόνων με λεζάντες (Vinyals et al., 2015) στοχεύει στη δημιουργία προτάσεων σε φυσική γλώσσα για την εκπαίδευση του μοντέλου RNN. Σημειώνεται ότι το μοντέλο RNN χρησιμοποιεί τον κανόνα frequent-first (πρώτα το συχνότερο) παρά τον κανόνα rare-first (πρώτα το σπανιότερο) που υιοθετείται από το μοντέλο RIA.

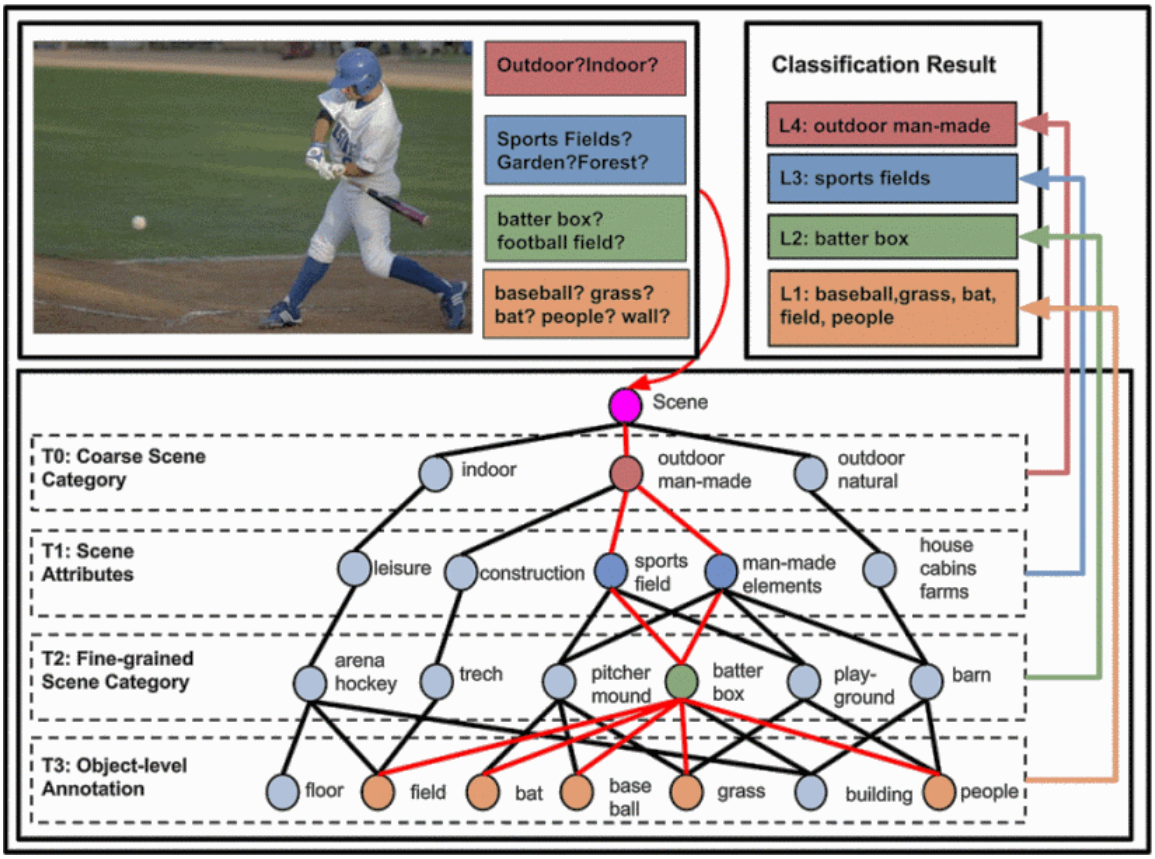
Το μοντέλο – βαθιάς μάθησης πολλαπλών στιγμιοτύπων – Deep Multiple Instance Learning (DMIL) (Wu et al., 2015) παρουσιάζει ένα πλαίσιο για τη μάθηση των αντιστοιχιών μεταξύ των περιοχών εικόνας και των λέξεων-κλειδιών. Στο DMIL, δύο ομάδες στιγμιοτύπων, προτάσεις αντικειμένων και λέξη-κλειδί, μαθαίνονται ταυτόχρονα από ένα κοινό βαθύ πλαίσιο μάθησης πολλαπλών στιγμιοτύπων. Συγκεκριμένα, το DMIL χρησιμοποιεί ένα CNN που περιέχει πέντε συνελικτικά στρώματα, ένα στρώμα συγκέντρωσης και τρία πλήρως συνδεδεμένα στρώματα για μάθηση οπτικής αναπαράστασης. Στη συνέχεια, χρησιμοποιεί ένα άλλο πλαίσιο βαθέως νευρωνικού δικτύου που περιέχει ένα στρώμα εισόδου, ένα κρυφό στρώμα και ένα στρώμα εξόδου με softmax για μάθηση πολλαπλών στιγμιοτύπων. Τέλος, συνδυάζει τόσο την εικόνα όσο και τις εξόδους κειμένου στο πλήρως συνδεδεμένο στρώμα.



Εικόνα 26. Απεικόνιση του μοντέλου DMIL για την από κοινού μάθηση περιοχών εικόνας και λέξεων-κλειδιών. Το P είναι ένα στρώμα συγκέντρωσης, το C ένα στρώμα συνέλιξης και το F για ένα πλήρως συνδεδεμένο στρώμα (Wu et al., 2015).

Το δομημένο νευρωνικό δίκτυο συμπερασμάτων (Structured Inference Neural Network - SINN) προτάθηκε από τους Hu et al. (2016) για προβλέψεις πολυεπίπεδων ετικετών. Η βασική ιδέα του SINN είναι ότι σε μια εικόνα με διάφορα αντικείμενα και άφθονες ιδιότητες είναι δυνατά διάφορα επίπεδα οπτικής κατηγοριοποίησης. Ας πάρουμε την εικόνα 27 ως παράδειγμα. Θα μπορούσαν να διαμορφωθούν διάφορα επίπεδα ερμηνείας για μια τέτοια εικόνα. Αυτή η εικόνα μίας σκηνής του μπέιζμπολ θα μπορούσε να περιγραφεί ως «εξωτερική εικόνα» σε ένα γενικό – αδρό – επίπεδο ή με μια πιο συγκεκριμένη έννοια όπως «αθλητικό γήπεδο» ή με μια ακόμα πιο λεπτομερή ετικέτα όπως «θέση του ροπαλοφόρου» και ετικέτες για τα αντικείμενα όπως γρασίδι, ρόπαλο, παίκτης.

Η προσέγγιση των Hu et al. (2016) χρησιμοποιεί ένα καινοτόμο νευρωνικό δίκτυο πρόβλεψης ετικετών, το οποίο καταγράφει τόσο τη σημασιολογία μεταξύ ετικετών όσο και μεταξύ των επιπέδων. Αρχικά, τα χαρακτηριστικά CNN εξάγονται ως οπτικές ενεργοποιήσεις σε κάθε στρώμα εννοιών. Τα στρώματα των εννοιών στοιβάζονται από λεπτά (λεπτομερή) έως πιο αδρά (γενικά) επίπεδα. Δεύτερον, οι σχέσεις μεταξύ ετικετών μεταξύ διαδοχικών στρωμάτων παράγονται ως ένας διαστρωματικός γράφος, όπου κάθε στρώμα εννοιών αντιπροσωπεύει ένα χρονικό βήμα του RNN. Οι συσχετίσεις μεταξύ των επιπέδων και οι σχέσεις εντός των επιπέδων λαμβάνονται σύμφωνα με τα χρονικά βήματα.

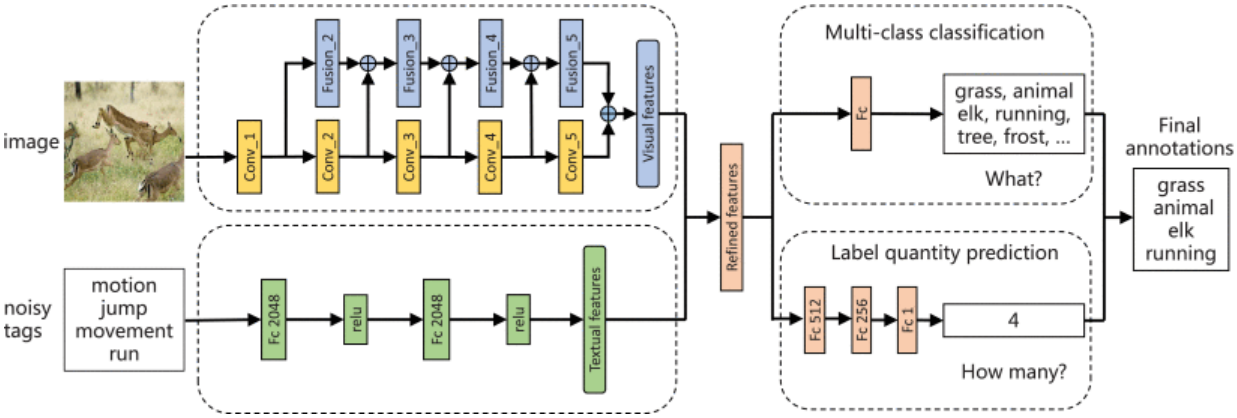


Εικόνα 27. Το μοντέλο SINN των Hu et al. (2016).

Αυτό το παράδειγμα εικόνας έχει οπτικές έννοιες σε διάφορα επίπεδα, από τον αθλητικό τομέα (σε υψηλό επίπεδο) έως το μπέιζμπολ και το άτομο (σε χαμηλότερο επίπεδο). Το μοντέλο SINN αξιοποιεί τις σχέσεις ετικέτας και από κοινού προβλέπει στρωματοποιημένες οπτικές ετικετών από μια εικόνα χρησιμοποιώντας ένα δομημένο νευρωνικό δίκτυο συμπερασμάτων. Στο γράφημα, οι έγχρωμοι κόμβοι αντιστοιχούν στις ετικέτες που σχετίζονται με την εικόνα και οι κόκκινες ακμές κωδικοποιούν τις σχέσεις ετικέτας (Hu et al., 2016).

Οι Niu et.al (2017) επικεντρώθηκαν κυρίως σε δύο ζητήματα για την επισημείωση εικόνας μεγάλης κλίμακας. Το πρώτο πώς να μάθουν μια πλούσια αναπαράσταση χαρακτηριστικών κατάλληλη για την πρόβλεψη ενός ποικίλου συνόλου οπτικών εννοιών που κυμαίνονται από το αντικείμενο, τη σκηνή έως πιο αφηρημένες έννοιες. Το δεύτερο ζήτημα είναι πως θα επισημειωθεί μία εικόνα με τον βέλτιστο αριθμό ετικετών κλάσης. Για να αντιμετωπιστεί το πρώτο ζήτημα, προτείνουν ένα νέο βαθύ μοντέλο πολλαπλών βαθμίδων για την εξαγωγή πλούσιων και διακριτικών χαρακτηριστικών ικανών να αντιπροσωπεύουν ένα ευρύ φάσμα οπτικών εννοιών. Συγκεκριμένα, προτείνεται μια νέα αρχιτεκτονική βαθέως νευρωνικού δικτύου δύο κλάδων, η οποία περιλαμβάνει ένα πολύ βαθύ κύριο κλάδο δικτύου και ένα συνοδευτικό τμήμα κλάδου δικτύου σύντηξης σχεδιασμένο για τη σύντηξη των πολλαπλών χαρακτηριστικών που υπολογίζονται από τον κύριο κλάδο. Το βαθύ μοντέλο είναι επίσης πολυτροπικό καθώς λαμβάνει θορυβώδεις ετικέτες που παρέχονται από το χρήστη ως εισόδους για να συμπληρώσει τις εικόνες εισόδου (του συνόλου δεδομένων εκπαίδευσης)

στην είσοδο του μοντέλου. Για την αντιμετώπιση του δεύτερου θέματος, εισάγουν μια βοηθητική εργασία πρόβλεψης ποσότητας των ετικετών στην κύρια εργασία πρόβλεψης των ετικετών για να εκτιμήσουν τον βέλτιστο αριθμό ετικετών για μια δεδομένη εικόνα. Σε αντίθεση με τις μεθόδους που βασίζονται στο RNN, η μέθοδος επισημείωσης των Niu et.al (2017) βλέπει την πρόβλεψη του αριθμού των ετικετών ως πρόβλημα παλινδρόμησης. Έχει αποδειχθεί ότι η πρόβλεψη της ποσότητας έχει μεγάλη επίδραση στην αποτελεσματικότητα της επισημείωσης εικόνας.



Εικόνα 28. Το διάγραμμα ροής του μοντέλου Niu et.al (2017) για επισημείωση εικόνας μεγάλης κλίμακας.

Το μοντέλο περιλαμβάνει τέσσερα συστατικά: μάθηση οπτικών χαρακτηριστικών, μάθηση λεκτικών χαρακτηριστικών, ταξινόμηση πολλαπλών κατηγοριών και πρόβλεψη ποσότητας ετικετών.

Κεφάλαιο 5
Μέθοδοι αξιολόγησης επισημείωσης εικόνας

Η σύγκριση και η ανάλυση της επίδοσης διαφορετικών μεθόδων προϋποθέτει την ύπαρξη καθιερωμένων μετρικών αξιολόγησης και βάσεων δεδομένων. Στο παρόν κεφάλαιο συνοψίζουμε βασικές μετρικές καθώς και δημόσια διαθέσιμες βάσεις δεδομένων αναφοράς για την αξιολόγηση μεθόδων αυτόματης επισημείωσης εικόνας.

5.1. Μέτρα αξιολόγησης συστημάτων αυτόματης επισημείωσης εικόνας

Οι λέξεις-κλειδιά που αντιστοιχίζονται σε μια εικόνα, αντιπροσωπεύουν σημασιολογικά τα περιεχόμενα της εικόνας. Όταν μια εικόνα έχει αντιστοιχιστεί σε μία μόνο λέξη-κλειδί, μπορεί να θεωρηθεί ως επισημείωση μονής ετικέτας ή απλά μια δυαδική ταξινόμηση όπου ένας ταξινομητής διαπιστώνει μόνο την παρουσία ή την απουσία λέξης-κλειδιού. Μόνο μία ετικέτα δεν μπορεί να αντιπροσωπεύει το πραγματικό περιεχόμενο της εικόνας. Ως εκ τούτου, οι μέθοδοι επισημείωσης εικόνας συνήθως εκχωρούν πολλές λέξεις-κλειδιά για να υποδείξουν την παρουσία πολλών αντικειμένων στην εικόνα. Αυτό ονομάζεται σύστημα επισημείωσης πολλαπλών ετικετών (multi-label) ή μπορεί επίσης να θεωρηθεί ως σύστημα ταξινόμησης πολλαπλών κλάσεων (multi-class). Για να διαπιστωθεί η ακρίβεια του συστήματος επισημείωσης, υπάρχουν δύο ευρείες κατηγορίες μέτρων αξιολόγησης: (i) ποιοτικά (qualitative) μέτρα και (ii) ποσοτικά (quantitative) μέτρα (Bhagat and Choudhary, 2018).

Τα ποιοτικά μέτρα χρησιμοποιούνται για την αξιολόγηση του συστήματος από τον ανθρώπινο παράγοντα. Ο ανθρώπινος παράγοντας καλείται να αξιολογήσει την απόδοση του συστήματος έτσι ώστε να μπορεί να σχηματιστεί μια πιο ολοκληρωμένη εικόνα του συστήματος επισημείωσης. Η ποσοτική αξιολόγηση πραγματοποιείται σε επίπεδο συστήματος όπου ένα σύνολο δεδομένων επαλήθευσης (ground truth) χρησιμοποιείται για να διαπιστωθεί η ακρίβεια του συστήματος. Για ένα σύστημα επισημείωσης μονής ετικέτας,

η ακρίβεια (accuracy) του συστήματος μπορεί να θεωρηθεί ως το βασικό κριτήριο αξιολόγησης των επιδόσεών του. Στην περίπτωση αυτή, η ακρίβεια αναφέρεται απλώς στο συνολικό ποσοστό σωστά ταξινομημένων εικόνων δοκιμής σε σχέση με τον συνολικό αριθμό εικόνων δοκιμής (Bhagat and Choudhary, 2018).

5.1.1. Μέτρα αξιολόγησης επισημείωσης μονής ετικέτας

Υπάρχουν αρκετές μετρικές για την αξιολόγηση της απόδοσης των διάφορων μεθόδων αυτόματης επισημείωσης εικόνας. Μεταξύ αυτών η ακρίβεια (precision), η ανάκληση (recall) και η βαθμολογία F1 (F1 score), είναι από τα πιο συνηθισμένα. Έστω ότι θέλουμε να αξιολογήσουμε τα αποτελέσματα του ταξινομητή σε ένα σύστημα επισημείωσης μονής ετικέτας για μία ετικέτα κλάσης. Ο ταξινομητής τροφοδοτείται με ένα σύνολο δεδομένων δοκιμής από τα οποία κάποιες εικόνες είναι επισημειωμένες με τη ground truth ετικέτα και θεωρούνται θετικά δείγματα ενώ άλλες δεν είναι επισημειωμένες με την ετικέτα και θεωρούνται αρνητικά δείγματα.

Τα αποτελέσματα που λαμβάνονται από τον ταξινομητή μπορούν να πάρουν τη μορφή ενός 2×2 πίνακα (πίνακας 7) που ονομάζεται confusion matrix. Οι όροι “true positive”, “true negative”, “false positive”, και “false negative” συγκρίνουν τα αποτελέσματα του ταξινομητή με τα ground truth δεδομένα. Οι όροι “positive” και “negative” αναφέρονται στην πρόβλεψη του ταξινομητή και οι όροι “true” και “false” αναφέρονται στο εάν αυτή η πρόβλεψη ανταποκρίνεται στα ground truth δεδομένα.

		Actual	
		positive	*negative*
Predicted / Classified	*positive*	**True Positive**	**False Positive**
	negative	**False Negative**	**True Negative**

Πίνακας 7. Confusion matrix για την αξιολόγηση των αποτελεσμάτων του ταξινομητή.

Η ακρίβεια (precision), η ανάκληση (recall) και η βαθμολογία F1 (F1 score) υπολογίζονται με τη βοήθεια του confusion matrix από τις εξισώσεις 5.1-5.3 (Bhagat and Choudhary, 2018).

$$\text{Precision: } P = \frac{TP}{TP + FP} \qquad (5.1)$$

$$Recall\text{: } R = \frac{TP}{TP + FN} \qquad (5.2)$$

$$F1\ score = 2 * \frac{P * R}{P + R} \qquad (5.3)$$

όπου:

- *TP (True Positive)*: Τόσο τα πραγματικά όσο και τα ληφθέντα αποτελέσματα είναι τα ίδια και υποδεικνύουν την παρουσία ετικέτας (υποδηλώνει τον αριθμό των εικόνων δοκιμής που επισημειώνονται σωστά με την ground truth ετικέτα).
- *TN (True Negative)* : Τόσο τα πραγματικά όσο και τα ληφθέντα αποτελέσματα είναι τα ίδια και υποδεικνύουν την απουσία ετικέτας (υποδηλώνει τον αριθμό των εικόνων δοκιμής που σωστά δεν επισημειώνονται με την ground truth ετικέτα).
- *FN (False Negative)*: Αν και η ετικέτα υπάρχει στα επισημειωμένα δεδομένα επαλήθευσης, το αποτέλεσμα που προκύπτει, αποκαλύπτει την απουσία ετικέτας (υποδηλώνει τον αριθμό των εικόνων δοκιμής που λανθασμένα δεν επισημειώνονται με την ground truth ετικέτα).
- *FP (False Positive)*: Τα αποτελέσματα που προέκυψαν, δείχνουν την ύπαρξη ετικέτας παρόλο που η ετικέτα απουσιάζει από το σύνολο επαλήθευσης (υποδηλώνει τον αριθμό των εικόνων δοκιμής που λανθασμένα επισημειώνονται με την ετικέτα).
- $TP + FP$: Ο αριθμός των εικόνων στο σύνολο δεδομένων δοκιμής που επισημειώνονται με ετικέτα.
- $TP + FN$: Ο αριθμός των εικόνων που επισημειώνονται με ετικέτα σύμφωνα τα ground truth δεδομένα.

Η ανάκληση μετρά τη δυνατότητα ανάκτησης των σχετικών πληροφοριών, ενώ η ακρίβεια μετρά την ικανότητα απόρριψης της μη συσχετισμένης πληροφορίας. Η ανάκληση και η ακρίβεια συνδυάζονται συνήθως για την αξιολόγηση της απόδοσης των μοντέλων AIA. Ωστόσο, είναι δύσκολο να αξιολογηθούν τα μοντέλα AIA μόνο με βάση την ανάκληση και την ακρίβεια, καθώς οι δύο μετρικές έρχονται σε σύγκρουση μεταξύ τους (Cheng et al., 2018). Ας σημειωθεί ότι οι δοκιμαστικές εικόνες συνήθως επισημειώνονται με k ετικέτες (συνήθως 5) με τις μεθόδους AIA, ακόμα και αν οι εικόνες είναι επισημειωμένες με λιγότερες ή περισσότερες ετικέτες στα ground-truth δεδομένα του συνόλου εκπαίδευσης. Επομένως, μπορεί να δώσουν προκατειλημμένες (biased) τιμές ανάκλησης και ακρίβειας ακόμη και αν ένα μοντέλο προβλέπει όλες τις ground-truth ετικέτες (Cheng et al., 2018).

Δεδομένου ότι είτε η ανάκληση είτε η ακρίβεια δεν επαρκούν για την πλήρη εκτίμηση της απόδοσης των μοντέλων AIA, έχουν ενσωματωθεί σε ένα μόνο δείκτη αξιολόγησης τη βαθμολογία F_1. Επιπλέον, η βαθμολογία F_1 μπορεί να χρησιμοποιηθεί για τη μέτρηση της ευρωστίας των μεθόδων AIA. Όσο μεγαλύτερη είναι η βαθμολογία F_1, τόσο πιο ισχυρό θα είναι το μοντέλο (Cheng et al., 2018).

Ένα άλλο μέτρο που χρησιμοποιείται συχνά για την αξιολόγηση της επίδοσης διάφορων μεθόδων AIA, είναι ο δείκτης N_+. Ο δείκτης N_+ χρησιμοποιείται για να δηλώσει τον αριθμό των λέξεων-κλειδιών που έχουν αντιστοιχιστεί σωστά σε τουλάχιστον μία εικόνα δοκιμής. Ο δείκτης σημαίνει επίσης τον αριθμό των λέξεων-κλειδιών με θετική ανάκληση. Μια υψηλή τιμή N_+ αποτελεί ένδειξη της καλής απόδοσης της μεθόδου AIA (Cheng et al., 2018).

5.1.2. Μέτρα αξιολόγησης επισημείωσης πολλαπλών ετικετών

Τα συστήματα μάθησης πολλαπλών ετικετών απαιτούν πιο πολύπλοκα κριτήρια αξιολόγησης των επιδόσεων του συστήματος από τα παραδοσιακά συστήματα μονής ετικέτας. Επίσης, πολλά συστήματα επισημείωσης βασίζονται σε κατάταξη όπου οι ετικέτες κατατάσσονται με βάση κάποιο συντελεστή εμπιστοσύνης που μπορεί να απαιτεί διαφορετικό σύνολο κριτηρίων αξιολόγησης (Bhagat and Choudhary, 2018).

Οι Schapire and Singer (2010) προτείνουν τρία μέτρα αξιολόγησης για επισημείωση πολλαπλών ετικετών. Το πρώτο, το «μοναδικό λάθος» (“one-error”) είναι μια εκδοχή του υπολογισμού της ακρίβειας του συστήματος επισημείωσης μιας ετικέτας για ένα σύστημα πολλαπλών ετικετών. Μετράει πόσες φορές οι ετικέτες με την υψηλότερη βαθμολογία δεν περιλαμβάνονται στο σύνολο πιθανών ετικετών, δηλαδή, μετρά τα ψευδώς θετικά (false-positive). Το one-error συνδέεται άμεσα με το λάθος εκπαίδευσης (training error). Τα άλλα δύο μέτρα αξιολόγησης, η κάλυψη (coverage) και η μέση ακρίβεια (average precision) βασίζονται σε μέτρα που χρησιμοποιούνται στην ανάκτηση πληροφορίας και χρησιμοποιούνται σε αλγορίθμους ταξινόμησης από την άποψη της κατάταξης των ετικετών τους. Ενώ το one-error αξιολογεί την απόδοση ενός συστήματος για τις κορυφαίες ετικέτες, η κάλυψη (coverage) αξιολογεί την απόδοση ενός συστήματος για όλες τις πιθανές ετικέτες. Η κάλυψη μετρά πόσο πολύ χρειαζόμαστε, κατά μέσο όρο, να κατέβουμε στη λίστα των ετικετών για να καλύψουμε όλες τις πιθανές ετικέτες που έχουν δοθεί σε μία εικόνα. Η κάλυψη σχετίζεται χαλαρά με την ακρίβεια στο επίπεδο της τέλειας ανάκλησης.

Τα παραπάνω μέτρα δεν είναι πλήρη για τα προβλήματα ταξινόμησης πολλαπλών ετικετών. Έτσι, το σύστημα μπορεί να έχει καλή (χαμηλή) κάλυψη αλλά να υποφέρει από υψηλά ποσοστά one-error και αντίστροφα. Η μέση ακρίβεια (average precision), ένας τρόπος μέτρησης της απόδοσης ενός συστήματος ανάκτησης εικόνων (Image Retrieval - IR), μπορεί να χρησιμοποιηθεί για την κατάταξη των ετικετών, η οποία υπολογίζει το μέσο κλάσμα των ετικετών που κατατάσσονται πάνω από σε μια «συγκεκριμένη» ταξινομημένη ετικέτα εξόδου. Εδώ, ο όρος «συγκεκριμένη» αναφέρεται σε μια ορισμένη τιμή η οποία χρησιμοποιείται ως τιμή κατωφλίου και όλες οι ετικέτες που κατατάσσονται πάνω από το όριο, θεωρούνται ως η έξοδος του συστήματος. Ας σημειωθεί, ωστόσο, ότι η μέση ακρίβεια χρησιμοποιείται συνήθως σε συστήματα IR για την αξιολόγηση της απόδοσης κατάταξης εικόνων για την ανάκτηση ερωτημάτων. Η μέθοδος αξιολόγησης της ανάκτησης εικόνας (IR) μετρά αν οι ανακτημένες εικόνες είναι σχετικές με την εικόνα της ερώτησης ή όχι, και πόσες ανακτημένες εικόνες σχετίζονται με το ερώτημα του χρήστη. Η μέθοδος αξιολόγησης της επισημείωσης εικόνας μετρά την πρόβλεψη ακριβών (σωστών) ετικετών. Ελέγχει πόσες ετικέτες έχουν προβλεφθεί με ακρίβεια, πόσες ετικέτες λείπουν από το αποτέλεσμα κ.λπ. για μια εικόνα ερωτήματος (Schapire and Singer, 2000).

Οι Zhang and Zhou (2010) χρησιμοποιούν δύο σύνολα κριτηρίων για την αξιολόγηση της απόδοσης της πρόβλεψης του συνόλου των ετικετών (label set prediction), καθώς και των επιδόσεων της κατάταξης των ετικετών (label ranking). Η πρώτη ομάδα κριτηρίων αξιολόγησης αφορά την απόδοση στην πρόβλεψη του συνόλου των ετικετών για κάθε στιγμιότυπο, με βάση τη συνάρτηση πρόβλεψης ετικέτας h: $X \rightarrow Y$ κάθε αλγορίθμου. Έστω m ο αριθμός των δειγμάτων του συνόλου εκπαίδευσης. Το σύνολο δεδομένων εισόδου έχει τη μορφή $\langle (X_1, Y_1), (X_2, Y_2), \ldots\ldots, (X_m, Y_m) \rangle$, όπου τα X_i και Y_i είναι τα σύνολα εισόδου και εξόδου αντίστοιχα για το $i - οστό$ στοιχείο δεδομένων εκπαίδευσης, το y_i^j είναι ένα στοιχείο του συνόλου Y_i και $h(X_i)$ είναι το δυαδικό διάνυσμα ετικέτας που προβλέπεται από έναν ταξινομητή πολλαπλών ετικετών για το στιγμιότυπο X_i (αντιπροσωπεύει το σύνολο των κορυφαίων k προβλεπόμενων ετικετών για το X_i) (Zhang and Zhou, 2010).

Η απώλεια Hamming (Hamming Loss - HL) μετράει την εσφαλμένη ταξινόμηση ενός ζευγαριού ετικέτας-εικόνας. Δηλαδή, παίρνει το μέσο όρο κάθε φορά που οι πραγματικές και προβλεπόμενες ετικέτες είναι διαφορετικές.

$$HL(h) = \frac{1}{m}\sum_{i=1}^{m}\frac{\|h(X_i) \oplus Y_i\|_1}{N} \qquad (5.4)$$

όπου $\oplus$ είναι ο τελεστής XOR και $\|.\|_1$ η $l1$-νόρμα.

Όσο μικρότερη είναι η τιμή, τόσο καλύτερη είναι η απόδοση. Αυτό είναι ένα από τα σημαντικότερα κριτήρια πολλαπλών ετικετών και έχει χρησιμοποιηθεί σε πολλές μελέτες (Zhang and Zhou, 2010).

Το "Macro-F1" υπολογίζει το μέσο όρο του μέτρου *F1* στις προβλέψεις διαφορετικών επισημειώσεων.

$$Macro\ F1(h) = \frac{1}{N}\sum_{i=1}^{N}\frac{2 * \sum_{j=1}^{m} h^i(X_j)y_j^i}{\sum_{j=1}^{m} y_j^i + \sum_{j=1}^{m} h^i(X_j)} \qquad (5.5)$$

όπου y^i το i-οστό στοιχείο του y και $h^i(x)$ το το i-οστό στοιχείο του $h(x)$.

Όσο μεγαλύτερη είναι η τιμή, τόσο καλύτερη είναι η απόδοση του συστήματος.

Το "Micro-F1" υπολογίζει το μέτρο *F1* στις προβλέψεις των διαφόρων ετικετών στο σύνολό τους.

$$Micro\ F1(h) = \frac{2 * \sum_{i=1}^{m}\|h(X_i) \cap Y_i\|_1}{\sum_{i=1}^{m}\|Y_i\|_1 + \sum_{i=1}^{m}\|h(X_i)\|_1} \qquad (5.6)$$

Όσο μεγαλύτερη είναι η τιμή, τόσο καλύτερη είναι η απόδοση του συστήματος.

Μία χαμηλή τιμή απώλειας Hamming και μια μεγάλη τιμή για τα micro-F1 και macro-F1 είναι απόδειξη μιας εξαιρετικής απόδοσης του συστήματος.

Η δεύτερη ομάδα κριτηρίων αξιολόγησης αφορά στην απόδοση κατάταξης της ετικέτας για κάθε στιγμιότυπο, με βάση την πραγματική συνάρτηση εξόδου του κάθε αλγορίθμου. Όπως σημειώθηκε προηγουμένως, υποθέτουμε ότι ένα σύστημα πολλαπλών ετικετών παράγει μία κατάταξη των πιθανών ετικετών για ένα συγκεκριμένο στιγμιότυπο. Δηλαδή, η έξοδος του συστήματος είναι μια συνάρτηση $f: X \times Y \rightarrow \mathcal{R}$ που ταξινομεί τις ετικέτες σύμφωνα με την $f(x,.)$ έτσι ώστε η ετικέτα l_1 να θεωρείται ότι είναι ψηλότερα στην κατάταξη από την l_2 εάν $f(x,l_1) > f(x,l_2)$. Η συνάρτηση f μπορεί να μετασχηματιστεί σε μια συνάρτηση κατάταξης $rank_f(x,.)$, η οποία αντιστοιχίζει τις εξόδους της $f(x,.)$ στο $\{1,2,3,\dots,M\}$ έτσι ώστε αν $f(x,l_1) > f(x,l_2)$ τότε $rank_f(x,l_1) < rank_f(x,l_2)$ (Schapire and Singer, 2000, Zhang and Zhou, 2010).

Η «απώλεια κατάταξης» (Ranking loss) χρησιμοποιείται για την αξιολόγηση της απόδοσης της κατάταξης της ετικέτας και υπολογίζει το μέσο κλάσμα των ζευγών ετικετών που δεν έχουν καταταχθεί σωστά.

$$RL(f) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{|Y_i||\overline{Y}_i|}|\{y_t, y_s \in Y_i \times \overline{Y}_i|\ (X_i, y_t) \leq h(X_i, y_s)\}| \qquad (5.7)$$

όπου, N είναι ο αριθμός ετικετών για μια εικόνα και $\|.\|_1$ είναι η νόρμα-l_1 ($\|x\|_1 = \sum_i |x_i| = |x_1| + |x_2| + \cdots + |x_i|$) και $\oplus$ είναι ο λογικός τελεστής XOR (αποκλειστική διάζευξη). Το συμπλήρωμα του Y_i συμβολίζεται με $\overline{Y}_i$.

Όσο μικρότερη είναι η τιμή, τόσο καλύτερη είναι η απόδοση του συστήματος.

Η «μέση ακρίβεια» (average precision) υπολογίζει το μέσο ποσοστό των ετικετών που κατατάσσονται πάνω από μια συγκεκριμένη ετικέτα $l' \in Y_i$ οι οποίες βρίσκονται στο Y_i.

$$AP(f) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{|Y_i|}\sum_{l\in Y_i}\frac{|\{l' \in Y_i|\ rank(X_i, l') \leq rank(X_i, l)\}|}{rank(X_i, l)} \qquad (5.8)$$

όπου $rank(X_i, l)$ είναι ο συντελεστής εμπιστοσύνης πραγματικής τιμής για την ετικέτα l από τις κορυφαίες προβλεπόμενες k ετικέτες για το X_i.

Όσο μεγαλύτερη είναι η τιμή, τόσο καλύτερη είναι η απόδοση του συστήματος. Σημειώνεται ότι είναι $AP(f) = 1$ για το σύστημα f το οποίο κατατάσσει τέλεια τις ετικέτες για όλες τις εικόνες (δεν υπάρχει εικόνα X_i για την οποία μια ετικέτα που δεν βρίσκεται στο Y_i να είναι υψηλότερα στην κατάταξη από μια ετικέτα στο Y_i) (Schapire and Singer, 2000).

