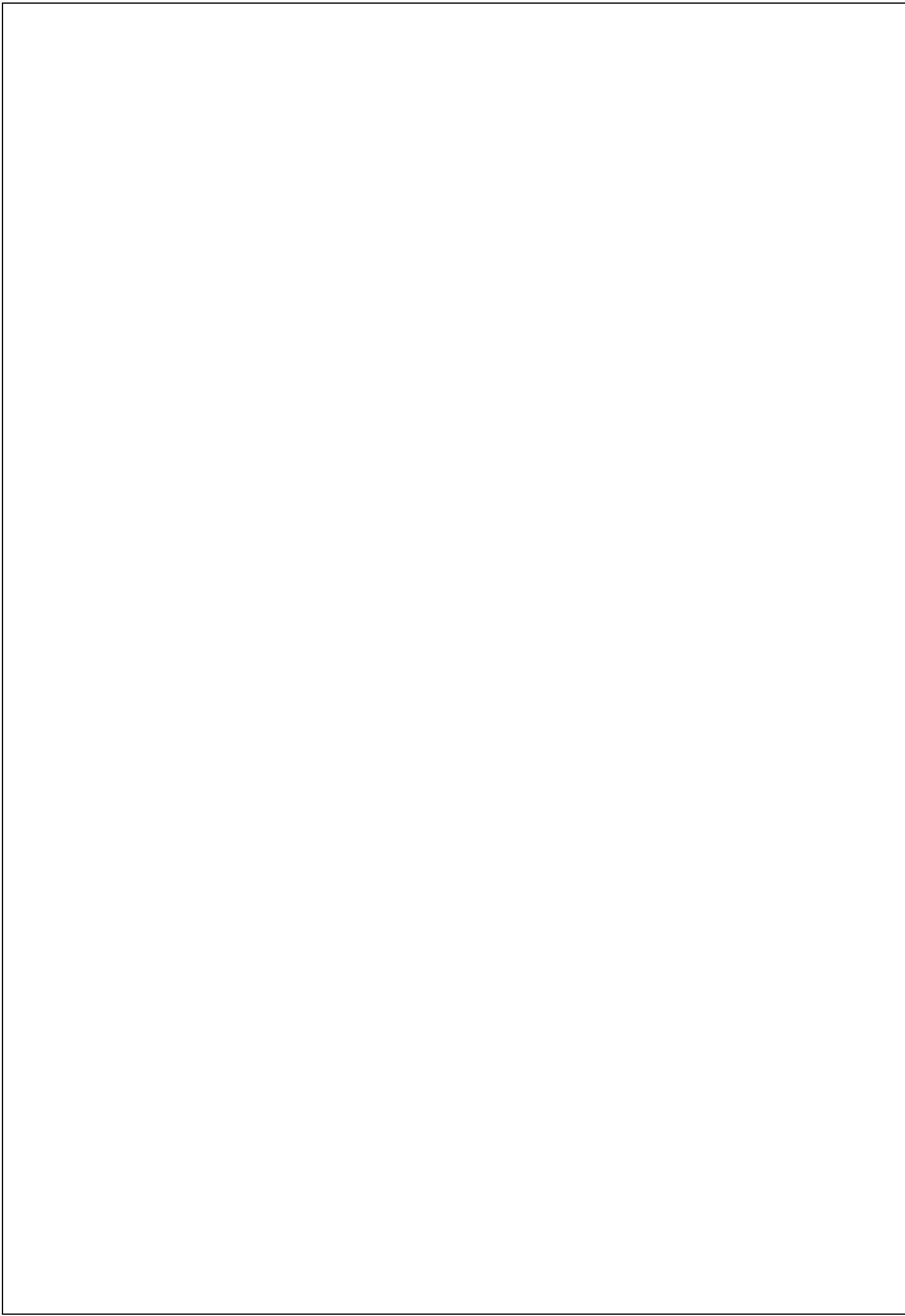




Ιόνιο Πανεπιστήμιο

Τμήμα Πληροφορικής



Πτυχιακή Εργασία

Ανάπτυξη συστημάτων υποστήριξης απόφασης για τη διάγνωση του καρκίνου του μαστού για το γενικό πληθυσμό και για τους γιατρούς

Φοιτητής
Τζιάλλας Ευθύμιος
Π2018015

Επιβλέπων Καθηγητής
Θεμιστοκλής Έξαρχος

Τριμελής Επιτροπή

Θεμιστοκλής Έξαρχος – Παναγιώτης Βλάμος – Αριστείδης Βραχάτης

Ιωάννινα 2023

Table of Contents

Εισαγωγή	6
Τεχνικές διάγνωσης	8
Τρόποι θεραπείας	10
Κεφάλαιο 1 ^ο	14
Κεφάλαιο 2 ^ο	24
Αλγόριθμοι.....	28
Αποτελέσματα.....	30
Ανάπτυξη εφαρμογής	33
Κεφάλαιο 3 ^ο	40
Αλγόριθμοι.....	43
Αποτελέσματα.....	43
Ανάπτυξη Ιστοσελίδας.....	46
Παραρτήματα	53
Παράρτημα 1 ^ο - Κώδικας εκπαίδευσης	53
Παράρτημα 2 ^ο - Κώδικας ανάπτυξης εφαρμογής.....	55
Παράρτημα 3 ^ο - Κώδικας εκπαίδευσης	59
Παράρτημα 4 ^ο - Κώδικας ανάπτυξης ιστοσελίδας	61
Συμπεράσματα.....	64
Αναφορές	65

Table of figures

Figure 1 - Breast-conserving surgery	10
Figure 2 - Total mastectomy	10
Figure 3 - Modified radical mastectomy	11
Figure 4 - Δείγματα υγρών	17
Figure 5 - WBCB Dataset.....	17
Figure 6 - BCCD Dataset.....	18
Figure 7 - Γράφημα Ακρίβειας.....	20
Figure 8 - SVM with training set.....	21
Figure 9 - SVM with testing set.....	21

Figure 10 - Χαρακτηριστικά του σετ δεδομένων 1	27
Figure 11 - Διάγραμμα αποτελεσμάτων Ακρίβειας 1.....	31
Figure 12 - Διάγραμμα αποτελεσμάτων Ακρίβειας 2.....	32
Figure 13 - Διαδικασία ανάπτυξης εφαρμογής 1	33
Figure 14 - Login Screen	34
Figure 15 - Αρχική Οθόνη	35
Figure 16 - Είσοδος Δεδομένων.....	35
Figure 17 - Εμφάνιση πρόβλεψης	36
Figure 18 - Εμφάνιση πρόβλεψης 2.....	36
Figure 19 - Πληροφορίες Γλυκόζης	37
Figure 20 - Πληροφορίες Ινσουλίνης	37
Figure 21 - Πληροφορίες Ρεζιστίνης.....	38
Figure 22 - Πληροφορίες MCP.1.....	39
Figure 23 - Χαρακτηριστικά του σετ δεδομένων 2.....	42
Figure 24 - Διάγραμμα αποτελεσμάτων Ακρίβειας 3.....	44
Figure 25 - Διάγραμμα αποτελεσμάτων Ακρίβειας 4.....	45
Figure 26 - Διαδικασία ανάπτυξης ιστοσελίδας.....	46
Figure 27 - Αρχική οθόνη	47
Figure 28 - Αρχική οθόνη 2	47
Figure 29 - Αρχική οθόνη 3	48
Figure 30 - Αρχική οθόνη 4	49
Figure 31 - Οθόνη Πρόβλεψης.....	50
Figure 32 - Μήνυμα λάθους.....	51
Figure 33 - Πρόβλεψη Malignant.....	51
Figure 34 - Πρόβλεψη Benign	52

List of Tables

Table 1 - Variables	15
Table 2 - Results	15
Table 3 - Results	16
Table 4 - Accuracy	18
Table 5 - AUC.....	19
Table 6 - Results	19
Table 7 - Results	20
Table 8 - Results	22
Table 9 - Classification Results	22

Table 10 - Results.....	23
Table 11 - 3-fold cross validation	23
Table 12 - Conventional Method.....	24
Table 13 - Ποσοτικά Χαρακτηριστικά 1.....	25
Table 14 - Labels 1	26
Table 15 - Αποτελέσματα	30
Table 16 - Labels 2.....	40
Table 17 - Ποσοτικά Χαρακτηριστικά 2.....	40
Table 18 - Αποτελέσματα	43

Εισαγωγή

Στον σύγχρονο κόσμο, ο καρκίνος του μαστού αναδεικνύεται ως μία από τις πιο συνηθισμένες και σοβαρές παθήσεις που αντιμετωπίζουν οι γυναίκες αλλά και οι άνδρες. Αν και οι επιστημονικές έρευνες και οι ιατρικές εξελίξεις έχουν καταλήξει σε σημαντικά επιτεύγματα όσον αφορά την θεραπεία, αυτή η νόσος παραμένει μία πρόκληση για την ιατρική κοινότητα και μία πηγή έντονης ανησυχίας για τους ασθενείς.

Είναι σημαντικό να επισημανθεί ότι η πρόληψη, η έγκαιρη διάγνωση και η αποτελεσματική θεραπεία παραμένουν κρίσιμοι παράγοντες για την αντιμετώπιση του καρκίνου του μαστού. Η ευαισθησία του κοινού και οι προληπτικές εξετάσεις, όπως η μαστογραφία, αποτελούν σημαντικά εργαλεία στην ανίχνευση πρόωρων σταδίων της νόσου. Παράλληλα, ο ρόλος της εκπαίδευσης και της ενημέρωσης είναι κρίσιμος. Οι γυναίκες και οι άνδρες πρέπει να γνωρίζουν τους παράγοντες κινδύνου, όπως η ηλικία, η οικογενειακή ιστορία καρκίνου του μαστού και άλλοι παράγοντες που μπορεί να συμβάλλουν στην εμφάνιση της νόσου. Η συνεχής έρευνα για νέες θεραπευτικές προσεγγίσεις, καθώς και η προώθηση της καινοτομίας στον τομέα της ιατρικής, αποτελούν βασικά κομμάτια της προσπάθειας για την καταπολέμηση αυτής της σοβαρής νόσου. Η ατομική προσέγγιση στη θεραπεία, που περιλαμβάνει τη γονιδιακή ανάλυση και την αξιολόγηση των χαρακτηριστικών του όγκου, επιτρέπει την προσαρμογή του προγράμματος θεραπείας σε κάθε ασθενή.

Ο καρκίνος του μαστού ξεκινά στον ιστό του μαστού όταν τα μεταλλαγμένα κύτταρα αναπτύσσονται ανεξέλεγκτα, δημιουργώντας τελικά μια μάζα ή όγκο. Τα κύτταρα στο ανθρώπινο σώμα έχουν έναν φυσιολογικό κυτταρικό κύκλο μέσω του οποίου αναπτύσσονται και διαιρούνται, δημιουργώντας αντίγραφα του εαυτού τους που απαιτείται για να αντικαταστήσουν τα παλιά ή μη φυσιολογικά κύτταρα. Λαμβάνουν σήματα για να αναπτυχθούν καθώς και για να πεθάνουν όταν καταστραφούν, αλλά τα καρκινικά κύτταρα δεν συμπεριφέρονται σαν υγιή κύτταρα. Αυτά τα ανώμαλα κύτταρα όχι μόνο επιβιώνουν, αλλά και διαιρούνται και πολλαπλασιάζονται – παρά το γεγονός ότι το σώμα δεν τα χρειάζεται. Αυτό δημιουργεί όλο και περισσότερα ανώμαλα κύτταρα, τα οποία στη συνέχεια σχηματίζουν όγκο ο οποίος μπορεί να είναι μη καρκινικός (καλοήθης) ή καρκινικός (κακοήθης). Οι μη καρκινικοί όγκοι αποτελούνται από κύτταρα που μοιάζουν πολύ με τα φυσιολογικά κύτταρα, αναπτύσσονται αργά και δεν εισβάλλουν σε κοντινούς ιστούς ούτε εξαπλώνονται σε άλλα μέρη του σώματος αντιθέτως οι καρκινικοί όγκοι, εάν αφεθούν ανεξέλεγκτοι, μπορούν τελικά να εξαπλωθούν πέρα από τον αρχικό όγκο σε άλλα μέρη του σώματος.

Ο καρκίνος του μαστού ξεκινά είτε στα κύτταρα των αδένων που παράγουν γάλα, που ονομάζονται λοβοί, είτε τις διόδους που αποστραγγίζουν το γάλα από τις λοβούς στη θηλή, που ονομάζονται αγωγοί. Λιγότερο συχνά, ο καρκίνος του μαστού μπορεί να ξεκινήσει τις λιπώδεις και ινώδεις συνδετικούς ιστούς του μαστού. Ανάλογα με το στάδιο του καρκίνου του μαστού, τα καρκινικά κύτταρα μπορούν να εισβάλουν στον κοντινό υγιή ιστό του μαστού και να εισέλθουν στους λεμφαδένες τις μασχάλης. Οι λεμφαδένες είναι μικρά όργανα που φιλτράρουν τις ξένες ουσίες στο σώμα. Εάν τα καρκινικά κύτταρα εισέλθουν στους λεμφαδένες, μπορούν να ταξιδέψουν μέσω του λεμφικού υγρού σε άλλα μέρη του σώματος.

Ο καρκίνος του μαστού προκαλείται πάντα από ένα λάθος στο γενετικό υλικό, που ονομάζεται γενετική ανωμαλία. Ωστόσο, μόνο το 5% έως το 10% των καρκίνων συνδέονται με γενετικές ανωμαλίες που κληρονομούνται από έναν γονέα. Περίπου το 85% των καρκίνων του μαστού προκαλούνται από γενετικές ανωμαλίες που είναι αποτέλεσμα της γήρανσης και της φθοράς της ζωής γενικότερα.

Σύμφωνα με την Αμερικανική Αντικαρκινική Εταιρεία, οποιαδήποτε από τις ακόλουθες ασυνήθιστες αλλαγές στο στήθος μπορεί να είναι σύμπτωμα καρκίνου του μαστού:

- πρήξιμο ολόκληρου ή μέρους του μαστού
- ερεθισμός του δέρματος ή λακκάκια
- πόνος στο στήθος
- πόνος στη θηλή ή στροφή της θηλής της προς τα μέσα
- ερυθρότητα, λέπια ή πάχυνση του δέρματος της θηλής ή του μαστού
- εκκρίσεις από τη θηλή εκτός από το μητρικό γάλα
- ένα εξόγκωμα στην περιοχή της μασχάλης

Αυτές οι αλλαγές μπορεί να είναι σημάδια λιγότερο σοβαρών καταστάσεων, μια μόλυνση ή μια κύστη. Αν και τα συμπτώματα του καρκίνου του μαστού ποικίλλουν ευρέως, πολλοί καρκίνοι του μαστού δεν έχουν καθόλου εμφανή συμπτώματα. Σε ορισμένες περιπτώσεις, ένα εξόγκωμα μπορεί να είναι πολύ μικρό για να γίνει αντιληπτό ή να προκαλέσει ασυνήθιστες αλλαγές. Συχνά, μια μη φυσιολογική περιοχή εμφανίζεται σε μια μαστογραφία προ συμπτωματικού ελέγχου (ακτινογραφία μαστού), η οποία οδηγεί σε πρόσθετες εξετάσεις.

Σε πολλές περιπτώσεις, το πρώτο σημάδι καρκίνου του μαστού είναι ένα νέο ογκίδιο ή μάζα στο στήθος που μπορεί να γίνει αντιληπτό είτε από τον ασθενή είτε από τον γιατρό. Ένα εξόγκωμα που είναι ανώδυνο, σκληρό και έχει ανομοιόμορφες άκρες είναι πιο πιθανό να είναι καρκίνος. Αλλά υπάρχουν και φορές που οι καρκίνοι μπορεί να είναι τρυφεροί, μαλακοί και στρογγυλεμένοι.

Η εκτέλεση μιας μηνιαίας αυτοεξέτασης μαστού είναι ο καλύτερος τρόπος για να παρατηρηθούν τυχόν αλλαγές. Άνθρωποι άνω των 40 ετών, θα πρέπει να προβαίνουν σε ετήσια μαστογραφία και φυσική εξέταση καθώς όσο πιο νωρίς εντοπιστεί και διαγνωστεί ο καρκίνος του μαστού, τόσο μεγαλύτερες είναι οι πιθανότητες για την επιτυχή θεραπεία του.

Τεχνικές διάγνωσης

Οι γιατροί συχνά χρησιμοποιούν εξετάσεις για να βρουν ή να διαγνώσουν τον καρκίνο του μαστού όπως:

Φυσική εξέταση και ιστορικό υγείας: Μια εξέταση του σώματος για τον έλεγχο των γενικών ενδείξεων υγείας, συμπεριλαμβανομένου του ελέγχου για σημεία ασθένειας, για εξογκώματα ή οτιδήποτε άλλο φαίνεται ασυνήθιστο. Έλεγχος ιστορικού των συνηθειών υγείας του ασθενούς και παλαιότερων ασθενειών και θεραπειών.

Κλινική εξέταση μαστού (CBE): Μια εξέταση του μαστού από γιατρό ή άλλο επαγγελματία υγείας. Ο γιατρός θα νιώσει και θα παρατηρήσει προσεκτικά τους μαστούς και κάτω από τα χέρια για εξογκώματα ή οτιδήποτε άλλο φαίνεται ασυνήθιστο.

Μαστογραφία: Η μαστογραφία είναι μια ειδική ιατρική εξέταση που χρησιμοποιείται για την απεικόνιση του ιστού του μαστού. Αποτελεί ένα σημαντικό εργαλείο στην πρόληψη και την έγκαιρη διάγνωση του καρκίνου του μαστού. Η διαδικασία περιλαμβάνει την προσπέλαση του μαστού και τη λήψη εικόνων ειδικά σχεδιασμένων για να απεικονίζουν τον ιστό του μαστού.

Η μαστογραφία χρησιμοποιείται κυρίως για τον έλεγχο των μαστών σε ανασφαλείς περιπτώσεις, τη διευκόλυνση της ανίχνευσης ύποπτων βλαβερών μορφών, και τον προσδιορισμό πιθανών κύστεων ή όγκων. Είναι ένα σημαντικό μέρος του προληπτικού ελέγχου για την ανίχνευση τυχόν προβλημάτων στον μαστό, ειδικά σε γυναίκες που έχουν υψηλό κίνδυνο εμφάνισης καρκίνου του μαστού λόγω οικογενειακής ιστορίας ή άλλων παραγόντων.

Εξέταση με υπερήχους: Μια διαδικασία κατά την οποία ηχητικά κύματα υψηλής ενέργειας (υπερηχογράφημα) αναπηδούν από εσωτερικούς ιστούς ή όργανα και κάνουν ηχώ. Η ηχώ σχηματίζει μια εικόνα των ιστών του σώματος που ονομάζεται υπερηχογράφημα. Η εικόνα μπορεί να εκτυπωθεί για να παρατηρηθεί αργότερα.

MRI (Magnetic Resonance Imaging): Μια διαδικασία που χρησιμοποιεί μαγνήτη, ραδιοκύματα και υπολογιστή για τη λήψη μιας σειράς λεπτομερών εικόνων και των δύο μαστών. Αυτή η διαδικασία ονομάζεται τις πυρηνική μαγνητική τομογραφία (NMRI).

Μελέτες χημείας αίματος: Μια διαδικασία κατά την οποία ένα δείγμα αίματος ελέγχεται για τη μέτρηση των ποσοτήτων ορισμένων ουσιών που απελευθερώνονται στο αίμα από όργανα και ιστούς του σώματος. Μια ασυνήθιστη, μεγαλύτερη ή χαμηλότερη από την κανονική, ποσότητα μιας ουσίας μπορεί να είναι σημάδι ασθένειας.

Βιοψία: Η αφαίρεση κυττάρων ή ιστών ώστε να μπορούν να προβληθούν στο μικροσκόπιο από έναν παθολόγο για να ελέγξει για σημάδια καρκίνου. Εάν εντοπιστεί ένα εξόγκωμα στο στήθος, μπορεί να γίνει βιοψία.

Υπάρχουν τέσσερις τύποι βιοψίας που χρησιμοποιούνται για τον έλεγχο του καρκίνου του μαστού:

- Εκτομή βιοψίας (Excisional biopsy): Η αφαίρεση ολόκληρου του όγκου ιστού.
- Βιοψία τομής (Incisional biopsy): Η αφαίρεση μέρους του όγκου ή του δείγματος ιστού.
- Βιοψία πυρήνα (Core biopsy): Η αφαίρεση ιστού με χρήση ευρείας βελόνας.
- Βιοψία με λεπτή βελόνα (Fine-needle aspiration (FNA) biopsy): Η αφαίρεση ιστού ή υγρού, χρησιμοποιώντας μια λεπτή βελόνα.

Εάν διαγνωστεί καρκίνος του μαστού, γίνονται εξετάσεις για να διαπιστωθεί εάν τα καρκινικά κύτταρα έχουν εξαπλωθεί εντός του μαστού ή σε άλλα μέρη του σώματος. Αυτή η διαδικασία ονομάζεται σταδιοποίηση (staging). Αν ο καρκίνος βρίσκεται μόνο στο στήθος, αν βρίσκεται στους λεμφαδένες κάτω από το μπράτσο ή έχει εξαπλωθεί εκτός του μαστού, καθορίζει το στάδιο του καρκίνου του μαστού. Ο τύπος και το στάδιο του καρκίνου του μαστού δείχνει στους γιατρούς τι είδους θεραπεία χρειάζεται.

Τρόποι θεραπείας

Υπάρχουν διάφοροι τύποι θεραπείας για ασθενείς με καρκίνο του μαστού, με τη χειρουργική επέμβαση να είναι η πιο συνηθισμένη. Οι τύποι χειρουργικής επέμβασης περιλαμβάνουν τα ακόλουθα:

Χειρουργική επέμβαση διατήρησης του μαστού (Breast-conserving surgery)

Αποτελεί μια επέμβαση για την αφαίρεση του καρκίνου και κάποιου φυσιολογικού ιστού γύρω από αυτόν, αλλά όχι του ίδιου του μαστού όπως φαίνεται και στην εικόνα **Figure 1 - Breast-conserving surgery**. Αυτός ο τύπος χειρουργικής επέμβασης μπορεί να ονομάζεται ογκεκτομή, μερική μαστεκτομή, τμηματική μαστεκτομή, τεταρτοτεκτομή ή χειρουργική επέμβαση διατήρησης του μαστού.

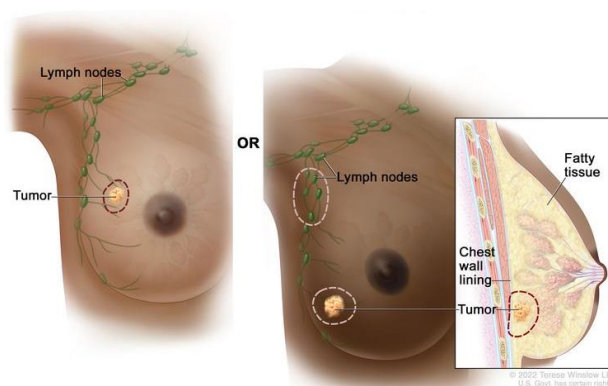


Figure 1 - Breast-conserving surgery

Ολική μαστεκτομή (Total mastectomy)

Χειρουργική επέμβαση για την αφαίρεση ολόκληρου του μαστού που έχει καρκίνο όπως φαίνεται και στην εικόνα **Figure 2 - Total mastectomy**. Αυτή η διαδικασία ονομάζεται απλή μαστεκτομή. Μερικοί από τους λεμφαδένες κάτω από τον βραχίονα μπορεί να αφαιρεθούν και να ελεγχθούν για καρκίνο. Αυτό μπορεί να γίνει ταυτόχρονα με τη χειρουργική επέμβαση μαστού ή μετά, μέσω ξεχωριστής τομής.

