ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Αξιοποίηση Εργαλείων και Εφαρμογών Τεχνητής Νοημοσύνης στη Διαχείριση και Αντιμετώπιση της Πανδημίας COVID-19

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γεωργία Α. Ξυπολιά

Επιβλέπων: Δημήτριος - Διονύσιος Κουτσούρης
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπουσα: Ουρανία Πετροπούλου
Ε.ΔΙ.Π. Ε.Μ.Π.

Αθήνα, Μάρτιος 2021

Η σελίδα αυτή είναι σκόπιμα λευκή.

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Αξιοποίηση Εργαλείων και Εφαρμογών Τεχνητής Νοημοσύνης στη Διαχείριση και Αντιμετώπιση της Πανδημίας COVID-19

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γεωργία Α. Ξυπολιά

Επιβλέπων: Δημήτριος - Διονύσιος Κουτσούρης
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπουσα: Ουρανία Πετροπούλου
Ε.ΔΙ.Π. Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 9η Μαρτίου 2021.

.....	.....	.....
Δημήτριος - Διονύσιος Κουτσούρης	Γεώργιος Ματσόπουλος	Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2021

.....

Γεωργία Α. Ξυπολιά

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Γεωργία Α. Ξυπολιά, 2021

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τη συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τη συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η πανδημία COVID-19, για την οποία είναι υπεύθυνος ο νέος κορωνοϊός SARS-CoV-2, έχει οδηγήσει στην πιο σοβαρή υγειονομική κρίση των τελευταίων 100 ετών, προκαλώντας έναν τρομακτικό και συνεχώς αυξανόμενο αριθμό θανάτων, μία πρωτόγνωρη πίεση στα συστήματα υγείας και τεράστιες οικονομικές και κοινωνικές συνέπειες παγκοσμίως.

Από την εμφάνιση του νέου ιού μέχρι και σήμερα, μια πληθώρα ερευνητών, οργανισμών και εταιρειών έχουν επιδοθεί σε έναν αγώνα δημιουργίας εφαρμογών που χρησιμοποιούν εργαλεία και μεθόδους τεχνητής νοημοσύνης, μηχανικής μάθησης και ανάλυσης μεγάλων δεδομένων, σε μια προσπάθεια να συνεισφέρουν στην αντιμετώπιση διαφορετικών πτυχών της πανδημίας. Μέσω της Διπλωματικής αυτής Εργασίας, επιχειρείται μια ανασκόπηση των υπαρχουσών εφαρμογών στις εξής τέσσερις περιοχές: την Επιδημιολογία, την Κλινική Ιατρική, τη Φαρμακολογία και την Επιδημία της Πληροφορίας (Infodemic).

Αρχικά, στον τομέα την Επιδημιολογίας, μελετάμε εφαρμογές που βοηθούν στην έγκαιρη ανίχνευση και προειδοποίηση, την αυτοματοποιημένη ιχνηλάτηση επαφών, την πρόβλεψη και την προσομοίωση της εξάπλωσης, την ομαδοποίηση γεωγραφικών περιοχών με βάση κοινά χαρακτηριστικά και την αξιολόγηση κινδύνου ανά περιοχή. Στην περιοχή της Κλινικής Ιατρικής, παρουσιάζουμε μεθόδους διάγνωσης της ασθένειας με χρήση ιατρικής απεικόνισης και άλλων εναλλακτικών πηγών πληροφορίας, πρόγνωσης της πορείας των ασθενών, καθώς και βέλτιστης διαχείρισης των υφιστάμενων πόρων και ανάπτυξης εξατομικευμένων θεραπειών. Σε ό,τι αφορά τη Φαρμακολογία, περιγράφουμε ενδιαφέρουσες εφαρμογές για την πρόβλεψη της τρισδιάστατης δομής των πρωτεϊνών, την ανακάλυψη θεραπειών και πιθανών στόχων εμβολίων, καθώς και εφαρμογές για τη βελτίωση των κλινικών δοκιμών. Τέλος, ορίζουμε την έννοια της Επιδημίας της Πληροφορίας και παρουσιάζουμε λύσεις που στοχεύουν στην ανίχνευση και τον υπολογισμό της εξάπλωσης των ψευδών ειδήσεων και της παραπληροφόρησης, αλλά και στην κατανόηση του συναισθήματος και των αντιδράσεων της κοινής γνώμης.

Καθοριστικό ρόλο στην επιτυχία αυτών των εφαρμογών τεχνητής νοημοσύνης και μηχανικής μάθησης διαδραματίζουν τα δεδομένα με τα οποία εκπαιδεύονται τα αντίστοιχα συστήματα. Για το λόγο αυτό, κάνουμε μια ανασκόπηση των διαφορετικών πηγών δεδομένων και παρουσιάζουμε δημοφιλή σύνολα δεδομένων, τα οποία είναι δημόσια διαθέσιμα και μπορούν να χρησιμοποιηθούν για την ανάπτυξη καινοτόμων λύσεων.

Ολοκληρώνουμε την Εργασία με την περιγραφή μιας σειράς κοινών προκλήσεων που έχουν να αντιμετωπίσουν οι πιο πάνω εφαρμογές σε σχέση με το πλήθος και την ποιότητα των δεδομένων, την προστασία ευαίσθητων και προσωπικών πληροφοριών, την ασφάλεια, την επεξηγησιμότητα, τη μεροληψία και άλλα ζητήματα ηθικής. Αναγνωρίζοντας τη σοβαρότητα τέτοιων προκλήσεων, προτείνουμε μια σειρά από κατευθύνσεις για την αντιμετώπισή τους, στοχεύοντας παράλληλα στη μεγιστοποίηση του οφέλους από τη χρήση των εφαρμογών αυτών.

Λέξεις Κλειδιά

COVID-19, SARS-CoV-2, πανδημία, κορωνοϊός, τεχνητή νοημοσύνη, μηχανική μάθηση, βαθιά μάθηση, μεγάλα δεδομένα, επιδημιολογία, ιχνηλάτηση επαφών, διάγνωση, πρόγνωση, ανάλυση ιατρικής εικόνας, ανακάλυψη φαρμάκων, ανακάλυψη εμβολίων, επιδημία πληροφορίας.

Abstract

The COVID-19 pandemic, for which the novel SARS-CoV-2 coronavirus is responsible, has led to the most serious health crisis of the last 100 years, causing a tremendous and constantly increasing number of deaths, an unprecedented pressure on healthcare systems and huge economic and social consequences worldwide.

Since the emergence of this new virus, a plethora of researchers, organizations and companies have engaged in a struggle to create applications that use Artificial Intelligence (AI) tools and methods, machine learning and big data analytics, in an effort to contribute to managing different aspects of the pandemic. Through this Diploma Thesis, we perform a review of existing AI applications in the following four areas: Epidemiology, Clinical Medicine, Pharmacology and the Epidemic of Information (Infodemic).

Beginning with the field of Epidemiology, we discuss applications for early detection and alerting of epidemic outbreaks, automated contact tracing, prediction and simulation of spread, clustering of different geographic regions based on common characteristics, as well as performing risk assessment by region. In the area of Clinical Medicine, we present methods for disease diagnosis, using medical imaging and other alternative sources of information, prognosis of the course of patients, as well as optimal management of existing resources and development of personalized treatments. Regarding Pharmacology, we describe interesting applications for predicting the three-dimensional structure of proteins, discovering therapies and potential vaccine targets, as well as applications for improving clinical trials. Finally, we define the concept of Infodemic and present solutions aimed at detecting and calculating the spread of fake news and misinformation, along with understanding sentiment and public reactions.

The data with which the respective systems are trained play a key role in the success of such applications utilizing artificial intelligence and machine learning. Due to this, we review the different types of data sources and present popular datasets, which are publicly accessible and can be used to develop innovative solutions.

In conclusion, we describe a series of common challenges facing the above applications related to the quantity and quality of data, the protection of sensitive and personal information, security, explainability, bias and other ethical issues. Recognizing the importance of overcoming these challenges, we propose several ways to address them, while aiming at maximizing the benefit and impact from the use of these applications.

Keywords

COVID-19, SARS-CoV-2, pandemic, coronavirus, artificial intelligence, machine learning, deep learning, big data, epidemiology, contact tracing, diagnosis, prognosis, medical image analysis, drug discovery, vaccine discovery, infodemic.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή κ. Δημήτριο - Διονύσιο Κουτσούρη για την εμπιστοσύνη που μου έδειξε με την ανάθεση αυτής της Διπλωματικής Εργασίας, δίνοντάς μου παράλληλα την ευκαιρία να ασχοληθώ με ένα τόσο ενδιαφέρον και επίκαιρο θέμα. Επίσης, θέλω να ευχαριστήσω ιδιαίτερα την Ε.ΔΙ.Π. κ. Ουρανία Πετροπούλου για την καθοδήγηση και την ενθάρρυνση που μου παρείχε καθ' όλη τη διάρκεια της εκπόνησης της Εργασίας αυτής.

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πίνακας Περιεχομένων

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή ____ *14*

1.1 Επισκόπηση της Πανδημίας COVID-19 ____ **14**

1.1.1 Τι Είναι η Ασθένεια COVID-19 ____ 14

1.1.2 Συμπτώματα, Μετάδοση και Βασικά Μέτρα Προστασίας ____ 15

1.1.3 Πώς η Ασθένεια COVID-19 Έγινε Πανδημία - Ένα Χρονοδιάγραμμα ____ 16

1.2 Προκλήσεις της Πανδημίας COVID-19 και Ευκαιρία για Αξιοποίηση της Τεχνητής Νοημοσύνης ____ **18**

1.2.1 Γιατί η Πανδημία COVID-19 Είναι Τόσο Επικίνδυνη ____ 18

1.2.2 Κίνητρα για Χρήση της Τεχνητής Νοημοσύνης ____ 20

1.3 Θεμελιώδεις Έννοιες της Τεχνητής Νοημοσύνης ____ **20**

1.3.1 Τεχνητή Νοημοσύνη ____ 20

1.3.2 Μηχανική Μάθηση ____ 21

1.3.3 Βαθιά Μάθηση και Επεξεργασία Φυσικής Γλώσσας ____ 24

1.3.4 Τεχνητά Νευρωνικά Δίκτυα ____ 25

1.3.5 Big Data και Σύγχρονες Αρχιτεκτονικές Τεχνητής Νοημοσύνης ____ 27

1.3.6 Τεχνητή Νοημοσύνη και Ανάλυση Μεγάλων Δεδομένων στο Χώρο της Υγείας και της Ιατρικής 29

1.4 Δομή της Διπλωματικής Εργασίας ____ **31**

ΚΕΦΑΛΑΙΟ 2: Εφαρμογές στην Επιδημιολογία ____ *33*

2.1 Ο Ρόλος της Τεχνητής Νοημοσύνης στην Επιδημιολογία ____ **33**

2.2 Έγκαιρη Ανίχνευση και Προειδοποίηση ____ **33**

2.2.1 Η Εφαρμογή της BlueDot ____ 34

2.2.2 Η Εφαρμογή της Metabiota ____ 35

2.2.3 Το Παρελθόν, το Παρόν και το Μέλλον στην Πρόβλεψη της Εξάπλωσης Μολυσματικών Ασθενειών ____ 36

2.3 Ιχνηλάτηση Επαφών ____ **38**

2.3.1 Ανάγκη για Έξυπνο Εντοπισμό και Ιχνηλάτηση Επαφών ____ 38

2.3.2 Τύποι Αρχιτεκτονικών των Εφαρμογών Ιχνηλάτησης Επαφών ____ 40

2.3.3 Παραδείγματα Εφαρμογών Ιχνηλάτησης Επαφών ____ 41

2.3.4 Μηχανική Μάθηση στην Ιχνηλάτηση Επαφών ____ 42

2.4 Επιδημιολογικά Μοντέλα - Πρόβλεψη και Προσομοίωση της Εξάπλωσης ____ **43**

2.4.1 Πρόβλεψη με Βάση Επίσημα Καταγεγραμμένα Στοιχεία ____ 43

2.4.2 Πρόβλεψη με Χρήση Επιπρόσθετων Online Πηγών ____ 46

2.5 Ομαδοποίηση Περιοχών ____ **48**

2.6 Εκτίμηση Κινδύνου ανά Περιοχή ____ **50**

ΚΕΦΑΛΑΙΟ 3: Εφαρμογές στην Κλινική Ιατρική ____ *53*

3.1 Εισαγωγή ____ **53**

3.2 Διάγνωση ____ **53**

3.2.1 Εισαγωγή στις Διαγνωστικές Μεθόδους για την Ασθένεια COVID-19 ____ 53

3.2.2 Διάγνωση μέσω Ιατρικής Απεικόνισης ____ 54

3.2.3 Ψηφιακές Μέθοδοι Ταχείας Διάγνωσης (Screening) με Χρήση Chatbots και Internet of Things 58

3.3 Πρόγνωση ____ **63**

3.3.1 Η Σημασία της Πρόγνωσης 63
3.3.2 Πρόβλεψη της Σοβαρότητας της Νόσου 63

3.4 Βέλτιστη Διαχείριση Ασθενών με Πεπερασμένους Πόρους 66
3.4.1 Οι Προκλήσεις της Πανδημίας COVID-19 στη Διαχείριση των Ασθενών 66
3.4.2 Ανάθεση Πόρων 67
3.4.3 Ανάπτυξη Εξατομικευμένων Πλάνων Διαχείρισης και Θεραπείας Ασθενών 70

ΚΕΦΑΛΑΙΟ 4: Εφαρμογές στη Φαρμακολογία 72

4.1 Εισαγωγή 72

4.2 Πρόβλεψη Δομής Πρωτεϊνών 72
4.2.1 Το Πρόβλημα Αναδίπλωσης Πρωτεϊνών 72
4.2.2 Εφαρμογές Τεχνητής Νοημοσύνης για την Πρόβλεψη της Δομής Πρωτεϊνών του SARS-CoV-2 75

4.3 Ανακάλυψη Θεραπειών και Εμβολίων 78
4.3.1 Ο Ρόλος της Τεχνητής Νοημοσύνης στην Ανάπτυξη Φαρμάκων 78
4.3.2 Επαναστόχευση Φαρμάκων 78
4.3.3 Ανακάλυψη Νέων Φαρμάκων 83
4.3.4 Ανακάλυψη Πιθανών Στόχων Εμβολίων 85

4.4 Βελτίωση Κλινικών Δοκιμών 88
4.4.1 Η Τρέχουσα Κατάσταση στις Κλινικές Δοκιμές 88
4.4.2 Πώς η Τεχνητή Νοημοσύνη Μπορεί να Μετασχηματίσει τις Κλινικές Δοκιμές 89
4.4.3 Μετασχηματισμός Κλινικών Δοκιμών στην Εποχή της Πανδημίας COVID-19 92

ΚΕΦΑΛΑΙΟ 5: Εφαρμογές στην Καταπολέμηση της Επιδημίας της Πληροφορίας 94

5.1 Η Επιδημία της Πληροφορίας (Infodemic) 94

5.2 Εφαρμογές Τεχνητής Νοημοσύνης για την Αντιμετώπιση της Επιδημίας της Πληροφορίας 95
5.2.1 Υπολογισμός του Δείκτη Κινδύνου Επιδημίας της Πληροφορίας 95
5.2.2 Μοντελοποίηση της Εξάπλωσης της Επιδημίας της Πληροφορίας 98
5.2.3 Κατανόηση Συναισθήματος και Αντιδράσεων της Κοινής Γνώμης 100

ΚΕΦΑΛΑΙΟ 6: Τύποι Δεδομένων και Ανοικτά Σύνολα Δεδομένων 103

6.1 Δεδομένα στην Εποχή της Πανδημίας COVID-19 103

6.2 Δεδομένα Κρουσμάτων 103

6.3 Δεδομένα Ιατρικής Απεικόνισης 105

6.4 Μοριακά Δεδομένα 107

6.5 Δεδομένα Επιστημονικής Βιβλιογραφίας 108

6.6 Δεδομένα Μέσων Κοινωνικής Δικτύωσης 110

ΚΕΦΑΛΑΙΟ 7: Προκλήσεις, Προτάσεις και Συμπεράσματα 114

7.1 Εισαγωγή 114

7.2 Προκλήσεις 114
7.2.1 Ποσότητα Δεδομένων, Ποιότητα Δεδομένων και Μεροληψία (Bias) 114
7.2.2 Μοντέλα «Μαύρου Κουτιού» 116
7.2.3 Έκτακτες Συνθήκες και Μεταβαλλόμενες Ανάγκες 116
7.2.4 Προσωπικά Δεδομένα 117

7.2.5 Ζητήματα Ηθικής 118

7.3 Προτάσεις 119

7.3.1 Κοινή Χρήση Δεδομένων και Ανοικτά Δεδομένα 119

7.3.2 Επεξηγησιμότητα και Δυνατότητα Ερμηνείας των Μοντέλων Τεχνητής Νοημοσύνης 119

7.3.3 Ηθική με Διαφάνεια και Αίσθηση του Επείγοντος 120

7.3.4 Αρχιτεκτονικές που Παρέχουν Ασφάλεια και Προστασία των Προσωπικών Δεδομένων 121

7.3.5 Επαλήθευση Μοντέλων σε Πραγματικές Συνθήκες 122

7.3.6 Διεπιστημονικές Συνεργασίες 123

7.3.7 Διεθνής Συνεργασία και Λύσεις για Αναπτυσσόμενες Χώρες 124

7.4 Συμπεράσματα 124

Βιβλιογραφικές Αναφορές 126

Πίνακας Σχημάτων

Σχήμα 1 *- Αναπαράσταση της δομής του SARS-CoV-2 και του τρόπου εισόδου του στο κύτταρο του ξενιστή _ 14*
Σχήμα 2 *- Μόλυνση από τον ιό SARS-CoV-2 – Ιικό Φορτίο και Απόκριση Αντισωμάτων σε σχέση με το χρόνο για Ήπιες - Μέτριες, Σοβαρές και Κρίσιμες περιπτώσεις συμπτωματικών ασθενών _____ 16*
Σχήμα 3 *- Αλληλεπίδραση ενός ευφυούς πράκτορα με το περιβάλλον του _____ 21*
Σχήμα 4 *- Οι τρεις κινητήριες δυνάμεις πίσω από την επανάσταση της τεχνητής νοημοσύνης _____ 21*
Σχήμα 5 *- Σύγκριση του παραδοσιακού προγραμματισμού και της μηχανικής μάθησης _____ 22*
Σχήμα 6 *- Επισκόπηση της επιτηρούμενης μάθησης _____ 22*
Σχήμα 7 *- Επισκόπηση της μη επιτηρούμενης μάθησης _____ 23*
Σχήμα 8 *- Αλληλεπίδραση ενός πράκτορα ενισχυτικής μάθησης με το περιβάλλον του _____ 24*
Σχήμα 9 *- Επισκόπηση ενός απλού τεχνητού νευρωνικού δικτύου _____ 26*
Σχήμα 10 *- Μια αρχιτεκτονική υψηλού επιπέδου ενός σύγχρονου συστήματος τεχνητής νοημοσύνης με δυνατότητα ανάλυσης Big Data _____ 28*
Σχήμα 11 *- Κατηγοριοποίηση των εφαρμογών τεχνητής νοημοσύνης στην παρούσα εργασία _____ 32*
Σχήμα 12 *- Διεπαφή web και εφαρμογή για κινητά τηλέφωνα του BlueDot Insights _____ 34*
Σχήμα 13 *- Διάγραμμα ροής ενός μοντέλου μη επιτηρούμενης ανάλυσης συναισθήματος _____ 37*
Σχήμα 14 *- Επισκόπηση της αυτοματοποιημένης ανίχνευσης επαφών με χρήση εφαρμογών για κινητά τηλέφωνα και Bluetooth _____ 39*
Σχήμα 15 *- Ενδεικτικά παραδείγματα εφαρμογών ιχνηλάτησης επαφών και αντίστοιχων αρχιτεκτονικών και πρωτοκόλλων _____ 42*
Σχήμα 16 *- Δομή ενός μοντέλου CNN για την πρόβλεψη του αριθμού των νέων ημερήσιων κρουσμάτων ___ 44*
Σχήμα 17 *- Διάγραμμα μοντέλου πρόβλεψης που συνδυάζει επίσημες με άλλες online πηγές δεδομένων ___ 47*
Σχήμα 18 *- Μέσος όρος κανονικοποιημένων ημερήσιων θανάτων για κάθε ομάδα χωρών ως προς το χρόνο (αριστερά) και κατανομή των ομάδων χωρών στον παγκόσμιο χάρτη (δεξιά) _____ 49*
Σχήμα 19 *- Ασαφείς κανόνες για τον υπολογισμό κινδύνου (αριστερά) και παραδείγματα εκτιμώμενων βαθμών κινδύνου 25 ημερών (με αρχή τις 20/08/2020) για διαφορετικές χώρες (δεξιά) _____ 51*
Σχήμα 20 *- Στιγμιότυπο του χάρτη εκτίμησης κινδύνου της πανδημίας COVID-19 στις ΗΠΑ μέσω της εφαρμογής α-Satellite (14-02-2021) _____ 52*
Σχήμα 21 *- Ορισμός της ευαισθησίας (sensitivity) και της ειδικότητας (specificity) σε προβλήματα απόφασης 55*
Σχήμα 22 *- Διάγραμμα ροής μιας μεθόδου βαθιάς μάθησης για τη διάγνωση COVID-19 με τη χρήση αξονικής τομογραφίας _____ 56*
Σχήμα 23 *- Παράδειγμα αρχιτεκτονικής ενός chatbot με δυνατότητες τεχνητής νοημοσύνης _____ 58*
Σχήμα 24 *- Το chatbot Symptoma στη διάγνωση της COVID-19 _____ 59*
Σχήμα 25 *- Αρχιτεκτονική της εφαρμογής AI4COVID-19 για διάγνωση COVID-19 από δείγματα βήχα _____ 61*
Σχήμα 26 *- Φορητές συσκευές ανίχνευσης συμπτωμάτων COVID-19 _____ 62*
Σχήμα 27 *- Μια αρχιτεκτονική για την πρόβλεψη της σοβαρότητας της COVID-19 βάσει αξονικής τομογραφίας θώρακος _____ 66*
Σχήμα 28 *- Παράδειγμα εισόδου και εξόδου της εφαρμογής CHIME για το σχεδιασμό της χωρητικότητας νοσοκομείων _____ 68*
Σχήμα 29 *- Η εφαρμογή CPAS και τα διαφορετικά επίπεδα πρόβλεψής της _____ 69*
Σχήμα 30 *- Αμινοξέα και πρωτεΐνες _____ 73*
Σχήμα 31 *- Η διαδικασία αναδίπλωσης μιας πρωτεΐνης _____ 74*
Σχήμα 32 *- Λειτουργικό διάγραμμα της λύσης AlphaFold 1 από την DeepMind _____ 76*
Σχήμα 33 *- 3D δομή της πρωτεΐνης ORF3a του ιού SARS-CoV-2 (μπλε: πρόβλεψη AlphaFold, πράσινο: πειραματική δομή) _____ 77*
Σχήμα 34 *- Δομή ενός γράφου γνώσης που χρησιμοποιείται για την επαναστόχευση φαρμάκων _____ 80*
Σχήμα 35 *- Μέρος του γράφου γνώσης της BenevolentAI _____ 82*
Σχήμα 36 *- Στιγμιότυπα αποτελεσμάτων της εφαρμογής IBM CogMol για την εύρεση νέων φαρμάκων κατά της COVID-19 _____ 85*
Σχήμα 37 *- Τρισδιάστατη δομή της πρωτεΐνης nsp3 του SARS-CoV-2 με επισήμανση των διαφορετικών επιτόπων - πιθανών στόχων εμβολίων _____ 87*

Σχήμα 38 *- Τυπικός κύκλος ανάπτυξης φαρμάκων στη Φαρμακοβιομηχανία* _____ *89*
Σχήμα 39 *- Παρακολούθηση ασθενών που συμμετέχουν σε κλινικές δοκιμές με τη χρήση wearable αισθητήρων και τεχνητής νοημοσύνης* _____ *91*
Σχήμα 40 *- Πώς εξαπλώνεται η επιδημία της πληροφορίας (infodemic)* _____ *96*
Σχήμα 41 *- Ένα στιγμιότυπο του δείκτη IRI ανά χώρα από την πλατφόρμα COVID19 Infodemics Observatory* _ *97*
Σχήμα 42 *- Διάστημα εμπιστοσύνης [5%, 95%] για το R_0, σύμφωνα με δύο διαφορετικά επιδημιολογικά μοντέλα, για πέντε πλατφόρμες μέσων κοινωνικής δικτύωσης* _____ *99*
Σχήμα 43 *- Μέσος συντελεστής ενίσχυσης για αναξιόπιστες (ε^U) και αξιόπιστες (ε^R) πηγές και συντελεστής σχετικής ενίσχυσης $\alpha=\varepsilon^U/\varepsilon^R$ σε διαφορετικά μέσα κοινωνικής δικτύωσης* _____ *100*
Σχήμα 44 *- Διάγραμμα ροής ενός μοντέλου ανάλυσης συναισθημάτων από μέσα κοινωνικής δικτύωσης* __ *101*
Σχήμα 45 *- Οπτική αναπαράσταση των νοσηλειών ανά 1 εκατομμύριο κατοίκους, ανά πολιτεία στις ΗΠΑ από το «The Covid Tracking Project»* _____ *104*
Σχήμα 46 *- Το σύνολο δεδομένων επιστημονικής βιβλιογραφίας CORD-19* _____ *108*
Σχήμα 47 *- Παράδειγμα ενός χάρτη εύρους μετακίνησης (movement range map) του Facebook* _____ *111*
Σχήμα 48 *- Χάρτης κίνησης του Facebook που δείχνει την ημερήσια κίνηση ανθρώπων μεταξύ του κεντρικού Λονδίνου και των γύρω περιοχών* _____ *112*
Σχήμα 49 *- Στιγμιότυπο από τις Αναφορές Κινητικότητας της Google για την Ελλάδα και συγκεκριμένα για την περιφέρεια Αττικής* _____ *113*
Σχήμα 50 *- Η έννοια της Επεξηγήσιμης Τεχνητής Νοημοσύνης (XAI)* _____ *120*

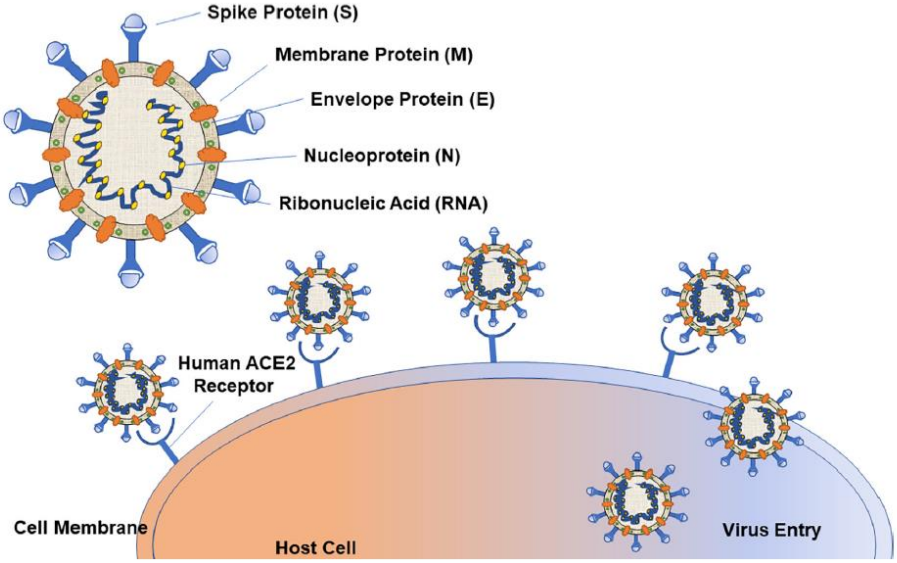
ΚΕΦΑΛΑΙΟ 1: Εισαγωγή

1.1 Επισκόπηση της Πανδημίας COVID-19

1.1.1 Τι Είναι η Ασθένεια COVID-19

Η ασθένεια COVID-19 (Corona Virus Disease 2019) είναι μια μολυσματική ασθένεια που προκαλείται από τον ιό SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus-2). Ο ιός SARS-CoV-2 αποτελεί ένα νέο στέλεχος κορωνοϊού, το οποίο ταυτοποιήθηκε σε ανθρώπους το Δεκέμβριο του 2019. Ο ιός αυτός ανιχνεύθηκε για πρώτη φορά στην πόλη Γουχάν της επαρχίας Χουμπέι στην Κίνα μέσω συρροής κρουσμάτων πνευμονίας, άγνωστης μέχρι τότε αιτιολογίας [1].

Οι κορωνοϊοί είναι μια ευρεία οικογένεια ιών και, παρότι κυκλοφορούν κυρίως μεταξύ των ζώων, σε σπάνιες περιπτώσεις ορισμένοι μπορούν επίσης να μολύνουν και τους ανθρώπους (ανθρώπινοι κορωνοϊοί). Προκαλούν αναπνευστικές, γαστρεντερικές, ηπατικές και νευρολογικές ασθένειες. Υπάρχουν επί του παρόντος επτά ανθρώπινοι κορωνοϊοί. Οι τέσσερις από αυτούς (HCoV-229E, HCoV-NL63, HCoV-OC43 και HCoV-HKU1) ευθύνονται για ήπιες λοιμώξεις του ανώτερου αναπνευστικού συστήματος, με συμπτώματα τύπου κοινού κρυολογήματος και απασχολούν τους ανθρώπους πάνω από έναν αιώνα. Σε παγκόσμιο επίπεδο, οι άνθρωποι μολύνονται με έναν ή περισσότερους από τους πιο πάνω ιούς σε κάποια φάση της ζωής τους. Οι άλλοι τρεις κορωνοϊοί προκαλούν πολύ πιο σοβαρές ασθένειες και έχουν προκύψει πιο πρόσφατα (τα τελευταία 20 χρόνια). Πρόκειται για τους ιούς SARS-CoV ή SARS-CoV-1 (που ήταν υπεύθυνος για το Σοβαρό Οξύ Αναπνευστικό Σύνδρομο το 2002), MERS-CoV (που ήταν υπεύθυνος για το Αναπνευστικό Σύνδρομο της Μέσης Ανατολής το 2012) και SARS-CoV-2 (που είναι υπεύθυνος για την ασθένεια COVID-19). Αξίζει να σημειωθεί ότι η γονιδιακή ακολουθία του νέου ανθρώπινου κορωνοϊού είναι παρόμοια κατά 80% με αυτήν του SARS-CoV, εξού και η ονομασία SARS-CoV-2, και κατά 96,2% με αυτήν του RaTG13, ενός κορωνοϊού των νυχτερίδων [2].



Σχήμα 1 *- Αναπαράσταση της δομής του SARS-CoV-2 και του τρόπου εισόδου του στο κύτταρο του ξενιστή* [3]

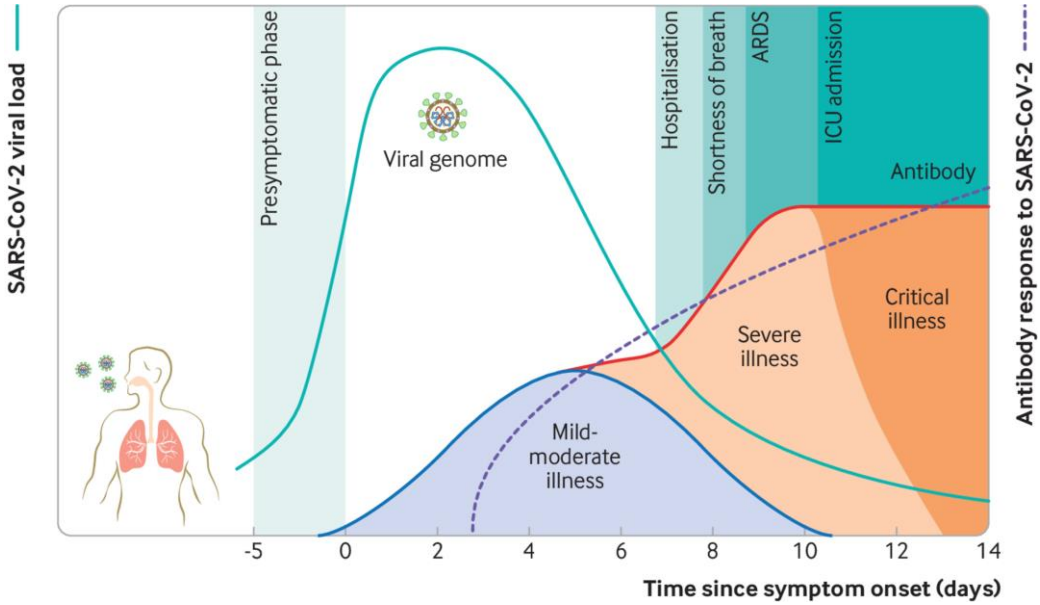
Ο ιός SARS-CoV-2 ανήκει στην οικογένεια των β-κορωνοϊών και διαθέτει έναν μεμβρανώδη εξωτερικό φάκελο, εντός του οποίου φέρει ως γονιδίωμα ένα μονόκλωνο RNA θετικής πολικότητας. Η πρωτεΐνη νουκλεοκαψιδίου (Nucleoprotein - N) προστατεύει το RNA του ιού και μπορεί να χρησιμεύσει ως διαγνωστικό αντιγόνο. Ο φάκελος του ιού είναι επικαλυμμένος με τις δομικές πρωτεΐνες ακίδας (Spike - S), μεμβράνης (Membrane - M) και φακέλου (Envelope - E). Η επιφανειακή γλυκοπρωτεϊνική ακίδα S διαδραματίζει βασικό ρόλο στη λοιμογόνο ικανότητα του ιού, καθώς διευκολύνει τη σύνδεση και κατ' επέκταση την είσοδο του ιού στο κύτταρο του ξενιστή (ανθρώπου). Πιο συγκεκριμένα, η ακίδα S μπορεί να εκτείνεται και να προσκολλάται σε ένα ένζυμο που βρίσκεται στην επιφάνεια των κυττάρων του αναπνευστικού συστήματος. Πρόκειται για το Μετατρεπτικό Ένζυμο της Αγγειοτενσίνης 2 (Angiotensin-Converting Enzyme 2 - ACE2), το οποίο αποτελεί τον ανθρώπινο κυτταρικό υποδοχέα του ιού SARS-CoV-2. Για το λόγο αυτό, η πρωτεϊνική ακίδα S θεωρείται πιθανός στόχος για αντιιικά φάρμακα και εμβόλια [3], [4].

1.1.2 Συμπτώματα, Μετάδοση και Βασικά Μέτρα Προστασίας

Η ασθένεια COVID-19 εκδηλώνεται ως λοίμωξη του αναπνευστικού συστήματος και ποικίλλει σημαντικά ως προς τη σοβαρότητά της. Μπορεί να υπάρξει πλήρης απουσία συμπτωμάτων (ασυμπτωματικοί φορείς) ή μπορεί να εκδηλωθούν ένα ή περισσότερα από τα εξής συμπτώματα: βήχας, πυρετός, δυσκολία στην αναπνοή, καταρροή, βουλωμένη μύτη, πονόλαιμος, πονοκέφαλος, ναυτία, μυϊκός πόνος και πόνος στις αρθρώσεις. Αξίζει να αναφερθεί ότι ένα άλλο, αρκετά κοινό, σύμπτωμα κατά τη διάρκεια της λοίμωξης του ανώτερου αναπνευστικού συστήματος είναι η απώλεια της αίσθησης της όσφρησης και της γεύσης για κάποιο χρονικό διάστημα.

Τα συμπτώματα αυτά μπορεί να εμφανιστούν 2 έως 14 ημέρες μετά την έκθεση στον ιό. Οι περισσότεροι άνθρωποι εκδηλώνουν συμπτώματα μετά από 5 περίπου ημέρες. Η περίοδος κατά την οποία ένα μολυσμένο άτομο είναι ικανό να μεταδώσει τον ιό (μολυσματική περίοδος) μπορεί να αρχίσει έως και δύο ημέρες πριν εμφανιστούν τα συμπτώματα, αλλά οι άνθρωποι είναι πιο μεταδοτικοί κατά τη διάρκεια της συμπτωματικής περιόδου, ακόμη και αν τα συμπτώματα είναι ήπια. Η μολυσματική περίοδος εκτιμάται ότι διαρκεί οκτώ έως δέκα ημέρες σε ήπιες - μέτριες περιπτώσεις και έως δύο εβδομάδες κατά μέσο όρο σε σοβαρές περιπτώσεις.

Συχνά τα συμπτώματα εμφανίζονται σταδιακά. Στην πράξη, οι περισσότεροι άνθρωποι έχουν ήπια - μέτρια συμπτώματα και μπορούν να αναρρώσουν στο σπίτι χωρίς επαγγελματική ιατρική περίθαλψη. Περίπου 10-15% των νοσούντων περνά μια σοβαρή μορφή της ασθένειας (σοβαρή πνευμονία, δύσπνοια, Σύνδρομο Οξείας Αναπνευστικής Δυσχέρειας) και χρειάζεται να νοσηλευθεί, και περίπου 2% περιέρχεται σε κρίσιμη κατάσταση (εισαγωγή σε μονάδα εντατικής θεραπείας) [5]. Στις ομάδες υψηλού κινδύνου περιλαμβάνονται οι ηλικιωμένοι καθώς και άνθρωποι όλων των ηλικιών με υποκείμενες παθολογικές καταστάσεις όπως η παχυσαρκία, η υπέρταση, ο διαβήτης, οι καρδιαγγειακές παθήσεις, οι χρόνιες αναπνευστικές παθήσεις ή το εξασθενημένο ανοσοποιητικό σύστημα.



Σχήμα 2 *- Μόλυνση από τον ιό SARS-CoV-2 – Ιικό Φορτίο και Απόκριση Αντισωμάτων σε σχέση με το χρόνο για Ήπιες - Μέτριες, Σοβαρές και Κρίσιμες περιπτώσεις συμπτωματικών ασθενών* [3]

Ο νέος ιός που προκαλεί την ασθένεια COVID-19 μεταδίδεται μέσω αναπνευστικών σταγονιδίων, τα οποία παράγονται όταν ένα μολυσμένο άτομο βήχει, φτερνίζεται ή απλώς μιλά. Όσοι βρίσκονται σε στενή επαφή (απόσταση μικρότερη των δύο μέτρων) εισπνέουν τα σταγονίδια αυτά και ενδέχεται να μολυνθούν με τον ιό. Η μετάδοση μέσω αναπνευστικών σταγονιδίων εκτιμάται ότι είναι ο κύριος τρόπος εξάπλωσης του ιού. Επίσης, υπάρχει και ένας άλλος, έμμεσος, τρόπος μετάδοσης του ιού, όταν σταγονίδια που φέρουν τον ιό επικάθονται σε επιφάνειες και αντικείμενα, με τα οποία θα έρθουν σε επαφή άλλα άτομα, προκαλώντας τους πιθανώς μόλυνση τη στιγμή που θα αγγίξουν τη μύτη, το στόμα ή τα μάτια τους [1].

Τα βασικά μέτρα για την αποτροπή της μόλυνσης από το νέο κορωνοϊό και της περαιτέρω διασποράς του είναι η χρήση μάσκας, η διατήρηση φυσικής απόστασης τουλάχιστον δύο μέτρων, η αποφυγή συνωστισμών και κοινωνικών συγκεντρώσεων, η εκτέλεση – όπου είναι εφικτό – δραστηριοτήτων σε εξωτερικούς αντί εσωτερικούς χώρους, καθώς και το συχνό πλύσιμο των χεριών.

1.1.3 Πώς η Ασθένεια COVID-19 Έγινε Πανδημία - Ένα Χρονοδιάγραμμα

Παρακάτω περιγράφουμε εν συντομία τη χρονολογική σειρά των πιο σημαντικών γεγονότων, από την καταγραφή των πρώτων κρουσμάτων του νέου ιού στα τέλη του 2019 μέχρι τη στιγμή που ο Παγκόσμιος Οργανισμός Υγείας (Π.Ο.Υ.) [6] κήρυξε την ασθένεια COVID-19 πανδημία [7], [8]. Πρόκειται για την πρώτη πανδημία που προκαλείται από κορωνοϊό.

Ημερομηνία	Περιγραφή
12/12/2019	Στη Γουχάν, υγειονομικοί υπάλληλοι αρχίζουν να ερευνούν μια συρροή κρουσμάτων ιογενούς πνευμονίας. Τελικά, διαπιστώνουν ότι οι περισσότεροι ασθενείς είχαν επισκεφθεί τη Χονδρική Αγορά Θαλασσινών Χουανάν. Η αγορά αυτή είναι γνωστή ως ένα σημείο πώλησης πουλερικών, νυχτερίδων, φιδιών και άλλων άγριων ζώων.

31/12/2019	Κινέζοι αξιωματούχοι από το χώρο της Υγείας ενημερώνουν τον Π.Ο.Υ. για μια συρροή 41 κρουσμάτων μιας μυστηριώδους πνευμονίας. Οι περισσότεροι ασθενείς συνδέονται με τη Χονδρική Αγορά Θαλασσινών Χουανάν.
01/01/2020	Η αγορά θαλασσινών Χουανάν κλείνει.
07/01/2020	Κινέζοι αξιωματούχοι ανακοινώνουν ότι έχουν ταυτοποιήσει σε ασθενείς στη Γουχάν ένα νέο κορωνοϊό, ο οποίος αποτελεί τον αιτιολογικό παράγοντα της πνευμονίας (με άρθρο που δημοσιεύθηκε 17 ημέρες αργότερα) [9].
10/01/2020	Τα δεδομένα της γονιδιακής ακολουθίας του νέου ιού δημοσιεύονται στο Virological.org από ερευνητές του Πανεπιστημίου Fudan της Σαγκάης. Άλλες τρεις γονιδιακές ακολουθίες δημοσιεύονται στην ιστοσελίδα του οργανισμού GISAID.
12/01/2020	Χρησιμοποιώντας τη γονιδιακή ακολουθία του νέου κορωνοϊού, η οποία τίθεται στη διάθεση του Π.Ο.Υ., τα εργαστήρια διαφόρων χωρών αρχίζουν να παράγουν ειδικά διαγνωστικά τεστ PCR. Η κινεζική κυβέρνηση αναφέρει ότι δεν υπάρχουν σαφείς αποδείξεις ότι ο ιός μεταδίδεται εύκολα από άνθρωπο σε άνθρωπο.
13/01/2020	Η Ταϊλάνδη ανακοινώνει την πρώτη μόλυνση από το νέο κορωνοϊό. Πρόκειται για την πρώτη περίπτωση μόλυνσης σε χώρα εκτός της Κίνας.
20/01/2020	Η Κίνα ανακοινώνει τρεις θανάτους και περισσότερες από 200 μολύνσεις. Επίσης, ανιχνεύονται περιστατικά εκτός της επαρχίας Χουμπέι (στο Πεκίνο, τη Σαγκάη και τη Σεντσέν). Αρκετές ασιατικές χώρες αρχίζουν να εισάγουν υποχρεωτικούς ελέγχους στα αεροδρόμια για όλες τις αφίξεις από περιοχές υψηλού κινδύνου της Κίνας.
22/01/2020	Μια αποστολή του κινεζικού γραφείου του Π.Ο.Υ. στη Γουχάν ανακοινώνει ότι υπάρχουν στοιχεία για μετάδοση του ιού από άνθρωπο σε άνθρωπο και τονίζει ότι απαιτείται περισσότερη έρευνα για την πλήρη κατανόηση του τρόπου μετάδοσής του.
23/01/2020	Οι κινεζικές αρχές θέτουν την πόλη Γουχάν σε καραντίνα και κάποιες ημέρες αργότερα κάνουν το ίδιο και στο υπόλοιπο της επαρχίας Χουμπέι. Σε μια επιτροπή που συγκαλείται εκτάκτως, ο Π.Ο.Υ. αναφέρει ότι ο νέος κορωνοϊός δεν αποτελεί ακόμη κατάσταση έκτακτης ανάγκης για τη δημόσια υγεία σε διεθνές επίπεδο.
24/01/2020	Σύμφωνα με μια μελέτη που δημοσιεύθηκε στο επιστημονικό περιοδικό «The Lancet», το πρώτο εργαστηριακά επιβεβαιωμένο κρούσμα στην πόλη Γουχάν ήταν ένας άνδρας, του οποίου τα συμπτώματα ξεκίνησαν την 1η Δεκεμβρίου 2019. Δε βρέθηκε επιδημιολογική σύνδεση με άλλα πρωτογενή κρούσματα. Κανείς από την οικογένειά του δεν αρρώστησε [10]. Επίσης, τουλάχιστον 830 περιπτώσεις μόλυνσης από το νέο κορωνοϊό έχουν διαγνωστεί σε εννέα χώρες: Κίνα, Ιαπωνία, Ταϊλάνδη, Νότια Κορέα, Σιγκαπούρη, Βιετνάμ, Ταϊβάν, Νεπάλ και Ηνωμένες Πολιτείες Αμερικής. Η Γαλλία ανακοινώνει τα τρία πρώτα επιβεβαιωμένα εισαγόμενα κρούσματα, τα οποία αποτελούν και τα πρώτα επίσημα κρούσματα στην Ευρωπαϊκή Ένωση.
30/01/2020	Ο Π.Ο.Υ. κηρύσσει παγκόσμια κατάσταση έκτακτης ανάγκης για τη δημόσια υγεία και συμβουλεύει «όλες οι χώρες πρέπει να είναι προετοιμασμένες για περιορισμό, συμπεριλαμβανομένης της ενεργού επιτήρησης, της έγκαιρης ανίχνευσης, της απομόνωσης και της διαχείρισης των κρουσμάτων, της ιχνηλάτησης επαφών και της πρόληψης της περαιτέρω εξάπλωσης της λοίμωξης από τον ιό 2019-nCoV, και όλοι να μοιράζονται τα πλήρη δεδομένα με τον Π.Ο.Υ.». Η Κίνα ανακοινώνει 7.711 κρούσματα και 170 θανάτους. Ο ιός έχει πλέον εξαπλωθεί σε όλες τις κινεζικές επαρχίες.
02/02/2020	Ανακοινώνεται ο πρώτος θάνατος εκτός Κίνας. Πρόκειται για το θάνατο στις Φιλιππίνες ενός Κινέζου από τη Γουχάν. Δύο ημέρες αργότερα ανακοινώνεται ακόμα ένας θάνατος στο Χονγκ Κονγκ.

11/02/2020	Σε λιγότερο από τρεις εβδομάδες μετά την εισαγωγή μέτρων μαζικής καραντίνας στην Κίνα, ο αριθμός των νέων ημερήσιων κρουσμάτων αρχίζει να μειώνεται. Ο Π.Ο.Υ. ανακοινώνει ότι η νέα μολυσματική ασθένεια θα ονομάζεται COVID-19 και ο νέος ιός που την προκαλεί SARS-CoV-2.
12/02/2020	Τα επιβεβαιωμένα κρούσματα κορωνοϊού αρχίζουν να αυξάνονται στη Νότια Κορέα.
19/02/2020	Αρχίζει η επιδημία κορωνοϊού στο Ιράν.
23/02/2020	Η Ιταλία ανακοινώνει 73 νέα κρούσματα, τα οποία αυξάνουν το σύνολο των επιβεβαιωμένων κρουσμάτων της σε 152, και έναν τρίτο θάνατο. Έτσι, καθίσταται η τρίτη χώρα στον κόσμο σε αριθμό κρουσμάτων, μετά την Κίνα και τη Νότια Κορέα.
03/03/2020	Τα κρούσματα κορωνοϊού αρχίζουν να αυξάνονται απότομα στην Ισπανία.
08/03/2020	Η Ιταλία θέτει σε καραντίνα και τα 60 εκατομμύρια των κατοίκων της.
11/03/2020	Με περισσότερα από 118.000 κρούσματα COVID-19 σε 114 χώρες και 4.291 θανάτους, ο Π.Ο.Υ. κηρύσσει την επιδημία του νέου κορωνοϊού πανδημία.

Κατά τη στιγμή της ολοκλήρωσης της συγγραφής της εργασίας αυτής (28-02-2021), το Κέντρο Επιστήμης και Μηχανικής Συστημάτων στο Πανεπιστήμιο John Hopkins των ΗΠΑ έχει αναφέρει 113.982.329 επιβεβαιωμένα κρούσματα και 2.528.948 θανάτους παγκοσμίως [11].

1.2 Προκλήσεις της Πανδημίας COVID-19 και Ευκαιρία για Αξιοποίηση της Τεχνητής Νοημοσύνης

1.2.1 Γιατί η Πανδημία COVID-19 Είναι Τόσο Επικίνδυνη

Στην ενότητα αυτή θα περιγράψουμε τους βασικούς λόγους για τους οποίους η πανδημία της νόσου COVID-19 θεωρείται εξαιρετικά επικίνδυνη, έχοντας ήδη προκαλέσει πρωτόγνωρα προβλήματα τόσο στα συστήματα υγείας και τους μηχανισμούς αντιμετώπισης μολυσματικών ασθενειών, όσο και σε ολόκληρο το φάσμα της οικονομικής και κοινωνικής δραστηριότητας διεθνώς.

Υψηλή μεταδοτικότητα

Ο κορωνοϊός SARS-CoV-2 είναι ένας νέος ιός για τον οποίο δεν υπάρχει ανοσία στον πληθυσμό. Αυτό πρακτικά σημαίνει ότι όλοι οι άνθρωποι είναι ευάλωτοι σε μόλυνση. Επιπλέον, δεν υπάρχει προς το παρόν μια αρκετά αποτελεσματική αντιιική θεραπεία, όπως έχουμε για παράδειγμα στην περίπτωση της γρίπης, και τα εμβόλια δεν είναι ακόμη ευρέως διαθέσιμα, καθώς μόλις άρχισαν να παράγονται.

Επίσης, αξίζει να αναφερθούμε ιδιαίτερα στους συμπτωματικούς ασθενείς που βρίσκονται στην προσυμπτωματική φάση και τους ασυμπτωματικούς φορείς, οι οποίοι παίζουν σημαντικό ρόλο στη ραγδαία εξάπλωση του ιού. Πιο συγκεκριμένα, στην περίπτωση των συμπτωματικών ασθενών, το ιικό φορτίο στην αναπνευστική τους οδό είναι μέγιστο τη στιγμή της εμφάνισης των συμπτωμάτων ή κατά την πρώτη εβδομάδα της ασθένειας και στη συνέχεια μειώνεται γρήγορα με την πάροδο του χρόνου (βλ. Σχήμα 2). Το γεγονός αυτό συνεπάγεται την υψηλότερη δυνατότητα μετάδοσης του ιού λίγο πριν την έναρξη των συμπτωμάτων ή εντός των πρώτων πέντε ημερών από την εκδήλωσή τους. Αυτό σημαίνει ότι οι άνθρωποι που έχουν μολυνθεί με τον ιό αλλά δεν έχουν εκδηλώσει ακόμη συμπτώματα (προσυμπτωματική φάση), μπορούν να είναι μεταδοτικοί. Αντίστοιχα, οι ασυμπτωματικοί φορείς, λόγω απουσίας συμπτωμάτων, δε θα αναζητήσουν ιατρική φροντίδα και έτσι δε θα διαγνωσθούν, με αποτέλεσμα να μεταδίδουν τον ιό χωρίς να το γνωρίζουν.

Μεγάλη πιθανότητα εκδήλωσης σοβαρής νόσου

Προκαταρκτικά στοιχεία από χώρες της Ευρωπαϊκής Ένωσης δείχνουν ότι περίπου 20-30% των επιβεβαιωμένων περιστατικών COVID-19 χρειάζεται νοσηλεία. Ωστόσο, είναι σημαντικό να σημειωθεί ότι τα άτομα που θα εκδηλώσουν σοβαρότερα συμπτώματα είναι πιο πιθανό να υποβληθούν σε τεστ σε σχέση με τα άτομα που θα εκδηλώσουν λιγότερο σοβαρά ή και καθόλου συμπτώματα. Συνεπώς, το πραγματικό ποσοστό των ατόμων που χρειάζεται εισαγωγή σε νοσοκομείο σε σχέση με τον αριθμό όλων των μολυσμένων ατόμων είναι χαμηλότερο. Πάντως, τα ποσοστά νοσηλείας είναι υψηλότερα για άτομα ηλικίας 60 ετών και άνω, καθώς και για άτομα με υποκείμενες παθολογικές καταστάσεις.

Για να αποκτήσουμε μια εικόνα της επικινδυνότητας της πανδημίας COVID-19 πρέπει να εκτιμήσουμε μια κρίσιμη παράμετρο που αφορά στη θνητότητα του ιού. Για το λόγο αυτό, κάνουμε χρήση των δεικτών IFR και CFR, δύο βασικών δεικτών της Επιδημιολογίας. Ο IFR (Infection Fatality Ratio) είναι ο δείκτης θνητότητας της λοίμωξης και υπολογίζει το ποσοστό θανάτων μεταξύ όλων των μολυσμένων ατόμων. Ο CFR (Case Fatality Ratio) είναι ο δείκτης θνητότητας των επιβεβαιωμένων κρουσμάτων και υπολογίζει το ποσοστό θανάτων μεταξύ όλων των διαγνωσμένων μολυσμένων ατόμων. Προφανώς, ο υπολογισμός του IFR αποτελεί μια σύνθετη διαδικασία στατιστικής έρευνας και εκτίμησης, καθώς δεν είναι εύκολο να μετρηθεί στην πράξη ο αριθμός όσων νόσησαν (συμπεριλαμβανομένων και των ασυμπτωματικών). Πάντως, ισχύει ότι ο IFR είναι μικρότερος από τον CFR λόγω του μεγαλύτερου παρονομαστή του πρώτου δείκτη.

Μια σημαντική μέθοδος που μας βοηθά να εκτιμήσουμε τον πραγματικό αριθμό των μολυσμένων ατόμων είναι η ανίχνευση ενδείξεων έκθεσης σε κάποιον παθογόνο παράγοντα μέσω του εργαστηριακού ελέγχου ενός αντιπροσωπευτικού τυχαίου δείγματος του πληθυσμού. Πολλές τέτοιες έρευνες διεξάγονται για το νέο ιό επί του παρόντος σε όλο τον κόσμο. Και ενώ στην αρχή της πανδημίας είχε προκληθεί τεράστια ανησυχία καθώς υπήρξε μια πρώτη εκτίμηση για ένα CFR 3,4%, κάποιες από τις πιο πάνω έρευνες έχουν μέχρι στιγμής αναδείξει ότι ο πραγματικός αριθμός των κρουσμάτων υποεκτιμάται σημαντικά, δηλαδή στην πραγματικότητα υπάρχουν πολύ περισσότερα κρούσματα από αυτά που καταγράφονται. Με βάση αυτές τις μελέτες, η εκτίμηση του IFR, και άρα της πραγματικής θνητότητας από το νέο κορωνοϊό, συγκλίνει περίπου στο 0,5 με 1% [12]. Όμως, αν αναλογιστούμε το μέγεθος του ανθρώπινου πληθυσμού, ακόμα και ένα IFR 0,5% συνεπάγεται έναν τρομακτικό αριθμό θανάτων. Το γεγονός αυτό αποτυπώνει και το μέγεθος της υγειονομικής κρίσης που έχουμε να αντιμετωπίσουμε παγκοσμίως.

Πρόκληση πίεσης στα συστήματα υγείας

Η πανδημία της νόσου COVID-19 αιφνιδίασε ολόκληρο τον κόσμο και πρωτίστως αποκάλυψε ότι τα συστήματα υγειονομικής περίθαλψης ανά την υφήλιο δεν έχουν σχεδιαστεί με τέτοιο τρόπο ώστε να είναι σε θέση να αντιμετωπίσουν αποτελεσματικά πανδημίες. Μέσα σε λίγες εβδομάδες από την έναρξη της πανδημίας, στην πλειονότητα των πληγεισών χωρών σημειώθηκε τεράστια υπερφόρτωση των νοσοκομείων, σημαντική έλλειψη υγειονομικών πόρων καθώς και υψηλός φόρτος εργασίας των επαγγελματιών υγείας [13]. Αυτό είχε ως αποτέλεσμα να δημιουργηθούν ακόμα και πολύ δυσάρεστες καταστάσεις, όταν οι γιατροί στα νοσοκομεία έπρεπε να επιλέξουν ποιοι ασθενείς θα νοσηλευθούν, ποιοι θα εισέλθουν σε μονάδες εντατικής θεραπείας και ποιοι θα λάβουν υποστήριξη οξυγόνου από αναπνευστήρα. Πολλές χώρες μάλιστα, προχώρησαν σε αναστολή της περίθαλψης και της φροντίδας των μη επειγόντων περιστατικών, γεγονός που επιβάρυνε σημαντικά την υγεία κυρίως ευάλωτων ομάδων ασθενών.

1.2.2 Κίνητρα για Χρήση της Τεχνητής Νοημοσύνης

Όλα τα πιο πάνω χαρακτηριστικά της πανδημίας COVID-19 επιβάλλουν άμεσες, αποτελεσματικές και στοχευμένες δράσεις σε διάφορους τομείς της Υγείας και της Ιατρικής. Οι παραδοσιακές μέθοδοι, που βασίζονται μόνο στην ανθρώπινη γνώση και ικανότητα, αδυνατούν να εκμεταλλευθούν την ποσότητα των δεδομένων που παράγεται καθημερινά σε όλο τον κόσμο. Δημιουργείται, λοιπόν, η ανάγκη για επεξεργασία και κατανόηση αυτού του τεράστιου όγκου δεδομένων με έξυπνες τεχνικές που δε βασίζονται μόνο στις υπάρχουσες γνώσεις και την ανθρώπινη παρέμβαση.

Η Τεχνητή Νοημοσύνη (Artificial Intelligence - AI), όπως θα δούμε με μεγαλύτερη λεπτομέρεια σε επόμενη παράγραφο, είναι ο τομέας εκείνος της επιστήμης των υπολογιστών που έχει στόχο να μιμηθεί τις ανθρώπινες γνωστικές λειτουργίες. Η τεχνητή νοημοσύνη έχει επιφέρει τεράστιες αλλαγές στο χώρο της Υγείας και της Ιατρικής και αναμένεται να διαδραματίσει βασικό ρόλο στη μάχη για την καταπολέμηση της πανδημίας COVID-19 αλλά και άλλων πανδημιών στο μέλλον.

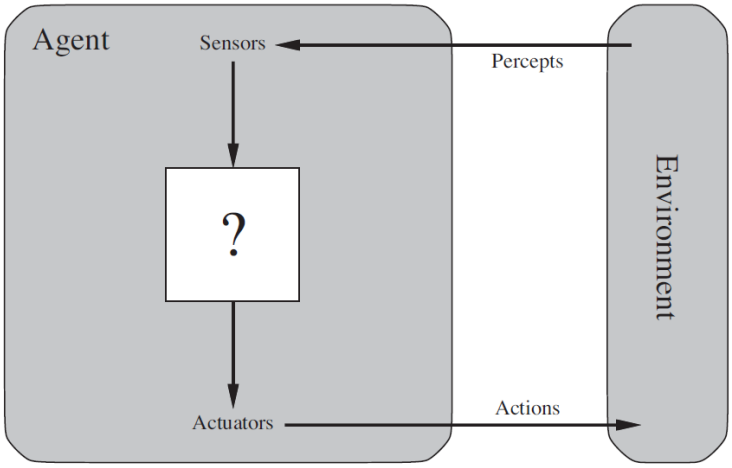
Η τεχνητή νοημοσύνη μπορεί να χρησιμοποιήσει εξελιγμένους αλγορίθμους για να «μάθει» και να αναγνωρίσει χαρακτηριστικά στοιχεία (μοτίβα) από ένα μεγάλο όγκο δεδομένων υγείας και, στη συνέχεια, να χρησιμοποιήσει τις γνώσεις που απέκτησε για να βελτιώσει για παράδειγμα τους διαγνωστικούς ελέγχους. Επίσης, ένα σύστημα τεχνητής νοημοσύνης μπορεί να βοηθήσει τους γιατρούς στην επιλογή του κατάλληλου θεραπευτικού σχήματος για τους ασθενείς, παρέχοντάς τους ενημερωμένες ιατρικές πληροφορίες από επιστημονικά περιοδικά, εγχειρίδια και ιατρικά κείμενα. Επιπλέον, ένα σύστημα τεχνητής νοημοσύνης μπορεί να συμβάλει στη μείωση των διαγνωστικών και θεραπευτικών σφαλμάτων που είναι αναπόφευκτα στην Κλινική Ιατρική. Επιπρόσθετα, ένα σύστημα τεχνητής νοημοσύνης μπορεί να αναλύσει και να εξάγει χρήσιμες πληροφορίες από ένα μεγάλο πληθυσμό ασθενών, και με αυτόν τον τρόπο, να βοηθήσει στην εξαγωγή συμπερασμάτων σε πραγματικό χρόνο σε σχέση με τον κίνδυνο και την πιθανή εξέλιξη της πορείας ενός ασθενούς με βάση την εικόνα του.

Εκτός της Κλινικής Ιατρικής, η τεχνητή νοημοσύνη βρίσκει εφαρμογή και σε άλλους κλάδους της Υγείας και της Ιατρικής, συμπεριλαμβανομένης της Επιδημιολογίας και της Φαρμακολογίας. Οι ραγδαίες εξελίξεις σε πολλούς τομείς των τεχνολογιών τεχνητής νοημοσύνης, όπως της μηχανικής μάθησης, των νευρωνικών δικτύων και της βαθιάς μάθησης, επιτρέπουν τη δημιουργία εφαρμογών και λύσεων που μπορούν να βοηθήσουν σημαντικά στην αντιμετώπιση πολλών προκλήσεων της νέας πανδημίας.

1.3 Θεμελιώδεις Έννοιες της Τεχνητής Νοημοσύνης

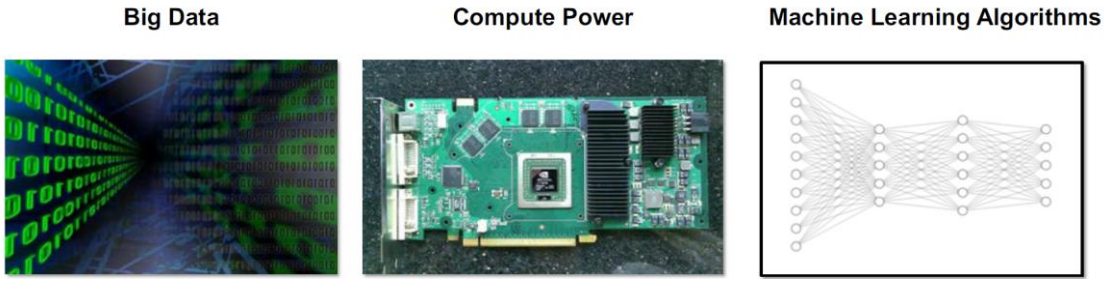
1.3.1 Τεχνητή Νοημοσύνη

Η **Τεχνητή Νοημοσύνη** είναι η νοημοσύνη που μπορούν να επιδείξουν οι μηχανές, σε αντίθεση με τη φυσική νοημοσύνη που επιδεικνύεται από τον άνθρωπο και τα ζώα. Κορυφαία βιβλία για την Τεχνητή Νοημοσύνη ορίζουν το συγκεκριμένο πεδίο ως τη μελέτη των «ευφυών πρακτόρων» (intelligent agents) [14]. Ένας ευφυής πράκτορας είναι οποιαδήποτε αυτόνομη οντότητα που αντιλαμβάνεται το περιβάλλον της μέσω αισθητήρων (sensors), κάνει συλλογισμούς και ενεργεί σε αυτό μέσω μηχανισμών δράσης (actuators). Οι δράσεις που αναλαμβάνει ένας ευφυής πράκτορας μεγιστοποιούν την πιθανότητα να επιτύχει ένα σύνολο από στόχους για τους οποίους έχει κατασκευαστεί. Με απλά λόγια, ο όρος Τεχνητή Νοημοσύνη χρησιμοποιείται συχνά για να περιγράψει την ικανότητα μιας μηχανής (ή υπολογιστή) να αναπαράγει τις γνωστικές (cognitive) λειτουργίες ενός ανθρώπου, όπως για παράδειγμα η μάθηση, ο σχεδιασμός και η επίλυση προβλημάτων.



