

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



Πτυχιακή Εργασία

Σχεδιασμός/Υλοποίηση Συστήματος Υποστήριξης Αποφάσεων για τη Διαχείριση Ασθενών
με Καρκίνο του Μαστού

Ιωάννα Παγούνη

Π2019052

Επιβλέπων: Θεμιστοκλής Έξαρχος

30 Σεπτεμβρίου 2024

Επιβλέπων

Θεμιστοκλής Έξαρχος, Αναπληρωτής Καθηγητής,
Ιόνιο Πανεπιστήμιο

Τριμελής Επιτροπή

Θεμιστοκλής Έξαρχος, Καθηγητής,
Ιόνιο Πανεπιστήμιο

Αριστείδης Βραχάτης, Επίκουρος Καθηγητής,
Ιόνιο Πανεπιστήμιο

Παναγιώτης Κουρουθανάσης, Καθηγητής,
Ιόνιο Πανεπιστήμιο

Περίληψη

Ο καρκίνος του μαστού, αποτελεί την δεύτερη πιο συχνή αιτία θανάτου στις γυναίκες ανά τον κόσμο. Ασκεί βαθιά επίδραση τόσο σε ατομικό, όσο και σε συλλογικό επίπεδο, ενώ εξετάζονται στρατηγικές για την αντιμετώπισή του. Τεχνολογίες όπως αναλυτική δεδομένων υγείας, μπορεί να συνεισφέρουν στην καλύτερη κατανόηση του καρκίνου του μαστού και στην αποτελεσματική καταπολέμησή του ασκώντας θετική επίδραση. Ο συνδυασμός κανόνων και διαφορετικών τεχνικών μηχανικής μάθησης αποδεικνύονται ισχυρά εργαλεία για την έγκαιρη διάγνωση ή πρόβλεψη καρκίνου του μαστού και συμβάλλουν εξίσου στην καλύτερη κατανόησή του. Ωστόσο, έχουν εμφανιστεί ποικίλα Συστήματα Υποστήριξης Κλινικών Αποφάσεων, τα οποία αποτελούν σημαντική πρόοδο στην ιατρική πρακτική και παρέχουν συστηματική βοήθεια για τη λήψη κλινικών αποφάσεων. Σε αυτή την εργασία, αναπτύξαμε και εκπαιδεύσαμε δύο διαφορετικά μοντέλα πρόβλεψης για την κατανόηση της διαγνωστικής ακρίβειας του καρκίνου του μαστού. Στη συνέχεια, υλοποιήσαμε δύο εφαρμογές πρόβλεψης μηχανικής μάθησης και ενσωματώσαμε τα ήδη εκπαιδευμένα μοντέλα εντός των εφαρμογών. Αυτά τα μοντέλα εκπαιδεύθηκαν για δύο σύνολα δεδομένων, τα σύνολα δεδομένων Wisconsin Diagnostic Breast Cancer Dataset και Breast Cancer Coimbra Dataset. Μετρικές και πίνακες σύγχυσης αναδείχθηκαν με σκοπό την αξιολόγηση των αποτελεσμάτων και του χρόνου εκπαίδευσης, εξετάζοντας διαφορετικές περιπτώσεις για το καθένα. Οι εφαρμογές που αναπτύξαμε στόχο έχουν να αποδώσουν αξιόπιστα αποτελέσματα τόσο στους ασθενείς, όσο και να διευκολύνουν ιατρικό προσωπικό και ειδικούς στην διάγνωση και πρόγνωση καρκίνου του μαστού χάρη στα εύκαμπτα εργαλεία που διαθέτουν.

Πρόλογος

Ο καρκίνος του μαστού, είναι μια από τις πιο συχνές μορφές καρκίνου που εντοπίζονται στις γυναίκες. Αποτελεί την δεύτερη κυριότερη αιτία θανάτου των γυναικών ανά τον κόσμο. Ωστόσο, μπορεί να εντοπιστεί και στους άντρες, με μικρότερη συχνότητα από αυτή των γυναικών. Είναι μια διαδεδομένη και επακόλουθη μορφή καρκίνου, η οποία επηρεάζει τις γυναίκες παγκοσμίως.

Παρόλο που έχει σημειωθεί σπουδαία πρόοδος επάνω στην περαιτέρω διερεύνηση βασικών ζητημάτων της νόσου, τη σήμερον ημέρα, εξακολουθεί να απασχολεί τόσο επιστήμονες, όσο και ασθενείς. Αρχές και διαδικασίες όπως έγκαιρη διάγνωση, πρόγνωση και παροχή σύγχρονων τεχνολογιών στη θεραπεία, έχουν αυξήσει τα ποσοστά επιβίωσης από καρκίνο του μαστού. Όχι όμως και την πλήρη εξάλειψή του. Ολοένα και περισσότεροι άνθρωποι διαθέτουν πρόσβαση σε φάρμακα ή θεραπείες για έγκαιρη πρόληψη και ανίχνευση καρκίνου του μαστού. Τελευταία, χάρη στη ραγδαία εξέλιξη της τεχνολογίας, αυτές οι δυνατότητες παρέχονται και σε ανθρώπους που κατοικούν σε απομακρυσμένες περιοχές, με σκοπό τη διεξαγωγή εξετάσεων ή αυτοέλεγχου. Επίσης, παρέχονται λοιπές εναλλακτικές μέθοδοι. Συμβάλλουν σε λεπτομερή κατανόηση πιθανών συμπτωμάτων καρκίνου του μαστού, για την αποτελεσματική αντιμετώπισή του.

Στον ιατρικό και υγειονομικό τομέα, μελετώνται διάφορα Συστήματα Υποστήριξης Κλινικών Αποφάσεων. Σχεδιάστηκαν για την αντιμετώπιση περίπλοκων ασθενειών και προκλήσεων. Είναι αρκετά διαδεδομένη και αποτελεσματική προσέγγιση. Με την χρήση τεχνολογιών μηχανικής μάθησης, ανάλυσης δεδομένων, και προσεγγίσεις τεχνητής νοημοσύνης, αποτελούν ισχυρά εργαλεία στη λήψη εξατομικευμένων αποφάσεων για την αντιμετώπιση καρκίνου του μαστού.

Η τρέχουσα εργασία, αναλύει την ανάπτυξη ενός Συστήματος Υποστήριξης Αποφάσεων για τη διαχείριση ασθενών με καρκίνο του μαστού. Για την επίτευξη αυτού του στόχου, αναλύθηκαν δύο σύνολα δεδομένων. Μελετήθηκε μια ποικιλία αλγορίθμων μηχανικής μάθησης εκπαιδεύοντας δύο μοντέλα υψηλής απόδοσης. Επιλέχθηκαν για τη διαδικασία ενσωμάτωσης στις εφαρμογές που θα αναπτυχθούν σε τελικό στάδιο. Και τα δύο μοντέλα, ενσωματώθηκαν εντός δύο εφαρμογών μηχανικής μάθησης. Τόσο το σύστημα, όσο και οι εφαρμογές, στόχο έχουν να βοηθήσουν το ιατρικό προσωπικό στη λήψη εξατομικευμένων αποφάσεων για έγκαιρη διάγνωση καρκίνου του μαστού, καθώς και τη διευκόλυνση ασθενών στην απόδοση των τελικών αποτελεσμάτων.

Κέρκυρα, 30 Σεπτεμβρίου 2024

Ιωάννα Παγούνη

Περιεχόμενα

1. Εισαγωγή.....	1
1.1. Τι είναι ο καρκίνος του μαστού;.....	1
1.2. Κατανόηση καρκίνου του μαστού στις μέρες μας.....	1
1.3. Παράγοντες που επηρεάζουν την εμφάνιση καρκίνου του μαστού	2
1.4. Δράσεις πρόληψης κατά του καρκίνου του μαστού	2
2. Αντίκτυπος του καρκίνου του μαστού.....	3
2.1. Πώς ανταποκρίνεται ο αντίκτυπος του καρκίνου του μαστού;	3
2.2. Ρόλος του προσυμπτωματικού ελέγχου στην πρόληψη	4
2.3. Μέθοδοι για την αξιολόγηση της αποτελεσματικότητας προσυμπτωματικού ελέγχου καρκίνου του μαστού.....	4
2.4. Ο ρόλος της μαστογραφίας στον προσυμπτωματικό έλεγχο	5
2.4.1.Κατανομή ποσοστών θνησιμότητας από καρκίνο του μαστού διεθνώς	5
2.4.2.Κατανομή ποσοστών θνησιμότητας από καρκίνο του μαστού στην Ευρώπη	7
2.5. Ηλικιακή κατανομή της μαστογραφίας	9
2.5.1.Τι ισχύει για τις νεαρές γυναίκες;	10
2.6. Πόσο κοστίζει μια εξέταση μαστογραφίας;.....	10
2.6.1.Τι ισχύει για τα ελληνικά δεδομένα;.....	10
2.6.2.Τι παρατηρείται διεθνώς;.....	11
2.7. Γενικές παρατηρήσεις για τον έλεγχο μαστογραφίας.....	12
3. Συνεισφορά της Αναλυτικής Δεδομένων Υγείας	13
3.1. Συνεισφορά τεχνολογιών ωμικής	13
3.2. Συνεισφορά τεχνολογιών μηχανικής μάθησης	14
3.2.1.Τεχνολογίες μηχανικής μάθησης στον υγειονομικό τομέα	14
3.2.2.Τεχνολογίες μηχανικής μάθησης στον ιατρικό τομέα	15
3.2.3.Εκπαίδευση μοντέλων μηχανικής μάθησης.....	15
3.2.4.Αξιοποίηση ταξινομητών μηχανικής μάθησης.....	16
3.3. Γενικές αρχές της συνεισφοράς της αναλυτικής δεδομένων υγείας.....	16

4. Περιγραφή των σημαντικότερων εργασιών από τη βιβλιογραφία.....	17
4.1. Ανάπτυξη μοντέλων πρόβλεψης για την απόδοση αποτελεσμάτων και χρόνου εκπαίδευσης σε τέσσερις διαφορετικούς ταξινομητές.....	17
4.2. Ανάπτυξη μοντέλων για την πρόβλεψη και απόδοση αποτελεσμάτων σε τέσσερις διαφορετικούς ταξινομητές	23
4.3. Τι ισχύει για τα συστήματα υποστήριξης κλινικών αποφάσεων;	27
4.3.1. Ανάπτυξη συστήματος υποστήριξης αποφάσεων για την πρόγνωση καρκίνου του μαστού σε πρώιμο στάδιο	28
4.4. Τι ισχύει για λοιπές τεχνολογίες/συστήματα;	33
4.4.1. Ανάπτυξη διαδικτυακής εφαρμογής για βελτίωση γνώσεων σχετικά με τον καρκίνο του μαστού	34
4.4.2. Ανάπτυξη εφαρμογής για κινητά σχετικά με την ευαισθητοποίηση και γνώση του καρκίνου του μαστού.....	35
4.5. Τι παρατηρούμε από τις εργασίες που έχουν συλλεχθεί;	36
5. Περιγραφή των σετ δεδομένων από την βιβλιογραφία	37
5.1. Περιγραφή του σετ δεδομένων Wisconsin Diagnostic Breast Cancer Dataset	37
5.2. Περιγραφή του σετ δεδομένων Breast Cancer Coimbra Dataset	38
6. Περιγραφή των βημάτων προεπεξεργασίας δεδομένων.....	39
6.1. Περιγραφή των βημάτων για το πρώτο σετ δεδομένων	39
6.2. Περιγραφή των βημάτων για το δεύτερο σετ δεδομένων.....	40
7. Περιγραφή της μεθόδου προεπεξεργασίας δεδομένων	41
7.1. Περιγραφή της μεθοδολογίας για το πρώτο σετ δεδομένων	41
7.2. Περιγραφή της μεθοδολογίας για το δεύτερο σετ δεδομένων.....	42
7.3. Κοινά βήματα και για τα δύο σετ δεδομένων.....	44
7.3.1. Απεικόνιση δέντρων αποφάσεων για το πρώτο σύνολο δεδομένων	45
7.3.2. Απεικόνιση δέντρων αποφάσεων για το δεύτερο σύνολο δεδομένων.....	48

8. Ανάλυση των αποτελεσμάτων των δεδομένων	51
8.1. Αποτελέσματα για το πρώτο σετ δεδομένων.....	51
8.1.1. Αποτελέσματα που προέκυψαν από τον ταξινομητή δέντρων αποφάσεων.....	51
8.1.2. Αποτελέσματα που προέκυψαν από τον ταξινομητή τυχαίων δασών	52
8.1.3. Αποτελέσματα που προέκυψαν από τον ταξινομητή ADABOOST.....	53
8.1.4. Αποτελέσματα που προέκυψαν από τον ταξινομητή Naïve Bayes	53
8.1.5. Αποτελέσματα που προέκυψαν από τον ταξινομητή XGBoost.....	54
8.2. Συμπεράσματα από τα αποτελέσματα του πρώτου σετ δεδομένων	55
8.3. Αποτελέσματα για το δεύτερο σετ δεδομένων	56
8.3.1. Αποτελέσματα που προέκυψαν από τον ταξινομητή δέντρων αποφάσεων.....	56
8.3.2. Αποτελέσματα που προέκυψαν από τον ταξινομητή τυχαίων δασών	57
8.3.3. Αποτελέσματα που προέκυψαν από τον ταξινομητή ADABOOST.....	58
8.3.4. Αποτελέσματα που προέκυψαν από τον ταξινομητή Naïve Bayes	58
8.3.5. Αποτελέσματα που προέκυψαν από τον ταξινομητή XGBoost.....	59
8.4. Συμπεράσματα από τα αποτελέσματα του δεύτερου σετ δεδομένων	60
9. Περιγραφή των εφαρμογών πρόβλεψης.....	61
9.1. Περιγραφή των βημάτων για τη σχεδίαση και των δύο εφαρμογών	61
9.2. Όνομα, προδιαγραφές και περιεχόμενα της εφαρμογής.....	62
9.2.1. Απεικόνιση διεπαφής για την πρώτη εφαρμογή	63
9.2.2. Απεικόνιση διεπαφής για την δεύτερη εφαρμογή.....	66
10. Συμπεράσματα	69
Βιβλιογραφία.....	71
Συντμήσεις	76
Γλωσσάρι Ξενικών Όρων	77
Ευρετήριο	79

Κατάλογος Εικόνων

Εικόνα 1: Παράγοντες που αυξάνουν τον κίνδυνο εμφάνισης καθώς και μέθοδοι πρόληψης καρκίνου του μαστού.	2
Εικόνα 2: Τα στάδια που εκτελούνται κατά τη διαδικασία της μηχανικής μάθησης.	16
Εικόνα 3: Κατανομή κατηγοριών διάγνωσης καρκίνου του μαστού με τη μέθοδο διαγράμματος. Απεικονίζονται ξεχωριστά διαγράμματα για έξι από τις συνολικά τριάντα στήλες του πρώτου σετ δεδομένων.	18
Εικόνα 4: Διαγράμματα διασποράς για την απεικόνιση των κατηγοριών διάγνωσης καρκίνου του μαστού.	18
Εικόνα 5: Σύγκριση των αποτελεσμάτων των μετρικών που προκύπτουν από τον ταξινομητή SVM, μεταξύ γραμμικού και πολυωνυμικού πυρήνα.	19
Εικόνα 6: Ανάλυση και σύγκριση ταξινομητών γραμμικής παλινδρόμησης και απόδοσης διασταυρούμενης επικύρωσης (cross-validation) με και χωρίς την μέθοδο αναδρομικής εξάλειψης χαρακτηριστικών (recursive feature elimination - RFE).	20
Εικόνα 7: Σύγκριση των αποτελεσμάτων των μετρήσεων που προκύπτουν από τη μέθοδο K-κοντινότερου γείτονα (KNN) σε βασική και υπερπαραμετρική κατηγορία.	21
Εικόνα 8: Απεικόνιση των μετρικών της ακρίβειας και του F1-score σε ραβδόγραμμα, κάνοντας χρήση διαφορετικών ταξινομητών εντός του πειράματος.	22
Εικόνα 9: Πίνακας σύγχυσης και διαγραμματική απεικόνιση για την συνολική πρόβλεψη καρκίνου του μαστού, χρησιμοποιώντας πέντε τεχνικές μηχανικής μάθησης.	25
Εικόνα 10: Συγκεντρωτικό ραβδόγραμμα όλων των μετρικών για την πρόβλεψη του καρκίνου του μαστού, χρησιμοποιώντας πέντε τεχνικές μηχανικής μάθησης.	26
Εικόνα 11: Περιγραφή της λειτουργικότητας ενός συστήματος υποστήριξης κλινικών αποφάσεων. .	27
Εικόνα 12: Κατανομή κατηγοριών διάγνωσης καρκίνου του μαστού κάνοντας χρήση της τεχνικής προεπεξεργασίας κωδικοποίησης ετικετών. (label encoding).	29
Εικόνα 13: Απεικόνιση της σχέσης ακρίβειας μοντέλου-αριθμού εποχών σε μορφή γραφικής παράστασης, για τον αλγόριθμο τεχνητών νευρωνικών δικτύων (artificial neural networks – ANN). 30	
Εικόνα 14: Απεικόνιση της σχέσης απώλειας-αριθμού εποχών σε μορφή γραφικής παράστασης, για τον αλγόριθμο τεχνητών νευρωνικών δικτύων (artificial neural networks – ANN).	31
Εικόνα 15: Συγκεντρωτικός-συγκριτικός πίνακας για την σύγκριση μεταξύ αλγορίθμων μηχανικής και βαθιάς μάθησης για την πρόβλεψη του καρκίνου του μαστού.	32

Εικόνα 16: Απεικόνιση της κεφαλίδας της διεπαφής της διαδικτυακής εφαρμογής “Breast-4Y”.....	34
Εικόνα 17: Επισκόπηση της διαδικτυακής εφαρμογής.....	35
Εικόνα 18: Δέντρο απόφασης για αναδρομική εξάλειψη χαρακτηριστικών και τον ταξινομητή που προέκυψε κατά τη διάρκεια προσαρμογής χαρακτηριστικών.....	45
Εικόνα 19: Δέντρο απόφασης για μη-αναδρομική εξάλειψη χαρακτηριστικών και τον αρχικό ταξινομητή που προέκυψε.	46
Εικόνα 20: Δέντρο απόφασης για αναδρομική εξάλειψη χαρακτηριστικών και τον ταξινομητή που προέκυψε στο στάδιο της εξάλειψης χαρακτηριστικών.....	47
Εικόνα 21: Δέντρο απόφασης για αναδρομική εξάλειψη χαρακτηριστικών για νέο ταξινομητή, χρησιμοποιώντας ως παράμετρο τον ταξινομητή που προέκυψε στο στάδιο της εξάλειψης χαρακτηριστικών.....	48
Εικόνα 22: Δέντρο απόφασης για ταξινομητή που προέκυψε πριν την διαδικασία εκπαίδευσης.	49
Εικόνα 23: Δέντρο απόφασης για ταξινομητή που προέκυψε κατά τη διάρκεια της διαδικασίας εκπαίδευσης.	50
Εικόνα 24: Απεικόνιση της διεπαφής της εφαρμογής στην οθόνη.	63
Εικόνα 25: Απεικόνιση της διεπαφής της εφαρμογής στην οθόνη.	64
Εικόνα 26: Απεικόνιση του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος στο κάτω μέρος της οθόνης.	64
Εικόνα 27: Απεικόνιση του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος, μαζί με το αποτέλεσμα της πρόβλεψης.	65
Εικόνα 28: Απεικόνιση του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος, μαζί με το αποτέλεσμα της πρόβλεψης.	65
Εικόνα 29: Απεικόνιση της διεπαφής της εφαρμογής στην οθόνη.	66
Εικόνα 30: Απεικόνιση της διεπαφής της εφαρμογής στην οθόνη.	67
Εικόνα 31: Απεικόνιση του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος στο κάτω μέρος της οθόνης.	67
Εικόνα 32: Απεικόνιση του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος, μαζί με το αποτέλεσμα της πρόβλεψης.	68
Εικόνα 33: Απεικόνιση του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος, μαζί με το αποτέλεσμα της πρόβλεψης.	68

Κατάλογος Πινάκων

Πίνακας 1: Στατιστικά στοιχεία για το 2018 σε επιλεγμένες χώρες ανά τον κόσμο.....	6
Πίνακας 2: Στατιστικά στοιχεία για τα έτη 2012 και 2017 σε ευρωπαϊκές χώρες.....	7
Πίνακας 3: Πίνακας σύγκρισης για τα αποτελέσματα μέσω μηχανών υποστήριξης διανυσμάτων (support vector machines – SVM).	23
Πίνακας 4: Πίνακας σύγκρισης για τα αποτελέσματα μέσω της μεθόδου K-κοντινότερου γείτονα (K-nearest neighbor – KNN).	23
Πίνακας 5: Πίνακας σύγκρισης για τα αποτελέσματα μέσω του ταξινομητή τυχαίων δασών (Random Forest).	24
Πίνακας 6: Πίνακας σύγκρισης για τα αποτελέσματα μέσω του ταξινομητή τεχνητών νευρωνικών δικτύων (artificial neural networks – ANN).	24
Πίνακας 7: Πίνακας σύγκρισης για τα αποτελέσματα μέσω γραμμικής παλινδρόμησης (linear regression – LR).....	24

Κεφάλαιο 1

Εισαγωγή

Ο ΚΑΡΚΙΝΟΣ του μαστού, συγκαταλέγεται μεταξύ των πιο συχνών ειδών καρκίνου, που εμφανίζονται στις γυναίκες παγκοσμίως. Αποτελεί την δεύτερη κύρια αιτία θανάτων στις γυναίκες. Ωστόσο, μπορεί να εμφανιστεί και στους άνδρες με μικρότερη συχνότητα από αυτή των γυναικών.

Τι είναι ο καρκίνος του μαστού;

Ο καρκίνος του μαστού, αναπτύσσεται σε πολλαπλά στάδια. Αναφέρεται ως μια διαδικασία, που περιλαμβάνει πολλαπλούς τύπους κυττάρων. Η πρόληψη του καρκίνου του μαστού, ως στάδιο, αποτελεί πρόκληση. Μια από τις πιο αποτελεσματικές προσεγγίσεις για την πρόληψή του, είναι η έγκαιρη διάγνωση ύπαρξης καρκίνου του μαστού, δηλαδή, εάν ανιχνεύεται ή μπορεί να ανιχνευθεί στον άνθρωπο. Σε βιολογικό και γονιδιακό επίπεδο, ο καρκίνος του μαστού ορίζεται ως συλλογή από διακριτές κακοήθειες, που εκδηλώνονται στους μαστικούς αδένες. Η πλειοψηφία των καρκίνων του μαστού απαρτίζεται από καρκινώματα, ενώ κατηγορίες σαρκωμάτων όπως φυλλώδεις όγκοι και αγγειοσαρκώματα εμφανίζονται σπάνια.

Είναι μεταστατικός καρκίνος, με αποτέλεσμα να μπορεί συνήθως να μεταφέρεται σε ποικίλα όργανα και συστήματα του ανθρώπινου σώματος. Κυριότερα παραδείγματα, αποτελούν τα οστά, το συκώτι, οι πνεύμονες και ο εγκέφαλος. Έγκαιρη διάγνωση της νόσου, στοχεύει στην αποτελεσματική θεραπεία της, καλή πρόγνωση, και υψηλά ποσοστά επιβίωσης.

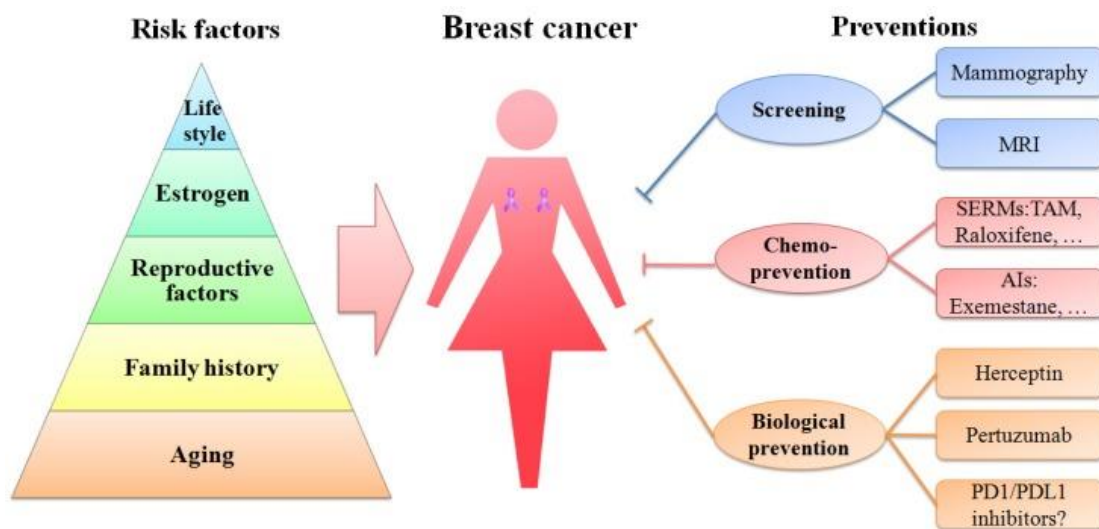
Κατανόηση καρκίνου του μαστού στις μέρες μας

Τελευταία, έχει σημειωθεί ραγδαία πρόοδος στην κατανόηση του καρκίνου του μαστού ως ασθένεια, καθώς και στην ανάπτυξη και εφαρμογή αποτελεσματικών μεθόδων πρόληψης. Όλο και σε περισσότερους ανθρώπους δίνεται πρόσβαση σε φάρμακα και θεραπείες για χημική πρόληψη και μεθόδους έγκαιρης ανίχνευσης συμπτωμάτων καρκίνου του μαστού. Έχουν ως κύριο στόχο τη διεξαγωγή εξετάσεων ή τον αυτοέλεγχο.

Για παράδειγμα, διεξαγωγή συχνών και τακτικών εξετάσεων σε εύλογο χρονικό διάστημα, καθώς και λοιπές εναλλακτικές μέθοδοι, όπως υποβολή σε μαστογραφία, υπέρηχο μαστών (ultrasound) ή εξέταση μαγνητικής τομογραφίας (MRI). Άλλες τεχνικές, έχουν επίσης μελετηθεί και επαληθευθεί.

Παράγοντες που επηρεάζουν την εμφάνιση καρκίνου του μαστού

Πολυάριθμοι παράγοντες μπορεί να αυξήσουν τον κίνδυνο εμφάνισης καρκίνου του μαστού. Μεταξύ αυτών, το φύλο, η ηλικία, ο τρόπος ζωής, το οικογενειακό ιστορικό¹, οιστρογόνα, ή γονιδιακές μεταλλάξεις.



Εικόνα 1: Παράγοντες που αυξάνουν τον κίνδυνο εμφάνισης καθώς και μέθοδοι πρόληψης καρκίνου του μαστού.

Πηγή: Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, Shi W, Jiang J, Yao PP, Zhu HP. 2017. Risk Factors and Preventions of Breast Cancer. *Int J Biol Sci.* 13(11):1387-1397.

Δράσεις πρόληψης κατά του καρκίνου του μαστού

Δράσεις όπως ευαισθητοποίηση και διαπαιδαγώγηση σχετικά με τον καρκίνο του μαστού, είναι σημαντικό βήμα για την αποτελεσματική καταπολέμηση της νόσου. Συνεχείς έρευνες και ραγδαία εξέλιξη της τεχνολογίας και της ιατρικής, έχουν στόχο την εξάλειψη των επιπτώσεων του καρκίνου του μαστού, τόσο σε άτομα ως μεμονωμένες οντότητες, όσο και στην κοινωνία ως σύνολο.

¹ εάν υπάρχει οικογενειακό ιστορικό σε μια γυναίκα

Κεφάλαιο 2

Αντίκτυπος του καρκίνου του μαστού

ΑΦΟΥ εξετάσαμε τις βασικές αρχές διάγνωσης και θεραπείας του καρκίνου του μαστού, καθώς και οι ευρύτερες έννοιες που τον διατρέχουν, είναι σημαντικό να εστιάσουμε στις πολυδιάστατες επιπτώσεις και τον αντίκτυπο που φέρει σε ατομικό και συλλογικό επίπεδο.

Πώς ανταποκρίνεται ο αντίκτυπος του καρκίνου του μαστού;

Ο καρκίνος του μαστού, αποτελεί την δεύτερη κυριότερη αιτία θανάτου στις γυναίκες. Είναι μια διαδεδομένη και επακόλουθη μορφή καρκίνου, η οποία επηρεάζει τις γυναίκες παγκοσμίως οδηγώντας σε σημαντικό αριθμό νεοδιαγνωσθέντων περιπτώσεων και θανάτων.

Ασκεί βαθιά επίδραση όχι μόνο σε ατομικό επίπεδο, αλλά και στην κοινωνία ως σύνολο. Αφενός, σε ατομικό επίπεδο, δεν περιορίζεται μόνο στη σωματική υγεία, καθώς επιβαρύνονται τόσο η ψυχική όσο και η συναισθηματική ευεξία των ασθενών και των οικογενειών τους. Αφετέρου, σε συλλογικό επίπεδο και την ευρύτερη κοινωνία, ο καρκίνος του μαστού έχει σοβαρές επιπτώσεις στον κοινωνικό και οικονομικό τομέα, εξαιτίας του κόστους εξέτασης και θεραπείας.

Αυτό το κεφάλαιο στοχεύει να αναλύσει διεξοδικά τον πολυεπίπεδο αντίκτυπο του καρκίνου του μαστού, εστιάζοντας τόσο στις άμεσες όσο και στις έμμεσες επιπτώσεις του, συμπεριλαμβανομένου των σωματικών, ψυχολογικών και οικονομικών βαρών που ασκούνται στους ασθενείς και τις οικογένειές τους. Επιπλέον, θα εξετάσουμε πιθανές στρατηγικές που αναπτύσσονται για την αποτελεσματική αντιμετώπιση και πρόληψη της νόσου. Δεδομένου ότι ο καρκίνος του μαστού είναι δύσκολο να προληφθεί, έγκαιρη ανίχνευση και αντιμετώπιση, μπορεί να βελτιώσει τα αποτελέσματα των ασθενών.

Ρόλος του προσυμπτωματικού ελέγχου στην πρόληψη

Ο προσυμπτωματικός έλεγχος του καρκίνου του μαστού, σε καθολικό επίπεδο, αποτελεί μια μέθοδο μείζονος σημασίας για την διάγνωση και θεραπεία συμπτωμάτων καρκίνου του μαστού. Στοχεύει στην ανίχνευση του καρκίνου του μαστού, σε πρώιμο στάδιο.

Η ακριβής διαφοροποίηση μεταξύ καλοηθών και κακοήθων όγκων βοηθά επίσης στην πρόληψη περιττών θεραπειών. Ως αποτέλεσμα, σημαντική έρευνα αφιερώνεται στην ακριβή διάγνωση και κατηγοριοποίηση καρκίνου του μαστού των ασθενών.

Μέθοδοι για την αξιολόγηση της αποτελεσματικότητας προσυμπτωματικού ελέγχου καρκίνου του μαστού

Οι μέθοδοι για την αξιολόγηση της αποτελεσματικότητας του προσυμπτωματικού ελέγχου του καρκίνου του μαστού, βασίζονται στην παρακολούθηση προσαρμοσμένων στην ηλικία ποσοστών εμφάνισης προχωρημένων καρκίνων. Αυτά τα ποσοστά θα πρέπει να μειωθούν μετά την εισαγωγή προσυμπτωματικού ελέγχου.

Ομοίως, τα ποσοστά θνησιμότητας του καρκίνου του μαστού, θα πρέπει να μειώνονται ταχύτερα σε περιοχές στις οποίες έχει υποβληθεί προσυμπτωματικός έλεγχος, παρά σε περιοχές με χαμηλότερα επίπεδα προσυμπτωματικού ελέγχου ή εάν δεν έχει πραγματοποιηθεί.

