



**Πανεπιστήμιο Πειραιώς**  
**Τμήμα Ψηφιακών Συστημάτων**

**Π.Μ.Σ. "Πληροφορικά Συστήματα & Υπηρεσίες"**

**Διαχείριση Ηλεκτρονικών Φακέλων Υγείας Διαβητικών και Covid-19  
Ασθενών με την χρήση υπηρεσιών Μηχανικής Μάθησης και  
Διαλειτουργικότητας Δεδομένων**

**Management of Electronic Health Records of Diabetic and Covid-19  
Patients using Machine Learning and Data Interoperability services**

**Κατεύθυνση:** Προηγμένα Πληροφορικά Συστήματα

**Διπλωματική Εργασία**

**Νικόλαος Ν. Ζαφειρόπουλος**

**Υπεύθυνος Καθηγητής:** Δημοσθένης Κυριαζής

ΠΕΙΡΑΙΑΣ, Φεβρουάριος 2023

## Περίληψη

Στην εποχή μας, είναι γεγονός ότι ο τομέας της υγείας και πιο συγκεκριμένα ο χώρος της υγειονομικής περίθαλψης, υστερεί σε θέματα ανάπτυξης έξυπνων εφαρμογών με μηχανισμούς Μηχανικής Μάθησης (Machine Learning - ML) και Διαλειτουργικότητας Δεδομένων (Data Interoperability). Στην καθημερινότητα ενός πολίτη, η χρήση και η εφαρμογή της πληροφορικής στην υγειονομική περίθαλψη παρέχει μια σειρά προνομίων με σημαντικά οφέλη. Αυτά μπορούν να αφορούν τόσο την καλύτερη και πιο άμεση εξυπηρέτηση των ασθενών, όσο και την διευκόλυνση του ιατρικού και νοσηλευτικού προσωπικού σε συνεργασία με τους υπεύθυνους χάραξης πολιτικών υγείας.

Ωστόσο, έπειτα από την εμφάνιση του νέου ιού στα τέλη του 2019 (SARS-CoV-2), άλλαξαν πολλές αντιλήψεις. Μια από αυτές σχετίζεται με την εφαρμογή και την κατάλληλη προσαρμογή των νέων τεχνολογιών στο χώρο της υγείας. Έτσι, έγινε αντιληπτό ότι υπήρξε μια σημαντική κινητικότητα στους τομείς έρευνας και ανάπτυξης του εν λόγω κλάδου.

Επομένως, αντικείμενο της παρούσας εργασίας είναι η εξέταση του Ηλεκτρονικού Φακέλου Υγείας (ΗΦΥ) ενός ασθενούς βάση του ιατρικού του ιστορικού. Το σύστημα που υποστηρίζει την εν λόγω αναφορά και το οποίο αναπτύχθηκε στα πλαίσια της διπλωματικής εργασίας είναι το πληροφοριακό σύστημα «**up-health**». Με το συγκεκριμένο σύστημα, το ιατρικό προσωπικό μπορεί να δει άμεσα και γρήγορα πληροφορίες του ΗΦΥ ενός ασθενούς. Τόσο η μοντελοποίηση των ΗΦΥ όσο και η ανάπτυξη των μοντέλων γίνονται μέσω των μηχανισμών Τεχνητής Νοημοσύνης (Artificial Intelligence - AI) για να προβλέψουν είτε με βάση το πρώτο σενάριο για το ποια είναι η καταλληλότερη αγωγή για τον διαβήτη είτε με το δεύτερο σενάριο για το ποια είναι η κατάσταση ενός ασθενούς, σχετικά με το αν πρέπει να εισαχθεί ή όχι στην Μονάδα Εντατικής Θεραπείας (ΜΕΘ), λόγω αναπνευστικών προβλημάτων (Covid-19). Επίσης, οι πληροφορίες που διαμοιράζονται εντός του συστήματος αλλά και με εξωτερικές οντότητες άλλων συστημάτων, θα μοντελοποιηθούν και θα ακολουθούν το πρότυπο HL7 FHIR (Health Level Seven Fast Healthcare Interoperability Resources). Όλες οι προαναφερόμενες υπηρεσίες, όπως και η Διεπαφή Χρήστη (User Interface) είναι φιλικά προς τον χρήστη μέσα από την χρήση νέων τεχνολογιών για την παροχή μιας κατάλληλης διαχείρισης των ΗΦΥ.

Επίσης, μέσα από το σύστημα «**up-health**», ο γιατρός μπορεί άμεσα να κλείσει ένα ραντεβού για τον ασθενή του, να εγγράψει έναν ασθενή, αλλά και να δει συγκεντρωτικά στοιχεία. Επιπλέον μπορεί να δει, τόσο την ανάλυση των σημαντικότερων χαρακτηριστικών που διατηρεί στο αρχείο του, όσο και όλων των χαρακτηριστικών που συνιστούν τον ΗΦΥ.

Πρόσθετα, πριν την ανάλυση του πληροφοριακού συστήματος «**up-health**» γίνεται μελέτη της διεθνούς βιβλιογραφίας που αφορά τον τομέα της Μηχανικής Μάθησης και κατά συνέχεια του δεύτερου τομέα προς μελέτη, αυτό της Διαλειτουργικότητας των Δεδομένων με λεπτομερής αναφορά, για τον ΗΦΥ.

Συνεπώς, μέσω του συγκεκριμένου συστήματος θα μπορούν να προσφερθούν λύσεις στον χώρο της υγείας. Θα λειτουργήσουν ως διαδικασίες για την ψηφιοποίηση και ηλεκτρονική καταχώρηση των στοιχείων ενός ασθενούς. Μέσω της προαναφερθείσας αναφοράς μειώνεται τόσο ο χρόνος όσο και το κόστος, που απαιτείται για την τήρηση ενός μηχανογραφημένου αρχείου, όπως επίσης μειώνεται και το συχνό φαινόμενο καθυστέρησης, όσον αφορά την διάγνωση μιας ασθένειας,

αλλά και στον άμεσο ορισμό ενός ραντεβού για την αποφυγή, τυχόν συγχύσεων στα ραντεβού από την πλευρά του ιατρικού προσωπικού, αλλά και των πολιτών για την αποφυγή άσκοπων μετακινήσεων.

**Λέξεις Κλειδιά:** Πληροφοριακό Σύστημα, Ηλεκτρονικός Φάκελος Υγείας, Μηχανική Μάθηση, Τεχνητή Νοημοσύνη, Διαλειτουργικότητα Δεδομένων.

## Abstract

In our time, it is a fact that the health sector, and more specifically the healthcare sector, lags behind in terms of developing smart applications with Machine Learning (ML) and Data Interoperability (DI) mechanisms. In the everyday life of a citizen, the use and application of Information Technology (IT) in health care provides a number of benefits with significant importance. These cannot be as better and more immediate for patients, as the facilitation of medical and nursing staff in collaboration with health policy makers.

However, following the emergence of the new virus in late 2019 (SARS-CoV-2), many perceptions have changed. One of them is related to the application and adaptation of new technologies in the field of health. Thus, it was realized that there has been a significant movement in the research and development areas of the sector in question.

Therefore, the subject of this work is the examination of the Electronic Health Record (EHR) of a patient based on his medical history. The system that supports this report and which was developed in the context of the thesis is the "**up-health**" information system. With this particular system, medical staff can instantly and quickly see a patient's EHR information. Both the modeling of the NHS and the development of the models are done through the mechanisms of Artificial Intelligence (AI) to predict either based on the first scenario which is the most suitable treatment for diabetes or with the second scenario which is the condition of a patient, regarding whether or not to be admitted to the Intensive Care Unit (ICU), due to respiratory problems (Covid-19). Also, the information that differs within the system but also with external entities of other systems, will be modeled and follow the HL7 FHIR (Health Level Seven Fast Healthcare Interoperability Resources) standard. The aforementioned services, as well as the User Interface, are user-friendly through the use of new technologies to provide appropriate management of the EHR.

Also, through the "**up-health**" system, the doctor can immediately make an appointment for his patient, register a patient, and see aggregate data. He can also see, both the analysis of the important features that you keep in his file, and all the features that make up the EHR.

Additionally, before the analysis of the "**up-health**" information system, a study is made of the international literature concerning the field of Machine Learning and then of the second field to be studied, that of Data Interoperability with a detailed report, for the EHR.

Therefore, through this specific system, solutions will be able to be offered in the field of health. They will act as procedures for the digitization and electronic registration of a patient's details. Through the aforementioned report, both the time and the cost required to maintain a computerized file are reduced, as well as the frequent phenomenon of delay, regarding the diagnosis of a disease, but also the immediate setting of an appointment to avoid, any confusion at appointments on the part of the medical staff, but also of citizens to avoid unnecessary travel.

**Keywords:** Information System, Electronic Health Record, Machine Learning, Artificial Intelligence, Data Interoperability.

## Ευχαριστίες

Κατ' αρχάς, θα ήθελα να εκφράσω τις ειλικρινείς ευχαριστίες μου στον επιβλέποντα καθηγητή, κ. Δημοσθένη Κυριαζή, για την αποδοχή, την ανάληψη και εκπόνηση της διπλωματικής μου διατριβής. Θα ήθελα επίσης να ευχαριστήσω, την μεταδιδάκτωρ κ. Αργυρώ Μαυρογιώργου για την υπομονή, τις καθοδηγήσεις της και την συνεχή υποστήριξη που μου παρείχε σε θέματα καθαρισμού και ανάλυσης δεδομένων. Επίσης να ευχαριστήσω θερμά και τον μεταδιδάκτωρ κ. Αθανάσιο Κιούρτη για τον χρόνο του και τις κατευθύνσεις στον τομέα της διαλειτουργικότητας των δεδομένων. Τέλος, θέλω να ευχαριστήσω την οικογένεια μου για την διαρκή τους υποστήριξη, που μου επέτρεψε την επιτυχή διεκπεραίωση των σπουδών μου.

## Πίνακας Περιεχομένων

Περίληψη .....	2
Abstract.....	4
Ευχαριστίες .....	5
Πίνακας Εικόνων .....	8
Κεφάλαιο 1 .....	10
Εισαγωγή.....	10
1.1.Μηχανική Μάθηση.....	10
1.2.Τομείς και Προκλήσεις Μηχανικής Μάθησης .....	11
1.3.Διαλειτουργικότητα Δεδομένων .....	12
1.4.Τομείς και Προκλήσεις Διαλειτουργικότητας Δεδομένων.....	13
1.5.Αντίκειμενο Διπλωματικής Εργασίας .....	13
1.6.Δομή Διπλωματικής.....	14
Κεφάλαιο 2 .....	15
Μηχανική Μάθηση.....	15
2.1.Προετοιμασία Δεδομένων.....	16
2.2.Εξαγωγή Χαρακτηριστικών .....	18
2.3.Οπτικοποίηση Δεδομένων .....	20
2.4.Ανάλυση Δεδομένων.....	22
2.5.Μοντέλα Μηχανικής Μάθησης.....	24
2.5.1.Bernoulli NB .....	24
2.5.2.Decision Tree Classifier .....	26
2.5.3.Gaussian NB .....	28
2.5.4.Gradient Boosting Classifier .....	29
2.5.5.KNeighbors Classifier .....	30
2.5.6.Logistic Regression.....	31
2.5.7.Neural Networks .....	32
2.5.8.Random Forest Classifier.....	34
2.5.9.Stochastic Gradient Descent.....	35
2.5.10.Support Vector Machine.....	37
2.6.Αξιολόγηση Μοντέλων Μηχανικής Μάθησης .....	37
2.6.1.Πίνακας Σύγκρισης.....	37
2.6.2.Μετρικές Αξιολόγησης .....	39

Κεφάλαιο 3 .....	40
Διαλειτουργικότητα Δεδομένων .....	40
3.1.Πρότυπα Διαλειτουργικότητας .....	40
3.1.1.ISO (International Organization for Standardization).....	40
3.1.2.W3C (World Wide Web Consortium).....	42
3.1.3.CEN (European Committee for Standardization) .....	43
3.2.Διαλειτουργικότητα στην Υγεία .....	44
3.3.Στάδια Προσέγγισης Υπηρεσιών Διαλειτουργικότητας .....	50
3.4.Πρότυπα Διαλειτουργικότητας στην Υγεία .....	52
3.4.1.SNOMED (Systematized Nomenclature of Medicine) .....	55
3.4.2.ICD-10 (International Classification of Diseases).....	56
3.4.3.HL7 v2 (Health Level Seven v2) .....	57
3.4.4.HL7 v3 (Health Level Seven v3) .....	57
3.4.5.HL7 FHIR (Health Level Seven Fast Healthcare Interoperability) .....	59
Κεφάλαιο 4 .....	62
Μελέτη Περίπτωσης.....	62
4.1.Αρχιτεκτονική Συστήματος .....	62
4.2.Σχετικές Τεχνολογίες.....	62
4.2.1. Υποστήριξη εφαρμογής μέσω Docker .....	65
4.3.Βάση Δεδομένων .....	68
4.4.Εγχειρίδιο Χρήσης Συστήματος.....	73
4.4.1. Unit Testing .....	92
4.4.2. Responsive Σχεδιασμός.....	95
Συμπεράσματα .....	99
Μελλοντικές Επεκτάσεις.....	100
Βιβλιογραφία .....	101

## Πίνακας Εικόνων

Εικόνα 1: Προετοιμασία Δεδομένων .....	17
Εικόνα 2: Εξαγωγή Χαρακτηριστικών με χρήση AutoML .....	19
Εικόνα 3: Εξαγωγή Χαρακτηριστικών .....	20
Εικόνα 4: Διαδικασία Οπτικοποίησης Δεδομένων .....	21
Εικόνα 5: Οπτικοποίηση Δεδομένων με την χρήση της "Python" .....	22
Εικόνα 6: Ανάλυση Δεδομένων .....	23
Εικόνα 7: Κατηγοριοποίηση Μοντέλων Μηχανικής Μάθησης .....	24
Εικόνα 8: Νευρωνικά Δίκτυα .....	32
Εικόνα 9: Αλγόριθμος SGD.....	35
Εικόνα 10: Πίνακας Σύγκρισης.....	39
Εικόνα 11: Αρχιτεκτονική Συστήματος "uphealth" .....	62
Εικόνα 12: Διεπαφή Χρήστη .....	63
Εικόνα 13: Υλοποίηση Backend Αρχιτεκτονικής.....	65
Εικόνα 14: Αρχιτεκτονική Dockerized Εφαρμογής .....	67
Εικόνα 15: Πίνακας Patient_Diabetes .....	69
Εικόνα 16: Πίνακας Patient_Covid.....	70
Εικόνα 17: Πίνακας Appointment_Diabetes.....	70
Εικόνα 18: Πίνακας Appointment_Covid.....	71
Εικόνα 19: Πίνακας DoctorInfo .....	72
Εικόνα 20: Πίνακας MIModel_Diabetes .....	72
Εικόνα 21: Πίνακας MIModel_Covid .....	73
Εικόνα 22: Κύρια Σελίδα .....	73
Εικόνα 23: Σελίδα Παροχής Υπηρεσιών.....	74
Εικόνα 24: Σελίδα Καινοτομιών .....	74
Εικόνα 25: Σελίδα Επικοινωνίας .....	75
Εικόνα 26: Σελίδα Επικοινωνίας (Επιτυχής Αποστολή) .....	75
Εικόνα 27: Σελίδα Log In .....	76
Εικόνα 28: Σελίδα Sign Up .....	76
Εικόνα 29: Σελίδα Log In (Στοιχεία Χρήστη) .....	77
Εικόνα 30: Σελίδα Επιλογής Κατηγορίας Ασθενών.....	77
Εικόνα 31: Σελίδα Αναζήτησης Διαβητικού Ασθενούς.....	78
Εικόνα 32: Αναζήτηση Ασθενούς "Vasilis".....	78
Εικόνα 33: Εκτεταμένος Φάκελος Υγείας Διαβητικού Ασθενούς.....	79
Εικόνα 34: Επιλογή Μοντέλου ML και Ορισμός Ραντεβού .....	79
Εικόνα 35: Επιλογή Μοντέλου ML Πρόβλεψης.....	80
Εικόνα 36: Σελίδα Ραντεβού (Διαβήτης) .....	80
Εικόνα 37: Σελίδα Ανάλυσης Δεδομένων (βήμα 1).....	81
Εικόνα 38: Εκτέλεση βήμα 1 .....	81
Εικόνα 39: Ανάλυση Δεδομένων (βήμα 2) .....	81
Εικόνα 40: Ανάλυση Δεδομένων (βήμα 3) .....	82
Εικόνα 41: Ανάλυση Δεδομένων (βήμα 4) .....	82



Εικόνα 42: Σελίδα Ανάλυσης Στατιστικών Στοιχείων ML Μοντέλων.....	83
Εικόνα 43: Σελίδα Ανάλυσης Στατιστικών Στοιχείων συγκεκριμένου ML Μοντέλου .....	83
Εικόνα 44: Σελίδα Data Profiling Διαβήτη .....	84
Εικόνα 45: Προσθήκη Διαβητικού Ασθενούς .....	84
Εικόνα 46: Προσθήκη Ραντεβού Διαβήτη .....	85
Εικόνα 47: Αναζήτηση Covid Ασθενούς.....	85
Εικόνα 48: Αναζήτηση Συγκεκριμένου Covid Ασθενούς.....	86
Εικόνα 49: Εκτεταμένος Φάκελος Υγείας Covid Ασθενούς .....	86
Εικόνα 50: Επιλογή μοντέλου ML Πρόβλεψης και Ορισμός Ραντεβού.....	86
Εικόνα 51: Πρόβλεψη εισαγωγής ασθενή στο νοσοκομείο βάσει συγκεκριμένου ML μοντέλου.....	87
Εικόνα 52: Στοιχεία Ραντεβού Covid Ασθενούς .....	87
Εικόνα 53: Προσθήκη Covid Ασθενούς.....	88
Εικόνα 54: Προσθήκη Ραντεβού Covid Ασθενούς.....	88
Εικόνα 55: Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 1) .....	89
Εικόνα 56: Σελίδα Ανάλυσης Δεδομένων Covid (εκτέλεση βήμα 1) .....	89
Εικόνα 57: Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 2) .....	90
Εικόνα 58: Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 3) .....	90
Εικόνα 59: Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 4) .....	90
Εικόνα 60: Σελίδα Ανάλυσης Στατιστικών Στοιχείων ML.....	91
Εικόνα 61: Σελίδα Ανάλυσης Στατιστικών Στοιχείων συγκεκριμένου ML .....	91
Εικόνα 62: Σελίδα Data Profiling Covid.....	92
Εικόνα 63: Έξοδος Συστήματος "uphealth" .....	92
Εικόνα 64: Unit Testing Angular.....	93
Εικόνα 65: Unit Testing HL7 FHIR.....	94
Εικόνα 66: Responsive Μενού Επιλογής Κατηγορίας Ασθενών .....	95
Εικόνα 67: Responsive Αναζήτηση Διαβητικού Ασθενούς .....	96
Εικόνα 68: Responsive Φάκελος Διαβητικού Ασθενούς .....	97
Εικόνα 69: Responsive Σελίδα Στατιστικών Μέτρων.....	98

## Κεφάλαιο 1

### Εισαγωγή

Στη σημερινή εποχή (εν έτη 2023) κρίνεται απαραίτητη η χρήση νέων τεχνολογιών στην καθημερινότητα. Αυτό συμβαίνει διότι την δεκαετία την οποία διανύουμε (2020-2030) έχει πολλαπλασιαστεί σε σημαντικό βαθμό όχι μόνο ο τεράστιος όγκος των δεδομένων που συνεχώς αυξάνεται αλλά και η πολυπλοκότητα των πληροφοριών που έρχονται προς επεξεργασία είτε μεμονωμένα είτε μαζικά από διάφορους οργανισμούς και επιχειρήσεις. Ωστόσο, η σημαντικότητα των νέων τεχνολογιών σε συνάρτηση με την βοήθεια που προσφέρουν διαδραματίζει σημαντικό ρόλο στην ταχύτητα λήψης και επεξεργασίας αυτών των πληροφοριών.

Επιπρόσθετα, στη σημερινή ψηφιακή εποχή κάθε οργανισμός χρειάζεται την τεχνολογία προκειμένου να αναπτυχθεί με απώτερο σκοπό την εξασφάλιση της απαραίτητης αξιοπιστίας, αποτελεσματικότητας αλλά και ευελιξίας για την εξυπηρέτηση των αναγκών των πελατών του. Επομένως, όπως εύκολα γίνεται αντιληπτό, η εξέλιξη αυτή έχει επηρεάσει και τον κλάδο της υγείας και ειδικότερα τον τομέα της υγειονομικής περίθαλψης των ασθενών σε κλινικές μονάδες και δημόσια νοσοκομεία. Όπως έχει παρατηρηθεί κατά καιρούς, υπάρχουν διάφορες μορφές ασυνέπειας του ιατρικού προσωπικού προς τους ασθενείς σε μηχανογραφικό επίπεδο ΗΦΥ. Μέχρι σήμερα δεν υπάρχει ένας σαφής ορισμός ή μια ξεκάθαρη άποψη για τα συστήματα ΗΦΥ. Στα συστήματα υγείας διαφόρων κρατών δεν υπάρχει ομοφωνία ως προς την έννοια του ΗΦΥ, με συνέπεια αυτός να αποδίδεται με διαφορετικές θεωρήσεις. Άλλοτε αποδίδεται ως αντίγραφο του χειρόγραφου φακέλου με διαδικασίες αυτόματης ανάγνωσης, άλλοτε ως αυτοματοποιημένος εργαστηριακός φάκελος Laboratory Medical Record (LMR) και άλλοτε ως ΗΦΥ.

Με χρήση των νέων τεχνολογιών, το ιατρικό προσωπικό θα έχει την άμεση πρόσβαση στα στοιχεία του κάθε ασθενή, ώστε να μπορεί να αξιοποιεί τα προσωπικά του στοιχεία καθ' όλη τη διάρκεια της νοσηλείας του με την χρήση του προτύπου HL7 FHIR για την διαλειτουργικότητα των δεδομένων και την ερμηνεία των δεδομένων με βάσει τους κανόνες που ορίζονται από το εν λόγω πρότυπο. Επίσης, μειώνεται μέσω των μηχανισμών AI που χρησιμοποιούν τα συστήματα, πολλά προβλήματα από αυτά που αντιμετωπίζει ο συγκεκριμένος κλάδος της υγείας.

Κλειδί για την εισαγωγή και την εφαρμογή των νέων τεχνολογιών, δεν είναι μόνο η προμήθεια των εξειδικευμένων προϊόντων πληροφορικής αλλά και η σωστή αξιοποίηση τους με σκοπό την πλήρη και παραγωγική λειτουργία τους. Τέλος αυτό, που θα πρέπει να λαμβάνεται υπόψη είναι η εναρμόνιση του ανθρώπινου δυναμικού-χρηστών των συστημάτων με την απαραίτητη τεχνική στελέχωση νέων μελών και η στενή συνεργασία των μελών της πληροφορικής σε συνεργασία με το ιατρικό προσωπικό κατά την ανάπτυξη των πληροφοριακών συστημάτων.

#### 1.1.Μηχανική Μάθηση

Η Μηχανική Μάθηση (ML) αποτελεί έναν τύπο της τεχνητής νοημοσύνης (AI), μέσω του οποίου παρέχεται στις εφαρμογές λογισμικού να εκτελούν πιο ακριβή αποτελέσματα στην πρόβλεψη των αποτελεσμάτων χωρίς να είναι να είναι ρητά προγραμματισμένες να το κάνουν. Οι αλγόριθμοι ML χρησιμοποιούν ιστορικά δεδομένα, όπως στην παρούσα περίπτωση ανάπτυξης της εφαρμογής διαχείρισης ασθενών, ως ιστορικά δεδομένα, ορίζεται ο ΗΦΥ του ασθενούς σε συνδυασμό με το

ιστορικό των εργαστηριακών εξετάσεών του. Έτσι, με τα δεδομένα αυτά να αποτελούν την είσοδο των απαραίτητων στοιχείων για την εκπαίδευση των αλγορίθμων, εξάγεται η τιμή ή οι τιμές πρόβλεψης που είναι και το αποτέλεσμα ή αλλιώς η έξοδος του εκάστοτε αλγορίθμου.

Η μηχανική μάθηση είναι σημαντική επειδή δίνει στους οργανισμούς μια άποψη για τις τάσεις στη συμπεριφορά των πελατών και τα επιχειρησιακά πρότυπα των επιχειρήσεων, ή όπως στην περίπτωση της παρούσας εργασίας δύναται μια πρόβλεψη σχετικά με την εξέλιξη και την πρόοδο του ασθενούς για την χορήγηση των απαραίτητων φαρμάκων και την καλύτερη παρακολούθηση από τον ειδικό στον χώρο της υγείας. Πολλές από τις κορυφαίες εταιρείες του σήμερα, όπως το Facebook, η Google και η Uber, κάνουν τη μηχανική μάθηση κεντρικό μέρος των εργασιών τους. Η μηχανική μάθηση έχει γίνει ένας σημαντικός ανταγωνιστικός παράγοντας διαφοροποίησης για πολλούς οργανισμούς.

Επίσης, υπάρχουν διάφοροι τύποι στην μηχανική μάθηση, μάθηση με επίβλεψη (supervised learning), μάθηση χωρίς επίβλεψη (unsupervised learning), ημι-επιτηρούμενη μάθηση (semi-supervised learning) και ενισχυτική μάθηση (reinforcement learning). Ο τύπος αλγορίθμου δεδομένων που επιλέγουν να χρησιμοποιήσουν οι επιστήμονες εξαρτάται από το είδος των δεδομένων που θέλουν να προβλέψουν [1].

Καθώς, η συνεχιζόμενη έρευνα για την συνεχή εξέλιξη του τομέα της μηχανικής μάθησης, παρατηρείται μια προσέγγιση ολοένα και περισσότερο στην ανάπτυξη πιο γενικών εφαρμογών με χρήση της μηχανικής μάθησης, προκειμένου αυτές οι λύσεις να είναι εύκολα εφαρμόσιμες σε μεγάλο πλήθος ανάπτυξης συστημάτων και προς την παροχή πιο άμεσων λύσεων. Τα σημερινά μοντέλα τεχνητής νοημοσύνης απαιτούν εκτενή εκπαίδευση προκειμένου να παραχθεί ένας αλγόριθμος που είναι εξαιρετικά βελτιστοποιημένος για την εκτέλεση μιας εργασίας [2]. Ωστόσο, ορισμένοι ερευνητές διερευνούν τρόπους για να κάνουν τα μοντέλα πιο ευέλικτα και αναζητούν τεχνικές που επιτρέπουν σε μια μηχανή να εφαρμόζει το πλαίσιο που έχει μάθει από μια εργασία σε μελλοντικές, διαφορετικές εργασίες.

## 1.2. Τομείς και Προκλήσεις Μηχανικής Μάθησης

Η μηχανική μάθηση παρέχει στους οργανισμούς τη γνώση ώστε να λαμβάνουν πιο ενημερωμένες, βασισμένες σε δεδομένα αποφάσεις που είναι ταχύτερες από τις παραδοσιακές προσεγγίσεις. Ωστόσο, δεν είναι η μυθική, μαγική διαδικασία που πολλοί την κατασκευάζουν. Η μηχανική μάθηση παρουσιάζει τις δικές της προκλήσεις. Ακολουθούν πέντε κοινά προβλήματα μηχανικής μάθησης σε συνδυασμό με την επίλυση των διάφορων θεμάτων που τίθενται. Αυτά είναι τα παρακάτω:

- **Κατανόηση διαδικασιών αυτοματισμού:** Γίνεται όλο και πιο δύσκολος ο διαχωρισμός των γεγονότων από τη φαντασία όσον αφορά τη μηχανική μάθηση σήμερα. Προτού αποφασίσετε ποια πλατφόρμα τεχνητής νοημοσύνης θα χρησιμοποιήσετε, πρέπει να αξιολογήσετε ποια προβλήματα επιδιώκετε να επιλύσετε. Οι πιο εύκολες διαδικασίες για αυτοματοποίηση είναι αυτές που γίνονται χειροκίνητα κάθε μέρα χωρίς μεταβλητή έξοδο. Οι περίπλοκες διαδικασίες απαιτούν περαιτέρω έλεγχο πριν από την αυτοματοποίηση. Ενώ η μηχανική μάθηση μπορεί σίγουρα να βοηθήσει στην αυτοματοποίηση ορισμένων

διαδικασιών, δεν χρειάζονται ωστόσο όλα τα προβλήματα αυτοματισμού τη μηχανική μάθηση.

- **Έλλειψη ποιοτικών δεδομένων:** Το κυριότερο πρόβλημα που αντιμετωπίζει η μηχανική μάθηση είναι η έλλειψη καλών δεδομένων. Ενώ η βελτίωση των αλγορίθμων καταναλώνει συχνά τον περισσότερο χρόνο των προγραμματιστών στο AI, η ποιότητα των δεδομένων είναι απαραίτητη για να λειτουργήσουν οι αλγόριθμοι όπως προβλέπεται. Τα θορυβώδη δεδομένα και τα ελλιπή δεδομένα είναι οι βασικοί εχθροί της ιδανικής μηχανικής μάθησης. Η λύση σε αυτό το αίνιγμα είναι να αφιερώσετε χρόνο για την αξιολόγηση και την εμπέδωση δεδομένων με σχολαστική διακυβέρνηση δεδομένων, ενοποίηση δεδομένων και εξερεύνηση δεδομένων μέχρι να λάβετε σαφή δεδομένα. Οπότε, κάτι τέτοιο αποτελεί μια εξορισμού ενέργεια κατά την έναρξη εκπαίδευσης των αλγορίθμων ML.
- **Ανεπαρκής υποδομή:** Η μηχανική μάθηση απαιτεί τεράστιες ποσότητες δυνατοτήτων των δεδομένων. Τα παλαιού τύπου συστήματα συχνά δεν μπορούν να χειριστούν τον φόρτο εργασίας και λειτουργούν υπό πίεση. Θα πρέπει να ελέγξετε εάν η υποδομή σας μπορεί να χειριστεί τη μηχανική μάθηση. Εάν δεν μπορεί, θα πρέπει να αναζητήσετε αναβάθμιση, με επιτάχυνση υλικού και ευέλικτο χώρο αποθήκευσης. Σαφώς, όλα τα παραπάνω που αναφέρονται πρέπει να ορίζονται πριν την έναρξη όλης αυτής της απαιτητικής διαδικασίας.
- **Εφαρμογή:** Οι οργανισμοί έχουν συχνά μηχανές ανάλυσης που συνεργάζονται μαζί τους μέχρι να επιλέξουν να κάνουν αναβάθμιση στη μηχανική μάθηση. Η ενσωμάτωση νεότερων μεθοδολογιών ML σε υπάρχουσες μεθοδολογίες είναι μια περίπλοκη εργασία. Η διατήρηση της σωστής ερμηνείας και τεκμηρίωσης συμβάλλει σε μεγάλο βαθμό στη διευκόλυνση της εφαρμογής. Η συνεργασία με έναν κατάλληλο συνεργάτη μπορεί να κάνει την υλοποίηση υπηρεσιών όπως η ανίχνευση ανωμαλιών, η προγνωστική ανάλυση και την μοντελοποίηση του συνόλου πολύ πιο εύκολη.

### 1.3. Διαλειτουργικότητα Δεδομένων

Η διαλειτουργικότητα δεδομένων υγειονομικής περίθαλψης είναι μια μέθοδος για τα συστήματα προκειμένου να παρέχεται προς τους χρήστες για πρόσβαση, κοινή χρήση, ενοποίηση, ερμηνεία και συνεργασία για τη χρήση διαφορετικών συστημάτων πληροφοριών, εφαρμογών και συσκευών με συντονισμένο τρόπο.

Η διαλειτουργικότητα εντός των συστημάτων υγειονομικής περίθαλψης μπορεί να χρησιμοποιηθεί εντός και πέρα από τα περιφερειακά, εθνικά και οργανωτικά όρια για να επιτευχθεί διαχρονική και απρόσκοπτη φορητότητα πληροφοριών. Βοηθά στη βελτιστοποίηση της υγείας των ασθενών παγκοσμίως. Με τη διαλειτουργικότητα δεδομένων υγειονομικής περίθαλψης, μπορείτε να έχετε πρόσβαση σε διεπαφές εφαρμογών, πρότυπα και αρχιτεκτονική ανταλλαγής δεδομένων και να μοιράζεστε δεδομένα με ασφάλεια σε ένα πλήρες φάσμα της υγειονομικής περίθαλψης με τους σχετικούς ενδιαφερόμενους φορείς.

Τα δεδομένα υγείας ήταν ανέκαθεν δύσκολη η πρόσβαση και η κοινή χρήση με ασφαλή τρόπο. Η φύση των δεδομένων υγείας δημιουργεί ένα παράδοξο: είναι δύσκολο να μοιραστούν επειδή είναι ευαίσθητα και απαιτεί υψηλό επίπεδο απορρήτου και ασφάλειας, ωστόσο η αδυναμία πρόσβασης σε αυτά όταν χρειάζεται μπορεί να προκαλέσει σημαντική βλάβη. Η έλλειψη

διαλειτουργικότητας μπορεί να οδηγήσει σε ελλιπή κατανόηση των αναγκών υγείας ενός ατόμου ή του πληθυσμού, γεγονός που μπορεί να οδηγήσει σε χειρότερα αποτελέσματα και υψηλότερο κόστος.

Γι' αυτό λοιπόν στην παρούσα εργασία έχει επιλεγεί το πρότυπο HL7 FHIR προκειμένου να επιτευχθεί η διαλειτουργικότητα των δεδομένων που ανταλλάσσονται μεταξύ του ΗΦΥ του εκάστοτε ασθενούς σε συνάρτηση με τις εργαστηριακές του εξετάσεις. Αυτό επιτυγχάνεται μέσω μοντέλων (models) και πόρων (resources) που εντοπίζουν την κατάλληλη πληροφορία και την επικοινωνούν είτε εντός ενός συστήματος είτε με εξωτερικά συστήματα, προκειμένου να γίνεται κατανοητή η ίδια πληροφορία προς όλα τα συστήματα και να υπάρχει η σωστή και ακριβή ερμηνεία των δεδομένων από όλους τους συμμετέχοντες.

#### 1.4. Τομείς και Προκλήσεις Διαλειτουργικότητας Δεδομένων

Ωστόσο, παρά τις γενικές πρωτοβουλίες παγκοσμίως, εξακολουθούν να υπάρχουν πολλές προκλήσεις, όπως η έλλειψη ενοποιήσεων συστημάτων, η ασύνδετη επιβολή των προτύπων διαλειτουργικότητας, η δυσκολία επιβολής προτύπων μέτρησης διαλειτουργικότητας σε έναν κλάδο, ο συντονισμός των ενδιαφερομένων σε ολόκληρο τον κλάδο και ο αποκλεισμός πληροφοριών [3]. Τα οφέλη της διαλειτουργικότητας είναι σαφή, με τους ηγέτες της υγειονομικής περίθαλψης να ζητούν από τον κλάδο να προσαρμοστεί σε αυτές τις προκλήσεις. Υπάρχουν εκατοντάδες πιστοποιημένα προϊόντα ΗΦΥ στην αγορά, το καθένα με τις δικές του κλινικές ορολογίες, τεχνικές προδιαγραφές και δυνατότητες. Αυτές οι διαφορές καθιστούν δύσκολη τη δημιουργία μιας τυπικής μορφής διαλειτουργικότητας για την κοινή χρήση δεδομένων μεταξύ οργανισμών και ιατρικού λογισμικού. Ακόμη και τα συστήματα ΗΦΥ που είναι κατασκευασμένα στην ίδια πλατφόρμα ενδέχεται να μην είναι διαλειτουργικά, καθώς συνήθως προσαρμόζονται στις ανάγκες ενός οργανισμού.

Οπότε παρακάτω αποτυπώνονται τα κύρια χαρακτηριστικά που αποτελούν πρόκληση στον χώρο της διαλειτουργικότητας των δεδομένων και αυτά έχουν ως εξής:

- Έλλειψη ενσωμάτωσης δεδομένων
- Ασυνεχής δομή πληροφορίας στον τομέα της υγείας
- Δυσκολία ενσωμάτωσης προτύπων διαλειτουργικότητας
- Προκλήσεις συντονισμού των ενδιαφερόμενων μερών σε ολόκληρο τον κλάδο
- Δυσκολία ανάκτησης πληροφοριών

#### 1.5. Αντίκειμενο Διπλωματικής Εργασίας

Σκοπός της παρούσας εργασίας είναι:

1. Να μελετήσει και να αναπτύξει τη διεθνή βιβλιογραφία σχετικά με τα πληροφοριακά συστήματα υγείας, αλλά και πιο συγκεκριμένα την υγειονομική περίθαλψη μέσω του ΗΦΥ. Πιο ειδικά, θα γίνει μελέτη και αναφορά στον ΗΦΥ, που συμπεριλαμβάνει τόσο κλινικά όσο και τα εργαστηριακά δεδομένα.

2. Να μελετήσει τη διεθνή βιβλιογραφία σχετικά με τη μηχανική μάθηση και την διαλειτουργικότητα των δεδομένων και μέσα από όλη αυτή τη βιβλιογραφία να αναπτυχθεί ένα πληροφοριακό σύστημα διαχείρισης ΗΦΥ για ασθενείς (διαβητικούς και covid-19) με τη χρήση νέων τεχνολογιών. Η εργασία αυτή παρουσιάζει πως ένα σύστημα υγείας καθώς και οι ιατρικές διαδικασίες μπορούν να αναβαθμιστούν σύμφωνα με τις εξελίξεις της τεχνολογίας.

### 1.6. Δομή Διπλωματικής

Η εργασία διαρθρώνεται σε τέσσερα μέρη. Το πρώτο μέρος αφορά την εισαγωγή για τον σκοπό διάρθρωσης της παρούσας εργασίας. Το δεύτερο μέρος αφορά την ανασκόπηση και μελέτη της διεθνούς βιβλιογραφίας σχετικά με τα πληροφοριακά συστήματα, τους ηλεκτρονικούς φακέλους υγείας, την μελέτη αλγορίθμων μηχανικής μάθησης για εξατομικευμένη φροντίδα στην παροχή των υπηρεσιών που προσφέρονται μέσω των ηλεκτρονικών φακέλων, καθώς και την επικοινωνία των δεδομένων προς διαχείριση μέσω ενός κοινού προτύπου επίτευξης διαλειτουργικότητας των δεδομένων.

Το τρίτο μέρος της εργασίας αναφέρει την αρχιτεκτονική στην οποία βασίστηκε η ανάπτυξη του πληροφοριακού συστήματος «**up-health**». Επίσης θα αναφερθούν τόσο οι τεχνολογίες που εφαρμόστηκαν τόσο στο «**frontend**» όσο και οι τεχνολογίες που εφαρμόστηκαν στο «**backend**».

Το τέταρτο και τελευταίο μέρος της εργασίας παρουσιάζει το πληροφοριακό σύστημα «**up-health**» με τις διάφορες διαδικασίες και ροές εργασιών που εκτελούνται από έναν ειδικό στον κλάδο της υγείας.

Επίσης, το πληροφοριακό σύστημα «**up-health**», μπορείτε να το βρείτε στο ηλεκτρονικό αποθετήριο του GitHub, κατόπιν συνεννόησης με τον υπεύθυνο της παρούσας εργασίας.

## Κεφάλαιο 2

### Μηχανική Μάθηση

Η υγειονομική περίθαλψη είναι ένας ευρύς όρος που αφορά ένα σύστημα που περιλαμβάνει τη βελτίωση των ιατρικών υπηρεσιών προκειμένου να εξυπηρετήσει τις ιατρικές απαιτήσεις των ανθρώπων. Στην υγειονομική περίθαλψη, καταβάλλονται προσπάθειες από ασθενείς, γιατρούς, πωλητές, εταιρείες υγείας και εταιρείες πληροφορικής για τη συντήρηση και την αποκατάσταση αρχείων υγείας. Κατά την τελευταία δεκαετία, η ινδική υγειονομική περίθαλψη είναι γνωστή ως μια από τις ταχέως αναπτυσσόμενες βιομηχανίες στον κόσμο [4]. Η ανάλυση της υγειονομικής περίθαλψης χειρίζεται διάφορους τύπους ασθενειών, όπως καρκίνο, διαβήτη, εγκεφαλικά επεισόδια και άλλες χρόνιες ασθένειες με τη χρήση μηχανικής μάθησης. Ο διαβήτης είναι μία από τις πιο θανατηφόρες ασθένειες.

Για τον χειρισμό του διαβήτη στην ανάλυση της υγειονομικής περίθαλψης, υπάρχει ανάγκη δημιουργίας σωστών και ποιοτικών δεδομένων. Σε αυτόν τον ανταγωνιστικό κόσμο, η υγειονομική περίθαλψη πρέπει να χρησιμοποιεί τα δεδομένα με τέτοιο τρόπο ώστε να υπάρχει πάντα αύξηση της ποιότητας της υγειονομικής περίθαλψης και μείωση του κόστους που απαιτείται για τον σκοπό της θεραπείας.

Τα τελευταία χρόνια, η έρευνα σχετικά με την υποστήριξη της υγειονομικής περίθαλψης με την χρήση της μηχανικής μάθησης αυξάνεται σταθερά. Λόγω της ποικιλίας των ιατρικών δεδομένων, συμπεριλαμβανομένων των κλινικών δεδομένων, των δεδομένων omics (γονιδιωματικά δεδομένα) ή των δεδομένων ΗΦΥ, είναι δύσκολο για τους ανθρώπους να εξάγουν συμπεράσματα για τα δεδομένα και να λάβουν αποφάσεις. Κατά συνέπεια, το ML σε συνδυασμό με την διαλειτουργικότητα δεδομένων έχουν προταθεί στην υγειονομική περίθαλψη για καλύτερη κατανόηση των δεδομένων και για καλύτερη διαδικασία λήψης αποφάσεων.

Επιπλέον, η δυσaréσκεια των κλινικών για τα συστήματα πληροφοριών περίθαλψης υποδηλώνει ότι πολλές προσπάθειες στον ηλεκτρονικό υπολογιστή των πληροφοριών υγειονομικής περίθαλψης δεν έχουν εκπληρώσει τις δυνατότητές τους, τουλάχιστον στον τομέα της περίθαλψης των ασθενών. Σε πρόσφατη έρευνα, τα ποσοστά χρήσης έχουν αναφερθεί σε ποσοστό άνω του 90%, αλλά η δυσaréσκεια για τον αντίκτυπο των συστημάτων πληροφοριών περίθαλψης στη ροή εργασίας και την απόδοση των ασθενών παραμένει υψηλή [5].

Η τρέχουσα προσέγγιση για την ανάπτυξη και το σχεδιασμό των συστημάτων πληροφοριών υγειονομικής περίθαλψης καθοδηγείται από μια λογική, τεχνοκρατική άποψη που αντικαθιστά την κρίση του σχεδιαστή ενός πληροφοριακού συστήματος με την κρίση του ιατρού. Οι σχεδιαστικές λύσεις που προκύπτουν από αυτό το ορθολογικό μοντέλο βασίζονται σε μεγάλο βαθμό σε υπερβολικά απλουστευμένους κανόνες υποστήριξης αποφάσεων που διαταράσσουν τις κλινικές ροές εργασίας και σε άκαμπτα πρότυπα και τυποποιημένα πρωτόκολλα που δεν ανταποκρίνονται στις πολύπλοκες και δυναμικές προκλήσεις που θέτουν τα διάφορα θέματα υγείας.

Τα ακόλουθα συστήματα δεν καταφέρνουν να λάβουν υπόψη τις απαιτήσεις της συλλογικής κλινικής εργασίας στην οποία οι επαγγελματίες του τομέα της υγειονομικής περίθαλψης πρέπει να συνειδητοποιήσουν και να δώσουν προτεραιότητα στη γνώση για να αντιμετωπίσουν πολλαπλές δυνατότητες ροής εργασίας και μη συνθήκες συνθήκες. Αυτό το χάσμα μεταξύ της

πραγματικότητας της κλινικής εργασίας και του τρόπου εξ ορθολογισμού του σχεδιασμού τεχνολογίας πληροφοριών οδηγεί σε ανάπτυξη συστημάτων που προκαλούν λάθη και δεν υποστηρίζουν σωστά την εργασία όπως προορίζεται.

Το Care Records Service είναι ένα ηλεκτρονικό αρχείο υγειονομικής περίθαλψης που αναπτύχθηκε για τη διαχείριση ιατρικών αρχείων για όλους τους ασθενείς [6]. Το γενικό όραμα για το Care Records Service εξελίχθηκε από την αναμφισβήτητη πεποίθηση ότι η υγειονομική περίθαλψη πρέπει να επωφεληθεί από τις νέες τεχνολογικές εξελίξεις. Η καλύτερη φροντίδα των ασθενών θα προέκυπτε φυσικά από την ανάπτυξη τεχνολογιών αυτοματισμού που θα εκτελούσαν συγκεκριμένα καθήκοντα υγειονομικής περίθαλψης και από μια σειρά τεχνολογιών που θα επέτρεπαν στους επαγγελματίες του τομέα της υγείας να έχουν πρόσβαση σε πρόσφατα ενημερωμένες πληροφορίες για τους ασθενείς [7].

Επίσης, οι υπεύθυνοι χάραξης πολιτικής υπέθεσαν ότι η μεταφορά εργασιών σε αυτοματοποίηση και η άμεση πρόσβαση σε τρέχουσες πληροφορίες για τον ασθενή θα βελτίωνε τις κλινικές αποφάσεις μέσω πιο αποτελεσματικών διαβουλεύσεων και, ως εκ τούτου, θα οδηγούσε σε μια πιο αποτελεσματική και πιο οικονομική φροντίδα, ενώ θα μείωνε τον κίνδυνο εσφαλμένων φαρμακευτικών αγωγών.

## 2.1. Προετοιμασία Δεδομένων

Διάφορες επιστημονικές μελέτες ενσωματώνουν μια μεγάλη ποικιλία και μεγάλο όγκο πηγών, τα οποία είναι τα κύρια χαρακτηριστικά των μεγάλων δεδομένων (Big Data). Εκτός από τις παραδοσιακές ιατρικές εγκαταστάσεις όπως νοσοκομεία, ασφαλιστικοί οργανισμοί, κέντρα φυσικοθεραπείας, γηροκομεία, η ανάπτυξη του διαδικτύου των πραγμάτων (Internet of Things - IoT) σε ιατρικές εγκαταστάσεις, θα εμπλουτίσει την ποικιλομορφία των δεδομένων που προετοιμάστηκαν, ιδίως για τις χρόνιες ασθένειες.

Ωστόσο, παρόλο που υπάρχει μεγάλο όφελος από τη διεξαγωγή πολυκεντρικών αναλυτικών δεδομένων Big Data και ενώ υπάρχουν πολλές διαθέσιμες πηγές δεδομένων, οι απαιτήσεις της ακριβούς απόκτησης κλινικών δεδομένων μέσω του συντονισμού και της συνεργασίας των πηγών δεδομένων εγκαίρως, κρίνεται απαραίτητη. Προβλέπεται ότι χωρίς προετοιμασία δεδομένων και χωρίς δυνατότητα ποιότητας και απορρήτου, η ποιότητα των δεδομένων που προετοιμάζονται για έρευνα αποτελεί μείζονα ανησυχία [8]. Παρά τις εκτιμήσεις ότι η κακή ποιότητα δεδομένων κοστίζει στον τομέα της υγειονομικής περίθαλψης δαπανώνται 3,1 τρισεκατομμύρια δολάρια ετησίως [9]. Οι περισσότεροι υπάρχοντες αλγόριθμοι εξόρυξης δεδομένων λειτουργούν με τις υποθέσεις ότι όλα τα δεδομένα είναι υψηλής ποιότητας και κάθε χαρακτηριστικό είναι σημαντικό, παραβλέποντας συχνά τα κρυμμένα πρότυπα και τους παράγοντες κινδύνου ζωτικής σημασίας για μακρο-αναλυτική ανάλυση δεδομένων. Λόγω της ευρείας χρήσης τους, οι ΗΦΥ αποτελούν τη βέλτιστη πηγή δεδομένων για μελέτες παρατήρησης και έχουν ήδη εφαρμοστεί για την αναγνώριση προτύπων.





*Εικόνα 1: Προετοιμασία Δεδομένων*

Δυστυχώς, η εγκυρότητα αυτών των μελετών παρατήρησης μπορεί να τεθεί υπό αμφισβήτηση ανάλογα με τη συνάφεια των ΗΦΥ με τις προδιαγραφές του ιατρικού ερευνητή. Το αποτέλεσμα της έρευνας βασίζεται αποκλειστικά στην ποιότητα των δεδομένων που προετοιμάζονται για ανάλυση, επομένως καθίσταται επιτακτική ανάγκη κάθε ΗΦΥ που διατίθεται τοπικά σε μια πηγή δεδομένων να επαληθεύεται, συχνά με μη αυτόματο τρόπο στον τομέα της υγείας, για να τηρεί τις παραμέτρους πέρα από ένα σημαντικό όριο προκειμένου να διασφαλιστεί ότι η μελέτη δεν διακυβεύεται. Αυτό έχει ως αποτέλεσμα, η προετοιμασία δεδομένων μελετών παρατήρησης των ΗΦΥ, από τοπικά αποθετήρια να είναι μια επίπονη και εκτεταμένη προσπάθεια που απαιτεί πολύ χρόνο (Εικόνα 1). Στην πραγματικότητα, το 40-60% της διάρκειας των μελετών αφιερώνεται επί του παρόντος αποκλειστικά στην προετοιμασία δεδομένων, δηλαδή στην ανάκτηση σχετικών ΗΦΥ από κάθε συμβαλλόμενο οργανισμό για ανάλυση [10].

Επιπλέον, οι λύσεις που χρησιμοποιούνται πιο πολύ όπως το RedCap [11] και το MDAnalyze [12], στοχεύουν στη διευκόλυνση της μη αυτόματης αναθεώρησης των μεγάλων βάσεων δεδομένων ΗΦΥ από ειδικούς στον τομέα για την ανάκτηση των ΗΦΥ και για την προετοιμασία δεδομένων είναι ανεπαρκείς. Δεδομένου ότι μόνο λίγες, εύκολες στην εκτέλεση εργασίες αυτοματοποιούνται από αυτά τα προγράμματα, αυτές οι προσεγγίσεις που βασίζονται στο λογισμικό δεν ελαχιστοποιούν τον χρόνο και τις δαπάνες προετοιμασίας δεδομένων σε σχέση με μια εντελώς χειροκίνητη προσέγγιση.

Λόγω αυτών των προκλήσεων, οι γιατροί περιορίζονται ήδη στο να ξοδεύουν μόνο το 27% του χρόνου τους για να βλέπουν ασθενείς και η χρήση ημι-αυτοματοποιημένων προσεγγίσεων για την ολοκλήρωση κλινικών και εργαστηριακών εξετάσεων είναι πολύ απίθανο να αυξήσει την αλληλεπίδραση των ασθενών με τους γιατρούς.

