

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



—Πτυχιακή Εργασία—

Συστήματα υποστήριξης αποφάσεων για ασθενείς με Χρόνια Αποφρακτική Πνευμονοπάθεια

Βλάχος Χαράλαμπος Γεώργιος

Επιβλέπον: Θεμιστοκλής Έξαρχος

Επιβλέπων

Θεμιστοκλής Έξαρχος Αναπληρωτής Καθηγητής

Ιόνιο Πανεπιστήμιο

Τριμελής Επιτροπή

Θεμιστοκλής Έξαρχος Αναπληρωτής Καθηγητής

Ιόνιο Πανεπιστήμιο

Κάτια-Λήδα Κερμανίδου Αναπληρώτρια Καθηγήτρια

Ιόνιο Πανεπιστήμιο

Αριστείδης Βραχάτης Επίκουρος Καθηγητής

Ιόνιο Πανεπιστήμιο

A'.1 Περίληψη

Η παρούσα εργασία αναφέρεται σε μια εφαρμογή για την διαχείριση ατόμων με Χρόνια αποφρακτική πνευμονοπάθεια (ΧΑΠ). Η ΧΑΠ είναι μια χρόνια πάθηση των βρόγχων και των πνευμόνων. Σύμφωνα με τον ΠΟΥ (Παγκόσμιο οργανισμό υγείας) 65 εκατομμύρια άνθρωποι παγκοσμίως πάσχουν από ΧΑΠ με μέτρια έως σοβαρά συμπτώματα. Η πάθηση αυτή είναι η τρίτη κυριότερη αιτία θανάτου, έχοντας 5% του συνολικού αριθμού θανάτων παγκοσμίως. Βασικός παράγοντας όξυνσης της πάθησης αυτής είναι το κάπνισμα. Παράλληλα, η νόσος αυτή είναι δύσκολο να ανιχνευτεί στα αρχικά της στάδια λόγω των μη διακριτών συμπτωμάτων της, με αποτέλεσμα αυτή η νόσος να χειροτερεύσει. Στην χώρα μας έχουμε πάνω από 700.000 ασθενείς, από τους οποίους οι 50 % δεν γνωρίζουν ότι πάσχουν από ΧΑΠ. Για την έγκυρη διάγνωση της πάθησης υπάρχουν ερωτηματολόγια σχεδιασμένα για αυτό το σκοπό. Τα ερωτηματολόγια αυτά είναι το COPD Assessment Test (CAT), το Saint George Respiratory Questionnaire (SGRQ) και η παραλλαγή του SGRQ-C. Παρόλα αυτά τα ερωτηματολόγια αυτά βασίζονται αποκλειστικά στην αυτοεκτίμηση του ασθενή και μπορούν να προκύψουν τυχών λάθη. Παράλληλα, για την διάγνωση των ασθενών που πάσχουν από αναπνευστικές παθήσεις όπως η ΧΑΠ υπάρχει ο έλεγχος σπιρομέτρησης. Ο έλεγχος μας δίνει σαν αποτέλεσμα της πνευμονικής λειτουργίας του ασθενή, και είναι πιο ακριβής από τα ερωτηματολόγια. Σκοπός της εργασίας μας είναι να δούμε ποια από τα ακόλουθα στοιχεία των ασθενών είναι σημαντικά για την διάγνωση της ΧΑΠ, και με την χρήση της μηχανικής μάθησης να φτιάξουμε ένα μοντέλο για την διάγνωση, βασισμένο σε πραγματικά δεδομένα ασθενών με ΧΑΠ. Για την δημιουργία του μοντέλου χρησιμοποιήθηκαν βιβλιοθήκες της γλώσσας Python καθώς και 98 δείγματα από ασθενής διαγνωσμένους με ΧΑΠ.

A'.2 Ευχαριστίες

Θέλω να ευχαριστήσω θερμά:

Τον επιβλέποντα καθηγητή της πτυχιακής μου εργασίας Δρ Θεμιστοκλή Έξαρχο για την καθοδήγηση, τις πολύτιμες συμβουλές, τις πολύτιμες συμβουλές και υποδείξεις του καθώς και τη βοήθεια του σε όλη τη διάρκεια της εκπόνησής της.

Περιεχόμενα

Περιεχόμενα

A' .1	Περίληψη.....	4
A'.2	Ευχαριστίες	4
	Κατάλογος Εικόνων	7
	Κατάλογος Πινάκων	8
A'.3	Εισαγωγή.....	9
A'.3.1	Αναπνευστικό σύστημα	9
A'.3.2	Χρόνια Αποφρακτική Πνευμονοπάθεια (ΧΑΠ ,COPD)	10
B'.	Βιβλιογραφική Ανασκόπηση	12
B'. 1	Εφαρμογή Συστημάτων Τεχνητής Νοημοσύνης	12
B'.2	Παρόμοιες εργασίες.....	12
B'.3	Άρθρο Kaplan et al [1]	13
B'.4	Άρθρο V. Nikolaou et al [2].....	13
B'.5	Άρθρο T. Goto et al (1) [3]	14
B'.6	Άρθρο P. Soltani Zarrin et al [4].....	15
B'.7	Άρθρο I. Mohamed et al [5].....	16
B'.8	Κείμενο Junfeng Peng et al [6]	18
B'.9	Κείμενο T. Goto et al (2)[7].....	19
B'.10	Κείμενο Wu et al [8]	20
B'.11	Κείμενο Xu Min, BinYu & FeiWang [9].....	21
B'.12	Κείμενο Dimitris Spathis, Panayiotis Vlamos [10]	23
Γ'.	Δεδομένα Μελέτης	25
Γ'.1	Αρχικά δεδομένα	25
Γ'2.	Επεξεργασία Δεδομένων	25
Γ'.2.1	Data cleaning.....	26
Γ'.2.2	Attribute Selection	26
Γ'.2.3	Εξισορρόπηση Δεδομένων	28
Γ'.3	Επιλογή Χαρακτηριστικών (Feature Selection).....	28
Γ'.3.1	Kendal Rank.....	28
Γ'.3.2.	Chi-Square	30
Δ'.	Αλγόριθμοι Μηχανικής Μάθησης.....	34
Δ'.1	Αλγόριθμοι Δέντρων απόφασης(Decision Trees Algorithms).....	34

Δ'.2	Αλγόριθμος CART(Classification and Regression Trees)	34
Δ'.3	Αλγόριθμος Adaboost	35
Δ'.3.1	Παράδειγμα αλγορίθμου	36
Δ'.4	Εκπαίδευση συστήματος	37
Δ'.5	Εργαλεία του προγράμματος	37
Ε'	Αποτελέσματα	38
Ε'.1	Πίνακας Σύγχυσης (Confusion Matrix)	38
Ε'.2	Μετρικές	39
Ε'.3	Δέντρα Αποφάσεων CART	40
Στ'	Περιγραφή Εφαρμογής	41
Στ'.1	Εργαλεία δημιουργίας	41
Στ'.2	Αρχική Σελίδα	41
Στ'.3	Έλεγχος	49
Στ'.4	Ερωτηματολόγιο CAT	50
Στ'.5	Ερωτηματολόγιο HAD	52
Στ'.6	Ερωτηματολόγιο SGRQ	53
Στ'.6.1	SGRQ-C	53
Στ'.7	Αποτελέσματα φόρμας	55
Ζ'	Συμπεράσματα	56
Η'	Παραρτήματα	56
	Παράρτημα Α Python	56
	Παράρτημα Β HTML	59
	Παράρτημα Γ CSS	63
	Παράρτημα Δ Javascript	68
	Παράρτημα Ε PHP	73
Θ'	Βιβλιογραφία	75

Κατάλογος Εικόνων

Εικόνα 1: Όργανα Αναπνευστικού Συστήματος.....	9
Εικόνα 2: Κυψελίδες.....	10
Εικόνα 3: Πίνακας Ταξινόμησης Φαινοτύπων ΧΑΠ	11
Εικόνα 4 : Σύγκριση αξίας των χαρακτηριστικών στο διάστημα ενός μήνα και τριών μηνών	18
Εικόνα 5 : Σύγκριση αξίας χαρακτηριστικών.....	19
Εικόνα 6 : Τρεις τύποι αναπαράστασης χρονικής διάρκειας συμβάντων α) Σειριακή Αναπαράσταση β) Αναπαράσταση Πίνακα γ) Αναπαράσταση Επεξεργασμένου Πίνακα.....	22
Εικόνα 7 : Αναπαράσταση των τιμών της κλάσης.....	27
Εικόνα 8 : Τύπος υπολογισμού δείκτη tau.....	29
Εικόνα 9 : Τύπος υπολογισμού αναμενόμενων τιμών.....	31
Εικόνα 10 : Τύπος υπολογισμού τιμών πίνακα chi-Square.....	32
Εικόνα 11 : Δέντρο απόφασης του παραδείγματος	34
Εικόνα 12 : Τύπος υπολογισμού Gini.....	35
Εικόνα 13 : Tree-Stump	36
Εικόνα 14 : Παράδειγμα δέντρων Adaboost.....	37
Εικόνα 15: Confusion Matrix Python.....	38
Εικόνα 16: Βέλτιστο δέντρο σύμφωνα με τον Adaboost	40
Εικόνα 17 : Αρχικά στοιχεία ασθενή	41
Εικόνα 18 : Ερωτήσεις για το ιστορικό του ασθενή.....	42
Εικόνα 19 : Συννοσηρότητες ασθενή	42
Εικόνα 20 : Αποτελέσματα ελέγχου του ασθενή	43
Εικόνα 21 : Αποτελέσματα Ερωτηματολογίων	44
Εικόνα 22 : Εξήγηση Pack Years	44
Εικόνα 23: Εξήγηση MWT1Best	45
Εικόνα 24 : Εξήγηση FEV1	46
Εικόνα 25 : Εξήγηση FEV1PRED.....	46
Εικόνα 26 : Εξήγηση FVC	46
Εικόνα 27 : Εξήγηση FVCPRED.....	47
Εικόνα 28 : Εξήγηση CAT	48
Εικόνα 29 : Εξήγηση HAD	48
Εικόνα 30 :Εξήγηση SGRQ	48
Εικόνα 31 : Είσοδος Λάθος Τιμής.....	49
Εικόνα 32 : Κενή Είσοδος	50
Εικόνα 33 : Ερωτηματολόγιο CAT	50
Εικόνα 34 : Ερωτηματολόγιο CAT, δεύτερη σελίδα.....	51
Εικόνα 35 : Αποτελέσματα ερωτηματολογίου CAT	51
Εικόνα 36 : Ερωτηματολόγιο HAD	52
Εικόνα 37 : Ερωτηματολόγιο SGRQ.....	55
Εικόνα 38 : Αποτελέσματα Ερωτηματολογίου SGRQ	55

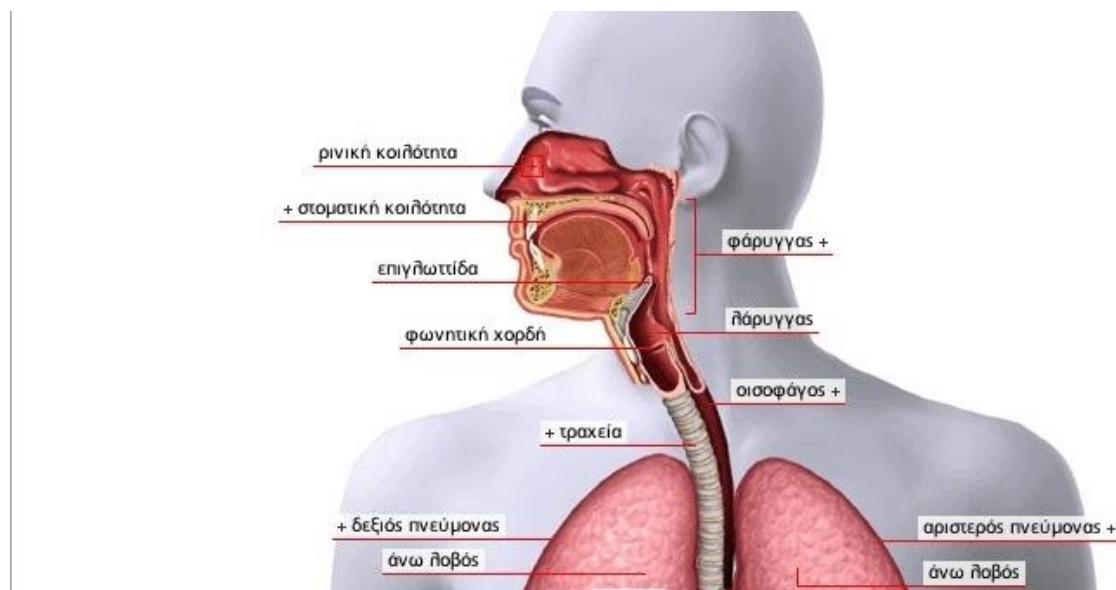
Κατάλογος Πινάκων

Πίνακας 1: Αποτελέσματα Απόδοσης Πληθυσμού 80 δειγμάτων.....	16
Πίνακας 2: Αποτελέσματα Απόδοσης Πληθυσμού 239 δειγμάτων.....	16
Πίνακας 3: Timeframe ενός μήνα	17
Πίνακας 4: Timeframe τριών μηνών	17
Πίνακας 5: Σύγκριση της απόδοσης των αλγορίθμων	18
Πίνακας 6: Πίνακας σύγκρισης μετρικής Precision και Fi	24
Πίνακας 7: Πίνακας σύγκρισης χαρακτηριστικών.....	24
Πίνακας 8 : Πίνακας χαρακτηριστικών και τιμής Tau	30
Πίνακας 9 : Πίνακας πραγματικών τιμών chi-Square	30
Πίνακας 10 : Πίνακας αναμενόμενων τιμών chi-Square.....	31
Πίνακας 11 : Πίνακας chi-Square	32
Πίνακας 12 : Αποτελέσματα feature selection για κατηγορικά χαρακτηριστικά	33
Πίνακας 13: Confusion Matrix Weka.....	39
Πίνακας 14 : Μετρικές Weka	39
Πίνακας 15: Μετρικές Python	39

Α'3 Εισαγωγή

Α'3.1 Αναπνευστικό σύστημα

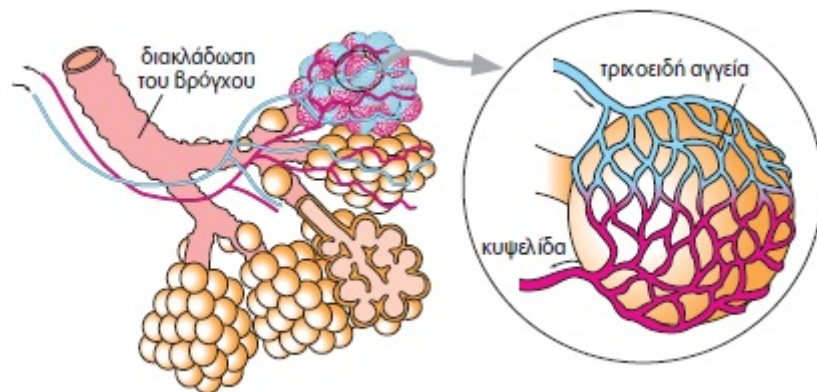
Η αναπνοή είναι μια πολύ σημαντική διαδικασία για την επιβίωση του οργανισμού. Η αναπνοή αποτελείται από δύο διαδικασίες την εισπνοή και την εκπνοή. Κατά την διαδικασία της εισπνοής εισάγουμε αέρα με οξυγόνο στους πνεύμονες και στην συνέχεια σε όλο το σώμα, εφοδιάζοντας τον οργανισμό μας με το απαραίτητο οξυγόνο. Αντίθετα, στην διαδικασία της εκπνοής αποβάλλουμε χρησιμοποιημένο αέρα με διοξείδιο του άνθρακα στο περιβάλλον.



Εικόνα 1: Όργανα Αναπνευστικού Συστήματος

<http://jvalak.blogspot.com/2013/01/blog-post.html>

Αναπνευστικό σύστημα είναι το σύνολο των οργάνων που συμμετέχουν στην διαδικασία της αναπνοής. Τα όργανα που εισέρχεται και αποβάλλεται ο αέρας είναι η ρινική κοιλότητα και η στοματική κοιλότητα. Σε κάθε εισπνοή και εκπνοή επιλέγεται μόνο ένα μονοπάτι από τον άνθρωπο. Οι δύο δρόμοι αυτοί ενώνονται στον φάρυγγα και στην συνέχεια οδηγούνται στον λάρυγγα. Στον λάρυγγα διαφοροποιείτε το πεπτικό σύστημα από το αναπνευστικό. Ο αέρας που μεταφέρεται μέσω του αναπνευστικού συστήματος περνάει από την τραχεία η οποία διακλαδίζεται σε δύο βρόγχους δεξιά και αριστερά. Κάθε βρόγχος οδηγεί στον αντίστοιχο πνεύμονα. Σε κάθε πνεύμονα ο κάθε βρόγχος διακλαδίζεται σε μικρότερους βρόγχους ,τα βρογχιόλια, σχηματίζοντας έτσι ένα βρογχικό δέντρο. Στις άκρες του δέντρου βρίσκονται οι κυψελίδες.



Εικόνα 2: Κυψελίδες

http://ebooks.edu.gr/ebooks/v/html/8547/2250/Biologia_A-Gymnasiou_html-empl/index4_4.html

Οι κυψελίδες είναι μικροσκοπικοί ελαστικοί σάκοι που περιβάλλονται από πολύ μικρά αιμοφόρα αγγεία, τα λεγόμενα τριχοειδή αγγεία. Στις κυψελίδες πραγματοποιείται η ανταλλαγή των αερίων, οξυγόνου και διοξειδίου του άνθρακα, από τα λευκά αιμοσφαίρια που βρίσκονται μέσα στα τριχοειδή αγγεία. Στην διαδικασία της εισπνοής κάθε κυψελίδα φουσκώνει και πραγματοποιείται η ανταλλαγή αερίων, με αυτό τον τρόπο τα ερυθρά αιμοσφαίρια ανταλλάσσουν το διοξείδιο του άνθρακα που μεταφέρουν με οξυγόνο. Στην συνέχεια τα ερυθρά αιμοσφαίρια μεταφέρουν οξυγόνο σε όλα τα μέλη του σώματος. Στην διαδικασία της εκπνοής οι σάκοι ξεφουσκώνουν και το διοξείδιο του άνθρακα αποβάλλεται από τον οργανισμό.

Α'3.2 Χρόνια Αποφρακτική Πνευμονοπάθεια (ΧΑΠ, COPD)

Η Χρόνια Αποφρακτική Πνευμονοπάθεια είναι μια ομάδα πνευμονικών παθήσεων, όπως το εμφύσημα, η χρόνια βρογχίτιδα και το άσθμα. Οι ασθενείς που νοσούν με ΧΑΠ παρουσιάζουν δυσκολία στην αναπνοή λόγω της στένωσης των αεραγωγών και την φλεγμονή των πνευμόνων. Αυτό συμβαίνει διότι τα τοιχώματα των αεραγωγών γίνονται παχύτερα και παράγεται μεγαλύτερη ποσότητα βλέννας απ' ότι χρειάζεται, δυσκολεύοντας την πορεία του αέρα. Παράλληλα, οι κυψελίδες χάνουν την ελαστικότητά τους με αποτέλεσμα να παγιδεύεται ο αέρας σε αυτές και να μην πραγματοποιείται η ανταλλαγή των αερίων και η μεταφορά του οξυγόνου στο σώμα. Τυπικά συμπτώματα της νόσου είναι βήχας με φλέματα, δύσπνοια, συριγμός, συχνές λοιμώξεις και σφίξιμο στο στήθος.

Η νόσος αυτή είναι πολύ δύσκολο να διαγνωστεί νωρίς. Αυτό συμβαίνει, διότι στα αρχικά στάδια της νόσου οι ασθενείς τις περισσότερες φορές παρουσιάζουν ήπια συμπτώματα ή δεν παρουσιάζουν συμπτώματα. Αυτό έχει σαν αποτέλεσμα η νόσος να μην ανιχνεύεται εγκαίρως, ώστε να αντιμετωπιστεί με κατάλληλες θεραπείες. Οι ασθενείς με σοβαρά συμπτώματα παρουσιάζουν σοβαρή δυσκολία στην αναπνοή και εξάντληση, ακόμα και όταν ασχολούνται με απλές καθημερινές δραστηριότητες,

όπως να ανεβαίνουν σκάλες ή να ντύνονται. Η ΧΑΠ είναι μία από τις πιο κοινές παθήσεις του αναπνευστικού συστήματος και προβλέπεται ότι θα γίνει κύρια αιτία θανάτου παγκοσμίως την τελευταία δεκαετία, εξαιτίας της δυσκολίας έγκαιρης και ακριβής διάγνωσης.

Η κύρια αιτία νόσησης είναι το κάπνισμα, το οποίο αποτελεί το 85-90% των περιστατικών. Επιπλέον παράγοντες που οφείλονται και χειροτερεύουν την νόσο είναι η ατμοσφαιρική ρύπανση και η συνεχόμενη έκθεση του οργανισμού σε σκόνη και αναθυμιάσεις.

Για την βελτίωση της διάγνωση της ΧΑΠ, ο βαθμός σοβαρότητας της νόσου έχει ταξινομηθεί σε τέσσερις κατηγορίες ανάλογα με τον φαινότυπο των ασθενών. Αυτοί οι φαινότυποι αξιολογούνται μέσω κλινικών εξετάσεων και ταξινομούνται σύμφωνα με το GOLD(Παγκόσμια αποφρακτική πνευμονοπάθεια). Το GOLD ταξινομεί τους ασθενείς στις τέσσερις κατηγορίες του παρακάτω πίνακα ανάλογα με τα συμπτώματα και το ιστορικό των παροξύνσεων των ασθενών.

Symptoms		
Moderate/severe exacerbation history	mMRC 0-1 and CAT<10	mMRC≥2 and CAT≥10
≥2 or ≥1 leading to hospital admission	C	D
0 or 1 not leading to hospital admission	A	B

mMRC, modified Medical Research Council dyspnea questionnaire; CAT, COPD assessment test.

Εικόνα 3: Πίνακας Ταξινόμησης Φαινοτύπων ΧΑΠ

<https://www.sciencedirect.com/science/article/pii/S095461112030233X>

Προκειμένου η κατάσταση της νόσου να καλυτερεύσει και οι ασθενείς να ζήσουν καλύτερα συνιστάτε η διακοπή του καπνίσματος και η αποφυγή παραγόντων παροξυσμού. Παράλληλα συστήνεται και η θεραπεία φαρμάκων ,διαφορετικά για κάθε κατηγορία GOLD, με την χρήση συσκευών εισπνοής. Για την κατηγορία Α(Ηπια συμπτώματα) προτείνεται χρήση φαρμάκου LAMA(Long-Acting Muscarinic Antagonists). Για την κατηγορία Β(Μέτρια συμπτώματα) προτείνεται χρήση φαρμάκου LAMA. Σε περίπτωση που τα συμπτώματα δεν καλυτερεύσουν άμεσα, πρέπει να εξετασθεί το ενδεχόμενο χρήσης φαρμάκων LABA(Long-Acting Beta Antagonists) και ICS(Inhaled Corticosteroids) καθώς και το ενδεχόμενο παραπομπής για πνευμονική αποκατάσταση και μεταμόσχευση πνεύμονα. Για τις κατηγορίες C και D (Σοβαρά συμπτώματα) συνιστάτε θεραπεία με ICS,LABA και LAMA μαζί με θεραπεία οξυγόνου. Παράλληλα, απαιτούνται συχνές παραπομπές για πνευμονική αποκατάσταση και πιθανή αξιολόγηση μεταμόσχευσης πνεύμονα.

B'. Βιβλιογραφική Ανασκόπηση

B'.1 Εφαρμογή Συστημάτων Τεχνητής Νοημοσύνης

Η τεχνολογία πλέον έχει αναπτυχθεί σε μεγάλο βαθμό και έχει εφαρμογή σε πολλούς τομείς της καθημερινής μας ζωής κάνοντας πολλές διαδικασίες να γίνονται χωρίς την παρέμβαση του ανθρώπου. Αξίζει να σημειωθεί ότι, η τεχνητής νοημοσύνης τα τελευταία χρόνια εφαρμόζεται συχνά στην η τέχνη και το marketing . Με την βοήθειά της τεχνητής νοημοσύνης μπορούμε να πραγματοποιήσουμε διαδικασίες οι οποίες παλαιότερα χρειαζόταν την παρουσία ανθρώπου. Οι εργασίες αυτές είναι επικίνδυνες(όπως εργασία σε εργοστάσια με ραδιενέργεια) ή επαναλαμβανόμενες (όπως η βίδωση μία βίδας).