Εάν ο στόχος ενός συστήματος πολλαπλών κλάσεων είναι η εκχώρηση μιας μόνο ετικέτας σε μία εικόνα, το one-error μετρά πόσες φορές η προβλεπόμενη ετικέτα δεν ήταν στο Y. Μπορούμε να ορίσουμε έναν ταξινομητή h: $X \rightarrow Y$ που αποδίδει μία ετικέτα σε μία εικόνα ως εξής: $h(x) = arg\max_{l\in y} f(x, y)$. Στη συνέχεια, για ένα σύνολο επισημειωμένων εικόνων το one-error (OE) ορίζεται ως:

$$OE(h) = \frac{1}{m}\sum_{i=1}^{m}[h(X_i) \notin Y_i] \qquad (5.9)$$

Για επισημείωση μονής ετικέτας το *one-error* είναι ίδιο με το συνηθισμένο λάθος. Όσο μικρότερη είναι η τιμή του, τόσο καλύτερη είναι η απόδοση του συστήματος.

Η «κάλυψη» (coverage) υπολογίζει το μέσο όρο του μέγιστου συντελεστή θετικής εμπιστοσύνης που υποδεικνύει τον τρόπο κάλυψης όλων των πιθανών ετικετών του συνόλου δεδομένων.

$$Coverage(h) = \frac{1}{m}\sum_{i=1}^{m} \max_{l \in Y_i} rank(X_i, l) - 1 \qquad (5.10)$$

Σε προβλήματα ταξινόμησης μιας ετικέτας, η κάλυψη είναι η μέση κατάταξη της σωστής ετικέτας και είναι μηδέν εάν το σύστημα δεν κάνει λάθη ταξινόμησης. Όσο μικρότερη είναι η τιμή της, τόσο καλύτερη είναι η απόδοση του συστήματος.

Για να αξιολογηθεί η απόδοση κατάταξης ανά ετικέτα, μελετάται επίσης η περιοχή κάτω από τις καμπύλες ROC (*Area Under ROC Curves*-AUC) για κάθε ετικέτα. Το πρόβλημα πολλαπλών ετικετών αντιμετωπίζεται ως πολλαπλά προβλήματα δυαδικής ταξινόμησης και υπολογίζεται η AUC για κάθε πρόβλημα. Στη συνέχεια καταγράφεται η μέση τιμή AUC για όλες τις ετικέτες (Zhang and Zhou ,2010).

Όλα αυτά τα μέτρα αξιολόγησης χρησιμοποιούνται συχνά για την αξιολόγηση της απόδοσης του συστήματος επισημείωσης. Σημειώνεται ότι τα οκτώ κριτήρια αξιολογούν την απόδοση των συστημάτων εκμάθησης πολλαπλών ετικετών από διαφορετικές απόψεις. Συνήθως λίγοι αλγόριθμοι υπερτερούν των υπολοίπων σε όλα αυτά τα κριτήρια (Zhang and Zhou, 2010).

Για την αξιολόγηση της απόδοσης ενός συστήματος επισημείωσης εικόνας μεγάλης κλίμακας, δύο συμπληρωματικά μέτρα αξιολόγησης επιβεβαιώνουν την ευρωστία (robustness) και τη σταθερότητα (stability) του συστήματος επισημείωσης. Για να εκτιμηθεί η ευρωστία του συστήματος επισημείωσης, οι Lin et al. (2016) πρότειναν ως μέτρο ακρίβειας επισημείωσης το zero-rate, που υπολογίζει τον αριθμό των λέξεων-κλειδιών που δεν έχουν προβλεφθεί με ακρίβεια. Για να εκτιμηθεί η σταθερότητα του συστήματος πρότειναν το συντελεστή διακύμανσης (coefficient of variation - CV) που μετράει τη διακύμανση της ακρίβειας επισημείωσης μεταξύ των λέξεων-κλειδιών. Για να είναι ένα σύστημα σταθερό, η τιμή του CV πρέπει να είναι χαμηλή, πράγμα που δείχνει ότι η ακρίβεια επισημείωσης δεν μεταβάλλεται σημαντικά από τη μία εικόνα στην άλλη. Αυτό μπορεί να βοηθήσει στην ανάκτηση παρόμοιων εικόνων, ενώ το ερώτημα μπορεί να είναι οποιαδήποτε από τις λέξεις-κλειδιά (Lin et al., 2016).

5.2. Βάσεις δεδομένων για την επισημείωση εικόνας

Υπάρχουν διάφορες διαθέσιμες στο κοινό βάσεις δεδομένων για την εκπαίδευση και αξιολόγηση συστημάτων επισημείωσης και ανάκτησης εικόνας. Για να είναι αποτελεσματικό το σύστημα επισημείωσης, είναι απαραίτητο το σύστημα να εκπαιδεύεται με μεγάλο αριθμό ισορροπημένων δεδομένων. Ωστόσο, έχουν καταβληθεί προσπάθειες για την εκπαίδευση του συστήματος, ακόμη και αν το σύνολο δεδομένων δεν είναι ισορροπημένο. Η δημιουργία ενός ισορροπημένου και χειροκίνητα επισημειωμένου συνόλου δεδομένων είναι μια δαπανηρή και χρονοβόρα διαδικασία. Έτσι, έχουν καταβληθεί προσπάθειες για την ανάπτυξη ενός ημι-εποπτευόμενου μοντέλου που μπορεί να εκπαιδευτεί χρησιμοποιώντας θορυβώδεις ή ελλιπείς ετικέτες. Παρόλα αυτά, για να εκπαιδεύσουμε ένα μοντέλο, απαιτούνται χειροκίνητα επισημειωμένες εικόνες με τον ένα ή τον άλλο τρόπο. Έχουν γίνει προσπάθειες από τις διάφορες κοινότητες για την ανάπτυξη μιας τυποποιημένης βάσης δεδομένων για την εκπαίδευση και την αξιολόγηση. Λεπτομέρειες σχετικά με ορισμένες από τις παγκοσμίως αποδεκτές και τυποποιημένες βάσεις δεδομένων δίνονται παρακάτω.

5.2.1. Η βάση δεδομένων Corel

Η βάση δεδομένων Corel[2] δημιουργήθηκε και συντηρείται από το Corel Photo Gallery. Διάφορες ερευνητικές ομάδες χρησιμοποίησαν δεδομένα Corel για την αξιολόγηση του συστήματος επισημείωσής τους. Οι εικόνες του σύνολου δεδομένων Corel είναι «χειροκίνητα» επισημειωμένες από εμπειρογνώμονες. Υπάρχουν διάφορες εκδόσεις του συνόλου δεδομένων Corel (Cheng et al., 2018).

5.2.1.1. Corel5K

Το Corel5 K περιέχει 5.000 εικόνες επισημειωμένες με το χέρι. Κάθε εικόνα είναι διάστασης είτε 192 × 128 είτε 128 × 192 pixels. Υπάρχουν συνολικά 371 μοναδικές λέξεις και κάθε εικόνα σχολιάζεται με 1 έως 5 λέξεις-κλειδιά. Το μέγεθος του συνόλου δεδομένων του Corel5K είναι μικρό. Επίσης, έχει ένα μικρό αριθμό λεξιλογίου. Ως εκ τούτου, όταν αξιολογείται ένα σύστημα επισημείωσης στο σύνολο δεδομένων του Corel5K, είναι δύσκολο να διαπιστωθεί αν το προτεινόμενο σύστημα έχει καλή ικανότητα γενίκευσης.

[2] https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval

5.2.1.2. Corel30K

Πρόκειται για μια επέκταση του συνόλου δεδομένων Corel5k που έχει 31.695 εικόνες. Οι εικόνες είναι διάστασης 384 × 256 ή 256 × 384 pixels. Το μέγεθος του λεξιλογίου έχει επίσης αυξηθεί στις 5.587 λέξεις. Κάθε εικόνα φέρει ετικέτα με 1 έως 5 λέξεις-κλειδιά. Ο μέσος αριθμός λέξεων ανά εικόνα είναι περίπου 3,6 (Bhagat and Choudhary, 2018).

5.2.1.3. Corel60K

Πρόκειται για άλλη μία επέκταση της βάσης δεδομένων Corel. Το Corel60K είναι ένα ισορροπημένο σύνολο δεδομένων. Περιέχει 60.000 εικόνες από 600 διαφορετικές κατηγορίες. Κάθε κατηγορία έχει περίπου 100 εικόνες που είναι διάστασης 384×256 ή 256×384 pixels. Υπάρχουν 417 διακριτές λέξεις-κλειδιά και κάθε εικόνα φέρει ετικέτα με 1 έως 7 λέξεις-κλειδιά.

Παρόλο που το σύνολο δεδομένων Corel είναι ένα τυποποιημένο σύνολο δεδομένων, οι Tang και Lewis (2007) επεσήμαναν μερικά από τα μειονεκτήματά του. Οι συγγραφείς εφάρμοσαν τρεις μεθόδους αυτόματης επισημείωσης εικόνας (CSD-prop, SvdCos και CSD-svm) στο σύνολο δεδομένων Corel και συνέκριναν τα αποτελέσματα με ορισμένες από τις state-of-art μεθόδους (μοντέλο μετάφρασης TM, μοντέλο CRM, Μοντέλο MBRM, μοντέλο MIX-Hier). Αυτοί προσπάθησαν να δείξουν ότι, όταν χρησιμοποιηθούν σύνολα εκπαίδευσης και δοκιμής από το ίδιο το σύνολο δεδομένων Corel, είναι σχετικά εύκολο να εκτελεστεί επισημείωση. Υποστήριξαν επίσης ότι το σύνολο δεδομένων Corel περιέχει περιττές πληροφορίες εκπαίδευσης και ότι ένα μοντέλο μπορεί να εκπαιδευτεί χρησιμοποιώντας μόνο το 25% του συνόλου εκπαίδευσης (Bhagat and Choudhary, 2018).

5.2.2. Η βάση δεδομένων του ImageNet

Από το 2010, το ImageNet[3] διοργανώνει κάθε χρόνο έναν διαγωνισμό για την ανίχνευση, την ταξινόμηση, τον εντοπισμό κ.λπ. των αντικειμένων από μια εικόνα. Το ILSVRC-10 είναι ένας διαγωνισμός που στοχεύει στην αξιολόγηση της αποτελεσματικότητας των μεθόδων επισημείωσης εικόνας (ταξινόμηση πολλαπλών ετικετών) όπου ο στόχος είναι να επισημειωθεί κάθε εικόνα με το πολύ πέντε (5) ετικέτες κατά φθίνουσα σειρά εμπιστοσύνης. Υπάρχουν 1.000 κατηγορίες αντικειμένων και οι ετικέτες οργανώνονται ιεραρχικά σε τρία

[3] http://www.image-net.org/challenges/LSVRC/

επίπεδα και περιέχουν 1.000 κόμβους στο επίπεδο των φύλλων χωρίς να επικαλύπτονται. Το σύνολο δεδομένων περιλαμβάνει 120.000 εκπαιδευτικές εικόνες με καθεμία να έχει το πολύ πέντε ετικέτες. Το σύνολο δεδομένων περιέχει 50.000 εικόνες επικύρωσης και 150.000 εικόνες δοκιμών που έχουν συλλεχθεί από το Flickr και άλλους ιστότοπους. Όλες οι εικόνες εκπαίδευσης επισημειώνονται χειροκίνητα χωρίς προηγούμενη τμηματοποίηση. Στο ILSVRC-2011, οι διοργανωτές πρόσθεσαν ένα επιπλέον έργο, δηλαδή την ταξινόμηση και τον εντοπισμό των αντικειμένων. Ο στόχος αυτής της νέας εργασίας είναι η πρόβλεψη των πέντε κορυφαίων ετικετών κλάσης και των πέντε πλαισίων οριοθέτησης, ένα για καθεμία ετικέτα κλάσης. Στο διαγωνισμό ILSVRC-2013 δύο νέες εργασίες ήταν στο επίκεντρο: η ανίχνευση αντικειμένων και η ταξινόμηση και ο εντοπισμός (όπως στο ILSVRC-2011). Ο στόχος της εργασίας ανίχνευσης αντικειμένων είναι να προσδιορίσει την κλάση του αντικειμένου (μεταξύ 200 κατηγοριών) από εικόνες με πλήρη χειροκίνητη επισημείωση με οριοθετημένα πλαίσια. Ανίχνευση αντικειμένων για εργασίες ταξινόμησης βίντεο και σκηνής συμπεριλήφθηκαν στον διαγωνισμό (ILSVRC-2015, ILSVRC-2016, ILSVRC-2017) (Bhagat and Choudhary, 2018).

5.2.3. Η βάση αναφοράς IAPR TC-12

Η βάση δεδομένων IAPR TC-12[4] αποτελείται από 20.000 φυσικές εικόνες που χρησιμοποιήθηκαν κατά τη διάρκεια της εκστρατείας αξιολόγησης ImageCLEF 2006-2008. Οι εικόνες στο σύνολο δεδομένων IAPR TC-12 έχουν πολλαπλά αντικείμενα και περιλαμβάνουν πλήρεις επισημειώσεις (πλήρες κείμενο καθώς και αγγλικά, γερμανικά και τυχαία) καθώς και μερικές επισημειώσεις (όλες οι επισημειώσεις εκτός από την περιγραφή). Υπάρχουν περίπου 291 μοναδικές ετικέτες και κάθε εικόνα έχει περίπου 1 έως 23 ετικέτες. Κατά μέσο όρο, υπάρχουν 5,7 ετικέτες ανά εικόνα και 153 έως 4.999 εικόνες ανά ετικέτα. Υπάρχουν 347 εικόνες ανά ετικέτα κατά μέσο όρο. Το γενικά χρησιμοποιούμενο σύνολο δεδομένων περιλαμβάνει συνολικά 19.627 εικόνες με 291 λέξεις-κλειδιά, δηλαδή 17.665 εικόνες για εκπαίδευση και 1.962 εικόνες για επικύρωση (Cheng et al., 2018).

Υπάρχει μια εκτεταμένη εκδοχή του IAPR TC-12 που ονομάζεται ταξινομημένο και επισημειωμένο IAPR TC-12 (SAIAPR TC-12) [91]. Το SAIAPR TC-12 περιλαμβάνει όλες τις εικόνες του IAPR TC-12 μαζί με την επιλεγμένη μάσκα και τις διαχωρισμένες εικόνες. Το

[4] https://www.imageclef.org/photodata

SAIAPR TC-12 περιλαμβάνει χαρακτηριστικά που εξάγονται από περιοχή, μαζί με ετικέτες που έχουν εκχωρηθεί σε κάθε περιοχή. Οι επισημειώσεις σε επίπεδο περιοχής σύμφωνα με την ιεραρχία και τις πληροφορίες για τις χωρικές σχέσεις είναι διαθέσιμες επίσης στο σύνολο δεδομένων. Τόσο το IAPR TC-12 όσο και το SAIAPR TC-12 είναι διαθέσιμα στο κοινό χωρίς κανένα περιορισμό πνευματικών δικαιωμάτων (Bhagat and Choudhary, 2018).

5.2.4. Επισημείωση εικόνων ImageCLEF

Το ImageCLEF[5] ξεκίνησε το 2003 ως μέρος του φόρουμ CLEF (Cross Language Evaluation Forum) για την αξιολόγηση της απόδοσης μεθόδων ανίχνευσης εννοιών, επισημείωσης και ανάκτησης και από το 2008 ξεκίνησε την διοργάνωση ενός διαγωνισμού οπτικής ανίχνευσης και επισημείωσης για φωτογραφικές εικόνες. Παρόλο που το ImageCLEF ξεκίνησε να οργανώνει διαγωνισμούς-προκλήσεις επισημείωσης και ανάκτησης ιατρικών εικόνων από το 2005, μόνο ένας μικρός αριθμός φωτογραφικών εικόνων ήταν διαθέσιμος για την εκπαίδευση ενός μοντέλου (1.800 στο ImageCLEF-2008, 5.000 στο ImageCLEF-2009, 8.000 στο ImageCLEF-2010 και στο ImageCLEF-2011) και όλες αυτές οι εικόνες επισημειώνονται χειροκίνητα. Αργότερα από το 2012, ο οργανισμός εισήγαγε μια νέα πρόκληση που ονομάζεται επεκτάσιμη εργασία επισημείωσης εικόνας (scalable image annotation). Η ιδέα είναι να ταξινομηθούν οι επισημειωμένες λέξεις-κλειδιά και να αποφασιστεί ο αριθμός των λέξεων-κλειδιών που μπορούν να ανατεθούν σε μια εικόνα. Επίσης, το σύνολο δεδομένων εκπαίδευσης περιέχει μόνο χαρακτηριστικά κειμένου (διεύθυνση URL, κείμενα κ.λπ.) που μπορούν να εξορύσσονται και να χρησιμοποιούνται ως ετικέτες. Το μέγεθος του εκπαιδευτικού συνόλου έχει επίσης αυξηθεί σημαντικά. Η εργασία επισημείωσης φωτογραφιών ImageCLEF-2013 αποτελεί σημείο αναφοράς για την επισημείωση, ανίχνευση οπτικών εννοιών και την ανάκτηση φωτογραφιών. Το σύνολο δεδομένων περιέχει 250.000 εικόνες εκπαίδευσης που έχουν ληφθεί από το διαδίκτυο, 1000 σύνολα εικόνων για ανάπτυξη των μοντέλων και 2.000 εικόνες δοκιμών που ανήκουν σε 95 κατηγορίες. Επισημειωμένες με ground truth ετικέτες παρέχονται οι εικόνες μόνο στο σύνολο ανάπτυξης.

Το σύνολο δεδομένων σχεδιάστηκε με πρόθεση να ελέγξει τη δυνατότητα κλιμάκωσης του προτεινόμενου συστήματος επισημείωσης. Ως εκ τούτου, ο κατάλογος των εννοιών είναι διαφορετικός για το εκπαιδευτικό σύνολο, το σύνολο ανάπτυξης και το σετ δοκιμών. Δεν

[5] https://www.imageclef.org

υπάρχουν χειροκίνητα επισημειωμένες εικόνες στο σύνολο εκπαίδευσης. Ο οργανισμός παρέχει λεκτικά και οπτικά χαρακτηριστικά με το σύνολο δεδομένων. Τα χαρακτηριστικά κειμένου περιλαμβάνουν τον πλήρη ιστό σε μορφή XML, μια λίστα με το ζεύγος λέξη-σκορ, τη διεύθυνση URL της εικόνας, την κατάταξη των εικόνων κατά την αναζήτηση της εικόνας μέσω μιας μηχανής αναζήτησης. Τα οπτικά περιλαμβάνουν τέσσερις τύπους χαρακτηριστικών SIFT (SIFT, C-SIFT, RGB-SIFT, OPPONENT-SIFT), δύο τύπους χαρακτηριστικών GIST (GIST, GIST2), LBP χαρακτηριστικά και δύο τύπους χαρακτηριστικών χρωμάτων (COLOR HIST, HSVHIST) και GETLF (Bhagat and Choudhary, 2018).

5.2.5. Βάση δεδομένων NUS-WIDE

Το σύνολο δεδομένων NUS-WIDE[6] δημιουργήθηκε και συντηρείται από το Εργαστήριο NUS για την αναζήτηση μέσων στο Εθνικό Πανεπιστήμιο της Σιγκαπούρης, για αναζήτηση μέσων για την επισημείωση και ανάκτηση από το Flickr. Το σύνολο δεδομένων περιέχει 269.648 εικόνες με 5.018 μοναδικές ετικέτες. Το σύνολο δεδομένων περιλαμβάνει επίσης έξι τύπους χαρακτηριστικών χαμηλού επιπέδου. Το σύνολο δεδομένων χωρίζεται σε 161.789 εικόνες εκπαίδευσης και 107.859 εικόνες δοκιμών. Για τον σκοπό της αξιολόγησης παρέχεται επίσης ground truth για 81 έννοιες. Τα έξι χαρακτηριστικά περιλαμβάνουν ένα ιστόγραμμα χρώματος 64-D, συσχέτιση χρωμάτων 144-D, ιστόγραμμα κατεύθυνσης ακμής 73-D, wavelet υφής 128-D, ροπές χρώματος 255-D και ένα 500-D bag-of-words που βασίζονται σε περιγραφές SIFT. Το σύνολο δεδομένων NUS-WIDE διατίθεται σε τρεις διαφορετικές εκδόσεις, την NUS-WIDE LITE μια ελαφρά έκδοση, το σύνολο δεδομένων NUS-WIDE OBJECT που περιέχει μόνο ένα αντικείμενο σε κάθε εικόνα και το σύνολο δεδομένων NUS-WIDE SCENE όπου κάθε εικόνα έχει μία σκηνή (Cheng et al., 2018, Bhagat and Choudhary, 2018).

5.2.6. Βάση δεδομένων παιχνιδιού ESP

Το σύνολο δεδομένων ESP αποτελείται από εικόνες που συλλέγονται από το ηλεκτρονικό παιχνίδι επισημείωσης ESP. Στο παιχνίδι ESP, δύο παίκτες κερδίζουν πόντους προβλέποντας την ίδια λέξη-κλειδί για μια εικόνα χωρίς να επικοινωνούν μεταξύ τους. Αυτό το σύνολο δεδομένων αποτελεί πρόκληση καθώς περιέχει μια μεγάλη ποικιλία εικόνων, όπως λογότυπα, σχέδια, τοπία και προσωπικές φωτογραφίες. Επίσης, συσσωρεύεται μια λίστα από λέξεις της καθομιλουμένης. Αυτό το σύνολο δεδομένων περιέχει 67.796 εικόνες συνολικά,

[6] https://lms.comp.nus.edu.sg/wp-content/uploads/2019/research/nuswide/NUS-WIDE.html

αλλά το μέρος που χρησιμοποιείται για πειράματα αποτελείται συνήθως από 20.770 εικόνες με 268 λέξεις-κλειδιά, συμπεριλαμβανομένων 18689 εκπαιδευτικών εικόνων και 2081 δοκιμαστικών εικόνων (Cheng et al.,2018, Bhagat and Choudhary, 2018).

5.2.7. Σύνολο δεδομένων MS-COCO

Το σύνολο δεδομένων της Microsoft (Microsoft Common Objects in COntext)[7] χρησιμοποιείται για την αναγνώριση, την τμηματοποίηση και την επισημείωση εικόνων. Το σύνολο δεδομένων περιέχει 91 κοινές κατηγορίες αντικειμένων με 82 από αυτές να έχουν πάνω από 5.000 επισημειωμένα στιγμιότυπα. Συνολικά, το σύνολο δεδομένων περιέχει 2,5 εκατομμύρια επισημειωμένα στιγμιότυπα σε 328 χιλιάδες εικόνες. Η δημιουργία του συνόλου δεδομένων προήλθε από την εκτεταμένη συμμετοχή πλήθους εργαζομένων μέσω νέων διεπαφών χρήστη. Αξίζει να σημειωθεί ότι ο αριθμός των εικόνων που χρησιμοποιούνται από το σύνολο δεδομένων MS-COCO, διαφέρει επίσης στις διάφορες μελέτες (Cheng et al., 2018).

Εκτός από τα προαναφερθέντα σύνολα δεδομένων εικόνων, υπάρχουν και άλλα σύνολα δεδομένων για την αξιολόγηση της απόδοσης των μεθόδων AIA. Το σύνολο δεδομένων MSRC, που παρέχεται από την ομάδα μηχανικής όρασης στο ερευνητικό κέντρο της Microsoft στο Cambridge, περιέχει 591 εικόνες επισημειωμένες από 23 κλάσεις. Το σύνολο δεδομένων Flickr30 είναι επίσης ένα σύνολο εικόνων του πραγματικού κόσμου που ανιχνεύεται από το Flickr και κατασκευάζεται με την υποβολή 30 μη-αφηρημένων εννοιών ως ερωτημάτων στο Flickr και στη συνέχεια τη συλλογή 1000 ανακτημένων εικόνων για κάθε έννοια. Ένα άλλο σύνολο δεδομένων που χρησιμοποιείται ευρέως στην μηχανική όραση είναι το σύνολο δεδομένων LabelMe, που είναι μια συλλογή από 72.852 εικόνες που περιέχουν πάνω από 10.000 έννοιες. Το σύνολο δεδομένων TRECVID χρησιμοποιείται επίσης ευρέως για επισημείωση εικόνας καθώς και επισημείωση βίντεο. Το σύνολο δεδομένων TRECVID 2005 περιέχει 137 βίντεο από 13 διαφορετικά προγράμματα στα αγγλικά, τα αραβικά και τα κινέζικα και είναι τμηματοποιημένα σε 74.523 φωτογραφίες.

[7] http://cocodataset.org/#home

5.2.8. Σύνοψη

Όλα τα σύνολα δεδομένων τα οποία παρουσιάστηκαν, μπορούν να χρησιμοποιηθούν για τον έλεγχο της ανταγωνιστικότητας των προτεινόμενων μεθόδων. Οι πίνακες 8 και 9 που ακολουθούν, παρουσιάζουν διάφορες κύριες στατιστικές των συνόλων δεδομένων που αναφέρονται στην ενότητα αυτή.

Σύνολο Δεδομένων	Αριθμός εικόνων	Μέγεθος λεξιλογίου	Αριθμός εικόνων εκπαίδευσης	Αριθμός εικόνων δοκιμής	Αριθμός λέξεων ανά εικόνα	Αριθμός εικόνων ανά λέξη
Corel 5K	5.000	260	4.500	500	3,4	58,6
ESP Game	20.770	268	18.689	2081	4,7	362,7
IAPR TC-12	19.627	291	17.665	1962	5,7	347,7

Πίνακας 8. Περιγραφικά στατιστικά στοιχεία των τριών συνόλων δεδομένων αναφοράς (Cheng et al., 2018).

Σύνολο Δεδομένων	Αριθμός εικόνων εκπαίδευσης	Αριθμός εικόνων επικύρωσης	Αριθμός εικόνων δοκιμής	Αριθμός εννοιών	Πλήρως χειροκίνητα επισημειωμένο	Μερικώς επισημειωμένο	Χωρίς επισημείωση
Corel5 K	5.000	–	–	371	Ναι	Όχι	Όχι
Corel30 K	31.695	–	–	5.587	Ναι	Όχι	Όχι
Corel60 K	60.000	–	–	417	Ναι	Όχι	Όχι
ILSVRC-10	1.2 million	50.000	15.000	1.000	Ναι	Όχι	Όχι
ILSVRC-11 μέχρι ILSVRC-14	1.2 m	50.000	10.000	1.000	Ναι	Όχι	Όχι
ImageCLEF-2008	1.800	1.000	2.000	16	Ναι	Όχι	Όχι
ImageCLEF-2009	5.000	1.300		53	Ναι	Όχι	Όχι
ImageCLEF-2010	8.000	10.000		93	Ναι	Όχι	Όχι
ImageCLEF-2011	8.000	10.000		99	Ναι	Όχι	Όχι
ImageCLEF-2012 to 2016	250.000	–	–	–	Όχι	Όχι	Ναι
IAPR TC-12	20.000	–	–	–	Ναι	Όχι	Όχι
SAIAPR TC-12	20.000	–	–	–	Ναι	Όχι	Όχι
NUS-WIDE	269.648	–	107.859	1.000	Ναι	Ναι	Ναι
ESP Game	67.769	–	–		Ναι	Όχι	Όχι

Πίνακας 9. Ορισμένες από τις κυριότερες βάσεις δεδομένων που χρησιμοποιούνται για την εκπαίδευση και την αξιολόγηση μεθόδων επισημείωσης εικόνας (Bhagat and Choudhary, 2018).

Κάθε προτεινόμενο σύστημα αυτόματης επισημείωσης εικόνας πρέπει να αξιολογείται με βάση ένα καλά ισορροπημένο αμερόληπτο σύνολο δεδομένων. Τα περισσότερα από τα διαθέσιμα σύνολα δεδομένων επισημειώνονται με το χέρι από έναν ειδικό ή από χρήστες. Οι

χειροκίνητα επισημειωμένες εικόνες είναι υποκειμενικές, δηλαδή, το σύνολο των ετικετών που έχουν εκχωρηθεί από ένα άτομο σε μια εικόνα, μπορεί να διαφέρει από άτομο σε άτομο. Καθώς ο αριθμός των εικόνων αυξάνεται, είναι ανάγκη να αναπτυχθεί ένα επεκτάσιμο σύστημα που χρησιμοποιεί μάθηση με μερική αλλά και μη επιβλεπόμενη μάθηση. Ένα επεκτάσιμο σύστημα επισημείωσης έχει τις δυνατότητες εύκολης αλλαγής ή κλιμάκωσης του αριθμού των λέξεων-κλειδιών που χρησιμοποιούνται για την επισημείωση της εικόνας. Για την ανάπτυξη ενός κλιμακούμενου συστήματος, υπάρχουν πολύ λίγα σύνολα δεδομένων (ImageCLEF-2012 έως 2016 και NUS-WIDE). Η ανάπτυξη περισσότερων συνόλων δεδομένων με στόχο ένα κλιμακούμενο σύστημα και μια μεγαλύτερη ποικιλία εικόνων, κρίνεται αναγκαία για την ανάπτυξη ενός πλήρους συστήματος επισημείωσης εικόνων (Bhagat and Choudhary, 2018).

Κεφάλαιο 6
Συγκριτική μελέτη μεθόδων επισημείωσης ιατρικής εικόνας

Στο κεφάλαιο αυτό παρουσιάζουμε συγκριτικά μια σειρά από μελέτες μεθόδων αυτόματης επισημείωσης ιατρικών εικόνων. Η σύγκριση των διαφόρων μεθόδων δεν λαμβάνει υπόψη την επίδοση του συστήματος, δηλαδή την ακρίβεια της ταξινόμησης ή της επισημείωσης. Η επιλογή αυτή έγινε γιατί η χρήση διαφορετικών οπτικών χαρακτηριστικών, ταξινομητών και συνόλου δεδομένων με διαφορετικό αριθμό παραδειγμάτων εκπαίδευσης και δοκιμής συνεπάγεται διαφορετικά αποτελέσματα τα οποία δεν είναι συγκρίσιμα.

Η αξιολόγηση της επίδοσης είναι ένα πολύ σημαντικό βήμα στην ανάπτυξη και διερεύνηση νέων μεθόδων έρευνας. Στην αναγνώριση ομιλίας, τη μηχανική μετάφραση και την ανάκτηση πληροφοριών, μεγάλης κλίμακας διαχειριζόμενα γεγονότα αξιολόγησης αποτελούν έναν κοινό τρόπο σύγκρισης των επιδόσεων των διαφόρων συστημάτων.

Το Φόρουμ Αξιολόγησης CLEF[8] (Cross Language Evaluation Forum) στοχεύει στην υποστήριξη παγκόσμιων εφαρμογών ψηφιακής βιβλιοθήκης αναπτύσσοντας μια υποδομή για τη δοκιμή, τον συντονισμό και την αξιολόγηση των συστημάτων ανάκτησης πληροφοριών. Αυτό επιτυγχάνεται με τη δημιουργία δοκιμαστικών μονάδων επαναχρησιμοποιήσιμων δεδομένων τα οποία μπορούν να αξιοποιηθούν από τους προγραμματιστές συστημάτων για τη συγκριτική αξιολόγηση των συστημάτων τους.

Η εργασία επισημείωσης ιατρικών εικόνων προστέθηκε στο διαγωνισμό ImageCLEF το 2005 παράλληλα με την υπάρχουσα εργασία ανάκτησης ιατρικής εικόνας και εξελίχθηκε περαιτέρω σε πέντε εκδόσεις μέχρι το 2009. Μια περιγραφή των πιο καινοτόμων μεθόδων που παρουσιάστηκαν στα πλαίσια του διαγωνισμού αυτόματης επισημείωσης του ImageCLEF, κρίθηκε αναγκαία.

[8] http://www.clef-initiative.eu/web/clef-initiative/home

6.1. Η αυτόματη επισημείωση ιατρικής εικόνας στο ImageCLEF

Ο σκοπός της αυτόματης επισημείωσης εικόνας είναι να περιγράψει το περιεχόμενο της εικόνας με βάση τα χαρακτηριστικά της, τόσο σε ένα αυστηρό πλαίσιο όσο και γενικευμένα, χρησιμοποιώντας μεθόδους από την αναγνώριση προτύπων και τη δομική ανάλυση. Αυτή η περιγραφή μπορεί στη συνέχεια να χρησιμοποιηθεί για να συγκριθεί μια νέα εικόνα με ένα γνωστό σύνολο δεδομένων που περιέχει μια ομάδα προκαθορισμένων κατηγοριών-κλάσεων και έτσι να εκχωρήσει τη σωστή ετικέτα στην εικόνα.

Στον ιατρικό τομέα, η αυτόματη ταξινόμηση εικόνων μπορεί να βοηθήσει στην εισαγωγή συμβατικών ακτινογραφιών σε ένα υπάρχον ηλεκτρονικό αρχείο χωρίς αλληλεπίδραση με το χρήστη. Άλλες εφαρμογές περιλαμβάνουν την αναζήτηση εικόνων σε μια βάση δεδομένων εικόνων ή τον περιορισμό του αριθμού των αποτελεσμάτων σε ένα ερώτημα του χρήστη, για παράδειγμα μετά από μια αναζήτηση εικόνων με κείμενο. Μπορεί ακόμη να είναι χρήσιμη για πολύγλωσση επισημείωση και διορθώσεις της κεφαλίδας DICOM ή ως ένα συστατικό τμήμα του συστήματος υποστήριξης διάγνωσης. Ο στόχος της ιατρικής επισημείωσης στο ImageCLEF είναι να αξιολογήσει και να προωθήσει τις τελευταίες τεχνικές για την αυτόματη επισημείωση των ιατρικών εικόνων. Μια βάση δεδομένων με πλήρως ταξινομημένες ακτινογραφίες τέθηκε στη διάθεση των συμμετεχόντων για την εκπαίδευση των συστημάτων επισημείωσης, παρέχοντας ένα σημείο αναφοράς. Ο σκοπός του διαγωνισμού έγκειται στην επισημείωση μιας σειράς μη επισημειωμένων εικόνων που παρέχονται σε μεταγενέστερο στάδιο για να αποφευχθεί η εκπαίδευση στα δεδομένα δοκιμών.