Σχήμα 3 *- Αλληλεπίδραση ενός ευφυούς πράκτορα με το περιβάλλον του* [14]

Οι πρώτες τεχνολογίες τεχνητής νοημοσύνης έκαναν την εμφάνισή τους περίπου πριν 50 χρόνια. Ωστόσο, η διαθεσιμότητα αναρίθμητων δεδομένων (που ονομάζονται επίσης μεγάλα δεδομένα - big data), η διαρκώς αυξανόμενη και ευρέως διαθέσιμη υπολογιστική ισχύς και η ανάπτυξη αλγορίθμων μηχανικής μάθησης (machine learning) επέτρεψαν την ταχύτατη ανάπτυξη της τεχνητής νοημοσύνης. Οι εφαρμογές τεχνητής νοημοσύνης έχουν γίνει αναπόσπαστο κομμάτι της καθημερινότητάς μας και αναμένεται να παίξουν βασικό ρόλο στον ψηφιακό μετασχηματισμό τομέων, όπως για παράδειγμα της ιατρικής, της δημόσιας διοίκησης, της αυτοκινητοβιομηχανίας, του ηλεκτρονικού εμπορίου, της εκπαίδευσης, των τραπεζικών και χρηματοπιστωτικών υπηρεσιών, των τηλεπικοινωνιών, της εφοδιαστικής αλυσίδας και των logistics, της ταξιδιωτικής βιομηχανίας, της μεταποίησης, των τυχερών παιχνιδιών.

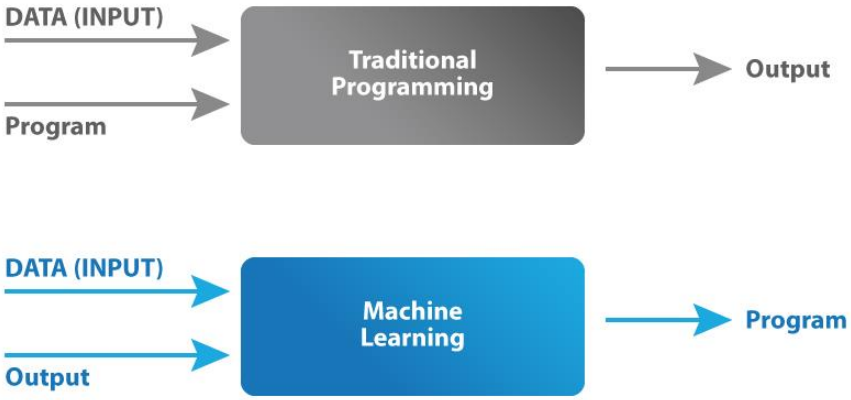


Σχήμα 4 *- Οι τρεις κινητήριες δυνάμεις πίσω από την επανάσταση της τεχνητής νοημοσύνης*

1.3.2 Μηχανική Μάθηση

Η **Μηχανική Μάθηση** (Machine Learning) αποτελεί έναν τομέα της τεχνητής νοημοσύνης, ο οποίος ασχολείται με τη μελέτη και την κατασκευή υπολογιστικών αλγορίθμων που μπορούν να βελτιώνονται αυτόματα μέσω της εμπειρίας. Οι αλγόριθμοι μηχανικής μάθησης δημιουργούν ένα μοντέλο βασισμένο σε δείγματα δεδομένων, γνωστά και ως δεδομένα εκπαίδευσης (training data), προκειμένου να κάνουν προβλέψεις ή να λάβουν αποφάσεις χωρίς να έχουν προγραμματιστεί ρητά για το πώς να το πράξουν. Τέτοιοι αλγόριθμοι χρησιμοποιούνται σε μια μεγάλη ποικιλία εφαρμογών, όπως για παράδειγμα είναι το φιλτράρισμα ανεπιθύμητων μηνυμάτων ηλεκτρονικού ταχυδρομείου (email spam filtering), η οπτική αναγνώριση χαρακτήρων (OCR), οι μηχανές αναζήτησης και η υπολογιστική όραση (computer vision). Σε αυτές τις περιπτώσεις, είναι δύσκολο ή και ανέφικτο να αναπτυχθούν συμβατικοί αλγόριθμοι για την εκτέλεση των απαραίτητων εργασιών. Μια

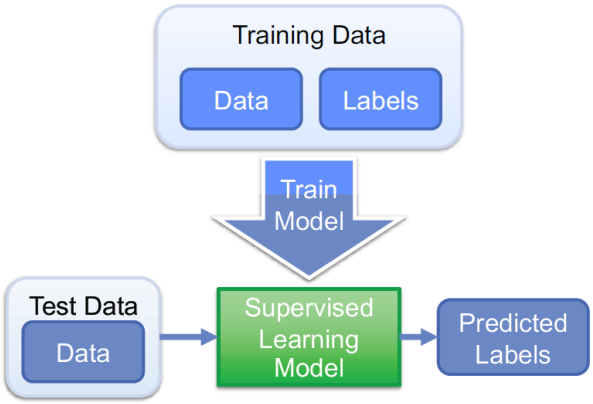
ενδιαφέρουσα σύγκριση ανάμεσα στον παραδοσιακό προγραμματισμό και τη μηχανική μάθηση παρουσιάζεται στο Σχήμα 5.



Σχήμα 5 *- Σύγκριση του παραδοσιακού προγραμματισμού και της μηχανικής μάθησης*

Υπάρχουν τρεις βασικές κατηγορίες μηχανικής μάθησης και κάθε μία εφαρμόζεται σε διαφορετικούς τύπους προβλημάτων [15]:

Η **Επιτηρούμενη Μάθηση** (Supervised Learning) είναι η κατηγορία μηχανικής μάθησης, όπου οι αλγόριθμοι τροφοδοτούνται με ένα σύνολο δεδομένων εκπαίδευσης (δεδομένα και αντίστοιχες ετικέτες), που ονομάζεται «αντικειμενική αλήθεια» («ground truth»), και εκπαιδεύονται με στόχο τη δημιουργία ενός μοντέλου, το οποίο μπορεί να χρησιμοποιηθεί για την πρόβλεψη ετικετών για καινούρια δεδομένα για τα οποία δεν υπάρχουν ετικέτες (test data).



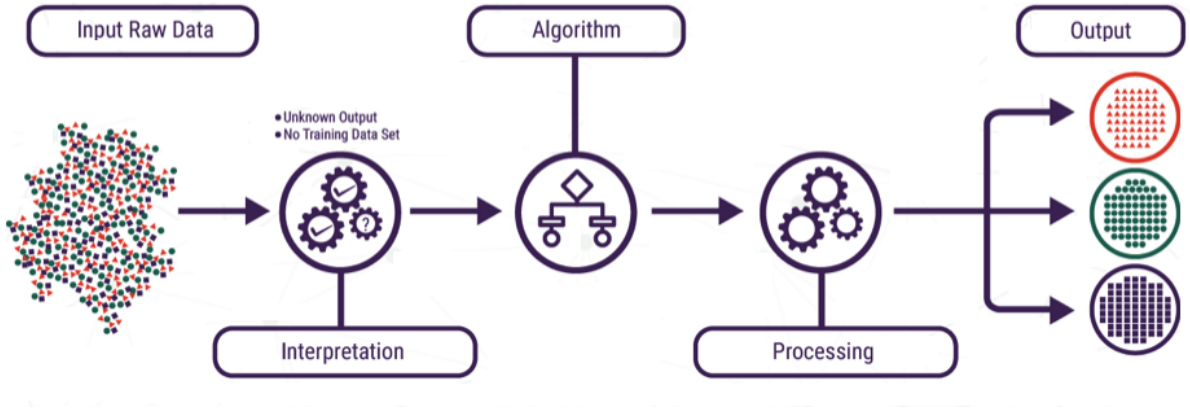
Σχήμα 6 *- Επισκόπηση της επιτηρούμενης μάθησης* [15]

Όπως όλοι οι αλγόριθμοι μηχανικής μάθησης, έτσι και η επιτηρούμενη μάθηση βασίζεται στην εκπαίδευση. Κατά τη διάρκεια της φάσης εκπαίδευσης, το μοντέλο καθοδηγείται από τα δεδομένα εκπαίδευσης για το ποια έξοδο πρέπει να παράξει για κάθε συγκεκριμένη είσοδο. Στη συνέχεια, το εκπαιδευμένο μοντέλο δέχεται ως είσοδο τα δεδομένα δοκιμής με στόχο την επαλήθευση του αποτελέσματος της εκπαίδευσης και τη μέτρηση της ακρίβειας της πρόβλεψης. Είναι σημαντικό τα δεδομένα δοκιμής να είναι διαφορετικά από τα δεδομένα εκπαίδευσης για να διασφαλιστεί ότι το μοντέλο δεν αντλεί απαντήσεις από την προηγούμενη εμπειρία του, και έτσι να μπορεί να γενικεύει

τη γνώση του σε νέα δεδομένα. Στις περισσότερες περιπτώσεις, το μεγαλύτερο μέρος του υπολογιστικού χρόνου δαπανάται στη φάση εκπαίδευσης του μοντέλου. Το μοντέλο που παράγεται είναι ουσιαστικά η αναπαράσταση του τι πιστεύει το σύστημα ότι συνδέει τις εισόδους με τις αντίστοιχες εξόδους.

Γενικά, στην επιτηρούμενη μάθηση εντάσσονται δύο κατηγορίες προβλημάτων: τα προβλήματα ταξινόμησης και τα προβλήματα παλινδρόμησης. Η ταξινόμηση (classification) χρησιμοποιείται για την πρόβλεψη της κλάσης ή της ετικέτας ενός νέου δείγματος (μοντέλα πρόβλεψης διακριτών τάξεων). Παραδείγματα τέτοιων προβλημάτων είναι τα φίλτρα spam (όπου οι είσοδοι είναι emails και οι κλάσεις είναι «spam» και «όχι spam») και η διάγνωση ασθενειών (όπου για παράδειγμα στην περίπτωση της νόσου COVID-19, οι είσοδοι είναι υπολογιστικές τομογραφίες θώρακος και οι κλάσεις είναι «ασθενής με COVID-19» και «υγιής»). Η παλινδρόμηση (regression) χρησιμοποιείται για την πρόβλεψη της μελλοντικής τιμής μιας συνεχούς μεταβλητής (μοντέλα πρόβλεψης αριθμητικών τιμών). Χαρακτηριστικό παράδειγμα προβλήματος παλινδρόμησης είναι η πρόγνωση του καιρού.

Η **Μη Επιτηρούμενη Μάθηση** (Unsupervised Learning) είναι η κατηγορία της μηχανικής μάθησης που έχει ως στόχο την ανακάλυψη μοτίβων (patterns) σε σύνολα μη χαρακτηρισμένων δεδομένων, δηλαδή δεδομένων τα οποία δεν έχουν προηγουμένως ταξινομηθεί σε κατηγορίες ούτε έχουν επισημανθεί με ετικέτες. Ουσιαστικά, ένα σύστημα τεχνητής νοημοσύνης με μη επιτηρούμενη μάθηση ομαδοποιεί/κατηγοριοποιεί τα δεδομένα εισόδου σύμφωνα με ομοιότητες και διαφορές, παρόλο που δεν παρέχονται οι αντίστοιχες ομάδες/κατηγορίες από το σχεδιαστή του συστήματος. Δηλαδή, σε αυτή την κατηγορία μηχανικής μάθησης δεν υπάρχει καμία εξωτερική καθοδήγηση (εκπαίδευση/επιτήρηση).



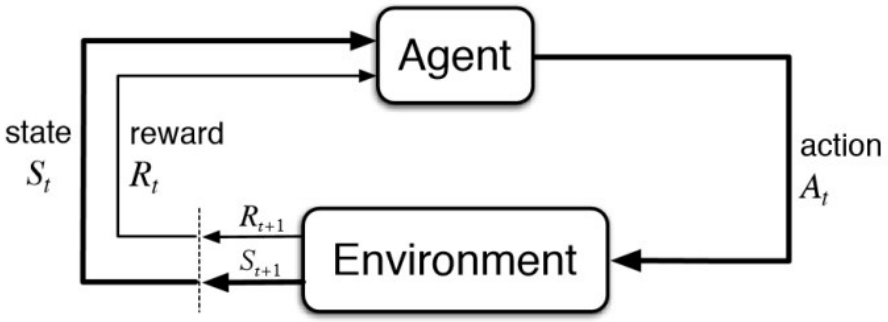
Σχήμα 7 *- Επισκόπηση της μη επιτηρούμενης μάθησης* [16]

Μια δημοφιλής οικογένεια τέτοιων αλγορίθμων είναι οι αλγόριθμοι ομαδοποίησης (clustering algorithms). Αυτοί ομαδοποιούν αντικείμενα του συνόλου δεδομένων, έτσι ώστε τα αντικείμενα μέσα σε μια ομάδα (cluster) να είναι πιο «όμοια» από εκείνα μεταξύ διαφορετικών ομάδων. Ο ορισμός της ομοιότητας εξαρτάται συχνά, σε μεγάλο βαθμό, από την εκάστοτε εφαρμογή. Χαρακτηριστικό παράδειγμα ομαδοποίησης συναντάμε στον τομέα της έρευνας αγοράς (market research), όπου οι καταναλωτές μπορούν να ομαδοποιηθούν με βάση τις προτιμήσεις τους σε προϊόντα.

Πολύ συχνά, και ως μέρος της αρχικής επεξεργασίας των δεδομένων σε μια ευρύτερη αλυσίδα συστημάτων τεχνητής νοημοσύνης, οι αλγόριθμοι μη επιτηρούμενης μάθησης χρησιμοποιούνται για την ανίχνευση των ακραίων τιμών (anomaly ή outlier detection) ή τη μείωση των διαστάσεων των δεδομένων (dimensionality reduction). Στη δεύτερη περίπτωση, τα δεδομένα εισόδου

απλοποιούνται και αντιστοιχίζονται σε ένα χώρο λιγότερων διαστάσεων. Ο χώρος διαστάσεων που προκύπτει εξακολουθεί να είναι επαρκής για την αποτελεσματική διάκριση των δεδομένων. Οι αλγόριθμοι ανίχνευσης ακραίων τιμών βρίσκουν εφαρμογή σε ιατρικά συστήματα παρακολούθησης (system health monitoring), όπου η ανίχνευση μη φυσιολογικών τιμών σε ζωτικές παραμέτρους θα σημάνει συναγερμό, ενώ οι αλγόριθμοι μείωσης διαστάσεων είναι ιδιαίτερα χρήσιμοι σε τομείς που διαχειρίζονται μεγάλο όγκο δεδομένων, όπως είναι η Βιοπληροφορική και η Νευροεπιστήμη.

Η **Ενισχυτική Μάθηση** (Reinforcement Learning) είναι η κατηγορία μηχανικής μάθησης που βασίζεται στην επιβράβευση των επιθυμητών συμπεριφορών ή και την τιμωρία των ανεπιθύμητων. Πιο συγκεκριμένα, ένας πράκτορας ενισχυτικής μάθησης (σύστημα) είναι σε θέση να αντιληφθεί και να ερμηνεύσει το περιβάλλον του, να αναλάβει δράσεις και να μάθει μέσω της δοκιμής και του λάθους (trial and error). Στην ενισχυτική μάθηση, οι προγραμματιστές επινοούν μια μέθοδο ανταμοιβής των επιθυμητών ενεργειών και τιμωρίας των ανεπιθύμητων. Αυτή η μέθοδος αντιστοιχίζει θετικές τιμές στις επιθυμητές ενέργειες για να ενθαρρύνει το σύστημα να τις επαναλάβει και αρνητικές τιμές στις ανεπιθύμητες για να το αποτρέψει από αυτές. Ουσιαστικά, αυτό έχει ως αποτέλεσμα το σύστημα να αναζητήσει μια μακροπρόθεσμη στρατηγική που θα του αποφέρει τη μέγιστη συνολική ανταμοιβή και κατ' επέκταση τη βέλτιστη λύση. Μάλιστα, αυτοί οι μακροπρόθεσμοι στόχοι αποτρέπουν το σύστημα από την προσήλωση σε μικρότερους, επιμέρους στόχους. Με τον καιρό, το σύστημα μαθαίνει να αποφεύγει τις αρνητικές και να αναζητά τις θετικές ενέργειες. Αυτή η μέθοδος μάθησης έχει προσελκύσει τεράστιο επιστημονικό ενδιαφέρον τα τελευταία χρόνια. Παραδείγματα εφαρμογών ενισχυτικής μάθησης μπορεί να συναντήσει κανείς στη ρομποτική, στην εκμάθηση ηλεκτρονικών παιχνιδιών (π.χ. Atari) από μηχανές και στη βελτίωση των επιδόσεων του υπολογιστή σε περίπλοκα στρατηγικά παιχνίδια (π.χ. επιτραπέζιο παιχνίδι Go).



Σχήμα 8 *- Αλληλεπίδραση ενός πράκτορα ενισχυτικής μάθησης με το περιβάλλον του*

1.3.3 Βαθιά Μάθηση και Επεξεργασία Φυσικής Γλώσσας

Η Μάθηση Αναπαράστασης (Representation Learning) ή Μάθηση Χαρακτηριστικών (Feature Learning) είναι ένα σύνολο μεθόδων που επιτρέπει σε μια μηχανή να τροφοδοτείται με ανεπεξέργαστα δεδομένα (raw data) και να ανακαλύπτει αυτόματα τις αναπαραστάσεις ή τα χαρακτηριστικά που απαιτούνται για την αναγνώριση ή την ταξινόμησή τους. Έτσι, η μάθηση αναπαράστασης αντικαθιστά τη χειροκίνητη διαδικασία εύρεσης χαρακτηριστικών (γνωστή και ως feature engineering) και επιτρέπει σε μια μηχανή να μαθαίνει τα χαρακτηριστικά των δεδομένων και ταυτόχρονα να τα χρησιμοποιεί για την εκτέλεση μιας συγκεκριμένης εργασίας.

Η **Βαθιά Μάθηση** (Deep Learning) είναι ένα υποσύνολο της μηχανικής μάθησης και περιλαμβάνει μεθόδους που βασίζονται στη μάθηση αναπαράστασης των δεδομένων με πολλαπλά επίπεδα αναπαράστασης. Τα επίπεδα αυτά λαμβάνονται από τη σύνθεση απλών αλλά μη γραμμικών οντοτήτων, κάθε μία από τις οποίες μετασχηματίζει την αναπαράσταση ενός επιπέδου (ξεκινώντας

από τα ανεπεξέργαστα δεδομένα εισόδου) σε μια αναπαράσταση ενός υψηλότερου, ελαφρώς πιο αφηρημένου (abstract) επιπέδου. Με τη σύνθεση πολλών τέτοιων μετασχηματισμών, το σύστημα βαθιάς μάθησης μπορεί να «μάθει» πολύ σύνθετες λειτουργίες. Για τις εργασίες ομαδοποίησης/ταξινόμησης (classification), τα υψηλότερα επίπεδα αναπαράστασης ενισχύουν πτυχές των δεδομένων εισόδου που είναι σημαντικές για τη διάκριση των χαρακτηριστικών και καταστέλλουν τις μη σημαντικές. Για παράδειγμα, στην επεξεργασία εικόνας (image processing) τα χαμηλότερα επίπεδα αναπαράστασης μπορούν να αναγνωρίσουν τις ακμές (edges) ενώ τα υψηλότερα επίπεδα μπορούν να αναγνωρίσουν ψηφία, γράμματα ή πρόσωπα.

Η βαθιά μάθηση έχει αποδειχθεί ιδιαίτερα αποτελεσματική στην ανακάλυψη περίπλοκων δομών σε δεδομένα πολλών διαστάσεων και ως εκ τούτου βρίσκει εφαρμογή σε πολλούς τομείς των επιστημών και της βιομηχανίας. Παραδείγματα τέτοιων εφαρμογών συναντάμε στην υπολογιστική όραση, την ανάλυση δεδομένων από επιταχυντές σωματιδίων, την εξυπηρέτηση πελατών μέσω chatbots, την αυτόνομη οδήγηση, την ανάλυση ιατρικής εικόνας, την πρόβλεψη της αποτελεσματικότητας πιθανών φαρμακευτικών συστατικών, την ανακατασκευή νευρωνικών κυκλωμάτων του εγκεφάλου (neural circuit reconstruction) και την πρόβλεψη ασθενειών λόγω μεταλλάξεων στο μη-κωδικοποιητικό (non-coding) DNA.

Ακόμα πιο εντυπωσιακά είναι τα ελπιδοφόρα αποτελέσματα της βαθιάς μάθησης στην επεξεργασία της φυσικής γλώσσας. Η **Επεξεργασία Φυσικής Γλώσσας** (Natural Language Processing) είναι ένας διεπιστημονικός τομέας της πληροφορικής, της τεχνητής νοημοσύνης και της υπολογιστικής γλωσσολογίας που ασχολείται με τις αλληλεπιδράσεις μεταξύ των υπολογιστών και των ανθρώπινων (φυσικών) γλωσσών. Πιο συγκεκριμένα, μελετά πώς μπορούν να προγραμματιστούν οι υπολογιστές έτσι ώστε να επεξεργάζονται και να αναλύουν μεγάλες ποσότητες φυσικών γλωσσικών δεδομένων. Μια ενδιαφέρουσα πρόκληση του τομέα αυτού είναι η κατανόηση φυσικής γλώσσας (natural language understanding), που αποτελεί την προσπάθεια να καταστούν ικανοί οι υπολογιστές να εξάγουν νοήματα από ανθρώπινα γλωσσικά δεδομένα. Μάλιστα, η κατανόηση φυσικής γλώσσας θεωρείται ένα από τα πιο δύσκολα προβλήματα της τεχνητής νοημοσύνης (AI-complete problem). Υπάρχει τεράστιο εμπορικό ενδιαφέρον στον τομέα αυτό, λόγω εφαρμογών όπως η μετάφραση κειμένου ή ομιλίας από μια γλώσσα σε μια άλλη (machine translation), η κατηγοριοποίηση εγγράφων (document classification), η αυτόματη περίληψη κειμένου (automatic summarization), η αυτόματη απάντηση ερωτήσεων (question answering) και η ανάλυση συναισθήματος (sentiment analysis ή emotion AI) [17].

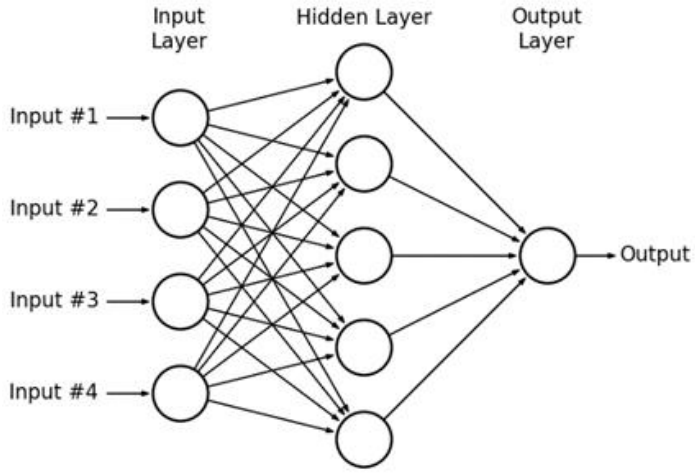
1.3.4 Τεχνητά Νευρωνικά Δίκτυα

Τα **Τεχνητά Νευρωνικά Δίκτυα** (Artificial Neural Networks - ANNs) ή απλά **Νευρωνικά Δίκτυα** (Neural Networks - NNs) είναι μια δημοφιλής τεχνική μηχανικής μάθησης που χρησιμοποιείται κυρίως για επιτηρούμενη μάθηση αλλά μπορεί να εφαρμοστεί και σε προβλήματα μη επιτηρούμενης μάθησης. Μιμούμενα τη λειτουργία των βιολογικών νευρώνων του εγκεφάλου και τη δομή των βιολογικών νευρωνικών δικτύων, τα υπολογιστικά αυτά συστήματα μαθαίνουν μέσα από επαναλαμβανόμενη εκπαίδευση να εκτελούν συγκεκριμένες εργασίες. Ουσιαστικά, τα τεχνητά νευρωνικά δίκτυα προσπαθούν να μιμηθούν τον τρόπο με τον οποίο ο ανθρώπινος εγκέφαλος μαθαίνει να επιλύει υπολογιστικά προβλήματα.

Ένα νευρωνικό δίκτυο αποτελείται από ένα σύνολο διασυνδεδεμένων κόμβων, που ονομάζονται τεχνητοί νευρώνες ή απλά νευρώνες. Οι νευρώνες αλληλεπιδρούν μεταξύ τους μέσω των ακμών τους. Κάθε ακμή, όπως και οι βιολογικές συνάψεις στον ανθρώπινο εγκέφαλο, μπορεί να μεταδώσει ένα σήμα σε άλλους νευρώνες. Κάθε φορά που ένας νευρώνας λαμβάνει ένα τέτοιο σήμα, το

επεξεργάζεται και στη συνέχεια μπορεί να το μεταδώσει σε άλλους νευρώνες που βρίσκονται άμεσα συνδεδεμένοι με αυτόν.

Υπάρχουν δύο τοπολογίες νευρωνικών δικτύων: τα δίκτυα πρόσθιας τροφοδότησης (feedforward neural networks) και τα δίκτυα ανατροφοδότησης (feedback neural networks). Στα δίκτυα πρόσθιας τροφοδότησης η ροή της πληροφορίας γίνεται προς μία μόνο κατεύθυνση, με αποτέλεσμα να μη δημιουργούνται βρόχοι ανατροφοδότησης. Οι είσοδοι και οι έξοδοι είναι πάντα σταθερές. Τα δίκτυα αυτά χρησιμοποιούνται σε προβλήματα αναγνώρισης/ταξινόμησης εικόνας. Αντίθετα, στα δίκτυα ανατροφοδότησης επιτρέπονται οι βρόχοι ανατροφοδότησης.



Σχήμα 9 *- Επισκόπηση ενός απλού τεχνητού νευρωνικού δικτύου*

Το νευρωνικό δίκτυο στο Σχήμα 9 αποτελείται από ένα στρώμα εισόδου, ένα κρυφό στρώμα και ένα στρώμα εξόδου, και αποτελεί ένα παράδειγμα ενός δικτύου πρόσθιας τροφοδότησης, όπου κάθε στρώμα συνδέεται μόνο με το επόμενο χωρίς να σχηματίζονται κύκλοι.

Γενικά, για ένα νευρωνικό δίκτυο k στρωμάτων, η εξίσωση που περιγράφει τη συμπεριφορά εισόδου-εξόδου μεταξύ οποιωνδήποτε δύο στρωμάτων (m και $m+1$) είναι η εξής:

$$y_{\mathrm{m+1}} = f\ (W_{\mathrm{m}} y_{\mathrm{m}} + b_{\mathrm{m}}),$$

όπου το y_m αντιστοιχεί στις εξόδους του στρώματος m, W_m είναι τα βάρη (weights) μεταξύ των στρωμάτων m και $m+1$ και b_m οι αντίστοιχες πολώσεις (biases). Σε αυτή την εξίσωση, η συνάρτηση $f()$ αντιστοιχεί σε μια μη γραμμική συνάρτηση ενεργοποίησης (activation function), όπως για παράδειγμα είναι η σιγμοειδής (sigmoid), ο ανορθωτής (rectifier ή rectified linear unit - ReLU) ή η συνάρτηση υπερβολικής εφαπτομένης (tanh).

Κατά τη διάρκεια της φάσης εκπαίδευσης ενός αλγορίθμου νευρωνικού δικτύου, τα βάρη W και οι πολώσεις b προσαρμόζονται επαναληπτικά προκειμένου να εκφράσουν με αντιπροσωπευτικό τρόπο τη σχέση μεταξύ των εισόδων και των εξόδων. Η παραπάνω διαδικασία επαναλαμβάνεται για όσο το δυνατόν περισσότερα δείγματα εκπαίδευσης (training samples), και όταν φτάσει να πληρείται ένα συγκεκριμένο κριτήριο σύγκλισης (όπως για παράδειγμα η ακρίβεια), το μοντέλο θεωρείται εκπαιδευμένο και μπορεί να χρησιμοποιηθεί για εξαγωγή συμπερασμάτων (inference).

Στην πράξη, είναι σύνηθες να διαιρείται το σύνολο των διαθέσιμων δεδομένων εκπαίδευσης σε δύο υποσύνολα: ένα υποσύνολο εκπαίδευσης (training subset) και ένα υποσύνολο επαλήθευσης (validation subset). Αυτό μπορεί να βοηθήσει στη μείωση της πιθανότητας υπερβολικής προσαρμογής (overfitting), στην οποία το εκπαιδευμένο μοντέλο λειτουργεί ιδιαίτερα καλά στα

δεδομένα της εκπαίδευσης, αλλά δε μπορεί να γενικεύσει τα συμπεράσματά του σε νέα δεδομένα. Όπως και σε άλλα μοντέλα μηχανικής μάθησης, το μεγαλύτερο μέρος του υπολογισμού λαμβάνει χώρα στη φάση εκπαίδευσης. Μόλις ένα μοντέλο εκπαιδευτεί και επαληθευτεί, είναι έτοιμο να χρησιμοποιηθεί για εξαγωγή συμπερασμάτων. Δηλαδή, το εκπαιδευμένο μοντέλο μπορεί να εφαρμοστεί σε νέα δεδομένα, προκειμένου να εξαγάγει συμπεράσματα και προβλέψεις.

Τα **Βαθιά Νευρωνικά Δίκτυα** (Deep Neural Networks - DNNs) είναι τα νευρωνικά δίκτυα που χαρακτηρίζονται από ένα μεγάλο αριθμό κρυφών στρωμάτων. Τα DNNs έχουν σημειώσει μεγάλη επιτυχία την τελευταία δεκαετία σε μια πληθώρα εφαρμογών και υποστηρίζονται από μια σειρά εργαλείων λογισμικού και υλικού. Η βαθιά μάθηση, που εξετάστηκε στην προηγούμενη ενότητα, μπορεί να οριστεί και ως ο τομέας της μηχανικής μάθησης που χρησιμοποιεί βαθιά νευρωνικά δίκτυα. Όπως έχει ήδη αναφερθεί, μια πολύ δημοφιλής εφαρμογή της βαθιάς μάθησης είναι η αναγνώριση/ταξινόμηση εικόνας. Τα νευρωνικά δίκτυα με πολλά στρώματα μπορούν να μοντελοποιήσουν πολύπλοκες σχέσεις μεταξύ της εισόδου και της εξόδου, αλλά προκειμένου να επιτύχουν βέλτιστη απόδοση απαιτούν περισσότερα δεδομένα, υπολογιστικό χρόνο ή προηγμένο αρχιτεκτονικό σχεδιασμό (σε σχέση με παραδοσιακές τεχνικές μηχανικής μάθησης). Όμως, οι σύγχρονες αρχιτεκτονικές επεξεργαστών μπορούν να παρέχουν την απαιτούμενη υπολογιστική ισχύ για τη βαθιά μάθηση.

Τα **Συνελικτικά Νευρωνικά Δίκτυα** (Convolutional Neural Networks - CNNs) είναι μια δημοφιλής αρχιτεκτονική DNNs που έχει σχεδιαστεί για την επεξεργασία δεδομένων που είναι οργανωμένα με ομοιόμορφη χωροταξία (όπως τα στοιχεία μιας ακολουθίας ή τα pixels μιας εικόνας). Οι δημιουργοί των CNNs εμπνεύστηκαν από βιολογικές διεργασίες του οπτικού φλοιού ανθρώπινων και ζωικών εγκεφάλων, όπου ο κάθε νευρώνας ενεργοποιείται μέσα στο δεκτικό του πεδίο όταν διεγείρεται. Ένα δεκτικό πεδίο ενός νευρώνα καλύπτει μια συγκεκριμένη υποπεριοχή του συνολικού οπτικού πεδίου και έτσι ολόκληρο το οπτικό πεδίο μπορεί να καταγραφεί από τη μερική επικάλυψη των δεκτικών πεδίων. Ένα CNN αποτελείται από πολλαπλά στρώματα, όπου κάθε νευρώνας ενός επόμενου (υψηλότερου) στρώματος συνδέεται με ένα υποσύνολο νευρώνων στο προηγούμενο (χαμηλότερο) στρώμα. Αυτό επιτρέπει στο δεκτικό πεδίο ενός νευρώνα ενός υψηλότερου στρώματος να καλύψει ένα μεγαλύτερο μέρος της εικόνας σε σύγκριση με αυτό ενός χαμηλότερου στρώματος.

Το υψηλότερο στρώμα είναι σε θέση να μάθει πιο αφηρημένα (abstract) χαρακτηριστικά των εικόνων σε σχέση με το χαμηλότερο στρώμα, λαμβάνοντας υπόψη τις χωρικές σχέσεις μεταξύ των διαφόρων δεκτικών πεδίων. Αυτή η χρήση των δεκτικών πεδίων επιτρέπει στα CNNs να αναγνωρίζουν οπτικά μοτίβα και να καταγράφουν χαρακτηριστικά από εικόνες χωρίς προηγούμενη γνώση ή χωρίς τον ορισμό χαρακτηριστικών (features) από ανθρώπους, όπως γίνεται στις παραδοσιακές προσεγγίσεις μηχανικής μάθησης. Αυτή η γενική αρχή εφαρμόζεται σε διάφορες αρχιτεκτονικές CNNs, αν και η κάθε μία επιμέρους αρχιτεκτονική μπορεί να διαφέρει ως προς τον αριθμό των στρωμάτων, τον αριθμό των νευρώνων σε κάθε στρώμα, την επιλογή των συναρτήσεων ενεργοποίησης και απωλειών (activation και loss functions), καθώς και την επιλογή αλγορίθμων κανονικοποίησης και εκπαίδευσης.

Αξιοποιώντας την υπολογιστική ισχύ, τα μεγάλα σύνολα δεδομένων και τα CNNs, η βαθιά μάθηση έχει μεταμορφώσει ολόκληρο το πεδίο της υπολογιστικής όρασης (computer vision), που όπως θα δούμε και στη συνέχεια, έφερε με τη σειρά της μια πραγματική επανάσταση στην ανάλυση ιατρικής εικόνας.

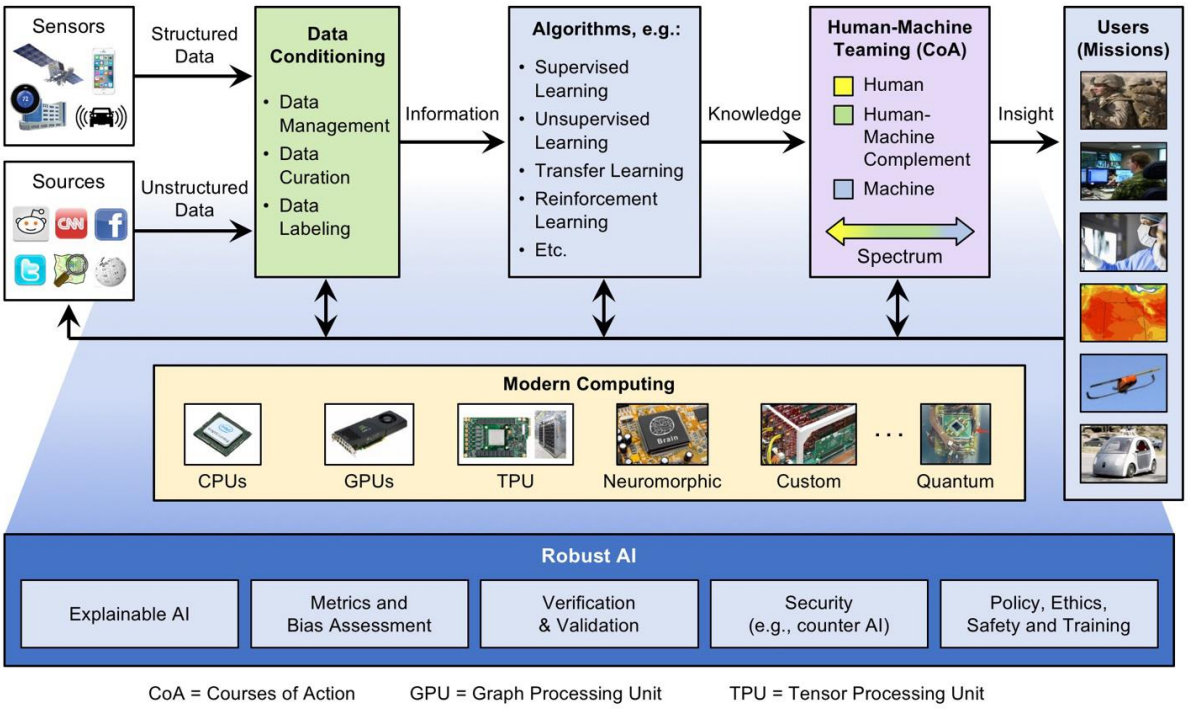
1.3.5 Big Data και Σύγχρονες Αρχιτεκτονικές Τεχνητής Νοημοσύνης

Ο όρος «Μεγάλα Δεδομένα» (Big Data) αναφέρεται σε σύνολα δεδομένων που είναι πολύ μεγάλα και εξαιρετικά πολύπλοκα, με αποτέλεσμα να μην μπορούν να αντιμετωπιστούν με τις παραδοσιακές

τεχνικές επεξεργασίας δεδομένων. Τα Big Data χαρακτηρίζονται συχνά από τα εξής 3 V's: το μεγάλο όγκο (Volume) των δεδομένων, την ευρεία ποικιλία (Variety) των τύπων δεδομένων και τη μεγάλη ταχύτητα (Velocity) με την οποία τα δεδομένα παράγονται, συλλέγονται και υποβάλλονται σε επεξεργασία. Αν και τα Big Data δεν ισοδυναμούν με συγκεκριμένο όγκο δεδομένων, οι εφαρμογές μεγάλων δεδομένων συχνά περιλαμβάνουν terabytes (TB), petabytes (PB), ακόμα και exabytes (EB) δεδομένων που συλλέγονται και υπόκεινται σε επεξεργασία σε ένα συγκεκριμένο χρονικό διάστημα.

Τα Big Data μπορεί να προέρχονται από πολλές διαφορετικές πηγές, όπως συστήματα επιχειρηματικών συναλλαγών, βάσεις δεδομένων πελατών, ιατρικά αρχεία, αρχεία καταγραφής της περιήγησης στο διαδίκτυο (clickstreams), κινητές εφαρμογές, κοινωνικά δίκτυα, online αποθετήρια (repositories) επιστημονικής έρευνας, δεδομένα που δημιουργούνται από μηχανές και αισθητήρες σε πραγματικό χρόνο και τα οποία χρησιμοποιούνται στο «Διαδίκτυο των Πραγμάτων» (Internet of Things), και πολλές άλλες. Επίσης, τα Big Data μπορεί να είναι διαφορετικού τύπου και συνήθως κατατάσσονται σε τρεις μεγάλες κατηγορίες: δομημένα (όπως αυτά που βρίσκονται σε βάσεις δεδομένων ή πίνακες υπολογιστικών φύλλων), αδόμητα (όπως αρχεία κειμένου, εικόνας, ήχου και βίντεο) και ημιδομημένα (όπως μια ιστοσελίδα HTML).

Η ανάλυση μεγάλων δεδομένων (Big Data Analytics) είναι ένας τομέας που ασχολείται με τρόπους ανάλυσης και συστηματικής εξαγωγής πληροφοριών και συμπερασμάτων από πηγές Big Data. Οι προκλήσεις στο χειρισμό μεγάλων δεδομένων περιλαμβάνουν την καταγραφή, την αποθήκευση, την ανάλυση, την αναζήτηση, την κοινή χρήση, τη μεταφορά, την απεικόνιση, την ενημέρωση καθώς και το σεβασμό στο απόρρητο και την ιδιωτική φύση ορισμένων πληροφοριών. Για να μπορέσει να ανταπεξέλθει στις προκλήσεις των Big Data, η ανάλυση μεγάλων δεδομένων χρησιμοποιεί συνήθως τεχνικές τεχνητής νοημοσύνης και αλγορίθμους μηχανικής μάθησης για την εξαγωγή χρήσιμων και εφαρμόσιμων συμπερασμάτων.



Σχήμα 10 *- Μια αρχιτεκτονική υψηλού επιπέδου ενός σύγχρονου συστήματος τεχνητής νοημοσύνης με δυνατότητα ανάλυσης Big Data* [15]

Το Σχήμα 10 παρουσιάζει μια ενδεικτική αρχιτεκτονική ενός συστήματος τεχνητής νοημοσύνης με δυνατότητες ανάλυσης Big Data, συμπεριλαμβανομένων όλων των σταδίων που απαιτούνται για την εξαγωγή πληροφορίας, γνώσης και συμπερασμάτων από τα δεδομένα.

Στην αριστερή πλευρά του Σχήματος, έχουμε δεδομένα που προέρχονται από διάφορες δομημένες και αδόμητες πηγές. Αυτά τα ανεπεξέργαστα δεδομένα συχνά τροφοδοτούν ένα υποσύστημα προετοιμασίας (conditioning) δεδομένων, στο οποίο συγχωνεύονται, δομούνται και μετατρέπονται σε πληροφορία. Οι τυπικές λειτουργίες που εκτελούνται στο πλαίσιο αυτού του υποσυστήματος περιλαμβάνουν την προτυποποίηση της μορφής των δεδομένων (standardization), την επισήμανση των δεδομένων (labelling), την ανάδειξη κενών ή ελλιπών δεδομένων καθώς και την εύρεση σφαλμάτων και μεροληψίας (bias) στα δεδομένα.

Οι πληροφορίες που δημιουργούνται από το υποσύστημα προετοιμασίας δεδομένων τροφοδοτούν συνήθως μια σειρά αλγορίθμων μηχανικής μάθησης. Οι αλγόριθμοι αυτοί χρησιμοποιούνται για τον εντοπισμό μοτίβων, την πρόβλεψη νέων γεγονότων, τη συμπλήρωση ελλιπών δεδομένων ή την αναζήτηση ομοιοτήτων μεταξύ των συνόλων δεδομένων. Ουσιαστικά, μετατρέπουν την πληροφορία εισόδου σε χρήσιμη και εφαρμόσιμη γνώση (knowledge). Ο όρος «γνώση» χρησιμοποιείται για να περιγράψει τη μετατροπή της πληροφορίας σε μια αναπαράσταση υψηλότερου επιπέδου, η οποία είναι κατάλληλη για εξαγωγή συμπερασμάτων.

Μετά την παραγωγή της γνώσης από τους αλγορίθμους, είναι σημαντικό να συμπεριληφθεί και ο άνθρωπος στη διαδικασία λήψης αποφάσεων. Αυτό γίνεται στη φάση της συνεργασίας ανθρώπου-μηχανής (Human-Machine Teaming). Αν και υπάρχουν μερικές εφαρμογές που μπορεί να είναι πλήρως αυτόνομες σε ό,τι αφορά τη λήψη αποφάσεων (όπως για παράδειγμα στο email spam filtering), πρόσφατη πρόοδος στον τομέα της τεχνητής νοημοσύνης έχει σημειωθεί σε τομείς όπου ο άνθρωπος είναι μέσα στη διαδικασία λήψης αποφάσεων (human in the loop). Η φάση της συνεργασίας ανθρώπου-μηχανής είναι κρίσιμη για τη σύνδεση των δεδομένων και των αλγορίθμων με τους τελικούς χρήστες αλλά και για την παροχή σε αυτούς στοχευμένων συμπερασμάτων. Η συνεργασία ανθρώπου-μηχανής είναι ουσιαστικά η φάση στην οποία η γνώση μετατρέπεται σε εφαρμόσιμη νοημοσύνη και απτά συμπεράσματα, μέσω της αποτελεσματικής χρήσης των κατάλληλων ανθρώπινων και υπολογιστικών πόρων.

Ως υπόβαθρο σε όλες αυτές τις φάσεις έχουμε την τεχνολογία των σύγχρονων υπολογιστικών συστημάτων που αποτελούνται από μια σειρά διαφορετικών υπολογιστικών μονάδων (όπως CPUs, GPUs και άλλες), η επιλογή των οποίων εξαρτάται κάθε φορά από τη φύση του προς επίλυση προβλήματος. Τέλος, με την έννοια της Ισχυρής (Εύρωστης) Τεχνητής Νοημοσύνης (Robust AI), αναφερόμαστε σε διάφορα επιθυμητά χαρακτηριστικά της Τεχνητής Νοημοσύνης που διασφαλίζουν τη σωστή λειτουργία τέτοιων σύγχρονων συστημάτων, όπως για παράδειγμα είναι η ασφάλεια, η επεξηγησιμότητα (explainability), η ηθική και άλλα, τα οποία θα εξετάσουμε σε μεγαλύτερο βάθος στο Κεφάλαιο 7.

1.3.6 Τεχνητή Νοημοσύνη και Ανάλυση Μεγάλων Δεδομένων στο Χώρο της Υγείας και της Ιατρικής

Η εξέλιξη της τεχνητής νοημοσύνης στο χώρο της υγείας και της ιατρικής

Ο χώρος της υγείας και της ιατρικής αναδείχθηκε πολύ νωρίς ανάμεσα στους πιο ελπιδοφόρους τομείς εφαρμογών της τεχνητής νοημοσύνης. Από τα μέσα του εικοστού αιώνα, οι ερευνητές πρότειναν και ανέπτυξαν πολλά κλινικά συστήματα υποστήριξης αποφάσεων (clinical decision support systems) με βάση αυστηρά ορισμένους κανόνες. Ενδεικτικές εφαρμογές αυτών των πρώτων προσπαθειών αφορούσαν στην ερμηνεία του Ηλεκτροκαρδιογραφήματος (ΗΚΓ), τη διάγνωση

ασθενειών, την επιλογή κατάλληλων θεραπειών καθώς και τον έλεγχο διαγνωστικών υποθέσεων για σύνθετες περιπτώσεις ασθενών [18]. Τα παραπάνω συστήματα, γνωστά και ως expert systems, ήταν δαπανηρά, χρονοβόρα στην κατασκευή και ευάλωτα σε αλλαγές, καθώς απαιτούσαν από τους επιστήμονες σαφείς ορισμούς των κανόνων λήψης αποφάσεων αλλά και τακτικές ενημερώσεις. Επίσης, η απόδοσή τους περιοριζόταν από την υπάρχουσα ιατρική γνώση.

Σε αντίθεση με αυτή την πρώτη γενιά των συστημάτων τεχνητής νοημοσύνης, πρόσφατη σχετική έρευνα εκμεταλλεύθηκε κατάλληλες μεθόδους μηχανικής μάθησης για την ανάλυση δεδομένων υγείας και την ανίχνευση μοτίβων σε αυτά, χωρίς να χρειάζεται εκ των προτέρων ο αυστηρός ορισμός κανόνων απόφασης για κάθε συγκεκριμένη περίπτωση [19]. Γενικότερα, η μηχανική μάθηση μπορεί να προβλέψει και να αναλύσει ασθένειες στα πρώτα στάδια, συμβάλλοντας έτσι στην έγκαιρη και έγκυρη διάγνωση και κατ' επέκταση στη βελτίωση των παρεχόμενων υπηρεσιών υγείας.

Ωστόσο, όπως έχει ήδη αναφερθεί, οι τεχνικές μηχανικής μάθησης απαιτούν μεγάλο όγκο δεδομένων προκειμένου τα αντίστοιχα μοντέλα/συστήματα να εκπαιδευτούν και να έχουν καλή απόδοση. Εξίσου σημαντική είναι και η ικανότητα επεξεργασίας και ανάλυσης αυτών των δεδομένων. Αξίζει να σημειωθεί ότι η αυξανόμενη υιοθέτηση από πολλές χώρες των συστημάτων Ηλεκτρονικού Φακέλου Ασθενών, όχι μόνο επιτάχυνε τη συλλογή κλινικών δεδομένων μεγάλης κλίμακας, αλλά ταυτόχρονα επέτρεψε και την ομαλότερη ενσωμάτωση των συστημάτων τεχνητής νοημοσύνης στην κλινική ροή εργασίας.

Υπάρχουν διάφοροι τρόποι για να ενσωματωθούν τα συστήματα υποστήριξης αποφάσεων στην κλινική ροή εργασίας, συμπεριλαμβανομένων διαφόρων επιπέδων αυτοματοποίησης. Κατ' αρχάς, τα συμβατικά συστήματα υποστήριξης αποφάσεων συλλέγουν πληροφορίες από τον ηλεκτρονικό φάκελο του ασθενούς και παρέχουν εισηγήσεις προς τους γιατρούς, οι οποίοι είναι και εκείνοι που τελικά λαμβάνουν τις αποφάσεις. Σε ένα δεύτερο επίπεδο, συναντάμε τα ολοκληρωμένα (integrated) συστήματα υποστήριξης αποφάσεων, τα οποία μπορούν αυτόματα να ζητούν σχετικές κλινικές πληροφορίες ή να συλλέγουν δεδομένα από τον ηλεκτρονικό φάκελο του ασθενούς, να εμφανίζουν τα αποτελέσματα στους γιατρούς και να τα αποθηκεύουν στον ηλεκτρονικό φάκελο. Και σε αυτή την περίπτωση, οι γιατροί εξακολουθούν να λαμβάνουν τις τελικές αποφάσεις. Τέλος, υπάρχουν και τα πλήρως αυτοματοποιημένα (αυτόνομα) κλινικά συστήματα, όπου ένα σύστημα συλλέγει πληροφορίες, λαμβάνει αποφάσεις και παράγει το αποτέλεσμα στον ηλεκτρονικό φάκελο του ασθενούς αυτόματα.

Τα μεγάλα δεδομένα στο χώρο της υγείας και της ιατρικής

Εξ ορισμού, τα μεγάλα δεδομένα στο χώρο της υγείας και της ιατρικής αναφέρονται σε ψηφιακά σύνολα δεδομένων που είναι τόσο μεγάλα και πολύπλοκα, γεγονός που καθιστά δύσκολη (ή και αδύνατη) τη διαχείρισή τους με παραδοσιακό λογισμικό και υλικό, αλλά και με παραδοσιακά εργαλεία και μεθόδους διαχείρισης δεδομένων. Τα μεγάλα δεδομένα υγείας δημιουργούν προκλήσεις όχι μόνο λόγω του όγκου τους, αλλά και εξαιτίας της ποικιλομορφίας και της ταχύτητας με την οποία πρέπει να γίνει η επεξεργασία και ανάλυσή τους.

Όπως προαναφέραμε, η ευρεία υιοθέτηση του ηλεκτρονικού φακέλου ασθενούς σε πολλά μέρη του κόσμου έχει δημιουργήσει τεράστια σύνολα δεδομένων. Τα δεδομένα του ηλεκτρονικού φακέλου μπορεί να είναι δομημένα, ημιδομημένα ή αδόμητα, διακριτά ή συνεχή και περιλαμβάνουν τα στοιχεία του ασθενούς, κλινικές σημειώσεις, διαγνώσεις, διαγράμματα, πίνακες, συνταγές, εργαστηριακές εξετάσεις, ιατρικές εικόνες (όπως για παράδειγμα μαγνητικές τομογραφίες (MRI), υπερηχογραφήματα, αξονικές τομογραφίες (CT) και άλλες). Κάποια δεδομένα συλλέγονται από εξειδικευμένες ιατρικές συσκευές παρακολούθησης ή από ένα δίκτυο αισθητήρων, όπως για

παράδειγμα είναι οι wearable συσκευές (συσκευές στο σώμα του ασθενούς). Ωστόσο, ακόμη και σήμερα, μεγάλο μέρος αυτού του πλούσιου συνόλου δεδομένων αντιμετωπίζεται σε πολλές περιπτώσεις ως υποπροϊόν των υπηρεσιών υγείας, και όχι τόσο ως βασικό εργαλείο για τη βελτίωση της ποιότητας και της αποτελεσματικότητάς τους [20].

Όμως, δεν είναι μόνο τα δεδομένα του ηλεκτρονικού φακέλου που αποτελούν τις πηγές των μεγάλων δεδομένων στο χώρο της υγείας. Ένα άλλο παράδειγμα είναι οι σύγχρονες τεχνολογίες γονιδιωματικής που παράγουν τεράστιες ποσότητες πρωτογενών δεδομένων σε σχέση με πολύπλοκες βιοχημικές διεργασίες στους ζωντανούς οργανισμούς [21].

Συνεπώς, η διαθεσιμότητα μεγάλης κλίμακας δεδομένων υγείας σε συνδυασμό με τη ραγδαία ανάπτυξη των τεχνολογιών της τεχνητής νοημοσύνης, μπορούν να επιφέρουν πραγματική επανάσταση στο χώρο της υγείας και να επαναπροσδιορίσουν το ρόλο των γιατρών στην καθημερινή πρακτική.

1.4 Δομή της Διπλωματικής Εργασίας

Στο Κεφάλαιο 2 αναφερόμαστε σε εφαρμογές στην Επιδημιολογία και συγκεκριμένα εξετάζουμε εφαρμογές έγκαιρης ανίχνευσης - προειδοποίησης, αυτοματοποιημένης ιχνηλάτησης των επαφών, επιδημιολογικά μοντέλα για την πρόβλεψη και την προσομοίωση της εξάπλωσης, καθώς και εφαρμογές ομαδοποίησης περιοχών και εκτίμησης κινδύνου ανά περιοχή.

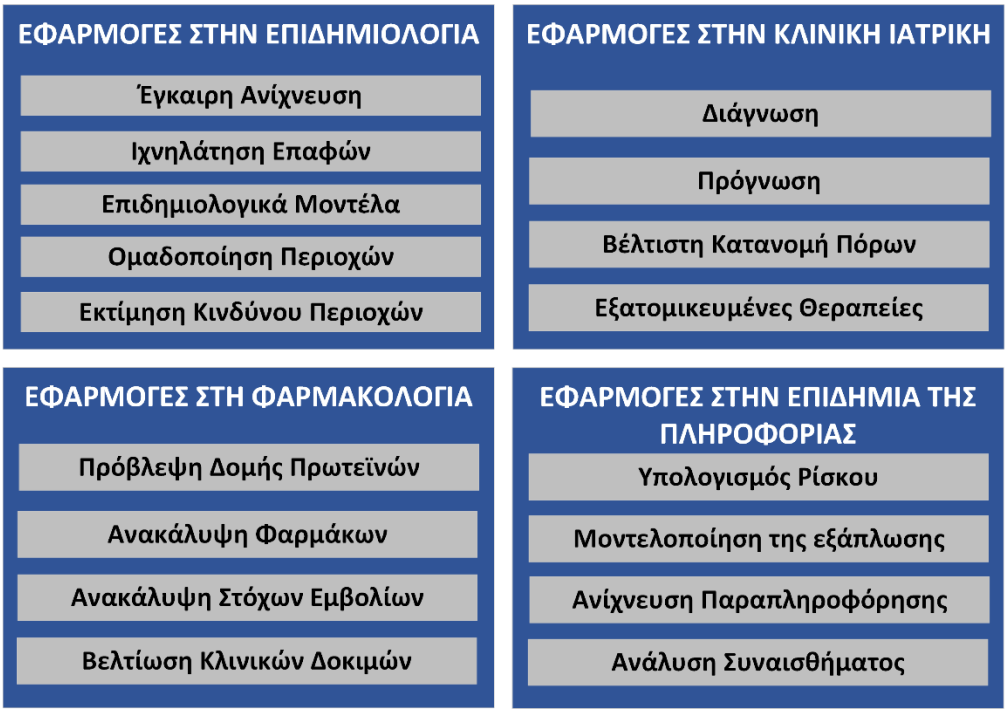
Στο Κεφάλαιο 3 παρουσιάζουμε εφαρμογές στην Κλινική Ιατρική, οι οποίες αφορούν τη διάγνωση της νόσου, την πρόγνωση της πορείας των ασθενών, καθώς και τον υπολογισμό του βέλτιστου τρόπου διαχείρισης των (πεπερασμένων) πόρων των συστημάτων υγείας αλλά και την παροχή εξατομικευμένων θεραπειών στους ασθενείς.

Το Κεφάλαιο 4 επικεντρώνεται σε εφαρμογές στη Φαρμακολογία. Ξεκινάμε με την πρόβλεψη της τρισδιάστατης δομής των πρωτεϊνών με βάση την ακολουθία των αμινοξέων τους, συνεχίζουμε με μεθόδους εύρεσης ασφαλών και αποτελεσματικών φαρμάκων, που περιλαμβάνουν είτε τη χρήση υφιστάμενων φαρμάκων (επαναστόχευση φαρμάκων) είτε την ανακάλυψη εντελώς νέων ουσιών. Το Κεφάλαιο ολοκληρώνεται με τη συνεισφορά της τεχνητής νοημοσύνης στην ανακάλυψη πιθανών στόχων εμβολίων αλλά και στη βελτίωση των κλινικών δοκιμών.

Στο Κεφάλαιο 5 εισάγουμε την έννοια της «Επιδημίας της Πληροφορίας» (Infodemic) και μελετάμε εφαρμογές που αναλύουν δεδομένα από μέσα κοινωνικής δικτύωσης και άλλες online πηγές, με σκοπό τον υπολογισμό του κινδύνου μιας τέτοιας επιδημίας, την ανίχνευση της παραπληροφόρησης και των ψευδών ειδήσεων καθώς και την κατανόηση των συναισθημάτων και των αντιδράσεων της κοινής γνώμης.

Στο Κεφάλαιο 6 περιγράφουμε τους διαφορετικούς τύπους δεδομένων που χρησιμοποιούνται στις πιο πάνω εφαρμογές τεχνητής νοημοσύνης, αναφερόμαστε στις πηγές αυτών των δεδομένων και στα χαρακτηριστικά τους και παρουσιάζουμε μερικά από τα πιο σημαντικά διαθέσιμα ανοικτά σύνολα δεδομένων.

Ολοκληρώνουμε τη διπλωματική εργασία με το Κεφάλαιο 7, παρουσιάζοντας τις προκλήσεις και τους περιορισμούς των εφαρμογών τεχνητής νοημοσύνης στα πλαίσια της πανδημίας COVID-19. Τέλος, προτείνουμε τρόπους για την αντιμετώπισή τους και αναφέρουμε τα συμπεράσματά μας.



Σχήμα 11 *- Κατηγοριοποίηση των εφαρμογών τεχνητής νοημοσύνης στην παρούσα εργασία*

ΚΕΦΑΛΑΙΟ 2: Εφαρμογές στην Επιδημιολογία

2.1 Ο Ρόλος της Τεχνητής Νοημοσύνης στην Επιδημιολογία

Ως Επιδημιολογία ορίζεται η Επιστήμη που μελετά την κατανομή και την εξέλιξη διαφόρων νοσημάτων ή χαρακτηριστικών σε συγκεκριμένους πληθυσμούς, καθώς και τους παράγοντες που τα διαμορφώνουν ή τα επηρεάζουν. Επίσης, εξετάζει χαρακτηριστικά που συνδέονται τόσο με τον περιβαλλοντικό όσο και με το γενετικό έλεγχο. Αντικείμενο της Επιδημιολογίας είναι και η περιγραφική κατανομή των παραγόντων που επηρεάζουν τον πληθυσμό, όπως για παράδειγμα η ηλικία, το φύλο, η φυλή, η γεωγραφική περιοχή και το επάγγελμα («Περιγραφική Επιδημιολογία»). Επιπλέον, οι αιτίες που προκαλούν ασθένειες ή αλλαγές στον ανθρώπινο πληθυσμό είναι αντικείμενο της «Επαγωγικής» ή «Αναλυτικής Επιδημιολογίας». Στην Επιδημιολογία κάθε νόσος περιγράφεται κυρίως από τη συχνότητα εμφάνισης, τη διάρκεια και την εξέλιξή της [22].

Η εξάπλωση του ιού SARS-CoV-2 σε όλο τον κόσμο έχει προκαλέσει τεράστια ανησυχία, με τις εθνικές και τοπικές πολιτικές να αλλάζουν ακόμη και καθημερινά σε πολλά μέρη του κόσμου, καθώς νέες πληροφορίες και προβλέψεις μοντέλων γίνονται διαθέσιμες. Η κατανόηση του τρόπου μετάδοσης του ιού και της πιθανής επίδρασής του σε γεωγραφικές τοποθεσίες με διαφορετικά δημογραφικά και άλλα χαρακτηριστικά είναι επομένως ζωτικής σημασίας για τις αρχές δημόσιας υγείας. Αυτές καλούνται να σχεδιάσουν δράσεις έτσι ώστε να προετοιμαστούν κατάλληλα και να διαχειριστούν με το βέλτιστο τρόπο την πανδημία. Για να γίνει αυτό, απαιτείται πρώτα να κατανοήσουν την τάση της επιδημιολογικής καμπύλης και να αξιολογήσουν το κατά πόσον επιτυγχάνουν ή όχι τον έλεγχο της εξάπλωσης. Η αξιολόγηση της επιδημιολογικής κατάστασης μπορεί επίσης να οδηγήσει σε συμπεράσματα σχετικά με την αποτελεσματικότητα των μέτρων που λαμβάνονται για τη μείωση ή την επιβράδυνση της εξάπλωσης του ιού.

Ο τομέας της Επιδημιολογίας βασιζόταν παραδοσιακά σε κλασικά στατιστικά μοντέλα, τα οποία μπορούν να προσαρμοστούν για την πανδημία COVID-19 και να χρησιμοποιηθούν για επιδημιολογική μοντελοποίηση. Σε αυτή την εργασία, έχουμε επιλέξει να εστιάσουμε σε εφαρμογές της Επιδημιολογίας που χρησιμοποιούν κυρίως τεχνολογίες τεχνητής νοημοσύνης και μηχανικής μάθησης.

Στις ενότητες που ακολουθούν, θα εξετάσουμε αρκετές εφαρμογές που καλύπτουν διάφορες πτυχές της επιδημιολογίας, από την έγκαιρη ανίχνευση και προειδοποίηση, μέχρι τον εντοπισμό και την ιχνηλάτηση επαφών, τα επιδημιολογικά μοντέλα για την πρόβλεψη και την προσομοίωση της εξάπλωσης, καθώς και την ομαδοποίηση περιοχών με βάση κοινά χαρακτηριστικά και την αξιολόγηση κινδύνου ανά περιοχή.

2.2 Έγκαιρη Ανίχνευση και Προειδοποίηση

Η έγκαιρη προειδοποίηση για νέες εστίες ασθενειών μπορεί να βοηθήσει ανθρώπους και κυβερνήσεις να σώσουν ζωές. Όπως αναφέρθηκε και στο εισαγωγικό Κεφάλαιο, η επίσημη προειδοποίηση που εξέδωσε ο Π.Ο.Υ. σχετικά με την εμφάνιση ενός νέου κορωνοϊού έγινε στις 9 Ιανουαρίου του 2020, αφού είχε λάβει αναφορές από επίσημες πηγές στα γραφεία του στην Κίνα [23]. Όπως διαπιστώθηκε αργότερα, ο ιός είχε εντοπιστεί για πρώτη φορά νωρίτερα, και συγκεκριμένα στις 8 Δεκεμβρίου 2019, όταν συνολικά έξι ασθενείς εισήχθησαν σε νοσοκομείο της Γουχάν, όπου υποβλήθηκαν σε θεραπεία και κατόπιν πήραν εξιτήριο. Περισσότερες από είκοσι μέρες αργότερα, και συγκεκριμένα στις 31 Δεκεμβρίου 2019, υγειονομικοί παράγοντες της Κίνας ανέφεραν

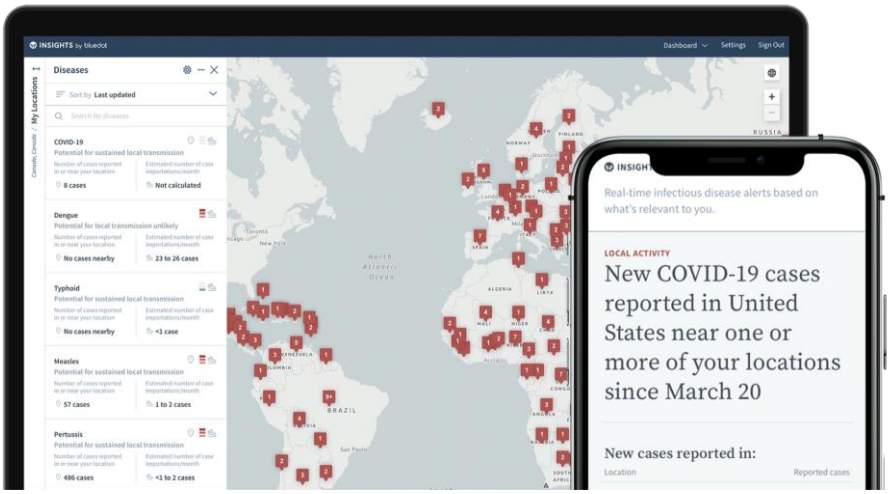
την ύπαρξη περιπτώσεων μιας μυστηριώδους πνευμονίας στον Π.Ο.Υ., πιθανώς γιατί μέχρι τότε δεν είχαν ακόμη κατανοήσει αν είχαν να αντιμετωπίσουν ένα νέο ιό.