Στον τομέα της Ιατρικής, η Τεχνητή Νοημοσύνη δεν μπορεί να αντικαταστήσει την γνώση και την εμπειρία ενός Γιατρού. Η Ιατρική είναι επιστήμη η οποία απαιτεί πολλά χρόνια αφοσίωσης και μελέτης προκειμένου να αποκτήσει κάποιος τις βασικές γνώσεις ώστε να εξασκήσει το επάγγελμα του ιατρού. Παράλληλα, η Ιατρική έχει πολλούς τομείς στους οποίους απαιτούνται παραπάνω χρόνια εκπαίδευσης προκειμένου ένας γιατρός να αποκτήσει εξειδίκευση σε κάποιον από αυτούς τους τομείς. Παρόλα αυτά, η τεχνητή νοημοσύνη βοηθάει στην μείωση και εξάλειψη του ανθρώπινου λάθους σε εργασίες που χρειάζονται ιδιαίτερη προσοχή και ακρίβεια. Παραδείγματα είναι η διάγνωση μιας ασθένειας, η επιλογή θεραπείας για συγκεκριμένο ασθενή και οι κλινικές δοκιμές ενός φαρμάκου. Στην εργασία αυτή παρουσιάζεται μια εφαρμογή που έφτιαξα για την πτυχιακή μου εργασία. Η εφαρμογή αυτή είναι ένα σύστημα υποστήριξης αποφάσεων για άτομα με χρόνια αποφρακτική πνευμονοπάθεια. Το σύστημα της εκπαιδευτική από ανωνυμοποιημένα δεδομένα ασθενών που βρίσκονται στο kaggle.com. Τα δεδομένα αυτά ανήκουν σε ασθενείς οι οποίοι είχαν προσκληθεί να συμμετάσχουν σε ένα πρόγραμμα αποκατάστασης για να βοηθήσουν την διαχείριση της ΧΑΠ τους.

B'.2 Παρόμοιες εργασίες

Η δημιουργία και εκπαίδευση συστημάτων τεχνητής νοημοσύνης για τον εντοπισμό πνευμονικών παθήσεων όπως η ΧΑΠ συνηθίζεται τα τελευταία χρόνια από για την διαχείριση ασθενών σε όλο τον κόσμο. Στις μελέτες αυτές οι ερευνητές παίρνοντας ανωνυμοποιημένα δεδομένα συνήθως από βάσεις δεδομένων νοσοκομείων και προσπαθούν να φτιάξουν κατάλληλο μοντέλο πρόβλεψης ή ταξινόμησης των ασθενών σε κατηγορίες ανάλογα τον σκοπό της έρευνας. Στα περισσότερα άρθρα σκοπός τους ήταν η πρόβλεψη.

B'3 Άρθρο Kaplan et al [1]

Το άρθρο μελετάει την αξία των εφαρμογών της τεχνητής νοημοσύνης στην ιατρική. Στο σημείο της πνευμονικής Ιατρικής πραγματοποιήθηκε ένα πείραμα για την ταξινόμηση πνευμονικών παθήσεων, όπως το ΑΣΘΜΑ και η ΧΑΠ. Στο πείραμα αυτό συμμετείχαν 120 πνευμονολόγοι και ένα σύστημα ΑΙ το οποίο είχε εκπαιδευτεί με δείγματα από 1430 ασθενών. Οι πνευμονολόγοι και το σύστημα καλούνται να ταξινομήσουν 50 δείγματα ασθενών σε 8 κατηγορίες έχοντας ως στοιχεία τα αποτελέσματα των πνευμονολογικών τεστ των ασθενών και κάποιες επιπλέον πληροφορίες (ιστορικό καπνιστή, βήχας, φλέματα, δύσπνοια).

Οι κατηγορίες αυτές:

- Υγής
- Ασθενής με Άσθμα
- Ασθενής με ΧΑΠ
- Διάμεση πνευμονοπάθεια
- Νευρομυική πάθηση
- Πνευμονική αγγειακή νόσος
- Νόσος του θώρακα
- Άλλες αποφρακτικές παθήσεις

Τα αποτελέσματα έδειξαν ότι τα ποσοστά sensitivity και positive predictive value του συστήματος ήταν μεγαλύτερα από αυτά των ιατρών. Παράλληλα οι μετρικές ταξινόμησης των δειγμάτων σε κατηγορίες (άτομα με αποφρακτικά προβλήματα, περιοριστικά προβλήματα και υγιά) και η απόδοση του συστήματος (100% και 82% αντίστοιχα) ήταν μεγαλύτερες από αυτές των γιατρών (56-88% και 24-62%). Σύμφωνα με τους συγγραφείς το σύστημα αυτό δέχτηκε στοιχεία με μεταβλητές ίδιες με αυτές που είχε εκπαιδευτεί και για αυτό μπορούσε εύκολα να ταξινομήσει νέα δείγματα με μεγάλη ακρίβεια σε αντίθεση με τους ιατρούς οι οποίοι δεν είχαν αρκετές πληροφορίες για την διάγνωση τους.

B'4 Άρθρο V. Nikolaou et al [2]

Στο κείμενο αυτό οι συγγραφείς σκοπεύουν να εκμεταλλευτούν την μηχανική μάθηση και την ανάλυση συστάδων με σκοπό τον χαρακτηρισμό φαινοτύπων της ΧΑΠ λαμβάνοντας υπόψη πρόσθετα χαρακτηριστικά όπως οι συννοσηρότητες, βιοδείκτες και γονιδιακές πληροφορίες. Αυτό θα βοηθούσε τους ερευνητές να εντοπίσουν νέους φαινοτύπους της ΧΑΠ και να χαρακτηρίσουν καλύτερα τους υπάρχοντες. Η έρευνά τους περιελάμβανε 117 άρθρα από τα όποια 14 από αυτά επιλέχθηκαν για την έρευνα. Ο πληθυσμός των ερευνών ήταν 65 με 30.961 ασθενείς. Σε όλες τις μελέτες οι φαινότυποι της ΧΑΠ έχουν επικυρωθεί με κλινικά αποτελέσματα όπως παροξύνσεις, ποιότητα ζωής που σχετίζεται με την υγεία, ποσοστό θνησιμότητας και ανταπόκριση στη θεραπεία.

Η έρευνα των ECLIPSE έχοντας 2746 δείγματα με την χρήση cross-validation έδειξαν ότι

- οι ασθενείς με ΧΑΠ είχαν συχνά συννοσηρότητες, ιδιαίτερα καρδιακές.
- Οι άνδρες με ΧΑΠ ήταν πιο ευαίσθητοι από τις γυναίκες στην καρδιαγγειακή συννοσηρότητα.

Στην έρευνα των Pikoula et al

- Στους ασθενείς με συννοσηρή καρδιαγγειακή νόσο και διαβήτη, ποσοστά εισαγωγής στο νοσοκομείο για οξείες παροξύνσεις ΧΑΠ ήταν υψηλά.
- Οι άνδρες είναι πιο πιθανό να πεθάνουν από καρδιαγγειακή νόσο από τις γυναίκες.

Σε μεταγενέστερες έρευνες οι ασθενείς καταχωρήθηκαν σε 3 κατηγορίες

- Αργή μείωση πνευμονικής λειτουργίας
- Γρήγορη μείωση πνευμονικής λειτουργίας
- Σταθερά μοτίβα.

Η τελευταία εργασία έδειξε ότι

- το εμφύσημα επηρέαζε σημαντικά την πτώση του FEV1
- επιπλέον 25% του πληθυσμού είχε συννοσηρότητα ΧΑΠ με άσθμα στην οποία οι ασθενείς χαρακτηρίζονταν από περισσότερες παροξύνσεις.

Οι αρθρογράφοι πιστεύουν πως η εργασία τους θα βοηθήσει τις μελλοντικές έρευνες δίνοντας στους επόμενους ερευνητές στην επιλογή μεταβλητών, στην εξωτερική επικύρωση και ελλειπόντων δεδομένων. Παράλληλα, αφήνει για τις μελλοντικές έρευνες τον χαρακτηρισμό φαινοτύπου νεαρών ασθενών με γρήγορη μείωση της πνευμονικής λειτουργίας και καρδιαγγειακή συννοσηρότητα.

B'5 Άρθρο T. Goto et al (1) [3]

Σκοπός του άρθρου αυτού είναι η δημιουργία μοντέλου διαλογής για την πρόβλεψη της κλινικής πορείας του άσθματος και παρόξυνσης της ΧΑΠ με δεδομένα επισκεπτών στα ΤΕΠ. Παράλληλα, γίνεται σύγκριση της λογιστικής παλινδρόμησης(Logistic Regression) με άλλους τέσσερεις αλγορίθμους με δεδομένα των ΕΣΙ. Το δείγμα περιελάμβανε 137 εκατομμύρια επισκέψεις ασθενών στα ΚΕΠ. Για την διαλογή χρησιμοποιήθηκαν οι παρακάτω προγνωστικοί παράγοντες:

- Ηλικία
- Φύλο
- Τρόπος άφιξης(εισαγωγή με ασθενοφόρο ή είσοδος)
- Ζωτικά σημεία (θερμοκρασία, ρυθμός παλμών, συστολική και διαστολική αρτηριακή πίεση, αναπνευστικός ρυθμός και κορεσμός οξυγόνου)
- Συχνές ενοχλήσεις (δύσπνοια, βήχας, πόνος στο στήθος)

- Κατάσταση άσθματος
- Κατάσταση ΧΑΠ
- Συννοσηρότητες

Αυτό που προσπαθούν να προβλέψουν είναι η κρίσιμη φροντίδα (εισαγωγή σε ειδική μονάδα εντατικής θεραπείας ή ενδονοσοκομειακός θάνατος) και νοσηλεία (εισαγωγή σε χώρο ενδονοσοκομειακής περίθαλψης είτε μεταφορά σε νοσοκομείο οξείας θεραπείας).

Για την πρόβλεψη και σύγκριση φτιάχτηκαν 5 μοντέλα βασισμένα σε αλγόριθμους μηχανικής μάθησης

- Logistic Regression
- Lasso Regression
- Random Forest
- Gradient Boosted Decision Tree
- Deep Neural Network.

Σε όλα τα μοντέλα μηχανικής μάθησης χρησιμοποιήθηκε το 70% των ασθενών του ΕΣΙ για εκπαίδευση και το 30% για testing. Το ίδιο πήραμα έγινε με τα δείγματα των ΤΕΠ. Το ποσοστό αποτελεσμάτων κρίσιμης φροντίδας ήταν 4%, το μοντέλο logistic regression είχε την χαμηλότερη ικανότητα διάκρισης(C-statistics) 0.68, οι υπόλοιποι αλγόριθμοι είχαν C-statistics 0.79,0.76, 0.80, 0.79 αντίστοιχα. Παράλληλα, ο αλγόριθμος gradient boosted decision tree είχε επίσης το μεγαλύτερο ποσοστό ευαισθησίας (sensitivity) 79%. Το ποσοστό κρίσιμης φροντίδας ήταν 26%, το μοντέλο με τον αλγόριθμο Random Forest είχε το υψηλότερο C-statistics 0.83% και υψηλότερη ευαισθησία 0.75%.

B'6 Άρθρο P. Soltani Zarrin et al [4]

Σκοπός της εργασίας αυτής είναι η χρήση εφαρμογή εργαλείων μηχανικής μάθησης για τον προσδιορισμό της ΧΑΠ με την χρήση δεδομένων από δείγματα σάλιου των ασθενών από υγειονομικό έλεγχο (Health Control HC). Για τον σκοπό έγινε σύγκριση 5 μοντέλων μηχανικής μάθησης ένα με κάθε αλγόριθμο από τους ακόλουθους XGBoost,SVM,GBN,LR και ANN. Συνολικά, συλλέχτηκαν 239 δείγματα σάλιου από δυο ομάδες ατόμων από 160 ασθενείς με έλεγχο υγείας και 79 ασθενείς με ΧΑΠ. Πληροφορίες που συλλέχτηκαν εκτός του σάλιου ήταν :

- Το φύλο
- Η ηλικία
- smoking status(καπνιστής, πρώην καπνιστής ή μη καπνιστής)
- Ημερομηνία ανίχνευσης
- Συνθήκες δειγματοληψίας
- Ειδικές σημειώσεις σχετικά με της μόλυνση σάλιου με αίμα.

Για τον υπολογισμό της απόδοσης πραγματοποιήθηκαν 2 πειράματα. Στο πρώτο πείραμα πήραμε 80 τυχαία δείγματα από τον πληθυσμό, το δεύτερο πείραμα έγινε με όλα τα δείγματα του πληθυσμού. Και στα δύο πειράματα χρησιμοποιήθηκε 5-fold cross validation για τον διαχωρισμό των testing και training δειγμάτων με αναλογία 20-80%. Όπως φαίνεται στον παρακάτω πίνακα το καλύτερο μοντέλο είναι αυτό με τον αλγόριθμο XGBoost έχοντας την μεγαλύτερη ακρίβεια(Accuracy) και ευαισθησία (Sensitivity) από τα υπόλοιπα μοντέλα.

Ταξινομητής	5-fold Accuracy	Sensitivity	Specificity	Precision
XGBoost	91.25%	100%	88.89%	87.5%
SVM	91.25%	100%	77.78%	77.78%
GNB	87.5%	85.71%	100%	100%
LR	90%	100%	11.11%	46.66%
ANN	73.75%	100%	44.45%	46.66%

Πίνακας 1: Αποτελέσματα Απόδοσης Πληθυσμού 80 δειγμάτων

Ταξινομητής	5-fold Accuracy	Sensitivity	Specificity	Precision
XGBoost	92.05%	95.24%	100%	100%
SVM	92.05%	90.47%	100%	100%
GNB	92.89%	95.24%	100%	100%
LR	91.65%	95.24%	96.3%	95.24%
ANN	93.33%	90.47%	96.3%	95%

Πίνακας 2: Αποτελέσματα Απόδοσης Πληθυσμού 239 δειγμάτων

B'.7 Άρθρο I. Mohamed et al [5]

Σκοπός της μελέτης αυτής είναι η ανάπτυξη μοντέλου μηχανικής μάθησης για την πρόβλεψη επανεισαγωγής ,ασθενών με ΧΑΠ υψηλού κινδύνου, στο νοσοκομείο μετά από 30 ημέρες από την μέρα που έφυγαν από αυτό .Τα δεδομένα της μελέτης συλλέχθηκαν από την GH και συλλέχθηκαν από την στιγμή της εισαγωγής, όλη την διάρκεια που βρίσκονταν στο νοσοκομείο, μέχρι 3 μήνες αφού έφυγαν. Συνολικά , συλλέχθηκαν 195 αρχεία ασθενών με ΧΑΠ το κάθε ένα με 32 μεταβλητές. Στην συνέχεια τα αρχεία αυτά χωρίστηκαν σε τρία μέρη με αναλογία 2:2:1. Το πρώτο μέρος χρησιμοποιήθηκε για την εκπαίδευση, το δεύτερο για την επικύρωση και το τρίτο για δοκιμή. Οι αλγόριθμοι που χρησιμοποιήθηκαν στην μελέτη ήταν:

- Artificial Neural Networks(ANNS)
- Decision Trees(DTS)
- Support Vector Machines(SVMS)

Για τον διαχωρισμό των δειγμάτων χρησιμοποιήθηκε 5-fold cross validation. Τα μοντέλα χρησιμοποιήθηκαν για την πρόβλεψη σε 2 timeframes, ενός μήνα(ημέρες 1-30) και τριών μηνών(ημέρα 31-90) μετά την έξοδο των ασθενών.

Model	ACC	AUC	Precision	F1 Score
Tree-As (CHAID)	88.7	0.77	0	0/0
Tree-AS(CHAID)	88.7	0.77	0	0/0
CHAID	89.1	0.75	0.31	0.41
SVM-Polynomial	88.3	0.73	0.11	0/0
SVM-RBF	87.4	0.73	0	0/0
C5	89.9	0.72	0.42	0.48
ANN-MLP	64.8	0.68	0.11	0/0
ANN-RBF	87.4	0.67	0.16	0/0
SVM-Linear	81.3	0.63	0.15	0.17
SVM-Sigmoid	88.7	0.38	0	0/0

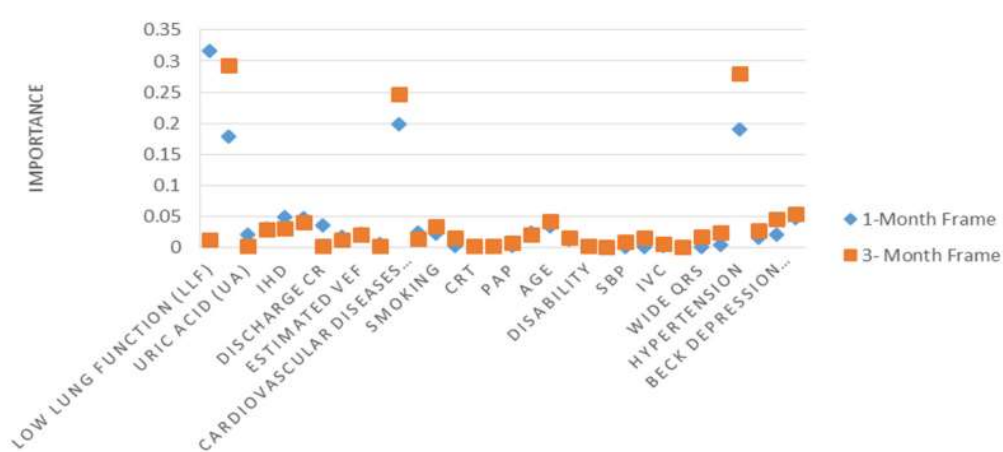
Πίνακας 3: Timeframe ενός μήνα

Model	ACC	AUC	Precision	F1 Score
Tree-As (CHAID)	67.7	0.62	0	0/0
Tree-AS(Exhaustive CHAID)	67.7	0.62	0.17	0/0
CHAID	53.0	0.46	0.24	0.25
SVM-Polynomial	53.0	0.45	0.23	0.24
SVM-RBF	60.0	0.47	0.11	0.14
C5	59.1	0.48	0.25	0.26
ANN-MLP	66.1	0.61	0.14	0.20
ANN-RBF	64.8	0.64	0.15	0/0
SVM-Linear	54.3	0.39	0.15	0.17
SVM-Sigmoid	54.3	0.46	0.08	0/0

Πίνακας 4: Timeframe τριών μηνών

Όπως φαίνεται στον πίνακα οι μετρικές ACC και AUC είναι υψηλότερες στο timeframe του ενός μήνα, καλύτερο μοντέλο στο πλαίσιο του ενός μήνα είναι το μοντέλο με τον C5 και στο πλαίσιο των τριών μηνών είναι αυτό με τον Tree-

As(CHAID). Στην συνέχεια κατέταξαν και τα χαρακτηρίστηκα ανάλογα με την αξία τους σε κάθε time-stamp.



Εικόνα 4 : Σύγκριση αξίας των χαρακτηριστικών στο διάστημα ενός μήνα και τριών μηνών

Β'8 Κείμενο Junfeng Peng et al [6]

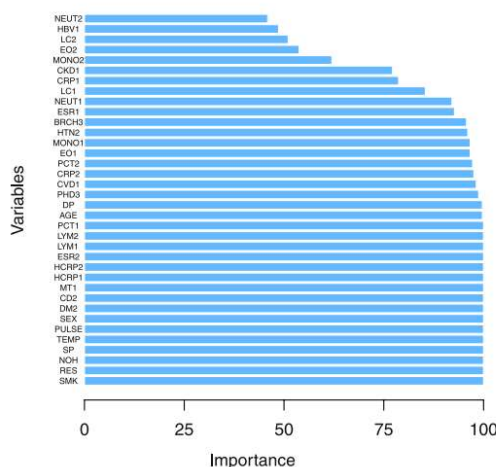
Στην εργασία αυτή στοχεύει στην δημιουργία μοντέλου για τον εντοπισμό της επιδείνωσης και τον κίνδυνο θανάτου σε ασθενείς με οξεία επιδείνωση της ΧΑΠ(AECOPD). Η μελέτη αυτή χρησιμοποίησε στοιχεία από 410 νοσηλευόμενους ασθενείς με AECOPD χρησιμοποιώντας 28 χαρακτηριστικά όπως ζωτικά σημεία, ιατρικό ιστορικό, συννοσηρότητες , και διάφορους φλεγμονώδεις δείκτες. Στην εργασία αυτή χρησιμοποιήθηκαν 4 διαφορετικοί αλγόριθμοι για την πρόβλεψη. Τα δείγματα χωρίστηκαν σε δεδομένα εκπαίδευσης/ δοκιμής 4 φορές για κάθε αλγόριθμο με αναλογία 90/10, 80/20, 70/30, 60/40, 50/50.

Μέθοδος	90/10	80/20	70/30	60/40	50/50	Μέση τιμή/απόκλιση
C5.0	0.725	0.803	0.730	0.724	0.702	0.737 +- 0.039
ABC4.5	0.65	0.704	0.713	0.699	0.629	0.679 +- 0.037
ABCA2	0.675	0.667	0.639	0.712	0.654	0.669 +- 0.027
ABID3	0.575	0.728	0.631	0.644	0.624	0.64 +- 0.0555

Πίνακας 5: Σύγκριση της απόδοσης των αλγορίθμων

Όπως φαίνεται στον πίνακα ο αλγόριθμος C5.0 είχε την καλύτερη απόδοση από τους υπόλοιπους και στους τέσσερεις διαχωρισμούς. Σύμφωνα με το κείμενο ο αλγόριθμος αυτός είχε απόδοση 80.3% και από αυτόν έχουν δημιουργηθεί 40 δέντρα αποφάσεων. Παράλληλα, φάνηκε ότι πολλές μεταβλητές ,όπως το ιστορικό καπνιστή, ο ρυθμός

αναπνοής, αριθμός νοσηλειών είχαν μεγάλη αξία για τον καθορισμό της Κλάσης στον αλγόριθμο C5.



Εικόνα 5 : Σύγκριση αξίας χαρακτηριστικών

Β'.9 Κείμενο T. Goto et al (2)[7]

Η εργασία αυτή γίνεται για την δημιουργία μοντέλου μηχανικής μάθησης για την πρόβλεψη επανεισαγάγεις ασθενών με ΧΑΠ στο νοσοκομείο εντός 30 ημερών, χρησιμοποιώντας δεδομένα από μια κρατική βάση υγειονομικής περίθαλψης. Η Βάση περιλαμβάνει τα εξής στοιχεία για κάθε ασθενή :

- Χαρακτηριστικά ασθενή (Πρωτογενείς και δευτερογενείς διαγνώσεις)
- Συννοσηρότητες
- Επιπλοκές κατά την διάρκεια της νοσηλείας με χρήση ICD-10
- Σοβαρότητα της νόσου (Hugh-Jones score)
- Επίπεδο συνείδησης κατά την εισαγωγή και την απαλλαγή
- Δραστηριότητες καθημερινής ζωής (ADL score)
- Επεμβάσεις
- Δόσεις φαρμάκων
- Ημερομηνία εξόδου

Από όλο τον πληθυσμό επέλεξαν τους ασθενείς οι οποίοι ήταν 40 ετών και άνω με πρωτογενή διάγνωση ΧΑΠ χρησιμοποιώντας τους κωδικούς ICD-10 και ασθενείς με πρωτογενή διάγνωση αναπνευστικής ανεπάρκειας και δευτερογενής διάγνωση ΧΑΠ. Συνολικά για την δημιουργία του μοντέλου χρησιμοποιήθηκαν 420 μεταβλητές. Τα ο πληθυσμός χωρίστηκε σε τυχαία σε τμήματα εκπαίδευσης και δοκιμής 70% και 30% αντίστοιχα. Για μοντέλο δοκιμής χρησιμοποίησαν τον αλγόριθμο logistic regression model. Παράλληλα για την σύγκριση χρησιμοποιήθηκαν 2 ακόμα μοντέλα το ένα με lasso regression και το άλλο με deep neural networks. Προκειμένου να υπολογίσουν τις μετρικές C-statistics, sensitivity και specificity χρησιμοποίησαν 5-fold cross validation. Από το 2011 με 2016 ο πληθυσμός περιελάμβανε 44.929 περιπτώσεις

νοσηλευόμενων από αυτά χρησιμοποιήθηκαν 3413 περιπτώσεις (7% των περιπτώσεων). Μετά το πείραμα φάνηκε ότι οι αλγόριθμοι Lasso Regression και Deep Neural Network είχαν μεγαλύτερο C-statistic και sensitivity αλλά είχαν μικρότερο specificity από τον Logistic regression. Παράλληλα, φάνηκε ότι το χαρακτηριστικό διάρκεια χρήσης σωλήνα τροφοδοσίας ήταν σημαντικό για την ταξινόμηση στο μοντέλο Lasso Regression με βαθμό 0,6, ακολουθούμενο με μεγάλη διαφορά από τα χαρακτηριστικά μετάγγιση αίματος(0.2) και θωρακοκέντηση(1.6).