Ξεκινώντας από το 2005, ο διαγωνισμός επισημείωσης εξελίχθηκε από μία απλή εργασία ταξινόμησης με 57 κλάσεις σε ένα έργο με σχεδόν 200 κλάσεις αφού πέρασε από ένα ενδιάμεσο στάδιο περίπου 120 κλάσεων. Από την αρχή, ωστόσο, ήταν σαφές ότι ο αριθμός των κλάσεων δεν μπορούσε να κλιμακωθεί επ' αόριστον. Ο αριθμός των πιθανών κατηγοριών που θα μπορούσαν να αναγνωριστούν στις ιατρικές εφαρμογές, είναι υπερβολικά υψηλός ώστε να συγκεντρωθούν επαρκή δεδομένα εκπαίδευσης για τη δημιουργία κατάλληλων ταξινομητών. Μια λύση για την αντιμετώπιση αυτού του ζητήματος είναι μια ιεραρχική δομή των κλάσεων, επειδή υποστηρίζει τη δημιουργία ενός συνόλου ταξινομητών για υποπροβλήματα. Ως εκ τούτου, από την αρχή η επισημείωση της εικόνας βασίστηκε στον

ιεραρχικό κώδικα ανάκτησης εικόνων σε ιατρικές εφαρμογές IRMA (Image Retrieval in Medical Applications). Μια αναλυτική περιγραφή της βάσης δεδομένων που αναπτύχθηκε για τους σκοπούς του διαγωνισμού αυτόματης επισημείωσης ιατρικής εικόνας ImageCLEF καθώς και του κώδικα IRMA παρουσιάζεται στο παράρτημα Α.

6.1.1. Επισημείωση ιατρικών εικόνων

Ο διαγωνισμός επισημείωσης ιατρικής εικόνας ImageCLEF προσέλκυσε τη συμμετοχή πολλών ερευνητικών ομάδων από όλο τον κόσμο από την πρώτη της έκδοση. Ορισμένες από τις ομάδες που συμμετείχαν είχαν εμπειρία σε συστήματα εξόρυξης δεδομένων (data mining) και ανάκτησης δεδομένων (retrieval systems), ενώ άλλες ειδικεύονταν στην αναγνώριση και την ανίχνευση αντικειμένων (object recognition and detection). Στη συνέχεια συγκρίνουμε τις διαφορετικές μεθόδους που παρουσιάστηκαν χρησιμοποιώντας τρία κριτήρια για την ανάλυσή τους: (1) την αναπαράσταση της εικόνας, (2) τις μεθόδους ταξινόμησης, (3) την αντιμετώπιση των προκλήσεων που τέθηκαν όπως η χρήση της ιεραρχίας και η μη ισορροπημένη κατανομή των κλάσεων.

6.1.1.1. Αναπαράσταση εικόνας

Ο τρόπος αναπαράστασης του περιεχομένου της εικόνας είναι το πρώτο πρόβλημα που πρέπει να αντιμετωπιστεί κατά την ανάπτυξη ενός συστήματος αυτόματης επισημείωσης. Υπάρχουν διαφορετικές προσεγγίσεις για την εξαγωγή χαρακτηριστικών από εικόνες, ανάλογα με το ποιες θεωρούνται οι πιο αντιπροσωπευτικές πληροφορίες που θα πρέπει να εξαχθούν. Δεδομένου ότι οι ακτινογραφίες δεν περιέχουν καμία πληροφορία χρώματος, τα χαρακτηριστικά ακμών, σχήματος και τα συνολικά χαρακτηριστικά υφής διαδραματίζουν σημαντικό ρόλο στην ταξινόμηση και επισημείωσή τους. Οι περισσότερες ερευνητικές ομάδες χρησιμοποίησαν κυρίως χαρακτηριστικά υφής και σχήματος είτε από ολόκληρη την εικόνα είτε από περιοχές της. Διάφορες μέθοδοι, που παρουσιάστηκαν στο διαγωνισμό του 2005, χρησιμοποιούν απευθείας τις τιμές των εικονοστοιχείων είτε χρησιμοποιώντας μοντέλα παραμόρφωσης στην πλήρη εικόνα (κλιμακούμενα σε σταθερό μέγεθος) είτε με τη χρήση αραιών δειγμάτων patches (Image Distortion Model - IDM) (Keysers et al., 2004, Lehmann et al., 2005). Οι μέθοδοι που προέρχονται από το πεδίο αναγνώρισης αντικειμένων, ακολουθούν κυρίως την ευρέως υιοθετημένη υπόθεση ότι ένα αντικείμενο στις εικόνες αποτελείται από τμήματα που μπορούν να μοντελοποιηθούν ανεξάρτητα. Έτσι, αυτές οι μέθοδοι εξάγουν τοπικά χαρακτηριστικά γύρω από τα σημεία ενδιαφέροντος και

ακολουθούν την προσέγγιση του σάκου-χαρακτηριστικών (bag-of-features) (Marée et al., 2005, Liu et al., 2007b, Tommasi et al., 2008a, Avni et al., 2009) για την αναπαράσταση της εικόνας. Σε γενικές γραμμές, η διάταξη των οπτικών λέξεων δεν λαμβάνεται υπόψη και χρησιμοποιείται μόνο η συχνότητα κάθε οπτικής λέξης για τη δημιουργία των διανυσμάτων χαρακτηριστικών. Ωστόσο, ορισμένες ομάδες πρόσθεσαν τις χωρικές πληροφορίες σε patches (περιοχές) που εξάγονται από τις εικόνες (Avni et al, 2009) αφού παρατήρησαν ότι οι ακτινογραφίες ενός συγκεκριμένου τμήματος του σώματος λαμβάνονται συνήθως στην ίδια θέση και συνεπώς παρουσιάζουν την ίδια χωρική διάταξη.

Άλλες μέθοδοι χρησιμοποιούν ποσοτικοποίηση των διαφορετικών επιπέδων του γκρι σε συνδυασμό με φίλτρα Gabor από το σύστημα ανάκτησης εικόνας medGIFT[9] (Müller et al., 2006). Δεδομένου ότι οι εικόνες δεν περιέχουν καμία πληροφορία χρώματος, χαρακτηριστικά υφής όπως τα χαρακτηριστικά υφής Tamura, οι οπτικοί περιγραφείς του προτύπου MPEG-7 ή τα χαρακτηριστικά Gabor χρησιμοποιήθηκαν από διάφορες ομάδες ως περιγραφείς της εικόνας (Lehmann et al., 2005, Müller et al., 2006, Xiong et al., 2006).

Κάνοντας μια σύντομη αποτίμηση των αποτελεσμάτων που έχουν δημοσιευθεί από το ImageCLEF και παρουσιάζονται στους πίνακες 10-14, παρατηρείται ότι το 2005 οι μέθοδοι που χρησιμοποιούν απευθείας τις τιμές των εικονοστοιχείων και τα μοντέλα παραμόρφωσης (Image Distortion Model - IDM) σημειώνουν καλύτερη επίδοση σε σχέση με τις περισσότερες άλλες μεθόδους. Οι μέθοδοι που προέρχονται από το πεδίο της αναγνώρισης αντικειμένων (Keysers et al., 2004, Marée et al., 2005, Lehmann et al., 2005), αποδίδουν επίσης πολύ καλά αν και δεν ήταν προσαρμοσμένες σε αυτό το έργο. Μπορεί επίσης να φανεί ότι οι μέθοδοι ανάκτησης εικόνων αποδίδουν καλά γι' αυτό το έργο, ειδικά εάν μπορεί να ενσωματωθεί η γνώση από το πεδίο της ιατρικής (Müller et al., 2006).

Μια άλλη σαφής παρατήρηση είναι ότι οι μέθοδοι που χρησιμοποιούν τοπικούς περιγραφείς εικόνων, ξεπερνούν σε επιδόσεις τις μεθόδους που χρησιμοποιούν συνολικούς περιγραφείς εικόνας. Οι μέθοδοι που σημειώνουν τις καλύτερες επιδόσεις από το 2006 έως το 2008, χρησιμοποιούν όλες είτε τοπικά χαρακτηριστικά εικόνας μόνο είτε τοπικά χαρακτηριστικά εικόνας σε συνδυασμό με ένα συνολικό περιγραφέα. Ο συνδυασμός διαφορετικών τοπικών

[9] Το GIFT (GNU Image Finding Tool) είναι ένα σύστημα ανάκτησης εικόνων με βάση το περιεχόμενο που αναπτύχθηκε στα πλαίσια του προγράμματος MedGIFT (που διήρκεσε από 2002 έως το 2007) στην Ιατρική Σχολή του Πανεπιστημίου της Γενεύης στην Ελβετία (http://medgift.hevs.ch/wordpress/).

και συνολικών περιγραφέων σε μια μοναδική αναπαράσταση χαρακτηριστικών αποτελεί μία άλλη ευρέως υιοθετημένη στρατηγική που συναντάται στις περισσότερες προσεγγίσεις (Liu et al., 2007b, Tommasi et al, 2008a).

2005					
Θέση	**Μελέτη**	**Ομάδα**	**Χαρακτηριστικά**	**Ταξινομητής**	**ER (%)**
1	Keysers et al. (2004)	RWTH-i6	thumb. X x 32 IDM	kNN, k=1	12.6
2	Lehman et al. (2005)	RWTH-mi	texture JSD + thumb. *X* x 32 IDM	kNN, k=1	13.3
3	Keysers et al. (2004)	RWTH-i6	image patches, BOW	log-linear model	13.9
4	Marée et al. (2005)	ULG	image patches	boosting	14.1
5	Lehman et al. (2005)	RWTH-mi	texture JSD + thumb. *X* x 32 IDM	kNN, k=1	14.6
6	Marée et al. (2005)	ULG	image patches	decision trees	14.1
7	Müller et al. (2005)	GE	texture, 8 grey lev.	GIFT + kNN, k=5	20.6
8	Xiong et al. (2006)	Infocomm	texture + LRPM, llc	SVM oa + RBF	20.6
9	Müller et al. (2006)	GE	texture, 16 grey lev.	GIFT + kNN, k=5	20.9
10	Xiong et al. (2006)	Infocomm	texture + LRPM, llc	SVM oa + RBF	20.6
11	Xiong et al. (2006)	Infocomm	texture + LRPM, llc	SVM oa + RBF	20.6
12	Müller et al. (2005)	GE	Texture, 8 grey lev.	GIFT + kNN, k=1	21.2
13	Müller et al. (2005)	GE	texture, 4 grey lev.	GIFT + kNN, k=10	21.3
14	Villena-Román et al. (2005)	MIRACLE	texture	GIFT + relev. N=20	21.4
15	Müller et al. (2005)	GE	Texture, 16 grey lev.	GIFT + kNN, k=1	21.7

Πίνακας 10. Η κατάταξη των ερευνητικών ομάδων που συμμετείχαν στο διαγωνισμό ImageCLEF 2005 με βάση το ποσοστό λάθους. Εμφανίζονται οι 15 πρώτες εκτελέσεις που υπέβαλαν οι ομάδες στο διαγωνισμό (Deselaers et al., 2006). (Συμβολισμοί: BOW: bag-of-words (σάκος οπτικών λέξεων), llc, hlc: low and high level cue combination (χαμηλή και υψηλή ενσωμάτωση χαρακτηριστικών))

2006					
Θέση	**Μελέτη**	**Ομάδα**	**Χαρακτηριστικά**	**Ταξινομητής**	**ER (%)**
1	Deselaers et al. (2007)	RWTH-i6	image patches + position, BOW	log-linear model	16.2
2	Setia et al. (2007)	UFR	local rel. coocc. matr. 1000 p.	SVM oa + HI	16.7
3	Deselaers et al. (2007)	RWTH-i6	image patches + position, BOW	SVM oa + HI	16.7
4	Florea et al. (2007)	CISMeF	local + global texture + PCA	SVM oa + RBF	17.2
5	Florea et al. (2007)	CISMeF	local + PCA	SVM oa + RBF	17.2
6	Chang et al. (2006)	MSRA	global, llc	SVM oo + SPM	17.6
7	Florea et al. (2007)	CISMeF	local + global texture + PCA	SVM oa + RBF	17.9
8	Setia et al. (2007)	UFR	local rel. coocc. matr. 800 p.	SVM oa + HI	17.9
9	Chang et al. (2006)	MSRA	image patches. BOW	SVM oo + SPM	18.2
10	Florea et al. (2007)	CISMeF	local + PCA	SVM oa + RBF	20.2
11	Deselaers et al. (2007)	RWTH-i6	thumb. *X* x 32 IDM	kNN, k=1	20.4
12	Güld et al. (2007)	RWTH-mi	texture JSD + thumb. *X* x 32 IDM	kNN, k=1	21.5
13	Güld et al. (2007)	RWTH-mi	texture JSD + thumb. *X* x 32 IDM '05	kNN, k=1	21.7
14	Rahman et al. (2007)	CINDI	local + global, hlc	SVM oo + RBF (+)	24.1
15	Rahman et al. (2007)	CINDI	local + global, hlc	SVM oo + RBF (x)	24.8

Πίνακας 11. Η κατάταξη των ερευνητικών ομάδων που συμμετείχαν στο διαγωνισμό ImageCLEF 2006 με βάση το ποσοστό λάθους. Εμφανίζονται οι 15 πρώτες εκτελέσεις που υπέβαλαν οι ομάδες στο διαγωνισμό (Müller et al., 2007). (Συμβολισμοί: SVM πυρήνες HI: histogram intersection, SPM: Spatial Pyramid Matching, RBF: Radial Basis Function, oa, oo: one-vs.-all and one-vs.-one SVM multi-class)

2007					
Θέση	**Μελέτη**	**Ομάδα**	**Χαρακτηριστικά**	**Ταξινομητής**	**ER (%)**
1	Tommasi et al. (2008a)	Idiap	local + global mlc	SVM oa + MCK χ^2	10.3
2	Tommasi et al. (2008a)	Idiap	local + global mlc	SVM oo + MCK χ^2	11.0
3	Tommasi et al. (2008a)	Idiap	local	SVM oo +χ^2	11.6
4	Tommasi et al. (2008a)	Idiap	local	SVM oa +χ^2	11.5
5	Tommasi et al. (2008a)	Idiap	local + global hlc	SVM oa +χ^2 DAS	11.1
6	Deselaers et al. (2006)	RWTH-i6	comb, rank 8, 10, 11, 12	log-linear model	13.2
7	Setia et al. (2007)	UFR	local rel. coocc. matr. 1000 p.	SVM + HI	12.1
8	Deselaers et al. (2006)	RWTH-i6	patches + position, BOW	log-linear model	11.9
9	Setia et al. (2007)	UFR	local rel. coocc. matr. 800 p.	SVM + HI	13.1
10	Deselaers et al. (2006)	RWTH-i6	patches + position, BOW	log-linear model	12.3
11	Deselaers et al. (2006)	RWTH-i6	patches + position, BOW	log-linear model	12.7
12	Deselaers et al. (2006)	RWTH-i6	patches + position, BOW	log-linear model	12.4
13	Deselaers et al. (2006)	RWTH-i6	patches + position, BOW	log-linear model	17.8
14	Setia et al. (2007)	UFR	local rel. coocc. matr.	SVM + HI AX	17.9
15	Setia et al. (2007)	UFR	local rel. coocc. matr.	decision tree	16.9

Πίνακας 12. Η κατάταξη των ερευνητικών ομάδων που συμμετείχαν στο διαγωνισμό ImageCLEF 2007 με βάση το ποσοστό λάθους. Εμφανίζονται οι 15 πρώτες εκτελέσεις που υπέβαλαν οι ομάδες στο διαγωνισμό (Deselaers et al., 2008). (Συμβολισμοί: llc, mlc, hlc: χαμηλή, μεσαία, υψηλή ενσωμάτωση χαρακτηριστικών, AX: τέσσερις διαφορετικές ταξινομήσεις, μία για κάθε άξονα του κώδικα IRMA εκτελούνται και στη συνέχεια συνδυάζονται)

2008					
Θέση	**Μελέτη**	**Ομάδα**	**Χαρακτηριστικά**	**Ταξινομητής**	**Score**
1	Tommasi et al. (2008b)	Idiap	local + global. llc + virt. img.	SVM oa + χ^2 comm.	74.9
2	Tommasi et al. (2008b)	Idiap	local + global. llc + virt. img.	SVM oa +χ^2	83.5
3	Tommasi et al. (2008b)	Idiap	local + global. llc	SVM oa + χ^2 comm.	83.8
4	Tommasi et al. (2008b)	Idiap	local + global, mlc + virt. img.	SVM + MCK oa χ^2 comm.	85.9
5	Tommasi et al. (2008b)	Idiap	local + global, llc	SVM oa +χ^2	93.2
6	Tommasi et al. (2008b)	Idiap	local	SVM oa +χ^2	100.3
7	Avni et al. (2008)	TAU	patches whole img. BOW	SVM oo + RBF	105.8
8	Avni et al. (2008)	TAU	patches mult. res. BOW, hlc	SVM oo + RBF (prob.)	105.9
9	Avni et al. (2008)	TAU	patches mult. res. BOW, hlc	SVM oo + RBF (vote)	109.4
10	Avni et al. (2008)	TAU	patches resized img. BOW	SVM oo + RBF	117.2
11	Tommasi et al. (2008b)	Idiap	local	SVM oa +χ^2	128.58
12	Güld et al. (2000)	RWTH-mi	texture JSD + thumb. *X* x 32 IDM	kNN, k=5 major.	182.8
13	Lana-Serrano et al. (2008b)	MIRACLE	local + global	kNN, k=3	187.9
14	Lana-Serrano et al. (2008b)	MIRACLE	local + global	kNN, k=2	190.4
15	Lana-Serrano et al. (2008b)	MIRACLE	local + global	kNN, k=2 + RF	190.4

Πίνακας 13. Η κατάταξη των ερευνητικών ομάδων που συμμετείχαν στο διαγωνισμό ImageCLEF 2008 με βάση τη συνολική βαθμολογία τους. Εμφανίζονται οι 15 πρώτες εκτελέσεις που υπέβαλαν οι ομάδες στο διαγωνισμό (Deselaers et al., 2009). (Συμβολισμοί: virt.im: χρήση εικονικών δειγμάτων που ορίζονται τροποποιώντας ελαφρώς τις αρχικές εικόνες, major: συνδυασμός με βάση την πλειοψηφία, prob: πιθανοτική ερμηνεία της εξόδου του SVM, vote: συνδυασμός των ψήφων σε one-vs-one multiclass SVM)

2009					
Θέση	**Μελέτη**	**Ομάδα**	**Χαρακτηριστικά**	**Ταξινομητής**	**Score**
1	Avni et al. (2009)	TAU	patches mult. res. BOW, llc	SVM oo	852.8
2	Tommasi et al. (2008b)	Idiap	local + global llc + virt. imm.	SVM oa $+\chi^2$ comm.	899.2
3	Tommasi et al. (2008b)	Idiap	local + global llc	SVM oa $+\chi^2$ comm.	899.4
4	Tommasi et al. (2008b)	Idiap	local + global llc	SVM oa $+\chi^2$	1039.6
5	Tommasi et al. (2008b)	Idiap	local + global llc + virt. imm.	SVM oa $+\chi^2$	1042.0
6	Dimitrovski et al. (2010)	FEITIJS	local + global llc	bagging, rand, forest, AX	1352.6
7	Ünay et al. (2009)	VPA-Sabanci	local + block position	SVM oa + RBF, AX	1456.2
8	Ünay et al. (2009)	VPA-Sabanci	local + block position	SVM oa + RBF	1513.9
9	Ünay et al. (2009)	VPA-Sabanci	local + block position freq.	SVM oa + RBF	1554.8
10	Ünay et al. (2009)	VPA-Sabanci	local + block position freq.	SVM oa + RBF	1581.7
11	Zhou et al. (2010)	GE	texture, 8 grey lev. vcad	GIFT + kNN, k=5	1633.3
12	Ünay et al. (2009)	GE	texture, 16 grey lev. vcad	GIFT + kNN, k=5	1633.3

Πίνακας 14. Η κατάταξη των ερευνητικών ομάδων που συμμετείχαν στο διαγωνισμό ImageCLEF 2008 με βάση τη συνολική βαθμολογία τους. Εμφανίζονται οι 15 πρώτες εκτελέσεις που υπέβαλαν οι ομάδες στο διαγωνισμό (Tommasi et al., 2010).

6.1.1.2. Μέθοδοι ταξινόμησης

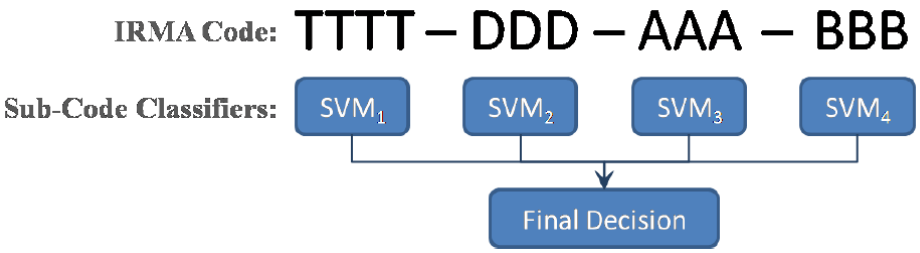
Η επιλογή της τεχνικής ταξινόμησης σημαίνει την επιλογή των κανόνων που αποτελούν τη βάση της διαδικασίας επισημείωσης. Έχουν εφαρμοστεί πολλές διαφορετικές στρατηγικές ταξινόμησης και ενώ τα πρώτα χρόνια οι προσεγγίσεις με βάση το μοντέλο k-πλησιέστερων γειτόνων με k μεταξύ 1 και 20, ήταν οι πιο συνηθισμένες και πιο επιτυχημένες (Keysers et al., 2004, Lehmann et al., 2005, Müller et al., 2006, Besancon and Millet, 2005), σταδιακά μετά το 2006, οι διακριτικές προσεγγίσεις όπως τα λογαριθμικά γραμμικά μοντέλα (log-linear models) (Deselaers et al, 2006) και τα δέντρα αποφάσεων (decision trees) (Marée et al., 2005, Setia et al, 2008, Dimitrovski et al. 2010) καθώς και οι μηχανές διανυσμάτων υποστήριξης (Support Vector Machines) έγιναν ολοένα και πιο κοινές και ξεπέρασαν τις προσεγγίσεις που βασίζονται στον πλησιέστερο γείτονα. Αναφορικά με τις μηχανές διανυσμάτων υποστήριξης (SVMs), η ποικιλία στην απόδοση είναι υψηλή και πιθανώς εξαρτάται από τις αναπαραστάσεις εικόνας και τους πυρήνες που χρησιμοποιούνται. Αλγόριθμοι μηχανικής μάθησης χρησιμοποιήθηκαν μόνο από συστήματα βασισμένα στο διακριτικό μοντέλο.

Αρκετές από τις μεθόδους που χρησιμοποίησαν τον χωρίς παραμέτρους ταξινομητή πλησιέστερων γειτόνων, αξιοποίησαν γνωστά συστήματα ανάκτησης εικόνων βάσει περιεχομένου (CBIR), όπως το GIFT (Müller et al, 2006), για τον προσδιορισμό των πλησιέστερων γειτόνων. Στα συστήματα αυτά οι εικόνες εκπαίδευσης χρησιμοποιούνται ως βάση δεδομένων εικόνων και οι εικόνες δοκιμής χρησιμοποιούνται ως εικόνες ερωτήματος για ανάκτηση. Για κάθε ερώτηση, οι εικόνες εκπαίδευσης ταξινομούνται ανάλογα με την

ομοιότητά τους και εφαρμόζεται ένας κανόνας απόφασης του πλησιέστερου γείτονα, δηλαδή η κλάση της πιο «όμοιας» εικόνας εκπαίδευσης επιλέγεται για κάθε εικόνα δοκιμής. Τέλος, ας σημειωθεί ότι ανάλογα με το συνδυασμό χαρακτηριστικών, ο συνδυασμός ταξινομητών ήταν επίσης ένας δημοφιλής τρόπος βελτίωσης της απόδοσης του συστήματος (Rahman et al, 2006, Tommasi et al., 2008a, Avni et al, 2009).

6.1.1.3. Ιεραρχική επισημείωση

Από το 2007, όταν χρησιμοποιήθηκε ολόκληρος ο κώδικας IRMA για την επισημείωση, οι περισσότερες από τις προτεινόμενες μεθόδους αντιμετωπίζουν την ιεραρχία εισάγοντας τέσσερις διαφορετικούς ταξινομητές, έναν για κάθε άξονα. Οι ληφθείσες ετικέτες από κάθε ταξινομητή συνδέονται στη συνέχεια για να δώσουν την τελική επισημείωση (Ünay et al., 2009, Setia et al., 2008), όπως φαίνεται στην εικόνα 29. Άλλες στρατηγικές στηρίζονται στον ορισμό ενός μοναδικού ταξινομητή ικανού να διαχειρίζεται τις γνώσεις που κωδικοποιούνται στην ιεραρχία των κλάσεων. Χαρακτηριστικά παραδείγματα είναι η εισαγωγή σταθμισμένων αποστάσεων σε ταξινομητές k-πλησιέστερων γειτόνων (Springmann and Schuldt, 2007) ή σταθμισμένων κανόνων διαχωρισμού σε δέντρα αποφάσεων που αντικατοπτρίζουν την ιεραρχική βαθμολογία του σφάλματος (Setia et al., 2008). Ορισμένες ομάδες πρότειναν επίσης τον συνδυασμό επισημείωσης με βάση τους άξονες και επίπεδης επισημείωσης (Deselaers et al. 2008) ή την ενσωμάτωση της εξόδου διαφορετικών ταξινομητών, λαμβάνοντας υπόψη την πλειοψηφία (majority voting) των χαρακτήρων σε κάθε θέση του κώδικα (Güld and Deserno, 2008).



Εικόνα 29. Απεικόνιση της συνολικής ταξινόμησης με βάση τους υποκώδικες IRMA. Για κάθε υποκώδικα εκπαιδεύεται ένας ξεχωριστός SVM και η τελική απόφαση σχηματίζεται με τη σύζευξη των προβλέψεων κάθε SVM (Ünay et al.,2009).

6.1.1.4. Μη ισορροπημένη κατανομή κλάσης

Μία από τις δυσκολίες της επισημείωσης ιατρικής εικόνας είναι η άνιση κατανομή δειγμάτων στις κλάσεις του συνόλου των εικόνων εκπαίδευσης. Έχουν υπάρξει μόνο λίγες προσπάθειες να αντιμετωπιστεί άμεσα η ανισορροπία κλάσεων. Μια από τις προσεγγίσεις επικεντρώθηκε

στον υπολογισμό των χαρακτηριστικών: ο αριθμός των patches που εξήχθησαν από κάθε εικόνα για την κατασκευή του λεξιλογίου οπτικών λέξεων ορίστηκε ως αντιστρόφως ανάλογος προς τον αριθμό των εικόνων στην κλάση (Marée et al., 2005). Μια άλλη προσέγγιση προσάρμοσε τον ταξινομητή k-πλησιέστερων γειτόνων (kNN) με διαφορετική τιμή k για κάθε κλάση, λαμβάνοντας υπόψη τη συχνότητα των εικόνων μέσα στο σύνολο εκπαίδευσης (Zhou et al., 2008). Η παρουσία των αραιών κλάσεων (κλάσεων με πολύ μικρό αριθμό εικόνων) στο αρχικό σύνολο εκπαίδευσης αντιμετωπίστηκε επίσης με το διαχωρισμό των δεδομένων σε υποομάδες με βάση τη συχνότητα και την εκπαίδευση ξεχωριστού SVM για καθένα μία από αυτές (Unay et al., 2009), αλλά και με τη δημιουργία εικονικών παραδειγμάτων (Tommasi et al., 2008b).

6.2. Συμπεράσματα από τη βιβλιογραφική ανασκόπηση

Στην ενότητα αυτή παρουσιάζουμε συνοπτικά τα βασικά συμπεράσματα που προέκυψαν από την βιβλιογραφική ανασκόπηση των επιστημονικών άρθρων που μελετήθηκαν στα πλαίσια της εργασίας μας και επικεντρώνονται στο πρόβλημα της επισημείωσης της ιατρικής εικόνας. Συνολικά 79 άρθρα, που δημοσιεύτηκαν από το 2003 έως το 2019, και έχουν ως κύριο θέμα τους την ταξινόμηση και επισημείωση ιατρικών εικόνων από διαφορετικές μεθόδους ιατρικής απεικόνισης (modality), από διαφορετικές ανατομικές περιοχές και βιολογικά συστήματα, μελετήθηκαν και παρουσιάζονται συγκριτικά στους Πίνακες 15 και 16.

Στον Πίνακα 15 καταγράφεται η προσέγγιση που ακολουθεί κάθε μελέτη για την αναπαράσταση της εικόνας με βάση οπτικά χαρακτηριστικά (χαμηλού επιπέδου) που εξάγονται είτε από ολόκληρη την εικόνα (συνολικά) είτε από περιοχές που περιγράφουν αντικείμενα και αντιστοιχούν σε σημασιολογικές έννοιες (τοπικά). Η εξαγωγή τοπικών χαρακτηριστικών σε αρκετές περιπτώσεις, εκτελείται μετά την τμηματοποίηση της εικόνας είτε σε περιοχές ενδιαφέροντος (ROI's) είτε σε σταθερού μεγέθους τμήματα (block) εφαρμόζοντας μία προσέγγιση πλέγματος. Εκτός από τα χαρακτηριστικά που εξάγονται σε κάθε μέθοδο για την αποτελεσματικότερη αναπαράσταση του περιεχομένου της εικόνας, καταγράφονται, όταν είναι διαθέσιμες, λεπτομέρειες σχετικά με τις διαστάσεις της εικόνας που υποβάλλεται σε επεξεργασία καθώς και η διάσταση του τελικού διανύσματος χαρακτηριστικών το οποίο χρησιμοποιείται για την διαδικασία της ταξινόμησης. Στον Πίνακα 16 καταγράφονται τα στοιχεία που αφορούν τη μέθοδο επισημείωσης που υιοθετεί κάθε

μελέτη, όπως τον ταξινομητή που χρησιμοποιείται ή /και τη μέθοδο μάθησης, τον αριθμό ετικετών που εκχωρούνται ανά εικόνα, τη δυνατότητα κλιμάκωσης (τον αριθμό των ετικετών / κλάσεων που μαθαίνει το σύστημα για επισημείωση) και το σύνολο δεδομένων που χρησιμοποιείται για την εκπαίδευση του μοντέλου.