Η μείωση της θνησιμότητας από καρκίνο του μαστού είναι παρόμοια σε περιοχές με πρώιμη εισαγωγή και υψηλή διείσδυση προσυμπτωματικού ελέγχου, καθώς και σε περιοχές με καθυστερημένη εισαγωγή και χαμηλή διείσδυση αντίστοιχα.

Επίσης, η ιατρική απεικόνιση είναι κρίσιμης σημασίας για την αύξηση της αποτελεσματικότητας της διάγνωσης του καρκίνου του μαστού, στα αρχικά στάδιά του. Παρά το γεγονός ότι ο καρκίνος του μαστού μπορεί να θεραπευθεί, η έγκαιρη αναγνώριση είναι ζωτικής σημασίας για μια ευνοϊκή πρόγνωση. Ανίχνευση καρκίνου του μαστού σε πρώιμα στάδια αυξάνει τις πιθανότητες ίασής του εκθετικά.

Ο ρόλος της μαστογραφίας στον προσυμπτωματικό έλεγχο

Μια από τις πιο ευρεία διαδεδομένες μεθόδους προσυμπτωματικού ελέγχου, αποτελεί η μαστογραφία. Χρησιμοποιείται ευρέως, με σκοπό την ανίχνευση καρκίνου του μαστού. Συμβάλλει στην μείωση της θνησιμότητας από καρκίνο του μαστού. Μέχρι πρόσφατα, ήταν ο μόνος τρόπος εντοπισμού πιθανών συμπτωμάτων καρκίνου του μαστού.

Είναι μια εξέταση με αδιαμφισβήτητη αξία. Μπορεί να αποβεί σωτήρια για τον καρκίνο του μαστού. Έχει ως σκοπό τον εντοπισμό του καρκίνου προτού γίνει ογκίδιο. Είναι πιο εύκολα αντιμετωπίσιμος, χωρίς να απαιτείται συνήθως καμία συστηματική θεραπεία, παρά μόνο η χειρουργική αφαίρεσή του. Εξετάζεται και πραγματοποιείται μέσω του μηχανήματος του ψηφιακού μαστογράφου.

Η διάχυση προσυμπτωματικού ελέγχου μαστογραφίας στους πληθυσμούς, έχει αποδειχθεί σαφής περίπτωση επιτυχίας σε ένα μεγάλο μέρος της κοινωνίας. Έχει επίσης συμβάλει σημαντικά στην ενίσχυση της ευαισθητοποίησης του μαστού μεταξύ γυναικών και επαγγελματιών υγείας στη βελτίωση της οργάνωσης των υπηρεσιών διαχείρισης ασθενών. Αυτή η πρόοδος έχει ωφελήσει όλους τους ασθενείς με καρκίνο του μαστού, ανεξάρτητα από το αν οι καρκίνοι είναι ανιχνευόμενοι ή συμπτωματικοί.

Κατανομή ποσοστών θνησιμότητας από καρκίνο του μαστού διεθνώς

Σε ένα σύνολο πληθυσμού, τα ποσοστά των γυναικών που υποβάλλονται σε εξέταση μαστογραφίας, ενδέχεται να διαφέρουν μεταξύ ηλικιακών ομάδων. Παραδείγματος χάρη, οι γυναίκες ηλικίας 40 ετών και άνω έχουν μεγαλύτερο ποσοστό σε σχέση με τις νεότερες γυναίκες. Αυτό όμως, διαφέρει ανά χρονική περίοδο.

Οι παρακάτω πίνακες, περιγράφουν τα ποσοστά θνησιμότητας από καρκίνο του μαστού που καταγράφηκαν για το έτος 2018².

² σε επιλεγμένες χώρες ανά τον κόσμο.

Global

Countries	Incidence Rate
Belgium	113.2
Luxembourg	109.3
The Netherlands	105.9
France	99.1
New Caledonia, France	98.0
Lebanon	97.6
Australia	94.5
United Kingdom	93.6
Italy	92.8
New Zealand	92.6
Ireland	90.3
Sweden	89.8

Πίνακας 1: Στατιστικά στοιχεία για το 2018 σε επιλεγμένες χώρες ανά τον κόσμο.

Πηγή: Global Cancer Observatory/Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, in press.

Finland	89.5
Denmark	88.8
Switzerland	88.1
Malta	87.6
Norway	87.5
Hungary	85.5
Germany	85.4
Iceland	85.2
United States	84.9
Canada	83.8
Cyprus	81.7
Samoa	80.1

Πίνακας 1: Στατιστικά στοιχεία για το 2018 σε επιλεγμένες χώρες ανά τον κόσμο.

Πηγή: Global Cancer Observatory/Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, in press.

Κατανομή ποσοστών θνησιμότητας από καρκίνο του μαστού στην Ευρώπη

Στην Ευρώπη, τα ποσοστά θνησιμότητας από καρκίνο του μαστού, ενδέχεται να διαφέρουν ανά χώρα. Αυτό οφείλεται στην ύπαρξη πολυάριθμων παραγόντων που επηρεάζουν τις διαφορές μεταξύ των ποσοστών θνησιμότητας. Μεταξύ αυτών, η ποιότητα των υγειονομικών υπηρεσιών, η αξιοπιστία των προγραμμάτων πρόληψης ενάντια στον καρκίνο του μαστού, ο τρόπος ζωής ή η στάση ορισμένων κρατών απέναντι στην υγεία και την ιατρική περίθαλψη ως οι κυριότεροι παράγοντες.

Οι παρακάτω πίνακες, περιγράφουν τα ποσοστά θνησιμότητας από καρκίνο του μαστού που καταγράφηκαν σε διάστημα πέντε ετών, από το 2012 ως και το 2017³. σε διάφορες ευρωπαϊκές χώρες.

Country	All ages			
	2012	2017	Deaths	% change
EU countries				
Austria	14.1	13.5	1566	-4.5
Belarus	12.8	11.4	1138	-10.7
Belgium	17.6	15.3	2227	-13.3
Bulgaria	15.2	14.9	1312	-2.2
Croatia	18.1	15.8	853	-12.7
The Czech Republic	13.3	12.1	1620	-9.3
Denmark	17.5	14.8	1069	-15.3
Estonia	14.5	13.8	260	-4.7
Finland	13.3	13.0	908	-2.6
France	15.4	14.7	12,434	-4.4
Germany	16.1	15.4	18,396	-4.3
Greece	13.8	14.3	2163	3.8

Πίνακας 2: Στατιστικά στοιχεία για τα έτη 2012 και 2017 σε ευρωπαϊκές χώρες.

³ ποσοστά για διάφορες ευρωπαϊκές χώρες: δεν υπάρχουν διαθέσιμα στατιστικά στοιχεία για τις χώρες Ουκρανία, Κύπρος, Αλβανία ή Βοσνία-Ερζεγοβίνη

Πηγή: P. Koczkodaj, U. Sulkowska, J. Gotlib, M. Mańczuk, Breast cancer mortality trends in Europe among women in perimenopausal and postmenopausal age (45+), Arch Med Sci, 16 (1) (2020), pp. 146-156

Hungary	17.7	17.1	2123	-3.2
Iceland	13.8	14.7	47	7.0
Ireland	15.2	16.6	678	8.9
Italy	14.8	14.2	12,616	-4.1
Latvia	16.8	17.0	445	1.5
Lithuania	15.6	13.9	521	-10.9
Luxembourg	15.2	14.6	89	-3.6
Malta	16.5	14.4	73	-12.9
Moldova	16.6	16.0	505	-4.0
The Netherlands	17.4	15.5	3107	-11.4

Πίνακας 2: Στατιστικά στοιχεία για τα έτη 2012 και 2017 σε ευρωπαϊκές χώρες.

Πηγή: P. Koczkodaj, U. Sulkowska, J. Gotlib, M. Mańczuk, Breast cancer mortality trends in Europe among women in perimenopausal and postmenopausal age (45+), Arch Med Sci, 16 (1) (2020), pp. 146-156

North Macedonia	17.1	N/A	N/A	N/A
Norway	12.1	11.2	623	-7.1
Poland	14.3	15.1	6670	5.1
Portugal	13.1	13.1	1769	-0.1
Romania	15.2	15.8	3495	4.6
Russia	16.1	N/A	N/A	N/A
Serbia	20.4	20.7	1766	1.2

Πίνακας 2: Στατιστικά στοιχεία για τα έτη 2012 και 2017 σε ευρωπαϊκές χώρες.

Πηγή: P. Koczkodaj, U. Sulkowska, J. Gotlib, M. Mańczuk, Breast cancer mortality trends in Europe among women in perimenopausal and postmenopausal age (45+), Arch Med Sci, 16 (1) (2020), pp. 146-156

Slovakia	15.9	N/A	N/A	N/A
Slovenia	15.0	14.7	433	-2.1
Spain	11.7	10.9	6489	-7.3
Sweden	12.7	11.3	1413	-11.1
United Kingdom	16.2	15.1	11,512	-6.9

Πίνακας 2: Στατιστικά στοιχεία για τα έτη 2012 και 2017 σε ευρωπαϊκές χώρες⁴

Πηγή: P. Koczkodaj, U. Sulkowska, J. Gotlib, M. Mańczuk, Breast cancer mortality trends in Europe among women in perimenopausal and postmenopausal age (45+), Arch Med Sci, 16 (1) (2020), pp. 146-156

Ηλικιακή κατανομή της μαστογραφίας

Αναθεωρώντας τους ορισμούς που περιγράφουν τη μαστογραφία ως εξέταση μέγιστης σημασίας, είναι σημαντικό να προσδιορίσουμε ότι κάθε γυναίκα πρέπει να εντάξει στην καθημερινότητά της, τουλάχιστον μια φορά τον χρόνο, την υποβολή σε εξέταση ή υπέρηχο μαστογραφίας.

Από μια ηλικία και μετά, οι γυναίκες μπορούν να υποβληθούν σε μαστογραφία. Κάποιες γυναίκες το αμελούν, ενώ οι περισσότερες από αυτές δεν έχουν υποβληθεί ποτέ σε ανάλογη εξέταση. Όλες οι γυναίκες, βάσει ηλικίας και οικογενειακού ιστορικού, πρέπει να υποβληθούν σε τακτικό έλεγχο των μαστών τους.

Γυναίκες που δεν διαθέτουν οικογενειακό ιστορικό, πρέπει να κάνουν την πρώτη μαστογραφία τους μεταξύ 35 και 40 ετών. Αυτή η εξέταση ονομάζεται μαστογραφία αναφοράς. Αν μια γυναίκα έχει οικογενειακό ιστορικό, θα πρέπει να υποβληθεί σε μαστογραφία δέκα χρόνια νωρίτερα από την ηλικία στην οποία εμφανίστηκε καρκίνος του μαστού σε συγγενικό της πρόσωπο.

⁴ ** no available status for the countries of Ukraine, Bosnia & Herzegovina, Albania or Cyprus

Τι ισχύει για τις νεαρές γυναίκες;

Οι νεαρές γυναίκες⁵ δεν μπορούν να υποβληθούν σε μαστογραφία για δύο κύριους λόγους που διατρέχουν την κατάσταση του μαζικού τους αδένα. Αρχικά, ο μαζικός αδένας είναι πολύ πυκνός, με αποτέλεσμα η εξέταση να είναι χαμηλής διαγνωστικής αξίας. Επίσης, είναι πιο ευαίσθητος στην ακτινοβολία. Πρέπει να αποφεύγεται η έκθεση σε επιφάνειες με ακτινοβολία, εάν δεν συντρέχουν σοβαροί λόγοι. Τέλος, οι νεαρές γυναίκες χωρίς οικογενειακό ιστορικό μπορούν να κάνουν κλινική εξέταση και υπερηχογράφημα μαστών.

Εκτός του ότι η μαστογραφία αποτελεί μια κρίσιμης σημασίας εξέταση για την γυναίκα, μπορεί να επιτευχθεί η είσπραξη ενός ποσού για την πραγματοποίηση εξέτασης μαστογραφίας. Μιας και παίζει ρόλο η ηλικία της γυναίκας, η οποία υποβάλλεται σε μαστογραφία, οι μαστογραφίες τις περισσότερες φορές ενδέχεται να είναι ακριβές.

Πόσο κοστίζει μια εξέταση μαστογραφίας;

Κάθε χρόνο, το κόστος μιας εξέτασης μαστογραφίας, ενδέχεται να διαφέρει, ανεξαρτήτου χώρας καταγωγής ή τόπου διαμονής του ασθενούς. Αυτό δεν οφείλεται αποκλειστικά σε αυτούς τους δύο βασικούς παράγοντες, εφόσον επηρεάζουν το κόστος δευτερεύοντες παράγοντες. Μεταξύ αυτών, η ύπαρξη τεχνολογικών ή/και διαγνωστικών επιτευγμάτων, αλλαγών στη νομοθεσία ή εξωτερικοί οικονομικοί παράγοντες, λειτουργική επάρκεια, πολιτικές υγειονομικής περίθαλψης ανά χώρα ή τοποθεσία, συγκαταλέγονται ως οι πιο γνωστοί δευτερεύοντες παράγοντες που ενδέχεται να επηρεάσουν το κόστος μιας εξέτασης μαστογραφίας.

Γενικά, μια γυναίκα μπορεί να υποβληθεί σε εξέταση μαστογραφίας, ψηφιακή ή μαγνητική, ανάλογα με το αν είναι ασφαλισμένη ή όχι στην κρατική ασφάλεια υγείας, ή εγγεγραμμένη σε κάποιο ασφαλιστικό πρόγραμμα από την ασφάλεια που χρηματοδοτείται. Με αποτέλεσμα να πραγματοποιήσει την εξέταση με μειωμένο κόστος, ή εντελώς δωρεάν.

Τι ισχύει για τα ελληνικά δεδομένα;

Στην Ελλάδα, το κόστος μιας εξέτασης μαστογραφίας διαφέρει⁶, ανάλογα με το αν ένα διαγνωστικό εργαστήριο είναι συμβεβλημένο με τον Εθνικό Οργανισμό Παροχής Υπηρεσιών Υγείας (ΕΟΠΥΥ). Μια εξέταση μαστογραφίας, ανεξαρτήτου ηλικίας, ανέρχεται στην τιμή των τριάντα ευρώ, σε

⁵ Γυναίκες ηλικίας από 25 ετών και κάτω

⁶ Με τα δεδομένα του έτους 2021

περίπτωση που πραγματοποιείται σε διαγνωστικό εργαστήριο, το οποίο έχει ασφαλιστεί μέσω του Εθνικού Οργανισμού Παροχής Υπηρεσιών Υγείας (ΕΟΠΥΥ).

Σε περίπτωση που πραγματοποιηθεί εξέταση μαστογραφίας σε ιδιωτικό διαγνωστικό εργαστήριο που δεν έχει ασφαλιστεί ακόμα μέσω Εθνικού Οργανισμού Παροχής Υπηρεσιών Υγείας (ΕΟΠΥΥ), το κόστος της εξέτασης μαστογραφίας κυμαίνεται ανά εργαστήριο⁷, με αποτέλεσμα να ανέρχεται από την τιμή των πενήντα ευρώ και άνω.

Επίσης, ένας αξιοσημείωτος παράγοντας που επηρεάζει το κόστος μιας εξέτασης μαστογραφίας, είναι το είδος εξέτασης⁸ στο οποίο μπορεί να υποβληθεί ένας ασθενής. Αυτό, έχει ως αποτέλεσμα τη διαφοροποίηση του κόστους ανά εξέταση. Παραδείγματος χάρη, εάν ένας ασθενής υποβάλλεται σε μαγνητική μαστογραφία, η τιμή της είναι πιο ακριβή από την υποβολή σε ψηφιακή μαστογραφία. Αντίθετα, εάν υποβληθεί ένας ασθενής σε ψηφιακή μαστογραφία, θα λάβει μικρότερη τιμή και φθηνότερο κόστος σε σχέση με την μαγνητική μαστογραφία.

Τι παρατηρείται διεθνώς;

Το ακόλουθο παράδειγμα, αναλύει τη διαφορά του κόστους μεταξύ μιας εξέτασης ψηφιακής μαστογραφίας και μαγνητικής μαστογραφίας στις Ηνωμένες Πολιτείες της Αμερικής σε τρεις ξεχωριστές περιπτώσεις.

Η πρώτη περίπτωση, εξετάζει τα δεδομένα που εκτιμήθηκαν για τον υπολογισμό του συνολικού μέσου κόστους εξετάσεων μαστογραφίας για το έτος 2019. Το μέσο κόστος μιας ψηφιακής εξέτασης μαστογραφίας, κυμάνθηκε σε εύρος τιμών μεταξύ 139 έως 360 δολαρίων, ενώ το κόστος μαγνητικής μαστογραφίας, σε εύρος τιμών μεταξύ 545 έως 2.439 δολαρίων. Ετησίως, ένα ποσοστό 50% του προσυμπτωματικού ελέγχου που εκτιμήθηκε για συγκεκριμένη ομάδα γυναικών, ανήλθε στην τιμή των 11,5 δισεκατομμυρίων δολαρίων, εντός του ορίου τιμών μεταξύ 10,37 έως 13,94 δισεκατομμυρίων δολαρίων. Συνολικά, κόστισε 77.824 δολάρια ανά καρκίνο του μαστού που εντοπίστηκε σε κάθε γυναίκα. Για το 100% του προσυμπτωματικού ελέγχου, που εκτιμήθηκε για συγκεκριμένη ομάδα γυναικών, το κόστος ανήλθε στην τιμή των 26,71 δισεκατομμυρίων δολαρίων, εντός του ορίου μεταξύ 24,61 έως 32,82 δισεκατομμυρίων δολαρίων.

⁷ Δεν είναι η τιμή ίδια ανά ιατρικό εργαστήριο

⁸ μαγνητική ή ψηφιακή μαστογραφία

Η δεύτερη περίπτωση, εξετάζει τα δεδομένα που εκτιμήθηκαν για τον υπολογισμό του συνολικού κόστους υβριδικής συμμετοχής σε εξετάσεις μαστογραφίας⁹. Το συνολικό κόστος υβριδικής συμμετοχής σε εξετάσεις μαστογραφίας, κόστισε 18,09 δισεκατομμύρια δολάρια, εντός ορίου τιμών μεταξύ 16,03 έως 21,16 δισεκατομμυρίων δολαρίων. Με βάση τα δεδομένα που δόθηκαν, το συνολικό κόστος κόστισε 7,57 δισεκατομμύρια δολάρια, εντός ορίου τιμών μεταξύ 6,97 έως 9,29 δισεκατομμυρίων δολαρίων. Για στρατηγικές που βασίζονται στον κίνδυνο¹⁰, το συνολικό κόστος κόστισε 9,60 δισεκατομμύρια δολάρια, εντός ορίου τιμών μεταξύ 8,88 έως 11,70 δισεκατομμυρίων δολαρίων.

Η τρίτη και τελευταία περίπτωση, εξετάζει τα δεδομένα που εκτιμήθηκαν για ηλικιακή ομάδα με όριο ελέγχου μέχρι τα 74 έτη. Το εκτιμώμενο μέσο κόστος ανά γυναίκα, ετησίως, κόστισε 22.315 δολάρια. Στην περίπτωση της υβριδικής συμμετοχής σε εξετάσεις μαστογραφίας, το μέσο κόστος κόστισε 16.552 δολάρια ετησίως. Αντίθετα, το συνολικό κόστος για το διάστημα 2019-2021, κόστισε 4,133 δολάρια. Για στρατηγικές που βασίζονται στον κίνδυνο, το μέσο κόστος¹¹ ανήλθε στην τιμή των 6.878 δολαρίων.

Το κόστος μπορεί να μειωθεί σημαντικά. Αυτό παρατηρείται στις περιπτώσεις που σημειώνονται τόσο σε επίπεδο διατίας όσο και σε στρατηγικές με βάση τον κίνδυνο εξέτασης.

Ο προσυμπτωματικός έλεγχος, βάσει κινδύνου, μείωσε το κόστος, ακόμη και με το κόστος των γενετικών δοκιμών πληθυσμού. Παράλληλα, διατηρήθηκε ο εντατικός έλεγχος για τις γυναίκες υψηλότερου κινδύνου.

Γενικές παρατηρήσεις για τον έλεγχο μαστογραφίας

Γενικά, έχει παρατηρηθεί ότι ο έλεγχος μαστογραφίας οδηγεί σε μετανάστευση σταδίου προς τα κάτω. Αντανακλάται από μικρότερα μεγέθη όγκου και λιγότερο εκτεταμένη κομβική συμμετοχή. Σχετίζεται με λιγότερο επιθετική χειρουργική θεραπεία καρκίνου του μαστού, συμπεριλαμβανομένων υψηλότερων ποσοστών διατήρησης του μαστού. Οι παρατηρούμενες αλλαγές στη χειρουργική αντιμετώπιση θα πρέπει να επηρεάσουν τη λήψη ατομικών αποφάσεων σχετικά με την προληπτική μαστογραφία.

⁹ για τη διατία από 2019 έως 2021

¹⁰ εξετάζοντας την περίπτωση της υβριδικής συμμετοχής

¹¹ δεδομένα για ηλικιακή ομάδα με όριο ελέγχου τα 74 έτη ηλικίας

Κεφάλαιο 3

Συνεισφορά της Αναλυτικής Δεδομένων Υγείας

ΚΑΘΩΣ αναλύσαμε ένα μεγάλο μέρος του αντικτύπου του καρκίνου του μαστού και τις πολλαπλές διαστάσεις του προβλήματος, είναι σημαντικό να εξετάσουμε τη συνεισφορά της αναλυτικής δεδομένων υγείας σε αυτή τη μάχη.

Τα τελευταία χρόνια, η τεχνολογία και η συλλογή μεγάλων δεδομένων, έχουν εξελιχθεί ραγδαία. Σαν αποτέλεσμα, έχουν εφαρμοστεί ριζικές αλλαγές στον τρόπο προσέγγισης έρευνας και θεραπείας του καρκίνου του μαστού. Η αναλυτική δεδομένων υγείας, χάρη στην αξιοποίηση προηγμένων αλγορίθμων και εξειδικευμένων τεχνικών ανάλυσης δεδομένων, παρέχει καινοτόμα εργαλεία και πληροφορίες. Αυτές οι τεχνικές, στόχο έχουν την συμβολή στην καλύτερη κατανόηση του καρκίνου, τη βελτίωση διαγνωστικών διαδικασιών και την ανάπτυξη πιο αποτελεσματικών και εξατομικευμένων θεραπευτικών προσεγγίσεων επάνω στο ζήτημα του καρκίνου του μαστού.

Σε αυτό το κεφάλαιο, θα εστιάσουμε σε δύο σημαντικούς άξονες: την διερεύνηση καινοτόμων προσεγγίσεων και την ανάλυση του τρόπου με τον οποίο, η αναλυτική δεδομένων υγείας, μπορεί να ασκήσει θετική επιρροή στην καταπολέμηση του καρκίνου του μαστού.

Συνεισφορά τεχνολογιών ωμικής

Ένα σημαντικό κεφάλαιο στο οποίο πρέπει να δοθεί προσοχή, είναι η συνεισφορά των τεχνολογιών ωμικής στα στάδια της πρόληψης, ανίχνευσης και θεραπείας της νόσου. Τα τελευταία χρόνια, χάρη στην υψηλή τους απόδοση, τεχνολογίες και πεδία ωμικής έχουν δείξει τις μεγάλες δυνατότητές τους στη δημιουργία νέων γνώσεων για την μελέτη του καρκίνου του μαστού. Η υψηλή απόδοσή τους, οφείλεται στην αναγνώριση νέων βιοδεικτών και μοριακού προφίλ με βάση το πεδίο.

Η δημιουργία νέων γνώσεων, στόχο έχει τη διάγνωση, πρόγνωση και πρόβλεψη της ανταπόκρισης στη θεραπεία κατά του καρκίνου του μαστού. Προσεγγίσεις που βασίζονται σε μεταβολομικό NMR, LC-MS και GC-MS, αποτελούν ισχυρά εργαλεία για την αναγνώριση πιθανών βιοδεικτών σε μια ποικιλία κλινικών πεδίων.

Επιτεύχθηκαν μέθοδοι διαχείρισης ασθενών με καρκίνο του μαστού με τη χρήση βιοδεικτών. Η εφαρμογή πεδίων μονής και πολλαπλής ωμικής, δίνει τη δυνατότητα εξονυχιστικής έρευνας και σαφούς

κατανόησης μιας περίπλοκης νόσου, όπως ο καρκίνος του μαστού. Υπόσχεται παροχή νέων γνώσεων, πιθανές θεραπευτικές επιλογές και προσαρμοσμένη θεραπεία για την ακριβή διάγνωση και πρόληψη μέσω των δυνατοτήτων που παρέχουν.

Στην πραγματικότητα, ιδίως στο πλαίσιο της πολλαπλής ωμικής, είναι προφανές ότι η χρήση ολοκληρωμένων εργαλείων και υπολογιστικών προσεγγίσεων είναι απαραίτητη για την βαθιά κατανόηση των βιολογικών μηχανισμών από μια προοπτική σε όλο το σύστημα. Συνεπώς, το κύριο πλεονέκτημα αποτελεί η ανάδειξη μοριακού προφίλ με το κλινικό αποτέλεσμα υπό μελέτη.

Συνεισφορά τεχνολογιών μηχανικής μάθησης

Εκτός από την αξιοποίηση τεχνικών μονής και πολλαπλής ωμικής, τα τελευταία χρόνια ιδιαίτερα, έχει δοθεί βάση και στην ανάπτυξη τεχνολογιών μηχανικής μάθησης. Αυτές οι προσεγγίσεις έχουν αποδειχθεί ισχυρές για την έγκαιρη διάγνωση ή πρόβλεψη καρκίνου του μαστού. Σε αυτό, οφείλεται η υψηλή τους ζήτηση στην προτίμηση των καταναλωτών. Συνήθως, απαιτούν αρκετό χρόνο επεξεργασίας και διακομιδής της εκπαιδευτικής διαδικασίας.

Δεν είναι μόνο οι προσεγγίσεις μηχανικής μάθησης που κέρδισαν έδαφος από τις ευρεία διαδεδομένες νέες τεχνολογίες. Αντίθετα, οι τεχνικές εξόρυξης δεδομένων¹², αποτελούν αποτελεσματικό τρόπο ταξινόμησης δεδομένων. Χρησιμοποιούνται για την εξαγωγή πολύτιμων πληροφοριών και γνώσεων από μεγάλα σύνολα δεδομένων. Με την ίδια φιλοσοφία, λειτουργεί και η διαδικασία της στατιστικής ανάλυσης για την προσέγγιση του ζητήματος μέσω ποσοστών, σε ένα μέρος του συνόλου.

Τεχνολογίες μηχανικής μάθησης στον υγειονομικό τομέα

Στον υγειονομικό τομέα, ολοένα και περισσότερο αυξάνεται ο αντίκτυπος της χρήσης τεχνολογιών τεχνητής νοημοσύνης και μηχανικής μάθησης, ιδίως τα τελευταία χρόνια. Αυτές οι τεχνικές, μπορούν να συνεισφέρουν σημαντικά στη διαδικασία πρόβλεψης και έγκαιρης διάγνωσης του καρκίνου του μαστού. Έχουν αποδειχθεί ισχυρές τεχνικές. Αυτό, έχει ως αποτέλεσμα, τα ιατρικά δεδομένα να αποθηκεύονται σε διαφορετικές βάσεις. Αυξάνονται ραγδαία τόσο ο αριθμός όσο και το μέγεθος των βάσεων δεδομένων.

Ο τεράστιος αυτός όγκος δεδομένων, υπερβαίνει την ικανότητα ανάλυσης και αναζήτησης των παραδοσιακών μεθόδων μέχρι πρότινος, επάνω σε πρότυπα και πληροφορίες που μπορούν να καταστούν ενδιαφέροντα. Ως εκ τούτου, αναπτύχθηκαν νέες τεχνικές και δεσμεύθηκαν καινοτόμα

¹² υποκατηγορία των τεχνολογιών μηχανικής μάθησης

εργαλεία για την αποτελεσματική ανακάλυψη χρήσιμων πληροφοριών, ενώ έγιναν πιο απαιτητικά στις ανάγκες των χρηστών.

Η διάγνωση πρώιμου διηθητικού καρκίνου του μαστού, βασίζεται στην ακτινολογία και την κλινική αξιολόγηση, που συμπληρώνονται από την επιβεβαίωση της βιοψίας. Οι πρόσφατες τεχνολογικές εξελίξεις στην ιατρική απεικόνιση υπόσχονται αντιμετώπιση κλινικών προκλήσεων στα στάδια ανίχνευσης καρκίνου, αξιολόγησης της ανταπόκρισης στη θεραπεία και παρακολούθησης της εξέλιξης της νόσου.

Καθεμία από τις τρεις περιπτώσεις που αναλύονται παρακάτω, αξιοποιεί και εστιάζει σε διαφορετικά εργαλεία μηχανικής μάθησης.

Τεχνολογίες μηχανικής μάθησης στον ιατρικό τομέα

Αρχικά, στην παρούσα περίπτωση, εξετάζονται οι τεχνολογίες μηχανικής μάθησης στον ιατρικό τομέα. Συγκεκριμένα, δίνεται έμφαση στα πεδία ωμικής, την παθολογία ή/και αντίστοιχες υποκατηγορίες. Για παράδειγμα, η ραδιενέργεια (radiomics), εξάγει χαρακτηριστικά από κλινικές εικόνες. Βασίζεται στην υπόθεση ότι τα εξαγόμενα δεδομένα προέρχονται από μηχανισμούς που λαμβάνουν χώρα σε γενετικό και μοριακό επίπεδο, παρέχοντας συλλογικά ολοκληρωμένο χαρακτηρισμό του όγκου. Αυτή η διαδικασία, ονομάζεται ραδιωμική υπογραφή του όγκου. Η ιστοπαθολογία χρησιμοποιείται για τη διάγνωση καρκίνου του μαστού.

Εκπαίδευση μοντέλων μηχανικής μάθησης

Στη συνέχεια, αναλύονται μοντέλα μηχανικής μάθησης. Τα μοντέλα αυτά έχουν προεκπαιδευτεί ή τηρούν υπό ανάπτυξη, σχεδιασμένα ειδικά για τέτοιες περιπτώσεις. Συνεισφέρουν στην πρόγνωση καρκίνου του μαστού με κριτήριο το τελικό αποτέλεσμα.

Ένα κοινό παράδειγμα μοντέλου μηχανικής μάθησης, αποτελεί το ευρέως χρησιμοποιούμενο μοντέλο μηχανικής μάθησης, το μηχανικό μοντέλο Gail. Έχει την ικανότητα να προβλέπει καρκίνο του μαστού. Για την διαδικασία πρόβλεψης καρκίνου του μαστού, χρησιμοποιεί απλές, στατιστικές αρχιτεκτονικές. Πρόσθετες εισροές στο μοντέλο, προέρχονται από κοστοβόρες ή/και επεμβατικές διαδικασίες.

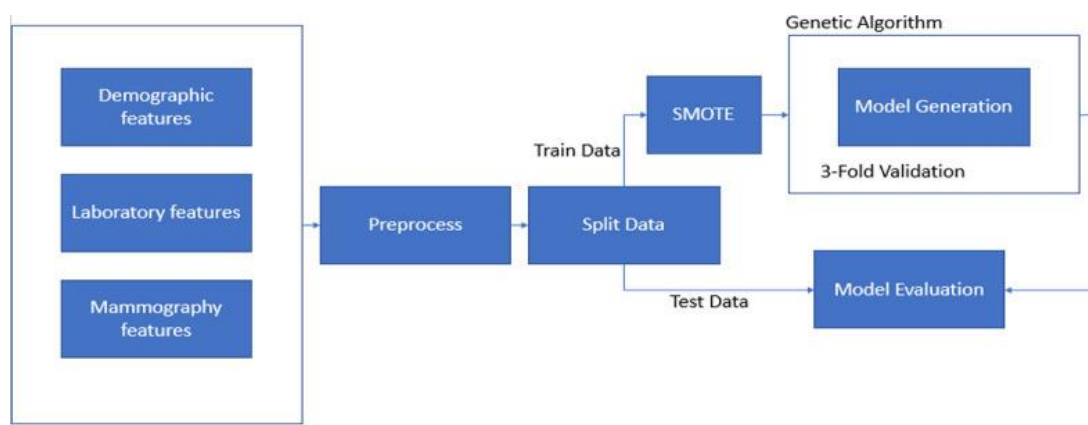
Γενικά, τα μοντέλα μηχανικής μάθησης διευκολύνουν τους γιατρούς να μειώσουν τον αριθμό των ψευδών αποφάσεων, θετικών ή αρνητικών, οι οποίες έχουν ληφθεί και έχουν ανιχνευθεί. Προσπαθούν να εκμεταλλευτούν πρότυπα και σχέσεις μεταξύ ενός μεγάλου αριθμού περιπτώσεων, καθώς και την πρόβλεψη της έκβασης μιας ασθένειας, χρησιμοποιώντας περιπτώσεις ιστορικών και έχουν αποθηκευθεί σε σύνολα δεδομένων.