Επιπλέον, υπάρχει μεγάλη ανησυχία για την προστασία των δεδομένων των ασθενών. Οι οδηγίες για το απόρρητο των ασθενών γίνονται όλο και πιο αυστηρές καθώς οι υποδομές για πρόσβαση σε αρχεία υγείας επεκτείνουν την εμβέλειά τους [13]. Οι νόμοι για τη φορητότητα και την ευθύνη για την ασφάλιση υγείας (HIPPA), τους κανονισμούς του θεσμικού οργάνου ελέγχου (IRB), της εποπτικής αρχής (SA) και τα περιφερειακά πρωτόκολλα απορρήτου θα δυσκολεύουν τις μεμονωμένες ιατρικές εγκαταστάσεις να συνεισφέρουν τα δεδομένα τους σε μια πολυκεντρική μελέτη λόγω των αυστηρών υγειονομικών πρωτοκόλλων [13]. Όσον αφορά τις κύριες αρχιτεκτονικές για την προετοιμασία κλινικών δεδομένων, το απόρρητο των ασθενών δεν μπορεί να εξαιρεθεί αποτελεσματικά λόγω διαφορετικών τοπικών προτύπων. Επιπλέον, καθώς περισσότερα ιατρικά αποθετήρια δεδομένων αποφασίζουν να συνεισφέρουν σε μια μελέτη,

απαιτείται επίσης ο συντονισμός και η συνεργασία μεταξύ ιατρικών πηγών δεδομένων για την ενσωμάτωση δεδομένων.

Η είσοδος και η έξοδος των ιατρικών πηγών δεδομένων πρέπει επίσης να διασφαλιστεί ώστε να μην σταματήσει ή ακόμη και να επηρεάσει το πρωτόκολλο συνεισφοράς σε άλλα αποθετήρια που επιχειρούν διανομή ΗΦΥ. Η συγκέντρωση των δεδομένων ΗΦΥ, πρέπει να αποφεύγεται ώστε να διατηρείται μια μελέτη επεκτάσιμη και σε εξέλιξη. Απαιτείται επίσης ο συντονισμός και η συνεργασία μεταξύ ιατρικών πηγών δεδομένων για την ενσωμάτωση των δεδομένων.

Οι προσεγγίσεις για την αυτοματοποίηση της προετοιμασίας δεδομένων ΗΦΥ για μελέτες παρατήρησης σχετικές με την προληπτική και εξατομικευμένη ιατρική πρέπει, όχι μόνο να διευκολύνουν την ανεξαρτησία στη λήψη αποφάσεων των ιατρικών οργανισμών που συνεισφέρουν κλινικά δεδομένα στην πολυκεντρική έρευνα, αλλά και να μετρά επακριβώς τη συνάφεια για τη διατήρηση της ποιότητας των δεδομένων που επιστρέφονται.

Στην παρούσα εργασία, η προετοιμασία των δεδομένων γίνεται με την χρήση της γλώσσας προγραμματισμού Python [14][14]. Οι τιμές που λείπουν είναι δεδομένα συνηθισμένων περιστατικών και αν δεν αντιμετωπιστούν στο σύνολο δεδομένων εκπαίδευσης, μπορεί να μειώσει την απόδοση προσαρμογής του μοντέλου ή μπορεί να οδηγήσει σε ένα υπό/υπέρ-προσαρμογής μοντέλο. Μπορεί επίσης, να οδηγήσει σε λάθος πρόβλεψη ή ταξινόμηση. Η τιμή που λείπει σημαίνει πολλά διαφορετικά πράγματα στα δεδομένα. Συνεπώς, στα δεδομένα που απουσιάζουν οι τιμές τους προστίθεται ο μέσος όρος κατά προσέγγιση [15]. Έτσι, λοιπόν εξασφαλίζεται, η αποφυγή κενών ή λανθασμένων τιμών.

## 2.2.Εξαγωγή Χαρακτηριστικών

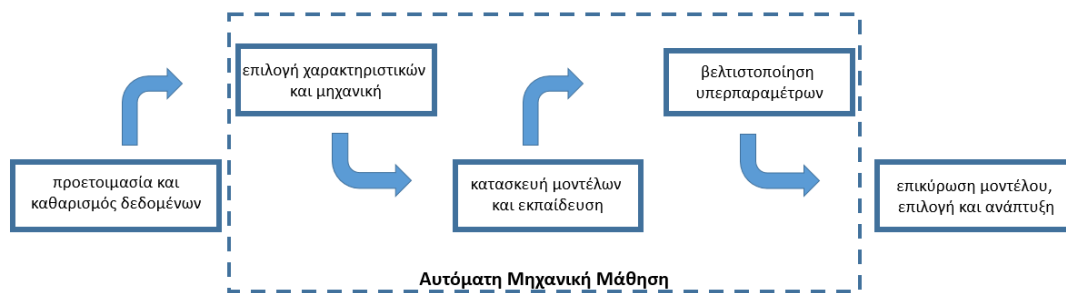
Η εκτεταμένη συλλογή δεδομένων για την υγεία μέσω ηλεκτρονικών αρχείων υγείας, γονιδιωματικής αλληλούχισης και ψηφιακών φορέων υγείας έχει οδηγήσει σε μια εκθετικά αυξανόμενη ποσότητα βιοϊατρικών «μεγάλων δεδομένων». Ο αριθμός των ψηφιακών πληροφοριών που διατίθενται στους κλινικούς ιατρούς γίνεται υπερβολικά μεγάλος για επεξεργασία: εντός του χρονικού διαστήματος των 20–40 λεπτών που αντιστοιχούν γενικά ανά επίσκεψη, είναι σχεδόν αδύνατο να αναθεωρηθούν 80+ megabyte (ισοδύναμο με 20.000+ σελίδες ελεύθερου κειμένου) με δεδομένα των ασθενών που καταγράφηκαν στο μέσο ατομικό ΗΦΥ [16].

Η μηχανική μάθηση, και πιο συγκεκριμένα η βαθιά μάθηση (Deep Learning - DL), είναι βασικές τεχνικές που έχουν δείξει την ικανότητα να μεταφράζονται αυτά τα μεγάλα σύνολα δεδομένων υγείας σε ενεργή γνώση. Γενικά, η χρήση μοντέλων μηχανικής μάθησης θα μπορούσε να βελτιώσει την ασφάλεια των ασθενών, την ποιότητα της περίθαλψης και να μειώσει το κόστος της υγειονομικής περίθαλψης. Συγκεκριμένα, η μηχανική μάθηση έχει τη δυνατότητα να αυξήσει το έργο των κλινικών για την επεξεργασία των δισεκατομμυρίων σημείων δεδομένων ασθενούς που είναι αποθηκευμένα σε ΗΦΥ και έχει ήδη εφαρμοστεί επιτυχώς σε πολλές κλινικές εφαρμογές, όπως η αναγνώριση ασθενών με υψηλό κίνδυνο μετακίνησης, διάγνωση αναπνευστικών καταστάσεων από ακτινογραφίες θώρακα και ανίχνευση πρώιμων σημάτων καρκίνου του πνεύμονα [16]. Ενώ η μηχανική μάθηση στην υγειονομική περίθαλψη είναι ένα πολύ ενεργό ερευνητικό θέμα, τα περισσότερα από τα δεδομένα υγείας που συλλέγονται δεν χρησιμοποιούνται ποτέ για τη δημιουργία μοντέλων πρόβλεψης που ενσωματώνονται διαδοχικά στο κλινικό

περιβάλλον με μόνο το 15% των νοσοκομείων επί του παρόντος να χρησιμοποιούν μηχανική μάθηση για περιορισμένους σκοπούς [16].

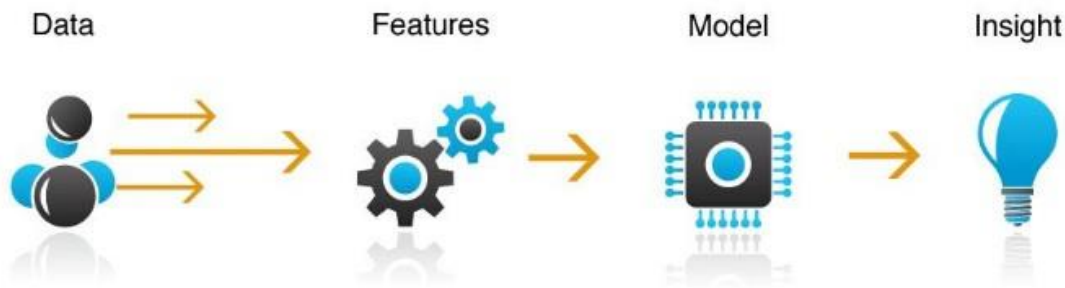
Επιπλέον, ενώ η μηχανική μάθηση έχει πολλά αποδεδειγμένα οφέλη, η επιτυχής χρήση της μηχανικής μάθησης απαιτεί μεγάλη προσπάθεια από ειδικούς ανθρώπους, δεδομένου ότι κανένας αλγόριθμος δεν μπορεί να επιτύχει καλή απόδοση σε όλα τα πιθανά προβλήματα. Παρόλο που οι ερευνητές υγειονομικής περίθαλψης είναι εξοικειωμένοι με τα κλινικά δεδομένα, εξακολουθούν να στερούνται συχνά τεχνογνωσία που απαιτείται για την εφαρμογή αυτών των τεχνικών σε μεγάλες πηγές δεδομένων. Οι ερευνητές υγειονομικής περίθαλψης μπορούν και συνεργάζονται με ειδικούς επιστήμονες δεδομένων, αλλά η διαδραστική διαδικασία απαιτεί γενικά πολύ χρόνο και προσπάθεια και από τα δύο μέρη. Όχι μόνο αυτό, αλλά γενικά τα δεδομένα και η ανθρώπινη εμπειρογνωμοσύνη δεν είναι άμεσα διαθέσιμα, ειδικά σε χώρους υγειονομικής περίθαλψης.

Επομένως, είναι δύσκολο να γίνει επινόηση και ανάπτυξη λύσεων μηχανικής μάθησης, καθώς ολόκληρη η έρευνα ξεκινά με μια μακρά διαδικασία παροχής δεδομένων, συνεχίζει με την εύρεση των σωστών συνεργατών και συνεπάγεται συνεχή εναλλαγή μεταξύ εμπειρογνομόνων ML και εμπειρογνομόνων στον κλάδο της υγείας. Η αυτοματοποίηση ορισμένων συστατικών που απαιτούν ανθρώπινη εμπειρογνωμοσύνη θα επέτρεπε στον κλάδο της υγειονομικής περίθαλψης να δημιουργήσει, να επικυρώσει και να αναπτύξει ταχύτερα λύσεις μηχανικής μάθησης και επομένως να αποκομίσει ευκολότερα τα οφέλη από τη βελτίωση της ποιότητας της υγειονομικής περίθαλψης για τους ασθενείς. Με κίνητρο αυτόν τον στόχο σε όλους τους κλάδους, η αυτοματοποιημένη μηχανική μάθηση (AutoML) έχει αναδειχθεί ως ένα νέο πεδίο έρευνας με στόχο την αυτόματη βελτιστοποίηση τμημάτων της μηχανικής μάθησης, όπως φαίνεται (Εικόνα 2).



Εικόνα 2: Εξαγωγή Χαρακτηριστικών με χρήση AutoML

Διάφορες λύσεις AutoML έχουν εμφανιστεί τα τελευταία χρόνια για τη βελτιστοποίηση ενός ή περισσοτέρων από αυτά τα στοιχεία, πολλά από τα οποία είναι προϊόν των διαγωνισμών AutoML Challenge μεταξύ 2015 και 2018 [17].



Εικόνα 3: Εξαγωγή Χαρακτηριστικών

Στην παρούσα εργασία, η διαδικασία της εξαγωγής χαρακτηριστικών (Εικόνα 3), πραγματοποιείται μέσω των κατάλληλων βιβλιοθηκών που διαθέτει η Python και οι περιπτώσεις που διακρίνονται είναι:

- **Επιλογή χαρακτηριστικών:** Όλες οι λειτουργίες δεν είναι ίδιες. Έχει να κάνει με την επιλογή ενός μικρού υποσυνόλου χαρακτηριστικών από μια μεγάλη ομάδα χαρακτηριστικών. Επιλέγονται εκείνα τα χαρακτηριστικά που εξηγούν καλύτερα τη σχέση μιας ανεξάρτητης μεταβλητής με τη μεταβλητή στόχου [18]. Υπάρχουν ορισμένα χαρακτηριστικά που είναι πιο σημαντικά από άλλα χαρακτηριστικά στην ακρίβεια του μοντέλου. Διαφέρει από τη μείωση της διαστατικότητας γιατί η μέθοδος μείωσης διαστάσεων το κάνει συνδυάζοντας τα υπάρχοντα χαρακτηριστικά, ενώ η μέθοδος επιλογής χαρακτηριστικών περιλαμβάνει ή εξαιρεί αυτά τα χαρακτηριστικά. Οι μέθοδοι επιλογής χαρακτηριστικών είναι η δοκιμή Chi-squared, οι βαθμολογίες του συντελεστή συσχέτισης, το LASSO, η παλινδρόμηση Ridge κ.ά [18].
- **Μετασχηματισμός χαρακτηριστικών:** Σημαίνει τη μετατροπή της αρχικής δυνατότητας στις λειτουργίες των πρωτογενών χαρακτηριστικών. Η κλιμάκωση, η διακριτοποίηση, η εκκαθάριση και η συμπλήρωση των τιμών δεδομένων που λείπουν είναι οι πιο κοινές μορφές μετασχηματισμού δεδομένων. Για να μειωθεί η σωστή ασυμμετρία των δεδομένων, χρησιμοποιείται το αρχείο καταγραφής [19].
- **Εξαγωγή χαρακτηριστικών:** Όταν τα δεδομένα για επεξεργασία μέσω αλγορίθμου είναι πολύ μεγάλα, θεωρούνται γενικά περιττά. Η ανάλυση με μεγάλο αριθμό μεταβλητών χρησιμοποιεί μεγάλη ισχύ υπολογισμού και μνήμη, επομένως θα πρέπει να μειωθεί η διάσταση αυτών των τύπων μεταβλητών. Είναι ένας όρος για την κατασκευή συνδυασμών των μεταβλητών. Για δεδομένα πίνακα, χρησιμοποιείται Principal Component Analysis (PCA) για να μειωθούν οι διαστάσεις [20].

### 2.3.Οπτικοποίηση Δεδομένων

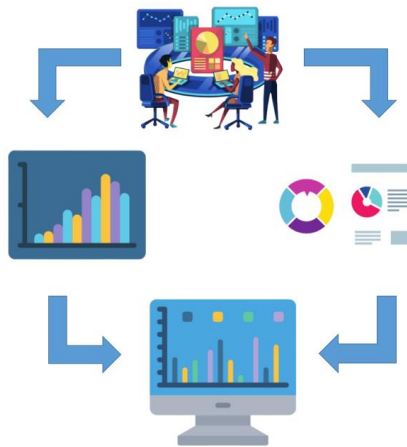
Έχουν αναπτυχθεί πολλές πλατφόρμες που εκτελούν οπτική ανάλυση σε δεδομένα υγειονομικής περίθαλψης ενός τύπου.

Η διαδικτυακή πλατφόρμα Ανί'ο [21] αναπτύχθηκε για να αναλύσει και να οπτικοποιήσει τα δεδομένα omics. Η πλατφόρμα εκτελεί μη εποπτευόμενη ανάλυση των ανεπεξέργαστων δεδομένων, ενώ προσφέρει τη δυνατότητα εποπτευόμενης ανάλυσης μέσω της διαδραστικής

οπτικοποίησης δεδομένων. Τα δεδομένα χρονικού συμβάντος παράγονται από καθημερινά κλινικά χαρακτηριστικά και δίκτυα αισθητήρων. Το DecisionFlow [22] παρέχει ένα πλαίσιο για οπτικές αναλύσεις σε δεδομένα διαστάσεων χρονικής εκδήλωσης υψηλής διάστασης.

Το Tableau, το Spotfire, το Google Charts, το Cubism.js και το Visual Sedimentation είναι καθιερωμένες πλατφόρμες οπτικής ανάλυσης ικανές να αναλύουν ένα ευρύ φάσμα δεδομένων [23]. Επιπλέον, το σύστημα Progressive Insights [23] προσφέρει προοδευτικά αναλυτικά στοιχεία για κάθε τύπο δεδομένων. Η μεθοδολογία ακολουθεί μια προσέγγιση δυναμική δεδομένου ότι παράγει και εμφανίζει ουσιαστικά αποτελέσματα. Επίσης, παρέχει τη δυνατότητα να παρέμβει στον αλγόριθμο και την εκ νέου ρύθμιση των παραμέτρων.

Οι αρχιτεκτονικές που βασίζονται στη ροή εργασίας έχουν ήδη χρησιμοποιηθεί σε υπάρχουσες πλατφόρμες ανάλυσης ή οπτικοποίησης. Το Orange [24] είναι μια αυτόνομη πλατφόρμα που επιτρέπει το σχεδιασμό μιας ροής εργασίας ανάλυσης και οπτικοποίησης, μέσω ενός διασθητικού γραφικού περιβάλλοντος χρήστη. Στοιχεία της ροής εργασίας, όπως μέθοδοι ανάλυσης και οπτικοποιήσεις, μπορούν να μεταφερθούν στην οθόνη, συνδέοντας την έξοδο ενός στοιχείου με την είσοδο ενός άλλου. Ο χρήστης μπορεί να διαμορφώσει κάθε στοιχείο δυναμικά και να δει άμεσα την επίδραση στα αποτελέσματα. Η οπτικοποίηση μέσω του Vega [24] επιτρέπει το σχεδιασμό προσαρμοσμένων οπτικοποιήσεων και αλληλεπιδράσεων μέσω αρχιτεκτονικής που βασίζεται σε σήματα. Η Vega επικεντρώνεται στην παροχή μηχανισμών για διαδραστική οπτικοποίηση, με λίγες μόνο λειτουργίες ανάλυσης δεδομένων.



Εικόνα 4: Διαδικασία Οπτικοποίησης Δεδομένων

Για εφαρμογές γενικού σκοπού, όπως το ReactiveX [24], υποστηρίζουν τη δημιουργία διαδικτυακών ροών εργασίας σε πραγματικό χρόνο, μέσω του ορισμού και της διασύνδεσης των ροών δεδομένων (Εικόνα 4). Τα στοιχεία μπορούν να ωθήσουν τιμές σε μια ροή δεδομένων, η οποία ακούγεται από άλλα στοιχεία. Τυχόν νέες τιμές που προστίθενται στη ροή μεταδίδονται αυτόματα στα στοιχεία που συνδέονται με αυτήν.

Το Healthvision [24] είναι μια διαδικτυακή πλατφόρμα ανάλυσης δεδομένων και διαδραστικής οπτικοποίησης που συνδυάζει τα χαρακτηριστικά των προαναφερόμενων πλατφορμών που επιτρέπουν το σχεδιασμό προσαρμοσμένων ροών εργασιών ανάλυσης δεδομένων και

οπτικοποίησης. Αυτή η μεθοδολογία μπορεί να διευκολύνει τη διαδικασία ανάλυσης δεδομένων για διαφορετικούς τύπους επιστημονικών ερωτημάτων (π.χ. μπορεί να χρησιμοποιηθεί για τον εντοπισμό της συσχέτισης μεταξύ των χαρακτηριστικών στα κλινικά αρχεία των ασθενών ή για την αναγνώριση των βιοδεικτών σε γενετικά δεδομένα). Η πλατφόρμα υποστηρίζει διαδραστική διάδοση αλλαγών τόσο σε στοιχεία οπτικοποίησης όσο και σε εργασίες ανάλυσης δεδομένων, ενώ παρέχει μια φιλική διεπαφή χρήστη για την προσαρμογή της ροής εργασίας των δεδομένων.

Η Python προσφέρει πολλές υπέροχες βιβλιοθήκες γραφικών που συνοδεύονται από πολλά διαφορετικά χαρακτηριστικά. Ανεξάρτητα από το εάν θέλετε να δημιουργήσετε διαδραστικά, ζωντανά ή πολύ προσαρμοσμένα γραφήματα, η Python διαθέτει εξαιρετικές βιβλιοθήκες (Εικόνα 5).



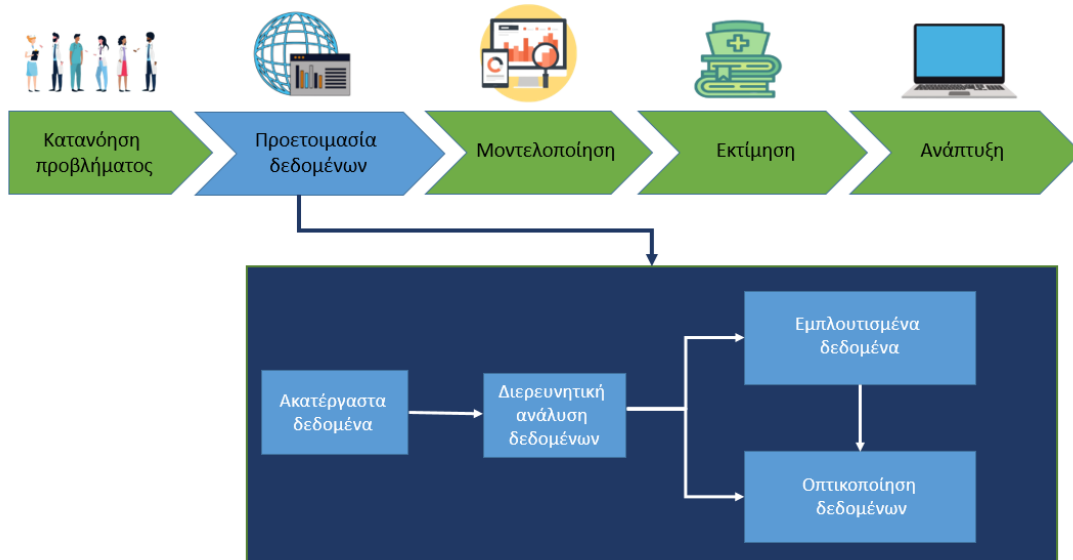
Εικόνα 5: Οπτικοποίηση Δεδομένων με την χρήση της "Python"

Για να δείτε μια μικρή επισκόπηση, εδώ είναι μερικές δημοφιλείς βιβλιοθήκες σχεδίασης, οι οποίες χρησιμοποιήθηκαν στα πλαίσια του συστήματος που παρουσιάζεται στην εν λόγω εργασία είναι:

- **[Matplotlib](#)**: Παροχή μεγάλης ελευθερίας ανάπτυξης κώδικα για στατιστικά στοιχεία
- **[Pandas Visualization](#)**: Εύχρηστη διεπαφή, βασισμένη στο Matplotlib
- **[Seaborn](#)**: Διεπαφή υψηλού επιπέδου, εξαιρετικά προεπιλεγμένα στυλ
- **[ggplot](#)**: Βάσει το ggplot2 του R, χρησιμοποιεί ειδικά γραφήματα
- **[Plotly](#)**: Δημιουργία διαδραστικών γραφημάτων

## 2.4.Ανάλυση Δεδομένων

Η Εξερευνητική Ανάλυση Δεδομένων (Exploratory Data Analysis - EDA) είναι ένα σημαντικό αρχικό βήμα για οποιαδήποτε διαδικασία ανακάλυψης γνώσης, στην οποία οι επιστήμονες δεδομένων διερευνούν διαδραστικά άγνωστα σύνολα δεδομένων εκδίδοντας μια ακολουθία λειτουργιών ανάλυσης (π.χ. φίλτρο, συγκέντρωση και οπτικοποίηση). Δεδομένου ότι το EDA είναι από καιρό γνωστό ως ένα δύσκολο έργο, που απαιτεί βαθιά αναλυτικές δεξιότητες, εμπειρία και γνώση τομέα, μια πληθώρα συστημάτων έχουν επινοηθεί την τελευταία δεκαετία προκειμένου να διευκολυνθεί το EDA [25].



Εικόνα 6: Ανάλυση Δεδομένων

Συγκεκριμένα, οι εξελίξεις στην έρευνα μηχανικής μάθησης δημιούργησαν συναρπαστικές ευκαιρίες, όχι μόνο για την καλύτερη διευκόλυνση του EDA, αλλά και για την πλήρη αυτοματοποίηση της διαδικασίας (Εικόνα 6). Ξεκινώντας από συστήματα σύστασης για την πρόταση μιας και μόνο διερευνητικής δράσης, περνώντας από ταξινομητές που βασίζονται σε kNN (k-Nearest Neighbors) και μεθόδους μάθησης για την πρόβλεψη των προτιμήσεων των ενδιαφερόντων των χρηστών και, τέλος, μέχρι την πλήρη αυτοματοποίηση του EDA χρησιμοποιώντας υπερσύγχρονες μεθόδους, όπως το Deep Learning [25].

Στην εξόρυξη δεδομένων, το EDA είναι μια προσέγγιση για την ανάλυση συνόλων δεδομένων για να συνοψίσει τα κύρια χαρακτηριστικά τους, συχνά με οπτικές μεθόδους. Το EDA χρησιμοποιείται για να δει τι μπορούν να πουν τα δεδομένα πριν από την εργασία μοντελοποίησης. Δεν είναι εύκολο να κοιτάξετε μια στήλη αριθμών ή ένα ολόκληρο υπολογιστικό φύλλο και να προσδιορίσετε σημαντικά χαρακτηριστικά των δεδομένων. Μπορεί να είναι κουραστικό ή βαρετό να αντλούνται πληροφορίες, κοιτώντας απλούς αριθμούς. Εξερευνητικές τεχνικές ανάλυσης δεδομένων έχουν επινοηθεί ως βοήθημα σε αυτήν την κατάσταση [25].

Οπότε, το EDA είναι ένα κρίσιμο βήμα που πρέπει να ακολουθείται πριν ξεκινήσετε τη μηχανική μάθηση ή τη στατιστική μοντελοποίηση, επειδή παρέχει το πλαίσιο που απαιτείται για την ανάπτυξη ενός κατάλληλου μοντέλου για το πρόβλημα που αντιμετωπίζετε και για την ορθή ερμηνεία των αποτελεσμάτων του. Το EDA είναι πολύτιμο για τον επιστήμονα δεδομένων για να βεβαιωθεί ότι τα αποτελέσματα που παράγουν είναι έγκυρα, σωστά ερμηνευμένα και εφαρμόσιμα στα επιθυμητά επιχειρηματικά περιβάλλοντα.

## 2.5. Μοντέλα Μηχανικής Μάθησης

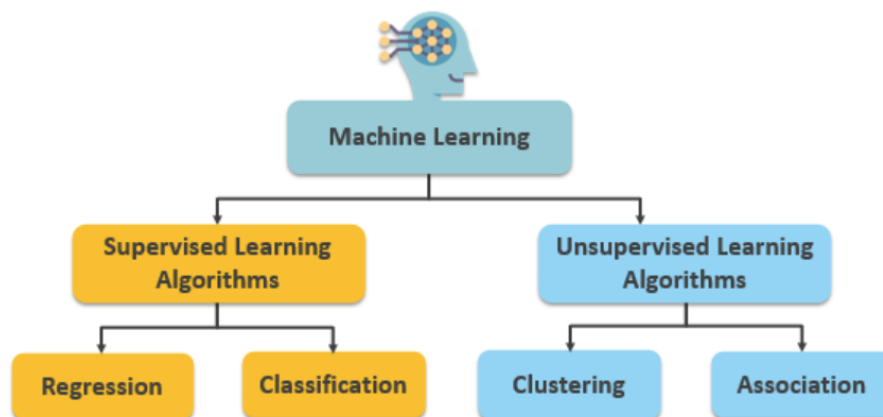
Σε γενικές γραμμές, υπάρχουν 2 τύποι αλγορίθμων μηχανικής εκμάθησης (Εικόνα 7):

### 1. Εποπτευόμενη μάθηση (Supervised Learning)

Αυτός ο αλγόριθμος αποτελείται από μια μεταβλητή στόχου/αποτελέσματος (ή εξαρτώμενη μεταβλητή) η οποία πρέπει να προβλεφθεί από ένα δεδομένο σύνολο προγνωστικών (ανεξάρτητες μεταβλητές). Χρησιμοποιώντας αυτά τα σύνολα μεταβλητών, δημιουργείται μια συνάρτηση που αντιστοιχίζει τις εισόδους στις επιθυμητές εξόδους. Η διαδικασία εκπαίδευσης συνεχίζεται έως ότου το μοντέλο επιτύχει ένα επιθυμητό επίπεδο ακρίβειας στα δεδομένα εκπαίδευσης. Παραδείγματα εποπτευόμενης μάθησης, αποτελούν η παλινδρόμηση, ο Decision Tree, ο Random Forest, ο KNN, ο Logistic Regression κ.ά [26].

### 2. Μη εποπτευόμενη μάθηση (Unsupervised Learning)

Σε αυτόν τον αλγόριθμο, δεν υπάρχει καμία μεταβλητή στόχου ή αποτελέσματος για την πρόβλεψη/εκτίμηση. Χρησιμοποιείται για τη συγκέντρωση πληθυσμού σε διαφορετικές ομάδες, το οποίο χρησιμοποιείται ευρέως για την τμηματοποίηση πελατών σε διαφορετικές ομάδες για συγκεκριμένη παρέμβαση. Παραδείγματα μη εποπτευόμενης μάθησης αποτελούν ο αλγόριθμος Apriori, K-mean κ.ά [26].



Εικόνα 7: Κατηγοριοποίηση Μοντέλων Μηχανικής Μάθησης

#### 2.5.1. Bernoulli NB

Ένα χρήσιμο μοντέλο του Bayes είναι το Bernoulli Naive Bayes (BernoulliNB). Αυτή η μέθοδος χρησιμοποιείται για διακριτά δεδομένα και βασίζεται στην κατανομή Bernoulli. Το κύριο χαρακτηριστικό του Bernoulli Naive Bayes είναι ότι λαμβάνει μόνο δυαδικές τιμές όπως true ή false, ναι ή όχι, επιτυχία ή αποτυχία, 0 ή 1 κ.ο.κ. Το `sklearn.naive_bayes` είναι μια λειτουργική μονάδα Scikit-learn που χρησιμοποιείται στην γλώσσα προγραμματισμού Python.

Επειδή έχουμε να κάνουμε με δυαδικούς αριθμούς, θεωρείται το 'p' ως πιθανότητα επιτυχίας και το 'q' ως την πιθανότητα αποτυχίας, με  $q=1-p$ . Επίσης, στην κατανομή Bernoulli, με μια τυχαία



μεταβλητή  $X$ , όπως φαίνεται στην παρακάτω εξίσωση (Εδώ το ' $x$ ' μπορεί να έχει μόνο δύο τιμές: 0 ή 1):

$$P(x) = P[X = x] = \begin{cases} q = 1 - p & x = 0 \\ p & x = 1 \end{cases} \quad (1)$$

Ο κανόνας για τον ταξινομητή Bernoulli Naive Bayes φαίνεται στην παρακάτω εξίσωση:

$$P(x_i / y) = p^{(i/y)} x_i + (1 - p^{(i/y)}) (1 - x_i) \quad (2)$$

Χρησιμοποιούμε αυτόν τον αλγόριθμο επειδή η ακρίβεια πρόβλεψης σε περίπτωση δεδομένων δυαδικής ταξινόμησης, καθώς έχουμε δύο προβλεπόμενες κατηγορίες που είναι είτε ο ασθενής πάσχει από διαβήτη είτε δεν πάσχει από αυτόν, αυτός ο αλγόριθμος είναι αποτελεσματικός για χρήση για αυτόν τον τύπο δεδομένων, επιπλέον έχουν διαφορετικές παρατηρήσεις που σημαίνει ότι οι παρατηρήσεις είναι μοναδικές αλλά ορισμένα χαρακτηριστικά σχετίζονται με το αποτέλεσμα. Αυτός ο αλγόριθμος χρησιμοποιεί τεχνική εκ των υστέρων πιθανοτήτων για να προβλέψει την έξοδο.

Οι Geetha και Prasad [27] πρότειναν μια θεωρία που επικεντρώθηκε στη δημιουργία υβριδικών μοντέλων που είναι εύκολο για τους γιατρούς να θεραπεύσουν τους διαβητικούς ασθενείς. Έχουν χρησιμοποιήσει μόνο δύο αλγόριθμους ML, τον Random Forest και τον Naive Bayes. Οι μετρήσεις απόδοσης που χρησιμοποιούνται για την ταξινόμηση είναι μετρήσεις σύγχυσης. Συγκρίνουν τα αποτελέσματα απόδοσης και των δύο αλγορίθμων. Οι Shetty et al. [28] πρότεινε ένα σύστημα πρόβλεψης διαβήτη το οποίο είναι έξυπνο και δίνει ανάλυση του διαβήτη χρησιμοποιώντας εξόρυξη δεδομένων [29]. Σε αυτό το σύστημα, χρησιμοποιούν αλγόριθμους όπως ο KNN και ο Naive Bayes για να εφαρμόσουν σε σύνολα δεδομένων ασθενών με διαβήτη και να τα αναλύσουν λαμβάνοντας μερικά ή όλα τα χαρακτηριστικά του συνόλου δεδομένων και ως εκ τούτου να προβλέψουν την ασθένεια [30]. Sisodia [31] σχεδίασε ένα μοντέλο που μπορεί να προβλέψει την ασθένεια με μέγιστη ακρίβεια. Έχουν χρησιμοποιήσει τρεις αλγόριθμους ταξινόμησης, δηλαδή το Decision Tree, το SVM και το Naive Bayes. Τα εργαλεία που χρησιμοποιούνται είναι το WEKA και η ταξινόμηση γίνεται με διαχωρισμό των δεδομένων στο 70% του συνόλου δεδομένων για εκπαίδευση. Επαλήθευσαν τα αποτελέσματα χρησιμοποιώντας την καμπύλη ROC. Οι Zeki et al. [32] πρότεινε ένα σύστημα δήλωσης ελέγχου που βασίζεται σε κανόνες. Έχουν χρησιμοποιήσει τρεις ενότητες για τρία στάδια, δηλαδή μπλοκ διάγραμμα, πίνακες αποφάσεων και Mockler Charts. Μετά από εξέταση πολλών παραγόντων και μεταβλητών, αυτό το σύστημα παρέχει μια διάγνωση διαβητικών [33].

Kumar et al. [34] εφάρμοσε ένα σύστημα χρησιμοποιώντας την τεχνική Hadoop και MapReduce για την επεξεργασία και την ανάλυση των δεδομένων του διαβήτη. Αυτό το μοντέλο προβλέπει επίσης τον τύπο του διαβήτη. Το σύστημά τους βασίζεται στο Hadoop και παρέχει έναν συστηματικό τρόπο αντιμετώπισης των ασθενών με διαθεσιμότητα και οικονομικά προσιτή τιμή. Οι Nnamoko et al. [35] χρησιμοποίησε πέντε ευρέως χρησιμοποιούμενους ταξινομητές που χρησιμοποιούνται για τα σύνολα και επίσης συγκεντρώνουν τα αποτελέσματα με τη βοήθεια μετα-ταξινομητών. Τα αποτελέσματα εξάγονται και συγκρίνονται με άλλες μελέτες που χρησιμοποιούν τα ίδια σύνολα δεδομένων. Αποδεικνύεται ότι με τη βοήθεια της προτεινόμενης μεθόδου, ο διαβήτης μπορεί να προβλεφθεί στην αρχική φάση με μεγαλύτερη ακρίβεια.

Οι Furqan Rustam et al. [36] έχουν αναπτύξει ένα μοντέλο που χρησιμοποιεί εποπτευόμενη μηχανική μάθηση για μελλοντικές προβλέψεις κρουσμάτων Covid-19. Προβλέπουν τον αριθμό των μελλοντικών ασθενών που θα επηρεαστούν από τον Covid-19. Η πρόβλεψη γίνεται με βάση τον αριθμό των πρόσφατα μολυσμένων κρουσμάτων, τον αριθμό των νεκρών και τις αναρρώσεις. Για την πρόβλεψη έχουν χρησιμοποιηθεί μοντέλα μάθησης όπως τα LR, SVM και LASSO. Από την άλλη πλευρά, οι Bastos et al. [37] έχουν αναπτύξει ένα μοντέλο πρόβλεψης χρησιμοποιώντας δεδομένα της Νότιας Αμερικής από τις ημερομηνίες 25 Φεβρουαρίου – 30 Μαρτίου 2020. Χτίστηκε γύρω από την παράμετρο της κοινωνικής απόστασης. Τα αποτελέσματα δείχνουν ότι η κοινωνική απόσταση θα μειώσει την εξάπλωση της μόλυνσης. Ωστόσο, εάν η κοινωνική απόσταση δεν εφαρμοστεί σωστά, η εξάπλωση της μόλυνσης θα αρχίσει να κερδίζει.

Akib Mohi Ud Din Khanday et al. [38] έχουν προτείνει ένα μοντέλο για την πρόβλεψη των κρουσμάτων Covid-19. Η πρόβλεψη γίνεται με βάση την ηλικία, το φύλο, το ποσοστό θανάτου στο νοσοκομείο και τα συμπτώματα των ασθενών. Οι προσεγγίσεις παλινδρόμησης για ανάλυση εκπαιδεύονται και ελέγχονται σε δεδομένα σε πραγματικό χρόνο χρησιμοποιώντας τα δεδομένα που δίνονται για τα αντίστοιχα.

Ένα άλλο μοντέλο εφαρμόζεται από τους Dehning et al. [39] που είναι συμπέρασμα Bayes με επιδημιολογικές παραμέτρους για την παρατήρηση της χρονικής εξάρτησης της αύξησης της νέας μεταδοτικής μόλυνσης του Covid-19 στη Γερμανία. Οι συγγραφείς ισχυρίστηκαν ότι το μοντέλο είναι πλήρως λειτουργικό και μπορεί να προσαρμοστεί από οποιαδήποτε περιοχή στον κόσμο.

### 2.5.2. Decision Tree Classifier

Ένα δέντρο αποφάσεων (Decision Tree Classifier) είναι ένας χάρτης των πιθανών αποτελεσμάτων μιας σειράς σχετικών επιλογών. Επιτρέπει σε ένα άτομο ή έναν οργανισμό να σταθμίζει πιθανές ενέργειες μεταξύ τους με βάση το κόστος, τις πιθανότητες και τα οφέλη του. Μπορούν να χρησιμοποιηθούν για την χαρτογράφηση ενός αλγορίθμου που προβλέπει μαθηματικά την καλύτερη επιλογή.

Ένα δέντρο αποφάσεων ξεκινά συνήθως με έναν μόνο κόμβο, ο οποίος διαιρείται σε πιθανά αποτελέσματα. Κάθε ένα από αυτά τα αποτελέσματα οδηγεί σε επιπλέον κόμβους, οι οποίοι διαχωρίζονται σε άλλες δυνατότητες.

Υπάρχουν τρεις διαφορετικοί τύποι κόμβων: κόμβοι πιθανότητας, κόμβοι αποφάσεων και κόμβοι τερματισμού. Ένας κόμβος πιθανότητας, που αντιπροσωπεύεται από έναν κύκλο, δείχνει τις πιθανότητες ορισμένων αποτελεσμάτων. Ένας κόμβος απόφασης, που αντιπροσωπεύεται από ένα τετράγωνο, δείχνει μια απόφαση που πρέπει να ληφθεί και ένας τελικός κόμβος δείχνει το τελικό αποτέλεσμα μιας διαδρομής αποφάσεων.

Η συμπεριφορά των ριζών και των κόμβων εξαρτάται από τις διαφορετικές μεθόδους:

1. CART: Αναφέρεται ως "Δέντρο ταξινόμησης και παλινδρόμησης". Χρησιμοποιείται για την ταξινόμηση των δυαδικών δεδομένων. Για την επιλογή των ριζών από τα χαρακτηριστικά δεδομένων, υπολογίζει τον δείκτη gini για κάθε μέθοδο χαρακτηριστικών, το χαρακτηριστικό που περιέχει τη μικρότερη τιμή gini που θα επιλεγεί ως ρίζα και, στη συνέχεια, τα χαρακτηριστικά δεδομένων θα διανεμηθούν για περαιτέρω υπορίζα.

2. ID3: Είναι μια άλλη μέθοδος, η οποία χρησιμοποιεί το κέρδος στη θέση του δείκτη gini, σε αυτήν την περίπτωση όσο μεγαλύτερη είναι η τιμή του κέρδους αυτό το χαρακτηριστικό θα επιλεγεί ως χαρακτηριστικό ρίζας. Αυτή η μέθοδος έχει επίσης ορισμένα μειονεκτήματα, η λήψη του δέντρου αποφάσεων με αυτήν τη μέθοδο δεν συναντά την τιμή που λείπει και μόνο ένα χαρακτηριστικό είναι υπεύθυνο για τη λήψη αποφάσεων.
3. C4.5: Αυτός ο αλγόριθμος είναι βασικά μια προσθήκη του ID3 που σημαίνει ότι χειρίζεται την τιμή που λείπει και χρησιμοποιείται επίσης για κατηγορικά δεδομένα.

Τα τελευταία χρόνια, πολλές μέθοδοι έχουν προταθεί και δημοσιευθεί για την πρόβλεψη του διαβήτη. Ένα πλαίσιο βασισμένο σε ML προτάθηκε στο [40] όπου οι συγγραφείς εφάρμοσαν τη Linear Discriminant Analysis (LDA) [41], την Qualitative Data Analysis (QDA) [42], την Naive Bayes (NB) [43], την Gauss Processing Classification (GPC) [44], Support Vector Machine (SVM) [45], Artificial Neural Network (ANN) [46], AdaBoost (AB) [47], Logistic Regression (LR) [48], Decision Tree (DT) [49] και Random Forest (RF) [50] με διαφορετικές τεχνικές μείωσης διαστάσεων και διασταυρούμενης επικύρωσης. Πραγματοποίησαν επίσης εκτεταμένα πειράματα σχετικά με την απόρριψη ακραίων τιμών και την συμπλήρωση των χαμένων τιμών για την ενίσχυση της απόδοσης του μοντέλου ML, όπου κατάφεραν να λάβουν την υψηλότερη δυνατή AUC των 0,930. Στο [51], οι συγγραφείς χρησιμοποίησαν τρεις διαφορετικούς ταξινομητές ML όπως DT, SVM και NB για να προβλέψουν την πιθανότητα διαβήτη με μέγιστη ακρίβεια. Έδειξαν ότι το NB είναι το μοντέλο με τις καλύτερες επιδόσεις με AUC 0,819. Οι τεχνικές AB και bagging ensemble χρησιμοποιώντας J48 (~4.5), ως βασικός εκπαιδευόμενος και ως αυτόνομη τεχνική εξόρυξης δεδομένων (J48), έχουν μελετηθεί και εφαρμοστεί στο [52] για την ταξινόμηση του σακχαρώδους διαβήτη. Τα πειραματικά αποτελέσματα τους αποδεικνύουν ότι η μέθοδος AB ensemble είναι καλύτερη από το bagging και αυτόνομη J48.

Ο γενετικός προγραμματισμός για την πρόβλεψη του διαβήτη είχε προταθεί στο [53] όπου το πλαίσιο είχε καλύτερες επιδόσεις σε σύγκριση με άλλες τεχνικές που εφαρμόστηκαν από αυτούς. Οι συγγραφείς στο [54], χρησιμοποίησαν τέσσερις μεθόδους ML, όπως DT, ANN, LR και NB για να ταξινομήσουν τον κίνδυνο σακχαρώδους διαβήτη, όπου ενίσχυσαν την ευρωστία με τεχνικές ενίσχυσης. Τα πειραματικά αποτελέσματα δείχνουν ότι ο αλγόριθμος RF δίνει τα βέλτιστα αποτελέσματα μεταξύ όλων των χρησιμοποιούμενων αλγορίθμων. Αν και υπάρχουν πολλά πλαίσια που έχουν ήδη δημοσιευθεί, τα τελευταία χρόνια, ωστόσο, η βελτίωση απαιτεί την ακρίβεια και την ευρωστία για την πρόβλεψη του διαβήτη.

Υπάρχουν πολλές εργασίες που αφορούν τον αλγόριθμο δέντρου αποφάσεων ID3 και τις βελτιώσεις τους όπως οι Es-Sabery et al. [55], Οι Yu-Xun et al. [56], Chai et al. [57], Elyassami et al. [58], Zou et al. [59], Srinivasan et al. [60], Chen et al. [61], Ding et al. [62], Zhu et al. [63], Kaewrod et al. [64], Rajeshkanna et al. [65]. Στο [55], προτείνεται ένα νέος βελτιωμένος αλγόριθμος δέντρου αποφάσεων ID3, ο οποίος ενσωματώνει τη σταθμισμένη θεωρία και το κριτήριο κέρδους πληροφοριών. Στη διαδικασία διαιρούμενης μάθησης, ο βελτιωμένος αλγόριθμος εξετάζει τη συσχέτιση μεταξύ του άμεσου χαρακτηριστικού υπό όρους, του χαρακτηριστικού απόφασης και των άλλων υπό όρους χαρακτηριστικών για να υπολογίσει το σταθμισμένο κέρδος πληροφοριών και, στη συνέχεια, να βρει το καλύτερο χαρακτηριστικό διαχωρισμού. Σε αντίθεση με το παραδοσιακό ID3, λαμβάνεται υπόψη μόνο τη συσχέτιση μεταξύ του τρέχοντος χαρακτηριστικού και του χαρακτηριστικού απόφασης για τον υπολογισμό του κέρδους πληροφοριών. Αυτός ο βελτιωμένος προτεινόμενος αλγόριθμος δενδρικής απόφασης

ID3 χρησιμοποιείται σε αυτήν την εργασία ως ταξινομητής. Elyassami et al. [58], οι συγγραφείς εφάρμοσαν τον αλγόριθμο Fuzzy ID3 για την πρόβλεψη τιμών που υλοποιείται συνδυάζοντας τις βασικές έννοιες του αλγορίθμου ID3 και τις αρχές της ασαφούς θεωρίας συνόλων παρέχοντας στο μοντέλο τη δυνατότητα να αντιμετωπίζει αβέβια και διαφορούμενα δεδομένα, τα οποία μπορούν να βελτιώσουν σημαντικά το ποσοστό ταξινόμησης. Οι Zou et al. [59] εφάρμοσαν την απόφαση του αλγορίθμου ID3 για την κατασκευή του μοντέλου ανίχνευσης. Η χρήση του ID3 απέδειξε ότι αυτός ο αλγόριθμος μπορούσε να εκτελέσει πολύ καλά την ταξινόμηση δεδομένων και παρέχει στους λήπτες αποφάσεων ένα σύνολο κανόνων απόφασης. Οι Srinivasan et al. [60] σχεδίασε τη μέθοδο ταξινόμησης Fast Fuzzy για καλύτερη απόδοση της ταξινόμησης. Έχουν επίσης ενσωματώσει τα πλεονεκτήματα του δέντρου αποφάσεων ID3 και του αλγορίθμου SVM, για τη βελτίωση της ακρίβειας και για τη λήψη ενός γρήγορου αποτελέσματος ταξινόμησης.

### 2.5.3. Gaussian NB

Ο Gaussian Naïve Bayes (GaussianNB) είναι ένας από τους πιο εποικοδομητικούς αλγόριθμους, βασίζεται στο θεώρημα bayes που εξαρτώνται από την πιθανοτική θεωρία. Είναι εποπτευόμενος αλγόριθμος μηχανικής μάθησης. Το αποτέλεσμα οποιασδήποτε παρατήρησης εξαρτάται από την προηγούμενη γνώση του συνόλου δεδομένων. Το μεγαλύτερο πλεονέκτημα αυτού του αλγορίθμου είναι ότι λαμβάνει υπόψη το αποτέλεσμα ενός συγκεκριμένου χαρακτηριστικού που είναι ανεξάρτητο από άλλα χαρακτηριστικά. Παρακάτω δίνεται η εξίσωση που εφαρμόζεται για την εξαγωγή του τελικού αποτελέσματος ως εξής:

$$P\left(\frac{m}{n}\right) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{\frac{-(m_i - \mu_n)^2}{2\sigma_n^2}} \quad (3)$$

Οι Giri και Todmal [66] πρότειναν ένα σύστημα για την πρόβλεψη του διαβήτη χρησιμοποιώντας τη νέα προσέγγιση που αποτελείται από δύο στάδια. Στο πρώτο στάδιο χρησιμοποιήθηκε η συνάρτηση Gauss για τη διανομή των δεδομένων και στο δεύτερο στάδιο χρησιμοποιήθηκαν δύο τεχνικές που ήταν η fuzzy logic και τα neural networks. Για το πείραμα χρησιμοποιήθηκε το σύνολο δεδομένων Pima Indians από το Πανεπιστήμιο της Καλιφόρνια. Ο Gaussian Kernel διανέμει τα δεδομένα με μεγάλη ακρίβεια και απαιτεί λιγότερο υπολογιστικό χρόνο. Αρχικά, υπολογίστηκε η τυπική απόκλιση που χρησιμοποιήθηκε από μια Gaussian συνάρτηση και χρησιμοποιήθηκε για την ταξινόμηση των δεδομένων. Το Fuzzy είναι σύστημα βασισμένο σε κανόνες και ο κανόνας είναι η κύρια λειτουργία στο σύστημα ασαφούς παρεμβολής. Λήφθηκαν βελτιωμένα αποτελέσματα χρησιμοποιώντας ασαφή σύνολα και το ANN αναγνωρίστηκε ως το καταλληλότερο για την τεχνική αναγνώρισης προτύπων. Σε αυτό το ήδη διανεμημένο σύστημα, τα δεδομένα τροφοδοτήθηκαν στο στρώμα εισόδου του ANN και αποτελείται από νευρώνες 'n'. Το κρυφό επίπεδο εκτελεί μια λειτουργία που βασίζεται στο επίπεδο εισόδου και αποτελείται από n+1 νευρώνες. Το στρώμα εξόδου αντιπροσωπεύει την έξοδο είτε ως 0 είτε ως 1. Αυτές οι τιμές υποδεικνύουν εάν ο ασθενής είναι διαβητικός ή όχι. Το συμπέρασμα του πειράματος ήταν ότι η ακρίβεια των συνδυασμένων μεθόδων βελτιώθηκε από τις μεμονωμένες μεθόδους.

Οι Li et al. [67] χρησιμοποίησε 3000 δημοσιεύσεις στο Weibo ως σύνολο δεδομένων. Τα δεδομένα έχουν ταξινομηθεί σε επτά κατηγορίες. Οι αναρτήσεις επισημαίνονται χειροκίνητα και βρίσκεται η τιμή K του Cohen 0,81. Τα SVM, Naïve Bayes και Random Forest χρησιμοποιούνται

στην πρόβλεψη, όπου το Random Forest υπερτερεί των άλλων. Barkur et al. [68] χρησιμοποίησε 24.000 tweets από το Twitter με τα #IndiaFightsCorona και #Indialockdown ως hashtags για ανάλυση. Κατέληξαν στο συμπέρασμα από το WordCloud ότι υπάρχει ισχυρό θετικό συναίσθημα για την ισοπέδωση της καμπύλης και ότι η κυβέρνηση έχει εμπιστοσύνη και υποστήριξη από Ινδούς πολίτες.

Οι Jelodar et al. [69] έχουν χρησιμοποιήσει 563.079 αναρτήσεις Reddit ως σύνολο δεδομένων για την ανάλυσή τους. Έχουν χρησιμοποιήσει το LDA για τη μοντελοποίηση θεμάτων. Έχουν ταξινομήσει τις δημοσιεύσεις σε πέντε βαθμολογίες συναισθήματος που κυμαίνονται από πολύ θετικές έως πολύ αρνητικές. Το LSTM συγκρίνεται με τα KNN, SVM, Naïve Bayes και την Logistic Regression στην οποία το LSTM υπερτερεί των άλλων με ακρίβεια 81,15%.

