



ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ  
ΙΑΤΡΙΚΗ ΣΧΟΛΗ, ΕΡΓΑΣΤΗΡΙΟ ΙΑΤΡΙΚΗΣ ΦΥΣΙΚΗΣ  
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ – ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ  
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΙΑΤΡΙΚΗ ΦΥΣΙΚΗ -  
ΑΚΤΙΝΟΦΥΣΙΚΗ

## ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Πρόβλεψη δόσης ακτινοβολίας ασθενών που υποβλήθηκαν σε Υπολογιστική Τομογραφία Κωνικής Δέσμης με χρήση τεχνητής νοημοσύνης  
Τσιρώνη Φερενίκη

A.M.20201368

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ:

ΚΑΘΗΓΗΤΗΣ ΔΑΜΗΛΑΚΗΣ Ι. (επιβλέπων)

ΚΑΘΗΓΗΤΗΣ ΜΑΡΗΣ Θ.

ΚΑΘΗΓΗΤΗΣ ΠΕΡΙΣΥΝΑΚΗΣ Κ.

ΗΡΑΚΛΕΙΟ, 2021-2022

## ΕΥΧΑΡΙΣΤΙΕΣ

Προτού ξεκινήσει η εργασία αυτή, οφείλω να ευχαριστήσω όλους εκείνους που συνέβαλαν σε αυτή την προσπάθεια. Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Ιωάννη Δαμηλάκη για τις πολύτιμες συμβουλές του και την καθοδήγησή του σε κάθε βήμα της προσπάθειας αυτής. Ήταν πραγματικά τιμή μου που με εμπιστεύθηκε αναθέτοντάς μου την παρούσα διπλωματική και κάνοντάς με μέλος της ομάδας του.

Το ευχαριστώ στον κ. Μάριο Μυρωνάκη ίσως είναι λίγο. Ήταν παρών καθ' όλη τη διάρκεια της εκπόνησης της διπλωματικής μου, παρέχοντάς μου συμβουλές και ακούγοντας με υπομονή κάθε απορία ή προβληματισμό μου, μεγάλο ή μικρό. Η συμβολή του σε αυτή την προσπάθεια ήταν τεράστια.

Θα ήθελα επίσης να ευχαριστήσω τον κ. Ιωάννη Στρατάκη ο οποίος ήταν πάντα πρόθυμος να βοηθήσει σε οποιοδήποτε πρόβλημα ή εμπόδιο αντιμετώπιζα στην πορεία αυτής της εργασίας.

Δεν μπορώ να μην ευχαριστήσω τα άλλα δύο μέλη της ομάδας που είχα την χαρά να γνωρίσω μέσα σε αυτόν τον χρόνο, την κα. Βαρβάρα Σωτηροπούλου και τον κ. Λευτέρη Τζανή. Η Βαρβάρα όχι μόνο με υποδέχτηκε με χαρά στην ομάδα αλλά με βοήθησε τόσο σε επιστημονικό, όσο και σε προσωπικό επίπεδο, και ο Λευτέρης είχε πάντα την όρεξη να συζητάει μαζί μου την πρόοδο της εργασίας.

Τις ευχαριστίες μου θα ήθελα να εκφράσω και στους καθηγητές της τριμελούς επιτροπής, κ. Θωμά Μαρή και κ. Κωνσταντίνο Περισυνάκη.

Τέλος, θα ήθελα να ευχαριστήσω τους δικούς μου ανθρώπους, οικογένεια και φίλους, για όλη την ηθική στήριξη που μου προσέφεραν. Κυρίως τον Χάρη, γιατί ποτέ δεν σταμάτησε να με στηρίζει, και την οικογένειά μου, που ακούραστα άκουγαν όλα τα άγχη και τα προβλήματα μου. Χωρίς εκείνους, πιθανόν να μην ήταν δυνατή η εκπόνηση αυτής της εργασίας.

## ΠΕΡΙΛΗΨΗ

Η Υπολογιστική Τομογραφία Κωνικής Δέσμης (CBCT) αναπτύχθηκε σύντομα μετά την Υπολογιστική Τομογραφία ώστε να επιφέρει τη λήψη τομογραφικών προβολών μέσω υπολογιστικής τομογραφίας μέσω μίας μόνο περιστροφής του ικριώματος (gantry).

Οι πρώτες προσπάθειες για την ανάπτυξη ενός συστήματος CBCT ξεκίνησαν την δεκαετία του 1970. Σήμερα η χρήση του είναι ευρέως διαδεδομένη, κυρίως για οδοντιατρικές εξετάσεις και στην ακτινοθεραπεία.

Σε συμβατικά συστήματα ακτινοθεραπείας, η τοποθέτηση του ασθενή γίνεται κυρίως με χρήση laser pointers και η παρακολούθηση της θέσης του γίνεται με σύστημα υπολογιστικής τομογραφίας fan beam, το οποίο, όμως, δεν παρέχει σε πραγματικό χρόνο, παρακολούθηση της θεραπείας και βελτίωση της τοποθέτησης του ασθενή. Τα προβλήματα αυτά ήρθε να λύσει η ενσωμάτωση του CBCT στον γραμμικό επιταχυντή.

Το μειονέκτημα της τεχνικής αυτής είναι η αύξηση της δόσης στον ασθενή, κάτι το οποίο κάνει την ανάγκη για δοσιμετρία μάλλον επιτακτική. Ένας αριθμός μεθόδων έχει αναπτυχθεί με σκοπό τον προσδιορισμό της δόσης βάση μετρήσεων ή υπολογισμών, με επικρατέστερη τον υπολογισμό του Computed Tomography Dose Index (CTDI), η οποία βασίζεται σε σταθερά πρωτόκολλα και ομοιογενή κυλινδρικά ομοιώματα ώστε να παράγει μια εκτίμηση της απορροφούμενης δόσης. Η μέθοδος αυτή οδηγεί σε απόκλιση μεταξύ της πραγματικής δόσης και της υπολογιζόμενης, λόγω ανατομικών διαφορών και διαφορών στη στοιχειακή σύσταση μεταξύ του ανθρώπινου σώματος και του ομοιώματος.

Μια ακόμα μέθοδος δοσιμετρίας είναι οι προσομοιώσεις Monte Carlo, η οποία επιτρέπει πιο εξατομικευμένες εκτιμήσεις της δόσης. Οι εικόνες της εξέτασης εισάγονται σε λογισμικό και εισάγοντας τα στοιχεία με τα οποία πραγματοποιήθηκε η εξέταση, υπολογίζεται η δόση pixel προς pixel.

Από τους παραπάνω λόγους προκύπτει η ανάγκη για προσωποποιημένη και σε πραγματικό χρόνο δοσιμετρία, η οποία να λαμβάνει υπόψη τα χαρακτηριστικά του σώματος του συγκεκριμένου εξεταζόμενου, τις ρυθμίσεις της έκθεσης και τα χαρακτηριστικά του εξοπλισμού που

χρησιμοποιείται. Η εισαγωγή της τεχνητής νοημοσύνης στον τομέα της ιατρικής φυσικής θα μπορούσε να εκμεταλλευθεί για την κάλυψη της ανάγκης αυτής.

Στην παρούσα εργασία πραγματοποιήθηκαν Monte Carlo προσομοιώσεις ακτινογραφικών και CBCT εξετάσεων θώρακα και άνω κοιλίας, όπως οι τελευταίες γίνονται στην ακτινοθεραπεία, ώστε να υπολογιστούν οι δόσεις ακτινοβολίας των οργάνων ενδιαφέροντος του εκάστοτε ασθενή. Τα αποτελέσματα που προέκυψαν για τις εξετάσεις CBCT θώρακα χρησιμοποιήθηκαν, μαζί με τους παράγοντες που συμβάλλουν στον προσδιορισμό της δόσης, για την ανάπτυξη και εκπαίδευση ενός μοντέλου τεχνητής νοημοσύνης. Οι Monte Carlo προσομοιώσεις έγιναν με χρήση του λογισμικού Impact MC σε εξετάσεις ασθενών που υποβλήθηκαν σε υπολογιστική τομογραφία διάγνωσης και ακτινοθεραπείας στο Πανεπιστημιακό Γενικό Νοσοκομείο Ηρακλείου (Πα.Γ.Ν.Η.) κατά το έτος 2020.

Στόχος ήταν ο αλγόριθμος αυτός να κάνει μια ικανοποιητική πρόβλεψη της δόσης που θα λάβει ο εκάστοτε ασθενής σε συγκεκριμένα όργανα ενδιαφέροντος, σε πραγματικό χρόνο.

## ABSTRACT

Cone beam computed tomography (CBCT) developed soon after the development of computed tomography address the acquiring tomographic projections with only one rotation of the gantry.

Early attempts for the development of a CBCT system started in the decade of 1970. Currently, use of CBCT is widely spread, mostly for dental examinations, and radiotherapy.

In conventional radiotherapy systems, patient's positioning is done mostly using laser pointers and the monitoring of his position is achieved using fan beam CT, which does not provide real time treatment monitoring for position improvement. These problems are solved with built-in CBCT systems.

The disadvantage of this technique is that the patient dose is increased, something that makes the need for dosimetry more than imperative. A number of methods have been developed to calculate dose based on measurements or computations. Computed Tomography Dose Index (CTDI) is the most prominent dose calculation method, which is based on standardized exposure protocols and homogeneous cylindrical phantoms to produce an estimate of absorbed dose. This method can lead to a deviation between the real dose and the calculated, due to anatomical and elemental composition differences between human body and phantom.

Another method of dosimetry is Monte Carlo simulations, which allows for more personalized dose estimation. Examination images and scan parameters are inserted in a software and the dose is calculated pixel by pixel.

For the above reasons, the need for personalized and real time dosimetry is implied that considers the specific patient body characteristics, the exposure settings and the system used. The advent of artificial intelligence in the medical physics field can be exploited towards for this purpose.

In this dissertation radiographic and CBCT thorax and abdomen examinations were simulated with the Monte Carlo technique, in order to calculate the dose that the patient's organs of interest have absorbed. The results for chest CBCT examinations, in combination with other dose predictors, were used for the development and training of an artificial intelligence model. The

simulations were done using Impact MC software and examinations that were occurred in diagnostic and radiotherapy CT in University General Hospital of Heraklion in 2020.

The aim of this algorithm was to predict in real time the dose of each patient in specific organs of interest.

## ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΧΑΡΙΣΤΙΕΣ .....	1
ΠΕΡΙΛΗΨΗ.....	2
ABSTRACT .....	4
ΓΕΝΙΚΟ ΜΕΡΟΣ.....	0
1. ΥΠΟΛΟΓΙΣΤΙΚΗ ΤΟΜΟΓΡΑΦΙΑ ΚΩΝΙΚΗΣ ΔΕΣΜΗΣ .....	1
1.1. Εισαγωγή.....	1
1.2. Το σύστημα CBCT ακτινοθεραπείας .....	1
1.3. Βασικές αρχές σάρωσης.....	4
2. ΔΟΣΙΜΕΤΡΙΑ ΣΤΟ CBCT .....	6
2.1. Εισαγωγή .....	6
2.2. Μέθοδοι δοσιμετρίας .....	6
2.2.1. Δοσιμετρία σε ομοίωμα .....	6
2.2.2. Δοσιμετρία με προσομοίωση Monte Carlo.....	7
2.2.2. Δοσιμετρία με χρήση Τεχνητής Νοημοσύνης.....	8
3. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ.....	10
3.1. Εισαγωγή .....	10
3.2. Απεικόνιση των δεδομένων .....	10
3.3. Διαχωρισμός του σετ δεδομένων.....	13
3.4. Training.....	13
3.4.1. Γραμμικά μοντέλα .....	14
3.4.2. Decision Tree Regressor .....	17
3.4.3. Ensemble τεχνικές.....	18
ΕΙΔΙΚΟ ΜΕΡΟΣ.....	21
4. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ .....	22
4.1. Εισαγωγή.....	22
4.2. Impact MC .....	22
4.2.1. Εικόνες σάρωσης.....	22
4.2.2. Παράμετροι σάρωσης .....	23
4.2.3. Παράμετροι προσομοίωσης .....	25
4.2.4. Εικόνες εξόδου .....	26
4.3. Αλγόριθμος τεχνητής νοημοσύνης .....	26

4.3.1.	Περιγραφική στατιστική .....	28
4.3.2.	Εκπαίδευση αλγορίθμου .....	31
4.3.3.	Tuning .....	36
4.3.4.	Validation του μοντέλου .....	38
5.	ΑΠΟΤΕΛΕΣΜΑΤΑ.....	41
5.1.	Αποτελέσματα προσομοιώσεων .....	41
5.1.1.	CBCT .....	41
5.1.2.	Ακτινογραφία.....	44
5.2.	Πρόβλεψη δόσης για CBCT θώρακος.....	51
5.2.1.	Οστά .....	52
5.2.2.	Οισοφάγος .....	53
5.2.3.	Μαστοί.....	54
5.2.4.	Καρδιά .....	55
5.2.5.	Πνεύμονες.....	56
5.2.6.	Δέρμα.....	57
5.2.7.	Integral.....	58
6.	ΣΥΖΗΤΗΣΗ.....	59
	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	61
	ΠΙΝΑΚΕΣ.....	62
	ΕΙΚΟΝΕΣ .....	63



# ΓΕΝΙΚΟ ΜΕΡΟΣ

# 1. ΥΠΟΛΟΓΙΣΤΙΚΗ ΤΟΜΟΓΡΑΦΙΑ ΚΩΝΙΚΗΣ ΔΕΣΜΗΣ

## 1.1. Εισαγωγή

Η υπολογιστική τομογραφία κωνική δέσμης (cone beam computed tomography ή CBCT) είναι μια μέθοδος λήψης εικόνων του ανθρώπινου σώματος μέσω CT, μόνο με μια περιστροφή του ικριώματος (gantry). Αυτό επιτυγχάνεται με τη χρήση ενός δισδιάστατου ενιαίου ανιχνευτή και την αντικατάσταση της δέσμης βεντάλιας (fan beam) από μια κωνική δέσμη που καλύπτει πλήρως την περιοχή ενδιαφέροντος.

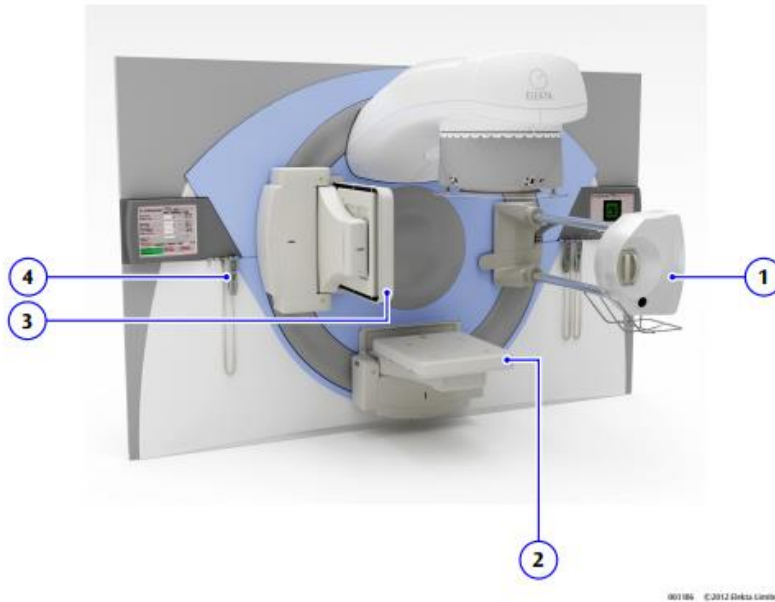
Οι πρώτες απόπειρες για τη δημιουργία ενός τέτοιου συστήματος ξεκίνησαν την δεκαετία του '80 από τη Mayo Clinic, χωρίς όμως να προκύψει ένα εμπορικό προϊόν. Οι προσπάθειες συνεχίστηκαν για τις επόμενες δύο, περίπου, δεκαετίες, μέχρι το 1999 που εισήχθησαν για πρώτη φορά ανιχνευτές πυριτίου, μεγάλου δυναμικού εύρους και η χρήση του CBCT διαδόθηκε ευρέως (Shaw, 2014).

Τα πλεονεκτήματα αυτής της τεχνικής είναι η μείωση του χρόνου σάρωσης, λόγω της απευθείας λήψης της τρισδιάστατης εικόνας, χωρίς το ενδιάμεσο στάδιο της ανακατασκευής, το χαμηλό κόστος της λυχνίας αλλά και η απλότητα της κατασκευής. Για τους λόγους αυτούς, το CBCT χρησιμοποιείται σε οδοντιατρικά, αγγειογραφικά και ακτινοθεραπευτικά συστήματα (Σειμένης, 2021).

Στις παρακάτω ενότητες θα περιγραφεί το σύστημα CBCT ακτινοθεραπείας Elekta XVI R5.0.4. και οι βασικές αρχές σάρωσης, δίνοντας έμφαση σε εξετάσεις θώρακος και κοιλίας. Οι εικόνες των εξαρτημάτων του συστήματος έχουν ληφθεί από το φυλλάδιο οδηγιών χρήσης του συστήματος.

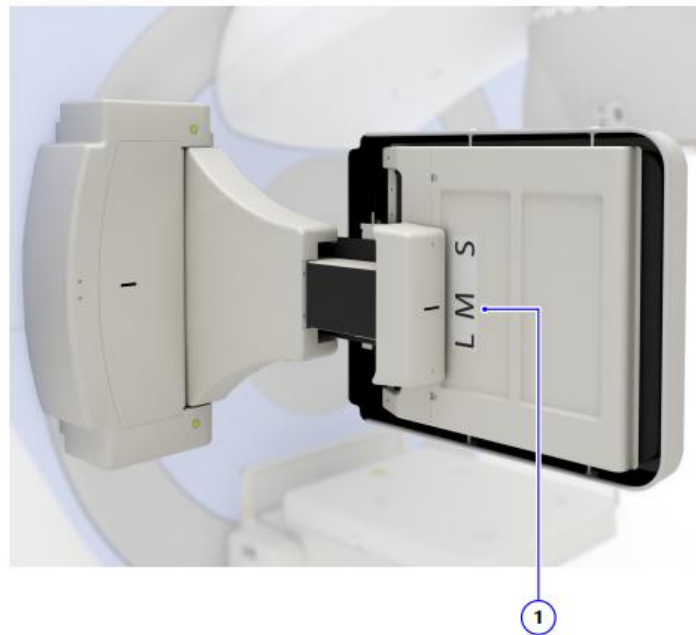
## 1.2. Το σύστημα CBCT ακτινοθεραπείας

Το σύστημα CBCT ακτινοθεραπείας αποτελείται από τέσσερα βασικά μέρη: την πηγή ακτίνων X (kV source arm) (1), δύο ανιχνευτές, έναν για ενέργειες της τάξης των MV (2) και έναν για της τάξης των kV (3) και έναν χειριστήριο (4).



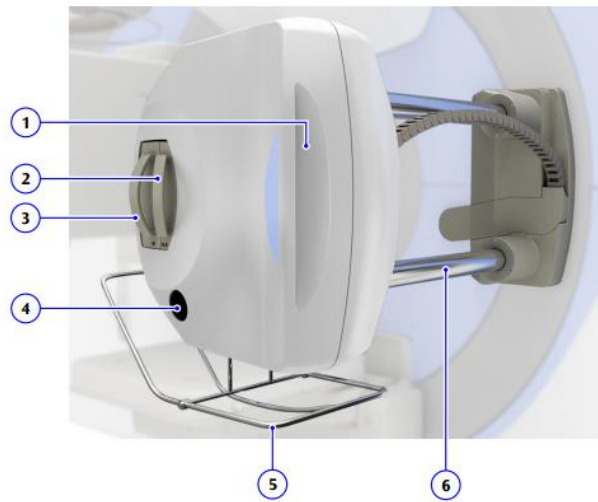
Εικόνα 1 το σύστημα CBCT XVI R5.0.4. της Elekta

Το ανιχνευτικό πάνελ kV είναι τοποθετημένο πάνω στο μπράτσο (arm) του ανιχνευτή kV, το οποίο μετακινείται οριζόντια ή κατακόρυφα, ώστε να γίνεται σωστή λήψη των εικόνων του ασθενή. Οι διαστάσεις του πάνελ μεταβάλλονται, ανάλογα με το επιθυμητό μέγεθος του πεδίου προβολής, από μικρό, μεσαίο ή μεγάλο.



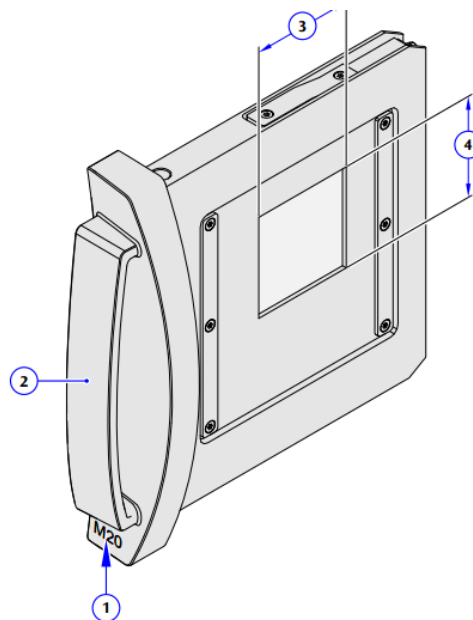
Εικόνα 2 ανιχνευτικό πάνελ kV

Στο μπράτσο της πηγής kV ακτίνων X εισάγονται ο κατευθυντήρας (2) και το φίλτρο (3) τα οποία αλλάζουν, ανάλογα το πεδίο απεικόνισης και τα σωματικά χαρακτηριστικά του ασθενή.



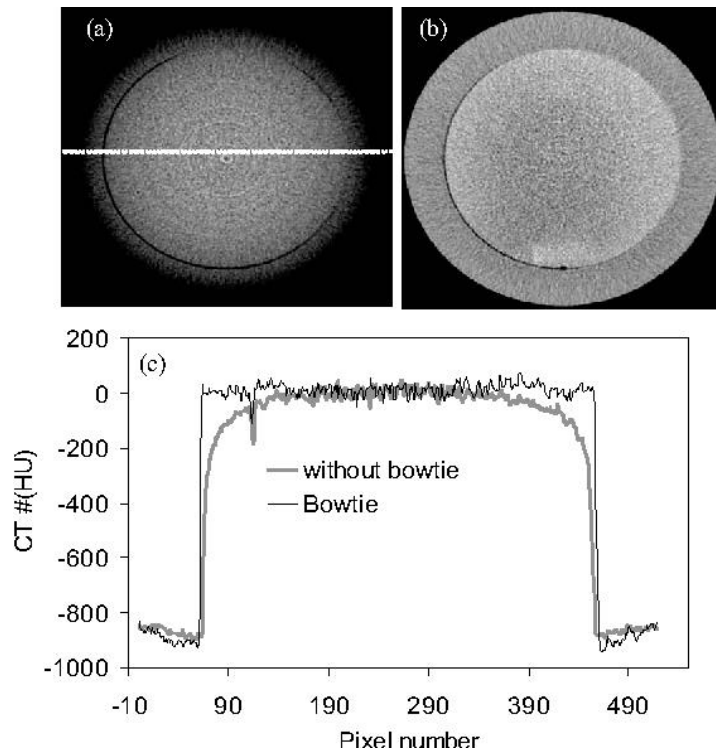
Εικόνα 3 το μπράτσο της πηγής kV

Οι κατευθυντήρες είναι φτιαγμένοι από μόλυβδο, εκτός από ένα άνοιγμα το οποίο επιτρέπει τη διέλευση των ακτίνων. Το μέγεθος και το σχήμα του ανοίγματος αυτού διαφέρει, ανάλογα με το επιθυμητό μέγεθος και σχήμα του πεδίου. Επίσης διαφέρει και η θέση του ανοίγματος, ώστε να επιτυγχάνεται η ακτινοβολή στο σωστό σημείο του σώματος του ασθενή.



Εικόνα 4 collimation cassette

Η χρησιμότητα των φίλτρων έγκειται στο να μειώνουν την ακτινοβολία που φτάνει στον ασθενή, με σκοπό και την ακτινοπροστασία του, αλλά και την αποτροπή του κορεσμού του πάνελ, άρα την βελτίωση της εικόνας. Το φίλτρο που χρησιμοποιείται στις περισσότερες εξετάσεις είναι το bowtie, το οποίο έχει μεγαλύτερο πάχος στα άκρα του. Λόγω του μεγαλύτερου πάχους που έχει το σώμα του ασθενή στο κέντρο του, οι ακτίνες που πέφτουν εκεί εξασθενούν περισσότερο απ' ότι οι περιφερειακές. Έτσι, χρησιμοποιώντας το bowtie φίλτρο επιτυγχάνεται ομοιόμορφη ακτινοβολήση του FOV.