Α/α	Μελέτη	Τμηματοποίηση	Χαρακτηριστικά	Διάσταση διανύσματος χαρακτηριστικών	Ανάλυση εικόνων εισόδου
2003					
1	Characterization of CT Liver Lesions Based on Texture Features and a Multiple Neural Network Classification Scheme, Mougiakakou et al. (2003)	No	Texture 5 feature sets (First Order Statistics, Spatial Gray Level Dependence Matrix-GLDM, Gray-Level Difference Matrix, Law's Texture Energy Measures, Fractal Dimension)	89	512x512
2004					
2	Classification of Medical Images Using Non-linear Distortion Models, Keysers et al. (2004)	No	IDM Edge (Sobel filter)	N/A	32 x 32
2005					
3	Automatic categorization of medical images for content-based retrieval and data mining, Lehmann et al. (2005)	No	IDM scaled to Xx32 Texture features (Tamura, Castelli (fractal dimension etc.), DCT-based) Edge structure	Various Tamura (384) Castelli (43)	256x256 Various scales
4	Biomedical Image Classification with Random Subwindows and Decision Trees, Marée et al. (2005)	B (randomly extracted subwindows)	Pixel values	256 values for each subwindow	Each subwindow scaled to 16x16
5	MIRACLE's naive approach to medical images annotation, Villena-Román et al. (2005)	B	GIFT features (Global and Local) Gabor Texture Filters	N/A	N/A
6	Data Fusion of Retrieval Results from Different Media: Experiments at ImageCLEF 2005, Besancon and Millet (2005)	B (4 blocks 50x50)	Color Histogram, Global Texture Histogram Edge Histogram (Sobel filter for edges)	Color Histogram: 64 Texture Histogram: 512 Edge Histogram: 400	100x100
7	Supervised Machine Learning Based Medical Image Annotation and Retrieval in ImageCLEFmed 2005, Rahman et al. (2005)	No	Texture (Haralick, 5 features), Shape (7 moments) Edge Histogram (Canny Edge detector)	99 (texture 20 + Shape 7 + Edge 72)	512 × 512 pixel bounding box (keeping aspect ratio), 256 gray values
2006					
8	The Use of MedGIFT and EasyIR for ImageCLEF 2005, Müller et al. (2006)	B	GIFT features(Global and Local) Gabor Texture Filters	N/A	different scales

9	Combining Visual Features for Medical Image Retrieval and Annotation, Xiong et al. (2006)	R (GMM and EM segmentation algorithms)	Global and local Color Histogram, Texture features (LRPM, texture histogram) Region shape (elliptical harmonics)	352	16x16
10	Combining Text and Image Queries at ImageCLEF2005, Chang et al. (2006)	B (32x32 blocks)	Block features: Average gray value for each block	1024	256x256
11	Combining Textual and Visual Features for Cross-Language Medical Image Retrieval, Cheng et al. (2006)	B	Facade scale feature (pixel values image scaled down to 8x8), coherence moment, fuzzy histogram, Gabor Texture Features	various	various scales
12	Categorizing and Annotating Medical Images by Retrieving Terms Relevant to Visual Features, Ballesteros and Petkova (2006)	B (5x5 grid)	Texture features (3 features), Gabor Energy	Tamura: 36, Gabor: 12	256x256
2007					
13	CYU_IM@ImageCLEF 2007: Medical Image Annotation Task, Cheng and Yang (2007)	B	Global and Local image features (Relative local image feature, Correlogram feature)	324	N/A
14	CINDI at ImageCLEF 2006: Image Retrieval and Annotation Tasks for the General Photographic and Medical Image Collections, Rahman et al. (2007)	B (4x4 grid, 5 overlapping subimages)	Global (Shape: MPEG-7 Edge Histogram Descriptor, MPEG-7 Color Layout Descriptor) Local (color moments) texture features from GLCM)	PCA	64x64
15	MedIC at ImageCLEF 2006: Automatic Image Categorization and Annotation Using Combined Visual Representations, Florea et al. (2007)	B (16 blocks 64x64)	global and local Texture (co-occurrence matrix, fractal dimension, Gabor filters, DCT etc.)	122 features for each block 2074-dimensional feature vector/ PCA -335	256×256
16	Medical Image Annotation and Retrieval Using Visual Features, Liu et al. (2007b)	B (64 blocks 8x8)	Global (gray-block features, block-wavelet features, features accounting for binarised images, edge histogram Local (400 patches from each image)	382-dimensional feature vector	N/A
17	A Refined SVM Applied in Medical Image Annotation, Qiu (2007)	R	LRPM (as color), texture features, salient map, salient point, SIFT, stripe	PCA	32x32
18	Medical Image Retrieval and Automated Annotation: OHSU at ImageCLEF 2006, Hersh et al. (2007)	No	partly localised GLCM features	Varying from 32 to 356.	256x256 scaled down to i. 16x16 and ii. 128x128

19	Image Retrieval and Annotation Using Maximum Entropy, Deselaers et al. (2007)	B	sparse histogram of image patches and absolute position	Histograms have 65536 or 4096 bins	32x32
20	Baseline Results for the ImageCLEF 2006 Medical Automatic Annotation Task, Güld et al. (2007)	No	global texture features TTM: Tamura, CTM: Castelli, CCF, IDM	TTM: 384 CTM: 43	32x32 for CCF, Xx32 for IDM
21	Grayscale Radiograph Annotation Using Local Relational Features, Setia et al. (2007)	No	Gradient-like features extracted over interest points (wavelet-based salient point detector)	a codebook of size 20 + co-occurrence matrix => 4D array per image which is flattened and used as the final feature vector	N/A
2008					
22	University and Hospitals of Geneva Participating at ImageCLEF2007, Zhou et al. (2008)	N/A	GIFT (GNU Image Finding Tool) features	N/A	N/A
23	Medical Image Retrieval and Automatic Annotation: OHSU at ImageCLEF 2007, Kalpathy-Cramer and Hersh (2008)	N/A	32 multiscale-oriented Gabor filters / SIFT descriptors	512-dimensional vector, PCA : 100-dimensional vector	N/A
24	Speeding Up IDM without Degradation of Retrieval Quality, Springmann and Schuldt (2008)	B	IMD with local context, edge features (Sobel filter)	N/A	X x 32 pixels
25	Discriminative cue integration for medical image annotation, Tommasi et al. (2008a)	No	Local (modSIFT) Global (raw pixels) BoVW (k=500) with k-means	2000+1024	32x32
26	MIRACLE at ImageCLEFanot 2007: Machine Learning Experiments on Medical Image Annotation, Lana-Serrano et al. (2008a)	N/A	i. Histogram (gray-value histograms, Tamura) ii. Vector (global texture features + Gabor features) iii. Comlpete (both)	Histogram: 768 Vector: 75 Complete: 843	N/A
27	Baseline Results for the ImageCLEF 2007 Medical Automatic Annotation Task Using Global Image Features, Güld and Deserno (2008)	No	Tamura texture histograms, image cross correlations, IDM (image deformation model)	Tamura: 384-dimensional histogram	256x256 downscaled to 32x32 for CCF, Xx32 for IDM
28	Automatic medical image categorization and annotation using LBP and MPEG-7 Edge Histograms, Tian et al. (2008)	B (16 blocks 64X64)	E (edge histogram MPEG7) T (text histogramLBP)	602 466 330 68 (PCA)	256x256
29	Automatic Multilevel Medical Image Annotation and Retrieval, Mueen et al. (2008)	B (4 non-overlap patches)	Local and global features: Texture (Gray level Co-occurrence matrix), Shape (Edge Histograms-Canny) + pixel values	490 (+PCA)	100x100 15x15
30	Grayscale medical image annotation using local	No	Salient Points, Texture Relational	All-invariant accumulator 4000	N/A

	relational features, Setia et al. (2008)		features (based on LBP), clustering	Rotation invariant accumulator 16000	
31	CLEF2008 Image Annotation Task: an SVM Confidence-Based Approach, Tommasi et al. (2008b)	B (4 parts)	Local descriptors (SIFT + Local Binary Pattern)	3240	N/A
32	TAU MIPLAB at ImageClef 2008, Avni et al. (2008)	B	BoVW- (700 visual words) sampled patches undergo PCA	6/7 features	N/A
33	MIRACLE at ImageCLEFannot 2008: Classification of Image Features for Medical Image Annotation, Lana-Serrano et al. (2008b)	B (64x64 pixel blocks)	global and local features (histogram, image statistics, Gabor, fractal dimension, DCT and DWT coefficients, Tamura, coocurrency matrix)	3741 features for each image	256x256
2009					
34	The MedGIFT Group at ImageCLEF 2008, Zhou et al. (2009)	B (four equally sized regions)	Color and Texture medGIFT (different descriptors, i.e. color histogram Gabor filters)	N/A	N/A
35	Baseline Results for the ImageCLEF 2008 Medical Automatic Annotation Task in Comparison over the Years, Güld et al. (2009)	B	Tamura Texture Measures, cross-correlation function (CCF), (IDM) Image Distortion Model	TTM: 384	256x256 downscaled to 32x32 for CCF, Xx32 for IDM
36	Automated X-Ray Image Annotation - Single versus Ensemble of Support Vector Machines, Ünay et al. (2009)	B (16 non overlapping blocks)	spatially enhanced local binary patterns (LBP)	944 features per image (59 per block)	N/A
37	Histopathology Image Classification Using Bagof Features and Kernel Functions, Caicedo et al. (2009)	B	Texton (raw pixel descriptor) SIFT (Bag-of-features)	SIFT: 128	N/A
2010					
38	Dense Simple Features for Fast and Accurate Medical X-Ray Annotation, Avni et al. (2009)	B	raw patches, raw patches with normalized variance, SIFT descriptors (+ spatial coordinates), bag-of-visual words	N/A Reduced with PCA	512x512
39	ImageCLEF 2009 Medical Image Annotation Task: PCTs for Hierarchical Multi-Label Classification, Dimitrovski et al. (2010)	B (16 non overlapping blocks)	global (edge histogram) and local (SIFT) features	2080	N/A
40	The MedGIFT Group at ImageCLEF 2009, Zhou et al. (2010)	B	GIFT / SIFT descriptors	SIFT: 128	N/A
41	Automatic pathology annotation on medical images: a statistical machine translation framework, Gong et al. (2010)	R	Shape + location (eccentricity, solidity, extent, skull)	N/A	N/A

42	Hierarchical Medical image annotation using SVM-based Approaches, Amaral et al. (2010)	B (32x32 non-overlapping subwindows)	C+T+S Global and local (MPEG-7 CLD, EHD, Tamura, GIST) Bag-of-Words (SURF)	954	Max 512x512
43	Content-Based Retrieval and Classification of Ultrasound Medical Images of Ovarian Cysts, Sohail et al. (2010)	R Manual segmentation Region of Interest (ROI)	Histograms moments, Gray Level Co-Occurrence Matrix (GLCM)	121 (64 moments as image histogram based features+ 57 texture features)	N/A
44	ImageCLEF 2010 Modality Classification in Medical Image Retrieval Multiple feature fusion with normalized kernel function, Han and Chen (2010)	B	Visual features: Color Histogram, Block-based Edge Histogram, Block-based variance histogram, SIFT Textual features: 90 vocabulary words	Visual: N/A Textual: 90	N/A
2011					
45	Hierarchical annotation of medical images, Dimitrovski et al. (2011)	B	C+T+S (Raw Pixel Representation, Edge Histogram, Local Binary Patterns, SIFT) Bag-of-Words	LBP:944, Edge Histogram: 80, SIFT BoW: 500 codeworks x 8= 4000	32x32
46	MRI brain classification using support vector machine, Otman et al. (2011)	No	DWT (discrete wavelet transformation)	approximate 17689 wavelet coefficients	256 x 256
47	Classification of MRI brain images using k-nearest neighbor and artificial neural network, Rajini and Bhavani (2011)	No	DWT (discrete wavelet transformation)	1024 (7 with PCA)	256 x 256
2012					
48	Automatic medical image annotation and keyword-based image retrieval using relevance feedback, Ko et al. (2012)	B	T (wavelet-based centre symmetric Local Binary Pattern (WCS-LBP))	768	N/A
49	Annotation of medical images using the surf descriptor, Wojnar, Pinheiro (2012)	B	Interest Points (SIFT / SURF)	SURF (64) SIFT (128)	N/A
50	Content-based medical image annotation and retrieval using perceptual hashing algorithm, Nagarajan and Saravanan (2012)	No	C (Hash value)	N/A	N/A
51	Medical X-ray Image Classification Using Gabor-Based CS-Local Binary Patterns, Ghofrani et al. (2012)	B (25 overlapping subimages)	Global+Local Edges and shape (Gabor filters, Local Binary Pattern (CS-LBP))	4800 (35 with PCA)	N/A
52	Medical X-ray Images Classification Based on Shape Features and Bayesian Rule, Fesharaki et al. (2012)	No	Shape features (include Fourier Descriptor (FD), Invariant Moments (IM), and Zernike Moments (ZM).	263	N/A

53	Novel shape-texture feature extraction for medical x-ray image classification, Mohammadi et al. (2012)	B (25 overlapping subimages)	Shape and texture (Gabor filters, Statistical: Invariant moments, Entropy)	2700 (240 with PCA)	N/A
54	Automatic Annotation of Radiological Observations in Liver CT Images, Gimenez et al. (2012)	R Manual segmentation Region of Interest (ROI)	Contrast, Texture (Haar wavelet transform, Gabor, Daubechies features), Edge, Shape	431	N/A
55	Automated Image Annotation for Semantic Indexing and Retrieval of Medical Images, Krishna and Prasad (2012)	No	Global Texture features (Gray level Co-occurrence matrix, Energy, Entropy, Contrast)	N/A	256x256
2013					
56	Automatic image annotation and semantic based image retrieval for medical domain, Burdescu et al. (2013)	R	C+T+S	N/A	N/A
57	Using a bag of words for automatic medical image annotation with a latent semantic, Bouslimi et al. (2013)	No	SIFT	128	N/A
58	Automatic medical x-ray image classification using annotation, Zare et al. (2013a)	No	SIFT	128	N/A
59	Classification Using k Nearest Neighbor for Brain Image Retrieval, Charde and Lokhande (2013)	No	Shape (Canny edge detection) and Texture (Gabor filters)	N/A	N/A
60	Medical X-ray Image Hierarchical Classification using a merging and splitting scheme in feature space, Fesharaki and Pourghassen (2013)	No	Shape (geometric, i.e. axis of least inertia (ALI), eccentricity, elongation regarding, circularity, invariant moments (IM) and Zernike moments (ZM), Fourier) and Texture (GLCM, Wavelet)	768 (48 with OFS method)	512 × 512
61	Automatic Classification Of Medical X-Ray Images, Zare et al. (2013b)	B	Canny, GLCM, Pixel, LPB / Bow (SIFT)	278 / 944	15x15
2014					
62	Medical image annotation and retrieval by using classification techniques, Abdulrazzaq et al. (2014)	B	C+T+S Local andGlobal (C:pixel values T:co-occurrence matrix, DWT, S:oriented gradient descriptor, SURF)	1785/PCA(20, 50, 100)	256x256

63	A new fusion model for classification of the lung diseases using genetic algorithm, Bhuvaneswi et al. (2014)	B (12 blocks)	Texture Fast Walsh–Hadamard transform (FWHT) and Gabor	60 features for each image	256 x 256
64	MRI brain cancer classification using Support Vector Machine, Nandpuru et al. (2014)	No	Gray scale (Standard Deviation, Skewness, Kurtosis), symmetrical (Exterior symmetry), texture (co-occurrence matrix, i.e. Entropy, Energy)	28 features for each Image (24 with PCA)	N/A
65	Augmenting Multi-Instance Multi-label Learning with Sparse Bayesian Models for Skin Biopsy Image Analysis, Zhang et al. (2014)	R (NCut)	Color (L,U,V means), DWT coefficients Graph representation	26	800x600
66	Liver CT annotation via generalized coupled tensor factorization, Ermis and Cemgil (2014)	R	Computer Generated-CoG	CoG: 446	various
67	Towards content-based image retrieval: From computer generated features to semantic descriptions of liver CT scans, Spanier and Joskowicz (2014)	R	Computer Generated-**CoG** a modified CoG list with 18 global image descriptors and 30 pathology descriptors	N/A	various
2015					
68	New approach for Automatic Medical Image Annotation Using the Bag-of-Words Model, Bouslimi and Akaichi (2015)	ROI's (detected using Maximally Stable Extremal Regions)	Bag-of-words (SIFT) Text descriptors (based on term frequency–inverse document frequency technique)	SIFT features: 128	N/A
69	Medical Image Classification via 2D color feature based Covariance Descriptors, Cirujeda and Binefa (2015)	No	Covariance-based descriptor (1st and 2nd order color features)	11x11 matrix	various
70	View classification of medical x-ray images using PNN classifier, decision tree algorithm and SVM classifier, Ganesan and Subashini (2015)	segmentation by Expectation Maximization (EM) algorithm	Discrete Wavelet Transform (DWT)	N/A	N/A
71	Automatic muscle perimysium annotation using deep convolutional neural network, Sapkota et al. (2015)	No	CNN features	N/A	32x32
72	Convolutional Neural Networks for Subfigure Classification, Lyndon et al. (2015)	No	CNN features	N/A	cropped to 500px and resized to 160x160px
73	Automatic annotation of liver CT image: Imageclef med 2015, Nedjar et al. (2015)	Method 1: R Method 2: B	Method 1: Shape and Texture features i. from liver ii. from lesion Method 2: Gabor	Method 1: Liver descriptor 111, Lesion descriptor 223 Method 2: 40x76	various
2016					

74	Adapting content-based image retrieval techniques for the semantic annotation of medical images, Kumar et al. (2016)	R	C+T+S Computer Generated-**CoG** – 60 features (3D object space properties, Gabor, Tamura, Haar, pixel intensity)/ SIFT BOW	CoG: 446 SIFT BoFV: 1000 Total: 1446	various
75	Multi-View Probabilistic Classification of Breast Microcalcifications, Bekker et al. (2016)	R (ROI's extracted)	Curvelet features	14	N/A
2017					
76	Computer-aided medical image annotation: preliminary results with liver lesions in CT, Marvasti et al. (2017)	R	Location +Shape +Texture (58 features)	N/A	N/A
77	Deep learning based feature representation for automated skin histopathological image annotation, Zhang et al. (2017)	R (Ncut)	CNN features	N/A	200x150 (padding)
78	Automatic Scoring of Multiple Semantic Attributes With Multi-Task Feature Leverage: A Study on Pulmonary Nodules in CT Images, Chen et al. (2017)	R	Stacked denoising autoencoder (SDAE), CNN, general low-level Haar-like and histogram of oriented gradients (HoG) features	100 SDAE features 1728 CNN features 50 Haar-like features 1296 HoG features	28x28
2019					
79	Medical image classification using synergic deep learning, Zhang et al. (2019)	No	CNN features	N/A	224x224x3

Πίνακας 15. Σύγκριση μελετών με βάση την αναπαράσταση εικόνας. (Σημείωση: Τμηματοποίηση: 'No' για καμία τμηματοποίηση), "B" για τμηματοποίηση βάσει Block, "R" για τμηματοποίηση βάσει περιοχής (Region-based), Χαρακτηριστικά: "C" για χρώμα (color), "T" για υφή (texture) και "S" σχήμα (shape)).

Α/α	**Μελέτη**	**Ταξινομητής**	**Αριθμός κλάσεων**	**Ετικέτες ανά εικόνα**	**Σύνολο Δεδομένων**
2003					
1	Characterization of CT Liver Lesions Based on Texture Features and a Multiple Neural Network Classification Scheme, Mougiakakou et al. (2003)	Multiple Neural Networks (5 NN)	4 (C1 – C4) (C1:normal, C2:hepatic cyst, C3:hemagioma, C4:carcinoma)	N/A	147 ROI's from CT images (C1:76, C2:19, C3:28, C4:24) Training set:83, Validation set :32, Testing set: 32
2004					
2	Classification of Medical Images Using Non-linear Distortion Models, Keysers et al. (2004)	k-NN (k=1)	57 (9 to 2563 images per class)	1 (class number)	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
2005					
3	Automatic categorization of medical images for content-based	k-NN (k=1)	57 (9 to 2563 images per class)	1 (class number)	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)

	retrieval and data mining, Lehmann et al. (2005)				
4	Biomedical Image Classification with Random Subwindows and Decision Trees, Marée et al. (2005)	Decision Tree (boosting)	57 (9 to 2563 images per class)	1 (class number)	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
5	MIRACLE's naive approach to medical images annotation, Villena-Román et al. (2005)	k-NN (k=20)	57 (9 to 2563 images per class)	1 (class number)	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
6	Data Fusion of Retrieval Results from Different Media: Experiments at ImageCLEF 2005, Besancon and Millet (2005)	k-NN (k=3)	57 (9 to 2563 images per class)	1 (class number)	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
7	Supervised Machine Learning Based Medical Image Annotation and Retrieval in ImageCLEF 2005, Rahman et al. (2005)	SVM	57 (9 to 2563 images per class)	1 (class number)	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
2006					
8	The Use of MedGIFT and EasyIR for ImageCLEF 2005, Müller et al. (2006)	k-NN (k=5)	57 (9 to 2563 images per class)	1 (class number)	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
9	Combining Visual Features for Medical Image Retrieval and Annotation, Xiong et al. (2006)	SVM	57 (9 to 2563 images per class)	1 (class number)	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
10	Combining Text and Image Queries at ImageCLEF 2005, Chang et al. (2006)	k-NN (k=1, k=2)	57 (9 to 2563 images per class)	1 (class number)	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
11	Combining Textual and Visual Features for Cross-Language Medical Image Retrieval, Cheng et al. (2006)	SVM	57 (9 to 2563 images per class)	1 (class number)	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
12	Categorizing and Annotating Medical Images by Retrieving Terms Relevant to Visual Features, Ballesteros and Petkova (2006)	k-NN	57 (9 to 2563 images per class)	1 (class number)	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
2007					
13	CYU_IM@ImageCLEF 2007: Medical Image Annotation Task, Cheng and Yang (2007)	K-NN	116	1 (class number)	IMAGECLEF2007 (10000 training + 1000 validation + 1000 test X-ray images)

14	CINDI at ImageCLEF 2006: Image Retrieval and Annotation Tasks for the General Photographic and Medical Image Collections, Rahman et al. (2007)	SVM	116	1 (class number)	ImageCLEF 2006 Dataset (11000 fully classified radiographs, 10000 training+1000 validation + 1000 testing)
15	MedIC at ImageCLEF 2006: Automatic Image Categorization and Annotation Using Combined Visual Representations, Florea et al. (2007)	SVM (RFB kernel)	116	1 (class number)	ImageCLEF 2006 Dataset (11000 fully classified radiographs, 10000 training+1000 validation + 1000 testing)
16	Medical Image Annotation and Retrieval Using Visual Features, Liu et al. (2007b)	SVM	116	1 (class number)	ImageCLEF 2006 Dataset (11000 fully classified radiographs, 10000 training+1000 validation + 1000 testing)
17	A Refined SVM Applied in Medical Image Annotation, Qiu (2007)	SVM	116	1 (class number)	ImageCLEF 2006 Dataset (11000 fully classified radiographs, 10000 training+1000 validation + 1000 testing)
18	Medical Image Retrieval and Automated Annotation: OHSU at ImageCLEF 2006, Hersh et al. (2007)	multilayer perceptron	116	1 (class number)	ImageCLEF 2006 Dataset (11000 fully classified radiographs, 10000 training+1000 validation + 1000 testing)
19	Image Retrieval and Annotation Using Maximum Entropy, Deselaers et al. (2007)	maximum entropy classifier	116	1 (class number)	ImageCLEF 2006 Dataset (11000 fully classified radiographs, 10000 training+1000 validation + 1000 testing)
20	Baseline Results for the ImageCLEF 2006 Medical Automatic Annotation Task, Güld et al. (2007)	k-NN	116	1 (class number)	ImageCLEF 2006 Dataset (11000 fully classified radiographs, 10000 training+1000 validation + 1000 testing)
21	Grayscale Radiograph Annotation Using Local Relational Features, Setia et al. (2007)	multi-class SVM (one-vs-rest) (histogram intersection kernel)	116	1 (class number)	ImageCLEF 2006 Dataset (11000 fully classified radiographs, 10000 training+1000 validation + 1000 testing)
2008					
22	University and Hospitals of Geneva Participating at ImageCLEF2007, Zhou et al. (2007)	No training **voting strategies** (similar to k-NN classifier)	116	1 (class number)	IMAGECLEF2007 (10000 training + 1000 validation + 1000 test X-ray images)
23	Medical Image Retrieval and Automatic Annotation: OHSU at ImageCLEF 2007, Kalpathy-Cramer and Hersh (2007)	neural networks (a SVM as a second level classifier to discriminate the two most difficult classes)	116	1 (class number)	IMAGECLEF2007 (10000 training + 1000 validation + 1000 test X-ray images)
24	Speeding Up IDM without Degradation of Retrieval Quality,	k-NN	116	1 (class number)	IMAGECLEF2007

	Springmann and Schuldt (2007)				(10000 training + 1000 validation + 1000 test X-ray images)
25	Discriminative cue integration for medical image annotation, Tommasi et al. (2008a)	SVM multi-class One-vs-one One-vs-all (exponential x^2 kernel)	196	N/A	IMAGECLEF2007 (10000 training + 1000 validation + 1000 test X-ray images)
26	MIRACLE at ImageCLEFanot 2007: Machine Learning Experiments on Medical Image Annotation, Lana-Serrano et al. (2008a)	k-NN (k=10)	116	1 (class number)	IMAGECLEF2007 (10000 training + 1000 validation + 1000 test X-ray images)
27	Baseline Results for the ImageCLEF 2007 Medical Automatic Annotation Task Using Global Image Features, Güld and Deserno (2008)	weighted combinations of k-NN (5 k-NN classifiers)	116	1 (class number)	IMAGECLEF2007 (10000 training + 1000 validation + 1000 test X-ray images)
28	Automatic medical image categorization and annotation using LBP and MPEG-7 Edge Histograms, Tian et al. (2008)	SVM multi-class One-against-one	196	N/A	IMAGECLEF2007 dataset (10000 training + 1000 validation + 1000 test X-ray images)
29	Automatic Multilevel Medical Image Annotation and Retrieval, Mueen et al. (2008)	SVM (3 classifiers)	57	N/A	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
30	Grayscale medical image annotation using local relational features, Setia et al. (2008)	SVM (one-vs-all)/ 4 multi-class SVM classifiers/ Binary classification tree (BCT)	196	N/A	IMAGECLEF2007 dataset (10000 training + 1000 validation + 1000 test X-ray images)
31	CLEF2008 Image Annotation Task: an SVM Confidence-Based Approach, Tommasi et al. (2008b)	SVM	196	N/A	IMAGECLEF2008 12076 X-ray images) (10000 training + 1000 validation + 1000 test)
32	TAU MIPLAB at ImageClef 2008, Avni et al. (2008)	SVMs one-vs-one technique for multi-class classification/ 4 SVMs on ¼ scaled down images for each IRMA sub-code	196	N/A	IMAGECLEF2008 12076 X-ray images) (10000 training + 1000 validation + 1000 test)
33	MIRACLE at ImageCLEFannot 2008: Classification of Image Features for Medical Image Annotation, Lana-Serrano et al. (2008b)	variation of the classical k-NN (+ Relevance Feedback)	196	N/A	IMAGECLEF2008 12076 X-ray images) (10000 training + 1000 validation + 1000 test)
2009					

34	The MedGIFT Group at ImageCLEF 2008, Zhou et al. (2009)	k-NN for the entire code) / classification per axis (akNN) / dynamic kNN classification per axis (adkNN)	196	N/A	IMAGECLEF2008 12076 X-ray images) (10000 training + 1000 validation + 1000 test)
35	Baseline Results for the ImageCLEF 2008 Medical Automatic Annotation Task in Comparison over the Years, Güld et al. (2009)	k-NN (k=1, k=5)	196	N/A	IMAGECLEF2008 12076 X-ray images) (10000 training + 1000 validation + 1000 test)
36	Automated X-Ray Image Annotation - Single versus Ensemble of Support Vector Machines, Ünay et al. (2009)	multi-class SVM one-vs-all / Ensemble SVMs (4 SVM for each IRMA sub-code)	193	N/A	ImageCLEF-2009 Medical Annotation dataset (12677 radiographs for training+2000 test set)
37	Histopathology Image Classification Using Bag of Features and Kernel Functions, Caicedo et al. (2009)	SVM	18	N/A	1502 histopathology images split in 2 sets: Training and Validation, testing
2010					
38	Dense Simple Features for Fast and Accurate Medical X-Ray Annotation, Avni et al. (2009)	SVM	193	N/A	ImageCLEF-2009 Medical Annotation dataset (12677 radiographs for training+2000 test set)
39	ImageCLEF 2009 Medical Image Annotation Task: PCTs for Hierarchical Multi-Label Classification, Dimitrovski et al. (2010)	Ensembles of PCTs (predictive clustering trees) Bagging and Random Forests	193	N/A	ImageCLEF-2009 Medical Annotation dataset (12677 radiographs for training+2000 test set)
40	The MedGIFT Group at ImageCLEF 2009, Zhou et al. (2010)	k-NN (k=5) / SVM	193	N/A	ImageCLEF-2009 Medical Annotation dataset (12677 radiographs for training+2000 test set)
41	Automatic pathology annotation on medical images: a statistical machine translation framework, Gong et al. (2010)	EM based statistical machine translation method (IBM Model 1)	N/A	N/A	500 CT images (450 training+50 testing)
42	Hierarchical Medical image annotation using SVM-based Approaches, Amaral et al. (2010)	multi-class SVM (LIBSVM implementation)	116	N/A	IRMA 2007 (12000 radiographs, 11000 training+1000 testing)

43	Content-Based Retrieval and Classification of Ultrasound Medical Images of Ovarian Cysts, Sohail et al. (2010)	Fuzzy k-NN (k=21)	3 classes (simple cyst, endometrioma, teratoma)	1	478 Ultrasound images (Simple Cyst (187 images), Endometrioma (154 images),Teratoma (137 images))
44	ImageCLEF 2010 Modality Classification in Medical Image Retrieval Multiple feature fusion with normalized kernel function, Han and Chen (2010)	SVM (Joint Kernel Equal Contribution-JKEC)	8	N/A	ImageCLEF-2010 Database (2390 annotated modality images (CT: 314; GX: 355; MR: 299; NM: 204; PET: 285; PX: 330; US: 307; XR:296) for training, 2620 evaluation set)
2011					
45	Hierarchical annotation of medical images, Dimitrovski et al. (2011)	Ensemble (Random Forests) of Predictive Clustering Trees (PCTs) SVMs (One-against-All)	116 (ImageCLEF2007) 193 (ImageCLEF2008)	N/A	ImageCLEF2007 (12339 training+1353 testing) ImageCLEF2008 (12667 training+1733 testing)
46	MRI brain classification using support vector machine, Otman et al. (2011)	SVM (RBF kernel)	2 classes ('normal' and 'abnormal')	1 label ('1' for 'normal' and '0' for 'abnormal')	Training 32 MRI (22 abnormal and 10 normal) Testing 60 MRI
47	Classification of MRI brain images using k-nearest neighbor and artificial neural network, Rajini and Bhavani (2011)	forward back propagation artificial neural network (FP-ANN), k-NN	2 classes ('normal' and 'abnormal')	1 label ('1' for 'normal' and '0' for 'abnormal')	50 MRI (20 normal and 30 abnormal)
2012					
48	Automatic medical image annotation and keyword-based image retrieval using relevance feedback, Ko et al. (2012)	Random Forests (Ensemble of Decision Trees) 120 trees, tree depth=20	30	8-10	2400 X-ray images from ImageCLEF2007 (900 training + 1500 testing)
49	Annotation of medical images using the SURF descriptor, Wojnar, Pinheiro (2012)	k-NN / SVM (quadratic kernel function)	2 (Lung/No-Lung)	1	IRMA dataset 4241 test images (594 lungs 3647 no-lungs) 65 training (21 lungs positive training set+44 no-lungs negative training set)
50	Content-based medical image annotation and retrieval using perceptual hashing algorithm, Nagarajan and Saravanan (2012)	k-NN	8	N/A	IRMA dataset (1926 images)
51	Medical X-ray Image Classification Using Gabor-Based CS-Local Binary Patterns, Ghofrani et al. (2012)	SVM (one-against-one) (Gaussian kernel)	15	1	1169 x-ray images from IRMA dataset
52	Medical X-ray Images Classification Based on Shape Features and Bayesian Rule, Fesharaki et al. (2012)	Bayesian	28	1	4937 X-ray images (4375 for training, 562 for test)
53	Novel shape-texture feature extraction for medical x-ray image	SVM, Euclidean distance, and	21	1	4402 X-ray images from IRMA dataset

	classification, Mohammadi et al. (2012)	Probabilistic neural network			(315 training, 4087 for test)
54	Automatic Annotation of Radiological Observations in Liver CT Images, Gimenez et al. (2012)	logistic regression, L1-regularized logistic regression (LASSO)	30	-	79 CT images of liver lesions
55	Automated Image Annotation for Semantic Indexing and Retrieval of Medical Images, Krishna and Prasad (2012)	Decision Tree (C4.5)	15	N/A (multi-label)	MR-T2 axial brain images (172 training set, 150 testing)
2013					
56	Automatic image annotation and semantic based image retrieval for medical domain, Burdescu et al. (2013)	CMRM CRM	N/A	N/A	Gastrolab image gallery (www.gastrolab.net) 400 color images of digestive system
57	Using a bag of words for automatic medical image annotation with a latent semantic, Bouslimi et al. (2013)	LSA	5	1	Military Tunisian Hospital (1000 radiology images)
58	Automatic medical x-ray image classification using annotation, Zare et al. (2013a)	PLSA / SVM	39	N/A	ImageCLEF2007 (564 training)
59	Classification Using k Nearest Neighbor for Brain Image Retrieval, Charde and Lokhande (2013)	k-NN	10	1	500 CT scan images of brain
60	Medical X-ray Image Hierarchical Classification using a merging and splitting scheme in feature space, Fesharaki and Pourghassen (2013b)	Multi-Layer Perceptron / k-NN	18	N/A	2158 X-ray images (from IMAGECLEF 2005 database) 1439 for training, 719 for test
61	Automatic Classification of Medical X-Ray Images, Zare et al. (2013b)	SVM (RBF kernel) / k-NN (with K = 9)	116	1	ImageCLEF2007 Dataset
2014					
62	Medical image annotation and retrieval by using classification techniques, Abdulrazzaq et al. (2014)	k-NN SVM	57	N/A	IMAGECLEF 2005 annotation dataset (9000 training, 1000 test X-ray images)
63	A new fusion model for classification of the lung diseases using genetic algorithm, Bhuvaneswi et al. (2014)	k-NN, Decision tree, Multi layer perceptron Neural Networks (MLP-NN)	4 (bronchitis, emphysema, pleural effusion, normal lung)	1	400 Lung CT scan dataset (100 images per class)
64	MRI brain cancer classification using Support Vector Machine, Nandpuru et al. (2014)	SVM	2 ('normal', 'abnormal')	1	50 MRI (46 training)
65	Augmenting Multi-Instance Multi-label Learning with Sparse	Sparse Bayesian MIML Model	15	N/A	12700 biopsy images (2048 × 1536

	Bayesian Models for Skin Biopsy Image Analysis, Zhang et al. (2014)				pixels with $24k$ colors) divided into training set and test set at a ratio 3:7
66	Liver CT annotation via generalized coupled tensor factorization, Ermis and Cemgil (2014)	matrix factorization models using GCTF framework	73	N/A	ImageCLEF2014 Liver CT Annotation Dataset (50 datasets of 3D CT images, 10 datasets for testing)
67	Towards content-based image retrieval: From computer generated features to semantic descriptions of liver CT scans, Spanier and Joskowicz (2014)	4 different classifiers: linear discriminant analysis (LDA), logistic regression (LR), KNN (k=5), SVM (RBF kernel)	50	N/A	ImageCLEF2014 Liver CT Annotation Dataset (50 datasets of 3D CT images, 10 datasets for testing)
2015					
68	New approach for Automatic Medical Image Annotation Using the Bag-of-Words Model, Bouslimi and Akaichi (2015)	LSA	5 categories, 5 types of reports for each type of image	N/A	1000 medical reports including radiology images classified in five categories: Thorax, abdomen, lumbar spine, pelvis and skull.
69	Medical Image Classification via 2D color feature based Covariance Descriptors, Cirujeda and Binefa (2015)	sparse classification method	30	N/A	ImageCLEF 2015 Medical Classification task dataset
70	View classification of medical x-ray images using PNN classifier, decision tree algorithm and SVM classifier, Ganesan and Subashini (2015)	Probabilistic Neural Network (PNN), Decision Tree, SVM	6	N/A	X-ray images from IRMA database and Department of Radiology of the Raja Muthaiah Medical College and Hospital (30 images for each category)
71	Automatic muscle perimysium annotation using deep convolutional neural network, Sapkota et al. (2015)	CNN (backpropagation with stochastic gradient descent)	2	1	83 skeletal muscle images
72	Convolutional Neural Networks for Subfigure Classification, Lyndon et al. (2015)	CNN (Softmax classifier)	N/A	N/A	ImageCLEF 2015 Medical Classification task dataset
73	Automatic annotation of liver CT image: Imageclefmed 2015, Nedjar et al. (2015)	1. random forest (RF) classifier 2. retrieval (Hamming distance)	66	N/A	ImageCLEF2014 Liver CT Annotation Dataset (50 datasets of 3D CT images, 10 datasets for testing)
2016					
74	Adapting content-based image retrieval techniques for the semantic annotation of medical images, Kumar et al. (2016)	WNN (Weighted Nearest-Neighbour), Multi-class SVM	65	2 – 10	ImageCLEF2014 Liver CT Annotation Dataset (50 CT 3D images)
75	Multi-View Probabilistic Classification of Breast Microcalcifications, Bekker et al. (2016)	Expectation-Maximization Logistic-Regression algorithm	2	1	DDSM dataset with both CC and MLO views (1410 images 705 pairs of CC+MLO views (372 benign and 333 malignant)

2017					
76	Computer-aided medical image annotation: preliminary results with liver lesions in CT, Marvasti et al. (2017)	Multiple RBF-SVMs, Bayesian Network Model	30	N/A	123 3D CT images of liver lesions (including 8 types of lesion diagnoses)
77	Deep learning based feature representation for automated skin histopathological image annotation, Zhang et al. (2017)	Sparse Bayesian MIML model (S-MIMLGP)	15	N/A	Dataset D1 (12600 skin biopsy images and their diagnosis descriptions in plain text) Dataset D2 (2828 with 4 different skin tissues)
78	Automatic Scoring of Multiple Semantic Attributes With Multi-Task Feature Leverage: A Study on Pulmonary Nodules in CT Images, Chen et al. (2017)	Multi-task Linear Regression, Random Forest Regression, Composite of Regression and Classification	9	-	LIDC dataset (more than 1,000 thoracic CT scans) 2400 pulmonary nodules: 982, 554, 441, and 423 nodules with one, two, three and four annotation instances from different radiologists
2019					
79	Medical image classification using synergic deep learning, Zhang et al. (2019)	SDL model 2DCNN's (ResNet-50)+Synergic Network	30	-	ImageCLEF-2015, ImageCLEF-2016, ISIC-2016, and ISIC-2017 datasets

Πίνακας 16. Σύγκριση μελετών με βάση το μοντέλο ταξινόμησης

Τα συμπεράσματα αφορούν δύο βασικά τμήματα της διαδικασίας επισημείωσης, την αναπαράσταση της εικόνας και τη μέθοδο ταξινόμησης.