#### 2.5.4. Gradient Boosting Classifier

Ο αλγόριθμος Gradient Boosting είναι μια εποπτευόμενη τεχνική εκμάθησης που βασίζεται στη μέθοδο gradient descent. Μπορεί να χρησιμοποιηθεί για ταξινόμηση αλλά και παλινδρόμηση. Έχει την ικανότητα να μειώνει τη διακύμανση και την προκατάληψη και βοηθά τον αδύναμο εκπαιδευόμενο να γίνει πιο δυνατός. Αυτό το μοντέλο βελτιώνει την ακρίβεια πρόβλεψης προσθέτοντας επιπλέον δέντρα για τη διόρθωση λαθών που έγιναν από προηγούμενα βασικά μοντέλα. Η αρχή του βασίζεται στην τεχνική Optimization που βρίσκει τις τιμές των συντελεστών της συνάρτησης ( $f$ ) που ελαχιστοποιούν το  $a$  ( $c$ ) συνάρτηση κόστους με συνάρτηση διαφορικής απώλειας. Όπως και άλλοι αλγόριθμοι ενίσχυσης, ένα σύνολο αδύναμων μαθητών γίνεται σύνολο σε έναν μόνο ισχυρό εκπαιδευόμενο διαδοχικά. Στην ενίσχυση κλίσης, κάθε δέντρο χτίζεται ένα προς ένα κάνοντας διόρθωση στο ποσοστό λάθους των προηγούμενων όπου όπως και στο Random Forest χτίζονται τυχαία. Η ενίσχυση κλίσης είναι χρονοβόρα σε σύγκριση με το Random Forest.

Ας υποθέσουμε ότι έχουμε το δείγμα εκπαίδευσης  $\{(x_i, y_i)\}_{i=1}^n$  και συνάρτηση απώλειας είναι  $L(y, F(x))$  με τον 1<sup>ο</sup> αριθμό επαναλήψεων. Εφαρμόζοντας τώρα την αρχή της εμπειρικής ελαχιστοποίησης κινδύνου με το μοντέλο που παίρνουμε να έχει ως εξής:

$$F_0(x) = \arg \min(\gamma) \sum_{i=1}^n L(y_i, \gamma) \quad (4)$$

Η Μηχανική Μάθηση όχι μόνο μπορεί να χρησιμοποιηθεί για την πρόγνωση και τη διάγνωση αλλά έχει και πολλές άλλες εφαρμογές στον ιατρικό τομέα. Το έμπειρο σύστημα που βασίζεται σε υπολογιστή για την καταγραφή και την ερμηνεία του πλαισίου του ασθενούς, δεν είναι μόνο χρήσιμο για τους ασθενείς αλλά και για τους γιατρούς. Το METABO είναι ένα τέτοιο σύστημα για το σύστημα παρακολούθησης και διαχείρισης του διαβήτη [70]. Αυτό το σύστημα αποτελείται από την κινητή συσκευή του ασθενούς, διαφορετικούς τύπους διακριτικών βιοαισθητήρων, ένα κεντρικό υποσύστημα που βρίσκεται εξ αποστάσεως στο νοσοκομείο και τον πίνακα ελέγχου που βοηθά στην ολοκληρωμένη παρακολούθηση των ενεργειών δεδομένων σχετικά με τον ασθενή και τη νόσο. Οι Breault et al. [71] δείχνει αυξανόμενη ανάπτυξη του διαβήτη στις ΗΠΑ, γεγονός που καθιστά πολύ σημαντικό θέμα που πρέπει να συλλογιστούμε. Διαφορετική τεχνική ταξινόμησης προτείνεται στο [72] όπως το DT, το SVM, το NB, όταν αξιολογήθηκαν για έλεγχο απόδοσης χωρίς ενίσχυση έδωσε ακρίβεια ως 76%, 79,68%, 78,1%, 74,47% αντίστοιχα και η

αξιολόγηση απόδοσης του Adaboost με DT, SVM, NB, DS ως βασικό ταξινομητή οδήγησε στη βελτίωση της ακρίβειας εκτός από το SVM που δεν δείχνει καμία εξέλιξη ακρίβειας [73].

Μετά την επιλογή σημαντικών μεταβλητών, τα δεδομένα πριν από τις 17 Φεβρουαρίου 2021 (N=769) χωρίστηκαν τυχαία σε σύνολα δεδομένων εκπαίδευσης και δοκιμής σε 80% και 20%, αντίστοιχα. Στη συνέχεια, το GBM και ένα μοντέλο LR που χρησιμοποιεί τη μέθοδο διασταυρούμενης επικύρωσης στρωματοποιημένης πτυχής K εφαρμόστηκαν στα δεδομένα (K=5). Σε σύγκριση με το μοντέλο LR, το Light GBM χρησιμοποίησε το "NaN" για να αναπαραστήσει τις τιμές που λείπουν και αντιμετωπίστηκαν χωριστά από το μηδέν, καθώς οι τιμές που λείπουν ερμηνεύτηκαν ότι περιείχαν πληροφορίες [74]. Η βελτιστοποίηση υπερπαραμέτρων πραγματοποιήθηκε χρησιμοποιώντας μια υλοποίηση που ονομάζεται "Optuna" για light GBM. Για τον LR, χρησιμοποιήθηκε μια στρατηγική αναζήτησης πλέγματος για να εντοπιστούν οι καλύτερες υπερπαραμέτροι συντονισμού [75].

### 2.5.5.KNeighbors Classifier

Σε πολλά ερευνητικά έργα, οι εφαρμογές υγειονομικής περίθαλψης αναλύονται χρησιμοποιώντας τεχνικές ταξινόμησης για να ανακαλύψουν λανθασμένες μορφές και σχέσεις μεταξύ των χαρακτηριστικών που μπορούν να καθοδηγήσουν την ιατρική διάγνωση και να ενισχύσουν την πτυχή της φροντίδας των ασθενών. Παρόλο που είμαστε σε θέση να λάβουμε σχετικές πληροφορίες από τα υπάρχοντα ιατρικά αρχεία, ωστόσο ενδέχεται να περιέχουν πολλά πρόσθετα και άσχετα χαρακτηριστικά.

Η μέθοδος K-Nearest Neighbor βασίζεται στην τεχνική της εποπτευόμενης μάθησης και είναι ένας από τους πιο βασικούς αλγόριθμους ML. Ο αλγόριθμος KNN είναι ένας μη παραμετρικός αλγόριθμος, που σημαίνει ότι δεν κάνει υποθέσεις σχετικά με τα υποκείμενα δεδομένα. Είναι επίσης γνωστός ως αλγόριθμος «τεμπέλης», καθώς δεν μαθαίνει από το σετ εκπαίδευσης αμέσως. Αντίθετα, αποθηκεύει το σύνολο δεδομένων και εκτελεί μια ενέργεια σε αυτό όταν έρθει η ώρα να το ταξινομήσει. Ο αλγόριθμος KNN μπορεί να χρησιμοποιηθεί τόσο για παλινδρόμηση όσο και για ταξινόμηση, αλλά χρησιμοποιείται πιο συχνά για προβλήματα ταξινόμησης. Για τον υπολογισμό της Ευκλείδειας απόστασης δίνεται η εξίσωση:

$$d_{Eucliden} = \sqrt{\sum_{i=0}^n (x_i + y_i)^2} \quad (5)$$

Η χρήση της μηχανικής μάθησης για την ταξινόμηση του διαβήτη έχει προταθεί σε αρκετές δημοσιεύσεις. ANN και adaptive neuro-fuzzy μοντέλα έχουν χρησιμοποιηθεί για τη σωστή ταξινόμηση του σακχαρώδους διαβήτη, με ακρίβεια 90,32% για τον τελευταίο [76]. Μια συγκριτική μελέτη των KNN, του J48, του RF και των SVM ανέφερε ότι οι KNN και RF ξεπέρασαν τους άλλους αλγόριθμους στην ταξινόμηση των διαβητικών ασθενών [77], με 100% ακρίβεια. Πιο πρόσφατα, το KNN, τα ANN και το J48 χρησιμοποιήθηκαν για την ανάπτυξη μοντέλων για την ταξινόμηση του σακχαρώδους διαβήτη [78]. Ο RF, με ακρίβεια 80%, ήταν ο αλγόριθμος με την καλύτερη απόδοση σε μια σύγκριση που περιελάμβανε επίσης τους Naive Bayes, ID3 και AdaBoost [79]. Ο Naive Bayes ήταν ο αλγόριθμος με την καλύτερη απόδοση, με ακρίβεια 76,3%, σε μια άλλη

πρόσφατη μελέτη [80]. Τέλος, οι SVM [81] και KNN [82] ήταν οι αλγόριθμοι που επιλέχθηκαν για την ταξινόμηση του σακχαρώδη διαβήτη. Αυτές οι μελέτες υπογραμμίζουν τη συνολική καλή απόδοση των KNN και RF στην ταξινόμηση του T2DM.

Οι Brinati et al. [83] ανέπτυξαν ένα μοντέλο για την πρόβλεψη της διάγνωσης του Covid-19. Χρησιμοποιήθηκε ένα σύνολο δεδομένων από 279 ασθενείς (177 θετικοί) από το IRCCS Ospedale San Raffaele της Ιταλίας. Η μελέτη χρησιμοποίησε διάφορες προσεγγίσεις ML: DT, KNN, RF, LR, NB, SVM και RF. Ο ταξινομητής RF βρέθηκε να επιτυγχάνει την καλύτερη απόδοση, με AUC 84%, ακρίβεια 82%, ευαισθησία 92% και ειδικότητα 65%. Τα πιο σημαντικά χαρακτηριστικά στην πρόβλεψη της διάγνωσης του COVID-19 ήταν: τα λεμφοκύτταρα, ηωσινόφιλα, τρανσαμινάση αλανίνης και ηλικία.

### 2.5.6. Logistic Regression

Ένα μοντέλο Logistic Regression (LR) είναι επίσης γνωστό ως μοντέλο logit. Είναι ένας αλγόριθμος ταξινόμησης μηχανικής μάθησης με τη βοήθεια του οποίου μπορούμε να προβλέψουμε την κλάση οποιασδήποτε τιμής εισόδου. Ο Logistic Regression ανακαλύπτει τη σχέση μεταξύ διαφορετικών χαρακτηριστικών του συνόλου δεδομένων. Χρησιμοποιείται επίσης για τη χαρτογράφηση της σχέσης μεταξύ ενός διπλού παράγοντα ενός τμήματος και ενός ελάχιστου ελεύθερου παράγοντα. Όλοι αυτοί οι παράγοντες συνδέονται με κάποια βάρη και στη συνέχεια συνοψίζονται. Το αποτέλεσμα θα εξαρτηθεί από μια συνάρτηση γνωστή ως σιγμοειδής συνάρτηση που θα επιστρέψει μια τιμή μεταξύ 0 και 1. Εάν η τιμή επιστροφής είναι μεγαλύτερη από 0,5, θα ταξινομηθεί ως 1, διαφορετικά 0. Η λογιστική παλινδρόμηση είναι προτιμότερη για δυαδική ταξινόμηση.

Ο διαβήτης προκαλεί μεγάλο αριθμό θανάτων κάθε χρόνο και ένας μεγάλος αριθμός ατόμων που ζουν με την ασθένεια δεν συνειδητοποιούν την κατάσταση της υγείας τους νωρίτερα. Σε αυτήν τη μελέτη, προτείνεται ένα μοντέλο βασισμένο στην εξόρυξη δεδομένων για έγκαιρη διάγνωση και πρόβλεψη του διαβήτη χρησιμοποιώντας το σύνολο δεδομένων Pima Indians Diabetes. Η παρούσα ανάλυση μιας μεγάλης κλινικής βάσης δεδομένων πραγματοποιήθηκε για την εξέταση ιστορικών προτύπων περίθαλψης διαβήτη σε ασθενείς με διαβήτη που εισήχθησαν σε νοσοκομείο των ΗΠΑ και για την ενημέρωση μελλοντικών κατευθύνσεων που θα μπορούσαν να οδηγήσουν σε βελτιώσεις στην ασφάλεια των ασθενών.

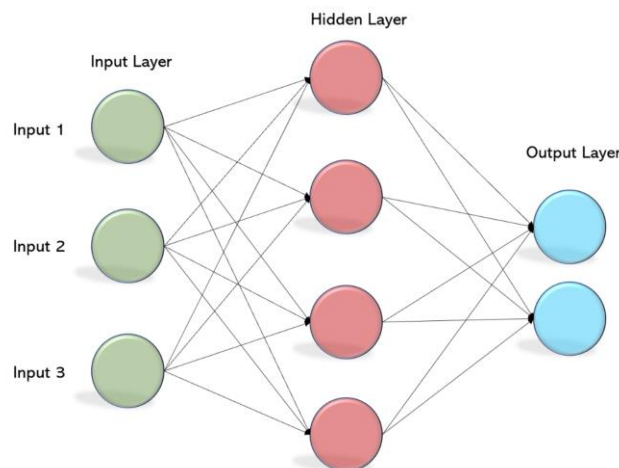
Πολλοί ερευνητές έχουν αναπτύξει διάφορα μοντέλα πρόβλεψης χρησιμοποιώντας εξόρυξη δεδομένων για την πρόβλεψη και τη διάγνωση του διαβήτη. Ο Iyer [84] στη μελέτη του πρότεινε τη χρήση του αλγόριθμου Naïve Bayes για την πρόβλεψη της εμφάνισης του διαβήτη. Η μελέτη έδωσε αποτέλεσμα ακρίβειας 79,56%. Ο Tarun [85] χρησιμοποίησε PCA και ένα SVM για την ταξινόμηση των διαβητικών ασθενών. Το πειραματικό αποτέλεσμα από τη μελέτη έδειξε ότι το προηγούμενο επίπεδο μπορεί να βελτιωθεί καθώς είχαν ακρίβεια ταξινόμησης 93,66%. Mustafa S. Kadhm [86] πρότεινε τη χρήση ενός DT για την ανάθεση κάθε δείγματος δεδομένων στην κατάλληλη κλάση του μετά την εφαρμογή του αλγόριθμου KNN για την εξάλειψη ανεπιθύμητων δεδομένων. Οι Han et al. [87] σχεδίασε ένα μοντέλο που χρησιμοποιεί τον αλγόριθμο k-means και τον αλγόριθμο LR για την πρόβλεψη του διαβήτη. Το μοντέλο πέτυχε ακρίβεια 95,42%.

Υπάρχουν τρεις μέθοδοι ταξινόμησης μεταξύ πολλών μεθόδων που χρησιμοποιούνται συχνά στην εξόρυξη δεδομένων. Δηλαδή Naive Bayes, K-Nearest Neighbor και Logistic Regression. Αρκετοί ερευνητές έχουν πραγματοποιήσει έρευνα που σχετίζεται με αυτές τις τρεις μεθόδους. Μεταξύ αυτών έγινε από τους Nova Tri Romadloni [88] και τους συνεργάτες του που συνέκριναν τους Naive Bayes, KNN και Decision tree για να αναλύσουν το συναίσθημα μεταφοράς, οι Daniela Xhemali et al [89] που εξέτασαν την απόδοση του Naive Bayes για την ταξινόμηση των ιστοσελίδων εκπαίδευσης και Van Der Heide et al [90] που συνέκριναν LR, Naive Bayes και RF μεθόδους για να προβλέψουν την επιβίωση των ζώων. Από αυτές τις μελέτες, δεν βρέθηκε έρευνα που να αναλύει την απόδοση της μεθόδου Naive Bayes, του KNN και της λογιστικής παλινδρόμησης για την πρόβλεψη ασθενών με υψηλή επικράτηση του Covid-19.

Η λογιστική παλινδρόμηση επομένως είναι ο αλγόριθμος για μια αποτελεσματική προγνωστική ανάλυση παλινδρόμησης. Η εφαρμογή του είναι αποτελεσματική όταν η εξαρτημένη μεταβλητή ενός συνόλου δεδομένων είναι διχοτομημένη (δυναδική). Η λογιστική παλινδρόμηση χρησιμοποιείται στην περιγραφή και ανάλυση δεδομένων προκειμένου να εξηγήσει τη σχέση μεταξύ μιας εξαρτημένης δυναδικής μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών.

### 2.5.7. Neural Networks

Ένα τεχνητό νευρωνικό δίκτυο (Neural Network) που ονομάζεται πολυστρωματικό perceptron (MultiLayer Perceptron - MLP) είναι ένας τύπος τεχνητού νευρωνικού δικτύου τροφοδοσίας (Artificial neural network - ANN). Ο ταξινομητής MLP, ο οποίος, όπως υποδηλώνει το όνομα, συνδέεται με ένα Νευρωνικό Δίκτυο. Η διάδοση προς τα πίσω είναι μια εποπτευόμενη τεχνική εκμάθησης που χρησιμοποιείται από το MLP κατά τη διάρκεια της εκπαίδευσης. Το MLP διακρίνεται από ένα γραμμικό perceptron από τα πολυάριθμα στρώματα και τη μη γραμμική ενεργοποίησή του. Μπορεί να διακρίνει τη διαφορά μεταξύ δεδομένων που δεν διαχωρίζονται γραμμικά. Το Perceptron αποτελείται από δύο εντελώς συνδεδεμένα στρώματα: ένα στρώμα εισόδου και ένα στρώμα εξόδου. Τα MLP έχουν τα ίδια επίπεδα εισόδου και εξόδου, αλλά όπως φαίνεται και στην ακόλουθη εικόνα, μπορεί να έχουν πολλά κρυφά στρώματα μεταξύ τους (Εικόνα 8).



Εικόνα 8: Νευρωνικά Δίκτυα



Μπορούμε να περιγράψουμε τον αλγόριθμο MLP ακολουθώντας τα παρακάτω βήματα:

1. Γίνεται φόρτωση των σχετικών βιβλιοθηκών και συναρτήσεων.
2. Ανάγνωση των δεδομένων και διεξαγωγή βασικών ελέγχων δεδομένων.
3. Δημιουργούνται πίνακες για τα χαρακτηριστικά και τη μεταβλητή απόκρισης.
4. Τα σύνολα δεδομένων εκπαίδευσης και δοκιμής δημιουργούνται.
5. Το μοντέλο νευρωνικών δικτύων κατασκευάστηκε, προβλέφθηκε και αξιολογήθηκε.

Abdullah Caliskan et al. [90] πρότεινε μια απλή στρατηγική εκπαίδευσης για ταξινομητή deep neural network χρησιμοποιώντας τον αλγόριθμο L-BFGS (Limited Broyden–Fletcher–Goldfarb–Shanno) όπου η αξιολόγηση έγινε με διάφορα σύνολα δεδομένων, συμπεριλαμβανομένων των δεδομένων διαβήτη Pima Indians. Πέτυχαν 77,09% ακρίβεια ταξινόμησης για το σύνολο δεδομένων διαβήτη Pima Indian. Στο [91] προτείνεται μια προσέγγιση βασισμένη σε deep convolutional neural network για την ταξινόμηση των περιπτώσεων διαβητικής αμφιβληστροειδοπάθειας με βάση δεδομένων εικόνας. Στο [92] ένα μοντέλο πρόβλεψης δημιουργείται χρησιμοποιώντας τεχνικές εξόρυξης δεδομένων χρησιμοποιώντας βελτιωμένο αλγόριθμο k-means και LR [93]. Αναφέρεται ότι πέτυχαν ακρίβεια και τιμή ανάκλησης 0,770 και 0,775 αντίστοιχα χρησιμοποιώντας τον αλγόριθμο Hoeffding Tree [94]. Ωστόσο, έχουν προταθεί αρκετές μελέτες όπου η ακρίβεια ταξινόμησης είναι μεταξύ 59,5% και 77,7% [95].

Έχει γίνει πολυάριθμος αριθμός μελετών για την προώθηση του αυτοματοποιημένου ελέγχου και διάγνωσης που βασίζεται στην ανάλυση των εικόνων CT θώρακα [96]. Η τεχνητή νοημοσύνη μπορεί να επιβληθεί στην περιοχή ηλεκτρονικής υγείας για να βοηθήσει στην έγκαιρη ανίχνευση του COVID-19 αναλύοντας αυτούς τους τρεις κύριους ήχους, δηλαδή τον βήχα, την αναπνοή και τη φωνή. Επιπλέον, οι αναπνευστικοί ήχοι μπορούν να φέρουν ποικίλες ενδείξεις για την κατάσταση της υγείας του ανθρώπου, οι οποίες μπορούν να αναγνωριστούν και να διαγνωστούν με την εφαρμογή αλγορίθμων μηχανικής μάθησης [97]. Έτσι, από τη μαζική έκρηξη του ιού COVID-19, επιστήμονες και ερευνητές εξετάζουν τώρα την ανίχνευση του COVID-19 από αναπνευστικούς ήχους [98]. Στο [99], προτείνεται μια χαμηλή καταναλωτική ισχύς και ένα σύγχρονο φορητό σύστημα για την ανίχνευση τόσο του άσθματος όσο και του συριγμού με βάση την ανάλυση των ηχητικών χαρακτηριστικών τους και τις συχνότητες των αναπνευστικών ήχων τους. Στο [100], τα CNN χρησιμοποιούνται για την ανίχνευση διαφορετικών τύπων βήχα με βάση την ανάλυση των εξαγόμενων χαρακτηριστικών ήχου τους. Επιπλέον, έχει προταθεί ένα σύστημα πρόβλεψης του COVID-19 χρησιμοποιώντας αλγόριθμους βαθιάς μάθησης και αρκετούς ταξινομητές όπως το CNN, το LSTM και τα ANN [101].

Η ανίχνευση της κατάστασης της υγείας των ασθενών με COVID-19 μπορεί να επιτευχθεί χρησιμοποιώντας τα σήματα ομιλίας τους. Επομένως, ένα σύστημα ανίχνευσης κατάστασης υγείας μπορεί να χρησιμοποιηθεί για την παρατήρηση και ανάλυση της ποιότητας του ύπνου, της σοβαρότητας της ασθένειας, της κόπωσης και του άγχους [102]. Ο βήχας ήταν πάντα σύμπτωμα πολλών ασθενειών. Στην πραγματικότητα, είναι δυνατό να γίνει διάκριση μεταξύ του βήχα και να εκτιμηθεί το είδος της ασθένειας δοκιμάζοντας τα ακουστικά χαρακτηριστικά χρησιμοποιώντας πολλαπλούς ταξινομητές [103].

### 2.5.8. Random Forest Classifier

Ο Random Forest (RF) είναι ένας δημοφιλής αλγόριθμος μηχανικής μάθησης που ανήκει στην εποπτευόμενη τεχνική μάθησης. Μπορεί να χρησιμοποιηθεί για προβλήματα ταξινόμησης και παλινδρόμησης στο ML. Βασίζεται στην έννοια του **ensemble learning**, η οποία είναι μια διαδικασία συνδυασμού πολλαπλών ταξινομητών για την επίλυση ενός σύνθετου προβλήματος και τη βελτίωση της απόδοσης του μοντέλου.

Όπως υποδηλώνει το όνομα, **"Το Random Forest είναι ένας ταξινομητής που περιέχει έναν αριθμό δέντρων αποφάσεων σε διάφορα υποσύνολα του δεδομένου συνόλου δεδομένων και παίρνει τον μέσο όρο για να βελτιώσει την προγνωστική ακρίβεια αυτού του συνόλου δεδομένων"**. Αντί να βασίζεται σε ένα δέντρο αποφάσεων, ο RF λαμβάνει την πρόβλεψη από κάθε δέντρο και βασίζεται στις πλειοψηφίες των προβλέψεων και μ' αυτόν τον τρόπο προβλέπει την τελική απόδοση. Συνεπώς, γίνεται αντιληπτό ότι οι τεχνικές μηχανικής μάθησης έχουν εφαρμοστεί σε διάφορους τομείς, ιδιαίτερα στον ιατρικό τομέα.

Ο RF αποτελείται από ένα μεγάλο αριθμό δέντρων απόφασης. Κάθε κόμβος του δέντρου αποφάσεων θέτει μια ερώτηση σχετικά με τα δεδομένα, με τους κλάδους να αντιπροσωπεύουν εναλλακτικές απαντήσεις. Εκατό δέντρα απόφασης συνδυάζονται σε μια τυχαία δασική προσέγγιση. Η δημοτικότητα των μοντέλων RFC πηγάζει από τη μεγάλη τους ακρίβεια και το φθινό υπολογιστικό κόστος. Ο RF είναι μια εποπτευόμενη μέθοδος μάθησης που μπορεί να χρησιμοποιηθεί σε μοντέλα ταξινόμησης και παλινδρόμησης. Ωστόσο, χρησιμοποιείται κυρίως για την επίλυση προβλημάτων ταξινόμησης. Τα βήματα που ακολουθούνται γενικά για τον αλγόριθμο RF είναι:

- Επιλέξτε ένα τυχαίο δείγμα από το σύνολο δεδομένων.
- Δημιουργήστε ένα δέντρο αποφάσεων για κάθε δείγμα. Στη συνέχεια, πάρτε το προβλεπόμενο αποτέλεσμα από κάθε δέντρο.
- Διάκριση για κάθε προβλεπόμενο αποτέλεσμα.

Το τελικό αποτέλεσμα θα προβλεφθεί από την πρόβλεψη με τις περισσότερες ψήφους.

Οι Butwall & Kumar [104] πρότειναν μια μεθοδολογία βασισμένη στον ταξινομητή Random Forest (RF) για να οραματιστεί τη συμπεριφορά του διαβήτη σε συμφωνία με συγκεκριμένες παραμέτρους του τρόπου ζωής, συμπεριλαμβανομένης της σωματικής δραστηριότητας και των συναισθηματικών καταστάσεων, ιδιαίτερα των ηλικιωμένων διαβητικών. Σε αυτήν την ερευνητική εργασία, χρησιμοποιήθηκε ο RF με διαφορετικές παραμέτρους δοκιμής στη βάση δεδομένων διαβητικών Pima Indian του UCI Machine Learning Lab. Διαπιστώθηκε ότι η RF είναι αποτελεσματική στη διάγνωση του σακχαρώδους διαβήτη όταν το άτομο παρέχει την απαιτούμενη τιμή χαρακτηριστικών. Οι Dewangan & Agrawal [105] έχουν υιοθετήσει διάφορες μεθόδους ταξινόμησης για να τις ανασυγκροτήσουν ώστε να δώσουν ένα νέο υβριδικό μοντέλο στην αναζήτηση της εύρεσης της καλύτερης απόδοσης. Τεχνικές μηχανικής μάθησης όπως το C4.5 και το Perceptron πολλαπλών επιπέδων έχουν εκπαιδευτεί στο σύνολο δεδομένων διαβήτη Pima Indian που συλλέγονται από το UCI Repository. Κύριος στόχος αυτής της εργασίας είναι η ανίχνευση του Σακχαρώδους Διαβήτη και η ταξινόμηση των δεδομένων ως διαβητικού και μη.

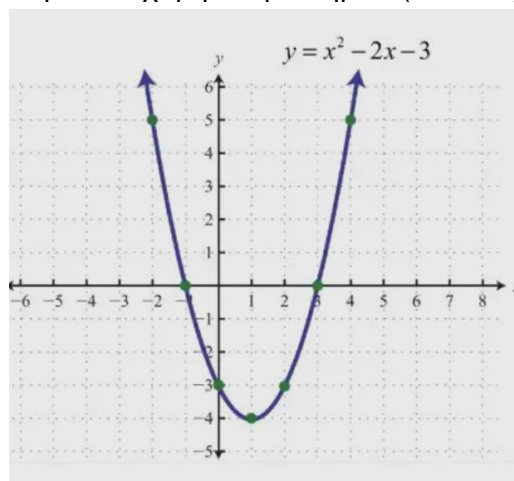
Στο [106] αλγόριθμοι μηχανικής μάθησης, συμπεριλαμβανομένων των C4.5 και Random Forest, έχουν χρησιμοποιηθεί για τη δημιουργία μοντέλων ταξινόμησης με σκοπό την πρόβλεψη των

διαβητικών ασθενών χρησιμοποιώντας πραγματικά σύνολα δεδομένων υγειονομικής περίθαλψης. Το Random Forest πέτυχε την καλύτερη απόδοση σε σύγκριση με άλλους ταξινομητές μηχανικής μάθησης. Επιπλέον, τα αποτελέσματα δείχνουν ότι το κατασκευασμένο μοντέλο θα μπορούσε να βοηθήσει τους παρόχους υγειονομικής περίθαλψης να λάβουν καλύτερες κλινικές αποφάσεις για τον εντοπισμό των διαβητικών ασθενών.

Όσον αφορά το σενάριο covid-19, οι Tang et al. [107] οι ερευνητές εφάρμοσαν τον αλγόριθμο RF ανάλυσης για την ανάλυση σοβαρότητας ασθενών με COVID-19 χρησιμοποιώντας τις σαρώσεις υπολογιστικής τομογραφίας (CT). Στο Khalifa et al. [108] οι συγγραφείς πρότειναν ένα προσαρμοσμένο μοντέλο βασισμένο στο Generative Adversarial Network για την ανίχνευση πνευμονίας από σαρώσεις με ακτίνες Χ θώρακα, το οποίο είναι ένα από τα συμπτώματα της λοίμωξης COVID-19. Στο Sujatha et al. [109], οι συγγραφείς πρότειναν μια μέθοδο που θα μπορούσε να είναι χρήσιμη για την πρόβλεψη της έκτασης του COVID-2019, εκτελώντας LR, και το μοντέλο MLP που θα μπορούσε να παρέχει μια προσδοκία σχετικά με τις πληροφορίες COVID-19, για την πρόβλεψη της επιδημιολογικό πρότυπο της νόσου και ποσοστό κρουσμάτων COVID-2019 στην Ινδία αλλά και σε άλλες χώρες ανά τον κόσμο.

### 2.5.9. Stochastic Gradient Descent

Ο Stochastic Gradient Descent (SGD) είναι ένας πολύ δημοφιλής και κοινός αλγόριθμος που χρησιμοποιείται σε διάφορους αλγόριθμους ML, και το πιο σημαντικό αποτελεί τη βάση των NNs. Κλίση, με απλούς όρους σημαίνει κλίση ή κλίση μιας επιφάνειας. Άρα κλίση κατάβασης κυριολεκτικά σημαίνει κατέβασμα μιας κλίσης για να φτάσετε στο χαμηλότερο σημείο αυτής της επιφάνειας. Ας φανταστούμε ένα δισδιάστατο γράφημα, όπως μια παραβολή στο παρακάτω σχήμα. Στο παρακάτω γράφημα, το χαμηλότερο σημείο της παραβολής εμφανίζεται στο  $x=1$ . Ο στόχος του αλγόριθμου καθόδου με κλίση είναι να βρει την τιμή του "x" έτσι ώστε το "y" να είναι ελάχιστο. Το "y" εδώ ορίζεται ως η αντικειμενική συνάρτηση στην οποία λειτουργεί ο αλγόριθμος καθόδου βαθμίδας, για να κατέβει στο χαμηλότερο σημείο (Εικόνα 9).



Εικόνα 9: Αλγόριθμος SGD

Αυτή είναι η ουσία του αλγορίθμου. Η γενική ιδέα είναι να ξεκινήσετε με ένα τυχαίο σημείο (στο παράδειγμά μας στην παραβολή ξεκινήστε με ένα τυχαίο "x") και να βρείτε έναν τρόπο να ενημερώνετε αυτό το σημείο με κάθε επανάληψη έτσι ώστε να κατεβαίνουμε την κλίση. Τα βήματα του αλγορίθμου είναι:

1. Βρείτε την κλίση της αντικειμενικής συνάρτησης **σε σχέση με κάθε παράμετρο/χαρακτηριστικό**. Με άλλα λόγια, να υπολογίσετε τη διαβάθμιση της συνάρτησης.
2. Επιλέξτε μια τυχαία αρχική τιμή για τις παραμέτρους. (Για να διευκρινίσουμε, στο παράδειγμα της παραβολής, διαφοροποιήστε το "y" σε σχέση με το "x". Αν είχαμε περισσότερα χαρακτηριστικά όπως x1, x2 κ.λπ., παίρνουμε τη μερική παράγωγο του "y" σε σχέση με καθένα από τα χαρακτηριστικά.)
3. Ενημερώστε τη λειτουργία gradient συνδέοντας τις τιμές παραμέτρων.
4. Υπολογίστε τα μεγέθη βημάτων για κάθε χαρακτηριστικό ως: **μέγεθος βήματος = κλίση \* ρυθμός εκμάθησης**.
5. Υπολογίστε τις νέες παραμέτρους ως : **new params = old params - step size**
6. Επαναλάβετε τα βήματα 3 έως 5 έως ότου η κλίση είναι σχεδόν 0.

Ο «ρυθμός μάθησης» που αναφέρεται παραπάνω είναι μια ευέλικτη παράμετρος που επηρεάζει σε μεγάλο βαθμό τη σύγκλιση του αλγορίθμου. Οι μεγαλύτεροι ρυθμοί εκμάθησης κάνουν τον αλγόριθμο να κάνει τεράστια βήματα κάτω από την κλίση και μπορεί να υπερπηδήσει το ελάχιστο σημείο με αποτέλεσμα να το χάσει. Επομένως, είναι πάντα καλό να επιμένετε σε χαμηλό ποσοστό μάθησης όπως το 0,01. Μπορεί επίσης να αποδειχθεί μαθηματικά ότι ο αλγόριθμος gradient descent κάνει μεγαλύτερα βήματα κάτω από την κλίση εάν το σημείο εκκίνησης είναι ψηλά και κάνει μικρά βήματα καθώς πλησιάζει στον προορισμό για να προσέξετε να μην τον χάσετε και επίσης να είναι αρκετά γρήγορος.

Η βελτιστοποίηση των τιμών δεδομένων μεμονωμένων ομάδων για την ελαχιστοποίηση των απωλειών μετά τις προβλέψεις από τις τιμές κλίσης έχει τεράστια επίδραση σε όλη τη ροή δεδομένων [110]. Η βέλτιστη λύση με υψηλή εκπαιδευμένη ακρίβεια δεδομένων επιτυγχάνεται χωρίς να περιλαμβάνεται στους άλλους δείκτες απόδοσης. Μεγάλος όγκος δεδομένων για εξομάλυνση γίνεται αυθόρμητα με την υλοποίηση του διαχωρισμού μπλοκ προς μπλοκ [111]. Επιπλέον, η ιδέα επεκτάθηκε για οποιοδήποτε δίκτυο συνέλιξης παρέχοντας μια λύση ελαχιστοποιημένης απώλειας με ελάχιστη απώλεια δεδομένων. Η στοχαστική κλίση δεδομένων με εύρος αβεβαιότητας παρέχει πολύτιμη ταξινόμηση για πολλές εφαρμογές [112].

Στο [113], έχει προταθεί μια ποσοτική μέθοδος για τον εντοπισμό και την αξιολόγηση του Covid-19 χρησιμοποιώντας τις εικόνες ακτίνων X θώρακα σε μια αρχιτεκτονική Convolutional Neural Network. Εκατόν εξήντα εικόνες ακτινογραφίας θώρακος του Covid-19 συλλέχθηκαν από το GitHub και ο ίδιος αριθμός κανονικών εικόνων ακτινογραφίας θώρακος συλλέχθηκαν από το αποθετήριο, το Kaggle. Κατά την εξέταση της διαδικασίας ανάλυσης του CNN, έχει πραγματοποιηθεί αρχικά η προεπεξεργασία του μοντέλου. Στη συνέχεια οι εικόνες έχουν αλλάξει μέγεθος σε 70×70 pixels, και η δημιουργία σεναρίου γίνεται αλλάζοντας τις παραμέτρους του μοντέλου. Ως εκ τούτου, οι αναλογίες δεδομένων άλλαξαν σε 80:20, 70:30, 60:40 και δοκιμάστηκαν χρησιμοποιώντας epochs των 20, 30 και 40. Στο στάδιο ταξινόμησης, το προτεινόμενο μοντέλο δοκιμάστηκε χρησιμοποιώντας εικόνες ακτινογραφίας θώρακα χωρίς ετικέτα. Αυτό το μοντέλο, το οποίο έχει προταθεί, είναι διαδοχικό, το Stochastic Gradient Descent (SGD) έχει χρησιμοποιηθεί ως βελτιστοποιητής και το Softmax ως συνάρτηση ταξινόμησης. Το προτεινόμενο μοντέλο αυτοματοποιημένης πρόβλεψης έχει παράγει καλή ακρίβεια απόδοσης 97,91%.

### 2.5.10.Support Vector Machine

Ο Support Vector Machine (SVM) είναι ένας εποπτευόμενος αλγόριθμος μάθησης. Η εφαρμογή αυτού του αλγορίθμου έχει ήδη δει στον τομέα της ταξινόμησης εικόνων, της χειρόγραφης αναγνώρισης χαρακτήρων και πολλά άλλα βρίσκονται ακόμη στη διαδικασία.

Είναι ένας αλγόριθμος ταξινόμησης, η κύρια στρατηγική που χρησιμοποιείται από αυτόν τον αλγόριθμο είναι να δημιουργήσει μέγιστο περιθώριο μεταξύ των κλάσεων, έτσι ώστε η έξοδος να μπορεί εύκολα να ταξινομηθεί. Θα θέλαμε να επιλέξουμε ένα υπερεπίπεδο που μεγιστοποιεί το περιθώριο μεταξύ των κλάσεων, η διάσταση του υπερεπίπεδου εξαρτάται από τον αριθμό των χαρακτηριστικών στο σύνολο δεδομένων, η διάσταση θα είναι πάντα ο αριθμός των χαρακτηριστικών μείον ένα. Τα διανύσματα που βρίσκονται στο επίπεδο είναι γνωστά ως 'Support Vectors', αυτά τα σημεία δεδομένων ή διανύσματα είναι η επιπλοκή του συνόλου εκπαίδευσης, με βάση αυτά τα σημεία δεδομένων αποφασίζεται το μέγιστο περιθώριο ή η βεβαιότητα του περιθωρίου. Η έννοια του υπερεπίπεδου είναι εφαρμόσιμη μόνο σε γραμμικά διαχωρίσιμα δεδομένα. Στην περίπτωση μη γραμμικών δεδομένων χρησιμοποιούμε μη γραμμική προσέγγιση πυρήνα για να διαχωρίσουμε τα δεδομένα.

Ο Kemal polat et al. (2008) [114] πρότεινε ένα νέο σύστημα μάθησης καταρράκτη που βασίζεται στη SVM και στη Generalized Discriminant Analysis (GDA). Το προτεινόμενο σύστημα αποτελείται από δύο φάσεις. Στο πρώτο στάδιο, χρησιμοποίησαν τη GDA για να διακρίνουν μεταβλητές χαρακτηριστικών μεταξύ ασθενούς με διαβήτη και υγιούς ασθενούς για τη διαδικασία προεπεξεργασίας. Σε δεύτερο στάδιο, χρησιμοποίησαν τον ταξινομητή SVM, προκειμένου να ταξινομήσουν τα δεδομένα του διαβήτη. Το SVM πέτυχε ακρίβεια ταξινόμησης 78,21% χρησιμοποιώντας 10-πλάσια διασταυρούμενη επικύρωση, το προτεινόμενο σύστημα που ονομάζεται GDA-LS-SVM πέτυχε ακρίβεια ταξινόμησης 82,05% χρησιμοποιώντας 10-πλάσια διασταυρούμενη επικύρωση. Τα πειραματικά αποτελέσματα αξιολογήθηκαν χρησιμοποιώντας ακρίβεια ταξινόμησης, μέθοδο διασταυρούμενης επικύρωσης k-fold και πίνακα σύγχυσης. Επιτεύχθηκε ακρίβεια ταξινόμησης 82,05% και είναι πολλά υποσχόμενο αποτέλεσμα σε σύγκριση με τις προαναφερθείσες τεχνικές ταξινόμησης.

Το 2020, οι Moraes Batista et al. [115] ερεύνησαν την ML για τη διάγνωση και την πρόβλεψη του COVID-19 για ασθενείς έκτακτης ανάγκης. Οι συγγραφείς στήριξαν την έρευνά τους στο ίδιο σύνολο δεδομένων των συγγραφέων του [116] και σε πέντε αλγόριθμους ML (NNs, GBT, RF, SVM, και LR) και εκπαίδευσαν το μοντέλο τους. Το καλύτερο προγνωστικό τους μοντέλο λήφθηκε από τον αλγόριθμο SVM (AUC: 0.85, sensitivity: 0.68, specificity: 0.85) που παρόλο του αποτελέσματος δεν είναι πολύ αξιόπιστος.

## 2.6.Αξιολογήση Μοντέλων Μηχανικής Μάθησης

### 2.6.1.Πίνακας Σύγχυσης

Ένας πίνακας σύγχυσης (Confusion Matrix - CM), όπως υποδηλώνει το όνομα, είναι ένας πίνακας αριθμών που αναφέρει τα σημεία στα οποία ένα μοντέλο ελέγχει εάν αποδίδει ή όχι και κατά πόσο επιτυγχάνεται μια «καλή» πρόβλεψη. Ο ορισμός του αν είναι καλό ή όχι εξαρτάται από τον κλάδο στον οποίο ανήκουν τα δεδομένα αλλά και ποιο είναι το τελικό αποτέλεσμα που θέλει κανείς να πετύχει. Οπότε, ο πίνακας σύγχυσης είναι μια ταξικά κατανομή της προγνωστικής απόδοσης ενός

μοντέλου ταξινόμησης - δηλαδή, ο πίνακας σύγχυσης είναι ένας οργανωμένος τρόπος χαρτογράφησης των προβλέψεων στις αρχικές κλάσεις στις οποίες ανήκουν τα δεδομένα.

Αυτό σημαίνει επίσης ότι οι πίνακες σύγχυσης μπορούν να χρησιμοποιηθούν μόνο όταν η κατανομή εξόδου είναι γνωστή, δηλαδή σε εποπτευόμενα πλαίσια μάθησης. Ο πίνακας σύγχυσης όχι μόνο επιτρέπει τον υπολογισμό της ακρίβειας ενός ταξινομητή, αλλά βοηθά επίσης στον υπολογισμό άλλων σημαντικών μετρήσεων που χρησιμοποιούν συχνά οι προγραμματιστές για να αξιολογήσουν τα μοντέλα τους. Ένας πίνακας σύγχυσης που υπολογίζεται για το ίδιο σύνολο δοκιμών ενός συνόλου δεδομένων, αλλά χρησιμοποιώντας διαφορετικούς ταξινομητές, μπορεί επίσης να βοηθήσει στη σύγκριση των σχετικών δυνατών και αδυναμιών τους και να εξαγάγει ένα συμπέρασμα για το πώς μπορούν να συνδυαστούν προκειμένου για να επιτευχθεί η βέλτιστη απόδοση [117].

Αν και οι έννοιες για τους πίνακες σύγχυσης είναι παρόμοιες ανεξάρτητα από τον αριθμό των κλάσεων στο σύνολο δεδομένων, είναι χρήσιμο να καταστεί κατανοητό ότι πρώτα εξάγεται τον πίνακα σύγχυσης για ένα σύνολο δεδομένων δυαδικών κλάσεων και στη συνέχεια να ενσωματώνονται αυτές οι ιδέες σε σύνολα δεδομένων με δύο (2) ή περισσότερες κλάσεις.

Ένα σύνολο δεδομένων δυαδικής κλάσης είναι ένα σύνολο δεδομένων που αποτελείται μόνο από δύο ξεχωριστές κατηγορίες δεδομένων. Όπως για παράδειγμα, τα επιλεγθέντα σενάρια της εφαρμογής είναι ο διαβήτης και η πνευμονία (covid-19). Στο πρώτο σενάριο ελέγχεται αν θα χορηγηθεί ινσουλίνη ή συνδυασμός ινσουλίνης με κάτι άλλο ως θεραπεία στον ασθενή. Ενώ στο άλλο σενάριο, ελέγχεται αν ο ασθενής εμφανίζει σημάδια ή όχι πνευμονίας. Οπότε, αυτές οι δύο κατηγορίες μπορούν να ονομαστούν "θετικές" και "αρνητικές" για λόγους απλότητας.

Για να γίνει πλήρως κατανοητή η ανάπτυξη του πίνακα σύγχυσης για αυτό το πρόβλημα ταξινόμησης δυαδικών κλάσεων, πρέπει πρώτα να εξοικειωθεί κάποιος με τους ακόλουθους όρους (Εικόνα 10):

- *To True Positive (TP)* αναφέρεται σε ένα δείγμα που ανήκει στη θετική κατηγορία που ταξινομείται σωστά.
- *To True Negative (TN)* αναφέρεται σε ένα δείγμα που ανήκει στην αρνητική κατηγορία που ταξινομείται σωστά.
- *To False Positive (FP)* αναφέρεται σε ένα δείγμα που ανήκει στην αρνητική κατηγορία αλλά έχει ταξινομηθεί εσφαλμένα ως ανήκει στη θετική τάξη.
- *To False Negative (FN)* αναφέρεται σε ένα δείγμα που ανήκει στη θετική τάξη αλλά έχει ταξινομηθεί εσφαλμένα ως ανήκει στην αρνητική τάξη.

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Εικόνα 10: Πίνακας Σύγχυσης

### 2.6.2. Μετρικές Αξιολόγησης

Σε συνδυασμό με την προηγούμενη ενότητα, μέσω των στοιχείων ενός πίνακα σύγχυσης εξάγονται ορισμένες μετρικές ποσοτικής αξιολόγησης και έχουν ως εξής [117]:

- Accuracy: Ο αριθμός των δειγμάτων που ταξινομήθηκαν σωστά από όλα τα δείγματα που υπάρχουν στο σετ δοκιμής.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (6)$$

- Precision: Ο αριθμός των δειγμάτων που ανήκουν στην πραγματικότητα στη θετική κατηγορία από όλα τα δείγματα που προβλεπόταν ότι ήταν της θετικής κατηγορίας από το μοντέλο.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

- Recall/Sensitivity: Ο αριθμός των δειγμάτων που προβλέφθηκε σωστά ότι ανήκουν στη θετική τάξη από όλα τα δείγματα που ανήκουν στην πραγματικότητα στη θετική τάξη.

$$\text{Recall/Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

- F1-score: Ο αρμονικός μέσος όρος των βαθμολογιών ακρίβειας και ανάκλησης που ελήφθησαν για τη θετική τάξη.

$$\text{F - Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

## Κεφάλαιο 3

### Διαλειτουργικότητα Δεδομένων

#### 3.1. Πρότυπα Διαλειτουργικότητας

Τα πρότυπα χαρακτηρίζονται από την γενική ανάπτυξη διαφόρων κανόνων ακολουθώντας ένα συγκεκριμένο υπόδειγμα για την εκάστοτε εφαρμογή ανάλογα τον τομέα εξειδίκευσης. Οπότε ένας ενδεικτικός ορισμός του προτύπου είναι εκείνο το σύνολο προδιαγραφών και κανόνων που περιγράφουν κατά συνέπεια και ακρίβεια για τον τρόπο μεταφοράς ή ακόμα και αποθήκευσης των δεδομένων που ακολουθούν το πρότυπο μέσα από την εφαρμογή κατάλληλων συνθηκών, σχέσεων, απαιτήσεων και των διαφόρων τρόπων για την αναπαράσταση της πληροφορίας. Οπότε με την χρήση των προτύπων παρέχεται ένα προσδιοριστικό για την πρόσβαση στα δεδομένα που εντοπίζονται εντός ενός συστήματος [118]. Ωστόσο, χρήζει σημαντικό να διευκρινιστεί ότι σε περίπτωση που ο τρόπος αποθήκευσης αλλά και μεταφοράς των δεδομένων αλλάξει, τότε δεν μπορεί να παρέχεται η δυνατότητα άντλησης της πληροφορίας. Άρα, με την σωστή χρήση των προτύπων μπορούν να αποφευχθούν διάφορων ειδών προβλήματα, όπως έχουν να κάνουν με ασυμβατότητα, ασάφεια και παρερμηνεία των δεδομένων [119].

Στο συγκεκριμένο κεφάλαιο θα τεκμηριωθούν τα κύρια πρότυπα διαλειτουργικότητας και στην συνέχεια θα πραγματοποιηθεί αναφορά στην ανάπτυξη της διαλειτουργικότητας στον τομέα της υγείας. Στην συνέχεια, θα τεκμηριωθούν τα πιο γνωστά πρότυπα διαλειτουργικότητας στον χώρο της υγείας, όπως είναι το SNOMED, το ICD-10, το HL7 [120]. Να διευκρινιστεί ότι το πρότυπο διαλειτουργικότητας HL7 διαθέτει ένα πλήθος εκδόσεων και στην παρούσα εργασία η ανάπτυξη του προτύπου διαλειτουργικότητας που επιλέχθηκε είναι το HL7 FHIR.

##### 3.1.1. ISO (International Organization for Standardization)

Ο οργανισμός ISO αποτελεί τον μεγαλύτερο οργανισμό στον κόσμο ή καλύτερα έναν από τους πιο δημοφιλείς οργανισμούς για την ανάπτυξη και έκδοση προτύπων με διεθνή πιστοποίηση [120]. Πιο συγκεκριμένα, ο οργανισμός του ISO λειτουργεί ως μια μη κυβερνητική υπηρεσία που ενώνει τον ιδιωτικό τομέα με το δημόσιο τομέα, προκειμένου να παρέχει την δυνατότητα επίτευξης συναίνεσης με λύσεις στις απαιτήσεις των οργανισμών όσο και στις περαιτέρω ανάγκες της κοινωνίας. Από τα βασικότερα και αναγνωρισμένα πρότυπα διαλειτουργικότητας που έχουν εγκριθεί είναι το ISO 15000 και το ISO/IEC 11179-1 [121].

##### **ISO 15000 Electronic business eXtensible Markup Language (ebXML)**

Η νέα ονομασία ISO 15000, υπό τον γενικό τίτλο, Electronic Business eXtensible Mark-up Language (ebXML), περιλαμβάνει τέσσερα (4) μέρη, καθένα από τα οποία αντιστοιχεί σε ένα από τα αρθρωτά πρότυπα της ebXML [122]:

- **ISO 15000-1:** Σύμβαση προφίλ συνεργατών ebXML
- **ISO 15000-2:** Προδιαγραφή υπηρεσίας μηνυμάτων ebXML
- **ISO 15000-3:** Μοντέλο πληροφοριών μητρώου ebXML



- **ISO 15000-4:** Προδιαγραφή υπηρεσιών μητρώου ebXML

Μέχρι τώρα, η τεχνολογία που ήταν διαθέσιμη για τις περισσότερες επιχειρήσεις για την ανταλλαγή δεδομένων ήταν η ηλεκτρονική ανταλλαγή δεδομένων (Electronic Data Interchange - EDI), η οποία συνέβαλε σημαντικά στην παραγωγικότητα και τον έλεγχο των αποθεμάτων. Πολλές εταιρείες, ωστόσο, θεωρούν ότι το EDI είναι ακριβό και δύσκολο να εφαρμοστεί. Η πρωτοβουλία ebXML, χρησιμοποιώντας τις οικονομίες κλίμακας που παρουσιάζονται από το διαδίκτυο, ξεπερνά αυτά τα εμπόδια.

Η ebXML παρέχει στις εταιρείες μια τυπική μέθοδο ανταλλαγής επιχειρηματικών μηνυμάτων, διεξαγωγής εμπορικών σχέσεων, επικοινωνίας δεδομένων με κοινούς όρους και καθορισμού και καταχώρησης επιχειρηματικών διαδικασιών [123]. Στοχεύει στο να διευκολύνει τους οργανισμούς να διασυνδέονται με άλλους εντός και εκτός του κλάδου τους, να ανοίξουν νέες αγορές με λιγότερη προσπάθεια από πριν και, ταυτόχρονα, να μειώσει το κόστος και να απλοποιήσει τη διαδικασία που σχετίζεται με την παραδοσιακή ανταλλαγή εγγράφων.