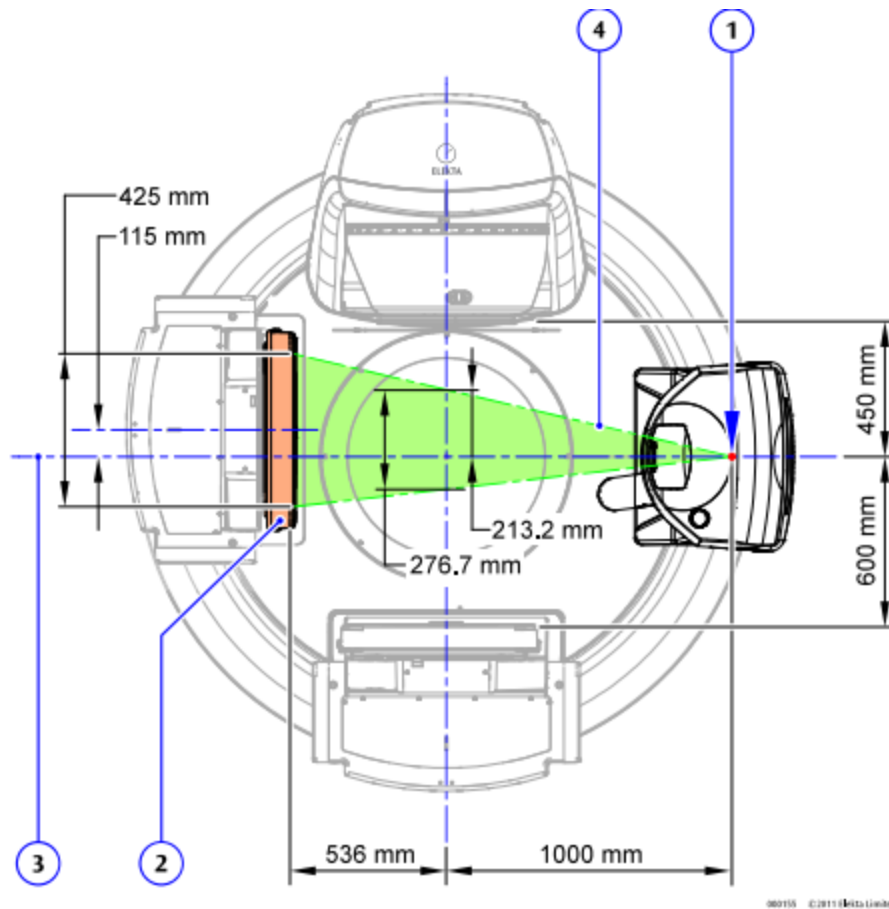


Εικόνα 5 η επίδραση του φίλτρου bowtie στην ποιότητα της εικόνας

### 1.3. Βασικές αρχές σάρωσης

Με το σύστημα του CBCT μπορούν να πραγματοποιηθούν απεικονίσεις θώρακα, άνω και κάτω κοιλίας, κεφαλής και λαιμού κ.α.. Στις επόμενες παραγράφους θα περιγραφούν οι απεικονίσεις θώρακα και άνω κοιλίας, μιας και αυτές χρησιμοποιήθηκαν για τις Monte Carlo προσομοιώσεις.

Και για τις δύο αυτές απεικονίσεις μέσω CBCT ακτινοθεραπείας, χρησιμοποιείται το μεσαίο FOV, όπως φαίνεται στην παρακάτω εικόνα.



Εικόνα 6 medium FOV

Η σήμανση (1) αναφέρεται στην σημειακή πηγή kV ακτίνων X η οποία απέχει  $1000mm$  απόσταση πηγής – ισοκέντρου |  $536mm$  απόσταση ισοκέντρου – πάνελ  $= 1536mm$  από το ανιχνευτικό πάνελ (2). Ο άξονας (3) αναφέρεται στον άξονα αναφοράς της δέσμης, ο οποίος είναι κάθετος στο επίπεδο του πάνελ. Τέλος, με τον αριθμό (4) επισημαίνεται η προβολή του πεδίου.

Από το σχήμα παρατηρείται ότι το πάνελ είναι τοποθετημένο κατά  $115mm$  έκκεντρα του άξονα της δέσμης. Αυτό συμβαίνει ώστε να αποφευχθούν συγκρούσεις μεταξύ των εξαρτημάτων του συστήματος.

Το πεδίο που σχηματίζεται είναι μεγέθους  $276mm \times 276mm$  στο ισοκέντρο και  $425mm \times 425mm$  στην απόσταση που βρίσκεται ο ανιχνευτής και το φίλτρο που χρησιμοποιείται είναι το bow-tie (Elekta, 2016).

## 2. ΔΟΣΙΜΕΤΡΙΑ ΣΤΟ CBCT

### 2.1. Εισαγωγή

Απόρροια της εισαγωγής του CBCT στην κλινική πράξη ήταν η ανάγκη για δοσιμετρία. Πολλές ήταν οι μέθοδοι που προτάθηκαν για την λύση αυτού του προβλήματος. Παρακάτω θα γίνει αναφορά στον υπολογισμό της δόσης μέσω του Computed Tomography Dose Index (CTDI) και του γινομένου δόσης μήκους πεδίου (Dose Length Product ή DLP), κάνοντας χρήση ομοιώματος, μέσω Monte Carlo προσομοιώσεων και μέσω τεχνητής νοημοσύνης.

### 2.2. Μέθοδοι δοσιμετρίας

#### 2.2.1. Δοσιμετρία σε ομοίωμα

Η παρακάτω περιγραφή είναι αυτή που προτείνει η κατασκευάστρια εταιρία και σύμφωνα με το πρωτόκολλο IEC 60601-2-44-A1 του 2012.

Αρχικά τοποθετείται το ομοίωμα διαμέτρου 160mm για εξετάσεις κεφαλής ή 320mm για σώματος, με το κέντρο του στο ισόκεντρο, και σημειώνονται οι οπές στις γωνίες 0 (η πιο κοντινή οπή στην πηγή), 90°, 180° και 270°. Έπειτα, ρυθμίζεται και διορθώνεται το ηλεκτρόμετρο ως προς την πίεση και την θερμοκρασία περιβάλλοντος. Ο θάλαμος ιονισμού τοποθετείται στην κεντρική οπή του ομοιώματος και στις υπόλοιπες τοποθετούνται ακρυλικές ράβδοι. Το μέγεθος του κατευθυντήρα θα πρέπει να είναι 20mm και να χρησιμοποιηθεί φίλτρο. Αφού ρυθμιστούν οι προηγούμενες παράμετροι, η ακτινοβολήση γίνεται τρεις φορές σύμφωνα με το πρωτόκολλο και λαμβάνεται η δόση από κάθε ακτινοβολήση. Έτσι, υπολογίζεται η μέση δόση  $\bar{D}$  και από το αυτήν το

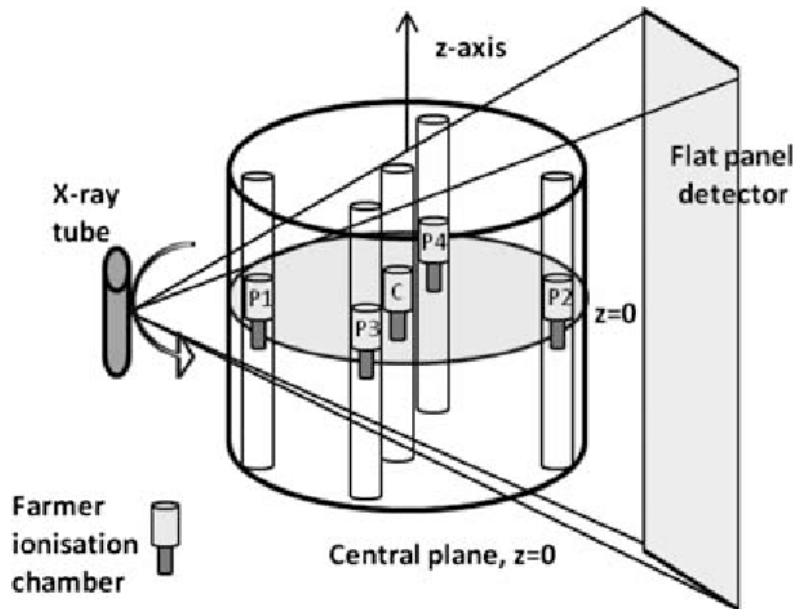
$CTDI_{100}$  σύμφωνα με τη σχέση  $CTDI_{100} = \int_{-50mm}^{50mm} \frac{D(z)}{100mm} dz = \frac{\bar{D}}{100mm}$ . Η διαδικασία επαναλαμβάνεται για κάθε γωνία που προαναφέρθηκε και υπολογίζεται το περιφερειακό CTDI σύμφωνα

με τη σχέση  $CTDI_{100(peripheral)} = \frac{1}{4}\{CTDI_{100(0)} + CTDI_{100(90^\circ)} + CTDI_{100(180^\circ)} + CTDI_{100(270^\circ)}\}$

και το σταθμισμένο CTDI σύμφωνα με τη σχέση  $CTDI_w = \frac{1}{3}CTDI_{100(center)} +$

$\frac{2}{3}CTDI_{100(peripheral)} = CTDI_{vol}$ .

Τελικά, το DLP θα είναι  $DLP = CTDI_w \times field\ length$  (Elekta, 2016).



Εικόνα 7 διάταξη για τον υπολογισμό του CTDI

Τα πλεονεκτήματα αυτής της μεθόδου δοσιμετρίας είναι ότι ο υπολογισμός της δόσης γίνεται βάση του εκάστοτε συστήματος (system specific) και κάνει μια καλή εκτίμησή της πριν πραγματοποιηθεί η εξέταση, οπότε μπορούν να αποφευχθούν τα καθορισμένα αποτελέσματα της ακτινοβολίας, μια και η εν λόγω απεικόνιση συνδυάζεται με την ακτινοθεραπεία.

Από την άλλη, κάθε ασθενής είναι διαφορετικός, γεγονός το οποίο δεν λαμβάνεται υπόψη, με αποτέλεσμα να υπάρχει πολλές φορές σημαντική απόκλιση από την πραγματική δόση που τελικά θα λάβει ο εξεταζόμενος.

### 2.2.2. Δοσιμετρία με προσομοίωση Monte Carlo

Η χρήση Monte Carlo τεχνικών προσομοίωσης είναι μια από τις πιο σύγχρονες και εξελιγμένες μορφές δοσιμετρίας. Εισάγοντας στο λογισμικό τις εικόνες που προέκυψαν από την σάρωση ενός ασθενούς καθώς και τα στοιχεία με τα οποία έγινε η εξέταση γίνεται υπολογισμός της δόσης μέσω τεχνικών Monte Carlo. Η κατανομή δόσης που προκύπτει είναι τρισδιάστατη, και βάση αυτής μπορεί να προσδιοριστεί η δόση στα όργανα και η ενεργός δόση. Λόγω του ότι η μέθοδος αυτή λαμβάνει υπόψη όλες τις μορφές αλληλεπιδράσεων μεταξύ φωτονίων και οργάνων, καθώς και όλα τα στοιχεία με τα οποία πραγματοποιείται η εξέταση (kV, mA, s κ.λπ.), δίνεται η



δυνατότητα υπολογισμού της δόσης που απορροφάται από κάθε ογκοστοιχείο (voxel) της δοσμένης εικόνας (© CT Imaging GmbH, 2016).

Το μεγάλο πλεονέκτημα αυτής της τεχνικής είναι η εξατομικευμένη δοσιμετρία, η οποία λαμβάνει υπόψη και τα χαρακτηριστικά του συστήματος που χρησιμοποιείται. Είναι, επίσης, μια εξαιρετικά ακριβής μέθοδος.

### 2.2.2. Δοσιμετρία με χρήση Τεχνητής Νοημοσύνης

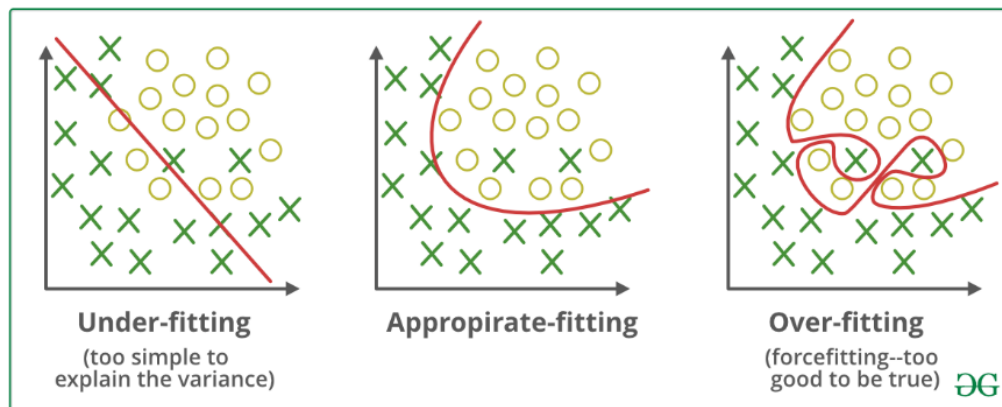
Τα τελευταία χρόνια, είναι ευρέως διαδεδομένη η χρήση της Τεχνητής Νοημοσύνης (Artificial Intelligence ή Α.Ι.) και του Machine Learning σε πάρα πολλούς τομείς της επιστήμης. Η εφαρμογή τους, ωστόσο, στην δοσιμετρία είναι σε πρώιμο επίπεδο, παρόλο που με αυτόν τον τρόπο μπορεί να γίνει εξατομικευμένη δοσιμετρία, η οποία να λαμβάνει υπόψη τα χαρακτηριστικά του συστήματος αλλά και του πρωτοκόλλου. Φαίνεται, επίσης, πως θα βοηθήσουν στο να υπερπηδηθούν εμπόδια, όπως αυτό της δοσιμετρίας σε πραγματικό χρόνο, δηλαδή αμέσως μετά την εξέταση (Damilakis, 2020).

Για να δημιουργηθεί ένας αλγόριθμος τεχνητής νοημοσύνης, ο οποίος να προβλέπει με ακρίβεια την δόση της οποία θα λάβει ο εκάστοτε ασθενής από μια απεικόνιση με CBCT ακτινοθεραπείας, απαιτείται ένας όγκος δεδομένων για την «εκπαίδευση» (training) του μοντέλου. Τα δεδομένα αυτά θα πρέπει να περιλαμβάνουν δόσεις οργάνων ενός πλήθους ασθενών που υποβλήθηκαν σε αυτή την εξέταση, σε συνδυασμό με τα στοιχεία του πρωτοκόλλου και του συστήματος που χρησιμοποιείται (dose predictors). Ο αλγόριθμος αυτός θα βρει τις συσχετίσεις μεταξύ των ανεξάρτητων μεταβλητών, δηλαδή των dose predictors, και της εξαρτημένης, δηλαδή της απορροφούμενης δόσης στο όργανο, και θα τις χρησιμοποιήσει ώστε να φτιάξει το μοντέλο πρόβλεψης.

Όμως, όταν ο αλγόριθμος κάνει χρήση γνωστών τιμών δόσης, υπάρχει ο κίνδυνος της υπερπροσαρμογής (overfitting). Αυτό, πρακτικά, σημαίνει ότι φτιάχνει μια καμπύλη προσαρμογής της εξαρτημένης μεταβλητής, η οποία περνάει από όλα τα σημεία των ανεξάρτητων μεταβλητών. Γι' αυτό, η χρήση ενός, άγνωστου στο μοντέλο, σετ δεδομένων κρίνεται απαραίτητη για την αξιολόγησή του (validation).

Από την άλλη, εάν τα δεδομένα δεν είναι αρκετά, τότε ο αλγόριθμος δεν θα είναι ικανός να μοντελοποιήσει όλες τις συσχετίσεις μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης και θα φτιάξει ένα πολύ απλό μοντέλο το οποίο, όταν του εισαχθούν άγνωστες τιμές, θα κάνει μια πρόβλεψη με πολύ μεγάλο σφάλμα. Αυτό είναι το φαινόμενο της υποπροσαρμογής (underfitting) (Chollet, 2018).

Για την αποφυγή των παραπάνω φαινομένων χρησιμοποιείται ένας ικανός αριθμός δεδομένων για το training του μοντέλου, αλλά και αρκετά δεδομένα ώστε να γίνει μια σωστή αξιολόγηση των προβλέψεων που κάνει.



Εικόνα 8 τα φαινόμενα underfitting και overfitting

Έπειτα, θα πρέπει να επιλεγεί η κατάλληλη τεχνική training. Στην επόμενη ενότητα θα γίνει περιγραφή αυτών των τεχνικών και πως εφαρμόζεται η κάθε μια στην πράξη.

Μπορεί να έχει επιλεχθεί η καλύτερη τεχνική για την εκπαίδευση του μοντέλου, όμως πάντα υπάρχουν περιθώρια βελτίωσης. Σε αυτό το σημείο γίνεται το λεγόμενο tuning του μοντέλου. Κάθε τεχνική training γίνεται λαμβάνοντας υπόψη κάποιες υπερπαραμέτρους, οι οποίες όμως δεν ταιριάζουν πάντα τέλεια στο dataset. Μεταβάλλοντάς τες, μπορεί να επιτευχθεί μια καλύτερη προσαρμογή των δεδομένων, άρα και μια πιο ακριβής πρόβλεψη.

Το τελευταίο στάδιο είναι το validation του μοντέλου. Εισάγοντας άγνωστα δεδομένα στον αλγόριθμο που έχει προκύψει, αξιολογείται η συμπεριφορά του με βάση την απόκλιση των τιμών που προβλέπει από τις πραγματικές. Όσο πιο μικρή είναι αυτή η απόκλιση, τόσο πιο ακριβές το μοντέλο (Brownlee, 2016).

## 3. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

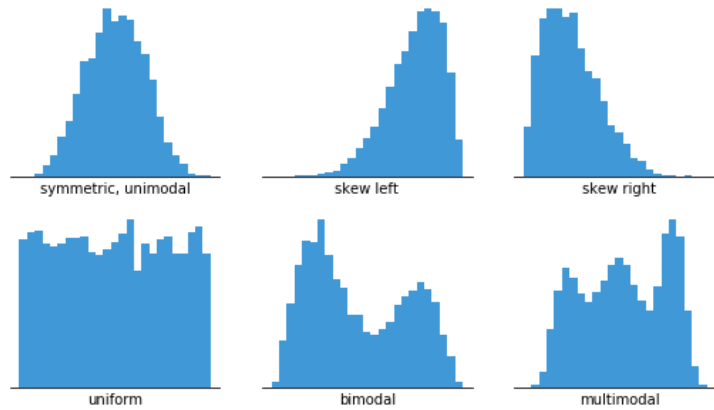
### 3.1. Εισαγωγή

Για να δημιουργηθεί ένας αλγόριθμος πρόβλεψης με χρήση τεχνητής νοημοσύνης αρχικά απαιτείται ένα σετ δεδομένων ικανού αριθμού, το οποίο να περιέχει της ανεξάρτητες μεταβλητές αλλά και την εξαρτημένη. Έπειτα, γίνονται διαγράμματα των μεταβλητών, ώστε να φανούν οι όποιες ιδιαιτερότητες υπάρχουν και να διορθωθούν στη συνέχεια (αλλαγή κλίμακας, σταθεροποίηση κ.α.). Διαγράμματα χρησιμοποιούνται επίσης και για την φανέρωση συσχετίσεων μεταξύ των μεταβλητών. Στη συνέχεια, το σετ δεδομένων χωρίζεται σε training και validation splits. Κάνοντας χρήση της γλώσσας προγραμματισμού Python και του πακέτο sci-kit learn και χρησιμοποιώντας γραμμικές, ensemble ή άλλες (neighbors, trees) τεχνικές εντοπίζεται ο καλύτερος αλγόριθμος συσχέτισης της εξαρτημένης μεταβλητής με τις ανεξάρτητες και σε αυτόν γίνεται βελτιστοποίηση (tuning), ώστε να δίνει το καλύτερο δυνατό αποτέλεσμα. Τέλος, γίνεται η επαλήθευση (validation) στο τμήμα του dataset το οποίο δεν έχει χρησιμοποιηθεί για την εκμάθηση του μοντέλου και προκύπτουν οι προβλεπόμενες τιμές.

Όλα τα παραπάνω θα αναλυθούν στις επόμενες υποενότητες.

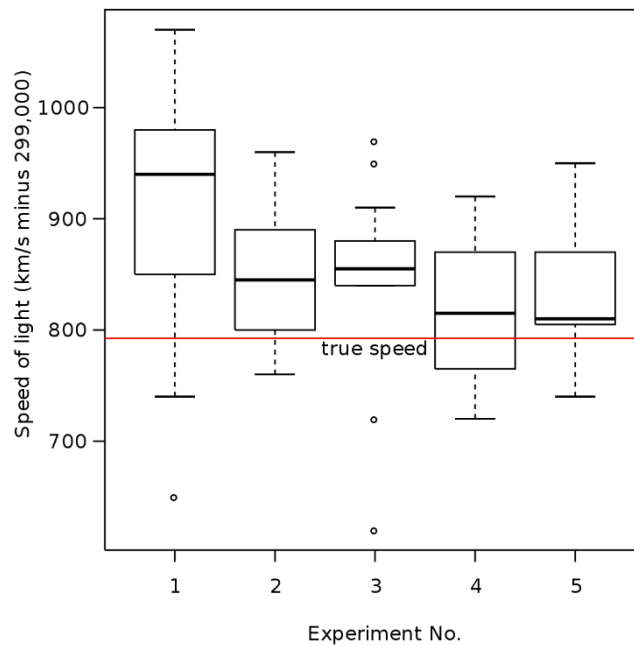
### 3.2. Απεικόνιση των δεδομένων

Το πρώτο βήμα είναι να γίνει σχεδιασμός γραφημάτων των μεταβλητών. Πολλές φορές είναι χρήσιμο να γίνονται μετασχηματισμοί των δεδομένων γύρω μια μέση τιμή ή μετατροπή των δεδομένων στην ίδια κλίμακα, διότι αυξάνεται η ακρίβεια του μοντέλου. Μέσω γραφημάτων όπως τα ιστογράμματα φαίνεται εάν υπάρχουν αποκλίσεις από την κανονική κατανομή, και σε ποιες μεταβλητές, ώστε να γίνουν σε επόμενο στάδιο οι κατάλληλοι μετασχηματισμοί.



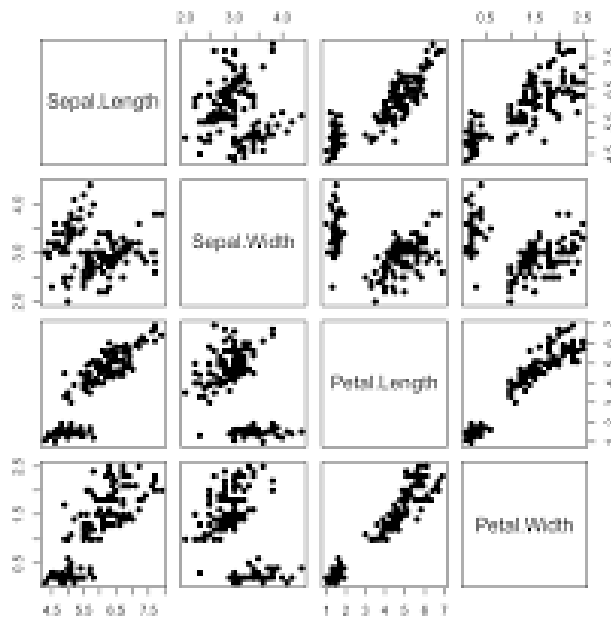
Εικόνα 9 παραδείγματα ιστογραμμάτων

Στο επόμενο διάγραμμα φαίνονται τα boxplots. Θεωρούνται ιδιαίτερα σημαντικά, καθώς δίνουν μια εικόνα για την μέση τιμή της κάθε μεταβλητής, αλλά για τα σημεία εκείνα τα οποία ξεφεύγουν της κατανομής και πιθανόν προέρχονται από σφάλματα κατά τη μέτρηση. Τα σημεία αυτά ονομάζονται outliers και θα πρέπει να αφαιρεθούν, καθώς οδηγούν σε εσφαλμένες προβλέψεις.



Εικόνα 10 παράδειγμα διαγράμματος boxplot. Με κύκλους φαίνονται τα outliers

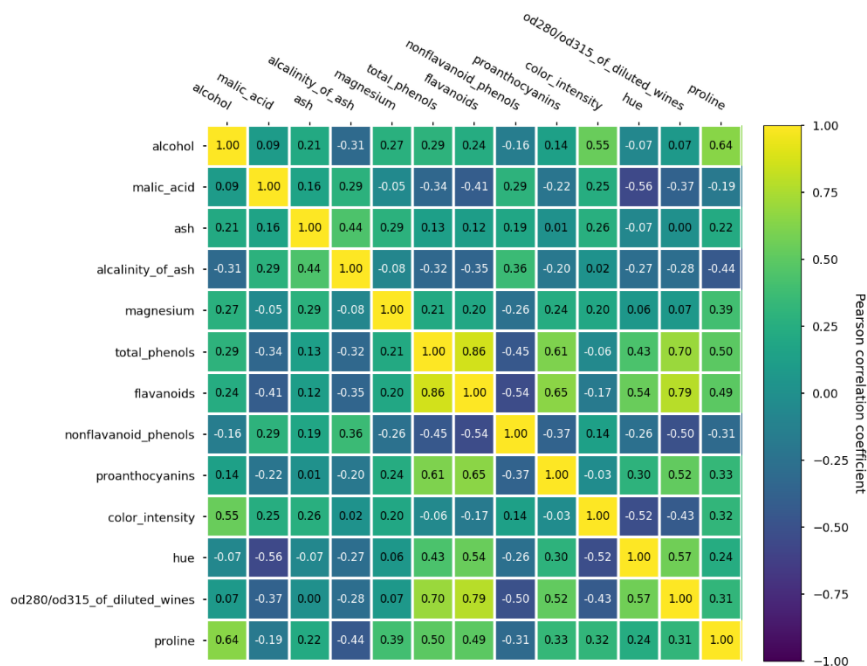
Τα διαγράμματα συσχετίσεων είναι επίσης σημαντικά. Δίνουν μια εικόνα για το κατά πόσο η εξαρτημένη μεταβλητή όντως εξαρτάται από τις ανεξάρτητες.



Εικόνα 11 scatter plots

Στην παραπάνω εικόνα φαίνεται ένα διάγραμμα διασποράς. Παρατηρείται ότι όλες οι μεταβλητές έχουν μια γραμμική εξάρτηση μεταξύ τους, εκτός από κάποια σημεία που ξεφεύγουν αυτής της γραμμικότητας, που πιθανόν είναι outliers και πρέπει να αφαιρεθούν.

Ο πίνακας συσχέτισης Pearson αποτυπώνει και αυτός τις συσχετίσεις με τη μέθοδο Pearson.



Εικόνα 12 παράδειγμα ενός correlation matrix

Στα δεξιά του γραφήματος φαίνονται τα χρώματα που μπορεί να πάρει το κάθε κελί, ανάλογα με την τιμή του συντελεστή Pearson. Για παράδειγμα, με κίτρινο και με μπλε παρουσιάζονται τα κελία εκείνα των οποίων οι αντίστοιχες μεταβλητές έχουν συντελεστή 1 ή -1, άρα είναι ανάλογες ή αντιστρόφως ανάλογες, αντίστοιχα, μεταξύ τους.

### 3.3. Διαχωρισμός του σετ δεδομένων

Το επόμενο βήμα είναι να χωριστεί το σετ αυτό σε δύο μέρη: το μεγαλύτερο μέρος θα χρησιμοποιηθεί για το training του μοντέλου, ενώ το μικρότερο για το validation. Αυτό γίνεται, όπως αναφέρθηκε και νωρίτερα, για την αποφυγή φαινομένων overfitting. Η αναλογία των ποσοστών των τμημάτων training και validation είναι τυπικά από 0.8 με 0.2, έως 0.65 με 0.35, ανάλογα με τα χαρακτηριστικά και το μέγεθος του αρχικού dataset (Brownlee, 2016).