B'.10 Κείμενο Wu et al [8]

Στόχος της μελέτης αυτής ήταν η ανάπτυξη ενός συστήματος για την έγκυρη πρόβλεψη των συμπτωμάτων της οξείας παρόξυνσης της ΧΑΠ(AECOPD) στις επόμενες 7 ημέρες χρησιμοποιώντας δεδομένα από τον τρόπο ζωής, συμπτώματα ασθενών των ασθενών και περιβαλλοντικούς παράγοντες που τους επηρεάζουν. Τα δεδομένα που σχετίζονταν με τον τρόπο ζωής(θερμοκρασία, υγρασία ,άλλα σωματίδια) συλλέχθηκαν με την χρήση φορητών συσκευών (Fit Versa) και συσκευών ανίχνευσης ποιότητας αέρα. Τα συμπτώματα της ΧΑΠ αξιολογήθηκαν από τους χρήστες με την χρήση ερωτηματολογίων. Για την πρόβλεψη χρησιμοποιήθηκαν τα μοντέλα random forest decision trees, k-nearest neighbor, linear discriminate analysis Adaboost και deep neural networks. Τα δεδομένα που χρησιμοποιήθηκαν ήταν από 67 ασθενείς με ΧΑΠ, σημειώνοντας 25 περιστατικά με AECOPD. Από την ανάλυση αυτή επιλεκτικέ το μοντέλο Deep Neural Networks με 92.1% ακρίβεια(accuracy), 94% ευαισθησία(sensitivity) ,90.4% ειδικότητα(specificity) και η μετρική AUROC(area under the receiver operating characteristic curve) ήταν 96.4%. Επιπλέον, οι σημαντικότερες μεταβλητές για το μοντέλο αυτό ήταν :

- Καθημερινά βήματα που ο ασθενής περπάτησε
- Πόσες σκάλες ανέβηκε μέσα στην μέρα

Οι ερευνητές αναφέρουν ότι η χρήση φορητών συσκευών και συσκευών ανίχνευσης αέρα για την συλλογή δεδομένων έδωσε μεγαλύτερη απόδοση από δεδομένα ερωτηματολογίων.

B'.11 Κείμενο Xu Min, BinYu & FeiWang [9]

Στόχος της μελέτης αυτής είναι η πρόβλεψη επανεισαγωγής στο νοσοκομείο εντός 30 ημερών από την ημέρα εξόδου. Τα χαρακτηρίστηκα των ασθενών καταχωρήθηκαν ως στοιχεία γνώσης(knowledge driven) είτε ως στοιχεία δεδομένων(data driven). Τα χαρακτηριστικά γνώσης είναι

- 4 στοιχεία από το Hospital Score(αριθμοί διαδικασιών κατά την παραμονή στο νοσοκομείο, αριθμός εισαγωγών σε νοσοκομεία το τελευταίο έτος, αριθμός νοσηλειών , τύπος εισαγωγής ευρετηρίου)
- Στοιχεία του Lace Score(διάρκεια διαμονής, οξεία παραδοχή, δείκτης συννοσηρότητας Charlson,αριθμός επισκέψεων στα ΤΕΠ εντός έξι μηνών)
- Επιπλέον 12 στοιχεία που συλλέχθηκαν(Όπως ηλικία, φύλο και διάρκεια διαμονής)

Τα χαρακτηριστικά δεδομένων είναι

- Στοιχεία διάγνωσης(τρεις πρώτοι κωδικοί του ICD-9, κωδικοί κλινικής εφαρμογής CSS, κωδικοί κατηγορίας ιεραρχικής κατάστασης).
- Διαδικασίες(κωδικοποιημένα στοιχεία των ασθενών από την εφαρμογή κλινικής ταξινόμησης CSS, κωδικοί τύπων επεξεργασίας Berenson-Eggers BETOS, κωδικοί εσόδων)
- Στοιχεία φαρμακείου (παγκόσμιος κωδικός φαρμάκου NDC, γενική θεραπευτική κατηγορία GTC)
- Τοποθεσίες
- Συχνότητα κάθε χαρακτηριστικού ανά χρονική περίοδο. Bag of words BoW
- Συχνότητα κάθε χαρακτηριστικού με συγκεκριμένες τιμές.Boolean Bag-of Words bBoW
- Συχνότητα όρου αντίστροφη συχνότητα εγγράφου TFIDF

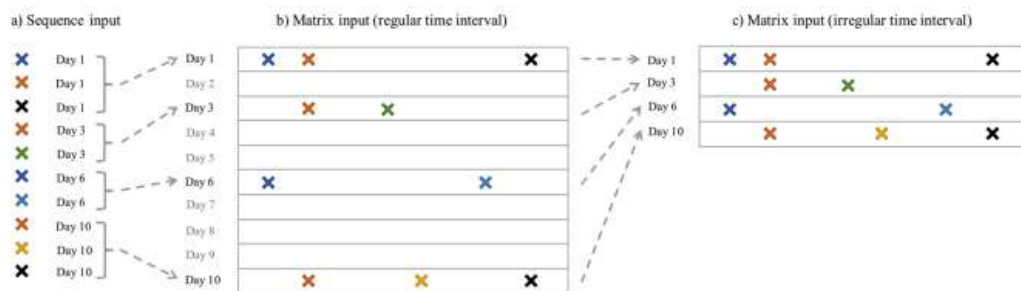
Στην εργασία αυτή χρησιμοποιήθηκαν τα παρακάτω μοντέλα κλασικής μάθησης:

- Λογιστική παλινδρόμηση(Logistic Regression) και παραδοχές με κοινωνικοποιήσεις κανόνων l1,l2.
- Random Forest (RF)
- Support Vector Machine(SVM)
- Gradient Boosting Decision Tree(GBDT)
- Multi-Layer Perception(MLP)

Επιπλέον χρησιμοποιήθηκαν τα μοντέλα βαθιάς μάθησης:

- Neural Networks (CNN)
- Recurrent Neural Networks (RNN)
- Long-Short Term Memory (LSTM)
- Gated Recurrent Unit (GRU)

Προκειμένου να διερευνηθούν ευκολότερα την χρονική διάρκεια κάθε συμβάντος στις αξιώσεις των ασθενών χρησιμοποιήθηκαν 3 τύποι αναπαράστασεων. Σειριακή αναπαράσταση, αναπαράσταση πίνακα, αναπαράσταση επεξεργασμένου πίνακα.



Εικόνα 6 : Τρεις τύποι αναπαράστασης χρονικής διάρκειας συμβάντων α) Σειριακή Αναπαράσταση β) Αναπαράσταση Πίνακα γ) Αναπαράσταση Επεξεργασμένου Πίνακα

Τα ακατέργαστα δεδομένα συνολικά ήταν 111.992. Για αυτό λοιπόν πραγματοποιήθηκε το ακόλουθη ακόλουθη επεξεργασία διαλογής 1) Κράτησαν ασθενής που το νοσοκομείο του δέχτηκε(MH approved) 2)Κράτησαν τους ασθενείς που είχαν συγκεκριμένη διάγνωση (τουλάχιστον έναν κωδικό από τους 491,492,496 ICD-9) 3) Κράτησαν τους ασθενείς που είναι τουλάχιστον 40 χρονών 4) Ασθενείς με καθορισμένο φύλο, 5) ήταν τουλάχιστον μια φορά εσωτερικοί ασθενείς στο νοσοκομείο 6) αυτούς με ιστορικό παρακολούθησης τουλάχιστον 60 ημέρες &) τουλάχιστον μια χρήση φαρμάκου στο ιστορικό τους. Αυτές οι διαδικασίες μείωσαν τον πληθυσμό στους 27.138 ασθενείς.

Η μέθοδος που χρησιμοποιήθηκε για τον διαχωρισμό των δεδομένων και τον υπολογισμό της απόδοσης των μοντέλων ήταν 5-fold cross validation.

Για την κλασική μάθηση

- Τα επιπλέον χαρακτηριστικά έχουν καλύτερη απόδοση από τις 2 βασικές μεθόδους των στοιχείων των Hospital Score και LACE score.
- Τα στοιχεία γνώσεων μαζί είχαν την καλύτερη απόδοση με μέσω όρο AUC 0.643 χρησιμοποιώντας GBDT.

Για την βαθιά μάθηση

- Η ταξινόμηση ακολουθίας χωρίς τον υπολογισμό του χρόνου δημιουργεί την χειρότερη απόδοση
- Η αναπαράσταση πίνακα είναι καλύτερη, ο ανεπτυγμένος πίνακας δεν βελτιώνει την απόδοση
- Την καλύτερη απόδοση την είχε το μοντέλο Gru με απόδοση 65% στην αναπαράσταση πίνακα.

Από την ανάλυση κατέληξαν στο συμπέρασμα ότι ο συνδυασμός των δύο κατηγοριών χαρακτηριστικών μπορεί να οδηγήσει στην καλύτερη απόδοση. Και η χρήση μεθόδων βαθιάς μάθησης (deep learning) δεν αύξησε ιδιαίτερα την απόδοση των κλασικών αλγορίθμων.

Β'.12 Κείμενο Dimitris Spathis, Panayiotis Vlamos [10]

Η μελέτη αυτή εξετάζει τους παράγοντες που χαρακτηρίζουν τη διάγνωση του άσθματος και της ΧΑΠ, με τεχνικές μηχανικής μάθησης. Στόχος της μελέτης αυτής είναι να γίνει ενσωμάτωση της σε συστήματα υποστήριξης κλινικών αποφάσεων (CDSS) για την διαχείριση ασθενών με άσθμα και ΧΑΠ . Στην έρευνα αυτή χρησιμοποιήθηκαν δεδομένα από επισκέψεις ασθενών από το ιατρείο ενός πνευμονολόγου στην χρονική περίοδο 2014-2015. Το δείγμα τους περιείχε στοιχεία από 132 ασθενείς κάθε ένα με 20 χαρακτηριστικά σχετικά με το δημογραφικό προφίλ, τις ιατρικές και ειδικές μετρήσεις των πνευμόνων , τις συνήθειες και τα συμπτώματα του ασθενή και τις εξαρτημένες μεταβλητές (Αν κάποιος πάσχει από ΧΑΠ ή Άσθμα). Για κάθε πάθηση πραγματοποιήθηκε διαφορετικό πείραμα. Μετά τον κατάλληλο καθαρισμό και ανωνυμοποίηση των δεδομένων το δείγμα δοκιμάστηκε με το λογισμικό στατιστικής ανάλυσης SPSS με την χρήση της λογιστικής παλινδρόμησης (Logistic Regression)έγινε μια πρώτη πρόβλεψη των εξαρτημένων μεταβλητών . Στην συνέχεια με την χρήση των τεχνικών cross validation (5 και 10 στρώσεις), δοκιμάστηκαν διαφορετικοί αλγόριθμοι μηχανικής μάθησης της βιβλιοθήκης Sklearn της γλώσσας Python. Οι αλγόριθμοι είναι οι παρακάτω:

- Naive Bayes
- Logistic Regression
- Neural Networks
- SVMs
- K-Nearest Neighbour
- Decision Trees
- Random Forest

Από την στατιστική ανάλυση ανακάλυψαν ότι τα περισσότερα στοιχεία ήταν από γυναίκες (65%) και το μεγαλύτερο ποσοστό ασθενών είναι υγιείς (45% είχαν άσθμα και 27% είχαν ΧΑΠ). Επιπλέον, παρατήρησαν ότι οι ασθενείς με ΧΑΠ οι οποίοι συνήθως είναι καπνιστές, εμφανίζουν χαμηλά αποτελέσματα σπιρομέτρησης, έχουν συχνό βήχα είναι άτομα μεγάλων ηλικιακών ομάδων. Από την ανάλυση των μοντέλων μηχανικής μάθησης πήραμε τα αποτελέσματα του παρακάτω πίνακα για την ΧΑΠ.

Classifiers	10-fold Precision	20% test Precision	20% test FI	5-fold Precision	20% test Precision	20% test FI
Naïve Bayes	96.9%	97%	96%	96.9%	97%	96%
Logistic Regression	96.2%	97%	96%	95.4%	97%	96%
Neural Network	81.8%	88%	83%	82.5%	89%	89%
SVM	93.9%	97%	96%	92.5%	97%	96%
K-Nearest Neighbour	76.5%	50%	58%	75.7%	63%	64%
Decision Tree	96.2%	97%	96%	94.6%	100%	100%
Random Forest	97.7%	100%	100%	96.9%	100%	100%

Πίνακας 6: Πίνακας σύγκρισης μετρικής Precision και Fi

Σύμφωνα με το παρακάτω δέντρο στην περίπτωση της ΧΑΠ, ο αλγόριθμος Random Forest είναι αυτό που επιτυγχάνει την καλύτερη απόδοση-precision(97.7%) με τον διαχωρισμό 10-fold cross validation, ακολουθούμενος από τον Naïve Bayes(96.9) και έπειτα Logistic Regression(96.2) και Decision Tree (96.2). Τέλος, αριθμήσαν τα χαρακτηριστικά με βάση την αξία τους στον καθορισμό της κλάσης με 2 μεθόδους, συσχετίσεις (Corelations) με την χρήση του δείκτη Pearson και απόκτηση πληροφοριών (Informations Gain). Στον παρακάτω πίνακα φαίνεται η κατάταξη των χαρακτηριστικών, για την ταξινόμηση των ασθενών με ΧΑΠ , ανάλογα με την αξία σύμφωνα με κάθε μέθοδο.

Rank	InfoGain	Corelation
1	Smoker	Smoker
2	FEV1	FEV1
3	Age	FVC
4	FVC	MEF50
5	Cough	AGE
6	Breath Shortness	MEF25
7	MEF50	MEF2575
8	Wheeze	PEF
9	MEF25	MEF75
10	Sex	Cough

Πίνακας 7: Πίνακας σύγκρισης χαρακτηριστικών

Όπως φάνηκε και στις δυο μεθόδους σημαντικό ρόλο για την ταξινόμηση των ασθενών με ΧΑΠ έχει αν ο ασθενής είναι καπνιστής αυτή την περίοδο(smoker), και τα αποτελέσματα σπιρομέτρησης FEV1 και FVC.

Γ'. Δεδομένα Μελέτης

Γ'.1 Αρχικά δεδομένα

Για την εκπαίδευση και δοκιμή του μοντέλου μηχανικής μάθησης της εργασίας μου χρειαζόμουν δείγματα από ασθενείς με ΧΑΠ προκειμένου το μοντέλο να δώσει αντικειμενικώς ορθά αποτελέσματα. Τα δεδομένα της εργασίας μου προήλθαν από το Kaggle.com. Τα δεδομένα αυτά ήταν από ασθενείς, οι οποίοι είχαν προσκληθεί να λάβουν μέρος σε ένα πρόγραμμα αποκατάστασης για να βοηθήσουν στη διαχείριση της ΧΑΠ τους. Συνολικά, το dataset περιελάμβανε στοιχεία 101 διαγνωσμένων ασθενών με ΧΑΠ, και 24 μεταβλητές. Σε αυτό το dataset υπάρχουν πληροφορίες σχετικά με τα χαρακτηριστικά των ασθενών, την σοβαρότητα της νόσου και τις συννοσηρότητες. Παράλληλα, περιέχει επίσης μετρήσεις της ικανότητας περπατήματος, της ποιότητας ζωής, και του άγχους και κατάθλιψης.

Γ'2. Επεξεργασία Δεδομένων

Τα δεδομένα αυτά προκειμένου να είναι κατάλληλα για την εκπαίδευση του συστήματος πρέπει να υποστούν την κατάλληλη επεξεργασία. Αν η επεξεργασία αυτή δεν πραγματοποιηθεί τα δεδομένα πιθανόν να δημιουργήσουν θόρυβο στο μοντέλο, σε περίπτωση που υπάρχουν ακραίες τιμές ή μη πραγματικές τιμές σε κάποια στοιχεία του πληθυσμού. Η διαγραφή αυτών των στοιχείων αυτών είναι απαραίτητη ώστε το μοντέλο να μην "παρασύρεται" από τα στοιχεία αυτά και οδηγεί σε λάθος αποτέλεσμα.

Στην περίπτωση των χαρακτηριστικών και αυτά μπορούν να δημιουργήσουν πρόβλημα, στην συγκεκριμένη περίπτωση μπορεί ένα χαρακτηριστικό να αντιστοιχίζεται άμεσα με όλες τις τιμές της κλάσης (με αυτή την λογική δεν χρειαζόμαστε μοντέλο μηχανικής μάθησης και δεν χρειαζόμαστε τα υπόλοιπα χαρακτηριστικά για την ταξινόμηση). Τέλος, πολλά χαρακτηριστικά μπορεί να είναι μη απαραίτητα ή να δημιουργούν θόρυβο (όπως το ID ή ο αριθμός γραμμής του δείγματος). Αυτά τα χαρακτηριστικά είναι αναγκαίο να αφαιρεθούν, προκειμένου το μοντέλο μας να είναι λειτουργικό.

Στην περίπτωση της κλάσης, υπάρχει πιθανότητα πολλά δείγματα να αντιστοιχούν σε μια τιμή της κλάσης. Αυτό θα έχει σαν αποτέλεσμα όταν το μοντέλο θα ταξινομήσει νέα στοιχεία, θα αντιστοιχεί σχεδόν όλα τα στοιχεία σε αυτή την τιμή αγνοώντας τις υπόλοιπες τιμές της κλάσης. Όταν αυτό συμβαίνει λέμε πως το μοντέλο έχει bias προς την τιμή αυτή. Για να αποφύγουμε αυτό το πρόβλημα υπάρχουν δυο μέθοδοι εξισορρόπησης των αποτελεσμάτων.

Η πρώτη μέθοδος είναι η μέθοδος oversampling, με αυτήν προσθέτονται ήδη υπάρχοντα στοιχεία μέσα στο δείγμα τα οποία έχουν τις τιμές της κλάσης που υστερούν. Η εναλλακτική μέθοδος ονομάζεται undersampling, με αυτήν αφαιρούμε στοιχεία τα οποία έχουν την τιμή της κλάσης που εμφανίζεται τις περισσότερες φορές, ώστε να υπάρχουν περίπου ίσα στοιχεία με τις υπόλοιπες τιμές της κλάσης. Και οι δύο αυτοί μέθοδοι έχουν ορισμένες αδυναμίες αρχικά η μέθοδος oversampling

προσθέτουμε τα ήδη υπάρχοντα στοιχεία ξανά θόρυβο ενώ στην διαδικασία undersampling χάνουμε σημαντική πληροφορία από τον πληθυσμό μας.

Γ'.2.1 Data cleaning

Αφαίρεση δειγμάτων: Αρχικά 1 στοιχείο διαγράφηκε γιατί οι μεταβλητές MWT1,MWT2,MWT1Best δεν ήταν συμπληρωμένες(Missing Value). Στην συνέχεια διαγράφηκε άλλο ένα στοιχείο του οποίου το CAT score>40, κάτι το οποίο είναι λάθος στο συγκεκριμένο τεστ (Noisy Data). Τέλος διαγράφηκε, για τον ίδιο λόγο, ένα ακόμα στοιχείο του οποίου το HAD score του ήταν μεγαλύτερο του 42. Με την διαδικασία αυτή αφαιρέσαμε συνολικά 3 δείγματα από τα 101.

Γ'.2.2 Attribute Selection

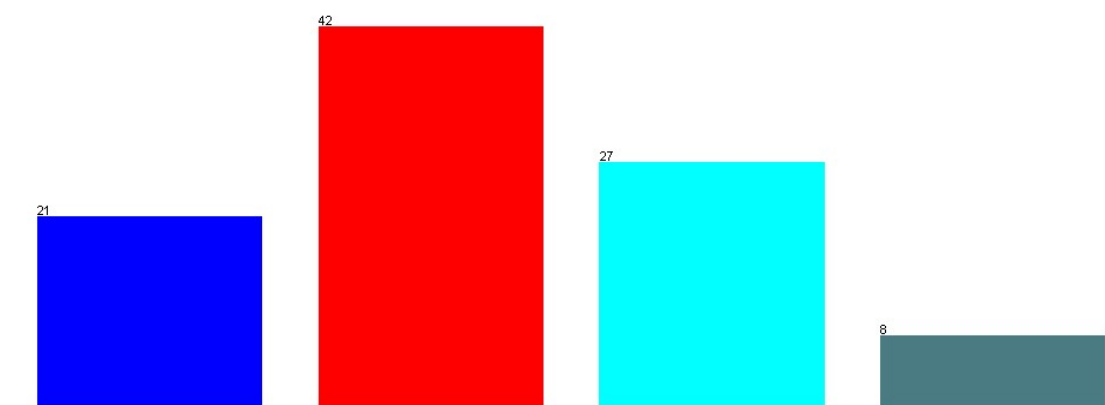
Αφαίρεση μεταβλητών: Οι μεταβλητές #(αριθμός στο dataset),ID(μοναδικό ID ασθενή) αφαιρέθηκαν γιατί ήταν μη απαραίτητες για την εργασία, και μπορούν αν προκαλέσουν θόρυβο γιατί οι τιμές τους δεν μπορούν σχετίζονται με την διάγνωση του ασθενή σε καμία περίπτωση. Παράλληλα οι μεταβλητές MWT1(απόσταση που διένυσε ο ασθενής, πρώτη προσπάθεια), MWT2(δεύτερη προσπάθεια) αφαιρέθηκαν ώστε να μην υπάρξει θόρυβος. Το MWT1Best(καλύτερη προσπάθεια) είναι πιο συγκεκριμένο χαρακτηριστικό και έχει τιμές και από τα δύο χαρακτηριστικά, οπότε μπορεί να τα εκπροσωπήσει.

Οι τιμές του COPD Severity (MILD, MODERATE, SEVERE, VERY SEVERE) αντιστοιχούν στις τιμές του χαρακτηριστικού copd (1,2,3,4). Επιπλέον αυτά τα χαρακτηριστικά πρέπει χρησιμοποιήσουμε ως κλάση. Οπότε το χαρακτηριστικό copd καθορίστηκε ως κλάση λόγω των στοιχείων του, και το χαρακτηριστικό COPD SEVERITY απορροφήθηκε. Παράλληλα, οι τιμές αυτών των δύο χαρακτηριστικών αντιστοιχούν στις τιμές που μπορεί να πάρει ο δείκτης GOLD για την διάγνωση της ΧΑΠ.

Από τις παραπάνω ενέργειες καταλήξαμε σε 98 δείγματα ασθενών, 18 μεταβλητές και μία κλάση η οποία εξαρτάτε από τις 18 μεταβλητές και αντιστοιχεί στον επίσημο όρο GOLD.

Μεταβλητές αυτές είναι:

- Age (Ηλικία του ασθενή)
- Pack History (Διεθνής όρος: υπολογίζεται ως αριθμός πακέτων την μέρα * τον χρόνο)
- MWT1Best (Απόσταση, σε μέτρα ,που διένυσε ο ασθενής σε 6 λεπτά, καλύτερη από τις 2 αποδώσεις)
- FEV1(Ποσότητα αέρα, σε λίτρα ,που ο ασθενής μπορεί να βγάλει από τους πνεύμονες βίαια)
- FEV1Pred(Πρόβλεψη του FEV1)
- FVC(Ποσότητα αέρα που ο ασθενής μπορεί να βγάλει με μια αναπνοή)
- FVCPRED(πρόβλεψη του FVC)
- CAT (Αποτέλεσμα του ερωτηματολογίου CAT, COPD Assessment Test)
- HAD (Αποτέλεσμα του ερωτηματολογίου HAD, Hospital Anxiety and Depression Scale)
- SGRQ (Αποτελέσματα του τεστ SGRQ, Saint George's Respiratory Questionnaire)
- AGEquartiles (Τεταρτημόριο ηλικίας, με τρεις άξονες 65, 70 και 75)
- Gender (Φύλο του ασθενή)
- Smoking (Κατάσταση καπνιστή,αν είναι αυτή την περίοδο καπνιστής)
- Muscular (Συννοσηρότητα με Μυοσκελετικά Προβλήματα)
- HID (Συννοσηρότητα Ισχαιμική Καρδιοπάθεια)
- AtrialFib (Συννοσηρότητα με Κολπική Μαρμαρυγή)
- Hypertension (Συννοσηρότητα με Αρτηριακή Υπέρταση)
- Diabetes (Συννοσηρότητα με Ζακχαρώδης Διαβήτης)



Εικόνα 7 : Αναπαράσταση των τιμών της κλάσης

Στην εικόνα αυτή φαίνεται πόσοι ασθενείς και σε ποιο βαθμό σοβαρότητας έχουν ΧΑΠ. Όπως φαίνεται το μεγαλύτερο ποσοστό(42 άτομα) έχει μέτρια συμπτώματα, ακολουθούν τα σοβαρά(27 άτομα) και τα ήπια(22 άτομα) συμπτώματα. Το μικρότερο ποσοστό είναι οι ασθενείς με πολύ σοβαρά συμπτώματα (8 άτομα).