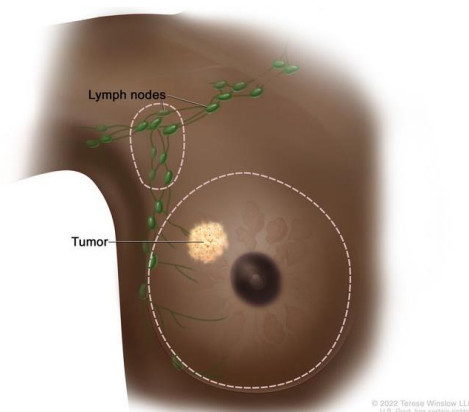


Figure 2 - Total mastectomy

Τροποποιημένη ριζική μαστεκτομή (Modified radical mastectomy)

Είναι χειρουργική επέμβαση για την αφαίρεση ολόκληρου του μαστού που έχει καρκίνο όπως φαίνεται και στην εικόνα **Figure 3 - Modified radical mastectomy****Error! Reference source not found..** Αυτό μπορεί να περιλαμβάνει αφαίρεση της θηλής, του σκουρόχρωμου δέρματος γύρω από τη θηλή και του δέρματος πάνω από το στήθος.

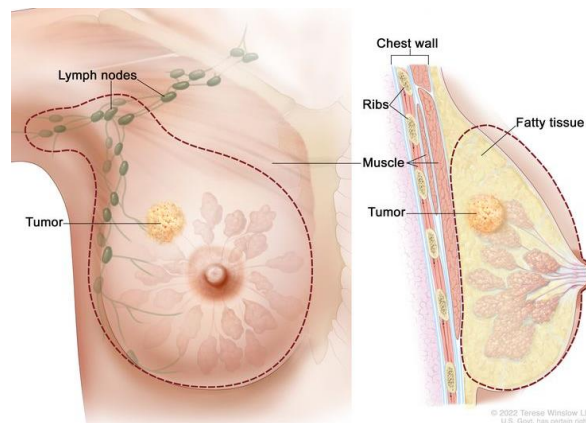


Figure 3 - Modified radical mastectomy

Μπορεί να χορηγηθεί χημειοθεραπεία πριν από τη χειρουργική επέμβαση για την αφαίρεση του όγκου. Όταν χορηγείται πριν από τη χειρουργική επέμβαση, η χημειοθεραπεία θα συρρικνώσει τον όγκο και θα μειώσει την ποσότητα του ιστού που πρέπει να αφαιρεθεί κατά τη διάρκεια της επέμβασης. Η θεραπεία που χορηγείται πριν από τη χειρουργική επέμβαση ονομάζεται προεγχειρητική θεραπεία (preoperative therapy) ή νεοεπικουρική θεραπεία (neoadjuvant therapy).

Αφού ο γιατρός αφαιρέσει όλο τον καρκίνο που μπορεί να παρατηρηθεί τη στιγμή της επέμβασης, σε ορισμένους ασθενείς μπορεί να χορηγηθεί ακτινοθεραπεία, χημειοθεραπεία, στοχευμένη θεραπεία (targeted therapy) ή ορμονοθεραπεία μετά την επέμβαση, για να σκοτωθούν τυχόν καρκινικά κύτταρα που έχουν απομείνει. Η θεραπεία που χορηγείται μετά το χειρουργείο, για τη μείωση του κινδύνου επανεμφάνισης, ονομάζεται μετεγχειρητική (postoperative therapy) θεραπεία ή επικουρική θεραπεία (adjuvant therapy).

Εάν ο ασθενής πρόκειται να υποβληθεί σε μαστεκτομή, μπορεί να εξεταστεί το ενδεχόμενο αποκατάστασης του μαστού (χειρουργική επέμβαση για την αποκατάσταση του σχήματος του μαστού μετά από μαστεκτομή). Η αποκατάσταση του μαστού μπορεί να γίνει κατά τη στιγμή της μαστεκτομής ή κάποια στιγμή μετά. Ο ανακατασκευασμένος μαστός μπορεί να γίνει με τον ιστό του ίδιου του ασθενούς (όχι του μαστού) ή με εμφυτεύματα γεμάτα με φυσιολογικό ορό ή σιλικόνη.

Ακτινοθεραπεία (Radiation therapy)

Η ακτινοθεραπεία είναι μια θεραπεία καρκίνου που χρησιμοποιεί ακτίνες X υψηλής ενέργειας ή τύπους ακτινοβολίας για να σκοτώσει τα καρκινικά κύτταρα ή να τα εμποδίσει να αναπτυχθούν. Υπάρχουν δύο τύποι ακτινοθεραπείας:

Η εξωτερική ακτινοθεραπεία, όπου χρησιμοποιεί ένα μηχάνημα έξω από το σώμα για να στείλει ακτινοβολία στην περιοχή του σώματος με καρκίνο και η εσωτερική ακτινοθεραπεία, που χρησιμοποιεί μια ραδιενεργή ουσία σφραγισμένη σε βελόνες, σπόρους, σύρματα ή καθετήρες που τοποθετούνται απευθείας μέσα ή κοντά στον καρκίνο.

Ο τρόπος που χορηγείται η ακτινοθεραπεία εξαρτάται από τον τύπο και το στάδιο του καρκίνου που αντιμετωπίζεται. Η εξωτερική ακτινοθεραπεία χρησιμοποιείται για τη θεραπεία του καρκίνου του μαστού. Η εσωτερική ακτινοθεραπεία χρησιμοποιείται για την ανακούφιση του πόνου των οστών που προκαλείται από καρκίνο του μαστού που έχει εξαπλωθεί στα οστά.

Χημειοθεραπεία (Chemotherapy)

Η χημειοθεραπεία χρησιμοποιεί φάρμακα για να σταματήσει την ανάπτυξη των καρκινικών κυττάρων, είτε σκοτώνοντας τα κύτταρα είτε εμποδίζοντας τη διαίρεση της. Η χημειοθεραπεία για τον καρκίνο του μαστού είναι συνήθως συστηματική, δηλαδή ενίεται σε φλέβα ή χορηγείται από το στόμα. Όταν χορηγούνται με αυτόν τον τρόπο, τα φάρμακα εισέρχονται στην κυκλοφορία του αίματος για να φτάσουν στα καρκινικά κύτταρα σε όλο το σώμα.

Ορμονική θεραπεία (Hormone therapy)

Η ορμονοθεραπεία, που ονομάζεται και ενδοκρινική θεραπεία, επιβραδύνει ή σταματά την ανάπτυξη των ορμονο-ευαίσθητων όγκων εμποδίζοντας την ικανότητα του σώματος να παράγει ορμόνες ή παρεμβαίνοντας στις επιδράσεις των ορμονών στα καρκινικά κύτταρα του μαστού. Ορισμένες ορμόνες μπορούν να προκαλέσουν την ανάπτυξη ορισμένων μορφών καρκίνου. Εάν οι εξετάσεις δείξουν ότι τα καρκινικά κύτταρα έχουν μέρη όπου μπορούν να προσκολληθούν οι ορμόνες (υποδοχείς), χρησιμοποιούνται φάρμακα, χειρουργική επέμβαση ή ακτινοθεραπεία για τη μείωση της παραγωγής ορμονών ή την παρεμπόδιση της λειτουργίας της.

Στοχευμένη θεραπεία (Targeted therapy)

Η στοχευμένη θεραπεία χρησιμοποιεί φάρμακα ή ουσίες για τον εντοπισμό και την επίθεση σε συγκεκριμένα καρκινικά κύτταρα. Ο γιατρός μπορεί να προτείνει τεστ βιοδεικτών για να βοηθήσει στην πρόβλεψη της ανταπόκρισής του ασθενούς σε ορισμένα φάρμακα στοχευμένης θεραπείας.

Ανοσοθεραπεία (Immunotherapy)

Η ανοσοθεραπεία βοηθά το ανοσοποιητικό σύστημα του ατόμου να καταπολεμήσει τον καρκίνο. Ο γιατρός μπορεί να προτείνει τεστ βιοδεικτών για να βοηθήσει στην πρόβλεψη της ανταπόκρισής του ασθενούς σε ορισμένα φάρμακα ανοσοθεραπείας.

Παγκοσμίως, ο γυναικείος καρκίνος του μαστού έχει πλέον ξεπεράσει τον καρκίνο του πνεύμονα ως ο πιο συχνά διαγνωσμένος καρκίνος. Υπολογίζεται ότι 2.261.419 νέες περιπτώσεις καρκίνου του μαστού διαγνώστηκαν σε γυναίκες σε όλο τον κόσμο το 2020.

Περισσότερες γυναίκες τις Ηνωμένες Πολιτείες διαγιγνώσκονται με καρκίνο του μαστού από οποιοδήποτε άλλο είδος καρκίνου, εκτός από τον καρκίνο του δέρματος. Η ασθένεια αντιπροσωπεύει 1 στους 3 νέους γυναικείους καρκίνους ετησίως. Το 2023, υπολογίζεται ότι 297.790 γυναίκες στις Ηνωμένες Πολιτείες θα διαγνωστούν με διηθητικό καρκίνο του μαστού (invasive), ένα στάδιο του καρκίνου του μαστού όπου τα καρκινικά κύτταρα επεκτείνονται εκτός των περιορισμένων περιοχών του μαστού και εισβάλλουν σε γύρω ιστούς, αιμοφόρα αγγεία ή/και λεμφαγγεία. Καθώς και 55.720 γυναίκες θα διαγνωστούν με μη διηθητικό (non-invasive) καρκίνο του μαστού, ένα στάδιο του καρκίνου του μαστού όπου τα καρκινικά κύτταρα παραμένουν εντός των περιορισμένων περιοχών, χωρίς να επεκτείνονται σε γύρω ιστάμενα.

Από τα μέσα τις δεκαετίας του 2000, ο διηθητικός καρκίνος του μαστού στις γυναίκες αυξήθηκε κατά περίπου μισό τοις εκατό κάθε χρόνο. Αυτό πιθανότατα προκλήθηκε από την αύξηση του υπερβολικού σωματικού βάρους στις γυναίκες συνολικά, καθώς και από την πτώση των ποσοστών γονιμότητας και την αύξηση της ηλικίας για τις πρώτες γεννήσεις. Όσον αφορά τον ανδρικό πληθυσμό, υπολογίζεται ότι 2.800 άνδρες στις Ηνωμένες Πολιτείες θα διαγνωστούν με διηθητικό καρκίνο του μαστού το 2023. Επί του παρόντος, υπάρχουν περισσότερες από 3,8 εκατομμύρια γυναίκες που ζουν με ή πέρα από τον καρκίνο του μαστού στις Ηνωμένες Πολιτείες και υπολογίζεται λόγω αυτού ότι 43.700 θάνατοι (43.170 γυναίκες και 530 άνδρες) θα συμβούν στις Ηνωμένες Πολιτείες το 2023. Παγκοσμίως, ο γυναικείος καρκίνος του μαστού είναι η Πέμπτη κύρια αιτία θανάτου.

Το 2020, υπολογίζεται ότι 684.996 γυναίκες σε όλο τον κόσμο πέθαναν από καρκίνο του μαστού. Ο καρκίνος του μαστού είναι η δεύτερη πιο κοινή αιτία θανάτου από καρκίνο στις γυναίκες μετά τον καρκίνο του πνεύμονα. Ωστόσο, ο αριθμός των γυναικών που πέθαναν από καρκίνο του μαστού μειώθηκε κατά 43% από το 2020 μετά την κορύφωση το 1989. Αυτό οφείλεται στην έγκαιρη ανίχνευση, στις βελτιώσεις στη θεραπεία και τη μεγαλύτερη ευαισθητοποίηση για τον καρκίνο του μαστού. Ως αποτέλεσμα, περισσότεροι από 460.000 θάνατοι από καρκίνο του μαστού είχαν αποτραπεί κατά τη διάρκεια της περιόδου. Αν και οι μαύρες γυναίκες διαγιγνώσκονται λιγότερο συχνά με καρκίνο του μαστού από τις λευκές γυναίκες, έχουν 40% υψηλότερο ποσοστό θνησιμότητας από τη νόσο.

Ο καρκίνος του μαστού συνήθως διαγιγνώσκεται σε γυναίκες μέσης και μεγαλύτερης ηλικίας. Η διάμεση ηλικία διάγνωσης είναι τα 63, αν και δεν είναι συχνό, ορισμένες γυναίκες που διαγιγνώσκονται με τη νόσο να είναι κάτω των 45 ετών. Τέλος, οι μαύρες γυναίκες έχουν περισσότερες πιθανότητες να αναπτύξουν καρκίνο του μαστού πριν από την ηλικία των 40 ετών σε σύγκριση με τις λευκές γυναίκες.

Κεφάλαιο 1^ο

ΑΙ στον καρκίνο του μαστού

Στο πρόσωπο της αυξανόμενης προκλητικότητας του καρκίνου του μαστού, η τεχνητή νοημοσύνη (ΑΙ) αναδεικνύεται ως ένα ισχυρό εργαλείο που ανοίγει νέες προοπτικές στην ιατρική κοινότητα. Ο συνδυασμός της αναλυτικής ικανότητας των υπολογιστών και της ευφύους επεξεργασίας δεδομένων με την εμπειρογνωμοσύνη των ιατρικών επιστημόνων, οδηγεί σε εξαιρετικά αποτελέσματα στον τομέα της καρκινοθεραπείας. Στον καρκίνο του μαστού, η εφαρμογή της Τεχνητής Νοημοσύνης αντιπροσωπεύει μια πρωτοποριακή προσέγγιση που επιτρέπει την ακριβή πρόβλεψη, την γρήγορη ανίχνευση και την εξατομικευμένη θεραπεία.

Η επιστημονική κοινότητα αναγνωρίζει την τεράστια σημασία της Τεχνητής Νοημοσύνης στη βελτίωση της ποιότητας της υγείας των ασθενών. Μέσω της ανάλυσης δεδομένων υψηλής πολυπλοκότητας, η Τεχνητή Νοημοσύνη επιτρέπει την ακριβή κατηγοριοποίηση των κρουσμάτων καρκίνου του μαστού και την πρόβλεψη της αντίδρασης σε διάφορες θεραπείες. Επιπλέον, διευκολύνει τη δημιουργία εξατομικευμένων προγραμμάτων θεραπείας, βασισμένων στις ειδικές ανάγκες κάθε ασθενούς.

Στην εποχή που η τεχνολογία συναντά την ιατρική επιστήμη, η εφαρμογή της Τεχνητής Νοημοσύνης στον καρκίνο του μαστού ανοίγει νέες προοπτικές για την πρόληψη, τη διάγνωση και τη θεραπεία αυτής της νόσου. Η συνεχής έρευνα και ανάπτυξη σε αυτόν τον τομέα αποτελεί κινητήριο δύναμη για τη μελλοντική πρόοδο της ιατρικής κοινότητας, προσφέροντας νέες ελπίδες και ενθουσιασμό για την υγεία και την ευημερία των ασθενών.

Σε αυτό το πλαίσιο, πραγματοποιήθηκε εκτενής έρευνα για τις προηγμένες τεχνικές μηχανικής μάθησης που μπορούν να εφαρμοστούν σε αυτό το πεδίο. Κατά τη διάρκεια αυτής της έρευνας, αναζητήθηκαν σχετικές πηγές και μελέτες για τον καρκίνο του μαστού, με σκοπό να κατανοηθεί η επιστημονική προσέγγιση στην πρόβλεψη του νοσήματος. Το αποτέλεσμα αυτής της έρευνας αναδεικνύει τη σημαντική συνεισφορά της τεχνητής νοημοσύνης στον αγώνα κατά του καρκίνου του μαστού και ανοίγει νέες προοπτικές για την βελτίωση της περίθαλψης και την αύξηση του ποσοστού επιβίωσης των ασθενών.

Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence

Ο κύριος στόχος αυτής της έρευνας είναι η σύγκριση των επιδόσεων τριών γνωστών αλγορίθμων στα δεδομένα μέσω της ευαισθησίας (sensitivity), της ιδιαιτερότητας (specificity) και της ακρίβειας (accuracy). Οι αλγόριθμοι: 1) Decision Tree (C4.5), 2) Support Vector Machine (SVM), 3) Artificial Neural Network (ANN). Το σύνολο δεδομένων περιείχε, όπως φαίνεται και στο **Table 1 - Variables**~~Error! Reference source not found.~~, 1189 εγγραφές, 22 μεταβλητές πρόβλεψης και μία μεταβλητή αποτελέσματος. Αντλήθηκε από το πρόγραμμα εγγεγραμμένων ασθενών του Ιρανικού Κέντρου για τον Καρκίνο του Μαστού (ICBC) από το 1997 έως το 2008.

Μετά τη διαδικασία της προ επεξεργασίας, έγινε αφαίρεση των ακατάλληλων περιπτώσεων από τα δεδομένα. Μέσω της διαδικασίας καθαρισμού δεδομένων και της εφαρμογής στρατηγικών προετοιμασίας, κατασκευάστηκε το συνολικό σύνολο δεδομένων. Συνολικά, αναλύθηκαν 547 περιπτώσεις, ενώ 642 εγγραφές αποκλείστηκαν λόγω απουσιάζουσων πληροφοριών

Table 1 - Variables

No	Variable Name	Definition
1.	Local Recurrence	Yes or No
2.	Age at Diagnosis	≤ 35, 35 to 44, 44-45, 55 ≥ years old
3.	Age at Menarche	≤ 12 to ≥ 12 years old
4.	Age at Menopause	≤ 50 to ≥ 50 years old
5.	Infertility	Yes or No
6.	Family History of Breast Cancer	Yes or No
7.	History of other Cancer (CA)	Yes or No
8.	Location	Upper outer Quadrant (UOQ), Upper inner Quadrant (UIQ), Lower outer Quadrant (LOQ), Lower inner Quadrant (LIQ), Central, Axilla, Upper half, Lateral half, Lower half
9.	Side	Left, Right, Bilateral
10.	Tumor Size	≤ 2cm to ≥ 5cm
11.	LN/Nexion	Lymph node involvement/number of removed nodes after surgery
12.	Metastasis	Bone, Liver, Lung, Brain, others
13.	NPositive	Number of Positive lymph node involvement
14.	B.Pathology	Results of Biopsy Pathology after
15.	Type of surgery	Mastectomy (Preservative or Bilateral)
16.	G (Grade)	1, 2 or 3
17.	Margin of Involvement	Free or ≥ 2cm
18.	Estrogen Receptor	Negative or Positive
19.	Progesterone Receptor	Negative or Positive
20.	Type of Chemotherapy	Adjuvant or Neoadjuvant
21.	Radiotherapy	Yes or No
22.	Hormone Therapy	Tamoxifen, Raloxifen, Femara, Aromasin or Megace
23.	Death	Realted to Breast Cancer or unrelated
24.	Her2	Negative or Positive

Variables used for recurrence modeling of breast cancer

Table 2 - Results

	Accuracy	Sensitivity	Specificity
Decision Tree (C4.5)	0.936	0.958	0.907
ANN (MLP)	0.947	0.956	0.928
SVM	0.957	0.971	0.945

Comparison of data mining models

Breast cancer detection by leveraging Machine Learning

Η έρευνα παρουσιάζει μια νέα μέθοδο για την ανίχνευση του καρκίνου του μαστού χρησιμοποιώντας τεχνικές Μηχανικής Μάθησης. Η προτεινόμενη μέθοδος απέδωσε εξαιρετικά ακριβή αποτελέσματα σε σύγκριση με τις υπάρχουσες μεθόδους. Οι συγγραφείς επέλεξαν να χρησιμοποιήσουν δημοφιλείς μεθόδους ανίχνευσης όπως τον Naïve Bayes Classifier, τον Support Vector Machine (SVM), τεχνικές Classifier, το Bi-clustering και τον

Ada Boost, καθώς και τον ταξινομητή R-CNN (συνελικτικά νευρωνικά δίκτυα) και τα αμφίδρομα επαναλαμβανόμενα νευρωνικά δίκτυα (HA-BiRNN).

Ωστόσο, οι υπάρχουσες μέθοδοι παράγουν εικόνες περιορισμένης ποιότητας και ενδέχεται να αντιμετωπίζουν ζητήματα απόδοσης. Συνεπώς, εισάγεται μια νέα μέθοδος, Deep Neural Network with Support Value (DNNS), προκειμένου να παράγονται εικόνες υψηλότερης ποιότητας και να διορθώνονται άλλες παράμετροι απόδοσης. Οι συγγραφείς προτείνουν έναν νέο αλγόριθμο μαζί με μαθηματικούς τύπους για την αξιολόγηση της απόδοσης και την εκτέλεση. Το σύνολο δεδομένων περιλαμβάνει 8009 δείγματα ιστοπαθολογικών εικόνων πάνω από 683 ασθενών σε διάφορα επίπεδα μεγέθυνσης. Το δεδομένο σύνολο διαθέτει μια ποικιλία εικόνων ιστοπαθολογίας, οι οποίες ταξινομούνται ως καλοήθεις και κακοήθεις όγκοι.

Η προτεινόμενη μεθοδολογία DNNS περιλαμβάνει τρεις φάσεις. Στη φάση της προεπεξεργασίας, οι κυτταρολογικές εικόνες εισόδου υποβάλλονται σε επεξεργασία για την αφαίρεση θορύβου, χρησιμοποιώντας αποτελεσματική τεχνική φιλτραρίσματος Gauss. Στη δεύτερη φάση, εξάγονται εντροπία, γεωμετρικά και υφικά χαρακτηριστικά από τις προεπεξεργασμένες εικόνες. Στην τρίτη φάση, το σύστημα τμηματοποιεί τον όγκο του μαστού από τις εξαγόμενες εικόνες, χρησιμοποιώντας ασαφή ομαδοποίηση βασισμένη στο ιστο-σιγμοειδές.

Τα πειραματικά αποτελέσματα όπως φαίνεται και στο **Table 3 - Results**, απέδειξαν ότι το προτεινόμενο DNNS είναι αρκετά ανώτερο από τις υπάρχουσες μεθόδους, πετυχαίνοντας Accuracy 97.21%, Precision 97.9% και Recall 97.01%. Επιβεβαιώνεται ότι ο προτεινόμενος αλγόριθμος είναι πλεονεκτικός τόσο στην απόδοση, την αποτελεσματικότητα όσο και στην ποιότητα των εικόνων, οι οποίες αποτελούν ζωτικό στοιχείο στα πιο πρόσφατα ιατρικά συστήματα.

Table 3 - Results

Methodology	Accuracy	Precision	Recall
Naïve Bayes classifier	95.61	95.65	93.61
SVM classifier	95.61	95.65	93.61
Bi-clustering and Ada boost techniques	95.75	95.72	96.26
RCNN classifier	91.3	91.3	89.3
Bidirectional Recurrent Neural Networks (HA-BiRNN)	82.50	80.09	79.03
Proposed Methodology - Deep Neural Network with Support Value (DNNS)	97.21	97.9	97.01

Analysis of the Wisconsin Breast Cancer Dataset and Machine Learning for Breast Cancer Detection

Η έρευνα συζητά μια τεχνική διάγνωσης που χρησιμοποιεί το FNA (Fine Needle Aspiration) με υπολογιστική ερμηνεία μέσω μηχανικής μάθησης και στοχεύει στη δημιουργία ενός classifier που παρέχει υψηλό επίπεδο ακρίβειας.

Το Dataset που χρησιμοποιήθηκε είναι δημόσια διαθέσιμο και δημιουργήθηκε από τον Dr. William H. Wolberg, γιατρό στο University of Wisconsin Hospital στο Madison, Wisconsin, ΗΠΑ. Για να δημιουργήσει το σύνολο δεδομένων, ο Δρ. Wolberg χρησιμοποίησε δείγματα υγρών, που ελήφθησαν από ασθενείς με συμπαγείς μάζες μαστού και ένα γραφικό πρόγραμμα

υπολογιστή που ονομάζεται Xcyt. Τα δεδομένα όπως φαίνεται και στο **Figure 4 - Δείγματα υγρών**, μπορούν να θεωρηθούν «noise-free» και έχουν 16 τιμές που λείπουν.