Αξιοποίηση ταξινομητών μηχανικής μάθησης

Τέλος, εξετάζεται η χρήση ταξινομητών μηχανικής μάθησης, με εφαρμογές στην αναγνώριση προτύπων, την προεπεξεργασία δεδομένων και την λήψη αποφάσεων. Έχει αναπτυχθεί μια μεγάλη ποικιλία ταξινομητών μηχανικής μάθησης, με αποτέλεσμα να συμβάλλουν στην έγκαιρη διάγνωση του καρκίνου του μαστού. Βασίζονται σε δημοφιλείς αλγοριθμικές τεχνικές και τεχνικές προεπεξεργασίας δεδομένων. Κύρια παραδείγματα ταξινομητών ή/και αλγορίθμων, αποτελούν οι μηχανές διανυσμάτων υποστήριξης (support vector machines - SVM), ο αλγόριθμος που δομεί τη μέθοδο k-πλησιέστερου γείτονα (k-nearest neighbor - KNN) και τη μέθοδο γραμμικής ανάλυσης διάκρισης (linear discrete analysis - LDA). Ωστόσο, η διακριτική ισχύς τους είναι περιορισμένη, εξαιτίας του υπολογιστικού κόστους προσδιορισμού τελικών χαρακτηριστικών, για τον καθορισμό, τον χαρακτηρισμό και τη βελτιστοποίηση υποσυνόλου.

Γενικές αρχές της συνεισφοράς της αναλυτικής δεδομένων υγείας

Ο συνδυασμός κανόνων και διαφορετικών τεχνικών μηχανικής μάθησης αποτελεί ακριβές και αξιόπιστο σύστημα για την έγκαιρη διάγνωση καρκίνου του μαστού. Η επίτευξη υψηλής ακρίβειας κατά την πρόβλεψη, είναι σημαντική για την ενημέρωση της πτυχής της θεραπείας και του προτύπου επιβίωσης των ασθενών. Είναι σημαντικό να γνωρίζουμε την ουσία της κατανόησης των παραγόντων που αυξάνουν τον κίνδυνο εμφάνισης καρκίνου του μαστού και την αξιοποίηση εργαλείων μηχανικής μάθησης για λήψη αποφάσεων και ποσοτικές εκτιμήσεις, έτσι ώστε να μπορέσουμε να συμβουλευόμαστε επαρκώς και να στοχεύουμε στρατηγικές για τους ασθενείς.



Εικόνα 2: Τα στάδια που εκτελούνται κατά τη διαδικασία της μηχανικής μάθησης.

Πηγή: Rabiei R, Ayyoubzadeh SM, Sohrabei S, Esmaili M, Atashi A. Prediction of Breast Cancer using Machine Learning Approaches. J Biomed Phys Eng. 2022 Jun 1;12(3):297-308. doi: 10.31661/jbpe.v0i0.2109-1403. PMID: 35698545; PMCID: PMC9175124.

Κεφάλαιο 4

Περιγραφή των σημαντικότερων εργασιών από τη βιβλιογραφία

Ο ΚΑΡΚΙΝΟΣ του μαστού θεωρείται ένας από τους πιο συχνούς καρκίνους που εντοπίζεται στις γυναίκες. Η μηχανική μάθηση έχει τη δυνατότητα να προβλέψει τον καρκίνο του μαστού βάσει των χαρακτηριστικών που εντοπίζονται στα δεδομένα.

Τα τελευταία χρόνια, η αξιοποίηση τεχνολογιών μηχανικής μάθησης (machine learning) και βαθιάς μάθησης (deep learning) έχουν προσφέρει εξαιρετικά αποτελέσματα, με μειωμένο κόστος και βελτιωμένη απόδοση πάνω σε εφαρμογές στον κλάδο της υγείας.

Παρακάτω, αναλύονται ξεχωριστές περιπτώσεις παραδειγμάτων από εργασίες που έχουν συλλεχθεί από τη βιβλιογραφία.

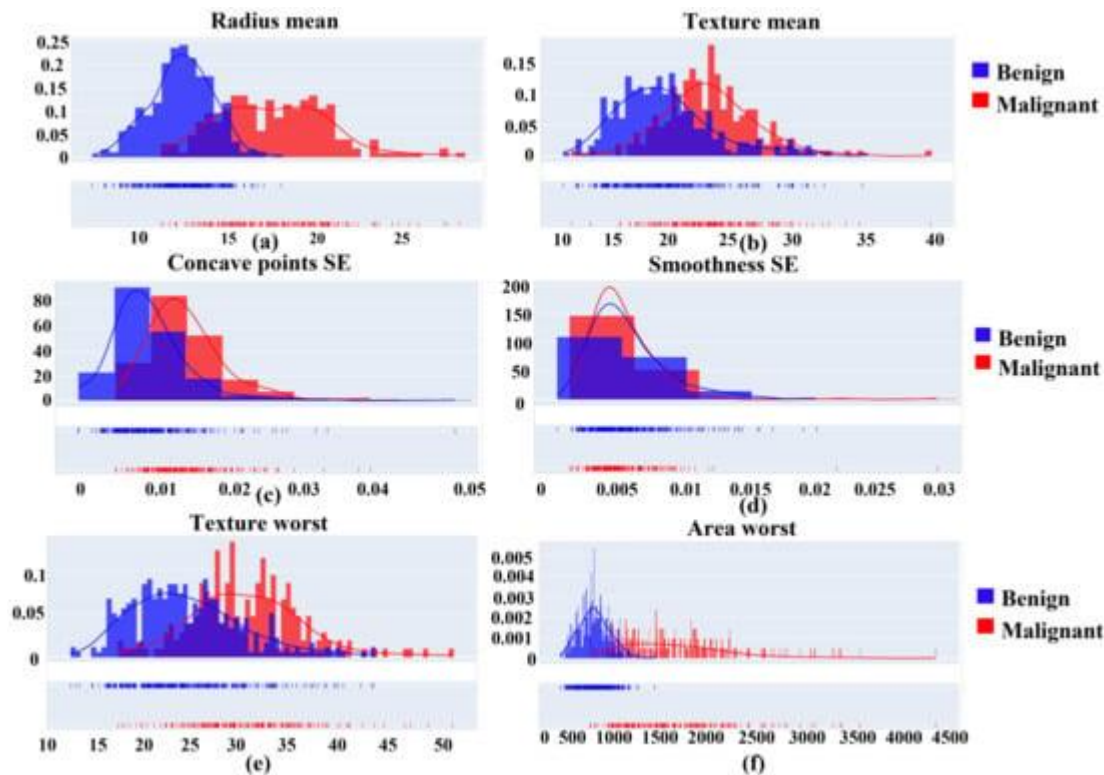
Ανάπτυξη μοντέλων πρόβλεψης για την απόδοση αποτελεσμάτων και χρόνου εκπαίδευσης σε τέσσερις διαφορετικούς ταξινομητές

Στο παράδειγμα των Rasool et al., αναπτύχθηκαν τέσσερα διαφορετικά μοντέλα πρόβλεψης επάνω σε δύο σύνολα δεδομένων· τα σύνολα δεδομένων Wisconsin Diagnostic Breast Cancer Dataset (WDBC) και Breast Cancer Coimbra Dataset (BCCD). Κάθε μοντέλο χρησιμοποιεί τέσσερις τεχνικές διερεύνησης δεδομένων¹³ και τέσσερις κοινούς ταξινομητές¹⁴, για την αξιολόγηση της αποτελεσματικότητας και του χρόνου εκπαίδευσης, για κάθε μία εκ των τεσσάρων τεχνικών.

Στην πρώτη περίπτωση, συγκεκριμένα, εξετάστηκε το σύνολο δεδομένων Wisconsin Diagnostic Breast Cancer (WDBC). Αποτελείται από 569 τιμές και 10 στήλες. Οι τιμές αντιστοιχούν στα δεδομένα που λήφθηκαν από τους ασθενείς συνολικά, ενώ οι στήλες στα χαρακτηριστικά που καταγράφηκαν. Χρησιμοποιήθηκαν διαφορετικοί ταξινομητές μηχανικής μάθησης για την προεπεξεργασία των δεδομένων του συνόλου.

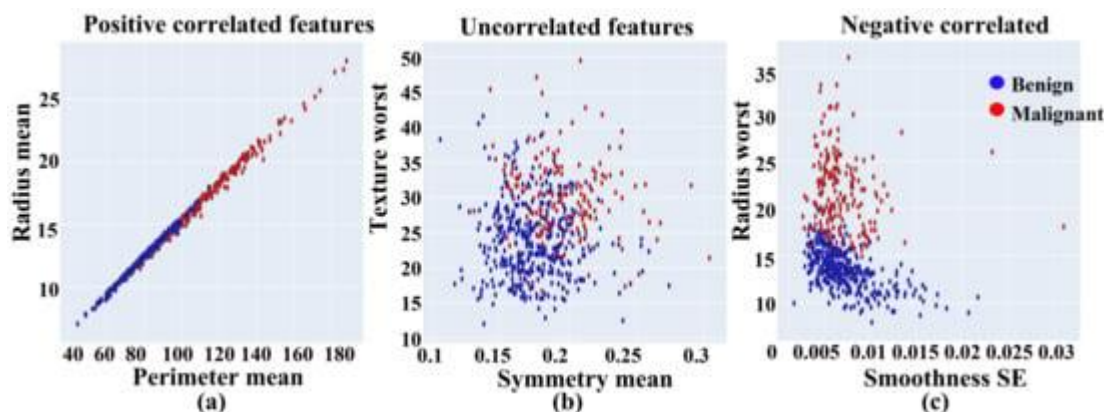
¹³ Κατανομή χαρακτηριστικών, συσχέτιση χαρακτηριστικών, αναδρομική εξάλειψη χαρακτηριστικών, βελτιστοποίηση υπερπαραμέτρων

¹⁴SVM, KNN, Linear Regression, ensemble classifier



Εικόνα 3: Κατανομή κατηγοριών διάγνωσης καρκίνου του μαστού με τη μέθοδο διαγράμματος. Απεικονίζονται ξεχωριστά διαγράμματα για έξι από τις συνολικά τριάντα στήλες του πρώτου σετ δεδομένων.

Πηγή: Rasool, A.; Bunternghit, C.; Tiejian, L.; Islam, M.R.; Qu, Q.; Jiang, Q. Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis. Int. J. Environ. Res. Public Health 2022, 19, 3211. <https://doi.org/10.3390/ijerph19063211>



Εικόνα 4: Διαγράμματα διασποράς για την απεικόνιση των κατηγοριών διάγνωσης καρκίνου του μαστού.

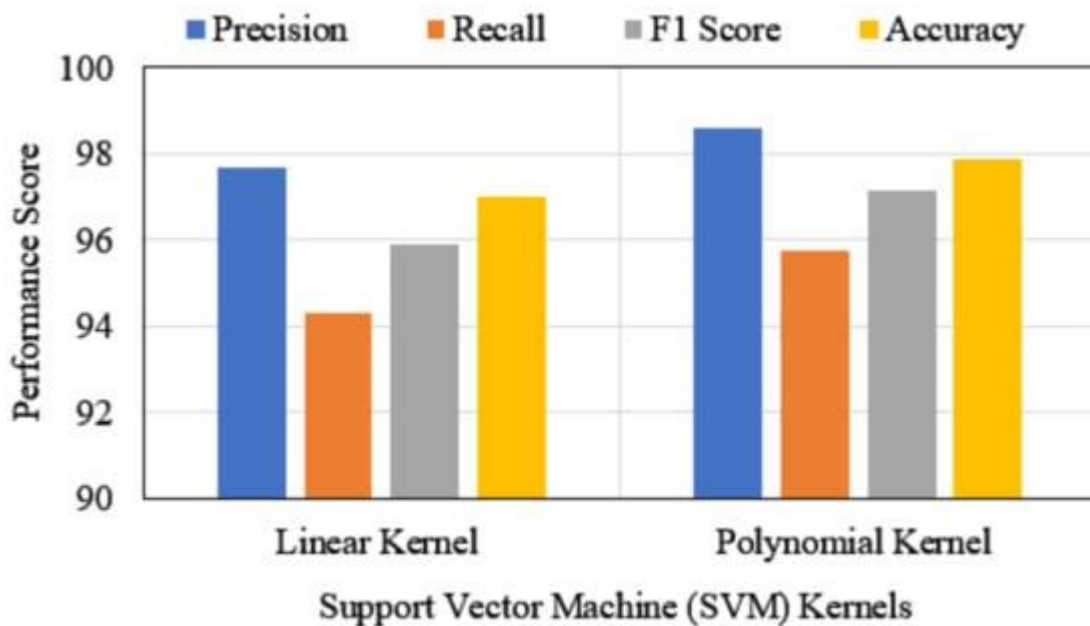
Πηγή: Rasool, A.; Bunternghit, C.; Tiejian, L.; Islam, M.R.; Qu, Q.; Jiang, Q. Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis. Int. J. Environ. Res. Public Health 2022, 19, 3211. <https://doi.org/10.3390/ijerph19063211>

Από τις διαγραμματικές απεικονίσεις σε ραβδογράμματα (εικόνα 3), διαγράμματα γραμμών (εικόνα 3) και διαγράμματα διασποράς (εικόνα 4), παρατηρούμε ότι κάθε μοντέλο πέτυχε διαφορετικό ποσοστό ακρίβειας, βάσει των ήδη καθορισμένων τεχνικών διερεύνησης δεδομένων.

Κατά τη χρήση πολυωνυμικού ταξινομητή μηχανών υποστήριξης διανυσμάτων (polynomial support vector machines classifier – polynomial SVM), επιτεύχθηκε ποσοστό ακρίβειας ίσο με 99,3%.

Κατά τη χρήση γραμμικού ταξινομητή μηχανών υποστήριξης διανυσμάτων (linear support vector machines classifier – linear SVM), επιτεύχθηκε ποσοστό ακρίβειας ίσο με 76.91%.

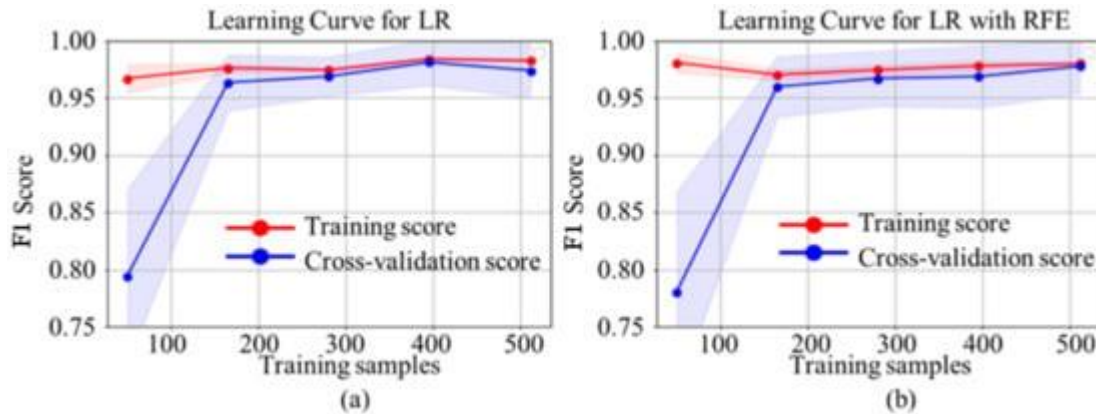
Στο παρακάτω σχήμα, απεικονίζεται ραβδόγραμμα και για τις δύο περιπτώσεις ταξινομητή μηχανών υποστήριξης διανυσμάτων, με βάση τις μετρικές που υπολογίστηκαν και αποδόθηκαν.



Εικόνα 5: Σύγκριση των αποτελεσμάτων των μετρικών που προκύπτουν από τον ταξινομητή SVM, μεταξύ γραμμικού και πολυωνυμικού πυρήνα.

Πηγή: Rasool, A.; Bunternghit, C.; Tiejian, L.; Islam, M.R.; Qu, Q.; Jiang, Q. Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis. Int. J. Environ. Res. Public Health 2022, 19, 3211. <https://doi.org/10.3390/ijerph19063211>

Κατά τη χρήση γραμμικής παλινδρόμησης (linear regression – LR) χωρίς την μέθοδο αναδρομικής εξάλειψης χαρακτηριστικών (recursive feature elimination – RFE), αναδείχθηκε ποσοστό ακρίβειας 98,06%. Ποσοστό χαμηλότερο από αυτό που αναδείχθηκε μέσω πολυωνυμικού ταξινομητή μηχανών υποστήριξης διανυσμάτων.



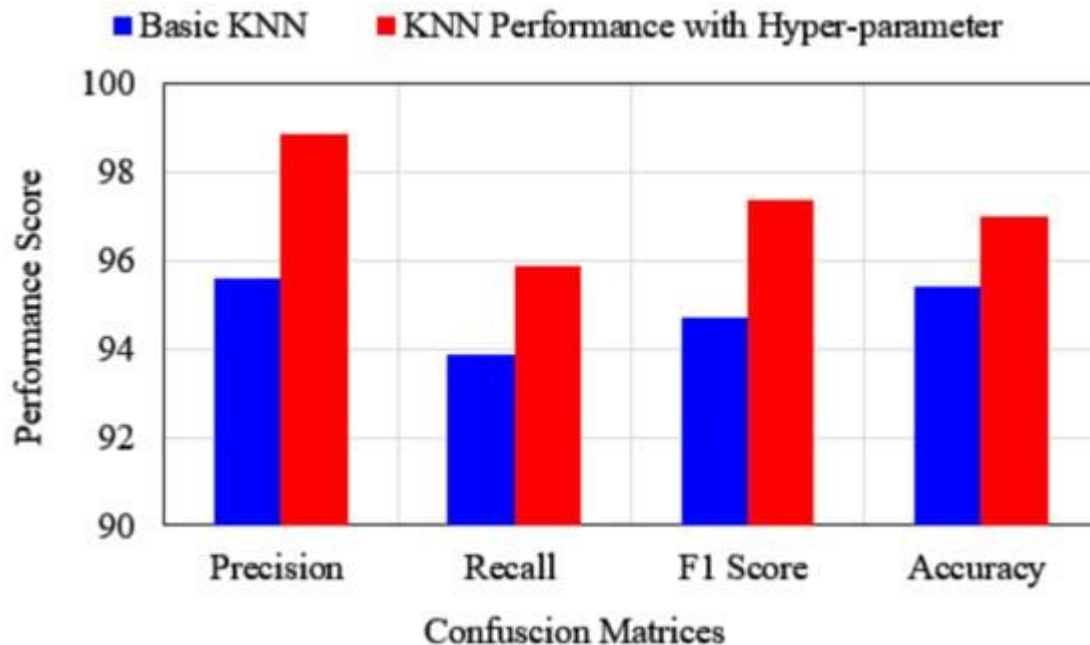
Εικόνα 6: Ανάλυση και σύγκριση ταξινομητών γραμμικής παλινδρόμησης και απόδοσης διασταυρούμενης επικύρωσης (cross-validation) με και χωρίς την μέθοδο αναδρομικής εξάλειψης χαρακτηριστικών (recursive feature elimination - RFE).

Πηγή: Rasool, A.; Bunternghit, C.; Tiejian, L.; Islam, M.R.; Qu, Q.; Jiang, Q. Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis. Int. J. Environ. Res. Public Health 2022, 19, 3211. <https://doi.org/10.3390/ijerph19063211>

Η χρήση της μεθόδου K-κοντινότερου γείτονα (K-nearest neighbor - KNN) πειραματίστηκε σε δύο κατηγορίες: βασική και υπερπαραμετρική.

Η βασική, λειτουργεί με τις προεπιλεγμένες παραμέτρους¹⁵. Ανέδειξε ποσοστό ακρίβειας ίσο με 95,43%.

Η υπερπαραμετρική¹⁶, αντίστοιχα, ανέδειξε ποσοστό ακρίβειας ίσο με 97,35%, επιτρέποντας συντονισμό παραμέτρων για την μέθοδο K-κοντινότερου γείτονα (KNN).



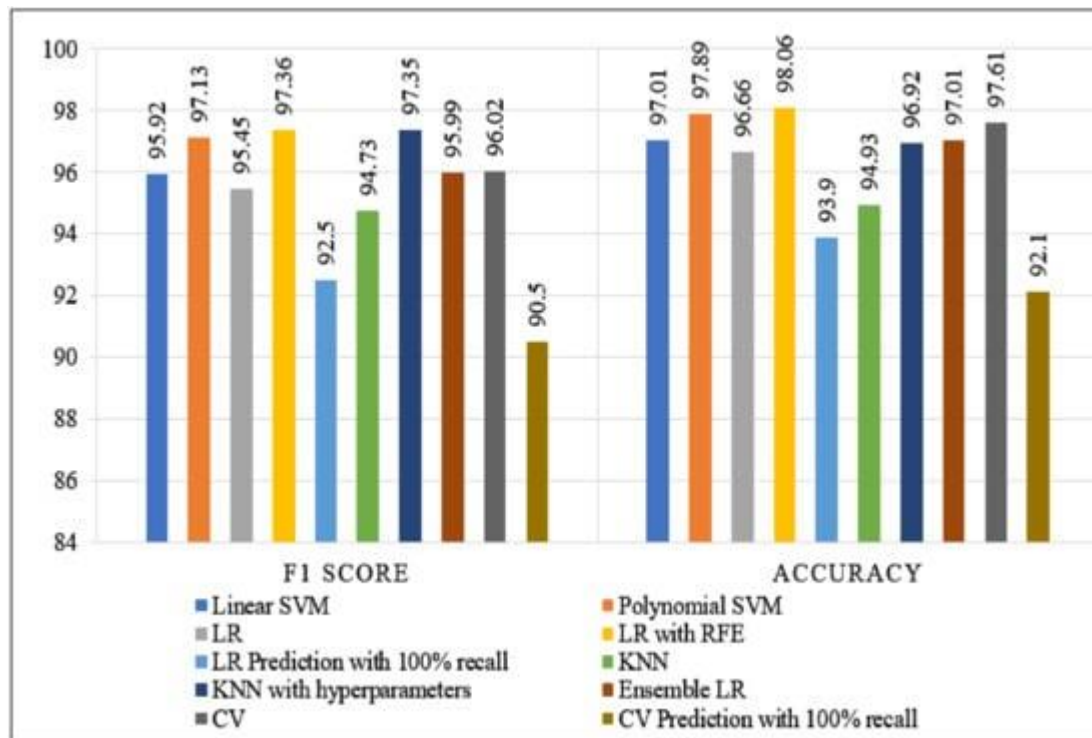
Εικόνα 7: Σύγκριση των αποτελεσμάτων των μετρήσεων που προκύπτουν από τη μέθοδο K-κοντινότερου γείτονα (KNN) σε βασική και υπερπαραμετρική κατηγορία.

Πηγή: Rasool, A.; Bunternghit, C.; Tiejian, L.; Islam, M.R.; Qu, Q.; Jiang, Q. Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis. Int. J. Environ. Res. Public Health 2022, 19, 3211. <https://doi.org/10.3390/ijerph19063211>

¹⁵ αυτόματα

¹⁶ χρησιμοποιείται μια ειδική παράμετρος για την βελτιστοποίησή της

Τέλος, ο ταξινομητής συνόλου (ensemble classifier – EC) ανέδειξε ποσοστό ακρίβειας ίσο με 97,61%. Ωστόσο, δεν έχουν γίνει γνωστά τα ποσοστά ακρίβειας που επιτεύχθηκαν για τη δεύτερη περίπτωση, για την οποία αντιστοιχεί το δεύτερο σετ δεδομένων.



Εικόνα 8: Απεικόνιση των μετρικών της ακρίβειας και του F1-score σε ραβδόγραμμα, κάνοντας χρήση διαφορετικών ταξινομητών εντός του πειράματος.

Πηγή: Rasool, A.; Bunternghit, C.; Tiejian, L.; Islam, M.R.; Qu, Q.; Jiang, Q. Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis. Int. J. Environ. Res. Public Health 2022, 19, 3211. <https://doi.org/10.3390/ijerph19063211>

Ανάπτυξη μοντέλων για την πρόβλεψη και απόδοση αποτελεσμάτων σε τέσσερις διαφορετικούς ταξινομητές

Στο παράδειγμα των Islam et al., χρησιμοποιήθηκαν πέντε διαφορετικοί ταξινομητές μηχανικής μάθησης¹⁷ επάνω σε ένα σύνολο δεδομένων, το σύνολο δεδομένων Wisconsin Diagnostic Breast Cancer (WDBC).

Για την πρόβλεψη και απόδοση των ήδη χρησιμοποιημένων τεχνικών και μετρικών αναδείχθηκε διαφορετικό ποσοστό ακρίβειας για κάθε ταξινομητή από τους πέντε συνολικά που χρησιμοποιήθηκαν.

Παρακάτω, ακολουθούν πίνακες σύγχυσης για κάθε ταξινομητή που έχει χρησιμοποιηθεί από τους πέντε συνολικά, και τα αντίστοιχα αποτελέσματά του για κάθε μία από τις μετρήσεις που έχουν υπολογιστεί στο πείραμα.

	Benign	Malignant
Benign	TP = 44 (62.86%)	FP = 1 (1.43%)
Malignant	FN = 0 (0.00%)	TN = 23 (32.85%)

Πίνακας 3: Πίνακας σύγχυσης για τα αποτελέσματα μέσω μηχανών υποστήριξης διανυσμάτων (support vector machines – SVM).

Πηγή: Islam, M.M., Haque, M.R., Iqbal, H. et al. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. SN COMPUT. SCI. 1, 290 (2020). <https://doi.org/10.1007/s42979-020-00305-w>

	Benign	Malignant
Benign	TP = 5 (64.29%)	TP = 45 (64.29%)
Malignant	FN = 1 (1.43%)	TN = 23 (32.85%)

Πίνακας 4: Πίνακας σύγχυσης για τα αποτελέσματα μέσω της μεθόδου Κ-κοντινότερου γείτονα (K-nearest neighbor – KNN).

Πηγή: Islam, M.M., Haque, M.R., Iqbal, H. et al. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. SN COMPUT. SCI. 1, 290 (2020). <https://doi.org/10.1007/s42979-020-00305-w>

¹⁷ SVM, KNN, random forest, ANN, linear regression

	Benign	Malignant
Benign	TP = 4 (62.86%)	TP = 44 (62.86%)
Malignant	FN = 2 (2.86)	TN = 23 (32.85%)

Πίνακας 5: Πίνακας σύγκρισης για τα αποτελέσματα μέσω του ταξινομητή τυχαίων δασών (Random Forest).

Πηγή: Islam, M.M., Haque, M.R., Iqbal, H. et al. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. SN COMPUT. SCI. 1, 290 (2020). <https://doi.org/10.1007/s42979-020-00305-w>

	Benign	Malignant
Benign	TP = 45 (64.29%)	TP = 45 (64.29%)
Malignant	FN = 0 (0.00%)	TN = 24 (34.28%)

Πίνακας 6: Πίνακας σύγκρισης για τα αποτελέσματα μέσω του ταξινομητή τεχνητών νευρωνικών δικτύων (artificial neural networks – ANN).

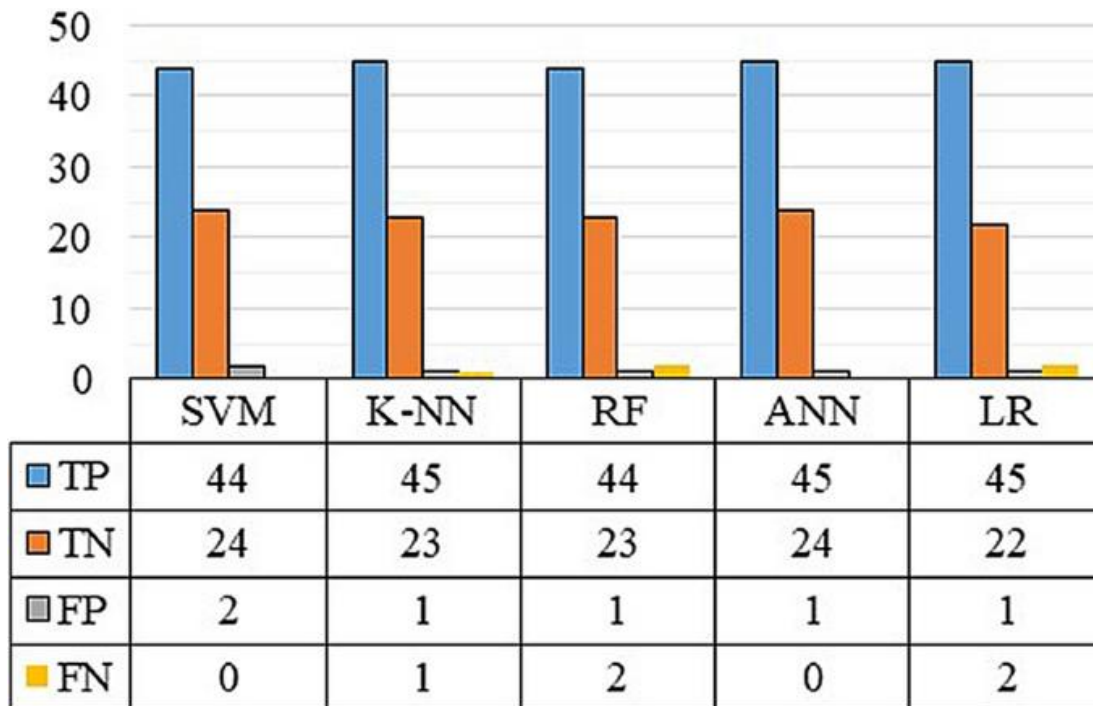
Πηγή: Islam, M.M., Haque, M.R., Iqbal, H. et al. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. SN COMPUT. SCI. 1, 290 (2020). <https://doi.org/10.1007/s42979-020-00305-w>

	Benign	Malignant
Benign	TP = 45 (64.29%)	TP = 45 (64.29%)
Malignant	FN = 2 (2.86%)	TN = 22 (31.42%)

Πίνακας 7: Πίνακας σύγκρισης για τα αποτελέσματα μέσω γραμμικής παλινδρόμησης (linear regression – LR).