Οι πιο βασικές δυνατότητες της ebXML, που αποτελούν σημαντικά στοιχεία για την διεξαγωγή επικοινωνιών στο ηλεκτρονικό εμπόριο μεταξύ των διαφορετικών επιχειρηματικών πλευρών. Συνεπώς, το ebXML περιλαμβάνει πέντε (5) τύπους προδιαγραφών που μπορούν να χρησιμοποιηθούν είτε μεμονωμένα είτε από κοινού και αυτοί ορίζονται ως εξής [124]:

- **Επιχειρηματική Διαδικασία (Business Process):** Το πρωτόκολλο μηνυμάτων περιγράφει πώς ανταλλάσσονται διαφορετικά κομμάτια πληροφοριών μεταξύ συνεργατών σε ένα επιχειρηματικό περιβάλλον.
- **Προφίλ Πρωτοκόλλου Συνεργασίας και Συμφωνίας (Collaboration Protocol Profile and Agreement):** Όταν οι άνθρωποι εμπορεύονται αγαθά ή υπηρεσίες, συχνά το κάνουν μέσω ηλεκτρονικών μηνυμάτων. Αυτό συμβαίνει γιατί είναι πιο εύκολο από το να συναντηθείτε προσωπικά και συνήθως είναι πιο γρήγορο από το να στείλετε μια επιστολή. Όταν δύο άτομα συναλλάσσονται μεταξύ τους μέσω ηλεκτρονικών μηνυμάτων, λέγεται ότι συνεργάζονται μέσω του ηλεκτρονικού επιχειρείν. Το τμήμα επικοινωνίας είναι σημαντικό γιατί σας βοηθά να δημιουργήσετε αλληλεπιδράσεις και σχέσεις με άλλα άτομα. Περιλαμβάνει λεπτομερείς πληροφορίες σχετικά με τα άτομα με τα οποία επικοινωνείτε, την ταυτότητά τους, τα στοιχεία επικοινωνίας τους και επιλογές ασφάλειας και εμπιστοσύνης.
- **Υπηρεσίες Μηνυμάτων (Messaging Services):** Η εταιρεία βοηθά στη μεταφορά και δρομολόγηση ηλεκτρονικών επιχειρηματικών μηνυμάτων μέσω του Διαδικτύου χρησιμοποιώντας τυπικές τεχνολογίες. Η προδιαγραφή ebMS (Employee Benefit Management Services) περιγράφει έναν τρόπο ασφαλούς αποστολής μηνυμάτων μεταξύ επιχειρήσεων. Ένα σύστημα αρχείων είναι ένα σύστημα για την οργάνωση και αποθήκευση αρχείων σε έναν υπολογιστή. Αυτό το σύστημα βοηθά στη διασφάλιση της αξιόπιστης μεταφοράς σημαντικών επιχειρηματικών πληροφοριών μεταξύ των υπολογιστών. Το ebMS είναι ένα σύστημα που βοηθά τις επιχειρήσεις να ανταλλάσσουν ηλεκτρονικά μηνύματα χρησιμοποιώντας τις ίδιες τυπικές μορφές. Αυτά τα μηνύματα χρησιμοποιούνται για διαφορετικούς σκοπούς. Ορισμένα μηνύματα μπορεί να χρησιμοποιηθούν για επικοινωνία με τον διακομιστή και άλλα για επικοινωνία με άλλους χρήστες. Οι εφαρμογές μπορούν να επικοινωνούν μεταξύ τους και με άλλα συστήματα για

να ολοκληρώσουν τα πράγματα. Για παράδειγμα, μια εφαρμογή μπορεί να ζητήσει μια υπηρεσία από άλλη εφαρμογή ή σύστημα. Αυτή η υπηρεσία μπορεί να βρίσκεται στην ουρά και να περιμένει να χρησιμοποιηθεί. Η εφαρμογή ενδέχεται να βρίσκεται σε διαδικασία ροής εργασίας που έχει ρυθμιστεί από κάποιον άλλο. Η εφαρμογή μπορεί να αναμένεται να κάνει κάτι συγκεκριμένο που έχει ήδη γίνει σε άλλα μέρη της επιχείρησης. Η εφαρμογή μπορεί να δρομολογηθεί μέσω ενός διαύλου υπηρεσιών, έτσι ώστε να μπορεί να φτάσει στα μέρη της εφαρμογής που χρησιμοποιούν πραγματικά οι χρήστες.

- **Μητρώα και Αρχεία Φύλαξης (Registries and Repositories):** Τα διαλειτουργικά μητρώα είναι ένας τρόπος παρακολούθησης διαφορετικών πληροφοριών (όπως διευθύνσεις), έτσι ώστε οι άνθρωποι να μπορούν να τις βρουν και να τις χρησιμοποιούν εύκολα. Τα μητρώα είναι επίσης υπεύθυνα για τη διατήρηση των πληροφοριών και τη διάθεσή τους σε άλλους. Αυτό σημαίνει ότι τα μητρώα μπορούν να είναι αξιόπιστα για να διατηρούν τις πληροφορίες ακριβείς και ενημερωμένες. Τέλος, οι θεματοφύλακες είναι υπεύθυνοι να διασφαλίζουν ότι οι πληροφορίες στα μητρώα είναι προσβάσιμες από τα κατάλληλα άτομα τη σωστή στιγμή.
- **Βασικά Στοιχεία (Core Components):** Υπάρχει ένας τρόπος για να δημιουργήσετε ένα σύνολο κοινών δομικών στοιχείων για την αναπαράσταση γενικών τύπων στα δεδομένα που λαμβάνονται από έναν οργανισμό. Αυτός ο τρόπος θα σας βοηθήσει να βεβαιωθείτε ότι τα δεδομένα σας είναι εύκολα κατανοητά και χρήσιμα. Παρέχουν επίσης μεθόδους για τη δημιουργία νέων σημασιολογικών λέξεων, αλλά και για την ανακατασκευή υφιστάμενων.

### 3.1.2.W3C (World Wide Web Consortium)

Η Κοινοπραξία του Παγκόσμιου Ιστού (W3C) [125] είναι μια ομάδα ειδικών που δημιουργούν πρότυπα που βοηθούν να γίνει ο Ιστός πιο διαλειτουργικός. Το κάνουν αυτό αναπτύσσοντας κατευθυντήριες γραμμές, εργαλεία και λογισμικό για να διασφαλίσουν ότι οι τεχνολογίες Ιστού συνεργάζονται. Επίσης, συνεργάζονται για να δημιουργήσουν τεχνολογίες που θα κάνουν το Διαδίκτυο πιο χρήσιμο και προσβάσιμο για όλους. Το W3C επιδιώκει να διασφαλίσει ότι διαφορετικές συσκευές και λογισμικό μπορούν να συνεργαστούν για πρόσβαση στον Ιστό, έτσι ώστε οι άνθρωποι να μην χρειάζεται να ανησυχούν για τις διαφορετικές συσκευές και το λογισμικό που χρησιμοποιούν διασπώντας την αγορά [125].

Οι τεχνολογίες του W3C εμπίπτουν σε κατηγορίες όπως Εφαρμογές Ιστού, Πρόσβαση στον Ιστό για φορητές συσκευές, Φωνή, Υπηρεσίες Ιστού, Σημασιολογικός Ιστός και Απόρρητο και Ασφάλεια. Αυτές οι τεχνολογίες βασίζονται στις θεμελιώδεις γλώσσες και πρωτόκολλα του W3C, όπως Uniform Resource Identifiers (URIs), Hypertext Transfer Protocol (HTTP), XML και χρησιμοποιούνται για την ανάπτυξη εφαρμογών που πληρούν τις απαιτήσεις προσβασιμότητας, διεθνοποίησης, ανεξαρτησίας συσκευών, πρόσβασης από κινητά και διασφάλισης ποιότητας [125].

Σε αυτό το σημείο θα παρουσιαστούν κάποια εκ των πιο γνωστών προτύπων διαλειτουργικότητας και οι αντίστοιχες προδιαγραφές τους είναι ως εξής [126]:

#### **XML (eXtensible Markup Language)**

Η XML είναι μια γλώσσα που χρησιμοποιείται για τη δημιουργία εγγράφων που μπορούν να διαβαστούν και να χρησιμοποιηθούν από άλλα προγράμματα.

### ***XML Schema (eXtensible Markup Language Schema)***

Ένα XML Schema είναι μια περιγραφή του τύπου ενός αρχείου XML, εκφράζοντας συνήθως τους περιορισμούς της δομής και του περιεχομένου του εγγράφου XML. Αυτοί οι περιορισμοί εκφράζονται χρησιμοποιώντας κανόνες γραμματικής που περιορίζουν το περιεχόμενο και τον τρόπο εμφάνισης των στοιχείων του εγγράφου XML.

### ***XSL (eXtensible Stylesheet Language)***

Το XSL είναι ένα εργαλείο που μπορεί να χρησιμοποιηθεί για την αλλαγή ενός εγγράφου XML σε άλλο έγγραφο XML ή σε μορφή κατανοητή από ένα πρόγραμμα περιήγησης Ιστού.

### ***XPath (XML Path Language)***

Η XPath είναι μια γλώσσα που σας επιτρέπει να πλοηγηθείτε σε ένα δέντρο εγγράφων XML. Μπορείτε να το χρησιμοποιήσετε για να επιλέξετε κόμβους με βάση μια ποικιλία κριτηρίων.

### ***SOAP (Simple Object Access Protocol)***

Το SOAP είναι ένα πρωτόκολλο που επιτρέπει την επικοινωνία μεταξύ διαφορετικών εφαρμογών σε ένα δίκτυο. Η αποστολή μηνυμάτων με χρήση XML σας επιτρέπει να τα οργανώσετε με τρόπο που είναι εύκολο να διαβάσετε και να κατανοήσετε. Χρησιμοποιεί επίσης άλλα πρωτόκολλα, όπως HTTP και SMTP, για να σας βοηθήσει να τα στείλετε στο σωστό άτομο. Το SOAP δεν απαιτεί συγκεκριμένο μοντέλο υλοποίησης ή προγραμματισμού, επομένως διαφορετικές εταιρείες μπορούν να δημιουργήσουν τις δικές τους υλοποιήσεις. Το σύστημα μονάδων σας επιτρέπει να πακετάρετε τον κώδικά σας με τρόπο που τον καθιστά εύκολο στην κατανόηση και τη χρήση του. Σας επιτρέπει επίσης να τοποθετείτε δεδομένα μέσα στις μονάδες κώδικα σας, ώστε να είναι εύκολη η πρόσβαση και η χρήση τους. Συνεπώς, το SOAP είναι ένα είδος στοιχείου που μπορεί να χρησιμοποιηθεί σε πολλά διαφορετικά συστήματα.

### ***WSDL (Web Services Description Language)***

Η γλώσσα περιγραφής των υπηρεσιών Ιστού (WSDL) είναι ένας τρόπος περιγραφής των υπηρεσιών Ιστού σε μορφή XML. Αυτό επιτρέπει στους άλλους να κατανοήσουν πώς λειτουργούν οι υπηρεσίες Ιστού και πώς να συνδεθούν σε αυτές. Οι λειτουργίες και τα μηνύματα περιγράφονται με γενικούς όρους και στη συνέχεια συνδέονται σε ένα συγκεκριμένο πρωτόκολλο δικτύου και μορφή μηνύματος για να ορίσουν ένα τελικό σημείο. Πολλές διαφορετικές υπηρεσίες συνδυάζονται για να δημιουργήσουν μια ολοκληρωμένη υπηρεσία.

### **3.1.3.CEN (European Committee for Standardization)**

Το CEN (ή το Ευρωπαϊκό Ινστιτούτο Προτύπων) είναι ένας μη κερδοσκοπικός οργανισμός που συμβάλλει στην προώθηση της ευρωπαϊκής οικονομίας διασφαλίζοντας ότι υπάρχουν συντονισμένα πρότυπα και προδιαγραφές μεταξύ οργανισμών και χωρών [126]. Λειτουργεί επίσης για να διασφαλίσει τη διατήρηση της ευημερίας των ευρωπαίων πολιτών και του περιβάλλοντος. Οι διεπαφές επιχειρησιακής διαλειτουργικότητας (Business Interoperability

Interface - BII), όπως και τα workshops είναι ένας τρόπος για να βοηθήσετε τις επιχειρήσεις αλλά και τους οργανισμούς στην Ευρώπη να ανταλλάσσουν πληροφορίες ομαλά και εύκολα. Ο κύριος στόχος του workshop είναι να δημιουργήσει ένα βασικό πλαίσιο για το πώς οι επιχειρήσεις/οργανισμοί μπορούν να συνεργαστούν σε όλη την Ευρώπη για την επίτευξη της διαλειτουργικότητας στα ανταλλασσόμενα δεδομένα.

Ειδικότερα, οι στόχοι των workshops καταγράφονται ως εξής [125]:

- Τα πρότυπα UN/CEFACT και UBL μπορούν να χρησιμοποιηθούν για να βοηθήσουν τις επιχειρήσεις να συνεννοούνται καλύτερα.
- Εντοπισμός και τεκμηρίωση των επιχειρηματικών διεπαφών που είναι απαραίτητες για τις πανευρωπαϊκές ηλεκτρονικές συναλλαγές στις δημόσιες συμβάσεις.
- Υποστήριξη προγραμμάτων που συνεργάζονται χρησιμοποιώντας συγκεκριμένες, τυποποιημένες τεχνικές προδιαγραφές. Αυτό θα διευκολύνει τη συνεργασία των προγραμμάτων και θα ξεπεράσει τυχόν τεχνικές δυσκολίες που θα μπορούσαν να τα εμποδίσουν να συνεργαστούν.

### 3.2. Διαλειτουργικότητα στην Υγεία

Διαλειτουργικότητα σημαίνει ότι τα συστήματα και οι συσκευές μπορούν να μοιράζονται δεδομένα εύκολα και να τα κατανοούν σωστά. Για να είναι δυνατή η επικοινωνία μεταξύ τους, δύο συστήματα πρέπει να μπορούν να μοιράζονται δεδομένα. Στη συνέχεια, μόλις γίνει κοινή χρήση των δεδομένων, τα συστήματα πρέπει να διευκολύνουν τους χρήστες να τα κατανοήσουν [126]. Επίσης, η διαλειτουργικότητα είναι η ικανότητα διαφορετικών συστημάτων πληροφορικής και λογισμικού να ανταλλάσσουν και να μοιράζονται δεδομένα από ποικίλες πηγές, συμπεριλαμβανομένων εργαστηρίων, κλινικών, φαρμακείων, νοσοκομείων και ιατρικών πρακτικών. Ακόμα, η διαλειτουργικότητα είναι ένας τρόπος σκέψης για το πώς συνεργάζονται διαφορετικά συστήματα υπολογιστών, αντί να βασίζεται σε διαφορετικά σύνολα συγκεκριμένων, τεχνικών προτύπων [127]. Επομένως, αυτό σημαίνει ότι οι γιατροί και οι νοσηλευτές μπορούν να συνεργαστούν για να βρουν τον καλύτερο τρόπο θεραπείας των ασθενών. Επίσης, μπορούν να αφιερώσουν περισσότερο χρόνο στις λεπτομέρειες της ασθένειας ή του τραυματισμού, πράγμα που σημαίνει ότι οι ασθενείς μπορούν να λάβουν την καλύτερη δυνατή φροντίδα. Η διαλειτουργικότητα της υγειονομικής περίθαλψης σημαίνει ότι οι άνθρωποι μπορούν εύκολα να έχουν πρόσβαση σε σημαντικές πληροφορίες σχετικά με την υγεία τους από διαφορετικές πηγές, αλλά λειτουργεί μόνο εάν οι εταιρείες ιατρικού λογισμικού συμφωνούν να μοιράζονται πληροφορίες στην αποκλειστική μορφή ενός συστήματος. Το σύστημα υγειονομικής περίθαλψης πρέπει να είναι σε θέση να μοιράζεται πληροφορίες ασθενών μεταξύ διαφορετικών βάσεων δεδομένων, προκειμένου οι άνθρωποι να έχουν πρόσβαση σε αυτές από διαφορετικές τοποθεσίες.

Οι πληροφορίες υγειονομικής περίθαλψης είναι συχνά πολύπλοκες, πράγμα που σημαίνει ότι είναι δύσκολο να τις μοιραστείτε μεταξύ διαφορετικών οργανισμών. Αυτός είναι ο λόγος για τον οποίο είναι σημαντικό να λαμβάνονται υπόψη οι διαφορετικοί τύποι ατόμων και οργανισμών που εμπλέκονται όταν προσπαθούμε να κάνουμε τις πληροφορίες υγειονομικής περίθαλψης

διαλειτουργικές. Οπότε παρακάτω αναφέρονται οι φορείς/οργανισμοί που συμμετέχουν στην ανάπτυξη ενός δυσλειτουργικού συστήματος:

- **Εταιρείες υγειονομικής περίθαλψης:** Η βιομηχανία ιατρικού εξοπλισμού περιλαμβάνει εταιρείες που πωλούν κλινικά συστήματα, συστήματα διαχείρισης πληροφορικής και ιατρικές συσκευές. Αυτή είναι μια κατηγορία επιχειρήσεων που έχουν πολλούς παγκόσμιους στόχους και μικρότερες επιχειρήσεις που επικεντρώνονται σε μια συγκεκριμένη περιοχή αγοράς.
- **Πάροχοι υπηρεσιών υγείας:** Οι γιατροί είναι άνθρωποι που εργάζονται στον τομέα της υγείας. Έχουν διαφορετικά επαγγέλματα, όπως γιατροί της ιατρικής (γιατροί που ειδικεύονται στη θεραπεία προβλημάτων υγείας), γιατροί οστεοπαθητικής (γιατροί που ειδικεύονται στη θεραπεία πόνου και διαταραχών του μυοσκελετικού συστήματος) και γιατροί οδοντιατρικής (γιατροί που ειδικεύονται στη θεραπεία προβλημάτων με δόντια και ούλα). Γιατροί, νοσηλευτές και τεχνικοί βοηθούν άτομα που είναι άρρωστα. Υπάρχουν πολύ λίγοι άνθρωποι που έχουν όλες τις απαραίτητες δεξιότητες για να εργαστούν σε έργα επιστήμης και τεχνολογίας υπολογιστών και τον χρόνο που χρειάζεται για να αφιερώσουν πόρους για την αντιμετώπιση ζητημάτων διαλειτουργικότητας ή τη συμμετοχή στην τυποποίηση της πληροφορικής και των προτύπων υγείας.
- **Προσωπικό πληροφορικής και διοικητικών στελεχών:** Τα μεγάλα ιδρύματα υγειονομικής περίθαλψης διαδραματίζουν σημαντικό ρόλο στην παροχή υγειονομικής περίθαλψης σε πολλούς ανθρώπους, αλλά δεν είναι πάντα η καλύτερη επιλογή για άτομα που χρειάζονται υγειονομική περίθαλψη σε ένα μικρό περιβάλλον. Πολύ λίγοι άνθρωποι εργάζονται για την ανάπτυξη προτύπων ή την ενσωμάτωση της τεχνολογίας στα συστήματα των διοικητικών στελεχών.
- **Υγειονομικές αρχές και κυβερνήσεις:** Το εθνικό σύστημα υγείας περιλαμβάνει διαφορετικούς οργανισμούς που διαχειρίζονται αποφάσεις και υπηρεσίες υγειονομικής περίθαλψης για όλους στη χώρα. Με τις εξελισσόμενες τεχνολογίες και πρότυπα, οι κυβερνήσεις και οι αρχές συνήθως επιβάλλουν τη διαλειτουργικότητα μέσω εθνικών προτύπων και στις περισσότερες περιπτώσεις, έχει αποδειχθεί αποτελεσματική.

Το σημερινό τοπίο σημαίνει ότι υπάρχουν πολλά διαφορετικά μέρη για να πάτε και πράγματα να κάνετε. Σήμερα, οι πάροχοι αντιμετωπίζουν πολλές προκλήσεις όσον αφορά την κοινή χρήση δεδομένων: περιορισμοί στην κοινή χρήση δεδομένων, έλλειψη τυποποιημένων μορφών δεδομένων σε διαφορετικά συστήματα, παλαιού τύπου συστήματα που αποθηκεύουν σημαντικά δεδομένα, ακριβά τέλη διασύνδεσης και μειωμένες αποδόσεις. Ο συνδυασμός αυτών των πραγμάτων μαζί σημαίνει ότι πρέπει να βρούμε τρόπους για να βελτιώσουμε τη φροντίδα που βασίζεται στο καλύτερο για τον ασθενή. Αυτό είναι ιδιαίτερα σημαντικό καθώς προσπαθούμε να αντιμετωπίσουμε τις προκλήσεις της βελτίωσης της φροντίδας. Τα εργαστήρια και τα συστήματα υγείας καταβάλλουν πολλά χρήματα και προσπάθεια για τη δημιουργία ενός δικτύου ώστε να μπορούν να επικοινωνούν με άλλα εργαστήρια και συστήματα υγείας. Τα συστήματα υγείας έχουν πολύπλοκες απαιτήσεις που μπορούν να καθυστερήσουν την εφαρμογή τους και να αυξήσουν το κόστος προσλαμβάνοντας ακριβό προσωπικό για την ολοκλήρωση έργων. Κάθε φορά που συνεργάζονται για να δημιουργήσουν νέες διεπαφές και να εκτελέσουν νέες υλοποιήσεις αυτών των ιδεών, χρειάζεται πολύς χρόνος και προσπάθεια για να αντιγράψετε όσα έχουν γίνει στο παρελθόν.

Τα οφέλη ωστόσο της διαλειτουργικότητας στον τομέα της υγείας είναι πολλά [127]. Ορισμένα από αυτά αναφέρονται παρακάτω:

- **Βελτιωμένη απόδοση:** Η διαλειτουργικότητα των συστημάτων υγειονομικής περίθαλψης σημαίνει ότι συνεργάζονται πιο αποτελεσματικά, ώστε οι άνθρωποι να μπορούν να λαμβάνουν τη φροντίδα που χρειάζονται πιο γρήγορα. Όταν τα δεδομένα εμφανίζονται με τρόπο εύκολα προσβάσιμο, οι επαγγελματίες μπορούν να δουν γρήγορα τις σημαντικές λεπτομέρειες προκειμένου να λάβουν τεκμηριωμένες αποφάσεις.
- **Ασφαλότερες μεταβάσεις φροντίδας:** Η συνέχεια της περίθαλψης είναι σημαντική για ασθενείς με χρόνιες παθήσεις ή που βρίσκονται σε κίνδυνο, είτε λαμβάνουν φροντίδα από μία υπηρεσία υγείας είτε από πολλές. Η διαλειτουργικότητα σημαίνει ότι διαφορετικοί πάροχοι υγειονομικής περίθαλψης μπορούν να συνεργαστούν για να παρέχουν καλύτερη φροντίδα στους ασθενείς. Αυτό βοηθά στην αποφυγή λαθών και στη βελτίωση των αποτελεσμάτων για όλους τους εμπλεκόμενους. Εάν ένας ασθενής είναι σε διακοπές και αρρωστήσει, μπορεί να μην είναι σε θέση να ενημερώσει τον γιατρό για το ιατρικό του ιστορικό. Αυτό μπορεί να είναι επικίνδυνο για τον ασθενή και για τον γιατρό. Πρόσφατα, ένα άρθρο στο *Medical Economics* ανέφερε μια περίπτωση όπου ένας ασθενής δεν μπορούσε να δώσει περισσότερες πληροφορίες σχετικά με τη χειρουργική επέμβαση. Εάν ο γιατρός είχε πρόσβαση στις πληροφορίες υγείας του ασθενούς από το ΗΦΥ στο σπίτι του, θα μπορούσε να αντιμετωπίσει το πρόβλημα υγείας του πιο γρήγορα και ολοκληρωμένα.
- **Μείωση κόστους:** Η διαλειτουργικότητα σημαίνει ότι πιο χρήσιμες πληροφορίες μπορούν να κοινοποιηθούν γρήγορα. Αυτό σημαίνει ότι τα δεδομένα από έναν ασθενή που έκανε εξέταση αίματος την περασμένη εβδομάδα μπορούν να χρησιμοποιηθούν σε άλλο χώρο, γεγονός που θα εξοικονομήσει χρόνο και χρήμα από τη διενέργεια νέων και περιττών εξετάσεων στο νοσοκομείο. Με την πιο αποτελεσματική κοινή χρήση πληροφοριών, μπορούμε να εξοικονομήσουμε χρόνο στις εξετάσεις μας. Αυτό, με τη σειρά του, σημαίνει περισσότερο χρόνο για να αφιερώσουμε σε άλλα πράγματα, όπως τη μάθηση. Η μείωση του χρόνου εξέτασης είναι σημαντική για τη μελλοντική υγεία ενός συστήματος υγείας. Με τη σύνδεση με άλλες επιχειρήσεις με έξυπνο τρόπο, μπορείτε να εξοικονομήσετε χρήματα. Ένα σύστημα υγειονομικής περίθαλψης που επιτρέπει σε όλες τις κλινικές συναλλαγές να πραγματοποιούνται μέσω ενός κεντρικού κόμβου είναι διαθέσιμο σήμερα. Αυτή η τεχνολογία επιτρέπει σε νοσοκομεία, εργαστήρια, ακτινολογικά κέντρα και κλινικές να χρησιμοποιούν τα ίδια εργαλεία και διεπαφές, εξοικονομώντας χρόνο και χρήμα. Αντί να δημιουργούν και να υλοποιούν προσαρμοσμένες διεπαφές, εστιάζουν στην εύκολη διαμόρφωση επιλογών από αναπτυσσόμενες λίστες, επιτρέποντας αστραπιαίες υλοποιήσεις - μερικές φορές μέσα σε λίγα λεπτά. Χρησιμοποιώντας μια πλατφόρμα ανάπτυξης front-end, η επιχείρησή σας μπορεί να εξοικονομήσει χρήματα σε μισθούς για προγραμματιστές front-end με υψηλούς μισθούς και ακριβές χρεώσεις.

### **Εμπόδια Διαλειτουργικότητας**

Τρεις κατηγορίες εμποδίων (εννοιολογικών, τεχνολογικών και οργανωτικών) προσδιορίζονται κατά την επίτευξη της διαλειτουργικότητας στον τομέα της Υγείας [126]:

- **Εννοιολογικά εμπόδια:** Τα εμπόδια αφορούν τους διαφορετικούς τρόπους με τους οποίους μπορούν να κοινοποιηθούν οι πληροφορίες. Αυτά τα εμπόδια αφορούν τον τρόπο με τον οποίο μοντελοποιούνται μοντέλα σε υψηλό επίπεδο αφαίρεσης (όπως τα επιχειρηματικά μοντέλα μιας επιχείρησης) και σε επίπεδο προγραμματισμού (όπως μοντέλα XML).
- **Τεχνολογικά εμπόδια:** Αυτές οι προκλήσεις προκαλούνται από τις διαφορές στον τρόπο λειτουργίας των διαφορετικών τεχνολογιών. Αυτά τα προβλήματα αφορούν τον τρόπο με τον οποίο οι υπολογιστές αποθηκεύουν, ανταλλάσσουν, επεξεργάζονται και επικοινωνούν δεδομένα.
- **Οργανωτικά εμπόδια:** Αυτά τα εμπόδια σχετίζονται με το ερώτημα ποιος είναι υπεύθυνος για τι και ποιος έχει την εξουσία να κάνει τι σε έναν οργανισμό. Αυτά τα ζητήματα μπορεί να είναι δύσκολο να ξεπεραστούν, καθώς διαφορετικές οργανωτικές δομές μπορεί να είναι ασυμβίβαστες.

Ανεξάρτητα από τις κατηγορίες αυτές, οι βασικές δυσκολίες και εμπόδια διαφαίνονται στα ακόλουθα [126]:

- **Ανεπαρκή πρότυπα δεδομένων:** Υπάρχουν πρότυπα για τον τρόπο με τον οποίο τα ηλεκτρονικά αρχεία υγείας πρέπει να επικοινωνούν μεταξύ τους, αλλά μερικοί άνθρωποι πιστεύουν ότι αυτά τα πρότυπα δεν αρκούν για να κάνουν τα αρχεία αυτά να συνεργάζονται. Η διαλειτουργικότητα δεδομένων είναι σημαντική, ώστε όλοι να μπορούν να μοιράζονται εύκολα πληροφορίες για την υγεία. Η επίτευξη ευρείας υιοθέτησης προτύπων δεδομένων είναι το πρώτο βήμα σε αυτή τη διαδικασία. Ορισμένοι οργανισμοί εξακολουθούν να πιστεύουν ότι είναι σημαντικό να διατηρούνται οι πληροφορίες των ασθενών σε τοπικό επίπεδο. Αυτό συμβαίνει επειδή μερικοί άνθρωποι πιστεύουν ότι αυτός είναι ο καλύτερος τρόπος προστασίας των πληροφοριών. Ωστόσο, ορισμένοι ειδικοί διαφωνούν. Μερικοί άνθρωποι μπορούν να μοιραστούν, αλλά δεν έχουν τη δυνατότητα να το κάνουν. Πολλοί ειδικοί έχουν τοποθετήσει το HL7 FHIR ως το πρωτόκολλο που θα επιτρέψει στους προμηθευτές να αγκαλιάσουν την απρόσκοπτη ανταλλαγή δεδομένων στο εγγύς μέλλον. Άλλες ομάδες εργάζονται για τρόπους για να διευκολύνουν τους ανθρώπους να μοιράζονται πληροφορίες με ασφάλεια. Το DirectTrust [127] το κάνει αυτό καθιστώντας εύκολη την αποστολή μηνυμάτων μεταξύ τους γρήγορα και εύκολα. Οι προμηθευτές και οι ιδιωτικές συνεργασίες εργάζονται για να δημιουργήσουν ένα πιο διαλειτουργικό οικοσύστημα, ώστε οι πάροχοι και οι ασθενείς να μπορούν να επωφεληθούν. Θα χρειαστεί λίγος χρόνος για να ωριμάσουν αυτές οι προσπάθειες. Θα πρέπει να είστε υπομονετικοί και να περιμένετε να τεθούν σε ισχύ.
- **Ανωριμότητα ΕΦΥ:** Όταν τα συστήματα υγείας συνεργάζονται, συχνά μπορούν να επιτύχουν πιο διαλειτουργική φροντίδα χρησιμοποιώντας ένα σύστημα παραπομπής. Αυτό σημαίνει ότι όταν ένα σύστημα γνωρίζει για ένα πρόβλημα υγείας ή ανάγκη, μπορεί εύκολα να παραπέμψει τον ασθενή σε ένα άλλο σύστημα που μπορεί να παρέχει καλύτερη φροντίδα. Τα συστήματα παραπομπής διευκολύνουν τους παρόχους υγειονομικής περίθαλψης να παραπέμπουν τους ασθενείς σε άλλους παρόχους. Αυτό βοηθά στη δημιουργία ισχυρών σχέσεων μεταξύ όλων των εμπλεκόμενων παρόχων, γεγονός που βοηθά στον συντονισμό και τον εξορθολογισμό της φροντίδας των ασθενών. Αυτό το σύστημα παραπομπής είναι συνδεδεμένο με έναν ΗΦΥ, επομένως μπορεί να

χρησιμοποιηθεί για την κοινή χρήση ιατρικών δεδομένων μεταξύ γιατρών και νοσοκομείων. Οι γιατροί και άλλοι πάροχοι υγειονομικής περίθαλψης δυσκολεύονται να λάβουν τα ηλεκτρονικά τους αρχεία υγείας (ΗΦΥ) να λειτουργούν καλά και να τους βοηθούν να διαχειρίζονται την υγεία των ασθενών τους, αλλά αυτό δεν είναι πάντα πρόβλημα. Μερικές φορές είναι απλώς θέμα να ζητηθεί από τους γιατρούς και άλλους παρόχους υγειονομικής περίθαλψης να μάθουν πώς να χρησιμοποιούν σωστά τα ΗΦΥ. Ωστόσο τα ηλεκτρονικά μητρώα υγείας (ΗΜΥ) εξακολουθούν να αναπτύσσονται, ενώ οι πάροχοι συχνά περιορίζονται από απαιτήσεις ποιοτικής αναφοράς, πρωτοβουλίες βελτίωσης ιδέας, συναγερμούς και ειδοποιήσεις που διακόπτουν τις ροές εργασίας και κανόνες ακεραιότητας δεδομένων. Τα δεδομένα μπορεί να είναι ελλιπή μόνο επειδή είναι συχνά ελλιπή. Πολλοί ασθενείς και κλινικοί ιατροί παραπονιούνται για την έλλειψη καλών αντιστοιχιών φαρμάκων, τη δύσκολη πρόσβαση στα αποτελέσματα των εξετάσεων και τις περίπλοκες διοικητικές διαδικασίες που καθιστούν δύσκολη την παροχή προσωπικής, προσεκτικής και ουσιαστικής φροντίδας.

- **Διακυμάνσεις των κανόνων περί απορρήτου του κράτους:** Εάν υπάρχουν διαφορετικοί κανόνες σχετικά με το πόσες πληροφορίες μπορούν να μοιραστούν οι άνθρωποι στο διαδίκτυο σε διαφορετικές πολιτείες ή από κράτος σε κράτος, τότε οι νομοθέτες θα μπορούσαν να βοηθήσουν ώστε οι κανόνες να γίνουν πιο συνεπείς. Μερικοί άνθρωποι λένε ότι είναι δύσκολο να μοιράζεστε πληροφορίες υγείας με άλλους παρόχους σε άλλες πολιτείες, έτσι ώστε να μπορούν να χρησιμοποιηθούν για να διατηρηθούν συνεπείς τα πράγματα σε ολόκληρη τη χώρα όσον αφορά τα ΗΜΥ. Εάν θέλετε να μοιραστείτε πληροφορίες υγείας με άλλους παρόχους, πρέπει πρώτα να λάβετε την άδεια των ασθενών. Διαφορετικά, οι πάροχοι ενδέχεται να είναι απρόθυμοι να μοιραστούν πληροφορίες μαζί σας. Ορισμένες από τις πιο σημαντικές πληροφορίες για κοινή χρήση είναι επίσης ευαίσθητες. Ορισμένες πολιτείες απαιτούν πρόσθετη άδεια για τους παρόχους να μοιράζονται δεδομένα ή πληροφορίες υγείας και τα ηλεκτρονικά εργαλεία δεν είναι αρκετά εξελιγμένα για να χειρίζονται σωστά αυτές τις εργασίες. Μια έρευνα διαπίστωσε ότι ορισμένοι άνθρωποι ανησυχούν για την κοινή χρήση πληροφοριών υγείας χωρίς την άδεια του ατόμου, επειδή θα μπορούσε να παραβιάσει το απόρρητό τους. Όταν οι άνθρωποι είναι προσεκτικοί, μπορεί να είναι δύσκολο να μοιραστούν πληροφορίες για την υγεία επειδή μπορεί να είναι δύσκολο να εμπιστευτούν ο ένας τον άλλον. Ωστόσο, όταν οι άνθρωποι έχουν μεγαλύτερη εμπιστοσύνη, μπορούν να επωφεληθούν από την ανταλλαγή πληροφοριών για την υγεία.
- **Ανεπαρκείς δυνατότητες αντιστοίχισης ασθενών:** Η αντιστοίχιση των ασθενών με τη σωστή θεραπεία είναι δύσκολη λόγω του τρόπου με τον οποίο έχει σχεδιαστεί σήμερα η τεχνολογία. Ακόμα κι αν οι πάροχοι μπορούσαν να καταλάβουν τις προκλήσεις της αποστολής ευαίσθητων πληροφοριών σε όλη την κατάσταση, δεν μπορούν να είναι σίγουροι ότι τα δεδομένα θα αποδοθούν στον σωστό ασθενή από την άλλη άκρη. Διαφορετικά συστήματα ΗΦΥ μπορούν να αποθηκεύουν διαφορετικούς τύπους πληροφοριών σχετικά με ασθενείς, γεγονός που καθιστά δύσκολο για τους παρόχους να ταξινομήσουν με ακρίβεια τα ΗΜΥ. Τα συστήματα υγείας χρησιμοποιούν διάφορους παράγοντες για να ταιριάξουν τους ασθενείς. Ορισμένα αρχεία υγείας βασίζονται σε διευθύνσεις και ονόματα, ενώ άλλα χρησιμοποιούν έναν εσωτερικό αριθμό περίπτωσης που λειτουργεί μόνο στο σύστημα του ίδιου του ατόμου. Μια μελέτη διαπίστωσε ότι ήταν



δύσκολο να επιτευχθεί διαλειτουργικότητα με άλλο οργανισμό, επειδή κάθε οργανισμός χρησιμοποιεί διαφορετικό τρόπο αντιστοίχισης ασθενών. Ένα εθνικό αναγνωριστικό ασθενούς ήταν δύσκολο να δημιουργηθεί και τόσες πολλές διαφορετικές ομάδες πρέπει να συνεργαστούν για να βρουν μια λύση.

- Οικονομικές ανησυχίες και κόστος επενδύσεων:** Η εύρεση χρημάτων για την υποστήριξη έργων πληροφορικής υγείας μπορεί να είναι δύσκολη. Ορισμένοι οργανισμοί υγειονομικής περίθαλψης έχουν ήδη επενδύσει πολλά χρήματα σε συστήματα ΗΜΥ που δεν πληρούν τις απαιτήσεις διαλειτουργικότητας δεδομένων υγείας. Είναι απρόθυμοι να ξοδέψουν περισσότερα χρήματα σε νέα συστήματα που μπορεί να μην είναι σε θέση να χειριστούν τα δεδομένα από όλους τους ασθενείς τους. Τα ενδιαφερόμενα μέρη και οι εκπρόσωποι λένε ότι το κόστος για την επίτευξη διαλειτουργικότητας μπορεί να είναι πολύ ακριβό, εν μέρει λόγω των δομών τιμολόγησης που έχουν δημιουργήσει οι προμηθευτές ΗΦΥ. Πολλά συστήματα ΗΦΥ απαιτούν διαφορετικούς τύπους συνδέσεων προκειμένου να συνεργαστούν με άλλους παρόχους υγειονομικής περίθαλψης και οργανισμούς. Η προσαρμογή του αρχείου υγείας σας μπορεί να είναι ακριβή, αλλά είναι συνήθως ευθύνη του αγοραστή ΗΜΥ. Οι πάροχοι συστημάτων υγείας μπορεί να χρειαστεί να πληρώσουν δικηγόρους για να διαπραγματευτούν συμφωνίες κοινής χρήσης δεδομένων μεταξύ τους. Οι οργανισμοί διαλειτουργικότητας εργάζονται για να διευκολύνουν τους ανθρώπους να χρησιμοποιούν τα ίδια ΗΜΥ καθιστώντας τις βασικές τεχνολογίες που συνδέουν τα συστήματα πληροφορικής υγείας πιο τυποποιημένες. Οι κρατικοί ή περιφερειακοί οργανισμοί ανταλλαγής πληροφοριών για την υγεία (ή SHINE [126]) είναι σαν σύνολο υπολογιστών. Βοηθούν τα μέλη τους να μοιράζονται πληροφορίες για την υγεία τους κάνοντας κάποιες από τις τεχνικές εργασίες για αυτά, με αντάλλαγμα μια αμοιβή κάθε χρόνο. Αυτό βοηθάει τα μέλη της ομάδας να μην χρειάζεται να ανακαλύπτουν ξανά τον τροχό κάθε φορά που ξεκινούν ένα νέο έργο. Ορισμένες ομάδες πιστεύουν ότι η χρήση αεροπορικών ταξιδιών για την προσαρμογή σε ένα νέο μέρος μπορεί να είναι ένας φθηνός και εύκολος τρόπος για να ξεκινήσετε. Τα API (Application Programming Interface) επιτρέπουν στους παρόχους να μοιράζονται δεδομένα υγείας εύκολα και με άλλους παρόχους που χρησιμοποιούν το ίδιο λογισμικό. Αυτός είναι ένας καλός τρόπος για τη βελτίωση της διαλειτουργικότητας των δεδομένων υγείας.
- Ανάγκη διακυβέρνησης πληροφόρησης, εμπιστοσύνης και συνεργασίας:** Αυτός είναι ο λόγος για τον οποίο πρέπει να συνεργαστούμε για να οικοδομήσουμε εμπιστοσύνη, ώστε να μπορούμε να επωφεληθούμε από όλες αυτές τις νέες τεχνολογικές εξελίξεις. Πρέπει οι βιομηχανικοί όμιλοι να καθησυχάσουν τους οργανισμούς υγειονομικής περίθαλψης ότι αυτές οι επενδύσεις αξίζουν τον κόπο, μέσω της ίδρυσής τους; Έχουμε έναν τρόπο να μοιραζόμαστε πληροφορίες για την υγεία που είναι σαφείς, ολοκληρωμένες και ωφέλιμες τόσο για εμάς όσο και για το άτομο με το οποίο τις μοιραζόμαστε. Αυτές οι πρακτικές διακυβέρνησης μπορούν να περιλαμβάνουν πράγματα όπως οργανωτικές πολιτικές που αφορούν το απόρρητο, την ασφάλεια των πληροφοριών, τη χρήση δεδομένων και άλλα ζητήματα που επηρεάζουν τον τρόπο με τον οποίο μοιράζονται οι πληροφορίες πέρα από τα όρια του οργανισμού. Προκειμένου να οικοδομηθεί εμπιστοσύνη στην κοινότητα της περίθαλψης, οι πάροχοι πρέπει να υιοθετήσουν ένα όραμα βασισμένο στην αξία, με επίκεντρο τον ασθενή για το ρόλο τους στο οικοσύστημα υγειονομικής περίθαλψης. Αυτό σημαίνει να βλέπεις τους ασθενείς ως άτομα και να

εστιάζεις στις ανάγκες τους, όχι στο τι μπορεί να πάρει ο πάροχος από αυτούς. Τα μεγάλα δεδομένα μπορούν να είναι χρήσιμα για μεμονωμένους οργανισμούς, αλλά είναι πραγματικά ισχυρά όταν μοιράζονται ελεύθερα μεταξύ ασθενών και γιατρών.

Οι άνθρωποι χρησιμοποιούν δεδομένα υγείας πιο συχνά για να βεβαιωθούν ότι όλοι λαμβάνουν τη φροντίδα που χρειάζονται. Οι πάροχοι αρχίζουν να συνεργάζονται για να μοιράζονται αυτά τα δεδομένα, ώστε όλοι να έχουν την καλύτερη δυνατή φροντίδα. Η αλλαγή από τη μια έκδοση μιας τεχνολογίας στην άλλη γίνεται αργά και μπορεί να είναι δύσκολη. Ωστόσο, τα στοιχεία για τη διαλειτουργικότητα συνεχίζουν να αυξάνονται και οι πάροχοι αλλάζουν την οπτική τους ανάλογα. Ο λόγος για τον οποίο η διαλειτουργικότητα των δεδομένων υγείας είναι τόσο σημαντική είναι ότι όλο και περισσότερες ιδιωτικές εταιρείες και ομοσπονδιακοί φορείς συνεργάζονται για να διασφαλίσουν ότι όλοι έχουν τις ίδιες πληροφορίες, ώστε να μπορούμε να συνεργαστούμε καλύτερα στο μέλλον. Το κύριο πρόβλημα δεν είναι τα ίδια τα συστήματα, αλλά η δυσκολία ανταλλαγής δεδομένων μεταξύ τους ή η ασυμβατότητα των δεδομένων.

### 3.3.Στάδια Προσέγγισης Υπηρεσιών Διαλειτουργικότητας Κατηγορίες Διαλειτουργικότητας στον τομέα της Υγείας

Για την ανταλλαγή πληροφοριών με άλλα συστήματα, πρέπει να γίνονται κατανοητά τα μηνύματα που στέλνονται μεταξύ ατόμων. Πρέπει να επιτευχθεί «Συντακτική Διαλειτουργικότητα» (την ικανότητα δύο ή περισσότερων συστημάτων να ανταλλάσσουν πληροφορίες μέσω καθορισμένων δομών μηνυμάτων) και «Σημασιολογική Διαλειτουργικότητα» (η ικανότητα των πληροφοριών που μοιράζονται τα συστήματα να γίνονται κατανοητές στο επίπεδο των εννοιών του ορισμού του αντικειμενικού τομέα). Η σημασιολογική διαλειτουργικότητα είναι σημαντική για την αυτόματη επεξεργασία πληροφοριών, η οποία θα επιτρέψει προηγμένες κλινικές εφαρμογές. Η υγειονομική περίθαλψη αφορά τη διαχείριση πληροφοριών και γνώσης. Ο στόχος είναι η ανταλλαγή πληροφοριών και η χρήση τους σε γιατρούς και ασθενείς. Σαν αποτέλεσμα αυτών, η διαλειτουργικότητα στον τομέα της υγείας μπορεί να επικεντρωθεί στους δύο ακόλουθους κύριους άξονες [127]:

#### **Άξονας 1: Διαλειτουργικότητα ανταλλαγής μηνυμάτων υγείας**

Οι διεπαφές βοηθούν στη σύνδεση διαφορετικών τύπων συστημάτων, έτσι ώστε οι πληροφορίες να ανταλλάσσονται γρήγορα και εύκολα. Οι διεπαφές επιτρέπουν στην κύρια εφαρμογή να συλλέγει δεδομένα και να τα στέλνει μέσω του δικτύου σε άλλη εφαρμογή. Όταν στέλνετε ένα μήνυμα, ο υπολογιστής του αποστολέα λαμβάνει το μήνυμα και το μεταφράζει σε μια μορφή που μπορεί να κατανοήσει ο παραλήπτης. Στη συνέχεια, ο υπολογιστής στέλνει το μήνυμα στην κύρια εφαρμογή του δέκτη για να το δουν και να το χρησιμοποιήσουν. Σε ένα σχέδιο όπως αυτό, ο αριθμός των διεπαφών που χρησιμοποιούνται σχετίζεται με τον αριθμό των εφαρμογών που χρησιμοποιούν τη διεπαφή για να επικοινωνούν μεταξύ τους. Προτείνεται να κωδικοποιήσετε τα μηνυμάτα σας με τρόπο που θα είναι κατανοητό από κάθε εφαρμογή που γνωρίζει αυτήν την κωδικοποίηση. Αυτό θα βοηθήσει στην επίλυση του προβλήματος. Ένα από τα πιο ευρέως χρησιμοποιούμενα πρότυπα κωδικοποίησης μηνυμάτων στην υγειονομική περίθαλψη είναι το HL7. Ωστόσο, εάν ένα σύστημα είναι συμβατό με την τρέχουσα έκδοση του HL7, δεν σημαίνει

απαραίτητα ότι μπορεί εύκολα να επικοινωνήσει με συστήματα που χρησιμοποιούν παλαιότερες εκδόσεις του HL7, καθώς υπάρχουν σημαντικά κενά στην αντιστοιχία των ορισμών της έκδοσης.

## **Άξονας 2: Διαλειτουργικότητα ΗΦΥ**

Το ΗΦΥ είναι ένα ψηφιακό αρχείο των πληροφοριών υγείας ενός ατόμου που χρησιμοποιείται από την ομάδα υγειονομικής περίθαλψης για να παρακολουθεί την υγεία και τη φροντίδα του ατόμου. Μπορεί επίσης να χρησιμοποιηθεί για εκπαίδευση, έρευνα και προστασία της ιδιωτικής ζωής του ασθενούς. Το γεγονός ότι οι πληροφορίες ΗΜΥ αποθηκεύονται σε διαφορετικά σημεία δεν σημαίνει ότι μπορούν εύκολα να μοιραστούν μεταξύ τους. Είναι σημαντικό οι υπηρεσίες υγείας να είναι συνεχείς, ώστε οι άνθρωποι που τις παρέχουν (γιατροί, νοσηλευτές κ.λπ.) να μπορούν πάντα να βοηθούν το άτομο που φροντίζουν. Ο γιατρός θα έχει πρόσβαση σε όλα τα ιατρικά δεδομένα του ασθενούς και το ιστορικό των κινήσεών του. Ο στόχος των προσπαθειών τυποποίησης είναι να διασφαλιστεί ότι τα ΗΜΥ από διαφορετικούς προμηθευτές είναι σε θέση να επικοινωνούν μεταξύ τους. Αυτό γίνεται μέσω προσπαθειών όπως το πρότυπο CEN/TC 251 EHRcom, το arepEHR και το πρότυπο HL7 CDA. Ακόμη και σήμερα, είναι δύσκολο να μοιραστούμε πληροφορίες για την υγεία μεταξύ διαφορετικών συστημάτων υγείας. Υπάρχουν πολλοί διαφορετικοί τρόποι για να κάνετε πράγματα, επομένως η χρήση ενός τρόπου δεν θα λειτουργήσει καλά για όλους.

Τα στάδια της διαλειτουργικότητας συνοψίζονται παρακάτω ως:

- **Νομική Διαλειτουργικότητα**

Η νομική διαλειτουργικότητα σημαίνει ότι διαφορετικά συστήματα πληροφοριών μπορούν να συνεργαστούν. Αυτό επιτρέπει στους ανθρώπους να μοιράζονται πληροφορίες πιο εύκολα, κάτι που μπορεί να τους βοηθήσει να λύσουν τα προβλήματα πιο γρήγορα. Αυτό το εργαλείο βοηθά τα συστήματα να συνδέονται μεταξύ τους, ώστε να μπορούν να μοιράζονται εύκολα δεδομένα. Καθορίζει τις συγκεκριμένες ανάγκες που απαιτούνται για να επικοινωνεί ένα σύστημα με ένα άλλο. Το σύστημα δεν χρειάζεται να καταλάβει τι σημαίνουν τα δεδομένα, καθώς είναι ήδη εύκολα κατανοητό.

- **Οργανωτική Διαλειτουργικότητα**

Οργανωτική διαλειτουργικότητα σημαίνει να έχεις ένα σύστημα που λειτουργεί καλά μαζί, ανεξάρτητα από το ποιος το κατασκευάζει, τι είναι ή πού βρίσκεται. Περιλαμβάνει πράγματα όπως οι τεχνικές πτυχές του συστήματος, καθώς και οι πολιτικές, κοινωνικές και οργανωτικές πτυχές του. Αυτά τα εργαλεία βοηθούν τους οργανισμούς να επικοινωνούν και να μοιράζονται εύκολα δεδομένα. Αυτό σημαίνει ότι η δουλειά σας μπορεί να γίνει με άλλα άτομα, ακόμα κι αν δεν βρίσκονται στο ίδιο γραφείο σας, μοιράζοντας πληροφορίες μαζί τους. Με αυτόν τον τρόπο, όλοι μπορούν να συνεργαστούν για να γίνουν τα πράγματα όσο το δυνατόν πιο γρήγορα και αποτελεσματικά.

- **Σημσιολογική Διαλειτουργικότητα**

Η σημσιολογική διαλειτουργικότητα είναι ο καλύτερος τρόπος για να παραμείνετε συνδεδεμένοι με άλλους. Οι πληροφορίες μπορούν εύκολα να ρέουν μεταξύ διαφορετικών συστημάτων ή τμημάτων συστημάτων. Αυτό σημαίνει ότι οι πάροχοι μπορούν να μοιράζονται δεδομένα ασθενών

χρησιμοποιώντας διαφορετικές λύσεις λογισμικού, χωρίς να χρειάζεται να ανησυχούν για τις συγκεκριμένες μορφές δεδομένων. Η ανταλλαγή δεδομένων χρησιμοποιείται για να περιλαμβάνει τυποποιημένες πληροφορίες με τρόπο που μπορούν να κατανοήσουν άλλα συστήματα. Αυτό επιτρέπει την πιο εύκολη επεξεργασία των δεδομένων. Αυτό σημαίνει ότι οι πάροχοι υγειονομικής περίθαλψης και οι ερευνητές μπορούν να συνεργαστούν για να μοιράζονται δεδομένα πιο εύκολα, κάτι που μπορεί να μας βοηθήσει να μάθουμε περισσότερα για ασθένειες και προβλήματα δημόσιας υγείας.