6.2.1. Αναπαράσταση εικόνας

Οι συλλογές ιατρικών εικόνων μπορούν να περιέχουν διάφορες εικόνες που λαμβάνονται με διαφορετικές τεχνικές απεικόνισης. Στα πλαίσια της παρούσας εργασίας μελετήθηκαν μέθοδοι ταξινόμησης και επισημείωσης εικόνων ακτίνων-Χ (ακτινογραφιών, μαστογραφιών), αξονικής (CT) και μαγνητικής (MRI) τομογραφίας, που είναι εικόνες αποχρώσεων του γκρι (grayscale) αλλά και έγχρωμες ιστοπαθολογικές εικόνες.

Η ακρίβεια της ταξινόμησης της ιατρικής εικόνας εξαρτάται κυρίως από την εξαγωγή χαρακτηριστικών. Όσο καλύτερη είναι η διακριτική ισχύς των εξαγόμενων χαρακτηριστικών τόσο καλύτερη είναι η απόδοση της ταξινόμησης (Mueen et al., 2008). Διαφορετικές τεχνικές εξαγωγής χαρακτηριστικών είναι σε θέση να συλλάβουν διάφορες πτυχές μιας εικόνας (π.χ. υφή, σχήματα, κατανομή χρώματος κ.λπ.).

Οι περισσότερες από τις μορφές ιατρικής απεικόνισης παράγουν ιατρικές εικόνες που είναι συνήθως σε αποχρώσεις του γκρι. Επομένως, τα χαρακτηριστικά χρώματος όπως τα ιστογράμματα χρώματος (color histograms) και τα χρωματικά διαγράμματα συσχέτισης

(color correlograms) δεν θα ήταν αποτελεσματικά στην ανάλυση του οπτικού περιεχομένου σε ιατρικές εικόνες. Χαρακτηριστικά χρώματος μπορούν να αξιοποιηθούν μόνο στις περιπτώσεις που έγχρωμες φωτογραφίες χρησιμοποιούνται για διάγνωση (Akgul et al., 2011). Έτσι στην πλειονότητα των μελετών, η περιγραφή της εικόνας πραγματοποιείται με την εξαγωγή πληροφορίας σε τρία επίπεδα, σε επίπεδο εικονοστοιχείου, η τιμή των pixel ως αυτόνομο χαρακτηριστικό, σε επίπεδο υφής και επίπεδο σχήματος, τοπικά ή συνολικά στην εικόνα.

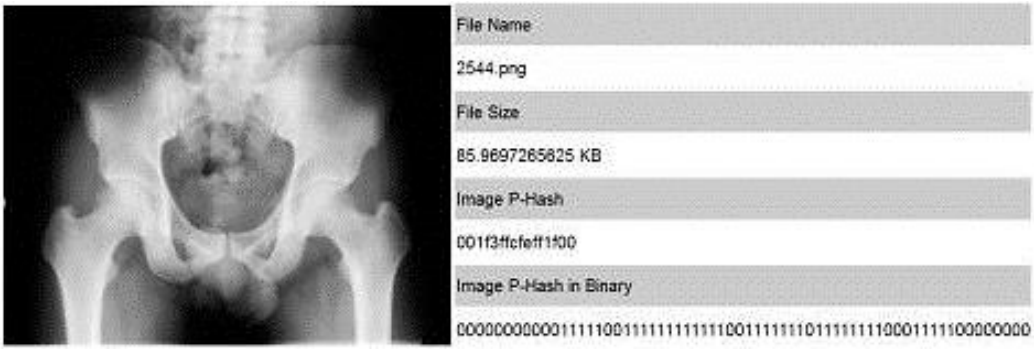
6.2.1.1. Αναπαράσταση σε επίπεδο εικονοστοιχείου

Η πιο απλή προσέγγιση στην ταξινόμηση εικόνων είναι η άμεση χρήση των τιμών των εικονοστοιχείων της εικόνας ως χαρακτηριστικών. Οι εικόνες κλιμακώνονται σε ένα μικρότερο μέγεθος και αντιπροσωπεύονται από ένα διάνυσμα χαρακτηριστικών που περιέχει τιμές εικονοστοιχείων εικόνας. Έχει αποδειχθεί ότι για την ταξινόμηση και την ανάκτηση των ιατρικών ακτινογραφιών, αυτή η μέθοδος αποτελεί μια βασική προσέγγιση. Διάφορες μελέτες (Tommasi et al., 2008a, Dimitrovski et al., 2011) προτείνουν την σμίκρυνση της αρχικής εικόνας σε ένα μέγεθος 32x32 pixels και τη χρήση των τιμών των 1024 pixel ως χαρακτηριστικά για την αναπαράσταση της εικόνας σε συνδυασμό πάντα με άλλα χαρακτηριστικά.

Ωστόσο, οι Marée et al. (2005) χρησιμοποιούν τις τιμές των εικονοστοιχείων ως αποκλειστικό περιγραφέα της εικόνας. Στη μέθοδό τους περιγράφουν την εικόνα με συνδυασμό ενός μεγάλου αριθμού τετραγώνων patches τυχαία εξαγόμενων από εικόνες (“random subwindows”), θεωρώντας ότι παρέχουν μια πλούσια αναπαράσταση εικόνων που αντιστοιχούν σε διάφορες επικαλυπτόμενες περιοχές τόσο τοπικές όσο και συνολικές. Επιπλέον, για να αποφευχθεί η απόρριψη χρήσιμων πληροφοριών και για να μπορέσουν να ταξινομήσουν ένα μεγάλο αριθμό κλάσεων, οι τιμές των εικονοστοιχείων αυτών των υποπεριοχών χρησιμοποιούνται για περιγραφή της εικόνας, κανονικοποιημένες ώστε να αποτελούν μια ισχυρή στις αλλαγές κλίμακας αναπαράσταση.

Οι Nagarajan and Saravanan (2012) προτείνουν τη χρήση ενός αλγορίθμου κατακερματισμού (hashing) για την αναπαράσταση της εικόνας που είναι ένας αξιόπιστος και ισχυρός αλγόριθμος που χρησιμοποιείται ευρέως για την αναγνώριση περιεχομένου. Στο σύστημά τους χρησιμοποιούν έναν αλγόριθμο κατακερματισμού (P-hash) ο οποίος παίρνει τις τιμές

των εικονοστοιχείων μιας εικόνας για να δημιουργήσει την τιμή (δυαδική συμβολοσειρά) κατακερματισμού. Αρχικά η εικόνα κλιμακώνεται σε καθορισμένο μικρό μέγεθος και στη συνέχεια υπολογίζεται η μέση τιμή από όλα τα εικονοστοιχεία της εικόνας. Κάθε bit στη τιμή κατακερματισμού υπολογίζεται συγκρίνοντας το μέσο όρο των συνολικών εικονοστοιχείων της εικόνας με την τιμή κάθε εικονοστοιχείου της. Έτσι το hashbit ορίζεται ίσο με 1 αν η τιμή του εικονοστοιχείου είναι μεγαλύτερη από τη μέση τιμή και 0 στην αντίθετη περίπτωση (εικόνα 30).



Εικόνα 30. Παράδειγμα δημιουργίας της τιμής hash μίας εικόνας ακτίνων-Χ (Nagarajan and Saravanan, 2012)

6.2.1.2. Χαρακτηριστικά υφής

Η υφή είναι ιδιαίτερα σημαντική, επειδή είναι δύσκολο να ταξινομηθούν οι ιατρικές εικόνες με τη χρήση πληροφοριών σχήματος ή γκρίζου επιπέδου. Η αποτελεσματική αναπαράσταση της υφής είναι απαραίτητη για να γίνει διάκριση μεταξύ εικόνων που έχουν ληφθεί με την ίδια μέθοδο (modality) και διάταξη (Dimitrovski et al., 2011)

Η υφή περιέχει σημαντικές πληροφορίες σχετικά με τη δομική διάταξη των επιφανειών μιας εικόνας. Ο πίνακας συνεμφάνισης επιπέδου γκρίζου (Grey Level Co-occurance Matrix - GLCM) είναι μία από τις στατιστικές τεχνικές στην ανάλυση της υφής για τον υπολογισμό των ιδιοτήτων της εικόνας που σχετίζονται με τα στατιστικά στοιχεία δεύτερης τάξης της εικόνας και εισήχθη από τους Haralick et al. (1973). Ο πίνακας συν-εμφάνισης κατασκευάζεται με τη λήψη πληροφοριών σχετικά με τον προσανατολισμό και την απόσταση μεταξύ των εικονοστοιχείων (Fesharaki and Pourghassen, 2013). Διάφορα στατιστικά μέτρα υφής μπορούν να υπολογιστούν απευθείας από τον πίνακα (Haralick et al., 1973). Σε αρκετές μελέτες, εξάγονται τα χαρακτηριστικά υφής του Haralick όπως η ενέργεια, η εντροπία, η τραχύτητα, η ομοιογένεια, η αντίθεση κ.λπ., από μια τοπική γειτονιά της εικόνας

(Mougiakakou et al., 2003, Rahman et al., 2005, Rahman et al., 2007, Florea et al., 2007, Mueen et al., 2008, Sohail et al., 2010, Fesharaki and Pourghassen, 2013, Zare et al., 2013b, Abdulrazzaq et al., 2014, Nandpuru et al., 2014).

Οι Krishna and Prasad (2012) χρησιμοποιούν αποκλειστικά συνολικά χαρακτηριστικά υφής για την περιγραφή εικόνων MRI (εντροπία, ενέργεια και αντίθεση) που εξάγονται από ένα σύνολο πινάκων GLCM που υπολογίζονται για διάφορες γωνιακές σχέσεις και αποστάσεις, καθώς θεωρούν ότι ανακριβής τμηματοποίηση μπορεί να οδηγήσει σε ανακριβή αναπαράσταση χαρακτηριστικών και επομένως να μειώσει την ακρίβεια επισημείωσης.

Τα τοπικά δυαδικά πρότυπα (Local Binary Pattern - LBP) είναι μια από τις καλύτερες αναπαραστάσεις του περιεχομένου υφής στις εικόνες. Η βασική ιδέα πίσω από την προσέγγιση LBP είναι να χρησιμοποιήσει τις πληροφορίες σχετικά με την υφή από μια τοπική γειτονιά. Είναι αμετάβλητα στις μονοτονικές αλλαγές σε εικόνες γκρι κλίμακας και υπολογίζονται γρήγορα. Επιπλέον, είναι σε θέση να ανιχνεύουν διαφορετικά μικρο-πρότυπα, όπως άκρα, σημεία και σταθερές περιοχές (Dimitrovski et al., 2011). Η εικόνα σαρώνεται με τον τελεστή LBP εικονοστοιχείο προς εικονοστοιχείο και οι έξοδοι συσσωρεύονται σε ένα διακριτό ιστόγραμμα. Ωστόσο, δεν περιέχουν όλοι οι κώδικες LBP την ίδια πληροφορία. Ορισμένοι κώδικες LBP καταγράφουν θεμελιώδεις ιδιότητες της υφής και ονομάζονται ομοιόμορφα πρότυπα επειδή αποτελούν τη συντριπτική πλειοψηφία, μερικές φορές πάνω από το 90%, όλων των προτύπων που υπάρχουν στις παρατηρούμενες υφές. Αυτά τα πρότυπα έχουν ως κοινό χαρακτηριστικό μια ομοιόμορφη κυκλική δομή που περιέχει πολύ λίγες χωρικές μεταβολές. Λειτουργούν ως πρότυπα για μικρο-δομές όπως φωτεινό σημείο, επίπεδη περιοχή ή σκοτεινό σημείο. Παραλλαγές των τοπικών δυαδικών προτύπων χρησιμοποιούνται ως περιγραφείς υφής σε πολλές μελέτες (Setia et al., 2008, Tian et al., 2008, Tommasi et al., 2008b, Ünay et al., 2009, Dimitrovski et al., 2011, Ko et al., 2012, Ghofrani et al., 2012).

Σε αρκετές μελέτες για την αναπαράσταση της υφής χρησιμοποιούνται τα χαρακτηριστικά που εισήγαγαν ο Tamura και ο Castelli (Lehmann et al., 2005, Ballesteros and Petkova, 2006, Güld et al., 2007, Lana-Serrano et al., 2008a, Güld and Deserno, 2008, Lana-Serrano et al., 2008b, Güld et al., 2009, Amaral et al., 2010).

Μια άλλη συνήθης πρακτική για την δημιουργία περιγραφέων με βάση την υφή είναι με τη χρήση μετασχηματισμών στο πεδίο των συχνοτήτων, όπως ο Διακριτός Μετασχηματισμός wavelets (Liu et al., 2007b, Otman et al., 2011, Rajini and Bhavani, 2011, Ganesan and Subashini, 2015) και τα φίλτρα Gabor (Villena-Roman et al., 2005, Müller et al, 2006, Cheng et al., 2006, Kalpathy-Cramer and Hersch, 2008, Zhou et al., 2009, Gimenez et al. 2012, Bhuvaneswi et al., 2014) σε τοπικά μπλοκ εικόνας.

6.2.1.3. Χαρακτηριστικά σχήματος και θέσης

Το σχήμα παρέχει γεωμετρικές πληροφορίες ενός αντικειμένου στην εικόνα, οι οποίες δεν αλλάζουν ακόμα και όταν αλλάζει η θέση, η κλίμακα και ο προσανατολισμός του αντικειμένου. Ο τελεστής ακμών Canny όπως και ο τελεστής Sobel χρησιμοποιούνται για τη δημιουργία ιστογραμμάτων ακμών (Keysers et al., 2004, Besancon and Millet, 2005, Rahman et al, 2005, Springmann and Schldt, 2008, Mueen et al., 2008, Charde and Lokhande, 2013, Zare et al., 2013b). Σε άλλες περιπτώσεις χρησιμοποιείται ο περιγραφέας Edge Histogram Descriptor του προτύπου MPEG-7 (Rahman et al., 2007, Tian et al., 2008, Amaral et al., 2010)

Χαρακτηριστικά θέσης χρησιμοποιούνται από κοινού με χαρακτηριστικά σχήματος για την επισημείωση περιοχών αλλοιώσεων σε εικόνες CT. Οι Gong et al. (2010) ανέπτυξαν ένα στατιστικό πλαίσιο για την επισημείωση περιοχών ενδιαφέροντος, όπως αιματωμάτων σε εικόνες CT του εγκεφάλου. Μετά την τμηματοποίηση των περιοχών των αιματωμάτων χρησιμοποιούν χαρακτηριστικά σχήματος και θέσης που θεωρούνται ιδιαίτερα σημαντικά για τη διάκριση των αιματωμάτων μεταξύ τους, όπως την εκκεντρότητα (ο λόγος της απόστασης μεταξύ των εστιών της έλλειψης και του μεγάλου μήκους του άξονα), τη στερεότητα (η αναλογία των εικονοστοιχείων στο κυρτό μέρος της καμπύλης που βρίσκονται επίσης στο ROI), την έκταση (το ποσοστό των εικονοστοιχείων στο πλαίσιο οριοθέτησης που βρίσκονται επίσης στο ROI) και το κρανίο (αν το ROI είναι δίπλα στο κρανίο ή όχι).

Οι Marvasti et al. (2017) σε μία πρόσφατη μελέτη για την επισημείωση αλλοιώσεων στο ήπαρ με 30 έννοιες διατυπωμένες στην οντολογία OLIRA (Ontology of Liver for Radiologists) χρησιμοποιούν 58 χαρακτηριστικά χαμηλού επιπέδου για την περιγραφή της θέσης, του σχήματος (όγκος, έκταση, στερεότητα, σφαιρικότητα κ.ά.) και της υφής (Gabor, Tamura, Haralick, Fourier Descriptors κ.ά.) του ήπατος και των αλλοιώσεών του χωρισμένα σε δύο ομάδες: συνολικά χαρακτηριστικά που συνοψίζουν τις γενικές οπτικές ιδιότητες του ήπατος,

των αρτηριών και των αλλοιώσεων και παθολογικά χαρακτηριστικά που αντανακλούν λεπτομερή επίπεδα οπτικής πληροφορίας σχετικά με ξεχωριστές αλλοιώσεις. Σε παλαιότερη μελέτη των Gimenez et al. (2012) μετά την τμηματοποίηση των ROI's ένα πλήθος χαρακτηριστικών έντασης, υφής (Gabor, Haar, εντροπία, διακύμανση κ.ά.), ακμών (ιστόγραμμα ακμών) και σχήματος συνενώνονται σε ένα διάνυσμα χαρακτηριστικών (διάστασης 431) για την ταξινόμηση των αλλοιώσεων σε 30 σημασιολογικές κατηγορίες.

Οι Kumar et al. (2016) για το ίδιο έργο και στα πλαίσια του διαγωνισμού επισημείωσης CT εικόνων του ήπατος ImageCLEF 2014, χρησιμοποίησαν ένα μέρος από 60 καθιερωμένα χαρακτηριστικά (συνολικής διάστασης 458) που παρείχε η βάση δεδομένων ImageCLEF 2014, τα οποία αναφέρονται ως CoG (Computer Generated features), για την περιγραφή της θέσης, του σχήματος και της υφής του ήπατος, των αρτηριών και των αλλοιώσεων αλλά και τον περιγραφέα SIFT, σε ένα πλαίσιο αναπαράστασης με βάση το σάκο οπτικών λεξεων (Bag-of-Words).

6.2.1.4. Αναπαράσταση εικόνας με βάση το σάκο χαρακτηριστικών

Τα τοπικά χαρακτηριστικά εικόνας είναι θεμελιώδη για την ερμηνεία της εικόνας. Ενώ τα συνολικά χαρακτηριστικά διατηρούν πληροφορίες για ολόκληρη την εικόνα, τα τοπικά χαρακτηριστικά καταγράφουν τις λεπτομέρειες. Επομένως, είναι πιο διακριτικά όσον αφορά το πρόβλημα της μεταβλητότητας μεταξύ και εντός των κλάσεων που αποτελεί πρόκληση κατά την αυτόματη επισημείωση των ιατρικών εικόνων (Dimitrovski et al., 2011). Σε πολλές μελέτες ο περιγραφέας SIFT χρησιμοποιείται για την εξαγωγή τοπικών χαρακτηριστικών της εικόνας (Qiu, 2007, Kalpathy-Cramer et al., 2008, Tommasi et al., 2008b, Anvi et al., 2009, Dimitrovski et al. 2010, Zhou et al., 2010, Zare et al., 2013a) ενώ υπάρχουν και κάποιες στις οποίες χρησιμοποιούνται παρόμοιοι περιγραφείς, όπως ο SURF (Wojnar and Pinheiro, 2012) και ο GIST (Amaral et al., 2010).

Η αναπαράσταση με βάση το σάκο χαρακτηριστικών (σάκο οπτικών λέξεων) είναι μια προσαρμοστική προσέγγιση για τη μοντελοποίηση της δομή της εικόνας με έναν εύρωστο τρόπο. Σε αντίθεση με την τμηματοποίηση της εικόνας, αυτή η προσέγγιση δεν επιχειρεί να εντοπίσει πλήρη αντικείμενα μέσα στις εικόνες, κάτι που μπορεί να είναι πιο δύσκολο από την ίδια την ταξινόμηση της εικόνας. Αντ' αυτού, η προσέγγιση του σάκου χαρακτηριστικών αναζητά μικρές χαρακτηριστικές περιοχές εικόνας, γύρω από σημεία ενδιαφέροντος που

επιτρέπουν την αναπαράσταση σύνθετων περιεχομένων της εικόνας χωρίς να μοντελοποιούνται ρητά αντικείμενα και οι σχέσεις τους (Caicedo et al., 2009). Η προσέγγιση του σάκου των χαρακτηριστικών επιτρέπει τον εντοπισμό οπτικών προτύπων που σχετίζονται με ολόκληρη τη συλλογή εικόνων, δηλαδή, τα μοτίβα που χρησιμοποιούνται στην αναπαράσταση μίας εικόνας, προέρχονται από την ανάλυση των προτύπων στην πλήρη συλλογή.

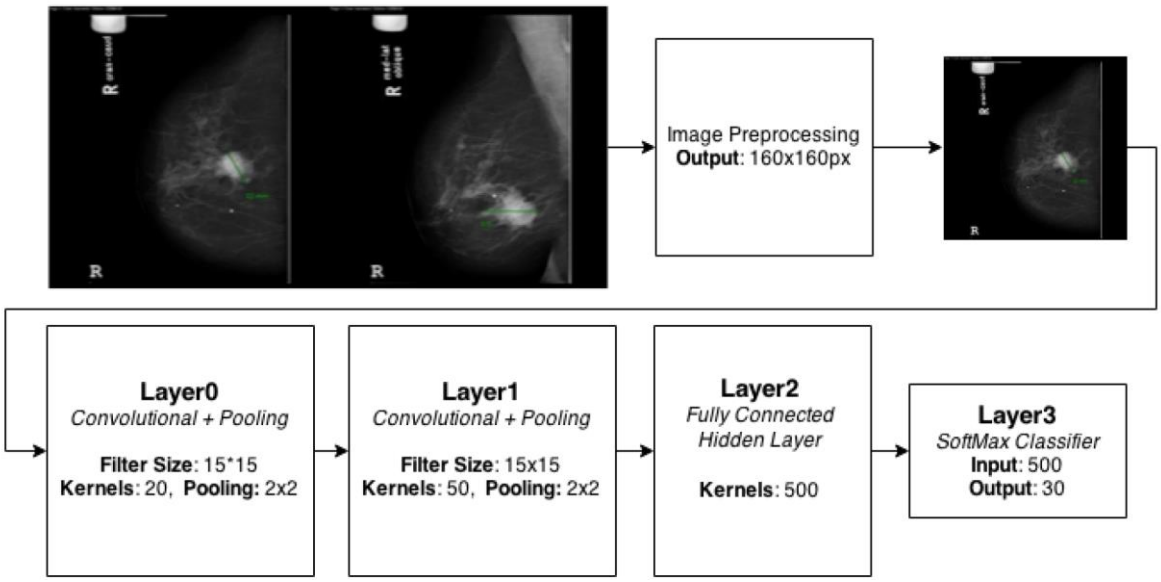
Στις μελέτες που εξετάστηκαν ακολουθούνται διαφορετικές προσεγγίσεις σχεδιασμού ενός περιγραφέα εικόνας στο πλαίσιο του σάκου χαρακτηριστικών καθεμία από τις οποίες οδηγεί σε διαφορετικές αναπαραστάσεις εικόνας που μπορεί να είναι περισσότερο ή λιγότερο διακριτικές. Η επικρατέστερη προσέγγιση είναι η κατασκευή του σάκου χαρακτηριστικών σε σημεία που εξάγονται με τον SIFT (Tommasi et al., 2008a, Caicedo et al., 2009, Dimitrovski et al., 2011, Bouslimi et al., 2013, Zare et al., 2013a, Bouslimi and Akaichi, 2015) ή τον SURF (Amaral et al., 2010).

6.2.1.5. Τεχνικές Βαθιάς μάθησης για την εξαγωγή χαρακτηριστικών

Όλες οι προαναφερόμενες μελέτες που έχουν ως στόχο την ταξινόμηση των ιατρικών εικόνων, βασίζονται στην εξαγωγή χειροποίητων (handcrafted) χαρακτηριστικών. Παρά την επιτυχία των μεθόδων αυτών, η εξαγωγή των βέλτιστων χαρακτηριστικών για μία εξειδικευμένη εργασία ταξινόμησης, είναι συνήθως μία δύσκολη διαδικασία. Τα τελευταία χρόνια, τεχνικές βαθιάς μάθησης, ειδικότερα βαθιά συνελικτικά νευρωνικά δίκτυα (Deep Convolutional Neural Networks - DCNN) έχουν οδηγήσει σε σημαντικά επιτεύγματα στην ταξινόμηση και την τμηματοποίηση ιατρικών εικόνων. Οι Zhang et al. (2019) προτείνουν, για το απαιτητικό έργο της ταξινόμησης ιατρικών εικόνων και την αντιμετώπιση του προβλήματος της μεγάλης διακύμανσης των εικόνων εντός μίας κλάσης και ταυτόχρονα της μεγάλης ομοιότητας μεταξύ διαφορετικών κλάσεων, τη χρήση ενός ζεύγους DCNNs που έχουν τη δυνατότητα να μαθαίνουν αμοιβαία το ένα από το άλλο. Για τη μελέτη τους χρησιμοποιούν δύο προ-εκπαιδευμένα residual neural networks 50 επιπέδων (ResNet-50) αν και σημειώνουν ότι οποιοδήποτε DCNN, όπως το AlexNet των Krizhnevsky et al. (2012), το VGGNet ή το GoogLeNet θα μπορούσαν να χρησιμοποιηθούν στο μοντέλο τους.

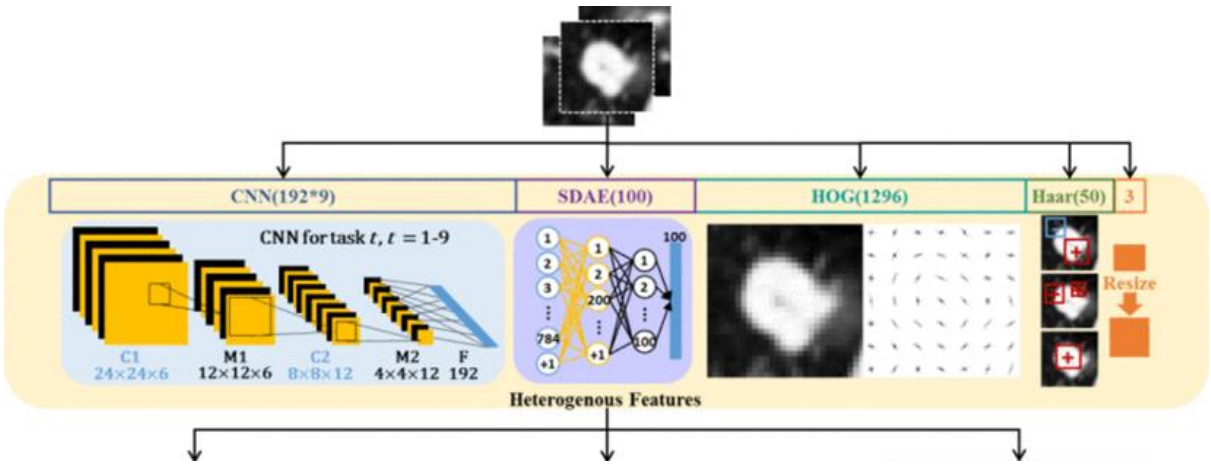
Στο ίδιο μήκος κύματος οι Lyndon et al. (2015), υπογραμμίζουν ότι η εγγενής ικανότητα των συνελικτικών νευρωνικών δικτύων (CNNs) να εξάγουν αυτόματα ιεραρχικές αναπαραστάσεις

από ακατέργαστα δεδομένα μπορεί να έχει μεγάλες δυνατότητες για τη ανάκτηση ιατρικών εικόνων. Συμμετέχοντας στην εργασία ταξινόμησης υποεικόνων του διαγωνισμού ImageCLEF 2015, προτείνουν ένα πλαίσιο βαθιάς μάθησης το οποίο μαθαίνει αναπαραστάσεις υψηλού επιπέδου εικόνων από διαφορετικές κατηγορίες απεικόνισης (modalities) και το χρησιμοποιούν για να ταξινομήσουν τον τρόπο απεικόνισης κάθε υποεικόνας. Η αρχιτεκτονική του CNN με το οποίο πειραματίζονται, αποτελεί μία απλοποιημένη εκδοχή του δικτύου LeNet-5 των LeCun et al. (1998). Το τροποποίησαν ώστε να δέχεται στην είσοδο μεγαλύτερες εικόνες και να έχει ως έξοδο περισσότερες κατηγορίες. Το δίκτυο αποτελείται από δύο συνελικτικά στρώματα ακολουθούμενα το καθένα από ένα στρώμα συγκέντρωσης και ένα πλήρως συνδεδεμένο κρυφό στρώμα, όπως αποτυπώνεται στην εικόνα 31. Τα χαρακτηριστικά που είναι η έξοδος του κρυφού στρώματος, χρησιμοποιούνται για την ταξινόμηση από ένα Softmax στρώμα ταξινόμησης.



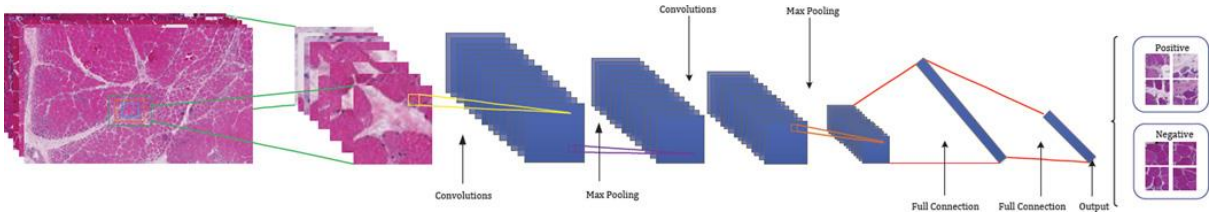
Εικόνα 31. Η αρχιτεκτονική του CNN που προτείνουν οι Lyndon et al. (2015) για την ταξινόμηση ιατρικών εικόνων προερχόμενων από διαφορετικές μεθόδους απεικόνισης.

Οι Chen et al. (2017) χρησιμοποιούν χαρακτηριστικά από μοντέλα βαθιάς μάθησης από κοινού με «χειροποίητα» χαρακτηριστικά για την περιγραφή πνευμονικών οζιδίων σε εικόνες CT με εννέα σημασιολογικές ετικέτες. Προτείνουν τη χρήση ενός multi-task πλαισίου μάθησης που αναμιγνύει ετερογενή χαρακτηριστικά που εξάγονται από ένα στοιβαγμένο αυτόματο κωδικοποιητή (SAE), ένα συνελικτικό νευρωνικό δίκτυο αλλά και χειροκίνητα (Haar, HoG descriptor). Η αρχιτεκτονική των δύο πλαισίων βαθιάς μάθησης που χρησιμοποιούν, παρουσιάζεται στην εικόνα 32.



Εικόνα 32. Η διαδικασία εξαγωγής χαρακτηριστικών που προτείνουν οι Chen et al. (2017).

Ένα συνελικτικό νευρωνικό δίκτυο ίδιας αρχιτεκτονικής με αυτό των Chen et al. (2017), με δύο συνελικτικά στρώματα ακολουθούμενα από στρώματα συγκέντρωσης, ένα πλήρως συνδεδεμένο στρώμα και ένα softmax στρώμα εξόδου χρησιμοποιούν οι Sapkota et al. (2015) για την αυτόματη ταξινόμηση έγχρωμων εικόνων σκελετικών μυών σε μία από δύο κατηγορίες μυϊκών ασθενειών, δερματομυοσίτιδα (DM) και πολυμυοσίτιδα (PM).



Εικόνα 33. Η αρχιτεκτονική του CNN δικτύου που χρησιμοποιούν οι Sapkota et al. (2015).

Οι Zhang et al. (2017) προτείνουν ένα πλαίσιο επισημείωσης για την αυτοματοποιημένη ανάλυση εικόνας βιοψίας δέρματος, η οποία χρησιμοποιεί ένα μοντέλο βαθιάς μάθησης χωρίς επίβλεψη για την αναπαράσταση των χαρακτηριστικών της εικόνας. Η μέθοδός τους τμηματοποιεί αρχικά μια ιστοπαθολογική εικόνα σε πλήρως διαχωρισμένες περιοχές με μια αυτο-προσαρμοστική στρατηγική, έτσι ώστε να μην ορίζει ρητά τον αριθμό των περιοχών που πρόκειται να δημιουργηθούν. Στη συνέχεια ένα συνελικτικό νευρωνικό δίκτυο (CNN) εφαρμόζεται σε κάθε περιοχή που παράγεται, για να μάθει την αναπαράσταση χαρακτηριστικών, που ονομάζεται επίσης κωδικοποίηση περιοχής (region encoding).