Το πρόγραμμα χρησιμοποιεί έναν αλγόριθμο προσαρμογής καμπύλης, για να υπολογίσει δέκα χαρακτηριστικά από κάθε ένα από τα κύτταρα στο δείγμα. Κάθε χαρακτηριστικό αξιολογείται σε μια κλίμακα από το 1 έως το 10, με το 1 να είναι το πλησιέστερο σε καλοήγη και 10 το πλησιέστερο σε κακοήγη.



Figure 4 - Δείγματα υγρών

Για την προεπεξεργασία των δεδομένων, χρησιμοποιώντας τα διαθέσιμα εργαλεία στο Weka 3.6. Οι classifiers που χρησιμοποιήθηκαν είναι ο «Bayesian Networks» και ο «J48». Ο πρώτος αλγόριθμος, Bayesian Networks, έδειξε καλή απόδοση με ακρίβεια 97,80%. Ο J48, είχε ως αποτέλεσμα ακρίβειας 96,05%. Η έρευνα κατέληξε με τα Bayesian Networks, με μια ακρίβεια ελαφρώς υψηλότερη.

Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction

Σε αυτή τη μελέτη χρησιμοποιήθηκαν πέντε διαφορετικά μοντέλα ταξινόμησης συμπεριλαμβανομένων των Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Neural Network (NN) και Logistics Regression (LR) για την ταξινόμηση δύο διαφορετικών συνόλων δεδομένων που σχετίζονται με τον καρκίνο του μαστού στο σύνολο δεδομένων για τον καρκίνο του μαστού Coimbra (BCCD) και στη βάση δεδομένων για τον καρκίνο του μαστού στο Wisconsin (WBCD).

Τα δεδομένα WBCD περιλαμβάνουν 699 περιπτώσεις, και 10 χαρακτηριστικά για κάθε περίπτωση ενώ τα δεδομένα BCCD περιέχουν 116 περιπτώσεις και υπάρχουν 10 προγνωστικοί παράγοντες, όλοι ποσοτικοί και μια δυαδική εξαρτημένη μεταβλητή, που υποδηλώνει την παρουσία ή την απουσία καρκίνου του μαστού. Ο στόχος αυτής της μελέτη είναι να βρεθεί το μοντέλο με την υψηλότερη ακρίβεια.

Σύμφωνα με τα αποτελέσματα των πειραμάτων, όπως φαίνεται και στο **Figure 5 - WBCB Dataset** και στο **Figure 6 - BCCD Dataset**, ο RF (Random Forest) αποκτά την υψηλότερη ακρίβεια και F-measure στα δύο σύνολα δεδομένων, που υποδηλώνει την υπεροχή του σε άλλες μεθόδους. Ως εκ τούτου, επιλέγεται το RF ως κύριο αναλυτικό μοντέλο.

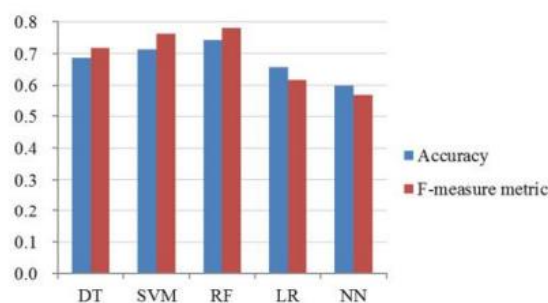


Figure 5 - WBCB Dataset

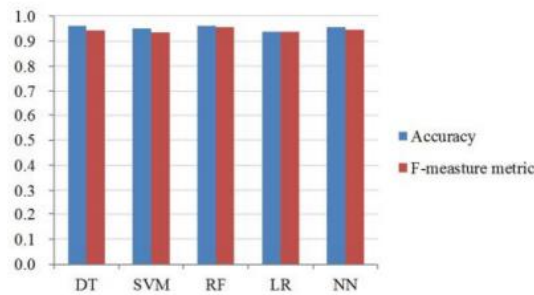


Figure 6 - BCCD Dataset

Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis

Ο κύριος στόχος της έρευνας είναι ο εντοπισμός του αποτελεσματικότερου και προγνωστικότερου αλγόριθμου για την ανίχνευση καρκίνου του μαστού. Χρησιμοποιήθηκαν οι εξής classifiers: Vector Machine (SVM), Random Forests, Logistic Regression, Decision tree (C4.5), K-Nearest Neighbors (KNN). Τα αποτελέσματα που προέκυψαν καθορίζουν ποιο μοντέλο παρέχει μεγαλύτερη ακρίβεια.

Τα δεδομένα που χρησιμοποιήθηκαν αντλήθηκαν από τα Νοσοκομεία του Πανεπιστημίου του Wisconsin της βάσης δεδομένων για τον καρκίνο του μαστού Madison. Το σύνολο δεδομένων έχει 569 περιπτώσεις (Καλοήθεις: 357 - Κακοήγη: 212), 2 κατηγορίες (62,74% καλοήγη και 37,26% κακοήγη) και 11 χαρακτηριστικά με ακέραια τιμή.

Το σύνολο δεδομένων υποβλήθηκε υπό επεξεργασία, η οποία περιλαμβάνει τέσσερα στάδια. Καθαρισμό, επιλογή χαρακτηριστικών, ορισμό ρόλου στόχου και εξαγωγή χαρακτηριστικών.

Για την κατασκευή του αλγόριθμου εκμάθησης χρησιμοποιήθηκε το 75% από το σύνολο δεδομένων ενώ το υπόλοιπο 25% χρησιμοποιήθηκε για την αξιολόγηση του μοντέλου. Όλα τα πειράματα στους αλγόριθμους μηχανικής μάθησης που περιγράφηκαν σε αυτήν την εργασία διεξήχθησαν χρησιμοποιώντας τη βιβλιοθήκη εκμάθησης Scikit και τη γλώσσα προγραμματισμού Python. Χρησιμοποιήθηκε Confusion Matrix, Accuracy, Precision, Sensitivity, F1 Score και AUC ως μετρήσεις απόδοσης για την αξιολόγηση και σύγκριση των μοντέλων.

Μετά από μια ακριβή σύγκριση μεταξύ των μοντέλων όπως φαίνεται και στο **Table 4 - AccuracyError! Reference source not found.**, βρέθηκε πως το Support Vector Machine πέτυχε υψηλότερη απόδοση με Ακρίβεια 97,2%, όπως επίσης και στο **Table 5 - AUCError! Reference source not found.**, με AUC 96,6%.

Table 4 - Accuracy

Algorithms	Accuracy Training Set (%)	Accuracy Testing Set (%)
SVM	98.4%	97.2%
Radom Forest	99.8%	96.5%
Logistic Regression	95.5%	95.8%
Decision Tree	98.8%	95.1%
K-NN	94.6%	93.7%

Table 5 - AUC

Algorithms	AUC (%)
SVM	0.966
Random Forests	0.960
Logistic Regression	0.947
Decision Tree	0.945
K-NN	0.952

Breast Cancer Prediction Using Machine Learning

Ο κύριος στόχος της έρευνας είναι η σύγκριση διαφόρων είδη αλγορίθμων ταξινόμησης όπως ο k Nearest Neighbors, Support Vector Machine, Logistic Regression και Gaussian Naive Bayes όσον αφορά τα: accuracy, precision, recall, f1-Score, και Jaccard index με σκοπό την εύρεση του ακριβότερου αλγορίθμου για την ανίχνευση του καρκίνου του μαστού.

Το σύνολο δεδομένων είναι ανοιχτά προσβάσιμο από το UCI Machine Learning αποθετήριο και αποτελείται από μερικά εκατοντάδες αρχεία δοκιμών ανθρώπινων κυττάρων καθένα από τα οποία περιέχει τις εκτιμήσεις ποιότητάς τους. Το σύνολο δεδομένων που έχει 10 χαρακτηριστικά. Οι τιμές αξιολογούνται από 1 έως 10, με 1 είναι το πιο κοντινό στον καλοήγη.

Σύμφωνα με τα αποτελέσματα της πρόβλεψης που φαίνονται στο **Table 6 - Results**, ο k Nearest Neighbors έχει την καλύτερη ακρίβεια για το σύνολο δεδομένων. Αυτό δείχνει πως ο k Nearest Neighbors, όπως φαίνεται και στο **Figure 7 - Γράφημα Ακρίβειας**, είναι καλύτερος για την πρόβλεψη του καρκίνου του μαστού σε σύγκριση με το Support Vector Machine, Logistic Regression, και Naive Bayes.

Table 6 - Results

Algorithms	Accuracy	Precision	Recall	F1-Score	Jaccard
kNN	0.99	0.98	0.99	0.98	0.98
SVM	0.96	0.95	0.97	0.96	0.96
LR	0.97	0.96	0.98	0.97	0.97
NB	0.95	0.94	0.96	0.95	0.94

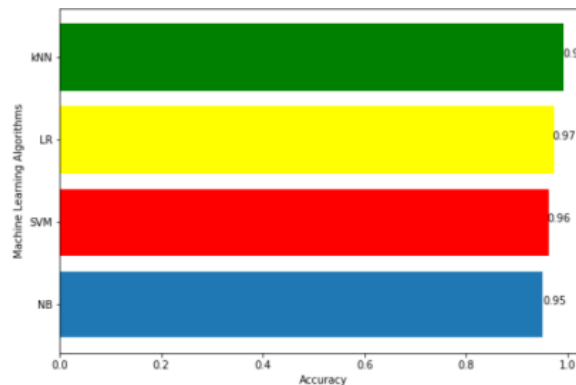


Figure 7 - Γράφημα Ακρίβειας

Prediction of Breast Cancer Using Supervised Machine Learning Techniques

Στόχος της συγκεκριμένης έρευνας είναι η μελέτη σχετικά με την υλοποίηση μοντέλων με χρήση Logistic Regression, Support Vector Machine (SVM) και K Nearest Neighbor (KNN), και σκοπό την εύρεση και την βαθμολογία των αποτελεσμάτων στο θέμα της ακρίβειας (accuracy, precision), ευαισθησία (sensitivity), ειδικότητα (specificity) και False Positive.

Οι δοκιμές εφαρμόστηκαν σε δύο σύνολα δεδομένων που αντλήθηκαν από το αποθετήριο UCI. Ένα σύνολο δεδομένων αυτών χρησιμοποιείται για ταυτοποίηση της νόσου όπως φαίνεται στο **Error! Reference source not found.** και ένα χρησιμοποιείται για την πρόβλεψη της όπως φαίνεται στο **Figure 9 - SVM with testing set.**

Τα αποτελέσματα, όπως φαίνεται στο **Table 7 - Results,Error! Reference source not found.** έδειξαν πως ο αλγόριθμος SVM (Support Vector Machine) δίνει την υψηλότερη ακρίβεια (92,7%) σε σύγκριση με τους άλλους δύο αλγορίθμους. Έτσι, προτείνεται ότι ο SVM είναι ο καταλληλότερος αλγόριθμος για την πρόβλεψη εμφάνισης καρκίνου του μαστού.

Table 7 - Results

Algorithm	Accuracy	Precision	Sensitivity	SP	FPR
Logistic Regression	92.10	95.31	91.0	93.61	6.3
K Nearest Neighbor	92.23	96.55	90.32	95.12	4.88
Support Vector Machine	92.78	95.94	91.07	95.14	4.86

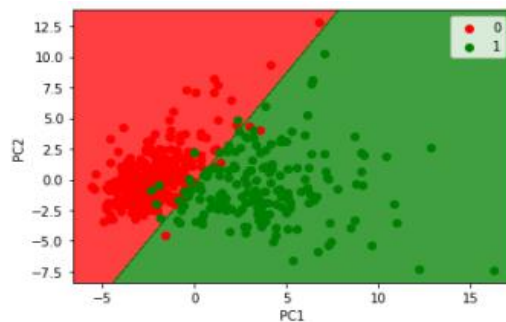


Figure 8 - SVM with training set

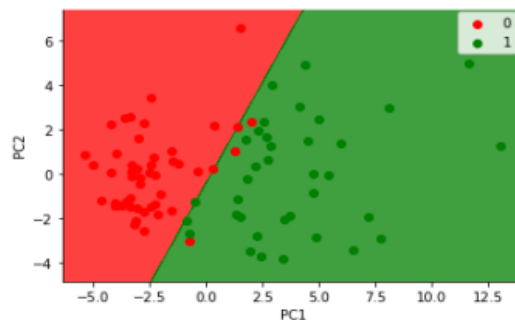


Figure 9 - SVM with testing set

Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis

Κύριος στόχος της έρευνας είναι η σύγκριση μεταξύ της απόδοσης τεσσάρων αλγορίθμων: Support Vector Machine (SVM) , Naive Bayes (NB), Decision tree (C4.5) και K Nearest Neighbor (KNN), όσον αφορά την ακρίβεια (accuracy, precision), ευαισθησία (sensitivity), ειδικότητα (specificity).

Τα σύνολα δεδομένων που χρησιμοποιήθηκαν στην έρευνα (Wisconsin Breast Cancer) αντλήθηκαν από το αποθετήριο μηχανικής μάθησης UCI . Περιέχουν 699 περιπτώσεις (καλοήθειες: 458 κακοήθειες: 241), 2 κατηγορίες (65,5% κακοήθειες και 34,5% καλοήθεις) και 11 χαρακτηριστικά με ακέραια τιμή.

Όλα τα πειράματα διεξήχθησαν χρησιμοποιώντας βιβλιοθήκες από το περιβάλλον μηχανικής μάθησης WEKA. Για την αξιολόγηση εφαρμόστηκε η τεχνική 10-fold cross validation

Τα αποτελέσματα των πειραμάτων όπως φαίνεται στο **Table 8 - Results, Error! Reference source not found.** έδειξαν ότι ο SVM επιτυγχάνει την υψηλότερη ακρίβεια (97,13%) με το χαμηλότερο ποσοστό σφάλματος (0,02%) σε αντίθεση με τον C4.5, τον Naïve Bayes και τον k-NN που έχουν ακρίβεια που κυμαίνεται μεταξύ 95,12 % και 95,28 % και ποσοστό σφάλματος που κυμαίνεται μεταξύ 0,03% και 0,06%.

Συμπερασματικά, ο SVM επιτυγχάνει την καλύτερη απόδοση όσον αφορά ακρίβεια, το χαμηλό ποσοστό σφάλματος και τη σωστή ταξινόμηση, όπως φαίνεται και στο **Table 9 - Classification Results**, στην πρόβλεψη του καρκίνου του μαστού.

Table 8 - Results

	TP	FP	Precision	Recall	F-Measure	Class
C4.5	0.95	0.05	0.96	0.95	0.96	Benign
	0.94	0.04	0.91	0.94	0.93	Malignant
SVM	0.97	0.03	0.98	0.97	0.97	Benign
	0.96	0.02	0.95	0.96	0.95	Malignant
NB	0.95	0.02	0.98	0.95	0.96	Benign
	0.97	0.04	0.91	0.97	0.94	Malignant
k-NN	0.97	0.08	0.95	0.97	0.96	Benign
	0.91	0.02	0.94	0.91	0.93	Malignant

Table 9 - Classification Results

	Benign	Malignant	class
C4.5	438	20	Benign
	14	227	Malignant
SVM	446	12	Benign
	9	232	Malignant
NB	436	22	Benign
	6	235	Malignant
k-NN	445	13	Benign
	20	221	Malignant

Breast Cancer Prediction using Machine Learning

Κύριος στόχος της έρευνας είναι η αξιολόγηση της ορθότητας στην ταξινόμηση των δεδομένων με βάση την ακρίβεια και την κατανάλωση χρόνου που χρησιμοποιείται για τον προσδιορισμό της απόδοσης των αλγορίθμων μεταξύ των SVM, Logistic Regression, Naïve Bayes και Random Forest.

Το αποθετήριο μηχανικής μάθησης UCI αποτελείται από το «The Wisconsin Breast Cancer» σύνολο δεδομένων που χρησιμοποιείται σε αυτή τη μελέτη. Υπάρχουν 699 περιπτώσεις στις οποίες 458 είναι καλοήθειες και 241 κακοήθειες. Επιπλέον, υπάρχουν δύο κατηγορίες κακοήθων που συνεισφέρουν στο 65,5% των δεδομένων και καλοήθειες 34,5%. Το σύνολο

δεδομένων για τον καρκίνο του μαστού λαμβάνεται σε μορφή csv από τη βάση δεδομένων τους.

Πραγματοποιήθηκε Data Visualization μέσω κάποιων βιβλιοθηκών της γλώσσας προγραμματισμού Python (Matplotlib και Seaborn). Από τον πίνακα παρατηρείται πως ο αλγόριθμος Random Forest απαιτεί 0,02 δευτερόλεπτα για τη δημιουργία μοντέλου σε αντίθεση με το Naïve Bayes που απαιτεί μόλις 0,01 δευτερόλεπτα, τον Logistic Regression με 0.02 και τέλος τον SVM που χρειάζεται είναι 0,03 δευτερόλεπτα. Η ακρίβεια του SVM είναι 97,41% που είναι υψηλότερη από του Naïve Bayes με ποσοστό ακρίβειας 94.83% αλλά αρκετά μικρότερη από του Logistic Regression και του Random Forest. Έτσι τα αποτελέσματα, όπως φαίνεται και στο **Table 10 - Results****Error! Reference source not found.**, έβγαλαν την απόδοση του Random Forest (99,76%) ως καλύτερη όταν πρόκειται για την ακρίβεια.

Table 10 - Results

Algorithm	Accuracy	Time taken
LR	99.06%	0.02s
SVM	98.59%	0.03s
NB	94.83%	0.01s
Random Forest	99.76%	0.02s

A Comparative Study of Machine Learning Algorithms applied to Predictive Breast Cancer Data

Κύριος στόχος της έρευνας είναι η εύρεση του καλύτερου μοντέλου μηχανικής μάθησης για τη διάγνωση του καρκίνου του μαστού. Τα δεδομένα για τον καρκίνο του μαστού που χρησιμοποιήθηκαν για αυτήν την εργασία ελήφθησαν από το σύνολο δεδομένων Wisconsin Breast Cancer Diagnostic από το UCI Machine Learning Repository.

Τα δεδομένα για τον καρκίνο του μαστού περιλαμβάνουν 569 περιπτώσεις βιοψιών καρκίνου. Κάθε εγγραφή έχει 32 χαρακτηριστικά. Ένα χαρακτηριστικό είναι ένας αριθμός αναγνώρισης, ένα άλλο είναι η διάγνωση του καρκίνου και 30 είναι εργαστηριακές μετρήσεις με αριθμητική αξία. Η διάγνωση κωδικοποιείται ως «M» για να υποδηλώνει κακοήγη ή «B» για να υποδηλώνει καλοήγη.

Έπειτα από μια προεργασία των δεδομένων, όπως η εξαίρεση του πρώτου χαρακτηριστικού ως μοναδικός αριθμός για κάθε ασθενή και η μετατροπή των χαρακτηριστικών διάγνωσης - "B" και "M" σε αριθμητικές τιμές 0 και 1 αντίστοιχα, έγιναν τα πειράματα με τη χρήση των Naïve Bayes, Artificial Neural Networks (ANN) και k-nearest neighbour (kNN).

Σε αυτή τη μελέτη, χρησιμοποιήθηκαν δύο μέθοδοι επαλήθευσης αποτελεσμάτων: η μέθοδος με ένα σετ εκπαίδευσης και ένα σετ δοκιμών (Conventional Method) και 3 - fold cross validation.

Τα αποτελέσματα δείχνουν ότι η ακρίβεια ταξινόμησης του ANN είναι υψηλότερη από αυτή των αλγορίθμων ταξινόμησης kNN και NB στη διάγνωση των προγνωστικών δεδομένων καρκίνου του μαστού από τη βάση δεδομένων UCI Wisconsin Breast Cancer. Ο αλγόριθμος ANN είναι ο πιο ακριβής με ακρίβεια 97,4% σε 3-fold cross validation, όπως φαίνεται και στο **Table 11 - 3-fold cross validation****Error! Reference source not found.**, και με ακρίβεια 100% στο Conventional Method, όπως φαίνεται στο **Table 12 - Conventional Method****Error! Reference source not found.** Ο kNN έρχεται δεύτερος και η ακρίβεια ταξινόμησης του ταξινομητή NB είναι η ελάχιστη και στις δύο περιπτώσεις.

Table 11 - 3-fold cross validation

Method	1 st Fold % accuracy	2 nd Fold % accuracy	3 rd Fold % accuracy	Averages % accuracy
k-Nearest Neighbors	96.3	96.8	94.8	96.0
Artificial Neural Networks	98.4	97.4	96.3	97.4
Naïve Bayes	93.4	92.4	95.8	93.9

Table 12 - Conventional Method

<i>Method</i>	<i>Accuracy (%)</i>
k-Nearest Neighbors	95.6
Artificial Neural Networks	100
Naïve Bayes	93.8

Κεφάλαιο 2^ο

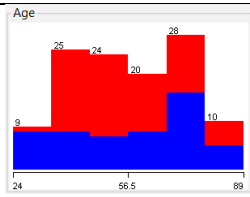
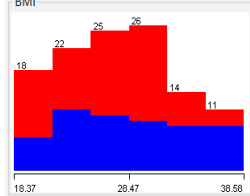
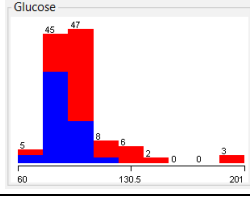
Dataset Breast Cancer Coimbra

Από το παρόν σημείο και εμπρός, προβαίνω σε μια εκτενή ανάλυση της αναπτυγμένης εφαρμογής που έχει δημιουργηθεί για την προηγμένη πρόβλεψη και διαχείριση του καρκίνου του μαστού. Στο πλαίσιο αυτό, επιδιώκεται η λεπτομερής παρουσίαση των λειτουργιών και των δυνατοτήτων της εν λόγω εφαρμογής, με έμφαση στη συμβολή της προς την προαγωγή της υγείας και την αποτελεσματική αντιμετώπιση του καρκίνου του μαστού.

Τα σετ δεδομένων τα οποία χρησιμοποιήθηκαν στα πειράματα επιλέχθηκαν έπειτα από συνεννόηση με τον καθηγητή. Το σύνολο δεδομένων αυτό εξετάζει τους προγνωστικούς παράγοντες, συγκεντρώνοντας ανθρωπομετρικά δεδομένα και παραμέτρους που ανακτώνται μέσω αναλύσεων ρουτίνας αίματος. Αυτά τα δεδομένα αναδεικνύουν τη σύνθεση του οργανισμού και αποκαλύπτουν φυσιολογικές διαταραχές, παρέχοντάς μας μια πολύτιμη εισαγωγή στην υγεία του ατόμου.

Το σετ δεδομένων περιέχει 116 περιπτώσεις και υπάρχουν 10 προγνωστικοί παράγοντες, όλοι ποσοτικοί, όπως φαίνεται και στο **Table 13 - Ποσοτικά Χαρακτηριστικά 1** και μια δυαδική εξαρτημένη μεταβλητή, που υποδηλώνει την παρουσία ή την απουσία καρκίνου του μαστού, όπως φαίνεται και στο **Table 14 - Labels 1**. Οι προγνωστικοί παράγοντες είναι ανθρωπομετρικά δεδομένα και παράμετροι που μπορούν να συγκεντρωθούν σε αναλύσεις ρουτίνας αίματος. Τα δεδομένα δεν περιέχουν missing values. Δηλαδή δεν υπάρχουν δεδομένα χωρίς να έχουν μια τιμή.

Table 13 - Ποσοτικά Χαρακτηριστικά 1

Age (years)	Περιγράφει την ηλικία του ατόμου.	
BMI (kg/m2)	Υπολογίζεται ως η αναλογία του βάρους προς το ύψος του ατόμου, προσδιορίζοντας τη σχέση μεταξύ βάρους και ύψους.	
Glucose (mg/dL)	Αναφέρεται στη συγκέντρωση της γλυκόζης στο αίμα, που είναι ένα σημαντικό δείκτης της σακχαρώδους διαβήτη.	