### 3.4. Training

Το κύριο μέρος του μοντέλου είναι το training, δηλαδή η εκπαίδευσή του. Γίνεται με το μεγαλύτερο κομμάτι του σετ δεδομένων, και σε αυτό το σημείο ο αλγόριθμος βρίσκει τον τρόπο που συσχετίζεται η δόση με τις υπόλοιπες μεταβλητές. Παρακάτω θα γίνει αναφορά στα κύρια μοντέλα εκπαίδευσης.

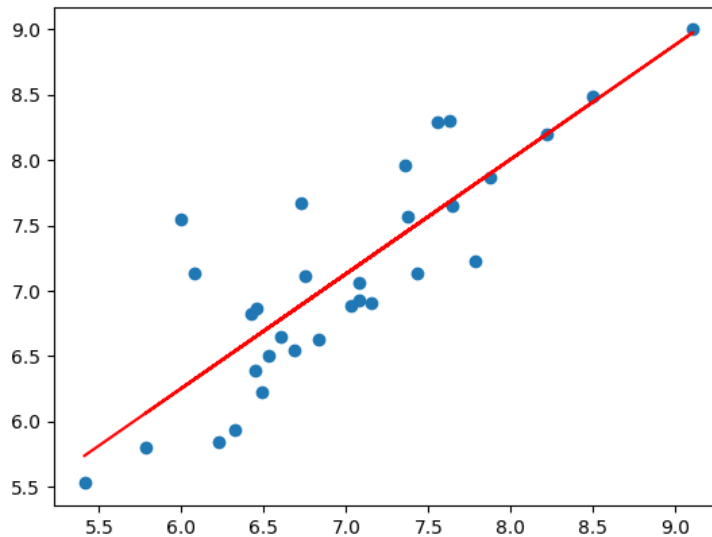
### 3.4.1. Γραμμικά μοντέλα

Τα γραμμικά μοντέλα συνδέουν τη μεταβλητή-στόχο ( $Y$ ) με τις υπόλοιπες μεταβλητές ( $X$ ) με γραμμικό τρόπο της μορφής  $Y = w_0 + w_1x_1 + \dots + w_px_p$ , όπου  $p$  ο αριθμός των δεδομένων, με σκοπό να ελαχιστοποιήσουν τον όρο των ελαχίστων τετραγώνων  $R^2 = 1 - \frac{\sum(Y_{true} - Y_{predicted})^2}{\sum(Y_{true} - \overline{Y_{true}})^2}$ , όπου  $Y_{true}$  η πραγματική τιμή της εξαρτημένης μεταβλητής,  $\overline{Y_{true}}$  η μέση τιμή της  $Y_{true}$  και  $Y_{predicted}$  η τιμή που προέβλεψε το μοντέλο. Η καλύτερη δυνατή τιμή είναι 1 για  $Y_{predicted} = Y_{true}$ , ενώ ένα μοντέλο το οποίο θα προβλέπει πάντα την σωστή τιμή  $Y$ , ανεξάρτητα των μεταβλητών του dataset θα έχει πάντα  $R^2 = 0$  (developers, 2007-2022).

Υποθέτουν ότι η κατανομή των μεταβλητών είναι κανονική, ότι η εξαρτημένη μεταβλητή όντως εξαρτάται από τις ανεξάρτητες και ότι οι ανεξάρτητες μεταβλητές δεν συνδέονται μεταξύ τους, διότι αυτό οδηγεί σε συγγραμμικότητα και κακή απόκριση του μοντέλου. Τα γραμμικά μοντέλα του sci kit learn είναι τα εξής: Linear Regression, Elastic Net και Lasso (Brownlee, 2016).

#### 3.4.1.1. Linear Regression

Το μοντέλο αυτό κάνει γραμμική προσαρμογή των δεδομένων, με συντελεστές ( $w_1, w_2, \dots, w_p$ ). Από μαθητικής άποψης, λύνει ένα πρόβλημα της μορφής  $\min_w \|Xw - Y\|_2^2$ . Είναι μια μέθοδος που απαιτεί ανεξαρτησία μεταξύ των τιμών του πίνακα  $X$ , αφού όταν αυτές συνδέονται γραμμικά τα ελάχιστα τετράγωνα γίνονται πολύ ευαίσθητα σε τυχαία σφάλματα στην εξαρτημένη μεταβλητή, με αποτέλεσμα να υπάρχει μεγάλη διακύμανση (developers, 2007-2022).

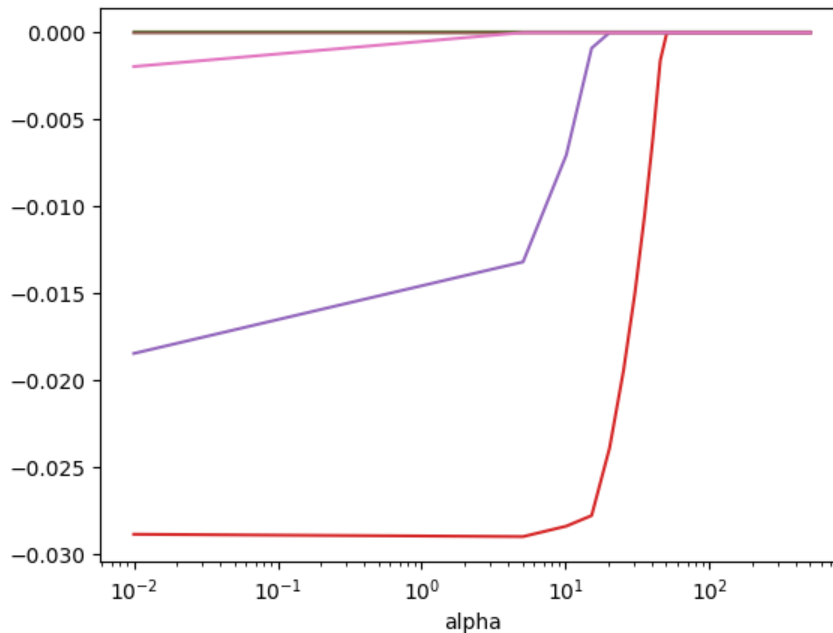


Εικόνα 13 linear regression με  $R^2=0.706$

#### 3.4.1.2. Lasso

Το Lasso είναι ένα γραμμικό μοντέλο το οποίο προσδιορίζει σποραδικούς συντελεστές. Λόγω του ότι επιλέγει λύσεις με λιγότερους μη μηδενικούς συντελεστές, καταφέρνει να μειώσει της μεταβλητές από τις οποίες εξαρτάται η μεταβλητή εκείνη για την οποία πρόκειται να γίνει η πρόβλεψη. Το μοντέλο χρησιμοποιεί την σχέση  $\min_w \left( \frac{1}{2n_{samples}} \|Xw - Y\|_2^2 + \alpha \|w\|_1 \right)$ , όπου  $\alpha$  μια σταθερά και  $\|w\|_1$  η απόλυτη τιμή του αθροίσματος των συντελεστών  $w$ , δηλαδή της κλίσης. Ο όρος αυτός ονομάζεται μέτρο l1 (l1 norm) (developers, 2007-2022).



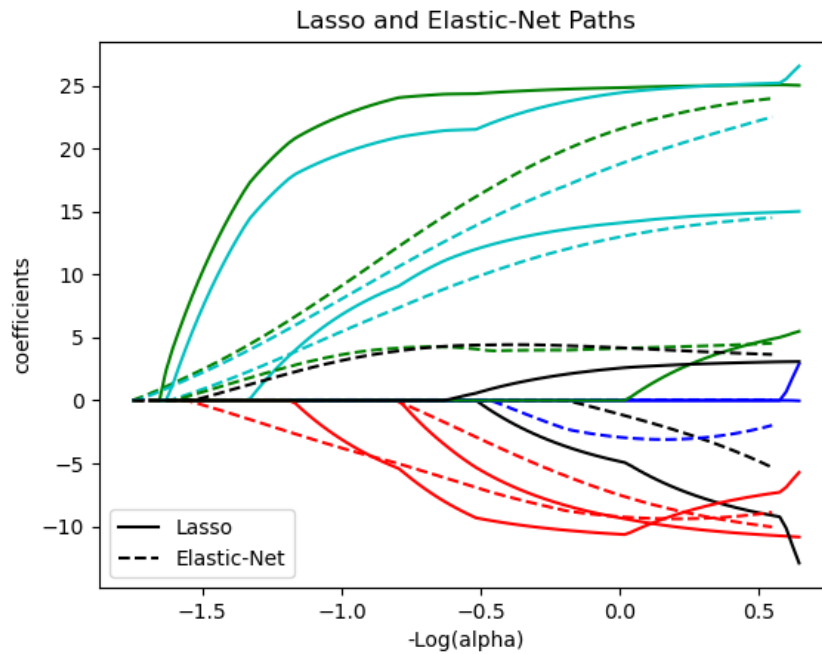


Εικόνα 14 γραφική παράσταση των συντελεστών του μοντέλου συναρτήσει του "α". Για α=0 το μοντέλο είναι απλό linear regression, ενώ για πολύ μεγάλα α όλοι οι συντελεστές μηδενίζονται

### 3.4.1.3. Elastic Net regression

Η τεχνική Elastic Net, όπως και η Lasso, προσδιορίζει σποραδικούς συντελεστές. Χρησιμοποιεί την l1 norm αλλά και την l2 norm η οποία ισούται με  $\|w\|_2^2$ . Έχει το πλεονέκτημα των λίγων μη μηδενικών συντελεστών που αναφέρθηκε στην προηγούμενη παράγραφο, αλλά είναι πιο σταθερή στην πολυσυγγραμμικότητα, λόγω του ότι όσο μεγαλύτερη βαρύτητα έχει η l2 norm, τόσο περισσότεροι συντελεστές θα μηδενιστούν και έτσι η μεταβλητή στόχος θα εξαρτάται από λιγότερες ανεξάρτητες μεταβλητές. Τα παραπάνω περιγράφονται μαθηματικά ως

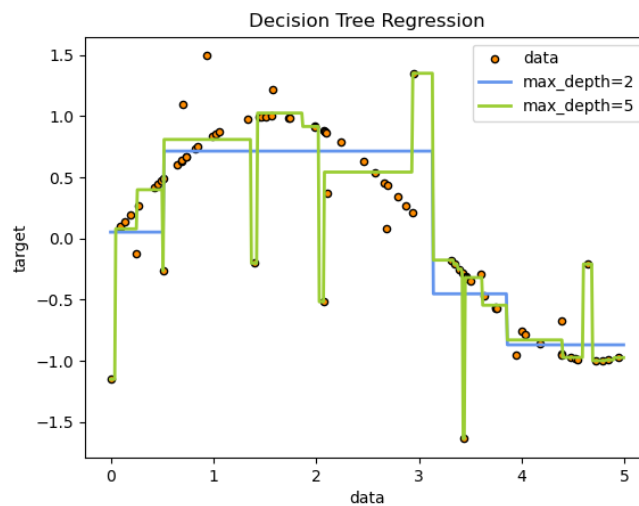
$$\min_w \left( \frac{1}{2n_{samples}} \|Xw - Y\|_2^2 + \alpha p \|w\|_1 + \frac{\alpha(1-p)}{2} \|w\|_2^2 \right) \text{ (developers, 2007 – 2022).}$$



Εικόνα 15 διάγραμμα των συντελεστών "α" των Lasso και Elastic Net, για τους ίδιους συντελεστές

### 3.4.2. Decision Tree Regressor

Τα Decision Trees είναι μια μη παραμετρική μέθοδος εκπαίδευσης. Κατασκευάζουν ένα μοντέλο το οποίο κάνει πρόβλεψη της εξαρτημένης μεταβλητής κάνοντας χρήση κανόνων που απορρέουν από τις ανεξάρτητες. Είναι σαν να γίνεται προσδιορισμός των σημείων τμηματικά.



Εικόνα 16 προσαρμογή δεδομένων χρησιμοποιώντας Decision Tree regression. όσο πιο βαθύ το δέντρο τόσο πιο περίπλοκος ο αλγόριθμος, αλλά τόσο καλύτερο το fitting

Τα πλεονεκτήματά του είναι πολλά. Πρώτα απ' όλα συνήθως χρειάζεται μικρή προετοιμασία των δεδομένων εν συγκρίσει με άλλα μοντέλα που χρειάζονται scaling, standardizing κ.α.. Είναι επίσης εύκολο να γίνει validation με απλά στατιστικά τεστ. Είναι εύκολο να ερμηνευθεί και συμπεριφέρεται γενικά καλά.

Στα μειονεκτήματά του περιλαμβάνεται η ευαισθησία του στο overfitting, εάν το μοντέλο που δημιουργηθεί είναι πολύ περίπλοκο. Υπάρχουν, ωστόσο τρόποι που αφορούν στο tuning του μοντέλου με τους οποίους το προηγούμενο πρόβλημα μπορεί να αποφευχθεί. Είναι, επίσης, αρκετά ασταθής, μιας και μικρές διαφοροποιήσεις στα αρχικά δεδομένα μπορούν να προκαλέσουν μεγάλες αλλαγές στο τελικό μοντέλο. Τέλος, πριν τη χρήση τους προτείνεται η ισοστάθμιση των δεδομένων (balancing), διότι όταν κάποιες μεταβλητές υπερισχύουν το δέντρο βασίζεται κυρίως σε αυτές (developers, 2007-2022).

#### 3.4.3. Ensemble τεχνικές

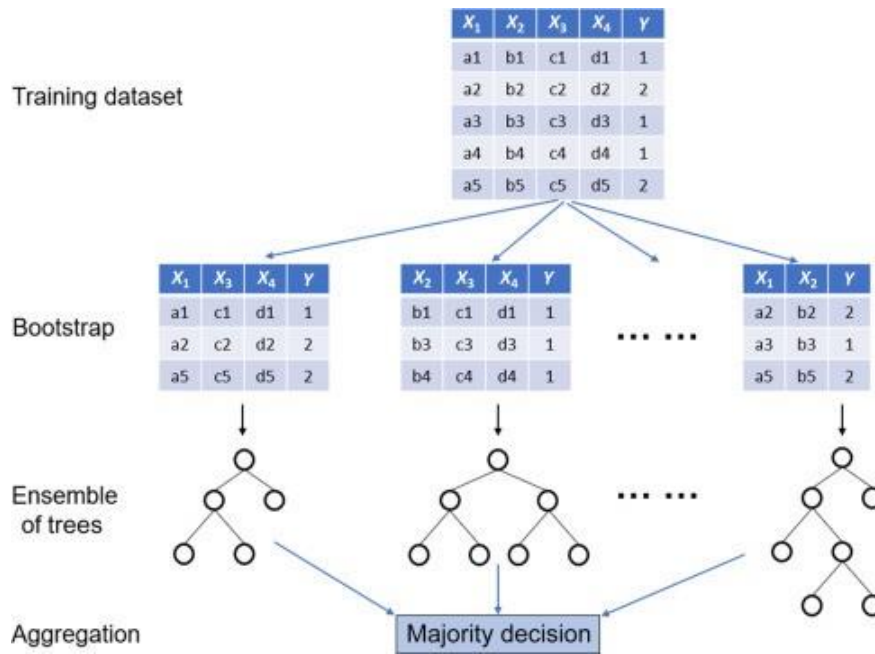
Στόχος των ensemble τεχνικών είναι να συνδυάσουν τις προβλέψεις από πολλούς παράγοντες που προκύπτουν από ένα συγκεκριμένο μοντέλο, με σκοπό να ενισχύσουν τη σταθερότητα του μοντέλου, σε σχέση με το να χρησιμοποιούνταν ένας μόνο παράγοντας.

Οι τεχνικές αυτές χωρίζονται σε δύο κατηγορίες: τις τεχνικές μέσου όρου (averaging) και τις τεχνικές ενίσχυσης (boosting). Στις πρώτες κατασκευάζονται πολλοί εκτιμητές παράγοντες ανεξάρτητα και στη συνέχεια υπολογίζεται ο μέσος όρος των προβλέψεών τους. Μια τέτοια τεχνική είναι ο Extremely Randomized Trees Regressor. Στις δεύτερες οι παράγοντες φτιάχνονται ο ένας μετά τον άλλον, και ένας από αυτούς έχει τον ρόλο του να διορθώνει την κλίση του συνδυασμένου μοντέλου. Παράδειγμα τέτοιας τεχνικής είναι ο Ada Boost Regressor.

##### 3.4.3.1. *Extremely Randomized Trees Regressor*

Σε αυτή τη μέθοδο κάθε δέντρο φτιάχνεται από δείγματα των αρχικών δεδομένων. Σε κάθε κόμβο ο διαχωρισμός κάθε μεταβλητής γίνεται τυχαία και επιλέγεται ο καλύτερος τρόπος διαχωρισμού ως κανόνας. Ο τρόπος αυτός, παρόλο που αυξάνει ελαφρώς την πολικότητα του μοντέλου, προκαλεί μείωση της διακύμανσης. Ο αριθμός των δέντρων καθώς και το μέγεθος των τυχαίων υποσυνόλων των μεταβλητών που λαμβάνονται υπόψη όταν γίνεται ο διαχωρισμός

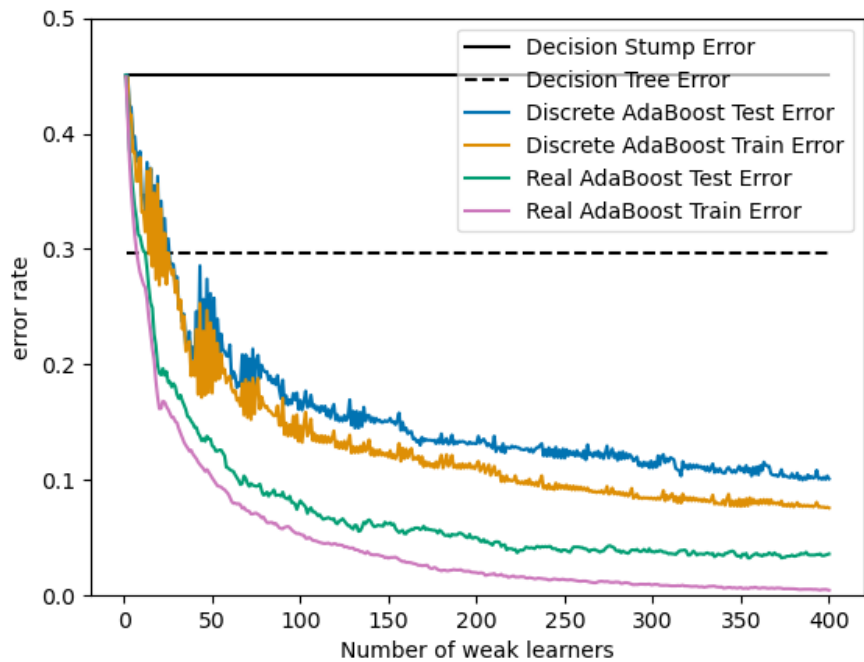
ενός κόμβου είναι οι κύριοι παράγοντες που επηρεάζουν την τεχνική αυτή (developers, 2007-2022).



Εικόνα 17 σχεδιάγραμμα του τρόπου λειτουργίας των Extremely Randomized Trees Regressor

### 3.4.3.2. Ada Boost Regressor

Πρόκειται για μια τεχνική η οποία χρησιμοποιεί πολλά αδύναμα μοντέλα, τα οποία δεν κάνουν αρκετά καλές προβλέψεις, και να τα ενισχύει κάνοντας επαναλαμβανόμενες μετατροπές στα δεδομένα. Οι προβλέψεις των παραπάνω μοντέλων αθροίζονται έχοντας η καθεμία διαφορετικό στατιστικό βάρος  $w_1, w_2, \dots, w_N$ , δημιουργώντας την τελική πρόβλεψη. Στην πρώτη προσέγγιση τα βάρη κάθε δείγματος είναι  $w_i = \frac{1}{N}$ , κάνοντας μια πρόβλεψη των αδύναμων μοντέλων. Μετά από κάθε επανάληψη οι βαρύτητες των δειγμάτων μεταβάλλονται και επαναλαμβάνεται η διαδικασία. Κάθε φορά αυξάνεται η βαρύτητα των δεδομένων εκείνων των οποίων η πρόβλεψη ήταν λανθασμένη, και μειώνεται για τα δεδομένα τα οποία προβλέφθηκαν σωστά, αναγκάζοντας έτσι το μοντέλο να επικεντρώνεται στα «προβληματικά» σημεία του αλγόριθμου (developers, 2007-2022).



Εικόνα 18 διάγραμμα του σφάλματος της πρόβλεψης συναρτήσει των αδύναμων μοντέλων που χρησιμοποιούνται. είναι εμφανής η μείωση του σφάλματος των εκάστοτε αλγορίθμων όταν χρησιμοποιούνται αυτόνομα και όταν συνδυάζονται στον AdaBoost Regressor

# ΕΙΔΙΚΟ ΜΕΡΟΣ

## 4. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

### 4.1. Εισαγωγή

Στο κεφάλαιο αυτό θα αναλυθούν τα λογισμικά, οι τεχνικές και οι αλγόριθμοι που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία. Συγκεκριμένα, θα γίνει αναφορά στο λογισμικό Impact MC και τις παραμέτρους του, όπως αυτά χρησιμοποιήθηκαν για να γίνουν οι προσομοιώσεις των εξετάσεων, καθώς και στις ίδιες τις εξετάσεις των ασθενών. Επίσης, θα παρουσιαστεί και θα εξηγηθεί ο αλγόριθμος τεχνητής νοημοσύνης που δημιουργήθηκε με τη γλώσσα προγραμματισμού Python και το πακέτο Scikit Learn για την πρόβλεψη της δόσης των οργάνων των ασθενών που υποβλήθηκαν σε εξετάσεις θώρακα σε CBCT ακτινοθεραπείας.

Για τις προσομοιώσεις χρησιμοποιήθηκαν εικόνες σάρωσης θώρακα 113 ασθενών, εικόνες σάρωσης άνω κοιλίας 89 ασθενών, οι οποίες παραχωρήθηκαν από το ΠΑ.Γ.Ν.Η..

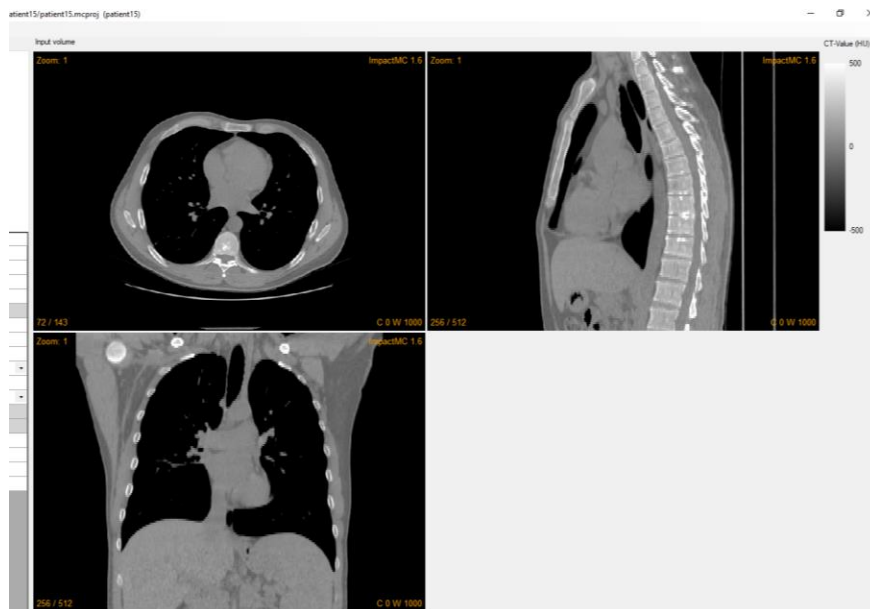
### 4.2. Impact MC

Το Impact MC είναι ένα λογισμικό Monte Carlo προσομοίωσης εικόνων Υπολογιστικής Τομογραφίας το οποίο υπολογίζει τη δόση που έλαβε ο εξεταζόμενος σε κάθε σημείο του FOV. Στο κεφάλαιο 2.2.2. εξηγήθηκε η χρήση Monte Carlo προσομοιώσεων και τα πλεονεκτήματά της. Στις επόμενες παραγράφους θα φανούν οι εικόνες που εισήχθησαν και οι παράμετροι των προσομοιώσεων, καθώς οι εικόνες δόσης που εξάγονται από το λογισμικό.

#### 4.2.1. Εικόνες σάρωσης

Το πρώτο βήμα για να πραγματοποιηθεί μια προσομοίωση με το Impact MC είναι να εισαχθεί μια εικόνα σάρωσης. Από τα dicom headers αυτής της εικόνας, το λογισμικό θα ανακτήσει πληροφορίες όπως το πάχος τομής, η απόσταση μεταξύ των pixels, τις διαστάσεις της εικόνας αλλά και τα ανατομικά χαρακτηριστικά του εξεταζόμενου, προκειμένου να γίνει όσο το δυνατόν πιο ακριβής υπολογισμός της δόσης σε κάθε voxel (Σωτηροπούλου, 2021).

Στις παρακάτω εικόνες φαίνεται το λογισμικό Impact MC όταν έχουν εισαχθεί οι εικόνες σάρωσης.