6.2.2. Μέθοδοι Επισημείωσης

Εάν ένα σύνολο ελεγχόμενου λεξιλογίου, όπως για παράδειγμα ο κώδικας IRMA (Lehmann et al., 2003) χρησιμοποιείται για να επισημειωθεί το περιεχόμενο μιας εικόνας, η διαδικασία της επισημείωσης εικόνας μπορεί να θεωρηθεί ως πρόβλημα επιβλεπόμενης ταξινόμησης· ένα σύνολο γνωστών παραδειγμάτων μπορεί να χρησιμοποιηθεί για την ανάπτυξη ενός ταξινομητή ή ενός μοντέλου για την πρόβλεψη άγνωστων δεδομένων. Αυτές οι ετικέτες κλάσης αντιστοιχούν στις λέξεις που περιέχονται στο ελεγχόμενο λεξιλόγιο. Λόγω της φύσης της εικόνας προς επισημείωση, έχουν προταθεί διάφορες εποπτευόμενες προσεγγίσεις ταξινόμησης για την επίτευξη του στόχου αυτού.

Οι μέθοδοι αυτόματης επισημείωσης ιατρικής εικόνας (Computer-aided Medical Image Annotation - CMIA) μπορούν ευρέως να κατηγοριοποιηθούν ως ανεξάρτητες (independent) και αλληλοεξαρτώμενες (inter-dependent) μέθοδοι. Οι ανεξάρτητες μέθοδοι υπολογίζουν την τιμή (επισημείωσης) για κάθε έννοια (σημασιολογική περιγραφή) χρησιμοποιώντας ανεξάρτητα χαρακτηριστικά εικόνας χαμηλού επιπέδου, ενώ οι αλληλεξαρτώμενες μέθοδοι εκμεταλλεύονται και τις αλληλεπιδράσεις των (σημασιολογικών) εννοιών (Marvasti et al. 2017).

Οι κυρίαρχες ανεξάρτητες μέθοδοι CMIA είναι ταξινομητές πολλαπλών κατηγοριών όπως οι μηχανές διανυσμάτων υποστήριξης (SVMs), οι τεχνικές που βασίζονται στην ανάκτηση παρόμοιων εικόνων, η λογιστική παλινδρόμηση, τα δέντρα απόφασης και τα τεχνητά νευρωνικά δίκτυα.

Οι μέθοδοι που επιδιώκουν να εκμεταλλευτούν τις αλληλεξαρτήσεις, μπορούν να ομαδοποιηθούν σε ιεραρχικούς ταξινομητές (hierarchical classifiers) και μάθηση πολλαπλών εργασιών (Multi-Task Learning - MTL). Στην πρώτη περίπτωση, χτίζεται μια ιεραρχία των εννοιών και ξεχωριστοί ταξινομητές εκπαιδεύονται για κάθε έννοια χρησιμοποιώντας τις εικόνες που σχετίζονται με αυτή την έννοια. Από την άλλη πλευρά, οι μέθοδοι μάθησης πολλαπλών εργασιών προσεγγίζουν ταυτόχρονα πολλαπλά αλληλεξαρτώμενα καθήκοντα, δηλαδή, σημασιολογικές επισημειώσεις, στην περίπτωσή μας.

Οι μηχανές διανυσμάτων υποστήριξης χρησιμοποιούνται ευρέως στην αναγνώριση και ταξινόμηση προτύπων λόγω της ισχυρής μαθηματικής τους θεμελίωσης, της δυνατότητας

γενίκευσης, της ακρίβειας ταξινόμησης και της αποτελεσματικότητας σε υψηλής διάστασης χώρους όπου η διαστασιμότητα είναι υψηλότερη από το μέγεθος του δείγματος. Παράλληλα είναι ευέλικτοι καθώς μπορεί να χρησιμοποιήσουν διαφορετικές συναρτήσεις πυρήνα για διαφορετικές εργασίες ταξινόμησης. Τα χαρακτηριστικά αυτά έχουν καταστήσει τους SVMs ιδιαίτερα δημοφιλείς που χρησιμοποιούνται ως ταξινομητές σε πάρα πολλές από τις μελέτες που εξετάστηκαν (Rahman et al., 2005, Xiong et al., 2006, Cheng et al., 2006, Rahman et al., 2007, Florea et al., 2007, Liu et al., 2007b, Qiu, 2007, Setia et al., 2007, Tommasi et al., 2008a, Tian et al., 2008, Mueen et al., 2008, Setia et al., 2008, Tommasi et al., 2008b, Avni et al. 2008, Ünay et al., 2009, Avni et al., 2009, Amaral et al., 2010, Han and Chen, 2010, Otman et al., 2011, Wojnar and Pinheiro, 2012, Ghofrani et al., 2012, Zare et al., 2013b, Nandpuru et al., 2014).

Σε κάποιες περιπτώσεις, που αφορούσαν ταξινόμηση σε δύο κλάσεις, χρησιμοποιήθηκε ο κλασικός δυαδικός ταξινομητής SVM. Οι Otman et al. (2011) ταξινομούν εικόνες MRI του εγκεφάλου σε δύο κλάσεις (“normal”, “abnormal”) χρησιμοποιώντας SVM με RBF πυρήνα, ενώ οι Nadpuru et al. (2014), για την ίδια εργασία δοκιμάζουν διαφορετικούς πυρήνες (γραμμικό, τετραγωνικό και πολυωνυμικό). Οι Wojnar and Pinheiro (2012) χρησιμοποιούν και αυτοί SVM πειραματιζόμενοι με διαφορετικούς πυρήνες (γραμμικό, τετραγωνικό, RBF, πολυωνυμικό 3^{ης} τάξης) για ταξινόμηση ακτινογραφιών πνευμόνων σε δύο κλάσεις (“lung”, “non-lung”). Οι περισσότερες μελέτες εξετάζουν το πρόβλημα της ταξινόμησης σε πολλές κλάσεις. Τα προβλήματα πολλαπλών κλάσεων συνήθως επιλύονται με τη χρήση μονάδων πολλαπλών SVMs. Οι Caicedo et al. (2009) ταξινομούν εικόνες ιστοπαθολογίας σε δεκαοκτώ κατηγορίες χρησιμοποιώντας για την αναπαράσταση της εικόνας το μοντέλο του σάκου χαρακτηριστικών. Η αναπαράσταση της εικόνας συνίσταται στη χρήση ιστογραμμάτων με τη συχνότητα εμφάνισης κάθε οπτικής λέξης. Έτσι επιλέγουν για τους SVMs, τον histogram intersection πυρήνα σε συνδυασμό και με RBF πυρήνα. Οι Setia et al. (2008) για ταξινόμηση ακτινογραφιών σε μία από τις 116 κλάσεις του κώδικα IRMA χρησιμοποιούν multi-class SVMs σε ένα πλαίσιο one-vs-rest. Παράλληλα εφαρμόζουν μια ιεραρχική μέθοδο ταξινόμησης για κάθε άξονα του κώδικα IRMA δημιουργώντας ένα δέντρο ταξινόμησης χρησιμοποιώντας τα εσωτερικά γινόμενα μεταξύ των SVM υπερεπιπέδων ως μέτρο ομοιότητας. Η ιεραρχική ταξινόμηση ακτινογραφιών, με βάση τους άξονες του κώδικα IRMA και τη χρήση multi-class SVMs χρησιμοποιούν τόσο οι Amaral et al. (2010) όσο και οι Tommasi et al. (2008a) και οι Mueen et al. (2008). Οι Mueen et al. (2008) αξιοποιούν την ιεραρχική δομή του κώδικα IRMA

για να επισημειώσουν εικόνες ακτίνων Χ εκπαιδεύοντας ξεχωριστούς SVM ταξινομητές για κάθε επίπεδο της ιεραρχίας. Οι Tommasi et al. (2008a) διαπιστώνουν την αδυναμία των SVM με αραιές κλάσεις, δηλαδή κλάσεις που περιέχουν πολύ λίγες εικόνες. Για να βελτιώσουν την αξιοπιστία της ταξινόμησης εμπλουτίζουν τις αραιές κλάσεις με εικονικά παραδείγματα, δηλαδή, με τροποποιημένα αντίγραφά τους. Επιπλέον χρησιμοποιούν τρεις στρατηγικές ταξινόμησης για τη σύντηξη των διανυσμάτων χαρακτηριστικών σε διαφορετικά στάδια της διαδικασίας επισημείωσης. Υψηλό επίπεδο ενσωμάτωσης όπου διαφορετικοί SVMs δέχονται διαφορετικές αναπαραστάσεις παράγοντας ο καθένας μια διαφορετική υπόθεση. Αυτές οι υποθέσεις συνδυάζονται για να επιτευχθεί η τελική απόφαση. Μεσαίο επίπεδο ενσωμάτωσης όπου διαφορετικοί περιγραφείς ενσωματώνονται σε έναν ταξινομητή παράγοντας την τελική υπόθεση. Στην προσέγγιση αυτή χρησιμοποιείται multi-class SVM με multi-cue πυρήνα. Στη χαμηλού επιπέδου ενσωμάτωση διαφορετικά διανύσματα χαρακτηριστικών συνδυάζονται σε ένα διάνυσμα που αποτελεί την είσοδο του ταξινομητή. Οι Han and Chen (2010) εφαρμόζουν σύντηξη πολλαπλών περιγραφέων χρησιμοποιώντας normalized συνάρτηση πυρήνα για ταξινόμηση με SVM εικόνων σε διαφορετικές κατηγορίες ιατρικής απεικόνισης.

Οι Kumar et al. (2016) αντιμετωπίζουν το πρόβλημα της ταξινόμησης εικόνων CT του ήπατος σε εξήντα πέντε κλάσεις σημασιολογικών εννοιών (επισημειώσεων) ως ένα πρόβλημα μάθησης πολλαπλών κλάσεων, πολλαπλών ετικετών. Προτείνουν ένα σχήμα ταξινόμησης δύο σταδίων με χρήση πολλαπλών SVM. Στο πρώτο στάδιο χρησιμοποιείται ένας 1-vs-all multi-class ταξινομητής για την εικόνα δοκιμής προκειμένου να εντοπιστούν ετικέτες από εικόνες που έχουν τα ίδια οπτικά χαρακτηριστικά. Αν μόνο ένας 1-vs-all SMV ταξινομητής επιστρέψει θετικό αποτέλεσμα, η ετικέτα που αντιστοιχεί σε αυτόν τον ταξινομητή, αποδίδεται στην εικόνα δοκιμής. Αν οι ταξινομητές σε αυτό το πρώτο στάδιο επιστρέψουν πολλές ετικέτες, (έστω l), τότε σε ένα δεύτερο στάδιο εκπαιδεύονται 1-vs-1 SVMs για κάθε ζευγάρι ετικετών μέσα σε αυτό το σύνολο πιθανών ετικετών (άρα $(l^2 - l)/2$ 1-vs-1 SVMs) και με ένα σχήμα ψηφοφορίας επιλέγεται η τελική επισημείωση.

Μια άλλη κατηγορία μεθόδων για την αυτόματη επισημείωση των ιατρικών εικόνων προέρχεται από τις τεχνικές ανάκτησης εικόνας με βάση το περιεχόμενο (Content-Based Image Retrieval - CBIR). Η βασική ιδέα πίσω από τις τεχνικές που βασίζονται στην ανάκτηση, είναι η αναγνώριση ή ανάκτηση μιας συλλογής από σημασιολογικά παρόμοιες με την εικόνα

για επισημείωση επισημειωμένων εικόνων, και η χρησιμοποίηση αυτής της συλλογής για να καθοριστούν οι καλύτερες σημασιολογικά επισημειώσεις για την εικόνα χωρίς ετικέτα. Ο ταξινομητής k πλησιέστερων γειτόνων (k-NN) είναι μια καθιερωμένη μέθοδος CBIR για τον εντοπισμό των k πιο όμοιων εικόνων με την εικόνα δοκιμής. Ο ταξινομητής k-NN είναι ένας συμβατικός, μη παραμετρικός επιβλεπόμενος ταξινομητής ο οποίος έχει καλή απόδοση για βέλτιστες τιμές του k. Η χρήση του είναι διαδεδομένη στον ιατρικό τομέα όπου η έλλειψη επισημειωμένων δεδομένων εκπαίδευσης αποτελεί μια σημαντική πρόκληση για την ανάπτυξη μεθόδων ταξινόμησης και επισημείωσης καθώς δεν επηρεάζεται από την έλλειψη εκπαιδευτικών δειγμάτων (Sohail et al., 2010, Wojnar and Pinheiro, 2012, Negarajan and Saravan, 2012, Charde and Lokhande, 2013, Fesharaki and Pourghassen, 2013, Abdulrazzaq et al., 2014, Bhuvaneswi et al.,2014, Spanier and Joskowicz, 2014, Kumar et al. 2016).

Ενώ οι πιο καθιερωμένες τεχνικές αυτόματης ταξινόμησης και επισημείωσης ιατρικών εικόνων είναι οι μηχανές διανυσμάτων υποστήριξης και οι k πλησιέστεροι γείτονες, έχουν προταθεί και άλλες μέθοδοι ταξινόμησης λόγω διάφορων περιορισμών που παρουσιάζουν, σε συγκεκριμένες περιπτώσεις. Η μία από αυτές είναι τα δέντρα απόφασης. Οι Ko et al. (2012) επιλέγουν για την ιεραρχική ταξινόμηση πολλαπλών ετικετών ακτινογραφιών σε 30 κλάσεις τυχαία δάση (random forests) αποτελούμενα από 120 δέντρα απόφασης θεωρώντας ότι οι multi-class SVM δεν αποτελούν την καλύτερη επιλογή όταν το διάνυσμα των χαρακτηριστικών έχει μεγάλη διάσταση και η βάση δεδομένων έχει περισσότερες από 1000 εικόνες λόγω της υπολογιστικής πολυπλοκότητας. Για την ίδια εργασία και οι Dimitrovski et al. (2011) χρησιμοποιούν πολλαπλά predictive clustering trees (PCT) για ιεραρχική ταξινόμηση πολλαπλών ετικετών. Οι Krisna and Prasad (2012) χρησιμοποιούν δέντρα απόφασης για ταξινόμηση πολλαπλών κλάσεων, πολλαπλών ετικετών, εικόνων MR-T2 του εγκεφάλου σε 15 σημασιολογικές κατηγορίες. Οι Chen et al. (2017) χρησιμοποιούν τυχαία δάση δέντρων παλινδρόμησης για τη βαθμολόγηση πνευμονικών οζιδίων σε εικόνες CT.

Για να μειωθούν τα προβλήματα που οφείλονται στην έλλειψη δεδομένων εκπαίδευσης, οι Gimenez et al. (2012) αποφεύγουν τις μεθόδους ταξινόμησης και αντ' αυτού επισημειώνουν εικόνες CT ήπατος χρησιμοποιώντας λογιστική παλινδρόμηση (logistic regression). Ωστόσο, η μέθοδός τους επισημειώνει μόνο δυαδικές σημασιολογικές εκβάσεις που θα μπορούσαν να παρουσιαστούν με θετικές ή αρνητικές παρατηρήσεις, για παράδειγμα, αν μια αλλοίωση ήταν ομοιογενής ή όχι.

Οι Bekker et al. (2016) χρησιμοποιούν, επίσης, λογιστική παλινδρόμηση για την ταξινόμηση των μαστογραφιών σε καλοήθεις (benign) και κακοήθεις (malignant). Για κάθε μαστό υπάρχουν μαστογραφίες από δύο όψεις (craniocaudal - CC, mediolateral obique - MLO). Τα δύο σύνολα εικόνων δίνονται ως είσοδοι σε δύο διαφορετικά μοντέλα λογιστικής παλινδρόμησης τα οποία παράγουν δύο τιμές ως απόφαση. Οι δύο αυτές τιμές συνδυάζονται στοχαστικά για την εξαγωγή της τελικής απόφασης επισημείωσης.

Υπάρχουν κάποιες μελέτες οι οποίες ακολουθούν διαφορετικές από τις κοινές προσεγγίσεις. Οι Bouslimi et al. (2013) και οι Zare et al. (2013a) χρησιμοποιούν την προσέγγιση του σάκου χαρακτηριστικών για τη δημιουργία ενός οπτικού λεξιλογίου και κατασκευάζουν ένα λεξιλόγιο εννοιών από τις επισημειώσεις του συνόλου εκπαίδευσης. Τα δύο λεξιλόγια συνδυάζονται χρησιμοποιώντας την τεχνική της λανθάνουσας σημασιολογικής ανάλυσης (LSA). Οι Zhang et al. (2014) προτείνουν μια καινοτόμο μέθοδο για το πρόβλημα της επισημείωσης εικόνων βιοψίας του δέρματος, το οποίο, όπως σημειώνουν, μπορεί να διατυπωθεί ως πρόβλημα μάθησης πολλαπλών κλάσεων πολλαπλών ετικετών (MIML), που βασίζεται σε ένα αραιό Bayesian μοντέλο. Οι Cirujeda and Binefa (2015) για το διαγωνισμό ταξινόμησης ιατρικών εικόνων ImageCLEF 2015 προτείνουν ένα μοντέλο αναπαράστασης δισδιάστατων εικόνων που μπορεί να είναι από ιατρικές εικόνες μέχρι βιοιατρικές δημοσιεύσεις και σχήματα, που χρησιμοποιεί $1^{ης}$ και $2^{ης}$ τάξης χαρακτηριστικά χρώματος από ολόκληρη την εικόνα υπό τη μορφή ενός πίνακα. Ο πίνακας αυτός που ονομάζεται Covariance matrix (πίνακας συνδιακύμανσης), αποτελεί τον περιγραφέα της εικόνας που χαρακτηρίζει την κατανομή των μεταβολών των χαρακτηριστικών στην εικόνα και όχι την απόλυτη τιμή τους, γεγονός που τον καθιστά αμετάβλητο στο μέγεθος της εικόνας και σε χωρικούς μετασχηματισμούς, όπως η περιστροφή. Έτσι το πρόβλημα της ταξινόμησης παίρνει τη μορφή ενός προβλήματος αραιής αναπαράστασης πίνακα.

Οι Burdescu et al. (2012) από την άλλη, εφαρμόζουν για την αυτόματη επισημείωση έγχρωμων φωτογραφιών καρκινωμάτων σε εννέα κλάσεις γνωστά μοντέλα επισημείωσης που έχουν προταθεί για γενικές εικόνες και συγκεκριμένα το μοντέλο CRM των Jeon et al. (2004) και το μοντέλο CMRM των Lavrenko et al. (2004). Οι Gong et al. (2010) τέλος, αναπτύσσουν ένα πλαίσιο επισημείωσης περιοχών ενδιαφέροντος σε εικόνες CT του εγκεφάλου που εμπνέεται από το Translation Model των Duygulu et al. (2002) για τη «μετάφραση» περιοχών αλλοιώσεων σε όρους παθολογίας.

Τα βαθιά συνελικτικά νευρωνικά δίκτυα (CNNs) χρησιμοποιούνται όλο και συχνότερα για την ταξινόμηση ιατρικών εικόνων. Οι Sapkota et al. (2015) χρησιμοποιούν ένα CNN τόσο για την εξαγωγή χαρακτηριστικών όσο και για την ταξινόμηση έγχρωμων ιατρικών εικόνων σε μία από δύο κατηγορίες. Οι Lyndon et al. (2015) χρησιμοποιούν ένα CNN για την ταξινόμηση ιατρικών εικόνων σε τριάντα διαφορετικές κατηγορίες ιατρικής απεικόνισης. Οι Zhang et al. (2019) εκπαιδεύουν ταυτόχρονα δύο βαθιά συνελικτικά δίκτυα ώστε να προβλέπουν αν ένα ζευγάρι εικόνων που δέχονται σαν είσοδο, ανήκουν ή όχι στην ίδια κλάση. Τέλος, οι Zhang et al. (2017) επιχειρούν να ενσωματώσουν ένα βαθύ συνελικτικό δίκτυο σε ένα μοντέλο μάθησης πολλαπλών στιγμιοτύπων, πολλαπλών ετικετών (MIML) για ταξινόμηση ιστοπαθολογικών εικόνων. Όμως, χρησιμοποιούν χωριστά κάθε στιγμιότυπο (περιοχή) για την εκπαίδευση του CNN και χρησιμοποιούν τα διανύσματα χαρακτηριστικών που παράγουν, με ένα MIML μοντέλο επισημείωσης, όπως το S-MIMLGP που έχουν αναπτύξει οι ίδιοι, και το γνωστό μοντέλο MIML ταξινόμησης MIMLBoost για σύγκριση.

6.2.3. Συμπεράσματα

Στην προηγούμενη ενότητα παρουσιάστηκε μια σειρά από μελέτες σχετικές με την ταξινόμηση και επισημείωση ιατρικών εικόνων που καλύπτουν το χρονικό διάστημα από το 2005 έως και σήμερα. Δε γίνεται ιδιαίτερη αναφορά σε προγενέστερες μελέτες, καθώς μέχρι το 2005, η αυτόματη ταξινόμηση ιατρικών εικόνων περιοριζόταν συχνά σε μικρό αριθμό κατηγοριών. Αρκετές μέθοδοι για την ανίχνευση προσανατολισμού ακτινογραφιών θώρακος είχαν προταθεί, όπου ο πλευρικός και ο μετωπικός προσανατολισμός εντοπιζόταν μέσω της ψηφιακής επεξεργασίας εικόνας. Γι’ αυτό το πρόβλημα δύο κλάσεων, τα ποσοστά σφάλματος ήταν κάτω από 1%. Οι Pinhas and Greenspan (2004) αναφέρουν ποσοστά σφάλματος κάτω του 1% για την αυτόματη ταξινόμηση 851 ακτινογραφιών σε οκτώ κατηγορίες ως προς τον προσανατολισμό τους, με χρήση ενός Gaussian mixture μοντέλου. Στο πιθανοτικό μοντέλο των Keysers et al. (2003) ορίζονται έξι κλάσεις σύμφωνα με το εξεταζόμενο μέρος του σώματος και για το σύνολο δοκιμών των 1617 εικόνων αναφέρεται ποσοστό σφάλματος 8%. Ωστόσο, ένας τέτοιος χαμηλός αριθμός κατηγοριών, όπως στις προαναφερόμενες μελέτες, δεν είναι κατάλληλος για εφαρμογές στην ιατρική βάσει αποδεικτικών στοιχείων (evidence-based medicine) ή στη συλλογιστική βάσει περιπτώσεων

(case-based reasoning), στις οποίες η κλάση της εικόνας πρέπει να καθοριστεί με περισσότερες λεπτομέρειες.

Ο διαγωνισμός επισημείωσης ιατρικής εικόνας ImageCLEF πρότεινε μια εργασία η οποία θα αντικατοπτρίζει τους πραγματικούς περιορισμούς της κατηγοριοποίησης εικόνας με βάση το περιεχόμενο σε ιατρικές εφαρμογές. Οι διοργανωτές κυκλοφόρησαν ένα μεγάλο και ετερογενές σώμα εικόνων ακτίνων-Χ και πρότειναν για την επισημείωσή τους τον κώδικα IRMA. Το 2005 και το 2006, οι 57 και 116 αντίστοιχα, κλάσεις καθορίστηκαν με την ομαδοποίηση παρόμοιων κωδικών σε μεμονωμένες κλάσεις και το ζητούμενο ήταν να προβλεφθεί η κλάση στην οποία ανήκει μια δοκιμαστική εικόνα. Το 2007 ο στόχος του έργου ήταν να βελτιωθεί η πρόβλεψη του πλήρους κώδικα IRMA. Στη συνέχεια χρησιμοποιήθηκε η ιεραρχική δομή για να περιγραφεί το περιεχόμενο της εικόνας, με το σχήμα αξιολόγησης να επιτρέπει μια λεπτότερη ακρίβεια στην ταξινόμηση. Το 2008, προστέθηκε το πρόβλημα της υψηλής ανισορροπίας των κλάσεων για να δοκιμαστεί η προηγούμενη γνώση που αποκτήθηκε στην ιεραρχική ταξινόμηση. Οι εικόνες στο σύνολο δοκιμών ήταν κυρίως από κλάσεις οι οποίες είχαν μόνο λίγα παραδείγματα στα δεδομένα εκπαίδευσης, καθιστώντας την επισημείωση σημαντικά δυσκολότερη. Το 2009 ο στόχος ήταν να συγκριθεί η δυνατότητα κλιμάκωσης διαφορετικών τεχνικών ταξινόμησης εικόνων με αυξανόμενο αριθμό κλάσεων, ιεραρχική ταξινόμηση και αραιές κλάσεις. Πολλές ερευνητικές ομάδες συμμετείχαν στο διαγωνισμό συγκρίνοντας τις μεθόδους τους και τα συμπεράσματα που προέκυψαν, εμπλούτισαν τη βιβλιογραφία και συνέβαλλαν στην πρόοδο που σημειώθηκε στο συγκεκριμένο τομέα. Η βάση δεδομένων που δημιουργήθηκε για τους σκοπούς του ImageCLEFmed αποτελεί έναν πολύτιμο πόρο για τη δημιουργία και τη δοκιμή αυτόματων συστημάτων επισημείωσης εικόνων ακτίνων-Χ και χρησιμοποιείται από πολλούς ερευνητές, μέχρι και σήμερα, ως βάση αναφοράς (benchmark) για την αξιολόγηση των συστημάτων τους.

Η ανάλυση των διαφορετικών μεθόδων αυτόματης επισημείωσης που χρησιμοποιούνται στις μελέτες που εξετάστηκαν, βασίστηκε σε δύο κριτήρια, την αναπαράσταση της εικόνας και τη μέθοδο ταξινόμησης. Όσον αφορά την αναπαράσταση της ιατρικής εικόνας οι προσεγγίσεις ποικίλλουν. Οι περισσότερες μελέτες χρησιμοποιούν χαρακτηριστικά που εξάγονται χειροκίνητα, αλλά τα τελευταία χρόνια τεχνικές βαθιάς μάθησης, ιδίως βαθιά συνελικτικά νευρωνικά δίκτυα (DCNNs), χρησιμοποιούνται για την εξαγωγή χαρακτηριστικών

απευθείας από τα ακατέργαστα εικονοστοιχεία της εικόνας.

Τα χειροκίνητα χαρακτηριστικά περιλαμβάνουν τις τιμές έντασης των εικονοστοιχείων της εικόνας καθώς και μετρήσεις που περιγράφουν την υφή και το σχήμα, καθώς οι ιατρικές εικόνες είναι στην πλειονότητα τους greyscale εικόνες (εικόνες ακτίνων-Χ, μαστογραφίες, εικόνες υπολογιστικής τομογραφίας, εικόνες υπερήχων, εικόνες μαγνητικού συντονισμού) και το χρώμα μπορεί να θεωρηθεί ως δευτερεύον χαρακτηριστικό. Η φύση κάθε κατηγορίας εικόνων καθορίζεται σε μεγάλο βαθμό από τη μέθοδο απεικόνισης (modality) περιπλέκοντας το πρόβλημα της επιλογής των οπτικών χαρακτηριστικών με τη μεγαλύτερη διακριτική ικανότητα για την αναπαράσταση της εικόνας. Για παράδειγμα, οι ενισχυτικές οθόνες που χρησιμοποιούνται για τη βελτίωση των ακτινογραφιών έχει ως αποτέλεσμα τη μείωση της χωρικής ανάλυσης, της αντίθεσης και της ευκρίνειας στην έξοδο. Έτσι, σε πολλές μελέτες οι ακτινογραφίες υποβάλλονται σε τεχνικές βελτίωσης (image enhancement), προκειμένου να βελτιωθεί η αντίθεση και η ευκρίνειά τους, πριν από το στάδιο της εξαγωγής χαρακτηριστικών προκειμένου να βελτιωθεί η ακρίβεια ταξινόμησης του μοντέλου.

Τα οπτικά χαρακτηριστικά εξάγονται είτε συνολικά, από ολόκληρη την εικόνα, είτε τοπικά από patches που λαμβάνονται με δειγματοληψία ή από περιοχές της εικόνας μετά από τμηματοποίηση ή από περιοχές γύρω από σημεία ενδιαφέροντος. Μια σαφής παρατήρηση είναι ότι οι μέθοδοι που χρησιμοποιούν τοπικούς περιγραφείς εικόνων, ξεπερνούν τις μεθόδους που χρησιμοποιούν συνολικούς περιγραφείς εικόνας. Τα περισσότερα συστήματα είτε χρησιμοποιούν αποκλειστικά τοπικά χαρακτηριστικά εικόνας είτε τοπικά χαρακτηριστικά εικόνας σε συνδυασμό με ένα συνολικό περιγραφέα. Οι μέθοδοι που χρησιμοποιούν τοπικά χαρακτηριστικά εξαγόμενα γύρω από σημεία ενδιαφέροντος και αναπαριστούν την εικόνα ως σάκο χαρακτηριστικών (bag-of-features), υιοθετούνται από αρκετές μελέτες τα τελευταία χρόνια. Μια μεγάλη ποικιλία προσεγγίσεων του σάκου-χαρακτηριστικών σημειώνεται στη βιβλιογραφία που διαφοροποιείται τόσο ως προς τη μέθοδο εξαγωγής των σημείων ενδιαφέροντος, όπου οι περιγραφείς SIFT και SURF αποδείχτηκαν ιδιαίτερα αποτελεσματικοί, όσο και ως προς τη δημιουργία του οπτικού λεξιλογίου. Σε άλλες μεθόδους η διάταξη των οπτικών λέξεων δεν λαμβάνεται υπόψη και χρησιμοποιείται μόνο η συχνότητα της κάθε οπτικής λέξης για τη δημιουργία των διανυσμάτων χαρακτηριστικών, ενώ άλλες ενσωματώνουν τις θέσεις όπου εξάγονται τα χαρακτηριστικά. Τέλος, έχει παρατηρηθεί ότι υπάρχει μια αυξανόμενη τάση προς τον

συνδυασμό των πολλαπλών χαρακτηριστικών σε διαφορετικά επίπεδα στη διαδικασία ταξινόμησης.

Όσον αφορά τις μεθόδους ταξινόμησης, οι προσεγγίσεις με βάση το μοντέλο των k πλησιέστερων γειτόνων (k-NN) που προέρχονται από το πλαίσιο CBIR, ήταν οι πιο συνηθισμένες τα πρώτα χρόνια, αφού μπορούσαν να επιτύχουν καλύτερα αποτελέσματα με τον ορισμό μιας κατάλληλης μετρικής απόστασης. Ωστόσο, οι διακριτικές προσεγγίσεις όπως οι μηχανές διανυσμάτων υποστήριξης (SVMs) και τα δέντρα αποφάσεων (DTs), έγιναν ολοένα και πιο κοινές και ξεπέρασαν τις προσεγγίσεις που βασίζονται στον πλησιέστερο γείτονα, καθώς αποδείχτηκε ότι μπορούν να ανταπεξέλθουν καλύτερα με μεγαλύτερο αριθμό κλάσεων από τον ταξινομητή k-NN. Η αλματώδης πρόοδος που συντελέστηκε στις τεχνικές βαθιάς μάθησης και η παράλληλη εξέλιξη στην ισχύ των υπολογιστικών συστημάτων επέτρεψαν την αξιοποίηση των βαθιών συνελικτικών νευρωνικών δικτύων στην αυτόματη επισημείωση έγχρωμων ιατρικών εικόνων. Το πρόβλημα μοντελοποιείται ως πρόβλημα ταξινόμησης με βάση τα εικονοστοιχεία, όπου ένα CNN εκπαιδεύεται με τις ακατέργαστες τιμές των εικονοστοιχείων της εικόνας και αυτόματα μαθαίνει ένα σύνολο ιεραρχικών χαρακτηριστικών για την ταξινόμηση (Sapkota et al., 2015).

Κεφάλαιο 7
Επίλογος

Η σημασία της ψηφιακής ιατρικής απεικόνισης στη σύγχρονη υγειονομική περίθαλψη έχει καταστήσει την ανάλυση της ιατρικής εικόνας αναπόσπαστο μέρος της κλινικής θεραπείας. Η ταξινόμηση των ιατρικών εικόνων, ένα θεμελιώδες βήμα στην ανάλυση της ιατρικής εικόνας, αποσκοπεί στη διάκριση των ιατρικών εικόνων σύμφωνα με ένα συγκεκριμένο κριτήριο, όπως είναι οι κλινικές παθολογίες ή οι μορφές απεικόνισης. Ένα αξιόπιστο σύστημα ταξινόμησης ιατρικών εικόνων είναι σε θέση να βοηθήσει τους γιατρούς στην ταχεία και ακριβή ερμηνεία τους.

Η αυτόματη επισημείωση εικόνας (AIA) αποτελεί ένα βήμα εμπρός προς αυτήν την κατεύθυνση, καθώς ο στόχος της είναι να αναγνωρίσει κάθε αντικείμενο που υπάρχει στην εικόνα και να αναθέσει αυτόματα τις αντίστοιχες ετικέτες που θα περιγράφουν το περιεχόμενό της. Οι τεχνικές AIA προσπαθούν να εκπαιδεύσουν ένα μοντέλο από τα δεδομένα εκπαίδευσης και στη συνέχεια να χρησιμοποιήσουν το εκπαιδευμένο μοντέλο για την αυτόματη ανάθεση σημασιολογικών ετικετών στη νέα εικόνα. Η AIA μπορεί να πραγματοποιηθεί χρησιμοποιώντας οπτικά χαρακτηριστικά ή οπτικά χαρακτηριστικά σε συνδυασμό με χαρακτηριστικά κειμένου ή χρησιμοποιώντας μεταδεδομένα που σχετίζονται με εικόνες.

Η ταξινόμηση των ιατρικών εικόνων έχει μελετηθεί διεξοδικά τις τελευταίες δεκαετίες με αποτέλεσμα ένα μεγάλο αριθμό λύσεων στη βιβλιογραφία, οι περισσότερες από τις οποίες βασίζονται στην εξαγωγή χαρακτηριστικών από τα πρωτογενή δεδομένα «χειροκίνητα». Τυπικά, ένα σύνολο χαρακτηριστικών όπως οι ακμές ή SIFT εξάγονται από τα ακατέργαστα δεδομένα και αυτά τα νέα χαρακτηριστικά αυτά καθαυτά ή σε συνδυασμό με τα πρωτότυπα πρωτογενή δεδομένα εισάγονται στον αλγόριθμο μηχανικής μάθησης. Αν και ορισμένες πτυχές της διαδικασίας μπορούν να αυτοματοποιηθούν ή να υλοποιηθούν με γνωστούς αλγορίθμους, ένα σημαντικό μειονέκτημα είναι ότι απαιτείται εξειδικευμένη γνώση του τομέα για να καθοριστεί ποια είναι τα βέλτιστα χαρακτηριστικά που θα πρέπει να

χρησιμοποιηθούν για μια συγκεκριμένη εργασία ταξινόμησης και να αξιολογηθεί η επιτυχία τους (Lyndon et al., 2015).