Insulin (μU/mL)	Αντιπροσωπεύει τη συγκέντρωση της ινσουλίνης στο αίμα, μια ορμόνη που ελέγχει την αναλογία του σακχάρου στο αίμα.	
HOMA	Είναι ένας υπολογισμός που αντικατοπτρίζει την αντίσταση της ινσουλίνης και την λειτουργία των κυττάρων του σώματος.	
Leptin (ng/mL)	Πρόκειται για μια ορμόνη που ρυθμίζει τον έλεγχο του βάρους.	
Adiponectin (μg/mL)	Είναι μια ορμόνη που σχετίζεται με τον έλεγχο του μεταβολισμού και την αντιστοίχιση του βάρους.	
Resistin (ng/mL)	Αναφέρεται σε μια ορμόνη που σχετίζεται με φλεγμονώδεις αντιδράσεις και την αντίσταση στην ινσουλίνη.	
MCP-1(pg/dL)	Παράγοντας που προκαλεί υπερφλεγμονώδη αντίδραση στον οργανισμό.	

Table 14 - Labels 1

1	Healthy controls – Υγιείς	
2	Patients – Ασθενείς	

Η εικόνα παρουσιάζει γραφικά τα ποσοτικά χαρακτηριστικά του σετ δεδομένων

Age	BMI	Glucose	Insulin	HOMA	Leptin	Adiponecti	Resistin	MCP.1	Classificati
48	23.5	70	2.707	0.467409	8.8071	9.7024	7.99585	417.114	1
83	20.69049	92	3.115	0.706897	8.8438	5.429285	4.06405	468.786	1
82	23.12467	91	4.498	1.009651	17.9393	22.43204	9.27715	554.697	1
68	21.36752	77	3.226	0.612725	9.8827	7.16956	12.766	928.22	1
86	21.11111	92	3.549	0.805386	6.6994	4.81924	10.57635	773.92	1
49	22.85446	92	3.226	0.732087	6.8317	13.67975	10.3176	530.41	1
89	22.7	77	4.69	0.890787	6.964	5.589865	12.9361	1256.083	1
76	23.8	118	6.47	1.883201	4.311	13.25132	5.1042	280.694	1
73	22	97	3.35	0.801543	4.47	10.35873	6.28445	136.855	1
75	23	83	4.952	1.013839	17.127	11.57899	7.0913	318.302	1
34	21.47	78	3.469	0.667436	14.57	13.11	6.92	354.6	1
29	23.01	82	5.663	1.145436	35.59	26.72	4.58	174.8	1
25	22.86	82	4.09	0.827271	20.45	23.67	5.14	313.73	1
24	18.67	88	6.107	1.33	8.88	36.06	6.85	632.22	1
38	23.34	75	5.782	1.06967	15.26	17.95	9.35	165.02	1
44	20.76	86	7.553	1.6	14.09	20.32	7.64	63.61	1
47	22.03	84	2.869	0.59	26.65	38.04	3.32	191.72	1
61	32.03896	85	18.077	3.790144	30.7729	7.780255	13.68392	444.395	1
64	34.52972	95	4.427	1.037394	21.2117	5.46262	6.70188	252.449	1
32	36.51264	87	14.026	3.00998	49.3727	5.1	17.10223	588.46	1
36	28.57668	86	4.345	0.921719	15.1248	8.6	9.1539	534.224	1
34	31.97501	87	4.53	0.972138	28.7502	7.64276	5.62592	572.783	1
29	32.27079	84	5.81	1.203832	45.6196	6.209635	24.6033	904.981	1
35	30.27682	84	4.376	0.906707	39.2134	9.048185	16.43706	733.797	1
54	30.48316	90	5.537	1.229214	12.331	9.73138	10.19299	1227.91	1
45	37.03561	83	6.76	1.383997	39.9802	4.617125	8.70448	586.173	1
50	38.57876	106	6.703	1.752611	46.6401	4.667645	11.78388	887.16	1
66	31.44654	90	9.245	2.05239	45.9624	10.35526	23.3819	1102.11	1
35	35.25076	90	6.817	1.513374	50.6094	6.966895	22.03703	667.928	1
36	34.17489	80	6.59	1.300427	10.2809	5.065915	15.72187	581.313	1
66	36.21228	101	15.533	3.869788	74.7069	7.53955	22.32024	864.968	1

Figure 10 - Χαρακτηριστικά του σετ δεδομένων I

Αλγόριθμοι

Στο προηγούμενο κεφάλαιο πραγματοποιήθηκε λεπτομερής ανάλυση του συνόλου δεδομένων που αφορά τον καρκίνο του μαστού. Στη συνέχεια, διεξήχθησαν πειράματα στο σύνολο αυτό με στόχο την εύρεση του βέλτιστου αλγορίθμου για την εκπαίδευση του μοντέλου πρόβλεψης. Κατά τη διάρκεια αυτών των πειραμάτων, εξετάστηκαν πολλοί αλγόριθμοι μηχανικής μάθησης, συμπεριλαμβανομένων του Logistic Regression, του Random Forest, του Decision Tree, του K-Nearest Neighbors (KNN), του Gradient Boosting και του Quadratic Discriminant Analysis (QDA).

Ο αλγόριθμος **Logistic Regression** χρησιμοποιείται για προβλέψεις σε δύο κατηγορίες. Αναλύει τη σχέση μεταξύ των εξαρτώμενων μεταβλητών και την πιθανότητα να ανήκουν σε μια συγκεκριμένη κατηγορία. Ο αλγόριθμος προσαρμόζει μια συνάρτηση, η οποία περιγράφει τις πιθανότητες των δύο κατηγοριών, με βάση τις εισόδους. Στη συνέχεια, χρησιμοποιεί αυτές τις πιθανότητες για να προβλέψει σε ποια κατηγορία ανήκει ένα νέο δείγμα.

Ο αλγόριθμος **Random Forest** είναι ένας ισχυρός αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται τόσο για προβλήματα ταξινόμησης όσο και παλινδρόμησης. Η ισχύς του πηγάζει από τον συνδυασμό πολλών απλών μοντέλων πρόβλεψης, γνωστών ως δέντρα αποφάσεων. Ο βασικός λόγος που ονομάζεται "Random Forest" είναι ότι δημιουργεί ένα δάσος (Forest) από δέντρα αποφάσεων. Κάθε δέντρο λαμβάνει ένα τυχαίο υποσύνολο των χαρακτηριστικών και δειγμάτων για την εκπαίδευσή του. Κατά την πρόβλεψη, κάθε δέντρο συμβάλλει με τη δική του πρόβλεψη, και η τελική απόφαση λαμβάνεται μέσω ψηφοφορίας ή μέσης των προβλέψεων των δέντρων.

Ο αλγόριθμος **Decision Tree** αναπαριστά ένα μονοπάτι αποφάσεων που διακλαδώνεται βάσει των χαρακτηριστικών των δεδομένων, βοηθώντας στην κατηγοριοποίηση των εισόδων σε διάφορες κατηγορίες. Κάθε κόμβος του δέντρου αποτελεί μια απόφαση βάσει ενός χαρακτηριστικού, και το δέντρο διαχωρίζει τα δεδομένα σε υποσύνολα με βάση αυτές τις αποφάσεις. Κάθε φύλλο του δέντρου αντιστοιχεί σε μια κατηγορία ή μια τιμή πρόβλεψης. Ο Decision Tree λειτουργεί με την αρχή του διαχωρισμού, προσπαθώντας να διαχωρίσει τα δεδομένα σε ομοιόμορφους και ξεκάθαρους τρόπους. Κατά την εκπαίδευση, το δέντρο επιλέγει τα χαρακτηριστικά που διαχωρίζουν καλύτερα τα δεδομένα, με στόχο να μειώσει την ανασφάλεια (uncertainty) για τις προβλέψεις του.

Ο **K-Nearest Neighbors (KNN)** βασίζεται στην απόσταση μεταξύ των δειγμάτων και ταξινομεί ένα νέο δείγμα βάσει των κοντινότερων (K) γειτονικών δειγμάτων από το σύνολο εκπαίδευσης. Λειτουργεί βασιζόμενος στην ιδέα ότι παρόμοια δείγματα ανήκουν σε παρόμοιες κατηγορίες. Για να κατατάξει ένα νέο δείγμα, βρίσκει τα κοντινότερα (K) δείγματα από τα αποθηκευμένα δεδομένα εκπαίδευσης. Η κατηγορία του νέου δείγματος καθορίζεται από την πλειοψηφία των κατηγοριών των K πλησιέστερων γειτόνων. Το συγκεκριμένο πείραμα πραγματοποιήθηκε με $K=5$.

Ο αλγόριθμος **Gradient Boosting** αποτελεί έναν αλγόριθμο μηχανικής μάθησης που χρησιμοποιείται για προβλήματα ταξινόμησης και παλινδρόμησης. Σε πολύ απλά λόγια, αυτός ο αλγόριθμος δημιουργεί πολλά μικρά μοντέλα πρόβλεψης και στη συνέχεια τα συνδυάζει για να δημιουργήσει ένα ισχυρότερο μοντέλο πρόβλεψης. Κάθε νέο μοντέλο επικεντρώνεται στα

σημεία που το προηγούμενο μοντέλο έκανε λάθη, επιτρέποντας μας να βελτιώσουμε συνεχώς την ακρίβεια των προβλέψεών μας.

Τέλος ο αλγόριθμος **Quadratic Discriminant Analysis** (QDA) αναπτύσσει μια κυρτή συνάρτηση απόφασης, η οποία είναι ένα πολυώνυμο δευτέρου βαθμού. Εξετάζει πιο πολύπλοκες καμπύλες για να προσαρμόσει τα δεδομένα. Αυτό επιτρέπει στον QDA να αντιμετωπίζει πιο σύνθετες σχέσεις μεταξύ των χαρακτηριστικών και των κατηγοριών, καθιστώντας τον κατάλληλο για δεδομένα όπου οι σχέσεις ανάμεσα στις μεταβλητές δεν είναι γραμμικές, αλλά περιέχουν πολυπλοκότερες κυρτές μορφές.

Κατά την αξιολόγηση της απόδοσης των αλγορίθμων, λήφθηκε υπόψη η ακρίβεια των προβλέψεών τους, προκειμένου να επιλεγεί ο βέλτιστος αλγόριθμος για την εκπαίδευση του τελικού μοντέλου πρόβλεψης του καρκίνου του μαστού.

Αποτελέσματα

Κατά τη διαδικασία αξιολόγησης της απόδοσης των αλγορίθμων, λήφθηκε υπόψη η ακρίβεια των προβλέψεών τους, με στόχο την επιλογή του βέλτιστου αλγορίθμου για την εκπαίδευση του τελικού μοντέλου πρόβλεψης του καρκίνου του μαστού. Κατά την αξιολόγηση της απόδοσης των αλγορίθμων πρόβλεψης καρκίνου του μαστού, ελήφθη υπόψη η χρήση διαφόρων μετρικών αξιολόγησης. Αυτές οι μετρικές περιλαμβάνουν την ακρίβεια (Accuracy), την εκτίμηση θετικών προβλέψεων (Precision), την ανάκληση (Recall), και το F1-score.

Η ακρίβεια (Accuracy) αποτελεί το ποσοστό των σωστών προβλέψεων συνολικά, εκφραζόμενη σε ποσοστό. Η εκτίμηση θετικών προβλέψεων (Precision) υπολογίζεται ως το ποσοστό των πραγματικά θετικών προβλέψεων ανάμεσα στις προβλέψεις που έγιναν ως θετικές. Η ανάκληση (Recall) αναφέρεται στο ποσοστό των πραγματικά θετικών προβλέψεων ανάμεσα στις πραγματικά θετικές περιπτώσεις. Το F1-score αποτελεί έναν συνδυασμό της ακρίβειας και της ανάκλησης, παρέχοντας έναν ενιαίο δείκτη ποιότητας της πρόβλεψης. Αυτές οι μετρικές παρέχουν εμβριθείς πληροφορίες σχετικά με την απόδοση των αλγορίθμων, καθορίζοντας τον βέλτιστο αλγόριθμο για την εκπαίδευση του τελικού μοντέλου πρόβλεψης του καρκίνου του μαστού.

Table 15 - Αποτελέσματα

Classifier	Accuracy	Precision	Recall	F1-score	Classification
Logistic Regression	71.67%	0.73	0.92	0.81	1
		0.89	0.67	0.76	2
Random Forest	61.22%	0.73	0.67	0.70	1
		0.69	0.75	0.72	2
KNN (5)	47.00%	0.42	0.42	0.42	1
		0.42	0.42	0.42	2
Decision Tree	55.44%	0.70	0.58	0.64	1
		0.64	0.75	0.69	2
Gradient Boosting	56.78%	0.67	0.67	0.67	1
		0.67	0.67	0.67	2
QDA	58.00%	0.64	0.75	0.69	1
		0.70	0.58	0.64	2

Τα αποτελέσματα, όπως φαίνεται και στο **Table 15 - Αποτελέσματα**, κατέληξαν σε μια ενδιαφέρουσα σύγκριση ακρίβειας μεταξύ των διαφόρων μεθόδων. Σύμφωνα με τα δεδομένα αυτά, ο Logistic Regression ξεχώρισε ως ο πιο αποτελεσματικός ταξινομητής με ακρίβεια

71.67%. Αντίθετα, ο αλγόριθμος K-Nearest Neighbors (KNN) παρουσίασε το χαμηλότερο ποσοστό ακρίβειας στα δεδομένα μας, με μόλις 47.00%. Παρόλο που ο KNN είναι γνωστός για την απλότητά του και την ευελιξία στην προσαρμογή σε διάφορα σύνολα δεδομένων, στην προκειμένη περίπτωση φάνηκε να μην αποδίδει τόσο αποτελεσματικά όσο οι υπόλοιποι ταξινομητές. Ο αλγόριθμος Random Forest επέδειξε μέτρια ακρίβεια με ποσοστό 61.22%, ενώ ο Decision Tree είχε ακρίβεια 55.44%. Ο Gradient Boosting παρουσίασε επίσης μέτρια ακρίβεια με ποσοστό 56.78%. Τέλος, ο αλγόριθμος Quadratic Discriminant Analysis (QDA) κατέγραψε ακρίβεια 58.00%. Οι παραπάνω αλγόριθμοι, παρά τις διαφορετικές τους προσεγγίσεις, δεν ξεπερνούν σε ακρίβεια τον Logistic Regression στο συγκεκριμένο σύνολο δεδομένων.

Η επιλογή του κατάλληλου ταξινομητή είναι κρίσιμη για την ακρίβεια των προβλέψεων σε ένα μοντέλο μηχανικής μάθησης. Ο Logistic Regression ξεχώρισε όχι μόνο λόγω της υψηλής ακρίβειάς του, αλλά και λόγω της ικανότητάς του να αντιμετωπίζει πολύπλοκες σχέσεις μεταξύ των μεταβλητών. Η απλότητα και η αποτελεσματικότητά του τον καθιστούν συχνά την πρώτη επιλογή σε πολλά πρακτικά προβλήματα. Παράλληλα, η σύγκρισή του με άλλους αλγορίθμους αποκαλύπτει τη σημασία της κατανόησης του προβλήματος και των χαρακτηριστικών των δεδομένων για την επιλογή της κατάλληλης μεθόδου.

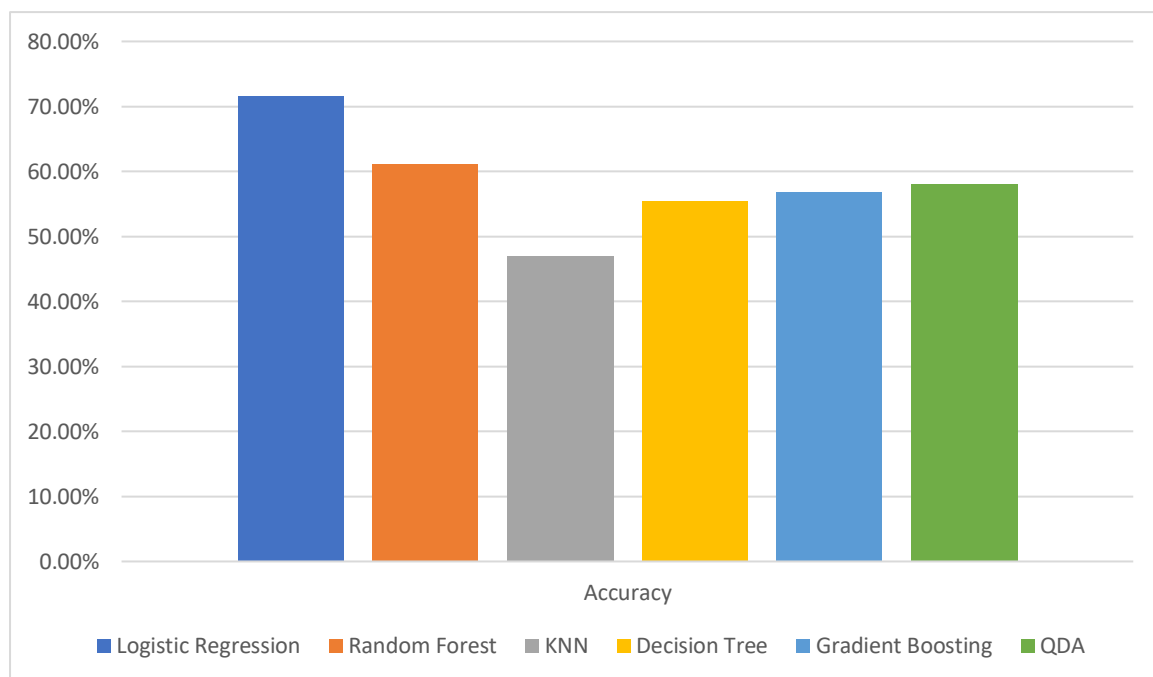


Figure 11 - Διάγραμμα αποτελεσμάτων Ακρίβειας I

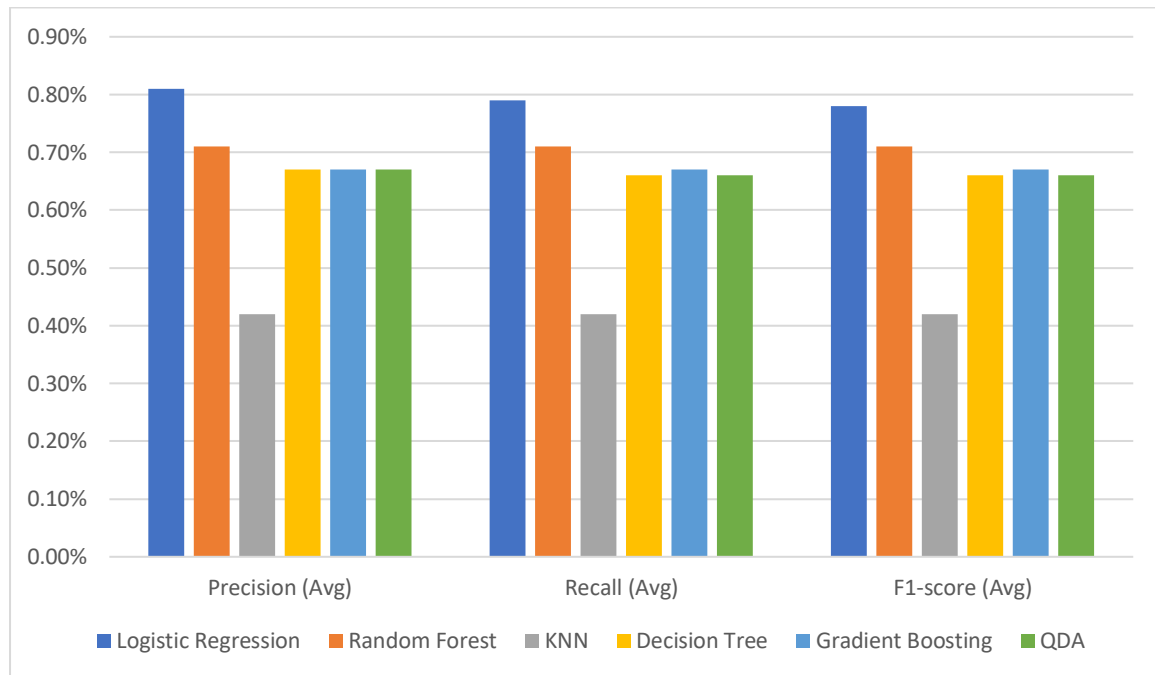


Figure 12 - Διάγραμμα αποτελεσμάτων Ακρίβειας 2

Ανάπτυξη εφαρμογής

Η android εφαρμογή που πήρε την ονομασία “B.C Compass”, δηλαδή Breast Cancer Compass, αναπτύχθηκε ειδικά για το ευρύ κοινό, με σκοπό να προσφέρει πρόσβαση σε προηγμένες επιστημονικές μεθόδους για την πρόληψη και τη διάγνωση του καρκίνου του μαστού. Μέσω των αποτελεσμάτων απλών εξετάσεων αίματος, η εφαρμογή παρέχει τη δυνατότητα αξιολόγησης του προσωπικού κινδύνου του χρήστη και παρέχει σημαντικές πληροφορίες σχετικά με τον κίνδυνο εμφάνισης καρκίνου του μαστού. Αναλύει τα αποτελέσματα των εξετάσεων αίματος χρησιμοποιώντας προηγμένες επιστημονικές μεθόδους και αλγόριθμους μηχανικής μάθησης. Αυτό επιτρέπει στον χρήστη να έχει μια σαφή εικόνα για την υγεία του και να λάβει τυχόν προληπτικά μέτρα, βασιζόμενος σε επιστημονικά αξιόπιστες πληροφορίες.

Η εφαρμογή αναπτύχθηκε χρησιμοποιώντας το περιβάλλον ανάπτυξης Android Studio. Η γραφική διεπαφή της εφαρμογής επιτρέπει στον χρήστη να εισάγει τιμές για τα επιλεγμένα χαρακτηριστικά. Κατόπιν, η εφαρμογή χρησιμοποιεί το εκπαιδευμένο μοντέλο για να κάνει προβλέψεις σχετικά με τον κίνδυνο εμφάνισης καρκίνου του μαστού βασιζόμενη στις εισόδους του χρήστη. Οι πληροφορίες σχετικά με τα χαρακτηριστικά επεξηγούνται λεπτομερώς μέσα από την εφαρμογή, παρέχοντας στον χρήστη σαφήνεια για τις προβλέψεις που λαμβάνει. Επιπλέον, παρέχεται δυνατότητα πρόσβασης σε περισσότερες πληροφορίες για κάθε χαρακτηριστικό μέσω διαφανών πλαισίων πληροφοριών.

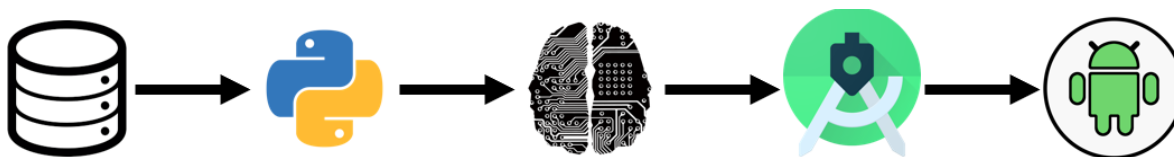


Figure 13 - Διαδικασία ανάπτυξης εφαρμογής 1

Ακολούθως, παρουσιάζονται διάφορα στιγμιότυπα οθόνης που αντιπροσωπεύουν τη λειτουργικότητα της εφαρμογής. Κάθε στιγμιότυπο προσφέρει λεπτομερή εικόνα της διεπαφής και των λειτουργιών της εφαρμογής, παρέχοντας μια επιστημονική οπτική προσέγγιση για τη χρήση του εργαλείου. Αυτή η σειρά εικόνων προσφέρει μια συνολική εικόνα του λειτουργικού περιβάλλοντος της εφαρμογής, εμβαθύνοντας στις λεπτομέρειες που καθορίζουν την απόδοση και την αποτελεσματικότητά της.