- **Τεχνική (Συντακτική) Διαλειτουργικότητα**

Η τεχνική (ή αλλιώς συντακτική) διαλειτουργικότητα είναι ο τρόπος με τον οποίο ανταλλάσσονται δεδομένα. Ο τρόπος με τον οποίο μορφοποιούνται τα μηνύματα μπορεί να βοηθήσει να διασφαλιστεί ότι οι πληροφορίες που αποστέλλονται προορίζονται για συγκεκριμένο σκοπό και δεν έχουν αλλάξει ή αλλοιωθεί. Σε αυτή την περίπτωση, μιλάμε για τους διαφορετικούς τύπους πεδίων δεδομένων σε μια βάση δεδομένων αρχείων ασθενών.

### 3.4.Πρότυπα Διαλειτουργικότητας στην Υγεία

#### Πρότυπα Διαλειτουργικότητας στον τομέα της Υγείας

Στις κοινωνίες του 21<sup>ου</sup> αιώνα βασιζόμαστε σε ορισμένα πράγματα για να είμαστε συνεπείς. Η συμμόρφωση με τα πρότυπα βοηθά στη διασφάλιση ότι τα προϊόντα είναι εναλλάξιμα και ότι τα προβλήματα μπορούν να επιλυθούν. Αυτό συμβάλλει στην ενθάρρυνση του εμπορίου και στη διασυννοριακή συμβατότητα των προϊόντων. Με την τυποποίηση των δεδομένων υγείας, γίνεται ευκολότερη η σύγκριση, η κατηγοριοποίηση και, τελικά, η καλύτερη υγειονομική περίθαλψη. Η συμμόρφωση με τα πρότυπα σημαίνει ότι μπορούμε να είμαστε σίγουροι ότι τα συστήματα που χρησιμοποιούμε είναι ασφαλή και λειτουργούν με τον ίδιο τρόπο κάθε φορά. Αυτό είναι σημαντικό γιατί η επιστήμη δεδομένων είναι ένας αναπτυσσόμενος τομέας που συχνά γίνεται με πολύπλοκο και απρόβλεπτο τρόπο. Η ύπαρξη προτύπων μας διευκολύνει να εργαζόμαστε με δεδομένα και να τα κατανοούμε πιο εύκολα.

Τα κλινικά δεδομένα βοηθούν τους γιατρούς να καταλάβουν τι συμβαίνει με έναν ασθενή και τι να κάνουν για αυτό. Τα ιατρικά αρχεία τηρούνται σε διαφορετικά συστήματα και μερικές φορές μπορεί να είναι ελλιπή ή ανακριβή.

Τα κλινικά δεδομένα οργανώνονται με τρόπο που τα καθιστά χρήσιμα για συγκεκριμένους σκοπούς, όπως η αναφορά, η επικοινωνία και η χρέωση. Όταν πρόκειται για την υγειονομική περίθαλψη, η επικοινωνία είναι το κλειδί - γι' αυτό οι κακώς δομημένες ιστορίες συχνά κερδίζουν περισσότερα δομημένα δεδομένα. Το κείμενο μιλάει για το πώς μπορούν να ομαδοποιηθούν διαφορετικοί τύποι πραγμάτων. Αυτό συμβαίνει γιατί όλα έχουν κάποιο κοινό χαρακτηριστικό. Η γραμματική και η ορθογραφία της δομής σας είναι αποδεκτά, παρόλο που μπορεί να υπάρχουν κάποιες μικρές διαφορές. Το ελεύθερο κείμενο δεν λειτουργεί καλά εάν τα άτομα που το χρησιμοποιούν τις λέξεις με διαφορετικούς τρόπους, πράγμα που σημαίνει ότι οι έννοιές τους δεν θα ταιριάζουν. Αφού επεξεργαστούμε τα δεδομένα, ενδέχεται να βρούμε πράγματα που δεν είχαμε δει πριν. Σε περιπτώσεις όπου τα κλινικά δεδομένα χρησιμοποιούνται για υποστήριξη αποφάσεων από υπολογιστή ή αναδρομική ανάλυση, μπορεί να είναι δύσκολο να βρεθούν οι

σωστές περιγραφές για τα δεδομένα. Αυτό μπορεί να οφείλεται στο ότι τα τοπικά λεξικά δεδομένων ενδέχεται να μην περιλαμβάνουν πληροφορίες σχετικά με τον συγκεκριμένο τύπο δεδομένων που χρησιμοποιούνται. Η προγνωστική ανάλυση μπορεί επίσης να παρεμποδιστεί από την έλλειψη σχετικών δεδομένων. Ανεξάρτητα από το είδος της χρήσης που έχετε για κλινικά δεδομένα (πρωτογενή ή δευτερογενή), πρέπει να ακολουθείτε κοινά πρότυπα. Τα πρότυπα αυτά πρέπει να περιγράφουν:

- Προέλευση δεδομένων (Δημιουργοί, χρόνοι δημιουργίας, λοιπές διαδικασίες).
- Πρότυπα πληροφοριών.
- Ορολογία οντολογιών.
- Σημασιολογική περιγραφή.
- Ορισμοί παραστατικών μονάδων.
- Επίσημη γλώσσα των παραπάνω.

Υπάρχει μεγάλο πρόβλημα με τον τομέα της υγείας γιατί δεν υπάρχει σύστημα που να μπορεί να κωδικοποιεί και να ταξινομεί μοναδικά διαφορετικές πτυχές της υγείας. Υπάρχουν πολλές πληροφορίες στον τομέα της υγείας που παράγονται. Η πηγή των πληροφοριών είναι ο διαχειριστής, το χρηματοπιστωτικό ίδρυμα, ο ιατρικός πάροχος ή άλλος οργανισμός που είναι υπεύθυνος για την παροχή της υπηρεσίας. Οι άνθρωποι έχουν διαφορετικές μορφές (π.χ. ήχος, βίντεο, εικόνες, κείμενα, βιοσήματα). Υπάρχουν πολλές διαφωνίες σχετικά με το πώς να κάνουμε πράγματα στον ιατρικό τομέα, γεγονός που καθιστά δύσκολη τη δημιουργία αυτοματοποιημένων συστημάτων και διαδικασιών. Η χρήση υπολογιστών για τη διαχείριση δεδομένων υγείας καθιστά ευκολότερο να δούμε πώς παρόμοια δεδομένα μπορούν να υποβληθούν σε επεξεργασία με διαφορετικούς τρόπους, καθώς απαιτείται από τους υπολογιστές να χρησιμοποιούν τυποποιημένες μεθόδους κωδικοποίησης και οργάνωσης δεδομένων.

Η κωδικοποίηση είναι ένας τρόπος για να διατηρούνται τα ιατρικά αρχεία οργανωμένα και ακριβή. Αυτό διευκολύνει τους γιατρούς να κάνουν διαγνώσεις και να παρέχουν καλύτερες υπηρεσίες υγειονομικής περίθαλψης. Η κωδικοποίηση είναι ένας τρόπος σύγκρισης και οργάνωσης δεδομένων για τη λήψη αποφάσεων σχετικά με τον τρόπο λειτουργίας του οργανισμού. Αυτό μας βοηθά να καταλάβουμε ποιες στρατηγικές θα χρησιμοποιήσουμε για να βελτιώσουμε τη δουλειά μας στο μέλλον. Το σύστημα κωδικοποίησης που χρησιμοποιείται για την οργάνωση δεδομένων υγείας είναι σαν ένα σύνολο αριθμών που μπορούν να συντομευθούν, να αποθηκευτούν και να προσπελαστούν εύκολα. Αυτός είναι ένας τρόπος αναγνώρισης ασθενειών, τραυματισμών ή διαδικασιών από τους ιατρικούς κώδικες. Ένα σύστημα κωδικοποίησης θα πρέπει να έχει έναν ενιαίο κωδικό για την καταγραφή όλων των ορολογιών που παρατίθενται, π.χ. Η κωδικοποίηση μπορεί να μας βοηθήσει να κατανοήσουμε και να αναπαραστήσουμε τις πληροφορίες με τρόπο που να έχει νόημα και χρήσιμο. Αυτό είναι σημαντικό για την επίτευξη σημασιολογικής διαλειτουργικότητας. Με την τυποποίηση των δεδομένων, διευκολύνουμε την επεξεργασία και την αναζήτηση. Αυτό βοηθά στη μείωση των σφαλμάτων και στη βελτίωση της ακρίβειας των δεδομένων. Επιπλέον, διευκολύνει τη σύγκριση δεδομένων από διαφορετικές πηγές.

Επίσης, η κωδικοποίηση είναι η διαδικασία μετάφρασης ενός συνόλου γραπτών οδηγιών σε μια μορφή που μπορεί να κατανοήσει ένας υπολογιστής. Η κωδικοποίηση είναι ένας τρόπος ομαδοποίησης πραγμάτων έτσι ώστε να είναι πιο κατανοητό και να θυμόμαστε. Υπάρχουν δύο

κύριοι τύποι κωδικοποίησης: η ταξινόμηση, η οποία εκχωρεί κωδικούς σε πράγματα για να τα κάνει πιο εύκολα αναγνωρίσιμα, και η ονοματολογία, η οποία ορίζει ονόματα για πράγματα ώστε να είναι ευκολότερα στη μνήμη. Το λεξικό είναι μια συλλογή λέξεων που σχετίζονται μεταξύ τους. Άλλοι τύποι κωδικοποίησης είναι η χρήση λέξεων για την αναπαράσταση πραγμάτων (θεωρία λέξεων), η ομαδοποίηση πραγμάτων (ταξονομία), και χρησιμοποιώντας συγκεκριμένους, επίσημους όρους (επίσημη ορολογία).

Η διεθνής συνεργασία είναι σημαντική στην προσπάθεια δημιουργίας ενιαίας ταξινόμησης και προτύπων για την υγεία. Αυτό συμβαίνει επειδή διευκολύνει τους ανθρώπους να κατανοήσουν τι είναι πληροφορίες που σχετίζονται με την υγεία και διευκολύνει τους γιατρούς να διαγνώσουν και να θεραπεύσουν ασθένειες.

Η Διεθνής Ταξινόμηση Νοσημάτων (ICD) είναι ένα σύστημα κατηγοριοποίησης ασθενειών. Το DICOM είναι ένα σύστημα αποθήκευσης ιατρικών εικόνων. Το HL7 είναι ένα σύστημα ανταλλαγής ιατρικών πληροφοριών [128]. Το SNOMED είναι ένα σύστημα ταξινόμησης ιατρικών όρων. Ο Παγκόσμιος Οργανισμός Υγείας έχει δημιουργήσει ένα σύστημα ταξινόμησης που ονομάζεται ICD-10. Ο τρόπος με τον οποίο πρέπει να χρησιμοποιείται η κωδικοποίηση δεν είναι μόνο για λόγους περιγραφής δεδομένων, αλλά και για πρακτικούς σκοπούς. Ο καλύτερος τρόπος για να διασφαλιστεί η ποιότητα και η αξιοπιστία των πληροφοριών σχετικά με την υγεία είναι η χρήση μιας πυκνής ορολογίας τόσο από την άποψη της ονομασίας των πραγμάτων (ονοματολογία) όσο και από την άποψη της κωδικοποίησης πληροφοριών (κωδικοποίηση). Η αξιοπιστία ενός κομματιού δεδομένων καθορίζεται από τη σταθερότητα, την αναπαραγωγικότητα και την ακρίβειά του. Η ποιότητα, από την άλλη πλευρά, καθορίζεται από τη σωστή απόδοση των δεδομένων, έτσι ώστε να μπορούν να ανακτηθούν με συνεπή τρόπο. Όλοι οι υπολογιστές του γραφείου συνεργάζονται για τη διαχείριση και την αποθήκευση δεδομένων, αλλά στέλνουν επίσης μηνύματα ο ένας στον άλλο για να συνεχίσουν να λειτουργούν ομαλά. Υπάρχουν διάφοροι τρόποι επεξεργασίας, αποθήκευσης και μεταφοράς δεδομένων σε διαφορετικά συστήματα. Για αυτό το λόγο, θα πρέπει να στηρίζονται σε ένα πρότυπο.

Στο χώρο της υγειονομικής περίθαλψης διακρίνεται μια σειρά βασικών κατηγοριών σε πρότυπα:

- Πρότυπα επικοινωνίας
- Πρότυπα αναπαράστασης κλινικών δεδομένων
- Πρότυπα αναγνώρισης
- Πρότυπα αρχιτεκτονικής ΗΦΥ
- Πρότυπα ασφάλειας των δεδομένων και εξασφάλισης του ιατρικού απορρήτου

Τα δεδομένα υγείας αποθηκεύονται πλέον σε υπολογιστές, επομένως είναι σημαντικό το σύστημα που κωδικοποιεί και αποθηκεύει αυτά τα δεδομένα να είναι συνεπές. Χωρίς μια προκαθορισμένη ορολογία είναι αδύνατο να γίνει μια αυτοματοποιημένη σύνθεση δεδομένων υγείας. Έστω ότι, ο πρώτος γιατρός κατέγραψε ότι ο ασθενής είχε δύσπνοια και ο δεύτερος γιατρός ότι ο ασθενής είχε δυσκολία στην αναπνοή. Ένα πρόγραμμα υπολογιστή μπορεί να μην είναι σε θέση να πει με βεβαιότητα ότι ο ασθενής έχει το ίδιο πρόβλημα και στις δύο περιπτώσεις. Υπάρχει μεγάλη τάση στην ανάπτυξη προτύπων στον τομέα της υγείας. Επίσης, υπάρχουν πολλοί διαφορετικοί τρόποι να κάνεις πράγματα, γιατί διαφορετικοί άνθρωποι πιστεύουν ότι αυτός είναι ο καλύτερος τρόπος για να κάνεις πράγματα. Τα άτομα που δημιουργούν αυτά τα πρότυπα προσπαθούν να

χρησιμοποιήσουν τις καλύτερες δυνατότητες από άλλα πρότυπα. Παρά τις καλύτερες προσπάθειες όλων, δεν υπάρχει ακόμη ένα συμφωνημένο πρότυπο που να διασφαλίζει τα οφέλη που θέλουν οι άνθρωποι. Το Ινστιτούτο Προτύπων της Ευρωπαϊκής Ένωσης, που ονομάζεται TC251 [128], είναι υπεύθυνο για τη διασφάλιση της τυποποίησης των μηνυμάτων τεχνολογίας πληροφοριών υγείας (IT) σε όλη την Ευρώπη.

Επί του παρόντος, υπάρχει χαμηλό ποσοστό υιοθέτησης προτύπων κλινικών δεδομένων. Η καθυστέρηση στη συλλογή κλινικών δεδομένων θα περιορίσει τη χρησιμότητά του για άλλους σκοπούς. Πρέπει να διορθωθεί η καθυστέρηση, ώστε τα κλινικά δεδομένα να χρησιμοποιηθούν πιο αποτελεσματικά. Η τυποποίηση είναι μια περίπλοκη διαδικασία που απαιτεί συνεργασία από πολλούς διαφορετικούς ανθρώπους. Απαιτεί από τους δημιουργούς και τους συντηρητές προτύπων να καταβάλουν μεγάλη προσπάθεια, καθώς και από τους χρήστες που θα χρησιμοποιήσουν το πρότυπο.

### 3.4.1. SNOMED (Systematized Nomenclature of Medicine)

Το SNOMED CT είναι μια πολύ μεγάλη και ολοκληρωμένη ορολογία υγειονομικής περίθαλψης που χρησιμοποιείται σε όλο τον κόσμο [129]. Η τυποποίηση του τρόπου με τον οποίο μιλάμε για πληροφορίες υγείας μπορεί να βοηθήσει να διασφαλίσουμε ότι η ποιότητα της περίθαλψης που λαμβάνουμε είναι συνεπής σε όλο τον κόσμο. Το SNOMED CT είναι ένα σύστημα ιατρικής ορολογίας που βοηθά τους γιατρούς και τους νοσηλευτές να τεκμηριώνουν και να μοιράζονται πληροφορίες σχετικά με τους ασθενείς. Τα κλινικά δεδομένα μπορούν να κωδικοποιηθούν, να ανακτηθούν και να αναλυθούν χρησιμοποιώντας ένα σύστημα κωδικοποίησης.

Επίσης, το SNOMED CT είναι μια βάση δεδομένων ιατρικών όρων που δημιουργήθηκε ως αποτέλεσμα της συγχώνευσης δύο άλλων βάσεων δεδομένων: SNOMED Reference Terminology (SNOMED RT) που αναπτύχθηκε από το College of American Pathologists (CAP) και οι Clinical Terms Version 3 (CTV3) που αναπτύχθηκε από την Εθνική Υπηρεσία Υγείας (NHS) του Ηνωμένου Βασιλείου [130]. Η ορολογία είναι ένας τρόπος για την ακριβή περιγραφή των κλινικών πληροφοριών στον τομέα της υγείας. Χρησιμοποιείται πολύ σε ηλεκτρονικά αρχεία υγείας. Οι εφαρμογές υγειονομικής περίθαλψης χρησιμοποιούν τη συλλογή κλινικών δεδομένων για να βοηθήσουν τους γιατρούς να κατανοήσουν καλύτερα την υγεία των ασθενών και να βρουν τις καλύτερες θεραπείες. Υπάρχουν πολλές πληροφορίες για την υγεία εκεί έξω, αλλά μπορεί να είναι δύσκολο να γνωρίζουμε πού να τις βρούμε. Μερικές φορές διαφορετικές πηγές δίνουν διαφορετικές πληροφορίες και μπορεί να είναι δύσκολο να καταλάβουμε τι είναι αλήθεια και τι όχι.

Το SNOMED CT είναι ένα σύστημα που βοηθά στην τυποποίηση ιατρικών πληροφοριών. Οι εφαρμογές χρησιμοποιούν έννοιες όπως ιεραρχία, συσχέτιση και ανάλυση για την κατανόηση δεδομένων. Κάνοντας έρευνα, οι επιστήμονες μπορούν να μάθουν πολλά για το τι λειτουργεί καλύτερα και τι είναι προσιτό. Αυτές οι πληροφορίες μπορούν να τους βοηθήσουν να παρέχουν υπηρεσίες υψηλής ποιότητας στους ανθρώπους. Η ονομασία των πραγμάτων με τον ίδιο τρόπο διευκολύνει τους επαγγελματίες υγείας να κατανοήσουν και να συνεργαστούν. Αυτό τους διευκολύνει να φροντίσουν για εσάς και να διασφαλίσουν ότι θα λάβετε την καλύτερη δυνατή θεραπεία. Οι άνθρωποι που αναπτύσσουν κλινικές εφαρμογές συνεργάζονται για να βοηθήσουν ασθενείς και διαχειριστές.

Συνεπώς, το SNOMED CT είναι ένας τρόπος παροχής πλήρεις και ακριβείς πληροφορίες σχετικά με ιατρικές καταστάσεις σε επαγγελματίες υγείας. Αυτό βοηθά να διασφαλιστεί ότι οι ασθενείς λαμβάνουν την καλύτερη δυνατή φροντίδα. Το ιατρικό ιστορικό σας περιλαμβάνει πληροφορίες για ασθένειες που είχατε στο παρελθόν. Αυτές οι πληροφορίες συχνά περιλαμβάνονται στα εργαστηριακά αποτελέσματα και στις θεραπείες που μπορεί να λάβετε. Φροντίζοντας την υγεία σας, μπορείτε να βελτιώσετε τις πιθανότητές σας να είστε υγιείς και να απολαύσετε τη ζωή. Όταν έχετε κλινικές πληροφορίες στο αρχείο σας, οι πάροχοι υγειονομικής περίθαλψης μπορούν εύκολα να εντοπίσουν εσάς και τα προβλήματά σας. Αυτό τους διευκολύνει να αποφασίσουν για την καλύτερη θεραπεία για εσάς. Έτσι, το SNOMED CT σχεδιάστηκε έτσι ώστε να είναι εύκολη η ηλεκτρονική κοινή χρήση αρχείων υγείας, χωρίς μεγάλες αποκλίσεις. Ο υπολογιστής μπορεί να κατανοήσει την ιατρική ορολογία επειδή έχει ένα λεξικό που τον βοηθά να δημιουργεί και να κατανοεί απλές φράσεις. Ωστόσο, επειδή έχει τόσους πολλούς όρους, δεν είναι εύκολο για τους ανθρώπους να το χρησιμοποιήσουν απευθείας. Ωστόσο, είναι πολύτιμο για τα αυτοματοποιημένα συστήματα κωδικοποίησης.

### 3.4.2. ICD-10 (International Classification of Diseases)

Το ICD είναι ένα σύστημα που βοηθά τους γιατρούς να κατανοήσουν πώς προκαλούνται οι ασθένειες και πώς είναι τα συμπτώματα. Το ICD-10 είναι η πιο πρόσφατη έκδοση του ICD, η οποία εστιάζει σε διαφορετικά συμπτώματα από τα παλαιότερα ICD [131].

Το σύστημα εντοπίζει ανώμαλα πράγματα ή πράγματα που ξεχωρίζουν από τον μέσο όρο και τα χαρακτηρίζει ως τέτοια. Το ICD-10 είναι ένα σύστημα καταγραφής και παρακολούθησης ασθενειών και προβλημάτων υγείας με τρόπο που είναι πιο ακριβής και αποτελεσματικός από τις προηγούμενες εκδόσεις του κώδικα. Αυτό το σύστημα χρησιμοποιείται για την καλύτερη κατανόηση της γενικής υγείας ενός πληθυσμού και για την πρόληψη της εξάπλωσης ασθενειών και προβλημάτων υγείας. Το σύστημα ταξινόμησης ασθενειών χρησιμοποιείται για την ταξινόμηση ασθενειών και άλλων προβλημάτων υγείας που καταγράφονται σε πολλούς τύπους ζωτικών αρχείων, συμπεριλαμβανομένων των πιστοποιητικών θανάτου και των αρχείων υγείας. Τα αρχεία υγείας είναι σημαντικά για να βοηθήσουν τους γιατρούς να διαγνώσουν ασθενείς και να παρακολουθούν την υγεία του πληθυσμού. Βοηθούν επίσης στην κατάρτιση εθνικών στατιστικών για τη θνησιμότητα και τη νοσηρότητα. Ο κατάλογος των ασθενειών περιλαμβάνει μολυσματικές, γενικές ασθένειες και ασθένειες που εξαρτώνται από την τοποθεσία στις ανατολικές χώρες. Περιλαμβάνονται επίσης διαταραχές ανάπτυξης και τραυματισμοί.

Το σύστημα ICD-10 είναι μια αναθεώρηση της Διεθνούς Ταξινόμησης Νοσημάτων, που δημοσιεύθηκε το 1992 [132]. Χρησιμοποιείται για την παρακολούθηση διαφορετικών τύπων ασθενειών. Η ταξινόμηση ICD είναι ένας τρόπος ομαδοποίησης ασθενειών χρησιμοποιώντας έναν απλό κώδικα. Αυτός ο κώδικας μπορεί περαιτέρω να χωριστεί σε δέκα υποκατηγορίες. Η δέκατη αναθεώρηση του συστήματος ICD (ICD-10) χρησιμοποιεί έναν κωδικό που αποτελείται από ένα γράμμα στην πρώτη θέση και έναν αριθμό στη δεύτερη, τρίτη και τέταρτη θέση. Ο τέταρτος χαρακτήρας είναι σαν ένας αριθμός με τελεία στο τέλος. Υπάρχουν πολλοί διαφορετικοί κωδικοί αριθμοί. Μερικά είναι 4000, άλλα 4999. Ένα νέο σύστημα χρησιμοποιεί νέες εμπειρίες και γνώσεις για να καλύψει ορισμένα κενά και παρουσιάζει μια νέα δομή. Χρησιμοποιεί έναν αλφαριθμητικό κωδικό, όπως το P. Υπάρχουν δύο ασθένειες με τον ίδιο κωδικό, ο ερυθματώδης



λύκος και ο ορός δερματικός ερυθματώδης λύκος. Το ένα είναι ασθένεια του δέρματος και το άλλο είναι ασθένεια του αίματος [131].

Συμπερασματικά, το ICD-10 είναι ένα σύστημα που σας επιτρέπει να κοιτάζετε διαφορετικά αρχεία, να κατανοείτε τι συνέβη με τους ανθρώπους σε αυτά και να τα συγκρίνετε. Αυτό μπορεί να σας βοηθήσει να καταλάβετε πόσο κοινές είναι οι διάφορες ασθένειες και οι θάνατοι και πώς αλλάζουν με την πάροδο του χρόνου.

### 3.4.3.HL7 v2 (Health Level Seven v2)

Το HL7 v2 επικεντρώθηκε στην ανταλλαγή μηνυμάτων μεταξύ διαφορετικών τοποθεσιών. Αυτό απαιτούσε κάθε τοποθεσία να προσαρμόζει τα μηνύματα με τρόπο που τα έκανε να φαίνονται διαφορετικά, γεγονός που οδήγησε σε κάποια ασυνέπεια μεταξύ των διαφορετικών υλοποιήσεων [133].

Το HL7 είναι ένα σύνολο προτύπων για την ανταλλαγή ιατρικών πληροφοριών. Αυτά τα πρότυπα αναπτύχθηκαν τη δεκαετία του 1980, όταν διαφορετικές ιατρικές τεχνολογίες χρησιμοποιούσαν διαφορετικές μορφές. Ωστόσο, τώρα που έχουμε πιο προηγμένες τεχνολογίες, τα πρότυπα του HL7 είναι ξεπερασμένα. Ο οργανισμός HL7 πιστεύει ότι η εξέλιξη των μηνυμάτων με τρόπο που να ανταποκρίνεται στις ατομικές ανάγκες είναι ο καλύτερος τρόπος για να το πετύχουμε. Οι σημασιολογικές πληροφορίες σχετικά με τις πληροφορίες δεν μπορούν να προσδιοριστούν αυτόματα. Η σύνταξη κωδικοποίησης για μηνύματα HL7 v2 είναι λίγο διαφορετική από αυτή που έχετε συνηθίσει. Αντί να χρησιμοποιούν γράμματα και αριθμούς, αυτά τα μηνύματα χρησιμοποιούν σύμβολα που δεν είναι εύκολα κατανοητά από τον άνθρωπο [134].

### 3.4.4.HL7 v3 (Health Level Seven v3)

Υπάρχουν ορισμένα προβλήματα με την έκδοση 2 του HL7, επομένως δημιουργήθηκε μια νέα έκδοση που ονομάζεται HL7 έκδοση 3 [134]. Το πρότυπο παρέχει έναν τρόπο να λέμε τα πράγματα με τρόπο που τα κάνει πιο κατανοητά. Τα μηνύματα HL7 v3 έχουν δύο μέρη: ένα μήνυμα εκπομπής (MSH [SFT] [MSA]) και ένα περιτύλιγμα πράξης ελέγχου συμβάντων ενεργοποίησης (EVN [ERR]). Το μήνυμα εκπομπής είναι παρόμοιο με το MSH [SFT] [MSA], αλλά περιλαμβάνει σφάλματα που σχετίζονται με το Αίτημα επιβεβαίωσης. Το νέο πρότυπο για την αναπαράσταση πληροφοριών και την ανταλλαγή μηνυμάτων είναι πιο προηγμένο από το παλιό. Έχει σχεδιαστεί για να είναι συμβατό με την ανάπτυξη του σημασιολογικού ιστού και βασίζεται σε νεότερη τεχνολογία.

Το HL7 v3 είναι μια πλατφόρμα που μπορεί να χρησιμοποιηθεί για την εφαρμογή σύγχρονων προτύπων για καταναμημένα συστήματα, όπως το eHealth. Αυτό σημαίνει ότι μπορεί να χρησιμοποιηθεί για πειραματισμό με νέα πρότυπα για καταναμημένα συστήματα, τα οποία μπορεί να είναι χρήσιμα και σε άλλους τομείς. Η εφαρμογή έχει σχεδιαστεί για να βοηθά με όλες τις πτυχές των ροών εργασίας της υγειονομικής περίθαλψης [133]. Το Διαδίκτυο άρχισε να αναπτύσσεται τη δεκαετία του 1990 και το 2005 δημιουργήθηκε μια τυποποιημένη έκδοσή του.

Οπότε, το πρότυπο V3 βασίζεται σε έναν τυπικό τρόπο να κάνεις πράγματα και να χρησιμοποιείς αντικειμενοστραφείς αρχές. Στα Μηνύματα του v2, εξακολουθούμε να κάνουμε τα μηνύματα

αναγνώσιμα από τον άνθρωπο, αλλά σε μορφή XML. Η προσέγγιση μοντελοποίησης HL7 v3 βασίζεται σε ένα τυποποιημένο μοντέλο πληροφοριών που βοηθά στον καθορισμό της κύριας επιχειρηματικής λογικής οποιοδήποτε περιβάλλοντος υγειονομικής περίθαλψης. Από αυτό το μοντέλο, μπορούν να οριστούν συγκεκριμένα μηνύματα και έγγραφα.

Τα βασικά χαρακτηριστικά του HL7 v3 είναι:

- Η επαναχρησιμοποίηση των δεδομένων και οι τύποι στοιχείων του εκάστοτε μηνύματος.
- Η σύνταξη που χρησιμοποιείται είναι μέσω XML (καθαρή και ευανάγνωστη δομή).
- Η εξαγωγή μεθόδων XML Schemas.
- Η υποστήριξη για την ανταλλαγή υπηρεσιών «μηνυμάτων» καθώς και μιας προσέγγισης ηλεκτρονικού εγγράφου (Clinical Document Architecture - CDA).
- Το πρότυπο έχει σχεδιαστεί για να βοηθά τους οργανισμούς υγειονομικής περίθαλψης να μοιράζονται εύκολα δεδομένα. Ο υπολογιστής μπορεί να βοηθήσει στην ανταλλαγή ιατρικών δεδομένων υψηλής πολυπλοκότητας, όπως για υποστήριξη αποφάσεων ή ηλεκτρονικά αρχεία υγείας.

### Διαφορές HL7 v2 και HL7 v3

Σε κάποιο βαθμό, το HL7 v3 είναι το HL7 v2 επόμενης γενιάς καθώς βασίζεται σε επαναχρησιμοποιήσιμες δομές, έναν υποκείμενο ορισμό τύπων δεδομένων και χρησιμοποιεί ένα βιομηχανικό πρότυπο ως μέθοδο κωδικοποίησης σύνταξης. Η έκδοση 3 HL7 είναι πιο προηγμένη από την έκδοση 2, κάτι που είναι καλό γιατί μπορεί να χειριστεί πιο περίπλοκα ιατρικά δεδομένα. Είναι επίσης κατάλληλο για την ανταλλαγή πληροφοριών μεταξύ οργανισμών. Η HL7 v2 σχεδιάστηκε για απλούστερες ανταλλαγές δεδομένων, αλλά δεν είναι κατάλληλη για πιο σύνθετες ανταλλαγές. Η έκδοση 3 έχει σχεδιαστεί για πιο σύνθετες ανταλλαγές.

Δεδομένου ότι το HL7 v3 σχεδιάστηκε για να συλλαμβάνει πιο πολύπλοκα κλινικά δεδομένα, η δομή του μηνύματος επιτρέπει σχέσεις μεταξύ δραστηριοτήτων. Για παράδειγμα, υποθέτοντας ότι θέλουμε να μεταφέρουμε ότι «το Χ άτομο υποβλήθηκε σε επείγουσα χειρουργική επέμβαση και έλαβε φάρμακα Χ και Υ αφού διαγνώστηκε με ρήξη πνεύμονα που προσδιορίστηκε με χρήση υπερήχων», απλά δεν μπορούμε να εκφράσουμε τις διάφορες σχέσεις (π.χ. Το HL7 είναι μια ομάδα ιατρικών εμπειρογνομόνων που συνεργάζονται για να βελτιώσουν τον τρόπο κοινής χρήσης ιατρικών πληροφοριών). Αυτό σημαίνει ότι οι ιατροί μπορούν να εργαστούν πιο αποτελεσματικά όταν έχουν πρόσβαση στις πιο πρόσφατες πληροφορίες.

Το HL7 v3 μπορεί να δείξει σχέσεις μεταξύ διαφορετικών πραγμάτων δημιουργώντας ένα γράφημα. Το HL7 v3 είναι ένα σύστημα λογισμικού που επιτρέπει στους επαγγελματίες υγείας να επικοινωνούν μεταξύ τους πιο εύκολα. Αυτό το σύστημα βασίζεται σε μοντέλα που δημιουργούνται χρησιμοποιώντας μια αντικειμενοστραφή προσέγγιση. Αυτά τα μοντέλα χρησιμοποιούνται στη συνέχεια για τη δημιουργία μοντέλων πληροφοριών αναφοράς (Reference Information Model - RIM), τα οποία χρησιμοποιούνται για να βοηθήσουν τους επαγγελματίες υγείας να κατανοήσουν ο ένας τις πληροφορίες του άλλου. Αυτό το μοντέλο βοηθά να βεβαιωθείτε ότι τα διαφορετικά μηνύματα στη ζωή σας είναι συνεπή.

Στη γλώσσα της υγειονομικής περίθαλψης, μια αφηρημένη ή γενική παρατήρηση είναι ένας τύπος ιατρικού φακέλου που καλύπτει ένα ευρύ φάσμα θεμάτων. Αυτές οι εγγραφές ορίζονται με

συγκεκριμένους όρους, ώστε να μπορείτε να βρείτε εύκολα πληροφορίες για οτιδήποτε ψάχνετε. Η ικανότητα πρόβλεψης του πώς θα λειτουργήσουν τα πράγματα επιτρέπει πιο συνεπή συμπεριφορά και πιο προσαρμοσμένα μοντέλα για συγκεκριμένες καταστάσεις.

### 3.4.5.HL7 FHIR (Health Level Seven Fast Healthcare Interoperability)

Το FHIR είναι ένα σύνολο προτύπων που βοηθούν στον καθορισμό των τύπων πληροφοριών που μπορούν να συμπεριληφθούν στα ηλεκτρονικά αρχεία υγείας. Αυτές οι πληροφορίες μπορούν στη συνέχεια να κοινοποιηθούν εύκολα μεταξύ διαφορετικών παρόχων υγειονομικής περίθαλψης και συστημάτων [135].

Το πρότυπο HL7 FHIR δημιουργήθηκε από τον οργανισμό προτύπων υγειονομικής περίθαλψης HL7 για να βοηθήσει τους παρόχους υγειονομικής περίθαλψης να επικοινωνούν πιο εύκολα. Το FHIR είναι μια μορφή δεδομένων που βασίζεται σε παλαιότερα πρότυπα, όπως το v2 και το v3. Αυτό σημαίνει ότι είναι ευκολότερο να κατανοηθεί και να χρησιμοποιηθεί από ορισμένες νεότερες μορφές. Το API είναι πιο εύκολο στη χρήση επειδή χρησιμοποιεί σύγχρονη τεχνολογία API που βασίζεται στον ιστό, συμπεριλαμβανομένου ενός πρωτότυπου πρωτοκόλλου RESTful που βασίζεται σε HTTP, HTML και CSS για την εννοποίηση της διεπαφής χρήστη και είτε JavaScript Object Notation (JSON) είτε XML για την αναπαράσταση δεδομένων. Στόχος αυτού του οργανισμού είναι να βοηθήσει τα συστήματα υγειονομικής περίθαλψης να συνεργάζονται πιο εύκολα, έτσι ώστε οι πάροχοι υγειονομικής περίθαλψης και οι άνθρωποι να μπορούν να λαμβάνουν πληροφορίες και ιατρικές εφαρμογές από μια ποικιλία συσκευών.

Το FHIR είναι ένα πρωτόκολλο επικοινωνίας που χρησιμοποιεί το πρωτόκολλο HTTPS. Αυτό σημαίνει ότι οι πλατφόρμες ανάλυσης δεδομένων μπορούν να συλλέγουν πληροφορίες για μηνύματα σε πραγματικό χρόνο. Οι οργανισμοί υγειονομικής περίθαλψης μπορούν να συλλέγουν δεδομένα σε πραγματικό χρόνο από συγκεκριμένα μέρη των μηνυμάτων FHIR καθώς ταξιδεύουν μέσω του δικτύου. Αυτά τα δεδομένα μπορούν να αποθηκευτούν σε υπολογιστή ή άλλη ηλεκτρονική συσκευή. Αυτό θα μας βοηθήσει να το κατανοήσουμε και να το χρησιμοποιήσουμε πιο εύκολα με άλλες πληροφορίες που έχουμε. Η εφαρμογή θα μπορούσε να χρησιμοποιηθεί για πράγματα όπως η επιδημιολογική επιτήρηση, η παρακολούθηση συνταγογράφησης φαρμάκων και οι ειδοποιήσεις ανεπιθύμητων αλληλεπιδράσεων φαρμάκων.

Το FHIR τεχνικά αποτελείται από ένα καλό υπόβαθρο, το οποίο έχει τα ακόλουθα χαρακτηριστικά [136]:

- Λεπτομερής τρόπος ανταλλαγής δεδομένων χωρίς την άκαμπτη ροή εργασίας του παραδοσιακού HL7.
- Αποφυγή επιβάρυνσης του SOAP αφού χρησιμοποιείται μια απλή προσέγγιση RESTful.
- Εφαρμογή συμμόρφωσης και αναφοράς ως μέρος της βασικής διαδικασίας.
- Επικέντρωση της λύσης του 80% των κοινών περιπτώσεων χρήσης αντί του 20% των εξαιρέσεων.
- Το FHIR συνδυάζει τα καλύτερα χαρακτηριστικά των HL7 v2, HL7 v3 και CDA, αξιοποιώντας τις πιο σύγχρονες τεχνολογίες υπηρεσιών ιστού.
- Το FHIR έχει σχεδιαστεί για να λειτουργεί με υπηρεσίες web, πράγμα που σημαίνει ότι είναι εύκολο στη χρήση και κατανοητό. Η πλειοψηφία των προφίλ ΗΦΥ βασίζεται σε

υπηρεσίες web SOAP, αλλά αυτό βασίζεται σε XML. Με τις υπηρεσίες ιστού RESTful, μπορείτε εύκολα να κάνετε βασικά πράγματα όπως δημιουργία, ανάγνωση, ενημέρωση και διαγραφή (CRUD) σε μια ιστοσελίδα.

- Το FHIR είναι ένα σύστημα που αποτελείται από διαφορετικά κομμάτια που ονομάζονται «πόροι». Αυτοί οι πόροι μπορούν να χρησιμοποιηθούν για την επίλυση προβλημάτων με πρακτικό τρόπο. Οι πόροι μπορούν να χρησιμοποιηθούν για τη δημιουργία πιο εξατομικευμένων λύσεων στην απαίτηση της υγειονομικής περίθαλψης για επιλογή και προσαρμογή. Αν θέλουμε να προσθέσουμε μια επέκταση στο σύστημά μας, μπορούμε να χρησιμοποιήσουμε την ίδια γλώσσα προγραμματισμού και τα ίδια εργαλεία που χρησιμοποιούμε για άλλους πόρους. Αυτό σημαίνει ότι το σύστημα μπορεί εύκολα να διαβάσει και να χρησιμοποιήσει την επέκταση.

Το FHIR είναι ένα σύστημα που δημιουργεί πόρους που αντιπροσωπεύουν πληροφορίες με τρόπο που είναι εύκολο να κατανοηθεί [136]. Κάθε πόρος περιλαμβάνει συγκεκριμένα πεδία δεδομένων και τοποθεσίες, καθιστώντας το εύκολο στην κατανόηση. Οι ορισμοί των κλινικών πόρων στο FHIR είναι συγκεκριμένες και εύκολα κατανοητές έννοιες, όπως συνταγές φαρμάκων, ανεπιθύμητες ενέργειες, διαδικασίες και καταστάσεις. Για παράδειγμα, ένας πόρος Συνταγογράφησης Φαρμάκων αναφέρει ρητά τον συνταγογράφο του (FHIR Practitioner), τον ασθενή του (FHIR Patient) και το φάρμακο που συνταγογραφήθηκε (FHIR Medication). Η προδιαγραφή FHIR περιλαμβάνει μια ενσωματωμένη δυνατότητα που επιτρέπει σε άλλους οργανισμούς να δημιουργούν τους δικούς τους ορισμούς πόρων για κλινικά αρχεία. Με αυτόν τον τρόπο, μπορείτε εύκολα να προσθέσετε συγκεκριμένες λεπτομέρειες σχετικά με τις εμπειρίες υγειονομικής περίθαλψης των ασθενών σας στο αποθετήριο FHIR.

Στην ανάκτηση πληροφοριών υγειονομικής περίθαλψης, οι ορισμοί των πόρων δεν εγγυώνται πάντα ότι τα δεδομένα που ανακτώνται θα είναι σημασιολογικά συνεπή. Αυτό σημαίνει ότι τα δεδομένα ενδέχεται να μην ταιριάζουν με τις προσδοκίες κάποιου που αναζητά πληροφορίες για ένα συγκεκριμένο θέμα. Για να κρατήσουμε διαφορετικά πράγματα ξεχωριστά, χρησιμοποιούμε διαφορετικά πράγματα για να τα διαχωρίσουμε. Για παράδειγμα, μπορεί να χρησιμοποιήσουμε ένα κομμάτι χαρτί για να κρατήσουμε ξεχωριστά πράγματα όπως το φαγητό και το ποτό. Συστήματα πληροφοριών υγείας (HMY, ροές εργασιών αναφοράς δημόσιας υγείας, κινητές συσκευές), και οι πόροι πληροφοριών υγείας (πόροι FHIR) μπορεί να χρησιμοποιούν διαφορετικούς όρους για να περιγράψουν το ίδιο πράγμα.

Για παράδειγμα, ένας ΗΦΥ μπορεί να χρησιμοποιεί τον όρο "ασθενής" για να αναφέρεται σε κάποιον που υποβάλλεται σε θεραπεία, ενώ ένας πόρος FHIR μπορεί να χρησιμοποιεί τον όρο "αρχείο ασθενούς" για να αναφέρεται σε όλες τις πληροφορίες σχετικά με έναν ασθενή, συμπεριλαμβανομένου του ιατρικού του ιστορικού. Η σημασιολογική συνέπεια βασίζεται σε ένα στρώμα αφαίρεσης που ονομάζεται προφίλ FHIR που βοηθά στον καθορισμό πόρων σε συγκεκριμένα περιβάλλοντα. Η δημιουργία προφίλ μπορεί να σας βοηθήσει να βεβαιωθείτε ότι διαφορετικά προγράμματα μπορούν να συνεργαστούν, ανακαλύπτοντας ποιες είναι οι απαιτήσεις τους, ποια σχήματα κωδικοποίησης πρέπει να χρησιμοποιηθούν και ποιους κανόνες πρέπει να ακολουθήσετε σχετικά με τον τρόπο αναπαράστασης των δεδομένων στους πόρους. Το FHIR API παρέχει μια σύγχρονη διεπαφή προσανατολισμένη στους πόρους για αναζήτηση,

δημιουργία, ανάγνωση, ενημέρωση και διαγραφή πόρων FHIR που αντιπροσωπεύουν κλινικά, διοικητικά δεδομένα ερωτημάτων σε επίπεδο πληθυσμού.

### **Οφέλη του FHIR**

Η πρόκληση για τα πρότυπα υγειονομικής περίθαλψης είναι πώς να διασφαλιστεί ότι όλοι λαμβάνουν την καλύτερη δυνατή φροντίδα, ακόμη και όταν η περίθαλψη γίνεται σε διαφορετικά μέρη. Καθώς οι προδιαγραφές γίνονται πιο λεπτομερείς, κοστίζει περισσότερο η εφαρμογή τους και γίνονται πιο περίπλοκες. Υπάρχουν δύο τρόποι να κάνουμε πράγματα: με τον τρόπο που τα κάνουμε συνήθως ή με έναν προσαρμοσμένο τρόπο. Αλλά ο προσαρμοσμένος τρόπος μπορεί να έχει πολλά προβλήματα. Το FHIR επιλύει αυτήν την πρόκληση δημιουργώντας ένα σύστημα κοινής χρήσης και προσαρμογής των υπαρχόντων πόρων. Με αυτόν τον τρόπο, ο καθένας μπορεί να χρησιμοποιήσει τους καλύτερους διαθέσιμους πόρους για να καλύψει καλύτερα τις ανάγκες του. Όλες οι επεκτάσεις (συμπεριλαμβανομένων αυτών που χρησιμοποιούνται από το σύστημα) είναι εύκολα προσβάσιμες και μπορούν να ανακτηθούν χρησιμοποιώντας τα ίδια εργαλεία με άλλους πόρους.

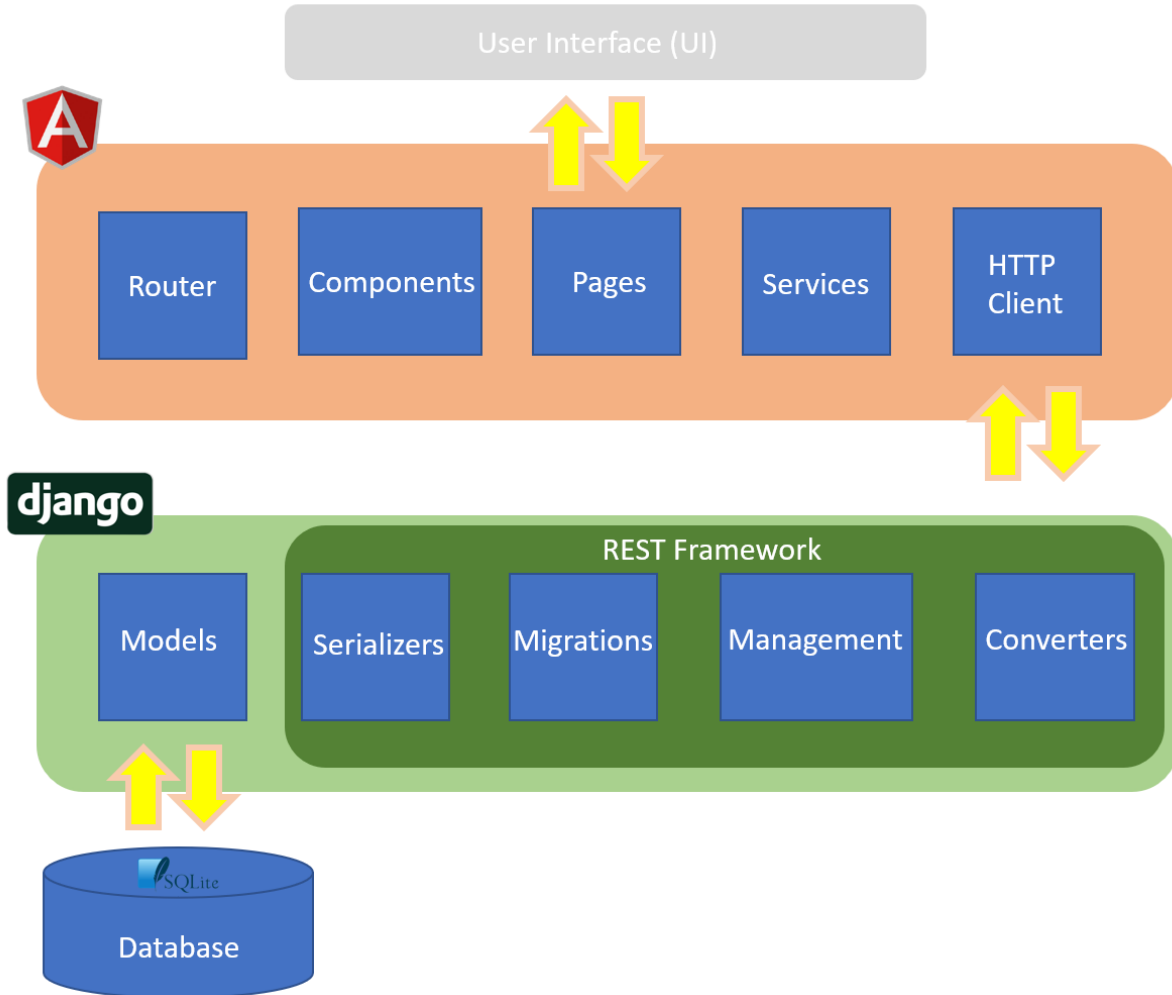
Μεταξύ άλλων, το FHIR προσφέρει πολλές βελτιώσεις σε σχέση με τα υπάρχοντα πρότυπα [135]:

- Γρήγορη και εύκολη υλοποίηση
- Πολλαπλές βιβλιοθήκες υλοποίησης, και πολλά παραδείγματα διαθέσιμα για την εκκίνηση ανάπτυξης
- Δωρεάν για χρήση χωρίς περιορισμούς
- Αναπτυξιακή πορεία εξέλιξης από τα πρότυπα HL7 v2 και CDA μπορεί να συνυπάρχουν και να αλληλοσυμπληρώνονται
- Διάθεση μίας ισχυρής βάσης στα πρότυπα XML, JSON, HTTP κ.ά.
- Υποστήριξη για RESTful αρχιτεκτονικές και επίσης απρόσκοπτη ανταλλαγή πληροφοριών με τη χρήση μηνυμάτων ή εγγράφων
- Διάθεση συνοπτικά και εύκολα κατανοητών προδιαγραφών
- Χρήση αναγνώσιμων μηνυμάτων από τον άνθρωπο
- Χρήση οντολογιών με αυστηρή και επίσημη αντιστοίχιση για ορθότητα

## Κεφάλαιο 4

### Μελέτη Περίπτωσης

#### 4.1. Αρχιτεκτονική Συστήματος



Εικόνα 11: Αρχιτεκτονική Συστήματος "urhealth"

#### 4.2. Σχετικές Τεχνολογίες

Στην παρούσα ενότητα θα παρουσιαστούν οι τεχνολογίες που χρησιμοποιήθηκαν τόσο στην ανάπτυξη του Frontend, όσο και για την ανάπτυξη του Backend. Ζητούμενο της σημερινής εποχής είναι να μπορούν οι ειδικοί του χώρου της πληροφορικής, στην προκειμένη περίπτωση οι προγραμματιστές να γνωρίζουν πλήθος νέων τεχνολογιών για την ανάπτυξη των πληροφοριακών συστημάτων. Αυτό συμβαίνει, καθώς είτε σε μια δημόσια υπηρεσία είτε σε μια ιδιωτική υπηρεσία, τα δεδομένα και οι πόροι που χρειάζονται, δεν μπορούν να υποστηριχθούν από παλαιότερες τεχνολογίες.