Όταν πρωτοεμφανίζεται μια νέα μεταδοτική ασθένεια, είναι αρχικά δύσκολο για τις κυβερνήσεις και τους αξιωματούχους της δημόσιας υγείας να συλλέξουν πληροφορίες γρήγορα και να συντονίσουν μια άμεση αντίδραση. Εκτός από τις επίσημες εκθέσεις της Κίνας και του Π.Ο.Υ. σχετικά με τον ιό, ορισμένες εφαρμογές που βασίζονται στην τεχνητή νοημοσύνη ήταν σε θέση να παρέχουν έγκαιρη ανίχνευση εστιών μετάδοσης του ιού, πριν ακόμη την επίσημη ενημέρωση από τον Π.Ο.Υ.. Στις συγκεκριμένες περιπτώσεις που παρουσιάζουμε και πιο κάτω, οι εφαρμογές αυτές κατάφεραν να προειδοποιήσουν άμεσα τους ταξιδιώτες που προβλέπονταν ότι διατρέχουν υψηλότερο κίνδυνο, με βάση τα στοιχεία πτήσης τους. Στην παρούσα ενότητα παρουσιάζουμε δύο τέτοιες εφαρμογές από τις εταιρείες BlueDot και Metabiota, μαζί με μια επισκόπηση αντίστοιχων προσπαθειών που είχαν γίνει στο παρελθόν, καθώς και μια αναφορά στην τρέχουσα έρευνα για την έγκαιρη ανίχνευση μολυσματικών ασθενειών.

2.2.1 Η Εφαρμογή της BlueDot

Μεταξύ εκείνων που εφάρμοσαν τεχνολογίες βασισμένες στην τεχνητή νοημοσύνη για την έγκαιρη ανίχνευση μολυσματικών ασθενειών και την αντίστοιχη αποστολή ειδοποιήσεων είναι η καναδική εταιρεία BlueDot [24]. Η BlueDot δημιούργησε ένα λογισμικό παρακολούθησης για την «πρόβλεψη των εστιών μολυσματικών ασθενειών, χρησιμοποιώντας την ανθρώπινη και την τεχνητή νοημοσύνη», όπως αναφέρει στην ιστοσελίδα της. Σε σχέση με την τρέχουσα πανδημία, η BlueDot εξέδωσε προειδοποιήσεις προς τους πελάτες / συνδρομητές της στις 31 Δεκεμβρίου του 2019, εννέα ημέρες πριν από την επίσημη ανακοίνωση του Π.Ο.Υ..

Ο μηχανισμός που έχει αναπτύξει συλλέγει δεδομένα από ειδησεογραφικούς ιστοτόπους διαφόρων χωρών σε διαφορετικές γλώσσες, πληροφορίες σχετικές με αεροπορικά εισιτήρια, δίκτυα σχετικά με γνωστούς φορείς ζωικών και φυτικών ασθενειών, καθώς και επίσημες ανακοινώσεις φορέων υγείας και άλλων υπηρεσιών, για την ανίχνευση νέων εστιών μολυσματικών ασθενειών. Αυτές οι πληροφορίες χρησιμοποιήθηκαν για την πρόβλεψη των περιοχών που ήταν επιρρεπείς στην εξάπλωση του ιού, με αφετηρία τις πληγείσες περιοχές στην Κίνα. Επίσης, η BlueDot, βασιζόμενη σε δεδομένα αεροπορικών ταξιδιών από τη Γουχάν, ήταν σε θέση να προβλέψει με ακρίβεια τις πόλεις που διέτρεχαν υψηλό κίνδυνο εμφάνισης νέων εστιών του ιού [25], [26].



Σχήμα 12 *- Διεπαφή web και εφαρμογή για κινητά τηλέφωνα του BlueDot Insights* [24]

Για να μπορέσει η BlueDot να παράξει προβλέψεις της εξάπλωσης και των νέων εστιών των μολυσματικών ασθενειών, βασίζεται σε μεγάλο βαθμό σε τεχνολογίες τεχνητής νοημοσύνης. Χρησιμοποιώντας διάφορους αλγορίθμους επεξεργασίας φυσικής γλώσσας, η εταιρεία συγκεντρώνει καθημερινά δεδομένα από ένα ευρύ φάσμα πηγών, συμπεριλαμβανομένων ιστοσελίδων 10.000 επίσημων υπηρεσιών και μέσων ενημέρωσης, σε περισσότερες από 60 διαφορετικές γλώσσες. Το σύνολο δεδομένων που αναλύει και επεξεργάζεται περιλαμβάνει επίσης πληροφορίες σχετικές με την πυκνότητα του πληθυσμού, οι οποίες αντλούνται από πηγές όπως οι εθνικές απογραφές, το World Factbook και εθνικές στατιστικές εκθέσεις, καθώς και δεδομένα κλιματικών συνθηκών σε πραγματικό χρόνο [27].

Αφού συλλέξει πληροφορίες από όλα τα πιο πάνω σύνολα δεδομένων, η εταιρεία χρησιμοποιεί στη συνέχεια εργαλεία φιλτραρίσματος για να περιορίσει τις περιοχές ενδιαφέροντος, όπως επίσης και μεθόδους ομαδοποίησης (clustering), για να προβεί στη γρήγορη ανακάλυψη περιοχών που θεωρεί ως hot spots ή cold spots αλλά και περιοχών που εμφανίζουν ασυνήθιστη συμπεριφορά (spatial outliers). Η εταιρεία εφαρμόζει επίσης τεχνολογίες μηχανικής μάθησης και συγκεκριμένα επεξεργασία φυσικής γλώσσας για την εκπαίδευση του συστήματός της, έτσι ώστε να είναι σε θέση να στέλνει τακτικές ειδοποιήσεις στους πελάτες της, ειδικά σε περιπτώσεις μιας νέας άγνωστης νόσου. Οι προειδοποιήσεις περιλαμβάνουν μια εκτίμηση του κινδύνου και τους πιθανούς προορισμούς όπου ενδέχεται να εμφανιστούν νέες εστίες μετάδοσης. Η φάση εκπαίδευσης του συστήματος περιλαμβάνει τη χρήση ενός μοντέλου εκτίμησης κινδύνου, το οποίο εκμεταλλεύεται τα μεγάλα σύνολα δεδομένων για να μπορέσει στη συνέχεια να προβεί στις κατάλληλες προβλέψεις.

Μόλις ολοκληρωθεί η αυτοματοποιημένη επεξεργασία των δεδομένων, αναλαμβάνουν δράση οι άνθρωποι. Συγκεκριμένα, εξειδικευμένοι επιδημιολόγοι ελέγχουν ότι τα συμπεράσματα που παρήγαγε το σύστημα τεχνητής νοημοσύνης έχουν νόημα από επιστημονική άποψη και στη συνέχεια, συντάσσονται εκθέσεις που αποστέλλονται σε πελάτες της εταιρείας, όπως δημόσιοι οργανισμοί, νοσοκομεία, καθώς και επιχειρήσεις.

Αξίζει να σημειωθεί ότι ακόμη και πριν από την τρέχουσα πανδημία, η BlueDot είχε επιβεβαιώσει την αποτελεσματικότητα της τεχνολογίας της τόσο κατά τη διάρκεια της πανδημίας γρίπης H1N1 το 2009, όσο και κατά την εξάπλωση του ιού Έμπολα το 2014 στη Δυτική Αφρική. Επίσης, το 2016 στην επιδημία του ιού Ζίκα στη Φλόριντα των ΗΠΑ, κατάφερε να προβλέψει την εμφάνιση του ιού έξι μήνες πριν εμφανιστούν τα πρώτα κρούσματα.

Τέλος, ένας από τους βασικούς στόχους της BlueDot είναι, μέσω της χρήσης κατάλληλων μοντέλων πρόβλεψης, να εξάγει πληροφορίες που θα είναι χρήσιμες στους επαγγελματίες υγείας για την έγκαιρη διάγνωση και πιθανή απομόνωση μολυσμένων και δυνητικά μεταδοτικών ατόμων, συνεισφέροντας έτσι στον περιορισμό της εξάπλωσης ενός νέου ιού.

2.2.2 Η Εφαρμογή της Metabiota

Μια άλλη εταιρεία που χρησιμοποιεί δυνατότητες τεχνητής νοημοσύνης είναι η Metabiota [28], με βάση το Σαν Φρανσίσκο στις ΗΠΑ. Στην περίπτωση της πανδημίας COVID-19, μέσω ανάλυσης μεγάλων δεδομένων από αεροπορικές πτήσεις, κατάφερε να προβλέψει με ακρίβεια ότι χώρες όπως η Ιαπωνία, η Ταϊλάνδη, η Ταϊβάν και η Νότια Κορέα διέτρεχαν κίνδυνο να αποτελέσουν νέες εστίες κορωνοϊού. Μάλιστα, οι προβλέψεις αυτές έγιναν αρκετές ημέρες πριν ανακοινωθούν τα πρώτα επιβεβαιωμένα κρούσματα στις χώρες αυτές [29].

Η Metabiota βασίζεται σε τεχνολογίες ανάλυσης μεγάλων δεδομένων και αλγορίθμους επεξεργασίας φυσικής γλώσσας για να κάνει προβλέψεις που αφορούν σε νέες εστίες μολυσματικών ασθενειών, την εξάπλωσή τους, καθώς και την εκτίμηση της σοβαρότητας της τρέχουσας κατάστασης [30]. Η

εταιρεία συγκεντρώνει μεγάλες ποσότητες δεδομένων τόσο από επίσημες όσο και από ανεπίσημες πηγές, συμπεριλαμβανομένων βιολογικών, κοινωνικοοικονομικών, πολιτικών και περιβαλλοντικών δεδομένων. Χρησιμοποιώντας προηγμένες μεθόδους ανάλυσης και οπτικοποίησης (visualization), μπορεί να προβεί σε προβλέψεις και να τις παρουσιάσει μέσω αντίστοιχων γραφημάτων.

Σε αντίθεση με την BlueDot, που ελάχιστα εξαρτάται από δεδομένα μέσων κοινωνικής δικτύωσης, η Metabiota βασίζεται σε μεγάλο βαθμό στη συλλογή και την ανάλυση των δεδομένων αυτών, που συνδυάζονται με δεδομένα μέσων ενημέρωσης, για να κάνει τις προβλέψεις της. Τα δεδομένα που συλλέγονται από όλες τις πηγές αναλύονται και κατηγοριοποιούνται χρησιμοποιώντας τεχνολογίες ανάλυσης και οπτικοποίησης σε ομάδες (clusters) ανάλογα με τη συχνότητα, τη σοβαρότητα και το χρόνο (διάρκεια) των εξάρσεων. Τα αποτελέσματα αυτά κοινοποιούνται στους πελάτες της ανάλογα με τις πληροφορίες που αυτοί αναζητούν.

Ακόμα, εκμεταλλευόμενη τη δύναμη της τεχνητής νοημοσύνης και της μηχανικής μάθησης, η Metabiota χρησιμοποιεί τα πιο πάνω δεδομένα για να κάνει προβλέψεις σχετικά με το πώς οι επιδημίες επιδρούν στην ανθρώπινη συμπεριφορά, καθώς και με τα επίπεδα φόβου που προκαλούν. Οι συγκεκριμένες προβλέψεις θεωρούνται πολύ χρήσιμες σύμφωνα με την εν λόγω εταιρεία, καθότι ανάμεσα στους κύριους πελάτες της είναι και ασφαλιστικές εταιρείες, που ενδιαφέρονται για τέτοιες πληροφορίες, ιδιαίτερα σε σχέση με το μετριασμό των επενδυτικών κινδύνων. Εκτός από τις ασφαλιστικές εταιρείες, οι προβλέψεις της Metabiota χρησιμοποιούνται από κρατικές υπηρεσίες, μη κερδοσκοπικούς οργανισμούς, καθώς και άλλα ιδρύματα και φορείς, που βασίζονται σε αυτές για να λαμβάνουν ορθότερες αποφάσεις σε περιπτώσεις εμφάνισης και εξάπλωσης μολυσματικών ασθενειών.

2.2.3 Το Παρελθόν, το Παρόν και το Μέλλον στην Πρόβλεψη της Εξάπλωσης Μολυσματικών Ασθενειών

Το 2009, ερευνητές από την Google δημοσίευσαν μια εργασία σχετικά με την ανίχνευση επιδημιών γρίπης χρησιμοποιώντας δεδομένα από μηχανές αναζήτησης [31]. Αυτή η εργασία αποτέλεσε την αρχή του ιστοτόπου Google Flu Trends (GFT). Η βασική ιδέα ήταν η εξής: όταν οι άνθρωποι είναι άρρωστοι με γρίπη, πολλοί από αυτούς αναζητούν πληροφορίες που σχετίζονται με τη γρίπη στο Google. Η υπόθεσή τους ήταν ότι ο όγκος των αντίστοιχων αναζητήσεων θα μπορούσε να παρέχει σχεδόν άμεση εικόνα για το βαθμό συνολικής εξάπλωσης της γρίπης. Στην εργασία αυτή οι ερευνητές απέδειξαν ότι τα δεδομένα αναζήτησης, αν ευθυγραμμιστούν κατάλληλα με τις επίσημες πληροφορίες παρακολούθησης της γρίπης από τα Κέντρα Ελέγχου και Πρόληψης Ασθενειών (Centers for Disease Control and Prevention - CDC), θα μπορούσαν να παράγουν ακριβείς εκτιμήσεις του βαθμού εξάπλωσης της γρίπης δύο εβδομάδες νωρίτερα από τα δεδομένα του CDC, μετατρέποντας ουσιαστικά τις αναζητήσεις των χρηστών σε δυνητικά σωτήρια πληροφορία.

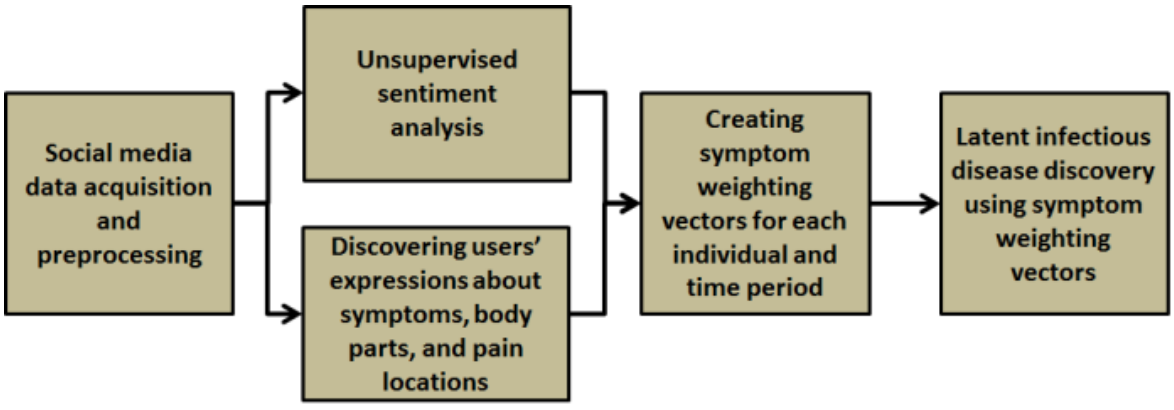
Δυστυχώς, το 2013, το GFT απέτυχε – κατά ένα μεγάλο ποσοστό (περίπου 140%) – να υπολογίσει το μέγεθος της κορύφωσης της περιόδου γρίπης στις ΗΠΑ. Ο λόγος για αυτή τη μεγάλη απόκλιση εξηγήθηκε αργότερα, με βάση ορισμένες ιδιαιτερότητες της ανάλυσης μεγάλων δεδομένων. Συγκεκριμένα, αυτό το είδος των μοντέλων πρόβλεψης πρέπει να επαναβαθμονομείται κάθε χρόνο ώστε να λαμβάνει υπόψη νέες συνθήκες και καταστάσεις. Το συγκεκριμένο έτος υπήρχαν πολλές αναζητήσεις από ανθρώπους που δεν ήταν άρρωστοι, αλλά απλά ήθελαν να βρουν περισσότερες πληροφορίες σχετικά με τη γρίπη, μετά την ενημέρωσή τους από τις ειδήσεις ότι επίκειται επιδημία γρίπης. Αυτό είχε ως αποτέλεσμα την υπερεκτίμηση της κορυφής της γρίπης από το GFT. Από τον Αύγουστο του 2015, ο ιστότοπος Google Flu Trends παραμένει κλειστός και αντ' αυτού η Google παρέχει πληροφορίες σε άλλα εξειδικευμένα ιδρύματα για τη δημιουργία των δικών τους μοντέλων πρόβλεψης. Θα συζητήσουμε ορισμένες από τις δυσκολίες και τις προκλήσεις της τεχνητής

νοημοσύνης και της ανάλυσης μεγάλων δεδομένων, καθώς και προτάσεις για την αντιμετώπισή τους, στο Κεφάλαιο 7.

Γενικότερα, η σάρωση μέσων κοινωνικής δικτύωσης, ειδήσεων και άλλων online πηγών δεδομένων μπορεί να χρησιμοποιηθεί για την ανίχνευση τοπικά εντοπισμένων εστιών ασθενειών πριν ακόμη φτάσουν στο επίπεδο των πανδημιών, κάτι που μπορεί να είναι ιδιαίτερα επωφελές στην προσπάθεια να περιοριστεί ένας ιός κατά τη διάρκεια της αρχικής φάσης εξάπλωσής του. Για παράδειγμα, το Google Trends χρησιμοποιήθηκε στο πρόσφατο παρελθόν για την ανίχνευση της επιδημίας του ιού Zika σε πληθυσμούς χρησιμοποιώντας δυναμικά μοντέλα πρόβλεψης [32]. Η χρήση δεδομένων ιατρικών αρχείων, καθώς και δορυφορικής απεικόνισης (π.χ. αυτοκίνητα που συνωστίζονται γύρω από ένα νοσοκομείο), είναι μερικοί άλλοι τρόποι με τους οποίους η ανάλυση μεγάλων δεδομένων έχει χρησιμοποιηθεί στο παρελθόν για τον εντοπισμό τοπικών εστιών.

Η ανάλυση συναισθήματος είναι μια ακόμα τεχνική που χρησιμοποιεί αλγορίθμους επεξεργασίας φυσικής γλώσσας (για παράδειγμα στα δεδομένα μέσων κοινωνικής δικτύωσης) για την εξαγωγή θετικών και αρνητικών συναισθημάτων. Σε μια ενδιαφέρουσα εργασία [33], οι συγγραφείς προτείνουν ένα μοντέλο μη επιτηρούμενης μάθησης, που έχει την ικανότητα να εντοπίζει λανθάνουσες μολυσματικές ασθένειες από την άντληση δεδομένων μέσων κοινωνικής δικτύωσης. Μια λανθάνουσα μολυσματική ασθένεια ορίζεται ως μια μεταδοτική ασθένεια που δεν έχει ακόμη ορισθεί επίσημα από τα εθνικά ιδρύματα δημόσιας υγείας και δεν έχει γνωστοποιηθεί η ύπαρξή της στο ευρύ κοινό.

Σε αντίθεση με άλλες μεθόδους, η προσέγγιση αυτή δε βασίζεται σε ήδη γνωστές πληροφορίες (όπως τα ονόματα των ασθενειών και τα συμπτώματά τους), αλλά χρησιμοποιεί ένα μη επιτηρούμενο μοντέλο ανάλυσης συναισθήματος. Οι εκφράσεις των χρηστών σε σχέση με τα συμπτώματα, τα μέρη του σώματος και τα σημεία του πόνου αναγνωρίζονται από τα δεδομένα των μέσων κοινωνικής δικτύωσης. Στη συνέχεια, δημιουργούνται διανύσματα στάθμισης συμπτωμάτων για κάθε άτομο και χρονική περίοδο, με βάση το συναίσθημά τους και τις εκφράσεις τους στα μέσα κοινωνικής δικτύωσης. Τέλος, οι πληροφορίες που σχετίζονται με τις λανθάνουσες μολυσματικές ασθένειες ανακτώνται από τους φορείς στάθμισης συμπτωμάτων των ατόμων.



Σχήμα 13 *- Διάγραμμα ροής ενός μοντέλου μη επιτηρούμενης ανάλυσης συναισθήματος* [33]

2.3 Ιχνηλάτηση Επαφών

2.3.1 Ανάγκη για Έξυπνο Εντοπισμό και Ιχνηλάτηση Επαφών

Σύμφωνα με τον Π.Ο.Υ., η ιχνηλάτηση επαφών είναι «η διαδικασία εντοπισμού, αξιολόγησης και διαχείρισης ατόμων που έχουν εκτεθεί σε μια ασθένεια, με στόχο την πρόληψη της περαιτέρω μετάδοσης». Έτσι, η ιχνηλάτηση επαφών για τη νόσο COVID-19 απαιτεί τον εντοπισμό των ατόμων που μπορεί να έχουν εκτεθεί στον ιό, και την παρακολούθησή τους καθημερινά για 14 ημέρες. Ο στόχος της ιχνηλάτησης επαφών είναι να περιοριστεί η εξάπλωση του ιού μέσω της μείωσης του αριθμού των ατόμων που κυκλοφορούν και είναι φορείς του ιού.

Όπως αναφέρθηκε και στην ενότητα 1.2.1, η ασθένεια COVID-19 είναι πολύ μεταδοτική. Επιπλέον, ένα ποσοστό των φορέων του ιού δεν εκδηλώνει ποτέ συμπτώματα (ασυμπτωματικοί φορείς) και όμως μπορεί να μεταδώσει τον ιό. Οι υπόλοιποι φορείς (συμπτωματικοί ασθενείς), μέχρι τη στιγμή που θα εκδηλώσουν συμπτώματα ή θα βρεθούν θετικοί σε τεστ, μπορεί να έχουν ήδη μεταδώσει τον ιό σε άλλα άτομα με τα οποία ήρθαν σε επαφή. Λόγω των παραπάνω χαρακτηριστικών μετάδοσης, η διαδικασία ανίχνευσης επαφών στην περίπτωση της πανδημίας COVID-19 είναι πολύ δύσκολη, αλλά ταυτόχρονα και πολύ σημαντική.

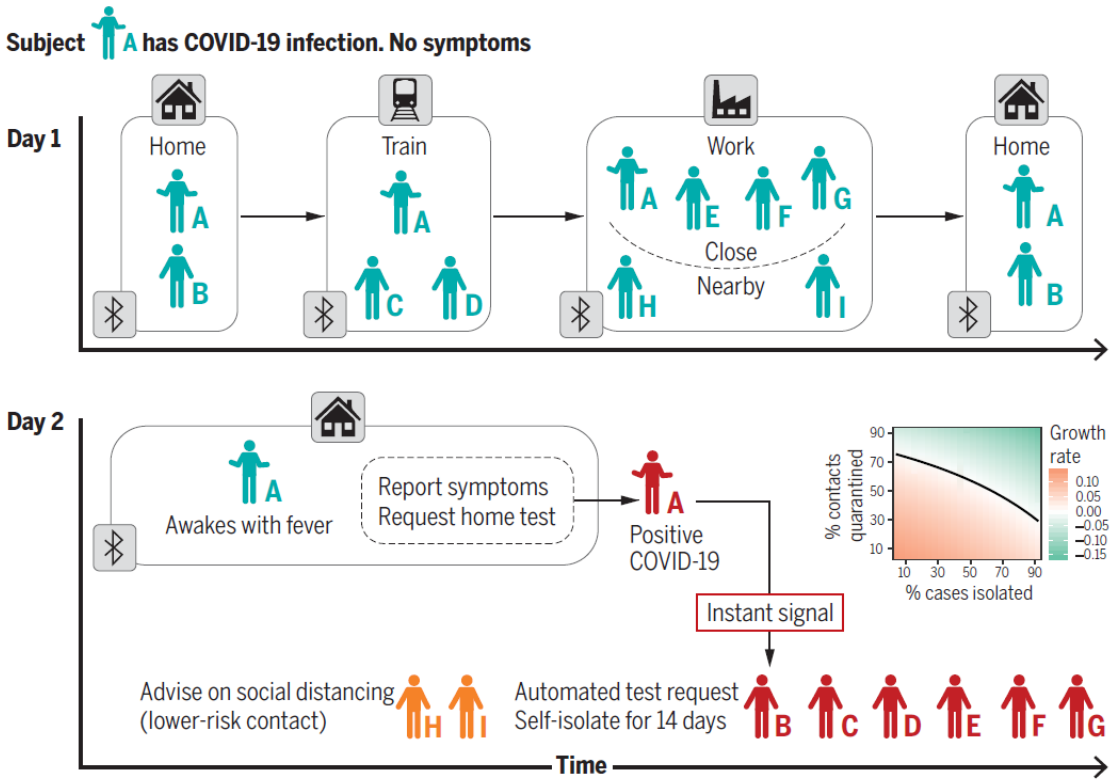
Ο παραδοσιακός τρόπος ιχνηλάτησης επαφών πραγματοποιείται συνήθως μέσω συνεντεύξεων των ατόμων που έχουν μολυνθεί, οι οποίες διεξάγονται από τις υγειονομικές αρχές. Στόχος μιας τέτοιας συνέντευξης είναι η συλλογή των επαφών που είχε το μολυσμένο άτομο με άλλα άτομα τις τελευταίες 14-21 ημέρες, οι οποίες προσδιορίζονται ως η περίοδος επώασης για το συγκεκριμένο ιό. Οι εργαζόμενοι στις υπηρεσίες υγείας μπορούν στη συνέχεια να χρησιμοποιήσουν αυτές τις πληροφορίες για να υπολογίσουν ένα βαθμό κινδύνου για κάθε μία από τις επαφές του εν λόγω ατόμου, με βάση το περιβάλλον (εσωτερικοί/εξωτερικοί χώροι), τη διάρκεια και την εγγύτητα (απόσταση) της κάθε συναναστροφής.

Όπως προκύπτει από τα παραπάνω, ο εντοπισμός των επαφών μόνο μέσω της μεθόδου των συνεντεύξεων έχει πολλούς περιορισμούς και ελλείψεις. Πρώτον, είναι δύσκολο για τους ανθρώπους να θυμούνται με ακρίβεια όλα τα άτομα με τα οποία έχουν συναντηθεί τις τελευταίες τρεις εβδομάδες. Εκτός αυτού, ένα μολυσμένο άτομο μπορεί να έχει μεταδώσει την ασθένεια σε πολλά άτομα που δε γνωρίζει, όπως για παράδειγμα μέσω της επαφής με αγνώστους που περιμένουν σε μια ουρά ενός ταμείου σουπερμάρκετ. Επιπλέον, το μεγάλο πλήθος των συνεντεύξεων απαιτεί σημαντικό εργατικό δυναμικό υπαλλήλων υγείας, εκπαιδευμένων στη διαδικασία ιχνηλάτησης επαφών. Κάτι τέτοιο καθίσταται προφανώς μη διαχειρίσιμο μετά από πολύ σύντομο χρονικό διάστημα από την έναρξη της εξάπλωσης της ασθένειας.

Στο πλαίσιο αυτό, πολλοί ερευνητές έχουν επικεντρωθεί σε τεχνολογικές λύσεις για την αυτοματοποίηση της διαδικασίας ιχνηλάτησης επαφών, με στόχο το γρήγορο και αξιόπιστο εντοπισμό ατόμων που μπορεί να διατρέχουν σημαντικό κίνδυνο μόλυνσης. Η γενική ιδέα είναι ότι μια εφαρμογή ιχνηλάτησης επαφών για έξυπνα κινητά τηλέφωνα αποθηκεύει στη μνήμη της τις επαφές ενός χρήστη. Αυτές ειδοποιούνται αμέσως σε περίπτωση που ο εν λόγω χρήστης βρεθεί θετικός στον ιό. Με αυτό τον τρόπο, μπορεί θεωρητικά να επιτευχθεί καλύτερος έλεγχος της επιδημίας, εάν οι εφαρμογές χρησιμοποιηθούν από την πλειοψηφία των ανθρώπων. Με κατάλληλη στόχευση των συστάσεων για απομόνωση μόνο σε όσους διατρέχουν πραγματικό κίνδυνο, οι επιδημίες θα μπορούσαν θεωρητικά να περιοριστούν χωρίς την επιβολή οριζόντιων μέτρων ολικής καραντίνας, τα οποία έχουν καταστροφικές συνέπειες για την οικονομία και την κοινωνία [34].

Η ευρεία υιοθέτηση των έξυπνων κινητών τηλεφώνων, η ικανότητά τους να γνωρίζουν ανά πάσα στιγμή τη θέση τους (μέσω τεχνολογιών GPS, κυψελωτών δικτύων και δικτύων WiFi), καθώς και η ενσωματωμένη διεπαφή Bluetooth που διαθέτουν και η οποία μπορεί να χρησιμοποιηθεί για την επικοινωνία και την ανίχνευση εγγύτητας με κοντινές συσκευές, τα καθιστά ιδανικά για τους σκοπούς της αυτοματοποιημένης και αξιόπιστης ιχνηλάτησης επαφών.

Στο παρελθόν έχουν υπάρξει διάφορες προσπάθειες χρήσης των πιο πάνω τεχνολογιών για την αυτοματοποιημένη ιχνηλάτηση επαφών, όπως για παράδειγμα το FluPhone Project [35], το οποίο χρησιμοποίησε επικοινωνίες Bluetooth για να εντοπίσει τις επαφές μεταξύ ανθρώπων στο Cambridge του Ηνωμένου Βασιλείου. Στην εργασία αυτή, οι επιστήμονες μελέτησαν πόσο συχνά διαφορετικοί άνθρωποι (που μπορεί να μη γνωρίζουν ο ένας τον άλλον) έρχονται σε κοντινή απόσταση μεταξύ τους ως μέρος της καθημερινότητάς τους. Για να γίνει αυτό, ζητήθηκε από κάποιους εθελοντές να εγκαταστήσουν μια εφαρμογή στα κινητά τους τηλέφωνα και να έχουν τις συσκευές μαζί τους κατά τη διάρκεια των κανονικών καθημερινών δραστηριοτήτων τους. Η εφαρμογή αυτή έψαχνε περιοδικά για άλλες κοντινές κινητές συσκευές χρησιμοποιώντας την επικοινωνία Bluetooth, κατέγραφε αυτές τις πληροφορίες και τις έστελνε πίσω στην ερευνητική ομάδα μέσω διαδικτύου. Οι συμμετέχοντες κλήθηκαν επίσης να παράσχουν στοιχεία σχετικά με συμπτώματα που έμοιαζαν με αυτά της γρίπης και που μπορεί να είχαν βιώσει κατά τη διάρκεια της περιόδου μελέτης. Με αυτό τον τρόπο, οι ερευνητές μπόρεσαν να αντιστοιχίσουν την εξάπλωση της γρίπης στο δίκτυο των επαφών που ανιχνεύθηκαν.



Σχήμα 14 *- Επισκόπηση της αυτοματοποιημένης ανίχνευσης επαφών με χρήση εφαρμογών για κινητά τηλέφωνα και Bluetooth* [34]

2.3.2 Τύποι Αρχιτεκτονικών των Εφαρμογών Ιχνηλάτησης Επαφών

Υπάρχουν τρεις διαφορετικές αρχιτεκτονικές συστημάτων που χρησιμοποιούνται συνήθως ή προτείνονται για την ανάπτυξη εφαρμογών αυτόματης ιχνηλάτησης επαφών στα πλαίσια της πανδημίας COVID-19. Πρόκειται για τη συγκεντρωτική (centralized), την αποκεντρωμένη (decentralized) και την υβριδική (hybrid) προσέγγιση. Η τελευταία συνδυάζει χαρακτηριστικά τόσο από τη συγκεντρωτική όσο και από την αποκεντρωμένη αρχιτεκτονική. Η διάκριση μεταξύ των τριών τύπων βασίζεται στον τρόπο με τον οποίο χρησιμοποιείται ο κεντρικός διακομιστής (server) και στο είδος των δεδομένων που χρησιμοποιούνται ή αποθηκεύονται από αυτόν [36].

Στη συγκεντρωτική προσέγγιση, ο χρήστης πρέπει να προεγγραφεί (preregister) στον κεντρικό διακομιστή. Ο διακομιστής δημιουργεί ένα προσωρινό αναγνωριστικό (TempID) για κάθε συσκευή, το οποίο δεν περιλαμβάνει τα προσωπικά στοιχεία του χρήστη. Αυτό το TempID κρυπτογραφείται στη συνέχεια με ένα μυστικό κλειδί (γνωστό μόνο στον κεντρικό διακομιστή) και αποστέλλεται στη συσκευή. Οι συσκευές ανταλλάσσουν αυτά τα TempIDs σε Bluetooth μηνύματα «συνάντησης» (encounter), όταν έρχονται σε κοντινή επαφή μεταξύ τους. Μόλις ένας χρήστης εξεταστεί και βρεθεί θετικός στον ιό, μπορεί εθελοντικά να ανεβάσει όλα τα αποθηκευμένα μηνύματα συνάντησής του στον κεντρικό διακομιστή. Στη συνέχεια, ο διακομιστής αντιστοιχίζει τα TempIDs των μηνυμάτων σε άτομα για να προσδιορίσει τις επαφές που διατρέχουν κίνδυνο και να τους αποστείλει τις κατάλληλες ειδοποιήσεις. Στη συγκεντρωτική αρχιτεκτονική, ο κεντρικός διακομιστής διαδραματίζει βασικό ρόλο στην εκτέλεση σημαντικών λειτουργιών, όπως την αποθήκευση κρυπτογραφημένων προσωπικών πληροφοριών (Personally Identifiable Information - PII), τη δημιουργία ανώνυμων TempIDs, τον υπολογισμό του κινδύνου και την αποστολή ειδοποιήσεων στις επαφές.

Σε αντίθεση με τη συγκεντρωτική, η αποκεντρωμένη αρχιτεκτονική ουσιαστικά μεταφέρει βασικές λειτουργίες στις συσκευές των τελικών χρηστών, αφήνοντας τον κεντρικό διακομιστή με ελάχιστη συμμετοχή στη διαδικασία ιχνηλάτησης επαφών. Η ιδέα πίσω από αυτή την αρχιτεκτονική είναι να ενισχυθεί η προστασία της ιδιωτικότητας των χρηστών με τη δημιουργία ανώνυμων αναγνωριστικών στις συσκευές, κρατώντας τις πραγματικές τους ταυτότητες μυστικές από τους άλλους χρήστες καθώς και από το διακομιστή. Επίσης, η επεξεργασία των ειδοποιήσεων για πιθανή έκθεση στον ιό γίνεται στις ίδιες τις συσκευές αντί στον κεντρικό διακομιστή. Η αποκεντρωμένη προσέγγιση δεν απαιτεί από τους χρήστες να προεγγραφούν (πριν από τη χρήση των εφαρμογών), αποφεύγοντας έτσι την αποθήκευση οποιωνδήποτε προσωπικών δεδομένων στο διακομιστή.

Στην αποκεντρωμένη αρχιτεκτονική, οι συσκευές παράγουν μοναδικούς τυχαίους κωδικούς που τις χαρακτηρίζουν (seeds), οι οποίοι χρησιμοποιούνται σε συνδυασμό με την τρέχουσα χρονική στιγμή ως είσοδοι σε μια γεννήτρια ψευδοτυχαίων αριθμών για τη δημιουργία ψευδωνύμων (chirps). Τα chirps έχουν πολύ σύντομη διάρκεια ζωής και συγκεκριμένα περίπου 1 λεπτό. Αυτά ανταλλάσσονται περιοδικά μεταξύ συσκευών οι οποίες έρχονται σε κοντινή επαφή. Μόλις ένας χρήστης διαγνωστεί θετικός με COVID-19, μπορεί εθελοντικά να ανεβάσει το seed του και τη σχετική πληροφορία του χρόνου μόλυνσής του στον κεντρικό διακομιστή. Ο κεντρικός διακομιστής λειτουργεί μόνο ως σημείο συνάντησης, παρόμοια με έναν πίνακα ανακοινώσεων, για τη δημοσιοποίηση των seeds των μολυσμένων χρηστών. Οι χρήστες της εφαρμογής μπορούν περιοδικά να κατεβάζουν αυτά τα seeds και να ανακατασκευάζουν τα αντίστοιχα chirps (χρησιμοποιώντας χρονικές σημάνσεις - timestamps) που στάλθηκαν από τους μολυσμένους χρήστες. Με αυτό τον τρόπο, μπορούν να ελέγχουν αν έχουν εκτεθεί για αρκετά μεγάλο χρονικό διάστημα στον ιό και να υπολογίζουν τον αντίστοιχο κίνδυνο μόλυνσης. Στην αρχιτεκτονική αυτή, ο κεντρικός διακομιστής, καθώς και οι υπόλοιποι χρήστες, δε μπορούν να ταυτοποιήσουν ένα συγκεκριμένο χρήστη με τη γνώση μόνο των seeds και των chirps.

Στην συγκεντρωτική αρχιτεκτονική, ο διακομιστής εκτελεί όλες τις σύνθετες εργασίες, π.χ. υπολογισμούς TempID, κρυπτογράφηση, αποκρυπτογράφηση, ανάλυση κινδύνου και αποστολή ειδοποιήσεων στις επαφές που διατρέχουν κίνδυνο. Αντίθετα, όλες αυτές οι λειτουργίες μεταβιβάζονται στις ίδιες τις συσκευές στην αποκεντρωμένη αρχιτεκτονική, διατηρώντας τον κεντρικό διακομιστή μόνο ως πίνακα ανακοινώσεων για σκοπούς αναζήτησης. Η υβριδική αρχιτεκτονική προτείνει οι πιο πάνω λειτουργίες να κατανέμονται μεταξύ του διακομιστή και των συσκευών. Πιο συγκεκριμένα, η παραγωγή και η διαχείριση των TempIDs παραμένουν αποκεντρωμένες (δηλαδή εκτελούνται από τις συσκευές) για να διασφαλιστεί η προστασία των προσωπικών δεδομένων και η ιδιωτικότητα, ενώ η ανάλυση κινδύνου και η αποστολή ειδοποιήσεων αποτελούν ευθύνη του κεντρικού διακομιστή.

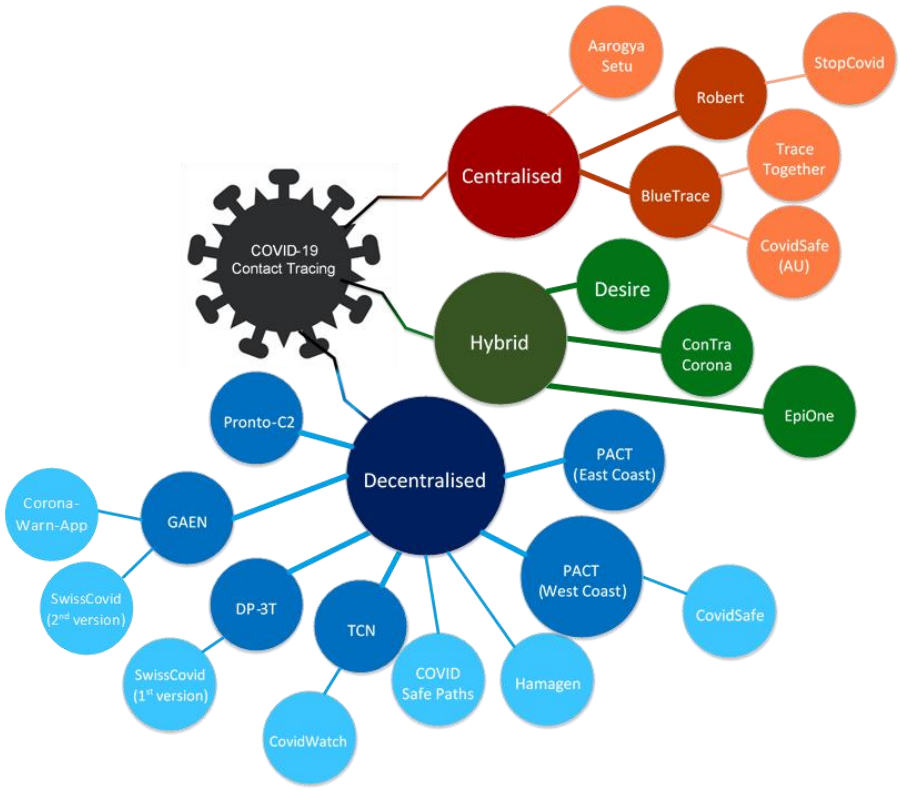
Ένα σημαντικό πλεονέκτημα της χρήσης μιας αρχιτεκτονικής που μεταφέρει τη λειτουργία ανάλυσης κινδύνου και αποστολής ειδοποιήσεων στον κεντρικό διακομιστή (χαρακτηριστικά τόσο της συγκεντρωτικής όσο και της υβριδικής αρχιτεκτονικής) είναι ότι οι υπεύθυνοι των πολιτικών δημόσιας υγείας μπορούν να αποφασίζουν σχετικά με την «ευαισθησία» των ειδοποιήσεων, ανάλογα με τις συνθήκες και το στάδιο της πανδημίας (π.χ. ανάλογα με τη διαθεσιμότητα εργαστηριακών τεστ). Ταυτόχρονα, οι υλοποιήσεις που βασίζονται σε αποκεντρωμένες και υβριδικές αρχιτεκτονικές στοχεύουν να κρατήσουν τις ταυτότητες των χρηστών μυστικές από τον κεντρικό διακομιστή, κάτι που δε συμβαίνει στη συγκεντρωτική αρχιτεκτονική. Μια παραβίαση της ασφάλειας του διακομιστή στη συγκεντρωτική αρχιτεκτονική θα είχε ως αποτέλεσμα πιθανή διαρροή προσωπικών και ευαίσθητων πληροφοριών.

2.3.3 Παραδείγματα Εφαρμογών Ιχνηλάτησης Επαφών

Υπάρχουν αρκετά πρωτόκολλα που υλοποιούν συγκεντρωτικές ή αποκεντρωμένες αρχιτεκτονικές, τα οποία χρησιμοποιούνται από διαφορετικές εφαρμογές κινητών τηλεφώνων για αυτοματοποιημένη ιχνηλάτηση επαφών σε όλο τον κόσμο. Υπάρχουν, επίσης, μερικά παραδείγματα χρήσης υβριδικών αρχιτεκτονικών. Ένας ενδεικτικός κατάλογος εφαρμογών ιχνηλάτησης επαφών και αντίστοιχων αρχιτεκτονικών και πρωτοκόλλων απεικονίζεται στο Σχήμα 15.

Παραδείγματα εφαρμογών για κινητές συσκευές που βασίζονται στη συγκεντρωτική αρχιτεκτονική είναι οι εφαρμογές TraceTogether (Σιγκαπούρη) και CovidSafe (Αυστραλία), οι οποίες χρησιμοποιούν το πρωτόκολλο Bluetrace, και η εφαρμογή StopCovid (Γαλλία), που χρησιμοποιεί το πρωτόκολλο ROBERT. Επίσης, η εφαρμογή Aarogya Setu (Ινδία) είναι μία επιπλέον υλοποίηση βασισμένη στη συγκεντρωτική αρχιτεκτονική και χρησιμοποιεί ένα συνδυασμό Bluetooth και GPS.

Όσον αφορά στην αποκεντρωμένη αρχιτεκτονική, υπάρχουν διάφορα πρωτόκολλα που έχουν αναπτυχθεί ή βρίσκονται υπό ανάπτυξη [36], όπως για παράδειγμα το DP-3T, το TCN και το PACT (east / west coast). Πέραν αυτών, ίσως το σημαντικότερο παράδειγμα πρωτοκόλλου αυτής της αρχιτεκτονικής είναι το GAEN (Google Apple Exposure Notification APIs), που αναπτύχθηκε μέσω συνεργασίας της Google με την Apple, σε μια προσπάθεια υποστήριξης της ιχνηλάτησης επαφών με ταυτόχρονη προστασία των προσωπικών δεδομένων. Το GAEN είναι ένα αποκεντρωμένο πρωτόκολλο που βασίζεται σε ένα συνδυασμό τεχνολογίας Bluetooth Χαμηλής Ενέργειας (Bluetooth Low Energy - BLE) και κρυπτογράφησης των προσωπικών δεδομένων. Το πρωτόκολλο αυτό εφαρμόζεται σε επίπεδο λειτουργικού συστήματος ως διεργασία παρασκηνίου (backend process), και με αυτό τον τρόπο επιτυγχάνει πιο αποδοτική λειτουργία. Οι εφαρμογές που το υλοποιούν παρέχουν στους χρήστες την επιλογή να ενεργοποιήσουν / απενεργοποιήσουν (opt-in / opt-out) τη λειτουργία αυτόματης ιχνηλάτησης και διατίθενται μόνο από εξουσιοδοτημένες υγειονομικές αρχές (μία ανά χώρα). Παραδείγματα εφαρμογών που χρησιμοποιούν αυτή την προσέγγιση (GAEN) είναι η SwissCovid (Ελβετία) και η Corona-Warn-App (Γερμανία).



Σχήμα 15 *- Ενδεικτικά παραδείγματα εφαρμογών ιχνηλάτησης επαφών και αντίστοιχων αρχιτεκτονικών και πρωτοκόλλων* [36]

2.3.4 Μηχανική Μάθηση στην Ιχνηλάτηση Επαφών

Η χρήση αλγορίθμων μηχανικής μάθησης μπορεί να ενισχύσει τη διαδικασία υπολογισμού του ρίσκου, δηλαδή της πιθανότητας μόλυνσης ενός ατόμου, στις εφαρμογές ιχνηλάτησης. Πιο συγκεκριμένα, λαμβάνοντας μετρήσεις για τις ορθώς θετικές, τις ψευδώς αρνητικές καθώς και τις ψευδώς θετικές προβλέψεις μόλυνσης, οι αλγόριθμοι μπορούν να προσαρμοστούν δυναμικά και να αυτοβελτιωθούν σε ό,τι αφορά την αξιοπιστία και την ακρίβειά τους.

Υπάρχει ήδη σημαντική έρευνα στον τομέα αυτό, κυρίως με την ενσωμάτωση της μηχανικής μάθησης στην αποκεντρωμένη αρχιτεκτονική. Ένα παράδειγμα τέτοιας πρωτοβουλίας ήταν ο πρόσφατος διαγωνισμός «Too-Close-For-Too-Long» (TC4TL) που διεξήγαγε το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας των ΗΠΑ (National Institute of Standards and Technology - NIST) [37]. Ο διαγωνισμός αυτός προσέλκυσε πολλές ερευνητικές ομάδες από όλο τον κόσμο, που μελέτησαν τους τρόπους με τους οποίους η μηχανική μάθηση θα μπορούσε να βελτιώσει τον υπολογισμό αποστάσεων στις εφαρμογές ιχνηλάτησης. Ανάμεσα στα συμπεράσματα του διαγωνισμού αναδείχθηκε η ανάγκη του συνδυασμού των μοτίβων σημάτων Bluetooth με δεδομένα από άλλους αισθητήρες των κινητών συσκευών, όπως επιταχυνσιόμετρα, γυροσκόπια και μαγνητόμετρα [38].

Σε μια ακόμα μελέτη για τη χρήση μηχανικής μάθησης στην ιχνηλάτηση επαφών [39], οι ερευνητές πρότειναν μια προσέγγιση βασισμένη σε γραμμική παλινδρόμηση (linear regression). Στόχος τους ήταν ο αξιόπιστος εντοπισμός των ατόμων που πέρασαν αρκετό χρόνο σε κοντινή απόσταση μεταξύ τους, γεγονός που θεωρητικά αυξάνει σημαντικά τον κίνδυνο να έχουν μολυνθεί. Πιο συγκεκριμένα, η προσέγγιση αυτή υπολογίζει κατά πόσον η ισχύς σήματος Bluetooth που λαμβάνεται από μία γειτονική συσκευή υπερβαίνει ένα κατώφλι «υψηλού κινδύνου». Παρόλο που η ισχύς του σήματος υπόκειται σε σημαντική μεταβλητότητα, οι ερευνητές διαπίστωσαν ότι το εκπαιδευμένο μοντέλο

τους είχε καλή απόδοση σε πειράματα που διεξήχθησαν μεταξύ 48 συμμετεχόντων και είχε υψηλή συσχέτιση με τον κίνδυνο αντικειμενικής αλήθειας (ground truth) όταν δοκιμάστηκε μεταξύ 392 επαφών που αντιστοιχούσαν στα δεδομένα επαλήθευσης του μοντέλου.

Οι ερευνητές ανέφεραν επίσης ότι η απόδοση του μοντέλου τους θα μπορούσε πιθανότατα να βελτιωθεί περαιτέρω, αν εκπαιδευτεί χρησιμοποιώντας ένα δείγμα των «πραγματικών γεγονότων μόλυνσης», όπως θα μπορούσαν να τα αναφέρουν εθελοντικά οι χρήστες της εφαρμογής. Επίσης, αν και η προσέγγιση αυτή δε λαμβάνει υπόψη σημαντικούς παράγοντες, όπως για παράδειγμα τη χρήση μάσκας, οι ερευνητές πιστεύουν ότι το σύστημα θα μπορούσε να βελτιωθεί με την εισαγωγή πρόσθετων πηγών δεδομένων, όπως απαντήσεις σε ερωτηματολόγια προς τους χρήστες, δεδομένα από GPS ή δεδομένα από αισθητήρα γυροσκοπίου της κινητής συσκευής [40].

2.4 Επιδημιολογικά Μοντέλα - Πρόβλεψη και Προσομοίωση της Εξάπλωσης

2.4.1 Πρόβλεψη με Βάση Επίσημα Καταγεγραμμένα Στοιχεία

Επαναχρησιμοποίηση υφιστάμενων μοντέλων πρόβλεψης

Μέθοδοι βασισμένες στην τεχνητή νοημοσύνη μπορούν να χρησιμοποιηθούν για να προβλέψουν πώς η ασθένεια COVID-19 θα εξαπλωθεί με την πάροδο του χρόνου και σε σχέση με το χώρο (διαφορετικές τοποθεσίες). Τέτοιες μέθοδοι χρησιμοποιήθηκαν στο παρελθόν και για άλλες πανδημίες. Για παράδειγμα, στην πανδημία του ιού Zika το 2015, ερευνητές ανέπτυξαν ένα δυναμικό νευρωνικό δίκτυο για να προβλέψουν την εξάπλωσή του [41]. Μοντέλα όπως αυτό, μπορούν να επαναχρησιμοποιηθούν στην τρέχουσα πανδημία για σκοπούς πρόβλεψης, επανεκπαιδεύοντάς τα με δεδομένα που αντιστοιχούν στην πανδημία COVID-19. Ένα ακόμα παράδειγμα είναι μια εργασία των ερευνητών του Πανεπιστημίου Carnegie Mellon [42], στην οποία αλγόριθμοι που είχαν χρησιμοποιηθεί στο παρελθόν για την πρόβλεψη της εποχικής γρίπης, εκπαιδεύονται τώρα εκ νέου για την πρόβλεψη της εξάπλωσης της νόσου COVID-19.

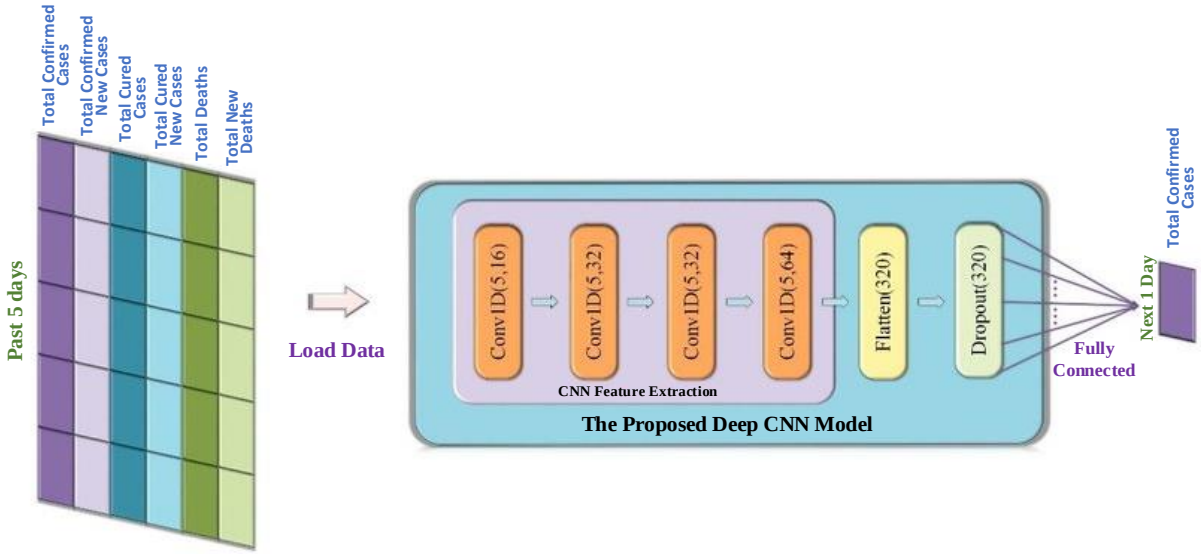
Στην ίδια κατηγορία εμπίπτει και η εργασία ερευνητών από το Πανεπιστήμιο του Σίδνεϋ στην Αυστραλία [43]. Στην εργασία αυτή, οι ερευνητές χρησιμοποίησαν ένα στοχαστικό μοντέλο διακριτού χρόνου βασισμένο σε πράκτορες (agents) με την ονομασία ACEMod (Australian Census-based Epidemic Model), το οποίο είχε αρχικά δημιουργηθεί για την προσομοίωση της πανδημίας της γρίπης. Στην συνέχεια, το προσάρμοσαν κατάλληλα για να μοντελοποιήσουν την πανδημία COVID-19 στην Αυστραλία και να υπολογίσουν την εξέλιξη του αριθμού των κρουσμάτων στην πάροδο του χρόνου.

Στο μοντέλο αυτό, κάθε πράκτορας αντιστοιχεί σε ένα άτομο με συγκεκριμένα χαρακτηριστικά, όπως η ηλικία, το φύλο, η ύπαρξη ευπαθειών ή υποκείμενων νοσημάτων, η ανοσία σε ασθένειες και η συχνότητα επαφής με άλλα άτομα. Το μοντέλο ACEMod προσαρμόστηκε κατάλληλα με παραμέτρους της πανδημίας COVID-19, όπως για παράδειγμα τα χαρακτηριστικά μετάδοσης του νέου ιού SARS-CoV-2. Μέσω της προσομοίωσης σεναρίων, αξιολογήθηκαν διάφορες στρατηγικές παρέμβασης, όπως μέτρα κοινωνικής αποστασιοποίησης (social distancing), κλείσιμο σχολείων, ταξιδιωτικές απαγορεύσεις και απομόνωση κρουσμάτων. Τα αποτελέσματα που προέκυψαν από τις προσομοιώσεις έδειξαν ότι απαιτείται συνδυασμός διαφόρων μέτρων για τον έλεγχο και την καταστολή της πανδημίας COVID-19. Μάλιστα, η καλύτερη στρατηγική που προτείνεται από το μοντέλο είναι ο συνδυασμός των εξής μέτρων: (α) περιορισμός στις αφίξεις από το εξωτερικό, (β) απομόνωση των κρουσμάτων και (γ) κοινωνική αποστασιοποίηση, για τουλάχιστον 13 εβδομάδες με το επίπεδο συμμόρφωσης να είναι μεγαλύτερο ή ίσο του 80%.

Δημιουργία νέων μοντέλων πρόβλεψης για την πανδημία COVID-19

Εκτός από την επαναχρησιμοποίηση και την τροποποίηση υφιστάμενων μοντέλων πρόβλεψης, πολλοί ερευνητές προσπαθούν να αναπτύξουν νέες τεχνικές μοντελοποίησης και αρχιτεκτονικές για την πρόβλεψη χαρακτηριστικών δεικτών της εξέλιξης της πανδημίας COVID-19, όπως ο συνολικός αριθμός των κρουσμάτων, τα ποσοστά θνησιμότητας και τα ποσοστά ανάρρωσης. Τα ακόλουθα παραδείγματα από τη βιβλιογραφία χρησιμοποιούν διαφορετικές μεθοδολογίες και τύπους μοντέλων μηχανικής μάθησης, που βασίζονται κυρίως σε αρχιτεκτονικές νευρωνικών δικτύων.

Για παράδειγμα, στο [44] οι ερευνητές χρησιμοποιούν ένα βαθύ συνελικτικό νευρωνικό δίκτυο (CNN) πολλαπλών εισόδων, στο οποίο αριθμητικά δεδομένα εισόδου συνδυάζονται και τοποθετούνται ως pixels εικόνας, με σκοπό την πρόβλεψη του αριθμού των κρουσμάτων σε διάφορες πόλεις της Κίνας. Ο αριθμός των κρουσμάτων την επόμενη ημέρα (έξοδος του μοντέλου) προβλέπεται λαμβάνοντας υπόψη τις εξής μετρήσεις για τις προηγούμενες 5 ημέρες (είσοδος του μοντέλου): συνολικός αριθμός των επιβεβαιωμένων κρουσμάτων μέχρι σήμερα, αριθμός νέων ημερήσιων επιβεβαιωμένων κρουσμάτων, συνολικός αριθμός κρουσμάτων που ανέρρωσαν πλήρως μέχρι σήμερα, αριθμός νέων ημερήσιων περιπτώσεων που ανέρρωσαν, συνολικός αριθμός θανάτων μέχρι σήμερα, αριθμός νέων ημερήσιων θανάτων. Τα αποτελέσματα της μελέτης αυτής έδειξαν ότι – σε σύγκριση με άλλες μεθόδους βαθιάς μάθησης – το προτεινόμενο μοντέλο CNN είχε τη μεγαλύτερη αποτελεσματικότητα πρόβλεψης.



Σχήμα 16 *- Δομή ενός μοντέλου CNN για την πρόβλεψη του αριθμού των νέων ημερήσιων κρουσμάτων* [44]

Ένας άλλος τύπος βαθέως νευρωνικού δικτύου, και συγκεκριμένα ένα δίκτυο τύπου Modified Stacked Autoencoder (MSA), χρησιμοποιήθηκε για την πρόβλεψη σε πραγματικό χρόνο των κρουσμάτων σε όλη την Κίνα [45]. Το συγκεκριμένο νευρωνικό δίκτυο αποτελείται από τέσσερα στρώματα, και συγκεκριμένα το στρώμα εισόδου, το πρώτο κρυφό στρώμα, το δεύτερο κρυφό στρώμα και το στρώμα εξόδου, με τον αριθμό των κόμβων σε κάθε στρώμα να είναι 8, 32, 4 και 1 αντίστοιχα. Μια χρονοσειρά 8 όρων (που αντιστοιχούν σε 8 ημέρες) χρησιμοποιείται ως είσοδος του δικτύου. Οι μεταβλητές που λαμβάνονται από το δεύτερο στρώμα του μοντέλου υποβάλλονται σε επεξεργασία πριν τροφοδοτηθούν σε αλγορίθμους ομαδοποίησης (clustering). Αυτό χρησιμεύει στην ομαδοποίηση των κρουσμάτων σε επαρχίες ή πόλεις και στη διερεύνηση της δυναμικής μετάδοσης της πανδημίας. Τα σφάλματα που προκύπτουν από το πιο πάνω μοντέλο είναι μικρά, γεγονός που

συνεπάγεται ότι το μοντέλο δύναται να αποτελέσει χρήσιμο εργαλείο για το σχεδιασμό κατάλληλων πολιτικών δημόσιας υγείας.

Ένα νέο μοντέλο πρόβλεψης του συνολικού αριθμού των κρουσμάτων αναπτύχθηκε στο [46]. Σε αυτή τη μελέτη, οι συγγραφείς συνδυάζουν ένα νευροασαφές σύστημα με την ονομασία ANFIS (Adaptive Neuro-Fuzzy Inference System) με ένα σύνθετο αλγόριθμο βασισμένο σε νευρωνικά δίκτυα με την ονομασία FPASSA (Flower Pollination Algorithm and Salp Swarm Algorithm) για τη βελτιστοποίηση των παραμέτρων του μοντέλου. Το ANFIS έχει εφαρμοστεί ευρέως σε προβλήματα πρόβλεψης χρονοσειρών και έχει αποδειχθεί πολύ αποδοτικό. Το μοντέλο που προκύπτει από την πιο πάνω σύνθεση (FPASSA-ANFIS) έχει ως βασικό χαρακτηριστικό τη δυνατότητα ανίχνευσης της μη γραμμικότητας στα δεδομένα των χρονοσειρών, καθώς συνδυάζει τις ιδιότητες τόσο των νευρωνικών δικτύων όσο και της ασαφούς λογικής (fuzzy logic). Το προτεινόμενο μοντέλο αξιολογήθηκε επιτυχώς, συγκρινόμενο με επίσημα στοιχεία του Π.Ο.Υ., για την πρόβλεψη των κρουσμάτων σε χρονικό βάθος δέκα ημερών στην Κίνα. Τέλος, πέραν της πανδημίας COVID-19, το μοντέλο δοκιμάστηκε και επαληθεύτηκε και στην περίπτωση της επιδημίας της γρίπης, με δεδομένα από το CDC των ΗΠΑ και από τον Π.Ο.Υ. για δύο διαφορετικές τετραετείς περιόδους.

Στο [47] οι ερευνητές υποστήριξαν ότι η παραδοσιακή μέθοδος SIR (Susceptible-Infected-Removed), που χρησιμοποιείται από πολλά κλασικά στατιστικά επιδημιολογικά μοντέλα, δεν είναι σε θέση να αποτυπώσει τις επιπτώσεις πολύπλοκων αλληλεπιδράσεων, όπως οι πολιτικές κοινωνικής αποστασιοποίησης και καραντίνας. Για το λόγο αυτό, πρότειναν την κωδικοποίηση της πολιτικής καραντίνας ως μια συνάρτηση ενίσχυσης (strength function), η οποία στη συνέχεια υπεισέρχεται σε ένα νευρωνικό δίκτυο για την πρόβλεψη της εξάπλωσης του ιού SARS-CoV-2 στη Γουχάν. Τα πειραματικά αποτελέσματα που προέκυψαν μέσω του πιο πάνω μοντέλου, με χρήση δημόσια διαθέσιμων στοιχείων από την Κινεζική Εθνική Επιτροπή Υγείας, απέδειξαν στην πράξη ότι η πολιτική καραντίνας διαδραματίζει πολύ σημαντικό ρόλο στον έλεγχο της επιδημίας COVID-19 και μάλιστα, ο αριθμός των κρουσμάτων μπορεί να αυξηθεί εκθετικά χωρίς τη σωστή πολιτική καραντίνας. Πιο συγκεκριμένα, αποδείχθηκε ότι τα μέτρα καραντίνας που εφαρμόστηκαν στη Γουχάν στις αρχές του 2020 μείωσαν τον πραγματικό ρυθμό μετάδοσης του ιού, R(t), από $R(t) > 1$ σε $R(t) < 1$ εντός ενός μήνα από την επιβολή τους. Με αυτό τον τρόπο, περιορίσθηκαν τα νέα κρούσματα και τελικά ελέγχθηκε η επιδημία στη Γουχάν.