Login Screen

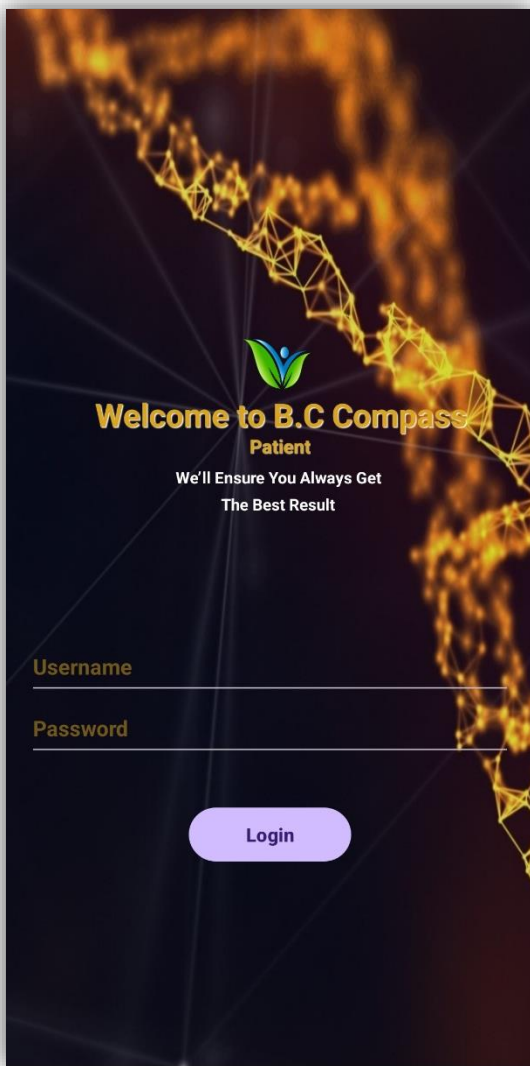


Figure 14 - Login Screen

Η αρχική οθόνη σύνδεσης της εφαρμογής B.C Compass Patient δημιουργήθηκε με έμφαση στην απλότητα και τη λειτουργικότητα. Όπως παρατηρείται και στο **Figure 14 - Login Screen**, οι χρήστες καλωσορίζονται με μια καθαρή διεπαφή, που τους δίνει άμεση πρόσβαση στον προσωπικό τους λογαριασμό. Το σύστημα εισόδου σχεδιάστηκε για να παρέχει ασφαλή σύνδεση, ενώ η απλότητα της διαδικασίας εξυπηρετεί την άμεση πρόσβαση στις λειτουργίες που αφορούν την εκτίμηση του προσωπικού κινδύνου και την αναζήτηση πληροφοριών για τον καρκίνο του μαστού. Με αυτόν τον τρόπο, οι χρήστες μπορούν να αξιοποιήσουν αποτελεσματικά τα χαρακτηριστικά της εφαρμογής για τη βελτίωση και την ενημέρωση της υγείας τους.

Αρχική Οθόνη

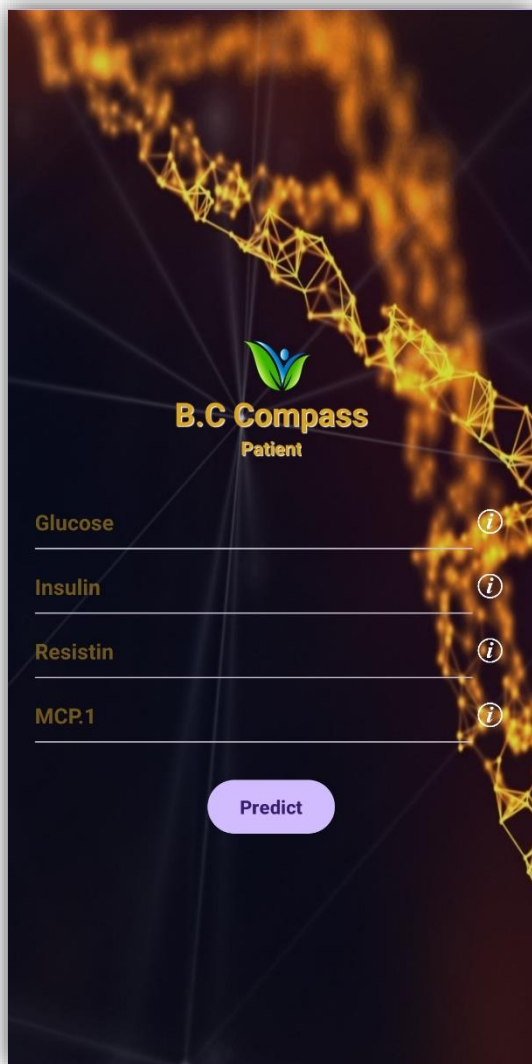


Figure 15 - Αρχική Οθόνη

Στο **Figure 16 - Είσοδος Δεδομένων** που παρουσιάζεται, παρατηρούμε την ενδεικτική εισαγωγή αριθμών στα πεδία Glucose, Insulin, Resistin και MCP.1. Αυτή η αναπαράσταση επισημαίνει τη δυνατότητα εισαγωγής τόσο ακέραιων όσο και δεκαδικών αριθμών, επιτρέποντας την πλήρη κάλυψη του φάσματος των δεδομένων και την ακριβή αξιολόγηση των παραμέτρων για κάθε χρήστη.

Στην αρχική οθόνη της εφαρμογής B.C Compass Patient παρουσιάζονται επίσημα τέσσερα πεδία εισαγωγής δεδομένων, κατηγοριοποιημένα ανά επίπεδο γλυκόζης, ινσουλίνης, αντίστασης και MCP.1 στο αίμα. Στο **Figure 15 - Αρχική Οθόνη** οι χρήστες έχουν τη δυνατότητα να καταχωρούν τα αποτελέσματά τους, παρέχοντας πολύτιμες πληροφορίες για την υγεία τους, ενώ ταυτόχρονα ενισχύεται η πληρότητα και η αξιοπιστία των δεδομένων που χρησιμοποιούνται για την εκτίμηση του κινδύνου εμφάνισης καρκίνου του μαστού.

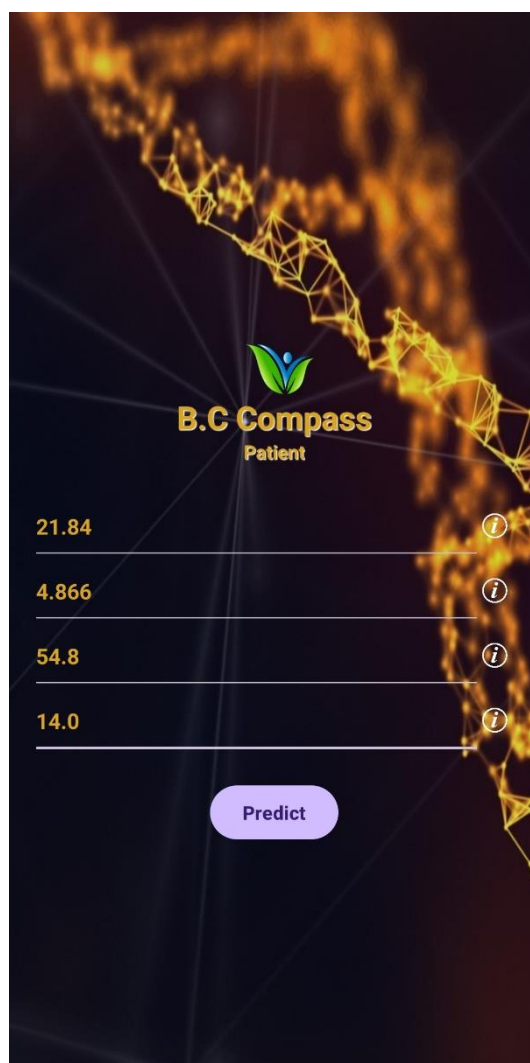


Figure 16 - Είσοδος Δεδομένων

Πρόβλεψη

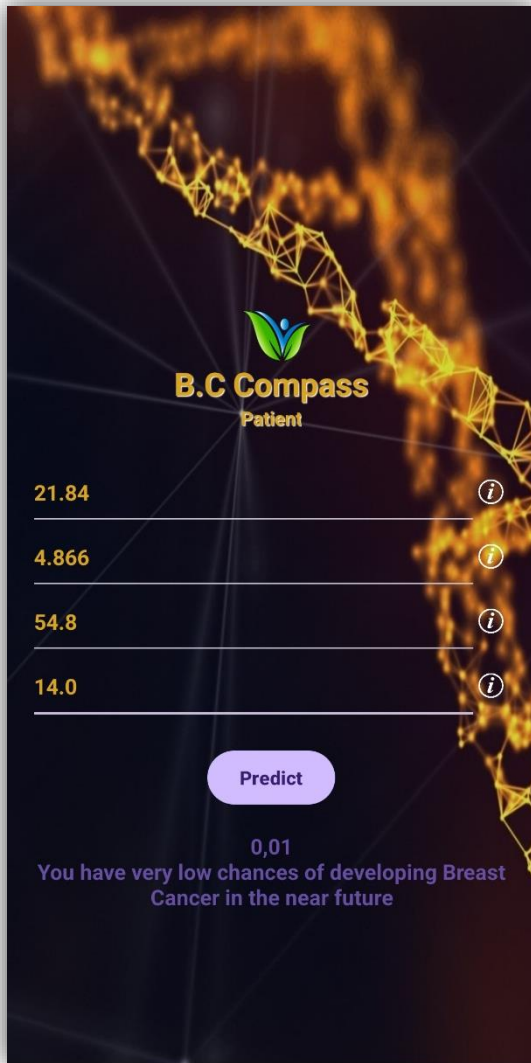


Figure 17 - Εμφάνιση πρόβλεψης

Μετά την πρόβλεψη του μοντέλου εμφανίζεται στην οθόνη του χρήστη, όπως φαίνεται στο **Figure 17 - Εμφάνιση πρόβλεψης** και **Figure 18 - Εμφάνιση πρόβλεψης 2**, το μήνυμα "Έχετε ... πιθανότητες να αναπτύξετε καρκίνο του μαστού στο κοντινό μέλλον.", καθώς και το ποσοστό εμφάνισης. Οι τρεις τελείες αντιπροσωπεύουν το ποσοστό πιθανότητας που προκύπτει από τα δεδομένα εισόδου που παρέχει ο χρήστης. Με αυτόν τον τρόπο, ο χρήστης λαμβάνει ένα σαφές μήνυμα σχετικά με τον εκτιμώμενο κίνδυνο και μπορεί να λάβει τυχόν προληπτικά μέτρα βασιζόμενος στην εκτίμηση αυτή. Στο συγκεκριμένο παράδειγμα βάσει των δεδομένων εισόδου παρατηρείται πως το ποσοστό είναι χαμηλό (0,01) και το μήνυμα αναφέρει πολύ χαμηλές πιθανότητες εμφάνισης καρκίνου του μαστού (very low chances).

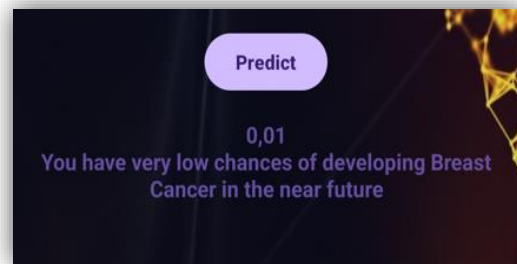


Figure 18 - Εμφάνιση πρόβλεψης 2

Επιπλέον, σε περίπτωση πιθανών λανθασμένων καταχωρήσεων ή ανακρίβειας των δεδομένων, η εφαρμογή εμφανίζει ένα σχετικό μήνυμα λάθους, προστατεύοντας την ακρίβεια των υπολογισμών. Ο χρήστης ενημερώνεται για το σφάλμα και καλείται να διορθώσει τυχόν ασάφειες προτού συνεχίσει την αξιολόγηση. Με αυτήν την προσέγγιση, διασφαλίζεται ότι ο χρήστης λαμβάνει αξιόπιστες και ενημερωμένες πληροφορίες για την υγεία του, παρέχοντας του ένα εργαλείο πρόβλεψης που βασίζεται σε ακριβή δεδομένα.

Πληροφορίες

Γλυκόζη

Στη **Figure 19 - Πληροφορίες Γλυκόζης** φαίνονται οι πληροφορίες για τα επίπεδα γλυκόζης, παρέχοντας στον χρήστη ένα σημαντικό δείκτη της υγείας του. Τα επίπεδα γλυκόζης στο αίμα αποτελούν κρίσιμο παράγοντα για τη διαχείριση της υγείας και την πρόληψη του καρκίνου του μαστού.

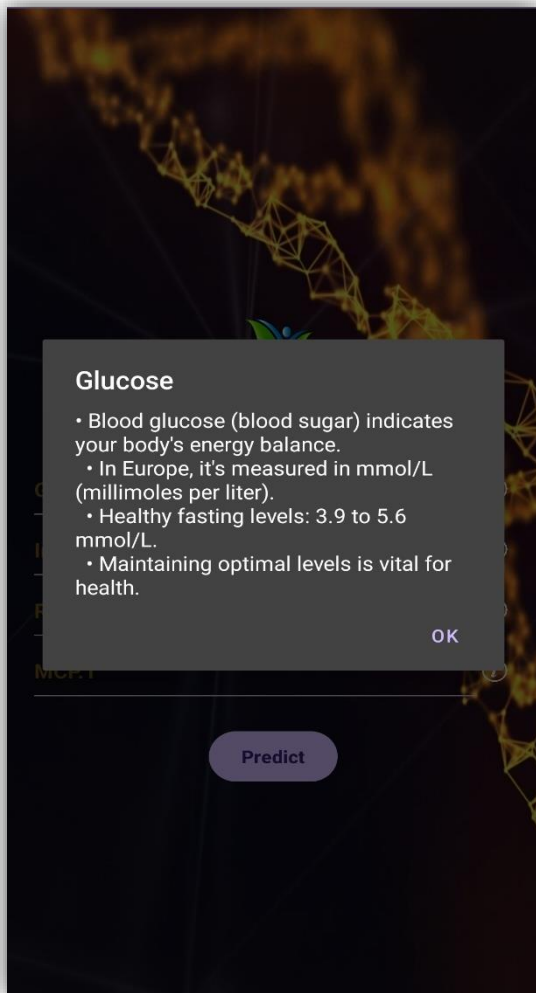


Figure 19 - Πληροφορίες Γλυκόζης

Η γλυκόζη αντιπροσωπεύει τη συγκέντρωση ζάχαρης στο αίμα και είναι καθοριστική για τη λειτουργία του οργανισμού. Υψηλά επίπεδα γλυκόζης μπορεί να υποδεικνύουν διαβήτη και άλλα προβλήματα υγείας. Η παρακολούθηση αυτών των επιπέδων είναι σημαντική για την πρόληψη και τη διαχείριση χρόνιων παθήσεων, συμπεριλαμβανομένου του καρκίνου του μαστού.

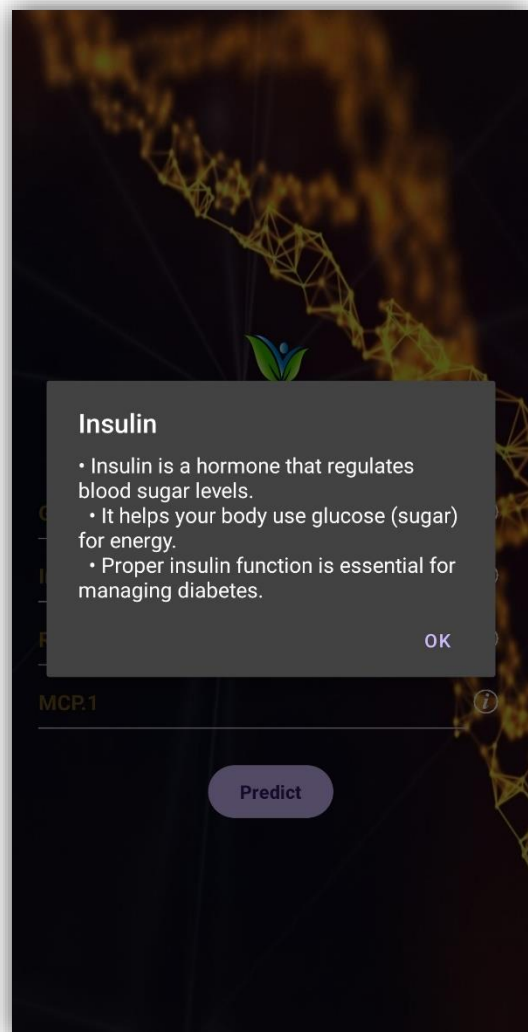


Figure 20 - Πληροφορίες Ινσουλίνης

Ινσουλίνη

Στη **Figure 20 - Πληροφορίες Ινσουλίνης**, φαίνονται οι πληροφορίες για τα επίπεδα ινσουλίνης, παρέχοντας στον χρήστη σημαντικές πληροφορίες για την υγεία του. Η ινσουλίνη είναι ουσιώδης για τη ρύθμιση της γλυκόζης και η παρακολούθησή της βοηθά στη διατήρηση της υγείας, ιδίως όταν συνδέεται με τον καρκίνο του μαστού.

Η ινσουλίνη είναι ορμόνη που επιτρέπει στα κύτταρα να απορροφούν τη γλυκόζη από το αίμα. Υψηλά επίπεδα ινσουλίνης μπορεί να υποδεικνύουν ανθεκτικότητα στην ινσουλίνη, συνδεδεμένη με τον κίνδυνο καρκίνου και άλλες παθήσεις.

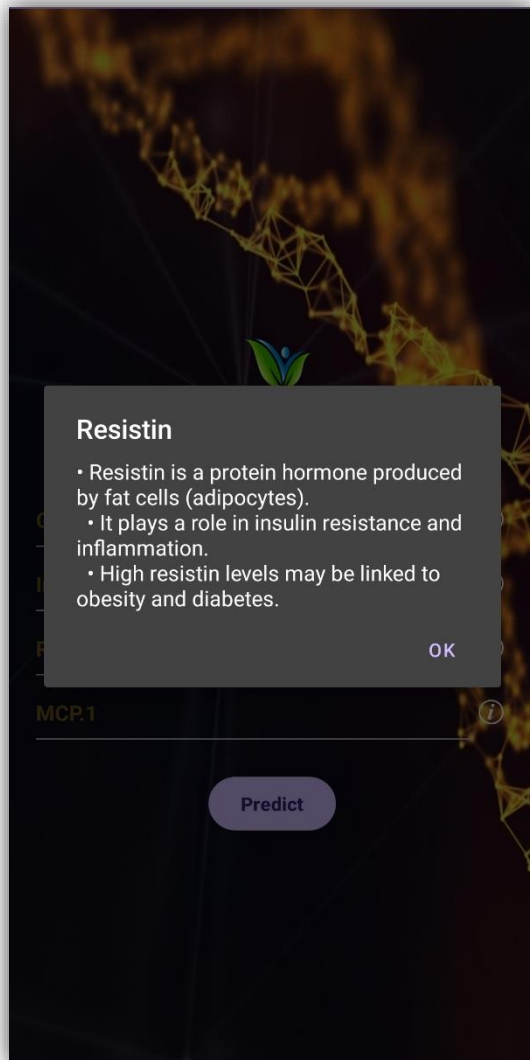


Figure 21 - Πληροφορίες Ρεζιστίνης

Ρεζιστίνη

Στο **Figure 21 - Πληροφορίες Ρεζιστίνης** παρουσιάζονται οι πληροφορίες σχετικά με τα επίπεδα ρεζιστίνης, παρέχοντας στον χρήστη σημαντικά δεδομένα σχετικά με την υγεία του. Η ρεζιστίνη είναι ένα ορμόνη που συνδέεται με την ανοσολογική απόκριση και τη ρύθμιση του μεταβολισμού.

Υψηλά επίπεδα ρεζιστίνης μπορεί να σχετίζονται με αντίσταση στην ινσουλίνη, η οποία μπορεί να οδηγήσει σε προβλήματα υγείας όπως ο σακχαρώδης διαβήτης. Η παρακολούθηση των επιπέδων ρεζιστίνης μπορεί να βοηθήσει στην πρόληψη και τη διαχείριση αυτών των προβλημάτων υγείας.

MCP.1

Στο **Figure 22 - Πληροφορίες MCP.1** παρουσιάζονται οι πληροφορίες σχετικά με τα επίπεδα MCP.1.

Το MCP.1, γνωστό επίσης ως Monocyte Chemoattractant Protein-1, αποτελεί έναν χημικό παράγοντα που συμμετέχει στον ρυθμιστικό έλεγχο της κυκλοφορίας των μονοκυττάρων, προσελκύοντας αυτά τα λευκοκύτταρα προς περιοχές φλεγμονής. Η παρακολούθηση των επιπέδων του MCP.1 αποτελεί σημαντικό εργαλείο για την αξιολόγηση της φλεγμονώδους αντίδρασης και της ανοσοκατάστασης. Εντός του πλαισίου της εφαρμογής B.C Compass Patient, η παρακολούθηση αυτού του παράγοντα συμβάλλει στην ολοκληρωμένη εκτίμηση του ατομικού κινδύνου.

Κατά το κλείσιμο αυτής της ανάλυσης, είναι εμφανές πως η εφαρμογή B.C Compass Patient διαδραματίζει έναν σημαντικό ρόλο στον τομέα της υγείας. Επωφελούμενη από προηγμένες επιστημονικές μεθόδους και τη χρήση της τεχνητής νοημοσύνης, η εφαρμογή παρέχει στους χρήστες σημαντικές πληροφορίες σχετικά με τον προσωπικό τους κίνδυνο για τον καρκίνο του μαστού.

Η δυνατότητα παρακολούθησης και αξιολόγησης των επιπέδων γλυκόζης, ινσουλίνης, ρεξιστίνης και MCP.1 προσφέρει στον χρήστη ένα ενημερωμένο βάθος κατανόησης της προσωπικής του υγείας. Μέσω της επιστημονικής ανάλυσης και των προηγμένων αλγορίθμων, η εφαρμογή συμβάλλει στην ενίσχυση της πρόληψης και της πρόγνωσης του καρκίνου του μαστού, ανοίγοντας νέες προοπτικές για τη βελτίωση της υγείας και της ευημερίας.

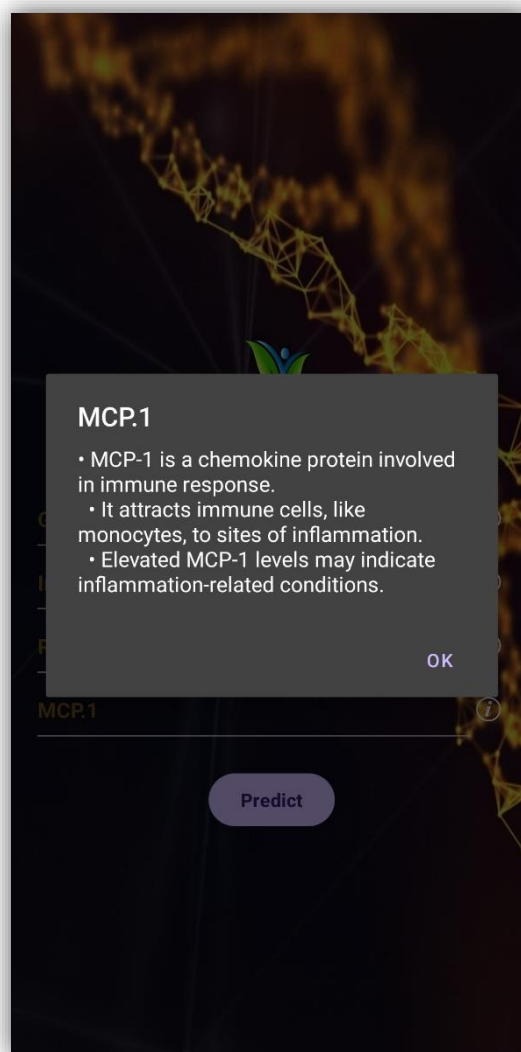


Figure 22 - Πληροφορίες MCP.1

Κεφάλαιο 3^ο

Dataset Breast Cancer Wisconsin (Diagnostic)

Το σύνολο δεδομένων αυτό προκύπτει από μια ψηφιοποιημένη εικόνα αναρρόφησης λεπτής βελόνας (FNA) μάζας μαστού. Αυτά τα δεδομένα δεν περιγράφουν απλά την εικόνα, αλλά αναδεικνύουν τα χαρακτηριστικά των κυτταρικών πυρήνων που υπάρχουν σε αυτήν. Αυτή η ψηφιακή αναπαράσταση προσφέρει εμβαθείς και λεπτομερείς αναλύσεις, επιτρέποντας την εξερεύνηση της δομής των κυττάρων και την ανακάλυψη σημαντικών πληροφοριών για πιθανές παθήσεις. Το σετ δεδομένων περιέχει 569 περιπτώσεις και 30 χαρακτηριστικά όπως παρατηρείται και στο **Table 17 - Ποσοτικά Χαρακτηριστικά 2**. Τα χαρακτηριστικά υπολογίζονται από μια ψηφιοποιημένη εικόνα αναρρόφησης λεπτής βελόνας (FNA) μάζας μαστού. Περιγράφουν χαρακτηριστικά των κυτταρικών πυρήνων που υπάρχουν στην εικόνα.