Γ'.2.3 Εξισορρόπηση Δεδομένων

Παρατηρούμε πως οι κατηγορίες του δείγματος δεν έχουν μεγάλη διαφορά. Αυτό σημαίνει ότι πιθανόν στην εκπαίδευση οι προβλέψεις δεν θα τείνουν στην κατηγορία με τον μεγαλύτερο πληθυσμό(την κατηγορία για τα μέτρια συμπτώματα). Οπότε δεν χρειάζεται να κάνουμε undersampling την κατηγορία με τα μέτρια συμπτώματα και να χάσουμε πληροφορία, είτε να κάνουμε oversampling την κατηγορία με τα πολύ σοβαρά συμπτώματα(κατηγορία 4) και να προσθέσουμε στοιχεία που ήδη υπάρχουν δημιουργώντας θόρυβο.

Γ'.3 Επιλογή Χαρακτηριστικών (Feature Selection)

Η διαδικασία αυτή είναι απαραίτητη σε περίπτωση που έχουμε πολλά χαρακτηριστικά και θέλουμε να κρατήσουμε μόνο τα απαραίτητα για τον καθορισμό της κλάσης. Για την συγκεκριμένη διαδικασία υπάρχουν πολλοί τρόποι με τους οποίους ορισμένα δεδομένα επιλέγονται ως πιο σημαντικά από άλλα. Στο πρόβλημα ταξινόμησης που έχουμε η κλάση είναι κατηγορηματική μεταβλητή. Παράλληλα, διαθέτουμε χαρακτηριστικά που ανήκουν και στις δύο κατηγορίες δεδομένων (αριθμητικά και κατηγορηματικά). Προκειμένου να ελέγξουμε αν υπάρχει συσχέτιση μεταξύ ποιόν μεταβλητών και της κλάσης, και πόσο δυνατή είναι η συσχέτιση μεταξύ τους πρέπει να πραγματοποιήσουμε έναν έλεγχο συσχέτισης(correlation testing).

Αρχικά χωρίζουμε τα χαρακτηριστικά σε κατηγορικά και αριθμητικά ώστε να πραγματοποιηθεί ο έλεγχος ξεχωριστά. Πρέπει να συμπεριληφθεί ότι το πρόγραμμα που κατασκεύασα δημιουργεί δέντρο αποφάσεων(decision tree) για την ταξινόμηση των ασθενών. Αυτό σημαίνει ότι χρειαζόμαστε μεθόδους συσχέτισης για ταξινόμηση με μη γραμμικούς αλγόριθμους(Nonlinear).

Ο έλεγχος στα αριθμητικά δεδομένα έγινε με την χρήση της μεθόδου συσχέτισης Kendall Rank Correlation Coefficient. Για την συγκεκριμένη μέθοδο υπάρχουν δύο μέθοδοι

Γ'.3.1 Kendal Rank

Η πρώτη μέθοδος είναι η παραμετρική συσχέτιση Pearson(Pearson Corelation): Η μέθοδο αυτή μετρά μια γραμμική εξάρτηση μεταξύ δύο μεταβλητών (x και y). Η μέθοδο αυτή εξαρτάτε από την κατανομή των δεδομένων και ο δείκτης είναι γνωστός ως παραμετρική.

Η δεύτερη μέθοδος είναι η μη παραμετρική συσχέτιση με συντελεστές Kendall(tau) και Spearman(rho). Αυτοί είναι συντελεστές συσχέτισης βάσει κατάταξης και είναι γνωστοί ως μη παραμετρικοί συσχετισμοί.

Γνωρίζοντας την φύση του προβλήματος στο πρόγραμμα μου χρησιμοποίησα την δεύτερη μέθοδο και χρησιμοποίησα τον συντελεστή Kendall(tau).Ο τύπος για τον υπολογισμό του συντελεστής συσχέτισης (τ-tau) είναι ο παρακάτω

$$\tau = \frac{\text{Number of concordant pairs} - \text{Number of discordant pairs}}{n(n-1)/2}$$

Εικόνα 8 : Τύπος υπολογισμού δείκτη tau

Όπου :

Concordant pairs (Σύμφωνα ζεύγη) : είναι τα ζεύγη παρατηρήσεων όπου για τα (x1,y1) και (x2,y2) ισχύει:

- $x_1 > x_2$ και $y_1 > y_2$ ή $x_1 < x_2$ και $y_1 < y_2$

Discordant Pairs (Μη Σύμφωνα ζεύγη) : είναι τα ζεύγη παρατηρήσεων όπου για τα (x1, y1) και (x2,y2) ισχύει:

- $x_1 < x_2$ και $y_1 > y_2$ ή $x_1 > x_2$ και $y_1 < y_2$

Η μεταβλητή n είναι ο αριθμός των δειγμάτων

Στα ζεύγη που ισχύει $x_1=x_2$ και $y_1=y_2$ δεν υπολογίζονται

Εφόσον ο αλγόριθμος βρει τις τιμές των παραπάνω μεταβλητών υπολογίζει την τιμή του δείκτη tau. Ο δείκτης αυτός παίρνει τιμές από -1 μέχρι 1 .Όσο μεγαλύτερος είναι αυτός ο δείκτης από το 0 και πλησιάζει στο 1 τόσο μεγαλύτερη συμφωνία έχουν οι δύο μεταβλητές x και y και τόσο μεγαλύτερη συσχέτιση έχουν . Σχετικά με τις αρνητικές τιμές, όσο μικρότερη είναι η τιμή του δείκτη από το 0 και πλησιάζει στο -1 τόσο μεγαλύτερη διαφωνία έχουν οι δύο μεταβλητές, παρόλα αυτά αυτό δείχνει πόσο μεγάλη αρνητική συσχέτιση έχουν.

Στην δική μας περίπτωση γίνεται συσχέτιση μεταξύ των τιμών των μεταβλητών AGE , PackHistory ,MWT1Best, FEV1,FEV1PRED,FVC,FVCPRED,CAT,HAD και SGRQ με τις τιμές της κλάσης του copd. Παρόλο που οι τιμές της κλάσης είναι κατηγορικές μπορεί να γίνει σύγκριση διότι παίρνει τιμές από 1 έως 4 ανάλογα με τα συμπτώματα. Αυτό σημαίνει ότι σε περίπτωση θετικής συσχέτισης η υψηλές τιμές των δεδομένων αντιστοιχούν με υψηλές τιμές στην κλάση το αντίθετο θα ισχύει στην περίπτωση αρνητικής συσχέτισης.

Στο πρόγραμμα μου, για τον υπολογισμό του δείκτη tau, χρησιμοποίησα το αρχείο kendalltau από την βιβλιοθήκη scipy.stats. Με την χρήση του αρχείου για τον υπολογισμό υπολογίστηκε ο δείκτης καθώς και η σχέση τους με την κλάση για κάθε χαρακτηριστικό που αναφέρθηκε παραπάνω. Στον παρακάτω πίνακα φαίνονται τα χαρακτηριστικά και η τιμή του δείκτη tau για κάθε χαρακτηριστικό.

Feature	Tau
AGE	0.03238
PackYears	0.03274
MWT1Best	-0.31947
FEV1	-0.64922
FEV1PRED	-0.76895
FVC	-0.44556
FVCPRED	-0.50058
CAT	0.27825
HAD	0.21073
SGRQ	0.23903

Πίνακας 8 : Πίνακας χαρακτηριστικών και τιμής Tau

Στον παραπάνω πίνακα βλέπουμε τον βαθμό συσχέτισης μεταξύ των χαρακτηριστικών και της κλάσης. Παρατηρούμε πως τα αποτελέσματα της σπιρομέτρησης (FEV1, FEV1PRED, FVC , FVCPRED) είναι αυτά που με τον μεγαλύτερο βαθμό συσχέτισης και μάλιστα αρνητικό. Αυτό σημαίνει ότι οι υψηλές τιμές σε αυτά τα χαρακτηριστικά αντιστοιχούν συνήθως σε μέτρια και ήπια συμπτώματα της πάθησης ενώ, οι χαμηλές τιμές σε αυτά τα χαρακτηριστικά αντιστοιχούν σε σοβαρά συμπτώματα της πάθησης. Δεδομένο που ισχύει γιατί τα παραπάνω χαρακτηρίστηκα είναι αποτελέσματα ελέγχου πνευμονικής λειτουργίας, εξαιτίας της πάθησης αυτής η απόδοση των πνευμόνων μειώνεται. Οπότε από την διαδικασία αυτή δεχόμαστε τα χαρακτηριστικά FEV1, FEV1PRED, FVC και FVCPRED , τα όποια έχουν απόλυτη τιμή δείκτη μεγαλύτερη από 0.4 .

Γ'.3.2. Chi-Square

Τα υπόλοιπα χαρακτηριστικά στο dataset μας είναι κατηγορηματικά (categorical). Οπότε χρειαζόμαστε διαφορετική μέθοδο για να βρούμε τα καλύτερα χαρακτηριστικά για την ταξινόμηση των κλάσεων. Για αυτή την διαδικασία χρησιμοποιήθηκε η μέθοδος υπολογισμού συσχέτισης chi-square του Pearson . Η μέθοδος αυτή είναι ένα τεστ στατιστικής υπόθεσης για την ανεξαρτησία μεταξύ κατηγορικών μεταβλητών. Βασικό εργαλείο της μεθόδου αυτής είναι ένα πίνακας λεγόμενος Contingency table(ή crosstab), αυτός χρησιμοποιείται για να φανεί η σχέση μεταξύ 2 κατηγορικών χαρακτηριστικών. Στο δικό μας παράδειγμα θα μπορούσε να διερευνηθεί αν υπάρχει συσχέτιση μεταξύ της συννοσηρότητας με Σακχαρώδη διαβήτη και του βαθμού σοβαρότητας της πάθησης ΧΑΠ. Σε αυτή την περίπτωση θα είναι ο ακόλουθος πίνακας.

	Mild	Moderate	Severe	Very-Severe	Total
Diabetes	4	7	9	1	21
Without Diabetes	17	35	18	7	77
Total	21	42	27	8	98

Πίνακας 9 : Πίνακας πραγματικών τιμών chi-Square

Γ'.3.2.1 Μηδενική υπόθεση (Null Hypothesis)

Η μηδενική υπόθεση (H0) είναι μια γενική στατιστική δήλωση ή υπόθεση σχετικά με μια παράμετρο πληθυσμού που υποθέτουμε ότι είναι αληθής, μέχρι να βρούμε επαρκή στοιχεία για να την απορρίψουμε.

Γ'.3.2.2 Εναλλακτική υπόθεση (Alternate Hypothesis)

Η εναλλακτική υπόθεση (H1) είναι μια γενική στατιστική υπόθεση η οποία ανταγωνίζεται την μηδενική υπόθεση. Οπότε, αν αυτή η υπόθεση ισχύει μπορούμε να απορρίψουμε την μηδενική υπόθεση. Η παραπάνω πρόταση είναι ο στόχος της μεθόδου αυτής.

Γ'.3.2.3 Τιμή P

Η τιμή-P (P-value) χρησιμοποιείται ως μέτρο απόδειξης έναντι της υπόθεσης H0. Σε περίπτωση που αυτή είναι μεγαλύτερη από το επίπεδο σημαντικότητάς μας (level of significance- alpha) τότε αποδεχόμαστε την μηδενική υπόθεση. Σε αντίθετη περίπτωση θα αποδεχτούμε την μη μηδενική υπόθεση και η μέθοδος μας επιτυγχάνει τον σκοπό της.

Στα δικά μας δεδομένα, στο παράδειγμα συσχέτισης της συννοσηρότητας με Ζαχαρώδη διαβήτη και σοβαρότητα της ΧΑΠ, αρχίζουμε με την μηδενική υπόθεση H0 η οποία υποδηλώνει ότι τα δύο χαρακτηριστικά αυτά είναι ανεξάρτητα. Αντίθετα, η εναλλακτική υπόθεση H1 υποδηλώνει ότι τα δύο χαρακτηριστικά είναι εξαρτημένα. Στόχος της μεθόδου είναι να δούμε ποια από τις υποθέσεις ισχύει. Για να το ελέγξουμε πρέπει να ρυθμίσουμε την τιμή του επιπέδου σημαντικότητας (έστω $\alpha=0.05$). Σε περίπτωση που η τιμή του p είναι μεγαλύτερη αυτού, τότε η υπόθεση H0 ισχύει και αποδεικνύεται ότι τα χαρακτηριστικά είναι ανεξάρτητα, σε περίπτωση που η τιμή του p είναι μικρότερη ή ίση με την τιμή του alpha τότε η υπόθεση H0 απορρίπτεται και αποδεικνύεται ότι τα χαρακτηριστικά είναι εξαρτημένα. Στην συνέχεια φτιάχνουμε ένα παρόμοιο πίνακα υπολογισμένων (ή αναμενόμενων) τιμών σε περίπτωση που οι μεταβλητές αυτές ήταν πράγματι ανεξάρτητες. Οι τιμές του πίνακα υπολογίζονται από τον παρακάτω τύπο και έχουν τα ακόλουθα αποτελέσματα που φαίνονται στον παρακάτω πίνακα.

$$\frac{\text{row total} * \text{column total}}{\text{grand total}}$$

Εικόνα 9 : Τύπος υπολογισμού αναμενόμενων τιμών

ακόλουθος πίνακας.

	Mild	Moderate	Severe	Very-Severe	Total
Diabetes	4.5	9	5.78	1.7	20.98
Without Diabetes	16.5	33	21.22	6.3	77.02
Total	21	42	27	8	98

Πίνακας 10 : Πίνακας αναμενόμενων τιμών chi-Square

Παρατηρείτε ότι οι τιμές του πίνακα αυτού είναι πολύ κοντά σε αυτές του πραγματικού. Θα μπορούσαμε να πούμε ότι φαίνεται με το μάτι ότι αυτές οι δύο μεταβλητές είναι ανεξάρτητες αλλά ας συνεχίσουμε το παράδειγμα.

Πίνακας Chi-Square

	Mild	Moderate	Severe	Very-Severe	Total
Diabetes	0.55	0.44	0.55	0.29	1.83
Without Diabetes	0.015	0.12	0.49	0.08	0.705
Total	0.565	0.56	1.04	0.37	2.535

Πίνακας 11 : Πίνακας chi-Square

Οι τιμές του πίνακα αυτός δημιουργούνται από τις τιμές των δυο προηγούμενων πινάκων βάση τον παρακάτω τύπο για κάθε τιμή που γίνεται σύγκριση.

$$\frac{(Observed_value - Calculated_value)^2}{Calculated_value}$$

Εικόνα 10 : Τύπος υπολογισμού τιμών πίνακα chi-Square

Η τιμή που υπολογίστηκε από τον πίνακα είναι η τιμή υπολογισμού του chi-square(calculated-value = 2.535). Τώρα χρειάζεται να βρούμε την κρίσιμη τιμή του chi-square. Αυτή μπορούμε να την λάβουμε από τον πίνακα διανομής του chi-square εφόσον γνωρίζουμε τους βαθμούς ελευθερίας των δεδομένων μας.

Οι βαθμοί ελευθερίας βασίζονται στον αριθμό των τιμών που οι μεταβλητές μας μπορούν να πάρουν ορίζονται ως: (αριθμός τιμών του x-1)*(αριθμός τιμών του y-1). Στο δικό μας παράδειγμα το x αντιστοιχεί στην συννοσηρότητα με διαβήτη , οπότε μπορεί να πάρει 2 τιμές και το y αντιστοιχεί στην κλάση μας που μπορεί να πάρει 4 τιμές. Οπότε, οι βαθμοί ελευθερίας είναι (4-1)*(2-1)=3 . Οπότε, σύμφωνα με τον πίνακα η κρίσιμη τιμή είναι critical-value=7.815.

Σε περίπτωση που η κρίσιμη τιμή (critical value) >= τιμή υπολογισμού(calculated value) τότε δεχόμαστε την H0. Σε αντίθετη περίπτωση απορρίπτουμε την υπόθεση H0 και δεχόμαστε την υπόθεση H1. Στο δικό μας παράδειγμα ισχύει η πρώτη περίπτωση, οπότε το χαρακτηριστικό που χρησιμοποιήσαμε ως x και η κλάση y είναι ανεξάρτητες μεταβλητές μεταξύ τους, Οπότε, το χαρακτηριστικό αυτό δεν επηρεάζει την κλάση μας.

Στο πρόγραμμα μου χρησιμοποιήθηκε το αρχείο chi2_contingency για τον υπολογισμό του p-value για κάθε ένα από τα χαρακτηριστικά AGEquartiles , gender, smoking , Diabetes, muscular, hypertension, AtrialFib, IHD, χρησιμοποιώντας $\alpha=0.05$ δίνοντας μας τα ακόλουθα αποτελέσματα.

Features	p-value	Accept H0
AGEquartiles	0.99	Accept
Gender	0.99	Accept
Smoking	1.0	Accept
Diabetes	0.98	Accept
Muscular	0.97	Accept
Hypertension	0.99	Accept
AtrialFib	0.99	Accept
IHD	0.94	Accept

Πίνακας 12 : Αποτελέσματα feature selection για κατηγορικά χαρακτηριστικά

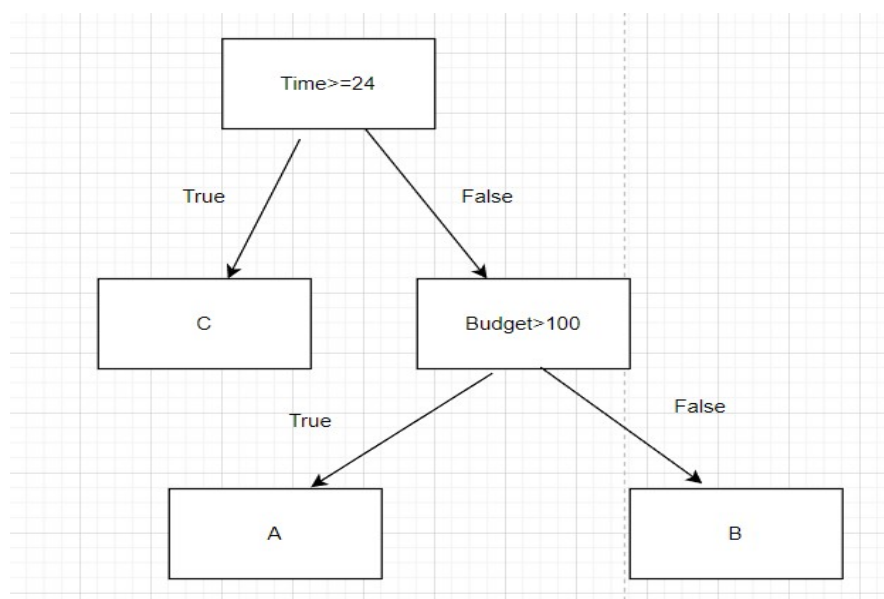
Όπως φαίνεται καμιά μεταβλητή δεν φάνηκε να έχει συσχέτιση με την κλάση. Αυτό σημαίνει ότι δεν θα χρησιμοποιήσουμε κανένα χαρακτηριστικό από τα παραπάνω για το feature selection.

Το feature selection με την χρήση δυο διαφορετικών μεθόδων μας έδωσε τα παρακάτω χαρακτηριστικά FEV1, FEV1PRED , FVC και FVCPRED. Παρατηρούμε ότι η ενέργεια αυτή μείωσε υπερβολικά τα χαρακτηριστικά των ασθενών. Αυτά τα χαρακτηριστικά δεν είναι αρκετά για την μελέτη μας. Οπότε, θα ήταν καλύτερο να μην ακολουθήσουμε αυτή την διαδικασία.

Δ'. Αλγόριθμοι Μηχανικής Μάθησης

Δ'.1 Αλγόριθμοι Δέντρων απόφασης(Decision Trees Algorithms)

Οι Αλγόριθμοι Δέντρων απόφασης χρησιμοποιούνται για την εξόρυξη δεδομένων. Οι αλγόριθμοι αυτοί χρησιμοποιούν τα δεδομένα του δείγματος που τους δόθηκαν για εκπαίδευση, βασισμένοι σε αυτά κατασκευάζουν ένα δέντρο απόφασης για την ταξινόμηση δειγμάτων. Το δέντρο αυτό μπορεί να χρησιμοποιηθεί για την ταξινόμηση νέων δειγμάτων. Τέτοιοι αλγόριθμοι είναι οι ID3, C4.5, C5.0 και CART. Τα δέντρα αποφάσεων είναι ένα ιεραρχικό μοντέλο υποστήριξης αποφάσεων που κατευθύνει τον χρήστη μέσω των κόμβων στα φύλα του(τελικοί). Η επιλογή μετάβασης από έναν κόμβο σε έναν άλλον κόμβο γίνεται με την αξιολόγηση των χαρακτηριστικών.



Εικόνα 11 : Δέντρο απόφασης του παραδείγματος

Όπως φαίνεται και στην εικόνα προκειμένου να μεταφερθούμε στους κόμβους πρέπει αρχικά να ελέγξουμε την τιμή των μεταβλητών Time και Budget στο δείγμα μας . Αρχίζοντας από την κορυφή του δέντρου(ρίζα) ελέγχουμε αν η τιμή Time είναι μεγαλύτερη ή ίση του 24. Σε περίπτωση που αυτό ισχύει θα μεταφερθούμε στον κόμβο με το γράμμα C, ο κόμβος αυτός ονομάζεται φύλο του δέντρου και είναι η απάντηση που το δέντρο αυτό μας δίνει. Εναλλακτικά, σε περίπτωση που η τιμή Time είναι μικρότερη του 24 θα μεταφερθούμε στον αριστερά κόμβο όπου θα πραγματοποιηθεί ακόμα ένας έλεγχος για την τιμή της μεταβλητής Budget αν αυτή είναι μεγαλύτερη του 100. Αν η σχέση αυτή ισχύει τότε θα μεταφερθούμε στον αριστερό κόμβο(παιδί) και θα λάβουμε απάντηση A σε αντίθετη περίπτωση θα μεταφερθούμε στον δεξί κόμβο και θα λάβουμε απάντηση B.

Δ'.2 Αλγόριθμος CART(Classification and Regression Trees)

Ο αλγόριθμος CART μπορεί να χρησιμοποιήσει δεδομένα και από τους δύο τύπους μεταβλητών (αριθμητικές και κατηγορικές), καθώς και να χειριστεί τόσο εργασίες

ταξινόμησης και όσο και παλινδρόμησης. Ο αλγόριθμος αυτός χρησιμοποιείται για την πρόβλεψη των τιμών της μεταβλητής στόχου (κλάσης) δημιουργώντας ένα δέντρο αποφάσεων. Κάθε διακλάδωση του χωρίζεται σε μια μεταβλητή πρόβλεψης. Οι μεταβλητές στόχοι βρίσκονται στο τέλος του δέντρου. Στο δέντρο αυτό οι κόμβοι χωρίζονται σε υποκόμβους με βάση μια τιμή κατωφλίου ενός χαρακτηριστικού. Ο κόμβος Ρίζα χωρίζεται σε δύο υποκόμβους λαμβάνοντας υπόψη το καλύτερο χαρακτηριστικό και την τιμή κατωφλίου. Με την ίδια λογική συνεχίζεται η διαδικασία αυτή στους υποκόμβους μέχρι να βρεθεί το τελευταίο καθαρό σύνολο κόμβων στο δέντρο ή μέχρι να φτάσουμε το μέγιστο αριθμό φύλων που εμείς έχουμε καθορίσει. Για την επιλογή των χαρακτηριστικών σύγκρισής στους κόμβους χρησιμοποιούμε τον δείκτη Gini. Ο δείκτης αυτός αποθηκεύει το άθροισμα των τετραγώνων πιθανοτήτων κάθε κλάσης, υπολογίζει τον βαθμό πιθανότητας μιας συγκεκριμένης μεταβλητής που ταξινομείτε λανθασμένα όταν επιλέγεται τυχαία.