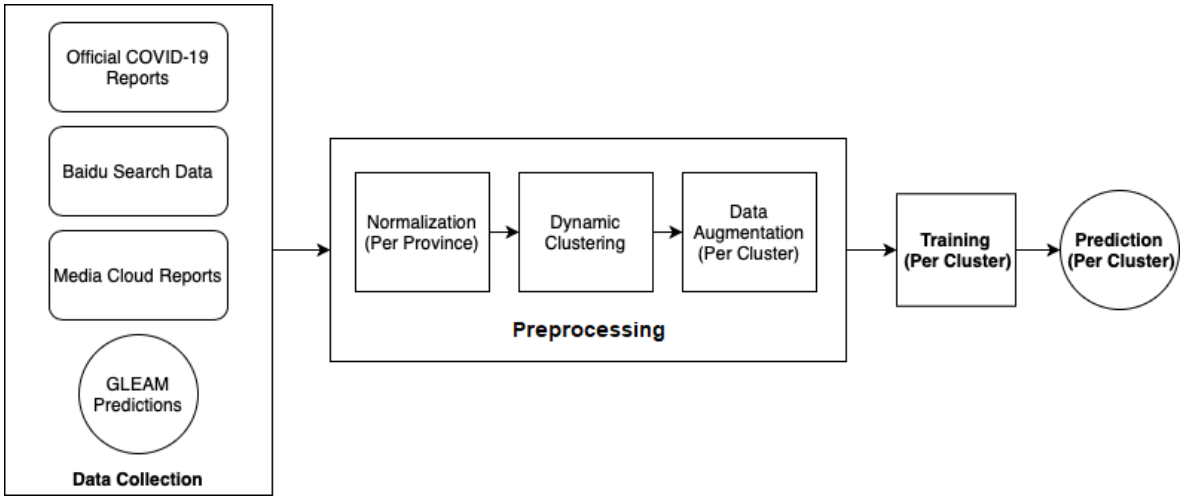
Ενδιαφέρον παρουσιάζει επίσης μια άλλη μελέτη [48], στην οποία ερευνητές χρησιμοποίησαν δεδομένα μετακινήσεων πληθυσμού εντός της Κίνας για να τροφοδοτήσουν ένα δυναμικό μοντέλο μετάδοσης του νέου κορωνοϊού. Το μοντέλο που χρησιμοποιήθηκε ονομάζεται SEIR (Susceptible-Exposed-Infectious-Removed) και συνδυάστηκε με ένα νευρωνικό δίκτυο, το οποίο εκπαιδεύτηκε με δεδομένα του ιού SARS-CoV-1, με σκοπό την πρόβλεψη της καμπύλης κρουσμάτων της πανδημίας COVID-19. Στα «τμηματικά» (compartmental) μοντέλα όπως το SEIR, οι πληθυσμοί χωρίζονται σε τμήματα και η ροή των ανθρώπων μεταξύ των τμημάτων μοντελοποιείται μέσω συστήματος διαφορικών εξισώσεων. Ο τύπος του νευρωνικού δικτύου που επιλέχθηκε ονομάζεται LTSM (Long Short-Term Memory) και ανήκει στην οικογένεια των Recurrent Neural Networks (RNN). Ο λόγος αυτής της επιλογής ήταν η καταλληλόλητα του συγκεκριμένου δικτύου για την πρόβλεψη χρονοσειρών, που αντιπροσωπεύουν εν προκειμένω την εξέλιξη των νέων κρουσμάτων ως συνάρτηση του χρόνου. Οι ερευνητές κατάφεραν και προέβλεψαν την κορύφωση της επιδημίας στην Κίνα στα τέλη Φεβρουαρίου και τη βαθμιαία αποκλιμάκωσή της μέχρι τα τέλη Απριλίου. Επίσης, απέδειξαν ότι μια καθυστέρηση πέντε ημερών στην εφαρμογή των αυστηρών περιοριστικών μέτρων από τις κινεζικές αρχές θα είχε ως αποτέλεσμα μια αύξηση του μεγέθους της επιδημίας κατά τρεις φορές. Τέλος, έδειξαν ότι μια ενδεχόμενη χαλάρωση / άρση των μέτρων αυτών θα είχε προκαλέσει ένα δεύτερο κύμα στην επαρχία Χουμπέι περί τα μέσα Μαρτίου.

Αξίζει να σημειωθεί ότι στα πρώτα στάδια της εμφάνισης μιας νέας μεταδοτικής ασθένειας, όπως η ασθένεια COVID-19, δεν υπάρχει επαρκής ποσότητα διαθέσιμων δεδομένων που να μπορούν να χρησιμοποιηθούν για την εκπαίδευση κατάλληλων αλγορίθμων μηχανικής μάθησης, γεγονός που δυσχεραίνει και καθυστερεί την ανάπτυξη αξιόπιστων μοντέλων. Πρόσθετες προκλήσεις αποτελούν η περιορισμένη γνώση της νέας ασθένειας, καθώς και οι πολύπλοκοι κοινωνικοπολιτικοί παράγοντες που πιθανώς επηρεάζουν την εξάπλωσή της. Αντιλαμβανόμενοι τα πιο πάνω, οι ερευνητές στο [49] πρότειναν μια μεθοδολογία με την ονομασία GROOMS (Group of Optimized and Multisource Selection). Η μεθοδολογία αυτή αποσκοπεί στη σύγκριση διαφορετικών μοντέλων μέσω κατάλληλων δοκιμών σε μικρά σύνολα δεδομένων. Με βάση τα αποτελέσματα της σύγκρισης, οι ερευνητές κατέληξαν σε ένα νευρωνικό δίκτυο PNN+cf (Polynomial Neural Network with corrective feedback), το οποίο κατάφερε να επιτύχει βέλτιστη απόδοση στην πρόβλεψη των ημερήσιων στατιστικών στοιχείων για μικρά σύνολα δεδομένων, που λήφθηκαν από τις κινεζικές υγειονομικές αρχές στα αρχικά στάδια της πανδημίας.

2.4.2 Πρόβλεψη με Χρήση Επιπρόσθετων Online Πηγών

Οι εφαρμογές που μελετήσαμε στην προηγούμενη ενότητα έκαναν χρήση δεδομένων από επίσημες πηγές, όπως για παράδειγμα του επίσημα δημοσιευμένου αριθμού των ημερήσιων επιβεβαιωμένων κρουσμάτων, για την πρόβλεψη της εξάπλωσης του ιού. Εκτός όμως από τα επίσημα δεδομένα, επιπρόσθετες πηγές, όπως τα μέσα κοινωνικής δικτύωσης και άλλες online πηγές δεδομένων, μπορούν να αποτελέσουν επίσης πολύτιμη πληροφορία για την κατανόηση της κοινής γνώμης και της συμπεριφοράς του πληθυσμού. Οι πληροφορίες αυτές μπορούν να ενσωματωθούν με τα υφιστάμενα επίσημα δεδομένα στα μοντέλα πρόβλεψης, με στόχο την καλύτερη κατανόηση των τρεχουσών συνθηκών και κατά συνέπεια την ακριβέστερη πρόβλεψη της εξάπλωσης.

Σε ένα παράδειγμα μιας αντίστοιχης μελέτης [50], οι ερευνητές συνδύασαν τη σχετική δραστηριότητα αναζήτησης στο διαδίκτυο (μέσω της μηχανής Baidu), δεδομένα από μέσα ενημέρωσης, επίσημα δεδομένα από το Κινεζικό Κέντρο Ελέγχου Ασθενειών και ημερήσιες προβλέψεις από το GLEAM (Global Epidemic And Mobility) [51]. Το GLEAM είναι ένα μηχανιστικό μοντέλο πρόβλεψης που βασίζεται σε ευφυείς πράκτορες (intelligent agents). Αρχικά, οι ερευνητές ομαδοποίησαν τις κινεζικές επαρχίες με βάση τις ομοιότητές τους σε σχέση με την εξάπλωση της ασθένειας COVID-19 και στη συνέχεια, εκπαίδευσαν ένα ξεχωριστό μοντέλο για κάθε ομάδα επαρχιών. Συγκεκριμένα, προσάρμοσαν ένα ήδη υπάρχον μοντέλο νευρωνικού δικτύου (Autoregressive Neural Network) για την πρόβλεψη των στατιστικών στοιχείων. Το μοντέλο αυτό ήταν σε θέση να παράγει σταθερές και ακριβείς προβλέψεις για μια σειρά διαφορετικών στατιστικών στοιχείων για τις επόμενες 2 ημέρες και αποδείχθηκε ακριβέστερο σε σχέση με άλλα μοντέλα για 27 από τις 32 κινεζικές επαρχίες. Ένα διάγραμμα ροής του συστήματος φαίνεται στο Σχήμα 17.



Σχήμα 17 *- Διάγραμμα μοντέλου πρόβλεψης που συνδυάζει επίσημες με άλλες online πηγές δεδομένων* [50]

Αντίστοιχα με την πιο πάνω μελέτη, δεδομένα που σχετίζονται με ερωτήματα αναζήτησης στο Google και με τον όγκο των ειδήσεων σε μέσα ενημέρωσης έχουν χρησιμοποιηθεί και σε άλλες εφαρμογές ως είσοδοι σε μοντέλα για την πρόβλεψη των ημερήσιων τάσεων. Για παράδειγμα, οι ερευνητές στο [52] αξιολογούν τη συχνότητα των αναζητήσεων για διαφορετικά συμπτώματα, όπως αυτά καταγράφονται σε μια έρευνα της Εθνικής Υπηρεσίας Υγείας του Ηνωμένου Βασιλείου για ασθενείς με COVID-19, με βάση αντίστοιχα ερωτηματολόγια. Στη μελέτη αυτή, οι ερευνητές υπολόγισαν αρχικά τις συχνότητες αναζήτησης για κάθε σύμπτωμα και στη συνέχεια, τις στάθμισαν σύμφωνα με την κατανομή πιθανοτήτων που προέκυψε από την πιο πάνω έρευνα της Εθνικής Υπηρεσίας Υγείας.

Χρησιμοποιώντας αυτά τα δεδομένα, μαζί με ημερήσια στατιστικά των προηγούμενων ημερών, οι ερευνητές εκπαίδευσαν ένα μοντέλο ElasticNet, το οποίο είναι ένα είδος μοντέλου παλινδρόμησης (regression), για την πρόβλεψη μελλοντικών τάσεων. Επιπλέον, διερεύνησαν τη δυνατότητα προσαρμογής των μοντέλων τους για εκτέλεση πρόβλεψης σε άλλες χώρες. Μια τέτοια προσέγγιση θα μπορούσε να είναι χρήσιμη για τη διερεύνηση του κατά πόσον η δημιουργία ενός μοντέλου σε χώρες πλούσιες σε δεδομένα θα μπορούσε να εφαρμοστεί σε άλλες χώρες όπου τα δεδομένα είναι ελάχιστα ή και ανύπαρκτα. Σε κάθε περίπτωση, κάποιες από τις παραμέτρους ενός τέτοιου μοντέλου θα πρέπει να προσαρμοστούν στις εκάστοτε τοπικές συνθήκες, δεδομένου ότι συνήθως υπάρχουν διαφορετικά δημογραφικά και άλλα κοινωνικοπολιτικά χαρακτηριστικά που επηρεάζουν την εξάπλωση του ιού σε κάθε χώρα.

Σε μια άλλη μελέτη [53], οι ερευνητές χρησιμοποίησαν μεθόδους ανάλυσης δεδομένων για να προβλέψουν τον αριθμό των νέων κρουσμάτων COVID-19. Πιο αναλυτικά, οι ερευνητές χρησιμοποίησαν μια σειρά από «Δείκτες αναζήτησης μέσων κοινωνικής δικτύωσης» (Social Media Search Indices - SMSI) για διάφορες λέξεις-κλειδιά, συμπεριλαμβανομένων των κλινικών συμπτωμάτων της ασθένειας COVID-19 (όπως ξηρός βήχας, πυρετός, δυσφορία στο στήθος και πνευμονία). Οι ερευνητές διαπίστωσαν ότι, με τη χρήση τεχνικών όπως η μέθοδος επιλογής υποσυνόλου (subset selection method), νέα κρούσματα COVID-19 θα μπορούσαν να ανιχνευθούν από 6 μέχρι και 10 ημέρες νωρίτερα. Η μέθοδος επιλογής υποσυνόλου αναφέρεται στη διαδικασία εύρεσης ενός μικρού υποσυνόλου διαθέσιμων χαρακτηριστικών (predictors) που οδηγούν σε ακριβή πρόβλεψη των μελλοντικών τιμών μιας μεταβλητής.

2.5 Ομαδοποίηση Περιοχών

Κατά τη διάρκεια πανδημιών όπως αυτή της COVID-19, κάθε χώρα βιώνει διαφορετικές περιόδους και ρυθμούς αύξησης της εξάπλωσης, όπως για παράδειγμα αποτυπώνεται στον αριθμό των επιβεβαιωμένων κρουσμάτων και των θανάτων, οι οποίοι αντικατοπτρίζουν τη σοβαρότητα της κατάστασης σε κάθε χώρα. Σε γενικές γραμμές, το πόσο επηρεάζεται μια χώρα από την πανδημία εξαρτάται από μια σειρά παραγόντων, συμπεριλαμβανομένων των διεθνών ταξιδιών, των δημογραφικών της στοιχείων, διαφόρων κοινωνικοοικονομικών παραγόντων, των χαρακτηριστικών του συστήματος υγειονομικής περίθαλψης καθώς και πολιτικών μέτρων και παρεμβάσεων. Αξιολογώντας τα κοινά χαρακτηριστικά ανάμεσα στις χώρες, σε ό,τι αφορά τις τάσεις μετάδοσης του ιού και της πιθανής συσχέτισής τους με άλλα δεδομένα τόσο σε επίπεδο χώρας όσο και σε περιφερειακό επίπεδο, θεωρητικά είναι δυνατή η κατηγοριοποίηση χωρών και περιφερειών σε ομάδες με παρόμοια επιδημιολογική συμπεριφορά. Κάτι τέτοιο μπορεί να είναι χρήσιμο για τον προσδιορισμό των χωρών που έχουν λάβει τα κατάλληλα μέτρα αναχαίτισης της εξάπλωσης του ιού, αλλά και για τη χρησιμοποίηση δεδομένων από μια περιοχή για την πρόβλεψη της εξέλιξης της πανδημίας σε μια άλλη με παρόμοια χαρακτηριστικά.

Αν και οι μελέτες σε αυτό τον τομέα είναι θεωρητικά χρήσιμες, ένας σημαντικός περιορισμός πολλών επιστημονικών άρθρων είναι οι διαφορετικοί τρόποι συλλογής δεδομένων και αναφοράς στατιστικών στοιχείων σε διαφορετικές χώρες. Οι διαφορές αυτές μπορεί να εστιάζονται στη συχνότητα των τεστ, στον τρόπο αναφοράς και παρακολούθησης των κρουσμάτων και στην αξιοπιστία των αντίστοιχων στοιχείων. Επίσης, όπως γίνεται εύκολα αντιληπτό, ορισμένα από τα επίσημα στοιχεία ενδέχεται να μην είναι ικανά από μόνα τους για την κατανόηση της επιδημιολογικής κατάστασης κάθε χώρας. Για παράδειγμα, ο ημερήσιος αριθμός επιβεβαιωμένων κρουσμάτων ανά 100.000 κατοίκους εξαρτάται από τη στρατηγική που ακολουθεί η κάθε χώρα σε ό,τι αφορά τα τεστ. Όμως, κάποιες άλλες μετρήσεις φαίνεται να είναι καταλληλότερες για σύγκριση ανάμεσα σε διαφορετικές χώρες, όπως ο ημερήσιος αριθμός θανάτων ανά 100.000 κατοίκους (με δεδομένο φυσικά ότι οι χώρες που συγκρίνονται καταγράφουν με τον ίδιο τρόπο θανάτους από τη νόσο COVID-19).

Σε μια μελέτη στην περιοχή αυτή [54], οι ερευνητές προσπαθούν να απαντήσουν σε δύο βασικά ερωτήματα: (α) ποιες χώρες έχουν πληγεί περισσότερο από την πανδημία COVID-19 και (β) ποιες κυβερνήσεις έχουν λάβει τα κατάλληλα μέτρα για να σταματήσουν τη διάδοση του ιού. Αρχικά, επιλέγουν το δείκτη εκείνο που θεωρούν ότι αντιπροσωπεύει καλύτερα το βαθμό στον οποίο μια χώρα έχει πληγεί από την πανδημία και καταλήγουν ότι αυτός είναι ο κανονικοποιημένος αριθμός των ημερήσιων θανάτων (αριθμός θανάτων ανά ημέρα, ανά 100.000 κατοίκους). Στη συνέχεια, εφαρμόζουν έναν αλγόριθμο ομαδοποίησης (clustering) μη επιτηρούμενης μάθησης για την ομαδοποίηση παρόμοιων καμπυλών χωρών, οι οποίες απεικονίζουν τους κανονικοποιημένους ημερήσιους θανάτους ως συνάρτηση του χρόνου. Η επιλεγμένη μέθοδος μηχανικής μάθησης είναι μια πολύ δημοφιλής μέθοδος ομαδοποίησης που ονομάζεται k-means.

Το αποτέλεσμα της εργασίας αυτής είναι η ομαδοποίηση των χωρών σε τρεις ομάδες (clusters): «πράσινη», «κόκκινη» και «μπλε». Η πράσινη ομάδα (158 χώρες) περιλαμβάνει τις χώρες με μικρό αριθμό κανονικοποιημένων ημερήσιων θανάτων, οι οποίες επηρεάστηκαν σε σχετικά μικρό έως μέτριο βαθμό από την πανδημία. Αντίθετα, η κόκκινη ομάδα (9 χώρες) είχε έναν τεράστιο αριθμό κανονικοποιημένων θανάτων κατά το πρώτο κύμα (Απρίλιος - Μάιος 2020), αλλά κατάφερε να φέρει την κατάσταση υπό έλεγχο, με αποτέλεσμα το ποσοστό θανάτων σχεδόν να μηδενιστεί. Τέλος, στη μπλε ομάδα (16 χώρες) ανήκουν χώρες οι οποίες από την αρχή της πανδημίας μέχρι και το πέρας της

πιο πάνω μελέτης κατέγραφαν συνεχώς ένα σημαντικό αριθμό θανάτων και η κατάστασή τους επιδεινωνόταν με την πάροδο του χρόνου.

Οι δημιουργοί της πιο πάνω εφαρμογής απεικόνισαν επίσης τα χρώματα των χωρών, με βάση τις ομάδες στις οποίες ανήκουν, στον παγκόσμιο χάρτη (βλ. Σχήμα 18) και έδειξαν ότι η κόκκινη ομάδα βρίσκεται στην Ευρώπη και οι χώρες της μπλε ομάδας συγκεντρώνονται κυρίως στη Βόρεια και τη Νότια Αμερική. Η κόκκινη ομάδα περιλαμβάνει την Ισπανία, τη Γαλλία, το Ηνωμένο Βασίλειο, την Ιταλία, το Βέλγιο, τη Σουηδία, την Ολλανδία, την Ιρλανδία και την Ανδόρα. Οι χώρες αυτές, αν και πιέστηκαν πολύ κατά τη διάρκεια του πρώτου κύματος της πανδημίας COVID-19, εντούτοις κατάφεραν να επιβραδύνουν το ρυθμό των θανάτων, εφαρμόζοντας περιορισμούς στην καθημερινή ζωή, με αποτέλεσμα να επιστρέψουν σε μια μορφή κανονικότητας μέχρι τον Ιούλιο του 2020. Αξίζει να σημειωθεί ότι η μελέτη αυτή καλύπτει μόνο το πρώτο κύμα της πανδημίας COVID-19. Με βάση αυτή την ανάλυση, οι συγγραφείς κατέληξαν στο συμπέρασμα ότι οι σκληροί κοινωνικοί περιορισμοί ήταν αυτοί που βοήθησαν τις περισσότερες ευρωπαϊκές χώρες να ελέγξουν την εξάπλωση του ιού.

Σχήμα 18 *- Μέσος όρος κανονικοποιημένων ημερήσιων θανάτων για κάθε ομάδα χωρών ως προς το χρόνο (αριστερά) και κατανομή των ομάδων χωρών στον παγκόσμιο χάρτη (δεξιά)* [54]

Μια ακόμα προσέγγιση για την ομαδοποίηση των χωρών με χρήση μη επιτηρούμενης μάθησης και συγκεκριμένα μέσω του αλγορίθμου k-means ακολουθήθηκε από τους ερευνητές στο [55]. Στην εργασία αυτή, ομαδοποιήθηκαν 155 χώρες με βάση δεδομένα σε σχέση με τη συχνότητα εμφάνισης ορισμένων νοσημάτων, την ποιότητα του αέρα, το κατά κεφαλήν ακαθάριστο εγχώριο προϊόν, το ποσοστό των ανδρών στον πληθυσμό και το δείκτη καθολικής υγειονομικής κάλυψης (universal health coverage index)[1]. Τα χαρακτηριστικά αυτά επιλέχθηκαν με βάση την κρίση των ερευνητών ότι πιθανώς να συνδέονται με την εξέλιξη της πανδημίας COVID-19, τόσο από την άποψη της κλινικής ιατρικής όσο και της ποιότητας παροχής υπηρεσιών υγείας σε κάθε χώρα. Οι ερευνητές διαπίστωσαν ότι το μοντέλο τους ήταν σε θέση να ομαδοποιήσει τις χώρες σε πέντε ομάδες με βάση παρόμοια συμπεριφορά στον αριθμό των επιβεβαιωμένων κρουσμάτων, αλλά δε μπόρεσε να τις ομαδοποιήσει σε ό,τι αφορά τον αριθμό των θανάτων ή το ποσοστό θνησιμότητας.

[1] Ο δείκτης αυτός διατηρείται από την Παγκόσμια Τράπεζα για κάθε χώρα στον κόσμο. Υπολογίζεται σε μια κλίμακα από 0 έως 100 και αντικατοπτρίζει την ποιότητα των παρεχόμενων υπηρεσιών υγείας σε κάθε χώρα με βάση διάφορα στατιστικά στοιχεία.

Κάποιες άλλες πιο προηγμένες προσεγγίσεις χρησιμοποιούν τα χαρακτηριστικά ενός τύπου νευρωνικών δικτύων και συγκεκριμένα των αυτοκωδικοποιητών (autoencoders), τόσο στην πρόβλεψη των ποσοστών μόλυνσης όσο και στον εντοπισμό ομάδων παρόμοιων περιοχών ή χωρών. Σε μια τέτοια μελέτη [56], οι ερευνητές χρησιμοποίησαν δεδομένα για τα συνολικά και νέα (ημερήσια) επιβεβαιωμένα κρούσματα σε 31 επαρχίες και πόλεις της Κίνας. Μετά την εκπαίδευση ενός νευρωνικού δικτύου τύπου Modified Autoencoder (MAE) για την πρόβλεψη σε πραγματικό χρόνο νέων κρουσμάτων, οι ερευνητές εξήγαγαν πληροφορία από τα εσωτερικά στρώματα του δικτύου για τον καθορισμό των πιο σημαντικών χαρακτηριστικών του μοντέλου για κάθε αναλυθείσα περιοχή. Στη συνέχεια, τα χαρακτηριστικά αυτά τροφοδοτήθηκαν σε έναν αλγόριθμο k-means, ο οποίος ομαδοποίησε παρόμοιες περιοχές για περαιτέρω ανάλυση. Το τελικό αυτό βήμα επέτρεψε την αποτελεσματικότερη διερεύνηση των περιοχών που παρουσίαζαν συγκεκριμένη συμπεριφορά σε σχέση με το ρυθμό εξέλιξης του αριθμού των κρουσμάτων.

Όπως και στην προηγούμενη προσέγγιση, ο ερευνητής στο [57] χρησιμοποίησε ένα εξειδικευμένο νευρωνικό δίκτυο τύπου Τοπολογικού Αυτοκωδικοποιητή (Topological Autoencoder - TA) για την απεικόνιση της παγκόσμιας τάσης μετάδοσης του ιού SARS-CoV-2. Το επιλεχθέν νευρωνικό δίκτυο εκπαιδεύτηκε με βάση τον αριθμό των κρουσμάτων COVID-19 σε 240 χώρες, χρησιμοποιώντας δεδομένα από το Κέντρο Επιστήμης και Μηχανικής Συστημάτων του Πανεπιστημίου Johns Hopkins των ΗΠΑ. Στη συνέχεια, επιλέχθηκαν κάποιες εσωτερικές μεταβλητές του TA για τη δημιουργία μιας δισδιάστατης ομαδοποίησης των χωρών και του αντίστοιχου τοπολογικού χάρτη. Στο χάρτη αυτό, χώρες με παρόμοια δυναμική μετάδοσης τοποθετούνται δίπλα η μία στην άλλη, ενώ χώρες με πολύ διαφορετική δυναμική βρίσκονται σε σημαντική απόσταση στο χάρτη.

Τα αρχικά αποτελέσματα της πιο πάνω μελέτης έδειξαν ότι μπορούμε να αποκτήσουμε κάποιες διαισθητικές γνώσεις σχετικά με τις ομοιότητες και τις διαφορές στη μετάδοση του νέου ιού μεταξύ χωρών. Επίσης, με την απεικόνιση των χαρτών στο χρόνο, μπορούμε να δούμε τις αλλαγές στην κατάσταση της κάθε χώρας, ως αποτέλεσμα για παράδειγμα στρατηγικών που ακολουθήθηκαν. Μια τέτοια διαισθητική / οπτική κατανόηση μέσω τοπολογικών χαρτών θα μπορούσε να είναι χρήσιμη στη σύγκριση στρατηγικών μεταξύ χωρών και στον επακόλουθο σχεδιασμό βέλτιστων στρατηγικών για την αναχαίτιση της μετάδοσης του νέου ιού.

2.6 Εκτίμηση Κινδύνου ανά Περιοχή

Οι εφαρμογές που παρουσιάστηκαν στις προηγούμενες ενότητες επικεντρώθηκαν στην πρόβλεψη τιμών για διάφορα στατιστικά μεγέθη σε διαφορετικές χώρες ή περιοχές, καθώς και στην ομαδοποίηση χωρών ή περιοχών με βάση παρόμοιες συμπεριφορές σε σχέση με τα μεγέθη αυτά. Ακόμα ένα πεδίο έρευνας εστιάζει στην πρόβλεψη του κινδύνου εξάρσεων της πανδημίας σε περιοχές ή χώρες, υπολογίζοντας ένα μοναδικό βαθμό (σκορ) κινδύνου. Το σκορ αυτό μπορεί να διευκολύνει την ταχεία εκτίμηση του κινδύνου σε διαφορετικές περιοχές και να χρησιμοποιηθεί ως αφετηρία για περαιτέρω διερεύνηση σε περιοχές αυξημένου κινδύνου.

Ακολουθώντας την πιο πάνω προσέγγιση, οι ερευνητές στο [58] πρότειναν τη χρήση ενός ρηχού νευρωνικού δικτύου LTSM για την πρόβλεψη της κατηγορίας κινδύνου μιας χώρας. Χρησιμοποίησαν μια bayesian μέθοδο για να βελτιστοποιήσουν και να σχεδιάσουν με αυτόματο τρόπο νευρωνικά δίκτυα προσαρμοσμένα στην κάθε χώρα. Κατά την αξιολόγηση των μεταβλητών που θα έπρεπε να συμπεριληφθούν στο μοντέλο, οι ερευνητές χρησιμοποιήσαν ένα απλό μοντέλο παλινδρόμησης με τη χρήση της μεθόδου των ελαχίστων τετραγώνων. Στη συνέχεια, το βελτιστοποιημένο νευρωνικό δίκτυο κάθε χώρας χρησιμοποιήθηκε για την πρόβλεψη του αριθμού των κρουσμάτων (k), του αριθμού των θανάτων (d) και του αριθμού των αναρρώσεων (r). Τέλος, τα δεδομένα αυτά

τροφοδότησαν ένα επιπλέον σύστημα για την εκτίμηση του βαθμού επικινδυνότητας της κάθε χώρας. Τα αποτελέσματά της μελέτης έδειξαν ότι το συγκεκριμένο σύστημα υπερτερεί έναντι εναλλακτικών μεθόδων με βάση τα δεδομένα 180 χωρών και μπορεί να αποτελέσει χρήσιμο εργαλείο για την κατηγοριοποίηση των χωρών με βάση τον κίνδυνο που διατρέχουν. Οι ερευνητές πειραματίστηκαν, επίσης, εισάγοντας και δεδομένα καιρού στην είσοδο του δικτύου, αλλά απέδειξαν ότι ο καιρός δε φαίνεται να διαδραματίζει σημαντικό ρόλο στην πρόβλεψη της εξέλιξης της πανδημίας.

Death Rate	Case Rate	Recovery Rate	Decision
High	High	Low	HR
Low	High	Low	HR
High	High	High	HR
Low	High	High	HR
High	Low	High	MR
High	Low	Low	MR
Low	Low	Low	MR
Low	Low	High	RE

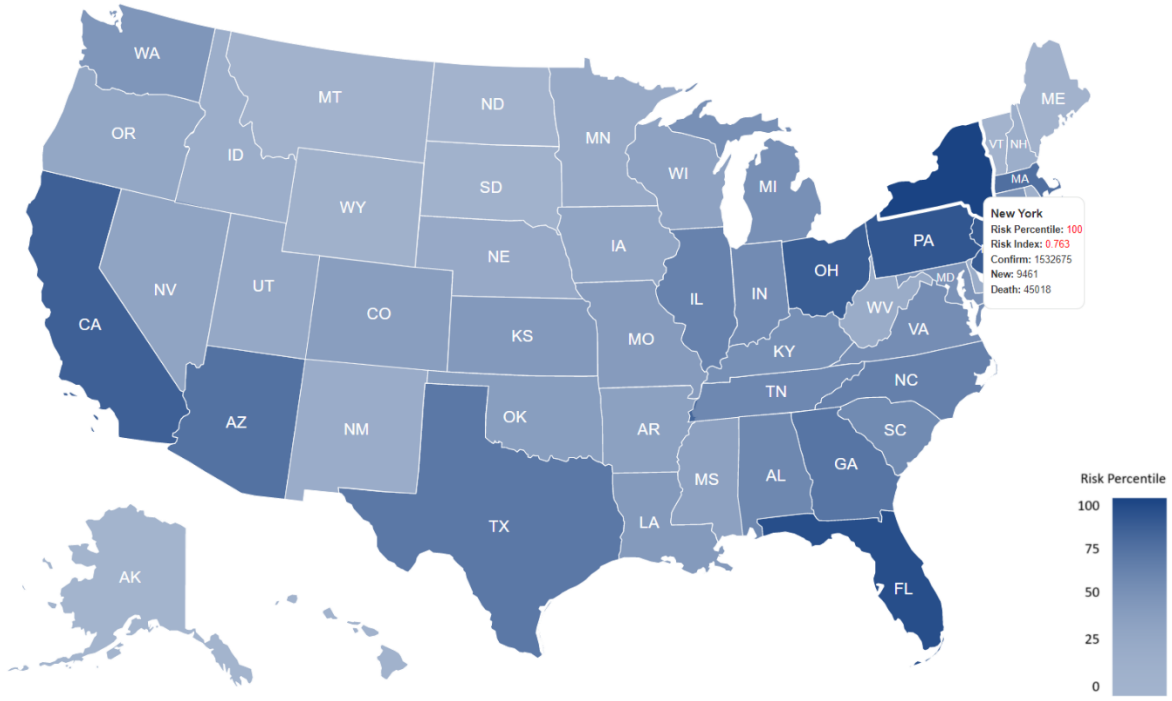
High Risk (HR)	Low Risk (LR)	Recovering (RE)
India	Slovakia	Greenland
USA	Malta	Uruguay
Russia	Denmark	Zimbabwe
Brazil	Switzerland	Japan
Mexico	Germany	Norway

Σχήμα 19 *- Ασαφείς κανόνες για τον υπολογισμό κινδύνου (αριστερά) και παραδείγματα εκτιμώμενων βαθμών κινδύνου 25 ημερών (με αρχή τις 20/08/2020) για διαφορετικές χώρες (δεξιά)* [58]

Μια παρόμοια μελέτη που πραγματοποιήθηκε στην Ιταλία εξέτασε τον Εγγενή Κίνδυνο Μετάδοσης (Inherent Risk of Contagion - IRC), ο οποίος ορίζεται ως ο μέσος βαθμός επιτάχυνσης της εξάπλωσης των κρουσμάτων COVID-19 σε μια γεωγραφική μονάδα (geo-unit), όπως μια επαρχία ή ένα δήμο [59]. Οι ερευνητές έκαναν την υπόθεση ότι ο IRC εξαρτάται από ένα δείκτη που ονομάζουν Δείκτη Κοινωνικής Συνοχής (Social Viscosity Index - SVI), ο οποίος είναι ένας μη γραμμικός συνδυασμός δημογραφικών και κοινωνικών χαρακτηριστικών κάθε γεωγραφικής περιοχής. Αρχικά, χρησιμοποίησαν έναν αλγόριθμο ομαδοποίησης k-means για τον εντοπισμό παρόμοιων περιοχών με βάση τον SVI. Στη συνέχεια, εκπαίδευσαν ένα πλήρως διασυνδεδεμένο νευρωνικό δίκτυο (Fully Connected Network - FCN) με δεδομένα από τη Λομβαρδία της Ιταλίας για την πρόβλεψη του IRC των υπόλοιπων επαρχιών και δήμων της χώρας. Οι ερευνητές επισήμαναν το γεγονός ότι, λόγω του επείγοντος χαρακτήρα της κατάστασης, οι διάφορες μέθοδοι μηχανικής μάθησης που εφάρμοσαν στην επεξεργασία των δεδομένων (ομαδοποίηση, ταξινόμηση και πρόβλεψη) δεν υποβλήθηκαν σε αυστηρές διαδικασίες αξιολόγησης και διασταύρωσης του επιπέδου ακρίβειάς τους.

Σε μια λεπτομερή μελέτη για τη δυναμική αξιολόγηση κινδύνου από την πανδημία COVID-19 στις ΗΠΑ [60], οι ερευνητές πρότειναν και ανέπτυξαν ένα σύστημα τεχνητής νοημοσύνης με την ονομασία α-Satellite. Πιο συγκεκριμένα, μέσω της αντίστοιχης εφαρμογής, οι χρήστες επιλέγουν αρχικά ένα σημείο ενδιαφέροντος (Point of Interest - POI) και στη συνέχεια, το σύστημα τους παρέχει αυτόματα δείκτες κινδύνου που σχετίζονται με ιεραρχικό τρόπο με το σημείο αυτό, ξεκινώντας από το ίδιο το σημείο και γενικεύοντας στην επαρχία και στην αντίστοιχη πολιτεία. Με αυτό τον τρόπο, οι χρήστες μπορούν να ρυθμίσουν κατάλληλα την καθημερινότητά τους, λαμβάνοντας για παράδειγμα τα κατάλληλα μέτρα προστασίας ανάλογα με την περιοχή που επιθυμούν να επισκεφθούν. Το μοντέλο αυτό καθώς και οι δείκτες κινδύνου που παράγει είναι δωρεάν προσβάσιμα μέσω της αντίστοιχης ιστοσελίδας [61]. Για τη σχεδίαση και τη δοκιμή αυτού του μοντέλου, οι ερευνητές χρησιμοποίησαν στοιχεία από τον Π.Ο.Υ., το CDC των ΗΠΑ, τοπικές κυβερνήσεις και άλλες πηγές σε ό,τι αφορά τα κρούσματα και τα ποσοστά θανάτου. Επίσης, έλαβαν υπόψη τους δημογραφικά στοιχεία, στοιχεία κινητικότητας καθώς και δεδομένα από μέσα κοινωνικής δικτύωσης για να κατανοήσουν τον τρόπο

με τον οποίο ο κόσμος αντιλαμβάνεται την έκτακτη αυτή κατάσταση. Ένα στιγμιότυπο της εφαρμογής αυτής για την εκτίμηση κινδύνου σε επίπεδο πολιτειών των ΗΠΑ, καθώς και κάποια περαιτέρω στατιστικά στοιχεία για την πολιτεία της Νέας Υόρκης, φαίνονται στο Σχήμα 20.



Σχήμα 20 *- Στιγμιότυπο του χάρτη εκτίμησης κινδύνου της πανδημίας COVID-19 στις ΗΠΑ μέσω της εφαρμογής α-Satellite (14-02-2021)* [61]

ΚΕΦΑΛΑΙΟ 3: Εφαρμογές στην Κλινική Ιατρική

3.1 Εισαγωγή

Στο Κεφάλαιο αυτό θα εξετάσουμε εφαρμογές της Τεχνητής Νοημοσύνης στο πεδίο της Κλινικής Ιατρικής. Θα ξεκινήσουμε με μεθόδους που αφορούν στη διάγνωση, δηλαδή στους τρόπους με τους οποίους μπορεί να διαπιστωθεί αν ένα άτομο έχει μολυνθεί με τον ιό SARS-CoV-2. Στη συνέχεια, θα παρουσιάσουμε εφαρμογές για την πρόβλεψη της πορείας ενός ασθενούς (πρόγνωση), θα περιγράψουμε τρόπους αποδοτικότερης διαχείρισης και αξιοποίησης των πεπερασμένων πόρων των συστημάτων υγείας και τέλος, θα αναφερθούμε σε μεθόδους εξατομικευμένης αντιμετώπισης και θεραπείας ασθενών.

Αναμφίβολα, η ακριβής και ταχεία διάγνωση της ασθένειας COVID-19 μπορεί να σώσει ζωές και να ανακόψει την εξάπλωσή της. Στα πλαίσια της παρούσας πανδημίας, ένας σημαντικός αριθμός εφαρμογών τεχνητής νοημοσύνης στην κλινική ιατρική έχει επικεντρωθεί στη διάγνωση με βάση την ιατρική απεικόνιση, χρησιμοποιώντας κυρίως εικόνες αξονικής τομογραφίας (Computed Tomography - CT) και ακτινογραφιών (X-Rays) θώρακος και πνευμόνων. Επιπρόσθετα, έχουν αναπτυχθεί εφαρμογές που χρησιμοποιούν μη επεμβατικές μεθόδους διάγνωσης, καθώς και διαφορετικούς τύπους δεδομένων εισόδου, όπως για παράδειγμα εφαρμογές κινητών συσκευών, οι οποίες θα μπορούσαν να εξυπηρετήσουν σκοπούς ταχείας και μαζικής ανίχνευσης ασθενών με COVID-19 (screening).

Από τη στιγμή που ένα άτομο διαγνωστεί θετικό στο νέο ιό, η πρόγνωση της πορείας του είναι καθοριστικής σημασίας τόσο για τον ίδιο τον ασθενή, όσο και για το σύστημα υγείας γενικότερα. Συνεπώς, οι εφαρμογές τεχνητής νοημοσύνης σε αυτό τον τομέα βοηθούν στην εύρεση των παραγόντων εκείνων που συνδέονται με τον κίνδυνο εμφάνισης σοβαρής νόσου και την ανάγκη νοσηλείας ή και εισαγωγής του ασθενούς σε μονάδα εντατικής θεραπείας. Κατ' επέκταση, οι εφαρμογές αυτές αποτελούν ένα χρήσιμο εργαλείο στα χέρια των ειδικών για τη λήψη ορθών και έγκαιρων αποφάσεων, τη βέλτιστη κατανομή των πόρων και την ανάπτυξη εξατομικευμένων πλάνων φροντίδας και θεραπείας των ασθενών.

3.2 Διάγνωση

3.2.1 Εισαγωγή στις Διαγνωστικές Μεθόδους για την Ασθένεια COVID-19

Αυτή τη στιγμή χρησιμοποιούνται διάφορες μέθοδοι για τη διάγνωση της COVID-19. Η πιο γνωστή και πιο ακριβής μέθοδος είναι η εργαστηριακή εξέταση RT-PCR (Reverse Transcription - Polymerase Chain Reaction). Η εξέταση αυτή, που ονομάζεται επίσης μοριακό τεστ, ανιχνεύει γενετικό υλικό του ιού SARS-CoV-2 σε κατάλληλο δείγμα που συλλέγεται από τη μύτη ή το λαιμό (ρινοφαρυγγικό ή στοματοφαρυγγικό επίχρισμα αντίστοιχα). Τα τεστ PCR είναι πολύ ακριβή και αξιόπιστα όταν εκτελούνται σωστά από έναν εκπαιδευμένο επαγγελματία υγείας. Επίσης, τα τεστ αυτά απαιτούν ειδικό εξοπλισμό μοριακών αναλυτών και εξειδικευμένα kits. Τα αποτελέσματά τους είναι συνήθως διαθέσιμα εντός μίας ημέρας, αλλά σε ορισμένες περιπτώσεις μπορεί να καθυστερήσουν αρκετές ημέρες, λόγω υπερφόρτωσης των κλινικών εργαστηρίων.

Μια δεύτερη διαγνωστική μέθοδος ανίχνευσης του ιού είναι το τεστ αντιγόνου ή rapid τεστ. Η εξέταση αυτή ανιχνεύει συγκεκριμένες πρωτεΐνες του ιού στο αντίστοιχο ρινοφαρυγγικό ή στοματοφαρυγγικό επίχρισμα. Ένα θετικό αποτέλεσμα στο τεστ αντιγόνου θεωρείται ακριβές όταν έχουν ακολουθηθεί προσεκτικά οι οδηγίες συλλογής του δείγματος, όμως υπάρχει μεγαλύτερη

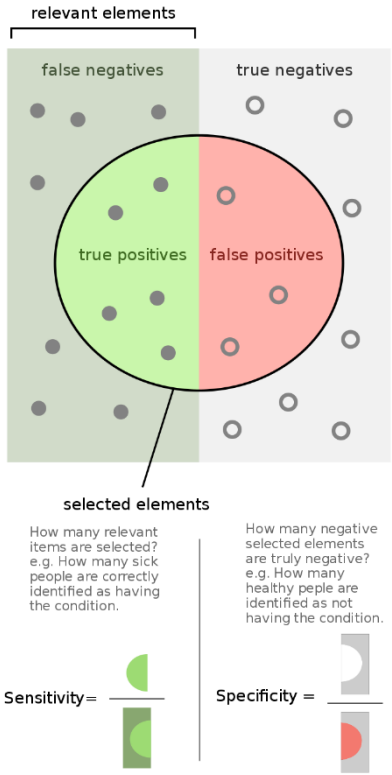
πιθανότητα ψευδώς αρνητικών αποτελεσμάτων σε σύγκριση με το τεστ RT-PCR. Αυτό σημαίνει ότι είναι πιθανό ένα μολυσμένο άτομο που κάνει το τεστ αντιγόνου να λάβει αρνητικό αποτέλεσμα. Συνήθως, τα τεστ αντιγόνου είναι πολύ γρήγορα και μπορούν να δώσουν αποτελέσματα σε 15-30 λεπτά.

Σε αυτή την ενότητα, θα περιγράψουμε εναλλακτικές μεθόδους διάγνωσης της ασθένειας COVID-19, οι οποίες χρησιμοποιούν διαφορετικά είδη δειγμάτων και σε συνδυασμό με τεχνικές τεχνητής νοημοσύνης μπορούν να προβλέψουν αν ένα άτομο έχει μολυνθεί με το νέο ιό. Οι μέθοδοι αυτές είναι ιδιαίτερα χρήσιμες κατά τις πρώτες φάσεις μιας νέας πανδημίας, όταν η δυνατότητα εξειδικευμένων εργαστηριακών δοκιμών είναι περιορισμένη, όπως συνέβη και στην περίπτωση της πανδημίας COVID-19.

3.2.2 Διάγνωση μέσω Ιατρικής Απεικόνισης

Η αυτοματοποιημένη διάγνωση με χρήση ιατρικής εικόνας αποτελεί μία από τις πιο επιτυχημένες εφαρμογές τεχνητής νοημοσύνης στον τομέα της Ιατρικής [18]. Πολλές ιατρικές ειδικότητες, συμπεριλαμβανομένης της ακτινολογίας, οφθαλμιατρικής, δερματολογίας και παθολογίας, βασίζονται σε διαγνώσεις μέσω ιατρικών εικόνων. Για παράδειγμα, οι ακτινολόγοι χρησιμοποιούν πολλά είδη ιατρικής εικόνας, όπως για παράδειγμα ακτινογραφίες ακτίνων Χ (X-Ray), αξονικές τομογραφίες (CT), απεικονίσεις μαγνητικού συντονισμού (MRI) και τομογραφίες εκπομπής ποζιτρονίων (PET), για τη διάγνωση ασθενειών. Οι μέθοδοι τεχνητής νοημοσύνης, ιδιαίτερα η βαθιά μάθηση, χρησιμοποιούνται ευρέως για την επεξεργασία και την ανάλυση δεδομένων ιατρικής απεικόνισης, βοηθώντας έτσι τους ιατρούς να βελτιώσουν την απόδοση των διαγνώσεών τους. Με την ενσωμάτωση σύγχρονων μεθόδων μηχανικής μάθησης, πολλές ακτινολογικές εφαρμογές, όπως η ανίχνευση οζιδίων των πνευμόνων με τη χρήση αξονικής τομογραφίας, η διάγνωση της φυματίωσης και άλλων πνευμονικών παθήσεων μέσω ακτινογραφίας θώρακος και η ανίχνευση όγκου μαστού με τη χρήση μαστογραφίας, έχουν πετύχει μεγάλη διαγνωστική ακρίβεια.

Ομοίως, αντίστοιχες μέθοδοι και τεχνολογίες τεχνητής νοημοσύνης έχουν προταθεί για τη διάγνωση της ασθένειας COVID-19, και συγκεκριμένα από ιατρικές εικόνες αξονικής τομογραφίας (CT) και ακτίνων Χ. Αν και οι ακτινολόγοι και οι κλινικοί ιατροί μπορούν να εκπαιδευθούν στην ανίχνευση περιπτώσεων COVID-19 μέσω αξονικής θώρακος, η διαδικασία αυτή είναι αρκετά χρονοβόρα, ιδιαίτερα όταν υπάρχουν πολλές περιπτώσεις υπό εξέταση. Σε μια σχετική μελέτη [62] έγινε ένα πείραμα, στο οποίο επτά ακτινολόγοι (τρεις από την Κίνα και τέσσερις από τις Ηνωμένες Πολιτείες) κλήθηκαν να μελετήσουν αξονικές θώρακος ασθενών και να διακρίνουν τις περιπτώσεις ασθενών με COVID-19 από άλλες περιπτώσεις ιογενούς πνευμονίας. Το δείγμα αποτελούνταν από 424 εικόνες, εκ των οποίων οι 219 αφορούσαν πνευμονία λόγω COVID-19 και οι υπόλοιπες 205 άλλου είδους ιογενή πνευμονία. Το πείραμα αυτό έδειξε ότι οι ακτινολόγοι κατάφεραν να επιτύχουν μια αρκετά καλή ειδικότητα (specificity) της τάξεως του 95%, η οποία αναφέρεται στο ποσοστό των αρνητικών περιπτώσεων που προσδιορίζονται ορθώς ως αρνητικές. Ωστόσο, η απόδοσή τους σε ό,τι αφορά την ευαισθησία (sensitivity), η οποία αναφέρεται στο ποσοστό των θετικών περιπτώσεων που προσδιορίζονται ορθώς ως θετικές, ήταν μέτρια και συγκεκριμένα λίγο κάτω από 80%. Το τελευταίο είναι καθοριστικής σημασίας, καθώς ουσιαστικά απέτυχαν να ανιχνεύσουν μία στις πέντε θετικές περιπτώσεις. Ο ορισμός της ευαισθησίας και της ειδικότητας σε προβλήματα απόφασης παρουσιάζεται στο Σχήμα 21.



$$\textbf{Sensitivity} = \frac{Number\ of\ True\ Positives}{Number\ of\ True\ Positives + Number\ of\ False\ Negatives}$$

$$= \frac{Number\ of\ True\ Positives}{Total\ Number\ of\ Individuals\ with\ the\ Illness}$$

$$\textbf{Specificity} = \frac{Number\ of\ True\ Negatives}{Number\ of\ True\ Negatives + Number\ of\ False\ Positives}$$

$$= \frac{Number\ of\ True\ Negatives}{Total\ Number\ of\ Individuals\ without\ the\ Illness}$$

Σχήμα 21 *- Ορισμός της ευαισθησίας (sensitivity) και της ειδικότητας (specificity) σε προβλήματα απόφασης*

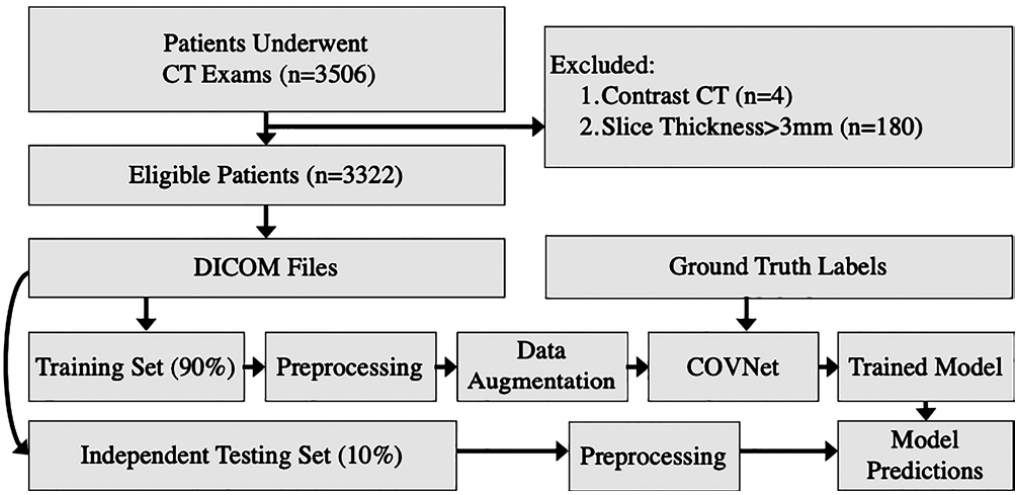
Διάγνωση με τη χρήση αξονικής τομογραφίας

Αρκετές εφαρμογές χρησιμοποιούν μηχανική μάθηση για τη διάγνωση της ασθένειας COVID-19 από αξονική τομογραφία μέσω ταξινόμησης των εικόνων σε διάφορες κλάσεις: υγιείς ασθενείς, ασθενείς με COVID-19 και ασθενείς με άλλους τύπους πνευμονίας. Ένα παράδειγμα αυτής της προσέγγισης ακολουθείται στο [63]. Σε αυτή τη μελέτη, μια τρισδιάστατη, επιτηρούμενη μέθοδος βαθιάς μάθησης, με την ονομασία COVNet (COVID-19 detection neural network) χρησιμοποιήθηκε για τη διάγνωση της ασθένειας COVID-19 με βάση ογκομετρικές εικόνες αξονικής θώρακος. Το COVNet είναι ένα συνελικτικό νευρωνικό δίκτυο τύπου ResNet-50, το οποίο λαμβάνει μια σειρά από τομές CT ως εισόδους και προβλέπει τις ετικέτες (κλάσεις) των αντίστοιχων εικόνων μέσω των εξόδων του (βλ. Σχήμα 22).

Για τη δοκιμή και την επιβεβαίωση της ικανότητας πρόβλεψης του προτεινόμενου μοντέλου, τρία είδη εικόνων CT από εξεταζόμενους σε νοσοκομεία της Κίνας, συμπεριλαμβανομένων εικόνων ασθενών με COVID-19, ασθενών με πνευμονία της κοινότητας (Community Acquired Pneumonia - CAP) και άλλων περιπτώσεων χωρίς πνευμονία, αναμείχθηκαν και δόθηκαν ως είσοδοι στο νευρωνικό δίκτυο. Η ακρίβεια αυτής της μεθόδου διάγνωσης αξιολογήθηκε με βάση το «εμβαδόν κάτω από την καμπύλη» (Area Under the Curve - AUC)[2]. Η υψηλή τιμή AUC που λήφθηκε (0,96),

[2] Το εμβαδόν κάτω από την καμπύλη (Area Under the Curve - AUC) είναι ένας δείκτης αξιολόγησης της απόδοσης ταξινόμησης (classification) ενός αλγορίθμου μηχανικής μάθησης και αντιπροσωπεύει ένα μέτρο της δυνατότητας διάκρισης μεταξύ των διαφορετικών κλάσεων. Όσο υψηλότερη είναι η τιμή του AUC (πιο κοντά στο 1) τόσο καλύτερο είναι το μοντέλο στη σωστή διάκριση των κλάσεων, εν προκειμένω των εξεταζόμενων ανάμεσα στις πιθανές διαγνώσεις.

κατέδειξε τη μεγάλη ικανότητα του προτεινόμενου μοντέλου για τη διάγνωση των περιπτώσεων ασθενών με COVID-19.



Σχήμα 22 *- Διάγραμμα ροής μιας μεθόδου βαθιάς μάθησης για τη διάγνωση COVID-19 με τη χρήση αξονικής τομογραφίας* [63]

Μια άλλη μέθοδος βαθιάς μάθησης που χρησιμοποιήθηκε για τη διάγνωση COVID-19 μέσω ταξινόμησης εικόνων CT θώρακος παρουσιάζεται στο [64]. Η εφαρμογή αυτή βασίζεται στη συνένωση μεταξύ δύο δικτύων, ενός Location Attention Network και ενός τρισδιάστατου συνελικτικού δικτύου τύπου ResNet-18. Το προτεινόμενο σύστημα βαθιάς μάθησης εκμεταλλεύεται τα ιδιαίτερα χαρακτηριστικά που διακρίνουν τις εικόνες CT ασθενών με COVID-19 έναντι άλλων ασθενών με διαφορετικούς τύπους ιογενούς πνευμονίας, όπως για παράδειγμα της πνευμονίας που προκαλεί η Γρίπη Α. Ένα σύνολο δεδομένων εισόδου που περιελάμβανε εικόνες CT ασθενών με COVID-19, ασθενών με ιογενή πνευμονία από Γρίπη Α, καθώς και περιπτώσεις υγιών ατόμων, χρησιμοποιήθηκε για την επικύρωση της απόδοσης της προτεινόμενης μεθόδου. Η συνολική ακρίβεια της μεθόδου (περίπου 86%) κρίνεται αρκετά ικανοποιητική και δείχνει την ικανότητά της να βοηθήσει τους κλινικούς ιατρούς στην έγκαιρη διάγνωση ασθενών με COVID-19.

Ορισμένες μελέτες υιοθετούν μια υβριδική προσέγγιση, συνδυάζοντας έτοιμο λογισμικό με εξειδικευμένα συστήματα μηχανικής μάθησης, με απώτερο στόχο την επίτευξη μεγαλύτερης διαγνωστικής ακρίβειας. Ένα παράδειγμα στην κατηγορία αυτή παρουσιάζεται στο [65], όπου ένα εμπορικό πρόγραμμα ιατρικής απεικόνισης (της εταιρείας RADlogics [66]) χρησιμοποιήθηκε για την αρχική επεξεργασία της ιατρικής εικόνας και στη συνέχεια, συνδυάστηκε με μια αλυσίδα μηχανικής μάθησης. Η αλυσίδα μηχανικής μάθησης αποτελούνταν από δύο βήματα. Το πρώτο αντιστοιχούσε σε ένα νευρωνικό δίκτυο αρχιτεκτονικής U-Net (ένα είδος συνελικτικού νευρωνικού δικτύου), το οποίο εκπαιδεύτηκε σε ιατρικές εικόνες που περιείχαν πνευμονικές ανωμαλίες, προκειμένου να εντοπίσει / απομονώσει τις περιοχές των πνευμόνων που παρουσίαζαν διαγνωστικό ενδιαφέρον. Το δεύτερο βήμα ήταν ένα νευρωνικό δίκτυο Resnet-50 που εκπαιδεύθηκε μέσω του ImagetNet (ενός μεγάλου online δικτύου εικόνων) και κατόπιν προσαρμόστηκε σε περιπτώσεις εικόνων ασθενών με COVID-19. Ο τελικός στόχος της συνολικής αλυσίδας συστημάτων ήταν η διάκριση των εικόνων μεταξύ ασθενών COVID-19 και υγιών ατόμων. Η προκύπτουσα αρχιτεκτονική ήταν σε θέση να εξάγει επιτυχώς τα σημαντικά χαρακτηριστικά των εικόνων και στη συνέχεια, να ανιχνεύσει περιπτώσεις πνευμονίας λόγω COVID-19. Η συγκεκριμένη εφαρμογή μπορεί να χρησιμοποιηθεί σε νοσοκομεία για την επιτάχυνση της διάγνωσης νέων περιπτώσεων.

Κάποιες μελέτες αναφέρουν εφαρμογή των μοντέλων τους σε πραγματικές συνθήκες, όπως για παράδειγμα στο [67]. Στη συγκεκριμένη εφαρμογή επιλέχθηκε η ενεργός συμμετοχή του ανθρώπου στη διαδικασία της διάγνωσης (human-in-the-loop), με ταυτόχρονη μείωση του απαιτούμενου χρόνου ανάλυσης των εικόνων μέσω τεχνολογιών μηχανικής μάθησης. Οι ερευνητές χρησιμοποίησαν ένα σύνολο από περίπου 2000 εικόνες, που είχαν επισημανθεί από ειδικούς με τις κατάλληλες ετικέτες (labels), για την εκπαίδευση ενός αρχικού μοντέλου βασισμένου στην αρχιτεκτονική V-Net (ένα είδος συνελικτικού νευρωνικού δικτύου που χρησιμοποιείται για την κατάτμηση ογκομετρικών εικόνων). Αυτό το μοντέλο ουσιαστικά προτείνει την κατάτμηση των αξονικών τομογραφιών, η οποία μπορεί στη συνέχεια να διορθωθεί από ακτινολόγους και να επανατροφοδοτήσει το μοντέλο, σε μια επαναληπτική διαδικασία. Μέσω αυτής της προσέγγισης ήταν δυνατή η ανάπτυξη ενός συστήματος βαθιάς μάθησης, τόσο για την αυτόματη κατάτμηση των εικόνων, όσο και για την καταμέτρηση των περιοχών μόλυνσης, και την επακόλουθη αξιολόγηση της σοβαρότητας της ασθένειας COVID-19 (μέσω υπολογισμού του ποσοστού μόλυνσης των πνευμόνων). Οι ερευνητές έδειξαν ότι το μοντέλο τους βελτίωσε σταδιακά την απόδοσή του και ταυτόχρονα ο χρόνος που απαιτούνταν για την ανάλυση νέων εικόνων από τους ειδικούς μειώθηκε από πάνω από 30 λεπτά αρχικά σε λιγότερο από 5 λεπτά μετά την εκπαίδευση του μοντέλου. Η μελέτη αυτή είναι ένα παράδειγμα μιας πολλά υποσχόμενης προσέγγισης που αξιοποιεί τη δύναμη της μηχανικής μάθησης παράλληλα με την εμπειρία του ανθρώπινου παράγοντα. Οι ρόλοι του ανθρώπου και της μηχανής σε εφαρμογές τεχνητής νοημοσύνης είναι συμπληρωματικοί και αμοιβαία επωφελείς, όπως θα συζητήσουμε και στο Κεφάλαιο 7.

Διάγνωση με τη χρήση ακτινογραφίας θώρακος

Η ιατρική απεικόνιση με τη χρήση ακτίνων Χ (X-Rays) και συγκεκριμένα οι ακτινογραφίες θώρακος μπορούν επίσης να χρησιμοποιηθούν για τη διάγνωση της ασθένειας COVID-19. Δεδομένης της μεγαλύτερης διαθεσιμότητας και της ευκολότερης φορητότητας του απαιτούμενου εξοπλισμού απεικόνισης σε σχέση με τους αξονικούς τομογράφους, οι ακτινογραφίες μπορούν να αποτελέσουν εναλλακτική λύση σε μέρη όπου η πρόσβαση σε προηγμένο ιατρικό εξοπλισμό είναι περιορισμένη. Επιπλέον, αν και οι ακτινογραφίες έχουν λιγότερη ευαισθησία από τις αξονικές τομογραφίες, είναι λιγότερο επεμβατικές, καθώς έχουν χαμηλότερη δόση ιοντίζουσας ακτινοβολίας. Έχει διαπιστωθεί ότι οι ασθενείς με COVID-19 παρουσιάζουν ανωμαλίες σε εικόνες ακτινογραφίας θώρακος που είναι χαρακτηριστικές της συγκεκριμένης νόσου και μπορούν να χρησιμοποιηθούν για τη διάγνωση της ασθένειας.

Χρησιμοποιώντας μια τέτοια λογική, ερευνητές ανέπτυξαν το COVID-Net, ένα βαθύ συνελικτικό νευρωνικό δίκτυο που μπορεί να διαγνώσει την ασθένεια COVID-19 από εικόνες ακτινογραφίας θώρακος [68]. Το δίκτυο αυτό εκπαιδεύτηκε σε δημόσια διαθέσιμα δεδομένα για περισσότερους από 13.000 ασθενείς με διάφορες πνευμονικές παθήσεις, συμπεριλαμβανομένης και της COVID-19. Όπως επισημαίνουν οι ερευνητές, αν και η λύση τους δεν είναι ακόμη έτοιμη για παραγωγική λειτουργία, εντούτοις παρέχει ένα επιπλέον εργαλείο στη διάθεση της επιστημονικής κοινότητας για περαιτέρω ανάπτυξη και βελτίωση, μέσω της διαθεσιμότητάς του σε μορφή ανοικτού κώδικα (open source). Στη λύση αυτή δόθηκε ιδιαίτερη σημασία στην πτυχή της επεξηγησιμότητας (explainability) του προτεινόμενου μοντέλου, με τον προσδιορισμό εκείνων των τμημάτων της ιατρικής εικόνας που επηρεάζουν περισσότερο την απόφαση του αλγορίθμου μηχανικής μάθησης. Για περισσότερες λεπτομέρειες σχετικά με την έννοια της επεξηγησιμότητας και το ρόλο που διαδραματίζει στη δημιουργία ισχυρών (εύρωστων) εφαρμογών τεχνητής νοημοσύνης (robust AI), μπορείτε να ανατρέξετε στην ενότητα 7.3.2.

Στην περιοχή αυτή εντάσσεται και η μελέτη ερευνητών του Ολλανδικού Πανεπιστημίου Delft που δημιούργησαν ένα μοντέλο βασισμένο σε τεχνητή νοημοσύνη για τη διάγνωση της ασθένειας

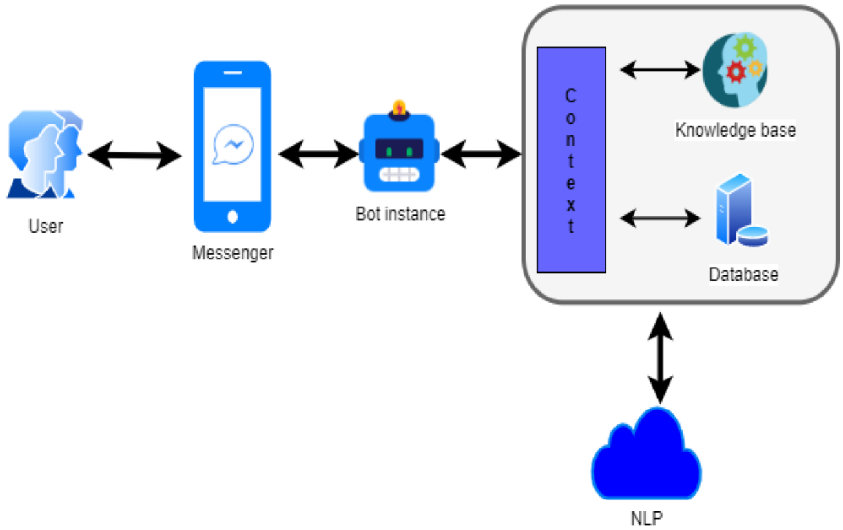
COVID-19 από ακτινογραφίες θώρακος [69]. Το μοντέλο αυτό, με την ονομασία CAD4COVID, βασίζεται σε προηγούμενα μοντέλα τεχνητής νοημοσύνης που αναπτύχθηκαν από το συγκεκριμένο πανεπιστήμιο για τη διάγνωση της φυματίωσης. Το εν λόγω σύστημα κατάφερε να ταξινομήσει ακτινογραφίες θώρακος ως πνευμονία λόγω COVID-19 με αρκετά ικανοποιητική ακρίβεια, όπως υποδεικνύεται από το εμβαδόν κάτω από την καμπύλη (AUC = 0,81), η οποία ήταν συγκρίσιμη με αυτήν που πέτυχαν έξι ανεξάρτητοι ειδικοί ακτινολόγοι.

Πέραν όμως των ερευνητικών εφαρμογών, υπάρχουν επίσης εμπορικές λύσεις λογισμικού που υλοποιήθηκαν από διάφορες εταιρείες, μερικές από τις οποίες είναι επίσης ελεύθερα διαθέσιμες για περιορισμένη χρήση, όπως για παράδειγμα το «Lunit Insight CXR for COVID-19». Η λύση αυτή επιτρέπει το δωρεάν ανέβασμα έως και 20 περιπτώσεων προς διάγνωση ημερησίως και έχει εγκριθεί και χρησιμοποιείται από διάφορα νοσοκομεία στην Κορέα [70]. Μια άλλη εμπορική λύση είναι το «VUN Med LungQuant» [71] που μπορεί να χρησιμοποιήσει τόσο εικόνες αξονικής τομογραφίας όσο και ακτινογραφίες.