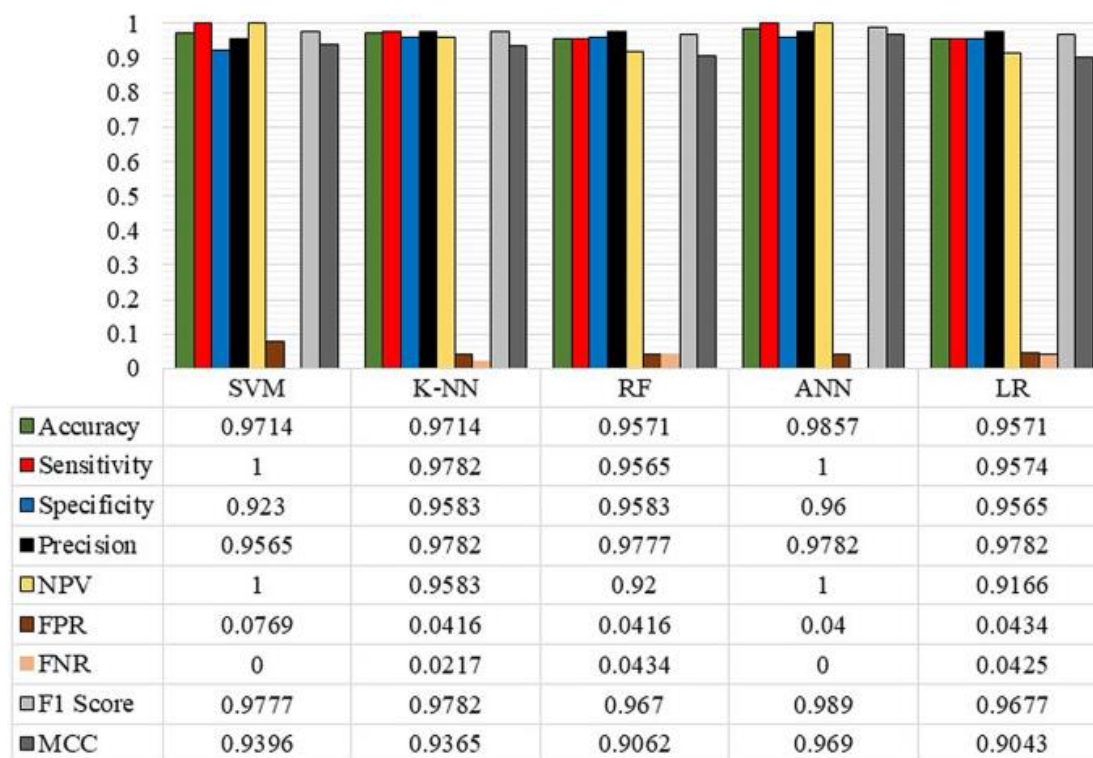
Πηγή: Islam, M.M., Haque, M.R., Iqbal, H. et al. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. SN COMPUT. SCI. 1, 290 (2020). <https://doi.org/10.1007/s42979-020-00305-w>

Επίσης, έχουν απεικονιστεί διαγραμματικές τεχνικές για καθένα από τα πέντε συνολικά αποτελέσματα που έχουν αναλυθεί κατά τη διάρκεια του πειράματος.



Εικόνα 9: Πίνακας σύγκρισης και διαγραμματική απεικόνιση για την συνολική πρόβλεψη καρκίνου του μαστού, χρησιμοποιώντας πέντε τεχνικές μηχανικής μάθησης.

Πηγή: Islam, M.M., Haque, M.R., Iqbal, H. et al. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. SN COMPUT. SCI. 1, 290 (2020). <https://doi.org/10.1007/s42979-020-00305-w>



Εικόνα 10: Συγκεντρωτικό ραβδόγραμμα όλων των μετρικών¹⁸ για την πρόβλεψη του καρκίνου του μαστού, χρησιμοποιώντας πέντε τεχνικές μηχανικής μάθησης

Πηγή: Islam, M.M., Haque, M.R., Iqbal, H. et al. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. SN COMPUT. SCI. 1, 290 (2020). <https://doi.org/10.1007/s42979-020-00305-w>

Τα τελικά αποτελέσματα, στις δύο παρακάτω περιπτώσεις που δόθηκαν, για δύο διαφορετικούς ταξινομητές αποκάλυψαν ότι μετά τη χρήση ταξινομητή τεχνητών νευρωνικών δικτύων (artificial neural networks – ANN) αναδείχθηκε ποσοστό ακρίβειας 98,57%.

Κατά τη χρήση διανυσματικών μηχανών υποστήριξης (support vector machines – SVM) αναδείχθηκε ποσοστό ακρίβειας ίσο με 97,14%.

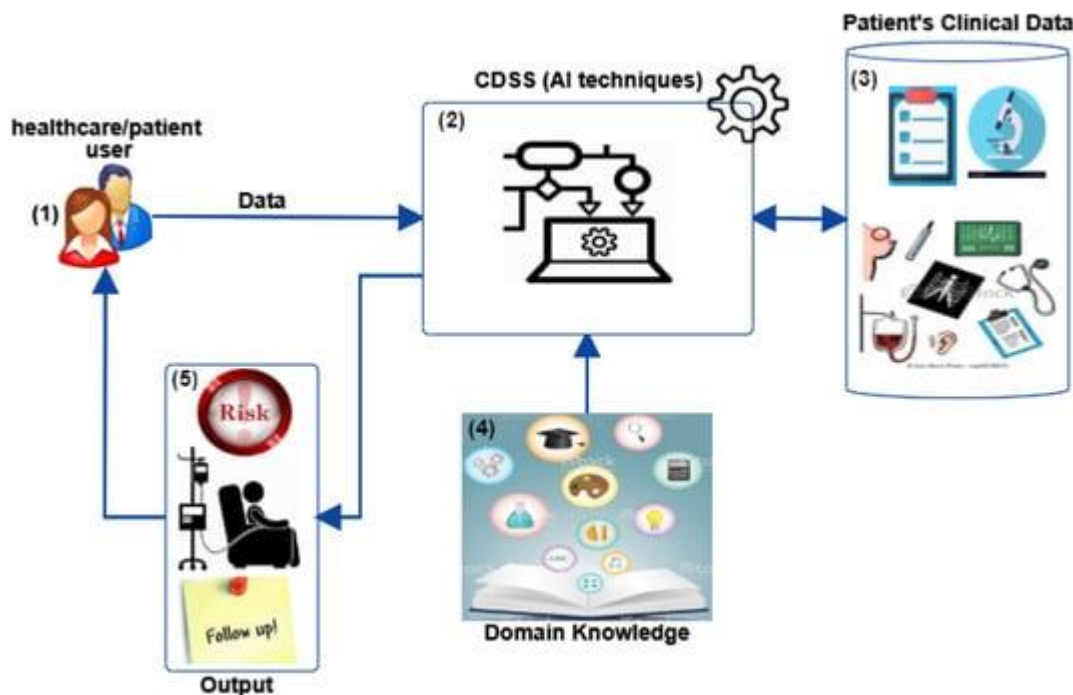
¹⁸ βασικών και σύνθετων

Τι ισχύει για τα συστήματα υποστήριξης κλινικών αποφάσεων;

Η εξατομικευμένη θεραπεία είναι κρίσιμη για τη βελτιστοποίηση αποτελεσμάτων για ασθενείς με καρκίνο του μαστού και τα συστήματα υποστήριξης κλινικών αποφάσεων μπορούν να διαδραματίσουν κρίσιμο ρόλο. Η ευρεία χρήση και ενσωμάτωση Συστημάτων Υποστήριξης Κλινικών Αποφάσεων (Clinical Decision Support Systems – CDSS) αποτελεί σημαντική πρόοδο στην ιατρική πρακτική.

Τα συστήματα αυτά, παρέχουν βοήθεια και υποστήριξη στο προσωπικό υγειονομικής περίθαλψης για τη λήψη κλινικών αποφάσεων. Ωστόσο, μπορεί να υπάρχουν διαφορές στον τρόπο με τον οποίο αναπτύσσονται, στην είσοδο χαρακτηριστικών ασθενών, στις πληροφορίες που περιλαμβάνουν, τις αποφάσεις που προτείνουν και τον τρόπο με τον οποίο χρησιμοποιούνται στην πράξη.

Είναι όμως συνηθισμένο ότι τα τελευταία χρόνια, τα περισσότερα Συστήματα Υποστήριξης Κλινικών Αποφάσεων, βασίζονται στη μηχανική μάθηση. Δεν είναι όμως βασικό στοιχείο ενός συστήματος υποστήριξης κλινικών αποφάσεων να αξιοποιείται μηχανική μάθηση ως κύρια μεθοδολογία ανάπτυξης. Ενίοτε, μπορεί να χρησιμοποιηθούν είτε παραπάνω από μία μεθοδολογίες ανάπτυξης ή συνδυασμός πολλαπλών μεθοδολογιών ταυτόχρονα.



Εικόνα 11: Περιγραφή της λειτουργικότητας ενός συστήματος υποστήριξης κλινικών αποφάσεων.

Πηγή: Mazo, C.; Kearns, C.; Mooney, C.; Gallagher, W.M. Clinical Decision Support Systems in Breast Cancer: A Systematic Review. *Cancers* 2020, 12, 369. <https://doi.org/10.3390/cancers12020369>.

Ορισμένα παραδείγματα μεθοδολογιών ή/και τεχνικών, αποτελούν η τεχνητή νοημοσύνη, η αναπαράσταση γνώσεων, η οπτικοποίηση πληροφοριών και δεδομένων, και η εξόρυξη κειμένων ως τα πιο γνωστά παραδείγματα. Έχει ως συμπέρασμα, το πλήθος τους στην αγορά να αυξάνεται.

Λίγα γεγονότα καθίστανται γνωστά για τα αποτελέσματα αυτών των συστημάτων όταν χρησιμοποιούνται συμπληρωματικά στην λήψη αποφάσεων από ανθρώπους.

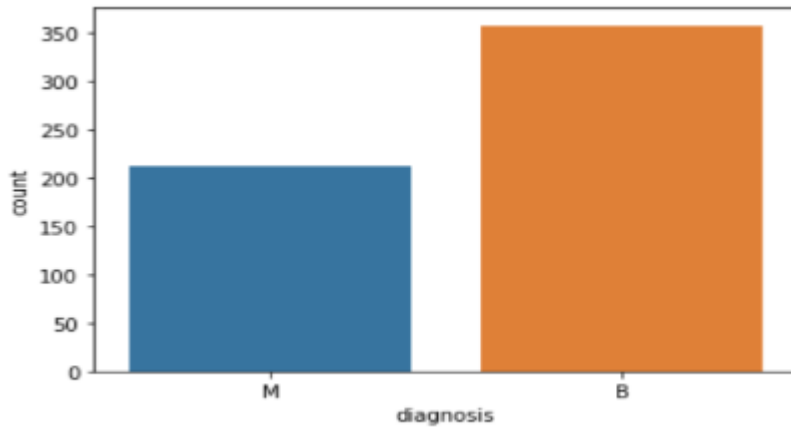
Ανάπτυξη συστήματος υποστήριξης αποφάσεων για την πρόγνωση καρκίνου του μαστού σε πρώιμο στάδιο

Στο παράδειγμα των Tiwari et al., εξετάζεται η υλοποίηση ενός συστήματος υποστήριξης κλινικών αποφάσεων. Το σύστημα, αφορά τη λήψη αποφάσεων κατά την πρόγνωση καρκίνου του μαστού σε πρώιμο στάδιο. Χρησιμοποιήθηκαν και δοκιμάστηκαν έξι διαφορετικοί ταξινομητές μηχανικής μάθησης¹⁹ επάνω σε ένα σύνολο δεδομένων, το σύνολο δεδομένων Wisconsin Diagnostic Breast Cancer (WDBC) που λήφθηκε από το απωθητήριο του Kaggle.

Για την προεπεξεργασία των δεδομένων, χρησιμοποιήθηκε το περιβάλλον ανάπτυξης Google Colab²⁰. Όλες οι τεχνικές, υλοποιήθηκαν σε γλώσσα προγραμματισμού Python. Κάθε αλγόριθμος που χρησιμοποιήθηκε κατά τη διάρκεια των πειραμάτων, έχει μετρηθεί και συγκριθεί σε σχέση με την ακρίβεια που ελήφθη από τον αντίστοιχο ταξινομητή.

¹⁹ Logistic Regression, SVM, KNN, multi-layer Perceptron, ANN, Random Forest, Naïve Bayes, decision tree

²⁰ Επιστημονικό Περιβάλλον Ανάπτυξης Python που βασίζεται στα πρότυπα του Jupyter Notebook.



Εικόνα 12: Κατανομή κατηγοριών διάγνωσης καρκίνου του μαστού κάνοντας χρήση της τεχνικής προεπεξεργασίας κωδικοποίησης ετικετών. (label encoding).

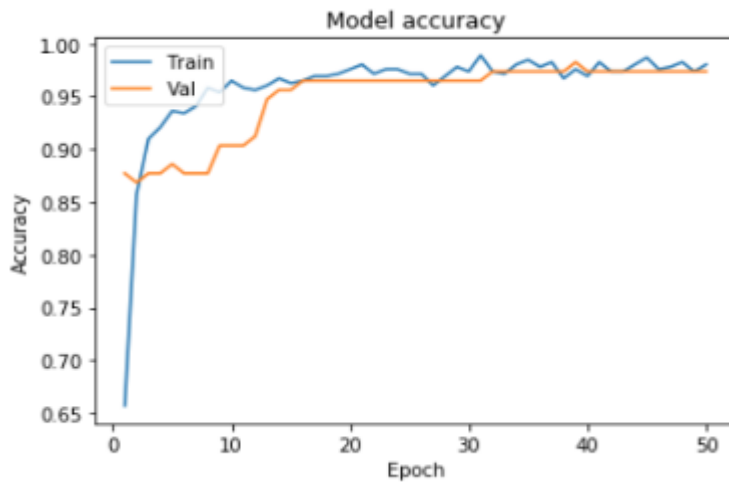
Πηγή: Tiwari, Monika and Bharuka, Rashi and Shah, Praditi and Lokare, Reena, Breast Cancer Prediction Using Deep Learning and Machine Learning Techniques (March 22, 2020), <http://dx.doi.org/10.2139/ssrn.3558786>

Τα τελικά αποτελέσματα των πειραμάτων, απέδειξαν ότι οι αλγόριθμοι διανυσματικών μηχανών υποστήριξης (support vector machines – SVM) και τυχαίων δασών (random forest), είναι οι καλύτεροι για προγνωστική ανάλυση των αποτελεσμάτων. Απέδωσαν ποσοστό ακρίβειας ίσο με 96,5%. Στην περίπτωση των αλγορίθμων διανυσματικών μηχανών υποστήριξης (SVM) και τυχαίων δασών (random forest) ταυτόχρονα, η μέγιστη ακρίβεια που επιτεύχθηκε και αποδόθηκε για τον κάθε αλγόριθμο ξεχωριστά, ανέδειξε ποσοστό ίσο με 96,5%.

Για να αυξηθεί το ποσοστό ακρίβειας, έχουν υλοποιηθεί, προστεθεί και εφαρμοστεί αλγόριθμοι βαθιάς μάθησης. Οι αλγόριθμοι αυτοί δίνουν έμφαση στα νευρωνικά δίκτυα. Συγκεκριμένα, χρησιμοποιήθηκαν οι αλγόριθμοι συνελκτικών νευρωνικών δικτύων (convolutional neural networks – CNN) και τεχνητών νευρωνικών δικτύων (artificial neural networks – ANN). Στην περίπτωση των αλγορίθμων συνελκτικών νευρωνικών δικτύων (convolutional neural networks – CNN), η ακρίβεια που επιτεύχθηκε κατά την χρήση τους, ανέδειξε ποσοστό ίσο με 97,3%. Κατά την χρήση τεχνητών νευρωνικών δικτύων (artificial neural networks – ANN), αναδείχθηκε ποσοστό ακρίβειας ίσο με 99,3%.

Και για τις δύο περιπτώσεις, τα ποσοστά ακρίβειας που επιτεύχθηκαν, ήταν πιο αποτελεσματικά σε σχέση με το ποσοστό που αναδείχθηκε από τους παραπάνω αλγορίθμους μηχανικής μάθησης, τους αλγορίθμους διανυσματικών μηχανών υποστήριξης (support vector machines) και τυχαίων δασών (random forest). Η σχέση αριθμού εποχών-ακρίβειας μοντέλου, κατά την εφαρμογή αλγορίθμου τεχνητών νευρωνικών δικτύων είναι ανάλογη.

Η εικόνα 13 απεικονίζει τη γραφική παράσταση του αριθμού των εποχών σε σχέση με την ακρίβεια του μοντέλου για τον αλγόριθμο τεχνητών νευρωνικών δικτύων.

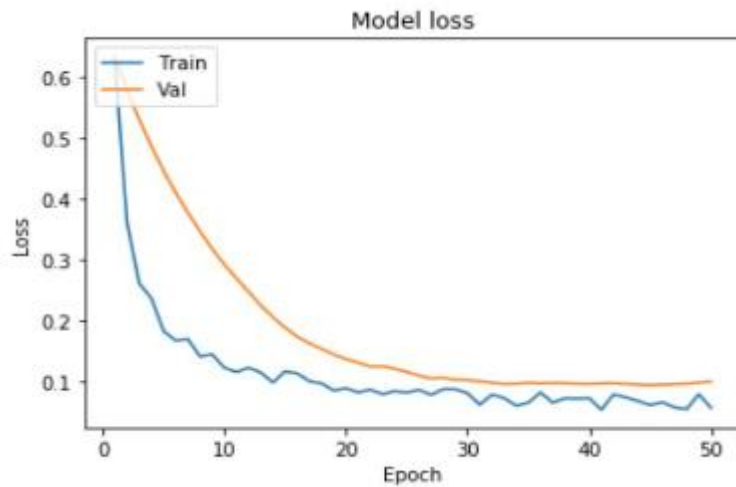


Εικόνα 13: Απεικόνιση της σχέσης ακρίβειας μοντέλου-αριθμού εποχών σε μορφή γραφικής παράστασης, για τον αλγόριθμο τεχνητών νευρωνικών δικτύων (artificial neural networks – ANN).

Πηγή: Tiwari, Monika and Bharuka, Rashi and Shah, Praditi and Lokare, Reena, Breast Cancer Prediction Using Deep Learning and Machine Learning Techniques (March 22, 2020), <http://dx.doi.org/10.2139/ssrn.3558786>

Από το σχήμα, παρατηρούμε ότι, καθώς αυξάνεται ο αριθμός εποχών του δικτύου, αυξάνεται παράλληλα και η ακρίβεια του μοντέλου. Αντίθετα, η σχέση αριθμού εποχών-απώλειας μοντέλου είναι αντιστρόφως ανάλογη.

Η εικόνα 14, απεικονίζει τη γραφική παράσταση του αριθμού των εποχών σε σχέση με την απώλεια του μοντέλου για τον αλγόριθμο τεχνητών νευρωνικών δικτύων.



Εικόνα 14: Απεικόνιση της σχέσης απώλειας-αριθμού εποχών σε μορφή γραφικής παράστασης, για τον αλγόριθμο τεχνητών νευρωνικών δικτύων (artificial neural networks – ANN).

Πηγή: Tiwari, Monika and Bharuka, Rashi and Shah, Praditi and Lokare, Reena, Breast Cancer Prediction Using Deep Learning and Machine Learning Techniques (March 22, 2020), <http://dx.doi.org/10.2139/ssrn.3558786>

Από το σχήμα, παρατηρούμε ότι, καθώς αυξάνεται ο αριθμός εποχών του δικτύου, μειώνεται η απώλεια του μοντέλου.

Έπειτα από τη χρήση αλγορίθμων και τεχνικών βαθιάς μάθησης, συμπεραίνουμε ότι οι τεχνικές βαθιάς μάθησης αποδείχθηκαν πιο αποτελεσματικές, σε σχέση με τεχνικές μηχανικής μάθησης. Αυτό οφείλεται στη χρήση συναρτήσεων ενεργοποίησης και πιθανοτικών μεθόδων, με αποτέλεσμα οι μετρικές ακρίβειας να αποδίδονται στον βέλτιστο βαθμό.

Στον ακόλουθο πίνακα, παρατίθενται αποτελέσματα που αναδείχθηκαν για τις μετρικές της συνολικής ακρίβειας, της ακρίβειας και της ευαισθησίας, συνδυάζοντας τεχνικές βαθιάς και μηχανικής μάθησης.

Algorithm	Accuracy	Precision	Sensitivity
KNN	0.95	0.95	0.99
SVM	0.96	0.98	0.97
Decision tree	0.95	0.99	0.93
Naïve Bayes	0.92	0.93	0.94
Logistic Regression	0.94	0.96	0.96
Random Forest	0.96	0.98	0.97
CNN	0.97	0.97	0.98
ANN	0.99	0.99	0.99

Εικόνα 15: Συγκεντρωτικός-συγκριτικός πίνακας για την σύγκριση μεταξύ αλγορίθμων μηχανικής και βαθιάς μάθησης για την πρόβλεψη του καρκίνου του μαστού.

Πηγή: Tiwari, Monika and Bharuka, Rashi and Shah, Praditi and Lokare, Reena, Breast Cancer Prediction Using Deep Learning and Machine Learning Techniques (March 22, 2020), <http://dx.doi.org/10.2139/ssrn.3558786>

Τι ισχύει για λοιπές τεχνολογίες/συστήματα;

Η αξιοποίηση τεχνολογιών και προσεγγίσεων μηχανικής μάθησης έχει κερδίσει τεράστιο έδαφος στο πεδίο της επιστήμης των υπολογιστών. Δίνεται έμφαση στην βιοπληροφορική και τις βιοϊατρικές εφαρμογές. Μεταξύ αυτών, αποτελεί παράδειγμα η ανάπτυξη λογισμικού και εφαρμογών για τη λήψη και υποστήριξη αποφάσεων²¹. Κύριος στόχος τους, αποτελεί η πρόγνωση καρκίνου του μαστού.

Η ραγδαία αύξηση τέτοιων εφαρμογών στην αγορά, παρατηρείται τόσο κατά το στάδιο της ανάπτυξης, όσο και κατά τη διάθεσή τους στο κοινό. Έχουν αναπτυχθεί εξελιγμένα προϊόντα, εφαρμογές ή αυτόνομο λογισμικό, τα οποία προορίζονται για προσωπική, ιδιωτική ή εμπορική χρήση.

Δεδομένου ότι τα περισσότερα διαθέσιμα προϊόντα διατίθενται επί πληρωμή για πλήρη χρήση, ολοένα και περισσότερο κοινό στρέφεται προς δωρεάν ή ανοιχτού κώδικα λογισμικό ή/και εφαρμογές. Προσφέρουν φιλική προς τους χρήστες διεπαφή, με εύχρηστα και εύκαμπτα εργαλεία. Τα εργαλεία που παρέχονται, προσπαθούν να επιδιώξουν το βέλτιστο δυνατό αποτέλεσμα για την αξιολόγηση κινδύνου εμφάνισης καρκίνου του μαστού ανά άτομο.

Πολλές μέθοδοι συνόλου έχουν εφαρμοστεί ευρέως στη διάγνωση του καρκίνου του μαστού. Αυτές οι μέθοδοι είναι ικανές να επιτύχουν και να αναδείξουν υψηλή ακρίβεια στο τελικό αποτέλεσμα.

Για παράδειγμα, συστήματα και εφαρμογές που βασίζονται σε τεχνολογίες υπολογιστικού νέφους (cloud computing) παίζουν ρόλο στην διάγνωση περίπλοκων και μη ασθενειών, ιδιαίτερα για κατοίκους που ζουν σε απομακρυσμένες περιοχές. Σκοπός τους είναι η παροχή υποστήριξης σε υπηρεσίες υγείας και διάγνωσης εξ' αποστάσεως.

²¹ Συστήματα υποστήριξης κλινικών αποφάσεων

Ανάπτυξη διαδικτυακής εφαρμογής για βελτίωση γνώσεων σχετικά με τον καρκίνο του μαστού

Στο παράδειγμα των Martin-Payo et al., εξετάζεται η ανάπτυξη διαδικτυακής εφαρμογής, σχεδιασμένη ειδικά για νεότερες γυναίκες. Σκοπός της εφαρμογής, η βελτίωση γνώσεων σχετικά με τους παράγοντες κινδύνου, τα πιθανά συμπτώματα, καθώς και προστατευτικές συμπεριφορές που υπόκεινται στη μείωση του κινδύνου εμφάνισης καρκίνου του μαστού. Παράλληλα, υπόσχεται αξιολόγηση χρηστικότητας και ποιότητας της εφαρμογής, η οποία προορίζεται για διαδικτυακή χρήση.

Για την ανάπτυξη της εφαρμογής, αναλύθηκαν τέσσερα βήματα (ανάλυση βιβλιογραφίας, σχεδιασμός διεπαφής, έλεγχος περιεχομένου από ειδικούς, αξιολόγηση). Αποτέλεσαν τα τελικά στάδια οριστικοποίησης της ιδέας της εφαρμογής. Ο προσδιορισμός του περιεχομένου που συμπεριελήφθη στην εφαρμογή, βασίστηκε σε δύο ενότητες. Οι δύο αυτές μεμονωμένες ενότητες, κατηγοριοποιούνται ως βιβλιογραφική ανασκόπηση και συστάσεις εμπειρογνομosύνης, για τις οποίες μπορούν να ληφθούν πληροφορίες σχετικά με παράγοντες κινδύνου καρκίνου του μαστού, διαδικασίες για την αύξηση της ικανότητας αντίληψης των γυναικών, συμπεριφορές υγιεινού τρόπου ζωής ή/και συμπτώματα έγκαιρης ανίχνευσης καρκίνου του μαστού.

Η εφαρμογή σχεδιάστηκε με κύριο σκοπό τη βελτίωση σωματικής και ψυχολογικής ικανότητας των γυναικών. Εκτός αυτού, τείνει να ενημερώσει τις γυναίκες σχετικά με την ευαισθησία εμφάνισης καρκίνου του μαστού, καθώς και να προσκομίσει οφέλη για την πρόληψή του, αναπτύσσοντας υγιεινές συμπεριφορές και στρατηγικές.



Εικόνα 16: Απεικόνιση της κεφαλίδας της διεπαφής της διαδικτυακής εφαρμογής “Breast-4Y”.

Πηγή: Ruben Martin-Payo, Claudia Leirós-Díaz, Jo Armes, Mei Rosemary Fu, Judit Cachero-Rodríguez, María del Mar Fernandez-Alvarez, The “Breast-4Y” web app for breast cancer prevention at young age: Development, evaluation, and validation, Internet Interventions, Volume 33, 2023, 100651, ISSN 2214-7829, <https://doi.org/10.1016/j.invent.2023.100651>.



Εικόνα 17: Επισκόπηση της διαδικτυακής εφαρμογής.

Πηγή: Ruben Martin-Payo, Claudia Leirós-Díaz, Jo Armes, Mei Rosemary Fu, Judit Cachero-Rodríguez, María del Mar Fernandez-Alvarez, The “Breast-4Y” web app for breast cancer prevention at young age: Development, evaluation, and validation, Internet Interventions, Volume 33, 2023, 100651, ISSN 2214-7829, <https://doi.org/10.1016/j.invent.2023.100651>.

Ανάπτυξη εφαρμογής για κινητά σχετικά με την ευαισθητοποίηση και γνώση του καρκίνου του μαστού

Στο παράδειγμα των Nasution et al., εξετάζεται η ανάπτυξη εφαρμογής ευαισθητοποίησης εξέτασης μαστού, σχεδιασμένη για κινητά.

Έπειτα από διεξαγωγή δειγματοληψίας, συγκεντρώθηκαν και ελήφθησαν πληροφορίες από δοκιμασμένες συνεντεύξεις. Αυτές οι πληροφορίες καταγράφηκαν, μεταγράφηκαν και οργανώθηκαν μεταξύ τους.

Για την οργάνωση των πληροφοριών, χρησιμοποιήθηκε το λογισμικό ανάλυσης δεδομένων NVIVO 11. Παράλληλα, συνέβαλαν στην τελική ανάπτυξη της εφαρμογής, ενώ στην εφαρμογή ενσωματώθηκαν θεωρίες υγείας για αποτελεσματική και σφαιρική κατανόηση του καρκίνου του μαστού.

Η βελτίωση γνώσης των παραγόντων κινδύνου, σημείων και συμπτωμάτων επηρεάζει θετικά την πρόθεση ενός ατόμου να αποκτήσει συμπεριφορές υγιεινού τρόπου ζωής. Οι συμπεριφορές αυτές αφορούν τα στάδια πρόληψης καρκίνου του μαστού, εγκαίρου εντοπισμού πιθανού κινδύνου εμφάνισής του και αναζήτησης συμβουλών υγείας.

Τι παρατηρούμε από τις εργασίες που έχουν συλλεχθεί;

Συνοψίζοντας, από όλες τις εργασίες που έχουν βρεθεί και αναλυθεί, έχει παρατηρηθεί ότι τα τελευταία χρόνια οι τεχνικές μηχανικής μάθησης έχουν συμβάλει σημαντικά στη διάγνωση του καρκίνου του μαστού, επιτυγχάνοντας αποτελέσματα υψηλής ακρίβειας κατά την πρόβλεψη.

Αλγοριθμικές τεχνικές με εφαρμογές στην μηχανική μάθηση, την τεχνητή νοημοσύνη, αυτόνομες τεχνικές εξόρυξης και αποθήκευσης δεδομένων έχουν αποδώσει αρκετά ικανοποιητικά αποτελέσματα στην προεπεξεργασία και αξιολόγηση δεδομένων, όχι μόνο στη φάση της πρόβλεψης. Αυτό οφείλεται στην επίτευξη υψηλής ακρίβειας, όπως και πολλών άλλων μετρικών. Ανεξάρτητα από τον επιλεγμένο ταξινομητή ή αντίστοιχη τεχνική μηχανικής μάθησης, έχει διασφαλιστεί αξιοπιστία στην απόδοση αποτελεσμάτων.

Είναι σύνηθες ότι όλο και περισσότερες εφαρμογές που σχετίζονται με την πρόβλεψη και τη διάγνωση του καρκίνου του μαστού έχουν αναπτυχθεί και έχουν δοκιμαστεί για προσωπικούς ή εμπορικούς σκοπούς. Συχνά τίθενται σε δημόσια ή παγκόσμια χρήση, ανεξάρτητα εάν διατίθενται επί πληρωμή, ανοιχτού κώδικα ή δωρεάν.

Έτσι, διαδίδονται ευρέως στο κοινό, ιδιαίτερα σε ασθενείς ή ιατρούς, χάρη στα αποτελεσματικά και εύκαμπτα εργαλεία που διευκολύνουν τη χρήση τέτοιων εφαρμογών.

Κεφάλαιο 5

Περιγραφή των σετ δεδομένων από την βιβλιογραφία

ΓΙΑ την ανάπτυξη ενός συστήματος υποστήριξης αποφάσεων που εξειδικεύεται τόσο στην πρόβλεψη καρκίνου του μαστού, όσο και τη διεξαγωγή πειραμάτων, αξιοποιήθηκαν δύο σετ δεδομένων, διαφορετικά μεταξύ τους.

Περιγραφή του σετ δεδομένων Wisconsin Diagnostic Breast Cancer Dataset

Το πρώτο σετ δεδομένων που αναλύεται κατά τη διάρκεια της εργασίας, είναι το σετ δεδομένων Wisconsin Diagnostic Breast Cancer Dataset.

Χρησιμοποιείται για να προβλέψει αν μια μάζα μαστού είναι καλοήθης ή κακοήθης. Διανεμήθηκε σε νοσοκομεία ανά την πόλη Μάντισον, στην πολιτεία Ουϊσκόνσιν των Ηνωμένων Πολιτειών.

Ελήφθη από το αποθητήριο του Kaggle. Διατίθεται στον ακόλουθο σύνδεσμο (<https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>).

Περιέχει 32 ιδιότητες. Ονομάζονται και χαρακτηριστικά του συνόλου δεδομένων. Τα 32 αυτά χαρακτηριστικά, απαρτίζουν τις στήλες που το συνθέτουν.

Συγκεκριμένα, προκύπτουν μία στήλη που αναγράφει τον αύξοντα αριθμό (ID) του ασθενούς από 569 περιπτώσεις, 30 χαρακτηριστικά εισαγωγής καθώς και μία στήλη που απεικονίζει την τελική διάγνωση του κάθε δείγματος: καλοήθης (benign – B) και κακοήθης (malignant – M).

Συμπερασματικά, 10 από τα συνολικά 32 χαρακτηριστικά, αποτελούν τα τελικά χαρακτηριστικά όγκου του μαστού. Το αποτέλεσμα στα δεδομένα, ελήφθη από δείγμα 569 ασθενών.

Περιγραφή του σετ δεδομένων Breast Cancer Coimbra Dataset

Το δεύτερο σετ δεδομένων που έχει χρησιμοποιηθεί στην τρέχουσα εργασία, είναι το σετ δεδομένων Breast Cancer Coimbra Dataset.