Το επίπεδο διαχωρισμού που περιγράφεται παραπάνω λήφθηκε χρησιμοποιώντας Multisurface Method-Tree (MSM-T) [K. P. Bennett, «Κατασκευή δέντρου αποφάσεων μέσω γραμμικού προγραμματισμού». Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society, σελ. 97-101, 1992], μια μέθοδος ταξινόμησης που χρησιμοποιεί γραμμικό προγραμματισμό για την κατασκευή ενός δέντρου αποφάσεων. Τα σχετικά χαρακτηριστικά επιλέχθηκαν χρησιμοποιώντας μια εξαντλητική αναζήτηση στο χώρο 1-4 χαρακτηριστικών και 1-3 διαχωριστικών επιπέδων.

Κατανομή κατηγορίας: Καλοήθεις: 357 - Κακοήθεις: 212

Table 16 - Labels 2

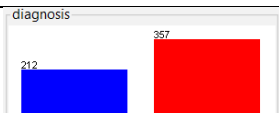
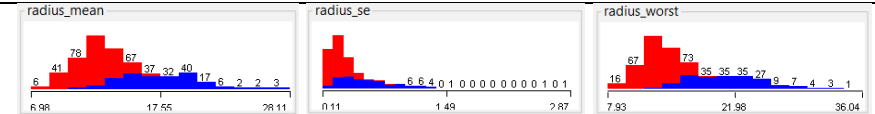
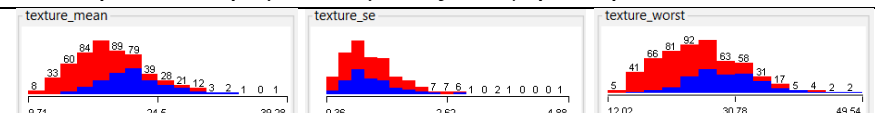
ID number	
Diagnosis (M = malignant(κακοήθεις), B = benign(καλοήθεις))	

Table 17 - Ποσοτικά Χαρακτηριστικά 2

Radius	Η μέση απόσταση από το κέντρο του πυρήνα στα σημεία της περιφέρειας
	
Texture	Τυπική απόκλιση τιμών κλίμακας του γκρι στην εικόνα
	

Perimeter	Το συνολικό μήκος της περιφέρειας του πυρήνα
	perimeter_mean perimeter_se perimeter_worst
Area	Η επιφανειακή έκταση του πυρήνα
	area_mean area_se area_worst
Smoothness	Τοπική διακύμανση στο μήκος της ακτίνας
	smoothness_mean smoothness_se smoothness_worst
Compactness	$(\text{Περίμετρος}^2)/\text{Εμβαδό} - 1.0$, μέτρηση του σχήματος του πυρήνα
	compactness_mean compactness_se compactness_worst
Concavity	Σοβαρότητα των κοίλων τμημάτων του περιγράμματος του πυρήνα
	concavity_mean concavity_se concavity_worst
Concave points	Ο αριθμός των κοίλων τμημάτων του περιγράμματος του πυρήνα
	points_mean points_se points_worst
Symmetry	Η αντιστοιχία στη δομή του πυρήνα
	symmetry_mean symmetry_se symmetry_worst
Fractal dimension	Απεικόνιση της "ακτινογραφίας της ακτής" του πυρήνα - 1
	fractal_dimension_mean fractal_dimension_se fractal_dimension_worst

Για κάθε εικόνα υπολογίστηκαν ο μέσος όρος, το τυπικό σφάλμα και το "χειρότερο" ή δηλαδή μεγαλύτερο (μέσος όρος των τριών μεγαλύτερων τιμών) από αυτά τα χαρακτηριστικά, με αποτέλεσμα να δημιουργηθούν 30 χαρακτηριστικά. Το σετ δεδομένων δεν περιέχει missing values, δηλαδή κάθε χαρακτηριστικό έχει από μία τιμή.

Η εικόνα παρουσιάζει γραφικά τα ποσοτικά χαρακτηριστικά του σετ δεδομένων

id	diagnosis	radius_me	texture_m	perimeter_area_mear	smoothness	compactness	concavity	points_me
842302	1	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001
842517	1	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869
84300903	1	19.69	21.25	130	1203	0.1096	0.1599	0.1974
84348301	1	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414
84358402	1	20.29	14.34	135.1	1297	0.1003	0.1328	0.198
843786	1	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578
844359	1	18.25	19.98	119.6	1040	0.09463	0.109	0.1127
84458202	1	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366
844981	1	13	21.82	87.5	519.8	0.1273	0.1932	0.1859
84501001	1	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273
845636	1	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299
84610002	1	15.78	17.89	103.6	781	0.0971	0.1292	0.09954
846226	1	19.17	24.8	132.4	1123	0.0974	0.2458	0.2065
846381	1	15.85	23.95	103.7	782.7	0.08401	0.1002	0.09938
84667401	1	13.73	22.61	93.6	578.3	0.1131	0.2293	0.2128
84799002	1	14.54	27.54	96.73	658.8	0.1139	0.1595	0.1639
848406	1	14.68	20.13	94.74	684.5	0.09867	0.072	0.07395
84862001	1	16.13	20.68	108.1	798.8	0.117	0.2022	0.1722
849014	1	19.81	22.15	130	1260	0.09831	0.1027	0.1479
8510426	2	13.54	14.36	87.46	566.3	0.09779	0.08129	0.06664
8510653	2	13.08	15.71	85.63	520	0.1075	0.127	0.04568
8510824	2	9.504	12.44	60.34	273.9	0.1024	0.06492	0.02956
8511133	1	15.34	14.26	102.5	704.4	0.1073	0.2135	0.2077
851509	1	21.16	23.04	137.2	1404	0.09428	0.1022	0.1097
852552	1	16.65	21.38	110	904.6	0.1121	0.1457	0.1525
852631	1	17.14	16.4	116	912.7	0.1186	0.2276	0.2229
852763	1	14.58	21.53	97.41	644.8	0.1054	0.1868	0.1425
852781	1	18.61	20.25	122.1	1094	0.0944	0.1066	0.149
852973	1	15.3	25.27	102.4	732.4	0.1082	0.1697	0.1683
853201	1	17.57	15.05	115	955.1	0.09847	0.1157	0.09875
853401	1	18.63	25.11	124.8	1088	0.1064	0.1887	0.2319

Figure 23 - Χαρακτηριστικά του σετ δεδομένων 2

Αλγόριθμοι

Όπως και προηγουμένως πραγματοποιήθηκε και σε αυτή τη φάση λεπτομερής ανάλυση του συνόλου δεδομένων που αφορά τον καρκίνο του μαστού. Στη συνέχεια, διεξήχθησαν επίσης πειράματα με στόχο την εύρεση του βέλτιστου αλγορίθμου για την εκπαίδευση του μοντέλου πρόβλεψης. Κατά τη διάρκεια αυτών των πειραμάτων, εξετάστηκαν πολλοί αλγόριθμοι μηχανικής μάθησης, συμπεριλαμβανομένων του **Logistic Regression**, του **Random Forest**, του **Decision Tree**, του **K-Nearest Neighbors (KNN)**, του **Gradient Boosting** και του **GaussianNB**.

Ο αλγόριθμος Gaussian Naive Bayes είναι ένας ταξινομητής βασισμένος στο θεώρημα του Bayes. Χρησιμοποιείται κυρίως για προβλήματα ταξινόμησης, και συγκεκριμένα για περιπτώσεις όπου τα χαρακτηριστικά που χρησιμοποιούνται για την κατηγοριοποίηση είναι ανεξάρτητα μεταξύ τους. Ένα από τα κύρια πλεονεκτήματα του Gaussian Naive Bayes είναι η απλότητα του μοντέλου και η γρήγορη εκπαίδευση. Κατά τη διάρκεια της εκπαίδευσης, ο αλγόριθμος υπολογίζει τις πιθανότητες εμφάνισης κάθε χαρακτηριστικού για κάθε κατηγορία. Στη συνέχεια, κατά τη διαδικασία πρόβλεψης, χρησιμοποιεί αυτές τις πιθανότητες για να κατατάξει νέα δεδομένα σε μια συγκεκριμένη κατηγορία. Η απλότητα του τον καθιστά αποτελεσματικό, ειδικά σε περιπτώσεις με μικρό αριθμό δειγμάτων.

Αποτελέσματα

Κατά τη διαδικασία αξιολόγησης της απόδοσης των αλγορίθμων, όπως και στο προηγούμενο κεφάλαιο λήφθηκε υπόψη η ακρίβεια των προβλέψεών τους, με στόχο την επιλογή του βέλτιστου αλγορίθμου για την εκπαίδευση του τελικού μοντέλου πρόβλεψης του καρκίνου του μαστού. Κατά την αξιολόγηση της απόδοσης των αλγορίθμων πρόβλεψης καρκίνου του μαστού, ελήφθη υπόψη η χρήση διαφόρων μετρικών αξιολόγησης. Αυτές οι μετρικές περιλαμβάνουν την ακρίβεια (Accuracy), την εκτίμηση θετικών προβλέψεων (Precision), την ανάκληση (Recall), και το F1-score.

Table 18 - Αποτελέσματα

Classifier	Accuracy	Precision	Recall	F1-score	Classification
Logistic Regression	79.48%	0.75	0.35	0.48	1
		0.78	0.95	0.85	2
Random Forest	87.68%	0.81	0.76	0.79	1
		0.90	0.93	0.91	2
KNN (5)	80.44%	0.76	0.76	0.76	1
		0.90	0.90	0.90	2
Decision Tree	82.21%	0.74	0.82	0.78	1
		0.92	0.88	0.90	2
Gradient Boosting	87.10%	0.81	0.76	0.79	1
		0.90	0.93	0.91	2
GaussianNB	78.11%	0.50	0.47	0.48	1
		0.78	0.80	0.79	2

Τα αποτελέσματα κατέληξαν σε μια ενδιαφέρουσα σύγκριση ακρίβειας μεταξύ των διαφόρων μεθόδων. Σύμφωνα με τα δεδομένα αυτά, όπως φαίνεται και στο **Table 18 - Αποτελέσματα**, ο Random Forest ξεχώρισε ως ο πιο αποτελεσματικός ταξινομητής με ακρίβεια 87.68%. Αντίθετα, ο αλγόριθμος K-Nearest Neighbors (KNN) παρουσίασε χαμηλότερο ποσοστό ακρίβειας στα δεδομένα, με μόλις 80.44%. Ο αλγόριθμος Logistic Regression επέδειξε μέτρια ακρίβεια με ποσοστό 79.48%, ενώ ο Decision Tree είχε ακρίβεια 82,21%. Ο Gradient Boosting παρουσίασε επίσης πολύ υψηλή ακρίβεια με ποσοστό 87.10%, λίγο χειρότερη από τον Random Forest κάνοντάς τον έτσι τον δεύτερο καλύτερο αλγόριθμο στην ακρίβεια από τους υπόλοιπους. Τέλος, ο αλγόριθμος GaussianNB κατέγραψε την χαμηλότερη ακρίβεια από όλους τους αλγορίθμους με ποσοστό 78.11%. Οι παραπάνω αλγόριθμοι, παρά τις διαφορετικές τους προσεγγίσεις, δεν ξεπερνούν σε ακρίβεια τον Random Forest στο συγκεκριμένο σύνολο δεδομένων.

Η επιλογή του κατάλληλου ταξινομητή είναι κρίσιμη για την ακρίβεια των προβλέψεων σε ένα μοντέλο μηχανικής μάθησης. Ο αλγόριθμος Random Forest καταδεικνύει την αρίστη του απόδοση με βάση τα προαναφερθέντα αποτελέσματα. Η χρήση διαφορετικών υποσυνόλων δεδομένων κατά την εκπαίδευση δημιουργεί ένα πολύπλοκο δίκτυο δέντρων πρόβλεψης, το οποίο συνεργάζεται ενισχύοντας τη σταθερότητα και την αξιοπιστία του μοντέλου. Η υψηλή ακρίβεια του, καθώς και η ικανότητά του να αντιμετωπίζει πολύπλοκες σχέσεις μεταξύ των μεταβλητών, καθιστούν το Random Forest επιλογή προτίμησης σε πρακτικά προβλήματα πρόβλεψης, όπου η ακρίβεια και η σταθερότητα αποτελούν ζωτικής σημασίας παράγοντες.

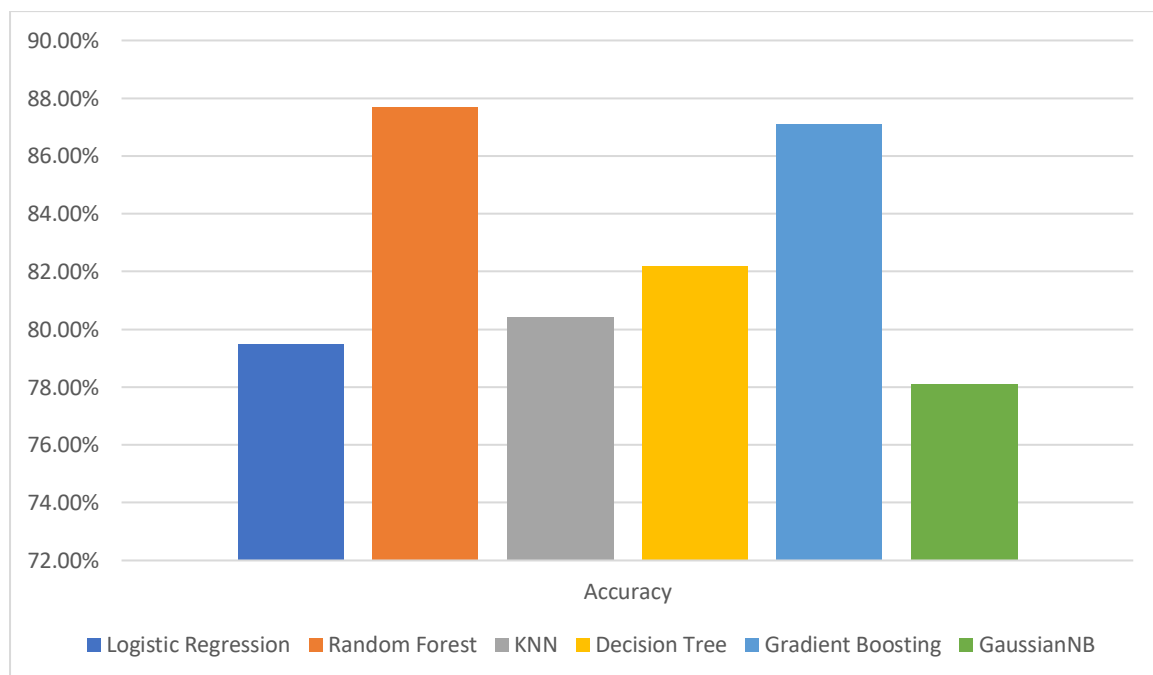


Figure 24 - Διάγραμμα αποτελεσμάτων Ακρίβειας 3

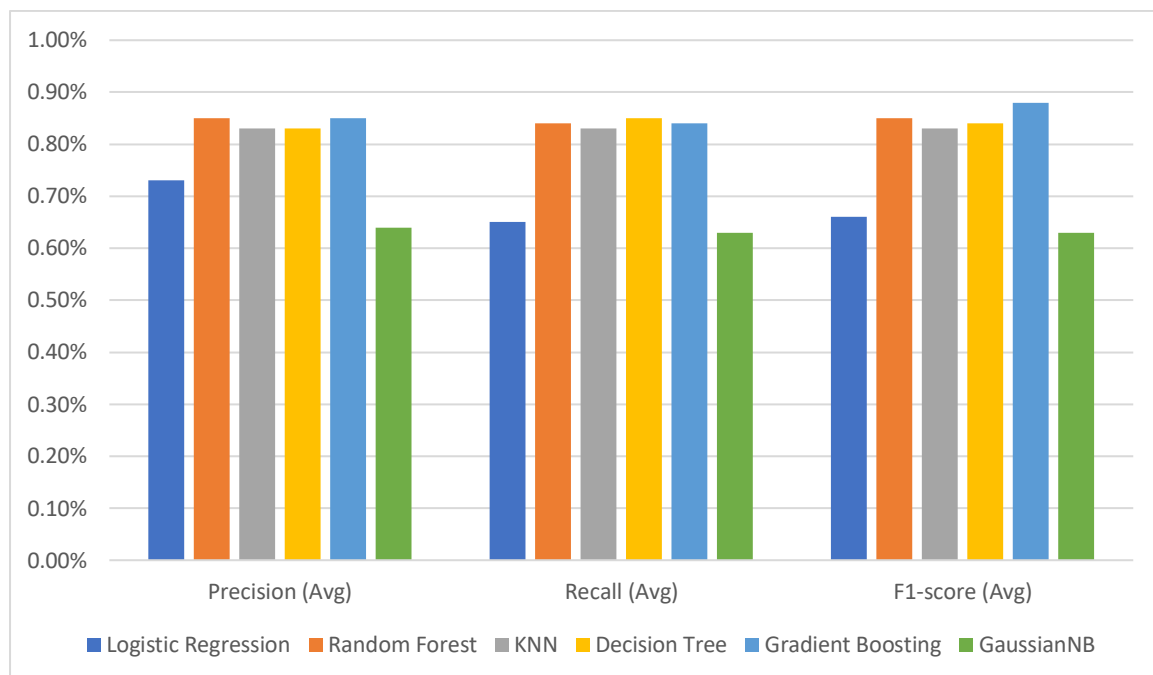


Figure 25 - Διάγραμμα αποτελεσμάτων Ακρίβειας 4

Ανάπτυξη Ιστοσελίδας

Η ιστοσελίδα B.C Compass αντιπροσωπεύει μια προηγμένη πλατφόρμα που σχεδιάστηκε ειδικά για τη χρήση από επαγγελματίες γιατρούς. Με επίκεντρο την εφαρμογή των προηγμένων μεθόδων πρόληψης και διάγνωσης του καρκίνου του μαστού, η ιστοσελίδα διευκολύνει τους γιατρούς στην αξιολόγηση του υγειονομικού κινδύνου των ασθενών τους. Το μοντέλο που χρησιμοποιείται εκπαιδεύτηκε με εξειδικευμένα σύνολα δεδομένων, που προέρχονται από ψηφιοποιημένες εικόνες αναρρόφησης λεπτής βελόνας (FNA) μάζας μαστού, που μπορούν να παρέχουν στους γιατρούς αξιόπιστα δεδομένα εισόδου για τη διαγνωστική διαδικασία.

Με την υποστήριξη του συστήματος WordPress, η ιστοσελίδα παρέχει μια φιλική περιβαλλοντική διεπαφή, καθιστώντας την χρηστική για τους επαγγελματίες του χώρου της υγείας. Ο συνδυασμός της προηγμένης τεχνολογίας μηχανικής μάθησης με την ευκολία της χρήσης συνιστά ένα ισχυρό εργαλείο για την ακριβή πρόληψη και διάγνωση του καρκίνου του μαστού, ενισχύοντας την ικανότητα των γιατρών να παρέχουν ατομική και προσαρμοσμένη φροντίδα στους ασθενείς τους.



Figure 26 - Διαδικασία ανάπτυξης ιστοσελίδας

Ακολουθώς, παρουσιάζονται διάφορα στιγμιότυπα οθόνης που αντιπροσωπεύουν τη λειτουργικότητα της ιστοσελίδας. Κάθε στιγμιότυπο προσφέρει λεπτομερή εικόνα της διεπαφής και των λειτουργιών της, παρέχοντας μια επιστημονική οπτική προσέγγιση για τη χρήση του εργαλείου. Αυτή η σειρά εικόνων προσφέρει μια συνολική εικόνα του λειτουργικού περιβάλλοντος της ιστοσελίδας, εμβαθύνοντας στις λεπτομέρειες που καθορίζουν την απόδοση και την αποτελεσματικότητά της.

Αρχική Οθόνη



Figure 27 - Αρχική οθόνη

Στην κεντρική σελίδα της ιστοσελίδας B.C Compass, όπως παρατηρείται και στο **Figure 27 - Αρχική οθόνη**, καλωσορίζεται ο χρήστης με το μήνυμα "Welcome to B.C Compass, we'll ensure you always get the best result." Αυτή η φιλόξενη και επαγγελματική καλωσόριση αντικατοπτρίζει την δέσμευση της εφαρμογής για υψηλή ποιότητα και ακρίβεια στα αποτελέσματα που παρέχει. Με ένα κομψό και διακριτικό σχεδιασμό, η σελίδα επιτρέπει εύκολη πλοήγηση και παρέχει στον χρήστη τη δυνατότητα να ξεκινήσει το τεστ με ένα απλό κλικ στο κουμπί "Try the Test". Αυτή η ευέλικτη διεπαφή ενθαρρύνει τον χρήστη να εξερευνήσει τις λειτουργίες της εφαρμογής και να έχει μια ευχάριστη εμπειρία πλοήγησης.