3.2.3 Ψηφιακές Μέθοδοι Ταχείας Διάγνωσης (Screening) με Χρήση Chatbots και Internet of Things

Chatbots με δυνατότητες τεχνητής νοημοσύνης

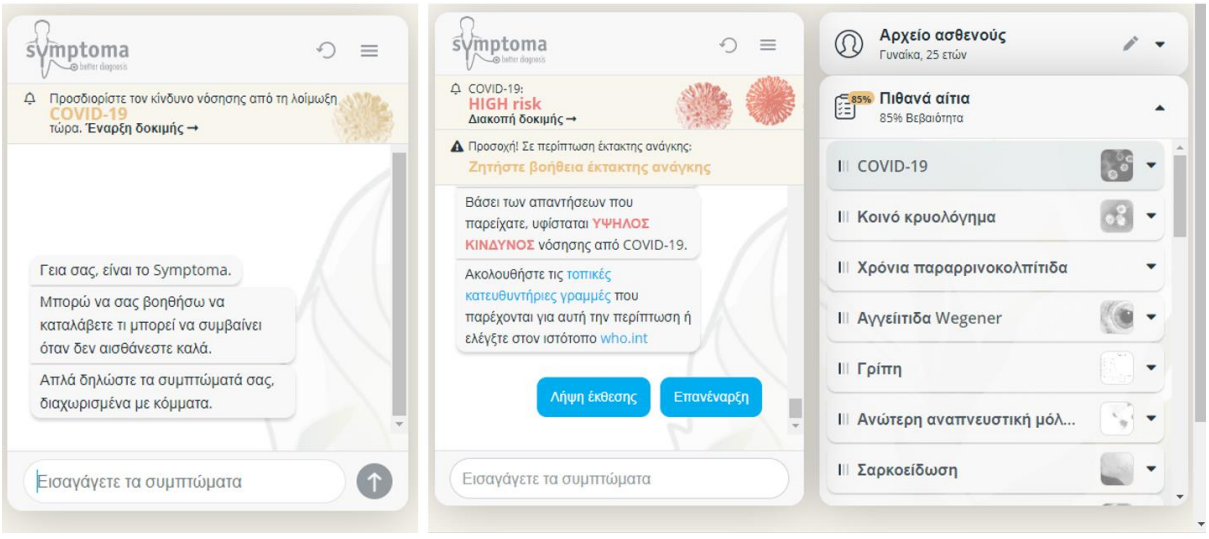
Στα πλαίσια της προσπάθειας για την αντιμετώπιση της πανδημίας COVID-19, οι περισσότερες κυβερνήσεις κρατών ανά τον κόσμο έχουν δημιουργήσει τηλεφωνικές γραμμές διαθέσιμες στο κοινό για την ανίχνευση πιθανών κρουσμάτων μέσω ερωτήσεων στους καλούντες. Αυτές οι ανοικτές γραμμές επικοινωνίας έχουν συνήθως να διαχειριστούν έναν τεράστιο όγκο κλήσεων, με αποτέλεσμα πολλές φορές ο χρόνος αναμονής να είναι πολύ μεγάλος και σε κάποιες περιπτώσεις να υπάρχει ακόμα και πλήρης αδυναμία επικοινωνίας με τις υγειονομικές αρχές. Τελικά, αυτό έχει ως αποτέλεσμα πολλά κρούσματα COVID-19 να μη διαγιγνώσκονται. Εναλλακτικές μέθοδοι, όπως για παράδειγμα οι ψηφιακοί βοηθοί (chatbots / virtual assistants) θα μπορούσαν δυνητικά να απορροφήσουν ένα μεγάλο μέρος της πίεσης από τα τηλεφωνικά κέντρα. Είναι προφανές ότι αν σχεδιαστεί και αναπτυχθεί αποτελεσματικά, ένα chatbot μπορεί να βοηθήσει τους πολίτες μέσω της παροχής ενός αρχικού ελέγχου για πιθανή νόσηση από COVID-19, της ενημέρωσης σε σχέση με κατάλληλα προστατευτικά μέτρα, καθώς και της παροχής ιατρικών συμβουλών και άλλων πληροφοριών σχετικά με το νέο ιό.



Σχήμα 23 *- Παράδειγμα αρχιτεκτονικής ενός chatbot με δυνατότητες τεχνητής νοημοσύνης* [72]

Οι ερευνητές στο [72] περιγράφουν το σχεδιασμό ενός chatbot με δυνατότητες τεχνητής νοημοσύνης για σκοπούς μιας αρχικής διάγνωσης καθώς και παροχής συμβουλών σε άτομα που πιθανώς έχουν εκτεθεί στον ιό. Αρχικά, το chatbot αξιολογεί τη σοβαρότητα της εκάστοτε περίπτωσης μέσω των απαντήσεων του χρήστη σε ένα προκαθορισμένο ερωτηματολόγιο. Στην συνέχεια, αφού αξιολογήσει την κατάσταση του χρήστη, παρέχει μια απάντηση με τη μορφή είτε γενικού κειμένου είτε ενός άρθρου που ανακτά από τη γνωσιακή του βάση (knowledge base). Το συγκεκριμένο chatbot χρησιμοποιεί κατάλληλη τεχνολογία για επεξεργασία φυσικής γλώσσας (NLP), έτσι ώστε να μπορεί να κατανοήσει την είσοδο που προέρχεται από τις απαντήσεις του χρήστη και στη συνέχεια, συντάσσει την κατάλληλη απάντηση. Στις περιπτώσεις εκείνες που το chatbot αξιολογεί ότι υπάρχουν συμπτώματα σοβαρής μορφής, μπορεί να φέρει τον αντίστοιχο χρήστη σε επαφή με το κατάλληλο υγειονομικό προσωπικό.

Η διαδικτυακή εφαρμογή Symptoma [73] είναι ένας επιπρόσθετος ψηφιακός βοηθός υγείας που μπορεί να ανιχνεύσει ασθένειες από τα συμπτώματά τους. Αναπτύχθηκε από ερευνητές στην Αυστρία και, σύμφωνα με τις μελέτες τους, μπορεί να ανιχνεύσει περισσότερες από 20.000 ασθένειες με ακρίβεια άνω του 90%. Το chatbot Symptoma θέτει στον χρήστη αρκετές ερωτήσεις, οι οποίες προσαρμόζονται σταδιακά καθώς ο χρήστης παρέχει απαντήσεις. Οι περισσότερες από τις ερωτήσεις αυτές απαντώνται διαλέγοντας μία από τις διαθέσιμες επιλογές ενώ κάποιες άλλες μέσω ελεύθερου κειμένου. Αφού ο χρήστης απαντήσει σε όλες τις ερωτήσεις, το Symptoma παρέχει τη διάγνωσή του μαζί με τον αντίστοιχο βαθμό βεβαιότητας, υπό τη μορφή μιας ταξινομημένης λίστας πιθανών αιτιών. Στο [74], οι ερευνητές αξιολόγησαν την απόδοση του Symptoma στη διάγνωση της ασθένειας COVID-19, χρησιμοποιώντας ένα σύνολο διαφορετικών κλινικών περιπτώσεων και απέδειξαν ότι το Symptoma μπορεί να ανιχνεύσει με ακρίβεια την ασθένεια COVID-19 στο 96,32% των περιπτώσεων. Σήμερα το Symptoma είναι διαθέσιμο σε 36 γλώσσες (συμπεριλαμβανομένων των ελληνικών) και εξυπηρετεί περίπου 10 εκατομμύρια χρήστες κάθε μήνα.



Σχήμα 24 *- Το chatbot Symptoma στη διάγνωση της COVID-19* [73]

Διάγνωση με χρήση εναλλακτικών πηγών δεδομένων

Υπάρχει μια σειρά από ερευνητικές προσπάθειες και αντίστοιχες εφαρμογές για την αρχική διάγνωση της ασθένειας COVID-19 που δεν απαιτούν εξειδικευμένο ιατρικό εξοπλισμό εργαστηριακών εξετάσεων ή απεικόνισης, αλλά βασίζονται σε δεδομένα που συλλέγονται από αισθητήρες έξυπνων

κινητών τηλεφώνων ή wearable συσκευών. Μερικά αντιπροσωπευτικά παραδείγματα τέτοιων εφαρμογών παρουσιάζονται στην παρούσα ενότητα. Αξίζει να σημειωθεί ότι οι προσεγγίσεις αυτές δεν αποσκοπούν στο να υποκαταστήσουν τις ιατρικές μεθόδους διάγνωσης, αλλά αντίθετα μπορούν να χρησιμοποιηθούν για ελέγχους μεγάλης κλίμακας μέσω της ανίχνευσης πιθανών περιπτώσεων COVID-19 που χρήζουν περαιτέρω διερεύνησης. Τα σύγχρονα κινητά τηλέφωνα, καθώς και πολλές wearables συσκευές (π.χ. έξυπνα ρολόγια) διαθέτουν ισχυρούς επεξεργαστές, μεγάλο χώρο μνήμης και κυρίως μεγάλο αριθμό αισθητήρων, συμπεριλαμβανομένων καμερών, μικροφώνου, αισθητήρων θερμοκρασίας, αδράνειας, εγγύτητας, χρώματος, υγρασίας. Επίσης, οι συσκευές αυτές έχουν δυνατότητες ασύρματης συνδεσιμότητας στο διαδίκτυο. Η αξιοποίηση των πλούσιων δεδομένων που συλλέγονται από τους συγκεκριμένους αισθητήρες και η επεξεργασία τους με αλγορίθμους μηχανικής μάθησης, είτε στην ίδια συσκευή είτε στο cloud, με στόχο την ανίχνευση συμπτωμάτων της ασθένειας COVID-19, αποτελεί αντικείμενο πολλών ερευνητικών έργων, μερικά παραδείγματα των οποίων περιγράφουμε παρακάτω.

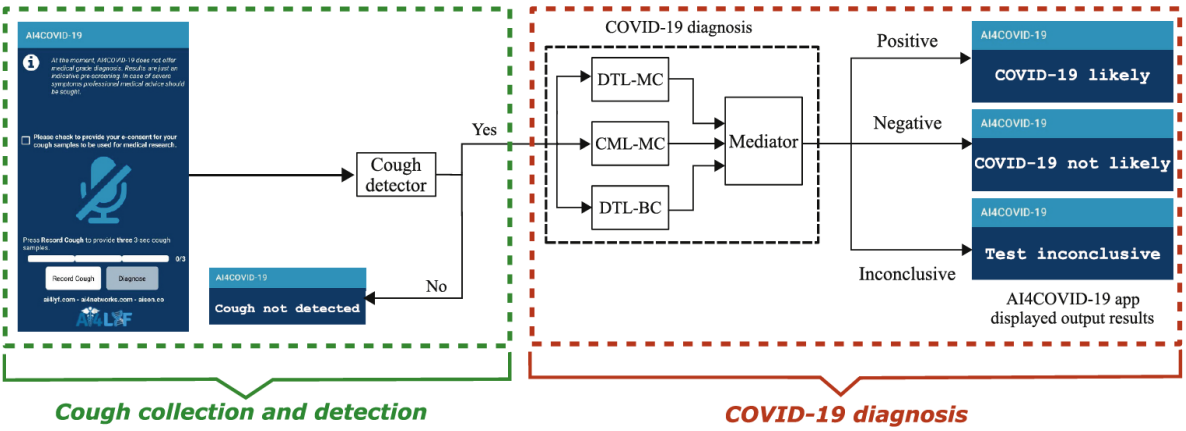
Όπως έχει ήδη αναφερθεί, τα πιο κοινά συμπτώματα της ασθένειας COVID-19 συνδέονται με μιας μορφής πνευμονία. Η ανάλυση του ήχου από την αναπνοή, την ομιλία και το βήχα αποτελεί μια πολλά υποσχόμενη μέθοδο διάγνωσης. Αντίστοιχες προσπάθειες είχαν λάβει χώρα και πριν την πανδημία COVID-19. Για παράδειγμα, οι ερευνητές στο [75] ανέπτυξαν μια χαμηλού κόστους εφαρμογή για κινητά τηλέφωνα με στόχο την ταχεία διάγνωση αναπνευστικών προβλημάτων. Το μοντέλο μηχανικής μάθησης της συγκεκριμένης εφαρμογής εκπαιδεύθηκε με 367 ήχους αναπνοής που συλλέχθηκαν από νοσοκομεία. Τα αποτελέσματα του μοντέλου έδειξαν ότι, ακόμα και παρουσία περιβαλλοντικών θορύβων, η μέθοδος ήταν ακριβής σε ποσοστό μεγαλύτερο του 90%, τόσο σε σχέση με την ευαισθησία όσο και με την ειδικότητά της, στην ανίχνευση πνευμονίας.

Μια εφαρμογή που εκμεταλλεύεται τις αλλαγές που προκαλεί η ασθένεια COVID-19 στα χαρακτηριστικά της αναπνοής, καθώς και στη χροιά της φωνής, παρουσιάζεται στο [76]. Σε αυτή τη μελέτη, οι ερευνητές αξιολογούν ποιες από τις μεθόδους υπολογιστικής ακοής (computer audition), που ορίζεται ως η ανάλυση της ομιλίας και των ήχων με τεχνικές τεχνητής νοημοσύνης, μπορούν να χρησιμοποιηθούν για την αρχική διάγνωση και την αξιολόγηση κινδύνου για την ασθένεια COVID-19. Επιπλέον, προτείνουν και άλλες εφαρμογές της μεθόδου αυτής, όπως για παράδειγμα την παρακολούθηση της εξάπλωσης στην κοινότητα με βάση την ακουστική ανάλυση προφορικών συνομιλιών μέσω δικτύων τηλεφωνίας. Συγκεκριμένα, σε ό,τι αφορά τη διάγνωση, οι ερευνητές πιστεύουν ότι μια τέτοια μέθοδος – αν και δεν υποκαθιστά τις επαγγελματικές μεθόδους διάγνωσης – είναι πολύ χρήσιμη, εφόσον είναι μη επεμβατική, είναι διαθέσιμη ουσιαστικά σε όλους μέσω κινητού τηλεφώνου και άμεση, καθώς παρέχει τα αποτελέσματά της σε πραγματικό χρόνο.

Εκτός από τις αλλαγές στη φωνή και την αναπνοή, και ο βήχας των ασθενών με COVID-19 έχει κάποια ιδιαίτερα χαρακτηριστικά που δυνητικά μπορούν να χρησιμοποιηθούν για σκοπούς αρχικής διάγνωσης. Όπως αναφέραμε και στο Κεφάλαιο 1, ένα σημαντικό ποσοστό των ανθρώπων που μολύνονται με τον ιό SARS-CoV-2 είναι ασυμπτωματικοί φορείς του ιού. Όπως είναι λογικό, τα άτομα αυτά είναι λιγότερο πιθανό να υποβληθούν σε κάποιο εργαστηριακό τεστ, με συνέπεια πολλές φορές να μεταδίδουν εν αγνοία τους τον ιό σε άλλους. Έχει φανεί, όμως, ότι ακόμα και αυτοί που είναι φαινομενικά ασυμπτωματικοί, μπορεί να μην είναι εντελώς απαλλαγμένοι από μεταβολές που προκαλούνται από τον ιό. Συγκεκριμένα, διάφοροι ερευνητές διαπίστωσαν ότι ακόμα και οι ασυμπτωματικοί φορείς μπορεί να διαφέρουν από τα πλήρως υγιή άτομα στον τρόπο με τον οποίο βήχουν. Αυτές οι διαφορές δε γίνονται συνήθως αντιληπτές από το ανθρώπινο αυτί, αλλά έχει αποδειχθεί ότι μπορούν να ανιχνευθούν με χρήση τεχνητής νοημοσύνης.

Στο [77] οι ερευνητές βασίστηκαν σε προηγούμενη έρευνα στον τομέα της διάγνωσης αναπνευστικών ασθενειών με βάση το βήχα και ανέπτυξαν μια μέθοδο ανίχνευσης της ασθένειας COVID-19 μέσω

μιας εφαρμογής για κινητά τηλέφωνα. Η εφαρμογή τους, με την ονομασία AI4COVID-19, καταγράφει και στέλνει 3 δείγματα ήχων βήχα διάρκειας 3 δευτερολέπτων το καθένα σε ένα μοντέλο μηχανικής μάθησης στο cloud, το οποίο επιστρέφει το αποτέλεσμα της διάγνωσης μέσα σε 2 λεπτά (βλ. Σχήμα 25). Για να ξεπεράσουν την έλλειψη δεδομένων εκπαίδευσης του μοντέλου με δείγματα βήχα ασθενών με COVID-19, οι ερευνητές εκμεταλλεύθηκαν μια τεχνική βαθιάς μάθησης που ονομάζεται μεταφορά μάθησης (transfer learning)[3]. Τα αποτελέσματά τους έδειξαν ότι η εφαρμογή AI4COVID-19 κατάφερε να διακρίνει το βήχα ασθενούς με COVID-19 από άλλους τύπους βήχα με αξιοσημείωτη ακρίβεια. Πέραν της χρήσης της ως μια γρήγορη, άμεση και χωρίς κόστος μέθοδος αρχικής διάγνωσης, η εφαρμογή αυτή μπορεί να συντελέσει και στην αποσυμφόρηση του συστήματος υγείας κατά τα αρχικά στάδια της πανδημίας, όπου οι ανάγκες για διαγνωστικά τεστ συχνά ξεπερνούν τις δυνατότητες των εργαστηριακών υποδομών.



Σχήμα 25 *- Αρχιτεκτονική της εφαρμογής AI4COVID-19 για διάγνωση COVID-19 από δείγματα βήχα* [77]

Σε μια ακόμα μελέτη που δημοσιεύθηκε πρόσφατα [78], μια ομάδα από το MIT δημιούργησε μια εφαρμογή τεχνητής νοημοσύνης βασισμένη σε μια αρχιτεκτονική συνελικτικού νευρωνικού δικτύου για τη διάγνωση ατόμων που έχουν μολυνθεί με τον ιό SARS-CoV-2. Το μοντέλο αυτό είναι σε θέση να διακρίνει ασυμπτωματικούς φορείς COVID-19 από υγιή άτομα μέσω δειγμάτων ήχου βήχα, τα οποία υποβάλλουν εθελοντικά οι χρήστες μέσω web browser. Το μοντέλο τους εκπαιδεύθηκε σε δεκάδες χιλιάδες δείγματα βήχα, καθώς και σε ήχους από ομιλία. Κατά τη φάση δοκιμών σε νέα δείγματα βήχα, η εφαρμογή κατάφερε να ανιχνεύσει με ακρίβεια το 98,5% των επιβεβαιωμένων περιπτώσεων COVID-19, συμπεριλαμβανομένου του 100% των ασυμπτωματικών φορέων (άτομα που ανέφεραν ότι δεν είχαν συμπτώματα, αλλά βρέθηκαν θετικοί στον ιό). Οι ερευνητές πρόκειται σύντομα να ενσωματώσουν το μοντέλο τους και σε εφαρμογές για κινητά τηλέφωνα.

Ακόμα μία ενδιαφέρουσα εφαρμογή αναπτύχθηκε από ερευνητές του Northwestern University και του Shirley Ryan AbilityLab στο Σικάγο. Αυτοί ανέπτυξαν κάποιες καινοτόμες wearable συσκευές σε συνδυασμό με ένα σύνολο αλγορίθμων ανάλυσης δεδομένων, που αποσκοπούν στην ανίχνευση πρώιμων συμπτωμάτων που σχετίζονται με την ασθένεια COVID-19, αλλά και στην παρακολούθηση των μολυσμένων ατόμων κατά τη διάρκεια της ασθένειάς τους [79]. Οι συσκευές που σχεδίασαν μπορούν να φοριούνται καθ' όλη τη διάρκεια της ημέρας, παράγοντας συνεχείς ροές δεδομένων. Κάνοντας χρήση τεχνητής νοημοσύνης μπορούν να αποκαλύψουν ήπια, αλλά δυνητικά σωτήρια

[3] Η μεταφορά μάθησης ή επαγωγική μάθηση (transfer or inductive learning) είναι μια τεχνική επιτηρούμενης μάθησης που επαναχρησιμοποιεί τμήματα ενός ήδη εκπαιδευμένου μοντέλου σε ένα νέο μοντέλο επιφορτισμένο με ένα διαφορετικό αλλά παρεμφερές πρόβλημα.

πρώιμα συμπτώματα. Οι αισθητήρες των συσκευών μετρούν συνεχώς το βήχα και την αναπνευστική δραστηριότητα με τρόπους που παραδοσιακά συστήματα παρακολούθησης αδυνατούν να πετύχουν. Η πρώτη συσκευή έχει το μέγεθος ενός γραμματοσήμου, είναι εύκαμπτη και διαθέτει δυνατότητα ασύρματης επικοινωνίας. Τοποθετείται ακριβώς κάτω από την υπερστερνική εντομή (όπως φαίνεται και στο Σχήμα 26). Από τη θέση αυτή, η συσκευή παρακολουθεί την ένταση και τα χαρακτηριστικά του βήχα, καθώς και τις κινήσεις του θωρακικού τοιχώματος (που υποδεικνύουν επίπονη ή ακανόνιστη αναπνοή), τους ήχους της αναπνοής, τον καρδιακό ρυθμό και τη θερμοκρασία του σώματος. Μια δεύτερη συσκευή φοριέται γύρω από το δάχτυλο και μετρά τα επίπεδα οξυγόνου στο αίμα και την πίεση του αίματος. Και οι δύο συσκευές μεταδίδουν ασύρματα τα κρυπτογραφημένα δεδομένα των μετρήσεών τους σε μια τοποθεσία στο cloud, όπου αλγόριθμοι μηχανικής μάθησης παράγουν διάφορες γραφικές απεικονίσεις των αποτελεσμάτων, με στόχο την εύκολη και γρήγορη απομακρυσμένη παρακολούθηση. Η συσκευή μπορεί να παρακολουθεί τους ασθενείς τόσο κατά τη διάρκεια της νοσηλείας τους όσο και στη συνέχεια, κατά τη διάρκεια της ανάρρωσής τους στο σπίτι, 24 ώρες το 24ωρο, 7 ημέρες την εβδομάδα.

Εφαρμογές όπως η πιο πάνω δημιουργούν νέες προοπτικές στον τομέα της τηλεϊατρικής σε ό,τι αφορά τη διαχείριση ασθενών με COVID-19, καθώς πλέον ένα μεγάλο ποσοστό των ασθενών δε θα χρειάζεται να πηγαίνει στο νοσοκομείο για παρακολούθηση. Επίσης, μπορούν να χρησιμοποιηθούν από τους ίδιους τους εργαζομένους στους τομείς υγείας πρώτης γραμμής, οι οποίοι διατρέχουν μεγάλο κίνδυνο να μολυνθούν με τον ιό, για να τους παρέχουν σήματα έγκαιρης προειδοποίησης πριν την εκδήλωση εμφανών συμπτωμάτων. Με αυτό τον τρόπο, θα έχουν τη δυνατότητα να λάβουν τα κατάλληλα μέτρα και να υποβληθούν σε τεστ το συντομότερο δυνατό.

Σχήμα 26 - *Φορητές συσκευές ανίχνευσης συμπτωμάτων COVID-19* [79]

Μια προτεινόμενη σχεδίαση εφαρμογής για κινητά τηλέφωνα που αποσκοπεί στην ανίχνευση πιθανών κρουσμάτων COVID-19 με χρήση ενσωματωμένων αισθητήρων, όπως κάμερες, μικρόφωνα, αισθητήρες θερμοκρασίας και αδράνειας, παρουσιάζεται στο [80]. Στη μελέτη αυτή, διάφορες μέθοδοι μηχανικής μάθησης χρησιμοποιούνται για την εκπαίδευση των μοντέλων με βάση δεδομένα που συλλέγονται από τους αισθητήρες και ενδέχεται να υποδηλώνουν συμπτώματα της νόσου. Η συγκεκριμένη προσέγγιση επαναχρησιμοποιεί διάφορα στοιχεία από εφαρμογές που έχουν χρησιμοποιηθεί στο παρελθόν για να λύσουν επιμέρους προβλήματα. Για παράδειγμα, τα δεδομένα που λαμβάνονται από τον αισθητήρα θερμοκρασίας / δακτυλικού αποτυπώματος χρησιμοποιούνται για την ανίχνευση πυρετού. Οι εικόνες και τα βίντεο που λαμβάνονται από την κάμερα των κινητών συσκευών και δεδομένα που συλλέγονται από τον αισθητήρα αδράνειας χρησιμοποιούνται για την ανίχνευση κόπωσης, που επίσης είναι ένα σύνηθες σύμπτωμα ασθενών με COVID-19.

Μια άλλη εφαρμογή για εναλλακτική διάγνωση της ασθένειας COVID-19 χρησιμοποιεί εικόνες από κάμερες βάθους της συσκευής Microsoft Kinect, καθώς και νευρωνικά δίκτυα τύπου GRU (Gated Recurrent Unit), για την ανίχνευση των χαρακτηριστικών αναπνοής των ασθενών [81]. Η μελέτη αυτή βασίστηκε σε πρόσφατα ευρήματα που έδειξαν ότι ένας ασθενής με COVID-19 έχει χαρακτηριστικά αναπνοής που είναι διακριτά από εκείνα της γρίπης και του κοινού κρυολογήματος, κυρίως σε ό,τι αφορά την εμφάνιση ταχύπνοιας. Η πρόβλεψη της ταχύπνοιας θα μπορούσε να είναι ένα πρώτο διαγνωστικό χαρακτηριστικό για χρήση σε ελέγχους μεγάλης κλίμακας.

3.3 Πρόγνωση

3.3.1 Η Σημασία της Πρόγνωσης

Αμέσως μόλις ένα άτομο διαγνωστεί θετικό στην ασθένεια COVID-19, το κύριο ερώτημα που ανακύπτει είναι πόσο σοβαρά θα επηρεαστεί από τον ιό. Όπως συζητήθηκε και στο Κεφάλαιο 1, τα άτομα που μολύνονται από το νέο κορωνοϊό εμφανίζουν διάφορα επίπεδα σοβαρότητας της νόσου, τα οποία κυμαίνονται από την πλήρη απουσία εμφανών συμπτωμάτων (ασυμπτωματικοί φορείς) μέχρι την εκδήλωση σοβαρής μορφής ασθένειας, που απαιτεί εισαγωγή σε μονάδα εντατικής θεραπείας και διασωλήνωση.

Είναι προφανές ότι ο έγκαιρος εντοπισμός των ασθενών που διατρέχουν μεγαλύτερο κίνδυνο εκδήλωσης σοβαρής νόσου έχει πολλαπλά οφέλη, συμπεριλαμβανομένης της βελτιωμένης κατανομής των πόρων του συστήματος υγείας και του προσωπικού, της δυνατότητας επικοινωνίας με άτομα που δεν έχουν εισαχθεί ακόμη στο νοσοκομείο αλλά χρήζουν διεξοδικής αξιολόγησης, καθώς και της μεταφοράς ασθενών που ήδη νοσηλεύονται σε άλλα νοσοκομεία που μπορούν να τους παρέχουν υψηλότερο επίπεδο φροντίδας και έχουν τη δυνατότητα να αντιμετωπίσουν δύσκολα περιστατικά.

Σε περιόδους όπου το σύστημα υγείας δέχεται μεγάλη πίεση λόγω του μεγάλου πλήθους των κρουσμάτων και των επακόλουθων εισαγωγών στα νοσοκομεία, η πρόβλεψη της πιθανής εξέλιξης (πρόγνωση) του κάθε ασθενούς πρέπει να γίνει γρήγορα και με όσο το δυνατό μεγαλύτερη ακρίβεια, αποτελώντας μια σημαντική πρόκληση για τους επαγγελματίες στον τομέα της υγείας. Σε αυτή τη μάχη, αυτοματοποιημένες προσεγγίσεις που εκμεταλλεύονται μεθόδους τεχνητής νοημοσύνης και συγκεκριμένα μηχανικής μάθησης μπορούν να συμβάλουν καθοριστικά. Στην ενότητα αυτή παρουσιάζουμε ορισμένες ενδιαφέρουσες ερευνητικές εργασίες και εφαρμογές στον τομέα της πρόγνωσης.

3.3.2 Πρόβλεψη της Σοβαρότητας της Νόσου

Πολλά συστήματα υποστήριξης αποφάσεων (decision support systems) βασισμένα στην τεχνητή νοημοσύνη υπήρχαν και πριν την εμφάνιση της ασθένειας COVID-19, κυρίως για μια σειρά χρόνιων ασθενειών. Τα συστήματα αυτά εκμεταλλεύονται την ικανότητα των μοντέλων μηχανικής μάθησης να «μαθαίνουν» πώς τα χαρακτηριστικά ενός ατόμου (παράγοντες κινδύνου), που περιλαμβάνουν κλινικές και άλλες πληροφορίες, μπορούν να αντιστοιχιστούν σε εξατομικευμένες προβλέψεις για την εξέλιξη μιας ασθένειας. Για παράδειγμα, διάφορες μελέτες έχουν εστιάσει στην εύρεση αλγοριθμικών εκτιμήσεων κινδύνου ασθενειών, όπως ο καρκίνος [82] και ο διαβήτης [83], με τη χρήση νευρωνικών δικτύων.

Επιπλέον, διάφοροι αλγόριθμοι μηχανικής μάθησης έχουν χρησιμοποιηθεί στο παρελθόν για την πρόβλεψη της εξέλιξης μεταδοτικών ασθενειών. Χαρακτηριστικό παράδειγμα αποτελεί μια μελέτη για την πρόγνωση ασθενών μου μολύνθηκαν με έναν άλλο κορωνοϊό, και συγκεκριμένα με τον

MERS-CoV [84]. Σε αυτή την εργασία, τα τέσσερα χαρακτηριστικά που αξιολογήθηκαν ως οι σημαντικότεροι παράγοντες πρόγνωσης (predictors) για την ανάκαμψη ενός ασθενούς ήταν η ηλικία του, η σοβαρότητα των συμπτωμάτων του κατά την εισαγωγή του στο κέντρο υγειονομικής περίθαλψης, το αν ο ασθενής ήταν εργαζόμενος στον τομέα της υγείας και η ύπαρξη συγκεκριμένων υποκείμενων νοσημάτων. Μια άλλη εφαρμογή με την ονομασία Ebola CARE (Computational Assignment of Risk Estimates) αναπτύχθηκε από μια άλλη ερευνητική ομάδα με στόχο την πρόβλεψη την έκβασης ενός ασθενούς που είχε μολυνθεί από τον ιό Έμπολα [85]. Η συγκεκριμένη εφαρμογή εντόπισε 24 κλινικές και εργαστηριακές παραμέτρους που δυνητικά επηρεάζουν την πρόγνωση ενός ασθενούς. Προσεγγίσεις σαν τις πιο πάνω θα μπορούσαν θεωρητικά να προσαρμοστούν για την πρόγνωση ασθενών με COVID-19. Στη συνέχεια, παρουσιάζουμε κάποιες ενδιαφέρουσες νέες εφαρμογές πρόγνωσης ασθενών με COVID-19 που κάνουν χρήση αλγορίθμων ανάλυσης δεδομένων και μηχανικής μάθησης.

Μια ευρέως διαδεδομένη πλατφόρμα λογισμικού που χρησιμοποιείται σήμερα, σε πάνω από 100 νοσοκομεία στις ΗΠΑ, είναι το Epic Deterioration Index (EDI) [86]. Πρόκειται για ένα εμπορικό πακέτο λογισμικού που βασίζεται στη μηχανική μάθηση και χρησιμοποιείται για την πρόβλεψη της θνητότητας από COVID-19, καθώς και για την παροχή πρακτικών συμβουλών στους νοσοκομειακούς ιατρούς. Το EDI συγκεντρώνει και αξιολογεί παράγοντες από τον ιατρικό φάκελο ενός νοσηλευόμενου ασθενούς, όπως η ηλικία, η κατάσταση των ζωτικών του λειτουργιών και τα αποτελέσματα των εργαστηριακών του εξετάσεων, παράγοντας τελικά μια συνολική βαθμολογία κινδύνου (risk score), που αντιστοιχεί στην πιθανότητα σοβαρής επιδείνωσης του ασθενούς. Οι κλινικοί ιατροί μπορούν να χρησιμοποιήσουν αυτή τη βοηθητική πληροφορία για να αποφασίσουν ποιοι ασθενείς θα πρέπει να παρακολουθούνται στενά και ποιοι θα πρέπει να μεταφερθούν σε μονάδες εντατικές θεραπείας.

Μια πρόσφατη ανεξάρτητη μελέτη από το Πανεπιστήμιο του Michigan αξιολόγησε την ακρίβεια και την αποτελεσματικότητα του EDI ανάμεσα σε ασθενείς με COVID-19 σε ένα μεγάλο νοσοκομείο των ΗΠΑ. Οι ερευνητές χρησιμοποίησαν το EDI, υπολογισμένο σε τακτά χρονικά διαστήματα των 15 λεπτών, για να προβλέψουν ένα συνδυαστικό αποτέλεσμα σε σχέση με την πιθανότητα (α) εισαγωγής σε ΜΕΘ, (β) μηχανικής υποστήριξης της αναπνοής και (γ) θανάτου. Τα αποτελέσματά τους έδειξαν ότι το EDI είχε σχετικά υψηλή θετική προγνωστική τιμή (πιθανότητα εμφάνισης του πρώτου σε σειρά αποτελέσματος που προέβλεπε το εργαλείο) και συγκεκριμένα γύρω στο 74% για ασθενείς που χαρακτηρίστηκαν ως υψηλού κινδύνου, ωστόσο, η ευαισθησία του ήταν μόλις 39% [87].

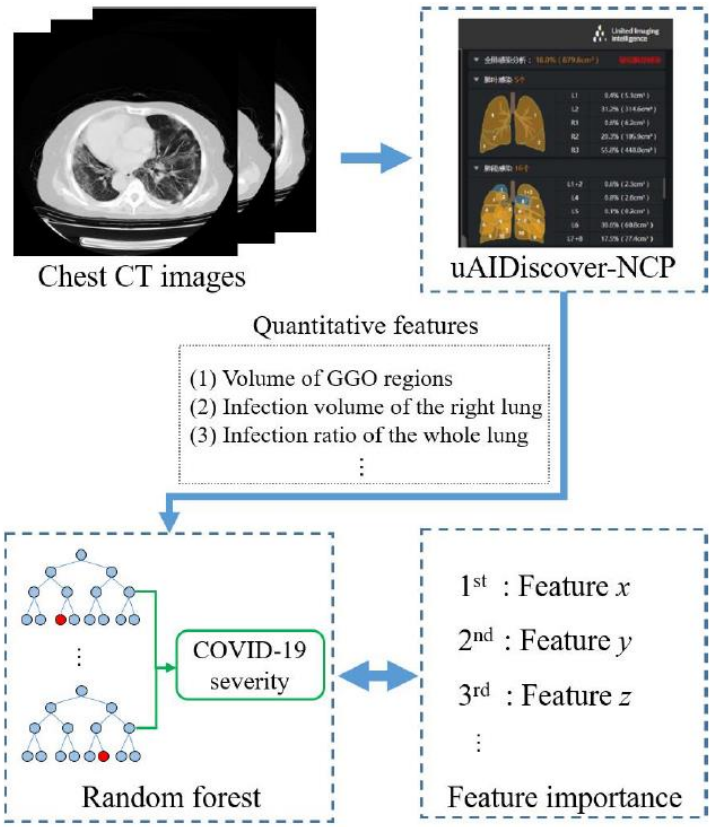
Μια ακόμα εφαρμογή για την πρόβλεψη του κινδύνου θανάτου ασθενών με COVID-19 βασισμένη στη μηχανική μάθηση περιγράφεται στο [88], η οποία επιτυγχάνει υψηλότερη ευαισθησία σε σχέση με την εφαρμογή EDI. Στη δημοσίευση αυτή, οι ερευνητές παρουσιάζουν μια αναδρομική μελέτη αξιολόγησης των εργαστηριακών δεδομένων και της αντίστοιχης θνητότητας ατόμων με θετικά αποτελέσματα σε PCR τεστ. Στόχος της μελέτης ήταν ο εντοπισμός προγνωστικών βιοδεικτών στο αίμα ασθενών που φαίνεται να συνδέονται με μεγαλύτερο κίνδυνο θανάτου. Για το σκοπό αυτό, ανέπτυξαν ένα μοντέλο μηχανικής μάθησης χρησιμοποιώντας πέντε εργαστηριακές παραμέτρους και συγκεκριμένα το CRP (C-Reactive Protein), το άζωτο ουρίας αίματος, το ασβέστιο ορού, την αλβουμίνη ορού και το γαλακτικό οξύ. Το δείγμα τους περιελάμβανε 398 ασθενείς (43 εκ των οποίων είχαν καταλήξει) και το μοντέλο τους κλήθηκε να προβλέψει την έκβαση του ασθενούς. Το προκύπτον μοντέλο επιτηρούμενης μάθησης (τύπου Support Vector Machine) πέτυχε 91% ευαισθησία, 91% ειδικότητα και εμβαδόν κάτω από την καμπύλη (AUC) 0,93 στην αντίστοιχη πρόβλεψη. Επίσης, οι ερευνητές εξέτασαν τη σημαντικότητα του κάθε χαρακτηριστικού, καθώς και

συνδυασμού χαρακτηριστικών, στην τελική έκβαση του ασθενούς, αναδεικνύοντας κάποια χαρακτηριστικά μοτίβα εργαστηριακών τιμών που φαίνεται να την επηρεάζουν περισσότερο.

Στο [89] παρουσιάζεται μια ακόμα μελέτη για τον προσδιορισμό των κλινικών χαρακτηριστικών ατόμων με COVID-19 που φαίνεται να επηρεάζουν την εξέλιξη της ασθένειάς τους, μέσω της ανάπτυξης ενός εργαλείου με δυνατότητες τεχνητής νοημοσύνης. Με βάση τα χαρακτηριστικά αυτά, το εργαλείο στοχεύει τελικά στο να προβλέψει ποιοι ασθενείς διατρέχουν μεγαλύτερο κίνδυνο σοβαρής νόσησης κατά την αρχική εμφάνισή τους στο νοσοκομείο. Τα προγνωστικά μοντέλα που αναπτύχθηκαν εκπαιδεύτηκαν μέσω ιστορικών δεδομένων στην πρόβλεψη της ανάπτυξης Συνδρόμου Οξείας Αναπνευστικής Δυσχέρειας (Acute Respiratory Distress Syndrome - ARDS), που είναι ένα από τα πιο σοβαρά συμπτώματα της ασθένειας COVID-19. Τα αποτελέσματά τους, με βάση δεδομένα ασθενών από δύο νοσοκομεία στην Κίνα, προσδιόρισαν τα εξής χαρακτηριστικά (με σειρά φθίνουσας σημαντικότητας) που φαίνεται να συνδέονται ισχυρά με τη μεταγενέστερη ανάπτυξη ARDS: (α) ήπια αυξημένη τιμή του ηπατικού ενζύμου ALT (alanine transaminase - αμινοτρανσφεράση της αλανίνης), (β) παρουσία μυαλγίας και (γ) αυξημένη αιμοσφαιρίνη. Η ακρίβεια που επιτεύχθηκε από το εν λόγω μοντέλο στην πρόβλεψη σοβαρών περιπτώσεων κυμάνθηκε από 70% έως 80%.

Εκτός από τις εφαρμογές που περιγράψαμε πιο πάνω, και οι οποίες χρησιμοποιούν κυρίως στοιχεία από τις εργαστηριακές εξετάσεις αίματος και χαρακτηριστικά της κλινικής εικόνας των ασθενών, άλλες ενδιαφέρουσες μελέτες εκμεταλλεύονται διάφορες μορφές ιατρικών εικόνων για σκοπούς πρόγνωσης. Για παράδειγμα στο [90], οι ερευνητές χρησιμοποίησαν τεχνικές μηχανικής μάθησης για να αξιολογήσουν τη σοβαρότητα ενός ασθενούς με COVID-19, με βάση εικόνες αξονικής τομογραφίας θώρακος. Ταυτόχρονα εξήγαγαν τα χαρακτηριστικά εκείνα των εικόνων που φαίνεται να επηρεάζουν περισσότερο την αντίστοιχη πρόβλεψη του μοντέλου τους.

Ως είσοδο στο μοντέλο χρησιμοποιήθηκαν αξονικές θώρακος 176 ασθενών με COVID-19, από τις οποίες υπολογίστηκαν 63 διαφορετικά ποσοτικά χαρακτηριστικά, όπως για παράδειγμα το ποσοστό μόλυνσης ανά μονάδα όγκου στους πνεύμονες και η εικόνα θαμβής υάλου (ground glass opacity - GGO). Αυτά τα ποσοτικά χαρακτηριστικά υπολογίστηκαν μέσω ενός εμπορικού εργαλείου ανάλυσης ιατρικών εικόνων που βασίζεται σε βαθιά μάθηση, και συγκεκριμένα του uAI-Discover-NCP της εταιρείας Shanghai United Imaging Intelligence. Στη συνέχεια, ένα μοντέλο μηχανικής μάθησης τύπου Random Forest (RF) εκπαιδεύτηκε για την ταξινόμηση της σοβαρότητας (σε μη σοβαρή ή σοβαρή ασθένεια) με βάση τα προεπιλεχθέντα ποσοτικά χαρακτηριστικά. Τέλος, μέσω του ίδιου μοντέλου, υπολογίστηκε η συσχέτιση του κάθε ποσοτικού χαρακτηριστικού με τη σοβαρότητα της ασθένειας COVID-19. Τα αποτελέσματα των υπολογισμών έδειξαν ότι ο όγκος των περιοχών GGO και η αναλογία τους σε σχέση με ολόκληρο τον όγκο των πνευμόνων σχετίζονται σε μεγάλο βαθμό με τη σοβαρότητα της νόσου. Μάλιστα, οι μετρήσεις που αφορούν στο δεξιό πνεύμονα φάνηκε να επηρεάζουν περισσότερο την αντίστοιχη πρόβλεψη σε σύγκριση με αυτές του αριστερού πνεύμονα. Το διάγραμμα του συνολικού συστήματος παρουσιάζεται στο Σχήμα 27.



Σχήμα 27 *- Μια αρχιτεκτονική για την πρόβλεψη της σοβαρότητας της COVID-19 βάσει αξονικής τομογραφίας θώρακος* [90]

Μια υβριδική προσέγγιση για την πρόβλεψη της σοβαρότητας της ασθένειας COVID-19 προτάθηκε από μια ομάδα ερευνητών, η οποία συνδυάζει στην είσοδο ενός μοντέλου εικόνες αξονικής τομογραφίας θώρακος, την ηλικία του ασθενούς, καθώς και τιμές εργαστηριακών εξετάσεων (γαλακτική αφυδρογονάση (Lactate dehydrogenase - LDH), CRP και αριθμό CD4 κυττάρων) που λήφθηκαν κατά την είσοδο των ασθενών στο νοσοκομείο [91]. Η μελέτη περιελάμβανε δείγματα από 196 νοσηλευόμενους ασθενείς με COVID-19, οι οποίοι ταξινομήθηκαν σε δύο ομάδες με βάση τη σοβαρότητα της νόσου που τελικά εκδήλωσαν (σοβαρές και μη σοβαρές περιπτώσεις). Για την πρόβλεψη της σοβαρότητας της ασθένειας COVID-19, χρησιμοποιήθηκε μια μέθοδος λογιστικής παλινδρόμησης και συγκεκριμένα ένα μοντέλο LASSO (Least Absolute Shrinkage and Selection Operator). Τα κλινικά χαρακτηριστικά που προσδιορίστηκαν σε αυτή τη μελέτη ως προγνώστες (predictors) για σοβαρής μορφής ασθένεια ήταν παρόμοια με αυτά των μελετών που προαναφέραμε. Το γεγονός ότι διαφορετικές μελέτες συγκλίνουν στα ίδια αποτελέσματα είναι ενθαρρυντικό για την περαιτέρω έρευνα και εφαρμογή των αντίστοιχων λύσεων στην κλινική πρακτική, ενώ ταυτόχρονα τονίζεται η σημασία της παρακολούθησης συγκεκριμένων κλινικών δεικτών.

3.4 Βέλτιστη Διαχείριση Ασθενών με Πεπερασμένους Πόρους

3.4.1 Οι Προκλήσεις της Πανδημίας COVID-19 στη Διαχείριση των Ασθενών

Τα συστήματα υγείας σε όλο τον κόσμο κλήθηκαν να αντιμετωπίσουν μια άνευ προηγουμένου κατάσταση, ειδικά στα αρχικά στάδια της πανδημίας, όπου η αύξηση των κρουσμάτων ήταν εκθετική, ενώ ταυτόχρονα οι πόροι για την αντιμετώπιση των ασθενών ήταν περιορισμένοι. Πολλές χώρες αντιμετώπισαν σημαντική έλλειψη πόρων, όπως για παράδειγμα kit δοκιμών, κρεβάτια μονάδων

εντατικής θεραπείας και αναπνευστήρες. Επίσης, όπως έχει ήδη αναφερθεί, η σοβαρότητα των περιστατικών COVID-19 ποικίλλει σημαντικά και οι επαγγελματίες υγείας καλούνται να βρουν τον καλύτερο τρόπο για την αντιμετώπισή τους.

Υπάρχει επομένως σαφής ανάγκη να αναπτυχθούν τεχνολογίες και εφαρμογές που μπορούν να βοηθήσουν το προσωπικό υγείας στην κατανομή των διαθέσιμων πόρων, να παρέχουν συμβουλές σχετικά με τις κατάλληλες θεραπευτικές προσεγγίσεις με βάση τα χαρακτηριστικά και την αναμενόμενη πορεία του κάθε ασθενούς, αλλά και να μειώσουν την πίεση που δέχονται τα νοσοκομεία, με το να είναι σε θέση να προβλέψουν πότε οι ασθενείς μπορούν με ασφάλεια να πάρουν εξιτήριο από το νοσοκομείο και να συνεχίσουν να παρακολουθούνται από το σπίτι. Στις ακόλουθες ενότητες παρουσιάζουμε ορισμένα παραδείγματα εφαρμογών σε αυτούς τους τομείς που ενσωματώνουν τεχνολογίες τεχνητής νοημοσύνης.

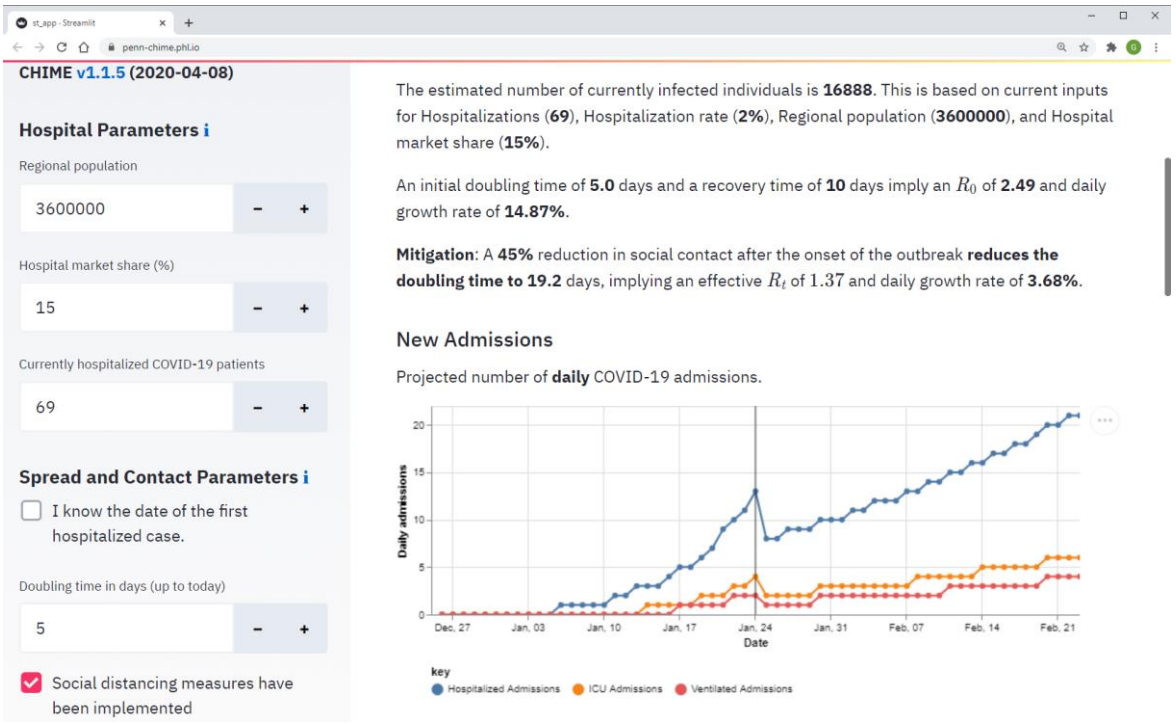
3.4.2 Ανάθεση Πόρων

Η πανδημία COVID-19 έχει δημιουργήσει σοβαρές προκλήσεις στις αλυσίδες εφοδιασμού υγειονομικού υλικού και εξοπλισμού. Στην κατηγορία αυτή περιλαμβάνεται εξοπλισμός ατομικής προστασίας, όπως μάσκες και ιατρικές ποδιές, kit εργαστηριακών εξετάσεων, καθώς και εξοπλισμός μονάδων εντατικής θεραπείας, όπως για παράδειγμα κρεβάτια και αναπνευστήρες. Η μηχανική μάθηση χρησιμοποιείται ήδη ευρέως στον τομέα των logistics και παρόμοιες τεχνικές μπορούν να εφαρμοστούν και σε ό,τι αφορά τη διαχείριση της εφοδιαστικής αλυσίδας των παραπάνω υγειονομικών πόρων.

Μετά το ξέσπασμα μιας πανδημίας όπως η COVID-19, πολιτικές κοινωνικής αποστασιοποίησης και άλλα μέτρα προστασίας σίγουρα βοηθούν στη μείωση της πίεσης στα συστήματα υγείας μέσω της αντίστοιχης μείωσης στην εξάπλωση της ασθένειας. Αυτό όμως δε σημαίνει ότι, ακόμα και έτσι, μεμονωμένες περιοχές ή μεμονωμένα νοσοκομεία δε θα πιεστούν πέραν των δυνατοτήτων τους. Επιπλέον, τα μέτρα αυτά αναγκαστικά χαλαρώνουν με την πάροδο του χρόνου σε μια προσπάθεια να μειωθούν οι οικονομικές και κοινωνικές επιπτώσεις τους. Είναι επομένως πολύ σημαντικό να διασφαλιστεί ότι τα νοσοκομεία διαθέτουν κατάλληλα συστήματα ανάλυσης δεδομένων που θα τα βοηθήσουν να διαχειριστούν τις περιόδους υψηλής ζήτησης για πόρους, όπως κρεβάτια ΜΕΘ ή αναπνευστήρες. Η έλλειψη πόρων ιατροφαρμακευτικής περίθαλψης επιδεινώνεται από την ανάγκη χρήσης τους τόσο για την αντιμετώπιση κρουσμάτων COVID-19 όσο και για την περίθαλψη και θεραπεία άλλων ασθενών.

Ερευνητές από το Πανεπιστήμιο της Πενσυλβάνια των ΗΠΑ δημιούργησαν μια εφαρμογή που ονομάζεται CHIME (COVID-19 Hospital Impact Model for Epidemics) στην προσπάθειά τους να βοηθήσουν τα νοσοκομεία και τους εργαζόμενους των υπηρεσιών υγείας να υπολογίσουν τις διάφορες ανάγκες τους, καθώς εξελίσσεται η πανδημία COVID-19 [92]. Η εφαρμογή CHIME επιτρέπει τον έγκαιρο σχεδιασμό της χωρητικότητας ενός νοσοκομείου (capacity planning), παρέχοντας εκτιμήσεις για τον αριθμό των νέων ημερήσιων νοσηλειών, το σύνολο των νοσηλευόμενων ασθενών, τον αριθμό εισαγωγών σε ΜΕΘ, καθώς και τον αριθμό των ασθενών που χρειάζονται αναπνευστήρα. Οι εκτιμήσεις αυτές δημιουργούνται χρησιμοποιώντας ένα μοντέλο SIR (Susceptible-Infected-Recovered), που αποτελεί μια κοινή τεχνική επιδημιολογικής μοντελοποίησης, όπως είδαμε και στο Κεφάλαιο 2. Μπορούν να χρησιμοποιηθούν για τη δημιουργία (υποθετικών) σεναρίων καλύτερης και χειρότερης περίπτωσης, βοηθώντας έτσι στον καλύτερο σχεδιασμό της χωρητικότητας του νοσοκομείου. Το μοντέλο μπορεί να προσαρμοστεί κατάλληλα, επιτρέποντας την εισαγωγή παραμέτρων εισόδου που αφορούν στο μέγεθος των νοσοκομείων, καθώς και στα χαρακτηριστικά εξάπλωσης του ιού. Όπως αναφέρουν οι ερευνητές, το μοντέλο τους αξιολογήθηκε θετικά από αρκετούς επιδημιολόγους. Σύμφωνα με την ιστοσελίδα τους, το CHIME διατίθεται ως λογισμικό

ανοιχτού κώδικα και χρησιμοποιείται από πολλούς οργανισμούς στις ΗΠΑ και σε άλλες χώρες του κόσμου. Ένα παράδειγμα της διεπαφής χρήστη της εφαρμογής CHIME φαίνεται στο Σχήμα 28.



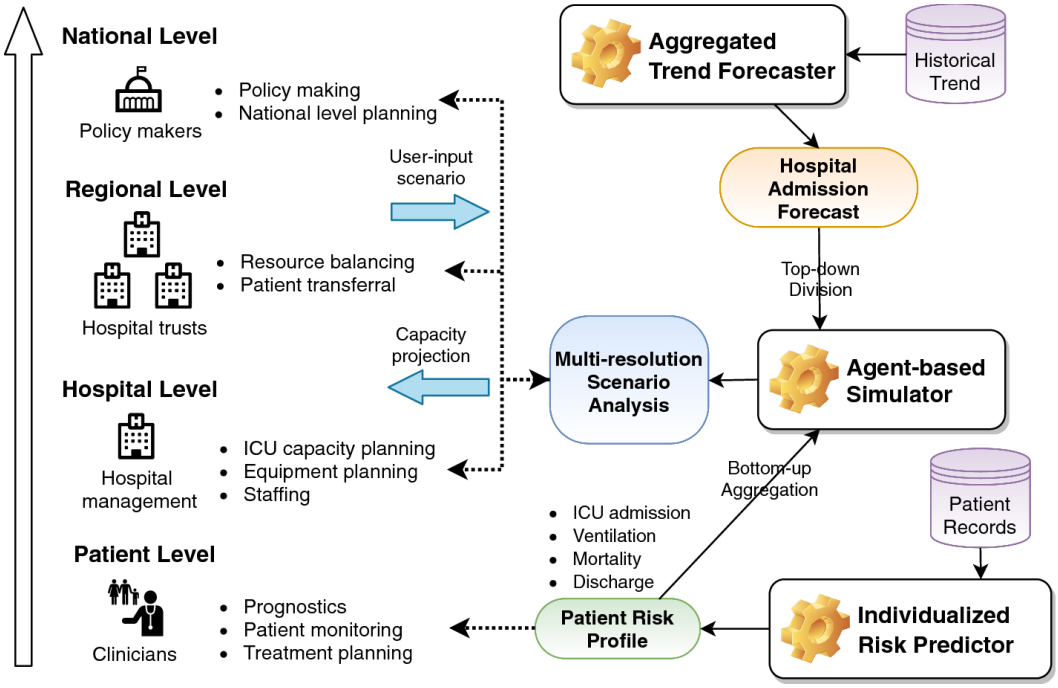
Σχήμα 28 *- Παράδειγμα εισόδου και εξόδου της εφαρμογής CHIME για το σχεδιασμό της χωρητικότητας νοσοκομείων* [92]

Σε μια παρόμοια προσπάθεια, ερευνητές του Πανεπιστημίου του Cambridge δημιούργησαν μια εφαρμογή με την ονομασία COVID-19 CPAS (Capacity Planning and Analysis System) [93], που ήδη λειτουργεί με επιτυχία σε πολλά νοσοκομεία στο Ηνωμένο Βασίλειο, η οποία κάνει χρήση μοντέλων μηχανικής μάθησης για την εξαγωγή προβλέψεων σε διάφορα επίπεδα. Η εφαρμογή αυτή χρησιμοποιεί ανώνυμα δεδομένα ασθενών από τον Οργανισμό Δημόσιας Υγείας της χώρας, όπως αποτελέσματα εργαστηριακών εξετάσεων, στοιχεία που αφορούν στη νοσηλεία των ασθενών, παράγοντες κινδύνου, καθώς και στοιχεία σε σχέση με την έκβαση των ασθενών [94] .

Στόχος των δημιουργών της εφαρμογής ήταν να προβλέψουν τη μελλοντική ανάγκη για διάφορους πόρους, όπως κρεβάτια ΜΕΘ και αναπνευστήρες, τόσο σε επίπεδο μεμονωμένου ασθενούς (με βάση τον υπολογισμό μιας βαθμολογίας κινδύνου) όσο και σε επίπεδο νοσοκομείου (με βάση τις συνολικές ανάγκες όπως προκύπτουν για όλους τους ασθενείς). Με αυτό τον τρόπο, κατάφεραν να βοηθήσουν τους επαγγελματίες υγείας στη λήψη τεκμηριωμένων αποφάσεων σχετικά με τη βέλτιστη κατανομή των περιορισμένων αυτών πόρων για την επίτευξη του καλύτερου αποτελέσματος προς όφελος των ασθενών. Επίσης, μέσω της έγκαιρης και ακριβούς πρόβλεψης για μελλοντικές ανάγκες, τα νοσοκομεία μπορούν να προβούν στις απαραίτητες ενέργειες για την προμήθεια του κατάλληλου υγειονομικού εξοπλισμού.

Η εφαρμογή CPAS σχεδιάστηκε για να εξυπηρετήσει τις ανάγκες διαφορετικών χρηστών που συμμετέχουν στο σχεδιασμό χωρητικότητας των νοσοκομείων σε διάφορα διοικητικά και γεωγραφικά επίπεδα. Πιο συγκεκριμένα, όπως φαίνεται στο Σχήμα 29, η εφαρμογή υπολογίζει τις ανάγκες για ΜΕΘ σε επίπεδο ασθενούς, σε επίπεδο νοσοκομείου, σε περιφερειακό και σε εθνικό επίπεδο. Στο υψηλότερο επίπεδο (εθνικό), η εφαρμογή βοηθά τους υπευθύνους χάραξης κεντρικής πολιτικής μιας χώρας να λαμβάνουν τεκμηριωμένες αποφάσεις προβλέποντας τις συνολικές τάσεις

για τις ανάγκες σε ΜΕΘ σε διάφορα υποθετικά σενάρια. Στο δεύτερο επίπεδο (περιφέρειας), η περιφερειακή ηγεσία των υπηρεσιών υγείας μπορεί να χρησιμοποιήσει το CPAS ως εργαλείο «εξισορρόπησης φορτίου», καθώς ορισμένα νοσοκομεία ενδέχεται να αντιμετωπίσουν υψηλότερη πίεση από άλλα. Μπορούν για παράδειγμα να λάβουν αποφάσεις σχετικά με τη μεταφορά ασθενών και πόρων ανάμεσα σε νοσοκομεία. Στο τρίτο επίπεδο (νοσοκομείου), οι διευθυντές των νοσοκομείων μπορούν να χρησιμοποιήσουν το CPAS για να προγραμματίσουν και να προετοιμάσουν έγκαιρα τους χώρους των νοσοκομείων, συμπεριλαμβανομένου του εξοπλισμό και του προσωπικού των ΜΕΘ. Τέλος, στο τέταρτο επίπεδο (ασθενούς), οι κλινικοί ιατροί πρώτης γραμμής μπορούν να χρησιμοποιήσουν το εργαλείο για να κατανοήσουν το προφίλ κινδύνου συγκεκριμένων ασθενών και να προτείνουν την κατάλληλη αγωγή.



Σχήμα 29 *- Η εφαρμογή CPAS και τα διαφορετικά επίπεδα πρόβλεψής της* [93]

Τα αποτελέσματα πρόβλεψης του μοντέλου που χρησιμοποιεί η εφαρμογή CPAS είναι αρκετά ενθαρρυντικά. Μετά την εκπαίδευσή του με στοιχεία ασθενών, το μοντέλο ήταν σε θέση να κάνει ιδιαίτερα ακριβείς προβλέψεις σε ένα σύνολο ασθενών που χρησιμοποιήθηκε για σκοπούς επαλήθευσης. Καθώς διοχετεύονται όλο και περισσότερα και καλύτερης ποιότητας δεδομένα στο μοντέλο, οι δημιουργοί του αναμένουν ότι η ακρίβειά του θα βελτιωθεί και, κατ' επέκταση, η χρησιμότητά του θα αυξηθεί. Σε σχέση με τα επόμενα βήματά τους, οι ερευνητές αναφέρουν ότι θα συνεχίσουν να επικυρώνουν τα μοντέλα που αναπτύσσουν και να βελτιώνουν την ετοιμότητά τους για χρήση σε πραγματικές συνθήκες από τους επαγγελματίες υγείας. Σχεδιάζουν επίσης να συμπεριλάβουν δεδομένα εισόδου ως συναρτήσεις του χρόνου, αντί απλών «στιγμιοτύπων» των μεταβλητών σε συγκεκριμένες χρονικές στιγμές όπως είναι σήμερα, κάτι που θα τους επιτρέψει να αποκτήσουν μια βαθύτερη εικόνα της εξέλιξης των ασθενών ενώ νοσηλεύονται. Με αυτό τον τρόπο, η πρόβλεψη κινδύνου που θα παράγει το μοντέλο θα λαμβάνει υπόψη την εξέλιξη της νόσου του κάθε ασθενούς.

3.4.3 Ανάπτυξη Εξατομικευμένων Πλάνων Διαχείρισης και Θεραπείας Ασθενών

Όπως είναι φανερό, λόγω της διαφορετικής πιθανής εξέλιξης των ασθενών με COVID-19, ενιαία σχέδια θεραπείας, που καθορίζονται σε επίπεδο πληθυσμού και όχι σε εξατομικευμένο επίπεδο ασθενούς, είναι πολύ πιθανό να είναι αναποτελεσματικά τόσο σε ό,τι αφορά την επίδρασή τους στους ασθενείς όσο και σε σχέση με την ορθή χρήση των περιορισμένων πόρων. Επίσης, μέχρι σήμερα, δεν έχει βρεθεί μια καθολικά εφαρμόσιμη φαρμακευτική αγωγή ή άλλη μέθοδος θεραπείας που να ισχύει για όλους τους ασθενείς σε οποιοδήποτε στάδιο της νόσου. Με βάση τα πιο πάνω, είναι απαραίτητο να αναπτυχθούν εξατομικευμένα πλάνα διαχείρισης και θεραπείας για κάθε ασθενή, ανάλογα με τα μοναδικά χαρακτηριστικά και τις ανάγκες του. Αυτό θα βοηθήσει τόσο στην καλύτερη αντιμετώπιση του κάθε περιστατικού αλλά και στη βέλτιστη χρήση των διαθέσιμων πόρων.

Μοντέλα που είναι βασισμένα στην τεχνητή νοημοσύνη προσπαθούν να απαντήσουν σε ερωτήσεις τύπου «What if» όπως για παράδειγμα: «τι θα συμβεί σε ένα συγκεκριμένο ασθενή αν καθυστερήσει να μπει σε μηχανική υποστήριξη αναπνοής για λίγες ώρες;» ή «ποια θα ήταν η πορεία ενός συγκεκριμένου ασθενούς αν αντί για υποστηρικτική αγωγή του χορηγηθεί μια συγκεκριμένη πειραματική θεραπεία μια εβδομάδα νωρίτερα;». Ακριβώς επειδή ο κάθε ασθενής είναι διαφορετικός, οι απαντήσεις σε υποθετικές ερωτήσεις, όπως οι πιο πάνω, απαιτούν αποτελεσματική ανάλυση τακτικά ενημερωμένων δεδομένων παρατήρησης από άλλους ασθενείς που έχουν υποβληθεί σε διάφορες θεραπείες από την αρχή της πανδημίας. Τα υποθετικά ερωτήματα που χρειάζεται να απαντηθούν δεν αφορούν μόνο ποια παρέμβαση θα πρέπει να εφαρμόζεται σε κάθε ασθενή, αλλά και πότε ο ασθενής χρειάζεται αυτή την παρέμβαση, καθώς διαφορετικά φάρμακα και μέθοδοι αντιμετώπισης μπορεί να είναι αποτελεσματικά σε συγκεκριμένα στάδια της νόσου. Ο χρόνος των παρεμβάσεων συνδέεται επίσης και με τη διαθεσιμότητα των αντίστοιχων πόρων που απαιτούνται, η οποία θα πρέπει να λαμβάνεται υπόψη από τα συγκεκριμένα μοντέλα.