Ο δείκτης αυτός παίρνει τιμές από 1 μέχρι 0

- Αν η τιμή του δείκτη είναι κοντά στο 0 φαίνεται ότι τα περισσότερα στοιχεία συνδέονται σε μια συγκεκριμένη κλάση, σε περίπτωση που είναι 0 σε κάποιο κόμβο ο κόμβος αυτός είναι η κλάση.
- Αν η τιμή του δείκτη είναι 1 ,ή κοντά σε αυτό, σημαίνει ότι όλα τα στοιχεία κατανέμονται τυχαία σε μεταξύ πολλών κλάσεων.
- Αν η τιμή είναι 0.5 τα στοιχεία κατανέμονται ομοιόμορφα σε ορισμένες κλάσεις.

Η τιμή του δείκτη Gini υπολογίζεται από τον παρακάτω τύπο όπου p_i είναι η πιθανότητα ταξινόμησης ενός αντικειμένου σε μια κλάση.

$$Gini = 1 - \sum_{i=1}^n (p_i)^2$$

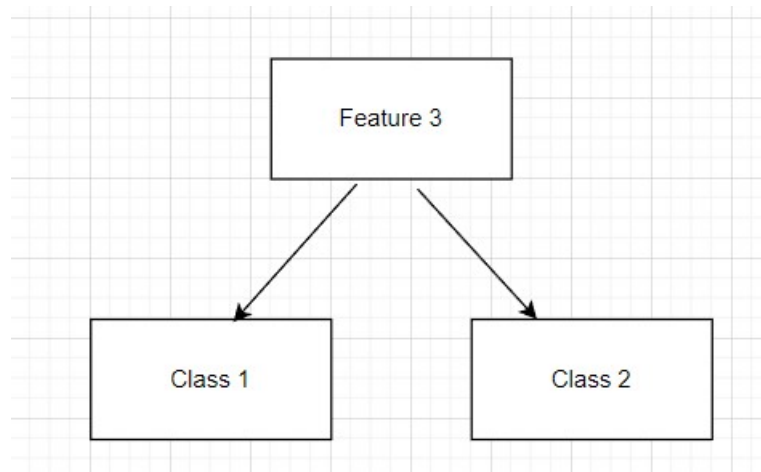
Εικόνα 12 : Τύπος υπολογισμού Gini

Στη δική μου εργασία για την δημιουργία του δέντρου χρησιμοποιήθηκε ο DecisionTreeClassifier από την βιβλιοθήκη Sklearn της Python.

Δ'3 Αλγόριθμος Adaboost

Ο αλγόριθμος Adaboost είναι ένας αλγόριθμος μηχανικής μάθησης συνόλου για προβλήματα ταξινόμησης. Είναι ένας αλγόριθμος που ανήκει στην κατηγορία boosting, ο οποίος προσθέτει νέα μοντέλα μηχανικής μάθησης, ώστε να διορθώσει τα λάθη των προηγούμενων μοντέλων. Ο αλγόριθμος αυτός σχεδιάστηκε για να

χρησιμοποιεί σύντομα μοντέλα αποφάσεων το καθένα με ένα μόνο σημείο απόφασης. Συγκεκριμένα, δέντρα που αποτελούνται μόνο από την ρίζα και 2 φύλλα [εικόνα].



Εικόνα 13 : Tree-Stump

Το πρώτο μοντέλο του αλγόριθμου κατασκευάζεται έχοντας όλες οι οντότητες το ίδιο βάρος και στις συνέχεια ελέγχεται αν ποια στοιχεία ταξινομήθηκαν σωστά. Σε περίπτωση λάθους ο αλγόριθμος αυξάνει το βάρος των στοιχείων που ταξινομήθηκαν λάθος, και φτιάχνει καινούριο δείγμα έχοντας περισσότερα στοιχεία από αυτά που ταξινομήθηκαν λάθος. Κάθε δέντρο που δημιουργείτε, σταθμίζεται με βάση τις δεξιότητές του στην πρόβλεψη όλων των στοιχείων των δειγμάτων. Το δέντρο που θα έχει την καλύτερη απόδοση θα επιλεγεί για την ταξινόμηση νέων χαρακτηριστικών .

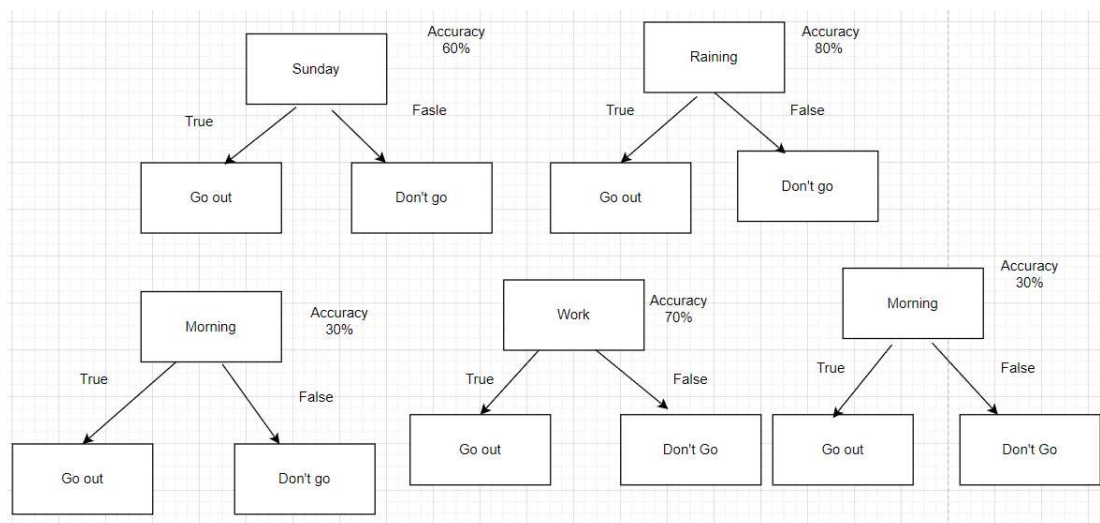
Δ'3.1 Παράδειγμα αλγορίθμου

Έστω ότι θέλουμε να φτιάξουμε ένα σύστημα υποστήριξης απόφασης για να καθορίσουμε αν θα βγούμε έξω από το σπίτι για περπάτημα με την χρήση του Adaboost και την χρήση Decision Stumps . Έστω ότι όλα τα παρακάτω χαρακτηριστικά που έχουμε είναι δυαδικά (έχουν τιμές μόνο ναι ή όχι).

- Sunday
- Raining
- Work
- Attitude
- Morning

Έχοντας βάλει σε όλα τα δείγματα το ίδιο βάρος για τον καθορισμό της κλάσης προκύπτει το πρώτο δέντρο. Το δείγμα χρησιμοποιείται για testing. Τα στοιχεία που ταξινομήθηκαν λανθασμένα σημειώνονται και στην συνέχεια, το βάρος τους μεγαλώνει. Στην συνέχεια, φτιάχνεται νέο δείγμα ίδιου μεγέθους με το προηγούμενο, μόνο που αυτή την φορά θα περιέχει περισσότερα στοιχεία από αυτά που ταξινομήθηκαν λάθος στο προηγούμενο testing. Η παραπάνω διαδικασία θα επαναληφθεί αρκετές φορές και θα φτιαχτούν καινούρια δέντρα. Στην συνέχεια όλα τα δέντρα θα δοκιμαστούν με όλα τα δείγματα που συμμετείχαν σε αυτή την

διαδικασία και θα υπολογιστεί η απόδοση που έχουν αυτά τα δέντρα. Το δέντρο που θα έχει την καλύτερη απόδοση θα είναι αυτό που θα επιλεγεί από τον αλγόριθμο.



Εικόνα 14 : Παράδειγμα δέντρων Adaboost

Στην παραπάνω εικόνα φαίνονται τα δέντρα που δημιουργήθηκαν από τον αλγόριθμο, και η απόδοσή τους στο σύνολο. Από αυτά θα επιλεγεί το Decision Stump που ελέγχει το χαρακτηριστικό Raining. Στην δική μου εργασία στην θέση των Decision Stump θα είναι CART Decision Trees, αυτό σημαίνει ότι θα έχουν περισσότερα επίπεδα και περισσότερα φύλλα.

Δ'4 Εκπαίδευση συστήματος

Για τον διαχωρισμό των δειγμάτων σε εκπαίδευση και δοκιμή χρησιμοποιήθηκε η μέθοδος stratified 10 fold cross –validation. Με την μέθοδο αυτή το 10% του πληθυσμού χρησιμοποιείται ως testing και το υπόλοιπο 90% ως training. Η μέθοδος αυτή επαναλαμβάνεται μέχρι να χρησιμοποιηθούν όλα τα δείγματα του πληθυσμού ως testing, στην περίπτωση μας 10 φορές. Σε κάθε επανάληψη επιλέγονται τυχαία δείγματα για testing. Τα δείγματα που έχουν επιλεγεί σε μια επανάληψη δεν μπορούν να επιλεγθούν σε επόμενες επαναλήψεις.

Δ'5 Εργαλεία του προγράμματος

Το πρόγραμμα μου έγινε με την χρήση της γλώσσας προγραμματισμού python. Η γλώσσα αυτή επιλέχθηκε λόγω της απλότητας της και λόγω της ευελιξίας της σε πολλούς τομείς, τόσο στην χρήση της στον τομέα της μηχανικής μάθησης όσο και στην επικοινωνία της με γλώσσες που χρησιμοποιούνται στην κατασκευή ιστοσελίδων όπως php και javascript. Η βασική βιβλιοθήκη της python που χρησιμοποιήθηκε ήταν η Sklearn για την δημιουργία των δέντρων και την εκπαίδευση των μοντέλων. Παράλληλα, οι βιβλιοθήκες pandas και numpy βοήθησαν στην διαχείριση των δειγμάτων ώστε να χρησιμοποιηθούν στην γλώσσα python και να ταιριάζουν στις απαιτήσεις των αλγορίθμων. Από την βιβλιοθήκη Sklearn χρησιμοποίησα τα αρχεία DecisionTreeClassifier για τον αλγόριθμο CART και AdaboostClassifier για τον αλγόριθμο Adaboost. Οι δύο αυτοί αλγόριθμοι είναι αυτοί που χάρις αυτούς

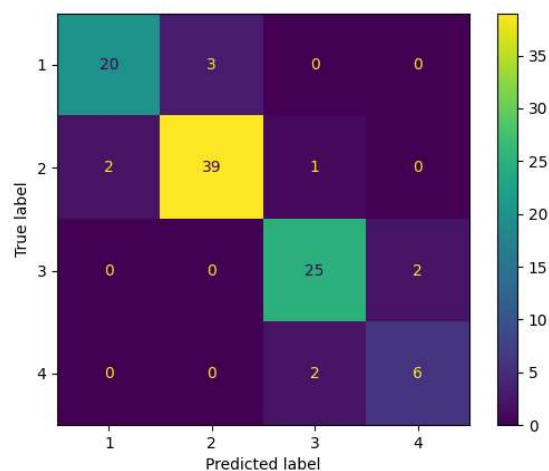
κατασκευάστηκαν τα δέντρα αποφάσεων. Τέλος από την βιβλιοθήκη Sklearn χρησιμοποιήθηκε η συνάρτηση StratifiedKfold για τον διαχωρισμό των δειγμάτων σε training και testing στην διαδικασία της εκπαίδευσης του μοντέλου.

Ε'. Αποτελέσματα

Μετά την επεξεργασία των δεδομένων πραγματοποιήθηκε η εκπαίδευση του μοντέλου. Το μοντέλο μας εκπαιδευτικό χρησιμοποιώντας τα παραπάνω δεδομένα ως training και testing χρησιμοποιώντας τον αλγόριθμο ταξινόμησης AdaBoost Classifier. Για τον διαχωρισμό του πληθυσμού χρησιμοποιήθηκε η μέθοδος 10-fold cross validation, χωρίζοντας έτσι σε κάθε επανάληψη τον πληθυσμό σε 90% ποσοστό εκπαίδευσης και 10% ποσοστό δοκιμής. Η διαδικασία αυτή έγινε χρησιμοποιώντας και τα 18 δεδομένα για τον καθορισμό της κλάσης. Τα αποτελέσματα του προγράμματος σε python της εργασίας συγκρίθηκαν με τα αποτελέσματα του weka χρησιμοποιώντας τον ίδιο αλγόριθμο.

Ε'.1 Πίνακας Σύγχυσης (Confusion Matrix)

Με την χρήση της γλώσσας python και των βιβλιοθηκών numpy και sklearn δημιουργήθηκε ο παρακάτω πίνακας σύγχυσης.



Εικόνα 15: Confusion Matrix Python

Με την χρήση του μοντέλου AdaBoost και την χρήση του 10-fold cross validation στο Weka πήραμε τον παρακάτω πίνακα σύγχυσης

Ταξινομήθηκε ως →	Ήπια συμπτώματα	Μέτρια συμπτώματα	Σοβαρά συμπτώματα	Πολύ-Σοβαρά συμπτώματα
Ήπια συμπτώματα	18	3	0	0
Μέτρια συμπτώματα	2	39	1	0
Σοβαρά συμπτώματα	0	0	26	1
Σοβαρά συμπτώματα	0	0	2	6

Πίνακας 13: Confusion Matrix Weka

Παρατηρούμε ότι οι δύο πίνακες μοιάζουν μεταξύ τους. Αυτό σημαίνει ότι ο κώδικας είναι κατάλληλος για την εκπαίδευση μοντέλου μηχανικής μάθησης.

Ε.2 Μετρικές

Στον παρακάτω πίνακα φαίνονται οι μετρικές του μοντέλου σύμφωνα με το weka. Θα εστιάσουμε στην μελέτη των μετρικών TP Rate, FP Rate και Precision και Recall.

Κλάση	TP RATE/Recall	FP RATE	PRECISION
1	0.864	0.013	0.95
2	0.952	0.053	0.93
3	0.963	0.042	0.897
4	0.75	0.011	0.857
Μέσος όρος	0.919	0.037	0.919

Πίνακας 14 : Μετρικές Weka

Στις παρακάτω εικόνες φαίνονται οι τιμές των μετρικών από το πρόγραμμα μου σε python.

Κλάση	TP RATE/Recall	FP RATE	PRECISION
1	0.863	0.013	0.95
2	0.928	0.053	0.93
3	0.925	0.042	0.897
4	0.75	0.011	0.857
Μέσος όρος	0.8665	0.02975	0.9085

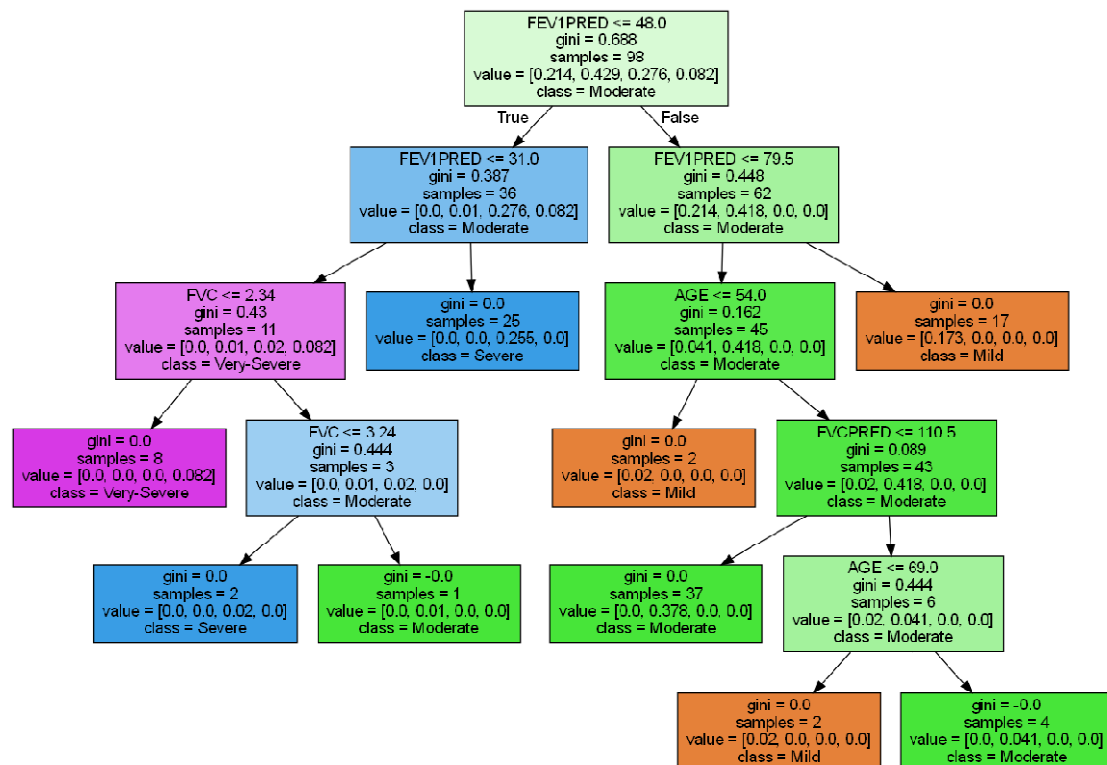
Πίνακας 15: Μετρικές Python

Όπως φαίνεται οι μετρικές που υπολόγισε το πρόγραμμα είναι σχεδόν πολύ κοντά με αυτές του Weka. Η μόνη εξαίρεση είναι η μετρική Precision για το 4 η οποία συγχωρείτε λόγο του μικρού πληθυσμού της κλάσης 4. Η απόδοση του συστήματος στο πρόγραμμα Python (90%) είναι λίγο μικρότερη από την απόδοση του μοντέλου στο Weka(91%).

Ε'3 Δέντρα Αποφάσεων CART

Ο αλγόριθμος AdaBoost χρησιμοποιεί σαν βάση του τον αλγόριθμο CART για να δημιουργήσει δυαδικά δέντρα αποφάσεων. Στο δικό μας πείραμα χρησιμοποιήσαμε τον αλγόριθμο CART για την δημιουργία των παρακάτω δέντρων. Στην συνέχεια ο αλγόριθμος Adaboost εφόσον φτιάξει τα δέντρα επιλέγει το βέλτιστο δέντρο το οποίο θα εκποσωπίσει τον αλγόριθμο για την ταξινόμηση νέων δειγμάτων .

Στην παρακάτω εικόνα φαίνεται το δέντρο το οποίο ο αλγόριθμος επέλεξε ως βέλτιστο.



Εικόνα 16: Βέλτιστο δέντρο σύμφωνα με τον Adaboost

Στ' Περιγραφή Εφαρμογής

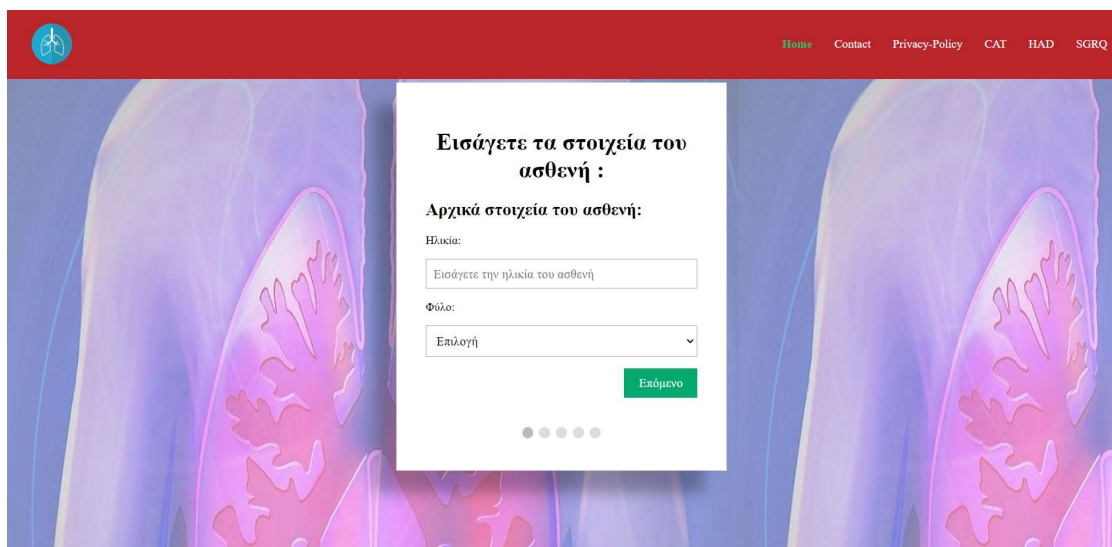
Σκοπός της εργασίας μου ήταν η δημιουργία εφαρμογής για την διάγνωση ατόμων με χρόνια αποφρακτική πνευμονοπάθεια. Προκειμένου το μοντέλο που φτιάχτηκε από το πρόγραμμα και επιλέχθηκε να εφαρμοστεί και να χρησιμοποιηθεί από άλλους χρήστες θα πρέπει να έχει και την κατάλληλη λειτουργικότητα και απεικόνιση, ώστε να ταιριάζει στις ανάγκες τους.

Στ'.1 Εργαλεία δημιουργίας

Ο σκελετός της εφαρμογής φτιάχτηκε με την χρήση κώδικα html και αποτελείτε από μια φόρμα την οποία οι χρήστες καλούνται να συμπληρώνουν. Η φόρμα αυτή ελέγχεται για την ορθότητά της με την χρήση συναρτήσεων σε γλώσσα javascript. Τα δεδομένα της φόρμας εφόσον έχουν συμπληρωθεί και ελεγχθεί για την ορθότητά τους μεταφέρονται μέσω της γλώσσας php σε ένα αρχείο python ώστε να πραγματοποιηθεί η διαδικασία ταξινόμησης των στοιχείων του και να λάβουμε απάντηση για την κατάσταση του. Τέλος, προκειμένου η εφαρμογή να είναι καλύτερη οπτικά στον χρήστη, χρησιμοποιήθηκε κώδικας CSS για την προσαρμογή των στοιχείων της html.

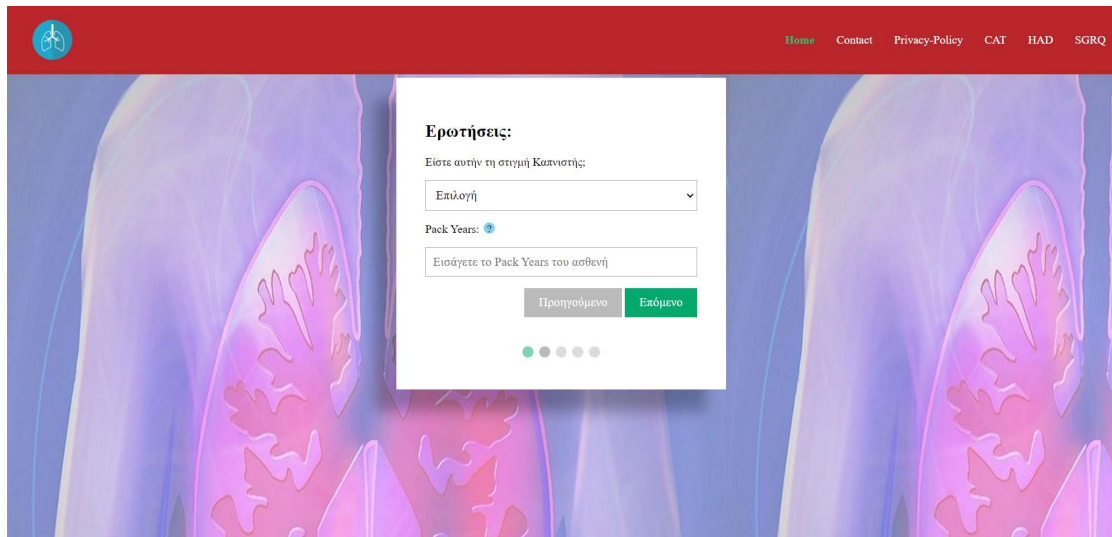
Στ'.2 Αρχική Σελίδα

Στην παρακάτω εικόνα φαίνεται η αρχική σελίδα της εφαρμογής. Η σελίδα αυτή αποτελείτε από μια φόρμα την οποία ο χρήστης συμπληρώνει και ένα μενού που οδηγεί στις υπόλοιπες σελίδες. Η φόρμα αυτή έχει 5 επίπεδα , το καθένα με το δικό του θέμα. Τα στοιχεία που ζητούνται από τον χρήστη σχετίζονται άμεσα με το μοντέλο που έχουμε αναπτύξει . Το πρώτο επίπεδο σχετίζεται με τα αρχικά εξωτερικά στοιχεία του ασθενή όπως φαίνεται και στην παρακάτω εικόνα. Τα στοιχεία αυτά είναι το φύλο και η ηλικία του.