Χρησιμοποιείται ευρέως για την πρόβλεψη παρουσίας ή απουσίας καρκίνου του μαστού σε μια γυναίκα. Συγκεντρώθηκε από το γυναικολογικό τμήμα του Πανεπιστημιακού Νοσοκομείου της πόλης Κοϊμπρα στην Πορτογαλία.

Ελήφθη από την ιστοσελίδα του αποθητηρίου μηχανικής μάθησης του πανεπιστημίου UCI, και διατίθεται στον ακόλουθο σύνδεσμο (<https://archive.ics.uci.edu/dataset/451/breast+cancer+coimbra>).

Αποτελείται από 10 ιδιότητες. Για την ακρίβεια, το σετ δεδομένων συντίθεται από 9 χαρακτηριστικά, τα οποία αποτελούν τους προγνωστικούς παράγοντες του σετ δεδομένων, καθώς και μια στήλη που απεικονίζει την τελική διάγνωση μέσω μιας εξαρτημένης μεταβλητής.

Αυτή η μεταβλητή, λαμβάνει τις τιμές '1' και '2'. Ας υποθέσουμε, στην περίπτωση που απουσιάζει διάγνωση καρκίνου του μαστού, λαμβάνει την τιμή '1'. Αντίθετα, λαμβάνει την τιμή '2' σε περίπτωση που εντοπιστεί παρουσία καρκίνου του μαστού στα δεδομένα.

Συμπερασματικά, το αποτέλεσμα συλλέχθηκε από 116 περιπτώσεις. Από τις 116 περιπτώσεις που εντοπίστηκαν συνολικά, συλλέχθηκαν δεδομένα τόσο για 64 γυναίκες με καρκίνο του μαστού, όσο και για 52 υγιείς γυναίκες, χωρίς να παρουσιάζεται καρκίνος του μαστού στην περίπτωση τους.

Κεφάλαιο 6

Περιγραφή των βημάτων προεπεξεργασίας δεδομένων

ΓΙΑ την προεπεξεργασία των δεδομένων, ακολουθήθηκε μια σειρά βημάτων. Τα παρακάτω βήματα, αποτελούν την τελική μεθοδολογία που εφαρμόστηκε και έχει προταθεί εξίσου για την επίλυση του προβλήματος.

Περιγραφή των βημάτων για το πρώτο σετ δεδομένων

Για το πρώτο σετ δεδομένων, τα βήματα, ακολουθούν ως εξής:

1. Λήψη του σετ δεδομένων από την πηγή, την ιστοσελίδα του αποθητηρίου που διατίθεται.
2. Εισαγωγή των απαραίτητων βιβλιοθηκών ή/και πακέτων της γλώσσας προγραμματισμού Python, στο περιβάλλον ανάπτυξης Visual Studio Code.
3. Φόρτωση συνόλου δεδομένων στο πρόγραμμα, χρησιμοποιώντας την βιβλιοθήκη Pandas της Python.
4. Εκτέλεση διερευνητικής ανάλυσης δεδομένων για τον εντοπισμό απόντων χαρακτηριστικών ή τιμών.
5. Κατηγοριοποίηση δεδομένων σε καλοήθη και κακοήθη.
6. Διαχωρισμός του πλαισίου δεδομένων σε σύνολο εκπαίδευσης-δοκιμής (train-test set) με ποσοστά 90% εκπαίδευση, 10% δοκιμή αντίστοιχα.
7. Πραγματοποίηση στρωματοποιημένης διαδικασίας διασταυρούμενης επικύρωσης (stratified k-fold cross validation) κατά 10 φορές²².
8. Εκτέλεση διαδικασίας αναδρομικής εξάλειψης χαρακτηριστικών (έξοδος: 15 χαρακτηριστικά)
9. Υλοποίηση και αρχικοποίηση ταξινομητών²³
10. Απόδοση πρόβλεψης²⁴ με διαφορετικές μετρικές και/ή πίνακες για κάθε ταξινομητή
11. Αξιολόγηση και σύγκριση αποτελεσμάτων.

²² Ten-fold cross validation

²³ decision trees, random forest, adaboost, naive bayes, xgboost

²⁴ εστίαση τόσο σε αναδρομική όσο και σε μη επαναλαμβανόμενη εξάλειψη χαρακτηριστικών

Περιγραφή των βημάτων για το δεύτερο σετ δεδομένων

Παρόμοια διαδικασία, υφίσταται και για το δεύτερο σετ δεδομένων. Στο δεύτερο σετ δεδομένων, ωστόσο, δεν μπορεί να πραγματοποιηθεί αναδρομική εξάλειψη χαρακτηριστικών εξαιτίας του μικρού μεγέθους στηλών που διαθέτει. Αντίθετα, παρατηρήθηκε αναντιστοιχία στα δεδομένα, με αποτέλεσμα να υποστεί υποδειγματοληψία.

Αυτό σημαίνει, ότι για να εξαλειφθεί η αναντιστοιχία στα δεδομένα, το δείγμα μιας από τις δύο κατηγορίες μειώθηκε σε 52 τιμές. Εν τέλει, σημειώνει 104 περιπτώσεις δειγμάτων μετά την διαδικασία της υποδειγματοληψίας.

Τα βήματα για το δεύτερο σετ δεδομένων, ακολουθούν ως εξής:

1. Λήψη του σετ δεδομένων από την πηγή, την ιστοσελίδα του αποθηκετηρίου που διατίθεται.
2. Εισαγωγή των απαραίτητων βιβλιοθηκών ή/και πακέτων της γλώσσας προγραμματισμού Python, στο περιβάλλον ανάπτυξης Visual Studio Code.
3. Φόρτωση συνόλου δεδομένων στο πρόγραμμα, χρησιμοποιώντας την βιβλιοθήκη Pandas της Python.
4. Εκτέλεση διερευνητικής ανάλυσης δεδομένων για τον εντοπισμό απόντων χαρακτηριστικών ή τιμών.
5. Κατηγοριοποίηση δεδομένων σε θετικό και αρνητικό, για το αρχικό πλαίσιο δεδομένων.
6. Εφαρμογή υποδειγματοληψίας σε μια από τις δύο κατηγορίες του σετ δεδομένων.
7. Δημιουργία ενός νέου πλαισίου δεδομένων με υποδειγματοληψία.
8. Εκτέλεση διερευνητικής ανάλυσης δεδομένων στο νέο πλαίσιο δεδομένων, για τον εντοπισμό απόντων χαρακτηριστικών ή τιμών.
9. Κατηγοριοποίηση δεδομένων σε θετικό και αρνητικό εκ νέου, στο νέο πλαίσιο δεδομένων.
10. Διαχωρισμός του πλαισίου δεδομένων σε σύνολο εκπαίδευσης-δοκιμής (train-test set) με ποσοστά 90% εκπαίδευση, 10% δοκιμή αντίστοιχα.
11. Πραγματοποίηση στρωματοποιημένης διαδικασίας διασταυρούμενης επικύρωσης (stratified k-fold cross validation) κατά 10 φορές²⁵.
12. Υλοποίηση και αρχικοποίηση ταξινομητών²⁶
13. Απόδοση πρόβλεψης με διαφορετικές μετρικές και/ή πίνακες για κάθε ταξινομητή.
14. Αξιολόγηση και σύγκριση αποτελεσμάτων.

²⁵ Ten-fold cross validation

²⁶ decision trees, random forest, adaboost, naive bayes, xgboost

Κεφάλαιο 7

Περιγραφή της μεθόδου προεπεξεργασίας δεδομένων

ΓΙΑ την αξιοποίηση βασικών γνώσεων πάνω στους αλγορίθμους ταξινόμησης μηχανικής μάθησης και την πρόβλεψη του αποτελέσματος για καθένα από αυτούς, εφαρμόστηκε μια αναλυτική και παράλληλα αυστηρή μεθοδολογία.

Αυτή η μεθοδολογία, περιλαμβάνει διαδικασίες προεπεξεργασίας δεδομένων, εκπαίδευσης μοντέλων και αξιολόγησης. Εξετάζοντας κάθε βήμα προσεκτικά από όσα αναλύθηκαν παραπάνω, στοχεύσαμε στον εντοπισμό της βέλτιστης προσέγγισης για αυτή την εργασία.

Περιγραφή της μεθοδολογίας για το πρώτο σετ δεδομένων

Αρχικά, πραγματοποιήθηκε λήψη του σετ δεδομένων Wisconsin Diagnostic Breast Cancer Dataset από την ιστοσελίδα του απωθητηρίου που διατίθεται. Αποτέλεσε το πρώτο βήμα, πριν τις φάσεις της κωδικοποίησης και της προεπεξεργασίας δεδομένων.

Στη συνέχεια, εισήχθησαν βασικές βιβλιοθήκες στον κώδικα, όπως οι βιβλιοθήκες `pandas`, `numpy` και `scikit-learn` της γλώσσας προγραμματισμού Python, καθώς και ενσωματωμένα πακέτα²⁷. Τόσο οι βιβλιοθήκες, όσο και τα πακέτα που εισήχθησαν, παρέχουν κατάλληλα εργαλεία για χειρισμό δεδομένων, ανάλυση, μοντελοποίηση και αξιολόγηση.

Μετά το στάδιο της εισαγωγής απαραίτητων βιβλιοθηκών εντός του κώδικα, εισήγαμε το σετ δεδομένων, χρησιμοποιώντας τη βιβλιοθήκη `pandas`. Το μετατρέψαμε σε μεμονωμένο πλαίσιο δεδομένων (`dataframe`), για πιο αποτελεσματική ανάγνωση των δεδομένων του.

Διεξήχθη διερευνητική ανάλυση δεδομένων, για την λεπτομερή κατανόηση των χαρακτηριστικών του. Έτσι, εντοπίσαμε απύσες τιμές ή χαρακτηριστικά σε έλλειψη.

²⁷ Πακέτα τα οποία αξιοποιούν επιπρόσθετες λειτουργίες της Python.

Πραγματοποιήθηκε κατηγοριοποίηση δεδομένων σε καλοήθη (benign) και κακοήθη (malignant). Ακολούθησε διαχωρισμός του πλαισίου δεδομένων σε σύνολο εκπαίδευσης-δοκιμής (train-test set) με ποσοστά 90% για την διαδικασία της εκπαίδευσης, 10% για τη διαδικασία της δοκιμής αντίστοιχα.

Έπειτα από την εφαρμογή διαχωρισμού εκπαίδευσης-δοκιμής, προέκυψε διαδικασία στρωματοποιημένης διασταυρούμενης επικύρωσης, κατά 10 φορές. Διασφαλίστηκε τόσο αξιόπιστη αξιολόγηση όσο και βέλτιστη απόδοση του μοντέλου. Πριν το στάδιο της αρχικοποίησης και εφαρμογής αλγορίθμων ταξινόμησης στο μοντέλο, έλαβε χώρα διαδικασία αναδρομικής εξάλειψης χαρακτηριστικών. Χρησιμοποιήθηκε για την μείωση των χαρακτηριστικών και καταληκτική επιλογή των 15 καλύτερων χαρακτηριστικών. Στόχος της, μια απλοποιημένη και ευνοϊκή διαδικασία μοντελοποίησης.

Ποικίλοι αλγόριθμοι ταξινόμησης, συμπεριλαμβανομένου των δέντρων αποφάσεων (decision trees), των αλγορίθμων τυχαίων δασών (random forest), του αλγορίθμου ADABOOST²⁸, του πιθανοτικού ταξινομητή Naïve Bayes²⁹ και του αλγορίθμου XGBoost³⁰, εφαρμόστηκαν στα δεδομένα.

Αποδόθηκε πρόβλεψη με διαφορετικές μετρικές ή/και πίνακες για κάθε αλγόριθμο που χρησιμοποιήθηκε στο πείραμα, εστιάζοντας τόσο σε αναδρομική όσο και σε μη επαναλαμβανόμενη εξάλειψη χαρακτηριστικών.

Τέλος, τα αποτελέσματα κάθε αλγορίθμου αξιολογήθηκαν και συγκρίθηκαν με βάση την απόδοση των μετρικών τους.

Περιγραφή της μεθοδολογίας για το δεύτερο σετ δεδομένων

Πραγματοποιήθηκε λήψη του σετ δεδομένων Breast Cancer Coimbra Dataset από την ιστοσελίδα του απωθητηρίου που διατίθεται. Αποτέλεσε το αρχικό βήμα, πριν τις φάσεις της κωδικοποίησης, της προεπεξεργασίας και της υποδειγματοληψίας στο πλήθος των δεδομένων του.

Στη συνέχεια, εισήχθησαν βασικές βιβλιοθήκες στον κώδικα, όπως οι βιβλιοθήκες pandas, numpy και scikit-learn της γλώσσας προγραμματισμού Python, καθώς και ενσωματωμένα πακέτα³¹ για αξιοποίηση επιπρόσθετων λειτουργιών.

²⁸ Αλγόριθμος προσαρμοστικής ενίσχυσης (adaptive boosting algorithm)

²⁹ Διατίθενται τρεις παραλλαγές του ταξινομητή Naïve Bayes, πολωνυμικός, γκαουσιανός και παραλλαγή Bernoulli

³⁰ Αλγόριθμος ακραίας ενίσχυσης κλίσης (extreme gradient boosting)

³¹ Πακέτα τα οποία αξιοποιούν επιπρόσθετες λειτουργίες της Python.

Μετά το στάδιο της εισαγωγής απαραίτητων βιβλιοθηκών εντός του κώδικα, εισήγαμε το σετ δεδομένων, χρησιμοποιώντας τη βιβλιοθήκη pandas. Το αρχικό σετ δεδομένων, μετατράπηκε σε μεμονωμένο πλαίσιο δεδομένων (dataframe), για πιο αποτελεσματική ανάγνωση των δεδομένων του.

Διεξήχθη διερευνητική ανάλυση δεδομένων, για την κατανόηση των χαρακτηριστικών στο αρχικό πλαίσιο δεδομένων, πριν την φάση της υποδειγματοληψίας σε μία εκ των δύο κατηγοριών του.

Παράλληλα, τα δεδομένα κατηγοριοποιήθηκαν σε θετικό (positive) και αρνητικό (negative), όσον αφορά την παρουσία ή απουσία καρκίνου του μαστού. Σε σχέση με το πρώτο σετ δεδομένων, στο δεύτερο σετ δεδομένων παρατηρείται ότι το μέγεθος των χαρακτηριστικών και των περιπτώσεων είναι μικρότερο. Για αυτό το λόγο, δεν μπορεί να πραγματοποιηθεί ομαλή κατανομή δεδομένων ή αναδρομική εξάλειψη χαρακτηριστικών για το τρέχον μέγεθός του³².

Πρόέκυψε υποδειγματοληψία για μία από τις δύο συνολικά κατηγορίες του, με αποτέλεσμα το μέγεθος του πλαισίου δεδομένων να μειωθεί από 116 περιπτώσεις σε 104. Οι 104 αυτές περιπτώσεις δειγμάτων, θα αποτελέσουν το νέο πλαίσιο δεδομένων που θα δημιουργηθεί έπειτα από υποδειγματοληψία στα δεδομένα του. Εν τέλει, δημιουργήθηκε νέο πλαίσιο δεδομένων, με 52 δείγματα σε κάθε κατηγορία του αρχικού πλαισίου δεδομένων.

Για το νέο πλαίσιο δεδομένων που υπέστηκε υποδειγματοληψία κατά 12 δείγματα σε μία από τις κατηγορίες του, διεξήχθη διερευνητική ανάλυση δεδομένων. Στόχος της, η κατανόηση των χαρακτηριστικών, για τον εντοπισμό απόντων τιμών ή/και χαρακτηριστικών.

Πρόέκυψε κατηγοριοποίηση δεδομένων εκ νέου έπειτα από υποδειγματοληψία. Ακολούθησε διαχωρισμός του νέου πλαισίου δεδομένων σε σύνολο εκπαίδευσης-δοκιμής (train-test set) με ποσοστά 90% για την διαδικασία της εκπαίδευσης, 10% για τη διαδικασία της δοκιμής αντίστοιχα.

Έπειτα από την εφαρμογή διαχωρισμού εκπαίδευσης-δοκιμής, ακολούθησε διαδικασία στρωματοποιημένης διασταυρούμενης επικύρωσης, κατά 10 φορές. Διασφαλίστηκε τόσο αξιόπιστη αξιολόγηση, όσο και ομαλή απόδοση του μοντέλου για λιγότερα χαρακτηριστικά που το διατρέχουν.

Στο τρέχον πλαίσιο δεδομένων, δεν έλαβε χώρα αναδρομική εξάλειψη χαρακτηριστικών. Αυτό οφείλεται στον αριθμό των στηλών που απαρτίζουν τα χαρακτηριστικά του. Όμως, αρχικοποιήθηκαν και εφαρμόστηκαν αλγόριθμοι ταξινόμησης στο μοντέλο κανονικά, δίχως να επηρεαστούν τα χαρακτηριστικά ή τα δείγματά του.

Πολυάριθμοι αλγόριθμοι ταξινόμησης, συμπεριλαμβανομένου των δέντρων αποφάσεων (decision trees), των αλγορίθμων τυχαίων δασών (random forest), του αλγορίθμου ADABOOST³³, του πιθανοτικού

³² (116 περιπτώσεις – 10 χαρακτηριστικά).

³³ Αλγόριθμος προσαρμοστικής ενίσχυσης (adaptive boosting algorithm)

ταξινομητή Naïve Bayes³⁴ και του αλγορίθμου XGBoost³⁵, εφαρμόστηκαν στο νέο πλαίσιο δεδομένων, το οποίο είναι και καταληκτικό για τα τελικά στάδια προεπεξεργασίας και αξιολόγησης.

Αποδόθηκε πρόβλεψη με διαφορετικές μετρικές ή/και πίνακες για κάθε αλγόριθμο που χρησιμοποιήθηκε για τις ανάγκες του πειράματος. Τέλος, τα αποτελέσματα κάθε αλγορίθμου αξιολογήθηκαν και συγκρίθηκαν με βάση την απόδοση των μετρικών τους.

Μετά το στάδιο της προεπεξεργασίας των δεδομένων, ακολούθησαν κάποια κοινά βήματα μεταξύ των δύο μοντέλων. Τα κοινά αυτά βήματα, στόχο έχουν την σχηματική απεικόνιση δέντρων αποφάσεων σε κάθε κατηγορία, ανά δείγμα, και την εφαρμογή κανόνων αποφάσεων για καθένα από τα δύο πλαίσια δεδομένων. Ωστόσο, οδήγησαν στην επιλογή του καλύτερου αλγορίθμου βάσει των αποτελεσμάτων που ανέδειξαν και την ανάπτυξη εφαρμογής πρόβλεψης για κάθε μοντέλο, ξεχωριστά.

Κοινά βήματα και για τα δύο σετ δεδομένων

Και τα δύο μοντέλα, υλοποιήθηκαν εξ' ολοκλήρου σε γλώσσα προγραμματισμού Python, αξιοποιώντας το πρόσθετο Jupyter Notebook εντός του περιβάλλοντος ανάπτυξης Visual Studio Code. Παράλληλα, αναπτύχθηκαν και μετατράπηκαν συμβατές εκδόσεις των μοντέλων και για το διαδικτυακό περιβάλλον ανάπτυξης Google Colab³⁶. Εκεί, υπέστηκαν ελάχιστες τροποποιήσεις στον κώδικα.

Κοινά βήματα και για τα δύο σετ δεδομένων, ακολουθούν ως εξής:

1. Απεικόνιση δέντρων αποφάσεων για κάθε μοντέλο ξεχωριστά, εξετάζοντας τόσο αναδρομική όσο και μη επαναλαμβανόμενη εξάλειψη χαρακτηριστικών.
2. Εισαγωγή βιβλιοθηκών απαραίτητες για διαδικασίες σειριοποίησης³⁷ και διασωλήνωσης³⁸
3. Αποθήκευση προεκπαιδευμένων μοντέλων χρησιμοποιώντας τη βιβλιοθήκη pickle της Python.
4. Φόρτωση προεκπαιδευμένων μοντέλων χρησιμοποιώντας τη βιβλιοθήκη joblib της Python.

Για να εμβαθύνουμε τη διαδικασία λήψης αποφάσεων, προχωρήσαμε στη φάση της οπτικοποίησης κάθε δέντρου αποφάσεων, τόσο για σενάρια αναδρομικής όσο και για μη αναδρομικής εξάλειψης χαρακτηριστικών. Αυτό, μας επέτρεψε να εξετάσουμε τους κανόνες απόφασης και το ρόλο που παίζουν τα χαρακτηριστικά σε κάθε κόμβο. Οι παρακάτω εικόνες, απεικονίζουν τα δέντρα απόφασης τόσο για το πρώτο, όσο και για το δεύτερο σύνολο δεδομένων που αναλύθηκαν στην τρέχουσα εργασία.

³⁴ Διατίθενται τρεις παραλλαγές του ταξινομητή Naïve Bayes, πολυωνυμικός, γκαουσιανός και παραλλαγή Bernoulli

³⁵ Αλγόριθμος ακραίας ενίσχυσης κλίσης (extreme gradient boosting)

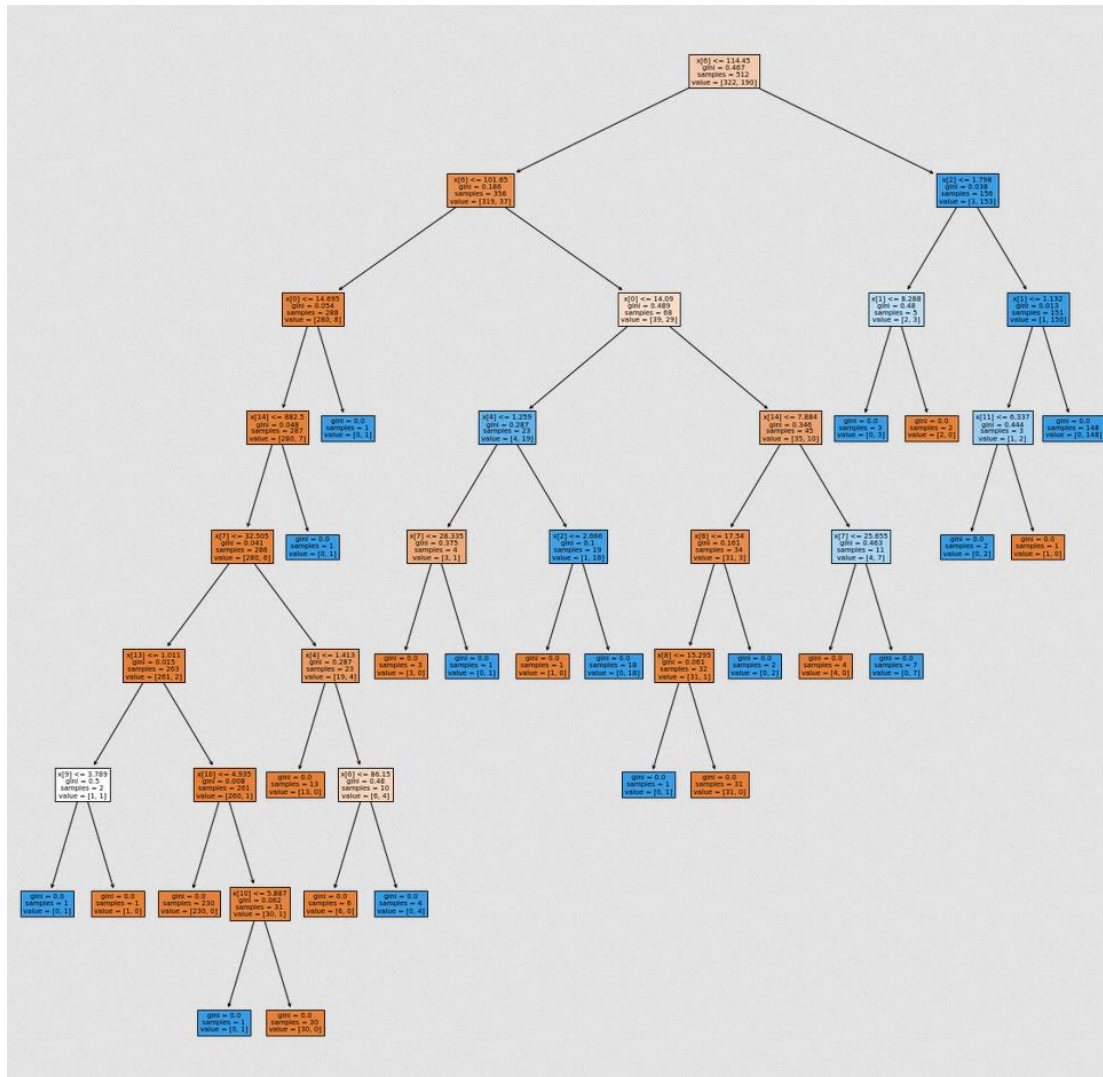
³⁶ Επιστημονικό Περιβάλλον Ανάπτυξης Python που βασίζεται στα πρότυπα και τη φιλοσοφία του Jupyter Notebook.

³⁷ Βιβλιοθήκη joblib της Python

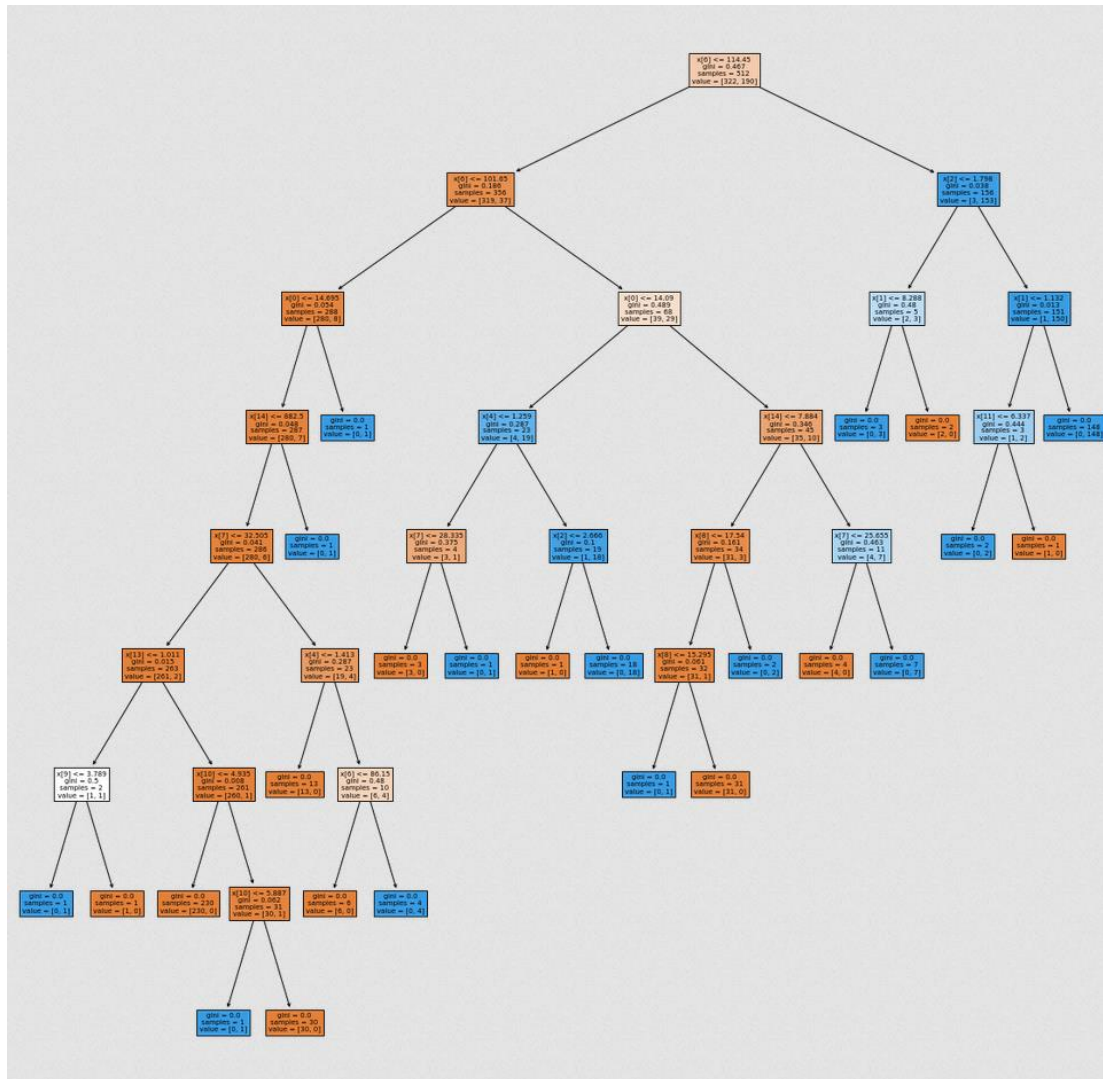
³⁸ Βιβλιοθήκες pickle και os της Python

Απεικόνιση δέντρων αποφάσεων για το πρώτο σύνολο δεδομένων

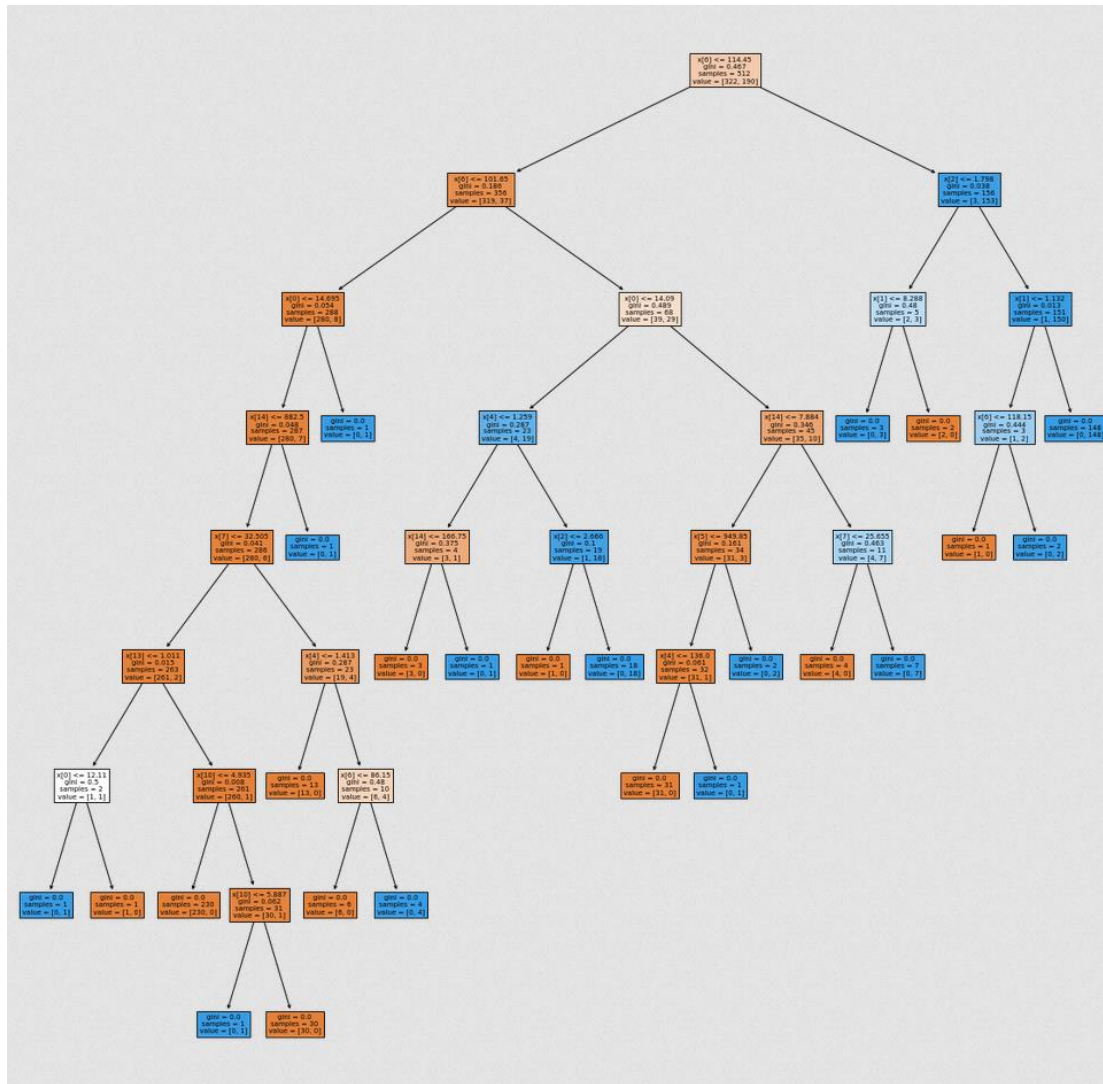
Για το πρώτο σύνολο δεδομένων, εξετάζονται περιπτώσεις για αναδρομική και μη-αναδρομική εξάλειψη χαρακτηριστικών ταυτόχρονα, καθώς και διαφορετικές υποπαραλλαγές για κάθε περίπτωση από τις τέσσερις που αναλύονται.