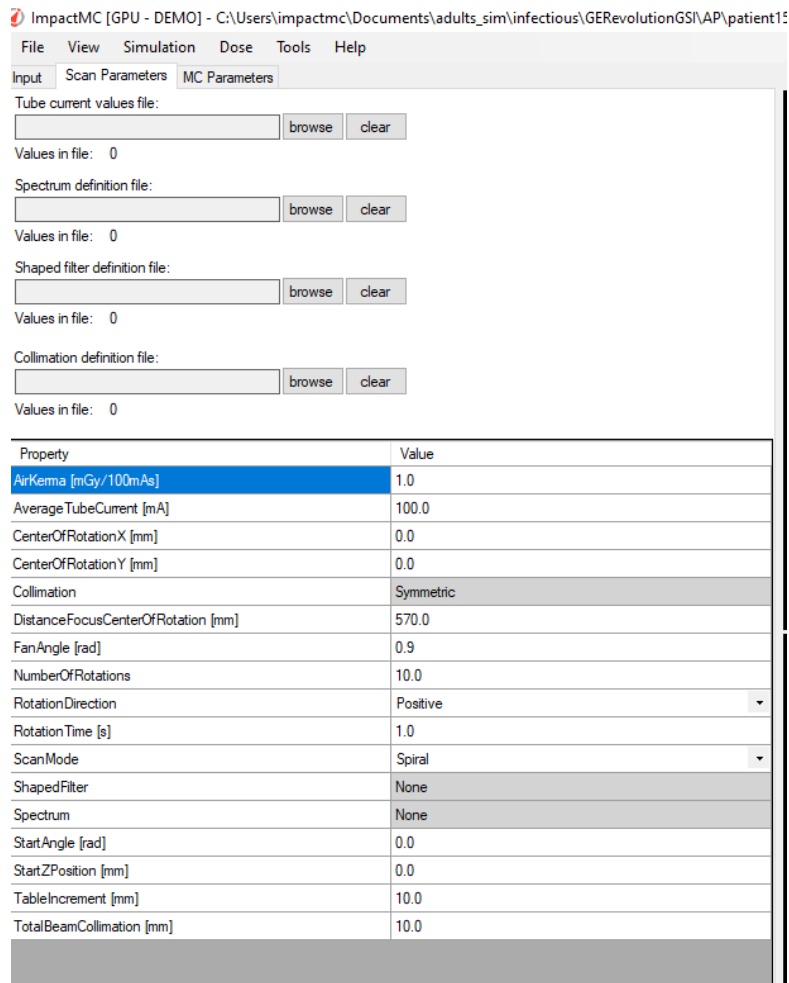
Εικόνα 19 εικόνα σάρωσης θώρακα στο Impact MC. Παρατηρείται ότι φαίνονται η axial, η sagittal και η coronal προβολή του όγκου

#### 4.2.2. Παράμετροι σάρωσης

Όπως πολλές φορές έχει αναφερθεί προηγουμένως, για να γίνει εξατομικευμένη και system-specific δοσιμετρία θα πρέπει να εισαχθούν στο λογισμικό όλα τα στοιχεία του πρωτοκόλλου με τα οποία έγινε η εξέταση. Αυτά περιλαμβάνουν την τιμή της τάσης της λυχνίας, την ένταση του ρεύματος και τον χρόνο που διήρκτησε η σάρωση. Θα πρέπει επίσης να ληφθεί υπόψιν και ο εξοπλισμός που χρησιμοποιήθηκε, δηλαδή τα φίλτρα και οι κατευθυντήρες. Όλα τα προηγούμενα αποτελούν τις παραμέτρους σάρωσης του Impact MC.

Στις παραμέτρους σάρωσης περιλαμβάνονται 4 παράμετροι που προσδιορίζονται από αρχείο και 17 που προσδιορίζονται με μια τιμή. Στην παρακάτω εικόνα φαίνεται αυτές οι παράμετροι.





Εικόνα 20 scan parameters

Για προσομοιώσεις CBCT θώρακα χρησιμοποιήθηκε το πρωτόκολλο που προτείνει η κατασκευαστρια εταιρία, βάση του οποίου εισάγεται τάση 120 kV, ρεύμα μέσης τιμής 20 mA και διάρκεια σάρωσης 13.2 s. Το Air Kerma υπολογίστηκε να είναι 4.92 mGy/100mAs ενώ η απόσταση της λυχνίας από το κέντρο περιστροφής ίση με 1000 mm. Τέλος, το φίλτρο με το οποίο γίνεται η σάρωση είναι το bow tie και ο κατευθυντήρας είναι ο M20 (μεσαίο FOV) (Elekta, 2016).

Οι παράμετροι center of rotation X και Y, καθώς και η start Z position προκύπταν βάση της εικόνας.

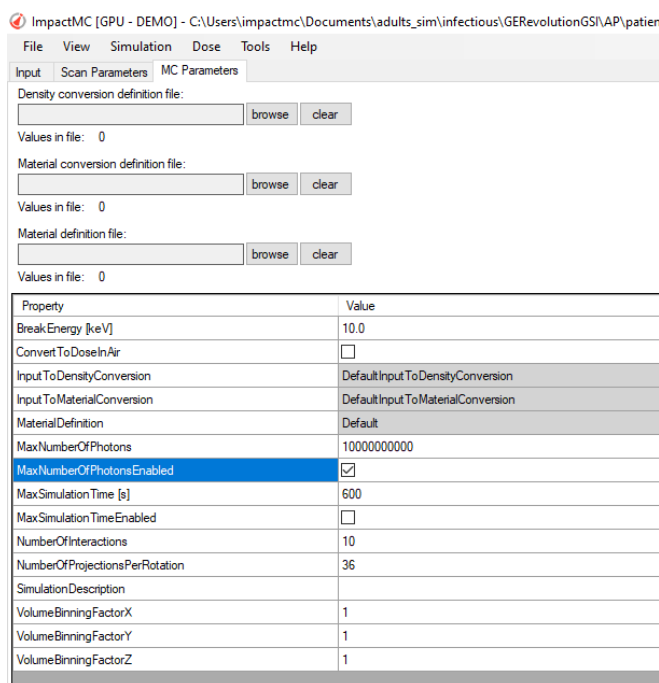
Τέλος, κάποιες παράμετροι ορίζονται από την φύση της εξέτασης, όπως το ότι ο αριθμός περιστροφών είναι πάντα 1, το scan mode είναι sequential και πως η μετατόπιση της τράπεζας είναι 0 (© CT Imaging GmbH, 2016).

Αυτά που αλλάζουν σε εξετάσεις άνω κοιλίας είναι η ένταση του ρεύματος, η οποία σε αυτή την περίπτωση είναι 40 mA και η διάρκεια σάρωσης είναι 26.4s (Elekta, 2016).

#### 4.2.3. Παράμετροι προσομοίωσης

Το επόμενο βήμα της προετοιμασίας μια Monte Carlo προσομοίωσης είναι να οριστούν οι παράμετροι με τις οποίες αυτή θα γίνει. Δηλαδή, θα πρέπει να οριστεί ο αριθμός των φωτονίων, ο αριθμός των προβολών ανά περιστροφή ή το ελάχιστο όριο τάσης της λυχνίας το οποίο θα λαμβάνεται υπόψη.

Οι default τιμές που πραγματοποιούνται οι προσομοιώσεις φαίνονται στην παρακάτω εικόνα.



Εικόνα 21 simulation parameters

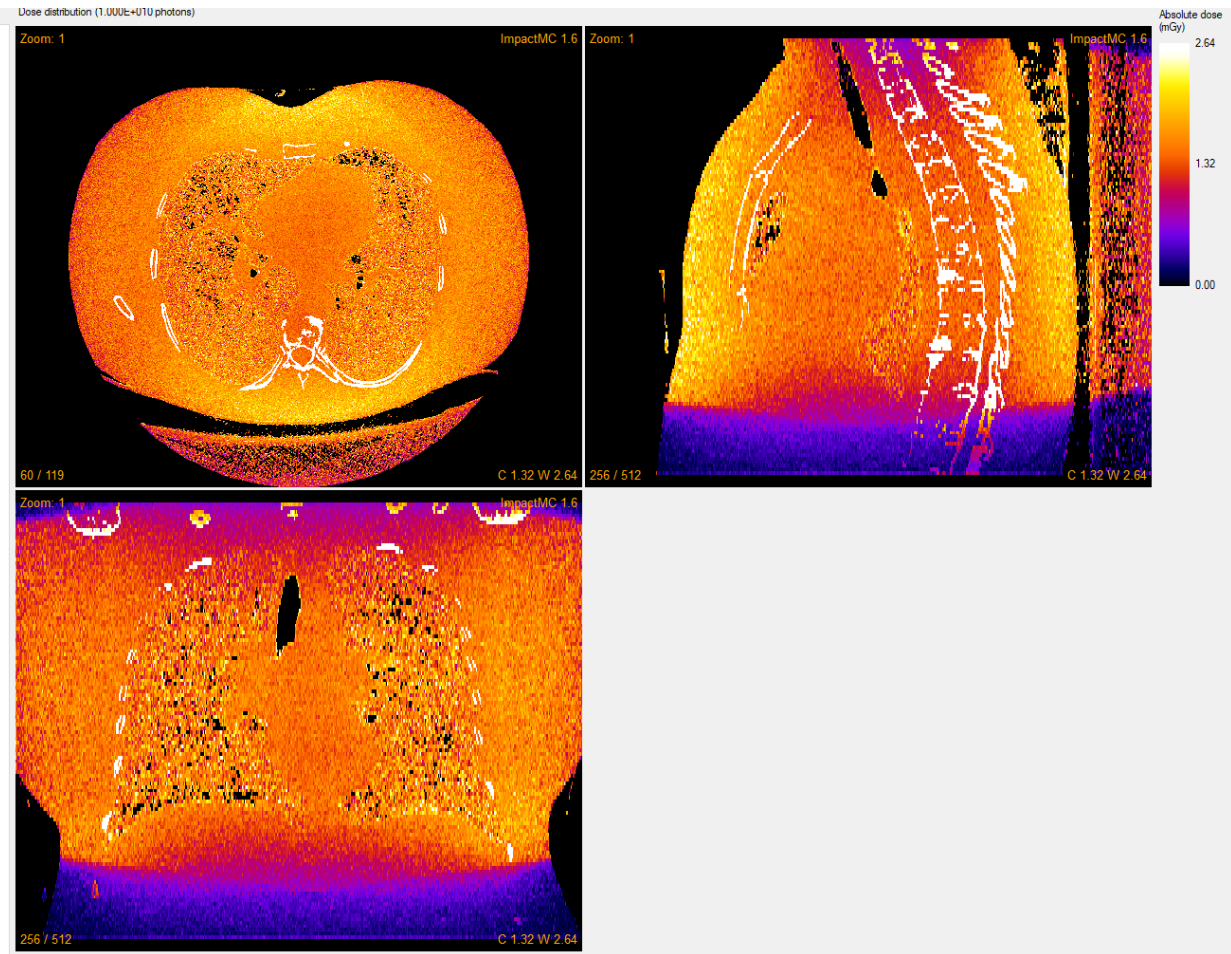
Από το πρωτόκολλο της Elekta η τιμή των προβολών ανά περιστροφή είναι 600 (Elekta, 2016).

Για καλύτερη στατιστική, ο αριθμός των φωτονίων με τα οποία γίνεται το simulation είναι  $10^{10}$ .

Οι υπόλοιπες τιμές μένουν ως έχουν.

#### 4.2.4. Εικόνες εξόδου

Μετά την ολοκλήρωση της προσομοίωσης προκύπτει μια εικόνα δόσης και κάθε voxel της έχει μια τιμή δόσης σε mGy. Η εικόνα αυτή αποθηκεύεται και, κάνοντας χρήση του λογισμικού ImageJ δημιουργούνται ROIs των οργάνων και υπολογίζεται η δόση σε καθένα από αυτά.



Εικόνα 22 εικόνα δόσης όπως προκύπτει από το Impact MC

#### 4.3. Αλγόριθμος τεχνητής νοημοσύνης

Παρακάτω θα φανεί ο αλγόριθμος που δημιουργήθηκε με τη γλώσσα προγραμματισμού Python και το πακέτο Scikit learn για την εκπαίδευση ενός μοντέλου τεχνητής νοημοσύνης το οποίο να προβλέπει τη δόση που λαμβάνουν τα όργανα ενός ασθενή που υποβάλλεται σε CBCT ακτινοθεραπείας για απεικονίσεις θώρακος και άνω κοιλίας. Ο αλγόριθμος αυτός χωρίζεται σε τρία μέρη: στο πρώτο γίνεται περιγραφική στατιστική του σετ δεδομένων με σκοπό να φανούν τυχόν ανωμαλίες, σημεία που οφείλονται σε λάθη κατά την προσομοίωση ή τον υπολογισμό της δόσης

και πρέπει να αφαιρεθούν (outliers) αλλά και οι συσχετίσεις μεταξύ της δόσης και των dose predictors που χρησιμοποιούνται. Οι dose predictors που θα χρησιμοποιηθούν είναι η τάση της λυχνίας (kV), η ένταση του ρεύματος (mA), η διάρκεια της σάρωσης (s), οι αξιαί διαστάσεις του σώματος του ασθενή (X και Y), το μέγεθος του οργάνου και το Water Equivalent Diameter (wed). Το δεύτερο μέρος αφορά το training, δηλαδή την επιλογή του καλύτερου μοντέλου πρόβλεψης και η βελτιστοποίησή του (tuning) είναι το τρίτο. Το τέταρτο και τελευταίο κομμάτι είναι το evaluation, ο έλεγχος, δηλαδή, της πρόβλεψης.

Πριν γίνουν τα παραπάνω, πρέπει να εισαχθούν όλες οι εντολές και τα πακέτα που θα χρησιμοποιηθούν και να οριστεί το dataset και οι τιμές των μεταβλητών. Αυτό γίνεται όπως φαίνεται παρακάτω.

```
import numpy
import statistics
import statsmodels.api as sm
import scipy
from scipy import stats
from numpy import arange
from matplotlib import pyplot
import pandas
from pandas import read_csv
from pandas import set_option
from pandas.plotting import scatter_matrix
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import RandomizedSearchCV
from sklearn.model_selection import RepeatedKFold
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn.linear_model import ElasticNet
from sklearn.linear_model import ElasticNetCV
from sklearn.tree import DecisionTreeRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.neighbors import LocalOutlierFactor
from sklearn.svm import SVR
from sklearn.pipeline import Pipeline
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.ensemble import ExtraTreesRegressor
from sklearn.ensemble import AdaBoostRegressor
from sklearn.metrics import mean_squared_error
filename=r"D:\thesis\csv_files\cbct\chest\csvs\heart.csv"
names=['kV', 'mA', 's', 'x', 'y', 'size', 'wed', 'dose']
dataset=read_csv(filename, names=names)
```

Εικόνα 23 εισαγωγή εντολών και dataset

#### 4.3.1. Περιγραφική στατιστική

Πρώτα απ' όλα, βοηθάει να δει κανείς τα αρχικά δεδομένα στην αρχική τους μορφή, ώστε να αποκτήσει μια εικόνα για την συνέχεια. Σημαντικά είναι και τα χαρακτηριστικά του dataset, καθώς και μια πρώτη εικόνα των συσχετίσεων μεταξύ των μεταβλητών.

```
cbct_heart.py - D:\thesis\εργασία\cbct_heart.py (3.10.5)
File Edit Format Run Options Window Help
print (dataset.head(20) )
print (dataset.describe() )
print (dataset.corr(method='pearson'))
```

Εικόνα 24 εντολές για προβολή των 20 πρώτων γραμμών του dataset, των χαρακτηριστικών του, αλλά και των συσχετίσεων μεταξύ των μεταβλητών βάση της μεθόδου Pearson

```
IDLE Shell 3.10.5
File Edit Shell Debug Options Window Help
      kV  mA  s      x      y      size      wed      dose
0  120  20  13.2  387.00  283.45  493000.0  286.94  7.4583
1  120  20  13.2  404.25  264.25  288000.0  292.86  6.9377
2  120  20  13.2  324.93  262.36  510000.0  260.31  8.6443
3  120  20  13.2  380.94  255.06  541000.0  274.59  7.5729
4  120  20  13.2  346.83  197.53  578000.0  230.98  9.8368
5  120  20  13.2  390.14  250.68  849000.0  282.23  7.4449
6  120  20  13.2  368.37  241.95  744000.0  261.71  8.6810
7  120  20  13.2  286.17  217.27  397000.0  223.54  10.1070
8  120  20  13.2  346.17  267.42  364000.0  264.41  8.9120
9  120  20  13.2  352.91  248.75  735000.0  262.44  8.7431
10 120  20  13.2  433.28  308.44  672000.0  316.02  6.4637
11 120  20  13.2  326.28  220.86  426000.0  269.57  7.6336
12 120  20  13.2  299.75  235.03  540000.0  258.88  8.2755
13 120  20  13.2  366.27  254.28  592000.0  281.32  7.4465
14 120  20  13.2  332.87  251.37  493000.0  246.96  10.0570
15 120  20  13.2  369.44  262.96  843000.0  287.58  6.9015
16 120  20  13.2  410.16  280.27  603000.0  310.89  6.5811
17 120  20  13.2  349.89  265.89  355000.0  266.11  8.5947
18 120  20  13.2  354.41  230.78  322000.0  255.29  9.3641
19 120  20  13.2  362.30  276.37  467000.0  283.24  7.0836
      kV  mA  s      x      y      size      wed      dose
count  110.0  110.0  110.0  ...  1.100000e+02  110.000000  110.000000
mean   120.0  20.0  13.2  ...  5.899273e+05  264.118364  8.526233
std     0.0  0.0  0.0  ...  1.639767e+05  31.290234  1.354770
min     120.0  20.0  13.2  ...  2.880000e+05  154.370000  5.406200
25%    120.0  20.0  13.2  ...  4.625000e+05  244.897500  7.584925
50%    120.0  20.0  13.2  ...  5.645000e+05  263.735000  8.584500
75%    120.0  20.0  13.2  ...  7.117500e+05  284.220000  9.446025
max     120.0  20.0  13.2  ...  1.100000e+06  350.130000  12.779000

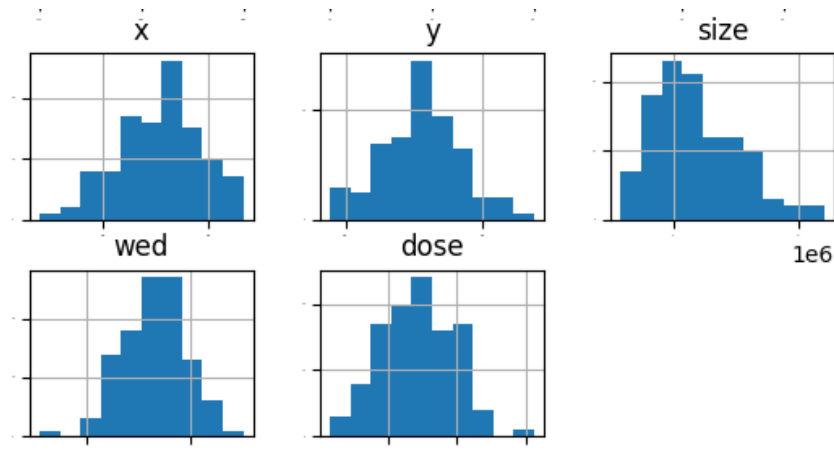
[8 rows x 8 columns]
      kV  mA  s      x      y      size      wed      dose
kV  NaN NaN NaN      NaN      NaN      NaN      NaN      NaN
mA  NaN NaN NaN      NaN      NaN      NaN      NaN      NaN
s   NaN NaN NaN      NaN      NaN      NaN      NaN      NaN
x   NaN NaN NaN  1.000000  0.763629  0.305675  0.810083 -0.791791
y   NaN NaN NaN  0.763629  1.000000  0.324573  0.799744 -0.683103
size NaN NaN NaN  0.305675  0.324573  1.000000  0.358695 -0.414426
wed NaN NaN NaN  0.810083  0.799744  0.358695  1.000000 -0.824523
dose NaN NaN NaN -0.791791 -0.683103 -0.414426 -0.824523  1.000000
```

Εικόνα 25 output των παραπάνω εντολών

Στην παράγραφο 3.2. φάνηκαν και εξηγήθηκαν τα διαγράμματα που είναι χρήσιμα για την καλύτερη κατανόηση του dataset. Τα διαγράμματα αυτά, καθώς και ο κώδικας με τον οποίο δημιουργήθηκαν φαίνονται παρακάτω.

```
dataset.hist(sharex=False, sharey=False, xlabelsize=1, ylabelsize=1)
plt.show()
dataset.plot(kind='box', subplots=True, layout=(4,4), sharex=False, sharey=False, fontsize=8)
plt.show()
scatter_matrix(dataset)
plt.show()
fig=plt.figure()
ax=fig.add_subplot(111)
cax=ax.matshow(dataset.corr(), vmin=-1, vmax=1, interpolation='none')
fig.colorbar(cax)
ticks=np.arange(0,5,1)
ax.set_xticks(ticks)
ax.set_yticks(ticks)
ax.set_xticklabels(names)
ax.set_yticklabels(names)
plt.show()
```

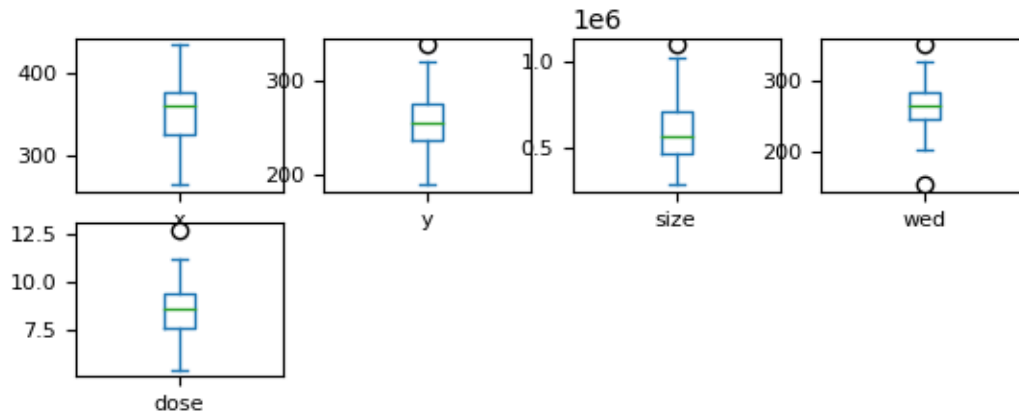
Εικόνα 26 εντολές για τη δημιουργία ιστογραμμάτων, boxplots, διαγραμμάτων διασποράς και διαγραμμάτων συσχετίσεων



Εικόνα 27 ιστογράμματα μεταβλητών

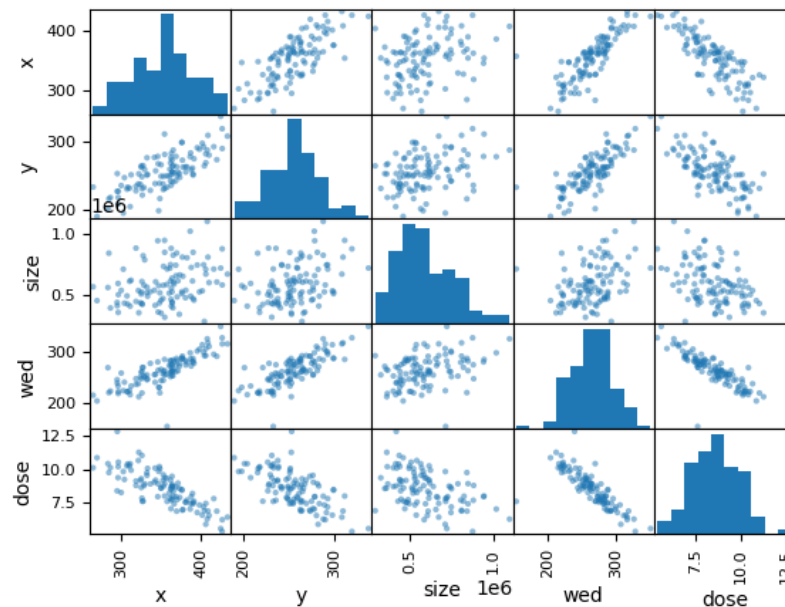
Από τό προηγούμενο διάγραμμα παρατηρείται πως πιθανόν να μην χρειαστεί κάποιος μετασχηματισμός, αφού όλες ακολουθούν την κανονική κατανομή.

Στο επόμενο διάγραμμα φαίνονται τα boxplots. Με κύκλους φαίνονται τα outliers.



Εικόνα 28 boxplots των μεταβλητών. Με κύκλους φαίνονται τα outliers

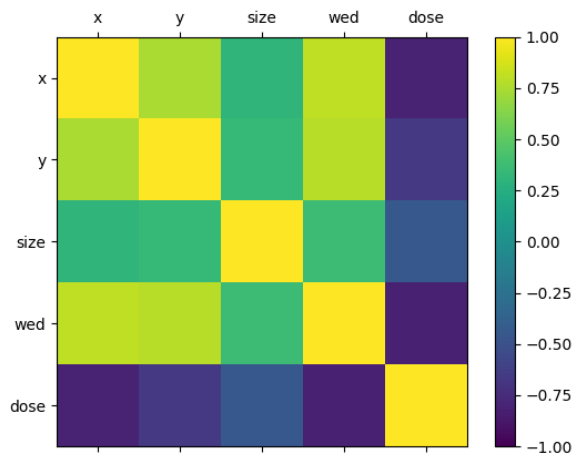
Το επόμενο διάγραμμα είναι ένα διάγραμμα διασποράς που περιγράφει τις συσχετίσεις Pearson.



Εικόνα 29 scatter plots

Παρατηρείται ότι η δόση έχει μια καλή γραμμική εξάρτηση με τις μεταβλητές x, y, wed, εκτός από κάποια σημεία τα οποία πιθανόν είναι outliers και θα πρέπει να αφαιρεθούν. Από τη διασπορά της μεταβλητής size δεν φαίνεται να υπάρχει μεγάλη συσχέτιση με τη δόση.

Ο πίνακας συσχέτισης Pearson αποτυπώνει και αυτός τις συσχετίσεις με τη μέθοδο Pearson.



Εικόνα 30 correlation matrix

Και από εδώ φαίνεται πως η δόση είναι αντιστρόφως ανάλογη των υπόλοιπων μεταβλητών.

#### 4.3.2. Εκπαίδευση αλγορίθμου

Αφού δόθηκε μια εικόνα των πρωτογενών δεδομένων, το επόμενο βήμα είναι η προετοιμασία αυτών και η εκπαίδευση του αλγορίθμου. Αρχικά θα πρέπει να οριστεί ποιες είναι οι ανεξάρτητες μεταβλητές (features) και ποια η εξαρτημένη (output). Όπως έχει ήδη αναφερθεί, τα features είναι οι dose predictors και η output μεταβλητή η δόση του οργάνου.