Εικόνα 17 : Αρχικά στοιχεία ασθενή

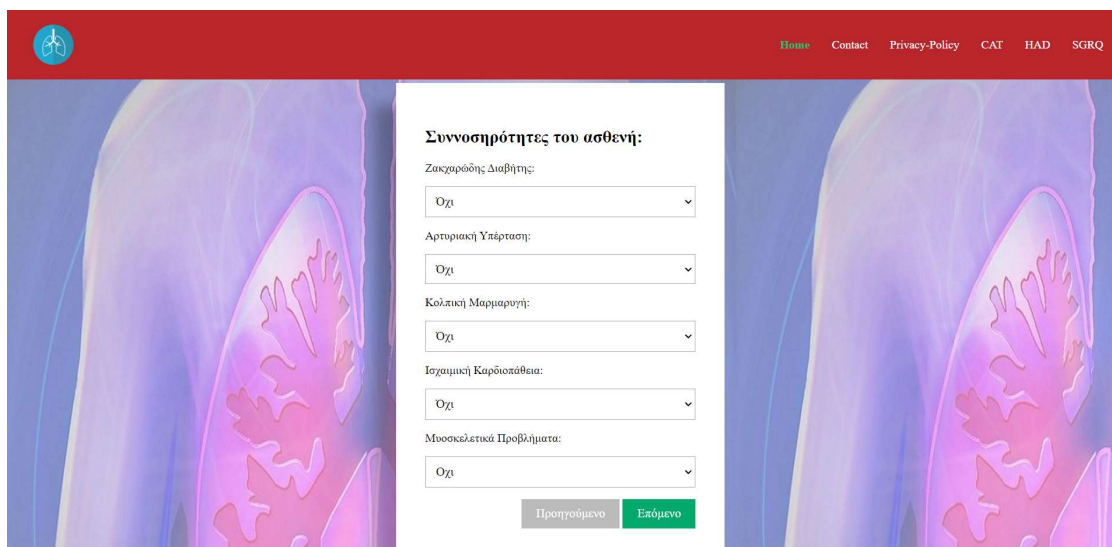
Το δεύτερο επίπεδο σχετίζεται με ερωτήσεις που κάνει ο πνευμονολόγος προκειμένου να μάθει την κατάσταση του ασθενή σχετικά με το κάπνισμα. Τα στοιχεία που ο χρήστης οφείλει να επιλέξει είναι αν ο ασθενής είναι ακόμα καπνιστής και τα Pack Years του ασθενή. Ο όρος Pack Years είναι το γινόμενο των χρόνων που είναι ο ασθενής καπνιστής * τον αριθμό πακέτων τσιγάρων που καπνίζει την ημέρα.



Εικόνα 18 : Ερωτήσεις για το ιστορικό του ασθενή

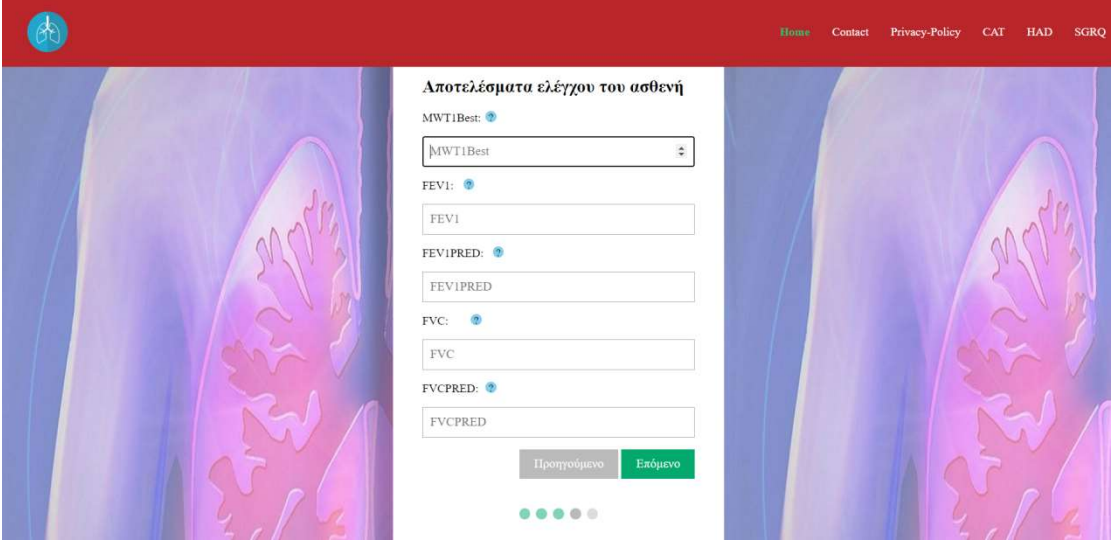
Το τρίτο επίπεδο σχετίζεται με τις συννοσηρότητες που πιθανόν να έχει ο ασθενής οι οποίες επηρεάζουν την κατάσταση της ΧΑΠ. Οι συννοσηρότητες αυτές είναι

- Ζαχαρώδης Διαβήτης
- Αρτηριακή Υπέρταση
- Κολπική Μαρμαρυγή
- Ισχαιμική Καρδιοπάθεια
- Διάφορα μυοσκελετικά προβλήματα



Εικόνα 19 : Συννοσηρότητες ασθενή

Το τέταρτο επίπεδο σχετίζεται με τα αποτελέσματα ελέγχου του ασθενή. Συγκεκριμένα, αποτελέσματα του ελέγχου MWT1. Σε αυτόν τον έλεγχο ο ασθενής πρέπει να περπατήσει για 6 λεπτά. Για τον έλεγχο αυτόν γίνονται δύο προσπάθειες, MWT1Best είναι η καλύτερη προσπάθεια από τις δύο. Τα υπόλοιπα αποτελέσματα είναι αποτελέσματα σπυρομέτρησης για την μέτρηση της πνευμονικής λειτουργίας. Το Fev1 είναι η ποσότητα αέρα που ο ασθενής μπορεί να βγάλει από τους πνεύμονές του σε ένα δευτερόλεπτο, μετριέται σε λίτρα. Το FVC είναι ο συνολικός όγκος αέρα που ο ασθενής μπορεί να αποβάλει με μια αναπνοή. Τα Fev1Pred και FVCPred είναι οι προβλέψεις των Fev1 και FVC αντίστοιχα.



Εικόνα 20 : Αποτελέσματα ελέγχου του ασθενή

Το τελευταίο επίπεδο σχετίζεται με τα αποτελέσματα ερωτηματολογίων σχετικά με την ΧΑΠ. Τα ερωτηματολόγια αυτά είναι :

- COPD Assessment Test(CAT), το οποίο χρησιμοποιείται από διαγνωσμένους ασθενείς με ΧΑΠ για την αξιολόγηση της κατάστασης της εξέλιξης της πνευμονικής νόσου
- Hospital Anxiety and Depression Scale (HADS), το οποίο σχετίζεται με την κατάσταση του άγχους και της κατάθλιψης, τις τελευταίες ημέρες.
- St. George's Respiratory Questionnaire (SGRQ), το οποίο χρησιμοποιείται για την μέτρηση του αντίκτυπου που έχουν διάφορες πνευμονικές παθήσεις στην ζωή των ασθενών.

Εικόνα 21 : Αποτελέσματα Ερωτηματολογίων

Κάθε στοιχείο που ζητάτε από τον χρήστη και δεν είναι ξεκάθαρο σε κάποιον που δεν έχει γνώσεις στον τομέα της πνευμονικής ιατρικής, διότι το στοιχείο αυτό μπορεί να είναι αποτελέσματα ελέγχων ή ερωτηματολογίων (Όπως FEV1 και CAT) ή διεθνείς όροι όπως Packyears . Ο χρήστης μπορεί να δει τι είναι κάθε στοιχείο του δείγματος και να μάθει πως υπολογίζεται.

Στην παρακάτω εικόνα φαίνεται η εξήγηση του Packyears.

Εικόνα 22 : Εξήγηση Pack Years

Στην συνέχεια φαίνονται οι εξηγήσεις των αποτελεσμάτων ελέγχου πνευμονικής λειτουργίας (MWT1Best, και τα αποτελέσματα σπироμέτρησης) του ασθενή.

Αποτελέσματα ελέγχου του ασθενή

MWT1Best: ?

M

Απόσταση που διανύει ο ασθενής σε 6 λεπτά σε μέτρα (καλύτερη προσπάθεια)

FEV1

FEV1

FEV1PRED: ?

FEV1PRED

FVC: ?

FVC

FVCPRED: ?

FVCPRED

Προηγούμενο

Επόμενο

Εικόνα 23: Εξήγηση MWT1Best

Αποτελέσματα ελέγχου του ασθενή

MWT1Best: ?

MWT1Best

FEV1: ?

Ποσότητα αέρα που ο ασθενής μπορεί να εξαναγκάσει από τους πνεύμονές του σε ένα δευτερόλεπτο σε λίτρα. Μέτρηση της πνευμονικής λειτουργίας,αποτέλεσμα σπιρομέτρησης

FVC: ?

FVC

FVCPRED: ?

FVCPRED

Προηγούμενο

Επόμενο

Εικόνα 24 : Εξήγηση FEV1

Αποτελέσματα ελέγχου του ασθενή

MWT1Best: ?

MWT1Best

FEV1: ?

FEV1

FEV1PRED: ?

FEV1

FVC: ?

FVC

FVC

FVCPRED: ?

FVCPRED

Προηγούμενο Επόμενο

● ● ● ● ●

Πρόβλεψη ποσότητας αέρα που ο ασθενής μπορεί να εξαναγκάσει από τους πνεύμονές του σε ένα δευτερόλεπτο σε λίτρα. Μέτρηση της πνευμονικής λειτουργίας,αποτέλεσμα σπιρομέτρησης

Εικόνα 25 : Εξήγηση FEV1PRED

Αποτελέσματα ελέγχου του ασθενή

MWT1Best: ?

MWT1Best

FEV1: ?

FEV1

FEV1PRED: ?

FEV1PRED

FVC: ?

FVC

FVC

FVCPRED: ?

FVCPRED

Προηγούμενο Επόμενο

● ● ● ● ●

Συνολικός όγκος αέρα που ο ασθενής μπορεί να εκπνεύσει με μία αναπνοή

Εικόνα 26 : Εξήγηση FVC

Αποτελέσματα ελέγχου του ασθενή

MWT1Best: ?

MWT1Best

FEV1: ?

FEV1

FEV1PRED: ?

FEV1PRED

FVC: ?

FVC

FVCPRED: ?

FV

Πρόβλεψη Συνολικού όγκου αέρα που ο ασθενής μπορεί να εκπνεύσει με μία αναπνοή

● ● ● ● ●

Εικόνα 27 : Εξήγηση FVCPRED

Παρακάτω φαίνονται οι σύντομες εξηγήσεις των ερωτηματολογίων CAT, HAD και SGRQ.

Αποτελέσματα Ερωτηματολογίων :

Αποτέλεσμα CAT: ?

CAT

Αποτέλεσμα HAD

HAD

Αποτέλεσμα SGRQ: ?

SGRQ

Προηγούμενο Υποβολή

● ● ● ● ●

Εικόνα 28 : Εξήγηση CAT

Αποτελέσματα Ερωτηματολογίων :

Αποτέλεσμα CAT: ?

CAT

Αποτέλεσμα HAD: ?

HAD

Αποτέλεσμα SGRQ: ?

SGRQ

Προηγούμενο Υποβολή

● ● ● ● ●

Εικόνα 29 : Εξήγηση HAD

Αποτελέσματα Ερωτηματολογίων :

Αποτέλεσμα CAT: ?

CAT

Αποτέλεσμα HAD: ?

HAD

Αποτέλεσμα SGRQ: ?

SGRQ

Αποτέλεσμα Ερωτηματολογίου ST GEORGE'S RESPIRATORY QUESTIONNAIRE

● ● ● ● ●

Εικόνα 30 :Εξήγηση SGRQ

Στ'.3 Έλεγχος

Σε περίπτωση εισόδου μη επιτρεπτής τιμής σε κάποιο στοιχείο της φόρμας. Η φόρμα μέσω της javascript ημερώνει τον χρήστη για το λάθος. Στην συνέχεια το στοιχείο που εισήγαγε ο χρήστης μη επιτρεπτή τιμή γίνεται κόκκινο μέχρι να πάρει επιτρεπτή τιμή και ο χρήστης να ζητήσει να μεταφερθεί στο επόμενο επίπεδο ή να υποβάλει την αίτηση.

Αποτελέσματα Ερωτηματολογίων :

Αποτέλεσμα CAT: ?

23

Αποτέλεσμα HAD: ?

50

Αποτέλεσμα SGRQ: ?

25.5

Προηγούμενο Υποβολή

● ● ● ● ●

Εικόνα 31 : Είσοδος Λάθος Τιμής

Ο χρήστης δεν μπορεί να προχωρήσει στο επόμενο επίπεδο εάν υπάρχει τουλάχιστον 1 στοιχείο της φόρμας που δεν έχει συμπληρωθεί ή έχει συμπληρωθεί με μη επιτρεπτή τιμή. Το ίδιο ισχύει σε περίπτωση που ο χρήστης βρίσκεται στο τελευταίο επίπεδο και θέλει να υποβάλει την φόρμα.

Ερωτήσεις:

Είστε αυτήν τη στιγμή Καπνιστής;

Καπνιστής

Pack Years: ?

Εισάγετε το Pack Years του ασθενή

Προηγούμενο Επόμενο

Εικόνα 32 : Κενή Είσοδος

Στ'.4 Ερωτηματολόγιο CAT

Το ερωτηματολόγιο αυτό χρησιμοποιείται σε διαγνωσμένους ασθενείς με ΧΑΠ. Οι ασθενείς μπορούν να αξιολογήσουν τα προβλήματα που η νόσος τους προκαλεί (όπως βήχας και σφίξιμο στο στήθος) και να λάβουν ως αποτέλεσμα την κατάσταση σοβαρότητας της νόσου, κινδύνους που πρέπει να αποφύγουν, καθώς και μέτρα τα οποία απαιτείται να πάρουν ώστε η κατάστασή τους να καλυτερέψει.

COPD Assessment Test (CAT) :

Βήχας :

Δεν βήχο καθόλου

Φλέγματα

Δεν έχω καθόλου φλέγμα (βλέννα) στο στήθος

Σφύξιμο στο Στήθος

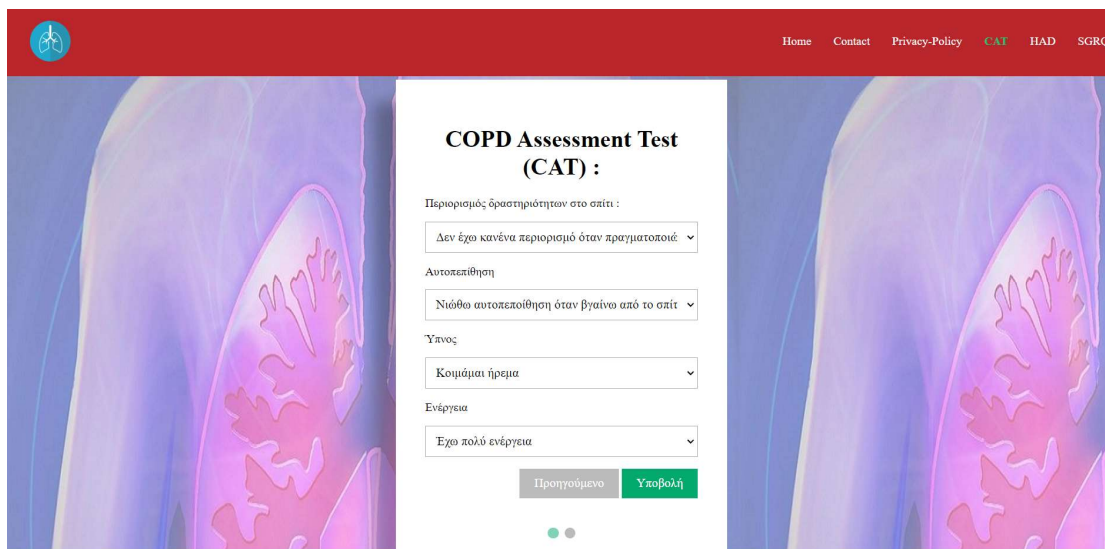
Δεν αισθάνομαι καθόλου σφίξιμο στο στήθος

Δυσπνοια

Δεν λαχανιάζω όταν περπατάω σε ανηφόρα ή ότι

Επόμενο

Εικόνα 33 : Ερωτηματολόγιο CAT



COPD Assessment Test (CAT) :

Περιορισμός δραστηριότητας στο σπίτι :

Δεν έχω κανένα περιορισμό όταν πραγματοποιώ

Αυτοπεπτιση

Νιώθω αυτοπεπτιση όταν βγαίνω από το σπίτι

Υπνος

Κοιμάμαι ήρεμα

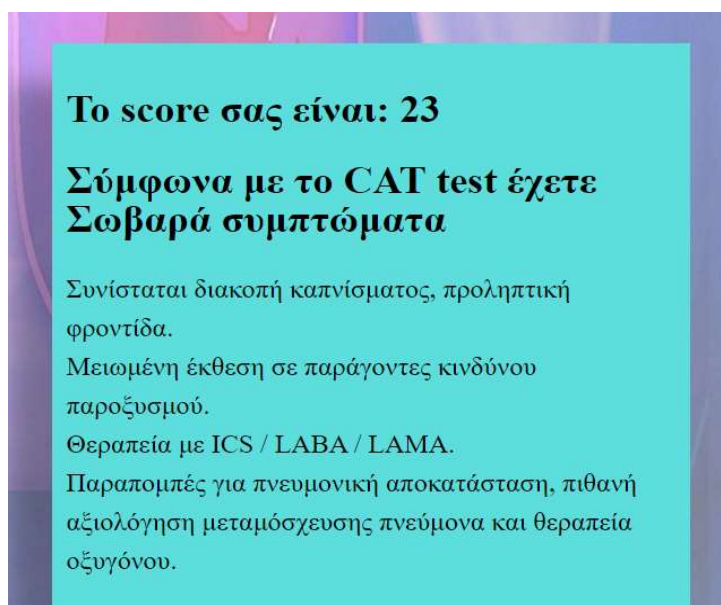
Ενέργεια

Έχω πολύ ενέργεια

Προηγούμενο Υποβολή

Εικόνα 34 : Ερωτηματολόγιο CAT, δεύτερη σελίδα

Παρακάτω φαίνεται το αποτέλεσμα που μας έδωσε το ερωτηματολόγιο από την δοκιμή του.



Το score σας είναι: 23

Σύμφωνα με το CAT test έχετε Σωβάρá συμπτώματα

Συνίσταται διακοπή καπνίσματος, προληπτική φροντίδα.

Μειωμένη έκθεση σε παράγοντες κινδύνου παροξυσμού.

Θεραπεία με ICS / LABA / LAMA.

Παραπομπές για πνευμονική αποκατάσταση, πιθανή αξιολόγηση μεταμόσχευσης πνεύμονα και θεραπεία οξυγόνου.

Εικόνα 35 : Αποτελέσματα ερωτηματολογίου CAT

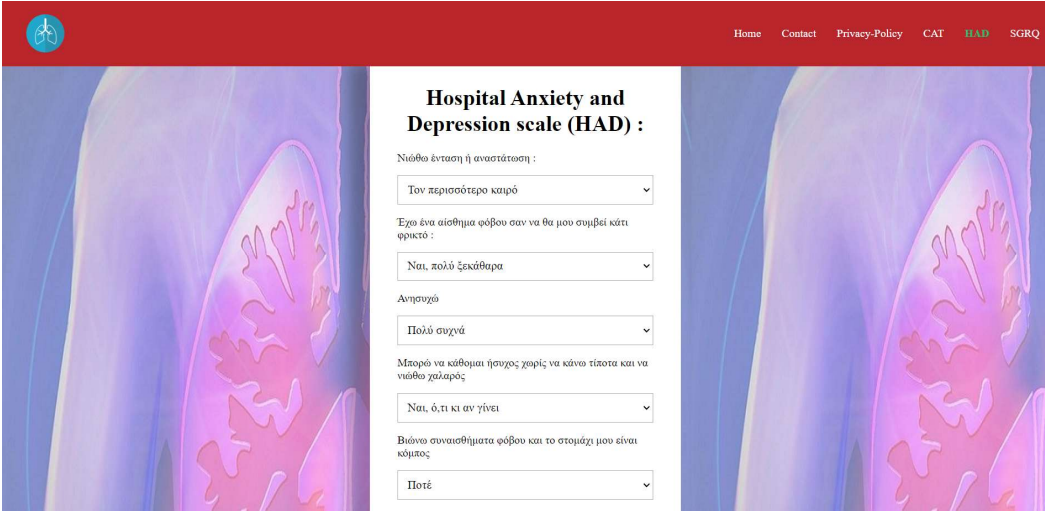
Το test αυτό λειτουργεί προσθετικά όλες οι ερωτήσεις έχουν το ίδιο βάρος. Κάθε απάντηση έχει διαφορετική βαθμολογία από 0-5. Σε κάθε ερώτηση η μεγαλύτερη βαθμολογία βρίσκεται στην χειρότερη περίπτωση. Από αυτό συνεπάγεται ότι όσο μεγαλύτερη βαθμολογία συμπληρώσει ο ασθενής τόσο χειρότερα είναι τα συμπτώματα της ΧΑΠ του με βάση το CAT.

Το τεστ αυτό μας δείχνει την κατάσταση της ΧΑΠ του ασθενή, έχοντας ως δεδομένα την κατάσταση των συμπτωμάτων και τα προβλήματα που δημιουργεί αυτή στον ασθενή. Παρόλα αυτά, στο ερωτηματολόγιο δεν αναφέρονται οι δραστηριότητες που ο χρήστης περιορίζεται να κάνει λόγω της πάθησής του και σχετικές συννοσηρότητες

που ο ασθενής μπορεί να έχει. Παράλληλα, όλα τα στοιχεία έχουν το ίδιο βάρος χωρίς να γίνεται προτίμηση σε σημαντικά στοιχεία για την διάγνωση. Τέλος, το ερωτηματολόγιο βασίζεται αποκλειστικά στην οπτική του ασθενή σε γενικές ερωτήσεις. Αυτό μπορεί να δημιουργήσει τον χρήστη να βαθμολογήσει λανθασμένα αρκετά τα στοιχεία με αποτέλεσμα να οδηγηθεί σε λάθος διάγνωση.

Στ'.5 Ερωτηματολόγιο HAD

Το ερωτηματολόγιο αυτό προσπαθεί να προβλέψει την ψυχολογική κατάσταση του ασθενή υπολογίζοντας το άγχος και την κατάθλιψη, αξιολογώντας συμπεριφορές (πιθανά συμπτώματα) που ο ασθενής αντιμετωπίζει τις τελευταίες ημέρες. Παρακάτω φαίνεται ένα δείγμα του ερωτηματολογίου αυτού.



Εικόνα 36 : Ερωτηματολόγιο HAD

Όπως και στο ερωτηματολόγιο CAT, το ερωτηματολόγιο HAD λειτουργεί προσθετικά, όλες οι ερωτήσεις του έχουν το ίδιο βάρος και η μεγαλύτερη βαθμολογία αντιστοιχεί σε χειρότερα συμπτώματα και για τις δύο περιπτώσεις του τεστ. Το ερωτηματολόγιο αυτό περιέχει 14 ερωτήσεις 7 ερωτήσεις για τον υπολογισμό του άγχους και 7 για τον υπολογισμό της κατάθλιψης. Όλες οι ερωτήσεις βαθμολογούνται από 0 έως 3.

Στοιχεία που σχετίζονται με την μέτρηση της κατάθλιψης είναι:

- Απολαμβάνω ακόμα τα ίδια πράγματα με το παρελθόν
- Μπορώ να γελάσω και να δω την θετική πλευρά των πραγμάτων
- Νιώθω χαρούμενος
- Νιώθω σαν να επιβραδύνομαι
- Έχω χάσει το ενδιαφέρον μου για την εμφάνιση μου
- Ανυπομονώ με ευχαρίστηση πράγματα να συμβούν
- Μπορώ να απολαύσω ένα καλό βιβλίο ή μια εκπομπή στο ραδιόφωνο ή την τηλεόραση

Στοιχεία που σχετίζονται με την μέτρηση του άγχους

- Αισθάνομαι ένταση ή κουρασμένος
- Φοβάμαι ότι κάτι απείσιο θα μου συμβεί
- Σκέφτομαι ανησυχητικές σκέψεις
- Μπορώ να καθίσω άνετα και να νιώσω χαλαρός
- Φοβάμαι και έχω κόμπους στο στομάχι
- Νιώθω ανήσυχος σαν να πρέπει να κάνω κάτι
- Ξαφνικά πανικοβάλλομαι

Οι μετρήσεις της κατάθλιψης και του άγχους μετριοούνται ξεχωριστά. Και στις δύο περιπτώσεις ξεχωριστά, αν το score είναι από 0-7 τότε τα συμπτώματα του ασθενή είναι φυσιολογικά. Σε περίπτωση που το score είναι από 8-10 τα συμπτώματα βρίσκονται στο όριο να χαρακτηριστούν μη φυσιολογικά. Σε περίπτωση που το score είναι από 11-21 τα συμπτώματα του ασθενή είναι μη φυσιολογικά. Το HAD score στο πρόγραμμα μου είναι το σύνολο των δύο αυτών score και παίρνει τιμές από 0-42.