Λαμβάνοντας υπόψη τα πιο πάνω, οι ερευνητές στο [95] πρότειναν τρόπους με τους οποίους η χρήση τεχνητής νοημοσύνης θα μπορούσε να βοηθήσει στην ανάπτυξη εξατομικευμένων πλάνων αντιμετώπισης και θεραπείας ασθενών με COVID-19. Στην ουσία, εισηγήθηκαν τη δημιουργία ενός μοντέλου που βασίζεται στη μηχανική μάθηση, το οποίο εκπαιδεύεται από δεδομένα παρατήρησης της πορείας των ασθενών σε συνδυασμό με τις θεραπείες που τους είχαν χορηγηθεί. Το μοντέλο αυτό μπορεί να «μάθει» τα αποτελέσματα των υφιστάμενων θεραπειών, δεδομένων των μοναδικών χαρακτηριστικών του κάθε ατόμου, που συμπεριλαμβάνουν και πιθανά υποκείμενα νοσήματα ή άλλους παράγοντες ευπάθειας. Για παράδειγμα, η προσέγγιση αυτή μπορεί να βοηθήσει στον καθορισμό βέλτιστων πρακτικών για την εφαρμογή μηχανικής υποστήριξης αναπνοής, καθώς και πειραματικών θεραπειών.

Υπάρχουν διάφορες προσεγγίσεις στην πρόσφατη επιστημονική βιβλιογραφία που έχουν ήδη αναπτυχθεί για άλλες ασθένειες ή ως γενικά πλαίσια πρόβλεψης εξατομικευμένων θεραπειών και θα μπορούσαν να προσαρμοστούν για χρήση σε ασθενείς με COVID-19. Αυτές οι προσεγγίσεις χρησιμοποιούν σύγχρονες εξελίξεις στον τομέα της μηχανικής μάθησης που βασίζονται σε εκπαίδευση μέσω δεδομένων παρατήρησης (observational data).

Για παράδειγμα στο [96], οι ερευνητές χρησιμοποίησαν μοντέλα που βασίζονται σε Gaussian διαδικασίες για να καταλάβουν τις επιδράσεις διαφόρων θεραπειών. Τα πειράματά τους έγιναν σε δύο ξεχωριστά σύνολα δεδομένων παρατήρησης. Το πρώτο σύνολο δεδομένων αφορούσε στην επίδραση ενός κοινωνικού προγράμματος που εφαρμόστηκε σε πρόωρα βρέφη για να ενισχύσει την κατάσταση της υγείας τους. Το δεύτερο σύνολο δεδομένων αφορούσε στην επίδραση της εφαρμογής μιας συγκεκριμένης συσκευής υποβοήθησης για καρδιακούς ασθενείς (Left Ventricular Assistance Device - LVAD), οι οποίοι βρίσκονταν σε λίστα αναμονής για μεταμόσχευση καρδιάς. Και στα δύο

πειράματα, οι ερευνητές έδειξαν ότι η προτεινόμενη μέθοδος πρόβλεψής τους υπερτερούσε σημαντικά σε σχέση με τα υφιστάμενα στατιστικά μοντέλα.

Σε μια άλλη παρόμοια μελέτη [97], οι ερευνητές εφάρμοσαν μια νέα μέθοδο για την εξαγωγή συμπερασμάτων σε σχέση με τις εξατομικευμένες επιδράσεις θεραπειών, βασιζόμενη σε νευρωνικά δίκτυα τύπου GAN (Generative Adversarial Networks). Το προτεινόμενο σύστημα επιχειρεί όχι μόνο να προσαρμόσει ένα μοντέλο στα πραγματικά δεδομένα παρατήρησης, αλλά και να προβλέψει συμβάντα που δεν έχουν παρατηρηθεί. Η συγκεκριμένη μέθοδος δοκιμάστηκε χρησιμοποιώντας τρία πραγματικά σύνολα δεδομένων και παρήγαγε πολύ καλά αποτελέσματα σε σύγκριση με άλλα μοντέλα πρόβλεψης. Οι ερευνητές στο [98], επιχείρησαν να αντιμετωπίσουν το ίδιο πρόβλημα πρόβλεψης χρησιμοποιώντας ένα άλλο μοντέλο που ονομάζεται DKLITE (Deep Kernel Learning for Individualized Treatment Effects), το οποίο βασίζεται επίσης σε βαθιά νευρωνικά δίκτυα.

ΚΕΦΑΛΑΙΟ 4: Εφαρμογές στη Φαρμακολογία

4.1 Εισαγωγή

Από την αρχή της πανδημίας COVID-19, η επιστημονική κοινότητα έχει επιδοθεί σε έναν αγώνα δρόμου για την εύρεση αποτελεσματικών θεραπειών και εμβολίων για την καταπολέμηση του νέου ιού. Ο τεράστιος όγκος των διαθέσιμων δεδομένων, καθώς και η πρόοδος των αλγορίθμων μηχανικής μάθησης οδήγησαν σε ευρεία υιοθέτηση της τεχνητής νοημοσύνης σε διάφορους τομείς της βιοϊατρικής και της φαρμακευτικής έρευνας.

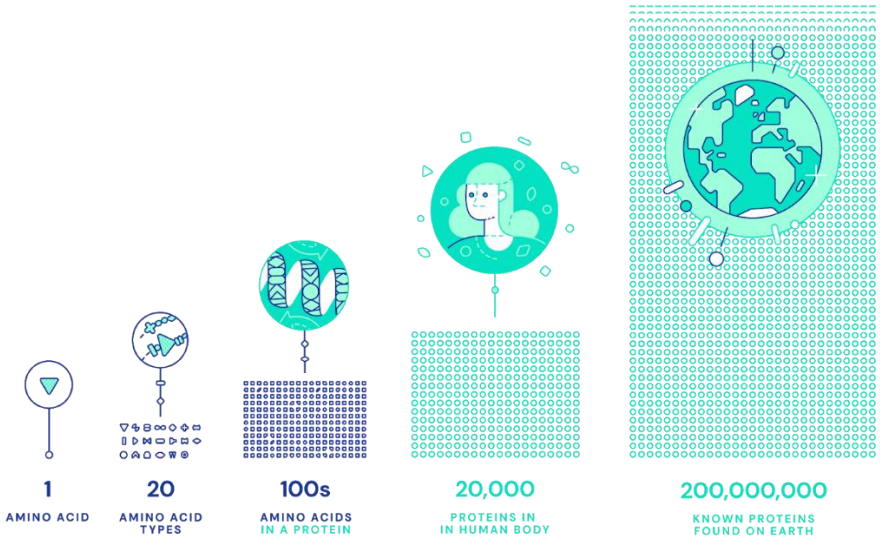
Σε αυτό το Κεφάλαιο, παρουσιάζουμε τις εφαρμογές τεχνητής νοημοσύνης στον τομέα της Φαρμακολογίας. Αρχικά, εξετάζουμε τρόπους με τους οποίους η τεχνητή νοημοσύνη μπορεί να χρησιμοποιηθεί για την πρόβλεψη της τρισδιάστατης δομής των πρωτεϊνών από την ακολουθία των αμινοξέων που τις αποτελούν, κάτι που ονομάζεται «πρόβλημα αναδίπλωσης πρωτεϊνών». Ιδιαίτερο ενδιαφέρον παρουσιάζει η πρόβλεψη της δομής σημαντικών πρωτεϊνών που συνθέτουν το νέο κορωνοϊό SARS-CoV-2, καθώς η ακριβής γνώση του τρισδιάστατου σχήματός τους είναι εξαιρετικά σημαντική για την καλύτερη κατανόησή τους και την επακόλουθη ανάπτυξη φαρμάκων και εμβολίων που μπορούν να δράσουν εναντίον τους. Στη συνέχεια, περιγράφουμε αρκετές εφαρμογές τεχνητής νοημοσύνης στην ανακάλυψη φαρμάκων και εμβολίων, είτε μέσω πρόβλεψης για το αν ένα υπάρχον φάρμακο μπορεί να είναι αποτελεσματικό κατά του νέου ιού («επαναστόχευση φαρμάκων») είτε μέσω της ανακάλυψης εντελώς νέων φαρμάκων και πιθανών στόχων εμβολίων. Ολοκληρώνουμε αυτό το Κεφάλαιο, εξετάζοντας τους τρόπους με τους οποίους η τεχνητή νοημοσύνη μπορεί να βοηθήσει στη βελτίωση και την επιτάχυνση των κλινικών δοκιμών.

4.2 Πρόβλεψη Δομής Πρωτεϊνών

4.2.1 Το Πρόβλημα Αναδίπλωσης Πρωτεϊνών

Οι πρωτεΐνες είναι μεγάλα και σύνθετα μόρια, που αποτελούνται από μία ή περισσότερες μεγάλες αλυσίδες αμινοξέων. Μέσα σε κάθε κύτταρο, χιλιάδες διαφορετικές πρωτεΐνες απαρτίζουν τους μηχανισμούς που διατηρούν ζωντανούς και λειτουργικούς όλους τους οργανισμούς, από ανθρώπους, ζώα και φυτά μέχρι μικροσκοπικά βακτήρια. Οι πρωτεΐνες είναι, δηλαδή, τα δομικά στοιχεία της ζωής. Στηρίζουν, όχι μόνο τις βιολογικές διεργασίες στο σώμα μας, αλλά και κάθε βιολογική διεργασία σε κάθε ζωντανό ον. Οι πρωτεΐνες εκτελούν ένα ευρύ φάσμα λειτουργιών μέσα στους οργανισμούς, συμπεριλαμβανομένης της δράσης τους ως καταλύτες σε διάφορες μεταβολικές αντιδράσεις, της αντιγραφής του DNA, της ανταπόκρισης σε ερεθίσματα, της παροχής δομής στα κύτταρα και τους οργανισμούς γενικότερα και της μεταφοράς μορίων από μία θέση σε κάποια άλλη. Επιπλέον, σχεδόν όλες οι ασθένειες, όπως ο καρκίνος, η άνοια, ακόμα και μολυσματικές ασθένειες όπως η ασθένεια COVID-19, σχετίζονται με τον τρόπο λειτουργίας των πρωτεϊνών.

Οι πρωτεΐνες διαφέρουν μεταξύ τους κυρίως στην ακολουθία των αμινοξέων τους, η οποία υπαγορεύεται από την ακολουθία των νουκλεοτιδίων στα γονίδια ενός οργανισμού, δηλαδή στο DNA του. Υπάρχουν 20 διαφορετικοί τύποι αμινοξέων που αποτελούν τα συστατικά στοιχεία των πρωτεϊνών. Επί του παρόντος, περίπου 200 εκατομμύρια πρωτεΐνες είναι γνωστές στον άνθρωπο ανάμεσα σε δισεκατομμύρια που πιθανολογούμε ότι υπάρχουν, με περίπου 30 εκατομμύρια νέες πρωτεΐνες να ανακαλύπτονται κάθε χρόνο. Μόνο στον ανθρώπινο οργανισμό υπάρχουν δεκάδες χιλιάδες πρωτεΐνες που επιτελούν διαφορετικές λειτουργίες [99].



Σχήμα 30 *- Αμινοξέα και πρωτεΐνες* [100]

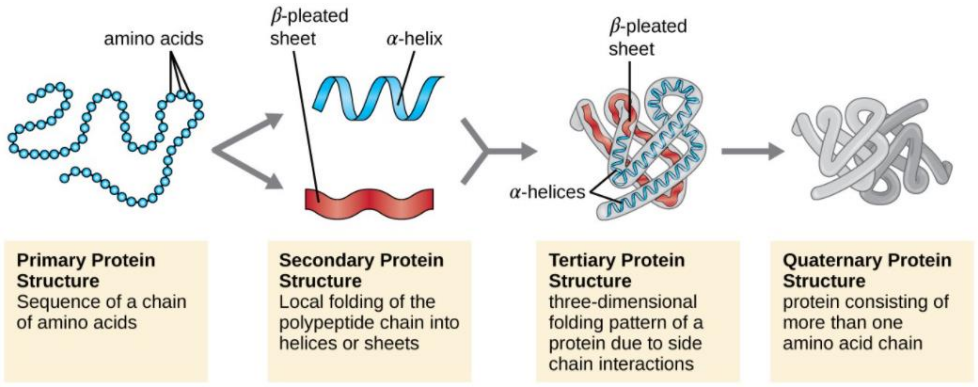
Κάθε πρωτεΐνη έχει μία μοναδική τρισδιάστατη δομή / σχήμα που καθορίζει τον τρόπο λειτουργίας της και το ρόλο της. Για παράδειγμα, οι πρωτεΐνες αντισωμάτων που παράγονται από το ανοσοποιητικό μας σύστημα είναι σχήματος Y, έτσι ώστε να μπορούν να σχηματίζουν «γάντζους» που ανιχνεύουν και εξουδετερώνουν βακτήρια και ιούς. Ακόμα και μικροσκοπικές αναδιατάξεις στα μόρια των πρωτεϊνών μπορεί να προκαλέσουν καταστροφικές επιπτώσεις στην υγεία μας. Συνεπώς, ένας από τους αποτελεσματικότερους τρόπους για να κατανοήσουμε καλύτερα τις ασθένειες αλλά και να ανακαλύψουμε τρόπους αντιμετώπισής τους είναι η μελέτη των πρωτεϊνών που εμπλέκονται σε αυτές.

Όπως προαναφέραμε, οι συνταγές για την παραγωγή των πρωτεϊνών είναι τα γονίδια που βρίσκονται στο DNA ενός οργανισμού. Το γεγονός όμως ότι μπορεί να γνωρίζουμε τη γενετική συνταγή για μία πρωτεΐνη δε σημαίνει αυτόματα ότι γνωρίζουμε και το τρισδιάστατο σχήμα της. Το DNA περιέχει μόνο πληροφορίες σχετικά με την ακολουθία των αμινοξέων, όχι όμως για το πώς αυτά τελικά αλληλεπιδρούν και δίνουν στην πρωτεΐνη το τελικό της σχήμα. Όσο μεγαλύτερη είναι μια πρωτεΐνη, τόσο δυσκολότερη είναι η εύρεση του τρισδιάστατου μοντέλου της, καθώς υπάρχουν περισσότερες αλληλεπιδράσεις μεταξύ των αμινοξέων που επηρεάζουν το τελικό της σχήμα.

Πολλές δομές πρωτεϊνών έχουν προσδιοριστεί μέσω πειραματικών προσεγγίσεων, όπως η κρυσταλλογραφία ακτίνων X (X-Ray Crystallography - XRC) και η μέθοδος πυρηνικού μαγνητικού συντονισμού (Nuclear Magnetic Resonance - NMR). Αυτές οι μέθοδοι καθώς και νεότερες μέθοδοι, όπως η χρήση εξειδικευμένων κρυο-ηλεκτρονικών μικροσκοπίων (cryo-electron microscopes), απαιτούν εξειδικευμένο εξοπλισμό, είναι δαπανηρές και χρονοβόρες. Ως αποτέλεσμα, μέχρι σήμερα γνωρίζουμε την ακριβή τρισδιάστατη δομή μόνο για ένα πολύ μικρό κλάσμα των πρωτεϊνών που είναι γνωστές στην επιστήμη [101].

Στην ομιλία του για την αποδοχή του βραβείου Νόμπελ Χημείας το 1972, ο Κρίστιαν Άνφινσεν είπε ότι, θεωρητικά η ακολουθία αμινοξέων μιας πρωτεΐνης θα πρέπει να καθορίζει πλήρως την (τρισδιάστατη) δομή της. Αυτό είναι γνωστό ως «δόγμα του Άνφινσεν» [102]. Η πρόβλεψη του τρόπου με τον οποίο οι πρωτεΐνες αναδιπλώνονται και λαμβάνουν το περίπλοκο τρισδιάστατο σχήμα τους είναι γνωστή ως «πρόβλημα αναδίπλωσης πρωτεϊνών». Στο πρόβλημα αυτό έχουν εργαστεί πάρα πολλοί επιστήμονες επί δεκαετίες, προσπαθώντας να βρουν μια ακριβή και γρήγορη μέθοδο που θα τους επέτρεπε να καθορίσουν αξιόπιστα το σχήμα οποιασδήποτε πρωτεΐνης μόνο από τη

γνώση της ακολουθίας των αμινοξέων της. Κατά τη διάρκεια των τελευταίων ετών, οι επιστήμονες έχουν στραφεί σε μεθόδους τεχνητής νοημοσύνης ως πιθανή εναλλακτική λύση σε αυτή τη μακρά και επίπονη διαδικασία εύρεσης της δομής δύσκολων πρωτεϊνών.



Σχήμα 31 *- Η διαδικασία αναδίπλωσης μιας πρωτεΐνης* [103]

Η ικανότητα να προβλεφθεί το σχήμα μιας πρωτεΐνης με καθαρά υπολογιστικό τρόπο από το γενετικό κώδικά της και μόνο – αντί μέσω δαπανηρών πειραματικών μεθόδων – θα μπορούσε να βοηθήσει στην επιτάχυνση της επιστημονικής έρευνας σε διάφορους τομείς. Παραδείγματα περιλαμβάνουν την εύρεση αλλά και το σχεδιασμό πρωτεϊνών, ικανών να συνδεθούν με συγκεκριμένους στόχους, να δράσουν ως καταλύτες σε νέες αντιδράσεις και γενικότερα να συντελέσουν στην πρόοδο σε διάφορους τομείς της βιολογίας και της βιοτεχνολογίας [104]. Επίσης, η περιορισμένη γνώση σχετικά με τις δομές των πρωτεϊνών αποτελεί σημαντικό εμπόδιο για την κατανόηση ορισμένων ασθενειών, οι οποίες επηρεάζουν τη ζωή εκατομμυρίων ανθρώπων και προκαλούν δεκάδες χιλιάδες θανάτους κάθε χρόνο. Τέλος, στον τομέα της Φαρμακολογίας ξοδεύονται πολλά χρήματα και πολλά χρόνια προσπαθειών για την ανάπτυξη ενός νέου φαρμάκου. Είναι βέβαιο ότι η γρήγορη και ακριβής πρόβλεψη των πρωτεϊνικών δομών θα μπορούσε να συμβάλει στην καλύτερη και αποδοτικότερη ανακάλυψη φαρμάκων, μέσω του προσδιορισμού της δομής πολλών πρωτεϊνών που εμπλέκονται στις ασθένειες.

Για να διευκολυνθεί η έρευνα και να αξιολογηθεί η πρόοδος σχετικά με τις σύγχρονες μεθόδους τεχνητής νοημοσύνης στη βελτίωση της ακρίβειας των προβλέψεων πρωτεϊνικών δομών, καθιερώθηκε από το 1994 ένας παγκόσμιος διαγωνισμός που ονομάζεται CASP (Critical Assessment of protein Structure Prediction), ο οποίος λαμβάνει χώρα κάθε δύο έτη. Ο διαγωνισμός αυτός έχει γίνει το «de facto standard» για την αξιολόγηση των υπολογιστικών τεχνικών πρόβλεψης και ταυτόχρονα αποτελεί μια παγκόσμια κοινότητα επιστημόνων που ενώνουν τις δυνάμεις τους σε μια κοινή προσπάθεια [105] .

Στις ερευνητικές ομάδες που συμμετέχουν στον CASP δίνονται διάφορες ακολουθίες αμινοξέων για ένα σύνολο πρωτεϊνών, των οποίων οι τρισδιάστατες δομές είναι ακόμη άγνωστες. Καθώς οι επιστήμονες μελετούν τις πρωτεΐνες στο εργαστήριο για να βρουν με πειραματικό τρόπο το σχήμα τους, οι ομάδες που συμμετέχουν στον CASP προσπαθούν να κάνουν το ίδιο με τη χρήση υπολογιστών. Τα αποτελέσματα όλων των ομάδων αξιολογούνται τελικά από ανεξάρτητους επιστήμονες. Ο βασικός δείκτης που χρησιμοποιείται από τον CASP για τη μέτρηση της ακρίβειας των προβλέψεων είναι το Global Distance Test (GDT), το οποία κυμαίνεται από 0 έως 100 και αντιπροσωπεύει ουσιαστικά το ποσοστό των αμινοξέων που βρίσκονται σε απόσταση μικρότερη ενός προκαθορισμένου κατωφλίου από τη σωστή θέση. Μια βαθμολογία GDT περίπου 90 και άνω

θεωρείται ανεπίσημα συγκρίσιμη με τα αποτελέσματα που λαμβάνονται από τις πειραματικές μεθόδους.

4.2.2 Εφαρμογές Τεχνητής Νοημοσύνης για την Πρόβλεψη της Δομής Πρωτεϊνών του SARS-CoV-2

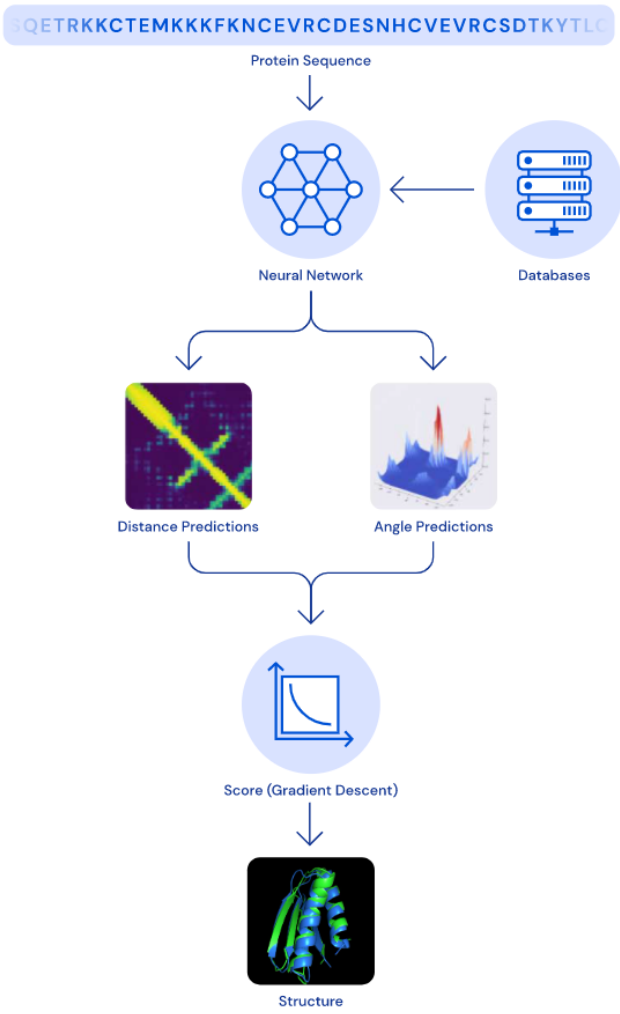
Ο ιός SARS-CoV-2 αποτελείται από περίπου 30 διαφορετικές πρωτεΐνες και μόνο περίπου δέκα από αυτές ήταν ελάχιστα κατανοητές μέχρι πρόσφατα. Η έρευνα στον τομέα της πρόβλεψης των πρωτεϊνικών δομών με τη χρήση τεχνητής νοημοσύνης, και συγκεκριμένα μηχανικής μάθησης, προσπαθεί να συμβάλει στην προσπάθεια καλύτερης κατανόησης περισσότερων πρωτεϊνών του ιού. Υπάρχουν δύο βασικές προσεγγίσεις σε ό,τι αφορά τις υπολογιστικές μεθόδους πρόβλεψης δομών πρωτεϊνών: (α) μοντελοποίηση με βάση πρότυπα (template-based modelling), η οποία προβλέπει τη δομή μιας πρωτεΐνης χρησιμοποιώντας άλλες παρόμοιες πρωτεΐνες, των οποίων η δομή έχει ήδη προσδιοριστεί πειραματικά και (β) μοντελοποίηση χωρίς τη χρήση προτύπων (template-free modelling), η οποία προβλέπει τη δομή πρωτεϊνών από το μηδέν, όταν δεν υπάρχουν δομές παρόμοιων πρωτεϊνών που θα μπορούσαν να χρησιμοποιηθούν ως πρότυπα.

Τα πιο ελπιδοφόρα αποτελέσματα στον τομέα της υπολογιστικής πρόβλεψης πρωτεϊνικών δομών – με βάση την ακρίβεια πρόβλεψης – προέρχονται από την DeepMind Technologies, μια εταιρεία με επίκεντρο την τεχνητή νοημοσύνη, η οποία ιδρύθηκε στο Λονδίνο το 2010 και εξαγοράστηκε από την Google το 2014. Η ερευνητική ομάδα της DeepMind, που ασχολείται με την πρόβλεψη των πρωτεϊνικών δομών, καθώς και το αντίστοιχο λογισμικό που αναπτύσσουν ονομάζονται AlphaFold. Η πρώτη έκδοση του AlphaFold παρουσιάστηκε στον 13^{ο} διαγωνισμό CASP το 2018 [106], όπου κατέλαβε την πρώτη θέση, προβλέποντας την ακριβέστερη δομή για τις δυσκολότερες πρωτεΐνες-στόχους. Η δεύτερη έκδοση του AlphaFold και η αντίστοιχη ομάδα της DeepMind πήραν και πάλι την πρώτη θέση στο 14^{ο} διαγωνισμό CASP τον Νοέμβριο του 2020, επιτυγχάνοντας ουσιαστικά μια βαθμολογία GDT πάνω από 90 για περίπου τα δύο τρίτα των πρωτεϊνών που εξετάστηκαν. Το εντυπωσιακό αυτό αποτέλεσμα αναγνωρίστηκε ως η σημαντικότερη πρόοδος στον τομέα της υπολογιστικής πρόβλεψης πρωτεϊνικών δομών και προκάλεσε μια ενθουσιώδη αντίδραση από ολόκληρη την επιστημονική κοινότητα και τον ίδιο τον CASP. Η μεγάλη ακρίβεια των προβλέψεων που πέτυχε το AlphaFold 2 ουσιαστικά έδωσε τη λύση σε ένα ανοιχτό πρόβλημα 50 ετών [99].

Το AlphaFold προσπαθεί να προβλέψει τις πρωτεϊνικές δομές από το μηδέν, χρησιμοποιεί δηλαδή την προσέγγιση της μοντελοποίησης χωρίς τη χρήση προτύπων (template-free modelling). Η πρώτη έκδοση του AlphaFold, για την οποία και έχουμε περισσότερες πληροφορίες σε σχέση με την αρχιτεκτονική της, κατάφερε να πετύχει υψηλό βαθμό ακρίβειας κατά την πρόβλεψη των φυσικών ιδιοτήτων μιας πρωτεΐνης και την επακόλουθη πρόβλεψη του τρισδιάστατου σχήματός της.

Το AlphaFold λύνει το πρόβλημα της πρόβλεψης σε δύο βασικά βήματα, όπως φαίνεται και στο Σχήμα 32. Το πρώτο βήμα περιλαμβάνει μια μέθοδο πρόβλεψης βασισμένη σε βαθιά μάθηση, και συγκεκριμένα σε ένα μοντέλο συνελικτικού νευρωνικού δικτύου (CNN) τύπου Res-Net, και το δεύτερο βήμα εφαρμόζει μια μέθοδο βελτιστοποίησης για την εύρεση του τελικού τρισδιάστατου σχήματος. Συνοπτικά, ο στόχος του πρώτου βήματος είναι να μετατρέψει την είσοδο (ακολουθία των αμινοξέων της πρωτεΐνης) σε δύο σημαντικούς πίνακες: έναν πίνακα κατανομής αποστάσεων και έναν πίνακα γωνιών στρέψης (torsion angles). Το βαθύ νευρωνικό δίκτυο του πρώτου βήματος χρησιμοποιεί δεδομένα για περισσότερες από 31.000 γνωστές πρωτεϊνικές δομές από την Τράπεζα Πρωτεϊνικών Δεδομένων (Protein Data Bank - PDB), χωρισμένες σε σύνολα εκπαίδευσης (training set) και δοκιμής (test set).

Μετά τον υπολογισμό αυτών των δύο πινάκων, το δεύτερο βήμα ουσιαστικά προβλέπει την τρισδιάστατη δομή της πρωτεΐνης και εκτελεί βελτιστοποίηση. Ο στόχος δηλαδή του δεύτερου βήματος είναι να μεταφραστούν οι δύο πίνακες που υπολογίστηκαν από το πρώτο βήμα σε μία τρισδιάστατη δομή. Η διαδικασία μετατροπής των πινάκων σε τρισδιάστατη δομή γίνεται με μία επαναληπτική μέθοδο καθόδου βασισμένη στην κλίση (gradient descent). Για να επιτευχθεί αυτός ο στόχος, ο αλγόριθμος ξεκινά με ένα «ομαλό» σχήμα τρισδιάστατης δομής που ενημερώνεται επαναληπτικά μέχρις ότου ο πίνακας αποστάσεων της προβλεπόμενης δομής να πλησιάσει όσο το δυνατόν περισσότερο τον αντίστοιχο πίνακα αποστάσεων που προήλθε από το πρώτο βήμα [107].

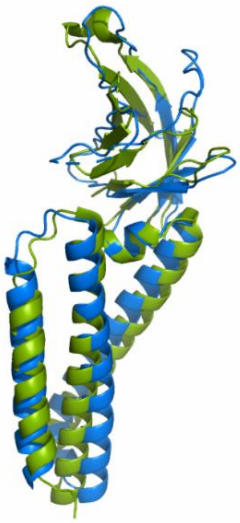


Σχήμα 32 *- Λειτουργικό διάγραμμα της λύσης AlphaFold 1 από την DeepMind* [107]

Τα ακριβή μοντέλα μηχανικής μάθησης πίσω από τη δεύτερη έκδοση του AlphaFold (AlphaFold 2) που παρουσιάστηκε στον 14ο CASP το Νοέμβριο του 2020 δεν είναι ακόμη γνωστά. Επί του παρόντος, γνωρίζουμε μόνο ότι επιτυγχάνει σημαντικά βελτιωμένη ακρίβεια σε σύγκριση με την πρώτη έκδοση, μέσω μιας νέας αρχιτεκτονικής βαθιάς μάθησης. Επίσης, η νέα έκδοση του AlphaFold λύνει το πρόβλημα σε ένα βήμα αντί σε δύο και βασίζεται σε ένα νευρωνικό δίκτυο με μηχανισμό προσοχής (attention-based network). Η συγκεκριμένη αρχιτεκτονική νευρωνικών δικτύων έγινε αρκετά δημοφιλής το 2015 μετά από ενθαρρυντικά αποτελέσματα στον τομέα της αναγνώρισης φυσικής γλώσσας και χρησιμοποιείται πλέον, όλο και περισσότερο, και στον τομέα της υπολογιστικής όρασης. Αναμένουμε σύντομα μια δημοσίευση που θα εξηγεί περισσότερες λεπτομέρειες για την αρχιτεκτονική του AlphaFold 2 [99].

Στα πλαίσια της συνεισφοράς προς την επιστημονική κοινότητα στη μάχη κατά της πανδημίας COVID-19, η DeepMind χρησιμοποίησε τα γονιδιώματα του ιού SARS-CoV-2 από ανοιχτά προσβάσιμες βάσεις δεδομένων για να προβλέψει υπολογιστικά τις δομές μερικών από τις πρωτεΐνες του ιού. Η πρώτη έκδοση των προβλέψεών της για διάφορες πρωτεΐνες του ιού, που δεν είχαν μελετηθεί εκτενώς μέχρι τότε, δόθηκε στη δημοσιότητα το Μάρτιο του 2020 [108]. Οι προβλέψεις της αφορούσαν στην πρωτεΐνη μεμβράνης (M), την πρωτεΐνη ORF3a, τις μη δομικές πρωτεΐνες (non-structural proteins) nsp2, nsp4 και nsp6, καθώς και την πρωτεάση PLpro (Papain-Like Protease).

Τα παραχθέντα τρισδιάστατα μοντέλα περιελάμβαναν βαθμούς εμπιστοσύνης για κάθε αμινοξύ στην αλυσίδα της πρωτεΐνης, ως ένδειξη της πιθανής ορθότητας των αντίστοιχων τμημάτων της δομής της. Αν και αυτές οι προβλέψεις είχαν ένα βαθμό αβεβαιότητας και το σύστημα AlphaFold ήταν ακόμη υπό ανάπτυξη, ο στόχος των ερευνητών από την ταχεία δημοσίευσή τους ήταν να συμβάλουν στην προσπάθεια της επιστημονικής κοινότητας για την καλύτερη κατανόηση της λειτουργίας του ιού και για την ανάπτυξη νέων θεραπειών. Μέχρι στιγμής, η ομάδα της DeepMind έχει δημοσιεύσει δύο ακόμα εκδόσεις των προβλέψεών της, τον Απρίλιο και τον Αύγουστο του 2020, με βελτιωμένη ακρίβεια. Στο Σχήμα 33 παρουσιάζεται η 3D αναπαράσταση της δομής της πρωτεΐνης ORF3a. Με μπλε χρώμα απεικονίζεται η πρόβλεψη της δομής από το AlphaFold και με πράσινο η πειραματικά εξακριβωμένη δομή.



Σχήμα 33 *- 3D δομή της πρωτεΐνης ORF3a του ιού SARS-CoV-2 (μπλε: πρόβλεψη AlphaFold, πράσινο: πειραματική δομή)* [108]

Εκτός από το AlphaFold της DeepMind, υπάρχουν και άλλες αξιόλογες ερευνητικές προσπάθειες για την υπολογιστική πρόβλεψη της δομής πρωτεϊνών και εν προκειμένω των πρωτεϊνών του ιού SARS-CoV-2. Για παράδειγμα, οι ερευνητές στο [99] χρησιμοποίησαν ένα τροποποιημένο νευρωνικό δικτύου τύπου ResNet, ως μέρος μιας ευρύτερης αλυσίδας συστημάτων με την ονομασία trRosetta (transform-restrained Rosetta), όπως ορίστηκε σε μια άλλη μελέτη για την πρόβλεψη πρωτεϊνικών δομών [109]. Στόχος των ερευνητών ήταν να προβλέψουν τη δομή των πρωτεϊνών που αναλύθηκαν από το AlphaFold, καθώς επίσης και των πρωτεϊνών ORF6, ORF7b, ORF8 και ORF10. Το δίκτυο trRosetta διαθέτει πολλαπλές κεφαλές εξόδου: μία που προβλέπει τις αποστάσεις μεταξύ των αμινοξέων στην ακολουθία αμινοξέων μιας πρωτεΐνης και άλλες που προβλέπουν τους προσανατολισμούς μεταξύ αυτών των αμινοξέων, όπως προσδιορίζονται από πέντε διαφορετικές γωνίες. Οι ερευνητές προσπάθησαν να βελτιώσουν τις προβλέψεις δομών του trRosetta και του AlphaFold 1 με τη χρήση προσομοιώσεων μοριακής δυναμικής και στη συνέχεια, συνέκριναν τα

αποτελέσματά τους με τις προβλέψεις δομών από μια τρίτη προσέγγιση που ονομάζεται C-I-TASSER. Το μοντέλο C-I-TASSER προέρχεται από μια άλλη ερευνητική ομάδα που κατατάχθηκε μεταξύ των κορυφαίων τριών συμμετεχόντων στους πρόσφατους διαγωνισμούς CASP. Ενώ οι ερευνητές διαπίστωσαν ότι οι προβλεπόμενες δομές μεταξύ των μοντέλων είχαν κάποιες διαφορές μεταξύ τους, εντούτοις συνέκλιναν σε ό,τι αφορά τη δομή της PLpro, τη δομή ενός τμήματος της nsp4 και τη δομή της πρωτεΐνης μεμβράνης M [110].

4.3 Ανακάλυψη Θεραπειών και Εμβολίων

4.3.1 Ο Ρόλος της Τεχνητής Νοημοσύνης στην Ανάπτυξη Φαρμάκων

Τα τελευταία χρόνια, όλο και περισσότερες φαρμακευτικές εταιρείες επενδύουν στην τεχνητή νοημοσύνη. Για παράδειγμα, η Pfizer χρησιμοποιεί το Watson, ένα σύστημα μηχανικής μάθησης της IBM, για να ενισχύσει την έρευνά της για ανοσο-ογκολογικά φάρμακα. Η Sanofi έχει συνάψει συνεργασία με τη βρετανική startup Excientia και χρησιμοποιεί την πλατφόρμα τεχνητής νοημοσύνης της για να ανακαλύψει θεραπείες μεταβολικών ασθενειών ενώ η Genentech (θυγατρική της Roche) χρησιμοποιεί ένα σύστημα τεχνητής νοημοσύνης από την εταιρεία GNS Healthcare για την αναζήτηση θεραπειών για τον καρκίνο. Οι περισσότερες σημαντικές φαρμακευτικές εταιρείες σήμερα έχουν παρόμοιες συνεργασίες ή εσωτερικά προγράμματα ανάπτυξης τεχνητής νοημοσύνης [111], [112].

Το εκτιμώμενο κόστος ανάπτυξης ενός νέου φαρμάκου με παραδοσιακές μεθόδους υπερβαίνει τα 2,5 δισεκατομμύρια δολάρια και ο εκτιμώμενος χρόνος πριν ένα φάρμακο καταλήξει στην αγορά είναι κατά μέσο όρο περισσότερο από 10 χρόνια [113], [114]. Το μεγαλύτερο μέρος των χρημάτων και του χρόνου δαπανάται σε πειραματικά φάρμακα που αποτυγχάνουν κάπου ανάμεσα στις κλινικές δοκιμές Φάσης I και στην τελική έγκρισή τους από τις ρυθμιστικές αρχές. Το κόστος αυτό συνήθως μετακυλίεται στους ασθενείς, καθιστώντας πολλά φάρμακα απρόσιτα σε ένα μεγάλο ποσοστό ανθρώπων και κυρίως στους λιγότερο εύπορους και στους ανασφάλιστους. Οι ανάγκες για βελτίωση της αποδοτικότητας και για μείωση του χρόνου και του κόστους έχουν ωθήσει τις εταιρείες του φαρμακευτικού κλάδου στην υιοθέτηση τεχνολογιών τεχνητής νοημοσύνης. Οι τεχνολογίες αυτές στοχεύουν στο να τις βοηθήσουν να αναπτύξουν φάρμακα με πιο αποτελεσματικό, λιγότερο δαπανηρό και ταχύτερο τρόπο, βελτιώνοντας ταυτόχρονα την ποιότητα και, σε ορισμένες περιπτώσεις, επιτρέποντας ακόμα και την ανάπτυξη εξατομικευμένων θεραπειών.

Σε αυτή την ενότητα, θα περιγράψουμε τον τρόπο με τον οποίο η πρόοδος στις εφαρμογές της τεχνητής νοημοσύνης στη φαρμακευτική βιομηχανία και τη Φαρμακολογία, γενικότερα, συμβάλλει στην καταπολέμηση της πανδημίας COVID-19, μέσω διαφορετικών μεθόδων για ανάπτυξη θεραπειών και εμβολίων.

4.3.2 Επαναστόχευση Φαρμάκων

Η επαναστόχευση φαρμάκων (drug repurposing) είναι μια στρατηγική για την εύρεση νέων ενδείξεων (σκοπών / εφαρμογών) για εγκεκριμένα ή πειραματικά φάρμακα που δεν έχουν ακόμη εγκριθεί πλήρως. Επειδή η ασφάλεια αυτών των φαρμάκων έχει ήδη δοκιμαστεί σε κλινικές δοκιμές για άλλες πιθανές εφαρμογές, η επαναστόχευση γνωστών φαρμάκων μπορεί να προσφέρει θεραπείες στους ασθενείς πολύ πιο γρήγορα και με μικρότερο κόστος, συγκρινόμενη με την ανάπτυξη νέων φαρμάκων. Εδώ και δεκαετίες, ακαδημαϊκά ιδρύματα καθώς και άλλοι ερευνητικοί οργανισμοί έχουν υποστηρίξει την ιδέα ότι η αναζήτηση ανάμεσα στο σύνολο των διαθέσιμων υφιστάμενων φαρμάκων και η εκτέλεση διαφόρων δοκιμών θα μπορούσαν να αποκαλύψουν νέες εφαρμογές για αυτά. Ήδη

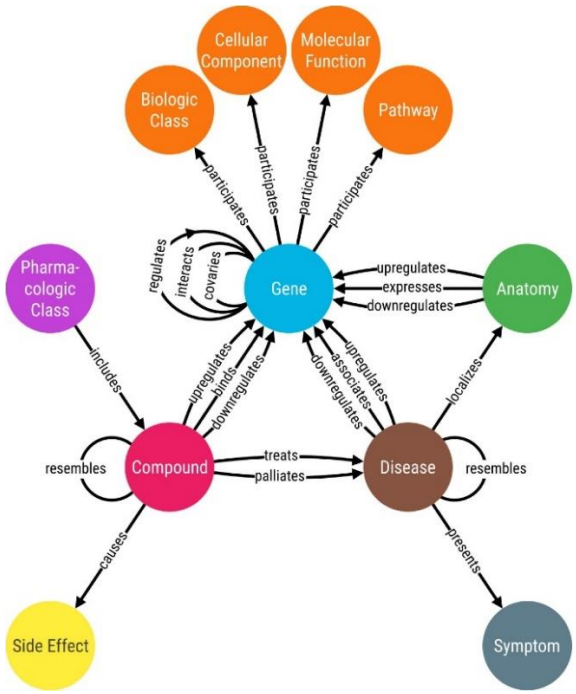
τέτοιες προσεγγίσεις έχουν οδηγήσει στην επιτυχή χρήση φαρμάκων, που αρχικά προορίζονταν για κάποιες ασθένειες, ως θεραπείες για άλλες.

Εκτός από τους παραπάνω λόγους, η επαναστόχευση φαρμάκων είναι μια βολική εναλλακτική λύση όταν η ανάγκη για φαρμακευτική αντιμετώπιση σε ένα απροσδόκητο σενάριο είναι επείγουσα, όπως συμβαίνει στην περίπτωση εμφάνισης νέων παθογόνων οργανισμών και των αντίστοιχων μεταδοτικών ασθενειών. Η ανάγκη αυτή προφανώς μεγεθύνεται σε συνθήκες πανδημίας, όπως η πανδημία COVID-19, λόγω της ταχείας εξάπλωσης της ασθένειας και της άμεσης ανάγκης για εύρεση αποτελεσματικών θεραπειών που θα μπορούσαν να μειώσουν τις συνέπειές της. Ο σημαντικός αριθμός των επιτυχώς επαναστοχευμένων φαρμάκων στο πρόσφατο παρελθόν οδήγησε πολλούς ερευνητές σε συστηματική αναζήτηση ανάμεσα σε γνωστά φάρμακα, με την ελπίδα ότι έτσι θα μπορέσουν να ανακαλύψουν κάποια που να αντιμετωπίζουν τη νόσο COVID-19.

Πολλές εφαρμογές τεχνητής νοημοσύνης στον τομέα της επαναστόχευσης φαρμάκων βασίζονται στα λεγόμενα δίκτυα φαρμάκων-στόχων (drug-target networks). Ένα δίκτυο φαρμάκων-στόχων είναι ουσιαστικά ένας διμερής γράφος που αποτελείται από εγκεκριμένα φάρμακα και πρωτεΐνες, που συνδέονται μεταξύ τους μέσω δυαδικών συσχετίσεων φαρμάκων-στόχων. Ωστόσο, επειδή οι στόχοι των φαρμάκων δε λειτουργούν μεμονωμένα από το πολύπλοκο σύστημα πρωτεϊνών που ορίζουν τις λειτουργίες των κυττάρων με τα οποία σχετίζονται, κάθε αλληλεπίδραση φαρμάκου-στόχου θα πρέπει να εξετάζεται σε ένα ευρύτερο πλαίσιο. Συγκεκριμένα, οι θεραπευτικές παρεμβάσεις πρέπει να εξετάσουν τις διαταραχές που προκαλούν οι ασθένειες, πέραν των γενετικών και γονιδιωματικών πληροφοριών. Αυτό ακριβώς είναι το αντικείμενο της δικτυακής ιατρικής (network medicine). Ουσιαστικά, η δικτυακή ιατρική είναι ένας κλάδος που επιδιώκει να επαναπροσδιορίσει τις ασθένειες και τις αντίστοιχες θεραπευτικές παρεμβάσεις μέσα από μία ολοκληρωμένη προσέγγιση, χρησιμοποιώντας αρχές από τη βιολογία καθώς και μεθοδολογίες από την επιστήμη των δικτύων (network science), ανοίγοντας έτσι νέες προοπτικές στο σχεδιασμό φαρμάκων. Με αυτό τον τρόπο, η γνώση των αλληλεπιδράσεων μεταξύ των στόχων, των φαρμάκων και των ανθρώπινων ασθενειών μπορεί να παρέχει ενδείξεις για πιθανή επαναστόχευση των φαρμάκων. Φάρμακα που στοχεύουν μια ασθένεια θα μπορούσαν να είναι αποτελεσματικά και σε μια άλλη μέσω ενός κοινού δικτύου αλληλεπίδρασης μεταξύ πρωτεϊνών [115].

Στην περίπτωση του ιού SARS-CoV-2, γνωρίζουμε ότι για τον επιτυχή πολλαπλασιασμό του κατά τη διάρκεια της μόλυνσης ενός ανθρώπου εκμεταλλεύεται συγκεκριμένους παράγοντες στα ανθρώπινα κύτταρα, όπως το ένζυμο ACE2, την πρωτεάση TMPRSS2 και τη φουρίνη (furin). Η συστηματική στόχευση των αλληλεπιδράσεων μεταξύ των πρωτεϊνών του ιού και των πρωτεϊνών των ανθρώπινων κυττάρων, που είναι γνωστές ως αλληλεπίδρωμα (interactome), προσφέρει μια νέα στρατηγική για αποτελεσματική επαναστόχευση φαρμάκων για την ασθένεια COVID-19. Σε αυτή την ενότητα, παρουσιάζουμε αρκετές εφαρμογές τεχνητής νοημοσύνης στην επαναστόχευση φαρμάκων για την αντιμετώπιση της ασθένειας COVID-19, που χρησιμοποιούν τους επονομαζόμενους ιατρικούς γράφους γνώσης (medical knowledge graphs), όπως περιγράφουμε με περισσότερες λεπτομέρειες παρακάτω.

Ο τομέας της δικτυακής ιατρικής, στον οποίο αναφερθήκαμε, περιλαμβάνει την κατασκευή αυτών των ιατρικών γράφων γνώσης που περιέχουν τις πολύπλοκες σχέσεις μεταξύ των διαφόρων οντοτήτων (π.χ. ασθένειες, φάρμακα, πρωτεΐνες) και στοχεύει στην πρόβλεψη νέων σχέσεων / συνδέσεων μεταξύ υφιστάμενων φαρμάκων και ασθενειών [116]. Χρησιμοποιώντας τα διανύσματα των χαρακτηριστικών των φαρμάκων και των ασθενειών, μπορούμε να υπολογίσουμε τις ομοιότητές τους και επομένως, να προσδιορίσουμε πιθανώς αποτελεσματικά φάρμακα για μια συγκεκριμένη ασθένεια. Ένα παράδειγμα της δομής ενός γράφου γνώσης φαίνεται στο Σχήμα 34.



Σχήμα 34 *- Δομή ενός γράφου γνώσης που χρησιμοποιείται για την επαναστόχευση φαρμάκων* [116]

Μια πρόκληση για τις μεθόδους που βασίζονται σε γράφους γνώσης είναι η κλιμάκωση. Στην πράξη, οι συγκεκριμένοι γράφοι είναι πολύ μεγάλοι και ο αριθμός των οντοτήτων που περιέχουν (κόμβοι και ακμές) μπορεί να αντιστοιχεί σε πολλά εκατομμύρια. Τα υπάρχοντα συστήματα μηχανικής μάθησης είναι κυρίως σχεδιασμένα για μεγάλα σύνολα δεδομένων με κανονικές δομές, αλλά όχι για γράφους μεγάλης κλίμακας. Για την αντιμετώπιση της αναγκαίας κλιμάκωσης, έχουν αναπτυχθεί διάφορες μέθοδοι, ειδικά σχεδιασμένες για τη μάθηση αναπαράστασης από γράφους μεγάλης κλίμακας. Για παράδειγμα, οι ερευνητές στο [117] ανέπτυξαν ένα σύστημα υψηλής απόδοσης που ονομάζεται GraphVite και το οποίο θα μπορούσε να χρησιμοποιηθεί για προβλήματα επαναστόχευσης φαρμάκων, καθώς δύναται να επεξεργαστεί αποτελεσματικά γράφους γνώσης με δεκάδες ή ακόμα και εκατοντάδες εκατομμύρια κόμβους.

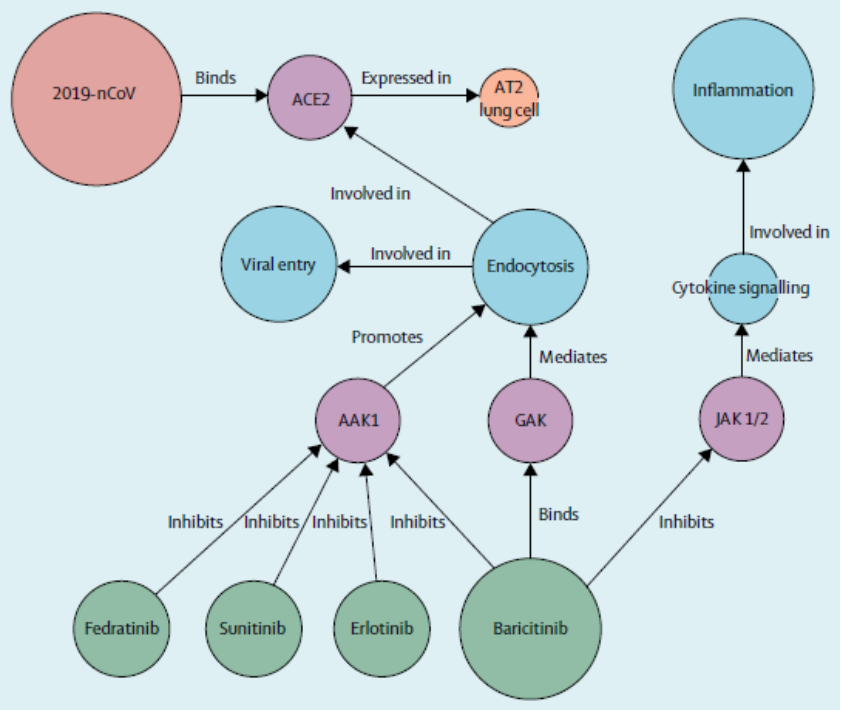
Ακόμη και πριν από την πανδημία COVID-19, υπήρχε αυξημένο ερευνητικό ενδιαφέρον γύρω από την ανάπτυξη τεχνικών μάθησης αναπαράστασης γράφων με στόχο την επαναστόχευση φαρμάκων. Για παράδειγμα, οι ερευνητές στο [118] κατασκεύασαν ένα γράφο ιατρικών γνώσεων που περιλαμβάνει φάρμακα, ασθένειες και γονίδια (ή αντίστοιχες πρωτεΐνες που τα κωδικοποιούν) από τη βιοϊατρική βιβλιογραφία και στη συνέχεια, χρησιμοποίησαν τεχνικές ενσωμάτωσης γράφων (graph embedding) για την πρόβλεψη των συνδέσεων μεταξύ φαρμάκων και ασθενειών. Επίσης, αποτύπωσαν την αβεβαιότητα των σχέσεων που προέρχονται από τη βιβλιογραφία και χρησιμοποίησαν τις σχέσεις που προέβλεψαν για να δημιουργήσουν υποθέσεις επαναστόχευσης φαρμάκων. Αυτές οι υποθέσεις μπορούν, στη συνέχεια, να επαληθευτούν ανεξάρτητα μέσω εξωτερικών πηγών βιβλιογραφίας.

Το ξέσπασμα της πανδημίας COVID-19 είχε ως αποτέλεσμα την επιτάχυνση της έρευνας σχετικά με την επαναστόχευση των φαρμάκων. Σε μια πρόσφατη μελέτη [119], οι ερευνητές ανέπτυξαν μια μέθοδο βασισμένη σε νευρωνικό δίκτυο γράφων (Graph Neural Network – GNN) και παρουσίασαν μια εφαρμογή της για την εύρεση θεραπειών έναντι του ιού SARS-CoV-2, προσδιορίζοντας 81 πιθανά φάρμακα ως υποψήφια για επαναστόχευση. Αρχικά, αποκάλυψαν τις πρωτογενείς συνέπειες του ιού

στους ιστούς των πνευμόνων, καθώς και παρατηρούμενες συννοσηρότητες που σχετίζονται με καρδιαγγειακές παθήσεις. Στη συνέχεια, προέβλεψαν ότι ο ιός θα μπορούσε να επηρεάσει και άλλους ιστούς. Με βάση αυτά τα ευρήματα, εξέτασαν τρεις διαφορετικές στρατηγικές επαναστόχευσης φαρμάκων και κατέταξαν όλα τα εγκεκριμένα φάρμακα με βάση την πιθανή αποτελεσματικότητά τους σε ασθενείς με COVID-19. Η χρησιμότητα των προβλέψεών τους επικυρώθηκε στην πράξη, καθώς κάποια από τα φάρμακα που προέβλεψαν βρίσκονται ήδη σε κλινικές δοκιμές.

Μια άλλη ενδιαφέρουσα εφαρμογή σχετικά με την επαναστόχευση φαρμάκων για την ασθένεια COVID-19 προέρχεται από τη BenevolentAI, μια εταιρεία που ιδρύθηκε το 2013. Η εταιρεία αυτή – σύμφωνα με την ιστοσελίδα της – χρησιμοποιεί κατάλληλη τεχνολογία για να σαρώσει τεράστιες ποσότητες ιατρικών δεδομένων, από κλινικές δοκιμές μέχρι επιστημονικές δημοσιεύσεις. Στόχος της είναι να βοηθήσει τους ερευνητές να βρουν νέους τρόπους για τη θεραπεία ασθενειών, καθώς και εξατομικευμένα φάρμακα ανάλογα με τα χαρακτηριστικά του κάθε ασθενούς [120]. Ο γράφος γνώσης της BenevolentAI είναι ένα μεγάλο σύνολο δομημένων ιατρικών πληροφοριών, συμπεριλαμβανομένων πολυάριθμων συνδέσεων μεταξύ διαφορετικών οντοτήτων που έχουν εξαχθεί μέσω τεχνικών μηχανικής μάθησης.

Σε ό,τι αφορά τον ιό SARS-CoV-2, μια ομάδα ερευνητών της BenevolentAI χρησιμοποίησε μια μορφή του ανωτέρω γράφου γνώσης για να αναζητήσει εγκεκριμένα φάρμακα που θα μπορούσαν να βοηθήσουν στην καταπολέμησή του, εστιάζοντας σε εκείνα που δυνητικά εμποδίζουν τη διαδικασία ιογενούς λοίμωξης [121]. Μέσω της διαδικασίας αυτής προσδιόρισαν ότι το baricitinib, ένα φάρμακο που χρησιμοποιείται για την αντιμετώπιση της ρευματοειδούς αρθρίτιδας, θα μπορούσε να είναι μια πιθανή θεραπεία για την ασθένεια COVID-19. Πιο συγκεκριμένα, προέβλεψαν ότι το baricitinib μπορεί να μειώσει την ικανότητα του ιού να μολύνει τα κύτταρα των πνευμόνων μέσω της αναστολής της κινάσης AAK1 (AP2-Associated protein Kinase 1), όπως φαίνεται και στο Σχήμα 35. Μετά από τυχαιοποιημένες κλινικές δοκιμές, βρέθηκε ότι ο συνδυασμός των φαρμάκων baricitinib και remdesivir εμφανίζεται να είναι καλύτερος συγκριτικά με τη χρήση μόνο του remdesivir. Συγκεκριμένα, ο συνδυασμός των δύο φαρμάκων φάνηκε να οδηγεί σε μείωση του χρόνου αποκατάστασης και να επιταχύνει τη βελτίωση της κλινικής εικόνας ασθενών με COVID-19, κυρίως εκείνων που λάμβαναν υψηλή ροή οξυγόνου (high flow oxygen) με μη επεμβατικό τρόπο [122]. Λίγο πριν από τη συγκεκριμένη δημοσίευση, ο Εθνικός Οργανισμός Τροφίμων και Φαρμάκων των ΗΠΑ (Food and Drug Administration - FDA) είχε ήδη παράσχει άδεια χρήσης έκτακτης ανάγκης για αυτό το κοκτέιλ φαρμάκων [123].



Σχήμα 35 *- Μέρος του γράφου γνώσης της BenevolentAI* [4] [121]

Μια άλλη ερευνητική ομάδα κατασκεύασε έναν αναλυτικό γράφο γνώσης για την ασθένεια COVID-19 (με την ονομασία CoV-KGE), ο οποίος περιλαμβάνει 15 εκατομμύρια ακμές που αντιστοιχούν σε 39 διαφορετικούς τύπους σχέσεων που συνδέουν φάρμακα, ασθένειες, πρωτεΐνες, γονίδια και εκφράσεις γονιδίων και πρωτεϊνών, αναλύοντας ένα μεγάλο όγκο 24 εκατομμυρίων επιστημονικών δημοσιεύσεων στο PubMed [124]. Χρησιμοποιώντας υπολογιστικούς πόρους στο cloud σε συνδυασμό με μία μέθοδο βαθιάς μάθησης δικτύων, οι ερευνητές προσδιόρισαν 41 υποψήφια φάρμακα (συμπεριλαμβανομένης της δεξαμεθαζόνης και της μελατονίνης) για τη θεραπεία της νόσου COVID-19 μέσω της επαναστόχευσης. Όπως ανέφεραν, ο στόχος τους δεν ήταν να προτείνουν συγκεκριμένα φάρμακα ως θεραπεία για την ασθένεια COVID-19, αλλά να δημιουργήσουν μία δομημένη μεθοδολογία βαθιάς μάθησης που μπορεί να καταδείξει συγκεκριμένα υπάρχοντα φάρμακα για περαιτέρω μελέτη.

Μια παρόμοια προσέγγιση για την κατασκευή ενός γράφου γνώσης που συνδέει τις ανθρώπινες πρωτεΐνες, τις πρωτεΐνες του ιού SARS-CoV-2, καθώς και φάρμακα μέσω της ανάλυσης βάσεων δεδομένων που καταγράφουν τις σχέσεις μεταξύ αυτών των οντοτήτων, παρουσιάζεται στο [125]. Ο προκύπτων γράφος χρησιμοποιήθηκε, στη συνέχεια, για την πρόβλεψη μιας λίστας δυνητικά αποτελεσματικών υποψηφίων φαρμάκων. Η λίστα αυτή περιορίστηκε περαιτέρω χρησιμοποιώντας ένα μοντέλο επεξεργασίας φυσικής γλώσσας που εφαρμόστηκε στη βάση δεδομένων του PubMed μαζί με κατάλληλα φίλτρα (σε σχέση με τα συστατικά των υποψηφίων φαρμάκων, τους κορωνοϊούς και τις αντίστοιχες πρωτεΐνες τους). Τέλος, οι ερευνητές προσδιόρισαν ένα συγκεκριμένο αναστολέα της πολυμεράσης PARP1 (Poly ADP-Ribose Polymerase 1) και συγκεκριμένα τον CVL218, ως ένα πολλά υποσχόμενο φάρμακο, το οποίο ήδη βρίσκεται σε φάση κλινικών δοκιμών.

[4] 2019-nCoV = ο νέος κορωνοϊός SARS-CoV2, AAK1 = Adaptor-Associated protein Kinase 1 ή AP2-Associated protein Kinase 1, GAK = Cyclin G-Associated Kinase, JAK 1/2 = Janus Kinase 1 ή 2.

4.3.3 Ανακάλυψη Νέων Φαρμάκων

Εκτός από τη μέθοδο επαναστόχευσης φαρμάκων που μελετήθηκε στην προηγούμενη ενότητα, αρκετοί ερευνητές εστιάζουν τις προσπάθειές τους στην ανακάλυψη εντελώς νέων χημικών ενώσεων για χρήση κατά του ιού SARS-CoV-2. Σε αυτή την ενότητα, θα παρουσιάσουμε παραδείγματα μελετών και εφαρμογών, οι οποίες έχουν ως βάση την τεχνητή νοημοσύνη και προηγμένες υπολογιστικές τεχνικές ανάλυσης δεδομένων, προκειμένου να επισπεύσουν την εξεύρεση νέων αποτελεσματικών φαρμάκων για την αντιμετώπιση της ασθένειας COVID-19.

Αν και ο αριθμός των χημικών ενώσεων που μπορούν θεωρητικά να παραχθούν είναι ουσιαστικά απεριόριστος, οι συμβατικές υπολογιστικές προσεγγίσεις σχεδιασμού φαρμάκων συνήθως περιλαμβάνουν μόνο έναν περιορισμένο αριθμό από πιθανά συστατικά ή χρησιμοποιούν εξελιγμένες μεθόδους αναζήτησης για να πάρουν κάποια δείγματα από μία συγκεκριμένη υποπεριοχή του συνόλου των χημικών ενώσεων. Για να μπορέσουν οι επιστήμονες να εκμεταλλευθούν το σύνολο του χώρου των χημικών ενώσεων που διαθέτουν πιθανές φαρμακολογικές ιδιότητες, έχουν αναπτυχθεί νέες υπολογιστικές μέθοδοι που αξιοποιούν την πρόσφατη πρόοδο στη βαθιά μάθηση και την τεχνητή νοημοσύνη γενικότερα. Τέτοιες τεχνικές μπορούν να εξάγουν αυτόματα λεπτομερείς πληροφορίες, χωρίς την ανάγκη για χειροκίνητη σχεδίαση χαρακτηριστικών (manual feature engineering), και να αντιληφθούν μη γραμμικές συσχετίσεις μεταξύ μοριακών δομών και των αντίστοιχων βιολογικών και φαρμακολογικών ιδιοτήτων τους [126].

Υπάρχουν διάφορες εφαρμογές που χρησιμοποιούν μηχανική μάθηση για την εύρεση νέων φαρμάκων κατά της ασθένειας COVID-19. Για παράδειγμα, ερευνητές από την εταιρεία Insilico Medicine στο Χονγκ Κονγκ εφάρμοσαν μεθόδους μηχανικής μάθησης για να βρουν αναστολείς μιας πρωτεΐνης του ιού SARS-CoV-2 και συγκεκριμένα της πρωτεάσης 3CLpro (3C-Like Protease) [127]. Τα μοντέλα τους χρησιμοποίησαν τρεις τύπους εισόδων: την κρυσταλλική δομή της πρωτεΐνης, τους προσδέτες (ligands)[5] και το τρισδιάστατο μοντέλο της πρωτεΐνης. Για κάθε τύπο εισόδου, οι ερευνητές προσπάθησαν να προσαρμόσουν 28 διαφορετικά μοντέλα νευρωνικών δικτύων, όπως για παράδειγμα Generative Autoencoders και Generative Adversarial Networks (GANs). Επίσης, διερεύνησαν πιθανά υποψήφια φάρμακα χρησιμοποιώντας μια μέθοδο ενισχυτικής μάθησης με μια συνάρτηση ανταμοιβής που πριμοδοτούσε ορισμένα επιθυμητά χαρακτηριστικά. Οι ερευνητές επιβεβαίωσαν ότι τα προτεινόμενα υποψήφια μόρια ήταν πράγματι διαφορετικά από τα υπάρχοντα φάρμακα. Όπως αναφέρουν, το έργο τους είναι ακόμη σε εξέλιξη, καθώς συνεχίζουν να συνθέτουν και να δοκιμάζουν τα παραγόμενα μόρια.

Η πιο πάνω εφαρμογή είναι ένα παράδειγμα μιας πολλά υποσχόμενης προσέγγισης στην ανακάλυψη φαρμάκων, καθώς αποδεικνύει στην πράξη πώς ανάλογες μέθοδοι τεχνητής νοημοσύνης μπορούν να εκμεταλλευθούν το μεγάλο χώρο των χημικών ουσιών με δυνητικά φαρμακολογικές ιδιότητες και να εξάγουν αυτόματα χρήσιμες πληροφορίες από δεδομένα πολλών διαστάσεων. Με αυτό τον τρόπο, είναι σε θέση να σχεδιάσουν νέα μόρια και να «μάθουν» τις σχέσεις μεταξύ των μοριακών δομών και των φαρμακολογικών ιδιοτήτων τους. Επίσης, οι προσεγγίσεις αυτές είναι οικονομικά αποδοτικές, γρήγορες και μπορούν να βοηθήσουν στην ανακάλυψη νέων φαρμακευτικών ουσιών στη μάχη κατά της πανδημίας COVID-19.