Στη συνέχεια, πρέπει να αφαιρεθούν τα outliers. Για τον σκοπό αυτό χρησιμοποιείται η εντολή Local Outlier Factor του πακέτου scikit learn, ο οποίος εντοπίζει και αφαιρεί τα σημεία εκείνα των οποίων η πυκνότητα γειτονικών σημείων είναι μικρότερη από αυτή των γειτόνων του (developers, 2007-2022).

Αφού γίνουν τα παραπάνω το dataset είναι έτοιμο να χωριστεί σε training και validation τμήματα, με ποσοστά 70% και 30%, αντίστοιχα.



```

---
array=dataset.values
X=array[:,0:7]
Y=array[:,7]
print(X.shape, Y.shape)
lof=LocalOutlierFactor()
yhat=lof.fit_predict(X)
mask=yhat !=-1
X, Y=X[mask, :], Y[mask]
print(X.shape, Y.shape)
validation_size=0.30
seed=7
X_train, X_validation, Y_train, Y_validation=train_test_split(X, Y, test_size=validation_size, random_state=seed)

```

Εικόνα 31 εφαρμογή του Local Outlier Factor και split του dataset

```

(110, 7) (110,)
(103, 7) (103,)

```

Εικόνα 32 output του Local Outlier Factor

Φαίνεται πως από τις 110 γραμμές που υπήρχαν αρχικά αφαιρέθηκαν 7, λόγω του ότι ένα ή περισσότερα features σε αυτή την γραμμή θεωρήθηκαν outliers.

Τώρα θα δοκιμαστούν όλα τα μοντέλα που αναφέρθηκαν στην παράγραφο 3.4., ώστε να βρεθεί εκείνο που δίνει το μικρότερο σφάλμα, και άρα κάνει και την καλύτερη πρόβλεψη. Το training set θα χωριστεί σε 10 επιπλέον τμήματα και σε κάθε επανάληψη θα δοκιμάζεται καθένα από αυτά σαν validation set, ενώ τα υπόλοιπα θα χρησιμοποιούνται σαν training set (developers, 2007-2022). Η αξιολόγηση της επίδοσης του μοντέλου θα γίνει με τα αρνητικά ελάχιστα τετράγωνα των σφαλμάτων. Όσο πιο κοντά η τιμή αυτή στο 0, τόσο καλύτερη η πρόβλεψη του μοντέλου.

Πολλές φορές αναφέρθηκε πως υπάρχει περίπτωση τα δεδομένα να χρειάζονται μετατροπές ώστε το μοντέλο να συμπεριφερθεί με τον καλύτερο δυνατό τρόπο. Γι' αυτό το λόγο, τα μοντέλα θα δοκιμαστούν και στα κανονικοποιημένα δεδομένα.

```

num_folds=10
seed=7
scoring='neg_mean_squared_error'
models=[]
models.append(('LR', LinearRegression()))
models.append(('LASSO', Lasso()))
models.append(('EN', ElasticNet()))
models.append(('KNN', KNeighborsRegressor()))
models.append(('CART', DecisionTreeRegressor()))
models.append(('SVR', SVR()))
results=[]
names=[]
for name, model in models:
    kfold=KFold(n_splits=num_folds, random_state=seed, shuffle=True)
    cv_results=cross_val_score(model, X_train, Y_train, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)
fig=pyplot.figure()
fig.suptitle('Algorithm Comparison')
ax=fig.add_subplot(111)
pyplot.boxplot(results)
ax.set_xticklabels(names)
pyplot.show()
pipelines=[]
pipelines.append(('ScaledLR', Pipeline([('Scaler', StandardScaler()), ('LR', LinearRegression())]))
pipelines.append(('ScaledLASSO', Pipeline([('Scaler', StandardScaler()), ('LASSO', Lasso())]))
pipelines.append(('ScaledEN', Pipeline([('Scaler', StandardScaler()), ('EN', ElasticNet())]))
pipelines.append(('ScaledKNN', Pipeline([('Scaler', StandardScaler()), ('KNN', KNeighborsRegressor())]))
pipelines.append(('ScaledCART', Pipeline([('Scaler', StandardScaler()), ('CART', DecisionTreeRegressor())]))
pipelines.append(('ScaledSVR', Pipeline([('Scaler', StandardScaler()), ('SVR', SVR())]))
results=[]
names=[]
for name, model in pipelines:
    kfold=KFold(n_splits=num_folds, random_state=seed, shuffle=True)
    cv_results=cross_val_score(model, X_train, Y_train, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg="%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)

```

---

```

names=[]
for name, model in pipelines:
    kfold=KFold(n_splits=num_folds, random_state=seed, shuffle=True)
    cv_results=cross_val_score(model, X_train, Y_train, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg="%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)
ensembles=[]
ensembles.append(('AB', AdaBoostRegressor()))
ensembles.append(('GBM', GradientBoostingRegressor()))
ensembles.append(('RF', RandomForestRegressor()))
ensembles.append(('ET', ExtraTreesRegressor()))
results=[]
names=[]
for name, model in ensembles:
    kfold=KFold(n_splits=num_folds, shuffle=True, random_state=seed)
    cv_results=cross_val_score(model, X_train, Y_train, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg="%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)
ensembles=[]
ensembles.append(('ScaledAB', Pipeline([('Scaler', StandardScaler()), ('AB', AdaBoostRegressor())]))
ensembles.append(('ScaledGBM', Pipeline([('Scaler', StandardScaler()), ('GBM', GradientBoostingRegressor())]))
ensembles.append(('ScaledRF', Pipeline([('Scaler', StandardScaler()), ('RF', RandomForestRegressor())]))
ensembles.append(('ScaledET', Pipeline([('Scaler', StandardScaler()), ('ET', ExtraTreesRegressor())]))
results=[]
names=[]
for name, model in ensembles:
    kfold=KFold(n_splits=num_folds, shuffle=True, random_state=seed)
    cv_results=cross_val_score(model, X_train, Y_train, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg="%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)

```

Εικόνα 33 δοκιμή των μοντέλων

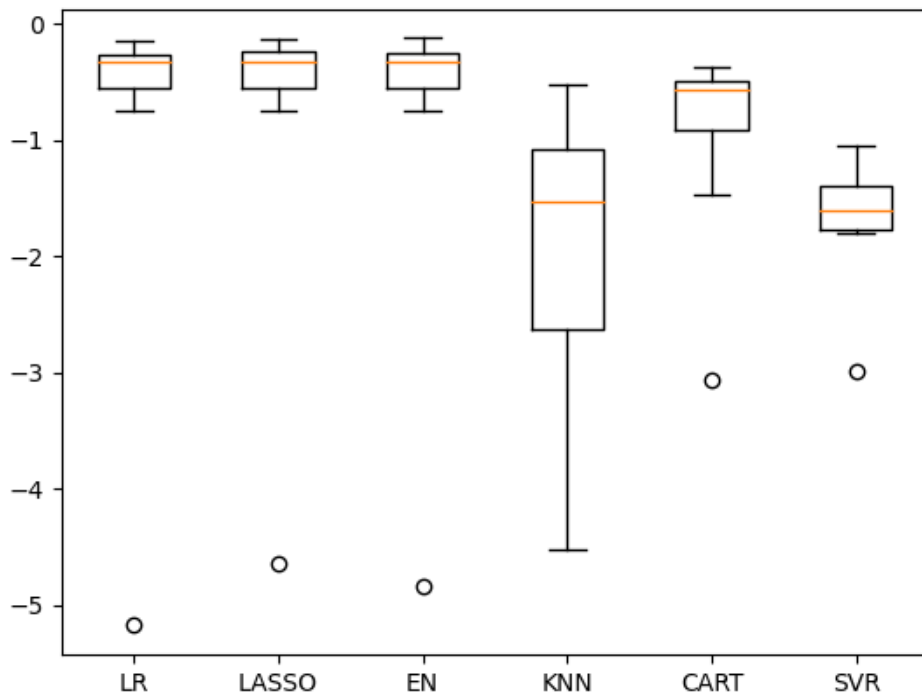
```

LR: -0.848084 (1.451966)
LASSO: -0.789161 (1.294339)
EN: -0.807728 (1.354150)
KNN: -1.922779 (1.175245)
CART: -0.899973 (0.754420)
SVR: -1.656813 (0.507253)
ScaledLR: -0.848084 (1.451966)
ScaledLASSO: -1.679914 (0.762400)
ScaledEN: -1.044241 (0.691033)
ScaledKNN: -0.651257 (0.798084)
ScaledCART: -0.919873 (0.758435)
ScaledSVR: -0.595706 (0.570449)
AB: -0.609447 (0.722361)
GBM: -0.740110 (0.661056)
RF: -0.639338 (0.748058)
ET: -0.607386 (0.647633)
ScaledAB: -0.674717 (0.722834)
ScaledGMB: -0.747751 (0.671164)
ScaledRF: -0.626093 (0.738303)
ScaledET: -0.596681 (0.645508)

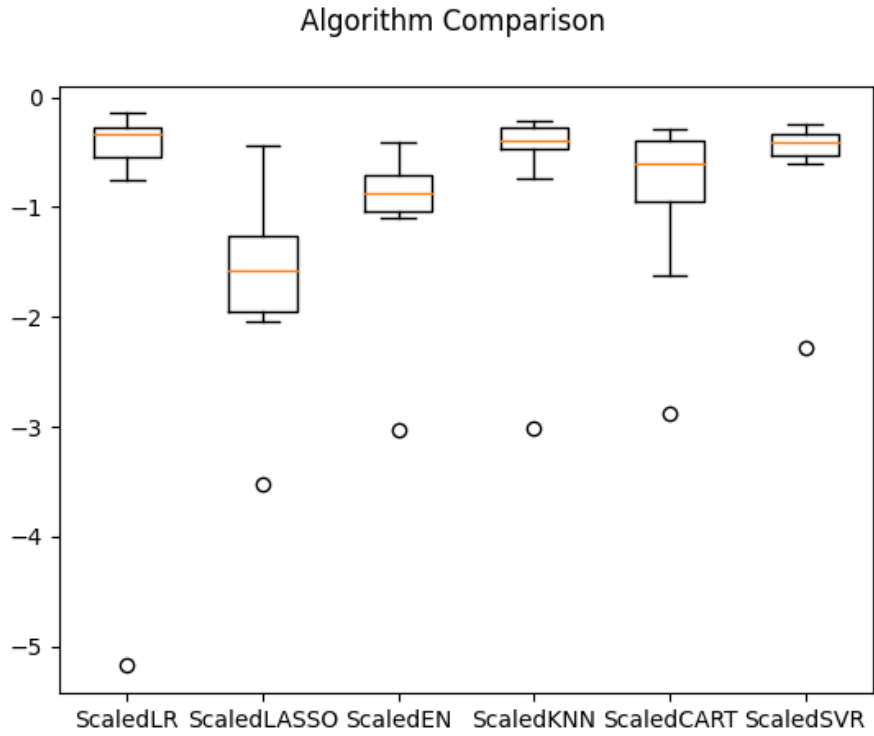
```

Εικόνα 34 negative mean squared errors των μοντέλων

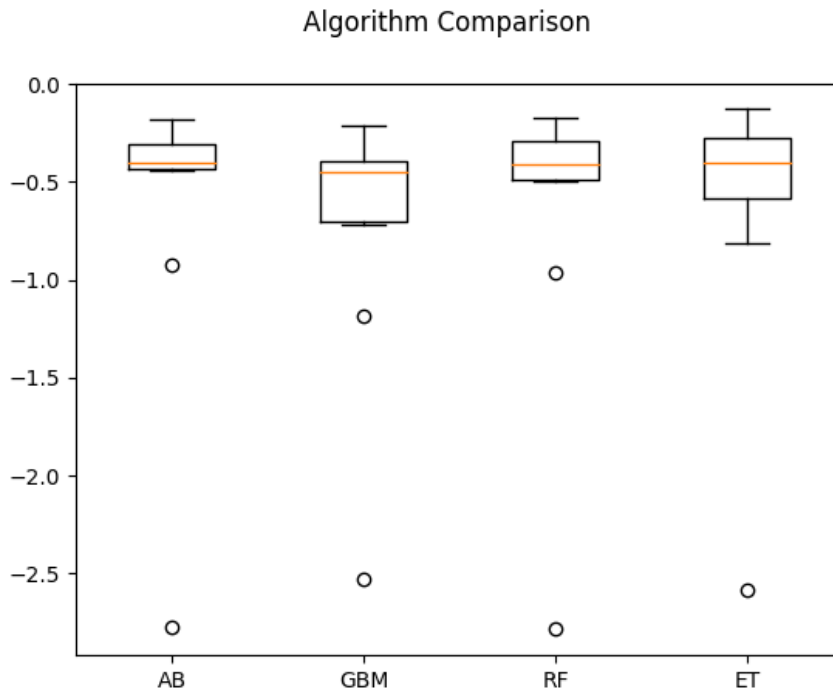
### Algorithm Comparison



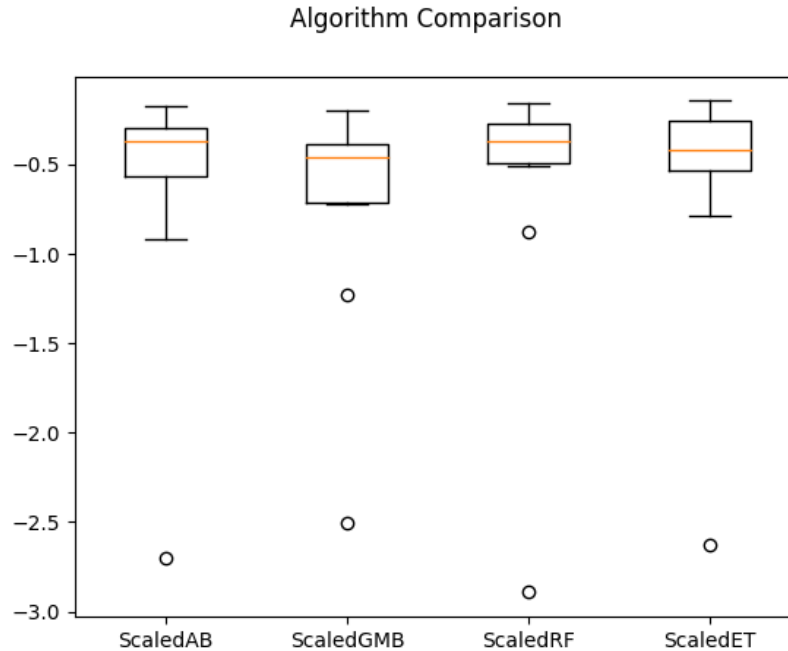
Εικόνα 35 boxplot των γραμμικών μοντέλων



Εικόνα 36 boxplot των γραμμικών μοντέλων για κανονικοποιημένα δεδομένα



Εικόνα 37 boxplot των ensemble μεθόδων



Εικόνα 38 boxplot των ensemble μεθόδων για κανονικοποιημένα δεδομένα

Παρατηρείται ότι όλες οι κατανομές του scoring των μοντέλων είναι πολύ κοντά στο μηδέν, ενώ είναι και αρκετά στενές (μικρή διακύμανση). Η εφαρμογή του γραμμικού μοντέλου Lasso πάνω στα κανονικοποιημένα δεδομένα έχει το πιο κοντινό στο μηδέν scoring, οπότε αυτό είναι που θα επιλεγεί ως μοντέλο εκπαίδευσης.

#### 4.3.3. Tuning

Όπως έχει ήδη γίνει γνωστό, κάθε μοντέλο επηρεάζεται από κάποιες υπερπαραμέτρους, οι οποίες δεν ταιριάζουν πάντα στο δεδομένο πρόβλημα. Η εύρεση των βέλτιστων υπερπαραμέτρων του μοντέλου και η αλλαγή τους μπορεί να βελτιώσει την επίδοση του αλγορίθμου. Η διαδικασία αυτή ονομάζεται tuning.

##### 4.3.3.1. Tuning των Lasso και Elastic Net

Τα δύο αυτά μοντέλα έχουν ως κύρια υπερπαραμέτρο την σταθερά «α». Επιστρέφοντας στο προηγούμενο παράδειγμα, προέκυψε πως η τεχνική Lasso με score -0.789 είναι μάλλον αυτή που ταιριάζει καλύτερα στο πρόβλημα. Μπορεί, όμως, να βελτιωθεί επιλέγοντας το κατάλληλο «α».

```

scaler=StandardScaler()
rescaledX=scaler.fit_transform(X_train)
rescaledValidationX=scaler.transform(X_validation)
model=Lasso()
grid=dict()
grid['alpha']=arange(0,1,0.01)
grid=GridSearchCV(model, grid, scoring=scoring, cv=kfold, n_jobs=-1)
grid_results=grid.fit(X_train, Y_train)
print(grid_results.best_params_)
print(grid_results.best_score_)

```

Εικόνα 39 data scaling και tuning του Lasso

```

{'alpha': 0.99}
-0.7894096083043991

```

Εικόνα 40 στην πρώτη γραμμή φαίνεται η καλύτερη τιμή για το "α", ενώ στη δεύτερη το καλύτερο score

Δεν φαίνεται να αλλάζει πολύ το score, ίσως επειδή το σφάλμα είναι ήδη μικρό, ωστόσο θα είναι αισθητή η διαφορά στο τελικό αποτέλεσμα, όπως θα φανεί στο τρίτο μέρος του αλγόριθμου.

Με τον ίδιο τρόπο γίνεται το tuning και του Elastic Net, αλλάζοντας την γραμμή του κώδικα model=Lasso() σε model=ElasticNet().

#### 4.3.3.2. Tuning του AdaBoost Regressor

Στην παράγραφο 3.4.3.2. έχει εξηγηθεί το πώς ο αριθμός των αδύναμων μοντέλων επηρεάζει τον AdaBoost. Αυτή είναι και η κύρια υπερπαραμέτρος του και ορίζεται ως n\_estimators. Στον παρακάτω αλγόριθμο, η τεχνική αυτή έδινε score 9.98 για n\_estimators=50 που είναι και η default τιμή. Το tuning έγινε δοκιμάζοντας τυχαίες τιμές της υπερπαραμέτρου, έως ότου να επιτευχθεί βελτίωση του score.

```

param_grid=dict(n_estimators=np.array([1, 100, 500, 1000, 10000, 150000]))
model=AdaBoostRegressor(random_state=seed)
kfold=KFold(n_splits=num_folds, random_state=seed, shuffle=True)
grid=GridSearchCV(estimator=model, param_grid=param_grid, scoring=scoring, cv=kfold)
grid_results=grid.fit(X_train, Y_train)
print("Best:%f using %s" %(grid_results.best_score_, grid_results.best_params_))

```

Εικόνα 41 κώδικας για το tuning του AdaBoost Regressor. Στην πρώτη σειρά φαίνονται οι τυχαίες τιμές που χρησιμοποιήθηκαν

```

Best:-8.131539 using {'n_estimators': 500}

```

Εικόνα 42 output του παρπάνω κώδικα

Είναι εμφανής η μείωση του σφάλματος με την αλλαγή του αριθμού των αδύναμων μοντέλων.

#### 4.3.3.3. Tuning του Extremely Randomized Trees Regressor

Σε αυτή τη μέθοδο, οι δύο κύριες υπερπαραμέτροι είναι ο αριθμός των δέντρων (`n_estimators`) και το μέγεθος των τυχαίων υποσυνόλων των μεταβλητών που λαμβάνονται υπόψη όταν γίνεται ο διαχωρισμός ενός κόμβου (`max_features`). Οι default τιμές είναι `n_estimators=100` και `max_features=1` (developers, 2007-2022). Για ένα συγκεκριμένο dataset, το score με αυτές τις τιμές είναι -1.63. δοκιμάζοντας τυχαίες τιμές εύρους από 1 μέχρι 100 με βήμα 1 και για τις δύο υπερπαραμέτρους γίνεται προσπάθεια βελτίωσης αυτής της τιμής.

```
model=ExtraTreesRegressor()
grid=GridSearchCV(
    estimator=model,
    param_grid={
        'n_estimators':range(1,100,1),
        'max_features':range(1,100,1),
    },
    scoring=scoring,
    cv=kfold
)
grid_results=grid.fit(X_train, Y_train)
print("best: %f using %s" %(grid_results.best_score_, grid_results.best_params_))
```

Εικόνα 43 κώδικας για το tuning του Extra Trees Regressor

Ο καλύτερος συνδυασμός προκύπτει να είναι `max_features=87` και `n_estimators=4` με score 1.34, οριακά μικρότερο από το προηγούμενο.

#### 4.3.4. Validation του μοντέλου

Το τελευταίο μέρος του αλγόριθμου είναι η εφαρμογή του βελτιστοποιημένου μοντέλου στο validation dataset για την πρόβλεψη της δόσης και η αξιολόγηση της τιμής που προκύπτει. Για τον λόγο αυτό γίνεται σύγκριση των τιμών δόσης που προβλέπονται με τις αρχικές. Ο πιο ενδεδειγμένος τρόπος σύγκρισης ενός μοντέλου με ένα μοντέλο αναφοράς είναι τα διαγράμματα Bland-Altman που παρουσιάζουν τις διαφορές των προβλέψεων από τις πραγματικές τιμές συναρτήσει των πραγματικών τιμών. Η ποσοστιαία διαφορά κάθε ζεύγους πραγματικής τιμής-πρόβλεψης ως  $difference = \frac{Y_{validation} - prediction}{Y_{validation}} 100\%$ . Στο διάγραμμα εμφανίζονται τρεις ευθείες: μία που δείχνει τη μέση τιμή των διαφορών και βρίσκεται στη μέση του διαγράμματος, και δύο που εμφανίζουν τις τιμές  $\pm 1.96SD$ , όπου SD η τυπική απόκλιση των διαφορών. Αυτές οι δύο είναι που καθορίζουν και τα όρια συμφωνίας των προβλέψεων με τις πραγματικές τιμές.

Για να θεωρηθεί πως υπάρχει συμφωνία θα πρέπει το 95% των σημείων να βρίσκεται μεταξύ αυτών των ορίων (Krouwer, 2008).

Η μέγιστη διαφορά παίζει επίσης σημαντικό ρόλο και θα πρέπει να ελέγχεται σε κάθε μοντέλο, καθώς και η περιγραφική στατιστική των συγκρινόμενων τιμών, ώστε να εντοπιστούν οι τυχόν ασυμφωνίες.

Ο κώδικας με τον οποίο κατασκευάζονται τα διαγράμματα και υπολογίζονται οι τιμές που αναφέρθηκαν φαίνονται στην παρακάτω εικόνα.

```
model.fit(X_train, Y_train)
predictions=model.predict(X_validation)
dif=(Y_validation-predictions)/Y_validation*100
msg="y_validation min: %f" % (min(Y_validation))
print(msg)
msg="y_validation max: %f" % (max(Y_validation))
print(msg)
msg="y_validation median: %f" % (statistics.median(Y_validation))
print(msg)
msg="y_validation mean: %f" % (statistics.mean(Y_validation))
print(msg)
msg="predictions min: %f" % (min(predictions))
print(msg)
msg="difference max: %f" % (max(dif))
print(msg)
x=(Y_validation)
y=(dif)
md=numpy.mean(y)
sd=numpy.std(y,axis=0)
CI_low= md - 1.96*sd
CI_high  = md + 1.96*sd
pyplot.scatter(x, y)
pyplot.axhline(md, color='gray', linestyle='-')
pyplot.axhline(md + 1.96*sd, color='gray', linestyle='--')
pyplot.axhline(md - 1.96*sd, color='gray', linestyle='--')
pyplot.title(r"$\mathbf{Bland-Altman}$" + " " + r"$\mathbf{Plot}$")
pyplot.xlabel("Y_validation")
pyplot.ylabel("% Difference")
xOutPlot = numpy.min(x) + (numpy.max(x)-numpy.min(x))*1.14
pyplot.text(xOutPlot, md - 1.96*sd,
            r'-1.96SD:' + "\n" + "%2f" % CI_low,
            ha = "center",
            va = "center",
            )
pyplot.text(xOutPlot, md + 1.96*sd,
            r'+1.96SD:' + "\n" + "%2f" % CI_high,
            ha = "center",
            va = "center",
            )

pyplot.text(xOutPlot, md,
            r'Mean:' + "\n" + "%2f" % md,
            ha = "center",
            va = "center",
            )
pyplot.subplots_adjust(right=0.85)
pyplot.show()
```

Εικόνα 44 πρόβλεψη δόσης και σύγκρισή της με τις πραγματικές τιμές



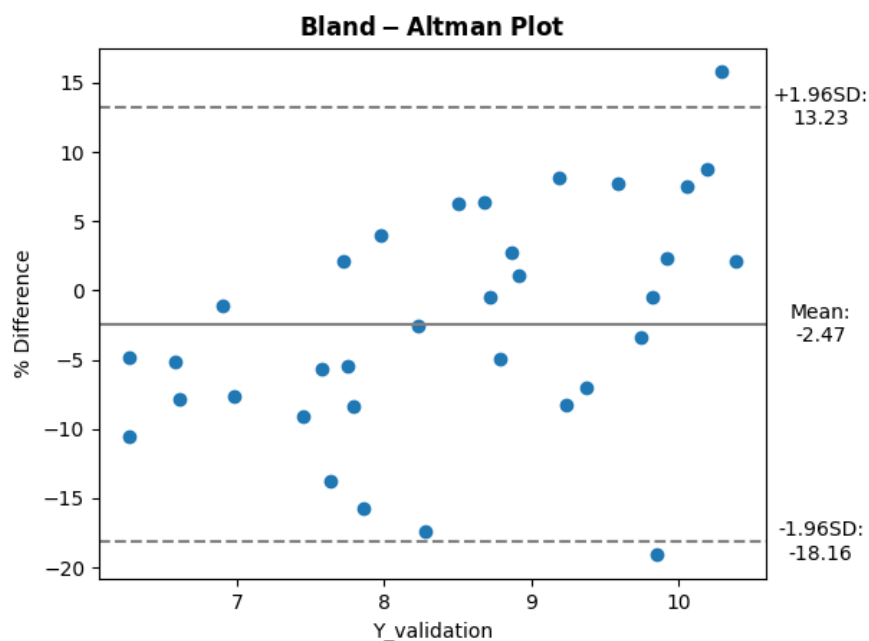
```

y_validation min: 6.267900
y_validation max: 10.387000
y_validation median: 8.590050
y_validation mean: 8.470109
predictions min: 6.575123
difference max: 15.777683

```

Εικόνα 45 περιγραφική στατιστική των τμών  $Y_{validation}$  και των προβλέψεων

Στον παραπάνω πίνακα φαίνεται ότι η μέγιστη τιμή των διαφορών είναι κοντά στο 16%, κάτι που προϋδεάζει για ύπαρξη σημείων εκτός των ορίων συμφωνίας. Στην παρακάτω εικόνα φαίνεται το διάγραμμα Bland-Altman βάσει του οποίου θα καθοριστεί εάν και κατά πόσο οι τιμές που προβλέφθηκαν συμφωνούν με τις πραγματικές.