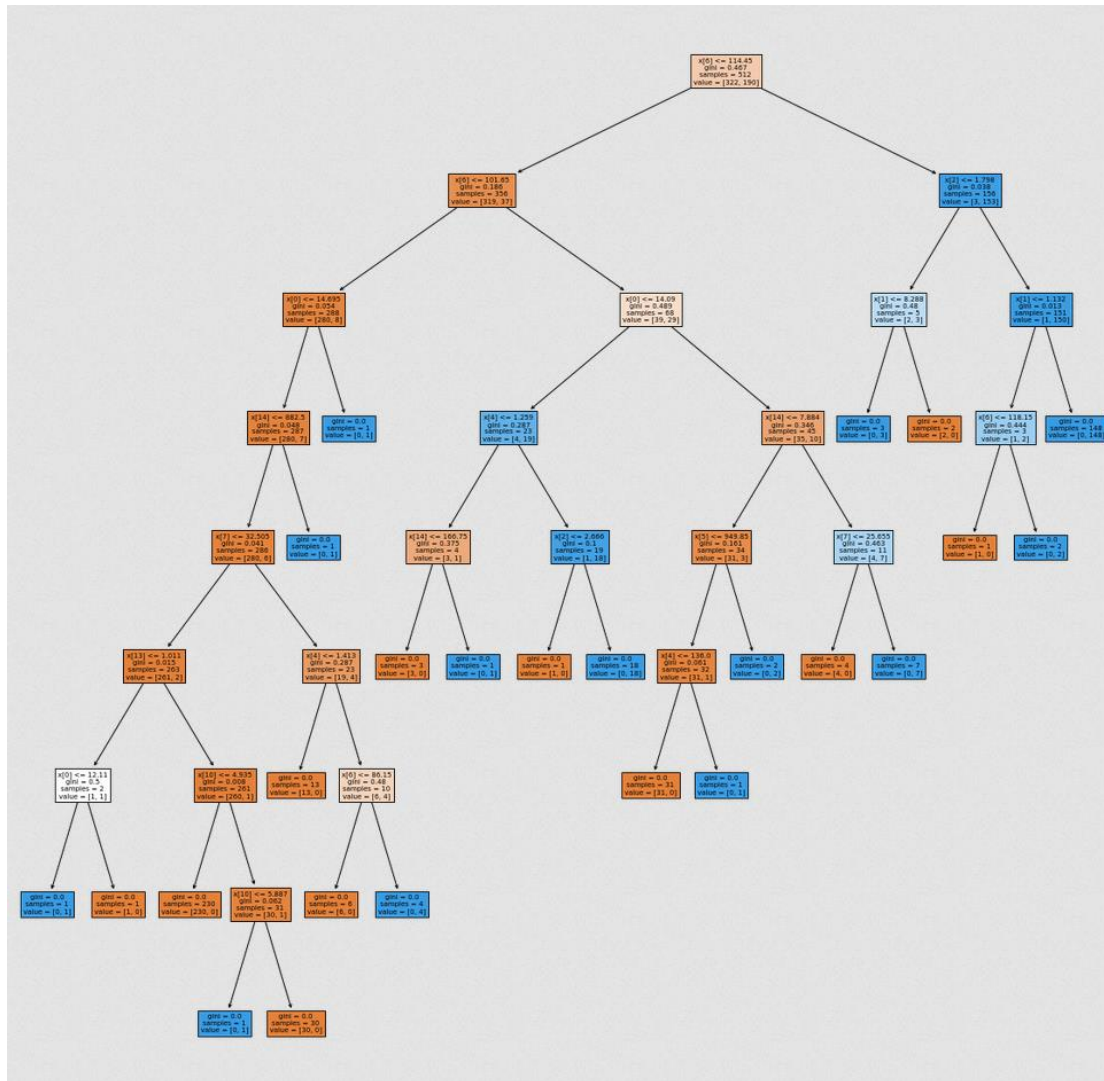
Εικόνα 18: Δέντρο απόφασης για αναδρομική εξάλειψη χαρακτηριστικών και τον ταξινομητή που προέκυψε κατά τη διάρκεια προσαρμογής χαρακτηριστικών.



Εικόνα 19: Δέντρο απόφασης για μη-αναδρομική εξάλειψη χαρακτηριστικών και τον αρχικό ταξινομητή που προέκυψε.



Εικόνα 20: Δέντρο απόφασης για αναδρομική εξάλειψη χαρακτηριστικών και τον ταξινομητή που προέκυψε στο στάδιο της εξάλειψης χαρακτηριστικών.

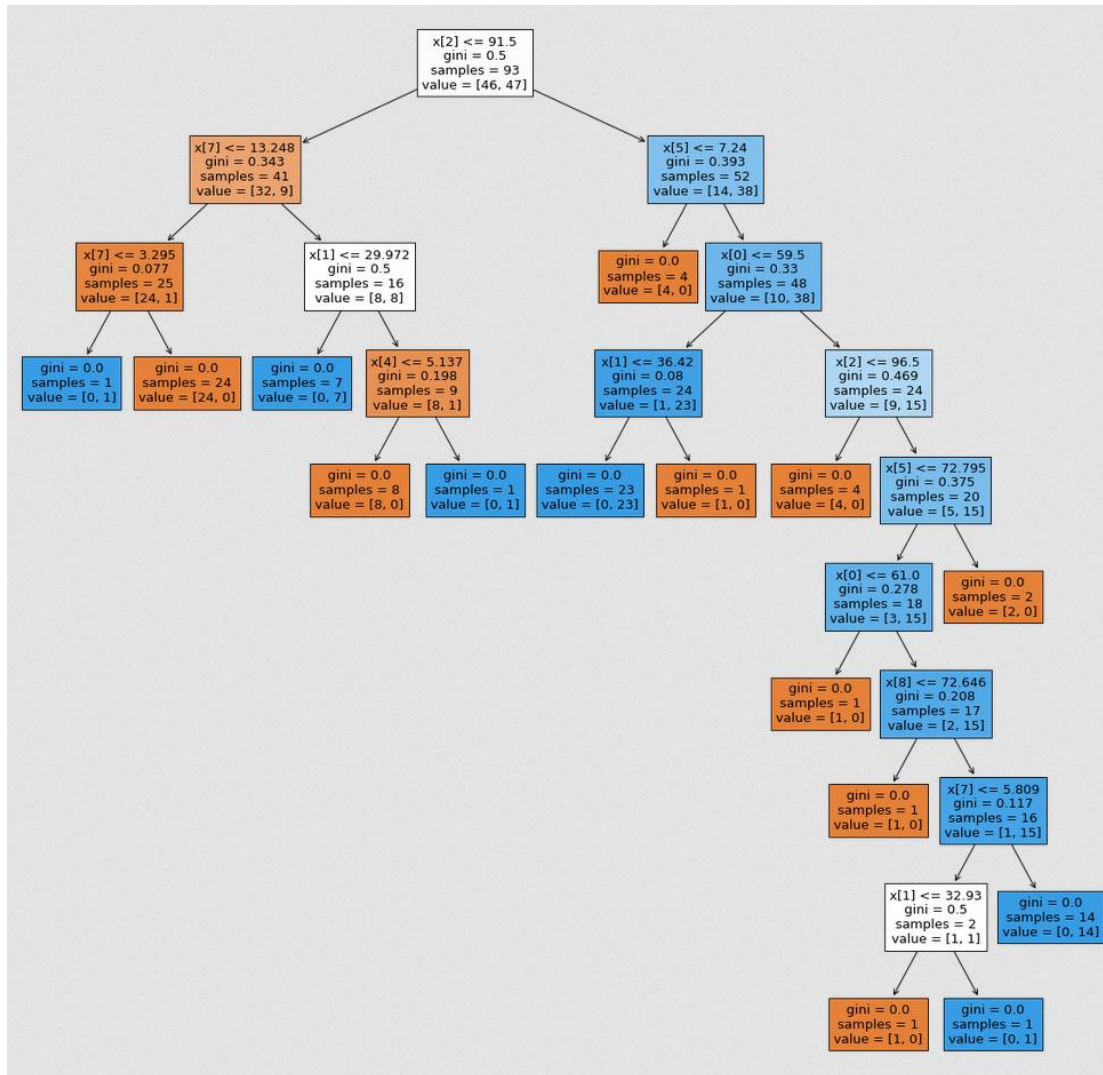


Εικόνα 21: Δέντρο απόφασης για αναδρομική εξάλειψη χαρακτηριστικών για νέο ταξινομητή, χρησιμοποιώντας ως παράμετρο τον ταξινομητή που προέκυψε στο στάδιο της εξάλειψης χαρακτηριστικών.

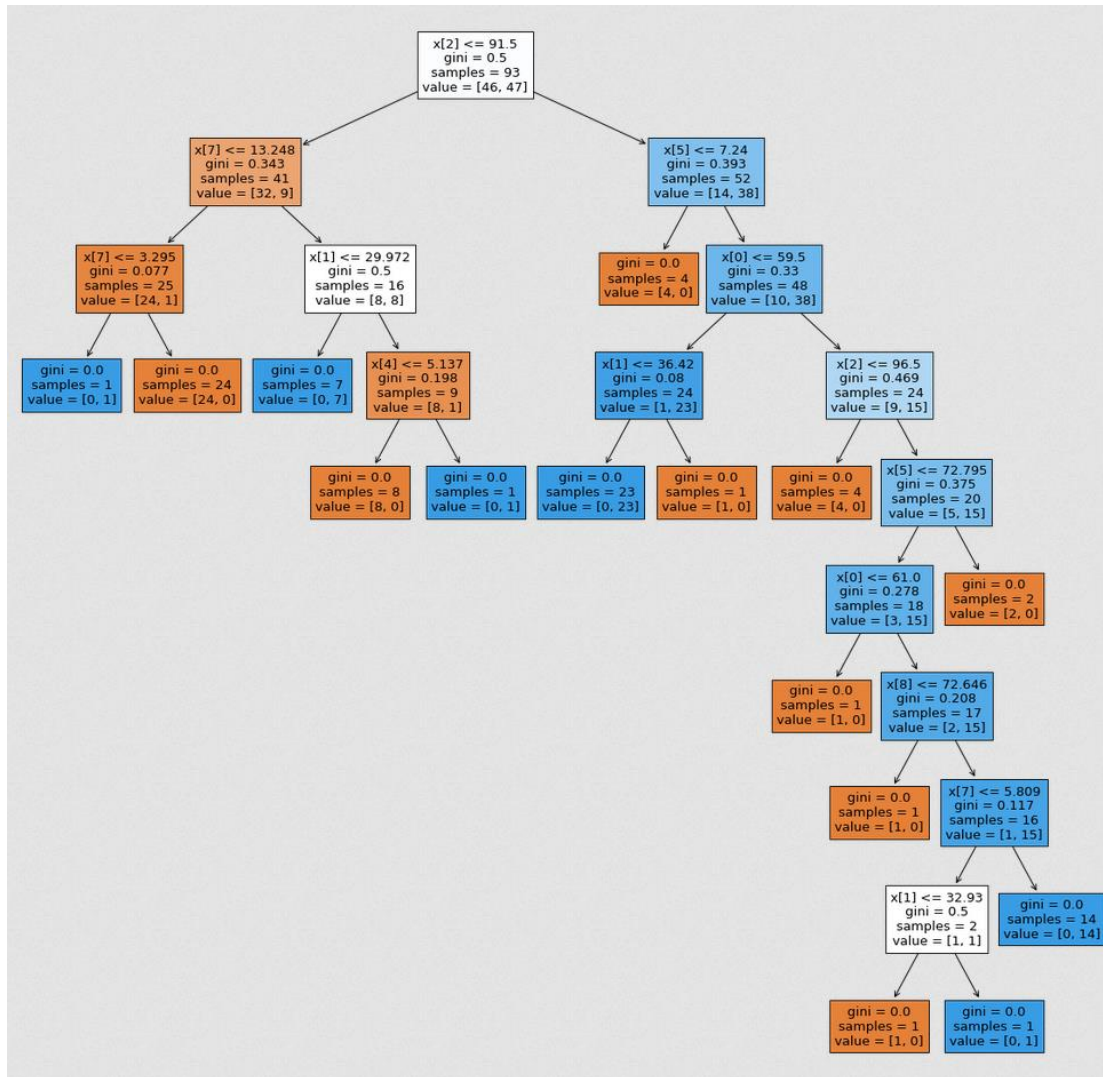
Απεικόνιση δέντρων αποφάσεων για το δεύτερο σύνολο δεδομένων

Αντίθετα, για το δεύτερο σετ δεδομένων, δεν υφίσταται αναδρομική ή μη-αναδρομική εξάλειψη χαρακτηριστικών. Δεν πραγματοποιήθηκε οπτικοποίηση δέντρου απόφασης για το αρχικό πλαίσιο δεδομένων, πριν την διαδικασία υποδειγματοληψίας.

Πραγματοποιήθηκε μόνο για το πλαίσιο δεδομένων στο οποίο υπέστηκε υποδειγματοληψία. Εστίασε αυστηρά και μόνο στους ταξινομητές που αρχικοποιήθηκαν πριν και κατά τη διάρκεια της διαδικασίας εκπαίδευσης.



Εικόνα 22: Δέντρο απόφασης για ταξινομητή που προέκυψε πριν την διαδικασία εκπαίδευσης.



Εικόνα 23: Δέντρο απόφασης για ταξινομητή που προέκυψε κατά τη διάρκεια της διαδικασίας εκπαίδευσης.

Στη συνέχεια, και πόσω μάλλον στις πρώτες γραμμές του κώδικα, εισήγαμε βιβλιοθήκες, απαραίτητες για διαδικασίες σειριοποίησης και διασωλήνωσης. Κυρίως, αξιοποιήθηκαν οι ενσωματωμένες βιβλιοθήκες pickle και joblib της Python.

Έτσι, δημιουργήσαμε και αποθηκεύσαμε προεκπαιδευμένα μοντέλα για μελλοντική χρήση. Υφίστανται ιδιαίτερα στην ανάπτυξη εφαρμογών μηχανικής μάθησης, με την ενσωματωμένη βιβλιοθήκη pickle.

Ενώ, με τη βιβλιοθήκη joblib, φορτώσαμε τα προεκπαιδευμένα μοντέλα, ειδικά όταν πρόκειται για μεγάλα σετ δεδομένων.

Ένα από τα πλεονεκτήματα που διαθέτει, είναι ότι δεν υφίσταται αυστηρά και μόνο για μικρά σετ δεδομένων σε μέγεθος ή πλήθος περιπτώσεων.

Κεφάλαιο 8

Ανάλυση των αποτελεσμάτων των δεδομένων

ΑΦΟΥ εκτελέσαμε τις φάσεις προεπεξεργασίας, ανάλυσης δεδομένων, αξιολόγησης και μοντελοποίησης, αμέσως επόμενο βήμα στην τρέχουσα εργασία, αποτελεί η διερεύνηση των αποτελεσμάτων που προέκυψαν από τους αλγορίθμους ταξινόμησης.

Προχωρώντας ταυτόχρονα σε ανάλυση μετρικών και βαθμολογιών για κάθε μοντέλο ξεχωριστά, τελικό στάδιο προκύπτει η σύγκριση και των δύο μοντέλων. Σε αυτό το σημείο, τείνουμε να συζητήσουμε την αποτελεσματικότητα των μετρικών για κάθε αλγόριθμο στην διαδικασία της πρόβλεψης καρκίνου του μαστού ενισχύοντας τις τρέχουσες γνώσεις μας.

Τα αποτελέσματα που ακολουθούν παρακάτω, υποδεικνύουν την αξιολόγηση των μετρικών που προέκυψαν από τους πέντε συνολικά αλγορίθμους ταξινόμησης. Και οι πέντε, εφαρμόστηκαν κατά τη διάρκεια του πειράματος.

Αποτελέσματα για το πρώτο σετ δεδομένων

Έχουν εξεταστεί δύο περιπτώσεις συνολικά. Αυτές οι δύο περιπτώσεις, εφαρμόζονται στην εφαρμογή αναδρομικής ή μη-επαναλαμβανόμενης εξάλειψης χαρακτηριστικών του τρέχοντος πλαισίου δεδομένων, και θα αναλυθούν ταυτόχρονα για κάθε αλγοριθμική τεχνική που υλοποιήθηκε.

Αποτελέσματα που προέκυψαν από τον ταξινομητή δέντρων αποφάσεων

Αυτή η εργασία έχει εξοπλίσει έναν ταξινομητή δέντρου απόφασης. Η αξιολόγηση των αποτελεσμάτων των μετρικών που προέκυψαν κατά την πρόβλεψη, δεν υφίσταται αποκλειστικά στην εξάλειψη χαρακτηριστικών ή όχι.

Για την περίπτωση αναδρομικής εξάλειψης χαρακτηριστικών, αναδείχθηκαν συνολική ακρίβεια (accuracy) ίση με 92,98% (περίπου 93%) καθώς και βαθμός διασταυρούμενης επικύρωσης (cross-validation score) ίσος με 91,98% (περίπου 92%).

Εξάιρεση, αποτελεί η μη-αναδρομική εξάλειψη χαρακτηριστικών, για την οποία ποικίλει η απόδοση μετρικών στον βαθμό διασταυρούμενης επικύρωσης (cross-validation score). Δεν είναι για κάθε μοιρασιά η ίδια τιμή. Ο μεγαλύτερος βαθμός ανέδειξε ποσοστό 94,11% και ο μικρότερος, ποσοστό ίσο με 84,31% από όσα έχουν διατεθεί.

Τα αποτελέσματα που προέκυψαν, αντιπροσωπεύουν σφαιρική³⁹ απόδοση μετρικών. Παρατηρούμε ότι, για κάθε κατηγορία, έχουν αποδοθεί διαφορετικά αποτελέσματα⁴⁰. Εδώ θα αναλυθούν δύο διαφορετικές περιπτώσεις για κάθε μία εκ των δύο κατηγοριών, καλοήγη (benign) και κακοήγη (malignant).

Η πρώτη περίπτωση, αναλύει την αξιολόγηση αποτελεσμάτων για την κατηγορία καλοήθους (benign) διάγνωσης. Δίνεται περισσότερο βάση σε μετρικές ακρίβειας (precision), ανάκλησης (recall), και βαθμολογίας F1-score. Αναδείχθηκαν ποσοστά⁴¹ ίσα με 94,28%. Η δεύτερη, αναλύει την αξιολόγηση αποτελεσμάτων για την κατηγορία κακοήθους (malignant) διάγνωσης. Αναδείχθηκε ποσοστό⁴² ίσο με 90,9%. Και για τις τρεις μετρικές συνολικά, παρατηρούνται ίδια αποτελέσματα μεταξύ τους.

Αποτελέσματα που προέκυψαν από τον ταξινομητή τυχαίων δασών

Ταυτόχρονα, έχει επίσης εξοπλιστεί και ένας ταξινομητής τυχαίων δασών (random forest). Οι παράμετροι που τον διατρέχουν, είναι η παρουσία εκτιμητών και κατάστασης τυχαιοποίησης. Οι τιμές τους ορίστηκαν ανάλογα με την απόδοση της εξόδου⁴³ του μοντέλου.

Για την αξιολόγηση των αποτελεσμάτων των μετρικών του, σε συνθήκες αναδρομικής εξάλειψης χαρακτηριστικών, αναδείχθηκαν συνολική ακρίβεια (accuracy) ίση με 96,49% και βαθμός διασταυρούμενης επικύρωσης (cross-validation score) ίσος με 95,7%. Πράγμα που καθιστά τα αποτελέσματά του αποδοτικά. Η απόδοση μετρικών στον βαθμό διασταυρούμενης επικύρωσης (cross-validation score) για την συνθήκη μη-επαναλαμβανόμενης εξάλειψης χαρακτηριστικών, ποικίλει⁴⁴. Ο μεγαλύτερος βαθμός ανέδειξε τιμή⁴⁵ ίση με 1. Ο μικρότερος βαθμός, ανέδειξε ποσοστό 88,23%.

Στην κατηγορία καλοήθους διάγνωσης, αναδείχθηκε ποσοστό⁴⁶ ίσο με 97,14%. Ενώ, για αυτή που εκπροσωπεί την κακοήγη διάγνωση, αναδείχθηκε ποσοστό ίσο με 95,45% και για τις τρεις μετρικές⁴⁷.

³⁹ Χωρίς να εμβαθύνει λεπτομερώς στις δύο κατηγορίες διάγνωσης

⁴⁰ Διαφέρουν από τις σφαιρικές αποδόσεις μετρικών για το μοντέλο.

⁴¹ Το ποσοστό αυτό ισχύει και για τις τρεις μετρικές, ακρίβειας, ανάκλησης και F1-score.

⁴² Ίδιο ποσοστό για τις τρεις μετρικές, ακρίβειας, ανάκλησης και F1-score.

⁴³ Αξιολογήθηκε για αριθμό 50 εκτιμητών.

⁴⁴ Η απόδοση είναι ίδια τόσο για τα δέντρα αποφάσεων όσο και για τα τυχαία δάση.

⁴⁵ Αποδόθηκε το 100% του βαθμού.

⁴⁶ Ισχύει για τις τρεις μετρικές ακρίβειας, ανάκλησης, F1-score (παρόμοια με πριν)

⁴⁷ ακρίβεια, ανάκληση, F1-score

Αποτελέσματα που προέκυψαν από τον ταξινομητή ADABOOST

Αξιοποιήθηκε και ένας αλγόριθμος ταξινόμησης ADABOOST⁴⁸. Εξαιτίας των συγκριτικών τους πλεονεκτημάτων σε σχέση με άλλους αλγορίθμους, εδώ, παρατηρούνται διαφορετικά αποτελέσματα στις μετρικές μεταξύ τους, ανεξάρτητα αν έχουν υποστεί εξάλειψη χαρακτηριστικών. Θα εξεταστούν δύο περιπτώσεις για κάθε συνθήκη, λαμβάνοντας υπόψη και τη διάγνωση.

Εξετάζοντας τη συνθήκη αναδρομικής εξάλειψης χαρακτηριστικών, έχουν αναδειχθεί συνολική ακρίβεια (accuracy) ίση με 96,49% (περίπου 96,5%) και βαθμός διασταυρούμενης επικύρωσης (cross-validation score) ίσος με 94,15%. Για την κατηγορία καλοήθους διάγνωσης, αναδείχθηκε ποσοστό⁴⁹ ίσο με 97,14%. Ενώ για την κατηγορία κακοήθους διάγνωσης, αποδόθηκε ποσοστό ίσο με 95,45% και για τις τρεις μετρικές⁵⁰.

Για τη συνθήκη μη-επαναλαμβανόμενης εξάλειψης χαρακτηριστικών, έχει αποδοθεί ποσοστό συνολικής ακρίβειας (accuracy) ίσο με 92,98% (περίπου 93%). Εντοπίζονται ιδιορρυθμίες στον βαθμό διασταυρούμενης επικύρωσης (cross-validation score) ανά μοίρασμα κατά δέκα φορές.

Μπορούμε να συμπεράνουμε ότι, ο μεγαλύτερος βαθμός ανέρχεται σε 98,03% και ο μικρότερος σε 90,19% από τα αποτελέσματα που δόθηκαν. Στην κατηγορία καλοήθους διάγνωσης, το ποσοστό⁵¹ ανέρχεται σε 94,28%. Για κακοήθη διάγνωση το ποσοστό ανέρχεται σε 90,9% και για τις τρεις μετρικές⁵².

Αποτελέσματα που προέκυψαν από τον ταξινομητή Naïve Bayes

Εκτός αυτού, έχουν χρησιμοποιηθεί πιθανολογικοί αλγόριθμοι. Κυριότερο παράδειγμα, αποτελεί ο αλγόριθμος Naïve Bayes. Θεωρήθηκε παραλλαγή Gauss (Gaussian Naïve Bayes) για την αξιολόγηση του μοντέλου. Στους πιθανολογικούς αλγορίθμους, τα αποτελέσματα τυχαίνει να διαφέρουν μεταξύ τους, εξαιτίας της παρουσίας τυχαιότητας.

Αρχές όπως τυχαιότητα ή αβεβαιότητα, τείνουν να επηρεάσουν την αξιολόγηση των αποτελεσμάτων. Ασχέτως αν εφαρμόστηκε εξάλειψη χαρακτηριστικών ή όχι, παρατηρούνται διαφορετικά αποτελέσματα μεταξύ των μετρικών τους.

⁴⁸ Αλγόριθμος προσαρμοστικής ενίσχυσης (adaptive boosting algorithm)

⁴⁹ Τδιο ποσοστό για τις τρεις μετρικές, ακρίβειας, ανάκλησης και F1-score.

⁵⁰ ακρίβεια, ανάκληση, F1-score

⁵¹ Και για τις τρεις μετρικές (ακρίβεια, ανάκληση, F1-score)

⁵² ακρίβεια, ανάκληση, F1-score

Εξετάζοντας τη συνθήκη αναδρομικής εξάλειψης χαρακτηριστικών, έχουν αναδειχθεί συνολική ακρίβεια (accuracy) ίση με 96,49% (περίπου 96,5%) και βαθμός διασταυρούμενης επικύρωσης (cross-validation score) ίσος με 90,4% (περίπου 90,5%).

Για την κατηγορία καλοήθους διάγνωσης, το ποσοστό και για τις τρεις μετρικές⁵³ αναδείχθηκε ίσο με 97,14%, ενώ για την κατηγορία κακοήθους διάγνωσης αποδόθηκε ποσοστό⁵⁴ ίσο με 95,45%.

Για τη συνθήκη μη-επαναλαμβανόμενης εξάλειψης χαρακτηριστικών, έχει αποδοθεί ποσοστό συνολικής ακρίβειας (accuracy), ίσο με 92,98% (περίπου 93%). Σε πιθανολογικό αλγόριθμο τέτοιο, όπως ο Naïve Bayes, εντοπίζονται ιδιορρυθμίες στον βαθμό διασταυρούμενης επικύρωσης (cross-validation score) ανά μοίρασμα.

Από τα αποτελέσματα που δόθηκαν, ας υποθέσουμε ότι ο μεγαλύτερος βαθμός ανέρχεται σε 96,07% και ο μικρότερος σε 88,23%. Στην κατηγορία καλοήθους διάγνωσης, το ποσοστό για τις τρεις μετρικές⁵⁵ ανέρχεται σε 94,28%, ενώ για κακοήθη διάγνωση, το ποσοστό⁵⁶ ανέρχεται σε 90,9%.

Γεγονός που καθιστά τα αποτελέσματα να μην είναι αρκετά αποδοτικά σε σχέση με αυτά που προέρχονται από τις τεχνικές που έχουν αναλυθεί ως τώρα.

Αποτελέσματα που προέκυψαν από τον ταξινομητή XGBoost

Τέλος, έχει εξοπλιστεί και ένας αλγόριθμος ταξινόμησης XGBoost⁵⁷. Αποτελεί τον πέμπτο από τους συνολικούς αλγορίθμους ταξινόμησης που χρησιμοποιήθηκαν κατά τη διάρκεια του πειράματος. Παρά την ευρωστία και την υπολογιστική απόδοσή του, τα αποτελέσματα των μετρικών διαφοροποιήθηκαν, ασχέτως αν εξετάζεται για δύο σενάρια⁵⁸.

Αναλύοντας τα αποτελέσματα που προέκυψαν μετά από αναδρομική εξάλειψη χαρακτηριστικών, αναδείχθηκε ποσοστό συνολικής ακρίβειας (accuracy) ίσο με 88,23% και βαθμού διασταυρούμενης επικύρωσης (cross-validation score) ίσο με 95,67%.

Για την κατηγορία καλοήθους διάγνωσης, η τιμή του ποσοστού⁵⁹ των μετρικών αναδείχθηκε ίση με 90,62%, ενώ για την κατηγορία κακοήθους διάγνωσης, αποδόθηκε ποσοστό 84,21% για τις τρεις μετρικές της ακρίβειας (precision), της ανάκλησης (recall) και της βαθμολογίας F1-score αντίστοιχα.

⁵³ (ακρίβεια, ανάκληση, F1-score)

⁵⁴ Και για τις τρεις μετρικές (ακρίβεια, ανάκληση, F1-score)

⁵⁵ (ακρίβεια, ανάκληση, F1-score)

⁵⁶ Και για τις τρεις μετρικές (ακρίβεια, ανάκληση, F1-score)

⁵⁷ Αλγόριθμος ακραίας ενίσχυσης κλίσης (extreme gradient boosting)

⁵⁸ εξάλειψη χαρακτηριστικών, αναδρομική ή μη-επαναλαμβανόμενη

⁵⁹ Ισχύει και για τις τρεις μετρικές (ακρίβεια, ανάκληση, F1-score), παρόμοια με πριν

Μετά από μη-επαναλαμβανόμενη εξάλειψη χαρακτηριστικών, τα αποτελέσματα ανέδειξαν ποσοστό για τη συνολική ακρίβεια ίσο με 96,07%. Παρατηρήθηκαν ιδιορρυθμίες εξίσου και στον βαθμό διασταυρούμενης επικύρωσης (cross-validation score) ανά μοίρασμα.

Ας υποθέσουμε, ο μεγαλύτερος βαθμός αποδίδεται με ποσοστό 96,10% και ο μικρότερος με ποσοστό 94,35%. Για την κατηγορία καλοήθους διάγνωσης, οι τιμές του ποσοστού και των τριών μετρικών⁶⁰ αναδείχθηκαν ίσες με 96,87%. Για την κατηγορία κακοήθους διάγνωσης, αποδόθηκε ποσοστό 94,73% αντίστοιχα, και στις τρεις μετρικές⁶¹.

Συμπεράσματα από τα αποτελέσματα του πρώτου σετ δεδομένων

Συμπεραίνουμε ότι, για το πρώτο σετ δεδομένων, τα αποτελέσματα αναδείχθηκαν αποδοτικά για σχεδόν κάθε τεχνική που αναλύθηκε.

Συγκρίνοντας τα αποτελέσματα των μετρικών και για τις δύο συνθήκες εξάλειψης χαρακτηριστικών, με βάση την απόδοση ακρίβειας στο καθένα, τα ποσοστά διαφέρουν μεταξύ τους ανά συνθήκη λαμβάνοντας υπόψη και την κατηγορία. Λίγες περιπτώσεις αποτελούν εξαίρεση, για των οποίων οι μετρικές ακρίβειας παρέμειναν σταθερές.

Προτεινόμενο μοντέλο στο τρέχον πείραμα, διατέλεσε ο ταξινομητής τυχαίων δασών. Ανεξάρτητα αν υπέστη εξάλειψη χαρακτηριστικών ή όχι, πέτυχε ακρίβεια ίση με 96,49%. Καθίσταται ως η υψηλότερη από όσες μετρικές συγκεντρώθηκαν και παρέμεινε σταθερή για κάθε περίπτωση που αναλύθηκε ξεχωριστά. Ισχύει για το πρώτο μοντέλο συνολικά.