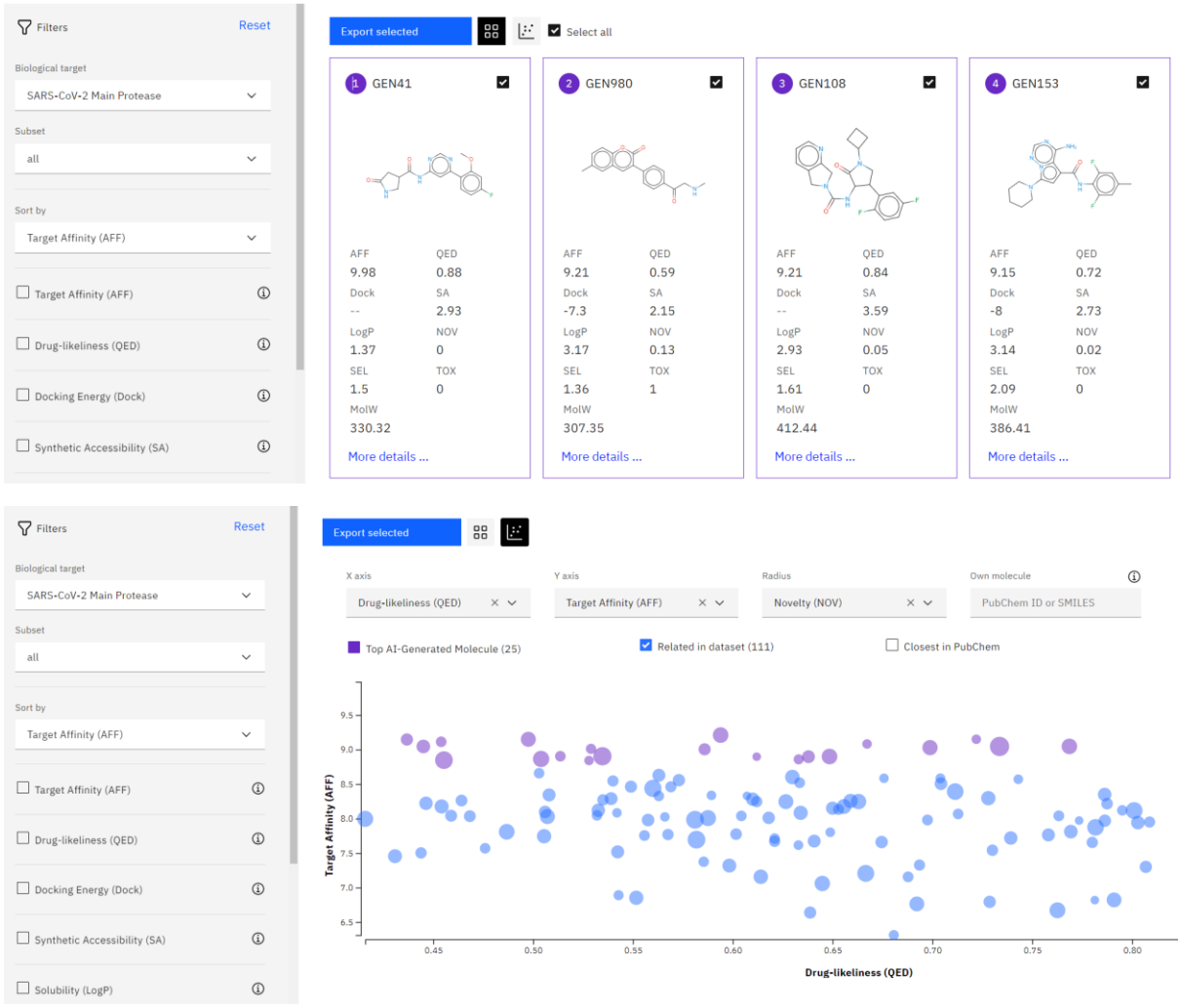
[5] Ένας προσδέτης (ligand) είναι ένα μικρό μόριο ικανό να συνδέεται με πρωτεΐνες μέσω αλληλεπιδράσεων όπως ιοντικοί δεσμοί, δεσμοί υδρογόνου, αλληλεπιδράσεις Van der Waals και υδρόφοβες αλληλεπιδράσεις. Σε ορισμένες περιπτώσεις, ένας προσδέτης χρησιμεύει επίσης ως μόριο ενεργοποίησης κάποιου βιολογικού σήματος.

Σε μια άλλη εφαρμογή [128], οι ερευνητές χρησιμοποίησαν μια άλλη μέθοδο ενισχυτικής μάθησης για την ανακάλυψη χημικών ενώσεων που δρουν ως αναστολείς κατά της ίδιας πρωτεάσης 3CLpro του ιού SARS-CoV-2. Αρχικά, δημιούργησαν μια λίστα με 284 μόρια που είναι γνωστό ότι δρουν ως αναστολείς ενάντια στον ιό SARS-CoV-1. Στη συνέχεια, κατάτμησαν τα πιο πάνω μόρια σε 316 τύπους θραυσμάτων (fragments), τα οποία συνδύασαν χρησιμοποιώντας ένα προηγμένο δίκτυο βαθιάς ενισχυτικής μάθησης που ονόμασαν ADQN-FBDD (Advanced Deep Q-learning Network with Fragment-Based Drug Design). Το συγκεκριμένο νευρωνικό δίκτυο ανταμείβει διάφορες πτυχές των ανακαλυφθέντων μορίων, όπως για παράδειγμα τη φαρμακο-ομοιότητα (drug-likeness)[6] και το κατά πόσον τα μόρια συμπεριλαμβάνουν προκαθορισμένα επιθυμητά θραύσματα. Μεταξύ των 4922 αρχικών υποψήφιων ενώσεων, οι 47 με τις υψηλότερες βαθμολογίες αξιολογήθηκαν με προσομοιώσεις μοριακής πρόσδεσης (molecular docking simulations). Από αυτές, οι ερευνητές τελικά επέλεξαν τη φαινομενικά πιο υποσχόμενη ένωση και παρήγαγαν διάφορες παραλλαγές της για περαιτέρω δοκιμές.

Μια ακόμα ερευνητική ομάδα από την εταιρεία IBM ανέπτυξε μια λύση που ονομάζεται CogMol (Controlled Generation of Molecules) [129], με στόχο να συμβάλει στην προσπάθεια των ερευνητών ανά τον κόσμο στο σχεδιασμό φαρμάκων και στην επιτάχυνση της εύρεσης θεραπειών μολυσματικών ασθενειών, όπως η ασθένεια COVID-19. Η συγκεκριμένη εφαρμογή βασίζεται σε γενετικούς αλγορίθμους βαθιάς μάθησης (generative modelling). Η άποψη των ερευνητών είναι ότι, αν και τα μοντέλα τεχνητής νοημοσύνης έχουν δείξει καλά αποτελέσματα στην αυτοματοποίηση της ανακάλυψης μορίων, εντούτοις δεν είναι το ίδιο αποτελεσματικά σε προβλήματα σχεδιασμού με πολλαπλούς περιορισμούς, έχουν περιορισμένες δυνατότητες επέκτασης και είναι πολύ δύσκολη και χρονοβόρα η επανεκπαίδευσή τους. Οι ερευνητές ανέπτυξαν την εφαρμογή CogMol με τρόπο ώστε να ξεπεράσουν τις πιο πάνω δυσκολίες.

Σε μια σχετική δημοσίευσή τους [130], οι δημιουργοί της εφαρμογής CogMol εφαρμόζουν τη λύση τους σε τρεις πρωτεΐνες του ιού SARS-CoV-2 και συγκεκριμένα στη ρεπλικάση nsp9, στην κύρια πρωτεάση Mpro (ή 3CLpro) και στην περιοχή πρόσδεσης στον υποδοχέα (receptor binding domain - RBD) της πρωτεΐνης ακίδας S. Με τον τρόπο αυτό, ανακαλύπτουν συγκεκριμένες χημικές ενώσεις - υποψήφια φάρμακα για την ασθένεια COVID-19, μέσω της δυνατότητας πρόσδεσής τους στις ανωτέρω πρωτεΐνες του ιού SARS-CoV-2. Τέλος, χρησιμοποιούν ένα ακόμα βαθύ νευρωνικό δίκτυο για να προβλέψουν την τοξικότητα των ανακαλυφθεισών ουσιών, προκειμένου να αποκλείσουν πιθανώς τοξικά υποψήφια φάρμακα που θα είχαν χαμηλή πιθανότητα επιτυχίας στις φάσεις κλινικών δοκιμών. Παρέχοντας ελεύθερη πρόσβαση στα αποτελέσματα πρόβλεψης της εφαρμογής CogMol μέσω διαδικτύου, η ερευνητική ομάδα της IBM θεωρεί ότι μπορεί να συμβάλει στην επιτάχυνση της ανακάλυψης υποσχόμενων νέων θεραπειών για την ασθένεια COVID-19, καθώς και για άλλες μολυσματικές ασθένειες στο μέλλον. Ένα στιγμιότυπο των αποτελεσμάτων αυτών φαίνεται στο Σχήμα 36.

[6] Η φαρμακο-ομοιότητα (drug-likeness) είναι μία ποιοτική έννοια που χρησιμοποιείται στο σχεδιασμό φαρμάκων και σχετίζεται με το κατά πόσον μια ουσία έχει ιδιότητες που την καθιστούν κατάλληλη για χρήση ως φάρμακο (π.χ. σε σχέση με τη διαλυτότητά της, το μοριακό της βάρος, κλπ.). Υπολογίζεται από τη μοριακή δομή της προτού ακόμη συντεθεί και δοκιμαστεί η ουσία.



Σχήμα 36 *- Στιγμιότυπα αποτελεσμάτων της εφαρμογής IBM CogMol για την εύρεση νέων φαρμάκων κατά της COVID-19* [131]

Σε μια άλλη μελέτη, οι ερευνητές προτείνουν τη χρήση ενός άλλου μοντέλου βασισμένου σε ένα βαθύ νευρωνικό δίκτυο, μέσω του οποίου κατάφεραν να ανακαλύψουν 31 νέα μόρια ικανά να αναστείλουν την κύρια πρωτεάση Mpro (ή 3CLpro) του ιού SARS-CoV-2 [132]. Χρησιμοποιώντας ειδικά φίλτρα σε σχέση με φυσικοχημικές ιδιότητες, οι συγγραφείς διασφάλισαν ότι τα παραγόμενα μόρια αντιστοιχούν πράγματι σε νέες χημικές ενώσεις και μάλιστα διαθέτουν φαρμακολογικές ιδιότητες. Αν και τα μόρια αυτά σχεδιάστηκαν ώστε να μπορούν να προσδεθούν στην αντίστοιχη περιοχή πρόσδεσης της πρωτεάσης Mpro, περαιτέρω ανάλυσή τους αποκάλυψε κάποια ομοιότητά τους με τους αναστολείς της πρωτεάσης του ιού HIV. Ενδιαφέρον επίσης παρουσιάζει το γεγονός ότι δύο από τα παραγόμενα μόρια είχαν υψηλή ομοιότητα με ένα φυσικό φυτικό προϊόν (Aurantiamide). Όπως αναφέρουν οι ερευνητές, το αποτέλεσμα αυτό είναι μάλλον ενθαρρυντικό καθώς χημικές ενώσεις που είναι παρόμοιες με φυσικά στοιχεία είναι συνήθως εύκολο να δημιουργηθούν στο εργαστήριο και τείνουν να έχουν λιγότερες παρενέργειες σε σύγκριση με μη φυσικές ενώσεις.

4.3.4 Ανακάλυψη Πιθανών Στόχων Εμβολίων

Η εξαιρετικά μεγάλη πολυπλοκότητα του ανθρώπινου ανοσοποιητικού συστήματος καθώς και οι διαφορές του μεταξύ ανθρώπων και πληθυσμιακών ομάδων καθιστούν την ανάπτυξη εμβολίων μία αρκετά επίπονη διαδικασία. Όλο και περισσότερο, στρατηγικές που εκμεταλλεύονται μεθόδους και

τεχνολογίες τεχνητής νοημοσύνης χρησιμοποιούνται σε διάφορα στάδια της ανάπτυξης εμβολίων, όπως για παράδειγμα στην εύρεση των καταλληλότερων στόχων έναντι του ιού SARS-CoV-2 [133].

Το ανθρώπινο ανοσοποιητικό σύστημα διαθέτει διάφορους μηχανισμούς για την καταπολέμηση των ιών. Κατά τη διαδικασία δημιουργίας μιας ανοσολογικής απόκρισης, τα Β-λεμφοκύτταρα ή αλλιώς Β-κύτταρα (ένας τύπος λευκών αιμοσφαιρίων του οργανισμού μας) παράγουν αντισώματα, τα οποία επιτίθενται σε ένα μέρος του παθογόνου (ιού) γνωστού ως αντιγόνο. Αυτός ο πρώτος μηχανισμός είναι γνωστός ως χυμική ανοσία (humoral immunity). Εκτός από την παραγωγή αντισωμάτων που εξουδετερώνουν τον ιό μέσω Β-λεμφοκυττάρων, όπως περιγράφηκε πιο πάνω, το σώμα χρησιμοποιεί επίσης Τ-λεμφοκύτταρα ή αλλιώς Τ-κύτταρα (ένα άλλο είδος λευκών αιμοσφαιρίων) για να επιτεθεί απευθείας στον ιό. Ένα υποσύνολο των Τ-λεμφοκυττάρων που ονομάζονται κύτταρα μνήμης, αναγνωρίζουν το αντιγόνο ενός πρόσφατα εξουδετερωμένου παθογόνου και με αυτό τον τρόπο μπορούν γρήγορα να ενεργοποιήσουν περισσότερα αποτελεσματικά Τ-λεμφοκύτταρα σε περίπτωση μεταγενέστερης έκθεσης στον ιό. Ο δεύτερος αυτός μηχανισμός ονομάζεται κυτταρική ανοσία (cellular immunity).

Οι πιο πάνω μηχανισμοί ανοσίας αποτελούν τη βάση για το σχεδιασμό εμβολίων. Σημαντικότατο ρόλο στη διαδικασία της ανοσολογικής απόκρισης διαδραματίζουν συγκεκριμένες βοηθητικές πρωτεΐνες που ονομάζονται πρωτεΐνες μείζονος συμπλέγματος ιστοσυμβατότητας (Major Histocompatibility Complex - MHC), οι οποίες «παρουσιάζουν» (εμφανίζουν στην επιφάνεια των κυττάρων) περιοχές πρόσδεσης των αντιγόνων, που ονομάζονται επίτοποι (epitopes), έτσι ώστε τα αντισώματα, τα Β-λεμφοκύτταρα ή τα Τ-λεμφοκύτταρα να μπορούν να τα δεσμεύσουν και να τα εξουδετερώσουν. Οι πρωτεΐνες αυτές χωρίζονται σε δύο ομάδες, MHC-I και MHC-II, και κωδικοποιούνται από τα συγκροτήματα γονιδίων του ανθρώπινου λευκοκυτταρικού αντιγόνου (Human Leukocyte Antigen - HLA), τα οποία ποικίλλουν από άτομο σε άτομο. Συνεπώς, για το σχεδιασμό αποτελεσματικών εμβολίων θα πρέπει να ληφθούν υπόψη οι εξής στόχοι: (α) να προσδιοριστούν κατάλληλοι επίτοποι για στόχευση και (β) να επιβεβαιωθεί ότι αυτοί οι επίτοποι μπορούν να «παρουσιαστούν» από τις πρωτεΐνες MHC που αντιστοιχούν σε διάφορους τύπους (alleles) HLA που εμφανίζονται στον πληθυσμό [134].

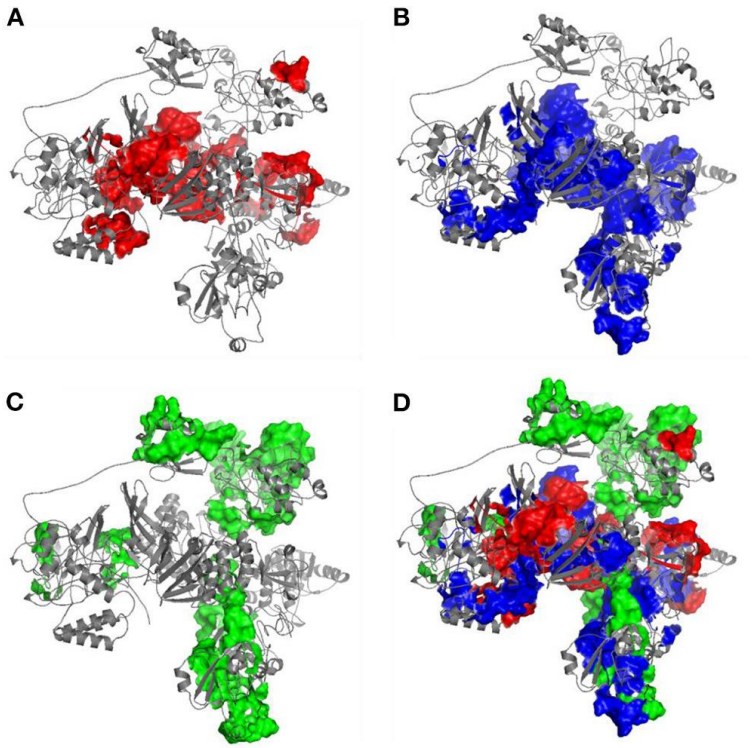
Σε μια ενδιαφέρουσα μελέτη ερευνητών από το Πανεπιστήμιο του Στάνφορντ, χρησιμοποιήθηκαν τεχνικές μηχανικής μάθησης μαζί με άλλες μεθόδους από τη δομική βιολογία για τον εντοπισμό επιτόπων του ιού SARS-CoV-2 για τα Τ-λεμφοκύτταρα και Β-λεμφοκύτταρα του ανθρώπινου οργανισμού [135]. Αρχικά, δύο αλγόριθμοι νευρωνικών δικτύων ανέλυσαν το γονιδίωμα του ιού SARS-CoV-2 για εύρεση υποψήφιων επιτόπων. Με αυτό τον τρόπο, προέβλεψαν 405 πιθανούς επιτόπους Τ-κυττάρων, με υψηλές βαθμολογίες παρουσίασης MHC-I και MHC-II, καθώς επίσης και δύο πιθανούς επιτόπους Β-κυττάρων στην πρωτεΐνη S. Αφού βρήκαν τους πιθανούς επιτόπους, οι ερευνητές εξέτασαν 68 διαφορετικές γενετικές παραλλαγές / μεταλλάξεις του ιού SARS-CoV-2 για να κατανοήσουν ποια τμήματα του ιού επηρεάζονται περισσότερο ή λιγότερο από αυτές τις μεταλλάξεις. Η ανάλυση αυτή έδειξε ότι δεν υπήρχαν μεταλλάξεις κοντά στην περιοχή πρόσδεσης του υποδοχέα (receptor binding domain) της πρωτεΐνης ακίδας S. Τα πιο πάνω ευρήματα συγκλίνουν στο ότι η πρωτεΐνη S φαίνεται να είναι ένας πολύ καλός στόχος για πιθανά εμβόλια.

Σε μια ακόμα προσπάθεια για το σχεδιασμό ενός καθολικά αποτελεσματικού εμβολίου, ερευνητές της εταιρείας NEC στην Ιαπωνία χρησιμοποίησαν μια πλατφόρμα βασισμένη σε τεχνολογίες τεχνητής νοημοσύνης [136]. Με τη χρήση της πλατφόρμας αυτής, κατάφεραν να αναλύσουν το σύνολο του πρωτεώματος[7] του ιού SARS-CoV-2 σε σχέση με τους 100 πιο κοινά εμφανιζόμενους τύπους HLA σε

[7] Το πρωτέωμα (proteome) είναι το σύνολο των πρωτεϊνών που εκφράζεται ή δύναται να εκφραστεί από ένα γονιδίωμα, κύτταρο, ιστό ή οργανισμό σε μια συγκεκριμένη στιγμή και υπό συγκεκριμένες συνθήκες.

ολόκληρο τον ανθρώπινο πληθυσμό και με αυτόν τον τρόπο, κατασκεύασαν ολοκληρωμένους χάρτες επιτόπων. Κατόπιν, οι ερευνητές ανέλυσαν περαιτέρω αυτούς τους χάρτες και προσδιόρισαν τις περιοχές ενδιαφέροντος επιτόπων (epitope hotspots) του ιού SARS-CoV-2 που είναι πιθανότερο να δημιουργήσουν ανοσολογική απόκριση σε ένα ευρύ φάσμα τύπων HLA και έτσι, να οδηγήσουν σε ισχυρές αποκρίσεις των Τ-κυττάρων για την πλειοψηφία του παγκόσμιου πληθυσμού.

Ένα ακόμα ενδιαφέρον παράδειγμα εφαρμογής στον τομέα της ανακάλυψης εμβολίων προέρχεται από μια ομάδα ερευνητών στο Πανεπιστήμιο του Michigan [137]. Η εφαρμογή αυτή, με την ονομασία Vaxign-ML, αξιοποιεί μοντέλα επιτηρούμενης μάθησης για σκοπούς ταξινόμησης, προσπαθώντας να προβλέψει ποιες από τις πρωτεΐνες του ιού μπορούν να αποτελέσουν αποτελεσματικούς στόχους εμβολίων. Αν και οι ερευνητές διαπίστωσαν ότι η πρωτεΐνη S είναι ο καλύτερος υποψήφιος στόχος για εμβόλια, προσδιόρισαν ακόμα πέντε μη δομικές πρωτεΐνες, εκ των οποίων οι nsp3 και nsp8 φαίνεται να είναι επίσης πολύ καλοί υποψήφιοι-στόχοι του εμβολίου. Σε αντίθεση με την πρωτεΐνη S, η πρωτεΐνη nsp3 δεν έχει δοκιμαστεί ακόμη σε κάποια μελέτη εμβολίων κατά της ασθένειας COVID-19 και επιλέχθηκε από τους ερευνητές για περαιτέρω μελέτη. Επίσης, η nsp3 βρέθηκε να διατηρείται περισσότερο σταθερή μεταξύ των τριών κορωνοϊών SARS-CoV-2, SARS-CoV και MERS-CoV σε σύγκριση με άλλους γνωστούς κορωνοϊούς που μολύνουν ανθρώπους και ζώα.



Σχήμα 37 *- Τρισδιάστατη δομή της πρωτεΐνης nsp3 του SARS-CoV-2 με επισήμανση των διαφορετικών επιτόπων - πιθανών στόχων εμβολίων* [137]

Στο Σχήμα 37 βλέπουμε την τρισδιάστατη δομή της πρωτεΐνης nsp3, όπου έχουν σημειωθεί: (A) οι επίτοποι MHC-I των Τ-κυττάρων (με κόκκινο χρώμα), (B) οι επίτοποι MHC-II των Τ-κυττάρων (με μπλε χρώμα), (C) οι επίτοποι των Β-κυττάρων (με πράσινο χρώμα) και (D) οι συνδυασμένοι επίτοποι. Όπως παρατηρούμε, οι επίτοποι των Β-κυττάρων εκτίθενται περισσότερο στην επιφάνεια της πρωτεΐνης, ενώ οι επίτοποι MHC-I και MHC-II των Τ-κυττάρων βρίσκονται περισσότερο εντός της πρωτεΐνης. Όπως αναφέρουν οι ερευνητές, οι προτεινόμενοι από το μοντέλο τους στόχοι παρέχουν τη δυνατότητα για ανάπτυξη αποτελεσματικών και ασφαλών εμβολίων κατά της ασθένειας COVID-19.

Προτείνουν, επίσης, ότι ένα εμβόλιο «κοκτέιλ» sp/nsp, το οποίο θα δρα έναντι μιας δομικής πρωτεΐνης (structural protein - sp) και μιας μη δομικής πρωτεΐνης (non-structural protein - nsp) του ιού παράγοντας συμπληρωματικές ανοσολογικές αντιδράσεις, ενδέχεται να είναι πολύ αποτελεσματικό.

4.4 Βελτίωση Κλινικών Δοκιμών

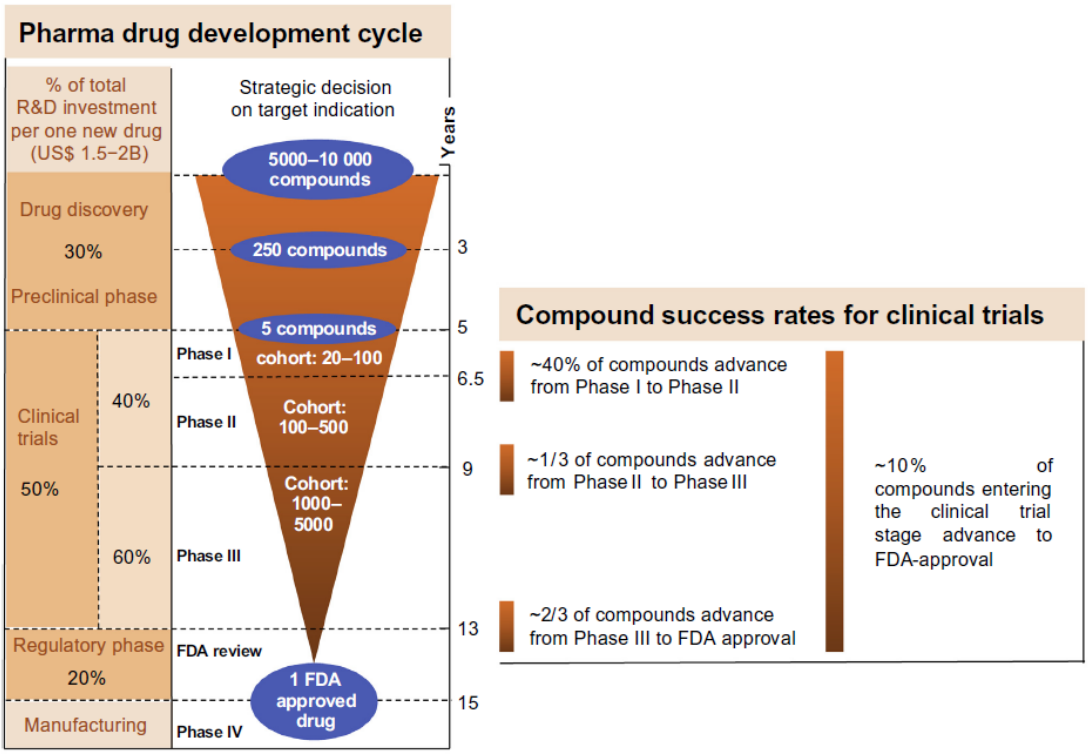
4.4.1 Η Τρέχουσα Κατάσταση στις Κλινικές Δοκιμές

Έχει παρατηρηθεί ότι στις διάφορες φάσεις των κλινικών δοκιμών καταναλώνεται περισσότερο από το μισό του συνολικού χρόνου και κόστους που σχετίζεται με τον κύκλο ανάπτυξης ενός νέου φαρμάκου μέχρι την τελική διάθεσή του στην αγορά. Το υπόλοιπο 50% των δαπανών έρευνας και ανάπτυξης καλύπτει τα προ-κλινικά στάδια της ανακάλυψης των συστατικών του φαρμάκου και των δοκιμών στο εργαστήριο (in-vitro), καθώς επίσης και διάφορες ρυθμιστικές διαδικασίες. Όπως είναι κατανοητό, μια αποτυχημένη κλινική δοκιμή επιφέρει όχι μόνο τη ζημία από τη χαμένη επένδυση στη δοκιμή αυτή καθ΄ εαυτή, αλλά και τη ζημία που αντιστοιχεί στα διάφορα στάδια της προ-κλινικής ανάπτυξης. Επίσης, παρά το γεγονός ότι τις τελευταίες δεκαετίες οι φαρμακευτικές εταιρείες και οι εταιρείες βιοτεχνολογίας αυξάνουν συνεχώς τις επενδύσεις τους για έρευνα και ανάπτυξη, έχει παρατηρηθεί ότι ο αριθμός των εγκεκριμένων νέων φαρμάκων ανά δισεκατομμύριο δολάρια που δαπανάται μειώνεται στο μισό περίπου κάθε 9 χρόνια [138].

Όπως αναφέρθηκε, ένα από τα βασικότερα εμπόδια στην αλυσίδα ανάπτυξης φαρμάκων είναι το υψηλό ποσοστό αποτυχίας των κλινικών δοκιμών. Έχει υπολογιστεί ότι το συντριπτικό ποσοστό των υποψήφιων φαρμάκων που εισέρχονται στις φάσεις των κλινικών δοκιμών (περίπου το 90%) δεν καταφέρνουν να φτάσουν μέχρι την τελική έγκριση. Συγκεκριμένα, όπως φαίνεται και στο Σχήμα 38, έχει υπολογιστεί ότι περίπου το 40% των υποψήφιων ουσιών της Φάσης I καταφέρνουν να περάσουν στη Φάση II. Επίσης, λιγότερο από το ένα τρίτο όλων των υποψήφιων φαρμάκων στη Φάση II συνεχίζουν στη Φάση III. Τέλος, περισσότερο από το ένα τρίτο όλων των υποψηφίων φαρμάκων που κατάφεραν να φτάσουν στη Φάση III δεν καταφέρνουν να λάβουν την τελική έγκριση από τις κανονιστικές αρχές [139].

Σημαντικοί παράγοντες για το υψηλό αυτό ποσοστό αποτυχίας είναι η μη βέλτιστη επιλογή κατάλληλων ομάδων εθελοντών και οι προβληματικές διαδικασίες πρόσληψής τους, οι οποίες αποτυγχάνουν να φέρουν τους καταλληλότερους ασθενείς σε μια δοκιμή εγκαίρως. Επιπλέον, η απουσία αξιόπιστου και αποτελεσματικού ελέγχου συμμόρφωσης ως προς τα πρωτόκολλα της δοκιμής, η ελλιπής – πολλές φορές – παρακολούθηση των εθελοντών κατά τη διάρκεια της θεραπείας τους, καθώς και η απουσία συστημάτων ανίχνευσης του κλινικού καταληκτικού σημείου (clinical endpoint)[8] συμβάλλουν με τη σειρά τους στο μεγάλο ποσοστό των αποτυχημένων κλινικών δοκιμών [140].

[8] Πολλές φορές μια κλινική δοκιμή επιχειρεί να αξιολογήσει τον πιθανό αντίκτυπο μιας ιατρικής παρέμβασης στην εμφάνιση μιας ασθένειας, όπως για παράδειγμα εκδηλώνεται μέσω συγκεκριμένου συμπτώματος. Η εμφάνιση ενός τέτοιου συμπτώματος σε έναν ασθενή κατά τη διάρκεια της δοκιμής σηματοδοτεί το κλινικό καταληκτικό σημείο (clinical endpoint) για τον εν λόγω ασθενή.



Σχήμα 38 - *Τυπικός κύκλος ανάπτυξης φαρμάκων στη Φαρμακοβιομηχανία* [139]

4.4.2 Πώς η Τεχνητή Νοημοσύνη Μπορεί να Μετασχηματίσει τις Κλινικές Δοκιμές

Για τη βελτίωση των κλινικών δοκιμών, πολλοί ερευνητές από την ακαδημαϊκή κοινότητα αλλά και τη φαρμακευτική βιομηχανία έχουν στραφεί στην τεχνητή νοημοσύνη, πιστεύοντας ότι με αυτό τον τρόπο μπορούν να ξεπεράσουν πολλά από τα προβλήματα που προαναφέρθηκαν. Οι δυνατότητες που παρέχει η τεχνητή νοημοσύνη μπορούν να χρησιμοποιηθούν για τη συσχέτιση μεγάλων και πολύπλοκων συνόλων δεδομένων, όπως για παράδειγμα ηλεκτρονικών φακέλων ασθενών, πληροφοριών από την ιατρική βιβλιογραφία, καθώς και βάσεων δεδομένων από παλαιότερες κλινικές δοκιμές για την εξαγωγή χρήσιμων συμπερασμάτων. Τα συμπεράσματα από αυτή την ανάλυση μπορούν να οδηγήσουν σε βελτιωμένη αντιστοίχιση μεταξύ εθελοντών και δοκιμών, έγκαιρη πρόσληψη των ατόμων πριν την έναρξη της δοκιμής, καθώς και αυτόματη και συνεχή παρακολούθησή τους κατά τη διάρκεια συμμετοχής τους στην κλινική δοκιμή. Στην ενότητα αυτή αναλύουμε περαιτέρω τα πιο πάνω οφέλη.

Επιλογή και πρόσληψη εθελοντών

Κάθε κλινική δοκιμή θέτει ειδικές απαιτήσεις στους συμμετέχοντες εθελοντές-ασθενείς όσον αφορά την επιλεξιμότητα και την καταλληλόλητά τους, και παρέχει συνήθως κίνητρα για τη συμμετοχή τους. Το ιατρικό ιστορικό ενός συγκεκριμένου ασθενούς μπορεί για παράδειγμα να τον καταστήσει μη επιλέξιμο. Επίσης, ακόμα και ένας αρχικά επιλέξιμος ασθενής μπορεί να μη βρίσκεται σε ένα συγκεκριμένο στάδιο της νόσου ή να μην ανήκει σε ένα συγκεκριμένο υπο-φαινότυπο στον οποίο στοχεύει το φάρμακο που πρόκειται να ελεγχθεί, καθιστώντας τον έτσι ακατάλληλο για τη συγκεκριμένη κλινική δοκιμή. Επιπρόσθετα, μπορεί να μη δίνονται τα κατάλληλα κίνητρα συμμετοχής στους επιλέξιμους και κατάλληλους ασθενείς. Όμως ακόμα και αν δίνονται, μπορεί τα συγκεκριμένα άτομα να μη γνωρίζουν ότι γίνεται η αντίστοιχη δοκιμή ή να θεωρούν τη διαδικασία συμμετοχής πολύ περίπλοκη και δύσκολη.

Όλα τα πιο πάνω εμπόδια οδηγούν σε μεγάλες καθυστερήσεις στις κλινικές δοκιμές και θέτουν σημαντικές προκλήσεις στην τήρηση των στενών χρονοδιαγραμμάτων που τις συνοδεύουν. Εργαλεία και μέθοδοι τεχνητής νοημοσύνης χρησιμοποιούνται ήδη για τον εντοπισμό ασθενών για κλινικές δοκιμές, συμπεριλαμβανομένων προηγμένων νευρωνικών δικτύων και άλλων συστημάτων μηχανικής μάθησης για την εξόρυξη μεγάλων συνόλων δεδομένων από ασθενείς.

Για παράδειγμα, οι φαρμακευτικές εταιρείες Novartis, Pfizer και Eli Lilly εκμεταλλεύθηκαν πρόσφατα μια τεχνολογική πλατφόρμα από την εταιρεία Antidote στο Ηνωμένο Βασίλειο, για να τις βοηθήσει στη βελτιστοποίηση των διαδικασιών των κλινικών δοκιμών, καθώς και στον εντοπισμό των κατάλληλων ασθενών. Η τεχνολογία της Antidote, που βασίζεται σε διάφορα μοντέλα μηχανικής μάθησης, χρησιμοποιείται για να αναλύσει δεδομένα από παλαιότερες κλινικές δοκιμές και να εξάγει πληροφορίες (που τελικά διαθέτει στις φαρμακευτικές εταιρείες) για τους συμμετέχοντες εθελοντές, συμπεριλαμβανομένων των σχετικών κριτηρίων επιλεξιμότητας. Συγκεκριμένα, η Antidote αναλύει δεδομένα που σχετίζονται με περίπου 14.000 κλινικές δοκιμές για περισσότερες από 700 ασθένειες και άλλες παθήσεις [141].

Μια ακόμα ενδιαφέρουσα εφαρμογή έχει αναπτυχθεί από την εταιρεία τεχνητής νοημοσύνης Deep 6 AI, η οποία χρησιμοποιεί μεθόδους επεξεργασίας φυσικής γλώσσας για να αναλύσει σημειώσεις ιατρών, αναφορές, διαγνώσεις και προτεινόμενες θεραπείες, καθώς και να ανιχνεύσει δυσεύρετα δεδομένα σχετικά με τον τρόπο ζωής, όπως το κάπνισμα και το ιστορικό σωματικής άσκησης, έτσι ώστε να αντιστοιχίσει τελικά ασθενείς σε κλινικές δοκιμές με βάση κατάλληλα κριτήρια [142].

Η Mendel.ai είναι μια ακόμα εταιρεία που αναπτύσσει λύσεις που βασίζονται στην τεχνητή νοημοσύνη για τη βελτίωση των κλινικών δοκιμών. Η πλατφόρμα της Mendel.ai χρησιμοποιεί έναν αλγόριθμο επεξεργασίας φυσικής γλώσσας για να ερμηνεύσει αδόμητα δεδομένα που προέρχονται από τον ιστότοπο clinicaltrials.gov και να αντιστοιχίσει συγκεκριμένους ασθενείς με τις αντίστοιχες κλινικές δοκιμές, με απώτερο στόχο τη μείωση του χρόνου και του κόστους των δοκιμών αυτών [143].

Παρακολούθηση ασθενών κατά τη διάρκεια των κλινικών δοκιμών

Η πρόσληψη των κατάλληλων ασθενών σε μια κλινική δοκιμή είναι μια σημαντική επένδυση τόσο σε χρόνο όσο και σε χρήμα. Η επένδυση αυτή μπορεί να αποδώσει μόνο με την επιτυχή ολοκλήρωση της δοκιμής. Συνεπώς, είναι επιτακτική ανάγκη οι ασθενείς να παραμείνουν στη δοκιμή και να τηρούν τις διαδικασίες και τους κανόνες της δοκιμής καθ' όλη τη διάρκειά της. Επίσης, είναι εξίσου σημαντικό όλα τα δεδομένα σε σχέση με την παρακολούθηση των επιπτώσεων ενός δοκιμαστικού φαρμάκου να συλλέγονται αξιόπιστα και με αποδοτικό τρόπο. Έρευνες έχουν δείξει ότι μόνο το 15% των κλινικών δοκιμών δεν παρουσιάζουν εγκαταλείψεις ασθενών, ενώ το μέσο ποσοστό αποχώρησης ασθενών σε όλες τις κλινικές δοκιμές είναι περίπου 30% [140]. Οι αποχωρήσεις που προκαλούνται από την έλλειψη τήρησης των πρωτοκόλλων της δοκιμής συνεπάγονται επιπρόσθετες προσπάθειες πρόσληψης εθελοντών, οι οποίες με τη σειρά τους οδηγούν σε επιπρόσθετες καθυστερήσεις και κόστη.

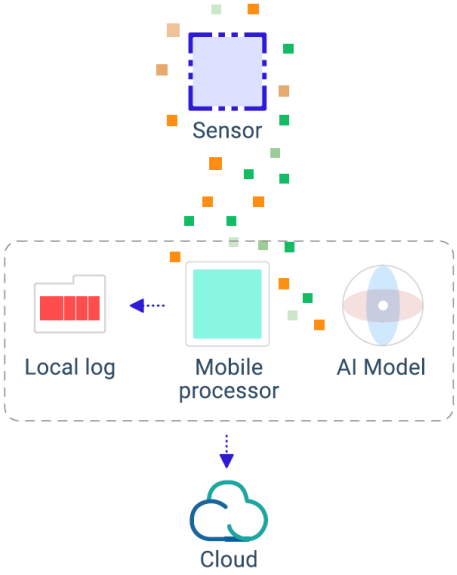
Για να ανταποκριθούν στα κριτήρια συμμόρφωσης που θέτουν οι κλινικές δοκιμές, οι ασθενείς υποχρεούνται να τηρούν λεπτομερή αρχεία σχετικά με την πρόσληψη των φαρμάκων τους, καθώς και για μια πληθώρα άλλων πληροφοριών που σχετίζονται με τις σωματικές λειτουργίες τους, την ανταπόκρισή τους στη φαρμακευτική αγωγή, κλπ. Η συλλογή και συστηματική καταγραφή όλων αυτών των στοιχείων είναι μια αρκετά χρονοβόρα διαδικασία, η οποία μπορεί να οδηγήσει πολλούς ασθενείς στο να μη συμμορφωθούν και να χρειαστεί να εγκαταλείψουν τη δοκιμή. Η σύγχρονη

ψηφιακή τεχνολογία μπορεί να διευκολύνει τους συμμετέχοντες στη συλλογή των δεδομένων και να καταστήσει εφικτές προηγμένες μεθόδους ανάλυσης και εξαγωγής συμπερασμάτων.

Για παράδειγμα, αισθητήρες από wearable συσκευές σε συνδυασμό με τεχνολογίες απομακρυσμένης παρακολούθησης μπορούν να χρησιμοποιηθούν για να συλλέξουν αυτόματα, και σχεδόν σε πραγματικό χρόνο, μετρήσεις διαφόρων λειτουργιών των ασθενών, χωρίς να χρειάζεται ο ασθενής να τις καταγράψει χειροκίνητα. Μοντέλα μηχανικής και κυρίως βαθιάς μάθησης μπορούν, στη συνέχεια, να χρησιμοποιηθούν για την ανάλυση αυτών των δεδομένων σε πραγματικό χρόνο, με στόχο τον εντοπισμό και την καταγραφή συμβάντων που χρήζουν προσοχής. Η προσέγγιση αυτή επιτρέπει τη δημιουργία «ημερολογίων ασθενειών» (disease diaries), τα οποία εξελίσσονται και προσαρμόζονται στον κάθε ασθενή, καθώς τα αντίστοιχα μοντέλα βαθιάς μάθησης επανεκπαιδεύονται περιοδικά με νέα δεδομένα μετρήσεων. Τέτοια ημερολόγια ασθενειών μπορούν να χρησιμεύσουν, επίσης, ως αποδεικτικά στοιχεία για την τήρηση ή τη μη τήρηση των κριτηρίων συμμόρφωσης με τη δοκιμή, καθώς και για να ανιχνεύσουν έγκαιρα μετρήσεις που πιθανώς να σχετίζονται με κάποιο κλινικό καταληκτικό σημείο.

Επίσης, μέθοδοι πρόβλεψης που βασίζονται στη μηχανική μάθηση μπορούν να χρησιμοποιηθούν για να προβλέψουν δυναμικά τον κίνδυνο εγκατάλειψης για ένα συγκεκριμένο ασθενή, μέσω της ανίχνευσης κάποιας συμπεριφοράς που υποδηλώνει ότι ο ασθενής μπορεί να αντιμετωπίζει προβλήματα με την τήρηση του πρωτοκόλλου της κλινικής δοκιμής. Σε αυτές τις περιπτώσεις, μπορεί να γίνει επικοινωνία με τον ασθενή για καλύτερη κατανόηση των προβλημάτων που αντιμετωπίζει και παροχή συμβουλών, που θα τον βοηθήσουν να παραμείνει στη δοκιμή.

Όπως φαίνεται στο Σχήμα 39, τα δεδομένα μετρήσεων συλλέγονται από έναν ή περισσότερους αισθητήρες σε wearable συσκευές και αναλύονται σε πραγματικό χρόνο στο σημείο συλλογής από έναν επεξεργαστή κινητής συσκευής, ο οποίος χρησιμοποιεί για την ανάλυση ένα μοντέλο τεχνητής νοημοσύνης. Στη συνέχεια, τα αποτελέσματα της ανάλυσης αποθηκεύονται σε ένα τοπικό αρχείο στην κινητή συσκευή και μεταδίδονται ταυτόχρονα στο cloud για σκοπούς απομακρυσμένης αποθήκευσης και περαιτέρω (κεντρικής) ανάλυσης.



Σχήμα 39 *- Παρακολούθηση ασθενών που συμμετέχουν σε κλινικές δοκιμές με τη χρήση wearable αισθητήρων και τεχνητής νοημοσύνης*

Ανάλυση των τελικών αποτελεσμάτων των δοκιμών

Μια άλλη σημαντική εφαρμογή της τεχνητής νοημοσύνης στον τομέα των κλινικών δοκιμών είναι η ανάλυση των τελικών αποτελεσμάτων των κλινικών δοκιμών και, ειδικότερα, ο προσδιορισμός των υποπληθυσμών που διαθέτουν παρόμοια χαρακτηριστικά και έχουν αντίστοιχη ανταπόκριση στη δοκιμαστική θεραπεία. Ο προσδιορισμός τέτοιων υποπληθυσμών είναι πολύ χρήσιμος, καθώς επιτρέπει την εξαγωγή συμπερασμάτων σε σχέση με τα διάφορα επίπεδα απόκρισης της υπό εξέταση θεραπείας σε διαφορετικές πληθυσμιακές ομάδες αλλά και σε σχέση με τους παράγοντες από τους οποίους αυτή φαίνεται να εξαρτάται. Επιπλέον, η ποσοτικοποίηση της αβεβαιότητας που συνδέεται με τη θεραπεία συγκεκριμένων υποπληθυσμών διευκολύνει στη στόχευση θεραπειών σε συγκεκριμένους υποπληθυσμούς, όπου υπάρχει μεγαλύτερος βαθμός βεβαιότητας για θετική απόκριση, καθώς και στην εύρεση θεραπειών που είναι κατάλληλες για το σύνολο του πληθυσμού.

4.4.3 Μετασχηματισμός Κλινικών Δοκιμών στην Εποχή της Πανδημίας COVID-19

Η πανδημία COVID-19 θέτει πολλές προκλήσεις στη διεξαγωγή κλινικών δοκιμών, όπως για παράδειγμα η μείωση της προθυμίας των εθελοντών ή του αντίστοιχου επιστημονικού προσωπικού να προσέλθουν στις κλινικές εγκαταστάσεις. Κάτι τέτοιο μπορεί να οδηγήσει στη διακοπή συλλογής δεδομένων μετρήσεων ή στην ανάγκη για απομακρυσμένη συλλογή δεδομένων, προκαλώντας καθυστερήσεις ή ακόμα και διακοπή των κλινικών δοκιμών. Η συνήθης προσέγγιση των Τυχαιοποιημένων Ελεγχόμενων Δοκιμών (Randomized Controlled Trials) περιλαμβάνει συνήθως χρονοβόρες και δύσκαμπτες διαδικασίες, που σε πολλές περιπτώσεις δε μπορούν να ανταποκριθούν στις απαιτήσεις για άμεσες και αποτελεσματικές λύσεις σε μια πανδημία, όπως αυτή που διανύουμε.

Σε ένα πρόσφατο επιστημονικό άρθρο [144], μια διεθνής ομάδα που αποτελείται από επιστήμονες δεδομένων (data scientists) και εμπειρογνώμονες από τη φαρμακευτική βιομηχανία περιγράφει τις προκλήσεις που θέτει η πανδημία COVID-19 στις κλινικές δοκιμές και εξηγεί πώς οι πρόσφατες εξελίξεις στον τομέα της μηχανικής μάθησης μπορούν να βοηθήσουν να ξεπεραστούν οι προκλήσεις αυτές. Το άρθρο καλύπτει τρεις τομείς κλινικών δοκιμών, στους οποίους η μηχανική μάθηση μπορεί να συμβάλει ενεργά: (α) σε δοκιμές για επαναστόχευση φαρμάκων για τη θεραπεία της νόσου COVID-19, (β) σε δοκιμές για νέα φάρμακα για τη θεραπεία της νόσου COVID-19 και (γ) σε εν εξελίξει κλινικές δοκιμές για φάρμακα που δε σχετίζονται με την ασθένεια COVID-19. Οι ερευνητές επισημαίνουν τρόπους με τους οποίους η χρήση μηχανικής μάθησης μπορεί να βοηθήσει τόσο στην αντιμετώπιση των προκλήσεων της παρούσας πανδημίας όσο και στον ευρύτερο μετασχηματισμό των κλινικών δοκιμών σε ένα πιο προηγμένο επίπεδο, καθιστώντας τις πιο αποτελεσματικές, ανθεκτικές και ευέλικτες. Συνοψίζουν τη δυνητική συμβολή της μηχανικής μάθησης στις ακόλουθες περιοχές:

- Υποστήριξη στη δημιουργία «εικονικών» ομάδων ελέγχου: Η ταχεία εξάπλωση της πανδημίας COVID-19 και η έλλειψη γνώσεων σε σχέση με αποτελεσματικά φάρμακα και θεραπείες έχουν οδηγήσει τα νοσοκομεία σε όλο τον κόσμο να πειραματιστούν με φάρμακα που δεν έχουν υποβληθεί σε κατάλληλες κλινικές δοκιμές και δεν προβλέπεται να υποβληθούν σε τέτοιες δοκιμές στο άμεσο μέλλον. Με τη συλλογή δεδομένων από νοσοκομεία, μέθοδοι ανάλυσης μεγάλων δεδομένων μπορούν να συγκρίνουν την εξέλιξη ασθενών που έλαβαν τυποποιημένες θεραπείες, σε σχέση με παρόμοιους (με βάση διάφορα χαρακτηριστικά) ασθενείς που έλαβαν συγκεκριμένες νέες πειραματικές θεραπείες. Με αυτό τον τρόπο, είναι εφικτός ο υπολογισμός της αποτελεσματικότητας των πειραματικών θεραπειών.

- Εξαγωγή γνώσεων από δεδομένα κλινικών δοκιμών που ανεστάλησαν λόγω της πανδημίας COVID-19 και αναπροσαρμογή του σχεδιασμού τους (συμπεριλαμβανομένων των πλάνων πρόσληψης εθελοντών, τα μεγέθη των δειγμάτων κλπ.).
- Βελτίωση του σχεδιασμού, της εκτέλεσης και της αξιολόγησης μεγάλων προσαρμοστικών (adaptive) κλινικών δοκιμών για την αξιολόγηση υποψήφιων επαναστοχευμένων φαρμάκων για την ασθένεια COVID-19.
- Κατάτμηση των μεγάλης διάρκειας και πολλαπλών φάσεων τυχαιοποιημένων ελεγχόμενων δοκιμών (RCTs) σε μια πιο αποδοτική, συνεχή, προσαρμοστική και επαναληπτική διαδικασία δοκιμής-συλλογής-επαναδοκιμής. Σε αυτή την περιοχή, η χρήση μεθόδων μηχανικής μάθησης μπορεί να βοηθήσει στην ταυτόχρονη εξαγωγή συμπερασμάτων σε σχέση με την τοξικότητα και την αποτελεσματικότητα ενός νέου φαρμάκου. Με αυτό τον τρόπο, μειώνεται σημαντικά ο συνολικός απαιτούμενος χρόνος των κλινικών δοκιμών, κάτι που είναι πολύ θετικό για την έγκαιρη ανακάλυψη θεραπειών ενάντια στην ασθένεια COVID-19.

ΚΕΦΑΛΑΙΟ 5: Εφαρμογές στην Καταπολέμηση της Επιδημίας της Πληροφορίας

5.1 Η Επιδημία της Πληροφορίας (Infodemic)

Η επιδημία της πληροφορίας (infodemic) ορίζεται ως η γρήγορη και εκτεταμένη εξάπλωση πληροφοριών – πολλές εκ των οποίων είναι ανακριβείς ή ψευδείς – για κάποιο θέμα, όπως για παράδειγμα μια ασθένεια. Καθώς τα πραγματικά γεγονότα, οι φήμες, οι φόβοι και οι ψευδείς πληροφορίες αναμειγνύονται και αναμεταδίδονται, καθίσταται πολύ δύσκολο για τους πολίτες να εντοπίσουν αξιόπιστες πηγές και να αποκτήσουν χρήσιμη και αντικειμενική γνώση για το εν λόγω θέμα [145]. Ο όρος infodemic είναι ουσιαστικά ένα λογοπαίγνιο μεταξύ των αγγλικών λέξεων information (πληροφορία) και epidemic (επιδημία), και αποδίδεται στον πολιτικό επιστήμονα David Rothkopf, ο οποίος τον επινόησε σε άρθρο του το 2003 για να περιγράψει τη μαζική παραπληροφόρηση γύρω από την επιδημία του ιού SARS-CoV και τις συνέπειες αυτής στην παγκόσμια οικονομία, πολιτική και ασφάλεια [146].

Η δημιουργία του όρου infodemic προέκυψε με βάση τεκμηριωμένες παρατηρήσεις επιστημόνων ότι υπήρχαν έντονες ομοιότητες μεταξύ του τρόπου με τον οποίο εξαπλώνεται μια μεταδοτική ασθένεια στον πληθυσμό και του μηχανισμού με τον οποίο μια ιδέα ή μια πληροφορία γενικότερα γινόταν δημοφιλής στο διαδίκτυο. Πιο συγκεκριμένα, ομάδα ερευνητών απέδειξε πως η διάδοση κοινωνικών αντιλήψεων, ιδεών ή συμπεριφορών ακολουθεί τα ίδια μοντέλα εξάπλωσης όπως ένας μεταδοτικός ιός [147].

Στην περίπτωση της πανδημίας COVID-19, η επιδημία της πληροφορίας περιλαμβάνει τόσο τη μη ηθελημένη παραπληροφόρηση (misinformation) όσο και εσκεμμένες προσπάθειες διάδοσης λανθασμένων πληροφοριών (disinformation) και ψευδών ειδήσεων (fake news), που θα μπορούσαν να υπονομεύσουν τις προσπάθειες για την προστασία της δημόσιας υγείας. Όσον αφορά στην ηθελημένη διάδοση λανθασμένων και ψευδών πληροφοριών (disinformation), συνήθως αποτελεί προϊόν προπαγάνδας και αποσκοπεί στην προώθηση συγκεκριμένων πολιτικών συμφερόντων ή και στην απόκτηση οικονομικού οφέλους από συγκεκριμένα άτομα ή ομάδες ατόμων.

Προφανώς, η παραπληροφόρηση, οι λανθασμένες και οι ψευδείς ειδήσεις δεν είναι κάτι που σχετίζεται μόνο με την πανδημία COVID-19, αλλά σε μια τέτοια έκτακτη κατάσταση που απαιτείται συντονισμένη προσπάθεια κρατών και πολιτών, μπορεί να αποβούν πάρα πολύ επικίνδυνες [148]. Όπως αναφέρουν διάφοροι αναλυτές, η επιδημία της πληροφορίας ενδέχεται να επιδεινώσει τις συνέπειες αυτής της υγειονομικής κρίσης, τόσο στον τομέα της δημόσιας υγείας όσο και στους τομείς της διεθνούς οικονομίας, της κοινωνικής συνοχής των κρατών και της παγκόσμιας γεωπολιτικής σταθερότητας [149]. Για παράδειγμα, η μετάδοση ψευδών ή παραπλανητικών πληροφοριών μπορεί να εμποδίσει την έγκαιρη και αποτελεσματική υιοθέτηση κατάλληλων συμπεριφορών και μέτρων προστασίας από τους πολίτες και έτσι να θέσει σε σημαντικό κίνδυνο τη δημόσια υγεία.

Με το ξέσπασμα μιας νέας πανδημίας, οι πολίτες καλούνται κατ' αρχάς να αντιμετωπίσουν μια νέα μολυσματική ασθένεια, η οποία αποτελεί πραγματική απειλή για την υγεία τους λόγω έλλειψης αποτελεσματικών θεραπειών και εμβολίων. Σχεδόν ταυτόχρονα, καλούνται επίσης να αντιμετωπίσουν την εξίσου επικίνδυνη εξάπλωση της επιδημίας της πληροφορίας, η οποία μάλιστα μεγεθύνεται όταν αξιόπιστες πηγές πληροφοριών αποτυγχάνουν να κερδίσουν την προσοχή και την εμπιστοσύνη ορισμένων ανθρώπων. Για αυτούς τους ανθρώπους, εναλλακτικές πηγές πληροφορίας αμφιβόλου ποιότητας είναι πιο ελκυστικές, καθώς ταιριάζουν καλύτερα με τις πεποιθήσεις ή τις προκαταλήψεις τους, ή φαίνονται πιο πειστικές εξαιτίας των υπεραπλουστευμένων μηνυμάτων τους.

Το Μάιο του 2020, τα κράτη μέλη του Π.Ο.Υ., αντιλαμβανόμενα ότι η διαχείριση της επιδημίας της πληροφορίας αποτελεί κρίσιμη προϋπόθεση για τον ευρύτερο έλεγχο της πανδημίας COVID-19, συνυπέγραψαν ένα νέο ψήφισμα. Στο ψήφισμα αυτό, τα κράτη μέλη συμφώνησαν να παρέχουν αξιόπιστες πληροφορίες σχετικά με την πανδημία COVID-19 και να λάβουν τα απαραίτητα μέτρα για την καταπολέμηση της παραπληροφόρησης και της διάδοσης ψευδών ειδήσεων, αξιοποιώντας τις δυνατότητες της ψηφιακής τεχνολογίας. Πιο συγκεκριμένα, το ψήφισμα αυτό κάλεσε τους διεθνείς οργανισμούς να αντιμετωπίσουν τις λανθασμένες και ψευδείς πληροφορίες που διακινούνται στον ψηφιακό κόσμο, να εργαστούν για την καταπολέμηση δραστηριοτήτων στον κυβερνοχώρο που υπονομεύουν την αντιμετώπιση της υγειονομικής κρίσης, καθώς και να συμβάλουν στην παροχή επιστημονικών δεδομένων στο κοινό.

Σε αυτή την ενότητα, παρουσιάζουμε ορισμένες ερευνητικές προσπάθειες και εφαρμογές που αφορούν στην κατανόηση, την ποσοτικοποίηση και τη μοντελοποίηση της διάδοσης πληροφοριών που σχετίζονται με την πανδημία. Η αντιμετώπιση αυτού του τεράστιου όγκου πληροφοριών απαιτεί την ανάπτυξη και την υιοθέτηση προηγμένων υπολογιστικών εργαλείων, κυρίως σε ό,τι αφορά την ανίχνευση και τη μέτρηση της εξάπλωσης λανθασμένων και ηθελημένα ψευδών πληροφοριών. Αν και τα τελευταία χρόνια, η τεχνητή νοημοσύνη και ειδικότερα οι μέθοδοι επεξεργασίας φυσικής γλώσσας έχουν σημειώσει μεγάλη πρόοδο πάνω στον τομέα αυτό, εξακολουθεί να υπάρχει η ανάγκη για καλύτερη κατανόηση και συστηματικότερη αντιμετώπιση τέτοιων κοινωνικών φαινομένων, τα οποία μάλιστα φαίνεται να επηρεάζουν σημαντικά την εξάπλωση μιας πανδημίας, όπως της πανδημίας COVID-19.

Τέλος, αξίζει να σημειωθεί ότι τα μέσα κοινωνικής δικτύωσης και άλλες διαδικτυακές πλατφόρμες διαδραματίζουν βασικό ρόλο στην ανταλλαγή πληροφοριών σε σχέση με το νέο ιό και την ασθένεια που αυτός προκαλεί. Αν και διάφοροι τοπικοί και διεθνείς οργανισμοί καθώς και κυβερνήσεις έχουν χρησιμοποιήσει τέτοιες πλατφόρμες στο παρελθόν για να επικοινωνήσουν χρήσιμη πληροφορία στους πολίτες, παρατηρούμε ότι η παραπληροφόρηση και η διάδοση ψευδών ειδήσεων είναι εντονότερη σε αυτές τις πλατφόρμες σε σχέση με παραδοσιακά μέσα ενημέρωσης.

5.2 Εφαρμογές Τεχνητής Νοημοσύνης για την Αντιμετώπιση της Επιδημίας της Πληροφορίας

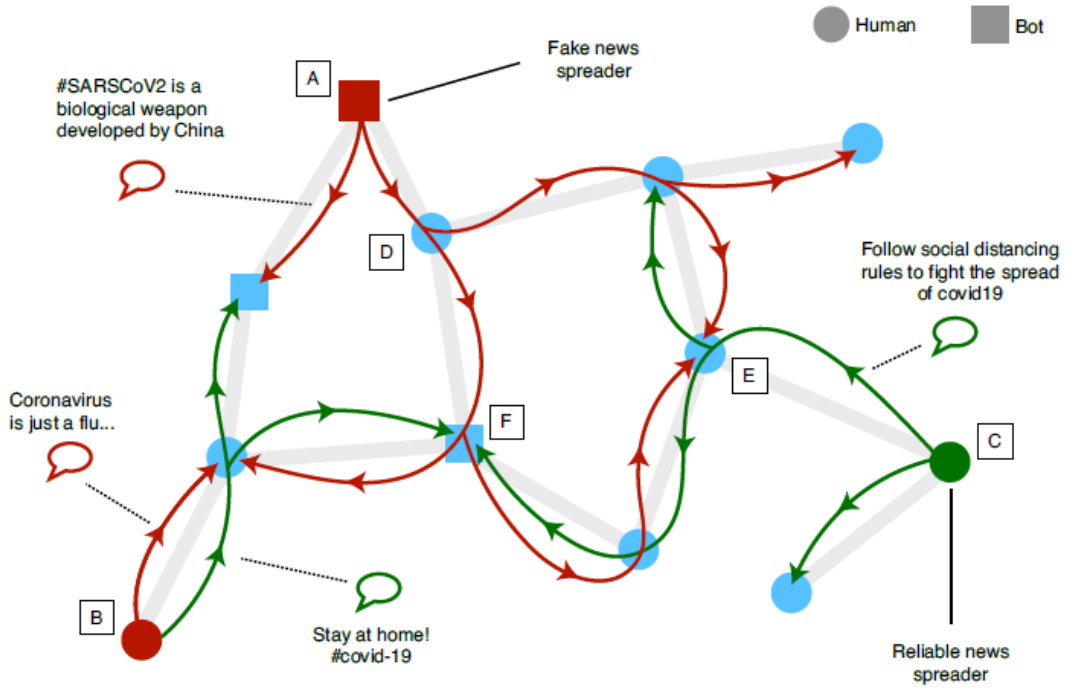
5.2.1 Υπολογισμός του Δείκτη Κινδύνου Επιδημίας της Πληροφορίας

Ο Δείκτης Κινδύνου Επιδημίας της Πληροφορίας (Infodemic Risk Index) μπορεί να οριστεί ως μια ποσότητα που είναι ανάλογη με το πλήθος των αναξιόπιστων πληροφοριών που διαδίδονται σε μια συγκεκριμένη χώρα ή τοποθεσία. Μια ενδιαφέρουσα μελέτη για την κατανόηση και τη μοντελοποίηση της δυναμικής της επιδημίας της πληροφορίας παρουσιάζεται στο [150]. Σε αυτή τη μελέτη, οι ερευνητές δανείστηκαν πολλές έννοιες από τη θεωρία της επιδημιολογίας και τις χρησιμοποίησαν σε συνδυασμό με τεχνικές επεξεργασίας φυσικής γλώσσας και εξόρυξης κειμένου (text mining) για την ανάλυση του περιεχομένου ενός δημοφιλούς μέσου κοινωνικής δικτύωσης, του Twitter. Χρησιμοποιώντας ως φίλτρο ένα σύνολο λέξεων που χρησιμοποιούνται συνήθως σε συζητήσεις ιατρικού περιεχομένου για την πανδημία COVID-19, οι ερευνητές συνέλεξαν περισσότερα από 112 εκατομμύρια μηνύματα σε 64 διαφορετικές γλώσσες και στη συνέχεια, τα ανέλυσαν.

Σύμφωνα με την έρευνά τους, και κατ' αντιστοιχία με τις επιδημίες ασθενειών, οι επιδημίες των πληροφοριών θα μπορούσαν να θεωρηθούν ως εστίες μετάδοσης φημών και ψευδών ή αναξιόπιστων ειδήσεων, οι οποίες και θα συνεισέφεραν στην περαιτέρω εξάπλωση των πραγματικών επιδημιών των ασθενειών. Μια επιδημία της πληροφορίας – ούσα και αυτή μιας μορφής επιδημία –

απαιτεί κατάλληλες παρεμβάσεις μέσω των οποίων θα καταστεί δυνατή η επιβράδυνση ή ο περιορισμός της εξάπλωσής της.

Οι ερευνητές παρουσιάζουν την επιδημία της πληροφορίας ως το αποτέλεσμα της ταυτόχρονης δράσης πολλαπλών ανθρώπινων και μη ανθρώπινων πηγών (bots) που αναμεταδίδουν πληροφορίες, συμπεριλαμβανομένων πολλών αναξιόπιστων, παραπλανητικών ή ψευδών πληροφοριών, σε περιόδους υπερπληθώρας ειδήσεων. Το γεγονός ότι ένας χρήστης μπορεί να δέχεται επανειλημμένα «αντίγραφα» συγκεκριμένου αναξιόπιστου μηνύματος από διαφορετικές πηγές λειτουργεί ως μια έμμεση επικύρωση της αξιοπιστίας και της σχετικότητάς του. Αυτό έχει ως αποτέλεσμα, ένας τέτοιος χρήστης να διαδίδει περαιτέρω τέτοια μηνύματα και να γίνεται με τη σειρά του άλλος ένας φορέας επικίνδυνα παραπλανητικών πληροφοριών.