Στ'.6 Ερωτηματολόγιο SGRQ

Το ερωτηματολόγιο αυτό σε αντίθεση με το HAD σχετίζεται με την διάγνωση του ΧΑΠ. Παρόλα αυτά δεν σχετίζεται αποκλειστικά με την πάθηση αυτή όπως το CAT αλλά σχετίζεται και με άλλες παθήσεις του αναπνευστικού. Το τεστ αυτό έχει συνολικά 42 ερωτήσεις και χωρίζεται σε 8 ενότητες και τρεις κατηγορίες συμπτώματα που έχει ο ασθενής, δραστηριότητες που περιορίζεται να κάνει λόγω της πάθησής του και σε ποιο βαθμό έχει επιπτώσεις η πάθηση τους. Όπως και τα άλλα δυο τεστ έτσι και σε αυτό προστίθενται οι απαντήσεις των ασθενών και όσο μεγαλύτερο είναι τόσο χειρότερη είναι η κατάσταση του ασθενή. Στο πρόγραμμα μου χρησιμοποίησα μια παραλλαγή του ερωτηματολογίου αυτού που είναι αποκλείστηκα για τους ασθενείς με ΧΑΠ, το SGRQ-C.

Στ'.6.1 SGRQ-C

Το ερωτηματολόγιο αυτό είναι παραλλαγή του ερωτηματολογίου SGRQ και σχετίζεται αποκλείστηκα με την ΧΑΠ. Παράλληλα, αποτελείτε από 14 ερωτήσεις οι οποίες χωρίζονται σε τρεις κατηγορίες ίδιες με αυτές του SGRQ.

Τα πρώτα 7 στοιχεία έχουν να κάνουν με τα συμπτώματα του ασθενή τις τελευταίες ευδομάδες.

- Βήχας
- Αναφορά φλεγμάτων
- Δύσπνοια
- Κρίσεις συριγμού
- Πόσες κρίσεις θωρακικού προβλήματος είχατε ;
- Πόσο συχνά έχετε καλές μέρες(με μικρές ενοχλήσεις στο στήθος);
- Εάν έχετε συριγμό είναι χειρότερος το πρωί;

Η κατηγορία σχετίζεται με τις δραστηριότητες περιλαμβάνει 2 στοιχεία

- Ερωτήσεις σχετικά με ποιες δραστηριότητες συνήθως σας κάνουν να αισθάνεστε δύσπνοια
- Πως σας επηρεάζει η κατάσταση του στήθους σας

Η Τρίτη κατηγορία σχετίζεται με τα συμπτώματα περιλαμβάνει τις υπόλοιπες 5 ερωτήσεις

- Πως περιγράφετε την κατάσταση του στήθους σας
- Ερωτήσεις σχετικά με Βήχα και δύσπνοια
- Άλλα θέματα του στήθους σας που σας ενοχλούν
- Ερωτήσεις για το πως οι δραστηριότητες μπορούν να επηρεάσουν την αναπνοή σας
- Πόσο συχνά το πρόβλημα στο στήθος σας επηρεάζει συνήθως στην καθημερινότητά σας

Κάθε μια από τις κατηγορίες αυτές έχει δικό της score και το άθροισμα αυτών αποτελεί το SGRQ score, παρόμοιος και το SGRQ-C score. Επειδή το SGRQ-C έχει λιγότερες ερωτήσεις από το SGRQ, χρησιμοποιείται μετατροπή των αποτελεσμάτων της παραδοχής σε αποτελέσματα του SGRQ με την χρήση των παρακάτω τύπων για κάθε κατηγορία.

Συμπτώματα: $\text{SGRQ score} = (\text{SGRQ-C score} * 0.99) + 0.94$

Δραστηριότητες: $\text{SGRQ score} = (\text{SGRQ-C score} * 0.87) + 7.01$

Επιπτώσεις: $\text{SGRQ score} = (\text{SGRQ-C score} * 0.88) + 2.18$

Συνολικά: $\text{SGRQ score} = (\text{SGRQ-C score} * 0.90) + 3.10$

Στην παρακάτω εικόνα φαίνεται ένα δείγμα του ερωτηματολογίου SGRQ-C στην εφαρμογή μου. Τα αποτελέσματα που φαίνονται στην εικόνα είναι τα αποτελέσματα του ερωτηματολογίου SGRQ-C ,που επέλεξε ο χρήστης στην εφαρμογή, και έχουν τροποποιηθεί σε αποτελέσματα του SGRQ.

Πως σας επηρεασαν τις τελευταίες εβδομάδες

Βήχας :

Σχεδόν κάθε μέρα

Αναφορά φλεγματίων :

Σχεδόν κάθε μέρα

Έχω δύσπνοια :

Σχεδόν κάθε μέρα

Έχω κρίσεις σπασμού

Σχεδόν κάθε μέρα

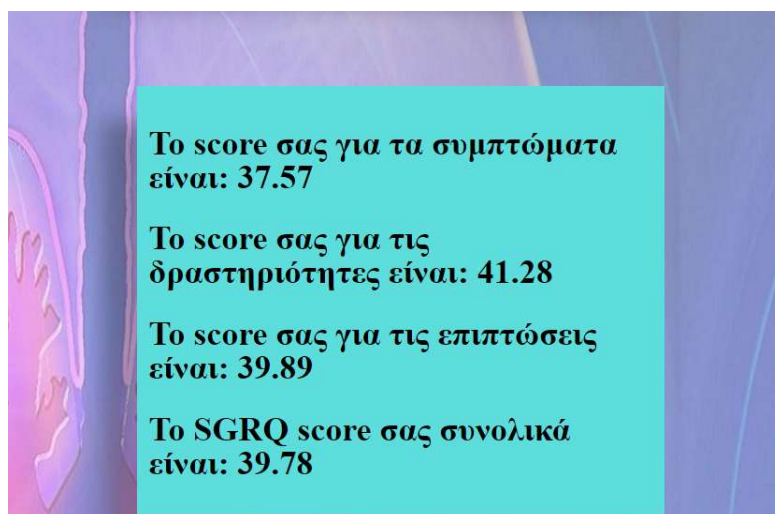
Πόσες κρίσεις θωρακικού προβλήματος είχατε;

3 ή περισσότερες

Πόσο συχνά έχετε καλές μέρες (με μικρές ενοχλήσεις στο στήθος);

Καθόλου

Εικόνα 37 : Ερωτηματολόγιο SGRQ



Εικόνα 38 : Αποτελέσματα Ερωτηματολογίου SGRQ

Τα αποτελέσματα και από τα τρία αυτά τεστ με την χρήση session αποθηκεύονται. Με αυτόν τον τρόπο μπορούν να μπουν αυτόματα στις κατάλληλες θέσεις στην φόρμα της αρχικής σελίδας, ώστε ο χρήστης να μην χρειάζεται να τα θυμάται.

Στ'.7 Αποτελέσματα φόρμας

Με την επιτυχής ολοκλήρωση της φόρμας ο χρήστης μπορεί να λάβει την διάγνωση του ασθενή (ή την διάγνωσή του αν είναι ο ασθενής), σύμφωνα με το σύστημα. Τα αποτελέσματα τις διάγνωσης αποτελούνται από τέσσερις διαφορετικές τιμές σύμφωνα με το σύστημα βαθμολόγησης GOLD. Σύμφωνα με αυτό οι ασθενείς ταξινομούνται σε τέσσερις διαφορετικές κατηγορίες ανάλογα με τον φαινότυπο της ΧΑΠ που έχουν. Σε κάθε βαθμό της ΧΑΠ προτείνουμε μέτρα για την καταπολέμηση της ΧΑΠ σύμφωνα με τα αποτελέσματα του

<https://www.mdcalc.com/calc/10161/copd-assessment-test-cat>

Στις 2 πρώτες κατηγορίες ασθενών(ήπια και μέτρια συμπτώματα) συνιστάται η διακοπή του καπνίσματος, προληπτική φροντίδα και αποφυγή κινδύνων παροξυσμού, όπως ψησταριές και τζάκια, σε περίπτωση που ο ασθενείς διαγνωστεί με σοβαρά συμπτώματα τα παραπάνω μέτρα απαιτούνται.

Σε περίπτωση που ο ασθενής διαγνωστεί με ήπια συμπτώματα συστήνεται η χρήση συσκευών εισπνοής και των φαρμάκων LAMA(Long-Acting Muscarinic Antagonist). Αν ο ασθενής διαγνωστεί με μέτρια συμπτώματα τότε μαζί με τα παραπάνω θα πρέπει εξετάσει το ενδεχόμενο για την χρήση φαρμάκων LABA(Long-Acting Beta Antagonists) και ICS(Inhaled Corticosteroids) καθώς και παραπομπές για πνευμονική αποκατάσταση.

Σε περίπτωση που τα συμπτώματα είναι σοβαρά ο ασθενής θα πρέπει να χρησιμοποιήσει και τα τρία αυτά φάρμακα και να εισέλθει προσέλθει σε παραπομπές για πνευμονική αποκατάσταση και πιθανή αξιολόγηση μεταμόσχευσης πνεύμονα.

Ζ' Συμπεράσματα

Η μελέτη αυτή εξετάζει την χρήση των συστημάτων μηχανικής μάθησης στην διάγνωση των ασθενών με ΧΑΠ. Με την δημιουργία του μοντέλου, καταφέραμε να λάβουμε υψηλή απόδοση του συστήματος με την χρήση των αλγορίθμων CART και ADABOOST για την δημιουργία δέντρου απόφασης, έχοντας συνολικό πληθυσμό 98 δείγματα από διαγνωσμένους ασθενείς. Παράλληλα καταφέραμε να φτιάξουμε μια δομημένη εφαρμογή την οποία κάθε χρήστης μπορεί να χρησιμοποιήσει ώστε να λάβει την διάγνωση ενός ασθενή. Εκτός, από την διάγνωση εφαρμογή αυτή δίνει αποτελέσματα με προτεινόμενες ενέργειες που ο ασθενής χρειάζεται να ακολουθήσει ώστε η κατάστασή του να καλυτερεύσει.

Κατά την διάρκεια δημιουργίας του μοντέλου , με την χρήση των μεθόδων επιλογής χαρακτηριστικών Feature Selection , βρήκαμε ότι τα σημαντικότερα χαρακτηριστικά στο δείγμα μας ήταν τα αποτελέσματα της σπυρομέτρησης FEV1,FEVPRED,FVC,FVCPRED.

Η εργασία αυτή είχε αρκετούς περιορισμούς. Αρχικά, όσον αφορά το μέγεθος του δείγματος, θεωρώ πως το δείγμα αυτό δεν είναι ικανοποιητικό για την δημιουργία συστήματος λήψης αποφάσεων ώστε να χρησιμοποιηθεί για γενική χρήση.

Παράλληλα, στην εργασία αυτή δεν αναφέρονται συγκρίσεις μεταξύ άλλων αλγορίθμων μηχανικής μάθησης όπως ο Adaboost ή ο C4.5 για την δημιουργία δέντρων απόφασης.

Παρόλα αυτά, το δείγμα μας είχε έναν ικανοποιητικό αριθμό χαρακτηριστικών (22) συμπεριλαμβανομένων αποτελεσμάτων ερωτηματολογίων, τεστ πνευμονικών λειτουργιών , ηλικία και ιστορικό καπνιστή. Επιπλέον, χρησιμοποιήθηκε ένας διαφορετικός αλγόριθμος για την δημιουργία μοντέλου(Adaboost), από αυτούς που χρησιμοποιούνται στις εργασίες της βιβλιογραφίας έχοντας ικανοποιητικά αποτελέσματα στην ταξινόμηση κάθε τιμής της κλάσης. Επιπλέον, με την χρήση της Python καταφέραμε να φτιάξουμε μοντέλο μηχανικής μάθησης, και με την χρήση των php,css, javascript να κατασκευάσουμε ένα φιλικό περιβάλλον για τον χρήστη.

Η' Παραρτήματα

Παράρτημα Α Python

Με την χρήση της γλώσσας προγραμματισμού Python φτιάξαμε και δοκιμάσαμε το μοντέλο μας. Αρχικά, χρησιμοποιήσαμε τις παρακάτω βιβλιοθήκες.

```
import pandas as pd
import numpy as np
from sklearn.ensemble import AdaBoostClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import StratifiedKFold
from sklearn import preprocessing
import matplotlib.pyplot as plt
from sklearn import metrics
```

Προκειμένου το αρχείο να μπορεί να χρησιμοποιηθεί από την Python. χρησιμοποιήθηκε η βιβλιοθήκη pandas με τον παρακάτω κώδικα.

```
df = pd.read_csv("./csvdataset1_featuresclass.csv", header=0, delimiter=';')  
Έχοντας λάβει το αρχείο καθορίζουμε τα χαρακτηριστικά και την κλάση από αυτό, με την χρήση της βιβλιοθήκης numpy.
```

```
x=np.array(df[['AGE', 'PackHistory', 'MWT1Best', 'FEV1', 'FEV1PRED', 'FVC', 'FVCPRED',  
, 'CAT', 'HAD', 'SGRQ', 'AGEquartiles', 'gender', 'smoking', 'Diabetes', 'muscular', 'hypertension', 'AtrialFib', 'IHD']])  
y=np.array(df['copd'])
```

Στη συνέχεια πήραμε τους αλγορίθμους που θα χρησιμοποιήσουμε από την βιβλιοθήκη Sklearn και βάλαμε τυπικούς περιορισμούς σε αυτούς.

```
dtree=DecisionTreeClassifier(max_depth=10, random_state=50,  
max_features=None,min_samples_leaf=8)  
adab=AdaBoostClassifier(n_estimators=50, base_estimator=dtree,learning_rate=1)
```

Έπειτα λάβαμε την μέθοδο διαχωρισμού των δεδομένων σε training και testing data από την βιβλιοθήκη sklearn. Και πραγματοποιήθηκε το πείραμα.

```
kf = StratifiedKFold(n_splits=10,shuffle=True)# X  
for train, test in kf.split(x,y):#X  
    #print("%s %s" % (train, test))  
    test_x=x[test]#X  
    test_y=y[test]  
    train_x=x[train]#X  
    train_y=y[train]  
    adab.fit(train_x,train_y)  
    y_pred= adab.predict(test_x)  
    acc = adab.score(test_x,test_y)  
    print(test_y)  
    print(y_pred)  
    print(acc)  
  
    for i in test_y:  
        real.append(i)  
  
    for i in y_pred:  
        pred.append(i)
```

Τα Αποτελέσματα αυτά μετά συλλέγονται για την δημιουργία του πίνακα σύγχυσης (Confusion Matrix).

Έχοντας μαζέψει τα αποτελέσματα αυτά και με την χρήση των βιβλιοθηκών Sklearn

και matplotlib. Φτιάχτηκε ο παρακάτω κώδικας. Οι πίνακες nreal και npredict είναι numpy arrays των λιστών real και predict αντίστοιχα.

```
confusion_matrix = metrics.confusion_matrix(nreal, npredict)

cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix =
confusion_matrix, display_labels = [1, 2, 3, 4])

cm_display.plot()
plt.show()
```

Για την αποθήκευση και χρήση του μοντέλου χρησιμοποιήθηκε η Βιβλιοθήκη joblib

```
import joblib
```

Για την αποθήκευση χρησιμοποιήθηκε η παρακάτω γραμμή κώδικα

```
joblib.dump(adab, open(filename, 'wb'))
```

Και για την φόρτωση και χρήση του μοντέλου χρησιμοποιήθηκαν οι παρακάτω γραμμές κώδικα

```
import joblib

test=[]

for i in range(17):
    Add= sys.argv[i+1]
    test.append(Add)

numpytest=np.array([test])
#print(numpytest)

filename = "C:\\xampp\\htdocs\\img\\SaveModel.sav"
loaded_model = joblib.load(open(filename, 'rb'))

y_pred= loaded_model.predict(numpytest)
print(str(y_pred[0]))
```

Παράλληλα ήταν απαραίτητες οι παρακάτω γραμμές για την επικοινωνία της με την php

```
#!C:\Python10\python.exe

#print("Content-Type: text/html\n")
```

Παράρτημα Β HTML

Με την χρήση της HTML φτιάχτηκε ο βασικός σκελετός της εφαρμογής, όπως η φόρμα, κουμπιά υποβολής και πλοήγησης.

Παρακάτω φαίνεται ο κώδικας δημιουργίας του σκελετού του μενού πλοήγησης. Από αυτό ο χρήστης μπορεί να περιηγηθεί στις υπόλοιπες σελίδες της εφαρμογής επιλέγοντας το κατάλληλο σύνδεσμο.

```
<nav class="nav">
  <div class="container">
    
    <ul>
      <li><a href="#" class="current">Home</a></li>
      <li><a href="contact.php">Contact</a></li>
      <li><a href="privacy.php">Privacy-Policy </a></li>
      <li><a href="CAT.php">CAT</a></li>
      <li><a href="HAD.php">HAD</a></li>
      <li><a href="SGRQ.php">SGRQ</a></li>
    </ul>
  </div>
</nav>
```

Παρακάτω φαίνεται ο κώδικας δημιουργίας της φόρμας για το πρώτο tab

```
<form id="regForm" action="" method="POST" onsubmit="myFunction()">
  <!-- One "tab" for each step in the form: -->

  <div class="tab"><h1>Εισάγετε τα στοιχεία του ασθενή :</h1>
    <h2>Αρχικά στοιχεία του ασθενή:</h2>

    <label for="AGE" class="col-form-label"
style="display:inline;">Ηλικία:</label>
    <p><input type="number" class="form-control" name="AGE" id="AGE"
placeholder="Εισάγετε την ηλικία του ασθενή"></p>

    <p><label for="Gender" class="col-form-label">Φύλο: </label></p>
    <p><select class="form-control" id="Gender" name="Gender">
      <option selected disabled value="">Επιλογή</option>
      <option value=1>Ανδρας</option>
      <option value=0>Γυναίκα</option>
    </select></p>

  </div>
```

Το Tab αυτό ασχολείται με τα αρχικά εξωτερικά στοιχεία του ασθενή, δηλαδή το φύλο και ηλικία του. Το επόμενο σχετίζεται με ερωτήσεις που κάνει ο γιατρός για να ξεκαθαρίσει το ιστορικό καπνιστή του ασθενή.

```
<div class="tab"><h2>Ερωτήσεις:</h2>
  <p><label for="smoking" class="col-form-label">Είστε αυτήν τη στιγμή
Καπνιστής; </label></p>
  <p><select class="custom-select" id="smoking" name="smoking">
    <option selected disabled>Επιλογή </option>
```

```

        <option value=2>Καπνιστής</option>
        <option value=1>Μη Καπνιστής</option>
    </select></p>

    <label for="PackHistory" class="col-form-label">Pack Years:
</label><div id="circle" class="circle" style="margin-right: 280px;">?</div>
    <span id="zi" class="triangle-isosceles top" style="margin-left:
24px;"> Ένα Pack Year ορίζετε ως 20 τσιγάρα(1 πακέτο) κάθε μέρα για ένα χρόνο
</span>

    <p><input type="number" class="form-control" name="PackHistory"
id="PackHistory" placeholder="Εισάγετε το Pack Years του ασθενή"></p>
</div>

```

Το τρίτο tab σχετίζεται με συννοσηρότητες που πιθανόν να πάσχει ο χρήστης όπως η αρτηριακή υπέρταση και ο σακχαρώδης διαβήτης.

```

<div class="tab"><h2>Συννοσηρότητες του ασθενή:</h2>
    <p><label for="Diabetes" class="col-form-label">Ζακχαρώδης Διαβήτης:
</label></p>
    <p><select class="custom-select" id="Diabetes" name="Diabetes">
        <!--<option selected disabled>Επιλογή </option>-->
        <option selected value=0>Όχι</option>
        <option value=1>Ναι</option>
    </select></p>

    <p><label for="hypertension" class="col-form-label">Αρτηριακή Υπέρταση:
</label></p>
    <p><select class="custom-select" id="hypertension" name="hypertension">

        <option selected value=0>Όχι</option>
        <option value=1>Ναι</option>
    </select></p>

    <p><label for="AtrialFib" class="col-form-label">Κολπική Μαρμαρυγή:
</label></p>
    <p><select class="custom-select" id="AtrialFib" name="AtrialFib">

        <option selected value=0>Όχι</option>
        <option value=1>Ναι</option>
    </select></p>

    <p><label for="IHD" class="col-form-label">Ισχαιμική Καρδιοπάθεια:
</label></p>
    <p><select class="custom-select" id="IHD" name="IHD">

        <option selected value=0>Όχι</option>
        <option value=1>Ναι</option>
    </select></p>

    <p><label for="muscular" class="col-form-label">Μυοσκελετικά
Προβλήματα: </label></p>

```

```

<p><select class="custom-select" id="muscular" name="muscular">

    <option selected value=0>Όχι</option>
    <option value=1>Ναι</option>

</select></p>
</div>

```

Στο τέταρτο tab περιέχονται τα αποτελέσματα πνευμονικών ελέγχων MWT1 και σπιρομέτρησης.

```

<div class="tab"><h2>Αποτελέσματα ελέγχου του ασθενή</h2>
    <label for="MWT1Best" class="col-form-label">MWT1Best: </label><div
id="circle" class="circle" style="margin-right: 270px;">?</div>
    <span id="zi" class="triangle-isosceles top" style="margin-left:
29px;">Απόσταση που διανύει ο ασθενής σε 6 λεπτά σε μέτρα (καλύτερη
προσπάθεια)</span>
    <p><input type="number" class="form-control" name="MWT1Best"
id="MWT1Best" placeholder="MWT1Best"></p>

    <label for="FEV1" class="col-form-label">FEV1: </label><div id="circle"
class="circle" style="margin-right: 300px;">?</div>
    <span id="zi" class="triangle-isosceles top">Ποσότητα αέρα που ο
ασθενής μπορεί να εξαναγκάσει από τους πνεύμονές του σε ένα δευτερόλεπτο σε
λίτρα. Μέτρηση της πνευμονικής λειτουργίας,αποτέλεσμα σπιρομέτρησης</span>
    <p><input type="number" class="form-control" name="FEV1" id="FEV1"
placeholder="FEV1"></p>

    <label for="FEV1PRED" class="col-form-label">FEV1PRED: </label><div
id="circle" class="circle" style="margin-right: 260px;">?</div>
    <span id="zi" class="triangle-isosceles top" style="margin-left:
40px;">Πρόβλεψη ποσότητας αέρα που ο ασθενής μπορεί να εξαναγκάσει από τους
πνεύμονές του σε ένα δευτερόλεπτο σε λίτρα. Μέτρηση της πνευμονικής
λειτουργίας,αποτέλεσμα σπιρομέτρησης</span>
    <p><input type="number" class="form-control" name="FEV1PRED"
id="FEV1PRED" placeholder="FEV1PRED"></p>

    <label for="FVC" class="col-form-label">FVC: </label><div id="circle"
class="circle" style="margin-right: 290px;">?</div>
    <span id="zi" class="triangle-isosceles top" style="margin-left:
10px;">Συνολικός όγκος αέρα που ο ασθενής μπορεί να εκπνεύσει με μία
αναπνοή</span>
    <p><input type="number" class="form-control" name="FVC" id="FVC"
placeholder="FVC"></p>

    <label for="FVCPRED" class="col-form-label">FVCPRED: </label><div
id="circle" class="circle" style="margin-right: 270px;">?</div>
    <span id="zi" class="triangle-isosceles top" style="margin-left:
30px;">Πρόβλεψη Συνολικού όγκου αέρα που ο ασθενής μπορεί να εκπνεύσει με μία
αναπνοή</span>
    <p><input type="number" class="form-control" name="FVCPRED"
id="FVCPRED" placeholder="FVCPRED"></p>

</div>

```

Παράλληλα, στο tag span περιέχεται η εξήγηση για κάθε στοιχείο. Κάθε tag αντιστοιχεί με ένα id="circle" που βρίσκεται από πάνω του ο ξεχωριστός κώδικας css χρειάζοταν ώστε να βρίσκονται στην κατάλληλη θέση.