⁶⁰ (ακρίβεια, ανάκληση, F1-score)

⁶¹ (ακρίβεια, ανάκληση, F1-score)

Αποτελέσματα για το δεύτερο σετ δεδομένων

Για το δεύτερο σετ δεδομένων, γνωρίζουμε ότι δεν έχει υποστεί εξάλειψη χαρακτηριστικών, εξαιτίας του μικρού μεγέθους που διαθέτει στα δεδομένα. Η αναντιστοιχία στα δείγματα⁶², αποτέλεσε αιτία να προκληθεί υποδειγματοληψία. Ως εκ τούτου, η αξιολόγηση των αποτελεσμάτων των μετρικών, προέκυψε για το πλαίσιο δεδομένων σφαιρικά. Ταυτόχρονα, θα αναλυθούν μεμονωμένες μετρικές για τις κατηγορίες πρόβλεψης.

Αποτελέσματα που προέκυψαν από τον ταξινομητή δέντρων αποφάσεων

Αυτή η μελέτη, χρησιμοποίησε έναν ταξινομητή δέντρων αποφάσεων. Από τα αποτελέσματα που δόθηκαν, το μεγαλύτερο ποσοστό αποδόθηκε σε μια τιμή ίση με 90,9%. Το μικρότερο, απέδωσε ποσοστό ίσο με 81,81%. Για κάθε μοίρασμα ξεχωριστά, παρατηρήθηκε αστάθεια στα αποτελέσματα της ακρίβειας. Εντοπίζεται και σε λοιπές μετρικές. Παραδείγματος χάρη, ο βαθμός διασταυρούμενης επικύρωσης (cross-validation score), έχει αναδείξει ποσοστό 90% ο μεγαλύτερος, 60% ο μικρότερος.

Τα αποτελέσματα που προέκυψαν, αντιπροσωπεύουν τη σφαιρική απόδοση μετρικών⁶³. Για κάθε κατηγορία, αποδίδονται διαφορετικά αποτελέσματα μετρικών⁶⁴. Θα αναλυθούν δύο διαφορετικές περιπτώσεις για κάθε μία εκ των δύο κατηγοριών πρόβλεψης για την ύπαρξη καρκίνου του μαστού, θετική (positive) και αρνητική (negative).

Η πρώτη περίπτωση, αναλύει την αξιολόγηση αποτελεσμάτων για την κατηγορία θετικής πρόβλεψης (positive). Δίνεται βάση σε μετρικές ακρίβειας (precision), ανάκλησης (recall), και βαθμολογίας F1-score. Παρατηρούνται διακυμάνσεις στα αποτελέσματα ανά μοίρασμα βάσει των αποτελεσμάτων που δόθηκαν, και θα αναλυθεί παρακάτω. Τα αποτελέσματα της ακρίβειας (precision) απέδωσαν ποσοστό ίσο με 83,33% για τη μικρότερη τιμή, ενώ για τη μεγαλύτερη, τιμή ίση με 1⁶⁵. Για την ανάκληση (recall), αποδόθηκε τιμή ίση με το 1⁶⁶ για την μεγαλύτερη τιμή. Το μικρότερο ποσοστό αναδείχθηκε σε 83,33%, και για την βαθμολογία F1-score, η μεγαλύτερη τιμή της, απέδωσε τιμή ίση με το 1⁶⁷. Η μικρότερη ποσοστιαία τιμή, πέτυχε ποσοστό ίσο με 80%.

⁶² Ανωμαλία σε μια από τις δύο κατηγορίες δεδομένων, με αποτέλεσμα το μέγεθος των δειγμάτων και στις δύο κατηγορίες, να παραμείνει ίδιο

⁶³ Χωρίς να εμβαθύνει λεπτομερώς στις δύο κατηγορίες πρόβλεψης

⁶⁴ Διαφέρουν από τις σφαιρικές αποδόσεις μετρικών για το μοντέλο.

⁶⁵ Αποδόθηκε το 100% της ακρίβειας.

⁶⁶ Αντιστοιχεί στο 100% της ανάκλησης.

⁶⁷ Αντιστοιχεί στο 100% της βαθμολογίας F1-score.

Η δεύτερη, αναλύει την αξιολόγηση αποτελεσμάτων για την κατηγορία αρνητικής (negative) πρόβλεψης. Τα αποτελέσματα της ακρίβειας (precision), απέδωσαν ποσοστό για την μεγαλύτερη τιμή ίσο με 83,33% και για την μικρότερη, ποσοστό ίσο με 80%. Για την ανάκληση (recall), η μεγαλύτερη ποσοστιαία τιμή, ανέδειξε τιμή ίση με το 1⁶⁸. Ενώ για τη μικρότερη, επιτεύχθηκε ποσοστό ίσο με 80%. Για τη βαθμολογία F1-score, το μεγαλύτερο ποσοστό αποδόθηκε σε 90,90%, ενώ το μικρότερο σε ποσοστιαία τιμή ίση με 80%.

Και για τις τρεις μετρικές μεμονωμένα, παρατηρούνται διακυμάνσεις στα αποτελέσματά τους, γεγονός που καθιστά τα αποτελέσματα που επιτεύχθηκαν να μην είναι αρκετά αποδοτικά. Αντίθετα, προσφέρονται σχετικά υψηλά αποτελέσματα στην πρόβλεψη.

Αποτελέσματα που προέκυψαν από τον ταξινομητή τυχαίων δασών

Ταυτόχρονα, έχει επίσης εξοπλιστεί και ένας ταξινομητής τυχαίων δασών (random forest). Αξιολογήθηκε για αριθμό 50 εκτιμητών, αποδίδοντας βέλτιστες επιδόσεις κατά τη διάρκεια της εκπαίδευσης. Για την αξιολόγηση των αποτελεσμάτων των μετρικών του, αναδείχθηκε συνολική ακρίβεια ίση με 72,72%. Εντοπίζονται ιδιορρυθμίες μεταξύ των αποτελεσμάτων του βαθμού διασταυρούμενης επικύρωσης (cross-validation score). Συμπεραίνουμε ότι, ο μεγαλύτερος βαθμός ανέρχεται σε 88,88% και ο μικρότερος σε 44,44% από τα αποτελέσματα που δόθηκαν.

Για την κατηγορία θετικής πρόβλεψης, το ποσοστό⁶⁹ για την ακρίβεια (precision) ανέρχεται σε 71,42%, την ανάκληση (recall) σε 83,33%, και τη βαθμολογία F1-score σε 76,92%. Για την κατηγορία αρνητικής πρόβλεψης, το ποσοστό για την ακρίβεια (precision), ανέρχεται σε 75%, την ανάκληση (recall) σε 60%, και τη βαθμολογία F1-score σε 66,66%.

Και στις τρεις μετρικές που εξετάστηκαν παρατηρήθηκαν διακυμάνσεις⁷⁰ στα αποτελέσματά τους, με εξαίρεση τις σφαιρικές μετρικές⁷¹.

⁶⁸ Απέδωσε το 100% της ανάκλησης.

⁶⁹ Διαφορετικά ποσοστά για κάθε κατηγορία, κάθε μεμονωμένη μετρική

⁷⁰ Για κάθε κατηγορία, μεμονωμένα

⁷¹ η συνολική ακρίβεια του μοντέλου (accuracy), να είναι μονίμως σταθερή

Αποτελέσματα που προέκυψαν από τον ταξινομητή ADABOOST

Αξιοποιήθηκε και ένας αλγόριθμος ταξινόμησης ADABOOST⁷². Παρά τα συγκριτικά του πλεονεκτήματα, είναι ο μοναδικός αλγόριθμος μέχρι στιγμής που διατηρεί σταθερές⁷³ τις τιμές του, στην παρούσα μελέτη.

Συγκεκριμένα, έχει αποδοθεί ποσοστό συνολικής ακρίβειας (accuracy) ίσο με 63,63%. Όμως, εντοπίζονται ιδιορρυθμίες στον βαθμό διασταυρούμενης επικύρωσης (cross-validation score). Για τον μεγαλύτερο βαθμό, αποδόθηκε ποσοστιαία τιμή ίση με 1⁷⁴. Ενώ, για τον μικρότερο, αναδείχθηκε ποσοστό ίσο με 50%, από τα αποτελέσματα που δόθηκαν.

Στην κατηγορία θετικής πρόβλεψης, το ποσοστό για την ακρίβεια (precision), την ανάκληση (recall), και τη βαθμολογία F1-score, ανέρχεται σε 66,66%. Για αρνητική πρόβλεψη, και για τις τρεις μετρικές, το ποσοστό ανέρχεται σε 60%.

Αποτελέσματα που προέκυψαν από τον ταξινομητή Naïve Bayes

Έχουν χρησιμοποιηθεί επίσης πιθανολογικοί αλγόριθμοι. Κυριότερο παράδειγμα, αποτελεί ο αλγόριθμος Naïve Bayes. Θεωρήθηκε παραλλαγή Gauss (Gaussian Naïve Bayes) για την αξιολόγηση του μοντέλου. Παρά την παρουσία τυχαιότητας, παρατηρείται σταθερότητα στα αποτελέσματά του, με ελάχιστες περιπτώσεις να αποτελούν εξαίρεση. Έχει αναδειχθεί συνολική ακρίβεια (accuracy) ίση με 63,63%. Και εδώ, εντοπίζονται ιδιορρυθμίες στον βαθμό διασταυρούμενης επικύρωσης (cross-validation score). Ο μεγαλύτερος βαθμός, απέδωσε ποσοστό 77,77%, ενώ ο μικρότερος 44,44%.

Για την κατηγορία θετικής πρόβλεψης, το ποσοστό για την ακρίβεια (precision) ανέρχεται σε 60%. Για την μετρική της ανάκλησης (recall) επιτεύχθηκε τιμή ίση με 1⁷⁵, ενώ για τη βαθμολογία F1-score, το ποσοστό ανέρχεται σε 75%. Για την κατηγορία αρνητικής πρόβλεψης, η μετρική της ακρίβειας (precision) ανέδειξε τιμή ίση με 1⁷⁶. Το ποσοστό για την ανάκληση (recall) ανέρχεται σε 20%, και για τη βαθμολογία F1-score σε 33,33%.

Και για τις τρεις μετρικές που εξετάστηκαν, παρατηρήθηκαν διακυμάνσεις στα αποτελέσματα, με εξαίρεση τις σφαιρικές μετρικές⁷⁷. Τα αποτελέσματα που προέρχονται από τον αλγόριθμο Naïve Bayes, δεν είναι αρκετά αποδοτικά για ένα μικρό σε μέγεθος σετ δεδομένων.

⁷² Αλγόριθμος προσαρμοστικής ενίσχυσης (adaptive boosting algorithm)

⁷³ Ελάχιστες περιπτώσεις υπ' εξαίρεση

⁷⁴ Αποδόθηκε το 100% του βαθμού.

⁷⁵ Αποδόθηκε το 100% της ανάκλησης.

⁷⁶ Αποδόθηκε το 100% της ακρίβειας.

⁷⁷ Αποτέλεσμα, η συνολική ακρίβεια του μοντέλου (accuracy), να είναι μονίμως σταθερή.

Αποτελέσματα που προέκυψαν από τον ταξινομητή XGBoost

Τέλος, έχει εξοπλιστεί και ένας αλγόριθμος ταξινόμησης XGBoost⁷⁸. Εδώ, παρατηρήθηκε σταθερότητα μόνο στο ποσοστό συνολικής ακρίβειας (accuracy) όσον αφορά μια σφαιρική ανάλυση αποτελεσμάτων.

Αναλύοντας τα αποτελέσματα που προέκυψαν, αναδείχθηκε ποσοστό συνολικής ακρίβειας (accuracy) ίσο με 66,66%. Παρατηρήθηκαν ιδιορρυθμίες και στον βαθμό διασταυρούμενης επικύρωσης (cross-validation score) ανά μοίρασμα. Ο μεγαλύτερος βαθμός, με βάση τα αποτελέσματα που δόθηκαν, αποδίδεται με ποσοστό 79,58%, και ο μικρότερος, με ποσοστό 70,69%.

Για την κατηγορία θετικής πρόβλεψης, το ποσοστό για την ακρίβεια (precision) ανέρχεται σε 66,66%, την ανάκληση (recall) σε 50% και τη βαθμολογία F1-score σε 57,14%.

Για την κατηγορία αρνητικής πρόβλεψης, το ποσοστό για την ακρίβεια (precision) ανέρχεται σε 66,66%, την ανάκληση (recall) σε 80% και τη βαθμολογία F1-score σε 57,14%. Και στις τρεις μετρικές που έχουν εξεταστεί, εντοπίζονται ιδιορρυθμίες στα αποτελέσματα.

Παρατηρούμε ότι, τα αποτελέσματα που προέρχονται από τον αλγόριθμο XGBoost, δεν είναι αρκετά αποδοτικά σε σχέση με τα αποτελέσματα που επέφεραν οι υπόλοιπες αλγοριθμικές τεχνικές. Βάσει των δοθέντων αποτελεσμάτων, δεν μπορεί να επιτευχθεί το 100% των μετρικών έπειτα από χρήση του αλγορίθμου ταξινόμησης XGBoost, ιδιαίτερα για ένα σετ δεδομένων με μικρό μέγεθος χαρακτηριστικών ή/και δειγμάτων.

⁷⁸ Αλγόριθμος ακραίας ενίσχυσης κλίσης (extreme gradient boosting)

Συμπεράσματα από τα αποτελέσματα του δεύτερου σετ δεδομένων

Συμπερασματικά, για το δεύτερο σετ δεδομένων, παρατηρείται αστάθεια στα αποτελέσματα των μετρικών του. Αυτό οφείλεται στο μικρό μέγεθος χαρακτηριστικών και δειγμάτων που διαθέτει, καθώς και στο γεγονός ότι έχει υποστεί υποδειγματοληψία⁷⁹ στο μέγεθος δειγμάτων του. λόγω αναντιστοιχίας στο πλήθος των δεδομένων.

Από τα αποτελέσματα που δόθηκαν, τα ποσοστά διαφέρουν μεταξύ τους τόσο σφαιρικά, όσο και ανά κατηγορία πρόβλεψης μεμονωμένα. Προέκυψαν έπειτα από σύγκριση των μετρικών των αλγορίθμων ταξινόμησης που έχουν χρησιμοποιηθεί. Λίγες περιπτώσεις παρουσίασαν σταθερότητα στις τιμές των μετρικών τους, χωρίς να ληφθεί υπόψη η κατηγορία πρόβλεψης που εξετάζεται.

Ειδικά, στους αλγορίθμους ταξινόμησης που προορίζονται σε ενίσχυση προσαρμοστικότητας, έχει αποδοθεί σταθερότητα στα αποτελέσματα της συνολικής ακρίβειας (accuracy) σφαιρικά. Εξαίρεση, αποτελούν οι μεμονωμένες κατηγορίες, οι οποίες παρουσιάζουν ποικίλες διακυμάνσεις στα αποτελέσματα των μετρικών τους.

Προτεινόμενο μοντέλο, εξετάστηκε και επιλέχθηκε ο ταξινομητής δέντρων αποφάσεων. Παρά την αστάθεια στις μετρικές που αποδόθηκαν τόσο σφαιρικά όσο και στο κομμάτι της πρόβλεψης για κάθε κατηγορία, ανέδειξε αποτελέσματα συνολικής ακρίβειας (accuracy) μεταξύ 81,81% και 90,9%. Η μεγαλύτερη τιμή συνολικής ακρίβειας (accuracy) που δόθηκε, απέδωσε ποσοστό ίσο με 90,9%. Προέκυψε από την αξιολόγηση των αποτελεσμάτων του ταξινομητή δέντρων αποφάσεων. Κατά συνέπεια, οδήγησε στην τελική επιλογή για την πρόβλεψη ύπαρξης καρκίνου του μαστού.

⁷⁹ λόγω αναντιστοιχίας στο πλήθος των δεδομένων

Κεφάλαιο 9

Περιγραφή των εφαρμογών πρόβλεψης

ΑΦΟΥ προβήκαμε στη φάση της αξιολόγησης αποτελεσμάτων μετρικών για κάθε αλγόριθμο ταξινόμησης, επόμενο βήμα αποτελεί η έναρξη της φάσης ανάπτυξης εφαρμογών. Θα περιγράψουμε τη διαδικασία σχεδίασης και υλοποίησης δύο ξεχωριστών εφαρμογών.

Η πρώτη προορίζεται ευρέως για διαδικτυακή χρήση, ενώ η δεύτερη, σχεδιάστηκε περισσότερο για κινητές συσκευές. Έχει προταθεί μια μεθοδολογία που εφαρμόστηκε για την σχεδίαση και υλοποίηση και των δύο εφαρμογών, χρησιμοποιώντας αναλυτικά και εξονυχιστικά βήματα.

Περιγραφή των βημάτων για τη σχεδίαση και των δύο εφαρμογών

Για την διαδικασία σχεδίασης και υλοποίησης και των δύο εφαρμογών, ακολουθήθηκε μια σειρά βημάτων. Τα βήματα που ακολουθούν, αποτελούν την τελική μεθοδολογία⁸⁰ που εφαρμόστηκε και προτάθηκε, ξεκινώντας από ένα αρχικό στάδιο και φτάνοντας στο τελικό αποτέλεσμα.

Κοινά βήματα και για τις δύο εφαρμογές, ακολουθούν ως εξής:

1. Εισαγωγή απαραίτητων βιβλιοθηκών ή/και πακέτων της γλώσσας προγραμματισμού Python.
2. Εισαγωγή βιβλιοθηκών για διαδικασίες σειριοποίησης⁸¹ και διασώληνωσης⁸².
3. Εισαγωγή βιβλιοθηκών για αποφυγή σφαλμάτων⁸³ και έλεγχο προειδοποιήσεων⁸⁴.
4. Φόρτωση προεκπαιδευμένου μοντέλου χρησιμοποιώντας τη βιβλιοθήκη joblib της Python.
5. Σχεδιασμός της διεπαφής χρησιμοποιώντας τη βιβλιοθήκη Streamlit της Python.
6. Δημιουργία διανύσματος χαρακτηριστικών.
7. Ενσωμάτωση διανύσματος χαρακτηριστικών στο κουμπί πρόβλεψης αποτελέσματος.
8. Εμφάνιση αποτελέσματος στην οθόνη.
9. Ενσωμάτωση εργαλείου προσαρμογής ορίου πιθανότητας, χρησιμοποιώντας εργαλεία της βιβλιοθήκης Streamlit της Python.

⁸⁰ Ισχύει και για τις δύο εφαρμογές.

⁸¹ Βιβλιοθήκη joblib της Python

⁸² Βιβλιοθήκη pickle της Python

⁸³ Βιβλιοθήκη os της Python

⁸⁴ Βιβλιοθήκη warnings της Python

Όνομα, προδιαγραφές και περιεχόμενα της εφαρμογής

Η εφαρμογή, ονομάζεται “BreastDiagApp”, και για τις δύο εκδοχές που υλοποιήθηκαν. Αναπτύχθηκε εξ’ ολοκλήρου σε γλώσσα προγραμματισμού Python.

Για τον σχεδιασμό της διεπαφής της εφαρμογής, χρησιμοποιήθηκε η βιβλιοθήκη ανοιχτού κώδικα Streamlit. Προσφέρει απλά και εύκαμπτα εργαλεία για τον σχεδιασμό αποτελεσματικών εφαρμογών πρόβλεψης δεδομένων.

Χρησιμοποιήθηκε μια σελίδα⁸⁵ για το σχεδιασμό της διεπαφής. Στόχο έχει, την παροχή μιας φιλικής προς τους χρήστες εφαρμογή.

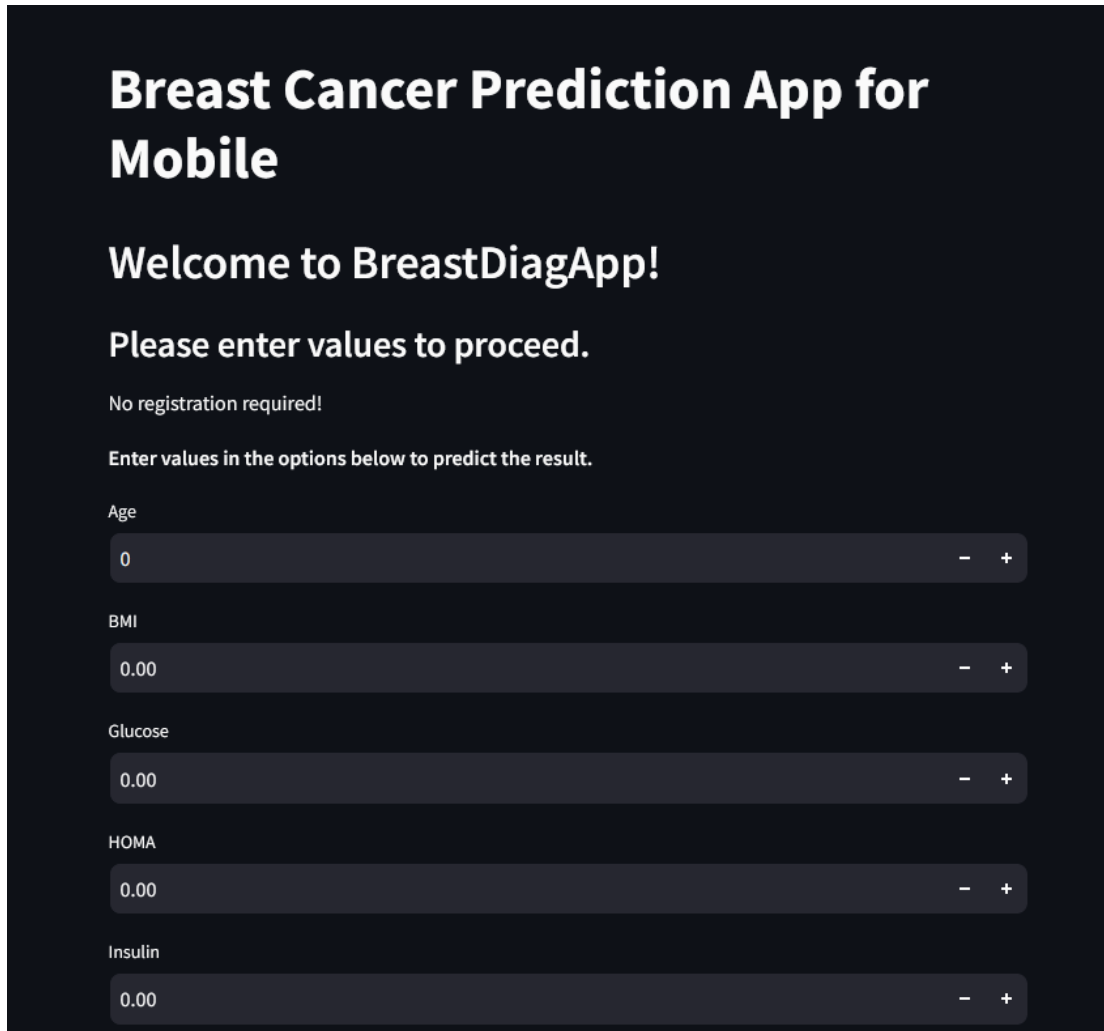
Έτσι, οι χρήστες μπορούν να εισάγουν τις επιθυμητές τιμές τους εντός της εφαρμογής, λαμβάνοντας άμεσα αποτελέσματα για την τελική πρόβλεψη.

⁸⁵ Για λόγους απλότητας

Απεικόνιση διεπαφής για την πρώτη εφαρμογή

Η πρώτη εφαρμογή έχει σχεδιαστεί και προορίζεται ευρέως για διαδικτυακή χρήση. Αυτό οφείλεται στο μεγάλο μέγεθος των δεδομένων⁸⁶ του μοντέλου.

Παρακάτω, παρατίθενται ενδεικτικές εικόνες απεικόνισης της διεπαφής της εφαρμογής στην οθόνη.

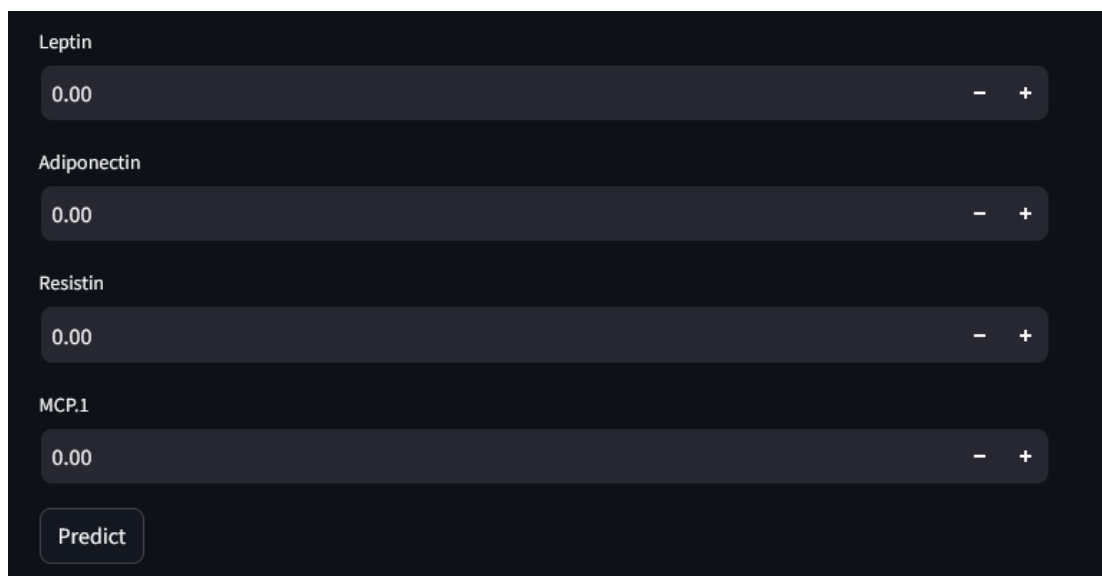


The screenshot displays the user interface of the 'Breast Cancer Prediction App for Mobile'. The title 'Breast Cancer Prediction App for Mobile' is prominently displayed at the top in a large, bold, white font. Below the title, a welcome message 'Welcome to BreastDiagApp!' is shown in a slightly smaller bold font. A prompt 'Please enter values to proceed.' is followed by a note 'No registration required!'. Below this, a instruction 'Enter values in the options below to predict the result.' is provided. The interface features five input fields, each with a label and a corresponding value: 'Age' (0), 'BMI' (0.00), 'Glucose' (0.00), 'HOMA' (0.00), and 'Insulin' (0.00). Each input field is a dark gray bar with a white minus sign on the left and a white plus sign on the right, indicating a range or increment/decrement functionality.

Εικόνα 24: Απεικόνιση της διεπαφής⁸⁷ της εφαρμογής στην οθόνη.

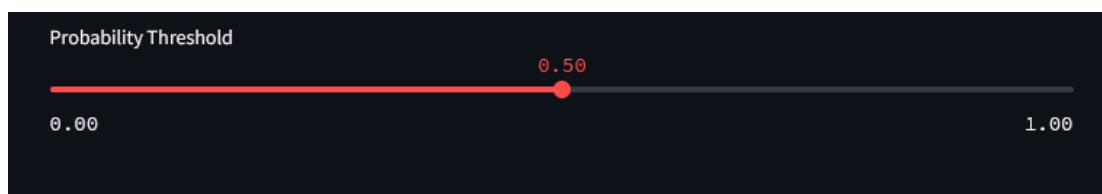
⁸⁶ Δείγματα και χαρακτηριστικά που το διατρέχουν

⁸⁷ Εξετάζεται η προβολή της για διαδικτυακή χρήση.



The screenshot displays a dark-themed user interface for a prediction application. It features four input fields, each labeled with a biomarker name: 'Leptin', 'Adiponectin', 'Resistin', and 'MCP.1'. Each field contains the numerical value '0.00' and is flanked by minus and plus icons for adjustment. Below these fields is a 'Predict' button.

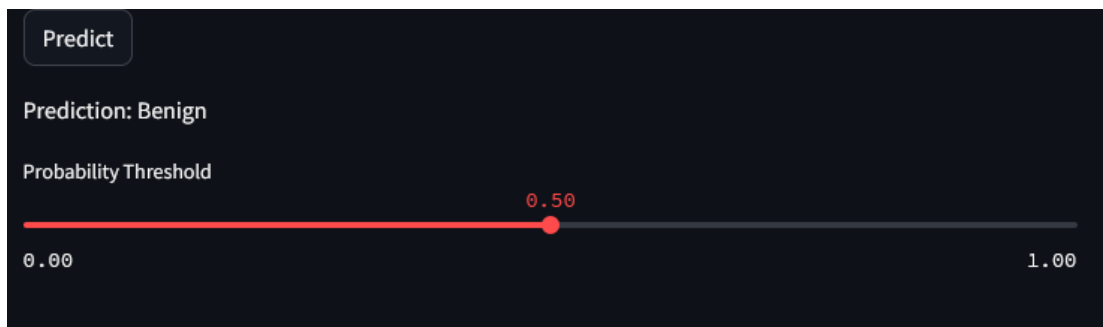
Εικόνα 25: Απεικόνιση της διεπαφής⁸⁸ της εφαρμογής στην οθόνη.



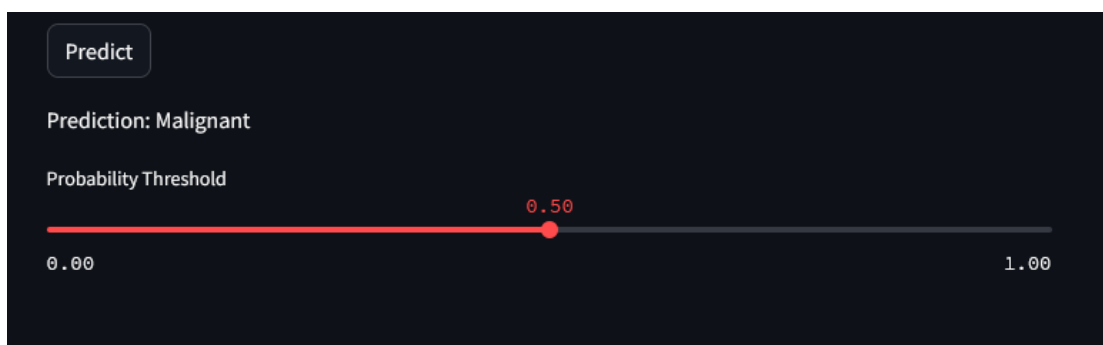
The screenshot shows a 'Probability Threshold' slider. The slider is a horizontal line with a red dot indicating the current threshold value of 0.50. The range is marked from 0.00 on the left to 1.00 on the right.

Εικόνα 26: Απεικόνιση του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος στο κάτω μέρος της οθόνης.

⁸⁸ Εξετάζεται η προβολή της για διαδικτυακή χρήση.



Εικόνα 27: Απεικόνιση⁸⁹ του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος, μαζί με το αποτέλεσμα της πρόβλεψης.



Εικόνα 28: Απεικόνιση⁹⁰ του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος, μαζί με το αποτέλεσμα της πρόβλεψης.