Εικόνα 46 οι προβλέψεις συναρτήσεϊ των πραγματικών τιμών (κόκκινη γραμμή), και η ευθεία  $y=x$  (μπλε γραμμή)

Από το διάγραμμα επιβεβαιώνεται πως υπάρχουν σημεία εκτός των ορίων, όμως το 95% των σημείων βρίσκεται εντός, οπότε μπορεί με ασφάλεια να θεωρηθεί πως οι προβλέψεις του μοντέλου συμφωνούν με τις πραγματικές τιμές δόσης.

## 5. ΑΠΟΤΕΛΕΣΜΑΤΑ

Σε αυτό το κεφάλαιο θα παρουσιαστούν τα αποτελέσματα των προσομοιώσεων Monte Carlo για τις εξετάσεις θώρακος και άνω κοιλίας με CBCT αλλά και με απλή ακτινογραφία σε προσθόπισθια (AP), οπισθοπρόσθια (PA), αριστερής και δεξιάς πλάγιας (LL) προβολής, για κάθε όργανο ξεχωριστά. Στο δεύτερο μέρος θα παρουσιαστούν τα αποτελέσματα του αλγόριθμου τεχνητής νοημοσύνης για τις εξετάσεις CBCT θώρακος και άνω κοιλίας, επίσης για κάθε όργανο.

Τα όργανα ενδιαφέροντος για τις εξετάσεις θώρακος είναι τα οστά, ο οισοφάγος, οι μαστοί (γυναίκες ασθενείς), η καρδιά, οι πνεύμονες, το δέρμα και το τμήμα του σώματος του ασθενή που υπάρχει στο FOV (για συντομία αναφέρεται ως integral).

Τα όργανα ενδιαφέροντος για τις εξετάσεις άνω κοιλίας είναι τα επινεφρίδια, τα νευρά, το στομάχι, το πάγκρεας, ο σπλήνας, το συκώτι και το integral.

Στο πρώτο μέρος θα παρουσιαστούν πίνακες με τα στοιχεία με τα οποία έγινε η κάθε προσομοίωση καθώς και η περιγραφική στατιστική των δόσεων σε κάθε όργανο.

Όσον αφορά το δεύτερο μέρος, για κάθε όργανο θα αναφέρεται το μοντέλο που χρησιμοποιήθηκε μαζί με τις υπερπαραμέτρους που προέκυψαν από το tuning και θα παρουσιάζονται ο πίνακας της περιγραφικής στατιστικής των αποτελεσμάτων, καθώς και οι διαφορές των προβλεπόμενων τιμών δόσης και των πραγματικών.

### 5.1. Αποτελέσματα προσομοιώσεων

#### 5.1.1. CBCT

##### 5.1.1.1. CBCT θώρακος

Τα στοιχεία του πρωτοκόλλου με τα οποία πραγματοποιήθηκε η εξέταση φαίνονται στον παρακάτω πίνακα.

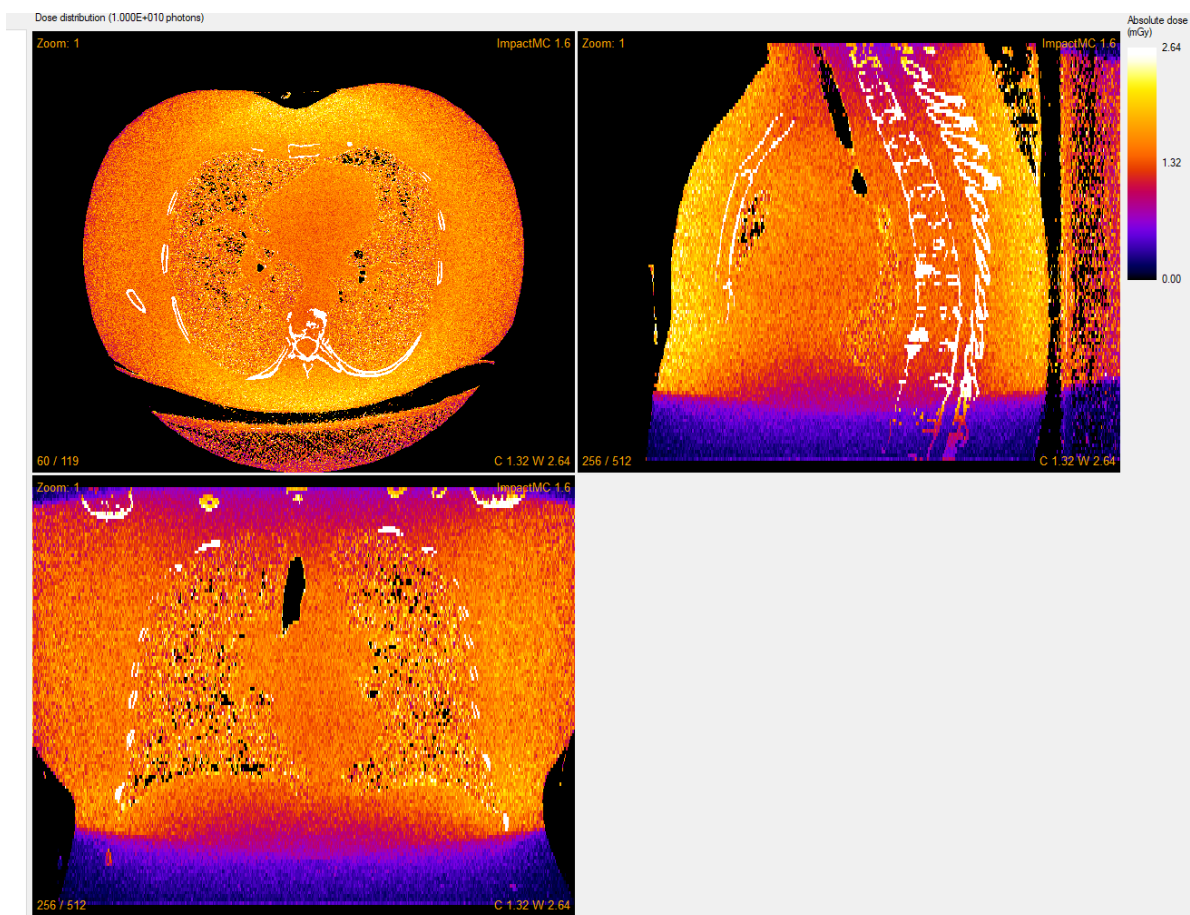
Τάση λυχνίας (kV)	Ένταση ρεύματος (mA)	Διάρκεια σάρωσης (s)	Air Kerma (mGy/100mAs)	Απόσταση λυχνίας-κέντρου περιστροφής (mm)	Αριθμός περιστροφών	Πρώθηση τράπεζας (mm)	Κατευθυντήρας	Φίλτρο
120	20	13.2	4.92	1000	1	0	M20	Bow tie

Table 1 στοιχεία πρωτοκόλλου για CBCT θώρακος

Τα αποτελέσματα που προέκυψαν φαίνονται στο παρακάτω πίνακα.

Δόση (mGy)	Όργανο	Οστά	Μαστοί	Οισοφάγος	Καρδιά	Πνεύμονες	Δέρμα	Integral
Min		6.236	7.202	4.413	5.406	3.220	3.489	5.052
Max		18.715	11.931	9.188	12.779	10.761	10.838	10.760
Median		12.374	9.674	6.991	8.585	7.624	6.937	7.217
Mean		12.427	9.680	6.944	8.526	7.621	6.955	7.284

Table 2 περιγραφική στατιστική των τιμών δόσης για CBCT θώρακα



Εικόνα 47 εικόνα δόσης από CBCT θώρακα όπως προέκυψε από το ImpactMC

### 5.1.1.2. CBCT άνω κοιλίας

Τα στοιχεία του πρωτοκόλλου με τα οποία πραγματοποιήθηκε η εξέταση φαίνονται στον παρακάτω πίνακα.

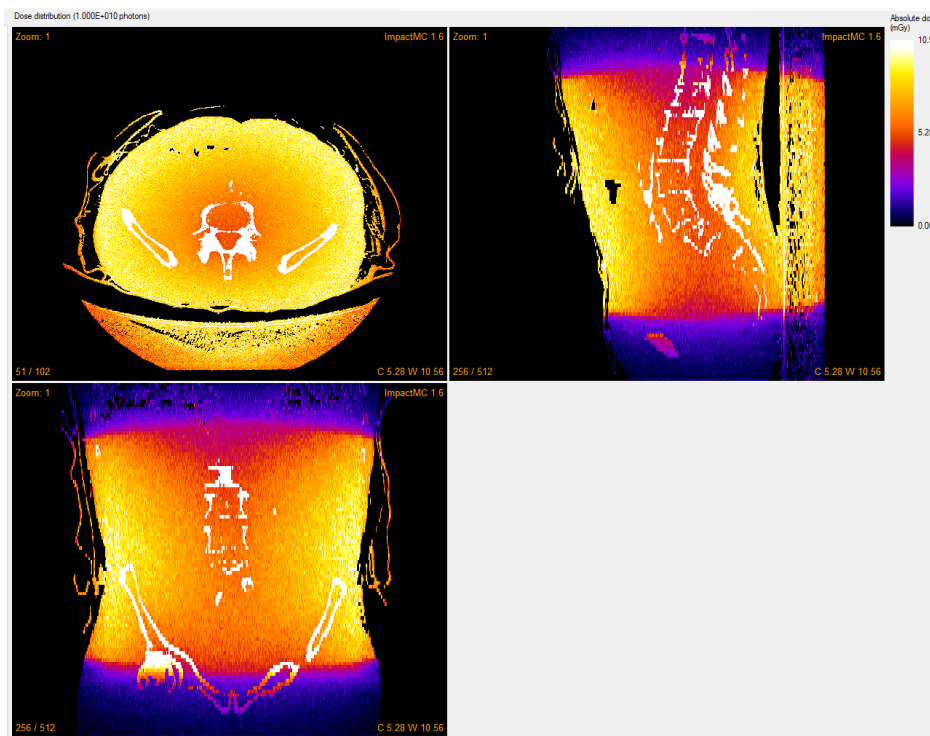
Τάση λυχνίας (kV)	Ένταση ρεύματος (mA)	Διάρκεια σάρωσης (s)	Air Kerma (mGy/100mAs)	Απόσταση λυχνίας-κέντρου περιστροφής (mm)	Αριθμός περιστροφών	Πρώθηση τράπεζας (mm)	Κατευθυντήρας	Φίλτρο
120	40	26.4	4.86	1000	1	0	M20	Bow tie

Table 3 στοιχεία πρωτοκόλλου για CBCT άνω κοιλίας

Τα αποτελέσματα που προέκυψαν φαίνονται στο παρακάτω πίνακα.

Δόση (mGy)	Όργανο	Νεφρά	Επινεφρίδια	Πάγκρεας	Συκώτι	Σπλήνας	Στομάχι	Integral
Min		15.246	10.825	11.958	12.729	12.418	8.686	13.071
Max		33.205	29.250	32.386	32.623	35.368	36.390	25.973
Median		23.105	18.110	21.041	20.893	23.462	19.641	17.259
Mean		23.585	18.416	21.231	21.217	23.732	20.075	18.219

Table 4 περιγραφική στατιστική των τιμών δόσης για CBCT άνω κοιλίας



Εικόνα 48 Εικόνα δόσης από CBCT άνω κοιλίας όπως προέκυψε από το ImpactMC

## 5.1.2. Ακτινογραφία

### 5.1.2.1. Ακτινογραφία θώρακος

#### 5.1.2.1.1. AP προβολή

Τα στοιχεία του πρωτοκόλλου με τα οποία πραγματοποιήθηκε η εξέταση φαίνονται στον παρακάτω πίνακα.

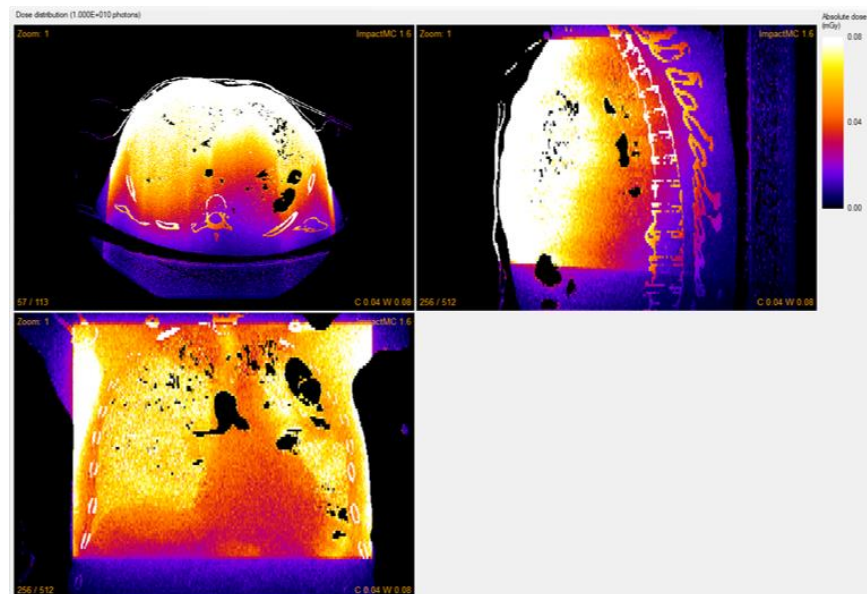
Τάση λυχνίας (kV)	Ένταση ρεύματος (mA)	Διάρκεια σάρωσης (s)	Απόσταση λυχνίας-ανιχνευτή (mm)	AEC
125	400	0.002 έως	1800	ON
137		0.013		

Table 5 στοιχεία πρωτοκόλλου για AP ακτινογραφία θώρακος

Τα αποτελέσματα που προέκυψαν φαίνονται στο παρακάτω πίνακα.

Δόση (mGy) \ Όργανο	Μαστοί	Οισοφάγος	Καρδιά	Πνεύμονες	Δέρμα	Integral
<b>Min</b>	0.132	0.033	0.078	0.045	0.042	0.045
<b>Max</b>	1.306	0.499	0.884	0.784	0.623	0.617
<b>Median</b>	0.539	0.183	0.370	0.251	0.246	0.260
<b>Mean</b>	0.563	0.199	0.376	0.271	0.259	0.261

Table 6 περιγραφική στατιστική των τιμών δόσης για AP ακτινογραφία θώρακα



Εικόνα 49 εικόνα δόσης για AP ακτινογραφία θώρακα όπως προέκυψε από το ImpactMC

### 5.1.2.1.2. PA προβολή

Τα στοιχεία του πρωτοκόλλου με τα οποία πραγματοποιήθηκε η εξέταση φαίνονται στον παρακάτω πίνακα (Long, Rollins, & Smith, 2016).

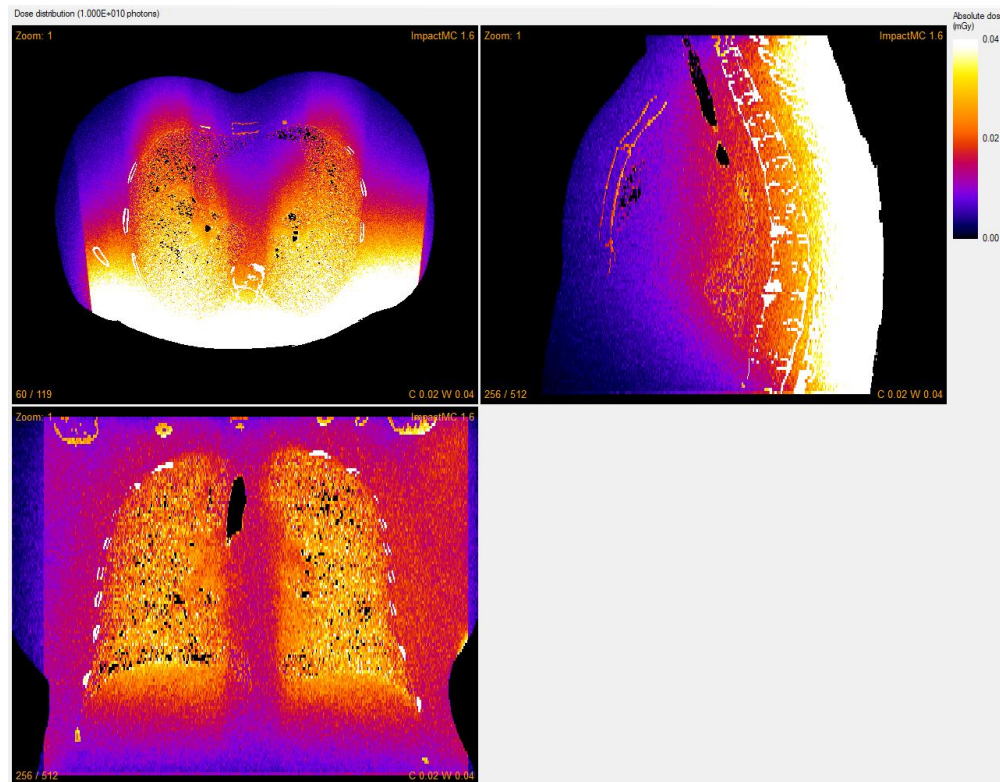
Τάση λυχνίας (kV)	Ένταση ρεύματος (mA)	Διάρκεια σάρωσης (s)	Απόσταση λυχνίας-ανιχνευτή (mm)	AEC
117	400	0.003 έως 0.013	1800	ON

Table 7 στοιχεία πρωτοκόλλου για PA ακτινογραφία θώρακος

Τα αποτελέσματα που προέκυψαν φαίνονται στο παρακάτω πίνακα.

Δόση (mGy)	Όργανο	Οστά	Μαστοί	Οισοφάγος	Καρδιά	Πνεύμονες	Δέρμα	Integral
Min		0.020	0.003	0.006	0.004	0.005	0.042	0.011
Max		0.151	0.030	0.043	0.038	0.072	0.623	0.069
Median		0.067	0.013	0.020	0.015	0.026	0.246	0.312
Mean		0.070	0.014	0.021	0.016	0.030	0.259	0.033

Table 8 περιγραφική στατιστική των τιμών δόσης για PA ακτινογραφία θώρακα



Εικόνα 50 εικόνα δόσης για PA ακτινογραφία θώρακα όπως προέκυψε από το ImpactMC

### 5.1.2.1.3. LL προβολές

Τα στοιχεία του πρωτοκόλλου με τα οποία πραγματοποιήθηκαν οι εξετάσεις φαίνονται στον παρακάτω πίνακα (Long, Rollins, & Smith, 2016).

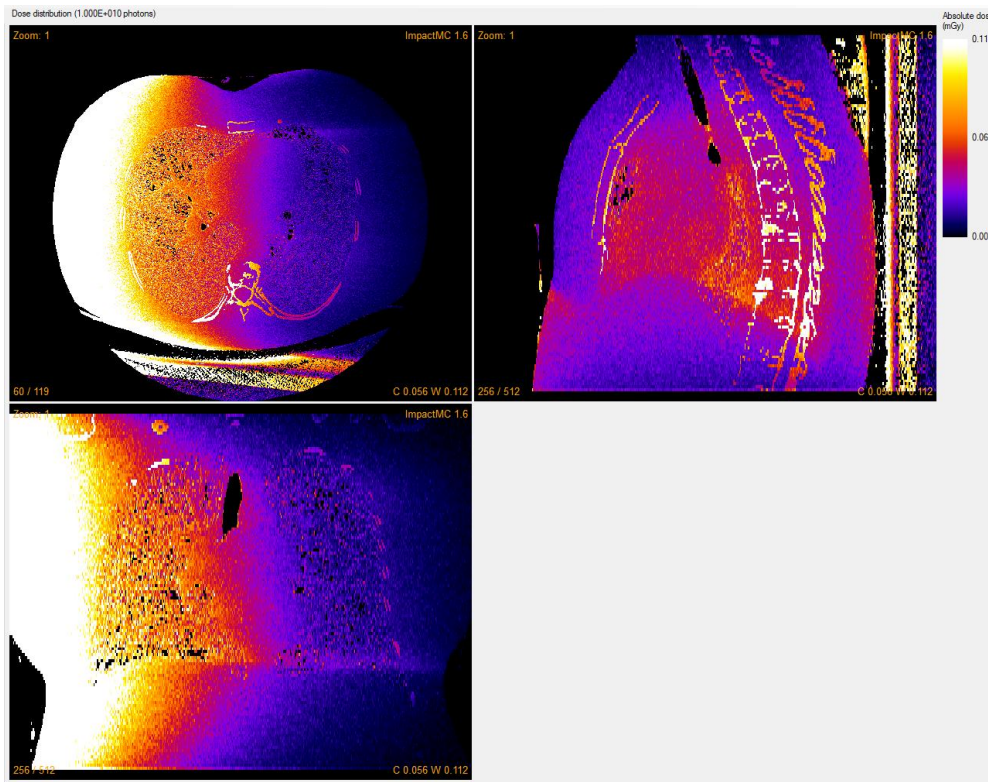
Τάση λυχνίας (kV)	Ένταση ρεύματος (mA)	Διάρκεια σάρωσης (s)	Απόσταση λυχνίας-ανιχνευτή (mm)	AEC
125	160	0.015 έως 0.251	1000	ON

Table 9 στοιχεία πρωτοκόλλου για PA ακτινογραφία θώρακος

Τα αποτελέσματα που προέκυψαν φαίνονται στους παρακάτω πίνακες.

Δόση (mGy)	Όργανο	Οστά	Μαστοί	Οισοφάγος	Καρδιά	Πνεύμονες	Δέρμα	Integral
<b>Min</b>		0.264	0.256	0.133	0.043	0.132	0.332	0.212
<b>Max</b>		7.165	5.696	5.058	0.384	4.888	7.377	4.710
<b>Median</b>		3.013	2.778	1.516	0.153	1.615	2.866	2.038
<b>Mean</b>		3.035	2.973	1.603	0.163	1.665	2.989	2.032

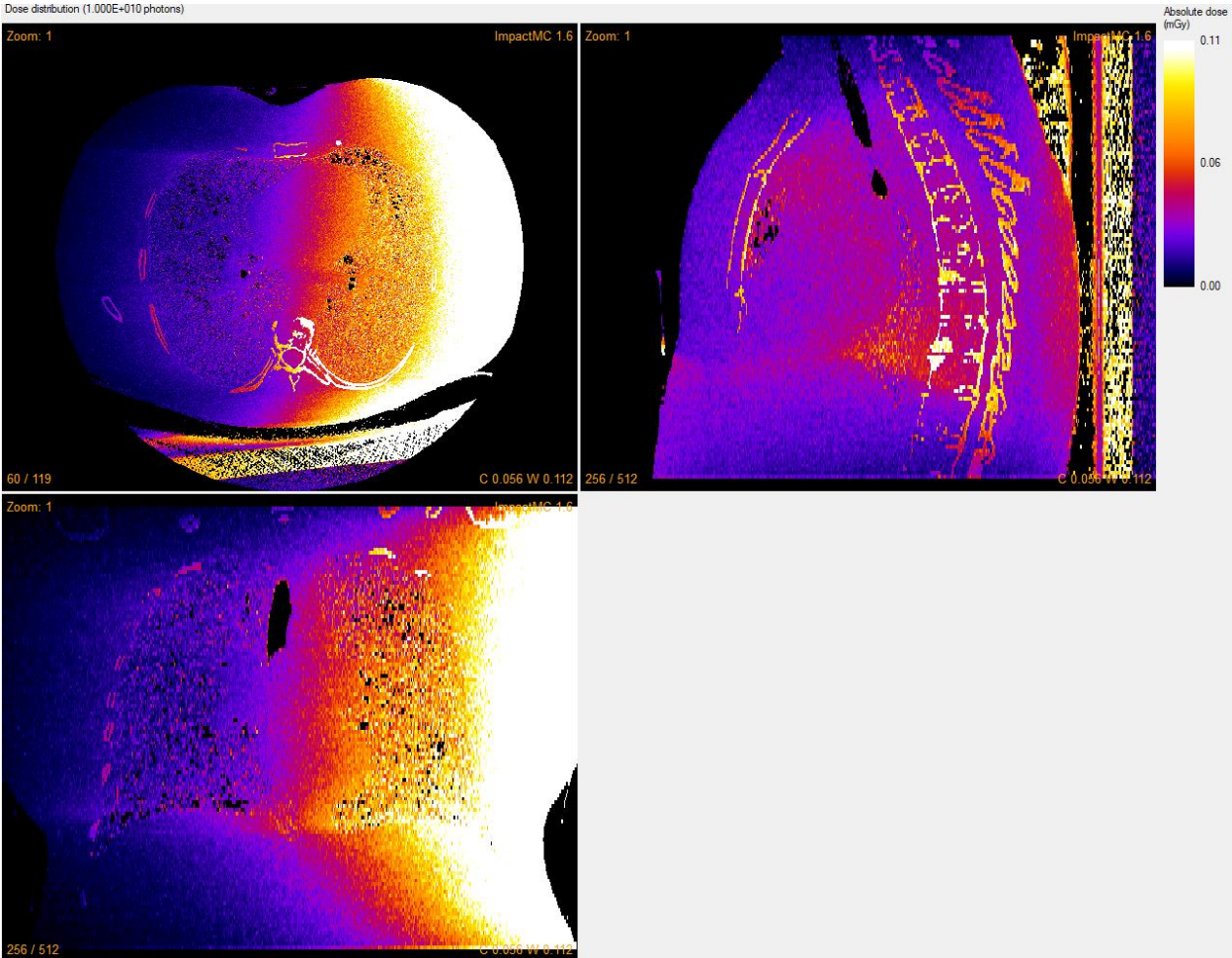
Table 10 περιγραφική στατιστική των τιμών δόσης για αριστερή LL ακτινογραφία θώρακος



Εικόνα 51 Εικόνα δόσης για αριστερή LL ακτινογραφία θώρακα όπως προέκυψε από το ImpactMC

Δόση (mGy) \ Όργανο	Οστά	Μαστοί	Οισοφάγος	Καρδιά	Πνεύμονες	Δέρμα	Integral
<b>Min</b>	0.24	0.251	0.138	0.194	0.161	0.366	0.219
<b>Max</b>	6.988	6.008	3.762	5.541	3.723	6.895	4.519
<b>Median</b>	3.063	2.341	1.340	2.001	1.360	3.048	2.042
<b>Mean</b>	3.000	2.477	1.457	2.111	1.502	2.998	1.959

Table 11 περιγραφική στατιστική των τιμών δόσης για δεξιά LL ακτινογραφία θώρακος



Εικόνα 52 εικόνα δόσης για δεξιά LL ακτινογραφία θώρακα όπως προεκυψε από το ImpactMC

5.1.2.2. Ακτινογραφία άνω κοιλίας

5.1.2.2.1. AP προβολή

Τα στοιχεία του πρωτοκόλλου με τα οποία πραγματοποιήθηκε η εξέταση φαίνονται στον παρακάτω πίνακα.