Συνεπώς στην παρούσα εργασία, για την ανάπτυξη του πληροφοριακού συστήματος «up-health», υποστηρίζεται η υγειονομική περίθαλψη διαβητικών και covid-19 ασθενών μέσω του ΗΦΥ. Έτσι, λοιπόν σε αυτό το σημείο θα παρουσιαστούν οι τεχνολογίες που χρησιμοποιήθηκαν και είναι οι εξής:

### Angular Framework (Frontend)

Η δημιουργία επεκτάσιμου λογισμικού είναι μια πρόκληση, καθώς οι σύγχρονες εφαρμογές διεπαφών χρήστη (Frontend), συνοδεύονται από αυξανόμενη πολυπλοκότητα, όλο και περισσότερους επιχειρηματικούς κανόνες, και αυξανόμενο όγκο δεδομένων που φορτώνονται στην εφαρμογή. Ιδιαίτερα στον χώρο της υγειονομικής περίθαλψης κρίνεται απαραίτητο να παρέχεται ένα εξατομικευμένο και προηγμένο περιβάλλον για την παροχή μιας φιλικής διεπαφής χρήστη. Προκειμένου να αντιμετωπιστούν οι αναφερόμενοι παράγοντες για τη διατήρηση υψηλής ποιότητας απόδοσης και χρήσης, απαιτείται μία ισχυρή και καλά στηριζόμενη αρχιτεκτονική. Βάσει των προαναφερθέντων, στα πλαίσια της Διεπαφής Χρήστη του «up-health» θα χρησιμοποιηθεί το framework της Angular, το οποίο ικανοποιεί εξ ολοκλήρου όλα αυτά τα χαρακτηριστικά [137]. Απώτερος σκοπός είναι να διατηρηθεί η ταχύτητα αποτελεσματικής και αποδοτικής ανάπτυξης και η ευκολία προσθήκης νέων χαρακτηριστικών μακροπρόθεσμα στα πλαίσια της πλατφόρμας του «up-health». Για να επιτευχθεί αυτό, η Διεπαφή Χρήστη θα διασπαστεί σε μικρότερα κομμάτια, μέσω των επιπέδων αφαίρεσης. Η παρακάτω εικόνα απεικονίζει τη γενική ιδέα αυτής της αποσύνθεσης. Η ιδέα είναι να τοποθετηθεί η σωστή ευθύνη στο σωστό επίπεδο του συστήματος, αναφερόμενοι στο επίπεδο πυρήνα, αφαίρεσης ή παρουσίασης (Εικόνα 12) [138].



Εικόνα 12: Διεπαφή Χρήστη

- **Επίπεδο Παρουσίασης:** Οι μόνες ευθύνες αυτού του επιπέδου είναι η παρουσίαση και η ανάθεση. Με άλλα λόγια, παρουσιάζει το περιβάλλον χρήστη και μεταβιβάζει τις ενέργειες

του χρήστη στο βασικό επίπεδο, μέσω του επιπέδου αφαίρεσης. Γνωρίζει τι πρέπει να εμφανίσει και τι ενέργειες να πραγματοποιήσει, αλλά δεν γνωρίζει πώς πρέπει να χειρίζονται οι αλληλεπιδράσεις του χρήστη με το σύστημα.

- **Επίπεδο Αφαίρεσης:** Το επίπεδο αφαίρεσης αποσυνδέει το επίπεδο παρουσίασης από το βασικό επίπεδο. Στην ουσία το εν λόγω επίπεδο εκθέτει τις ροές της κατάστασης και της διεπαφής για τα στοιχεία στο επίπεδο παρουσίασης, αντιπροσωπεύοντας το ρόλο της πρόσοψης.
- **Επίπεδο Πυρήνα:** Το τελευταίο επίπεδο, το επίπεδο πυρήνα, αποτελεί το βασικό επίπεδο του συστήματος. Εδώ εφαρμόζεται η βασική λογική εφαρμογών, όπου λαμβάνουν χώρα όλοι οι χειρισμοί των δεδομένων και η επικοινωνία με τον εξωτερικό κόσμο.

Εφαρμόζοντας την παραπάνω αρχιτεκτονική στη Διεπαφή Χρήστη του «up-health», εξασφαλίζεται η ομαλή και αποθεματική λειτουργία της με την πάροδο του χρόνου.

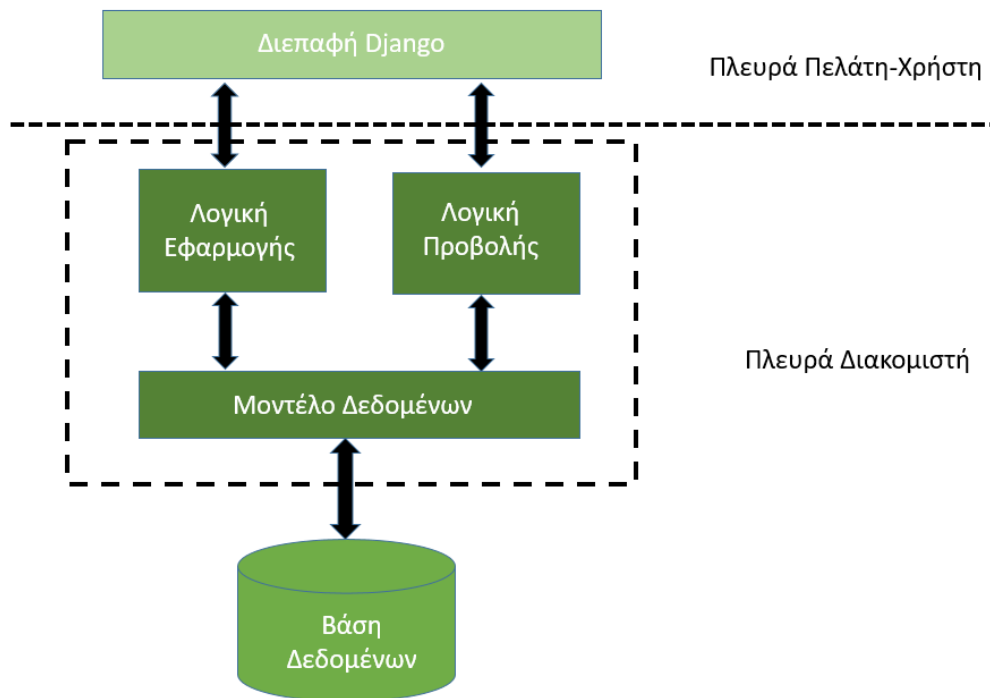
### Django Framework (Backend)

Το Django είναι ένα δομημένο πλαίσιο βασισμένο στην ευκολότερη και διαθέσιμη γλώσσα προγραμματισμού (Python). Επίσης, το Django ακολουθεί μια αρχιτεκτονική Model-View-Controller (MVC), η οποία χωρίζεται σε τρία διαφορετικά μέρη:

- Το **Μοντέλο (Model)** είναι η λογική δομή δεδομένων πίσω από ολόκληρη την εφαρμογή και αντιπροσωπεύεται από μια βάση δεδομένων (γενικά σχεσιακές βάσεις δεδομένων όπως την MySQL, Postgres).
- Η **Προβολή (View)** είναι το περιβάλλον εργασίας χρήστη, αυτό που βλέπετε στο πρόγραμμα περιήγησής σας όταν επισκέπτεστε έναν ιστότοπο. Αυτά αντιπροσωπεύονται από αρχεία HTML / CSS / Javascript.
- Ο **Ελεγκτής (Controller)** είναι ο μεσάζων που συνδέει την προβολή και το μοντέλο μαζί, πράγμα που σημαίνει ότι είναι αυτό που διαβιβάζει δεδομένα από το μοντέλο στην προβολή.

Με το MVC, η εφαρμογή σας θα περιστρέφεται γύρω από το μοντέλο, είτε θα το εμφανίζει είτε θα το χειρίζεται (Εικόνα 13) [139].





Εικόνα 13: Υλοποίηση Backend Αρχιτεκτονικής

Έτσι λοιπόν, έστω ότι ένας χρήστης θα εισάγει μια διεύθυνση URL στο πρόγραμμα περιήγησής του. Στην συνέχεια, το αίτημα θα περάσει από τα πρωτόκολλα διαδικτύου, στον διακομιστή σας, ο οποίος θα καλέσει το Django. Στη συνέχεια, το Django θα επεξεργαστεί τη δεδομένη διαδρομή URL και εάν ταιριάζει με μια διαδρομή URL που έχετε δηλώσει ρητά, θα καλέσει τον **Ελεγκτή**, ο οποίος θα εκτελέσει μια συγκεκριμένη ενέργεια, όπως λήψη μιας καταχώρισης από το **Μοντέλο** (βάση δεδομένων) και στη συνέχεια, θα δώσει μια **Προβολή** (δηλαδή ένα κείμενο JSON, HTML / CSS / JavaScript) [140].

#### 4.2.1. Υποστήριξη εφαρμογής μέσω Docker

Το Docker, ένα υποσύνολο του έργου Moby, καθώς αποτελεί ένα πλαίσιο λογισμικού για τη δημιουργία, τη λειτουργία και τη διαχείριση κοντέινερ σε διακομιστές και στο cloud. Ο όρος "docker" μπορεί να αναφέρεται είτε στα εργαλεία (τις εντολές) είτε στη Dockerfile μορφή αρχείου. Για την ιστορία το Moby είναι ένα ανοιχτό πλαίσιο που δημιουργήθηκε από την Docker για τη συναρμολόγηση εξειδικευμένων συστημάτων εμπορευματοκιβωτίων χωρίς την επανεφεύρεση του τροχού.

Η τεχνολογία container μπορεί να θεωρηθεί ως τρεις διαφορετικές κατηγορίες:

- **Builder:** ένα εργαλείο ή μια σειρά εργαλείων που χρησιμοποιούνται για την κατασκευή ενός container, όπως το distrobuilder για LXC ή ένα Dockerfile για το Docker.

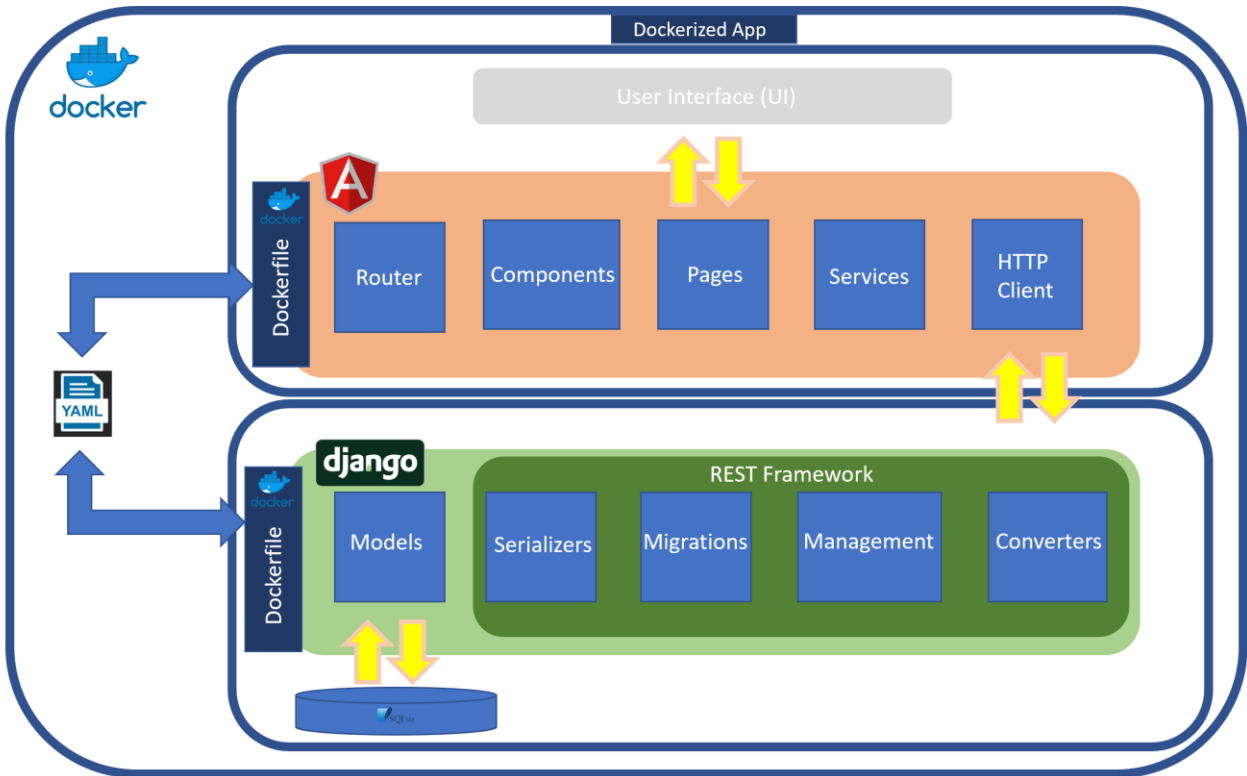
- Engine: μια εφαρμογή που χρησιμοποιείται για τη λειτουργία ενός container. Για το Docker, αυτό αναφέρεται στην εντολή **docker dockerd**. Για άλλους, αυτό μπορεί να αναφέρεται στον **containerd** και τις σχετικές εντολές (όπως το **podman**).
- Εννοχρήστρωση: τεχνολογία που χρησιμοποιείται για τη διαχείριση πολλών container, συμπεριλαμβανομένων των Kubernetes και OKD.

Στην παρούσα εργασία η χρήση του Docker, έγινε με την υλοποίηση του docker-compose.yml αρχείου, που πρακτικά αναφέρονται τα properties των containers που εφαρμόζονται επί της εφαρμογής. Στο παρών σύστημα που αναπτύχθηκε, υλοποιούνται 2 containers. Αυτά είναι:

1. **Angular\_web**: Σκοπός του παρόντος container είναι να επικοινωνήσει με το αντίστοιχο dockerfile που υπάρχει εντός του sub-project της Angular, προκειμένου να τραβήξει όλα τα dependencies και requirements που απαιτούνται για να τρέξει η εφαρμογή στον χρήστη. Αυτό έχει ως αποτέλεσμα της αποφυγή σπατάλης χρόνου για εγκαταστάσεις προγραμμάτων και βιβλιοθηκών από τον χρήστη που λαμβάνει την εφαρμογή.
2. **Django\_web**: Σκοπός του παρόντος container είναι να επικοινωνήσει με το αντίστοιχο dockerfile που υπάρχει εντός του sub-project της Django, προκειμένου να τραβήξει όλα τα dependencies και requirements που απαιτούνται για να τρέξει η εφαρμογή στον χρήστη.

Έτσι, αποφεύγει ο χρήστης την σπατάλη χρόνου για εγκαταστάσεις και ενσωμάτωση βιβλιοθηκών που απαιτούνται για να είναι λειτουργική η εφαρμογή στην απέναντι πλευρά του χρήστη που επιθυμεί να δοκιμάσει το run του project «up-health». Αυτό γίνεται με 2 απλές εντολές, θα πρέπει ο χρήστης να κάνει clone το project τοπικά στον υπολογιστή του και στην συνέχεια να τρέξει την εντολή docker-compose up.

Η αρχιτεκτονική του project με docker, που περιεγράφηκε παραπάνω, αποτυπώνεται στην εικόνα που ακολουθεί (Εικόνα 14).



Εικόνα 14: Αρχιτεκτονική Dockerized Εφαρμογής

### 4.3.Βάση Δεδομένων

Στην ανάπτυξη του πληροφοριακού συστήματος «ur-health», χρησιμοποιήθηκε η σχεσιακή βάση δεδομένων SQLite. Η SQLite είναι μια ενσωματωμένη μηχανή βάσης δεδομένων SQL.

Σε αντίθεση με τις περισσότερες άλλες βάσεις δεδομένων SQL, η SQLite δεν διαθέτει ξεχωριστή διαδικασία διακομιστή. Η SQLite διαβάζει και γράφει απευθείας σε συνηθισμένα αρχεία δίσκου. Μια πλήρης βάση δεδομένων SQL με πολλούς πίνακες, δείκτες, κανόνες ετικέτας και προβολές, περιέχεται σε ένα αρχείο δίσκου. Η μορφή αρχείου βάσης δεδομένων είναι πολλαπλής πλατφόρμας και μπορείτε ελεύθερα να αντιγράψετε μια βάση δεδομένων μεταξύ συστημάτων 32-bit και 64-bit ή μεταξύ αρχιτεκτονικών big-endian και small-endian. Αυτές οι δυνατότητες καθιστούν την SQLite μια δημοφιλή επιλογή ως μορφή αρχείου εφαρμογής.

Έτσι λοιπόν παρακάτω παρουσιάζονται οι πίνακες που δημιουργήθηκαν στην βάση δεδομένων του πληροφοριακού συστήματος ως εξής:

**Patient\_Diabetes** (Διαβητικός Ασθενής): Αποτελείται από τα χαρακτηριστικά που συνιστούν τον ΗΦΥ του κάθε ασθενούς. Παρακάτω δίνονται όλα τα χαρακτηριστικά αναφορικά ως εξής:

- Id
- first\_name
- last\_name
- age
- gender
- polyuria
- polydipsia
- sudden\_weight\_loss
- weakness
- polyphagia
- genital\_thrush
- visual\_blurring
- itching
- irritability
- delayed\_healing
- partial\_paresis
- muscle\_stiffness
- alopecia
- obesity
- class\_predict
- created\_at
- updated\_at

Παρακάτω παρουσιάζονται σε μορφή πίνακα, τα πρώτα χαρακτηριστικά από τυχαίους ασθενείς του πληροφοριακού συστήματος «ur-health» (Εικόνα 15).

id	first_name	age	gender	polyuria	polydipsia	sudden_weight_loss	weakness	polyphagia	genital_thrush	visual_blurring	itching	irritability	delayed_heal
1	Vasilis	40	Male	0	0	1	0	0	0	0	0	0	0
2	Athanasios	58	Female	1	1	1	1	1	0	1	0	0	0
3	Panos	41	Male	1	1	1	1	1	1	1	1	1	0
4	Dimitris	45	Male	0	0	0	0	1	1	0	0	0	0
5	Katerina	60	Male	0	0	1	0	0	0	0	0	0	0
6	Konstantina	55	Female	1	1	1	0	1	0	0	1	0	1
7	Athanasios	57	Male	0	0	0	0	1	0	1	0	0	0
8	Christos	66	Male	0	0	0	1	1	0	1	1	0	1
9	Christos	67	Male	0	0	0	1	0	0	0	1	0	1
10	Dimitris	70	Male	1	0	0	0	1	0	1	1	0	1
11	Christos	44	Male	0	0	0	0	0	0	0	0	0	0
12	Nikolas	38	Male	0	0	0	0	0	0	0	0	0	0

Εικόνα 15: Πίνακας Patient\_Diabetes

**Patient\_Covid** (Covid-19 Ασθενής): Αποτελείται από τα χαρακτηριστικά που συνιστούν τον ΗΦΥ του κάθε ασθενούς. Παρακάτω δίνονται όλα τα χαρακτηριστικά αναφορικά ως εξής:

- id
- first\_name
- last\_name
- patient\_id
- patient\_age\_quantile
- hematocrit
- hemoglobin
- platelets
- red\_blood\_Cells
- lymphocytes
- leukocytes
- basophils
- eosinophils
- monocytes
- serum\_Glucose
- neutrophils
- urea
- proteina
- creatinine
- potassium
- sodium
- alanine\_transaminase
- aspartate\_transaminase
- label\_predict
- created\_at
- updated\_at

Παρακάτω παρουσιάζονται σε μορφή πίνακα, τα πρώτα χαρακτηριστικά από τυχαίους ασθενείς του πληροφοριακού συστήματος «ur-health» (Εικόνα 16).

id	first_name	last_name	patient_id	patient_age_quantile	hematocrit	hemoglobin	platelets	red_blood_Cells	lymphocytes	leukocytes
1	Athanasios	Sloukas	acd761fe16b5d0f	17	-3.2425479889	-2.7792034149	-1.7735939026	-3.3182849884	-1.8309534788	-1.733674764
2	Aggeliki	Diamantidou	2697fdccbfeb7f7	19	0.6942868829	0.5415639281	-0.9068291187	0.5780237913	-0.2957255244	-1.288428068
3	Other	Other	126e9dd13932f68	17	0.2365154475	-0.0223402679	-0.5174130201	0.1020041853	0.3183657527	-0.094610348

Εικόνα 16: Πίνακας Patient\_Covid

**Appointment\_Diabetes** (Ραντεβού Διαβητικού Ασθενούς): Αποτελείται από τα κλινικά και εργαστηριακά στοιχεία μιας επίσκεψης ενός ασθενούς στον γιατρό του.

- id
- patient
- doctor
- arterial\_pressure
- respiratory
- digestive
- neurological
- sugar
- hemoglobin
- creatinine
- uric
- acid
- potassium
- sodium
- cholesterol
- triglycerides
- agdl
- ldl
- hematocrit
- general\_urine
- sgot
- sgdp
- ggd
- tsg
- appointment\_date

Παρακάτω παρουσιάζονται σε μορφή πίνακα, ένα συγκεκριμένο πλήθος ραντεβού που έχουν οριστεί για διάφορους ασθενείς (μπορεί και παραπάνω του ενός ραντεβού για έναν συγκεκριμένο ασθενή) από τον εκάστοτε γιατρό (Εικόνα 17).

id	arterial_pressure	respiratory	digestive	neurological	sugar	hemoglobin	creatinine	uric	acid	potassium	sodium	cholesterol	triglycerides	agdl	ldl	her
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	1	1	1	1	1	1	1	1	1	1	1	1	1	15	1	1
4	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
5	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11

Εικόνα 17: Πίνακας Appointment\_Diabetes

**Appointment\_Covid** (Ραντεβού Covid-19 Ασθενούς): Αποτελείται από τα κλινικά και εργαστηριακά στοιχεία μιας επίσκεψης ενός ασθενούς στον γιατρό του.

- id
- patient\_covid
- doctor
- tlc
- neutrophil
- lymphocyte
- monocyte
- eosinophil
- basophil
- hemoglobin
- platelets
- inr
- glucose
- urea
- creatine
- albumin
- alt
- ast
- ldh
- bilirubin
- appointment\_date

Παρακάτω παρουσιάζονται σε μορφή πίνακα, ένα συγκεκριμένο πλήθος ραντεβού που έχουν οριστεί για διάφορους ασθενείς (μπορεί και παραπάνω του ενός ραντεβού για έναν συγκεκριμένο ασθενή) από τον εκάστοτε γιατρό (Εικόνα 18).

id	tlc	neutrophil	lymphocyte	monocyte	eosinophil	basophil	hemoglobin	platelets	inr	glucose	urea	creatinine	albumin	alt	ast	ldh	bilirubin	appointment_date
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10	1	2022-01-15 1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2022-01-20 1

Εικόνα 18: Πίνακας Appointment\_Covid

**DoctorInfo** (Πληροφορίες Ιατρού): Αποτελείται από τα κυριότερα στοιχεία ενός γιατρού, όπως είναι ο μοναδικός του κωδικός στο πληροφοριακό σύστημα, καθώς και η ειδικότητά του.

- id
- speciality
- doctor\_id

Παρακάτω παρουσιάζονται σε μορφή πίνακα, ένα συγκεκριμένο πλήθος του ιατρικού προσωπικού που έχουν εγγραφεί και έχουν πρόσβαση στο πληροφοριακό σύστημα του «ur-health» (Εικόνα 19).

id	speciality	doctor_id
1	Dermatologist	2
2	Dermatologist	3
3	ophthalmologist	4
4	dermatologist	5

Εικόνα 19: Πίνακας DoctorInfo

**Mlmodel** (Μοντέλα Μηχανικής Μάθησης Διαβήτη): Αποτελείται από τα μοντέλα μηχανικής μάθησης που χρησιμοποιούνται είτε για την πρόβλεψη της καταλληλότερης φαρμακευτικής αγωγής.

- id
- name
- file
- algorithm
- classification\_result

Παρακάτω παρουσιάζονται σε μορφή πίνακα, τα διάφορα μοντέλα μηχανικής μάθησης που χρησιμοποιήθηκαν κατά την ανάπτυξη του πληροφοριακού συστήματος (Εικόνα 20).

id	name	file	classification_result	algorithm
1	patient_readmission_LR	LogisticRegression_model.pkl	readmission	LogisticRegression
2	patient_readmission_RFC	RandomForestClassifier_model.pkl	readmission	RandomForestClassifier
3	patient_readmission_DTC	DecisionTreeClassifier_model.pkl	readmission	DecisionTreeClassifier
4	patient_readmission_BNB	BernoulliNB_model.pkl	readmission	BernoulliNaiveBayes
5	patient_readmission_GNB	GaussianNB_model.pkl	readmission	GaussianNaiveBayes
6	patient_readmission_GBC	GradientBoostingClassifier_model.pkl	readmission	GradientBoostingClassifier
7	patient_readmission_KNC	KNeighborsClassifier_model.pkl	readmission	KNeighborsClassifier
8	patient_readmission_NN	Neural_Nets_model.pkl	readmission	NeuralNetworks
9	patient_readmission_SGD	Stochastic_Gradient_Descent_model.pkl	readmission	StochasticGradientDescent
10	patient_readmission_SVM	SupportVectorMachine_model.pkl	readmission	SupportVectorMachine

Εικόνα 20: Πίνακας MlModel\_Diabetes

**Mlmodel\_Covid** (Μοντέλα Μηχανικής Μάθησης Covid-19): Αποτελείται από τα μοντέλα μηχανικής μάθησης που χρησιμοποιούνται είτε για την πρόβλεψη εισαγωγής ενός ασθενούς στο νοσοκομείο και πιο συγκεκριμένα σε ΜΕΘ.

- id
- name
- file
- algorithm
- classification\_result

Παρακάτω παρουσιάζονται σε μορφή πίνακα, τα διάφορα μοντέλα μηχανικής μάθησης που χρησιμοποιήθηκαν κατά την ανάπτυξη του πληροφοριακού συστήματος (Εικόνα 21).



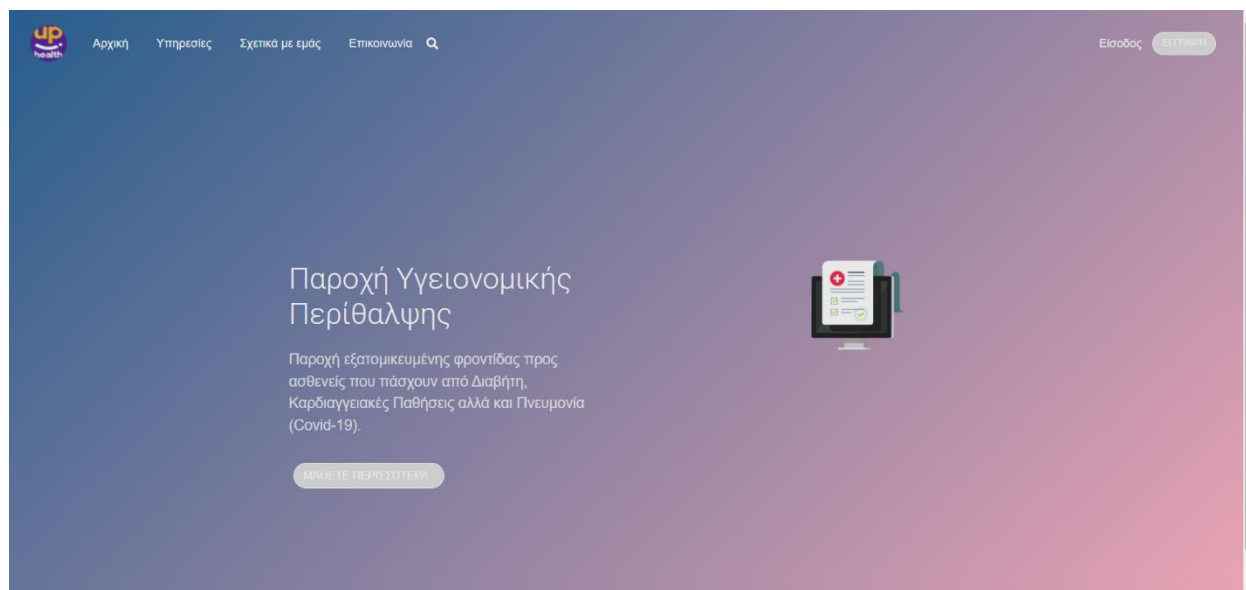
id	name	algorithm	classification_result	file
1	patient_covid_LR	LogisticRegression	covid	LogisticRegression_covid_model.pkl
2	patient_covid_RFC	RandomForestClassifier	covid	RandomForestClassifier_covid_model.pkl
3	patient_covid_DTC	DecisionTreeClassifier	covid	DecisionTreeClassifier_covid_model.pkl
4	patient_covid_BNB	BernoulliNaiveBayes	covid	BernoulliNB_covid_model.pkl
5	patient_covid_GNB	GaussianNaiveBayes	covid	GaussianNB_covid_model.pkl
6	patient_covid_GBC	GradientBoostingClassifier	covid	GradientBoostingClassifier_covid_model.pkl
7	patient_covid_KNC	KNeighborsClassifier	covid	KNeighborsClassifier_covid_model.pkl
8	patient_covid_NN	NeuralNetworks	covid	Neural_Nets_covid_model.pkl
9	patient_covid_SGD	StochasticGradientDescent	covid	Stochastic_Gradient_Descent_covid_model.pkl
10	patient_covid_SVM	SupportVectorMachine	covid	SupportVectorMachine_covid_model.pkl

Εικόνα 21: Πίνακας MIModel\_Covid

#### 4.4.Εγχειρίδιο Χρήσης Συστήματος

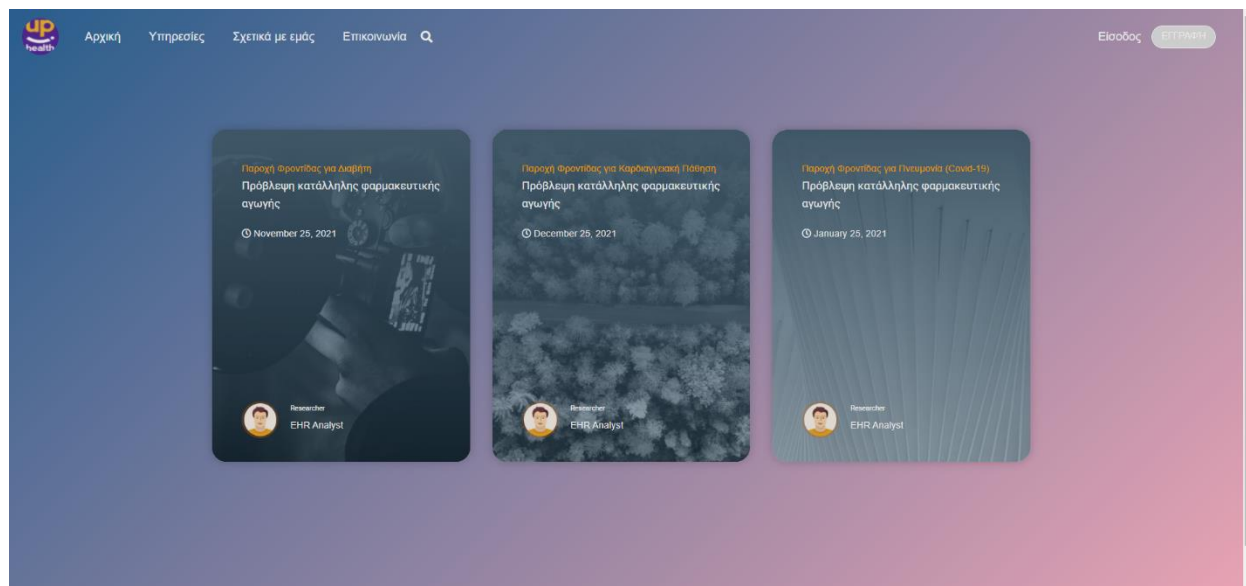
Στην παρούσα ενότητα θα αναλυθεί μέσω στιγμιότυπων οθόνης οι διαφορετικές σελίδες του πληροφοριακού συστήματος «ur-health» με τις αντίστοιχες επεξηγήσεις τους.

Στο παρακάτω στιγμιότυπο οθόνης, παρουσιάζεται η κεντρική σελίδα (**Main Page**) του πληροφοριακού συστήματος «ur-health» (Εικόνα 22).



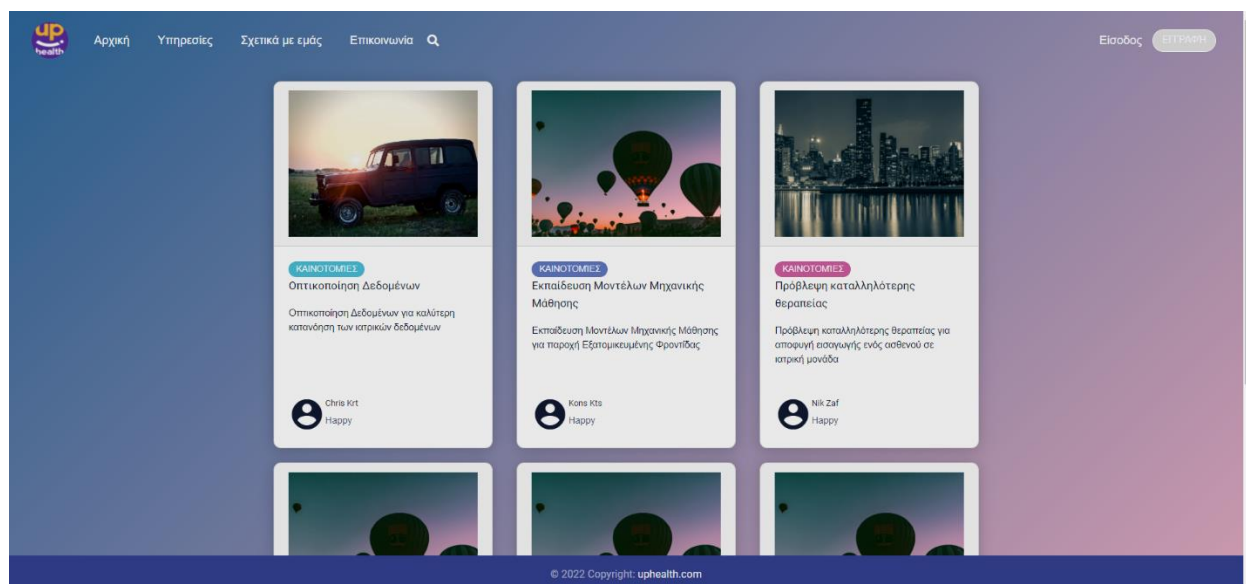
Εικόνα 22: Κύρια Σελίδα

Στην συνέχεια ακολουθεί το στιγμιότυπο οθόνης από την διεπαφή της σελίδας «**Services**», που περιέχει πληροφορίες σχετικά με τις λειτουργίες που υποστηρίζει το πληροφοριακό σύστημα (Εικόνα 23).



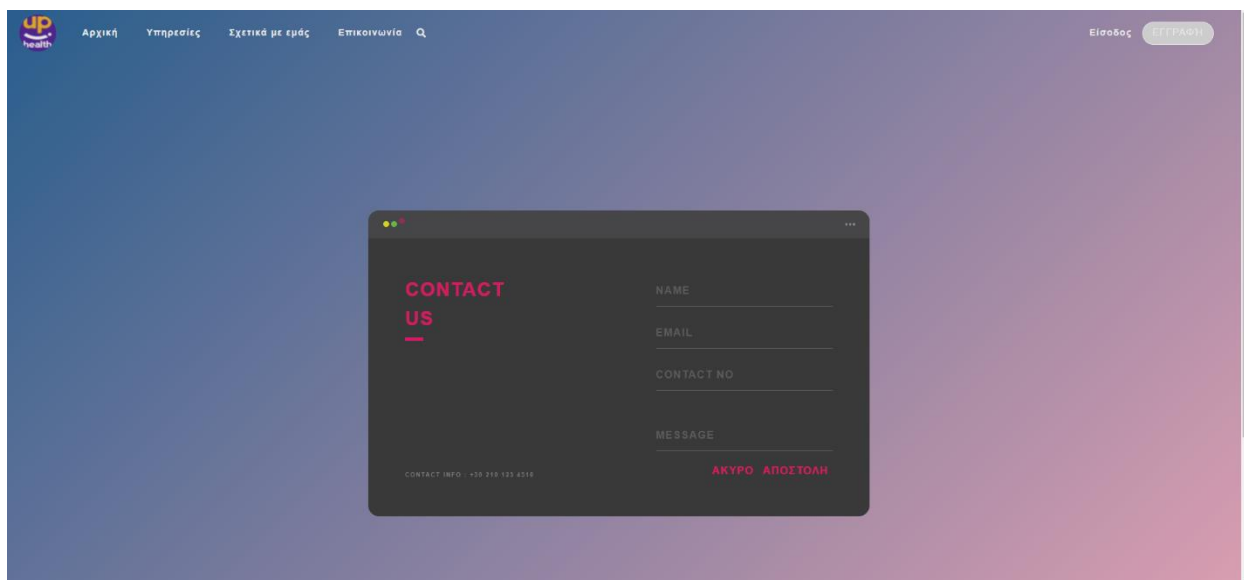
Εικόνα 23: Σελίδα Παροχής Υπηρεσιών

Στην συνέχεια ακολουθεί το στιγμιότυπο οθόνης από την διεπαφή της σελίδας «**About Us**», που περιέχει πληροφορίες σχετικά με τις καινοτομίες που υποστηρίζει το πληροφοριακό σύστημα (Εικόνα 24).

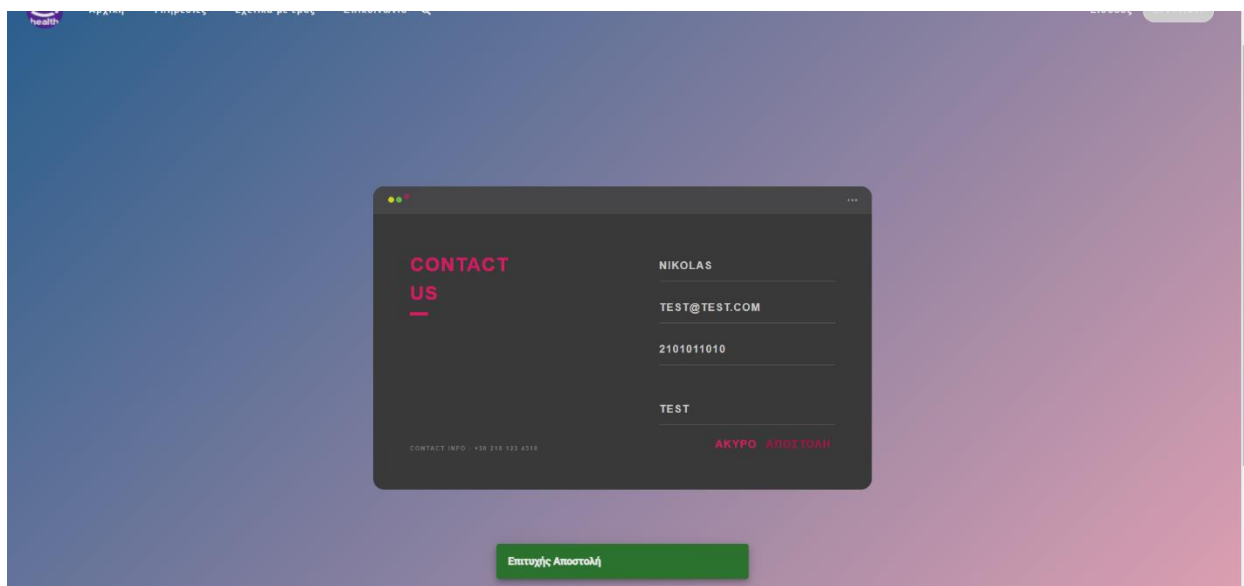


Εικόνα 24: Σελίδα Καινοτομιών

Έπειτα, ακολουθεί η οθόνη της διεπαφής του «**Contact Us**», σε περίπτωση που κάποιος επιθυμεί να επικοινωνήσει με τα μέλη της πλατφόρμας, στέλνοντας τα απαραίτητα στοιχεία όπως αποτυπώνονται στην παρακάτω εικόνα (Εικόνα 25).

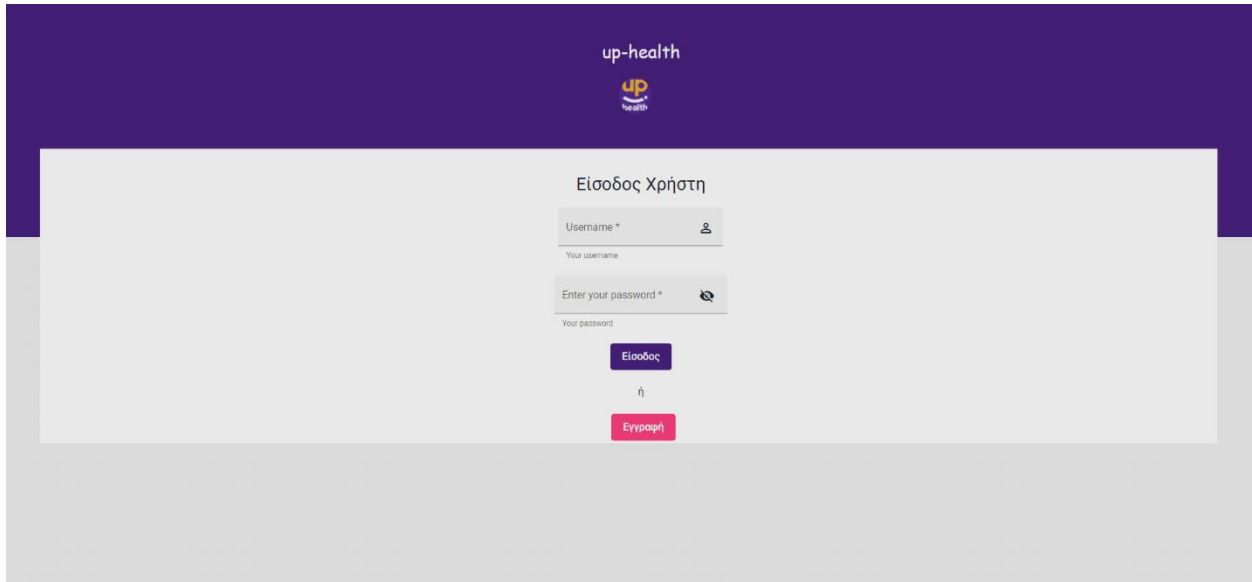


Εικόνα 25: Σελίδα Επικοινωνίας



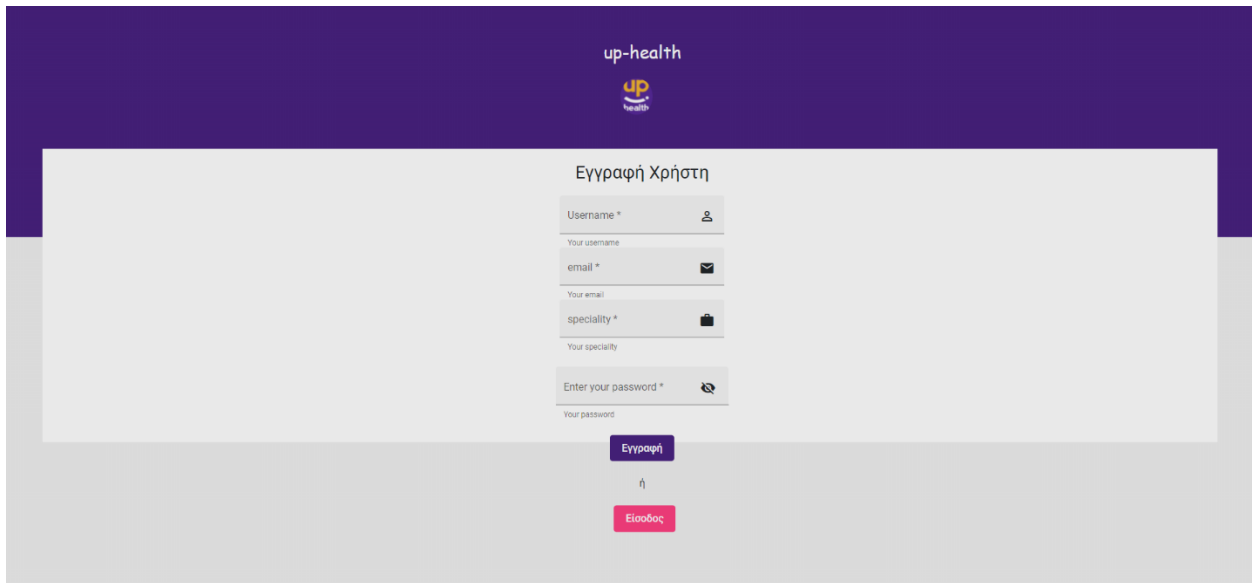
Εικόνα 26: Σελίδα Επικοινωνίας (Επιτυχής Αποστολή)

Σε περίπτωση που ο χρήστης (ιατρός ή πιο γενικά το ιατρικό προσωπικό) επιθυμεί να κάνει είσοδο και είναι ήδη εγγεγραμμένο μέλος τότε, μπορεί να βάλει το **Όνομα Χρήστη** και τον **Κωδικό** και να εισέλθει στο εσωτερικό περιβάλλον του πληροφοριακού συστήματος «up-health» (Εικόνα 27).



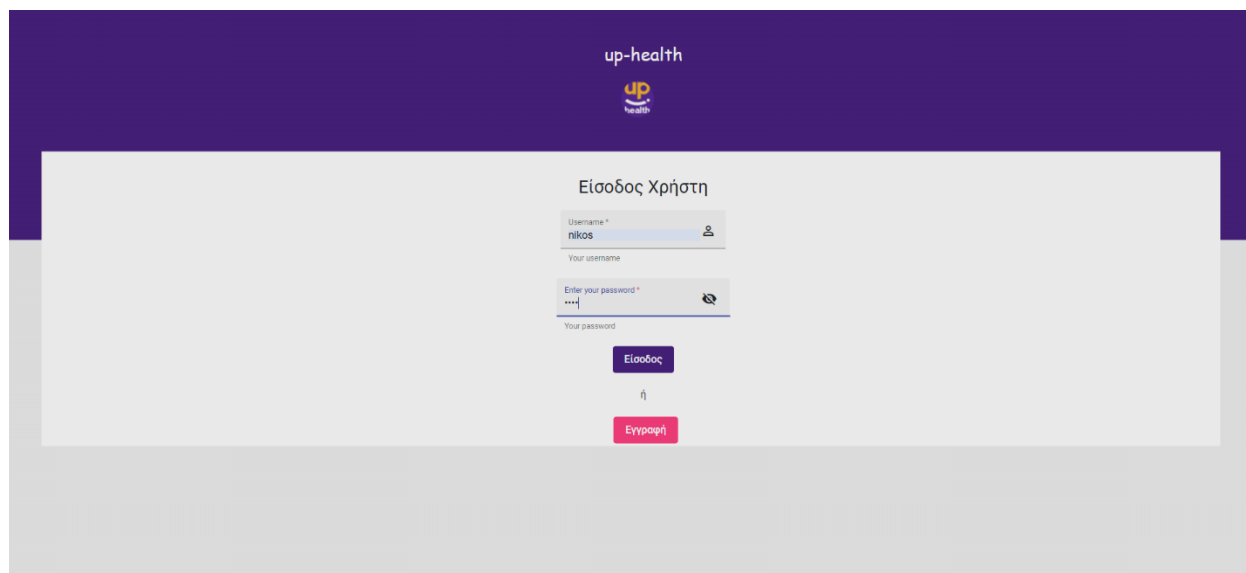
Εικόνα 27: Σελίδα Log In

Παρακάτω, στο στιγμιότυπο οθόνης που ακολουθεί, ένας νέος χρήστης μπορεί να κάνει την εγγραφή του στο πληροφοριακό σύστημα του «up-health», με την συμπλήρωση του **Όνοματος Χρήστη**, το **Προσωπικό του Ταχυδρομείο**, την **Ειδικότητα** του και τον **Κωδικό Πρόσβασης** του (Εικόνα 28).



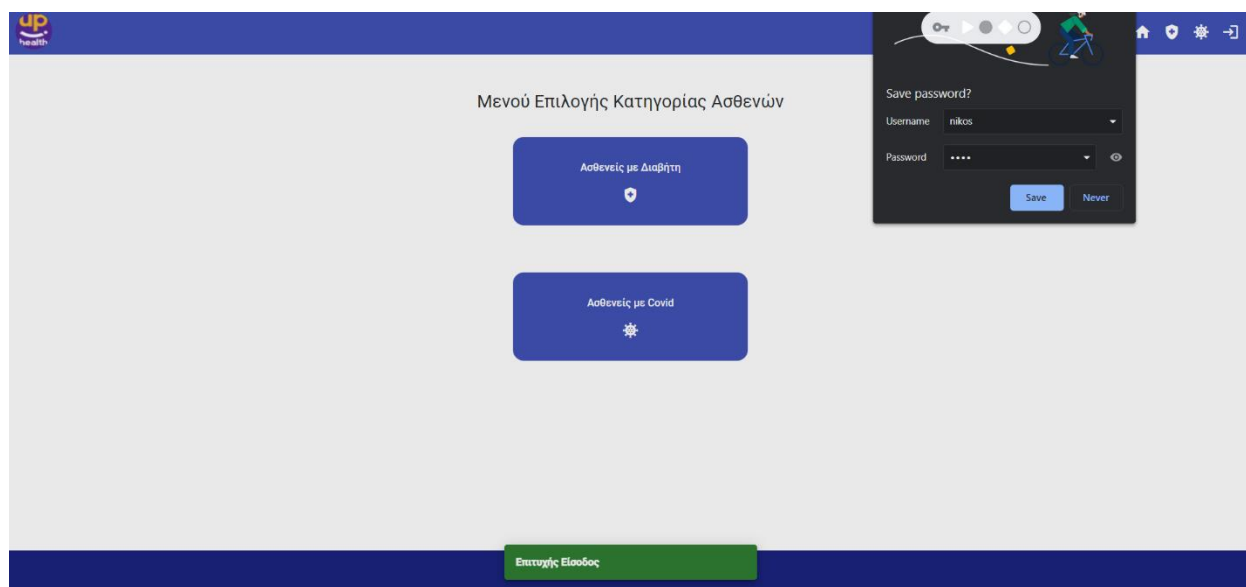
Εικόνα 28: Σελίδα Sign Up

Έτσι λοιπόν, παρατηρείται στην ακόλουθη εικόνα έναν χρήστη του πληροφοριακού συστήματος που προσπαθεί να εισέλθει με τα στοιχεία του (Εικόνα 29).



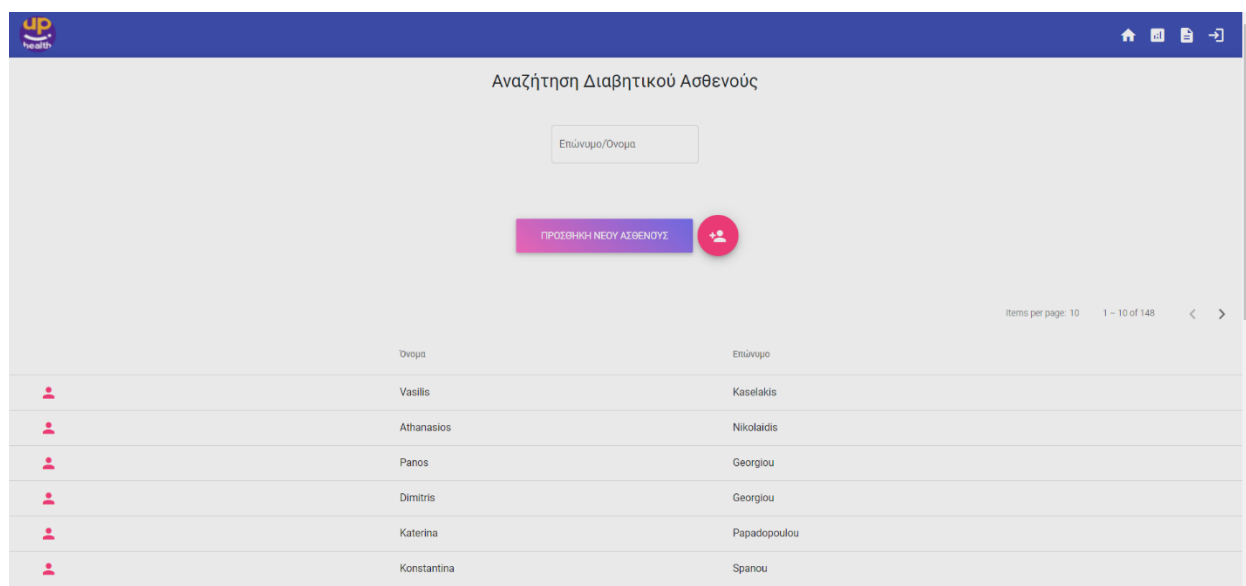
Εικόνα 29: Σελίδα Log In (Στοιχεία Χρήστη)

Στην συνέχεια βλέπουμε την κύρια σελίδα (**Home Info Page**) με το που εισέλθει ο χρήστης (γιατρός) και βλέπει το μενού επιλογής κατηγορίας ασθενών στο ψηφιακό αρχείο του συστήματός του (Εικόνα 30).