Στο τελευταίο Tab περιέχονται τα αποτελέσματα των ερωτηματολογίων του ασθενή

```
<div class="tab"><h2>Αποτελέσματα Ερωτηματολογίων :</h2>
    <label for="CAT" class="col-form-label">Αποτελέσμα CAT: </label><div
id="circle" class="circle" style="margin-right: 235px;">?</div>
    <span id="zi" class="triangle-isosceles top" style="margin-left:
64px;">Αποτελέσμα Ερωτηματολογίου COPD Assessment Test</span>
    <p><input type="number" class="form-control" name="CAT" id="CAT"
placeholder="CAT" value=?php
    if(isset($_SESSION['CAT']))){
        print_r($_SESSION['CAT']);

    }else{
        echo("");
    }

    ?></p>

    <label for="HADi" class="col-form-label">Αποτελέσμα HAD: </label><div
id="circle" class="circle" style="margin-right: 225px;">?</div>
    <span id="zi" class="triangle-isosceles top" style="margin-left:
75px;">Αποτελέσμα Ερωτηματολογίου Hospital Anxiety and Depression Scale</span>
    <p><input type="number" class="form-control" name="HADi" id="HADi"
placeholder="HAD" value=?php
    if(isset($_SESSION['HAD']))){
        print_r($_SESSION['HAD']);

    }else{
        echo("");
    }

    ?></p>

    <label for="SGRQ" class="col-form-label">Αποτελέσμα SGRQ: </label> <div
id="circle" class="circle" style="margin-right: 215px;">?</div>
    <span id="zi" class="triangle-isosceles top" style="margin-left:
85px;">Αποτελέσμα Ερωτηματολογίου ST GEORGE'S RESPIRATORY QUESTIONNAIRE</span>
    <p><input type="number" class="form-control" name="SGRQ" id="SGRQ"
placeholder="SGRQ" value=?php
    if(isset($_SESSION['SGRQ']))){
        print_r($_SESSION['SGRQ']);

    }else{
        echo("");
    }

    ?> </p>
</div>
```

Σε κάθε ένα από αυτά τα πεδία υπάρχει κώδικας php ώστε να συμπληρωθεί αυτόματα σε περίπτωση που ο χρήστης έχει ήδη συμπληρώσει τα αντίστοιχα ερωτηματολόγια. Παρακάτω φαίνεται ο κώδικας δημιουργίας των κουμπιών υποβολής και πλοήγησης στην φόρμα. Παράλληλα φαίνεται και ο χάρτης για τα tabs(δείχνει πια είναι συμπληρωμένα και πια ασυμπλήρωτα).

```
<div style="overflow:auto;">
  <div style="float:right;">
    <button type="button" id="prevBtn" onclick="nextPrev(-1)">Προηγούμενο</button>
    <button type="button" id="nextBtn" onclick="nextPrev(1)">Επόμενο</button>
  </div>
</div>
<!-- Circles which indicates the steps of the form: -->
<div style="text-align:center;margin-top:40px;">
  <span class="step"></span>
  <span class="step"></span>
  <span class="step"></span>
  <span class="step"></span>
  <span class="step"></span>
</div>
</form>
```

Παράρτημα Γ CSS

Αρχικά φαίνονται μικρά μέρη κώδικα σχετικά με τη γραμματοσειρά και το φόντο.

```
* {
  box-sizing: border-box;
  font-family: serif;
}

body {
  background-color: #f1f1f1;
  background-image: url(backgroundpurple50better.jpg);
}
```

Παρακάτω φαίνεται ο κώδικας για την μορφή της φόρμας. Σε αυτό συμπεριλαμβάνονται τα στοιχεία της φόρμας, τα κουμπιά πλοήγησης της φόρμας, το κουμπί υποβολής, και τα tabs. Τα tabs αρχικά είναι κριμένα και καθορίζεται από την javascript ποιο απ' όλα θα εμφανίζεται στον χρήστη.

```
#regForm {

  background-color: #ffffff;
  margin: 100px auto;
  /*font-family: Raleway;*/
  padding: 40px;
  width: 30%;
  min-width: 300px;
  box-shadow: -30px 30px 20px rgba(0, 0, 0, 0.3);
}
```

```
h1 {
  text-align: center;
}

select {
  padding: 10px;
  width: 100%;
  font-size: 17px;
  /*font-family: Raleway;*/
  border: 1px solid #aaaaaa;
}
select.invalid {
  background-color: #ffdddd;
}

input {
  padding: 10px;
  width: 100%;
  font-size: 17px;
  /*font-family: Raleway;*/
  border: 1px solid #aaaaaa;
}

textarea{
  padding: 10px;
  width: 100%;
  font-size: 17px;
  border: 1px solid #aaaaaa;
  height: fit-content;
  min-height: 280px;
}

/* Mark input boxes that gets an error on validation: */
input.invalid {
  background-color: #ffdddd;
}

/* Hide all steps by default: */
.tab {
  display: none;
}
button {
  background-color: #04AA6D;
  color: #ffffff;
  border: none;
  padding: 10px 20px;
  font-size: 17px;
  /*font-family: Raleway;*/
  cursor: pointer;
}

button:hover {
  opacity: 0.8;
}
```

```

#prevBtn {
  background-color: #bbbbbb;
}

/* Make circles that indicate the steps of the form: */
.step {
  height: 15px;
  width: 15px;
  margin: 0 2px;
  background-color: #bbbbbb;
  border: none;
  border-radius: 50%;
  display: inline-block;
  opacity: 0.5;
}

.step.active {
  opacity: 1;
}

/* Mark the steps that are finished and valid: */
.step.finish {
  background-color: #04AA6D;
}

```

Στη συνέχεια φαίνεται ο κώδικας για την εμφάνιση του μενού

```

.logo{
  font-size: 46px;
  width: 1.2em;
  margin-left: 0.8em;
}

.nav {
  position: fixed;
  background-color: #B8252A;
  top: 0;
  left: 0;
  right: 0;
  transition: all 0.3s ease-in-out;
}

.nav .container {
  display: flex;
  justify-content: space-between;
  align-items: center;
  padding: 20px 0;
  transition: all 0.3s ease-in-out;
}

.nav ul {
  display: flex;
  list-style-type: none;
  align-items: center;
}

```

```

    justify-content: center;
}

.nav a {
    color: #fff;
    text-decoration: none;
    padding: 7px 15px;
    transition: all 0.3s ease-in-out;
}

.nav.active {
    background-color: #fff;
    box-shadow: 0 2px 10px rgba(0, 0, 0, 0.3);
}

.nav.active a {
    color: #000;
}

.nav.active .container {
    padding: 10px 0;
}

.nav a.current,
.nav a:hover {
    color: #1DBF73;
    font-weight: bold;
}

```

Έπειτα, φαίνεται ο κώδικας παρουσίασης των αποτελεσμάτων που δίνονται από την php.

```

.phpoutput {
    margin-left: 27.5em;
    background-color: rgb(93, 222, 220);
    padding: 10px;
    width: 30%;
    min-width: 300px;
    box-shadow: -30px 30px 20px rgba(0, 0, 0, 0.3);
    font-size: larger;
    line-height: 150%;
    /*visibility: hidden;*/
}

br {
    line-height: 300%;
}

```

Τέλος, φαίνεται ο κώδικας παρουσίασης των εξηγήσεων των στοιχείων της φόρμας.

```

.circle{
    width:15px;
    height:15px;
    border-radius:15px;
}

```

```

font-size:15px;
color:#131212;
/*line-height:25px;*/
text-align:center;
background:#83ceec;
display:inline;
float:right;
}

span {
  display: none;
}
.circle:hover+span{
  display: block;
  position: relative;
  max-width: 20em;
  box-shadow: 0 0.125rem 0.5rem rgba(0, 0, 0, .3), 0 0.0625rem 0.125rem rgba(0, 0, 0, .2);
}

.triangle-isosceles.top{
  display:none;
  background: #83ceec;
}

.triangle-isosceles.top:after{
  top: -15px;
  left: 50px;
  bottom: auto;
  right: auto;

  border-bottom: 15px solid #83ceec;
  border-left: 15px solid transparent;
  border-right: 15px solid transparent;
  border-top: 0;
}

.triangle-isosceles{
  position: relative;
  padding: 15px;
  margin: 1em 0 4em;

  color: #000;
  background: linear-gradient(#f9d835, #f3921c);
  border-radius: 10px;
  /*margin-bottom: 2em;*/
}

.triangle-isosceles:after{
  content: "";
  position: absolute;

```

```

bottom: -15px;
left: 50px;
width:0;
height: 0;

border-top: 15px solid #f3921c;

border-left: 15px solid transparent;
border-right: 15px solid transparent;
}

#zi {
position: absolute;
z-index: 10;
padding-top: 30px;
}

```

Παράρτημα Δ Javascript

Η γλώσσα προγραμματισμού Javascript χρησιμοποιήθηκε ώστε η εφαρμογή να γίνει πιο δυναμική.

Συγκεκριμένα χρησιμοποιήθηκε για την διαχείριση των tabs με την χρήση του παρακάτω κώδικα

```

const myDiv = document.getElementById('hidden');
const main = document.getElementsByClassName('circle');

var currentTab = 0; // Current tab is set to be the first tab (0)
showTab(currentTab); // Display the current tab
sub=0;

function showTab(n) {// This function will display the specified tab of the
form...
var x = document.getElementsByClassName("tab");
x[n].style.display = "block";

if (n == 0) {//... and fix the Previous/Next buttons:
document.getElementById("prevBtn").style.display = "none";
} else {
document.getElementById("prevBtn").style.display = "inline";
}
if (n == (x.length - 1)) {
document.getElementById("nextBtn").innerHTML = "Υποβολή";
} else {
document.getElementById("nextBtn").innerHTML = "Επόμενο";
}
// ... and run a function that will display the correct step
indicator:
fixStepIndicator(n)
}

function nextPrev(n) {// This function will figure out which tab to display

```

```

        var x = document.getElementsByClassName("tab");// Exit the function if
any field in the current tab is invalid:

        if (n == 1 && !validateForm()){
            return false;
        }
        x[currentTab].style.display = "none"; // Hide the current tab:

        currentTab = currentTab + n; // Increase or decrease the current tab by
1:

        if (currentTab >= x.length) { // if you have reached the end of the
form...

            var inv=document.getElementById("phpoutput");
            inv.style.display = 'block';
            document.getElementById("regForm").submit();// ... the form gets
submitted:
            sub=1

            return false;
        }

        showTab(currentTab);// Otherwise, display the correct tab:
    }
function fixStepIndicator(n) {
    // This function removes the "active" class of all steps...
    var i, x = document.getElementsByClassName("step");
    for (i = 0; i < x.length; i++) {
        x[i].className = x[i].className.replace(" active", "");
    }
    //... and adds the "active" class on the current step:
    x[n].className += " active";
}

```

Και τον έλεγχο της φόρμας χρησιμοποιώντας τον παρακάτω κώδικα

```

function validateForm() { // This function deals with validation of the form
fields

    var valid = true;
    x = document.getElementsByClassName("tab");
    y = x[currentTab].getElementsByTagName("input");
    z = x[currentTab].getElementsByTagName("select");

```

```

        for (i = 0; i < y.length; i++) { // A loop that checks every input
field in the current tab:

            if (y[i].value == "") { // If a field is empty...
                y[i].className += " invalid"; // add an "invalid" class to the
field:

                valid = false; // and set the current valid status to false
            }else if(y[i].className.includes(" invalid") == true){
                y[i].className -= " invalid";
            }

            if (currentTab==0 && (Number(y[i].value) <= 0 || Number(y[i].value)
>= 100 )){// this is for AGE on first tab
                alert("Παρακαλώ εισάγετε πραγματική Ηλικία");
                y[i].className += " invalid"; // add an "invalid" class to the
field:

                valid = false;
            }else if(currentTab==0 && y[i].className.includes(" invalid") ==
true){
                y[i].className -= " invalid";
            }

            if (currentTab==1 && Number(y[i].value) <= 0 ){// this is for Pack
History
                alert("Παρακαλώ εισάγετε πραγματικά Packyears στο παιδίο
Ιστορικό καπνιστή");
                y[i].className += " invalid"; // add an "invalid" class to the
field:

                valid = false;
            }else if(currentTab==1 && y[i].className.includes(" invalid") ==
true){
                y[i].className -= " invalid";
            }

            if (currentTab==3 && i==0 && (Number(y[i].value) <= 0 ||
Number(y[i].value) >= 900)){ // For MWT1Best
                alert("Παρακαλώ εισάγετε πραγματικό MWT1Best");
                y[i].className += " invalid"; // add an "invalid" class to the
field:

                valid = false;
            }else if(currentTab==3 && i==0 && y[i].className.includes("
invalid") == true){
                y[i].className -= " invalid";
            }

            if (currentTab==3 && i==1 && (Number(y[i].value) <= 0 ||
Number(y[i].value) >= 10)){ // For FEV1
                alert("Παρακαλώ εισάγετε πραγματικό FEV1");
                y[i].className += " invalid"; // add an "invalid" class to the
field:

                valid = false;
            }else if(currentTab==3 && i==1 && y[i].className.includes("
invalid") == true){
                y[i].className -= " invalid";
            }

```

```

        if (currentTab==3 && i==2 && (Number(y[i].value) <= 0 ||
Number(y[i].value) >= 150)){ // For FEV1Pred
            alert("Παρακαλώ εισάγετε πραγματικό FEV1Pred");
            y[i].className += " invalid"; // add an "invalid" class to the
field:
                valid = false;
            }else if(currentTab==3 && i==2 && y[i].className.includes("
invalid") == true){
                y[i].className -= " invalid";
            }

        if (currentTab==3 && i==3 && (Number(y[i].value) <= 0 ||
Number(y[i].value) >= 15)){ // For FCV
            alert("Παρακαλώ εισάγετε πραγματικό FCV");
            y[i].className += " invalid"; // add an "invalid" class to the
field:
                valid = false;
            }else if(currentTab==3 && i==3 && y[i].className.includes("
invalid") == true){ // for FCV
                y[i].className -= " invalid";
            }

        if (currentTab==3 && i==4 && (Number(y[i].value) <= 0 ||
Number(y[i].value) >= 150)){ // For FCPred
            alert("Παρακαλώ εισάγετε πραγματικό FCPred");
            y[i].className += " invalid"; // add an "invalid" class to the
field:
                valid = false;
            }else if(currentTab==3 && i==4 && y[i].className.includes("
invalid") == true){
                y[i].className -= " invalid";
            }

        if (currentTab==4 && i==0 && (Number(y[i].value) < 0 ||
Number(y[i].value) >= 41)){// For CAT
            alert("Παρακαλώ εισάγετε πραγματικό CAT score");
            y[i].className += " invalid"; // add an "invalid" class to the
field:
                valid = false;
            }else if(currentTab==4 && i==0 && y[i].className.includes("
invalid") == true){
                y[i].className -= " invalid";
            }

        if (currentTab==4 && i==1 && (Number(y[i].value) < 0 ||
Number(y[i].value) >= 43)){// For HAD
            alert("Παρακαλώ εισάγετε πραγματικό HAD score");
            y[i].className += " invalid"; // add an "invalid" class to the
field:
                valid = false;
            }else if(currentTab==4 && i==1 && y[i].className.includes("
invalid") == true){
                y[i].className -= " invalid";
            }

```

```

        if (currentTab==4 && i==2 && (Number(y[i].value) < 0 ||
Number(y[i].value) >= 120)){
            alert("Παρακαλώ εισάγετε πραγματικό SGRQ score");
            y[i].className += " invalid"; // add an "invalid" class to the
field:
            valid = false;
        }else if(currentTab==4 && i==2 && y[i].className.includes("
invalid") == true){
            y[i].className -= " invalid";
        }
    }

    for (i = 0; i< z.length; i++){

        if (z[i].value == "") {
            z[i].className += " invalid";
            valid = false;
        }else if (z[i].className.includes(" invalid") == true) {
            z[i].className -= " invalid";
        }
    }

    // If the valid status is true, mark the step as finished and valid:
    if (valid) {
        document.getElementsByClassName("step")[currentTab].className += "
finish";
    }

    return valid; // return the valid status
}

```

Παράρτημα Ε PHP

Η γλώσσα προγραμματισμού PHP χρησιμοποιήθηκε για την επικοινωνία με την Python όπου το μοντέλο φορτώνεται και την εμφάνιση του αποτελέσματος της φόρμας.

```
<div class="phpoutput" Id="phpoutput" style="visibility: hidden">
<?php
if ($_SERVER["REQUEST_METHOD"] == "POST") {

    echo '<style>#phpoutput{visibility: visible !important;}</style>';

    $AGE = $_POST['AGE'];
    $PackHistory = $_POST['PackHistory'];
    $MWT1Best = $_POST['MWT1Best'];
    $FEV1 = $_POST['FEV1'];
    $FEV1PRED = $_POST['FEV1PRED'];
    $FVC = $_POST['FVC'];
    $FVCPRED = $_POST['FVCPRED'];
    $CAT = $_POST['CAT'];
    $SGRQ = $_POST['SGRQ'];
    $gender = $_POST['Gender'];
    $smoking = $_POST['smoking'];
    $Diabetes = $_POST['Diabetes'];
    $muscular = $_POST['muscular'];
    $hypertension = $_POST['hypertension'];
    $AtrialFib = $_POST['AtrialFib'];
    $IHD= $_POST['IHD'];

    if($AGE<=65) {

        $AGEquartiles=1;
    }elseif($AGE<=70) {
        $AGEquartiles=2;
    }elseif($AGE<=75) {
        $AGEquartiles=3;
    }else {
        $AGEquartiles=4;
    }

    $output=shell_exec("python ./backend.py ". $AGE . " " . $PackHistory . " "
    . $MWT1Best . " " . $FEV1 . " " . $FEV1PRED . " " . $FVC . " " . $FVCPRED . " " .
    $CAT . " " . $SGRQ . " " . $AGEquartiles . " " . $gender . " " . $smoking . " " .
    $Diabetes . " " . $muscular . " " . $hypertension . " " . $AtrialFib . " " . $IHD);

    #echo "<p style='font-family: 'Mosterrat'; color:red; text-align:center;
width:320px; padding:10px; border:5px solid gray; margin: 0;
width:100%;'>$output</p>";

    /*echo($output);*/

    if ($output == 1) {
        echo("<h2>Ηπια συμπτώματα</h2>");
    }
}
```

```

        $text="Συνιστάται διακοπή του καπνίσματος, προληπτική φροντίδα.<br>
Μειωμένη έκθεση σε παράγοντες κινδύνου παροξυσμού.<br> Συστήνεται χρήση
συσκευών εισπνοής και φαρμάκων LAMA";

    }elseif ($output == 2){
        echo("<h2>Μέτρια συμπτώματα</h2>");
        $text="Συνιστάται διακοπή του καπνίσματος, προληπτική φροντίδα.<br>
Μειωμένη έκθεση σε παράγοντες κινδύνου παροξυσμού.<br> Χρήση συσκευών εισπνοής
και φαρμάκων LAMA.<br> Εξετάστε το ενδεχόμενο χρήσης φαρμάκων ICS και LABA.<br>
Παραπομπές για πνευμονική αποκατάσταση και πιθανή αξιολόγηση μεταμόσχευσης
πνεύμονα.";

    }elseif ($output == 3){
        echo("<h2>Σωβαρά συμπτώματα</h2>");
        $text=" Συνιστάται διακοπή καπνίσματος, προληπτική φροντίδα.<br>
Μειωμένη έκθεση σε παράγοντες κινδύνου παροξυσμού.<br> Θεραπεία με ICS / LABA /
LABA.<br> Παραπομπές για πνευμονική αποκατάσταση, πιθανή αξιολόγηση
μεταμόσχευσης πνεύμονα και θεραπεία οξυγόνου.";

    }elseif ($output == 4){
        echo("<h2>Πολύ Σοβαρά συμπτώματα</h2>");
        $text="Συνιστάται διακοπή καπνίσματος, προληπτική φροντίδα.<br>
Μειωμένη έκθεση σε παράγοντες κινδύνου παροξυσμού.<br> Θεραπεία με ICS / LABA /
LABA.<br> Παραπομπές για πνευμονική αποκατάσταση, πιθανή αξιολόγηση
μεταμόσχευσης πνεύμονα και θεραπεία οξυγόνου.";

    }

    echo("<p>$text</p>");

}
?></div>

```

Επιπλέον, χρησιμοποιήθηκαν Session Keys στην php για την αυτόματη συμπλήρωση των πεδίων CAT,HAD,SGRQ στην φόρμα.

```

?php
// Start the session
session_start();
?>

```

```
$_SESSION["CAT"] = $output;
```

```
$_SESSION["HAD"] = $output;
```

```
$_SESSION["SGRQ"] = $output;
```

Θ' Βιβλιογραφία

1. Xu Min, BinYu & FeiWang. Predictive Modeling of the Hospital Readmission Risk from Patients' Claims Data Using Machine Learning: A Case Study on COPD. Scientific reports. (2019) 9:2362.
2. Alan Kaplan, Hui Cao, J. Mark FitzGerald, Nick Iannotti, Eric Yang, Janwillem W.H. Kocks, Konstantinos Kostikas, David Price, Helen K. Reddel Ioanna Tsiligianni, Claus F. Vogelmeier, Pascal Pfister and Paul Mastoridis. Artificial Intelligence/Machine Learning in Respiratory Medicine and Potential Role in Asthma and COPD Diagnosis. J ALLERGY CLIN IMMUNOL PRACT VOLUME 9, NUMBER 6.
3. Vasilis Nikolaou, Sebastiano Massaro, Masoud Fakhimi, Lampros Stergioulas, David Price. COPD phenotypes and machine learning cluster analysis: A systematic review and future research agenda. Respiratory Medicine 171 (2020) 106093
4. Tadahiro Goto, MD, MPH *, Carlos A. Camargo Jr., MD, DrPH, Mohammad Kamal Faridi, MPH, Brian J. Yun, MD, MBA, MPH, Kohei Hasegawa, MD, MPH. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. American Journal of Emergency Medicine 36 (2018) 1650-1654
5. POUYA SOLTANI ZARRIN, NIELS ROECKENDORF, AND CHRISTIAN WENGER. In-Vitro Classification of Saliva Samples of COPD Patients and Healthy Controls Using Machine Learning Tools. IEEE Access Volume: 8 168053-168060
6. ISRAA MOHAMED, MOSTAFA M. FOUDA (Senior Member, IEEE), AND KHALID M. HOSNY (Senior Member, IEEE). Machine Learning Algorithms for COPD Patients Readmission Prediction: A Data Analytics Approach. IEEE Access Volume: 10 15279-15287
7. Junfeng Peng, ChuanChen, Mi Zhou, XiaohuaXie, Yuqi Zhou & Ching-Hsing Luo. A Machine-learning Approach to Forecast Aggravation Risk in Patients with Acute Exacerbation of Chronic Obstructive Pulmonary Disease with Clinical Indicators. Scientific Reports | (2020) 10:3118
8. Tadahiro Goto, Taisuke Joo, Hiroki Matsui, Kiyohide Fushimi, Hiroyuki Hayashie, and Hideo Yasunaga. Machine Learning-Based Prediction Models for 30-Day Readmission after Hospitalization for Chronic Obstructive Pulmonary Disease. COPD: JOURNAL OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE 2019, VOL. 16, NOS. 5–6, 338–343
9. Chia-Tung Wu, MSc; Guo-Hung Li, MSc; Chun-Ta Huang, MD, PhD; Yu-Chieh Cheng, BSc; Chi-Hsien Chen, MD, PhD; Jung-Yien Chien, MD, PhD; Ping-Hung Kuo, MD; Lu-Cheng Kuo, MD; Feipei Lai, PhD. Acute Exacerbation of a Chronic Obstructive Pulmonary Disease Prediction System Using Wearable Device Data, Machine Learning, and Deep Learning: Development and Cohort Study. JMIR Mhealth Uhealth 2021 | vol. 9 | iss. 5 | e22591

10. Dimitris Spathis, Panayiotis Vlamos. Diagnosing asthma and chronic obstructive pulmonary disease with machine learning. *Health Informatics Journal* 2019, Vol. 25(3) 811–827.
11. Marc Decramer, Cristopher B Cooper. Treatment of COPD: the sooner the better?. *Thorax* 2010;65:837e841. doi:10.1136/thx.2009.133355
12. Miravittles and Ribera Respiratory Research. Understanding the impact of symptoms on the burden of COPD. DOI 10.1186/s12931-017-0548-3
13. ΚΟΥΛΙΔΟΥ ΤΡΙΑΝΤΑΦΥΛΛΙΑ. Η Φυσιολογία του Αναπνευστικού Συστήματος στην Ηρεμία και Κατά τη Διάρκεια της Άσκησης. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης Τμήμα Επιστήμης Φυσικής Αγωγής και Αθλητισμού.