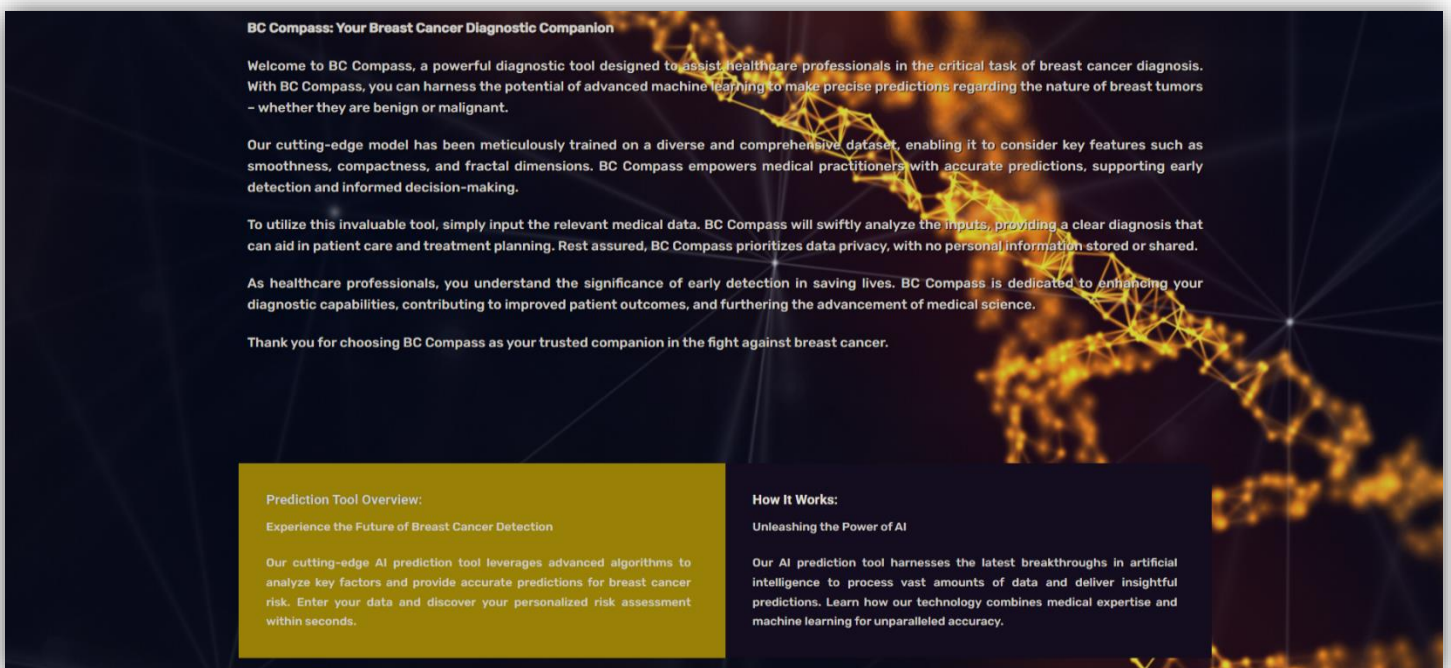


Figure 28 - Αρχική οθόνη 2

Στη συνέχεια της αρχικής οθόνης, όπως δείχνει το **Figure 28 - Αρχική οθόνη 2**, η ιστοσελίδα παρέχει μια σύντομη περιγραφή που προσφέρει την επισκόπηση του προγράμματος πρόβλεψης. Με λίγα λόγια, εξηγεί τον σκοπό της ιστοσελίδας, επισημαίνοντας τη σημασία του Prediction Tool. Ειδικότερα, περιγράφει τον τρόπο λειτουργίας του εργαλείου πρόβλεψης και πώς μπορεί να παρέχει προσωπικές εκτιμήσεις του κινδύνου εμφάνισης καρκίνου του μαστού. Αυτό το σύντομο αλλά ενημερωτικό εισαγωγικό κείμενο παρέχει στον χρήστη μια γρήγορη επισκόπηση των δυνατοτήτων της ιστοσελίδας και του τρόπου που μπορεί να επωφεληθεί από αυτήν.

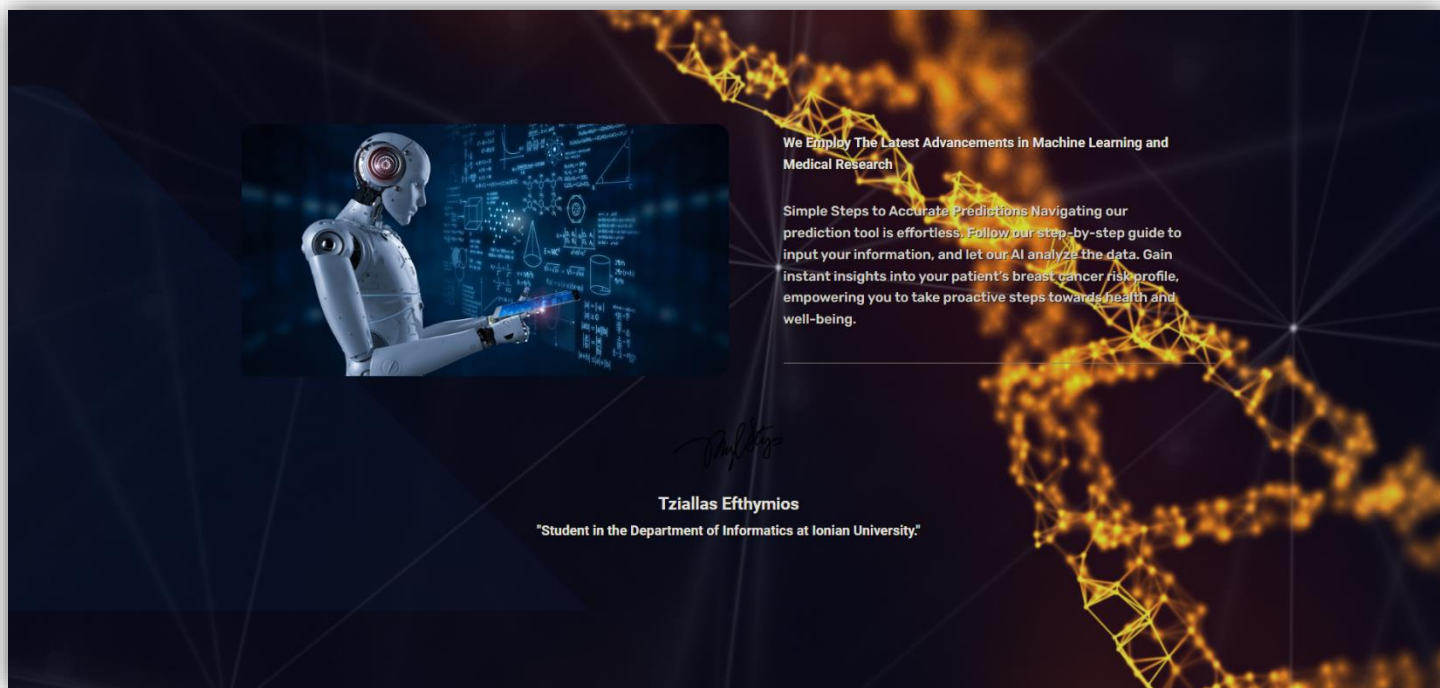


Figure 29 - Αρχική οθόνη 3

Στο **Figure 29 - Αρχική οθόνη 3** της ιστοσελίδας αναδεικνύεται η τεχνολογική πρωτοπορία και η εξειδίκευση της B.C Compass. Αναφέρει πως με την έμφαση στη χρήση των τελευταίων επιτευγμάτων στη Μηχανική Μάθηση και την Ιατρική Έρευνα, η ιστοσελίδα παρέχει έναν απλό και αποτελεσματικό τρόπο για τους ιατρούς να προβλέπουν τον κίνδυνο καρκίνου του μαστού για τους ασθενείς τους. Ο οδηγός βήμα προς βήμα καθιστά εύκολο τον χειρισμό του Prediction Tool, ενώ η είσοδος των πληροφοριών από τον ιατρό ενεργοποιεί το μοντέλο AI για την ανάλυση των δεδομένων. Οι άμεσες ενδείξεις για τον κίνδυνο του καρκίνου επιτρέπουν στους ιατρούς να λαμβάνουν προληπτικά μέτρα για την υγεία και την ευημερία των ασθενών.

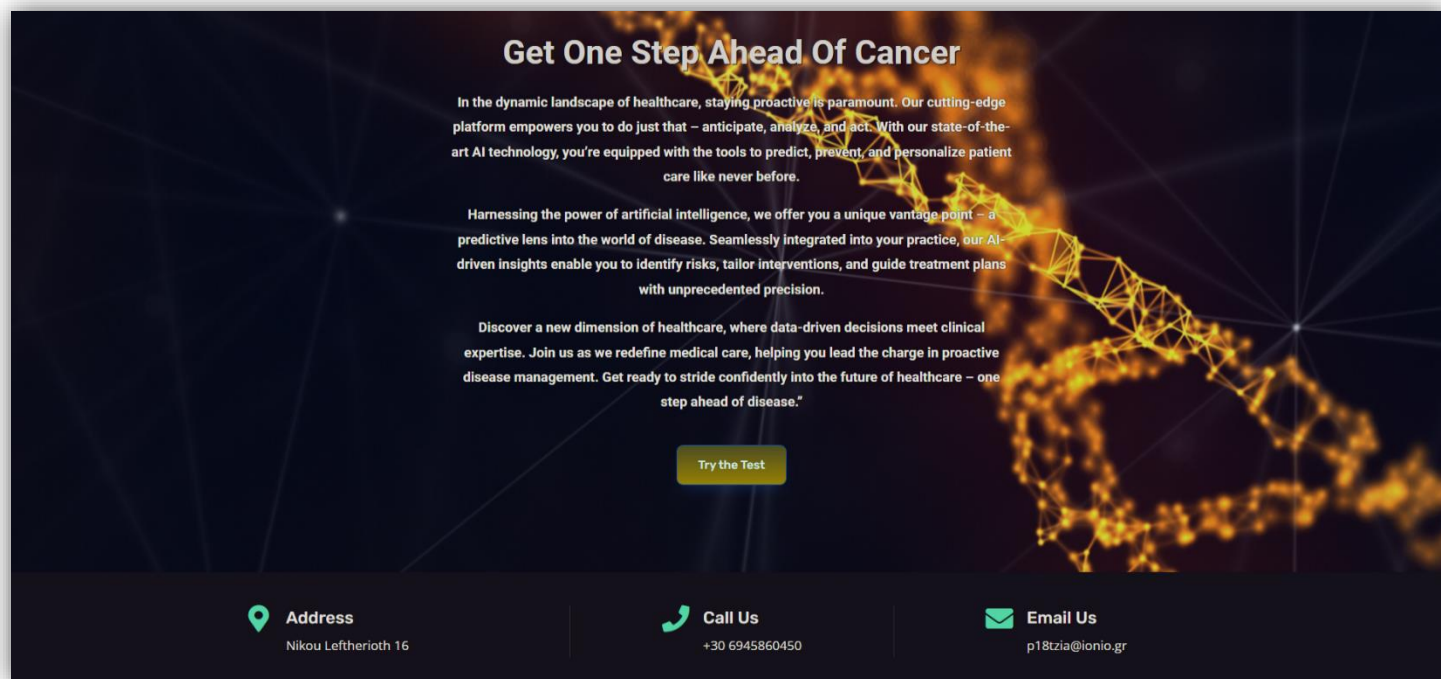


Figure 30 - Αρχική οθόνη 4

Η τελευταία φωτογραφία της αρχικής οθόνης, όπως φαίνεται και στο **Figure 30 - Αρχική οθόνη 4**, αναδεικνύει τη φιλοσοφία και τον στόχο της πλατφόρμας "B.C Compass Doctor". Με την επιγραφή "Get One Step Ahead Of Cancer", παρουσιάζεται η αποφασιστικότητα της πλατφόρμας να παρέχει εξειδικευμένα εργαλεία για την πρόληψη και την αντιμετώπιση του καρκίνου. Ο κείμενος επισημαίνει την αξία της προορατικότητας στον τομέα της υγείας και περιγράφει την πλατφόρμα ως κορυφαία στη χρήση τεχνολογίας τελευταίας γενιάς και τεχνητής νοημοσύνης.

Το κείμενο επικεντρώνεται στη δυνατότητα του χρήστη να προβλέπει, προλαμβάνει, και εξατομικεύει τη φροντίδα των ασθενών με τη χρήση προηγμένης τεχνολογίας τεχνητής νοημοσύνης. Επισημαίνεται η ουσιαστική προσθήκη αξίας στην κλινική πρακτική μέσω των αναλυτικών εργαλείων που παρέχει η πλατφόρμα. Το κείμενο καλεί τους επαγγελματίες της υγείας να ανακαλύψουν ένα νέο επίπεδο υγειονομικής φροντίδας, όπου οι αποφάσεις βασίζονται σε δεδομένα και συναντούν την κλινική εμπειρογνωμοσύνη. Συνοψίζοντας, προτρέπει τους χρήστες της, δηλαδή τους γιατρούς, να προετοιμαστούν να προχωρήσουν με αυτοπεποίθηση στο μέλλον της υγείας, ένα βήμα μπροστά από την ασθένεια.

Πρόβλεψη

The image shows a web application titled "BC Compass" with a background featuring a glowing DNA double helix and a network of nodes. The interface includes a section titled "Enter Feature Values" with seven input fields for the following features: Smoothness Mean, Fractal Dimension Mean, Compactness SE, Fractal Dimension SE, Smoothness Worst, Compactness Worst, and Fractal Dimension Worst. A "Predict" button is located below the input fields. At the bottom left, it says "Made with Streamlit".

Figure 31 - Οθόνη Πρόβλεψης

Στη φωτογραφία που παρουσιάζεται στην ιστοσελίδα, **Figure 31 - Οθόνη Πρόβλεψης**, εμφανίζονται τα πλαίσια για την εισαγωγή των επτά κρίσιμων δεδομένων για την αξιολόγηση του κινδύνου καρκίνου του μαστού. Η καταχώρηση των δεδομένων Smoothness Mean, Fractal Dimension Mean, Compactness Standard Error, Fractal Dimension Standard Error, Smoothness Worst, Compactness Worst, και Fractal Dimension Worst επιτρέπει στους χρήστες να παρέχουν τις απαραίτητες πληροφορίες που αφορούν την πυκνότητα, την διασπορά, και τις διαστάσεις των κυττάρων του μαστού.

Η διεπαφή είναι σχεδιασμένη για ευκολία χρήσης, προσφέροντας σαφή και ευανάγνωστα πεδία για την εισαγωγή των σημαντικών δεδομένων. Οι εκάστοτε χρήστες, δηλαδή οι γιατροί, μπορούν να ακολουθήσουν απλά βήματα για την αξιολόγηση του κινδύνου και τη λήψη σημαντικών αποφάσεων βασισμένων σε ακριβείς προβλέψεις του συστήματος.

The screenshot shows the 'BC Compass' web application interface. The title 'BC Compass' is at the top in yellow. Below it is the section 'Enter Feature Values'. There are seven input fields arranged in three rows: 'Smoothness Mean' (0.1134), 'Fractal Dimension Mean' (0.0696), 'Compactness SE' (0.01099), 'Fractal Dimension SE' (0.005953), 'Smoothness Worst' (0.1275), 'Compactness Worst' (0.09866), and 'Fractal Dimension Worst' (empty). A red 'Predict' button is below the inputs. A red error message box at the bottom says 'Please provide valid inputs for all features.' The background features a glowing DNA double helix and a network graph.

Figure 32 - Μήνυμα λάθους

Στην εικόνα, **Figure 32 - Μήνυμα λάθους**, προβάλλεται η συμπλήρωση των πεδίων δεδομένων, εκτός από ένα που παραμένει κενό. Κατά το πάτημα του κουμπιού "Predict" χωρίς πλήρη παροχή δεδομένων, εμφανίζεται το μήνυμα "Please provide valid inputs for all features" υπογραμμίζοντας την αναγκαιότητα πλήρους και έγκυρης εισόδου για την ακριβή πρόβλεψη. Εάν διαπιστωθεί χρήση μη αποδεκτών χαρακτήρων, το σύστημα ανταποκρίνεται με το ίδιο μήνυμα, ενισχύοντας τη σημασία της πλήρους, ακριβούς και έγκυρης εισόδου δεδομένων για να διασφαλιστεί η ακρίβεια των προβλέψεων του συστήματος.

The screenshot shows the 'BC Compass' web application interface with the same layout as Figure 32. All seven input fields are now filled with numerical values: 'Smoothness Mean' (0.1122), 'Fractal Dimension Mean' (0.0659), 'Compactness SE' (0.02648), 'Fractal Dimension SE' (0.003601), 'Smoothness Worst' (0.1851), 'Compactness Worst' (0.4061), and 'Fractal Dimension Worst' (0.1031). The red 'Predict' button is still present. Below the inputs, the section 'Diagnosis Prediction' shows a red box with the text 'Malignant'. The background features a glowing DNA double helix and a network graph.

Figure 33 - Πρόβλεψη Malignant

Στην εικόνα, **Figure 33 - Πρόβλεψη Malignant**, παρατηρείται η συμπλήρωση όλων των πεδίων δεδομένων, με το σύστημα να εμφανίζει το μήνυμα "Malignant" μετά το πάτημα του κουμπιού "Predict". Αυτή η απόκριση υποδηλώνει ότι το μοντέλο πρόβλεψης αναγνώρισε τα

παρεχόμενα δεδομένα ως σχετικά με μια κακοήγη κατάσταση. Η ένδειξη "Malignant" παρέχει στον χρήστη μια σαφή και άμεση ενημέρωση σχετικά με το αποτέλεσμα της πρόβλεψης, επιτρέποντάς του να λάβει άμεσα τα απαραίτητα μέτρα ή να συμβουλευτεί με άλλους ειδικούς υγείας.

The screenshot displays the 'BC Compass' web application interface. The background features a glowing DNA double helix and a network of nodes. The interface includes a section titled 'Enter Feature Values' with seven input fields for different features. Below these fields is a 'Predict' button. Underneath the button, the 'Diagnosis Prediction' section shows the result 'Benign' in a green box.

Feature	Value
Smoothness Mean	0.1134
Fractal Dimension Mean	0.0696
Compactness SE	0.01099
Fractal Dimension SE	0.005953
Smoothness Worst	0.1275
Compactness Worst	0.09866
Fractal Dimension Worst	0.0802

Predict

Diagnosis Prediction

Benign

Figure 34 - Πρόβλεψη Benign

Τέλος όπως φαίνεται και στο **Figure 34 - Πρόβλεψη Benign** που παρουσιάζεται, όλα τα πεδία έχουν συμπληρωθεί πλήρως. Με το πάτημα του κουμπιού "Predict," το μοντέλο αναγνωρίζει τα δεδομένα και εμφανίζει το αποτέλεσμα ως "Benign". Αυτή η αναγνώριση υποδηλώνει ότι, με βάση τα εισαγόμενα δεδομένα, ο αλγόριθμος προβλέπει μια καλοήγη κατάσταση. Η παροχή αυτού του αποτελέσματος στον γιατρό επιτρέπει μια γρήγορη και εύκολη αξιολόγηση, επιτρέποντάς του να λάβει άμεσες αποφάσεις ή να προχωρήσει σε περαιτέρω διαγνωστικές διαδικασίες με βάση την πρόβλεψη του μοντέλου.

Παραρτήματα

Παράρτημα 1^ο - Κώδικας εκπαίδευσης

Παρακάτω παρουσιάζεται ο κώδικας που χρησιμοποιήθηκε για την εύρεση του καλύτερου αλγορίθμου που αναφέρθηκε προηγουμένως, την εκπαίδευση του μοντέλου πρόβλεψης καρκίνου του μαστού, καθώς και την αξιολόγηση των διαφόρων αλγορίθμων μηχανικής μάθησης.

```
1 import pandas as pd
2 from sklearn.model_selection import train_test_split, cross_val_score
3 from sklearn.ensemble import RandomForestClassifier
4 from sklearn.metrics import classification_report
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.ensemble import GradientBoostingClassifier
7 from sklearn.neighbors import KNeighborsClassifier
8 from sklearn.tree import DecisionTreeClassifier
9 from sklearn.naive_bayes import QuadraticDiscriminantAnalysis
10 from joblib import dump
11
12 class CancerPredictionApp:
13     def run(self):
14         df =
15         pd.read_csv(r"C:\Users\efthi\OneDrive\Υπολογιστής\Σχολή\Εαρινό
16         Εξάμηνο 2023-2024\Πτυχιακή\SmallData.csv")
17
18         selected_features = ['Glucose', 'Resistin', 'Insulin', 'MCP.1']
19
20
21         X = df[selected_features]
22         y = df['Classification']
23
24         X_train, X_test, y_train, y_test = train_test_split(X, y,
25         test_size=0.1, random_state=42)
26
27         classifier = LogisticRegression()
28
29         classifier.fit(X_train, y_train)
30
31         y_pred = classifier.predict(X_test)
32
33         accuracies = cross_val_score(classifier, X_train, y_train, cv=10)
34
35         average_accuracy = accuracies.mean()
36         print(f"Average Accuracy with selected features: {average_accuracy *
37         100:.2f}%")
38
39         accuracy = (y_pred == y_test).mean()
40         print(f"Accuracy on Test Set: {accuracy * 100:.2f}%")
41
42         report = classification_report(y_test, y_pred)
43         print("Classification Report on Test Set:")
44         print(report)
45         print(classifier)
46
47         #Αποθήκευση Μοντέλου σε joblib
```

```
48 classifier.fit(X_train, y_train)
49 dump(classifier, 'BigModel5.joblib')
50 print("Model has been trained and saved as 'BigModel5.joblib'.")
51
52 if __name__ == "__main__":
53     app = CancerPredictionApp()
54     app.run()
```

Αρχικά, εισάγονται οι απαραίτητες βιβλιοθήκες, μεταξύ των οποίων περιλαμβάνονται η pandas για τη διαχείριση δεδομένων, η scikit-learn για την εκπαίδευση και αξιολόγηση μοντέλων, καθώς και η joblib για την αποθήκευση εκπαιδευμένων μοντέλων.

Στη συνέχεια, πραγματοποιείται η φόρτωση των δεδομένων από το Dataset Breast Cancer Coimbra σε μορφή CSV. Επιλέγονται συγκεκριμένα χαρακτηριστικά, όπως 'Glucose', 'Resistin', 'Insulin', και 'MCP.1', τα οποία θα χρησιμοποιηθούν για την εκπαίδευση του μοντέλου.

Οι δεδομένα διαχωρίζονται σε σύνολα εκπαίδευσης (X_train, y_train) και δοκιμής (X_test, y_test) μέσω της συνάρτησης train_test_split της scikit-learn. Στη συνέχεια, επιλέγεται ο αλγόριθμος Logistic Regression ως μοντέλο ταξινόμησης και πραγματοποιείται η εκπαίδευσή του με χρήση του συνόλου εκπαίδευσης.

Η απόδοση του μοντέλου αξιολογείται μέσω της χρήσης cross-validation για τον υπολογισμό του μέσου όρου ακρίβειας. Επιπλέον, υπολογίζεται η ακρίβεια στο σύνολο δοκιμής και εκτυπώνεται.

Τέλος, το εκπαιδευμένο μοντέλο αποθηκεύεται στο αρχείο 'BigModel5.joblib' για χρήση στην δημιουργία της εφαρμογής.

Παράρτημα 2^ο - Κώδικας ανάπτυξης εφαρμογής

Παρακάτω παρουσιάζεται ο κώδικας Java που δημιουργήθηκε για την ανάπτυξη της εφαρμογής σε περιβάλλον Android Studio. Η εφαρμογή προορίζεται για χρήση σε κινητές συσκευές Android και χρησιμοποιεί τη βιβλιοθήκη TensorFlow Lite για την πρόβλεψη κινδύνου εμφάνισης καρκίνου του μαστού.