Όλα τα υφιστάμενα οπτικά χαρακτηριστικά έχουν περιορισμούς στην περιγραφή της εικόνας και κανένα από αυτά δεν είναι αρκετά ισχυρό για να αντιπροσωπεύσει τη μεγάλη ποικιλία των ιατρικών εικόνων. Κοινή πρακτική είναι ο συνδυασμός διάφορων τύπων χαρακτηριστικών για την αναπαράσταση όσο το δυνατόν περισσότερων εικόνων. Ωστόσο, η επεξεργασία και η ανάλυση διανυσμάτων χαρακτηριστικών υψηλής διάστασης είναι περίπλοκο ζήτημα. Λόγω της «κατάρας της διαστατικότητας» (curse of dimensionality), η απόδοση των ταξινομητών υποβαθμίζεται δραματικά όταν η διάσταση των διανυσμάτων χαρακτηριστικών είναι πολύ υψηλή για το δεδομένο αριθμό δειγμάτων. Επομένως, τα χαρακτηριστικά πρέπει να μειώνονται περαιτέρω ώστε να επιλεγεί ο σωστός αριθμός χαρακτηριστικών αλλά και τα σωστά χαρακτηριστικά για τη διαδικασία της επισημείωσης (Zhang et al., 2012).

Μία λύση στο συγκεκριμένο πρόβλημα έδωσαν οι τεχνικές βαθιάς μάθησης, ειδικά τα βαθιά συνελικτικά νευρωνικά δίκτυα (DCNN). Τα CNN μοιράζονται τα κοινά χαρακτηριστικά όλων των μοντέλων βαθιάς μάθησης: στοιβαγμένα στρώματα μη γραμμικών μονάδων επεξεργασίας πληροφοριών που μαθαίνουν ιεραρχικές αναπαραστάσεις (επιτρέποντας την κατανόηση των δεδομένων σε διάφορα επίπεδα αφαίρεσης, μεμονωμένα ή σε συνδυασμό), την ικανότητα να εκτελούνται με επίβλεψη ή χωρίς επίβλεψη προ-εκπαιδευμένα σε μη επισημειωμένα δεδομένα και τη δυνατότητα παραλληλοποίησης σε πολλαπλούς πολυπύρηνους επεξεργαστές (Lyndon et al., 2015). Η επιτυχής εφαρμογή τους σε μια ποικιλία γενικευμένων εργασιών απεικόνισης, όπως στην ταξινόμηση ιατρικών εικόνων αλλά και στον τμηματοποίηση της ιατρικής εικόνας υποδεικνύει τις δυνατότητες που έχουν για την αυτόματη επισημείωση εικόνας (Bhagat and Choudhary, 2018, Zhang et al., 2019).

Το μεγαλύτερο πλεονέκτημά τους, ωστόσο, είναι ότι μπορούν να εξάγουν αυτόματα χαρακτηριστικά από ακατέργαστα δεδομένα. Με την τροφοδοσία των δεδομένων διαδοχικά μέσω πολλών διαδοχικών στρωμάτων υπομονάδων, τα υψηλότερα επίπεδα του συστήματος είναι σε θέση να κατανοήσουν τα δεδομένα με όρους αφηρημένων αναπαραστάσεων. Έτσι τα μοντέλα DCNN παρέχουν ένα ενοποιημένο πλαίσιο, εξαγωγής χαρακτηριστικών –

ταξινόμησης, απαλλαγμένο από την προβληματική «χειροκίνητη» εξαγωγή χαρακτηριστικών για την επισημείωση ιατρικών εικόνων (Lyndon et al., 2015).

Ωστόσο, αν και αυτές οι μέθοδοι των DCNNs είναι ακριβέστερες από τις προσεγγίσεις που βασίζονται σε χειροκίνητα χαρακτηριστικά, δεν έχουν σημειώσει την ίδια επιτυχία στην ταξινόμηση ιατρικών εικόνων με αυτή που σημείωσε το δίκτυο AlexNet των Krizhevsky et al. (2012) στο διαγωνισμό ImageNet του 2012 (Zhang et al., 2019). Ένα σημαντικό εμπόδιο για την ανάπτυξή τους στον ιατρικό τομέα είναι η σχετική έλλειψη μεγάλων βάσεων ιατρικών εικόνων σε σύγκριση με τις γενικές βάσεις εικόνων αναφοράς, όπως του ImageNET (Lyndon et al., 2015). Μία από τις βασικές απαιτήσεις της βαθιάς μάθησης είναι η ύπαρξη μεγάλου αριθμού δεδομένων εκπαίδευσης. Τα DCNNs περιέχουν πολλά εκατομμύρια εσωτερικών παραμέτρων που πρέπει να εκτιμηθούν από τα δεδομένα. Πολύ λίγα δεδομένα μπορούν να έχουν ως αποτέλεσμα η ενεργοποίηση των νευρώνων υψηλότερου επιπέδου να είναι το αποτέλεσμα σημαντικών χαρακτηριστικών του εκπαιδευτικού συνόλου και να μην αντικατοπτρίζει τις υψηλού επιπέδου αναπαραστάσεις. Αν παρουσιαστεί αυτή η «υπερπροσαρμογή», τότε η ικανότητα του συστήματος να γενικεύει σε νέα δεδομένα είναι σοβαρά μειωμένη (Lyndon et al., 2015, Zhang et al., 2019).

Επίσης, η εκπαίδευση ενός μοντέλου βαθιάς μάθησης είναι μια πολύ χρονοβόρα διαδικασία. Ακόμη και η απλούστερη υλοποίηση ενός CNN απαιτεί θεμελιώδεις επιλογές για τον αριθμό και τον τύπο των στρωμάτων, το μέγεθος φίλτρου και τον αριθμό των φίλτρων ανά στρώμα και τον ρυθμό εκμάθησης. Ενώ υπάρχουν κατευθυντήριες γραμμές για τις επιλογές αυτές στη βιβλιογραφία, η δυσκολία ακόμη και μιας μικρής αναζήτησης παραμέτρων επιδεινώνεται από τις αυξημένες υπολογιστικές απαιτήσεις της εκπαίδευσης του συστήματος (Lyndon et al., 2015, Bhagat and Choudhary, 2018).

Προκειμένου να αντιμετωπιστεί το πρόβλημα της έλλειψης ενός μεγάλου αριθμού δεδομένων εκπαίδευσης στον ιατρικό τομέα και το πρόβλημα χρόνου εκπαίδευσης ενός DCNN, υιοθετούνται προ-εκπαιδευμένα βαθιά μοντέλα, για την εργασία ταξινόμησης ιατρικών εικόνων. Έχει αναγνωριστεί ευρέως ότι η ικανότητα αναπαράστασης εικόνας ενός DCNN που εκπαιδεύεται από σύνολα δεδομένων μεγάλης κλίμακας, όπως το ImageNet, μπορεί να μεταφερθεί αποτελεσματικά σε γενικές εργασίες οπτικής αναγνώρισης, όπου τα δεδομένα εκπαίδευσης είναι περιορισμένα (Lyndon et al., 2015, Cheng et al., 2018).

Η έλλειψη μιας κοινά αποδεκτής βάσης δεδομένων αναφοράς (benchmark) για την εκπαίδευση και αξιολόγηση μοντέλων αυτόματης επισημείωσης ιατρικής εικόνας αποτελεί ένα σημαντικό πρόβλημα. Όλες οι τυπικές μέθοδοι αυτόματης επισημείωσης απαιτούν μεγάλο αριθμό δειγμάτων επισημειωμένων εικόνας για την εκπαίδευση του μοντέλου. Αυτή τη στιγμή, διάφορες μέθοδοι AIA χρησιμοποιούν διαφορετικά σύνολα δεδομένων εικόνας για εκπαίδευση και αξιολόγηση, καθιστώντας δύσκολη τη σύγκριση της απόδοσης. Το θέμα της βάσης δεδομένων σχετίζεται στενά με το ζήτημα της έλλειψης ταξινομίας (ελεγχόμενου λεξιλογίου). Εάν είναι διαθέσιμη μια τυπική ταξινομία (για παράδειγμα ο κώδικας IRMA) της σημασιολογίας της εικόνας, μπορεί να δημιουργηθεί και μια τυποποιημένη βάση δεδομένων.

Το ελεγχόμενο λεξιλόγιο διαδραματίζει ζωτικό ρόλο στη διαδικασία επισημειώσεων ιατρικών εικόνων, επειδή μπορεί όχι μόνο να επιτρέπει στους χρήστες να καθορίζουν τα ερωτήματά τους από μια λίστα προκαθορισμένων λέξεων, αλλά και να μειώνουν την πολυπλοκότητα της επισημείωσης εικόνας. Παρόλο που έχουν δημιουργηθεί αρκετοί πόροι ιατρικών θησαυρών για ιατρική χρήση, το πρότυπο ελεγχόμενο λεξιλόγιο για κάθε τύπο ιατρικής εικόνας πρέπει ακόμα να αναγνωριστεί από τους επαγγελματίες του ιατρικού τομέα. Επιπλέον, το πρότυπο ελεγχόμενο λεξιλόγιο θα πρέπει να περιέχει ακριβείς ορισμούς του λεξιλογίου και των χαρακτηριστικών τους (Zhang et al., 2012).

Εκτός από τα επιμέρους ζητήματα που αναφέρθηκαν, η μεγαλύτερη πρόκληση μιας τεχνικής AIA είναι να μειώσει το σημασιολογικό κενό μεταξύ οπτικών χαρακτηριστικών χαμηλού επιπέδου της εικόνας που συλλαμβάνονται από μηχανές και έννοιες σημασιολογίας υψηλού επιπέδου που αντιλαμβάνεται ένας άνθρωπος (Cheng et al., 2018). Το ερώτημα συνεπώς είναι πώς θα οικοδομήσουμε ένα αποτελεσματικό μοντέλο επισημείωσης. Τα περισσότερα υπάρχοντα μοντέλα αυτόματης επισημείωσης ιατρικής εικόνας προσπαθούν να μάθουν τη σημασιολογία υψηλότερου επιπέδου από χαρακτηριστικά χαμηλού επιπέδου (οπτικά) χαρακτηριστικά μόνο. Επομένως, πληροφορίες κειμένου ή τα μεταδεδομένα που σχετίζονται με την εικόνα θα πρέπει να χρησιμοποιούνται, εάν είναι διαθέσιμα, για να ξεπεραστούν οι περιορισμοί των οπτικών χαρακτηριστικών προκειμένου να βελτιωθεί η ακρίβεια των επισημειώσεων. Η ανάλυση των μεταδεδομένων είναι ένα πολύπλοκο θέμα και θέτει ένα άλλο μεγάλο πρόβλημα στην επισημείωση εικόνας (Zhang et al., 2012).

Μοντέλα που βασίζονται στη μάθηση χωρίς επίβλεψη μπορούν να παράγουν ετικέτες μεταβλητού πλήθους και δεν απαιτούν επισημειωμένα δεδομένα. Η εκπαίδευση του μοντέλου πάνω στα μεταδεδομένα εικόνας είναι ένα δύσκολο έργο, καθώς τα περισσότερα από τα μεταδεδομένα είτε δεν είναι ακριβή είτε δεν επαρκούν και οι ετικέτες πρέπει να εξορύσσονται από αυτά τα θορυβώδη μεταδεδομένα και να διορθώνονται. Παρόλο που η διαδικασία αυτή παρουσιάζει προκλήσεις, έχει καταστεί δημοφιλής, καθώς ταιριάζει με την πραγματική κατάσταση σε σχέση με τις βάσεις δεδομένων μεγάλης κλίμακας και τις εικόνες του Ιστού (Bhagat and Choudhary, 2018). Προς την κατεύθυνση αυτή το ImageCLEF από το 2012, διοργανώνει επεκτάσιμη (scalable) εργασία επισημείωσης εικόνας, όπου η πρόκληση είναι να βρεθεί μια ισχυρή σχέση μεταξύ των εικόνων και του γύρω κειμένου. Η εστίαση δίνεται κυρίως στην απόκτηση της επισημείωσης από τα μεταδεδομένα του διαδικτύου. Το κίνητρο είναι να σχεδιαστεί ένα σύστημα επισημείωσης εικόνας όπου ο αριθμός των λέξεων-κλειδιών είναι επεκτάσιμος.

Μια άλλη κατεύθυνση στην οποία έχει στραφεί η έρευνα είναι η περιγραφή των εικόνων με προτάσεις. Οι εικόνες συνήθως φέρουν ετικέτες με καθορισμένες λέξεις-κλειδιά. Ωστόσο, οι προτάσεις έχουν πλουσιότερο περιεχόμενο και πιο συμπαγείς και λεπτές αναπαραστάσεις πληροφοριών παρά οι διακεκριμένες λέξεις-κλειδιά (Cheng et al., 2018). Οι μέθοδοι επισημείωσης σταθερού πλήθους ετικετών προσπάθησαν να λύσουν το πρόβλημα της ανισορροπίας κλάσεων, του θορυβώδους συνόλου δεδομένων, των ελλιπών ετικετών κ.λπ. αξιοποιώντας τη συσχέτιση μεταξύ ετικετών και οπτικών χαρακτηριστικών. Μία μέθοδος επισημείωσης βάσει μεταβλητού πλήθους ετικετών είναι ένας ρεαλιστικός τρόπος για την επισημείωση ιατρικών εικόνων, καθώς επιτρέπει την επισημείωση όλων των σχετικών αντικειμένων που υπάρχουν στην εικόνα. Πρόσφατα, τεχνικές βαθιάς μάθησης χρησιμοποιούνται για την πρόβλεψη αυθαίρετου πλήθους ετικετών. Μία μέθοδος, εμπνευσμένη από τις μεθόδους δημιουργίας υποτίτλων εικόνων, χρησιμοποιεί ένα αναδρομικό νευρωνικό δικτύο (RNN) στην έξοδο ενός κλασικού DCNN δικτύου (Vinyals et al., 2015).

Η αυτόματη επισημείωση ιατρικής εικόνας εξακολουθεί να είναι ένας πολύ δύσκολος ερευνητικός χώρος που θέτει συνεχώς νέες προκλήσεις. Αυτή η μελέτη επιχείρησε να αποτυπώσει την αλματώδη ανάπτυξη των μεθόδων αυτόματης επισημείωσης εικόνας αναλύοντας τις καθιερωμένους τεχνικές, εκθέτοντας τα προβλήματα και τους περιορισμούς

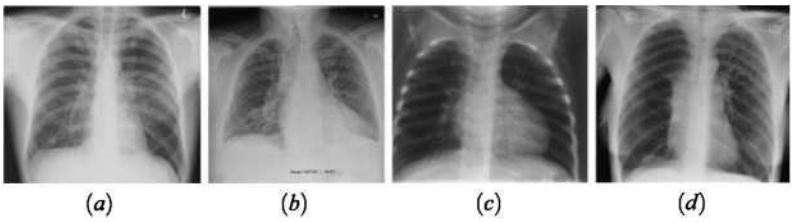
που συνάντησε η ερευνητική κοινότητα και αποκαλύπτοντας τις αναδυόμενες κατευθύνσεις που θα συγκεντρώσουν μελλοντικά την προσοχή των ερευνητών, με στόχο να προσφέρει τις απαραίτητες γνώσεις για τη διερεύνηση του τομέα της αυτόματης επισημείωσης ιατρικής εικόνας.

Παράρτημα Α

Η βάση δεδομένων ImageCLEF medical annotation 2005-2009

Η βάση δεδομένων που χρησιμοποιήθηκε για την εργασία επισημείωσης ιατρικής εικόνας του διαγωνισμού ImageCLEF[10] δημιουργήθηκε με εικόνες που παρείχε η ομάδα IRMA (Image Recovery for Medical Applications) του Πανεπιστημιακού Νοσοκομείου RWTH του Άαχεν της Γερμανίας. Αποτελείται από ιατρικές ακτινογραφίες που επιλέχτηκαν τυχαία από την καθημερινή εργασία στο Τμήμα Διαγνωστικής Ακτινολογίας. Οι περισσότερες εικόνες είναι δευτερεύουσες ψηφιοποιημένες εικόνες από απλές ακτινογραφίες, αλλά η βάση δεδομένων περιλαμβάνει επίσης εικόνες από άλλες μεθόδους απεικόνισης, όπως εικόνες CT και εικόνες υπερήχων (Deselaers et al, 2008).

Το σύνολο δεδομένων περιέχει μεγάλη μεταβλητότητα: εικόνες διαφόρων τμημάτων του σώματος, ασθενών διαφορετικών ηλικιών, διαφορετικού φύλου, από διαφορετική οπτική γωνία με ή χωρίς παθολογίες. Επιπλέον, η ποιότητα των ακτινογραφιών ποικίλει σημαντικά και υπάρχει μεγάλη μεταβλητότητα εντός της κλάσης μαζί με μια ισχυρή οπτική ομοιότητα μεταξύ πολλών εικόνων που ανήκουν σε διαφορετικές κατηγορίες (εικόνες 34 (a) – (d)).



Εικόνα 34. Εικόνες από τη βάση IRMA που χρησιμοποιήθηκαν για το διαγωνισμό ImageCLEF (Deselaers et al, 2008).

Παρατηρείται υψηλή οπτική ομοιότητα μεταξύ των εικόνων. Κάθε μία από αυτές ανήκει σε διαφορετική κλάση. Όλες έχουν σαν τρόπο λήψης την ετικέτα-έννοια “high beam energy” (ενέργεια υψηλής ακτινοβολίας), ως περιοχή σώματος “chest unspecified” (μη προσδιορισμένο στήθος), ως βιολογικό σύστημα “unspecified” (απροσδιόριστο), αλλά διαφέρουν ως προς τον προσανατολισμό του σώματος: α) “PA unspecified” (μη καθορισμένο), β) “PA expiration” (εκπνοή), (γ) “AP inspiration” (εισπνοή), (δ) “AP supine” (ύφεση).

[10] https://www.imageclef.org/

Όλες οι εικόνες παρέχονται ως αρχεία PNG, κλιμακούμενα ώστε να ταιριάζουν σε ένα πλαίσιο οριοθέτησης 512×512 εικονοστοιχείων (διατηρώντας τον λόγο διαστάσεων) χρησιμοποιώντας 256 αποχρώσεις του γκρι.

IRMA: 1121-220-230-700
T - ακτινογραφία, ακτινογραφία προβολής, αναλογική, εικόνα
D - οβελιαία, αριστερά-δεξιά πλευρική
A - κρανίο, νευρο κρανίου
B - μυοσκελετικό σύστημα

Εικόνα 35. Παράδειγμα ακτινογραφίας επισημειωμένης με τον κώδικα IRMA (Deselaers et al, 2008).

Καθώς οι επισημειώσεις των ιατρικών εικόνων απαιτούν ονοματολογία συγκεκριμένων όρων για την περιγραφή του περιεχομένου τους, οι εικόνες ταξινομήθηκαν χειροκίνητα από ειδικούς ιατρούς χρησιμοποιώντας τον κώδικα IRMA. Ο κώδικας IRMA (Image Recovery for Medical Applications) είναι ένα μονο-ιεραρχικό πολυαξονικό πρότυπο για την ταξινόμηση των ιατρικών εικόνων, το οποίο έχει σχεδιαστεί για την αποφυγή ελλείψεων, ασάφειας και έλλειψης αιτιώδους συνάφειας που παρατηρούνται σε άλλα σχήματα ταξινόμησης. Το πρότυπο IRMA θεωρεί ότι οι μόνες πιθανές σχέσεις μεταξύ στοιχείων κώδικα και υποκώδικα είναι της μορφής “is a” «είναι ένα» και “part of” «μέρος του».

Αποτελείται από τέσσερις ανεξάρτητους άξονες που περιγράφουν διαφορετικό περιεχόμενο μέσα στην εικόνα: ο κώδικας του τεχνικού (T - Technical) άξονα περιγράφει τον τρόπο λήψης (modality) της εικόνας, ο κώδικας του άξονα κατεύθυνσης (D - Directional) περιγράφει τον προσανατολισμό του σώματος, ο κώδικας του ανατομικού (A - Anatomical) άξονα την εξεταζόμενη περιοχή του σώματος και ο κώδικας του βιολογικού (B - Biological) άξονα το εξεταζόμενο βιολογικό σύστημα (Lehmann et al., 2003).

Κάθε ένας από αυτούς συνδέεται με μια ετικέτα με τρεις έως τέσσερις χαρακτήρες από το σύνολο {0, ..., 9, a, ..., z}, όπου το “0” υποδηλώνει “μη καθορισμένο” για να καθορίσει το τέλος μιας διαδρομής κατά μήκος ενός άξονα. Σε αυτήν την ιεραρχία, όσο περισσότερο διαφέρει η θέση του κώδικα από το “0”, τόσο πιο λεπτομερής είναι η περιγραφή. Έτσι, ο

πλήρης κώδικας IRMA είναι μια συμβολοσειρά 13 χαρακτήρων TTTT-DDD-AAA-BBB, μια δομή που μπορεί εύκολα να επεκταθεί εισάγοντας χαρακτήρες σε μια συγκεκριμένη θέση κώδικα εάν εισαχθούν νέοι τρόποι απεικόνισης (Lehmann et al., 2003). Ένα μικρό απόσπασμα από τον άξονα ανατομίας του κώδικα IRMA δίνεται στον Πίνακα 17. Ένα παράδειγμα εικόνας από τη βάση δεδομένων μαζί με τις λεκτικές ετικέτες και τον πλήρη κωδικό τους δίνονται στο σχήμα 35.

Κωδικός	Λεκτική περιγραφή
000	not further specified
…	
400	upper extremity (arm)
410	upper extremity (arm); hand
411	upper extremity (arm); hand; finger
412	upper extremity (arm); hand; middle hand
413	upper extremity (arm); hand; carpal bones
420	upper extremity (arm); radio carpal join
430	upper extremity (arm); forearm
431	upper extremity (arm); forearm; distal forearm
432	upper extremity (arm); forearm; proximal forearm
440	upper extremity (arm); elbow
…	

Πίνακας 17. Παραδείγματα από τον κώδικα IRMA, ανατομικός άξονας (Deselaers et al, 2008).

Το 2005, δημιουργήθηκε μια βάση δεδομένων 10.000 εικόνων. Οι εικόνες ομαδοποιήθηκαν σύμφωνα με την επισημείωση IRMA που έφεραν σε ένα αδρό επίπεδο λεπτομέρειας συγκροτώντας 57 κλάσεις. Από αυτές 9.000 τυχαία επιλεγμένες εικόνες επελέγησαν ως δεδομένα εκπαίδευσης και δόθηκαν πριν από την αξιολόγηση. Ένα υπόλοιπο σύνολο, 1.000 εικόνων, δημοσιεύθηκε αργότερα ως δεδομένα δοκιμών χωρίς πληροφορίες κλάσης. Σε όλες τις επόμενες εκδόσεις του διαγωνισμού ImageCLEF, η βάση δεδομένων χτίστηκε πάνω στη βάση του προηγούμενου έτος. Το 2006, το σύνολο των 10.000 εικόνων του 2005 χρησιμοποιήθηκε για την εκπαίδευση και συλλέχθηκε μια νέα ομάδα 1.000 εικόνων για δοκιμές. Ο αριθμός των κλάσεων υπερδιπλασιάστηκε και βάσει του κώδικα IRMA ορίστηκαν 116 κατηγορίες. Το 2007 υιοθετήθηκε η ίδια διαδικασία, προστέθηκε ένα νέο σύνολο 1.000 εικόνων δοκιμών και οι 11.000 εικόνες από το 2006 χρησιμοποιήθηκαν ως δεδομένα εκπαίδευσης. Ο αριθμός των κλάσεων παρέμεινε σταθερός στις 116 αλλά αυτή τη φορά το ζητούμενο δεν ήταν να προβλεφθεί η ακριβής κλάση, αλλά να προβλεφθεί ο κώδικας IRMA και καθορίστηκε ένα κριτήριο αξιολόγησης που θα αναγνωρίζει την ιεραρχία. Το 2008 τα

δεδομένα που κυκλοφόρησαν, αποτελούνταν από 12.076 εκπαιδευτικές εικόνες (11.000 εκπαιδευτικές εικόνες του 2007 + 1.000 εικόνες δοκιμών του 2007 + 76 νέες εικόνες) και ένα νέο σύνολο δοκιμών 1.000 εικόνων, όλες τους επισημειωμένες με συνολικά, 196 μοναδικούς κωδικούς.

Σε όλες τις χρησιμοποιούμενες βάσεις δεδομένων οι κλάσεις ήταν άνισα κατανεμημένες αντικατοπτρίζοντας την ακτινολογική ρουτίνα της λήψης. Ωστόσο, στις τρεις πρώτες εκδόσεις του διαγωνισμού, κάθε κλάση περιείχε τουλάχιστον 10 εικόνες. Το 2005, η μεγαλύτερη κλάση είχε το 28,6% (2.860 εικόνες) του πλήρους συνόλου δεδομένων, η δεύτερη το 9,6% (959 εικόνες) της συλλογής και υπήρχαν αρκετές κατηγορίες που συνιστούσαν μόνο μεταξύ 0,1% και 0,2% (10 με 20 εικόνες) του πλήρους συνόλου. Το 2006 οι δύο πιο μεγάλες κλάσεις είχαν αντίστοιχα 19,3% και 9,2% του συνόλου δεδομένων, ενώ έξι κλάσεις είχαν μόνο 1% ή λιγότερο.

Η ανισορροπία των κλάσεων επιδεινώθηκε το 2008. Από το σύνολο των 196 κωδικών που υπήρχαν στο στάδιο της εκπαίδευσης, μόνο 187 εμφανίστηκαν στο σύνολο δοκιμών. Η πιο συχνή κλάση στα δεδομένα εκπαίδευσης περιλάμβανε περισσότερες από 2300 εικόνες αλλά τα δεδομένα των δοκιμών είχαν μόνο ένα παράδειγμα από αυτή την κλάση. Η κατανομή των δεδομένων των δοκιμών ήταν σχεδόν ομοιόμορφη ενώ για τα δεδομένα εκπαίδευσης η κατανομή κορυφώθηκε σε ορισμένες κλάσεις.

Τέλος, το 2009 δημιουργήθηκε μια βάση δεδομένων με 12.677 ακτινογραφίες που είχαν ταξινομηθεί πλήρως ως σύνολο εκπαίδευσης. Οι εικόνες παρέχονται με ετικέτες σύμφωνα με τα συστήματα ταξινόμησης των εργασιών επισημείωσης από το 2005-2008:

- 57 κλάσεις όπως και το 2005 (12.631 εικόνες) + μια κλάση “clutter” C (46 εικόνες).
- 116 κλάσεις όπως και το 2006 (12.334 εικόνες) + μια κλάση “clutter” C (343 εικόνες).
- 116 κωδικοί IRMA όπως και το 2007 (12.334 εικόνες) + μια κλάση “clutter” C (343 εικόνες).
- 193 κωδικοί IRMA όπως και το 2008 (12.677 εικόνες).

Η κλάση “clutter’” για ένα συγκεκριμένο σύνολο δεδομένων περιείχε όλες τις εικόνες που δεν ήταν αναγνωρίσιμες εκείνη τη χρονιά, αλλά επισημειώθηκαν με ένα υψηλότερο επίπεδο λεπτομέρειας κώδικα κατά τα επόμενα έτη. Τα δεδομένα δοκιμής αποτελούνταν από 1.733

εικόνες. Δεν είχαν όλες οι κλάσεις του εκπαιδευτικού συνόλου παραδείγματα σε αυτό το σύνολο:

- Ετικέτες 2005: 55 κλάσεις (από 57) με 1.639 εικόνες + κλάση C με 94 εικόνες.
- Ετικέτες 2006: 109 κλάσεις (από 116) με 1.353 εικόνες + κλάση C με 380 εικόνες.
- Ετικέτες 2007: 109 κωδικοί IRMA (από 116) με 1.353 εικόνες + κλάση C με 380 εικόνες.
- Ετικέτες 2008: 169 κωδικοί IRMA (από 193) με 1.733 εικόνες.

Βιβλιογραφία

Abdulrazzaq, M., Mohd, S. and Fadhil, M. (2014). Medical Image Annotation and Retrieval by Using Classification Techniques. In: *3rd International Conference on Advanced Computer Science Applications and Technologies*. New Jersey: IEEE, pp.32-36.

Amaral, I., Coelho, F., da Costa, J. and Cardoso, J. (2010). Hierarchical medical image annotation using SVM-based approaches. *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine*.

Avni, U., Goldberger, J. and Greenspan, H. (2008). TAU MIPLAB at ImageClef 2008. *CLEF*.

Avni, U., Greenspan, H. and Goldberger, J. (2009). Dense Simple Features for Fast and Accurate Medical X-Ray Annotation. *Lecture Notes in Computer Science*, pp.239-246.

Bakliwal, P. and Jawahar, C. (2015). Active learning based image annotation. In: *Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*. New Jersey: IEEE.

Ballesteros, L. and Petkova, D. (2006). Categorizing and Annotating Medical Images by Retrieving Terms Relevant to Visual Features.

Bay, H., Tuytelaars, T. and Van Gool, L. (2006). SURF: Speeded Up Robust Features. *Computer Vision – ECCV 2006*, pp.404-417.

Bekker, A., Shalhon, M., Greenspan, H. and Goldberger, J. (2016). Multi-View Probabilistic Classification of Breast Microcalcifications. *IEEE Transactions on Medical Imaging*, 35(2), pp.645-653.

Besançon, R. and Millet, C. (2005). Data Fusion of Retrieval Results from Different Media: Experiments at ImageCLEF 2005. *Accessing Multilingual Information Repositories*, pp.622-631.

Bhagat, P. and Choudhary, P. (2018). Image annotation: Then and now. *Image and Vision Computing*, 80, pp.1-23.

Bhuvaneswari, C., Aruna, P. and Loganathan, D. (2014). A new fusion model for classification of the lung diseases using genetic algorithm. *Egyptian Informatics Journal*, 15(2), pp.69-77.

Blei, D. and Jordan, M. (2003). Modeling annotated data. In: *26th Annual International ACMSIGIR Conference on Research and Development in Information Retrieval*. NewYork: ACM, pp.127-134.

Bouslimi, R. and Akaichi, J. (2015). New approach for automatic medical image annotation using the bag-of-words model. In: *6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*. New Jersey: IEEE, pp.1088-1093.

Bouslimi, R., Messaoudi, A. and Akaichi, A. (2013). Using a Bag of Words for Automatic Medical Image Annotation with a Latent Semantic. *International Journal of Artificial Intelligence & Applications*, 4(3), pp.51-60.

Burdescu, D., Mihai, C., Stanescu, L. and Brezovan, M. (2012). Automatic image annotation and semantic based image retrieval for medical domain. *Neurocomputing*, 109, pp.33-48.

Βερύκιος, Β., Καγκλής, Β. και Σταυρόπουλος, Η. (2019). *Η επιστήμη των δεδομένων μέσα από τη γλώσσα R*. [online] Hdl.handle.net. Available at: http://hdl.handle.net/11419/2965 [Accessed 5 Dec. 2019].

Caicedo, J., Cruz, A. and Gonzalez, F. (2009). Histopathology Image Classification Using Bag of Features and Kernel Functions. *Artificial Intelligence in Medicine*, pp.126-135.

Calonder, M., Lepetit, V., Strecha, C. and Fua, P. (2010). BRIEF: Binary Robust Independent Elementary Features. *Computer Vision – ECCV 2010*, pp.778-792.

Carneiro, G., Chan, A., Moreno, P. and Vasconcelos, N. (2007). Supervised Learning of Semantic Classes for Image Annotation and Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3), pp.394-410.

Carson, C., Belongie, S., Greenspan, H. and Malik, J. (2002). Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8), pp.1026-1038.

Chan, T. and Vese, L. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, 10(2), pp.266-277.

Chang, Y., Lin, W. and Chen, H. (2006). Combining Text and Image Queries at ImageCLEF2005. *CLEF*.

Channin, D., Mongkolwat, P., Kleper, V., Sepukar, K. and Rubin, D. (2009). The caBIG™ Annotation and Image Markup Project. *Journal of Digital Imaging*, 23(2), pp.217-225.

Chapelle, O., Haffner, P. and Vapnik, V. (1999). Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks*, 10(5), pp.1055-1064.

Charde, P. and Lokhande, S. (2013). Classification Using K Nearest Neighbor for Brain Image Retrieval. *International Journal of Scientific & Engineering Research*, 4(8), pp.760- 765.

Chen, S., Qin, J., Ji, X., Lei, B., Wang, T., Ni, D. and Cheng, J. (2017). Automatic Scoring of Multiple Semantic Attributes With Multi-Task Feature Leverage: A Study on Pulmonary Nodules in CT Images. *IEEE Transactions on Medical Imaging*, 36(3), pp.802-814.

Cheng, P. and Yang, W. (2007). CYU_IM@ImageCLEF 2007: Medical image annotation task.

Cheng, P., Chien, B., Ke, H. and Yang, W. (2006). Combining Textual and Visual Features for Cross-Language Medical Image Retrieval. *Accessing Multilingual Information Repositories*, pp.712-723.

Cheng, Q., Zhang, Q., Fu, P., Tu, C. and Li, S. (2018). A survey and analysis on automatic image annotation. *Pattern Recognition*, 79, pp.242-259.

Chi, A. and Christante, B. (2015). *Image Retrieval Discussion*.

Cirujeda, P. and Binefa, X. (2015). Medical Image Classification via 2D color feature based Covariance Descriptors.

Clouard, R., Renouf, A. and Revenu, M. (2010). An Ontology-Based Model for Representing Image Processing Objectives. *International Journal of Pattern Recognition and Artificial Intelligence*, 24(08), pp.1181-1208.

Cortes, C. and Vapnik, V. (1995). *Machine Learning*, 20(3), pp.273-297.

Datta, R., Joshi, D., Li, J. and Wang, J. (2008). Image retrieval. *ACM Computing Surveys*, 40(2), pp.1-60.

Deng, Y. and Manjunath, B. (2001). Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8), pp.800-810.

Deselaers, T. and Deserno, T. (2009). Medical Image Annotation in ImageCLEF 2008. *Lecture Notes in Computer Science*, pp.523-530.