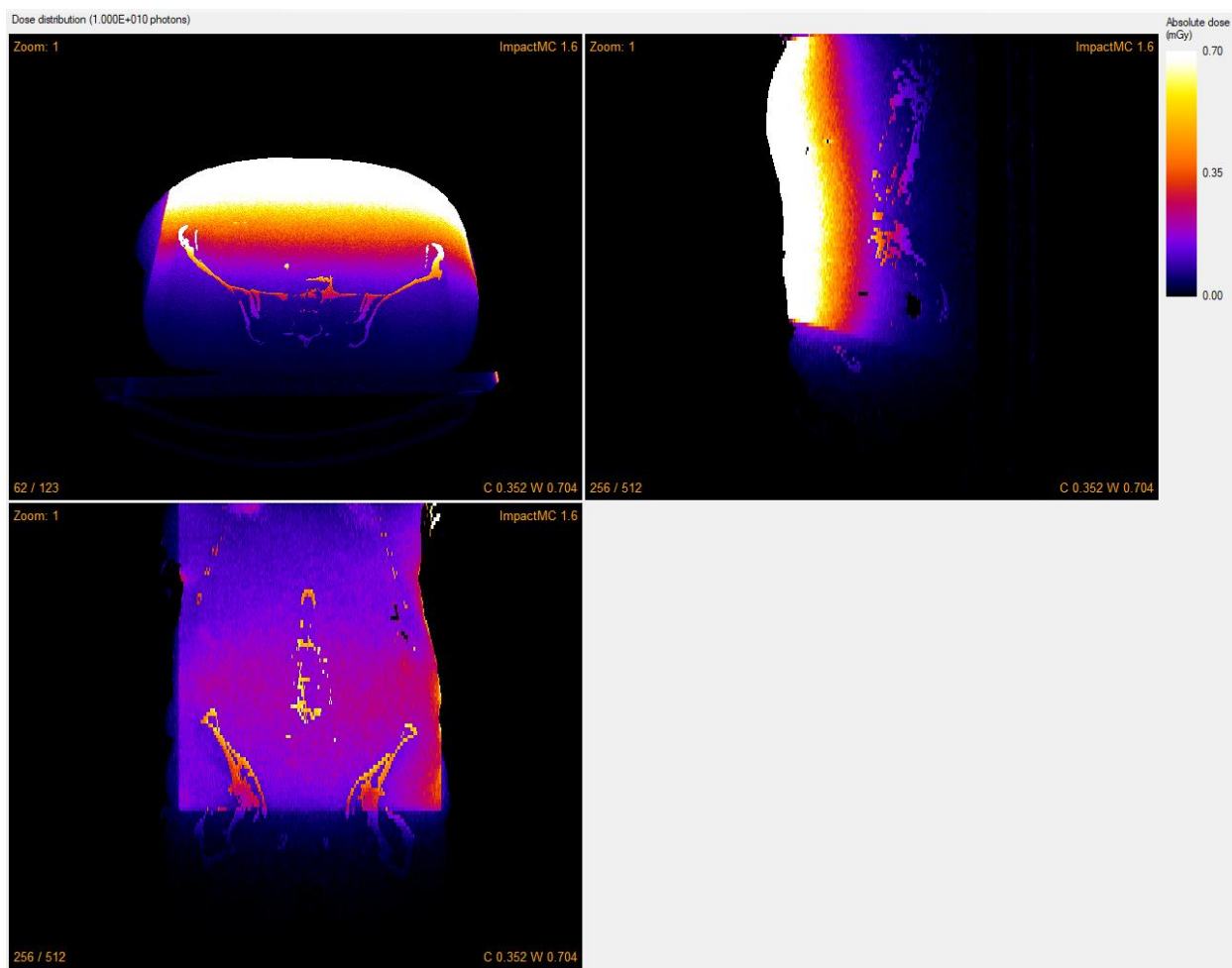
Τάση λυχνίας (kV)	Ένταση ρεύματος (mA)	Διάρκεια σάρωσης (s)	Απόσταση λυχνίας-ανιχνευτή (mm)	AEC
75	640	0.021 έως 0.468	1000	ON

Table 12 στοιχεία πρωτοκόλλου για AP ακτινογραφία άνω κοιλίας

Τα αποτελέσματα που προέκυψαν φαίνονται στο παρακάτω πίνακα.

Όργανο	Νεφρά	Επινεφρίδια	Πάγκρεας	Συκώτι	Σπλήνας	Στομάχι	Integral
Δόση (mGy)							
Min	0.678	0.644	1.345	12.729	0.234	1.908	1.794
Max	22.165	21.384	41.587	32.623	25.861	67.025	29.136
Median	4.899	5.558	11.471	20.893	3.123	16.483	17.376
Mean	5.755	7.232	13.797	21.217	4.498	19.895	14.802

Table 13 περιγραφική στατιστική των τιμών δόσης για AP προβολή άνω κοιλίας



Εικόνα 53 εικόνα δόσης για AP ακτινογραφία άνω κοιλίας όπως προέκυψε από το ImpactMC

### 5.1.2.2.2. PA προβολή

Τα στοιχεία του πρωτοκόλλου με τα οποία πραγματοποιήθηκε η εξέταση φαίνονται στον παρακάτω πίνακα.

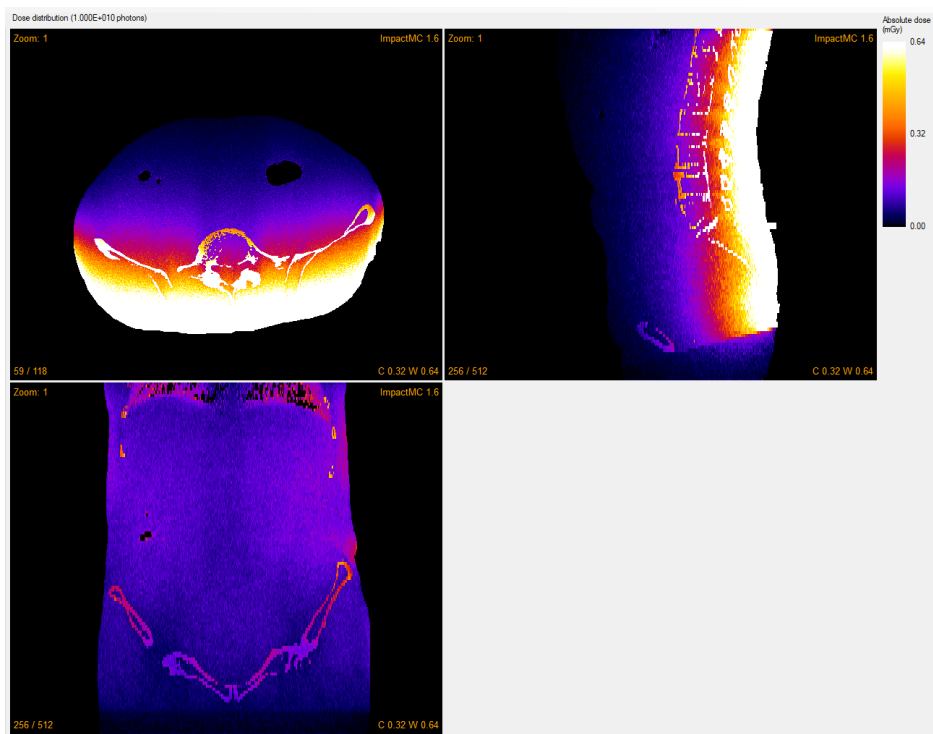
Τάση λυχνίας (kV)	Ένταση ρεύματος (mA)	Διάρκεια σάρωσης (s)	Απόσταση λυχνίας-ανιχνευτή (mm)	AEC
75	640	0.005 έως 0.500	1000	ON
82				

Table 14 στοιχεία πρωτοκόλλου για PA ακτινογραφία άνω κοιλίας

Τα αποτελέσματα που προέκυψαν φαίνονται στο παρακάτω πίνακα.

Δόση (mGy)	Όργανο	Νεφρά	Επινεφρίδια	Πάγκρεας	Συκώτι	Σπλήνας	Στομάχι	Integral
Min		0.431	0.126	0.071	0.078	0.270	0.053	0.237
Max		63.927	33.840	23.349	31.015	73.366	21.859	34.577
Median		15.968	10.142	5.392	7.724	20.981	4.126	14.424
Mean		19.822	12.925	7.125	10.405	22.698	5.407	14.276

Table 15 περιγραφική στατιστική των τιμών δόσης για PA προβολή άνω κοιλίας



Εικόνα 54 εικόνα δόσης για PA ακτινογραφία άνω κοιλίας όπως προέκυψε από το ImpactMC

### 5.1.2.2.3. LL προβολές

Τα στοιχεία του πρωτοκόλλου με τα οποία πραγματοποιήθηκαν οι εξετάσεις φαίνονται στον παρακάτω πίνακα (Long, Rollins, & Smith, 2016).

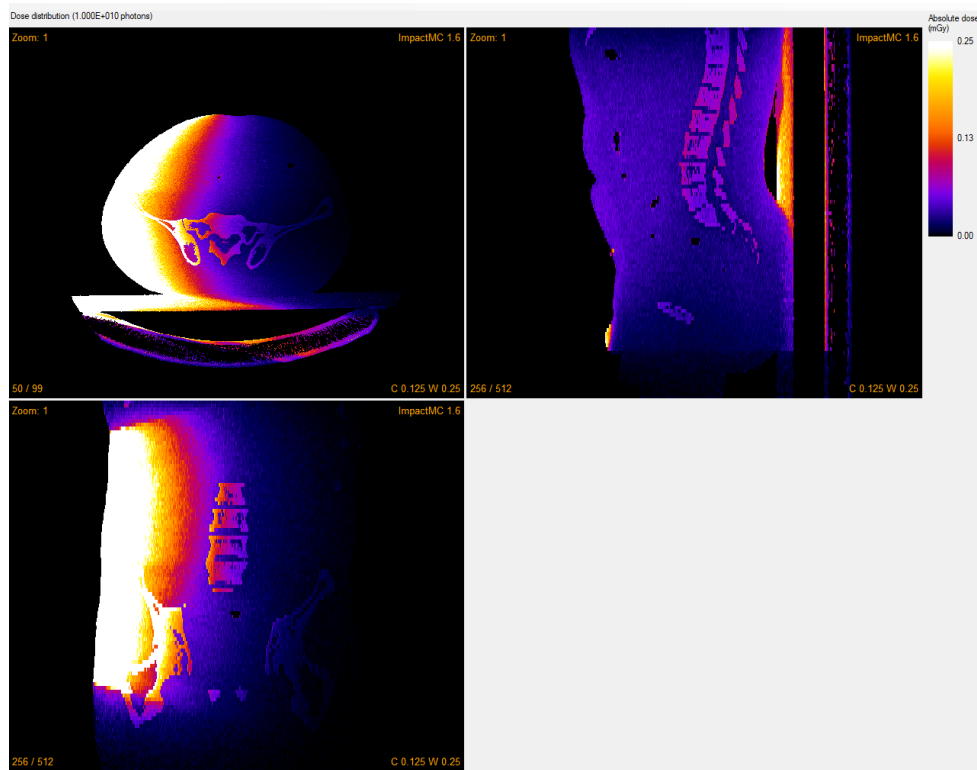
Τάση λυχνίας (kV)	Ένταση ρεύματος (mA)	Διάρκεια σάρωσης (s)	Απόσταση λυχνίας-ανιχνευτή (mm)	AEC
90	250	0.005 έως 0.500	1000	ON

Table 16 στοιχεία πρωτοκόλλου για LL ακτινογραφία άνω κοιλίας

Τα αποτελέσματα που προέκυψαν φαίνονται στους παρακάτω πίνακες.

Δόση (mGy)	Όργανο	Νεφρά	Επινεφρίδια	Πάγκρεας	Συκώτι	Σπλήνας	Στομάχι	Integral
<b>Min</b>		0.140	0.039	0.223	0.223	0.005	0.019	0.130
<b>Max</b>		16.020	10.514	28.175	28.768	16.380	6.298	14.081
<b>Median</b>		5.572	1.903	9.232	9.610	0.405	1.340	7.664
<b>Mean</b>		5.658	2.769	11.684	11.940	0.794	1.723	7.370

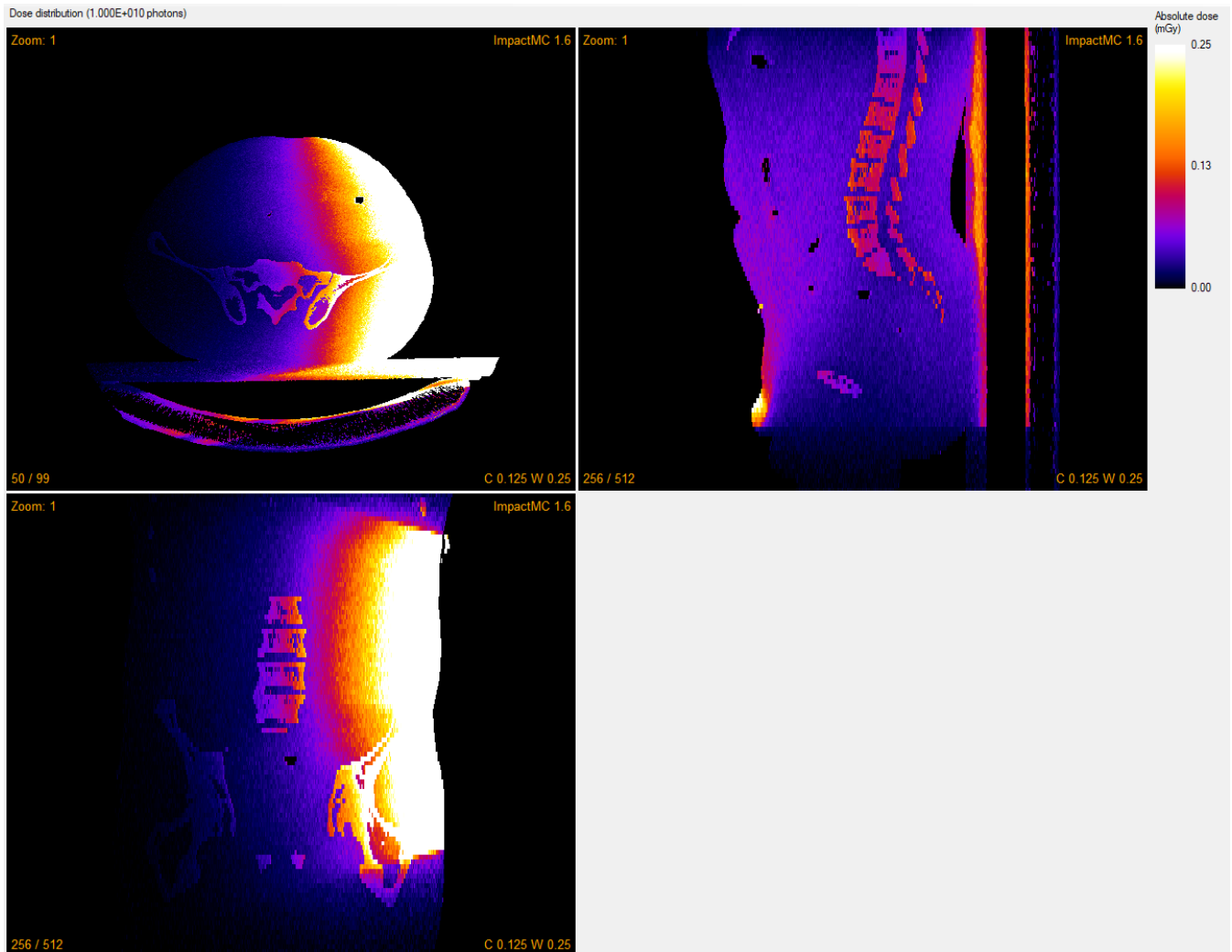
Table 17 περιγραφική στατιστική των τιμών δόσης για αριστερή LL ακτινογραφία άνω κοιλίας



Εικόνα 55 εικόνα δόσης για αριστερή LL ακτινογραφία άνω κοιλίας όπως προέκυψε από το ImpactMC

Όργανο \ Δόση (mGy)	Νεφρά	Επινεφρίδια	Πάγκρεας	Συκώτι	Σπλήνας	Στομάχι	Integral
<b>Min</b>	0.203	0.057	0.100	0.012	0.233	0.052	0.135
<b>Max</b>	17.637	10.514	10.095	28.768	23.322	10.052	13.298
<b>Median</b>	6.769	2.324	3.083	1.282	2.253	2.588	7.437
<b>Mean</b>	6.433	3.001	3.366	8.792	5.496	3.174	7.227

Table 18 περιγραφική στατιστική των τιμών δόσης για δεξιά LL ακτινογραφία άνω κοιλίας



Εικόνα 56 εικόνα δόσης για δεξιά LL ακτινογραφία άνω κοιλίας όπως προέκυψε από το ImpactMC

## 5.2. Πρόβλεψη δόσης για CBCT θώρακος

Στο τελευταίο μέρος των αποτελεσμάτων, το οποίο ήταν και ο στόχος αυτής της εργασίας, θα παρουσιαστούν τα αποτελέσματα των προβλέψεων των αλγορίθμων τεχνητής νοημοσύνης για

την πρόβλεψη της δόσης που απορροφήθηκε από κάθε όργανο ενδιαφέροντος. Η αξιολόγηση του μοντέλου θα γίνει με τη χρήση διαγραμμάτων Bland-Altman

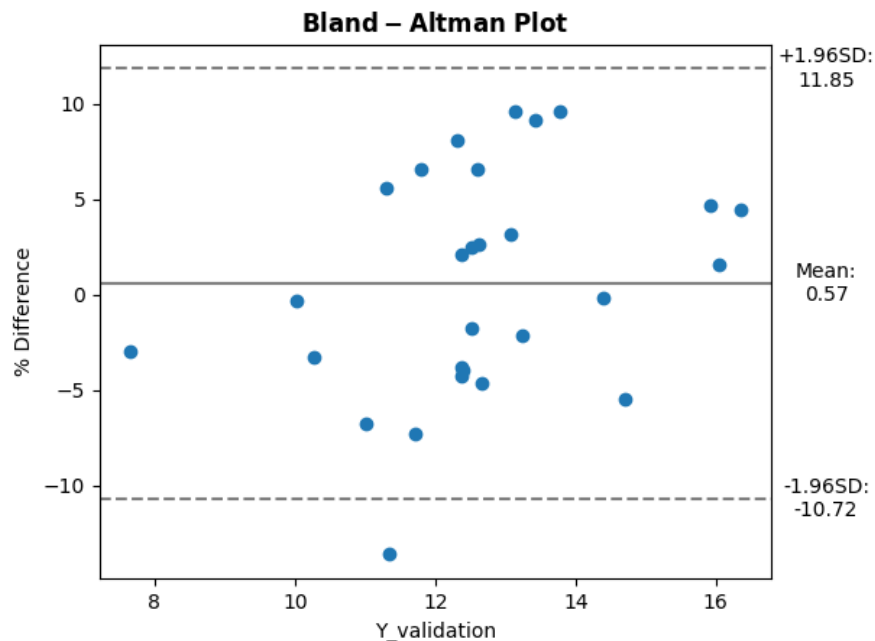
### 5.2.1. Οστά

*Lasso (alpha=0.0)*

*MSE=1.602*

	<b>Y_validation</b>	<b>Predictions</b>
Min	7.654	8.009
Max	16.353	15.716
Median	12.511	12.495
Mean	12.641	12.531

*Table 19 αποτελέσματα για τα οστά*



*Εικόνα 57 διάγραμμα Bland-Altman για τα οστά*

Παρατηρείται ότι υπάρχουν τιμές της προβλεπόμενης δόσης οι οποίες απέχουν από την πραγματική τιμή, με μέγιστη διαφορά κοντά στο 13.5%. Ωστόσο, μόνο ένα σημείο δεν βρίσκεται εντός

των ορίων συμφωνίας, οπότε συμπεραίνεται πως οι προβλέψεις συμφωνούν με τις πραγματικές τιμές.

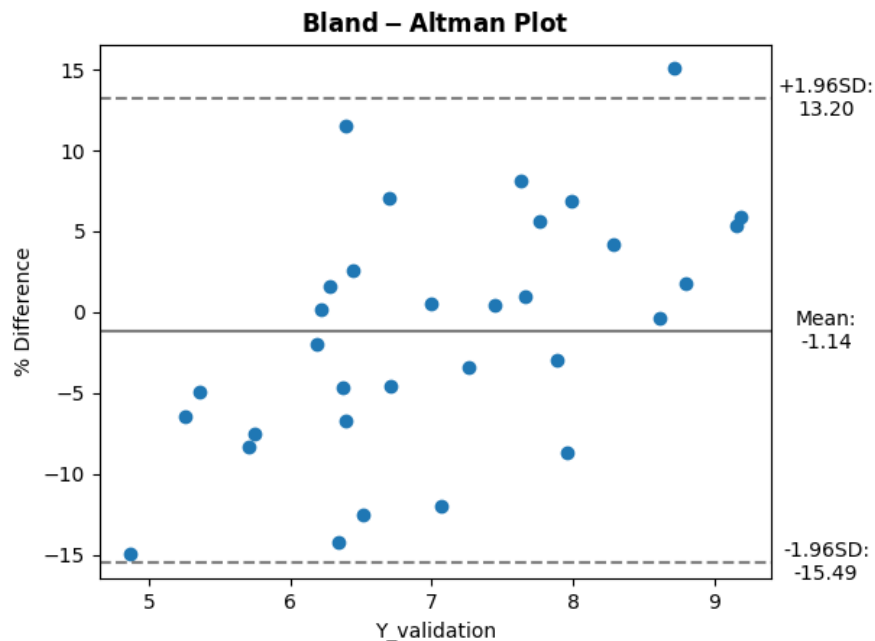
### 5.2.2. Οισοφάγος

*AdaBoostRegressor(n\_estimators=52)*

*MSE=0.246*

	Y_validation	Predictions
Min	4.865	5.593
Max	9.188	8.662
Median	6.851	7.127
Mean	7.059	7.093

Table 20 αποτελέσματα για τον οισοφάγο



Εικόνα 58 διάγραμμα Bland-Altman για τον οισοφάγο

Στην περίπτωση αυτή οι τιμές δόσης που προβλέπει το μοντέλο έχουν μέγιστη διαφορά από τις πραγματικές 15%, και το 95% των σημείων βρίσκονται μέσα στο διάστημα  $\pm 1.96SD$ , οπότε με ασφάλεια μπορεί να θεωρηθεί πως το μοντέλο κάνει σωστές προβλέψεις.

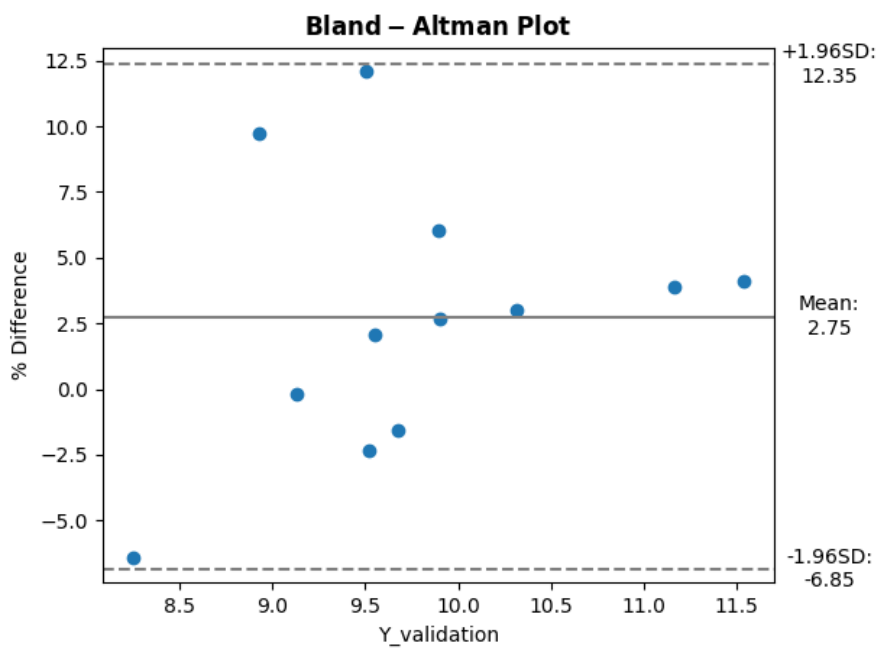
### 5.2.3. Μαστοί

*ElasticNet (alpha=0.45)*

*MSE=0.452*

	Y_validation	Predictions
Min	8.253	8.061
Max	11.536	11.064
Median	9.612	9.495
Mean	9.781	9.501

Table 21 αποτελέσματα για τους μαστούς



Εικόνα 59 διάγραμμα Bland-Altman για τους μαστούς

Καλές προβλέψεις κάνει ο αλγόριθμος και σε αυτή την περίπτωση, όπου η μέγιστη διαφορά μεταξύ προβλέψεων και πραγματικών τιμών είναι 12%, ενώ όλα τα σημεία είναι μεταξύ των ορίων.