Σημαντικό είναι ότι το αρχικό μοντέλο, αποθηκευμένο σε μορφή joblib, μετατράπηκε σε μοντέλο TensorFlow προτού ενσωματωθεί στην εφαρμογή. Αυτή η διαδικασία προσφέρει συμβατότητα με το TensorFlow Lite και εξασφαλίζει αποτελεσματική εκτέλεση σε κινητές συσκευές.

```
1 package com.example.myapplication;
2
3 import androidx.appcompat.app.AppCompatActivity;
4 import android.content.res.AssetFileDescriptor;
5 import android.os.Bundle;
6 import android.widget.Button;
7 import android.widget.EditText;
8 import android.widget.TextView;
9 import org.tensorflow.lite.Interpreter;
10 import java.io.FileInputStream;
11 import java.nio.MappedByteBuffer;
12 import java.nio.channels.FileChannel;
13 import android.widget.ImageView;
14 import android.content.Intent;
15 import android.app.AlertDialog;
16 import android.content.DialogInterface;
17 import android.view.View;
18
19
20
21
22 public class MainActivity extends AppCompatActivity {
23
24     private Interpreter tflite;
25     private ImageView infoIcon1;
26
27
28
29     @Override
30     protected void onCreate(Bundle savedInstanceState) {
31         super.onCreate(savedInstanceState);
32         setContentView(R.layout.activity_main);
33
34
35         // Initialize the TFLite interpreter
36         tflite = new Interpreter(loadModelFile(), new
37             Interpreter.Options());
38
39         final EditText editText1 = findViewById(R.id.editText1);
40         final EditText editText2 = findViewById(R.id.editText2);
41         final EditText editText3 = findViewById(R.id.editText3);
42         final EditText editText4 = findViewById(R.id.editText4);
```

```

43     Button predictButton = findViewById(R.id.predictButton);
44     final TextView resultTextView =
45     findViewById(R.id.resultTextView);
46
47     ImageView infoIcon1 = findViewById(R.id.infoIcon1);
48     ImageView infoIcon2 = findViewById(R.id.infoIcon2);
49     ImageView infoIcon3 = findViewById(R.id.infoIcon3);
50     ImageView infoIcon4 = findViewById(R.id.infoIcon4);
51
52     infoIcon1.setOnClickListener(new View.OnClickListener() {
53         @Override
54         public void onClick(View view) {
55             String featureInfo = "• Blood glucose (blood sugar)
56             indicates your body's energy balance.\n" +
57             "    • In Europe, it's measured in mmol/L
58             (millimoles per liter).\n" +
59             "    • Healthy fasting levels: 3.9 to 5.6
60             mmol/L.\n" +
61             "    • Maintaining optimal levels is vital for
62             health.";
63             showInformationDialog("Glucose", featureInfo);
64         }
65     });
66
67     infoIcon2.setOnClickListener(new View.OnClickListener() {
68         @Override
69         public void onClick(View view) {
70             String featureInfo = "• Insulin is a hormone that regula
71             blood sugar levels.\n" +
72             "    • It helps your body use glucose (sugar) for
73             energy.\n" +
74             "    • Proper insulin function is essential for
75             managing diabetes.";
76             showInformationDialog("Insulin", featureInfo);
77         }
78     });
79
80     infoIcon3.setOnClickListener(new View.OnClickListener() {
81         @Override
82         public void onClick(View view) {
83             String featureInfo = "• Resistin is a protein hormone
84             produced by fat cells (adipocytes).\n" +
85             "    • It plays a role in insulin resistance and
86             inflammation.\n" +
87             "    • High resistin levels may be linked to
88             obesity and diabetes.";
89             showInformationDialog("Resistin", featureInfo);
90         }
91     });
92
93     infoIcon4.setOnClickListener(new View.OnClickListener() {
94         @Override
95         public void onClick(View view) {
96             String featureInfo = "• MCP-1 is a chemokine protein
97             involved in immune response.\n" +
98             "    • It attracts immune cells, like monocytes,
99             to sites of inflammation.\n" +
100            "    • Elevated MCP-1 levels may indicate
101            inflammation-related conditions.\n";
102            showInformationDialog("MCP.1", featureInfo);

```

```

103         }
104     });
105
106
107
108
109     predictButton.setOnClickListener(v -> {
110         String input1Text = editText1.getText().toString();
111         String input2Text = editText2.getText().toString();
112         String input3Text = editText3.getText().toString();
113         String input4Text = editText4.getText().toString();
114
115         if (input1Text.isEmpty() || input2Text.isEmpty() ||
116             input3Text.isEmpty() || input4Text.isEmpty()) {
117             resultTextView.setText("Please provide values for all
118                 inputs before predicting");
119             return;
120         }
121
122         float input1 = Float.parseFloat(input1Text);
123         float input2 = Float.parseFloat(input2Text);
124         float input3 = Float.parseFloat(input3Text);
125         float input4 = Float.parseFloat(input4Text);
126
127         float[] inputArray = {input1, input2, input3, input4};
128         float[][] outputArray = new float[1][1];
129
130         tflite.run(inputArray, outputArray);
131
132         float rawPrediction = outputArray[0][0];
133         String predictionLabel;
134
135         if (rawPrediction >= 0 && rawPrediction < 0.2) {
136             predictionLabel = "You have very low chances of
137                 developing Breast Cancer in the near future";
138         } else if (rawPrediction >= 0.2 && rawPrediction < 0.4) {
139             predictionLabel = "You have low chances of developing
140                 Breast Cancer in the near future";
141         } else if (rawPrediction >= 0.4 && rawPrediction < 0.7) {
142             predictionLabel = "You have high chances of developing
143                 Breast Cancer in the near future";
144         } else {
145             predictionLabel = "You have very high chances of
146                 developing Breast Cancer in the near future";
147         }
148
149
150         resultTextView.setText(getString(R.string.
151             prediction_result_with_label, rawPrediction,
152             predictionLabel));
153     });
154 }
155
156 private MappedByteBuffer loadModelFile() {
157     try (AssetFileDescriptor fileDescriptor =
158         getAssets().openFd("breast_cancer_model.tflite");
159         FileInputStream inputStream = new
160             FileInputStream(fileDescriptor.getFileDescriptor())) {
161
162         FileChannel fileChannel = inputStream.getChannel();

```

```

163         return fileChannel.map(FileChannel.MapMode.READ_ONLY,
164             fileDescriptor.getStartOffset(), fileDescriptor.getLength());
165     }
166     } catch (Exception e) {
167         e.printStackTrace();
168         return null;
169     }
170 }
171
172 private void showInformationDialog(String featureName, String
173     featureInfo) {
174     AlertDialog.Builder builder = new AlertDialog.Builder(this);
175     builder.setTitle(featureName);
176     builder.setMessage(featureInfo);
177     builder.setPositiveButton("OK", new
178         DialogInterface.OnClickListener() {
179         public void onClick(DialogInterface dialog, int id) {
180             dialog.dismiss();
181         }
182     });
183     AlertDialog dialog = builder.create();
184     dialog.show();
185 }
186 }

```

Στη μέθοδο onCreate, που καλείται κατά την εκκίνηση της εφαρμογής, πραγματοποιείται η αρχικοποίηση του χρησιμοποιούμενου μοντέλου TensorFlow Lite με τη χρήση του Interpreter. Έπειτα, ορίζονται τα στοιχεία του UI όπως EditText για τα χαρακτηριστικά εισόδου, ένα κουμπί πρόβλεψης και ένα πεδίο κειμένου για την εμφάνιση των αποτελεσμάτων.

Επίσης, δημιουργούνται εικονίδια πληροφοριών (infoIcon1, infoIcon2, infoIcon3, infoIcon4) που προσφέρουν επεξηγηματικές πληροφορίες για κάθε χαρακτηριστικό με τη χρήση διαλόγων πληροφοριών.

Όσον αφορά τη λειτουργία πρόβλεψης, όταν ο χρήστης πατάει το κουμπί "Πρόβλεψη", οι τιμές εισόδου λαμβάνονται από τα πεδία κειμένου, ελέγχονται για συμπληρωμένα πεδία, μετατρέπονται σε αριθμούς και προωθούνται στο μοντέλο TensorFlow Lite για την εκτέλεση της πρόβλεψης. Τα αποτελέσματα παρουσιάζονται στο χρήστη με βάση τις κατηγορίες πιθανότητας.

Στο κομμάτι της διαχείρισης του μοντέλου TensorFlow Lite, η μέθοδος loadModelFile() υπεύθυνη για το φόρτωμα του μοντέλου από τον πόρο assets. Τέλος, η μέθοδος showInformationDialog() δημιουργεί διάλογο πληροφοριών για κάθε χαρακτηριστικό.

Ο κώδικας ακολουθεί βέλτιστες πρακτικές για τη δημιουργία της Android εφαρμογής και προσφέρει ένα συστηματικό πλαίσιο για την χρήση μοντέλων μηχανικής μάθησης στο πεδίο της υγείας.

Παράρτημα 3^ο - Κώδικας εκπαίδευσης

Παρακάτω παρουσιάζεται ο κώδικας που χρησιμοποιήθηκε για την εύρεση του καλύτερου αλγορίθμου, την εκπαίδευση του μοντέλου πρόβλεψης καρκίνου του μαστού, καθώς και την αξιολόγηση των διαφόρων αλγορίθμων μηχανικής μάθησης.

```
1 import pandas as pd
2 from sklearn.model_selection import train_test_split, cross_val_score
3 from sklearn.ensemble import RandomForestClassifier
4 from sklearn.metrics import classification_report
5 from sklearn.linear_model import LogisticRegression
6 from sklearn.ensemble import GradientBoostingClassifier
7 from sklearn.neighbors import KNeighborsClassifier
8 from sklearn.tree import DecisionTreeClassifier
9 from sklearn.naive_bayes import GaussianNB
10 from joblib import dump
11
12 class CancerPredictionApp:
13     def run(self):
14         df =
15         pd.read_csv(r"C:\Users\efthi\OneDrive\Υπολογιστής\Σχολή\Εαρινό
16         Εξάμηνο 2023-2024\Πτυχιακή\BigDataa.csv")
17
18         selected_features = ['smoothness_mean',
19                             'fractal_dimension_mean', 'compactness_se',
20                             'fractal_dimension_se', 'smoothness_worst', 'compactness_worst',
21                             'fractal_dimension_worst']
22
23         X = df[selected_features]
24         y = df['diagnosis']
25
26         X_train, X_test, y_train, y_test = train_test_split(X, y,
27         test_size=0.1, random_state=42)
28
29         classifier = RandomForestClassifier()
30         classifier.fit(X_train, y_train)
31
32         y_pred = classifier.predict(X_test)
33
34         accuracies = cross_val_score(classifier, X_train, y_train,
35         cv=10)
36
37         average_accuracy = accuracies.mean()
38         print(f"Average Accuracy with selected features:
39         {average_accuracy * 100:.2f}%")
40
41         accuracy = (y_pred == y_test).mean()
42         print(f"Accuracy on Test Set: {accuracy * 100:.2f}%")
43
44         report = classification_report(y_test, y_pred)
45         print("Classification Report on Test Set:")
46         print(report)
47         print(classifier)
```

```

48
49     #Αποθήκευση Μοντέλου σε joblib
50     classifier.fit(X_train, y_train)
51     dump(classifier, 'BigModel6.joblib')
52     print("Model has been trained and saved as
53           'BigModel6.joblib'.")
54
55 if __name__ == "__main__":
56     app = CancerPredictionApp()
57     app.run()

```

Αρχικά εισάγονται οι απαραίτητες βιβλιοθήκες, περιλαμβάνοντας τη pandas για τη διαχείριση των δεδομένων και τη scikit-learn για την εκπαίδευση και αξιολόγηση μοντέλων. Επιπλέον, χρησιμοποιείται η βιβλιοθήκη joblib για την αποθήκευση εκπαιδευμένων μοντέλων.

Τα δεδομένα Breast Cancer Wisconsin (Diagnostic) φορτώνονται σε μορφή CSV, περιλαμβάνοντας παράλληλα την επιλογή συγκεκριμένων χαρακτηριστικών, όπως 'smoothness_mean', 'fractal_dimension_mean', 'compactness_se', 'fractal_dimension_se', 'smoothness_worst', 'compactness_worst', 'fractal_dimension_worst', για την εκπαίδευση του μοντέλου.

Τα δεδομένα διαχωρίζονται σε σύνολα εκπαίδευσης (X_train, y_train) και δοκιμής (X_test, y_test) χρησιμοποιώντας τη συνάρτηση train_test_split. Ο ταξινομητής Random Forest Classifier επιλέγεται για την εκπαίδευση, ενώ η απόδοση του μοντέλου αξιολογείται μέσω μετρικών όπως η ακρίβεια και ο μέσος όρος ακρίβειας με χρήση cross-validation.

Τέλος, το εκπαιδευμένο μοντέλο αποθηκεύεται σε αρχείο με την ονομασία 'BigModel6.joblib', επιτρέποντας έτσι τη χρήση του στην ανάπτυξη της ιστοσελίδας.

Παράρτημα 4^ο- Κώδικας ανάπτυξης ιστοσελίδας

Παρακάτω παρουσιάζεται ο κώδικας, γραμμένος σε Python, χρησιμοποιώντας τη βιβλιοθήκη Streamlit, για την ανάπτυξη της ιστοσελίδας.

```
1 import streamlit as st
2 import joblib
3 import pandas as pd
4
5 model_path = r'C:\Users\efthi\OneDrive\Υπολογιστής\Σχολή\Ιστοσελίδα\
6             PythonScripts\BigModel6.joblib'
7 classifier = joblib.load(model_path)
8
9 def main():
10     # Set page title and favicon
11     st.set_page_config(page_title="BC Compass", page_icon="")
12
13     # Create a header with a custom background color and title
14
15
16
17     st.markdown(
18         f"""
19         <style>
20         .stApp {{
21             background: url("https://www.zipitwireless.com/hubfs/connectivity-1-1.
22             background-size: cover
23         }}
24
25
26         </style>
27         """,
28         unsafe_allow_html=True
29     )
30
31     st.markdown(
32         "<h1 class='header' style='color: #DAA520;'>BC Compass</h1>",
33         unsafe_allow_html=True
34     )
35
36     # Create a container for feature input
37     st.subheader("Enter Feature Values")
38     feature_cols = st.columns(3)
39
40     with feature_cols[0]:
41         smoothness_mean = st.text_input("Smoothness Mean", "")
42
43     with feature_cols[1]:
44         fractal_dimension_mean = st.text_input("Fractal Dimension Mean", "")
45
46     with feature_cols[2]:
47         compactness_se = st.text_input("Compactness SE", "")
48
49     with feature_cols[0]:
```

```

50     fractal_dimension_se = st.text_input("Fractal Dimension SE", "")
51
52     with feature_cols[1]:
53         smoothness_worst = st.text_input("Smoothness Worst", "")
54
55     with feature_cols[2]:
56         compactness_worst = st.text_input("Compactness Worst", "")
57
58     with feature_cols[0]:
59         fractal_dimension_worst = st.text_input("Fractal Dimension Worst", "")
60
61     # Create a button for prediction
62     if st.button("Predict"):
63         # Check if any of the feature values is empty
64         invalid_inputs = (
65             not smoothness_mean.replace(".", "").isdigit() or
66             not fractal_dimension_mean.replace(".", "").isdigit() or
67             not compactness_se.replace(".", "").isdigit() or
68             not fractal_dimension_se.replace(".", "").isdigit() or
69             not smoothness_worst.replace(".", "").isdigit() or
70             not compactness_worst.replace(".", "").isdigit() or
71             not fractal_dimension_worst.replace(".", "").isdigit()
72         )
73
74         if invalid_inputs:
75             st.error("Please provide valid inputs for all features.")
76         else:
77             # Convert the input to numbers (floats)
78             smoothness_mean = float(smoothness_mean)
79             fractal_dimension_mean = float(fractal_dimension_mean)
80             compactness_se = float(compactness_se)
81             fractal_dimension_se = float(fractal_dimension_se)
82             smoothness_worst = float(smoothness_worst)
83             compactness_worst = float(compactness_worst)
84             fractal_dimension_worst = float(fractal_dimension_worst)
85
86             # Create a DataFrame with the user's input
87             user_input = pd.DataFrame({
88                 'smoothness_mean': [smoothness_mean],
89                 'fractal_dimension_mean': [fractal_dimension_mean],
90                 'compactness_se': [compactness_se],
91                 'fractal_dimension_se': [fractal_dimension_se],
92                 'smoothness_worst': [smoothness_worst],
93                 'compactness_worst': [compactness_worst],
94                 'fractal_dimension_worst': [fractal_dimension_worst]
95             })
96
97             # Make predictions
98             prediction = classifier.predict(user_input)
99
100            # Display the prediction
101            st.subheader("Diagnosis Prediction" )
102            if prediction[0] == 1:
103                st.error("Malignant")
104            elif prediction[0] == 2:
105                st.success("Benign")
106
107    # Run the app
108    if __name__ == "__main__":
109        main()

```

Βασικό στοιχείο είναι πως ενώ το υπόλοιπο UI της εφαρμογής έχει δημιουργηθεί μέσω του WordPress σε ένα τοπικό περιβάλλον, εδώ γίνεται χρήση της βιβλιοθήκης Streamlit για την δημιουργία μιας διαδικτυακής διεπαφής για την πρόβλεψη με την προσθήκη του σχετικού κώδικα.

Αυτή η προσέγγιση επιτρέπει την ολοκλήρωση του μοντέλου μηχανικής μάθησης στην υπάρχουσα ιστοσελίδα WordPress. Ο χρήστης μπορεί να εισάγει τα χαρακτηριστικά, να πατήσει το κουμπί "Πρόβλεψη", και να λάβει τη διάγνωση χωρίς να χρειάζεται να πλοηγηθεί σε διαφορετικά περιβάλλοντα.

Με αυτόν τον τρόπο, επιτυγχάνεται μια ολοκληρωμένη λειτουργικότητα πρόβλεψης καρκίνου στην υπάρχουσα ιστοσελίδα WordPress, προσφέροντας έναν ομαλό και ομοιογενή χώρο για τους χρήστες.

Στην αρχή, γίνεται η εισαγωγή των απαραίτητων βιβλιοθηκών όπως το Streamlit, το joblib για το φορτωμένο του μοντέλου, και το pandas για τη διαχείριση των δεδομένων. Στη συνέχεια, καθορίζεται το μονοπάτι (path) προς το αρχείο που περιέχει το εκπαιδευμένο μοντέλο, και το μοντέλο φορτώνεται με τη χρήση του joblib. Η διεπαφή της εφαρμογής δημιουργείται με το Streamlit, περιλαμβάνοντας έναν τίτλο ("BC Compass") και πεδία εισαγωγής για τα χαρακτηριστικά του μοντέλου. Για κάθε χαρακτηριστικό, υπάρχει ένα πεδίο εισαγωγής (text_input) που επιτρέπει στον χρήστη να εισάγει τιμές. Επίσης υπάρχει ένα κουμπί "Predict" που, όταν πατηθεί, πραγματοποιεί έλεγχο εγκυρότητας των εισαγμένων τιμών και στη συνέχεια προβαίνει σε πρόβλεψη χρησιμοποιώντας το φορτωμένο μοντέλο. Το αποτέλεσμα της πρόβλεψης (Malignant ή Benign) εμφανίζεται στην διεπαφή, χρησιμοποιώντας διάφορα χρώματα για τη διακριτικότητα.

Συμπεράσματα

Ο σκοπός της εργασίας ήταν πολύπλευρος, εστιάζοντας ταυτόχρονα στην ενημέρωση του κοινού και την υποστήριξη των ιατρών, με την ανάπτυξη μιας ολοκληρωμένης πλατφόρμας για την πρόβλεψη του καρκίνου του μαστού, στηριζόμενη σε τεχνητή νοημοσύνη. Η δημιουργία της εφαρμογής Android B.C Compass Patient και της ιστοσελίδας B.C Compass, σχεδιάστηκε για να καλύψει δύο κύριους στόχους:

Πρώτον, η παροχή ενός εργαλείου πρόβλεψης καρκίνου του μαστού για το ευρύ κοινό, προωθώντας την επίγνωση και την πρόληψη της νόσου. Με την απλότητα της διαδικασίας εισόδου δεδομένων και την άμεση προβολή των προβλέψεων, οι χρήστες μπορούν να αξιολογήσουν τον κίνδυνο πιθανού καρκίνου του μαστού.

Δεύτερον, η δημιουργία μιας προηγμένης πλατφόρμας για επαγγελματίες της υγείας, οι οποίοι έχουν πρόσβαση σε προσαρμοσμένες πληροφορίες και προβλέψεις για τους ασθενείς τους. Η συνεργασία της κλινικής εμπειρίας με την ισχύ της τεχνητής νοημοσύνης βοηθά στην καλύτερη κατανόηση του κινδύνου, παρέχοντας προσαρμοσμένες θεραπευτικές προσεγγίσεις και βοηθώντας στην εξέλιξη της περίθαλψης.

Με τη συνολική προσέγγιση του έργου, επιδιώκεται να δημιουργηθεί ένα ολοκληρωμένο σύστημα που συνδυάζει τη δύναμη της τεχνητής νοημοσύνης με την ανθρώπινη εμπειρία, προωθώντας την πρόληψη, την προσωπική περίθαλψη και την κλινική ακρίβεια. Προβλέποντας το μέλλον, υπάρχουν πολλές προοπτικές βελτιώσεις που μπορούν να ενισχύσουν την αποτελεσματικότητα και τη χρησιμότητα της πλατφόρμας.

Η διαδικασία της βελτίωσης της πλατφόρμας B.C Compass είναι συνεχής και προβλέπεται ότι θα περιλαμβάνει εκτενή πειράματα και εξελίξεις σε διάφορους τομείς. Πρώτον, προγραμματίζονται περαιτέρω πειράματα με ποικίλους αλγορίθμους μηχανικής μάθησης, με στόχο την επιλογή του βέλτιστου μοντέλου για τις προβλέψεις όσον αφορά την ακρίβεια. Επίσης η πραγματοποίηση πειραμάτων με περισσότερα και πιο ποικίλα σύνολα δεδομένων θα μπορεί να συμβάλει στην πιο ολοκληρωμένη κατανόηση των παραμέτρων πρόβλεψης αλλά και στην καλύτερη εκπαίδευση του μοντέλου.

Επιπρόσθετα, σχεδιάζεται η ενίσχυση της πλατφόρμας με επιλογές περισσότερων μοντέλων ανάλογα με τον τύπο της ασθένειας που εξετάζεται. Αυτό θα επιτρέψει στους χρήστες, είτε είναι ασθενείς είτε γιατροί, να προσαρμόσουν τις αναζητήσεις τους με βάση τις συγκεκριμένες τους ανάγκες και ενδιαφέροντα. Η επέκταση των περιοχών έρευνας σε προγνώσεις για άλλες ασθένειες και η εφαρμογή προηγμένων τεχνικών επεξεργασίας εικόνας θα αποτελέσουν μέρος της συνεχούς προσπάθειας για επίτευξη ακόμη υψηλότερης ακρίβειας και χρηστικότητας. Αυτές οι βελτιώσεις θα οδηγήσουν σε μια πλήρη, εξελιγμένη και προσαρμόσιμη πλατφόρμα, εξυπηρετώντας καλύτερα τις ανάγκες των χρηστών και ενισχύοντας τον τομέα της υγείας.

Αναφορές

1. Ahmad, L.G., Eshlaghy, A.T., Poorebrahimi, A., Ebrahimi, M. and Razavi, A.R., 2013. Using three machine learning techniques for predicting breast cancer recurrence. *J Health Med Inform*, 4(124), p.3.
2. Vaka, A.R., Soni, B. and Reddy, S., 2020. Breast cancer detection by leveraging Machine Learning. *Ict Express*, 6(4), pp.320-324.
3. Borges, L.R., 1989. Analysis of the wisconsin breast cancer dataset and machine learning for breast cancer detection. *Group*, 1(369), pp.15-19.
4. Li, Y. and Chen, Z., 2018. Performance evaluation of machine learning methods for breast cancer prediction. *Appl Comput Math*, 7(4), pp.212-216.
5. Naji, M.A., El Filali, S., Aarika, K., Benlahmar, E.H., Abdelouhahid, R.A. and Debauche, O., 2021. Machine learning algorithms for breast cancer prediction and diagnosis. *Procedia Computer Science*, 191, pp.487-492.
6. Singh, G., 2020. Breast cancer prediction using machine learning. *Int. J. Sci. Res. Comput. Sci., Eng. Inf. Technol.*, 8(4), pp.278-284.
7. Shravya, C., Pravalika, K. and Subhani, S., 2019. Prediction of breast cancer using supervised machine learning techniques. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(6), pp.1106-1110.
8. Asri, H., Mousannif, H., Al Moatassime, H. and Noel, T., 2016. Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, 83, pp.1064-1069.
9. Sivapriya, J., Kumar, A., Sai, S.S. and Sriram, S., 2019. Breast cancer prediction using machine learning. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(4), pp.4879-4881.
10. Potdar, K. and Kinnerkar, R., 2016. A comparative study of machine learning algorithms applied to predictive breast cancer data. *International Journal of Science and Research*, 5(9), pp.1550-1553.
11. www.breastcancer.org. (n.d.). About Breast Cancer. [online] Available at: <https://www.breastcancer.org/about-breast-cancer>.
12. www.breastcancer.org. (n.d.). *Signs and Symptoms of Breast Cancer*. [online] Available at: <https://www.breastcancer.org/signs-symptoms>.
13. www.cancer.gov. (2020). *Breast Cancer Treatment (Adult) (PDQ®)–Patient Version - National Cancer Institute*. [online] Available at: <https://www.cancer.gov/types/breast/patient/breast-treatment-pdq>.
14. Cancer.net (2019). *Breast Cancer - Statistics*. [online] Cancer.net. Available at: <https://www.cancer.net/cancer-types/breast-cancer/statistics>.