Deselaers, T., Deserno, T. and Müller, H. (2008). Automatic medical image annotation in ImageCLEF 2007: Overview, results, and discussion. *Pattern Recognition Letters*, 29(15), pp.1988-1995.

Deselaers, T., Müller, H., Clough, P., Ney, H. and Lehmann, T. (2006). The CLEF 2005 Automatic Medical Image Annotation Task. *International Journal of Computer Vision*, 74(1), pp.51-58.

Deselaers, T., Weyand, T. and Ney, H. (2007). Image Retrieval and Annotation Using Maximum Entropy. *Evaluation of Multilingual and Multi-modal Information Retrieval*, pp.725-734.

Dimitrovski, I., Kocev, D., Loskovska, S. and Džeroski, S. (2010). ImageCLEF 2009 Medical Image Annotation Task: PCTs for Hierarchical Multi-Label Classification. *Lecture Notes in Computer Science*, pp.231-238.

Dimitrovski, I., Kocev, D., Loskovska, S. and Džeroski, S. (2011). Hierarchical annotation of medical images. *Pattern Recognition*, 44(10-11), pp.2436-2449.

Duygulu, P., Barnard, K., Freitas, J. and Forsyth, D. (2002). Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. In: *Seventh European Conference on Computer Vision*. Springer, pp.349-354.

Ermis, B., Cemgil, A., Marvasti, N. and Acar, B. (2014). Liver CT Annotation via Generalized Coupled Tensor Factorization. *CLEF*.

Feng, S., Manmatha, R. and Lavrenko, V. (2004). Multiple Bernoulli relevance models for image and video annotation. In: *Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* New Jersey: IEEE.

Fesharaki, N. and Pourghassem, H. (2012). Medical X-ray Images Classification Based on Shape Features and Bayesian Rule. *2012 Fourth International Conference on Computational Intelligence and Communication Networks*.

Fesharaki, N. and Pourghassem, H. (2013). Medical X-ray Image Hierarchical Classification Using a Merging and Splitting Scheme in Feature Space. *Journal of Medical Signals & Sensors*, 3(3), p.150.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Qian Huang, Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D. and Yanker, P. (1995). Query by image and video content: the QBIC system. *Computer*, 28(9), pp.23-32.

Florea, F., Rogozan, A., Barbu, E., Bensrhair, A. and Darmoni, S. (2007). MedIC at ImageCLEF 2006: Automatic Image Categorization and Annotation Using Combined Visual

Representations. *Evaluation of Multilingual and Multi-modal Information Retrieval*, pp.670-677.

Frate, F., Pacifici, F., Schiavon, G. and Solimini, C. (2007). Use of Neural Networks for Automatic Classification from High-Resolution Images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4), pp.800-809.

Ganesan, S. and Subashini, T. (2015). View classification of medical x-ray images using PNN classifier, decision trees algorithm and SVM classifier. *International Journal of Research in Engineering and Technology*, 4(11), pp.277-282.

Gevers, T. and Smeulders, A. (2000). PicToSeek: combining color and shape invariant features for image retrieval. *IEEE Transactions on Image Processing*, 9(1), pp.102-119.

Ghofrani, F., Helfroush, M., Danyali, H. and Kazemi, K. (2012). Medical X-ray Image Classification Using Gabor-Based CS-Local Binary Patterns.

Gimenez, F., Xu, J., Liu, Y., Liu, T., Beaulieu, C., Rubin, D. and Napel, S. (2012). Automatic Annotation of Radiological Observations in Liver CT Images. In: *AMIA- Annual Symposium proceedings*, pp.257-263.

Goh, K., Chang, E. and Li, B. (2005). Using one-class and two-class SVMs for multiclass image annotation. *IEEE Transactions on Knowledge and Data Engineering*, 17(10), pp.1333-1346.

Gong, T., Li, S., Tan, C., Pang, B., Lim, C., Lee, C., Tian, Q. and Zhang, Z. (2010). Automatic Pathology Annotation on Medical Images: A Statistical Machine Translation Framework. *2010 20th International Conference on Pattern Recognition*.

Gong, Y., Jia, Y., Leung, T., Toshev, A. and Ioffe, S. (2013). Deep Convolutional Ranking for Multilabel Image Annotation.

Gonzalez, R. and Woods, R. (2014). *Digital image processing*. New Delhi: Dorling Kindersley.

Guillaumin, M., Mensink, T., Verbeek, J. and Schmid, C. (2009). TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation. In: *12th International Conference on Computer Vision*. New Jersey: IEEE, pp.309–316.

Güld, M. and Deserno, T. (2008). Baseline Results for the ImageCLEF 2007 Medical Automatic Annotation Task Using Global Image Features. *Lecture Notes in Computer Science*, pp.637-640.

Güld, M., Kohnen, M., Keysers, D., Schubert, H., Wein, B., Bredno, J. and Lehmann, T. (2002). Quality of DICOM header information for image categorization</title>. *Medical Imaging 2002: PACS and Integrated Medical Information Systems: Design and Evaluation*.

Güld, M., Thies, C., Fischer, B. and Deserno, T. (2007). Baseline Results for the ImageCLEF 2006 Medical Automatic Annotation Task. *Evaluation of Multilingual and Multi-modal Information Retrieval*, pp.686-689.

Güld, M., Welter, P. and Deserno, T. (2009). Baseline Results for the ImageCLEF 2008 Medical Automatic Annotation Task in Comparison over the Years. *Lecture Notes in Computer Science*, pp.752-755.

Gupta, A. and Jain, R. (1997). Visual information retrieval. *Communications of the ACM*, 40(5), pp.70-79.

Han, X. and Chen, Y. (2010). ImageCLEF 2010 Modality Classification in Medical Image Retrieval: Multiple Feature Fusion with Normalized Kernel Function. *CLEF*.

Haralick, R., Shanmugam, K. and Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6), pp.610-621.

Hersh, W., Kalpathy-Cramer, J. and Jensen, J. (2007). Medical Image Retrieval and Automated Annotation: OHSU at ImageCLEF 2006. *Evaluation of Multilingual and Multi-modal Information Retrieval*, pp.660-669.

Hou, Y. and Lin, Z. (2015). Image tag completion and refinement by subspace clustering and matrix completion. *2015 Visual Communications and Image Processing (VCIP)*.

Hsu, C., Chu, W. and Taira, R. (1996). A knowledge-based approach for retrieving images by content. *IEEE Transactions on Knowledge and Data Engineering*, 8(4), pp.522-532.

Hu, H., Zhou, G., Deng, Z., Liao, Z. and Mori, G. (2016). Learning structured inference neu- ral networks with label relations. In: *Conference on Computer Vision and Pattern Recognition*. New Jersey: IEEE, pp.2960–2968.

Huang, J., Kumar, S., Mitra, M., Zhu, W. and Zabih, R. (1997). Image indexing using color correlograms. In: *Computer Society Conference on Computer Vision and Pattern Recognition*. New Jersey: IEEE, pp.762–765.

Huang, J., Kumar, S. and Zabih, R. (1998). An automatic hierarchical image classification scheme. *Proceedings of the sixth ACM international conference on Multimedia - MULTIMEDIA '98.*

Huang, J., Kumar, S., Mitra, M. and Zhu, W. (1999). Spatial color indexing and applications. *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*.

Nedjar I., Saïd, M., S., Chikh, M., Abi-Yad, K. and Bouafia, Z. (2015). Automatic Annotation of Liver CT Image: ImageCLEFmed 2015.

Islam, M., Zhang, D. and Lu, G. (2009). Region Based Color Image Retrieval Using Curvelet Transform. *Computer Vision – ACCV 2009*, pp.448-457.

Jeon, J., Lavrenko, V. and Manmatha, R. (2004). Automatic image annotation and retrieval using cross-media relevance models. In: *26th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.119–126.

Jin, J. and Nakayama, H. (2016). Annotation order matters: recurrent image annotator for arbitrary length image tagging. In: *23rd International Conference on Pattern Recognition (ICPR)*, pp.2452–2457.

Johnson, J., Ballan, L. and Fei-Fei, L. (2015). Love Thy Neighbors: Image Annotation by Exploiting Image Metadata. *2015 IEEE International Conference on Computer Vision (ICCV)*.

Kalpathy-Cramer, J. and Hersh, W. (2008). Medical Image Retrieval and Automatic Annotation: OHSU at ImageCLEF 2007. *Lecture Notes in Computer Science*, pp.623-630.

Keysers, D., Dahmen, J. and Ney, H. (2003). Statistical framework for model-based image retrieval in medical applications. *Journal of Electronic Imaging*, 12(1), p.59.

Keysers, D., Gollan, C. and Ney, H. (2004). Classification of Medical Images Using Non-linear Distortion Models. *Informatik aktuell*, pp.366-370.

Ko, B., Lee, J. and Nam, J. (2012). Automatic medical image annotation and keyword-based image retrieval using relevance feedback. *Journal of Digital Imaging*, 25(4), pp.454-465.

Krishna, A. and Prasad, B. (2012). Automated Image Annotation for Semantic Indexing and Retrieval of Medical Images. *International Journal of Computer Applications*, 55(3), pp.26-33.

Krizhevsky, A., Sutskever, I. and Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), pp.84-90.

Kumar, A., Dyer, S., Kim, J., Li, C., Leong, P., Fulham, M. and Feng, D. (2016). Adapting content-based image retrieval techniques for the semantic annotation of medical images. *Computerized Medical Imaging and Graphics*, 49, pp.37-45.

Kuroda, K. and Hagiwara, M. (2002). An image retrieval system by impression words and specific object names–IRIS. *Neurocomputing*, 43(1-4), pp.259-276.

Lana-Serrano, S., Villena-Román, J., González-Cristóbal, J. and Goñi-Menoyo, J. (2008). MIRACLE at ImageCLEFanot 2007: Machine Learning Experiments on Medical Image Annotation. *Lecture Notes in Computer Science*, pp.597-600.

Lana-Serrano, S., Villena-Román, J., González-Cristóbal, J. and Goñi-Menoyo, J. (2009). MIRACLE at ImageCLEFannot 2008: Nearest Neighbour Classification of Image Feature Vectors for Medical Image Annotation. *Lecture Notes in Computer Science*, pp.728-731.

Lavrenko, V., Manmatha, R. and Jeon, J. (2004). A Model for Learning the Semantics of Pictures. *NIPS*.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature*, 521(7553), pp.436-444.

Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), pp.2278-2324.

Lehmann, T., Güld, M., Deselaers, T., Keysers, D., Schubert, H., Spitzer, K., Ney, H. and Wein, B. (2005). Automatic categorization of medical images for content-based retrieval and data mining. *Computerized Medical Imaging and Graphics*, 29(2-3), pp.143-155.

Lehmann, T., Schubert, H., Keysers, D., Kohnen, M. and Wein, B. (2003). The IRMA code for unique classification of medical images. *Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation*.

Li, X., Shen, B., Liu, B. and Zhang, Y. (2016). A Locality Sensitive Low-Rank Model for Image Tag Completion. *IEEE Transactions on Multimedia*, 18(3), pp.474-483.

Li, X., Zhang, Y., Shen, B. and Liu, B. (2014). Image tag completion by low-rank factorization with dual reconstruction structure preserved. In: *International Conference on Image Processing (ICIP)*. New Jersey: IEEE, pp.3062–3066.

Lienhart, R., Romberg, S. and Hörster, E. (2009). Multilayer pLSA for multimodal image retrieval. In: *International Conference on Image and Video Retrieval*. New York: ACM.

Lin, W., Ke, S. and Tsai, C. (2016). Robustness and reliability evaluations of image annotation. *The Imaging Science Journal*, 64(2), pp.94-99.

Lin, Z., Ding, G., Hu, M., Lin, Y. and Sam Ge, S. (2014). Image tag completion via dual-view linear sparse reconstructions. *Computer Vision and Image Understanding*, 124, pp.42-60.

Lin, Z., Ding, G., Hu, M., Wang, J. and Sun, J. (2012). Automatic image annotation using tag-related random search over visual neighbors. In: *21st ACM international conference on Information and knowledge management*. ACM, pp.1784-1788.

Lin, Z., Ding, G., Hu, M., Wang, J. and Ye, X. (2013). Image Tag Completion via Image-Specific and Tag-Specific Linear Sparse Reconstructions. (2013). In: *Conference on Computer Vision and Pattern Recognition*. New Jersey: IEEE, pp.1618-1625.

Liu, J., Hu, Y., Li, M., Ma, S. and Ma, W. (2007). Medical Image Annotation and Retrieval Using Visual Features. *Evaluation of Multilingual and Multi-modal Information Retrieval*, pp.678-685.

Liu, J., Wang, B., Li, M., Li, Z., Ma, W., Lu, H. and Ma, S. (2007). Dual cross-media relevance model for image annotation. *Proceedings of the 15th international conference on Multimedia - MULTIMEDIA '07*.

Liu, Y., Zhang, D. and Lu, G. (2008). Region-based image retrieval with high-level semantics using decision tree learning. *Pattern Recognition*, 41(8), pp.2554-2570.

Llorente, A., Manmatha, R. and Rüger, S. (2010). Image retrieval using Markov Random Fields and global image features. *Proceedings of the ACM International Conference on Image and Video Retrieval - CIVR '10*.

Long, F., Zhang, H. and Feng, D. (2003). Fundamentals of Content-Based Image Retrieval. *Multimedia Information Retrieval and Management*, pp.1-26.

Lowe, D. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), pp.91-110.

Lyndon, D., Kumar, A., Kim, J., Leong, P. and Feng, D. (2015). Convolutional Neural Networks for Subfigure Classification. *CLEF*.

Ma, W. and Manjunath, B. (1997). NeTra: a toolbox for navigating large image databases. *Proceedings of International Conference on Image Processing*.

Maher, M. (2017). A Survey on Content-based Image Retrieval. *International Journal of Advanced Computer Science and Applications*, 8(5).

Makadia, A., Pavlovic, V. and Kumar, S. (2008). A New Baseline for Image Annotation. *Lecture Notes in Computer Science*, pp.316-329.

Malik, J., Belongie, S., Leung, T. and Shi, J. (2001). *International Journal of Computer Vision*, 43(1), pp.7-27.

Manjunath, B. and Ma, W. (1996). Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8), pp.837-842.

Manjunath, B., Salembier, P. and Sikora, T. (2002). *Introduction to MPEG-7*. Chichester: Wiley.

Marée, R., Geurts, P., Piater, J. and Wehenkel, L. (2005). Biomedical Image Classification with Random Subwindows and Decision Trees. *Computer Vision for Biomedical Image Applications*, pp.220-229.

Martinet, J. and Elsayad, I. (2012). Mid-Level Image Descriptors. In: L. Yan and Z. Ma, ed., *Intelligent Multimedia Databases and Information Retrieval*, 1st ed. Hershey: Information Science Reference, pp.46-60.

Marvasti, N., Yoruk, E. and Acar, B. (2017). Computer-Aided Medical Image Annotation: Preliminary Results with Liver Lesions in CT. *IEEE Journal of Biomedical and Health Informatics*, 22(5), pp.1561-1570.

Mayhew, M., Chen, B. and Ni, K. (2016). Assessing semantic information in convolutional neural network representations of images via image annotation. In: *International Conference on Image Processing (ICIP)*. New Jersey: IEEE, pp.2266–2270.

Mohammadi, S., Helfroush, M. and Kazemi, K. (2012). Novel shape-texture feature extraction for medical x-ray image classification. *International Journal of Innovative Computing, Information and Control*, 8(1), pp.659-676.

Mori, Y., Takahashi, H. and Oka, R. (1999). Image-to-word Transformation Based on Dividing and Vector Quantizing Images with Words.

Mori, Y., Takahashi, H. and Oka, R. (1999). Image-to-word transformation based on dividing and vector quantizing images with words. In: *Proceedings of the Seventh ACM International Conference on Multimedia*. ACM Press.

Mougiakakou, S., Valavanis, I., Nikita, K., Nikita, A. and Kelekis, D. (2003). Characterization of CT liver lesions based on texture features and a multiple neural network classification scheme. *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*.

Mueen, A., Zainuddin, R. and Baba, M. (2008). Automatic Multilevel Medical Image Annotation and Retrieval. *Journal of Digital Imaging*, 21(3), pp.290-295.

Mukherjea, S., Hirata, K. and Hara, Y. (1999). *World Wide Web*, 2(3), pp.115-132.

Müller, H., Deselaers, T., Deserno, T., Clough, P., Kim, E. and Hersh, W. (2007). Overview of the ImageCLEFmed 2006 Medical Retrieval and Medical Annotation Tasks. *Evaluation of Multilingual and Multi-modal Information Retrieval*, pp.595-608.

Müller, H., Geissbühler, A., Marty, J., Lovis, C. and Ruch, P. (2006). The Use of MedGIFT and EasyIR for ImageCLEF 2005. *Accessing Multilingual Information Repositories*, pp.724-732.

Murthy, V., Maji, S. and Manmatha, R. (2015). Automatic Image Annotation using Deep Learning Representations. *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval - ICMR '15*.

Nagarajan, S. and Saravanan, S. (2012). Content-based Medical Image Annotation and Retrieval using Perceptual Hashing Algorithm. *IOSR Journal of Engineering*, 02(04), pp.814-818.

Nalini, P. and Malleswari, B. (2017). An empirical study and comparative analysis of medical image retrieval and classification techniques. *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*.

Nandpuru, H., Salankar, S. and Bora, V. (2014). MRI brain cancer classification using Support Vector Machine. In: *IEEE Students' Conference on Electrical, Electronics and Computer Science*. New Jersey: IEEE.

Nedjar, I., Mahmoudi, S., Chikh, A., Abi-Yad, K. and Bouafia, Z. (2015). Automatic Annotation of Liver CT Image: ImageCLEFmed 2015. *CLEF*.

Niu, Y., Lu, Z., Wen, J., Xiang, T. and Chang, S. (2017). Multi-Modal Multi-Scale Deep Learning for Large-Scale Image Annotation. *IEEE Transactions on Image Processing*, 28(4), pp.1720-1731.

Ojala, T., Pietikainen, M. and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), pp.971-987.

Othman, M., Abdullah, N. and Kamal, N. (2011). MRI brain classification using support vector machine. *2011 Fourth International Conference on Modeling, Simulation and Applied Optimization*.

Park, K., Jeong, J. and Lee, D. (2007). OLYBIA: Ontology-Based Automatic Image Annotation System Using Semantic Inference Rules. *Advances in Databases: Concepts, Systems and Applications*, pp.485-496.

Pass, G. and Zabih, R. (1996). Histogram refinement for content-based image retrieval. In: *Third IEEE Workshop on Applications of Computer Vision. WACV'96*. New Jersey: IEEE, pp.96-102.

Pentland, A., Picard, R. and Sclaroff, S. (1996). Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3), pp.233-254.

Petrakis, E. and Faloutsos, A. (1997). Similarity searching in medical image databases. *IEEE Transactions on Knowledge and Data Engineering*, 9(3), pp.435-447.

Pinhas, A. and Greenspan, H. (2004). A continuous and probabilistic framework for medical image representation and categorization. *Medical Imaging 2004: PACS and Imaging Informatics*.

Putthividhy, D., Attias, H. and Nagarajan, S. (2010). Topic regression multi-modal Latent Dirichlet Allocation for image annotation. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. New Jersey: IEEE, pp.3408–3415.

Qi, X. and Han, Y. (2007). Incorporating multiple SVMs for automatic image annotation. *Pattern Recognition*, 40(2), pp.728-741.

Qiu, B. (2007). A Refined SVM Applied in Medical Image Annotation. *Evaluation of Multilingual and Multi-modal Information Retrieval*, pp.690-693.

Rahman, M., Desai, B. and Bhattacharya, P. (2005). Supervised Machine Learning Based Medical Image Annotation and Retrieval in ImageCLEFmed 2005. *Accessing Multilingual Information Repositories*, pp.692-701.

Rahman, M., Sood, V., Desai, B. and Bhattacharya, P. (2007). CINDI at ImageCLEF 2006: Image Retrieval & Annotation Tasks for the General Photographic and Medical Image Collections. *Evaluation of Multilingual and Multi-modal Information Retrieval*, pp.715-724.

Rajini, N. and Bhavani, R. (2011). Classification of MRI brain images using k-nearest neighbor and artificial neural network. *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*.

Rosten, E. and Drummond, T. (2006). Machine Learning for High-Speed Corner Detection. *Computer Vision – ECCV 2006*, pp.430-443.

Rublee, E., Rabaud, V., Konolige, K. and Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision*.

Rui, Y., Huang, T., Ortega, M. and Mehrotra, S. (1998). Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5), pp.644-655.

Sapkota, M., Xing, F., Su, H. and Yang, L. (2015). Automatic muscle perimysium annotation using deep convolutional neural network. *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*.

Schapire, R. and Singer, Y. (2000). *Machine Learning*, 39(2/3), pp.135-168.

Schmid, C. and Mohr, R. (1997). Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5), pp.530-535.

Sebe, N., Tian, Q., Loupias, E., Lew, M. and Huang, T. (2003). Evaluation of salient point techniques. *Image and Vision Computing*, 21(13-14), pp.1087-1095.

Setia, L., Teynor, A., Halawani, A. and Burkhardt, H. (2007). Grayscale Radiograph Annotation Using Local Relational Features. *Evaluation of Multilingual and Multi-modal Information Retrieval*, pp.644-651.

Setia, L., Teynor, A., Halawani, A. and Burkhardt, H. (2008). Grayscale medical image annotation using local relational features. *Pattern Recognition Letters*, 29(15), pp.2039-2045.

Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), pp.888-905.

Smeulders, A., Worring, M., Santini, S., Gupta, A. and Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), pp.1349-1380.

Smith, J. and Chang, S. (1997). VisualSEEk. *Proceedings of the fourth ACM international conference on Multimedia - MULTIMEDIA '96*.

Sohail, A., Bhattacharya, P., Mudur, S., Krishnamurthy, S. and Gilbert, L. (2010). Content-Based Retrieval and Classification of Ultrasound Medical Images of Ovarian Cysts. *Artificial Neural Networks in Pattern Recognition*, pp.173-184.

Spanier, A. and Joskowicz, L. (2014). Towards Content-Based Image Retrieval: From Computer Generated Features to Semantic Descriptions of Liver CT Scans. *CLEF*.

Springmann, M. and Schuldt, H. (2008). Speeding Up IDM without Degradation of Retrieval Quality. *Lecture Notes in Computer Science*, pp.607-614.

Stricker, M. and Orengo, M. (1995). Similarity of color images. *Storage and Retrieval for Image and Video Databases III*.

Sumana, I., Islam, M., Zhang, D. and Lu, G. (2008). Content based image retrieval using curvelet transform. *2008 IEEE 10th Workshop on Multimedia Signal Processing*.

Swain, M. and Ballard, D. (1991). Color indexing. *International Journal of Computer Vision*, 7(1), pp.11-32.

Swets, D. and Weng, J. (1996). Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8), pp.831-836.

Tamura, H., Mori, S. and Yamawaki, T. (1978). Textural Features Corresponding to Visual Perception. *IEEE Transactions on Systems, Man, and Cybernetics*, 8(6), pp.460-473.

Tang, J. and Lewis, P. (2007). A Study of Quality Issues for Image Auto-Annotation with the Corel Dataset. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(3), pp.384-389.

Tao, W., Jin, H. and Zhang, Y. (2007). Color Image Segmentation Based on Mean Shift and Normalized Cuts. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 37(5), pp.1382-1389.

Tian, F. and Shen, X. (2014). Learning Label Set Relevance for Search Based Image Annotation. In: *International Conference on Virtual Reality and Visualization*. New Jersey: IEEE, pp.260-265.

Tian, G., Fu, H. and Dagan Feng, D. (2008). Automatic medical image categorization and annotation using LBP and MPEG-7 edge histograms. *2008 International Conference on Technology and Applications in Biomedicine*.

Tommasi, T., Caputo, B., Welter, P., Güld, M. and Deserno, T. (2010). Overview of the CLEF 2009 Medical Image Annotation Track. *Lecture Notes in Computer Science*, pp.85-93.

Tommasi, T., Orabona, F. and Caputo, B. (2008). CLEF2008 Image Annotation Task: an SVM Confidence-Based Approach. *CLEF*.

Tommasi, T., Orabona, F. and Caputo, B. (2008). Discriminative cue integration for medical image annotation. *Pattern Recognition Letters*, 29(15), pp.1996-2002.

Tsai, C. and Hung, C. (2008). Automatically Annotating Images with Keywords: A Review of Image Annotation Systems. *Recent Patents on Computer Sciencee*, 1(1), pp.55-68.

Tuytelaars, T. and Van Gool, L. (1999). Content-Based Image Retrieval Based on Local Affinely Invariant Regions. *Visual Information and Information Systems*, pp.493-500.

Ünay, D., Soldea, O., Ozogur-Akyuz, S., Cetin, M. and Ercil, A. (2009). Automated X-Ray Image Annotation. *Lecture Notes in Computer Science*, pp.247-254.

Verma, Y. and Jawahar, C. (2012). Image Annotation Using Metric Learning in Semantic Neighbourhoods. *Computer Vision – ECCV 2012*, pp.836-849.

Villena-Román, J., Gonzalez, J., Goñi-Menoyo, J. and Martínez-Fernán, J. (2005). MIRACLE's naive approach to medical images annotation. *CLEF*.

Vinyals, O., Toshev, A., Bengio, S. and Erhan, D. (2015). Show and tell: a neural image caption generator. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. New Jersey: IEEE, pp.3156–3164.

Wang, C., Yan, S., Zhang, L. and Zhang, H. (2009). Multi-label sparse coding for automatic image annotation. In: *Conference on Computer Vision and Pattern Recognition*. New Jersey: IEEE, pp.1643–1650.

Wang, C., Yan, S., Zhang, L. and Zhang, H. (2009). Multi-label sparse coding for automatic image annotation. *2009 IEEE Conference on Computer Vision and Pattern Recognition*.

Wang, H., Huang, H. and Ding, C. (2011). Image annotation using bi-relational graph of images and semantic labels. In: *Computer Vision and Pattern Recognition*. New Jersey: IEEE, pp.793-800.

Wang, H., Huang, H. and Ding, C. (2011). Image annotation using bi-relational graph of images and semantic labels. *CVPR 2011*.

Wang, J., Wiederhold, G., Firschein, O. and Xin Wei, S. (1998). Content-based image indexing and searching using Daubechies' wavelets. *International Journal on Digital Libraries*, 1(4), pp.311-328.

Wang, J., Yang, Y., Mao, j., Huang, Z., Huang, C. and Xu, W. (2016). CNN-RNN: A Unified Framework for Multi-label Image Classification. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. New Jersey: IEEE, pp.2285-2294.

Wang, R., Xie, Y., Yang, J., Xue, L., Hu, M. and Zhang, Q. (2017). Large scale automatic image annotation based on convolutional neural network. *Journal of Visual Communication and Image Representation*, 49, pp.213-224.

Wang, X., Zhang, L., Jing, F. and Ma, W. (2008). AnnoSearch: Image Auto-Annotation by Search. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*.

Wei, C. and Chen, S. (2012). Annotation of Medical Images. In: L. Yan and Z. Man, ed., *Intelligent Multimedia Databases and Information Retrieval*, 1st ed. Hershey: Information Science Reference, pp.74-90.

Wennerberg, P., Schulz, K. and Buitelaar, P. (2011). Ontology modularization to improve semantic medical image annotation. *Journal of Biomedical Informatics*, 44(1), pp.155-162.

Wojnar, A. and Pinheiro, A. (2012). Annotation of medical images using the SURF descriptor. *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*.

Wu, B., Lyu, S., Hu, B. and Ji, Q. (2015). Multi-label learning with missing labels for image annotation and facial action unit recognition. *Pattern Recognition*, 48(7), pp.2279-2289.

Wu, L., Hoi, S., Jin, R., Zhu, J. and Yu, N. (2009). Distance metric learning from uncertain side information with application to automated photo tagging. *Proceedings of the seventeen ACM international conference on Multimedia - MM '09*.

Wu, L., Jin, R. and Jain, A. (2013). Tag Completion for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3), pp.716-727.

Wu, P., Hoi, S., Zhao, P. and He, Y. (2011). Mining social images with distance metric learning for automated image tagging. In: *Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM*, pp.197-206.

Xia, T., Tao, D., Mei, T. and Zhang, Y. (2010). Multiview Spectral Embedding. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(6), pp.1438-1446.

Xiang, Y., Zhou, X., Chua, T. and Ngo, C. (2009). A revisit of generative model for automatic image annotation using markov random fields. In: *Conference on Computer Vision and Pattern Recognition*. New Jersey: IEEE, pp.1153–1160.

Xie, B., Mu, Y., Tao, D. and Huang, K. (2010). m-SNE: Multiview Stochastic Neighbor Embedding. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(4), pp.1088-1096.

Xiong, W., Qiu, B., Tian, Q., Xu, C., Ong, S. and Foong, K. (2006). Combining Visual Features for Medical Image Retrieval and Annotation. *Accessing Multilingual Information Repositories*, pp.632-641.

Xu, C., Tao, D. and Xu, C. (2013). A Survey on Multi-view Learning. (CoRRabs/1304.5634, arXvi:1304.5634).

Yang, Y., Zhang, W. and Xie, Y. (2015). Image automatic annotation via multi-view deep representation. *Journal of Visual Communication and Image Representation*, 33, pp.368-377.

Yu, H., Li, M., Zhang, H. and Feng, J. (2002). Color texture moments for content-based image retrieval. In: *Proceedings. International Conference on Image Processing*. New Jersey: IEEE.

Yu, J., Rui, Y., Tang, Y. and Tao, D. (2014). High-Order Distance-Based Multiview Stochastic Learning in Image Classification. *IEEE Transactions on Cybernetics*, 44(12), pp.2431-2442.

Zare, M., Mueen, A. and Seng, W. (2013). Automatic Medical X-ray Image Classification using Annotation. *Journal of Digital Imaging*, 27(1), pp.77-89.

Zare, M., Seng, W. and Mueen, A. (2013). Automatic Classification Of Medical X-Ray Images. *Malaysian Journal of Computer Science*, 26(1), pp.9-22.

Zha, Z., Hua, X., Mei, T., Wang, J., Qi, G. and Wang, Z. (2008). Joint multi-label multi-instance learning for image classification. In: *Conference on Computer Vision and Pattern Recognition*. New Jersey: IEEE.

Zhang, D. and Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, 37(1), pp.1-19.

Zhang, D., Islam, M. and Lu, G. (2012). A review on automatic image annotation techniques. *Pattern Recognition*, 45(1), pp.346-362.

Zhang, G., Hsu, C., Lai, H. and Zheng, X. (2017). Deep learning based feature representation for automated skin histopathological image annotation. *Multimedia Tools and Applications*, 77(8), pp.9849-9869.

Zhang, G., Yin, J., Su, X., Huang, Y., Lao, Y., Liang, Z., Ou, S. and Zhang, H. (2014). Augmenting Multi-Instance Multilabel Learning with Sparse Bayesian Models for Skin Biopsy Image Analysis. *BioMed Research International*, 2014, pp.1-13.

Zhang, J., Xie, Y., Wu, Q. and Xia, Y. (2019). Medical image classification using synergic deep learning. *Medical Image Analysis*, 54, pp.10-19.

Zhang, M. and Zhou, Z. (2014). A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8), pp.1819-1837.

Zhang, Y. and Zhou, Z. (2010). Multilabel dimensionality reduction via dependence maximization. *ACM Transactions on Knowledge Discovery from Data*, 4(3), pp.1-21.

Zhou, X., Eggel, I. and Müller, H. (2010). The MedGIFT Group at ImageCLEF 2009. *Lecture Notes in Computer Science*, pp.211-218.

Zhou, X., Gobeill, J. and Müller, H. (2009). The MedGIFT Group at ImageCLEF 2008. *Lecture Notes in Computer Science*, pp.712-718.

Zhou, X., Gobeill, J., Ruch, P. and Müller, H. (2008). University and Hospitals of Geneva Participating at ImageCLEF 2007. *Lecture Notes in Computer Science*, pp.649-656.

Zhou, Z. and Zhang, M. (2007). Multi-instance multi-label learning with application to scene classification. In: B. Schoikopf, J. Platt and T. Hoffman, ed., *Advances in Neural Information Processing Systems 19*, 1st ed. MIT Press, pp.1609–1616.

Zhu, S. and Yuille, A. (1996). Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9), pp.884-900.

Διαδικτυακές πηγές

Anon, (2019). *NUS-WIDE*. [online] Available at: https://lms.comp.nus.edu.sg/wp-content/uploads/2019/research/nuswide/NUS-WIDE.html [Accessed 5 Oct. 2019].

Clef-initiative.eu. (2019). *The CLEF Initiative (Conference and Labs of the Evaluation Forum) - Homepage*. [online] Available at: http://www.clef-initiative.eu/web/clef-initiative/home [Accessed 5 Dec. 2019].

Cocodataset.org. (2019). *COCO - Common Objects in Context*. [online] Available at: http://cocodataset.org/#home [Accessed 5 Oct. 2019].

Imageclef.org. (2019). *ImageCLEF - The CLEF Cross Language Image Retrieval Track | ImageCLEF / LifeCLEF - Multimedia Retrieval in CLEF*. [online] Available at: https://www.imageclef.org [Accessed 5 Nov. 2019].

Imageclef.org. (2019). *IAPR TC-12 Benchmark | ImageCLEF / LifeCLEF - Multimedia Retrieval in CLEF*. [online] Available at: https://www.imageclef.org/photodata [Accessed 5 Oct. 2019].

Imageclef.org. (2019). *ImageCLEF - The CLEF Cross Language Image Retrieval Track | ImageCLEF / LifeCLEF - Multimedia Retrieval in CLEF*. [online] Available at: https://www.imageclef.org [Accessed 5 Oct. 2019].

Image-net.org. (2019). *ImageNet Large Scale Visual Recognition Competition (ILSVRC)*. [online] Available at: http://www.image-net.org/challenges/LSVRC/ [Accessed 5 Oct. 2019].

MedGIFT. (2019). *MedGIFT*. [online] Available at: http://medgift.hevs.ch/wordpress/ [Accessed 5 Dec. 2019].

Sites.google.com. (2019). *The COREL Database for Content based Image Retrieval - dct-research*. [online] Available at: https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval [Accessed 5 Oct. 2019].