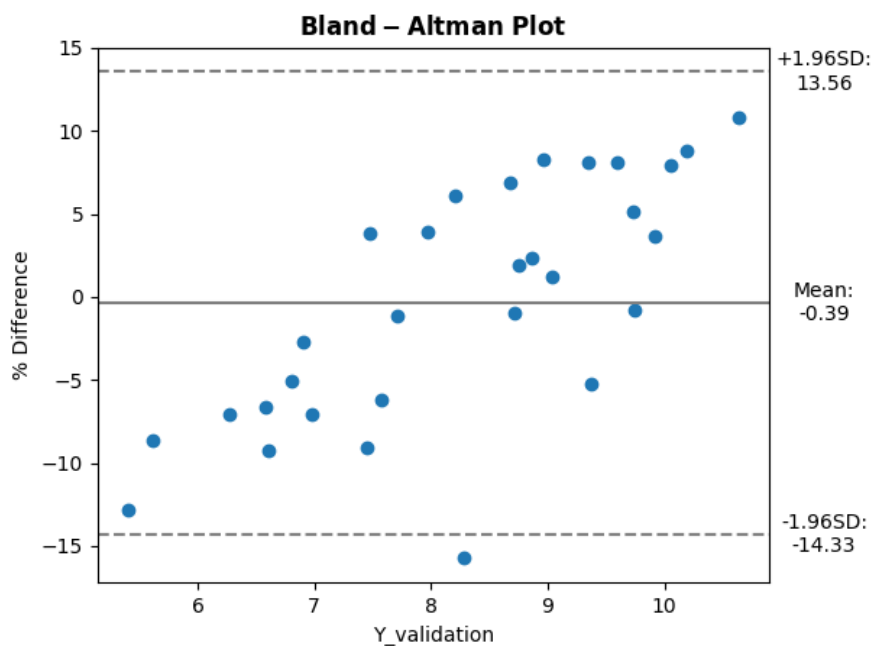
#### 5.2.4. Καρδιά

*Lasso(alpha=0.99)*

*MSE=0.730*

	Y_validation	Predictions
Min	5.406	6.100
Max	10.636	8.172
Median	8.478	8.643
Mean	8.248	8.206

Table 22 αποτελέσματα για την καρδιά



Εικόνα 60 διάγραμμα Bland-Altman για την καρδιά

Οι προβλέψεις της δόσης για την καρδιά φτάνουν να διαφέρουν από τις πραγματικές τιμές δόσης μέχρι περίπου 16%. Το γεγονός αυτό πιθανόν οφείλεται στο ότι το dataset περιέχει τιμές δόσης για άνδρες και για γυναίκες, μη λαμβάνοντας, έτσι, υπόψη ότι το όργανο αυτό στις γυναίκες απορροφά μικρότερη δόση σε σχέση με τους άνδρες, λόγω της παρουσίας των μαστών.



Ωστόσο, το 95% των σημείων βρίσκεται εντός των ορίων συμφωνίας, οπότε οι προβλέψεις συμφωνούν με τις πραγματικές τιμές.

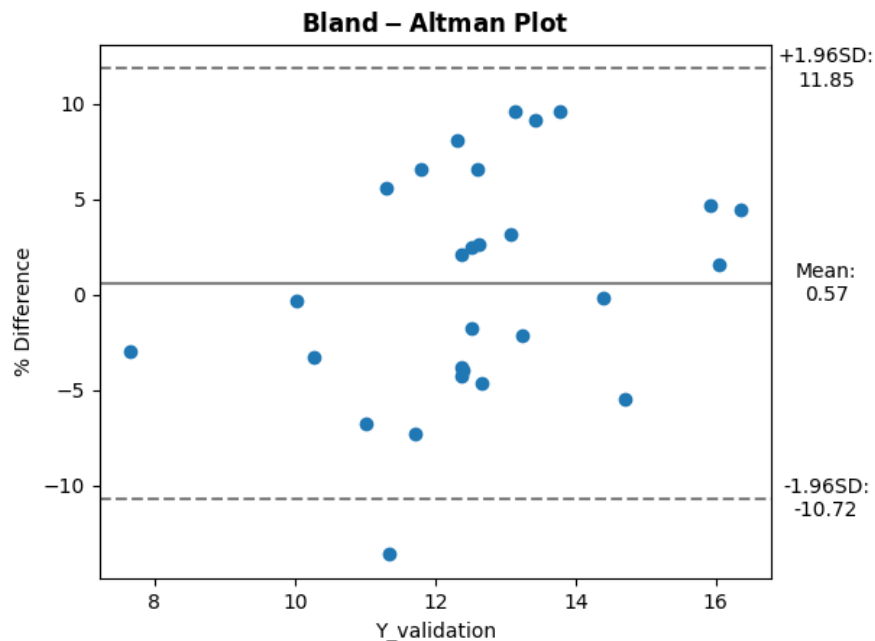
### 5.2.5. Πνεύμονες

*ExtraTreesRegressor (n\_estimators=31, max\_features=88)*

*MSE=1.271*

	Y_validation	Predictions
Min	5.669	6.334
Max	10.434	9.891
Median	7.899	7.279
Mean	7.960	7.648

*Table 23 αποτελέσματα για τους πνεύμονες*



*Εικόνα 61 διάγραμμα Bland-Altman για τους πνεύμονες*

Παρατηρείται η μέγιστη διαφορά να φτάνει στο 20.25%, πρόβλημα το οποίο υπάρχει κυρίως στις ακραίες τιμές δόσης. Ωστόσο, ένα μόνο σημείο βρίσκεται εκτός των ορίων, οπότε και εδώ

οι προβλέψεις συμφωνούν με τις πραγματικές τιμές. Και εδώ, όπως και στην καρδιά, η διαφορά έγκειται στην παρουσία των μαστών.

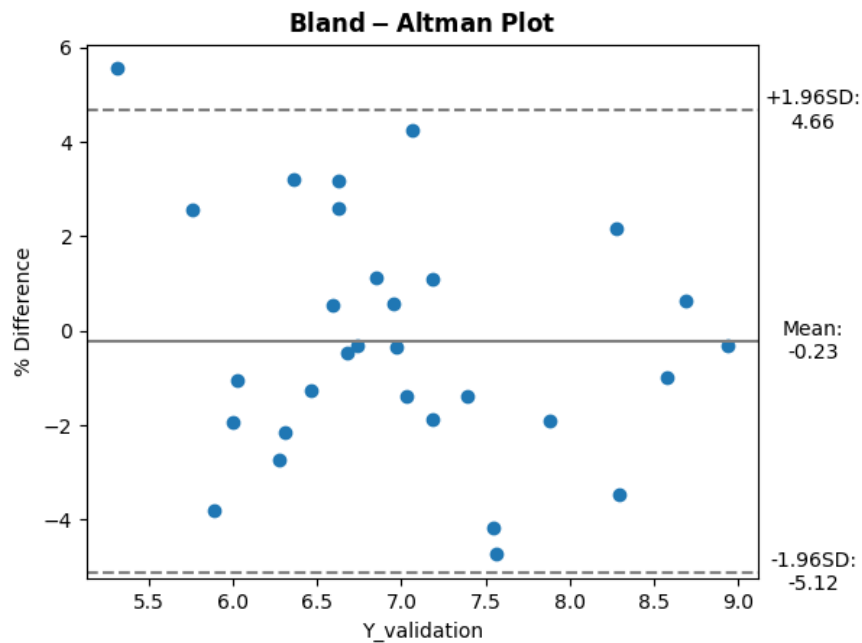
### 5.2.6. Δέρμα

*ElasticNet(alpha=0.99)*

*MSE=0.072*

	<b>Y_validation</b>	<b>Predictions</b>
Min	5.312	5.017
Max	8.939	8.969
Median	6.903	6.770
Mean	7.002	7.023

Table 24 αποτελέσματα για το δέρμα



Εικόνα 62 διάγραμμα Bland-Altman για το δέρμα

Εδώ φαίνεται η μικρότερη μέγιστη διαφορά η οποία είναι λιγότερο από 9%. Και εδώ υπάρχει συμφωνία μεταξύ των συγκρινόμενων τιμών.

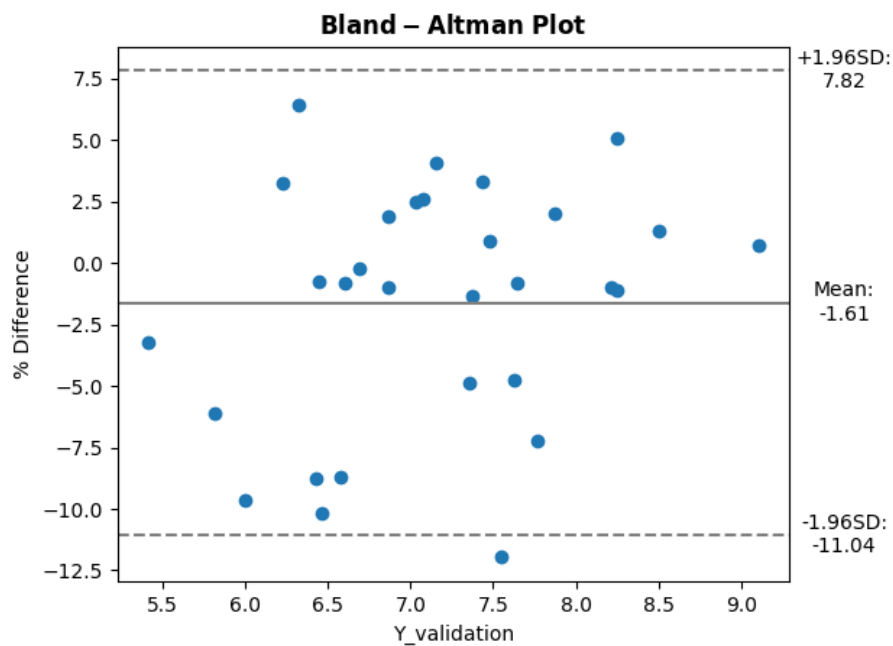
### 5.2.7. Integral

*LinearRegression()*

*MSE=0.241*

	Y_validation	Predictions
Min	5.411	5.587
Max	9.107	9.043
Median	7.116	7.130
Mean	7.145	7.250

*Table 25 αποτελέσματα για το integral*



*Εικόνα 63 διάγραμμα Bland-Altman για το integral*

Από το διάγραμμα παρατηρείται μια, γενικά, καλή συμφωνία μεταξύ των δύο ευθειών, διότι, παρόλο που υπάρχουν τιμές που αποκλίνουν (μέγιστη διαφορά περίπου 12%), λιγότερα από το 5% των σημείων βρίσκεται εκτός ορίων.

## 6. ΣΥΖΗΤΗΣΗ

Όπως έχει ήδη αναφερθεί, η υπολογιστική τομογραφία κωνικής δέσμης (CBCT) αποτελεί μια από τις πιο σύγχρονες μεθόδους απεικόνισης του ανθρώπινου σώματος. Η εφαρμογή της στην ακτινοθεραπεία κρίνεται ως μεγάλης σημασίας, καθώς δίνει την δυνατότητα real time παρακολούθησης και βελτίωσης της τοποθέτησης του ασθενή κατά τη διάρκεια της θεραπείας. Ωστόσο, αυτό οδηγεί σε αύξηση της δόσης, άρα και του κινδύνου στοχαστικών ή καθορισμένων αποτελεσμάτων της ακτινοβολίας. Έτσι, η δοσιμετρία είναι μάλλον απαραίτητη, πόσο μάλλον σε πραγματικό χρόνο που μπορεί να μειώσει κατά πολύ αυτόν τον κίνδυνο. Στο πρώτο μέρος της παρούσας εργασίας έγινε Monte Carlo δοσιμετρία σε ασθενείς που υποβλήθηκαν σε ακτινογραφικές εξετάσεις και εξετάσεις CBCT ακτινοθεραπείας σε θώρακα (113 ασθενείς) και άνω κοιλία (89 ασθενείς). Στη συνέχεια χρησιμοποιήθηκαν οι δόσεις που προέκυψαν για τις εξετάσεις CBCT θώρακος μαζί με παράγοντες που τις επηρεάζουν (dose predictors) ώστε να δημιουργηθεί ένας αλγόριθμος τεχνητής νοημοσύνης που να προβλέπει την δόση από την εξέταση αυτή.

Στο πρώτο μέρος όπου έγιναν οι προσομοιώσεις των παραπάνω εξετάσεων, έγινε με χρήση του λογισμικού Impact MC. Οι παράμετροι που χρησιμοποιήθηκαν βασίστηκαν στα Dicom Headers εξετάσεων που πραγματοποιήθηκαν στο Πανεπιστημιακό Γενικό Νοσοκομείο Ηρακλείου, καθώς και σε υπάρχοντα πρωτόκολλα.

Όσον αφορά τις παραμέτρους που χρησιμοποιήθηκαν για τις προσομοιώσεις των CBCT εξετάσεων, βασίστηκαν στα πρωτόκολλα της κατασκευάστριας εταιρίας Elekta. Οι δόσεις για τα όργανα ενδιαφέροντος αποθηκεύτηκαν και δημιουργήθηκαν αρχεία για κάθε όργανο που να περιέχουν την τάση της λυχνίας, την ένταση του ρεύματος, τη διάρκεια της σάρωσης, τις axial διαστάσεις x και y του ασθενή, το water equivalent diameter (WED), το μέγεθος του οργάνου και την δόση του οργάνου για κάθε ασθενή.

Στο δεύτερο μέρος της εργασίας χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python για τη δημιουργία αλγορίθμων τεχνητής νοημοσύνης, οι οποίοι, χρησιμοποιώντας τα δεδομένα από τα παραπάνω αρχεία, να προβλέπουν τη δόση σε κάθε όργανο ενδιαφέροντος.

Οι προβλέψεις των αλγορίθμων ήταν καλές, με μέσες τιμές διαφορών των τιμών που προβλέπονται από το μοντέλο και των πραγματικών τιμών να κυμαίνονται από 2.5% έως 13%, οι οποίες σε επίπεδο mGy θεωρούνται αμελητέες. Η μέγιστη διαφορά εμφανίστηκε στους πνεύμονες, λόγω του ότι το dataset περιείχε τιμές δόσης για άνδρες και γυναίκες, χωρίς να λαμβάνεται υπόψη η παρουσία των μαστών στις γυναίκες. Ένας διαχωρισμός των φύλων θα βελτίωνε το αποτέλεσμα. Παρόλα αυτά, οι αξιολογήσεις των μοντέλων βάσει των γραφημάτων Bland-Altman έδειξαν πως οι τιμές της δόσης που προβλέπουν οι αλγόριθμοι για κάθε όργανο, συμφωνούν με αυτές που προέκυψαν από τις Monte Carlo προσομοιώσεις.

Συμπερασματικά, οι αλγόριθμοι που δημιουργήθηκαν στην παρούσα εργασία μπορούν να χρησιμοποιηθούν με ασφάλεια για την πρόβλεψη της δόσης ασθενών που υποβλήθηκαν σε εξετάσεις θώρακα CBCT ακτινοθεραπείας, ωστόσο μια εκτενέστερη έρευνα με μεγαλύτερο δείγμα ασθενών σίγουρα θα βελτίωνε το αποτέλεσμα.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- © CT Imaging GmbH. (2016). User Guide ImpactMC. Germany.
- Brownlee, J. (2016). *Machine Learning Mastery With Python*. Jason Brownlee.
- Chollet, F. (2018). *Deep Learning with Python*. Manning Publications Co.
- Damilakis, J. (2020). CT Dosimetry: What Has Been Achieved and What Remains to Be Done. *Investigative Radiology*.
- developers, s.-l. (2007-2022). *scikit learn user guide*. Ανάκτηση από scikit learn: [scikit-learn.org](https://scikit-learn.org)
- Elekta. (2016). Elekta Medical Linear Accelerator XVI R5.0.4 Instructions for Use for: Elekta Synergy®, Elekta Axesse™, Elekta Infinity™, Versa HD™.
- Krouwer, J. S. (2008). Why Bland–Altman plots should use  $X$ , not  $(Y + X)/2$  when  $X$  is a reference method. *Statistics in Medicine*.
- Long, B. W., Rollins, J. H., & Smith, B. J. (2016). *Merill's Atlas of Radiographic Positioning & Procedures*. Elsevier.
- Shaw, C. C. (2014). *Cone Beam Computed Tomography*. CRC Press.
- Sykes, J. R. (2013). Dosimetry of CBCT: methods, doses and clinical consequences. *Journal of Physics*.
- Σεϊμένης, Ι. (2021). Υπολογιστική Τομογραφία, σημειώσεις για το μάθημα Διαγνωστική Ακτινολογία του Δ.Δ.Μ.Π.Σ. Ιατρικής Φυσικής-Ακτινοφυσικής.
- Σωτηροπούλου, Β. (2021). Εξατομικευμένη δοσιμετρική σύγκριση οργάνων από ασθενείς που έχουν υποβληθεί σε εξετάσεις Υπολογιστικής Τομογραφίας για διάγνωση και θεραπεία μέσω εξομοιώσεων Monte Carlo. *Διπλωματική εργασία*.

## ΠΙΝΑΚΕΣ

Table 1 στοιχεία πρωτοκόλλου για CBCT θώρακος .....	42
Table 2 περιγραφική στατιστική των τιμών δόσης για CBCT θώρακα.....	42
Table 3 στοιχεία πρωτοκόλλου για CBCT άνω κοιλίας.....	43
Table 4 περιγραφική στατιστική των τιμών δόσης για CBCT άνω κοιλίας.....	43
Table 5 στοιχεία πρωτοκόλλου για AP ακτινογραφία θώρακος .....	44
Table 6 περιγραφική στατιστική των τιμών δόσης για AP ακτινογραφία θώρακα .....	44
Table 7 στοιχεία πρωτοκόλλου για PA ακτινογραφία θώρακος .....	45
Table 8 περιγραφική στατιστική των τιμών δόσης για PA ακτινογραφία θώρακα .....	45
Table 9 στοιχεία πρωτοκόλλου για PA ακτινογραφία θώρακος .....	46
Table 10 περιγραφική στατιστική των τιμών δόσης για αριστερή LL ακτινογραφία θώρακος.....	46
Table 11 περιγραφική στατιστική των τιμών δόσης για δεξιά LL ακτινογραφία θώρακος.....	47
Table 12 στοιχεία πρωτοκόλλου για AP ακτινογραφία άνω κοιλίας.....	48
Table 13 περιγραφική στατιστική των τιμών δόσης για AP προβολή άνω κοιλίας.....	48
Table 14 στοιχεία πρωτοκόλλου για PA ακτινογραφία άνω κοιλίας.....	49
Table 15 περιγραφική στατιστική των τιμών δόσης για PA προβολή άνω κοιλίας.....	49
Table 16 στοιχεία πρωτοκόλλου για LL ακτινογραφία άνω κοιλίας.....	50
Table 17 περιγραφική στατιστική των τιμών δόσης για αριστερή LL ακτινογραφία άνω κοιλίας.....	50
Table 18 περιγραφική στατιστική των τιμών δόσης για δεξιά LL ακτινογραφία άνω κοιλίας.....	51
Table 19 αποτελέσματα για τα οστά.....	52
Table 20 αποτελέσματα για τον οισοφάγο .....	53
Table 21 αποτελέσματα για τους μαστούς .....	54
Table 22 αποτελέσματα για την καρδιά.....	55
Table 23 αποτελέσματα για τους πνεύμονες .....	56
Table 24 αποτελέσματα για το δέρμα.....	57
Table 25 αποτελέσματα για το integral.....	58

## ΕΙΚΟΝΕΣ

Εικόνα 1 το σύστημα CBCT XVI R5.0.4. της Elekta .....	2
Εικόνα 2 ανιχνευτικό πάνελ kV .....	2
Εικόνα 3 το μπράτσο της πηγής kV .....	3
Εικόνα 4 collimation cassette .....	3
Εικόνα 5 η επίδραση του φίλτρου bowtie στην ποιότητα της εικόνας .....	4
Εικόνα 6 medium FOV .....	5
Εικόνα 7 διάταξη για τον υπολογισμό του CTDI .....	7
Εικόνα 8 τα φαινόμενα underfitting και overfitting .....	9
Εικόνα 9 παραδείγματα ιστογραμμάτων .....	11
Εικόνα 10 παράδειγμα διαγράμματος boxplot. Με κύκλους φαίνονται τα outliers.....	11
Εικόνα 11 scatter plots .....	12
Εικόνα 12 παράδειγμα ενός correlation matrix.....	13
Εικόνα 13 linear regression με $R^2=0.706$ .....	15
Εικόνα 14 γραφική παράσταση των συντελεστών του μοντέλου συναρτήσε του "α". Για $\alpha=0$ το μοντέλο είναι απλό linear regression, ενώ για πολύ μεγάλα α όλοι οι συντελεστές μηδενίζονται .....	16
Εικόνα 15 διάγραμμα των συντελεστών "α" των Lasso και Elastic Net, για τους ίδιους συντελεστές .....	17
Εικόνα 16 προσαρμογή δεδομένων χρησιμοποιώντας Decision Tree regression. όσο πιο βαθύ το δέντρο τόσο πιο περίπλοκος ο αλγόριθμος, αλλά τόσο καλύτερο το fitting.....	17
Εικόνα 17 σχεδιάγραμμα του τρόπου λειτουργίας των Extremely Randomized Trees Regressor .....	19
Εικόνα 18 διάγραμμα του σφάλματος της πρόβλεψης συναρτήσε των αδύναμων μοντέλων που χρησιμοποιούνται. είναι εμφανής η μείωση του σφάλματος των εκάστοτε αλγορίθμων όταν χρησιμοποιούνται αυτόνομα και όταν συνδυάζονται στον AdaBoost Regressor.....	20
Εικόνα 19 εικόνα σάρωσης θώρακα στο Impact MC. Παρατηρείται ότι φαίνονται η axial, η sagittal και η coronal προβολή του όγκου .....	23
Εικόνα 20 scan parameters.....	24
Εικόνα 21 simulation parameters .....	25
Εικόνα 22 εικόνα δόσης όπως προκύπτει από το Impact MC.....	26
Εικόνα 23 εισαγωγή εντολών και dataset .....	27
Εικόνα 24 εντολές για προβολή των 20 πρώτων γραμμών του dataset, των χαρακτηριστικών του, αλλά και των συσχετίσεων μεταξύ των μεταβλητών βάση της μεθόδου Pearson .....	28
Εικόνα 25 output των παραπάνω εντολών .....	28
Εικόνα 26 εντολές για τη δημιουργία ιστογραμμάτων, boxplots, διαγραμμάτων διασποράς και διαγραμμάτων συσχετίσεων .....	29
Εικόνα 27 ιστόγραμμα μεταβλητών.....	29
Εικόνα 28 boxplots των μεταβλητών. Με κύκλους φαίνονται τα outliers.....	30
Εικόνα 29 scatter plots .....	30
Εικόνα 30 correlation matrix .....	31
Εικόνα 31 εφαρμογή του Local Outlier Factor και split του dataset.....	32
Εικόνα 32 output του Local Outlier Factor .....	32
Εικόνα 33 δοκιμή των μοντέλων.....	33
Εικόνα 34 negative mean squared errors των μοντέλων .....	34
Εικόνα 35 boxplot των γραμμικών μοντέλων.....	34



Εικόνα 36	boxplot των γραμμικών μοντέλων για κανονικοποιημένα δεδομένα.....	35
Εικόνα 37	boxplot των ensemble μεθόδων.....	35
Εικόνα 38	boxplot των ensemble μεθόδων για κανονικοποιημένα δεδομένα.....	36
Εικόνα 39	data scaling και tuning του Lasso.....	37
Εικόνα 40	στην πρώτη γραμμή φαίνεται η καλύτερη τιμή για το "α", ενώ στη δεύτερη το καλύτερο score .....	37
Εικόνα 41	κώδικας για το tuning του AdaBoost Regressor. Στην πρώτη σειρά φαίνονται οι τυχαίες τιμές που χρησιμοποιήθηκαν.....	37
Εικόνα 42	output του παρπάνω κώδικα .....	37
Εικόνα 43	κώδικας για το tuning του Extra Trees Regressor.....	38
Εικόνα 44	πρόβλεψη δόσης και σύγκρισή της με τις πραγματικές τιμές .....	39
Εικόνα 45	περιγραφική στατιστική των τμών Y_validation και των προβλέψεων.....	40
Εικόνα 46	οι προβλέψεις συναρτήσει των πραγματικών τιμών (κόκκινη γραμμή), και η ευθεία $y=x$ (μπλε γραμμή) .....	40
Εικόνα 47	εικόνα δόσης από CBCT θώρακα όπως προέκυψε από το ImpactMC .....	42
Εικόνα 48	Εικόνα δόσης από CBCT άνω κοιλίας όπως προέκυψε από το ImpactMC .....	43
Εικόνα 49	εικόνα δόσης για AP ακτινογραφία θώρακα όπως προέκυψε από το ImpactMC.....	44
Εικόνα 50	εικόνα δόσης για PA ακτινογραφία θώρακα όπως προέκυψε από το ImpactMC.....	45
Εικόνα 51	Εικόνα δόσης για αριστερή LL ακτινογραφία θώρακα όπως προέκυψε από το ImpactMC.....	46
Εικόνα 52	εικόνα δόσης για δεξιά LL ακτινογραφία θώρακα όπως προέκυψε από το ImpactMC.....	47
Εικόνα 53	εικόνα δόσης για AP ακτινογραφία άνω κοιλίας όπως προέκυψε από το ImpactMC.....	48
Εικόνα 54	εικόνα δόσης για PA ακτινογραφία άνω κοιλίας όπως προέκυψε από το ImpactMC.....	49
Εικόνα 55	εικόνα δόσης για αριστερή LL ακτινογραφία άνω κοιλίας όπως προέκυψε από το ImpactMC .....	50
Εικόνα 56	εικόνα δόσης για δεξιά LL ακτινογραφία άνω κοιλίας όπως προέκυψε από το ImpactMC.....	51
Εικόνα 57	διάγραμμα Bland-Altman για τα οστά .....	52
Εικόνα 58	διάγραμμα Bland-Altman για τον οισοφάγο.....	53
Εικόνα 59	διάγραμμα Bland-Altman για τους μαστούς.....	54
Εικόνα 60	διάγραμμα Bland-Altman για την καρδιά .....	55
Εικόνα 61	διάγραμμα Bland-Altman για τους πνεύμονες .....	56
Εικόνα 62	διάγραμμα Bland-Altman για το δέρμα .....	57
Εικόνα 63	διάγραμμα Bland-Altman για το integral .....	58