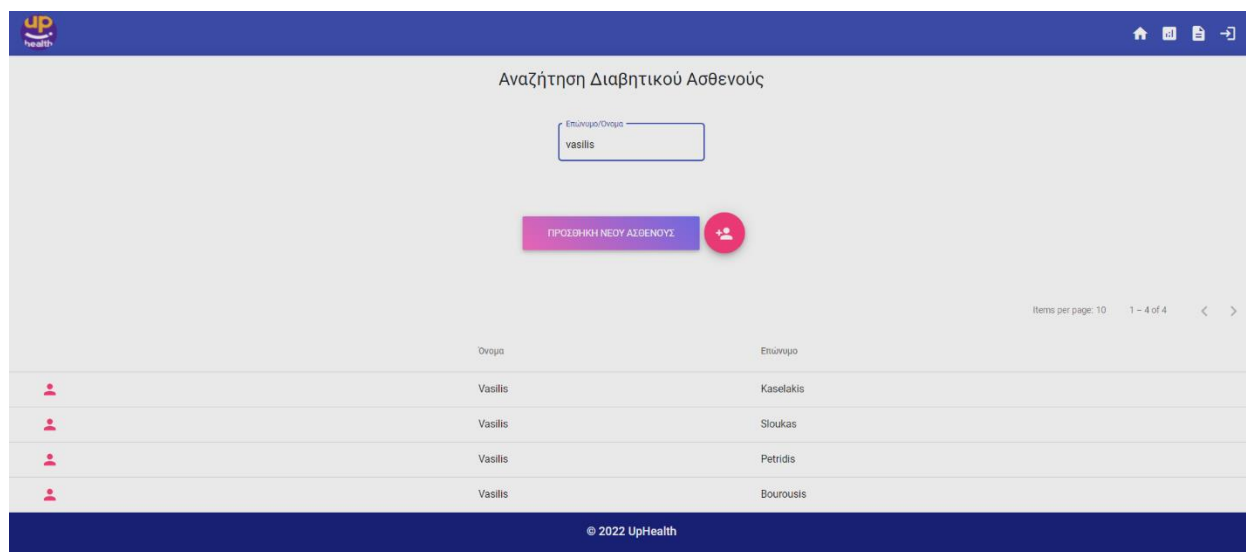
Εικόνα 30: Σελίδα Επιλογής Κατηγορίας Ασθενών

Στην συνέχεια έστω ότι επιλέγεται η κατηγορία ασθενών «Διαβήτη», τότε παρακάτω φαίνεται η κύρια σελίδα (**Home Diabetes Page**) με το που εισέλθει ο χρήστης (γιατρός) και βλέπει τον κατάλογο των ασθενών του, που διατηρεί στο ψηφιακό αρχείο του συστήματός του (Εικόνα 31).



Εικόνα 31: Σελίδα Αναζήτησης Διαβητικού Ασθενούς

Μετάπειτα, ο γιατρός μπορεί να πληκτρολογήσει το Όνομα ή το Επώνυμο ή και τα δύο, ενός ασθενή και μέσω του πεδίου αναζήτησης που του παρέχεται να βρει τον ασθενή. Επίσης, για παροχή πιο άμεσης πρόσβασης, το πεδίο αναζήτησης λειτουργεί με «**auto-filtering**», οπότε ο γιατρός πληκτρολογώντας τον ασθενή που αναζητά, του επιστρέφονται αυτόματα αποτελέσματα, βάσει των στοιχείων που έχει γράψει στο πεδίο αναζήτησης (Εικόνα 32).



Εικόνα 32: Αναζήτηση Ασθενούς "Vasilis"

Παρακάτω στο στιγμιότυπο οθόνη που ακολουθεί, παρατηρείτε συμπληρωμένα τα στοιχεία του ηλεκτρονικού φακέλου υγείας του ασθενούς. Το πλήθος των στοιχείων που αποτελούν τον ΗΦΥ, ποικίλει καθώς από έναν μοναδικό αριθμό που χαρακτηρίζει τον ασθενή και τα στοιχεία του όπως Όνομα και Επώνυμο, επεκτείνεται μέχρι βαθύτερες έννοιες της υγειονομικής περιθαλψης ενός ασθενούς (Εικόνα 33).

Πληροφορίες Ασθενούς

Αλληγορησιακό Ασθενή 1	Όνομα Vasilis	Επίσημο Kaselalis	Ηλικία 40
Φύλο Male	Πολυκορία No	Πολυθερία No	Σοφική απώλεια βάρους Yes
Αδυναμία No	Πολικρασία No	Μόλυνση Σιμής No	Οπτικό θόρυβος No
Φωνισμός No	Ειερβιστία No	Καθυστέρησή επώδυνη No	Μερική παράλυση No
Μυϊκή δυσκαμψία No	Ακμωπείωση No	Παρανορία Yes	

SAVE

Εικόνα 33: Εκτεταμένος Φάκελος Υγείας Διαβητικού Ασθενούς

Στην συνέχεια, παρατηρούνται και 2 επιπλέον δράσεις. Η μια είναι για τον έλεγχο της καταλληλότερης θεραπείας (αγωγής) για τον διαβητικό ασθενή. Αυτό γίνεται μέσα από τους ήδη εκπαιδευμένους αλγόριθμους Μηχανικής Μάθησης (Machine Learning). Η άλλη δράση αφορά τον ορισμό ή την ενημέρωση ενός ραντεβού (Εικόνα 34).

Έλεγχος Καταλληλότερης Θεραπείας

Μοντέλο πρόβλεψης καταλληλότερ...  
Logistic Regression

Selected: LR

ΠΡΟΒΛΕΨΗ

Appointment + Προσθήκη νέου appointment

	appointment_date	respiratory	digestive
	07 Jan 2022 17:20	1	1
	15 Jan 2022 12:54	2	2
	22 Jan 2022 00:59	1	1
	15 Jan 2022 19:01	10	10

Επιστροφή στον Κατάλογο Διαβητικών Ασθενών

Επιστροφή

Εικόνα 34: Επιλογή Μοντέλου ML και Ορισμός Ραντεβού

Στο ακόλουθο στιγμιότυπο, αποτυπώνεται μέσω των μοντέλων μηχανικής μάθησης, για το ποια είναι η καταλληλότερη αγωγή βάσει των στοιχείων του ασθενούς στον ηλεκτρονικό φάκελο υγείας. Έτσι λοιπόν, με βάση τα στοιχεία του ασθενούς στον προσωπικό του φάκελο και με το αντίστοιχο μοντέλο μηχανικής μάθησης που χρησιμοποιήθηκε, δείχνει ότι ο χρήστης θα πρέπει να κάνει χορήγηση της ινσουλίνης σε συνδυασμό με κάτι άλλο (Εικόνα 35).

The image shows a patient data entry form with the following fields:

Ανογνωτικό Ασθενή 1	Όνομα Vasilis	Επώνυμο Kaselakis	Ηλικία 40
Φύλο Male	Πολυνοσία No	Πολύνομία No	Εμφανική απώλεια βάρους Yes
Αδυναμία No	Πολιφαγία No	Μόλυνση Σιμής No	Οπτικό θάμπωμα No
Φαγοόρα No	Εμερήσιο No	Καθυσταρμένη επούλωση No	Μερική παράλυση No
Μυϊκή δυσκαμψία No	Αλωπεκίαση No	Παχυσαρκία Yes	

Below the form is a section titled "Έλεγχος Καταλληλότερης Θεραπείας" (Check Best Therapy). It includes a dropdown menu for "Μοντέλο πρόβλεψης καταλληλότερης" (Best prediction model) with "Logistic Regression" selected. Below this, it says "Selected: LR" and a "ΠΡΟΒΛΕΨΗ" (PREDICT) button. At the bottom, a message states: "Προτείνεται ως καταλληλότερη θεραπεία η ινσουλίνη σε συνδυασμό με κάτι άλλο" (Insulin is recommended as the best therapy in combination with something else).

Εικόνα 35: Επιλογή Μοντέλου ML Πρόβλεψης

Στην συνέχεια, στην παρακάτω εικόνα αποτυπώνονται οι κλινικές και εργαστηριακές εξετάσεις του ασθενούς. Αυτές πραγματοποιούνται με ραντεβού που κλείνει ο ασθενής με τον γιατρό του κατόπιν συνεννόησης. Στην συνέχεια, ο γιατρός περνάει τα στοιχεία του ασθενούς κατά την επίσκεψη του και έτσι ο γιατρός μπορεί και διατηρεί τις εξετάσεις που πραγματοποιούνται σε κάθε επίσκεψη του ασθενούς στο ιατρείο του (Εικόνα 36).

The image shows a form titled "Κλινικές και Εργαστηριακές Εξετάσεις" (Clinical and Laboratory Tests). It contains the following fields:

Ασπριλική Πίλη 1	Ασπιρικό 1	Ρετινικό 1
Νεφρολογικό 1	Σίκαρο 1.00	Αμυγραμμία 1.00
Κρεατινίνη 1.00	Ουρικό 1.00	Οξύ 1.00
Ηβικα 1.00	Νάτριο 1.00	Καλσιμίνης 1.00
Τριγλυκερίδια 1.00	ΑΔΟΛ 15.00	LDL 1.00
Αιμοκρίτης 1.00	Γενική αίμα 1.00	gHbT 1.00
gSDP 1.00	gSD 1.00	TSO 1.00

At the bottom, there are fields for "Ημερομηνία" (Date) set to 22/01/2022 and "Στα Ραντεβού" (Appointment) set to 00:59. A "ΑΠΟΘΗΚΕΥΣΗ" (SAVE) button is located below these fields.

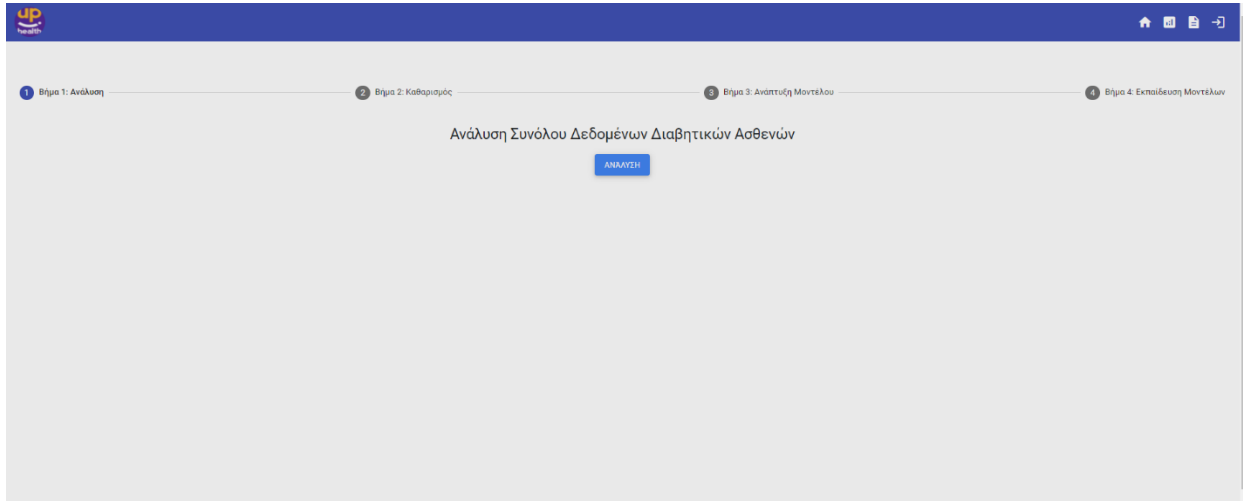
Εικόνα 36: Σελίδα Ραντεβού (Διαβήτης)

Έπειτα, ο χρήστης (Γιατρός) μπορεί να πραγματοποιήσει ανάλυση των δεδομένων των διαβητικών ασθενών. Συνεπώς, στην παρακάτω εικόνα διακρίνει κανείς 4 βήματα:

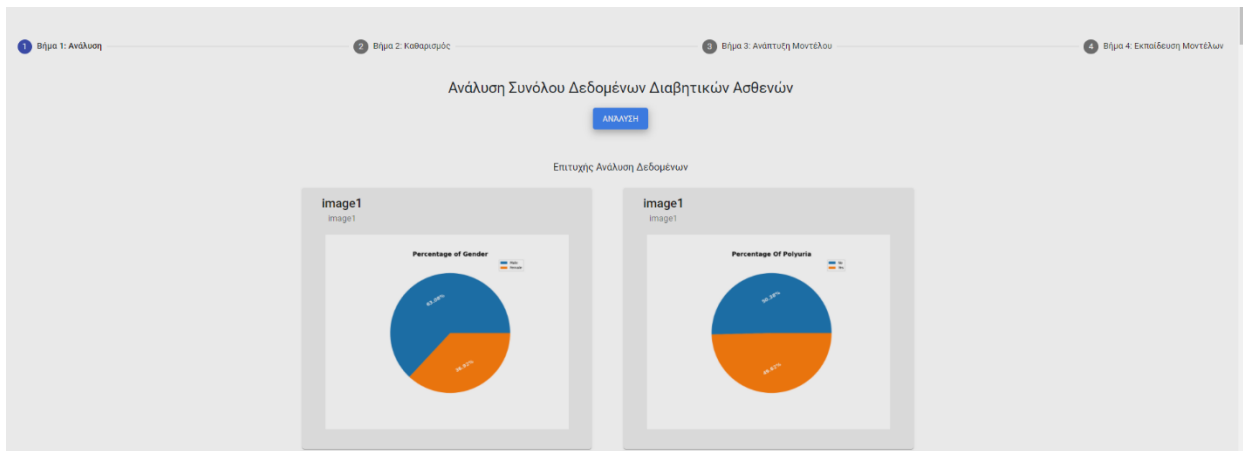
- Ανάλυση Δεδομένων (Εικόνα 37, Εικόνα 38)
- Καθαρισμός Δεδομένων (Εικόνα 39)



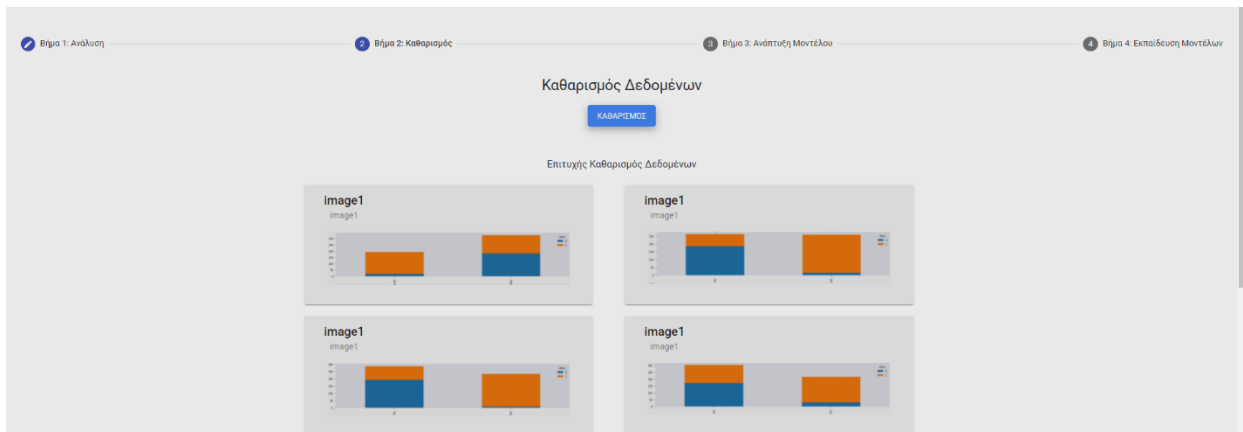
- Ανάπτυξη Βασικού Μοντέλου (Εικόνα 40)
- Εκπαίδευση Μοντέλων Μηχανικής Μάθησης (Εικόνα 41)



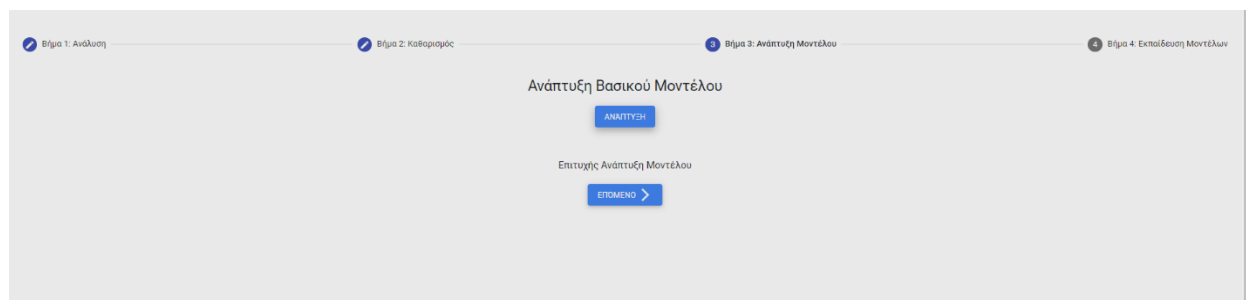
Εικόνα 37: Σελίδα Ανάλυσης Δεδομένων (βήμα 1)



Εικόνα 38: Εκτέλεση βήμα 1



Εικόνα 39: Ανάλυση Δεδομένων (βήμα 2)

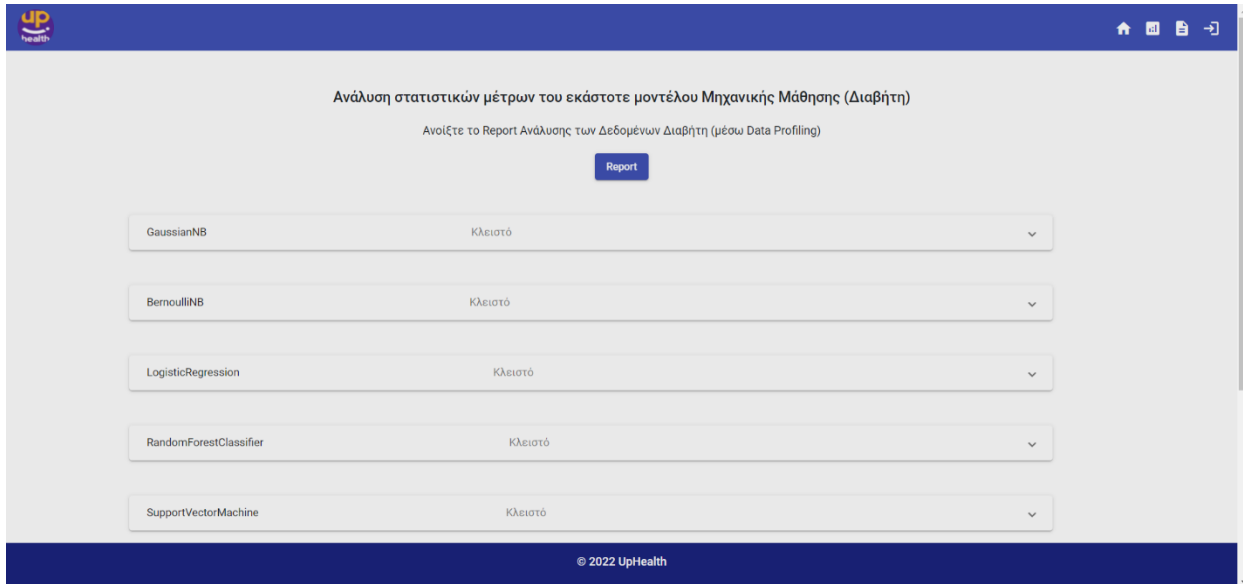


Εικόνα 40: Ανάλυση Δεδομένων (βήμα 3)



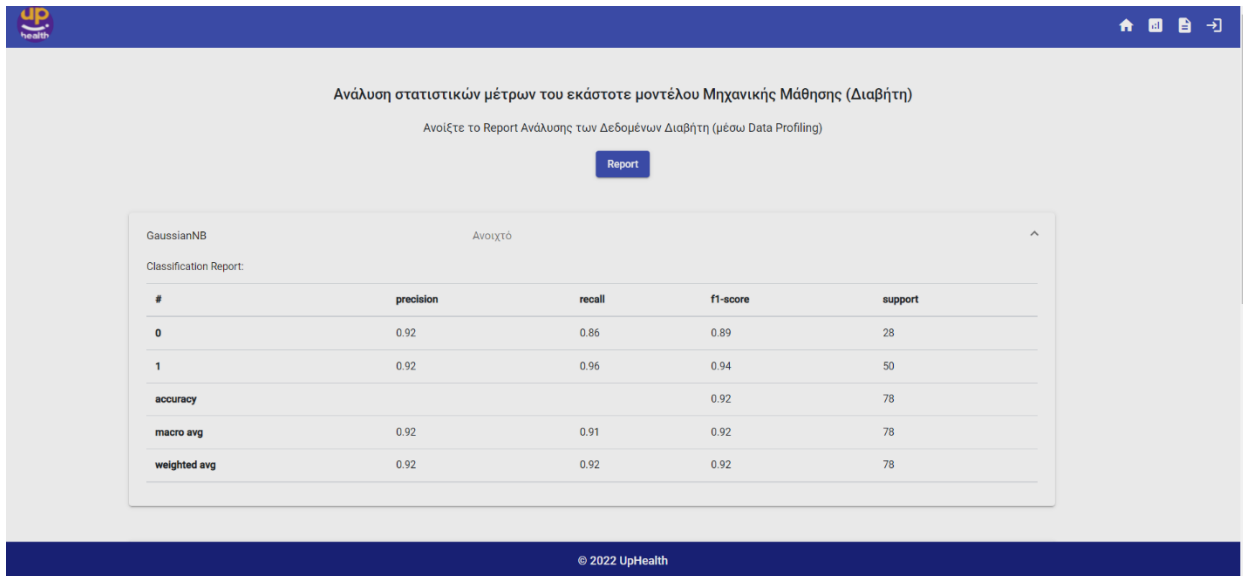
Εικόνα 41: Ανάλυση Δεδομένων (βήμα 4)

Στην ακόλουθη εικόνα, ο χρήστης μπορεί να μελετήσει τα στατιστικά μέτρα για το εκάστοτε μοντέλο Μηχανικής Μάθησης, ωστόσο μπορεί να δει παράλληλα και το γεωγρ ανάλυσης με επιπλέον στοιχεία για το κάθε χαρακτηριστικό που συνιστά τον φάκελο υγείας του διαβητικού ασθενούς (Εικόνα 42).



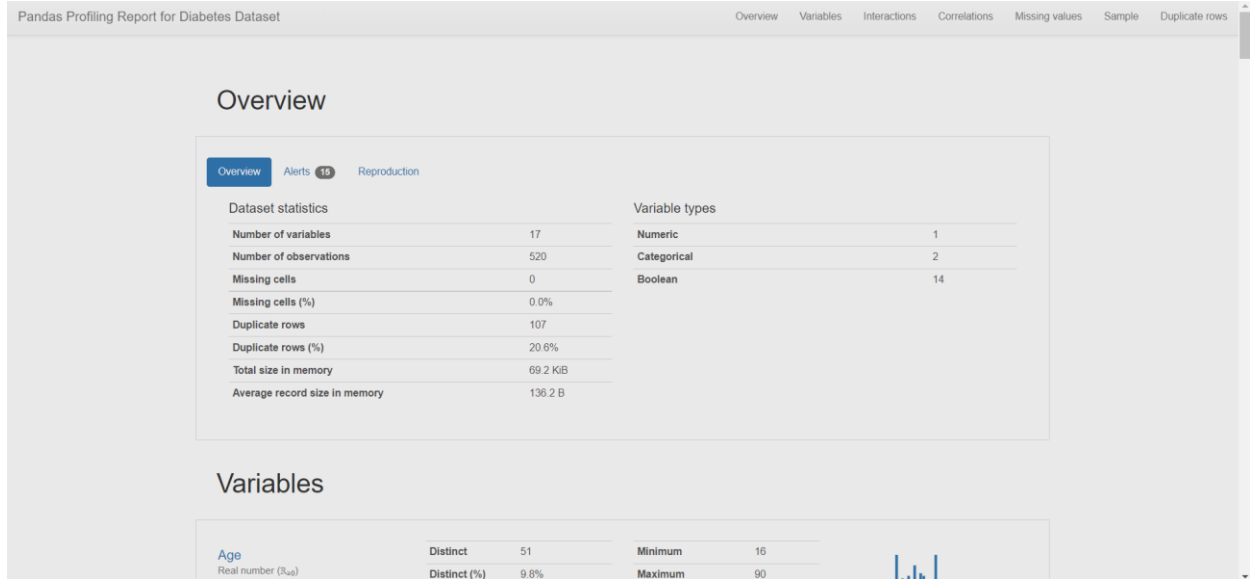
Εικόνα 42: Σελίδα Ανάλυσης Στατιστικών Στοιχείων ML Μοντέλων

Δειγματοληπτικά, επιλέχθηκε να ανοιχθεί ένα από τα μοντέλα Μηχανικής Μάθησης, παρουσιάζοντας τον πίνακα των στατιστικών μέτρων, που συνιστούν το εν παρουσίαση μοντέλο της παρακάτω εικόνας (Εικόνα 43).



Εικόνα 43: Σελίδα Ανάλυσης Στατιστικών Στοιχείων συγκεκριμένου ML Μοντέλου

Στην συνέχεια, ακολουθεί ένα στιγμιότυπο οθόνης από την σελίδα του report ανάλυσης των ασθενών με Διαβήτη (Εικόνα 44).



Εικόνα 44: Σελίδα Data Profiling Διαβήτη

Παρακάτω ακολουθούν 2 στιγμιότυπα:

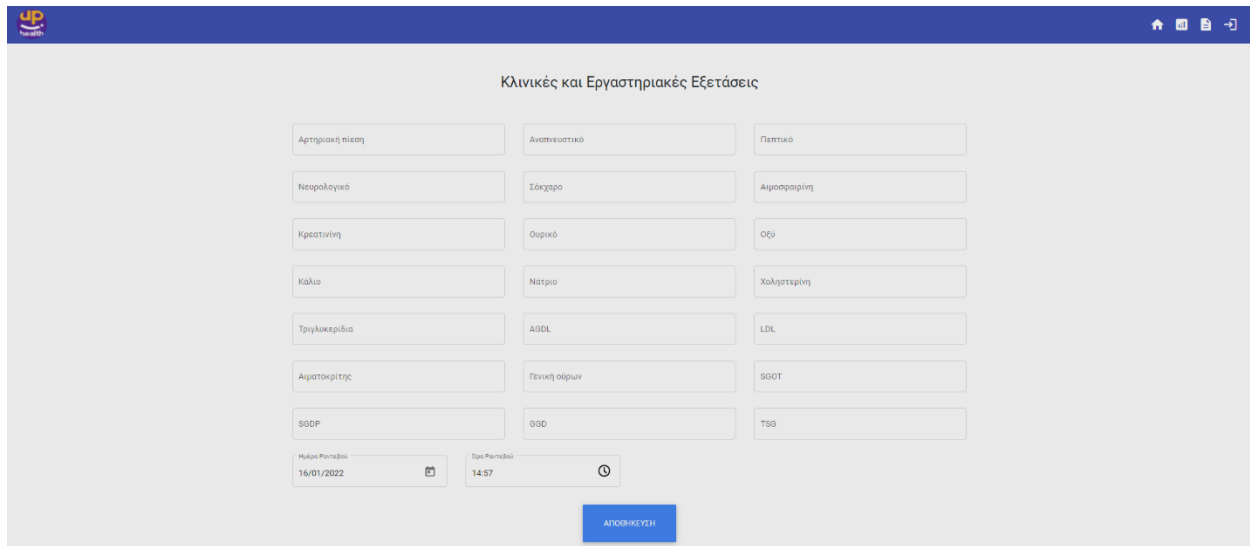
- Προσθήκη νέου διαβητικού ασθενούς (Εικόνα 45)
- Προσθήκη νέου ραντεβού για έναν συγκεκριμένο διαβητικό ασθενή (Εικόνα 46)

**Πληροφορίες Ασθενούς**

Αναγνωριστικό Ασθενή	Όνομα	Επίπλευμα	Ηλικία
Φύλο	Πολιουρία	Πολιδιψία	Ξαφρική απώλεια βάρους
Αδυναμία	Πολυφαγία	Μόλυνση ζύμης	Οπτικό θάμπωμα
Φαγούρα	Ευερέθιστο	Καθυστερημένη επούλωση	Μερική παράλυση
Μυϊκή δυσκαμψία	Αλιπεκίαση	Παχυσαρκία	

SAVE

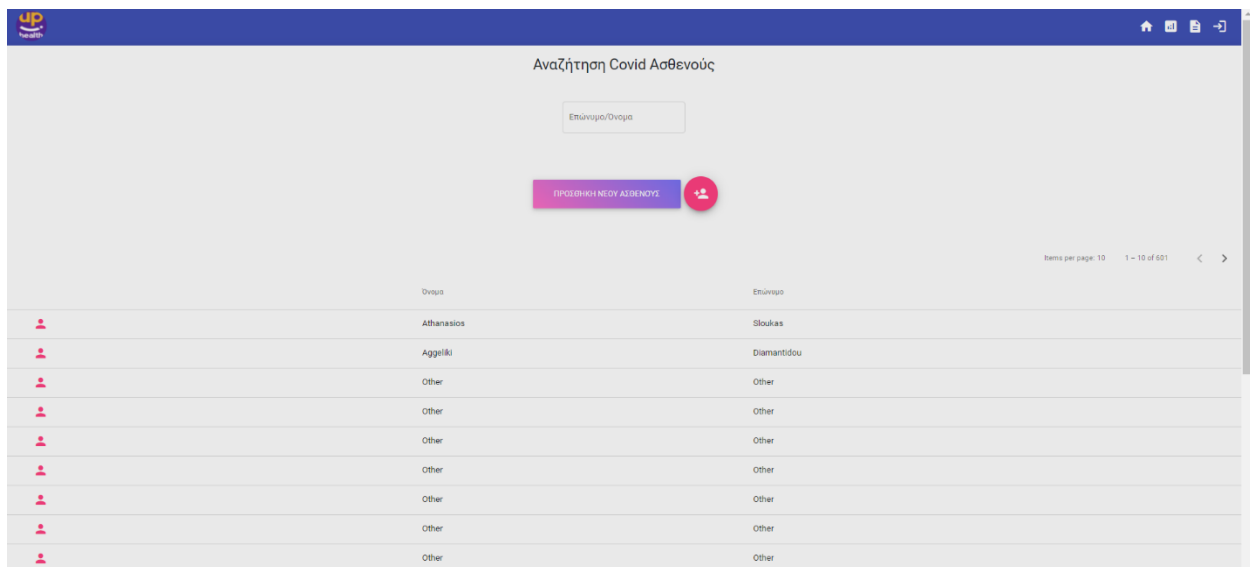
Εικόνα 45: Προσθήκη Διαβητικού Ασθενούς



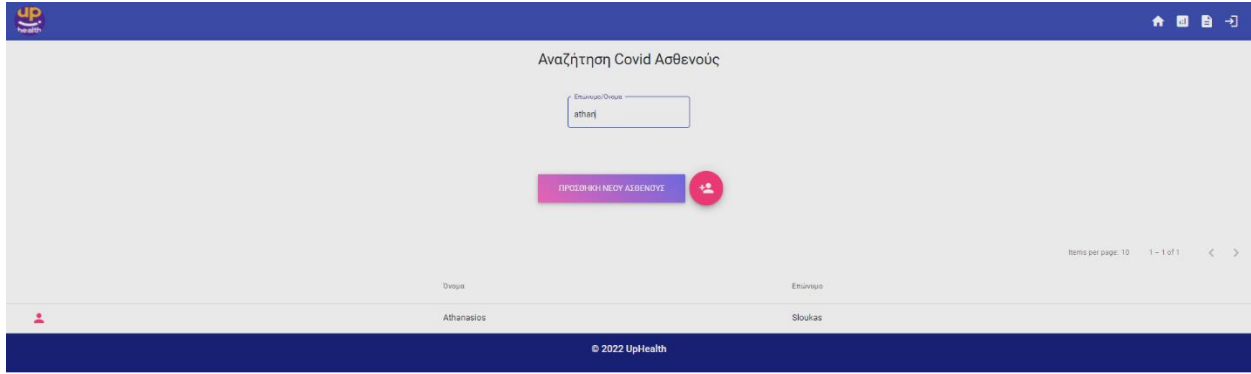
Εικόνα 46: Προσθήκη Ραντεβού Διαβήτη

Παρακάτω ακολουθούν, στιγμιότυπα οθόνης από την δεύτερη κατηγορία ασθενών, αυτών με covid-19 και έχουν ως εξής:

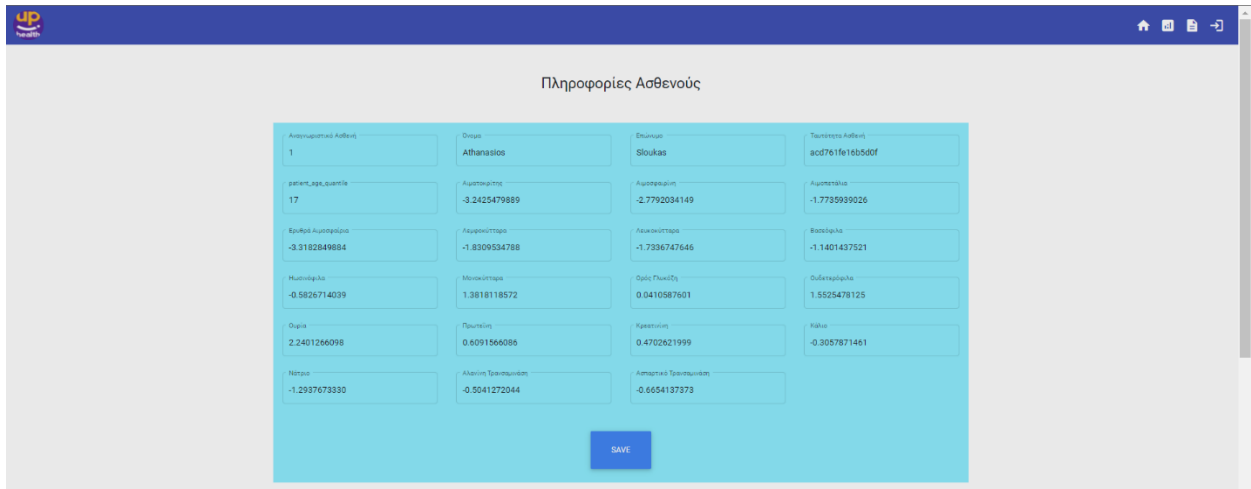
- Αναζήτηση Covid Ασθενούς (Εικόνα 47)
- Αναζήτηση Συγκεκριμένου Covid Ασθενούς (Εικόνα 48)
- Εκτεταμένος Φάκελος Υγείας Covid Ασθενούς (Εικόνα 49) Εικόνα 49: Εκτεταμένος Φάκελος Υγείας Covid Ασθενούς
- Επιλογή μοντέλου ML Πρόβλεψης και Ορισμός Ραντεβού (Εικόνα 50)
- Πρόβλεψη εισαγωγής ασθενή στο νοσοκομείο βάσει συγκεκριμένου ML **μοντέλου** (Εικόνα 51)
- Στοιχεία Ραντεβού Covid Ασθενούς (Εικόνα 52)



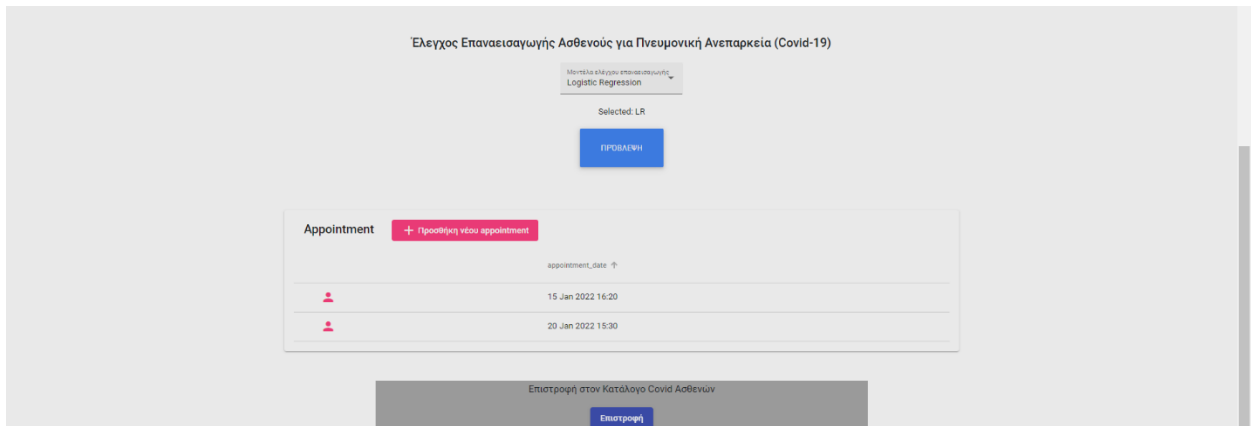
Εικόνα 47: Αναζήτηση Covid Ασθενούς



Εικόνα 48: Αναζήτηση Συγκεκριμένου Covid Ασθενούς



Εικόνα 49: Εκτεταμένος Φάκελος Υγείας Covid Ασθενούς



Εικόνα 50: Επιλογή μοντέλου ML Πρόβλεψης και Ορισμός Ραντεβού

Εικόνα 51: Πρόβλεψη εισαγωγής ασθενή στο νοσοκομείο βάσει συγκεκριμένου ML μοντέλου

Εικόνα 52: Στοιχεία Ραντεβού Covid Ασθενούς

Παρακάτω ακολουθούν 2 στιγμιότυπα:

- Προσθήκη νέου Covid ασθενούς (Εικόνα 53)
- Προσθήκη νέου ραντεβού για έναν συγκεκριμένο Covid ασθενή (Εικόνα 54)

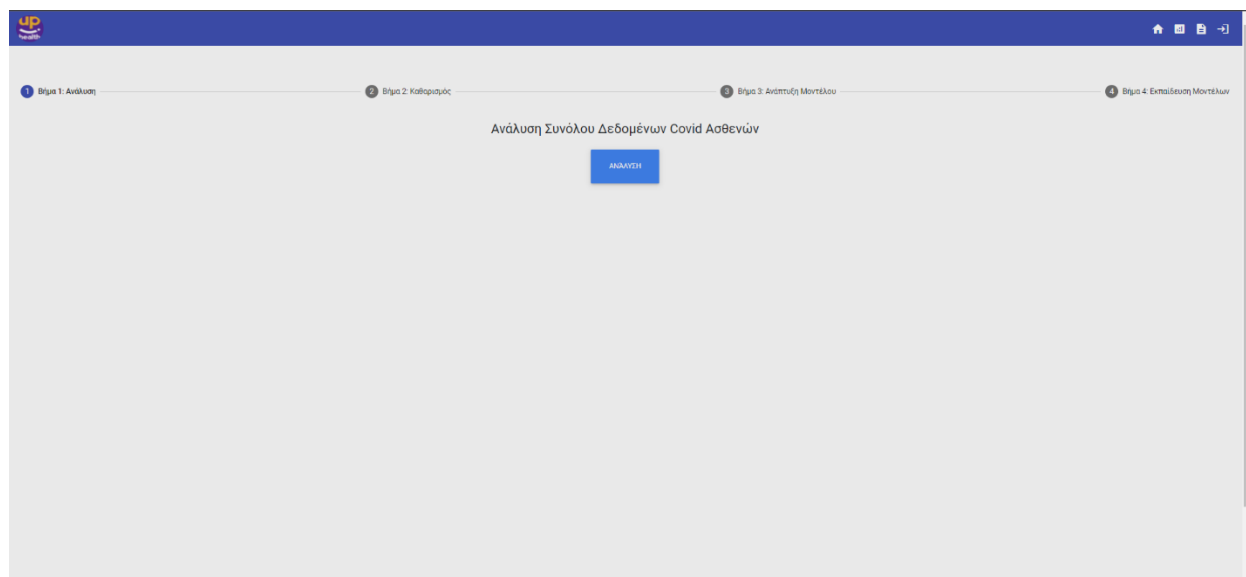
Εικόνα 53: Προσθήκη Covid Ασθενούς

Εικόνα 54: Προσθήκη Ραντεβού Covid Ασθενούς

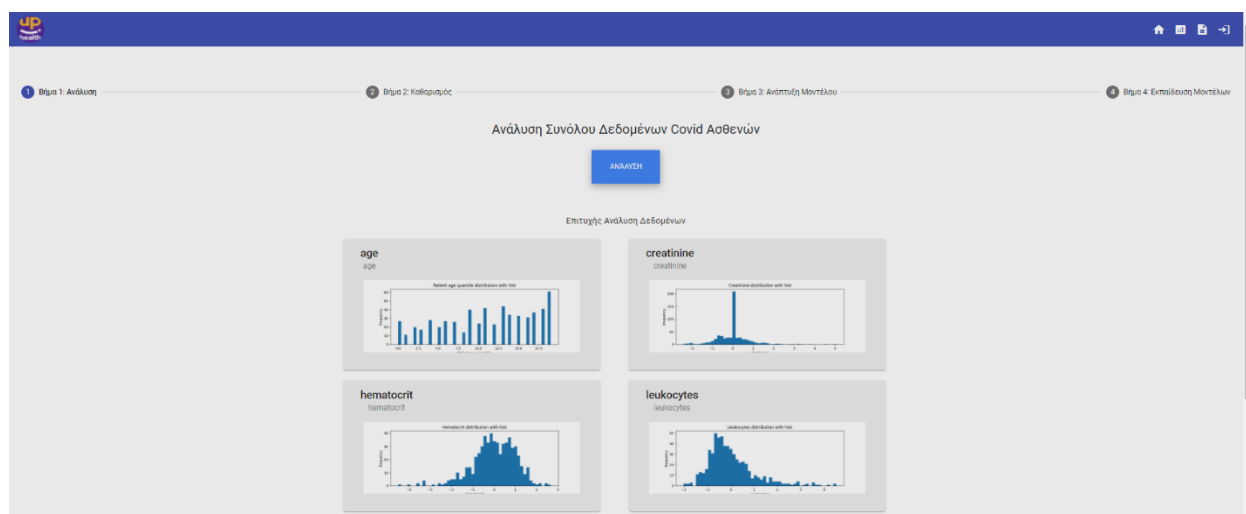
Στα επόμενα στιγμιότυπα οθόνης ακολουθεί η ανάλυση του συνόλου δεδομένων:

- Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 1) (Εικόνα 55, Εικόνα 56)
- Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 2) (Εικόνα 57)
- Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 3) (Εικόνα 58)
- Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 4) (Εικόνα 59)

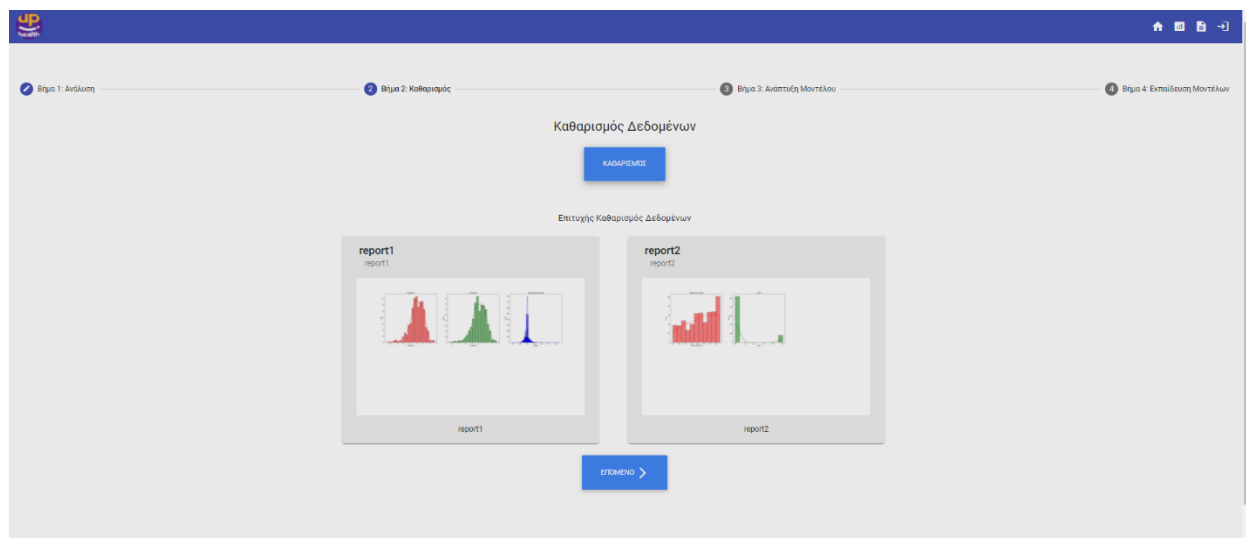




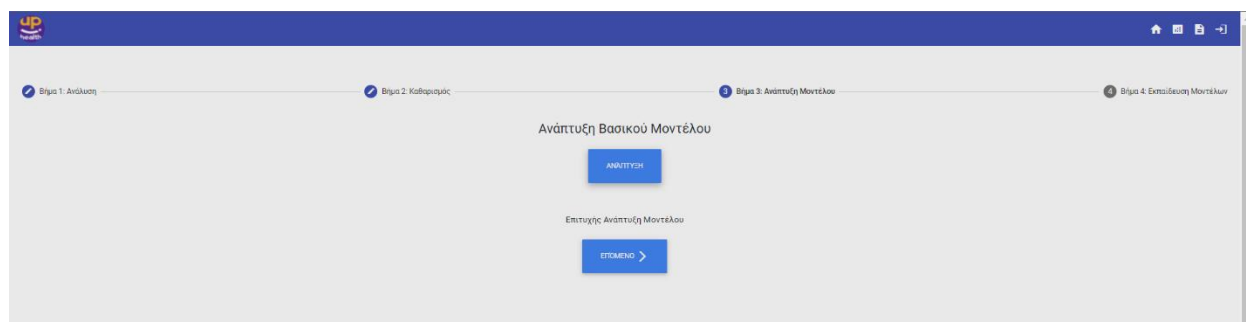
Εικόνα 55: Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 1)



Εικόνα 56: Σελίδα Ανάλυσης Δεδομένων Covid (εκτέλεση βήμα 1)



Εικόνα 57: Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 2)



Εικόνα 58: Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 3)



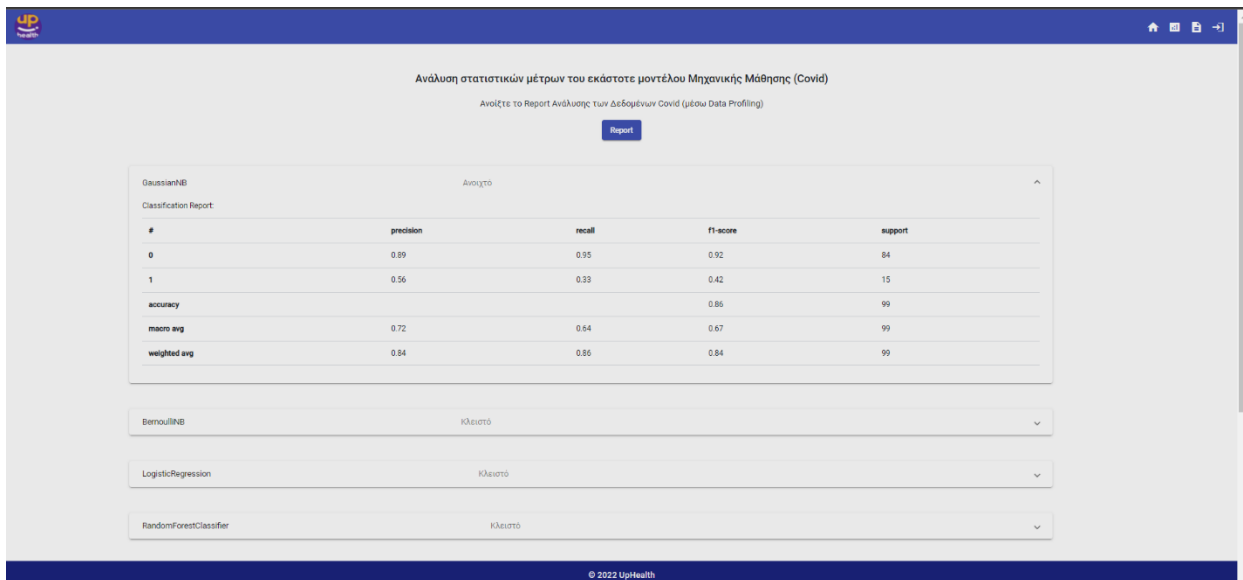
Εικόνα 59: Σελίδα Ανάλυσης Δεδομένων Covid (βήμα 4)

Στην εικόνα που ακολουθεί, παρουσιάζεται η σελίδα της ανάλυσης στατιστικών στοιχείων (Εικόνα 60).



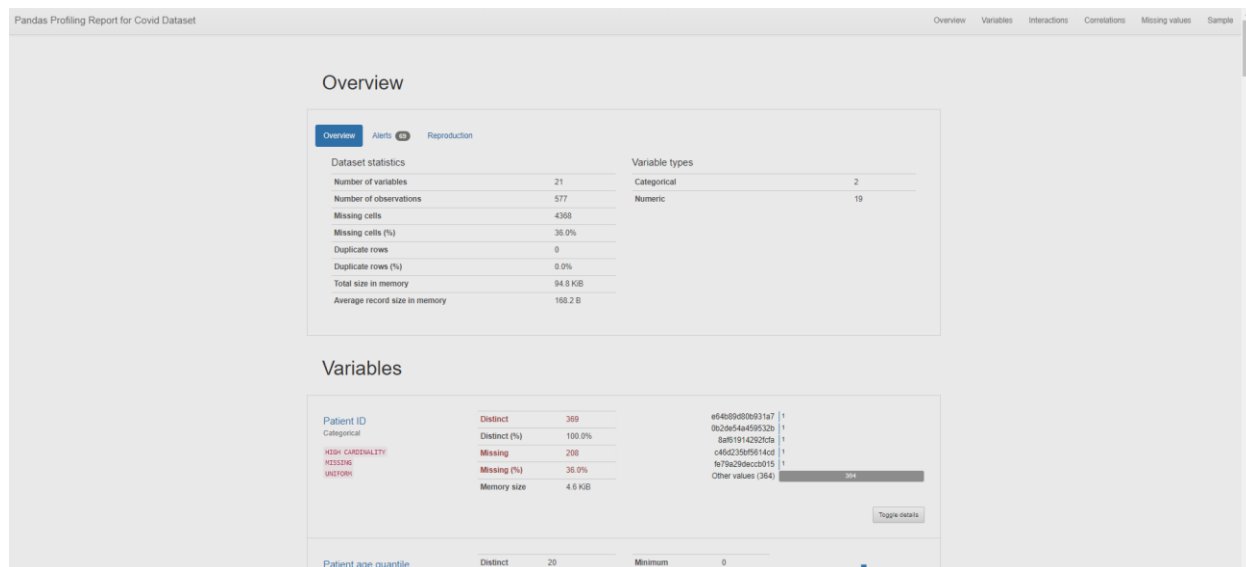
Εικόνα 60: Σελίδα Ανάλυσης Στατιστικών Στοιχείων ML

Στην συνέχεια, το ακόλουθο στιγμιότυπο οθόνης δείχνει την άνωθεν σελίδα, με την επιπλέον πληροφορία, ότι έχει επιλεγεί ένα εκ των μοντέλων μηχανικής μάθησης (Εικόνα 61).