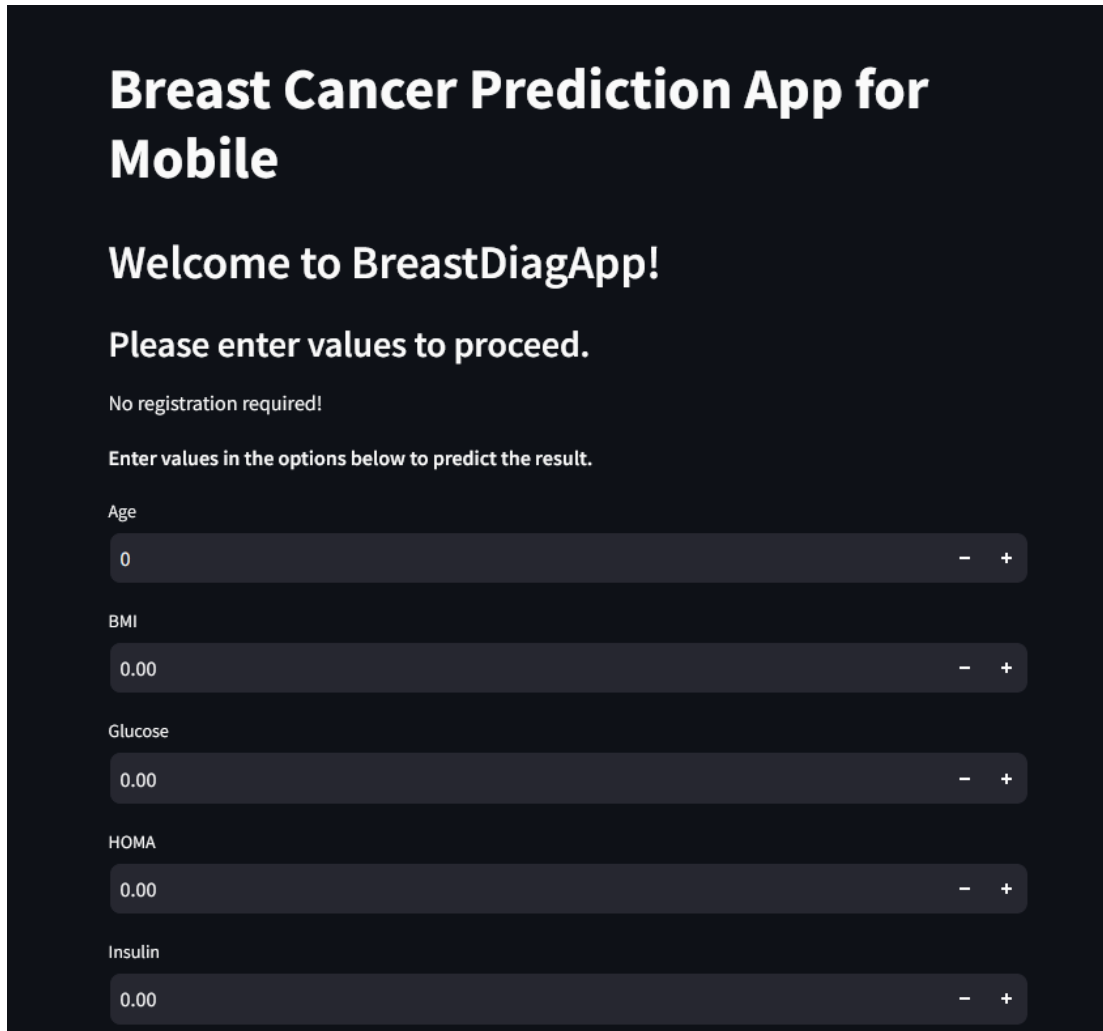
⁸⁹ Στην παρούσα περίπτωση, εξετάζεται για καλοήγη διάγνωση καρκίνου του μαστού.

⁹⁰ Εξετάζεται για κακοήγη διάγνωση καρκίνου του μαστού.

Απεικόνιση διεπαφής για την δεύτερη εφαρμογή

Η δεύτερη εφαρμογή, έχει σχεδιαστεί και προορίζεται κυρίως για κινητές συσκευές. Αυτό, οφείλεται στο μικρό μέγεθος των δεδομένων⁹¹ του μοντέλου και του σετ δεδομένων που εξετάστηκε.

Παρακάτω, παρατίθενται ενδεικτικές εικόνες απεικόνισης της διεπαφής της εφαρμογής στην οθόνη.

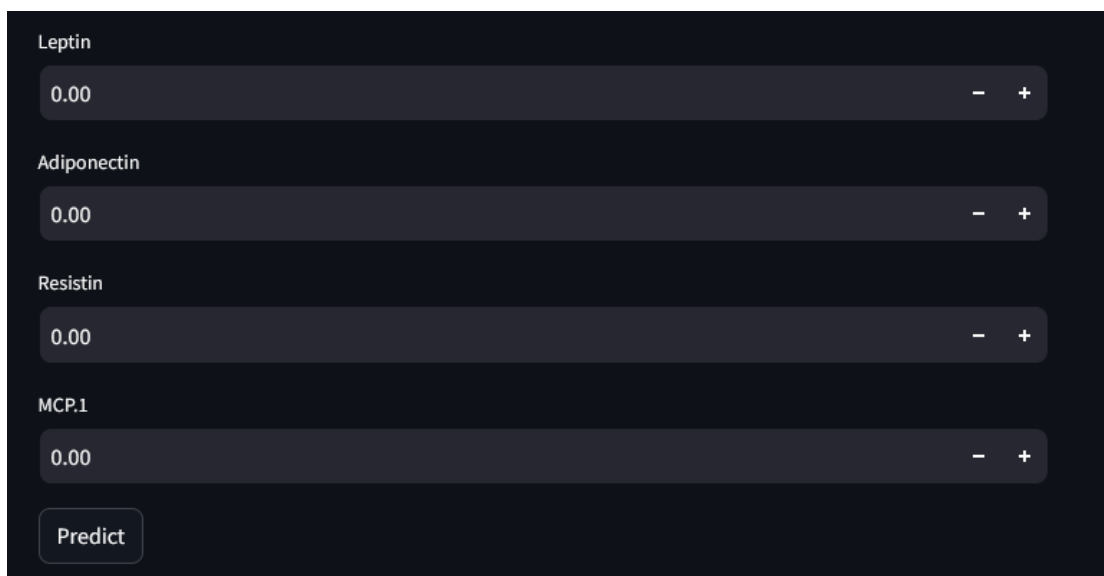


The screenshot displays the user interface of the 'Breast Cancer Prediction App for Mobile'. The title 'Breast Cancer Prediction App for Mobile' is prominently displayed at the top. Below it, a welcome message 'Welcome to BreastDiagApp!' is shown. The app prompts the user to 'Please enter values to proceed.' and notes that 'No registration required!'. It instructs the user to 'Enter values in the options below to predict the result.' The interface features five input fields, each with a label and a numeric value: 'Age' (0), 'BMI' (0.00), 'Glucose' (0.00), 'HOMA' (0.00), and 'Insulin' (0.00). Each input field is accompanied by minus and plus buttons for adjustment.

Εικόνα 29: Απεικόνιση της διεπαφής⁹² της εφαρμογής στην οθόνη.

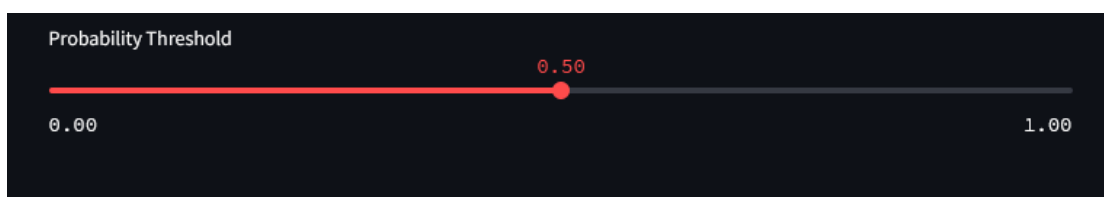
⁹¹ Δείγματα και χαρακτηριστικά που το διατρέχουν

⁹² Εξετάζεται η προβολή της για διαδικτυακή χρήση.



The screenshot displays a dark-themed user interface for a prediction application. It features four input fields, each labeled with a biomarker name: 'Leptin', 'Adiponectin', 'Resistin', and 'MCP1'. Each field contains the numerical value '0.00' and is flanked by minus and plus icons for adjustment. Below these fields is a 'Predict' button.

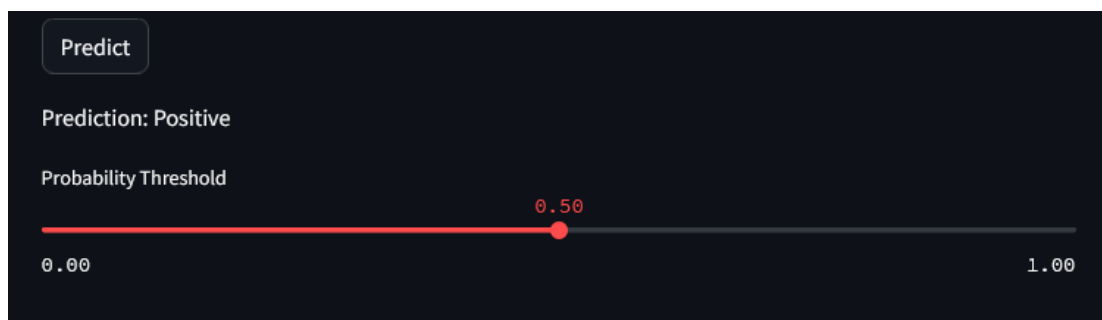
Εικόνα 30: Απεικόνιση της διεπαφής⁹³ της εφαρμογής στην οθόνη.



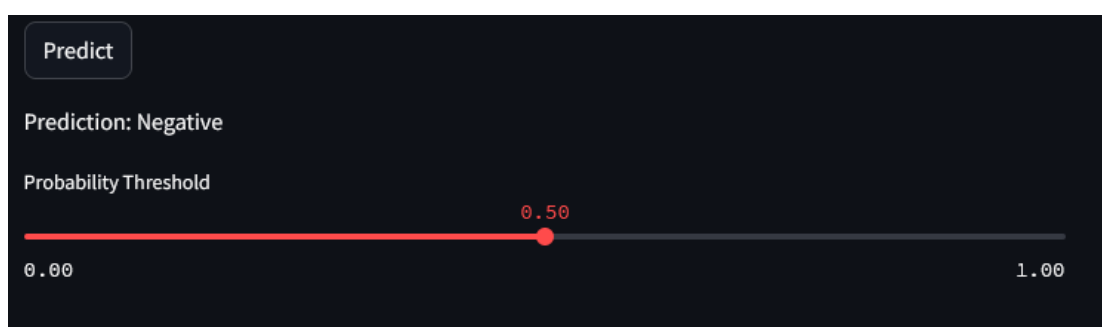
The screenshot shows a 'Probability Threshold' slider. A horizontal red line represents the range from 0.00 to 1.00. A red dot is positioned at the 0.50 mark, indicating the current threshold value.

Εικόνα 31: Απεικόνιση του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος στο κάτω μέρος της οθόνης.

⁹³ Εξετάζεται η προβολή της για διαδικτυακή χρήση.



Εικόνα 32: Απεικόνιση⁹⁴ του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος, μαζί με το αποτέλεσμα της πρόβλεψης.



Εικόνα 33: Απεικόνιση⁹⁵ του εργαλείου προσαρμογής ορίου πιθανότητας αποτελέσματος, μαζί με το αποτέλεσμα της πρόβλεψης.

⁹⁴ Στην παρούσα περίπτωση, εξετάζεται για θετική πρόβλεψη καρκίνου του μαστού.

⁹⁵ εξετάζεται για αρνητική πρόβλεψη καρκίνου του μαστού

Κεφάλαιο 10

Συμπεράσματα

ΕΝ κατακλείδι, συμπεραίνουμε ότι και οι δύο εφαρμογές που αναπτύχθηκαν σ' αυτή την εργασία, επέφεραν αποδοτικά και πολλά υποσχόμενα αποτελέσματα. Αποδείχθηκε τόσο στην ακριβή πρόβλεψη όσο και την απόδοση των κατάλληλων αποτελεσμάτων στους ασθενείς.

Μέχρι στιγμής, τα στάδια που έλαβαν χώρα έχουν αποδώσει καρπούς στην ενδυνάμωση ατόμων για την κατανόηση των κινδύνων εμφάνισης και την έγκαιρη ανίχνευση και πρόληψη καρκίνου του μαστού.

Σε καμία περίπτωση δεν πρέπει να αντικατασταθεί η κλασική ψηφιακή μαστογραφία, από νέες τεχνολογίες. Οι νέες τεχνολογίες περιλαμβάνουν μαγνητικές μαστογραφίες. Καθίσταται σαφές ότι κάθε μέθοδος διαφέρει μεταξύ της, όχι μόνο τεχνολογικά, αλλά και αποτελεσματικά.

Η ερμηνεία των αποτελεσμάτων θα πρέπει να γίνεται με συνδυασμό και των τριών μεθόδων⁹⁶ όπως και εντοπισμός ανάλογων παθήσεων στο μαστό. (όγκοι κτλ.). Είναι προφανές ότι κάθε μέθοδος συμπληρώνει την άλλη μεταξύ τους.

Η εξέταση της μαστογραφίας είναι σημαντική για τις γυναίκες. Μπορεί να σώσει τη ζωή τους πολύ πριν εντοπιστεί ένδειξη καρκίνου του μαστού. Η διεξαγωγή υπερήχων μαστού, ανταποκρίνεται περισσότερο σε νεαρές γυναίκες, που βρίσκονται σε αναπαραγωγική ηλικία. Έτσι, αποφεύγεται η ακτινοβολή του μαστικού αδένα των γυναικών. Πραγματοποιείται έλεγχος μαστών με τη χρήση υπερήχων, διότι είναι ακίνδυνη μέθοδος.

Η χρήση εννοιών και μεθόδων βιοπληροφορικής, στοχεύει στην αναλυτική κατανόηση του ζητήματος. Απευθύνεται πιο ευέλικτα και αποτελεσματικά στο κοινό. Τελευταία, με την άνθηση της τεχνητής νοημοσύνης, έχουν κερδίσει μεγάλο έδαφος τεχνικές μηχανικής μάθησης για την πρόγνωση καρκίνου του μαστού.

⁹⁶ μαγνητική/κλασική/ψηφιακή μαστογραφία

Έχει παρατηρηθεί ότι οι μέθοδοι μηχανικής μάθησης έχουν αποδώσει καρπούς σε εργασίες εποπτευόμενης μάθησης, πόσω μάλλον και στη διάγνωση καρκίνων. Δεν ισχύει αυστηρά για τον ιατρικό τομέα.

Εκτός αυτού, χάρη στα πλεονεκτήματα που παρέχονται από την τεχνητή νοημοσύνη και την μηχανική μάθηση, ανάπτυξη και χρήση συστημάτων υποστήριξης κλινικών αποφάσεων με έμφαση στην διάγνωση συμπτωμάτων ή πιθανών ενδείξεων καρκίνου του μαστού έχουν αυξήσει την δημοτικότητά τους και στρέφονται περισσότερο στο κοινό.

Τέλος, οι έννοιες «διάγνωση» και «πρόγνωση», δεν πρέπει να συγχέονται μεταξύ τους, ούτε πρέπει να συγχέεται η χρονολογική τους σειρά.

Βιβλιογραφία

- [1] Yixiao Feng, Mia Spezia, Shifeng Huang, Chengfu Yuan, Zongyue Zeng, Linghuan Zhang, Xiaojuan Ji, Wei Liu, Bo Huang, Wenping Luo, Bo Liu, Yan Lei, Scott Du, Akhila Vuppalapati, Hue H. Luu, Rex C. Haydon, Tong-Chuan He, Guosheng Ren, Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis, *Genes & Diseases*, Volume 5, Issue 2, 2018, Pages 77-106, ISSN 2352-3042, <https://doi.org/10.1016/j.gendis.2018.05.001>.
- [2] Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, Shi W, Jiang J, Yao PP, Zhu HP. Risk Factors and Preventions of Breast Cancer. *Int J Biol Sci*. 2017 Nov 1;13(11):1387-1397. doi: 10.7150/ijbs.21635. PMID: 29209143; PMCID: PMC5715522.
- [3] Akram, M., Iqbal, M., Daniyal, M. *et al*. Awareness and current knowledge of breast cancer. *Biol Res* **50**, 33 (2017). <https://doi.org/10.1186/s40659-017-0140-9>
- [4] Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, Shi W, Jiang J, Yao PP, Zhu HP. 2017. Risk Factors and Preventions of Breast Cancer. *Int J Biol Sci*. 13(11):1387-1397.
- [5] F. Sun and X. Yang, "Advancing Breast Cancer Diagnosis: A Comprehensive Study of Machine Learning Algorithms on Histological Tumor Characteristics," 2023 International Conference on Data Science & Informatics (ICDSI), Bhubaneswar, India, 2023, pp. 302-306, doi: 10.1109/ICDSI60108.2023.00064. keywords: {Logistic regression;Machine learning algorithms;Databases;Soft sensors;Feature extraction;Breast cancer;Linear discriminant analysis;Breast cancer;machine learning;Wisconsin breast cancer database},
- [6] T. Zhu, X. Fu and Y. Huang, "The impact and prospects of computer-aided technology for promoting universal breast cancer screening," 2023 2nd International Conference on Health Big Data and Intelligent Healthcare (ICHIH), Zhuhai, China, 2023, pp. 152-159, doi: 10.1109/ICHIH60370.2023.10395941. keywords: {Computational modeling;Sociology;Medical services;Big Data;Breast cancer;Statistics;Diseases;breast cancer;universal screening;computer-aided technology;influencing factors},
- [7] N. Panwar and K. Sethi, "Breast cancer :Early diagnosis by using artificial intelligence," 2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, 2022, pp. 1-4, doi: 10.1109/CISCT55310.2022.10046467. keywords: {Training;Support vector machine classification;Feature extraction;Breast cancer;Real-time systems;Reliability;Artificial intelligence;Breast Cancer;Support Vector;KNN;Naïve Bayes;Accuracy;Precision},

- [8] S. Ruban, M. Moosa Jabeer and R. Shenoy Basti, "Improvising Breast Cancer detection using CNN, VGG and SSD Algorithms," 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICT), Kannur, India, 2022, pp. 1772-1776, doi: 10.1109/ICICT54557.2022.9918009. keywords: {Deep learning;Visualization;Hospitals;Training data;Detectors;Breast cancer;Mammography;Mammography;Breast Cancer Detection;Artificial Intelligence;ConvNet;SSD},
- [9] Philippe Autier, Mathieu Boniol, Mammography screening: A major issue in medicine, *European Journal of Cancer*, Volume 90, 2018, Pages 34-62, ISSN 0959-8049, <https://doi.org/10.1016/j.ejca.2017.11.002>.
- [10] Global Cancer Observatory/Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, in press.
- [11] P. Koczkodaj, U. Sulkowska, J. Gotlib, M. Mańczuk, Breast cancer mortality trends in Europe among women in perimenopausal and postmenopausal age (45+), *Arch Med Sci*, 16 (1) (2020), pp. 146-156
- [12] Kimberly Badal et al., Cost of breast cancer screening in the USA: Comparison of current practice, advocated guidelines, and a personalized risk-based approach.. *JCO* 41, e18917-e18917(2023). DOI:10.1200/JCO.2023.41.16_suppl.e18917
- [13] Løberg, M., Lousdal, M.L., Bretthauer, M. et al. Benefits and harms of mammography screening. *Breast Cancer Res* 17, 63 (2015). <https://doi.org/10.1186/s13058-015-0525-z>
- [14] Silva C, Perestrelo R, Silva P, Tomás H, Câmara JS. Breast Cancer Metabolomics: From Analytical Platforms to Multivariate Data Analysis. A Review. *Metabolites*. 2019 May 22;9(5):102. doi: 10.3390/metabo9050102. PMID: 31121909; PMCID: PMC6572290.
- [15] Neagu AN, Whitham D, Buonanno E, Jenkins A, Alexa-Stratulat T, Tamba BI, Darie CC. Proteomics and its applications in breast cancer. *Am J Cancer Res*. 2021 Sep 15;11(9):4006-4049. PMID: 34659875; PMCID: PMC8493401.

- [16] James, T.A., Wade, J.E. and Sprague, B.L. (2016), The impact of mammographic screening on the surgical management of breast cancer. *J. Surg. Oncol.*, 113: 496-500.
<https://doi.org/10.1002/jso.24184>
- [17] Rossi, C., Cicalini, I., Cufaro, M.C. et al. Breast cancer in the era of integrating “Omics” approaches. *Oncogenesis* 11, 17 (2022). <https://doi.org/10.1038/s41389-022-00393-8>
- [18] Mohammed Amine Naji, Sanaa El Filali, Kawtar Aarika, EL Habib Benlahmar, Rachida Ait Abdelouhahid, Olivier Debauche, Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis, *Procedia Computer Science*, Volume 191, 2021, Pages 487-492, ISSN 1877-0509,
<https://doi.org/10.1016/j.procs.2021.07.062>.
- [19] Hiba Asri, Hajar Mousannif, Hassan Al Moatassime, Thomas Noel, Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis, *Procedia Computer Science*, Volume 83, 2016, Pages 1064-1069, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2016.04.224>.
- [20] Chaurasia V, Pal S, Tiwari B. Prediction of benign and malignant breast cancer using data mining techniques. *Journal of Algorithms & Computational Technology*. 2018;12(2):119-126.
doi:10.1177/1748301818756225
- [21] Laura M. Bozzuto, Breast cancer risk reduction: who, why, and what?, *Best Practice & Research Clinical Obstetrics & Gynaecology*, Volume 83, 2022, Pages 36-45, ISSN 1521-6934,
<https://doi.org/10.1016/j.bpobgyn.2021.11.012>.
- [22] Wang, J., Yang, X., Cai, H. et al. Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning. *Sci Rep* 6, 27327 (2016). <https://doi.org/10.1038/srep27327>
- [23] Tagliafico AS, Piana M, Schenone D, Lai R, Massone AM, Houssami N. Overview of radiomics in breast cancer diagnosis and prognostication. *Breast*. 2020 Feb;49:74-80. doi: 10.1016/j.breast.2019.10.018. Epub 2019 Nov 6. PMID: 31739125; PMCID: PMC7375670.
- [24] Saxena S, Gyanchandani M. Machine Learning Methods for Computer-Aided Breast Cancer Diagnosis Using Histopathology: A Narrative Review. *J Med Imaging Radiat Sci*. 2020 Mar;51(1):182-193. doi: 10.1016/j.jmir.2019.11.001. Epub 2019 Dec 26. PMID: 31884065.
- [25] Stark GF, Hart GR, Nartowt BJ, Deng J. Predicting breast cancer risk using personal health data and machine learning models. *PLoS One*. 2019 Dec 27;14(12):e0226765. doi: 10.1371/journal.pone.0226765. PMID: 31881042; PMCID: PMC6934281.
- [26] Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. *Technol Health Care*. 2016;24(1):31-42. doi: 10.3233/THC-151071. PMID: 26409558.

- [27] Rabiei R, Ayyoubzadeh SM, Sohrabei S, Esmaeili M, Atashi A. Prediction of Breast Cancer using Machine Learning Approaches. *J Biomed Phys Eng*. 2022 Jun 1;12(3):297-308. doi: 10.31661/jbpe.v0i0.2109-1403. PMID: 35698545; PMCID: PMC9175124.
- [28] Chugh, G., Kumar, S. & Singh, N. Survey on Machine Learning and Deep Learning Applications in Breast Cancer Diagnosis. *Cogn Comput* 13, 1451–1470 (2021). <https://doi.org/10.1007/s12559-020-09813-6>
- [29] Mazo, C.; Kearns, C.; Mooney, C.; Gallagher, W.M. Clinical Decision Support Systems in Breast Cancer: A Systematic Review. *Cancers* **2020**, *12*, 369. <https://doi.org/10.3390/cancers12020369>
- [30] Rasool, A.; Bunternghit, C.; Tiejian, L.; Islam, M.R.; Qu, Q.; Jiang, Q. Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis. *Int. J. Environ. Res. Public Health* 2022, *19*, 3211. <https://doi.org/10.3390/ijerph19063211>
- [31] Islam, M.M., Haque, M.R., Iqbal, H. et al. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. *SN COMPUT. SCI.* 1, 290 (2020). <https://doi.org/10.1007/s42979-020-00305-w>
- [32] Massafra R., Latorre A., Fanizzi A., Bellotti R., Didonna V., Giotta F., La Forgia D., Nardone A., Pastena M., Ressa CM, Rinaldi L., Russo AOM, Tamborra P., Tangaro S., Zito A., Lorusso V., A Clinical Decision Support System for Predicting Invasive Breast Cancer Recurrence: Preliminary Results, *Frontiers in Oncology*, Volume 11, 2021, ISSN 2234-943X, doi: 10.3389/fonc.2021.576007
- [33] Vasey B, Ursprung S, Beddoe B, et al. Association of Clinician Diagnostic Performance With Machine Learning–Based Decision Support Systems: A Systematic Review. *JAMA Netw Open*. 2021;4(3):e211276. doi:10.1001/jamanetworkopen.2021.1276
- [34] Tiwari, Monika and Bharuka, Rashi and Shah, Praditi and Lokare, Reena, Breast Cancer Prediction Using Deep Learning and Machine Learning Techniques (March 22, 2020), <http://dx.doi.org/10.2139/ssrn.3558786>
- [35] Ioannis Sechopoulos, Jonas Teuwen, Ritse Mann, Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art, *Seminars in Cancer Biology*, Volume 72, 2021, Pages 214-225, ISSN 1044-579X, <https://doi.org/10.1016/j.semcancer.2020.06.002>.
- [36] Lahoura, V.; Singh, H.; Aggarwal, A.; Sharma, B.; Mohammed, M.A.; Damaševičius, R.; Kadry, S.; Cengiz, K. Cloud Computing-Based Framework for Breast Cancer Diagnosis Using Extreme Learning Machine. *Diagnostics* 2021, *11*, 241, <https://doi.org/10.3390/diagnostics11020241>

- [37] Sutong Wang, Yuyan Wang, Dujuan Wang, Yunqiang Yin, Yanzhang Wang, Yaochu Jin, An improved random forest-based rule extraction method for breast cancer diagnosis, *Applied Soft Computing*, Volume 86, 2020, 105941, ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2019.105941>.
- [38] Ruben Martin-Payo, Claudia Leirós-Díaz, Jo Armes, Mei Rosemary Fu, Judit Cachero-Rodríguez, María del Mar Fernandez-Alvarez, The “Breast-4Y” web app for breast cancer prevention at young age: Development, evaluation, and validation, *Internet Interventions*, Volume 33, 2023, 100651, ISSN 2214-7829, <https://doi.org/10.1016/j.invent.2023.100651>.
- [39] Nasution A, Yusuf A, Lean Keng S, Rasudin NS, P Iskandar YH, Ab Hadi IS. Development of Mobile App for Breast Examination Awareness Using Health Belief Model: A Qualitative Study. *Asian Pac J Cancer Prev*. 2021 Oct 1;22(10):3151-3163. doi: 10.31557/APJCP.2021.22.10.3151. PMID: 34710991; PMCID: PMC8858252
- [40] <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>
- [41] <https://archive.ics.uci.edu/dataset/451/breast+cancer+coimbra>

Συντμήσεις

ADABoost	Adaptive boosting
ANN	Artificial Neural Networks
BCCD	Breast Cancer Coimbra Dataset
CDSS	Clinical Decision Support Systems
CNN	Convolutional Neural Networks
GC-MS	Gas chromatography–mass spectrometry
ID	Identity
KNN	K-nearest neighbor
LC-MS	Liquid chromatography–mass spectrometry
LR	Linear Regression
MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
RFE	Recursive Feature Elimination
SVM	Support Vector Machines
WDBC	Wisconsin Diagnostic Breast Cancer
XGBoost	Extreme gradient boosting
ΕΟΠΥΥ	Εθνικός Οργανισμός Παροχής Υπηρεσιών Υγείας

Γλωσσάρι Ξενικών Όρων

Accuracy	Ακρίβεια
Adaptive Boosting	Προσαρμοστική ενίσχυση
Artificial Neural Networks	Τεχνητά νευρωνικά δίκτυα
Benign	Καλοήθης
Clinical Decision Support Systems	Συστήματα Υποστήριξης Κλινικών Αποφάσεων
Cloud Computing	Υπολογιστικό νέφος
Convolutional Neural Networks	Συνελικτικά νευρωνικά δίκτυα
Cross-Validation	Διασταυρούμενη επικύρωση
Cross-Validation Score	Βαθμός διασταυρούμενης επικύρωσης
Dataframe	Πλαίσιο δεδομένων
Decision Support Systems	Συστήματα Υποστήριξης Αποφάσεων
Decision Trees	Δέντρα αποφάσεων
Deep Learning	Βαθιά μάθηση
Extreme Gradient Boosting	Ακραία ενίσχυση κλίσης
F1-score	Βαθμολογία F1
K-Nearest Neighbor	K-κοντινότερος γείτονας
Linear Regression	Γραμμική παλινδρόμηση
Linear Support Vector Machines	Γραμμικές διανυσματικές μηχανές υποστήριξης
Machine Learning	Μηχανική μάθηση
Malignant	Κακοήθης

Negative	Αρνητικό
Polynomial Support Vector Machines	Πολυωνυμικές διανυσματικές μηχανές υποστήριξης
Positive	Θετικό
Precision	Ακρίβεια
Radiomics	Ραδιενέργεια
Random Forest	Τυχαία δάση
Recall	Ανάκληση
Recursive Feature Elimination	Αναδρομική εξάλειψη χαρακτηριστικών
Stratified K-fold Cross Validation	Στρωματοποιημένη διαδικασία διασταυρούμενης επικύρωσης
Support Vector Machines	Διανυσματικές μηχανές υποστήριξης
Train-Test Set	Σύνολο εκπαίδευσης-δοκιμής

Ευρετήριο

ADABoost, v, 42, 43, 53, 58

F1-score, vi, 22, 52, 53, 54, 55, 56, 57, 58, 59

Naïve Bayes, v, 28, 42, 44, 53, 54, 58, 71

Python, 28, 39, 40, 41, 42, 44, 50, 61, 62

random forest, 23, 29, 39, 40, 42, 43, 52, 57, 75

ακρίβεια, 28, 29, 30, 33, 38, 52, 53, 54, 55, 57, 58, 59

αναδρομική εξάλειψη χαρακτηριστικών, 42, 51

ανάκληση, 52, 53, 54, 55, 56, 57, 58, 59

αναλυτική δεδομένων υγείας, 13

ανάπτυξη εφαρμογών, 61

ανάπτυξη εφαρμογών πρόβλεψης, 36

ανάπτυξη λογισμικού για τη λήψη και υποστήριξη αποφάσεων, 33

ανταπόκριση αντικτύπου καρκίνου του μαστού, 3

αντίκτυπος του καρκίνου του μαστού, 3

αξιολόγηση προσυμπτωματικού ελέγχου καρκίνου του μαστού, 4

ανταπόκριση αντικτύπου καρκίνου του μαστού, 3

ρόλος προσυμπτωματικού ελέγχου στην πρόληψη, 4

αξιολόγηση προσυμπτωματικού ελέγχου καρκίνου του μαστού, 4

βαθιά μάθηση, 17

βαθμός διασταυρούμενης επικύρωσης, 51, 52, 53, 54, 56

βήματα, 39

βιοϊατρικές εφαρμογές, 33

δέντρα αποφάσεων, 45, 51, 56

διαδικτυακή χρήση, 63

διεπαφή, 62

εκπαίδευση μοντέλων μηχανικής μάθησης, 15

εξέταση μαστογραφίας, 10

εφαρμογές λήψης αποφάσεων, 33

ιατρική απεικόνιση, 4

ιστοπαθολογία, 15

κινητές συσκευές, 66

κλινική εξέταση, 10

κόστος, 10

κόστος είδους εξέτασης, 11

λήψη αποφάσεων, 33

λογισμικό, 33

μαστογραφία, 5

μαστογραφία αναφοράς, 9

μεθοδολογία, 41

μηχανική μάθηση, 14

μοντέλα μηχανικής μάθησης, 15

οικογενειακό ιστορικό, 9

οπτικοποίηση, 44

παράγοντες, 2

πλαίσιο δεδομένων, 41

πρόβλεψη, 44

προεπεξεργασία δεδομένων, 39

προσεγγίσεις μηχανικής μάθησης, 33

προσυμπτωματικός έλεγχος, 4

ρόλος προσυμπτωματικού ελέγχου στην πρόληψη, 4

σετ δεδομένων, 37

σημαντικότερες εργασίες, 17

στρωματοποιημένη διασταυρούμενη επικύρωση, 42, 43

συνολική ακρίβεια, 51, 52, 53, 54, 55, 57, 58

σύνολο εκπαίδευσης-δοκιμής, 39

συστήματα υποστήριξης κλινικών αποφάσεων, 27

ταξινομητές μηχανικής μάθησης, 16

τεχνητή νοημοσύνη, 14

τεχνικές εξόρυξης δεδομένων, 14

τεχνολογίες μηχανικής μάθησης, 14, 15

τεχνολογίες ωμικής, 13

τυχαία δάση, 52, 57

υπερηχογράφημα μαστών, 10

υποδειγματοληψία, 40, 43

υπολογιστικό νέφος, 33

ψηφιακή μαστογραφία, 10