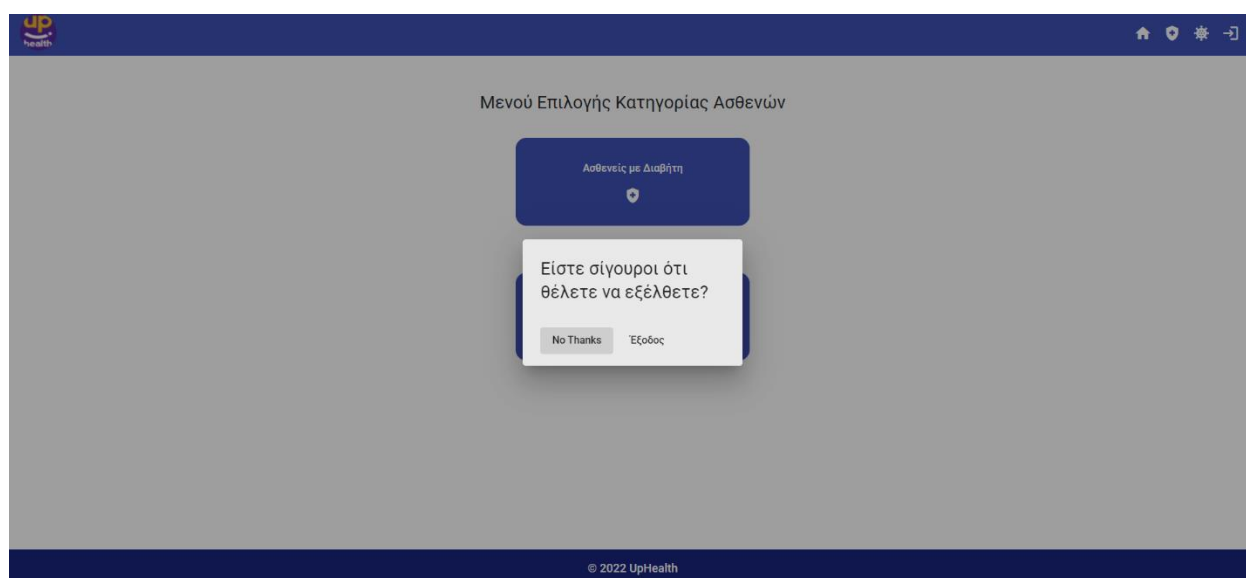
Εικόνα 61: Σελίδα Ανάλυσης Στατιστικών Στοιχείων συγκεκριμένου ML

Επίσης, παρακάτω ακολουθεί η εικόνα της σελίδας data profiling για το σύνολο δεδομένων covid (Εικόνα 62).



Εικόνα 62: Σελίδα Data Profiling Covid

Τέλος, προτού, ο χρήστης (γιατρός) εξέλθει του συστήματος, εμφανίζεται ένα «pop-up» μήνυμα που ρωτάει, αν ο χρήστης επιθυμεί την έξοδο του από το πληροφοριακό σύστημα (Εικόνα 63).



Εικόνα 63: Έξοδος Συστήματος "urhealth"

#### 4.4.1. Unit Testing

Τα στοιχεία (components) είναι μια από τις βασικές έννοιες των Angular εφαρμογών. Επομένως, εάν θέλετε να βεβαιωθείτε ότι η εφαρμογή σας είναι όσο το δυνατόν υψηλής ποιότητας, πρέπει να δώσετε λίγο περισσότερο κόπο σε αυτά τα βασικά δομικά στοιχεία (Εικόνα 64). Όσον αφορά την ποιότητα του λογισμικού, το unit testing είναι πολύ βασικό, οπότε στην παρακάτω εικόνα παρατηρείται πως μπορεί να διασφαλίσει κάποιος ότι τα εξαρτήματά λειτουργούν τέλεια (Δημιουργία dynamic form fields). Επίσης, και σχετικά με την επικοινωνία του μοντέλου «Patient»


του προτύπου HL7 FHIR ελέγχεται με την χρήση της μηχανισμού unit testing ελέγχοντας τα δεδομένα 2 ασθενών, πραγματοποιείται η διασύνδεση μεταξύ ενός συγκεκριμένου πεδίου και πιο συγκεκριμένα ο πόρος προέλευσης του μοντέλου, το οποίο περνάει επιτυχώς τον έλεγχο, διότι η πληροφορία από 2 διαφορετικούς ασθενείς είναι όμοια, οπότε τα δεδομένα επικοινωνούν ως προς το συγκεκριμένο πεδίο χωρίς κάποιο πρόβλημα διαλειτουργικής επικοινωνίας (Εικόνα 65).

```
dynamic-form-field.component.spec.ts
1  import { ComponentFixture, TestBed } from '@angular/core/testing';
2
3  import { DynamicFormFieldComponent } from './dynamic-form-field.component';
4
5  describe('DynamicFormFieldComponent', () => {
6    let component: DynamicFormFieldComponent;
7    let fixture: ComponentFixture<DynamicFormFieldComponent>;
8
9    beforeEach(async () => {
10     await TestBed.configureTestingModule({
11       declarations: [ DynamicFormFieldComponent ]
12     })
13     .compileComponents();
14   });
15
16   beforeEach(() => {
17     fixture = TestBed.createComponent(DynamicFormFieldComponent);
18     component = fixture.componentInstance;
19     fixture.detectChanges();
20   });
21
22   it('should create', () => {
23     expect(component).toBeTruthy();
24   });
25 });
26
```

Εικόνα 64: Unit Testing Angular

SMART FHIR Client

Component of Unit Testing into a patient model following prototype of  
 HL7 FHIR



```
[ ] from unittest import TestCase
    from unittest import main

import json

class Test(TestCase):
    def test_1(self):
        with open("dict1.json") as jsonFile: # Replace "path/to/file.json" with the path to your JSON file
            dict1 = json.load(jsonFile)

            with open("dict2.json") as jsonFile:
                dict2 = json.load(jsonFile)

            # resourceType
            self.assertEqual(dict1['resourceType'], dict2['resourceType'], "Patient")
            print("\nUnit Testing #1")
            print("\nresourceType")
            print("\nresourceType (Patient #1)':" + dict1['resourceType'] + "," + "resourceType (Patient #2)':" + dict2['resourceType'] + "," + "HL7 FHIR'")

if __name__ == '__main__':
    main(argv=['first-arg-is-ignored'], exit=False)
```

```
Unit Testing #1
resourceType
'resourceType (Patient #1)':Patient,'resourceType (Patient #2)':Patient,'HL7 FHIR'
-----
Ran 1 test in 0.009s
OK
```

Εικόνα 65: Unit Testing HL7 FHIR

## Πλεονεκτήματα του Unit Testing

- Βελτίωση του σχεδιασμού των υλοποιήσεων

Όταν κάποιος ξεκινάει να κωδικοποιεί ένα χαρακτηριστικό χωρίς να το σκεφτεί πολύ για το σχεδιασμό είναι ένα πολύ κοινό λάθος μεταξύ των προγραμματιστών. Η χρήση δοκιμών μονάδας θα αναγκάσει κάποιον να σκεφτεί και να ξανασκεφτεί το σχέδιο, και εάν χρησιμοποιηθεί το Test Driven Development (TDD) ο αντίκτυπος είναι ακόμη μεγαλύτερος.

- Ανακατασκευή υλοποίησης

Εφόσον έχουν γίνει ήδη δοκιμές που διασφαλίζουν ότι όλα λειτουργούν όπως αναμένεται, μπορεί κανείς εύκολα να προσθέσει αλλαγές στον κώδικα με τη βεβαιότητα ότι δεν πρόσθεσε σφάλματα.

- Προσθήκη νέων δυνατοτήτων χωρίς περαιτέρω συνέπειες

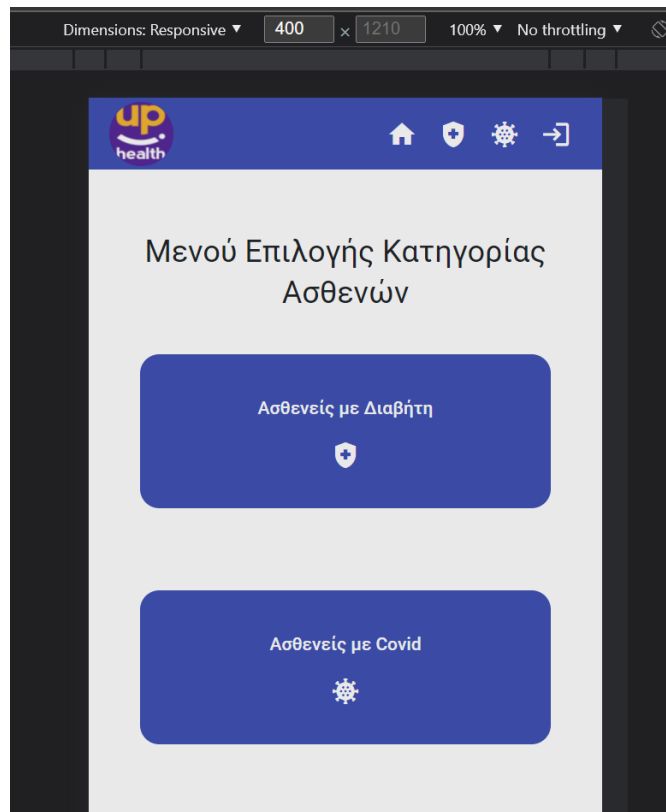
Όταν προστίθεται μια νέα δυνατότητα, μπορεί κάποιος να εκτελέσει τις δοκιμές για να βεβαιωθεί ότι δεν έχει σπάσει οποιοδήποτε άλλο τμήμα της εφαρμογής.

#### 4.4.2. Responsive Σχεδιασμός

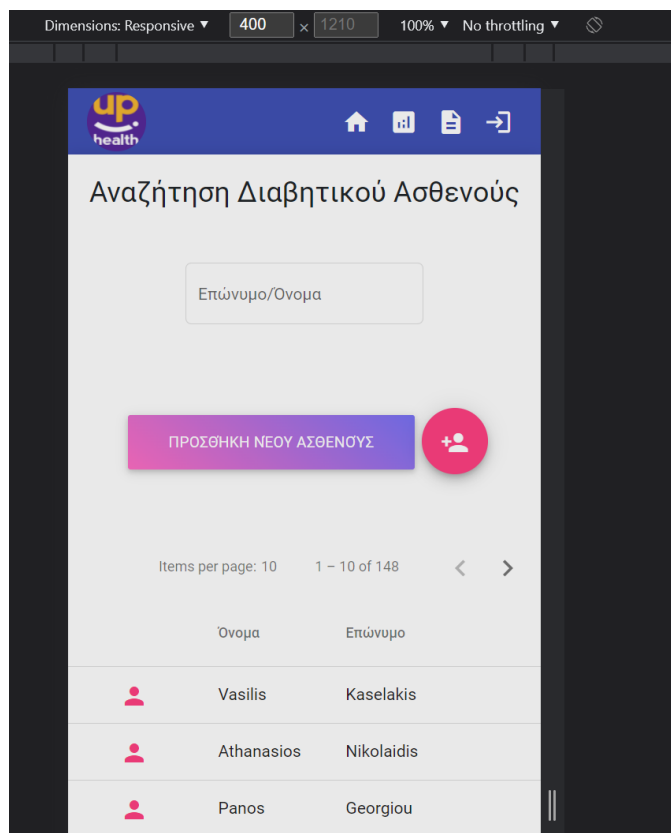
Το Responsive Web Design είναι η προσέγγιση που προτείνει ότι ο σχεδιασμός και η ανάπτυξη πρέπει να ανταποκρίνονται στη συμπεριφορά και το περιβάλλον του χρήστη με βάση το μέγεθος της οθόνης, την πλατφόρμα και τον προσανατολισμό (Εικόνα 66, Εικόνα 67, Εικόνα 68, Εικόνα 69).

Η πρακτική αποτελείται από έναν συνδυασμό ευέλικτων πλεγμάτων και διατάξεων, εικόνων και μια έξυπνη χρήση ερωτημάτων μέσω CSS. Καθώς ο χρήστης αλλάζει από το φορητό υπολογιστή του στο iPad, ο ιστότοπος θα πρέπει να μεταβαίνει αυτόματα για να προσαρμόζεται σε ανάλυση, μέγεθος εικόνας και ικανότητες προσαρμογής των ενεργειών που συνιστούν την διεπαφή χρήστη.

Παρακάτω ακολουθούν ενδεικτικά μερικά στιγμιότυπα οθόνης από τον responsive σχεδιασμό που υποστηρίζει το υλοποιημένο σύστημα σε όλη την έκτασή του. Για ευνότητους λόγους, συμπεριλαμβάνονται ενδεικτικά κάποιες σελίδες, μέσω των οποίων παρουσιάζεται ο responsive σχεδιασμός.

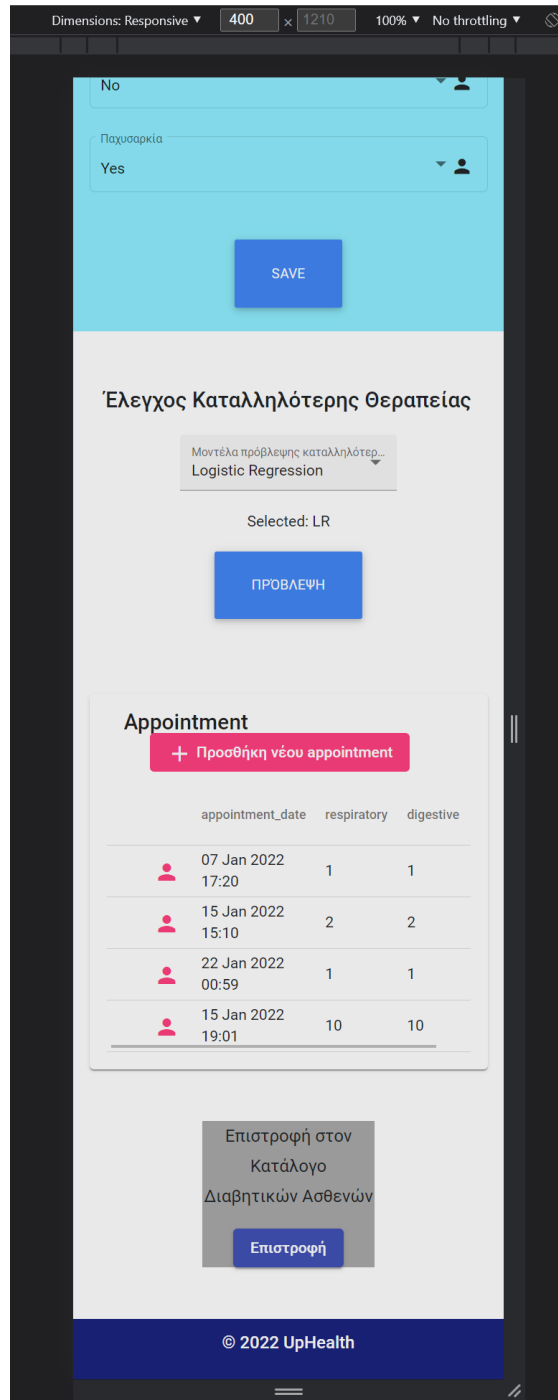


Εικόνα 66: Responsive Μενού Επιλογής Κατηγορίας Ασθενών

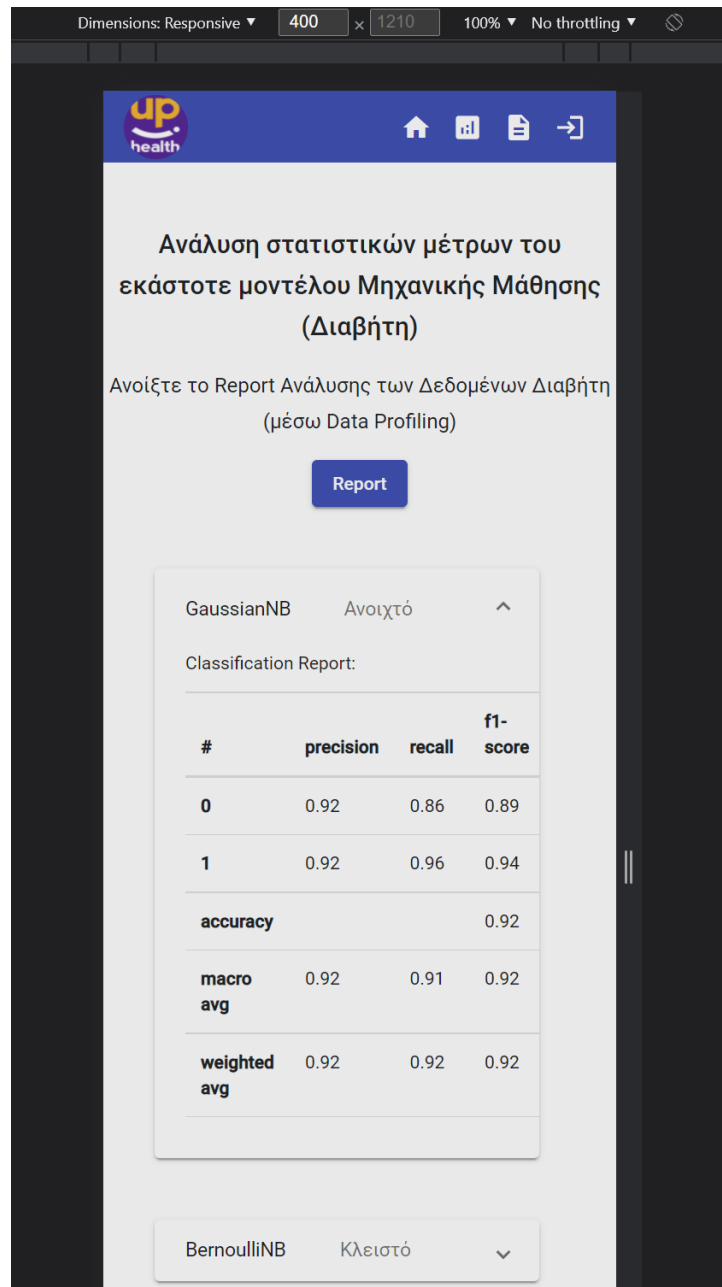


Εικόνα 67: Responsive Αναζήτηση Διαβητικού Ασθενούς





Εικόνα 68: Responsive Φάκελος Διαβητικού Ασθενούς



Εικόνα 69: Responsive Σελίδα Στατιστικών Μέτρων

## Συμπεράσματα

Η ανάγκη ανάπτυξης συστημάτων και υπηρεσιών ηλεκτρονικής υγείας καθίσταται επιτακτική καθώς οι σύγχρονες απαιτήσεις στον τομέα της υγειονομικής περίθαλψης και γενικότερα στον τομέα της υγείας διαρκώς μεταβάλλονται και εξελίσσονται τόσο σε υποδομή όσο και σε «γνώση».

Η ανάπτυξη της παραπάνω εφαρμογής παρέχει την δυνατότητα της πρόβλεψης και της εύρεση της καταλληλότερης θεραπείας στην εμφάνιση νόσων διαβήτη και Covid-19. Η βελτίωση και η εξέλιξη των μοντέλων μηχανικής μάθησης που χρησιμοποιήθηκαν τόσο μέσω τροποποιήσεων των παραμέτρων τους όσο και μέσω της συνεχούς εκπαίδευσης τους από πληθώρα εγγραφών δεδομένων θα οδηγήσει σε σύντομο χρονικό διάστημα στην επίτευξη καλύτερων και πιο αποτελεσματικών έγκαιρων διαγνώσεων και θεραπειών.

Συνεπώς, η μηχανική μάθηση και η πληροφορική μπορούν να αναβαθμίσουν τον τομέα της υγειονομικής περίθαλψης. Η εξειδικευμένη μελέτη κάθε ασθένειας οδήγησε τους σύγχρονους ερευνητές στον συνδυασμό όλων των υπαρχόντων δεδομένων (κλινικών και εργαστηριακών) με προσδοκώμενο αποτέλεσμα μια αποτελεσματική θεραπεία. Αν σε αυτό προστεθεί η μοντελοποίηση των αλγορίθμων μηχανικής μάθησης για κάθε ασθένεια, τότε θα υπάρξει η δυνατότητα της πλέον αποτελεσματική θεραπεία.

Επίσης, η εξοικονόμηση πόρων (εξοπλισμός ιατρικών εργαστηρίων, μελέτες, θεραπείες, δοκιμές) η οποία επιτυγχάνεται με χρήση των παραπάνω μοντέλων δεν θεωρείται καθόλου αμελητέα. Περαιτέρω, η ανακάλυψη μέσω των αλγορίθμων νέων δεδομένων που θα συμβάλλουν στην ταχύτερη και αποτελεσματικότερη διάγνωση της νόσου, εξασφαλίζει την άριστα παρεχόμενη ποιότητα υπηρεσιών στον ασθενή (αποφυγή χρήσης λανθασμένων φαρμάκων, νοσηλεία, περιττές εξετάσεις).

Συμπερασματικά, αν και οι περισσότεροι συμφωνούν ότι οι μηχανές θα αργήσουν να αντικαταστήσουν τους γιατρούς στην κλινική πρακτική, στον τομέα της διάγνωσης ασθενειών σαφώς και θα γίνουν αρωγοί τους. Και όπως ήδη αποδεικνύεται, θα τους βοηθήσουν να γίνουν καλύτεροι στα καθήκοντά τους, προς όφελος των ασθενών.

## Μελλοντικές Επεκτάσεις

Ο χώρος της υγειονομικής περίθαλψης εξελίσσεται ταχέως, ειδικά μετά την εμφάνιση του νέου ιού Covid-19. Οπότε, μέσω της παρούσας έρευνας δύναται σε κάποιον ειδικό στον χώρο της υγείας να διαθέτει ένα έξυπνο εργαλείο διαχείρισης των ασθενών του. Ωστόσο, η συγκεκριμένη υλοποίηση διαθέτει αρκετούς μηχανισμούς για την αυτοματοποίηση μερικών διαδικασιών με την χρήση της μηχανικής μάθησης και της διαλειτουργικότητας των δεδομένων.

Ως επέκταση της εφαρμογής που αναπτύχθηκε στα πλαίσια της διπλωματικής εργασίας, αποτελεί η ενοποίηση περισσότερων πηγών δεδομένων και χρήση ακόμα πιο έξυπνων μηχανισμών μηχανικής μάθησης με περαιτέρω εξειδίκευση σε παραμέτρους των αλγορίθμων για την επίτευξη ακόμα καλύτερων αποτελεσμάτων σε συνάρτηση με την χρήση περισσότερων κανόνων διαλειτουργικότητας των δεδομένων. Ο συνδυασμός μηχανικής μάθησης και διαλειτουργικότητας των δεδομένων θα επιτύχει την προς φιλική παράδοση συστημάτων εύκολα και κατανοήσιμα από τους ειδικούς στον χώρο της υγείας.

Κλείνοντας, η μηχανική μάθηση έχει ήδη αρχίσει να εκπληρώνει τις δυνατότητές της για την υγειονομική περίθαλψη, από τη διευκόλυνση της αποτελεσματικότερης έρευνας και ανάπτυξης φαρμάκων μέχρι τη φροντίδα ασθενών και τις διοικητικές διαδικασίες. Τα επόμενα χρόνια, είναι πιθανή η ευρεία υιοθέτηση της μηχανικής μάθησης και άλλων τεχνολογιών τεχνητής νοημοσύνης. Αντί να αντικαθιστούν πλήρως τους κλινικούς γιατρούς, αυτές οι τεχνολογίες είναι πιθανό να συμπληρώνουν και να ενισχύουν τους ρόλους τους. Τα μακροπρόθεσμα αποτελέσματα θα μπορούσαν να περιλαμβάνουν καλύτερη ποιότητα περίθαλψης και ένα πιο αποτελεσματικό και οικονομικά αποδοτικό σύστημα υγειονομικής περίθαλψης.

Από την πλευρά της διαλειτουργικότητας των δεδομένων, κάθε οργανισμός υγειονομικής περίθαλψης σε όλον τον κόσμο θα πρέπει να εφαρμόζει το πρότυπο FHIR. Καθώς, η υιοθέτηση του FHIR εξαπλώνεται, πλησιάζουμε σε ένα μέλλον όπου ένα πρότυπο γίνεται πράγματι ένα ακολουθούμενο πρότυπο και η πραγματική διαλειτουργικότητα στην υγειονομική περίθαλψη μπορεί να επιτευχθεί σε διεθνή κλίμακα.

## Βιβλιογραφία

- [1] Brunton, S. L., Noack, B. R., & Koumoutsakos, P. (2020). Machine learning for fluid mechanics. *Annual review of fluid mechanics*, 52, 477-508.
- [2] Expert.ai, <https://www.expert.ai/blog/machine-learning-definition/>
- [3] Carecloud, <https://www.carecloud.com/continuum/why-interoperability-is-such-a-challenge/>
- [4] Deshmukh, S. G., & Haleem, A. (2020). Framework for manufacturing in post-COVID-19 world order: An Indian perspective. *International Journal of Global Business and Competitiveness*, 15(1), 49-60.
- [5] Gardner, R. L., Cooper, E., Haskell, J., Harris, D. A., Poplau, S., Kroth, P. J., & Linzer, M. (2019). Physician stress and burnout: the impact of health information technology. *Journal of the American Medical Informatics Association*, 26(2), 106-114.
- [6] Vora, J., Nayyar, A., Tanwar, S., Tyagi, S., Kumar, N., Obaidat, M. S., & Rodrigues, J. J. (2018, December). BHEEM: A blockchain-based framework for securing electronic health records. In *2018 IEEE Globecom Workshops (GC Wkshps)* (pp. 1-6). IEEE.
- [7] Paul, S., Riffat, M., Yasir, A., Mahim, M. N., Sharnali, B. Y., Naheen, I. T., ... & Kulkarni, A. (2021). Industry 4.0 applications for medical/healthcare services. *Journal of Sensor and Actuator Networks*, 10(3), 43.
- [8] Mansour, R. F., El Amraoui, A., Nouaouri, I., Díaz, V. G., Gupta, D., & Kumar, S. (2021). Artificial intelligence and internet of things enabled disease diagnosis model for smart healthcare systems. *IEEE Access*, 9, 45137-45146.
- [9] Insidebigdata, <https://insidebigdata.com/2022/06/05/bad-data-costs-u-s-companies-trillions-how-data-quality-audits-can-help/>
- [10] Rostamzadeh, N., Abdullah, S. S., & Sedig, K. (2021, February). Visual analytics for electronic health records: a review. In *Informatics* (Vol. 8, No. 1, p. 12). MDPI.
- [11] Redcap, <https://www.project-redcap.org/>
- [12] MDAnalyze, <https://www.mdanalysis.org/>
- [13] Mansour, R. F., El Amraoui, A., Nouaouri, I., Díaz, V. G., Gupta, D., & Kumar, S. (2021). Artificial intelligence and internet of things enabled disease diagnosis model for smart healthcare systems. *IEEE Access*, 9, 45137-45146.
- [14] Python, <https://www.python.org/>
- [15] Rostamzadeh, N., Abdullah, S. S., & Sedig, K. (2021, February). Visual analytics for electronic health records: a review. In *Informatics* (Vol. 8, No. 1, p. 12). MDPI.
- [16] Harmonyhit, <https://www.harmonyhit.com/health-data-volumes-skyrocket-legacy-data-archives-rise-hie/>
- [17] Guyon, I., Sun-Hosoya, L., Boullé, M., Escalante, H. J., Escalera, S., Liu, Z., ... & Viegas, E. (2019). Analysis of the automl challenge series. *Automated Machine Learning*, 177.
- [18] Saberi-Movahed, F., Rostami, M., Berahmand, K., Karami, S., Tiwari, P., Oussalah, M., & Band, S. S. (2022). Dual regularized unsupervised feature selection based on matrix factorization and minimum redundancy with application in gene selection. *Knowledge-Based Systems*, 256, 109884.
- [19] Baradaran Rezaei, H., Amjadian, A., Sebt, M. V., Askari, R., & Gharaei, A. (2022). An ensemble method of the machine learning to prognosticate the gastric cancer. *Annals of Operations Research*, 1-42.
- [20] Sharma, G., & Patil, G. R. (2021). Public transit accessibility approach to understand the equity for public healthcare services: A case study of Greater Mumbai. *Journal of Transport Geography*, 94, 103123.
- [21] Anvio, <https://anvio.org/>
- [22] Decisionflow, <https://www.decisionflowgroup.com/>
- [23] Capterra, <https://www.capterra.com/data-visualization-software/compare/176573-81277/Google-Charts-vs-TIBCO-Spotfire>
- [24] Progressive, <https://www.progressive.com/resources/insights/>

- [25] Raval, M., Sivashanmugam, P., Pham, V., Gohel, H., Kaushik, A., & Wan, Y. (2021). Automated predictive analytics tool for rainfall forecasting. *Scientific Reports*, 11(1), 1-13.
- [26] Rashidi, H. H., Tran, N., Albahra, S., & Dang, L. T. (2021). Machine learning in health care and laboratory medicine: General overview of supervised learning and Auto-ML. *International Journal of Laboratory Hematology*, 43, 15-22.
- [27] Geetha, G., & Prasad, K. M. (2020). Prediction of diabetics using machine learning. *IJRTE (International Journal of Recent Technology and Engineering)*, 8(5), 1119-1124.
- [28] Shetty, D., Rit, K., Shaikh, S., & Patil, N. (2017, March). Diabetes disease prediction using data mining. In 2017 international conference on innovations in information, embedded and communication systems (ICIIECS) (pp. 1-5). IEEE.
- [29] Sarwar, M. A., Kamal, N., Hamid, W., & Shah, M. A. (2018, September). Prediction of diabetes using machine learning algorithms in healthcare. In 2018 24th international conference on automation and computing (ICAC) (pp. 1-6). IEEE.
- [30] Mujumdar, A., & Vaidehi, V. (2019). Diabetes prediction using machine learning algorithms. *Procedia Computer Science*, 165, 292-299.
- [31] Sisodia, D., & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia computer science*, 132, 1578-1585.
- [32] Jayashree, J., & Kumar, S. A. (2018). Linear discriminant analysis based genetic algorithm with generalized regression neural network—a hybrid expert system for diagnosis of diabetes. *Programming and Computer Software*, 44(6), 417-427.
- [33] Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval* (Vol. 39, pp. 234-265). Cambridge: Cambridge University Press.
- [34] Eswari, T., Sampath, P., & Lavanya, S. J. P. C. S. (2015). Predictive methodology for diabetic data analysis in big data. *Procedia Computer Science*, 50, 203-208.
- [35] Nnamoko, N., Hussain, A., & England, D. (2018, July). Predicting diabetes onset: an ensemble supervised learning approach. In 2018 IEEE Congress on Evolutionary Computation (CEC) (pp. 1-7). IEEE.
- [36] Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B. W., Aslam, W., & Choi, G. S. (2020). COVID-19 future forecasting using supervised machine learning models. *IEEE access*, 8, 101489-101499.
- [37] Bastos, S. B., & Cajueiro, D. O. (2020). Modeling and forecasting the Covid-19 pandemic in Brazil. *arXiv preprint arXiv:2003.14288*
- [38] Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., Rouf, N., & Mohi Ud Din, M. (2020). Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology*, 12(3), 731-739.
- [39] Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., & Priesemann, V. (2020). Inferring COVID-19 spreading rates and potential change points for case number forecasts. *arXiv preprint arXiv:2004.01105*
- [40] M. Maniruzzaman, M. J. Rahman, M. Al-MehediHasan, H. S. Suri, M. M. Abedin, A. El-Baz, et al., "Accurate diabetes risk stratification using machine learning: Role of missing value and outliers", *J. Med. Syst.*, vol. 42, no. 5, pp. 92, May 2018.
- [41] G. J. McLachlan, "Discriminant analysis and statistical pattern recognition", *J. Roy. Stat. Soc. Ser. A Statist. Soc.*, vol. 168, no. 3, pp. 635-636, Jun. 2005.
- [42] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition", *IEEE Trans. Electron. Comput.*, vol. EC, no. 3, pp. 326-334, Jun. 1965.
- [43] G. I. Webb, J. R. Boughton and Z. Wang, "Not so naive bayes: Aggregating one-dependence estimators", *Mach. Learn.*, vol. 58, no. 1, pp. 5-24, Jan. 2005.
- [44] S. Brahim-Belhouari and A. Bermak, "Gaussian process for nonstationary time series prediction", *Comput. Statist. Data Anal.*, vol. 47, no. 4, pp. 705-712, Nov. 2004.
- [45] C. Cortes and V. Vapnik, "Support-vector networks", *Mach. Learn.*, vol. 20, pp. 237-297, Sep. 1995.

- [46] A. Reinhardt and T. Hubbard, "Using neural networks for prediction of the subcellular location of proteins", *Nucleic Acids Res.*, vol. 26, no. 9, pp. 2230-2236, May 1998.
- [47] B. Kégl, "The return of AdaBoost.MH: Multi-class Hamming trees", arXiv:1312.6086, 2013, [online] Available: <http://arxiv.org/abs/1312.6086>.
- [48] B. P. Tabaei and W. H. Herman, "A multivariate logistic regression equation to screen for diabetes: Development and validation", *Diabetes Care*, vol. 25, no. 11, pp. 1999-2003, Nov. 2002.
- [49] I. Jenhani, N. B. Amor and Z. Elouedi, "Decision trees as possibilistic classifiers", *Int. J. Approx. Reasoning*, vol. 48, no. 3, pp. 784-807, Aug. 2008.
- [50] L. Breiman, "Random forests", *Mach. Learn.*, vol. 45, no. 1, pp. 5-32, Oct. 2001.
- [51] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms", *Procedia Comput. Sci.*, vol. 132, pp. 1578-1585, Jan. 2018.
- [52] S. Perveen, M. Shahbaz, A. Guergachi and K. Keshavjee, "Performance analysis of data mining classification techniques to predict diabetes", *Procedia Comput. Sci.*, vol. 82, pp. 115-121, 2016.
- [53] M. Pradhan and G. R. Bamnote, "Design of classifier for detection of diabetes mellitus using genetic programming", *Proc. 3rd Int. Conf. Frontiers Intell. Comput. Theory Appl.*, pp. 763-770, Nov. 2015.
- [54] N. Nai-arun and R. Moungrmai, "Comparison of classifiers for the risk of diabetes prediction", *Procedia Comput. Sci.*, vol. 69, pp. 132-142, Dec. 2015.
- [55] F. Es-SABERY and A. Hair, "An improved ID3 classification algorithm based on correlation function and weighted attribute", *Proc. Int. Conf. Intell. Syst. Adv. Comput. Sci. (ISACS)*, pp. 1-8, Dec. 2019.
- [56] L. Yuxun and X. Niuniu, "Improved ID3 algorithm", *Proc. 3rd Int. Conf. Comput. Sci. Inf. Technol.*, pp. 465-468, Jul. 2010.
- [57] R.-M. Chai and M. Wang, "A more efficient classification scheme for ID3", *Proc. 2nd Int. Conf. Comput. Eng. Technol.*, pp. 329-332, 2010.
- [58] S. Elyassami and A. Idri, "Applying fuzzy ID3 decision tree for software effort estimation" in arXiv:1111.0158, 2011, [online] Available: <http://arxiv.org/abs/1111.0158>.
- [59] K. Zou, W. Sun, H. Yu and F. Liu, "ID3 decision tree in fraud detection application", *Proc. Int. Conf. Comput. Sci. Electron. Eng.*, pp. 399-402, Mar. 2012.
- [60] V. Srinivasan, G. Rajenderan, J. Vandar Kuzhali and M. Aruna, "Fuzzy fast classification algorithm with hybrid of ID3 and SVM", *J. Intell. Fuzzy Syst.*, vol. 24, no. 3, pp. 555-561, 2013.
- [61] X. J. Chen, Z. G. Zhang and Y. Tong, "An improved ID3 decision tree algorithm", *Adv. Mater. Res.*, vol. 962, pp. 2842-2847, Jun. 2014.
- [62] B. Ding, Y. Zheng and S. Zang, "A new decision tree algorithm based on rough set theory", *Proc. Asia-Pacific Conf. Inf. Process.*, pp. 326-329, Jul. 2009.
- [63] L. Zhu and Y. Yang, "Improvement of decision tree ID3 algorithm" in *Collaborate Computing: Networking Applications and Worksharing*, Cham, Switzerland:Springer, vol. 201, pp. 595-600, 2017.
- [64] N. Kaewrod and K. Jearanaitanakij, "Improving ID3 algorithm by ignoring minor instances", *Proc. 22nd Int. Comput. Sci. Eng. Conf. (ICSEC)*, pp. 1-5, Nov. 2018.
- [65] A. Rajeshkanna and K. Arunesh, "ID3 decision tree classification: An algorithmic perspective based on error rate", *Proc. Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, pp. 787-790, Jul. 2020.
- [66] Tejashri, N., & Satish, R. (2015). Prognosis of diabetes using neural network, fuzzy logic, gaussian kernel method. *International Journal of Computer Applications*, 124(10), 33-36.
- [67] L. Li et al., "Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on Weibo", *IEEE Trans. Comput. Social Syst.*, vol. 7, no. 2, pp. 556-562, Mar. 2020.
- [68] G. Barkur and G. B. Kamath, "Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India", *Asian J. Psychiatry*, vol. 51, Jun. 2020.

- [69] H. Jelodar, Y. Wang, R. Orji and S. Huang, "Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach", *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2733-2742, Oct. 2020.
- [70] Georga E et al (2009) Data mining for blood glucose prediction and knowledge discovery in diabetic patients: the METABO diabetes modeling and management system. In: 31st Annual international conference. IEEE EMBS Minneapolis, Minnesota, USA, vol 43100, pp 5633–5636
- [71] Breault, J. L., Goodall, C. R., & Fos, P. J. (2002). Data mining a diabetic data warehouse. *Artificial intelligence in medicine*, 26(1-2), 37-54.
- [72] Vijayan V, Ravi K (2015) Prediction and diagnosis of diabetes mellitus—a machine learning approach, December, pp 122–127
- [73] Polat, K., Güneş, S., & Arslan, A. (2008). A cascade learning system for classification of diabetes disease: Generalized discriminant analysis and least square support vector machine. *Expert systems with applications*, 34(1), 482-487.
- [74] Zein, J. G., Wu, C. P., Attaway, A. H., Zhang, P., & Nazha, A. (2021). Novel machine learning can predict acute asthma exacerbation. *Chest*, 159(5), 1747-1757.
- [75] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*, 2(1), 56-67.
- [76] Swain A, Mohanty S N, Das AC (2016) Comparative risk analysis on prediction of diabetes mellitus using machine learning approach. In: 2016 international conference on electrical, electronics, and optimization Techniques (ICEEOT), pp 3312–3317
- [77] Pradeep Kandhasamy J, Balamurali S (2015) Performance analysis of classifier models to predict diabetes mellitus. *Proc Comput Sci* 47:45–51
- [78] Zou Q, Qu K, Luo Y, Yin D, Ju Y, Tang H (2018) Predicting diabetes mellitus with machine learning techniques. *Frontiers in Genetics* 9:515
- [79] Xu W, Zhang J, Zhang Q, Wei X (2017) Risk prediction of type II diabetes based on random forest model. In: 3rd international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB), pp 382–386, vol 2017
- [80] Sisodia D, Sisodia DS (2018) Prediction of diabetes using classification algorithms. *Proc Comput Sci* 132:1578–1585
- [81] Ribeiro AC, Barros AK, Santana E, Príncipe JC (2015) Diabetes classification using a redundancy reduction preprocessor
- [82] Panwar M, Acharyya A, Shafik R A, Biswas D (2016) K-nearest neighbor based methodology for accurate diagnosis of diabetes mellitus. In: 2016 6th international symposium on embedded computing and system design (ISED), pp 132–136
- [83] Brinati, D., Campagner, A., Ferrari, D., Locatelli, M., Banfi, G., & Cabitza, F. (2020). Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. *Journal of medical systems*, 44(8), 1-12.
- [84] Iyer, A., Jeyalatha, S., & Sumbaly, R. (2015). Diagnosis of diabetes using classification mining techniques. *arXiv preprint arXiv:1502.03774*.
- [85] Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.
- [86] Kadhm, M. S., Ghindawi, I. W., & Mhawi, D. E. (2018). An accurate diabetes prediction system based on K-means clustering and proposed classification approach. *International Journal of Applied Engineering Research*, 13(6), 4038-4041.
- [87] Diabetesdaily, <https://www.diabetesdaily.com/learn-about-diabetes/basics/what-is-diabetes/how-many-people-have-diabetes/>
- [88] Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). Comparison of Naive Bayes KNN And Decision Tree Methods To Sentiment Analysis Of Commuter Line KRL Transport. *J. IKRA-ITH Inform.*, 3(2), 1-9.
- [89] Xhemali, D., Hinde, C. J., & Stone, R. (2009). Naïve bayes vs. decision trees vs. neural networks in the classification of training web pages.



- [90] Van der Heide, E. M. M., Veerkamp, R. F., Van Pelt, M. L., Kamphuis, C., Athanasiadis, I., & Ducro, B. J. (2019). Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle. *Journal of dairy science*, 102(10), 9409-9421.
- [91] Caliskan, A., Yuksel, M. E., Badem, H., & Basturk, A. (2018). Performance improvement of deep neural network classifiers by a simple training strategy. *Engineering Applications of Artificial Intelligence*, 67, 14-23.
- [92] Wan, S., Liang, Y., & Zhang, Y. (2018). Deep convolutional neural networks for diabetic retinopathy detection by image classification. *Computers & Electrical Engineering*, 72, 274-282.
- [93] Wu, H., Yang, S., Huang, Z., He, J., & Wang, X. (2018). Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, 10, 100-107.
- [94] Mercaldo, F., Nardone, V., & Santone, A. (2017). Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. *Procedia computer science*, 112, 2519-2528.
- [95] Duch, W. (2000). Datasets used for classification: comparison of results. <http://158.75.5.90/kmk/projects/datasets.html>.
- [96] Farooq, M., & Hafeez, A. (2020). Covid-resnet: A deep learning framework for screening of covid19 from radiographs. *arXiv preprint arXiv:2003.14395*.
- [97] Pramono, R. X. A., Bowyer, S., & Rodriguez-Villegas, E. (2017). Automatic adventitious respiratory sound analysis: A systematic review. *PloS one*, 12(5), e0177926.
- [98] Schuller, B. W., Schuller, D. M., Qian, K., Liu, J., Zheng, H., & Li, X. (2021). Covid-19 and computer audition: An overview on what speech & sound analysis could contribute in the sars-cov-2 corona crisis. *Frontiers in digital health*, 14.
- [99] Li, S. H., Lin, B. S., Tsai, C. H., Yang, C. T., & Lin, B. S. (2017). Design of wearable breathing sound monitoring system for real-time wheeze detection. *Sensors*, 17(1), 171.
- [100] Bales, C., Nabeel, M., John, C. N., Masood, U., Qureshi, H. N., Farooq, H., ... & Imran, A. (2020, October). Can machine learning be used to recognize and diagnose coughs?. In *2020 International Conference on e-Health and Bioengineering (EHB)* (pp. 1-4). IEEE.
- [101] Alakus, T. B., & Turkoglu, I. (2020). Comparison of deep learning approaches to predict COVID-19 infection. *Chaos, Solitons & Fractals*, 140, 110120.
- [102] Deshpande, G., & Schuller, B. (2020). An overview on audio, signal, speech, & language processing for COVID-19. *arXiv preprint arXiv:2005.08579*.
- [103] Brown, C., Chauhan, J., Grammenos, A., Han, J., Hasthanasombat, A., Spathis, D., ... & Mascolo, C. (2020). Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. *arXiv preprint arXiv:2006.05919*.
- [104] Butwall, M., & Kumar, S. (2015). A data mining approach for the diagnosis of diabetes mellitus using random forest classifier. *International Journal of Computer Applications*, 120(8).
- [105] Kumar Dewangan, A., & Agrawal, P. (2015). Classification of diabetes mellitus using machine learning techniques. *International Journal of Engineering and Applied Sciences*, 2(5), 257905.
- [106] Daghistani, T., & Alshammari, R. (2016). Diagnosis of diabetes by applying data mining classification techniques. *International Journal of Advanced Computer Science and Applications*, 7(7).
- [107] Tang, Z., Zhao, W., Xie, X., Zhong, Z., Shi, F., Liu, J., & Shen, D. (2020). Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images. *arXiv preprint arXiv:2003.11988*.
- [108] Khalifa, N. E. M., Taha, M. H. N., Hassanien, A. E., & Elghamrawy, S. (2023). Detection of coronavirus (covid-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest x-ray dataset. In *International Conference on Advanced Intelligent Systems and Informatics* (pp. 234-247). Springer, Cham.

- [109] Sujath, R. A. A., Chatterjee, J. M., & Hassanien, A. E. (2020). A machine learning forecasting model for COVID-19 pandemic in India. *Stochastic Environmental Research and Risk Assessment*, 34(7), 959-972.
- [110] Jiang, W. E. I., & Siddiqui, S. (2020). Hyper-parameter optimization for support vector machines using stochastic gradient descent and dual coordinate descent. *EURO Journal on Computational Optimization*, 8(1), 85-101.
- [111] Mareček, J., Richtárik, P., & Takáč, M. (2015). Distributed block coordinate descent for minimizing partially separable functions. In *Numerical Analysis and Optimization* (pp. 261-288). Springer, Cham.
- [112] Parnell, T., Dünner, C., Atasu, K., Sifalakis, M., & Pozidis, H. (2020). Tera-scale coordinate descent on GPUs. *Future Generation Computer Systems*, 108, 1173-1191.
- [113] James, R. M., & Arief, M. R. (2020, October). Classification of x-ray covid-19 image using convolutional neural network. In *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)* (pp. 1-6). IEEE.
- [114] Polat, K., Güneş, S., & Arslan, A. (2008). A cascade learning system for classification of diabetes disease: Generalized discriminant analysis and least square support vector machine. *Expert systems with applications*, 34(1), 482-487.
- [115] de Moraes Batista, A. F., Miraglia, J. L., Donato, T. H. R., & Chiavegatto Filho, A. D. P. (2020). COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. *MedRxiv*.
- [116] Banerjee, A., Ray, S., Vorselaars, B., Kitson, J., Mamalakis, M., Weeks, S., ... & Mackenzie, L. S. (2020). Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population. *International immunopharmacology*, 86, 106705.
- [117] Pourhomayoun, M., & Shakibi, M. (2021). Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health*, 20, 100178.
- [118] Jabbar, R., Fetais, N., Krichen, M., & Barkaoui, K. (2020, February). Blockchain technology for healthcare: Enhancing shared electronic health record interoperability and integrity. In *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)* (pp. 310-317). IEEE.
- [119] Alamri, A. (2018). Ontology middleware for integration of IoT healthcare information systems in EHR systems. *Computers*, 7(4), 51.
- [120] Kouroubali, A., & Katehakis, D. G. (2019). The new European interoperability framework as a facilitator of digital transformation for citizen empowerment. *Journal of biomedical informatics*, 94, 103166.
- [121] Jung, H., & Jeong, D. (2021). Blockchain Implementation Method for Interoperability between CBDCs. *Future Internet*, 13(5), 133.
- [122] Biswas, P. EHR-Interoperability study of HL7 amid the COVID-19 epidemic.
- [123] Lee, E., Seo, Y. D., Oh, S. R., & Kim, Y. G. (2021). A Survey on Standards for Interoperability and Security in the Internet of Things. *IEEE Communications Surveys & Tutorials*, 23(2), 1020-1047.
- [124] Satti, F. A., Ali, T., Hussain, J., Khan, W. A., Khattak, A. M., & Lee, S. (2020). Ubiquitous Health Profile (UHP): a big data curation platform for supporting health data interoperability. *Computing*, 102(11), 2409-2444.
- [125] Lehne, M., Sass, J., Essenwanger, A., Schepers, J., & Thun, S. (2019). Why digital medicine depends on interoperability. *NPJ digital medicine*, 2(1), 1-5.
- [126] Powell, K., & Alexander, G. (2019). Mitigating barriers to interoperability in health care. *On-Line Journal of Nursing Informatics*, 23(2).
- [127] Shehzad, H. M. F., Ibrahim, R. B., Yusof, A. F., Khaidzir, K. A. M., Iqbal, M., & Razzaq, S. (2021). The role of interoperability dimensions in building information modelling. *Computers in Industry*, 129, 103444.
- [128] Saripalle, R., Runyan, C., & Russell, M. (2019). Using HL7 FHIR to achieve interoperability in patient health record. *Journal of biomedical informatics*, 94, 103188.

- [129] Bodenreider, O., Cornet, R., & Vreeman, D. J. (2018). Recent developments in clinical terminologies—SNOMED CT, LOINC, and RxNorm. *Yearbook of medical informatics*, 27(01), 129-139.
- [130] Wang, K. C. (2018). Standard lexicons, coding systems and ontologies for interoperability and semantic computation in imaging. *Journal of digital imaging*, 31(3), 353-360.
- [131] Rico, J., Echevarría-González de Garibay, L. J., García-López, M., Guardiola-Villarraig, S., Maceda-Roldán, L. A., Zurriaga, Ó., & Cavero-Carbonell, C. (2021). The interoperability between the Spanish version of the International Classification of Diseases and ORPHAcodes: towards better identification of rare diseases. *Orphanet journal of rare diseases*, 16(1), 1-13.
- [132] Juric, D., Geleta, D., McKay, G., & Stoilos, G. (2021). A Platform and Algorithms for Interoperability Between Clinical Coding Systems. *Procedia Computer Science*, 192, 563-572.
- [133] Gómez, D., Romero, J., López, P., Vázquez, J., Cappel, C., Pinto, D., & Villalba, C. (2022). Cloud architecture for electronic health record systems interoperability. *Technology and Health Care*, 30(3), 551-564.
- [134] Bae, S., & Yi, B. K. (2022). Development of eClaim system for private indemnity health insurance in South Korea: Compatibility and interoperability. *Health Informatics Journal*, 28(1), 14604582211071019.
- [135] Osterman, T. J., Terry, M., & Miller, R. S. (2020). Improving cancer data interoperability: the promise of the Minimal Common Oncology Data Elements (mCODE) initiative. *JCO Clinical Cancer Informatics*, 4, 993-1001.
- [136] Saripalle, R. K. (2019). Fast Health Interoperability Resources (FHIR): current status in the healthcare system. *International Journal of E-Health and Medical Communications (IJEHMC)*, 10(1), 76-93.
- [137] The Clean Architecture, <https://8thlight.com/blog/uncle-bob/2012/08/13/the-clean-architecture.html>
- [138] Angular v2 Archive, <https://v2.angular.io/docs/ts/latest/guide/architecture.html>
- [139] Dev Academy, <https://dev-academy.com/angular-architecture-best-practices/>
- [140] <https://docs.djangoproject.com/en/1.11/intro/>