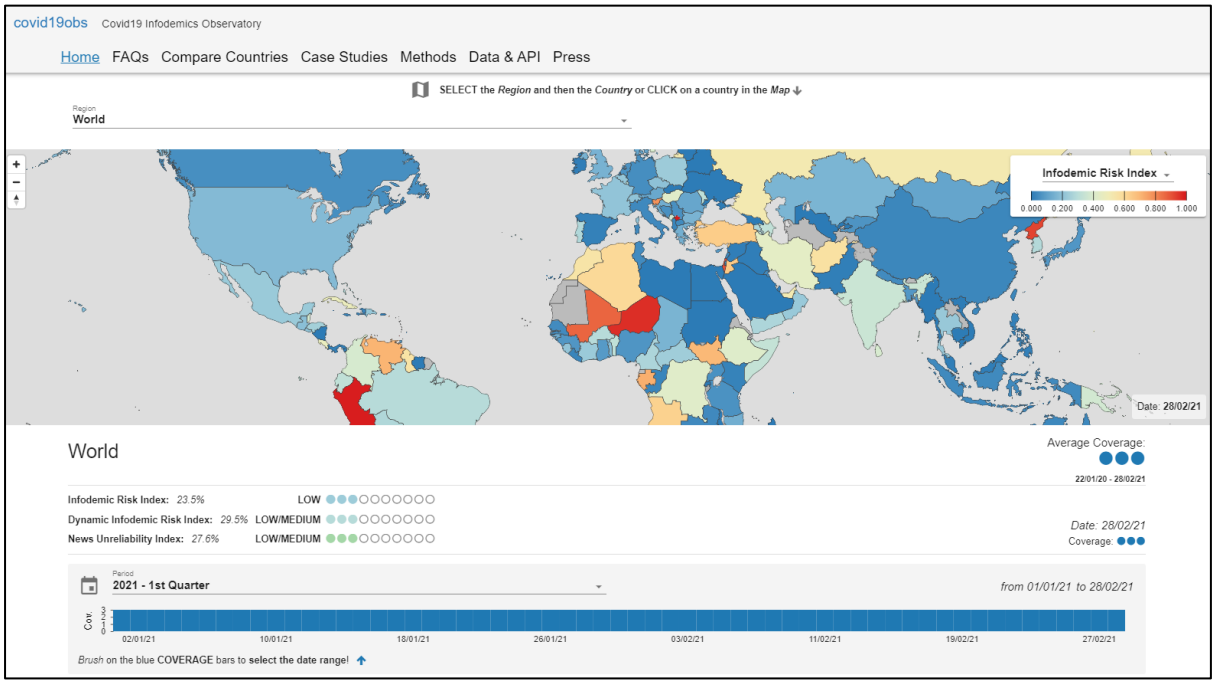


Σχήμα 40 *- Πώς εξαπλώνεται η επιδημία της πληροφορίας (infodemic)* [150]

Όπως φαίνεται και στο Σχήμα 40, οι ανθρώπινοι (κύκλοι) και οι μη ανθρώπινοι / bots (τετράγωνα) λογαριασμοί χρηστών συμμετέχουν στη διάδοση των ειδήσεων σε ένα κοινωνικό δίκτυο. Ορισμένοι χρήστες (A και B) δημιουργούν αναξιόπιστο περιεχόμενο, όπως ψευδείς ή παραπλανητικές ειδήσεις καθώς και μη τεκμηριωμένες πληροφορίες, ενώ άλλοι χρήστες (C) δημιουργούν περιεχόμενο που στηρίζεται σε αξιόπιστες πηγές. Όταν ένα θέμα προσελκύει την παγκόσμια προσοχή, όπως στην περίπτωση της πανδημίας COVID-19, ο όγκος των πληροφοριών που κυκλοφορούν καθιστά πολλές φορές δύσκολη τη διάκριση και τον προσδιορισμό των αξιόπιστων πηγών. Πράγματι, ορισμένοι χρήστες (D) ενδέχεται να εκτίθενται μόνο σε αναξιόπιστες πληροφορίες, ενώ άλλοι (E και F) ενδέχεται να λαμβάνουν αντιφατικές πληροφορίες και να μην είναι βέβαιοι ως προς τις πληροφορίες που πρέπει να εμπιστευτούν και στη συνέχεια να αναμεταδώσουν. Το φαινόμενο αυτό επιδεινώνεται όταν παρουσιάζονται πολλές ταυτόχρονες διαδικασίες εξάπλωσης, και έτσι ορισμένοι χρήστες ενδέχεται να εκτεθούν πολλές φορές στην ίδια πληροφορία ή σε διαφορετική πληροφορία που δημιουργείται από ξεχωριστούς λογαριασμούς χρηστών.

Οι ερευνητές υπολογίζουν το Δείκτη Κινδύνου Επιδημίας της Πληροφορίας (IRI) ως το ρυθμό με τον οποίο ένας μέσος χρήστης σε μια συγκεκριμένη χώρα ή περιοχή εκτίθεται σε αναξιόπιστες δημοσιεύσεις από διαφορετικές κατηγορίες χρηστών, και συγκεκριμένα από επιβεβαιωμένους λογαριασμούς ανθρώπων, ανεπιβεβαίωτους λογαριασμούς ανθρώπων, επιβεβαιωμένους λογαριασμούς bots και ανεπιβεβαίωτους λογαριασμούς bots. Ο IRI λαμβάνει υπόψη τον αριθμό των ακολούθων (followers) των χρηστών κάθε κατηγορίας, τον αριθμό των μηνυμάτων που δημοσιεύουν οι χρήστες, καθώς και την αξιοπιστία τους (όπως μετράται από συγκεκριμένα fact-checking εργαλεία που εφαρμόζονται σε δείγματα μηνυμάτων του χρήστη). Ο IRI είναι συνάρτηση του χρόνου και κυμαίνεται από 0 (που υποδηλώνει μηδενικό κίνδυνο επιδημίας της πληροφορίας) έως 1 (που υποδηλώνει μέγιστο κίνδυνο επιδημίας της πληροφορίας).

Το COVID19 Infodemics Observatory είναι μια ψηφιακή πλατφόρμα που επιτρέπει την απεικόνιση του IRI κάθε χώρας στον κόσμο και είναι δημόσια διαθέσιμη μέσω διαδικτύου [151].



Σχήμα 41 *- Ένα στιγμιότυπο του δείκτη IRI ανά χώρα από την πλατφόρμα COVID19 Infodemics Observatory* [151]

Η πιο πάνω μελέτη έδειξε ότι οι χώρες του κόσμου που χαρακτηρίζονται από διαφορετικά επίπεδα κινδύνου παρουσιάζουν πολύ διαφορετικά προφίλ ειδήσεων, τα οποία φαίνεται να μη σχετίζονται αυστηρά με το επίπεδο κοινωνικοοικονομικής ανάπτυξης των χωρών. Σε χώρες χαμηλού κινδύνου, όπως ο Καναδάς και η Νότια Κορέα, το επίπεδο του IRI παραμένει μικρό καθ' όλη τη διάρκεια της εν λόγω μελέτης (22/1/2020 - 10/3/2020), εκτός κάποιων μεμονωμένων αιχμών (spikes), που συνδέονται κυρίως με μη επιβεβαιωμένες πηγές. Στις χώρες αυτές, καθώς η επιδημία της ασθένειας COVID-19 εξαπλώνεται, ο κίνδυνος επιδημίας της πληροφορίας μειώνεται, υποδηλώνοντας την αυξανόμενη στροφή του κοινού σε αξιόπιστες πηγές ειδήσεων.

Αντίθετα, σε μια χώρα υψηλού κινδύνου όπως η Βενεζουέλα, το επίπεδο επιδημίας της πληροφορίας είναι υψηλό καθ' όλη τη διάρκεια της παρατήρησης, και φαίνεται ότι, εκτός από την αναμενόμενη δραστηριότητα από μη επιβεβαιωμένους λογαριασμούς, ακόμα και επιβεβαιωμένοι λογαριασμοί χρηστών συμβάλλουν σε μεγάλο βαθμό στη διασπορά αναξιόπιστων ειδήσεων. Τέλος, σε μια χώρα σχετικά υψηλού κινδύνου όπως η Ρωσία, το IRI παρουσιάζει μια κάπως ακανόνιστη μορφή ως προς

το χρόνο, με κάποιες αιφνίδιες, πολύ υψηλές, αιχμές. Και στην περίπτωση αυτή, φαίνεται ότι οι επιβεβαιωμένοι λογαριασμοί χρηστών διαδραματίζουν σημαντικό ρόλο στην εξάπλωση αναξιόπιστων πληροφοριών.

Συνολικά, οι ερευνητές παρατήρησαν ότι το παγκόσμιο επίπεδο κινδύνου επιδημίας της πληροφορίας τείνει να μειωθεί καθώς η πανδημία COVID-19 εξαπλώνεται σε όλο τον κόσμο, γεγονός που υποδηλώνει ότι τα πολλά καθημερινά κρούσματα COVID-19 οδήγησαν τους ανθρώπους να αναζητήσουν σχετικά πιο αξιόπιστες πηγές. Φάνηκε επίσης ότι οι επιβεβαιωμένοι «influencers», δηλαδή λογαριασμοί χρηστών με πολλούς ακολούθους, σταδιακά άρχισαν να αναμεταδίδουν πιο αξιόπιστες ειδήσεις. Ωστόσο, η πραγματική ταχύτητα προσαρμογής στη μετάδοση αξιόπιστων ειδήσεων καθορίζει αν μια αρχικά περιορισμένη κατάσταση θα μετεξελιχθεί σε ελεγχόμενη επιδημία ή παγκόσμια πανδημία.

Επίσης, τονίζεται ότι σε μια ψηφιακή και δικτυακά διασυνδεδεμένη παγκόσμια κοινωνία, η επιδημία της ασθένειας και η επιδημία της πληροφορίας αλληλοεξαρτώνται. Δεδομένης της έλλειψης αποτελεσματικών θεραπειών για την καταπολέμηση της νόσου COVID-19, τουλάχιστον στην αρχική φάση της πανδημίας, οι υπεύθυνες συμπεριφορές που βασίζονται σε αξιόπιστες πληροφορίες είναι καίριας σημασίας για το μετριασμό των δυσμενών επιπτώσεων της πανδημίας. Συνεπώς, είναι σημαντικό να αναπτυχθούν ολοκληρωμένες προσεγγίσεις δημόσιας υγείας, όπου οι υγειονομικές μαζί με τις επικοινωνιακές πτυχές μιας πανδημίας λαμβάνονται εξίσου σοβαρά υπόψη και τυγχάνουν κατάλληλης διαχείρισης μέσω του προσεκτικού σχεδιασμού αποτελεσματικών παρεμβάσεων.

5.2.2 Μοντελοποίηση της Εξάπλωσης της Επιδημίας της Πληροφορίας

Όπως με τις επιδημίες ασθενειών, έτσι και με την επιδημία της πληροφορίας, η μοντελοποίηση και ο υπολογισμός της εξάπλωσής της έχει προσελκύσει σημαντικό ερευνητικό ενδιαφέρον. Σε μια πρόσφατη μελέτη [152], οι ερευνητές παρείχαν μια ανάλυση της κοινωνικής δυναμικής με βάση την αλληλεπίδραση χρηστών σε μερικές από τις πιο δημοφιλείς πλατφόρμες μέσων κοινωνικής δικτύωσης. Αναλύοντας θέματα συζητήσεων που σχετίζονταν με την πανδημία COVID-19, προσπάθησαν να μοντελοποιήσουν την εξάπλωση της επιδημίας της πληροφορίας. Συγκεκριμένα, συνέλεξαν οκτώ εκατομμύρια σχόλια και δημοσιεύσεις σε διάστημα 45 ημερών από τρεις mainstream κοινωνικές πλατφόρμες (Twitter, Instagram και YouTube), καθώς και από δύο λιγότερο ρυθμισμένες (regulated) πλατφόρμες (Reddit και Gab)[9].

Οι ερευνητές πραγματοποίησαν μια συγκριτική ανάλυση της δυναμικής της διάδοσης της πληροφορίας σε περιβάλλοντα κοινωνικών δικτύων με διαφορετικά χαρακτηριστικά αλληλεπίδρασης και ακροατήρια (προφίλ) χρηστών. Συνέλεξαν όλο το περιεχόμενο που δημοσίευσαν οι χρήστες σε σχέση με τη νέα ασθένεια από την 1η Ιανουαρίου του 2020 έως τις 14 Φεβρουαρίου του 2020. Τα δεδομένα επιλέχθηκαν με ένα φίλτρο με λέξεις-κλειδιά που βασίστηκαν σε δημοφιλή ερωτήματα χρηστών στη μηχανή αναζήτησης της Google, όπως αναλύονται από την ιστοσελίδα Google Trends, και περιελάμβανε λέξεις όπως: «coronavirus», «coronavirus outbreak», «ncov», «ncov-19», «pandemic», «wuhan». Το προκύπτον σύνολο δεδομένων αποτελούνταν από 1.342.103 δημοσιεύσεις (posts) και 7.465.721 σχόλια που παρήχθησαν από συνολικά 3.734.815 διαφορετικούς χρήστες.

[9] Το Gab είναι ένα χρηματοδοτούμενο από τους χρήστες (crowd-funded) μέσο κοινωνικής δικτύωσης, του οποίου η δομή και τα χαρακτηριστικά είναι εμπνευσμένα από το Twitter. Εκτελεί πολύ λίγο έλεγχο στο περιεχόμενο που δημοσιεύουν οι χρήστες του. Το Reddit είναι ένας αμερικανικός ιστότοπος συγκέντρωσης ειδήσεων (news aggregation), αξιολόγησης περιεχομένου και συζητήσεων.

Για να αποκτήσουν μια εικόνα των συζητήσεων ανάμεσα στους χρήστες σχετικά με την επιδημία της νέας νόσου, οι ερευνητές στη συνέχεια εκμεταλλεύθηκαν τεχνικές επεξεργασίας φυσικής γλώσσας. Συγκεκριμένα, χρησιμοποίησαν νευρωνικά δίκτυα και μεθόδους ομαδοποίησης βασισμένες σε μη επιτηρούμενη μάθηση, για να εξάγουν τελικά ομάδες λέξεων και θεμάτων που φάνηκε να σχετίζονται περισσότερο με συζητήσεις χρηστών γύρω από τη νέα πανδημία.

Οι ερευνητές υπολόγισαν το ενδιαφέρον των χρηστών για θέματα που σχετίζονταν με την πανδημία και συνέκριναν την εξέλιξη των συζητήσεων σε κάθε πλατφόρμα κοινωνικής δικτύωσης. Το ενδιαφέρον των χρηστών μετρήθηκε χρησιμοποιώντας τον αριθμό των δημοσιεύσεων και τον αριθμό των αντιδράσεων (π.χ. σχόλια, likes, κλπ.) σε αυτές τις δημοσιεύσεις κατά τη διάρκεια της περιόδου των 45 ημερών. Για όλες τις πλατφόρμες, βρήκαν μια σημαντική αλλαγή συμπεριφοράς γύρω στις 20 Ιανουαρίου, δηλαδή την ημέρα που ο Π.Ο.Υ. εξέδωσε την πρώτη έκθεση κατάστασης για το νέο κορωνοϊό.

Στη συνέχεια, οι ερευνητές προσπάθησαν να προσομοιώσουν τη διάδοση πληροφοριών στα μέσα κοινωνικής δικτύωσης μέσω παραλλαγών διαφόρων γνωστών επιδημιολογικών μοντέλων. Συγκεκριμένα, χρησιμοποίησαν δύο μοντέλα, ένα μοντέλο εκθετικής αύξησης και ένα δεύτερο μοντέλο βασισμένο στο γνωστό επιδημιολογικό μοντέλο SIR, όπως περιγράφεται στην ενότητα 2.4.1. Με αυτό τον τρόπο, μοντελοποίησαν ουσιαστικά την αύξηση του αριθμού των ατόμων που κάνουν μια δημοσίευση (post) σχετική με ένα θέμα ως μια μολυσματική διαδικασία. Επίσης, μελέτησαν το μέσο αριθμό δευτερογενών «κρουσμάτων» για κάθε μία από τις πλατφόρμες κοινωνικής δικτύωσης ως το πλήθος των χρηστών που αρχίζουν να αναμεταδίδουν δημοσιεύσεις σχετικές με την ασθένεια COVID-19, οι οποίες είχαν προηγουμένως δημιουργηθεί από ένα «μολυσματικό» άτομο.

Όπως και σε άλλες επιδημιολογικές μελέτες ασθενειών, οι ερευνητές προσομοίωσαν την πιθανότητα μιας επιδημίας της πληροφορίας, στην οποία το μέγεθος της συζήτησης σχετικά με το νέο ιό αυξάνεται εκθετικά και άρα, ο βασικός αριθμός αναπαραγωγής R_0 είναι μεγαλύτερος του 1, τουλάχιστον στα αρχικά στάδια. Οι εκτιμώμενες τιμές για το R_0 με βάση τα δύο διαφορετικά μοντέλα εξάπλωσης για τις πέντε διαφορετικές πλατφόρμες μέσων κοινωνικής δικτύωσης παρουσιάζονται στο Σχήμα 42. Όπως παρατηρούμε, και στις πέντε πλατφόρμες και για τα δύο επιδημιολογικά μοντέλα, το R_0 είναι πάντα μεγαλύτερο του 1, γεγονός που δηλώνει ότι η εξάπλωση της πληροφορίας σχετικά με το νέο κορωνοϊό είχε χαρακτηριστικά εκθετικής αύξησης κατά τη διάρκεια του διαστήματος παρατήρησης.

	Gab	Reddit	YouTube	Instagram	Twitter
R_0^{EXP}	[1.42, 1.52]	[1.44, 1.51]	[1.56, 1.70]	[2.02, 2.64]	[1.65, 2.06]
R_0^{SIR}	[2.2, 2.5]	[2.4, 2.8]	[3.2, 3.5]	$[1.1 \times 10^2, 1.6 \times 10^2]$	[4.0, 5.1]

Σχήμα 42 *- Διάστημα εμπιστοσύνης [5%, 95%] για το R_0, σύμφωνα με δύο διαφορετικά επιδημιολογικά μοντέλα, για πέντε πλατφόρμες μέσων κοινωνικής δικτύωσης* [152]

Επιπλέον, οι ερευνητές εξέτασαν τον τρόπο εξάπλωσης της παραπληροφόρησης, την οποία προσδιόρισαν χρησιμοποιώντας εξωτερικούς οργανισμούς ελέγχου γεγονότων (fact-checkers) για τις διάφορες δημοσιεύσεις. Αρχικά, διαπίστωσαν ότι σε ό,τι αφορά τις τρεις μεγάλες πλατφόρμες (Instagram, Twitter, YouTube), οι πληροφορίες τόσο από αξιόπιστες όσο και από αναξιόπιστες πηγές, διαδίδονται με παρόμοιο τρόπο. Όμως, η αλληλεπίδραση των χρηστών με δημοσιεύσεις από λιγότερο αξιόπιστες πηγές είναι γενικά χαμηλότερη σε σχέση με την αλληλεπίδρασή τους με περιεχόμενο από αξιόπιστες πηγές. Ειδικότερα, παρατήρησαν ότι σε αυτά τα μέσα κοινωνικής

δικτύωσης, ο αριθμός των δημοσιεύσεων που παράγονται από αμφιλεγόμενες πηγές είναι μικρός συγκρινόμενος με τον αριθμό των δημοσιεύσεων που παράγονται από αξιόπιστες πηγές, και μάλιστα κάτι αντίστοιχο παρατηρήθηκε και στο Reddit. Μεταξύ των λιγότερο ρυθμισμένων (regulated) μέσων κοινωνικής δικτύωσης, παρατηρήθηκε ένα αξιοπρόσεκτο φαινόμενο στο Gab. Ενώ ο όγκος των δημοσιεύσεων από αμφιλεγόμενες πηγές αντιστοιχούσε περίπου στο 70% του όγκου των δημοσιεύσεων από αξιόπιστες πηγές, ο όγκος των αντιδράσεων (likes και comments) στο αμφιλεγόμενο περιεχόμενο ήταν περίπου τρεις φορές μεγαλύτερος από τον αντίστοιχο όγκο αντιδράσεων στο αξιόπιστο περιεχόμενο. Το αποτέλεσμα αυτό φανερώνει πιθανότατα ότι, ανάλογα με τη φύση του κάθε μέσου κοινωνικής δικτύωσης, οι χρήστες τείνουν να αντιδρούν με διαφορετικό τρόπο σε πληροφορίες που παράγονται από αξιόπιστες ή αμφιλεγόμενες πηγές.

Για την περαιτέρω διερεύνηση του παραπάνω φαινομένου, οι ερευνητές όρισαν τον παράγοντα ενίσχυσης ε ως το μέσο αριθμό αντιδράσεων σε μια δημοσίευση, που αντιπροσωπεύει ουσιαστικά το βαθμό στον οποίο μια δημοσίευση «ενισχύεται» σε ένα μέσο κοινωνικής δικτύωσης. Επίσης, όρισαν το συντελεστή σχετικής ενίσχυσης $\alpha = \varepsilon^U/\varepsilon^R$ για να μετρήσουν πόσες φορές περισσότερο ενισχύεται μια αναξιόπιστη δημοσίευση σε σχέση με μια αξιόπιστη. Όπως φαίνεται στο Σχήμα 43, ο παράγοντας ενίσχυσης αναξιόπιστου περιεχομένου (που παράγεται από αμφιλεγόμενες πηγές) και ο αντίστοιχος παράγοντας ενίσχυσης αξιόπιστου περιεχομένου (που παράγεται από αξιόπιστες πηγές) διαφέρουν σημαντικά από τη μία πλατφόρμα μέσων κοινωνικής δικτύωσης στην άλλη. Συγκεκριμένα, λαμβάνουν τις μεγαλύτερες τιμές τους στο YouTube και τις χαμηλότερες στο Gab.

Οι ερευνητές συνέκριναν επίσης τη σχετική συμπεριφορά μιας πλατφόρμας σε δημοσιεύσεις από αμφιλεγόμενες και αξιόπιστες πηγές, όπως αυτή χαρακτηρίζεται από το συντελεστή σχετικής ενίσχυσης α και ανακάλυψαν ότι, μεταξύ των mainstream μέσων κοινωνικής δικτύωσης, το Twitter φαίνεται να είναι το πιο ουδέτερο, ενώ το YouTube φαίνεται να ενισχύει περισσότερο τις αξιόπιστες πηγές. Μεταξύ των λιγότερο δημοφιλών μέσων κοινωνικής δικτύωσης, το Reddit φαίνεται να μειώνει τον αντίκτυπο των αμφιλεγόμενων πηγών, ενώ το Gab τις ενισχύσει σημαντικά σε σχέση με τις αξιόπιστες πηγές.

	$\mathcal{E}^U$	$\mathcal{E}^R$	α
Gab	5.6	1.4	3.9
Reddit	22.7	40.1	0.55
Twitter	15.1	15.6	0.97
YouTube	1.4×10^4	3.9×10^4	0.35

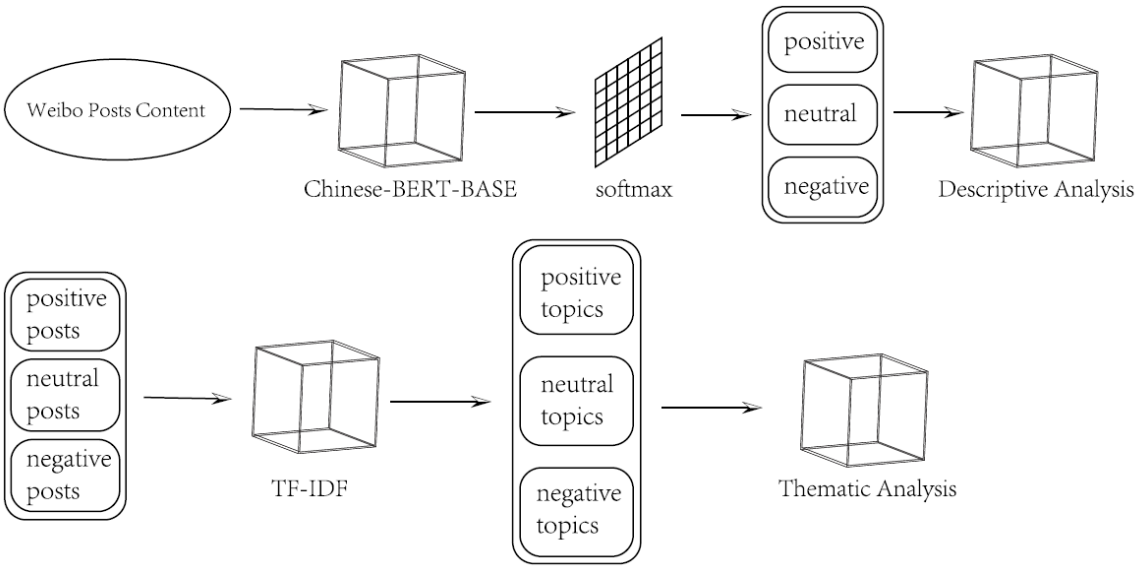
Σχήμα 43 *- Μέσος συντελεστής ενίσχυσης για αναξιόπιστες (ε^U) και αξιόπιστες (ε^R) πηγές και συντελεστής σχετικής ενίσχυσης $\alpha=\varepsilon^U/\varepsilon^R$ σε διαφορετικά μέσα κοινωνικής δικτύωσης* [152]

5.2.3 Κατανόηση Συναισθήματος και Αντιδράσεων της Κοινής Γνώμης

Όπως έχει ήδη αναφερθεί, θέματα που σχετίζονται με τη νέα πανδημία COVID-19 αποτελούν σημαντικό μέρος της δημόσιας συζήτησης στα διάφορα μέσα κοινωνικής δικτύωσης. Η καλύτερη κατανόηση αυτών των συζητήσεων, μέσω των θεμάτων και των αντίστοιχων συναισθημάτων που εκφράζονται από τους χρήστες των μέσων κοινωνικής δικτύωσης, μπορεί να παρέχει χρήσιμες πληροφορίες σε άτομα, οργανισμούς και κυβερνήσεις και να βοηθήσει στη λήψη κατάλληλων αποφάσεων και δράσεων.

Μια ενδιαφέρουσα μελέτη στον τομέα αυτό περιγράφεται στο [153], όπου οι ερευνητές συνέλεξαν δεδομένα από το Sina Weibo, μια δημοφιλή πλατφόρμα κοινωνικής δικτύωσης στην Κίνα, για να αναλύσουν δημοσιεύσεις που φαίνεται να υποδηλώνουν αρνητικά συναισθήματα. Συγκεκριμένα, ανέλυσαν 999.978 τυχαία επιλεγμένες δημοσιεύσεις σχετικές με τον ιό και τη νέα ασθένεια που προκαλεί, από την 1η Ιανουαρίου 2020 έως τις 18 Φεβρουαρίου 2020. Προκειμένου να ταξινομήσουν τα συναισθήματα σε θετικά, ουδέτερα και αρνητικά, χρησιμοποίησαν ένα μοντέλο μη επιτηρούμενης μάθησης με την ονομασία BERT (Bidirectional Encoder Representations from Transformers). Επίσης, χρησιμοποίησαν ένα δεύτερο μοντέλο με την ονομασία TF-IDF (Term Frequency - Inverse Document Frequency) για να συνοψίσουν τα θέματα των δημοσιεύσεων.

Οι ερευνητές πραγματοποίησαν ανάλυση τάσεων (trend analysis) και θεματική ανάλυση (thematic analysis) για τον προσδιορισμό των χαρακτηριστικών των αρνητικών συναισθημάτων. Είναι γενικώς αποδεκτό ότι το συναίσθημα των ανθρώπων στα μέσα κοινωνικής δικτύωσης τείνει να αντανακλά την ευρύτερη ψυχολογική κατάστασή τους. Επίσης, η μεγάλη διάδοση δημοσιεύσεων με αρνητικό συναίσθημα μπορεί να ενισχύσει την κοινωνική αναστάτωση και να θέσει επιπλέον προκλήσεις στον αγώνα για τον περιορισμό της εξάπλωσης της νέας ασθένειας. Για τους πιο πάνω λόγους, οι ερευνητές ανέλυσαν περαιτέρω 11 κυρίαρχα διαφορετικά θέματα συζητήσεων από τις δημοσιεύσεις των χρηστών που περιείχαν αρνητικό συναίσθημα και διερεύνησαν τις τάσεις εκδήλωσης αρνητικών συναισθημάτων καθώς και τη σχέση τους με τα αντίστοιχα θέματα συζητήσεων, με τελικό στόχο να κατανοήσουν τις πραγματικές ανησυχίες των χρηστών. Τέτοιου είδους γνώση θα μπορούσε να χρησιμοποιηθεί από τις δημόσιες αρχές και τις υπηρεσίες υγείας για τη δημιουργία στοχευμένων ενημερωτικών εκστρατειών.



Σχήμα 44 *- Διάγραμμα ροής ενός μοντέλου ανάλυσης συναισθημάτων από μέσα κοινωνικής δικτύωσης* [153]

Δυστυχώς, πολλές φορές το διαδίκτυο και τα μέσα κοινωνικής δικτύωσης χρησιμοποιούνται από συγκεκριμένα άτομα ή ομάδες ατόμων για τη διάδοση ηθελημένα ψευδούς και επικίνδυνου ή προσβλητικού περιεχομένου. Ένα τέτοιο παράδειγμα είναι η διάδοση θεωριών συνωμοσίας και ρητορικής μίσους προς συγκεκριμένες εθνικές ομάδες, ιδιαίτερα προς τον κινεζικό λαό, με αφορμή το γεγονός ότι ο ιός που προκαλεί την ασθένεια COVID-19 φαίνεται να προήλθε από την Κίνα. Είναι προφανές ότι η καλύτερη κατανόηση και η επακόλουθη καταπολέμηση της διάδοσης τέτοιας προβληματικής πληροφορίας είναι ζωτικής σημασίας για την προστασία των δικαιωμάτων των

ανθρώπων, και κυρίως των πιο ευάλωτων και περιθωριοποιημένων ομάδων. Μόνο έτσι μπορούν να σχεδιαστούν οι κατάλληλοι μηχανισμοί παρέμβασης.

Οι ερευνητές στο [154] συνέλεξαν δύο μεγάλα σύνολα δεδομένων από το Twitter και το 4Chan για μια χρονική περίοδο συνολικά πέντε μηνών, με στόχο τη μέτρηση της σινοφοβικής[10] (sinophobic) συμπεριφοράς που εκδηλώθηκε με αφορμή την πανδημία. Αφού ανέλυσαν τα πιο πάνω δεδομένα, χρησιμοποιώντας μοντέλα νευρωνικών δικτύων, διαπίστωσαν ότι η εμφάνιση της πανδημίας COVID-19 οδήγησε πράγματι στην άνοδο της σινοφοβίας στο διαδίκτυο και μάλιστα, η διάδοση σινοφοβικού περιεχομένου είναι ένα φαινόμενο που φαίνεται να παρατηρείται σε όλες τις πλατφόρμες κοινωνικής δικτύωσης.

[10] Η σινοφοβία είναι ένας όρος που χρησιμοποιείται για να περιγράψει αρνητικά συναισθήματα, όπως φόβο και μίσος, για την Κίνα, τους κατοίκους της και τον πολιτισμό της. Η σινοφοβία αναζωπυρώθηκε με αφορμή την πανδημία COVID-19, η οποία ξεκίνησε από την Κίνα.

ΚΕΦΑΛΑΙΟ 6: Τύποι Δεδομένων και Ανοικτά Σύνολα Δεδομένων

6.1 Δεδομένα στην Εποχή της Πανδημίας COVID-19

Η επιτυχία των εφαρμογών τεχνητής νοημοσύνης, που παρουσιάστηκαν στα προηγούμενα Κεφάλαια, για την αντιμετώπιση των διαφορετικών πτυχών της πανδημίας COVID-19 στηρίζεται σε μεγάλο βαθμό στην επαρκή πρόσβαση σε αξιόπιστα δεδομένα. Εκτός από την υπολογιστική ισχύ, η μηχανική μάθηση και κυρίως η βαθιά μάθηση απαιτούν μεγάλες ποσότητες ποιοτικών δεδομένων, προκειμένου να εκπαιδευθούν σωστά νέοι αλγόριθμοι και μοντέλα νευρωνικών δικτύων.

Όπως αναφέρθηκε και στο εισαγωγικό Κεφάλαιο, στον κλάδο της υγείας, οι πηγές κλινικών δεδομένων περιλαμβάνουν αρχεία νοσοκομείων, ιατρικούς φακέλους ασθενών, αποτελέσματα ιατρικών εξετάσεων και μετρήσεις από συσκευές, οι οποίες αποτελούν μέρος του «διαδικτύου των πραγμάτων». Ο κλάδος της βιοϊατρικής έρευνας παράγει επίσης ένα σημαντικό μέρος των μεγάλων δεδομένων (big data) που σχετίζονται με το χώρο της υγείας. Όλες οι πιο πάνω πηγές δεδομένων απαιτούν κατάλληλη διαχείριση, επεξεργασία και ανάλυση, έτσι ώστε να μπορούν να μετατραπούν σε χρήσιμη γνώση.

Στην προ-COVID-19 εποχή, υπήρξαν κάποιες προσπάθειες για την οργάνωση της συλλογής μεγάλου όγκου ιατρικών δεδομένων, κυρίως μέσω της θέσπισης ορισμένων μελετών μεγάλης κλίμακας, όπως για παράδειγμα ο «Άτλας του Καρκινικού Γονιδιώματος» (Cancer Genome Atlas) [155] και η βάση δεδομένων UK Biobank στο Ηνωμένο Βασίλειο [156], καθώς και μέσω της δημιουργίας κάποιων συνόλων δεδομένων, όπως η πλατφόρμα Broad Bioimage Benchmark Collection (συλλογή εικόνων από μικροσκόπιο) [157] και η πλατφόρμα Image Data Resources (συλλογή βιοϊατρικών εικόνων από επιστημονικές δημοσιεύσεις) [158].

Η πανδημία COVID-19 έχει επιταχύνει τις προσπάθειες συλλογής διαφόρων τύπων δεδομένων σε παγκόσμιο επίπεδο μέσω πρωτοβουλιών, τόσο από δημόσιους φορείς όσο και από οργανισμούς του ιδιωτικού τομέα. Σε αυτό το Κεφάλαιο, παρουσιάζουμε αρχικά τους διάφορους τύπους δεδομένων που χρησιμοποιούνται από τις εφαρμογές τεχνητής νοημοσύνης, τις οποίες μελετήσαμε στα προηγούμενα Κεφάλαια. Επίσης, αναφέρουμε και περιγράφουμε επιγραμματικά διάφορα σύνολα δεδομένων ανοικτής πρόσβασης που είναι διαθέσιμα στο διαδίκτυο. Ένας πλούσιος κατάλογος πόρων ανοικτής πρόσβασης διαφορετικών τύπων δεδομένων σχετικών με την αντιμετώπιση της πανδημίας COVID-19, ο οποίος συγκεντρώνει πηγές που παρέχονται από δημόσιες υπηρεσίες καθώς και από ιδιωτικές πρωτοβουλίες, είναι διαθέσιμος στο [159].

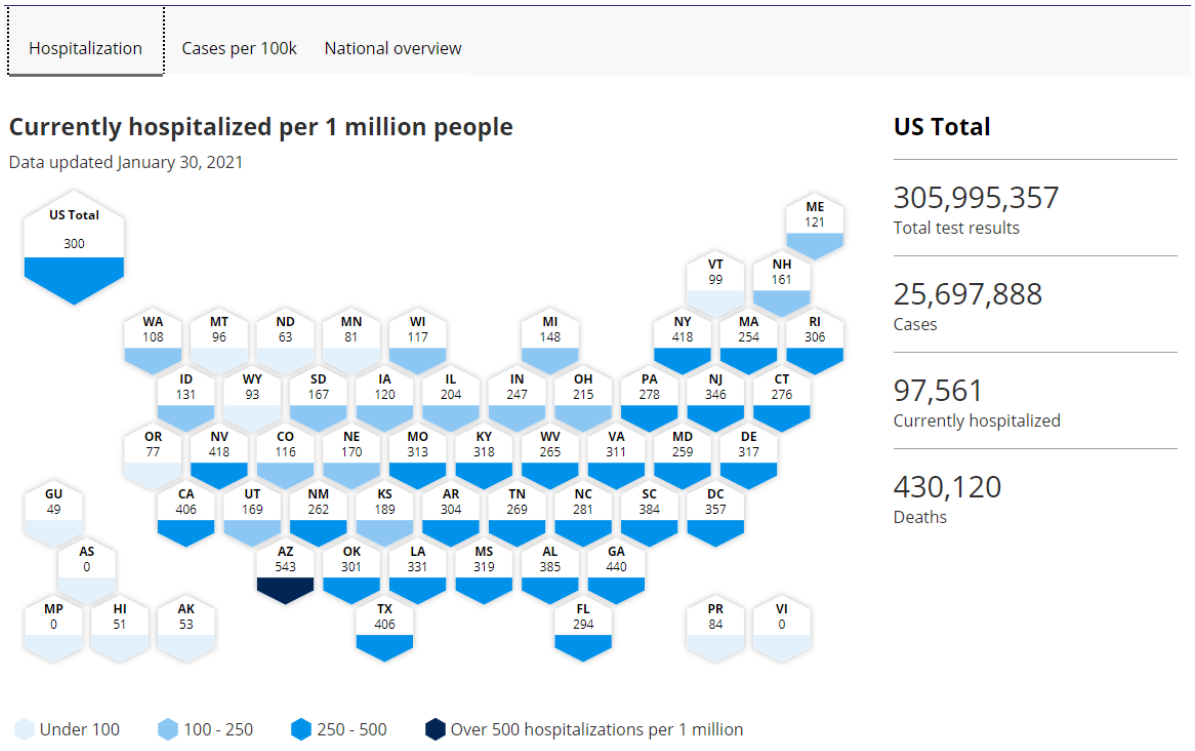
6.2 Δεδομένα Κρουσμάτων

Η γνώση του αριθμού των κρουσμάτων COVID-19 σε συνδυασμό με την αντίστοιχη γεωγραφική τους τοποθεσία μπορεί να βοηθήσει στην παρακολούθηση της εξάπλωσης της πανδημίας και της γεωγραφικής κατανομής των κρουσμάτων. Οι περισσότερες χώρες του κόσμου συλλέγουν και δημοσιοποιούν πληροφορίες σε σχέση με τα κρούσματα. Η συνεπής και ακριβής καταγραφή των επιδημιολογικών πληροφοριών είναι σημαντική για την κατανόηση της μεταδοτικότητας του ιού, του κινδύνου της γεωγραφικής του εξάπλωσης, καθώς και άλλων παραγόντων που πιθανώς επηρεάζουν τον κίνδυνο αυξημένης μετάδοσης. Επίσης, τα δεδομένα κρουσμάτων είναι η βάση για τη λειτουργία των επιδημιολογικών μοντέλων που μελετήσαμε στο Κεφάλαιο 2 και αποτελούν αναπόσπαστο κομμάτι στο σχεδιασμό των δημόσιων πολιτικών και των μέτρων για τη μείωση της εξάπλωσης της πανδημίας. Ένα πλήθος διαφορετικών αλγορίθμων τεχνητής νοημοσύνης μπορεί να εφαρμοστεί σε αυτό τον τύπο δεδομένων, όπως για παράδειγμα αλγόριθμοι πρόβλεψης χρονοσειρών για την

πρόβλεψη της εξέλιξης των κρουσμάτων σε παγκόσμια ή τοπική κλίμακα, όπως έχουμε ήδη περιγράψει λεπτομερώς στην Ενότητα 2.4.

Ένα από τα ευρέως χρησιμοποιούμενα σύνολα δεδομένων κρουσμάτων έχει δημιουργηθεί και ενημερώνεται καθημερινά από το Πανεπιστήμιο John Hopkins των ΗΠΑ, και περιέχει τον ημερήσιο αριθμό κρουσμάτων, τον αριθμό των ασθενών που έχουν ήδη αναρρώσει, καθώς και τα ποσοστά θνησιμότητας σε επίπεδο χώρας, για όλες τις χώρες του κόσμου για τις οποίες υπάρχουν δημοσιευμένα δεδομένα και επιπρόσθετα, σε επίπεδο πολιτείας/επαρχίας για τις ΗΠΑ και την Κίνα [160]. Το Πανεπιστήμιο έχει αναπτύξει επίσης ένα διαδραστικό online dashboard για την απεικόνιση και την παρακολούθηση των προαναφερθέντων δεικτών σε σχεδόν πραγματικό χρόνο. Το dashboard αυτό δημοσιεύθηκε για πρώτη φορά στο διαδίκτυο στις 22 Ιανουαρίου 2020.

Όλα τα δεδομένα που συλλέγονται και εμφανίζονται από το Πανεπιστήμιο John Hopkins είναι ελεύθερα διαθέσιμα σε ένα online αποθετήριο (repository) στο GitHub. Σε ό,τι αφορά τις πηγές από τις οποίες το John Hopkins αντλεί στοιχεία, ο ιστότοπος DXY [161] αποτελεί την κύρια πηγή δεδομένων κρουσμάτων για όλες τις επαρχίες της ηπειρωτικής Κίνας, το Χονγκ Κονγκ, το Μακάου και την Ταϊβάν. Επιπρόσθετες πηγές χρησιμοποιούνται για την άντληση δεδομένων κρουσμάτων σε περιοχές εκτός Κίνας, όπως υπηρεσίες υγείας των ΗΠΑ, καθώς και άλλοι ιστότοποι που συγκεντρώνουν στοιχεία κρουσμάτων από όλο τον κόσμο όπως οι 1point3acres [162], Worldometers.info [163] και COVID Tracking Project [164] (για στοιχεία τεστ και νοσηλειών). Οι ιστότοποι αυτοί βασίζονται με τη σειρά τους σε ένα συνδυασμό ημερήσιων εκθέσεων κρουσμάτων από υπηρεσίες υγείας ανά τον κόσμο, καθώς και σε στοιχεία από διάφορα μέσα ενημέρωσης.



Σχήμα 45 *- Οπτική αναπαράσταση των νοσηλειών ανά 1 εκατομμύριο κατοίκους, ανά πολιτεία στις ΗΠΑ από το «The Covid Tracking Project»* [164]

Σε μια παρόμοια προσπάθεια, οι New York Times δημιούργησαν ένα σύνολο δεδομένων για τις πολιτείες των ΗΠΑ, που αποτελείται από τους συνολικούς αριθμούς κρουσμάτων σε επίπεδο

πολιτείας και κομητείας με την πάροδο του χρόνου. Αυτά τα στοιχεία χρονοσειρών συλλέχθησαν από τις κυβερνήσεις των πολιτειών και τις τοπικές αρχές, καθώς επίσης και από τις υπηρεσίες υγείας, σε μια προσπάθεια να δημιουργηθεί ένα πλήρες αρχείο της τρέχουσας πανδημίας αλλά και της εξέλιξης της εξάπλωσής της στις ΗΠΑ [165]. Εκτός από την αποτύπωση αυτών των δεδομένων σε χάρτες και αναφορές που είναι διαθέσιμες στην αντίστοιχη ιστοσελίδα των New York Times, τα ίδια τα δεδομένα σε ανεπεξέργαστη (raw) μορφή είναι επίσης διαθέσιμα στο κοινό με τη μορφή ενός αποθετηρίου Git. Με αυτό τον τρόπο, ερευνητές, επιστήμονες καθώς και κυβερνητικοί αξιωματούχοι έχουν πρόσβαση σε αυτά τα δεδομένα για την καλύτερη κατανόηση της εξάπλωσης, καθώς και για πιθανή δημιουργία δικών τους εφαρμογών, οι οποίες θα εκμεταλλεύονται τα συγκεκριμένα δεδομένα.

Στην Ελλάδα, το COVID-19 Response Greece [166] αποτελεί μια ανοιχτή κοινότητα περισσότερων από 250 εθελοντών με επιστημονικό και τεχνολογικό υπόβαθρο, οι οποίοι ένωσαν τις δυνάμεις τους για την έρευνα και το σχεδιασμό ανοικτών και πρακτικών λύσεων που μπορούν να συμβάλουν στην αντιμετώπιση των προκλήσεων που θέτει η πανδημία COVID-19. Μία από τις δράσεις της συγκεκριμένης κοινότητας είναι η Ανάλυση Δεδομένων (Data Analytics), η οποία επικεντρώνεται στη συλλογή δεδομένων σε σχέση με τη διάδοση της πανδημίας COVID-19 στην Ελλάδα μέσω διαφόρων πηγών και στη δημιουργία μιας οργανωμένης και ενημερωμένης βάσης δεδομένων που είναι διαθέσιμη για χρήση από τους ερευνητές. Η συγκεκριμένη βάση δεδομένων ενημερώνεται αυτόματα σε καθημερινή βάση και συνδυάζει δεδομένα από διάφορες πηγές, όπως το Κέντρο Επιστήμης και Μηχανικής Συστημάτων του Πανεπιστημίου Johns Hopkins, τον Εθνικό Οργανισμό Δημόσιας Υγείας (ΕΟΔΥ) [167] και τον Παγκόσμιο Οργανισμό Υγείας.

Σε αυτή την κατεύθυνση, έχει επίσης δημιουργηθεί μια δημόσια διαθέσιμη προγραμματιστική διεπαφή (Application Programming Interface - API) με στόχο την παροχή πρόσβασης στα δεδομένα σε σχεδόν πραγματικό χρόνο. Επιπλέον, οι εθελοντές του COVID-19 Response Greece δημιούργησαν dashboards που απεικονίζουν την εξάπλωση της πανδημίας COVID-19 στην Ελλάδα και ανέπτυξαν λογισμικό ανοιχτού κώδικα, το οποίο είναι δωρεάν διαθέσιμο σε ερευνητικά ιδρύματα, κρατικές υπηρεσίες και άλλες ενδιαφερόμενες ομάδες.

6.3 Δεδομένα Ιατρικής Απεικόνισης

Όπως συζητήθηκε στο Κεφάλαιο 3, ιατρικές εικόνες με τη μορφή αξονικών τομογραφιών και ακτινογραφιών θώρακος χρησιμοποιούνται από μεθόδους αυτοματοποιημένης διάγνωσης και πρόγνωσης της ασθένειας COVID-19. Στο πλαίσιο αυτό, έχουν καταβληθεί αρκετές προσπάθειες για συστηματική συλλογή και διάθεση αυτών των δεδομένων απεικόνισης, είτε ως αποτέλεσμα εργασίας μεμονωμένων ερευνητών είτε ως πρωτοβουλίες ευρύτερων κοινοτήτων χρηστών (με τη μορφή crowd-sourcing). Οι ακτινολογικές εικόνες είναι, βασικά, δεδομένα πολλών διαστάσεων που απαιτούν – όπως είδαμε και στο Κεφάλαιο 3 – δυνατότητες επεξεργασίας μεθόδων βαθιάς μάθησης, στις οποίες τα βαθιά νευρωνικά δίκτυα και κυρίως τα μοντέλα που βασίζονται σε μορφές συνελικτικών δικτύων (CNN) διαδραματίζουν έναν πρωτεύοντα ρόλο.

Ένα αρκετά μεγάλο σύνολο δεδομένων ιατρικής απεικόνισης είναι το Covid-Chestxray-Dataset [168]. Αυτό το σύνολο δεδομένων περιέχει επί του παρόντος εκατοντάδες ακτινογραφίες θώρακος, καθιστώντας το ένα χρήσιμο πόρο για την ανάπτυξη και την αξιολόγηση εργαλείων και εφαρμογών διάγνωσης και πρόγνωσης της ασθένειας COVID-19. Τα δεδομένα αυτά συγκεντρώθηκαν από αποθετήρια διαφόρων σχετικών επιστημονικών δημοσιεύσεων, καθώς και από άλλα αποθετήρια διαθέσιμα στο διαδίκτυο και διατίθενται σε μια μορφή φιλική προς επεξεργασία από μεθόδους μηχανικής μάθησης. Οι δημιουργοί αυτού του συνόλου δεδομένων συνέλεξαν εικόνες εμπρόσθιας και πλευρικής όψης καθώς και πολλά χρήσιμα μεταδεδομένα (metadata) που τις συνοδεύουν, όπως

ο χρόνος από την εμφάνιση των πρώτων συμπτωμάτων, το κατά πόσον ο εν λόγω ασθενής εισήλθε σε ΜΕΘ και αν διασωληνώθηκε, αποτελέσματα εξετάσεων αίματος, κλινικές σημειώσεις σε μορφή ελεύθερου κειμένου για την κάθε εικόνα και την κάθε περίπτωση ασθενούς, αν τελικά ο ασθενής επιβίωσε, καθώς και η γεωγραφική τοποθεσία του νοσοκομείου όπου νοσηλεύθηκε. Στην αντίστοιχη ερευνητική εργασία που δημοσίευσαν οι δημιουργοί του συγκεκριμένου συνόλου δεδομένων [169], προτείνουν πολλαπλές πιθανές εφαρμογές μέσω της χρήσης των δεδομένων αυτών, όπως η πρόβλεψη της ανάγκης για εισαγωγή του ασθενούς σε ΜΕΘ, η πρόβλεψη της επιβίωσης του ασθενούς, καθώς και η καλύτερη κατανόηση της πορείας ενός ασθενούς κατά τη διάρκεια της θεραπείας του.

Ένα άλλο ενδιαφέρον σύνολο δεδομένων είναι η εθνική βάση δεδομένων ιατρικής απεικόνισης θώρακος του Ηνωμένου Βασιλείου (National COVID-19 Chest Imaging Database - NCCID) [170], η οποία περιλαμβάνει ακτινογραφίες, αξονικές τομογραφίες και μαγνητικές τομογραφίες θώρακος, καθώς και άλλες σχετικές πληροφορίες ασθενών με υποψία COVID-19. Τα δεδομένα αυτά συλλέγονται από περίπου 20 νοσοκομεία στο Ηνωμένο Βασίλειο. Η βάση δεδομένων NCCID δημιουργήθηκε με στόχο να επιτρέψει την ανάπτυξη και την επικύρωση των αυτοματοποιημένων τεχνολογιών ανάλυσης ιατρικής εικόνας για την αντιμετώπιση ασθενών με COVID-19, αλλά και να επιταχύνει την έρευνα για την καλύτερη κατανόηση της νόσου. Εκτός από τα δεδομένα απεικόνισης, συλλέγονται επίσης και σχετικά κλινικά δεδομένα για όλους τους ασθενείς που βρέθηκαν θετικοί στην ασθένεια COVID-19, καθώς και στοιχεία για ένα μικρότερο δείγμα αρνητικών περιπτώσεων. Οι κατηγορίες δεδομένων που συλλέγονται είναι οι εξής:

- συνήθη δημογραφικά δεδομένα
- δεδομένα καρδιοαναπνευστικής εξέτασης ρουτίνας κατά τη στιγμή της εισαγωγής του ασθενούς στο νοσοκομείο
- όλες οι ακτινογραφίες θώρακος που πραγματοποιήθηκαν κατά τη διάρκεια της παραμονής του ασθενούς στο νοσοκομείο
- όλες οι αξονικές τομογραφίες που πραγματοποιήθηκαν κατά τη διάρκεια της παραμονής του ασθενούς στο νοσοκομείο
- όλες οι ιατρικές εικόνες θώρακος που πραγματοποιήθηκαν κατά τη διάρκεια των τριών ετών που προηγήθηκαν της πρώτης απεικονιστικής εξέτασης που σχετίζεται με την ασθένεια COVID-19 (μόνο για ασθενείς με COVID-19)
- επιπρόσθετες ιατρικές εικόνες (συμπεριλαμβανομένων και μαγνητικών τομογραφιών) που πραγματοποιήθηκαν για τη διερεύνηση καρδιακής βλάβης που πιθανώς σχετίζεται με την ασθένεια COVID-19
- βιοχημικά και αιματολογικά δεδομένα που συλλέγονται συστηματικά
- δεδομένα σε σχέση με την έκβαση του ασθενούς, όπως ο χρόνος αναπνευστικής υποστήριξης, η ημερομηνία εξιτηρίου ή θανάτου

Τα αιτήματα πρόσβασης στη βάση δεδομένων NCCID αξιολογούνται από επιτροπή εμπειρογνωμόνων των τομέων της υγείας και της τεχνολογίας, βάσει διαφόρων κριτηρίων που αναφέρονται στον αντίστοιχο δικτυακό τόπο.

Μια ακόμα ενδιαφέρουσα πρωτοβουλία, που προέρχεται από την Ευρωπαϊκή Ένωση, στην περιοχή των δεδομένων και των εφαρμογών ιατρικής εικόνας είναι το Covid19Challenge.EU [171]. Κατανοώντας ότι η ιατρική απεικόνιση είναι δυνητικά χρήσιμη για τη διάγνωση της ασθένειας COVID-19, η πρωτοβουλία αυτή έχει στόχο να συνδέσει επιστήμονες και μηχανικούς τεχνητής νοημοσύνης για την υποστήριξη των ιατρών στην επίλυση σύνθετων προβλημάτων. Μία από τις πρώτες φάσεις αυτού του έργου αφορά στη συλλογή δεδομένων από διάφορους παγκόσμιους

κλινικούς φορείς - συνεργάτες. Οι συνεργάτες (ακαδημαϊκά ιδρύματα που παρέχουν δεδομένα ασθενών) έχουν το δικαίωμα πρόσβασης σε ολόκληρη τη βάση δεδομένων προκειμένου να μπορέσουν να πραγματοποιήσουν την έρευνά τους. Οι κλινικοί ιατροί και οι επιστήμονες υγείας αναπτύσσουν στόχους και υποθέσεις που καθορίζουν τα προς επίλυση προβλήματα.

6.4 Μοριακά Δεδομένα

Οι εφαρμογές τεχνητής νοημοσύνης στον τομέα της Φαρμακολογίας, όπως περιγράφονται στο Κεφάλαιο 4, χρησιμοποιούν διάφορα είδη δεδομένων σε μοριακό επίπεδο, όπως για παράδειγμα γονιδιωματικές ακολουθίες και πρωτεϊνικές δομές. Σε αυτή την περιοχή, υπάρχουν πολλά διαθέσιμα σύνολα δεδομένων είτε από ήδη υπάρχουσες βάσεις δεδομένων είτε από νέες που δημιουργήθηκαν ειδικά στα πλαίσια της πανδημίας COVID-19.

Η αλληλουχία του γονιδιώματος του ιού SARS-CoV-2 είναι ζωτικής σημασίας για το σχεδιασμό και την αξιολόγηση των ειδικών διαγνωστικών δοκιμών (τεστ), την ιχνηλάτηση των διαφόρων στελεχών του ιού, καθώς και για τον εντοπισμό των κατάλληλων στόχων για πιθανές θεραπείες και εμβόλια. Το GISAID (Global Initiative on Sharing Avian Influenza Data) [172] είναι μια παγκόσμια επιστημονική πρωτοβουλία και ταυτόχρονα μια βάση πρωτογενών δεδομένων που παρέχει ανοικτή πρόσβαση σε γονιδιωματικά δεδομένα των ιών της γρίπης και του νέου κορωνοϊού SARS-CoV-2. Στις 10 Ιανουαρίου του 2020, οι πρώτες αλληλουχίες ολόκληρου του γονιδιώματος του SARS-CoV-2 δημοσιεύθηκαν στο GISAID, γεγονός που σήμανε την έναρξη της έρευνας σε διάφορους τομείς, καθώς και την ανάπτυξη των πρώτων υποψήφιων εμβολίων κατά του ιού και των ειδικών διαγνωστικών δοκιμών για την ανίχνευσή του. Το GISAID διευκολύνει επίσης τον τομέα της γονιδιωματικής επιδημιολογίας, μέσω της παρακολούθησης σε πραγματικό χρόνο πιθανής εμφάνισης νέων στελεχών του ιού SARS-CoV-2 σε ολόκληρο τον πλανήτη.

Ένας από τους σημαντικότερους πόρους σε ό,τι αφορά γνωστές πρωτεϊνικές δομές, οι οποίες χρησιμοποιούνται από εφαρμογές τεχνητής νοημοσύνης, είναι η Τράπεζα Δεδομένων Πρωτεϊνών (Research Collaboratory for Structural Bioinformatics, Protein Data Bank – RCSB PDB) [173]. Οι γνωστές αυτές δομές χρησιμοποιούνται κατά τη φάση της εκπαίδευσης των κατάλληλων μοντέλων μηχανικής μάθησης σε εφαρμογές όπως η πρόβλεψη νέων πρωτεϊνικών δομών, καθώς και η ανακάλυψη φαρμάκων και εμβολίων. Η βάση δεδομένων της PDB περιέχει πληροφορίες για την τρισδιάστατη δομή μεγάλων βιολογικών μορίων, όπως πρωτεΐνες και νουκλεϊνικά οξέα (DNA και RNA). Οι τρισδιάστατες αυτές δομές λαμβάνονται συνήθως με κρυσταλλογραφία ακτίνων-Χ (X-ray crystallography), φασματοσκοπία πυρηνικού μαγνητικού συντονισμού (Nuclear Magnetic Resonance - NMR - spectroscopy), ή, όλο και συχνότερα, με κρυο-ηλεκτρονική μικροσκοπία (cryo-electron microscopy). Τα δεδομένα αυτά προέρχονται από βιολόγους και βιοχημικούς από όλο τον κόσμο και είναι ελεύθερα προσβάσιμα στο διαδίκτυο μέσω των ιστοσελίδων των οργανώσεων - μελών της PDB. Όπως ήδη αναφέρθηκε στο Κεφάλαιο 4, η γνώση της τρισδιάστατης δομής ενός βιολογικού μακρομορίου είναι απαραίτητη για την κατανόηση του ρόλου του στην υγεία και στις ασθένειες των ανθρώπων και των ζώων.

Μια άλλη ενδιαφέρουσα πηγή δεδομένων ανοικτής πρόσβασης δημιουργήθηκε από την υπηρεσία Chemical Abstracts Service (CAS), που αποτελεί τμήμα της Αμερικανικής Χημικής Εταιρείας (American Chemical Society). Η CAS δημοσίευσε ένα σύνολο δεδομένων με χημικές ενώσεις που αποτελούν υποψήφια αντιιικά φάρμακα κατά της ασθένειας COVID-19 [174]. Το σύνολο αυτό περιέχει πληροφορίες για σχεδόν 50.000 χημικές ουσίες, όπως γνωστά αντιιικά φάρμακα καθώς και παρεμφερείς χημικές ενώσεις, ενώσεις δηλαδή που παρουσιάζουν δομικές ομοιότητες με γνωστά αντιιικά φάρμακα. Παρέχοντας αυτό το σύνολο υποψηφίων μορίων, η CAS υποστηρίζει τους

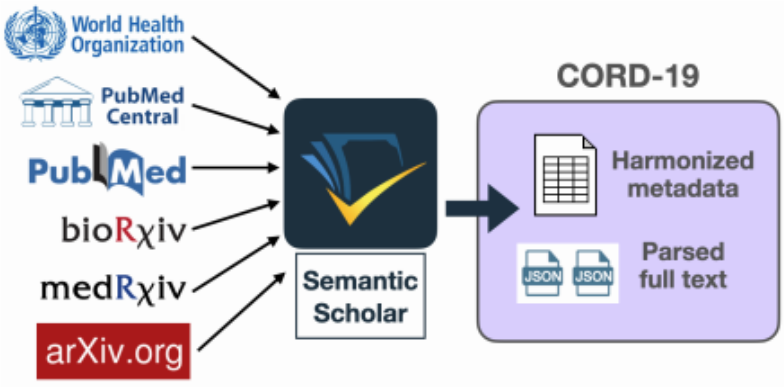
ερευνητές στην ανακάλυψη νέων ή επαναστοχευμένων θεραπειών κατά της νόσου. Εκτός από το χημικό τύπο της κάθε ουσίας, παρέχονται επίσης σχετικά μεταδεδομένα, όπως ο αριθμός μητρώου CAS, οι φυσικές ιδιότητες και ένας πίνακας σύνδεσης για κάθε ουσία[11].

6.5 Δεδομένα Επιστημονικής Βιβλιογραφίας

Από την αρχή της εμφάνισης της νόσου COVID-19, η ερευνητική κοινότητα έχει υπάρξει πάρα πολύ δραστήρια, κάτι που αποδεικνύεται στην πράξη από το μεγάλο αριθμό επιστημονικών δημοσιεύσεων που παράγονται καθημερινά. Η πληθώρα αυτών των άρθρων μπορεί να αναλυθεί χρησιμοποιώντας τεχνικές μηχανικής μάθησης για να δώσει απαντήσεις σε διάφορα ερευνητικά ερωτήματα. Κατανοώντας τη δυνατότητα αυτή, αρκετοί οργανισμοί έχουν αναλάβει πρωτοβουλίες για τη συγκέντρωση του σχετικού επιστημονικού έργου και τη διάθεσή του στην κατάλληλη μορφή για περαιτέρω υπολογιστική ανάλυση.

Σε αυτή την προσπάθεια, ο Π.Ο.Υ. συγκέντρωσε όλα τα πρόσφατα διεθνή επιστημονικά ευρήματα σε διάφορες γλώσσες και την αντίστοιχη γνώση σχετικά με την πανδημία COVID-19 σε μια βάση δεδομένων με την ονομασία «WHO COVID-19 database». Αυτή η βάση δεδομένων ενημερώνεται καθημερινά από αυτόματες αναζητήσεις βιβλιογραφικών βάσεων δεδομένων, αναζητήσεις άρθρων από ειδικούς καθώς και από την προσθήκη άλλων επιστημονικών άρθρων [175].

Ένα από τα πιο σημαντικά και πιο πλήρη σύνολα δεδομένων επιστημονικής βιβλιογραφίας είναι το σύνολο δεδομένων CORD-19. Πρόκειται για το αποτέλεσμα της συλλογικής εργασίας από την ομάδα Semantic Scholar του Allen Institute for AI, το τμήμα έρευνας της Microsoft (Microsoft Research), την Εθνική Βιβλιοθήκη Ιατρικής των Εθνικών Ινστιτούτων Υγείας των ΗΠΑ (National Institutes of Health - NIH) και το Γραφείο Επιστήμης και Τεχνολογίας του Λευκού Οίκου (Office of Science and Technology - OSTP) [176]. Κατά τη στιγμή της συγγραφής αυτής της διπλωματικής εργασίας, το CORD-19 περιλαμβάνει περισσότερα από 200.000 επιστημονικά άρθρα σε μορφή αναγνώσιμη από υπολογιστή (machine-readable).



Σχήμα 46 *- Το σύνολο δεδομένων επιστημονικής βιβλιογραφίας CORD-19* [176]

Ο στόχος της δημιουργίας του CORD-19 ήταν να διευκολύνει την ιατρική κοινότητα καθώς και την κοινότητα της μηχανικής μάθησης στην έρευνά τους για την εφαρμογή σύγχρονων τεχνικών

[11] Ένας πίνακας σύνδεσης είναι μια μήτρα που παρέχει πληροφορίες που χρειάζεται ένας υπολογιστής για τη δημιουργία ενός μοριακού γράφου. Συγκεκριμένα, καθορίζει ποια είναι τα άτομα του μορίου και πώς συνδέονται μεταξύ τους, δηλαδή τους κόμβους και τις ακμές του γράφου.

επεξεργασίας φυσικής γλώσσας και την εξαγωγή χρήσιμης γνώσης από την επιστημονική βιβλιογραφία, ικανής να συμβάλει στην καταπολέμηση της ασθένειας COVID-19.

Το CORD-19 συλλέγει άρθρα από πολλές επιμέρους online βιβλιοθήκες επιστημονικής βιβλιογραφίας όπως οι PMC, BioRxiv, MedRxiv και άλλες, μέσω ερωτημάτων αναζήτησης που σχετίζονται με την πανδημία COVID-19 (και περιλαμβάνουν λέξεις-κλειδιά όπως «COVID-19», «Coronavirus», «Corona virus», «2019-nCoV», κλπ.). Περιλαμβάνει επίσης πρόσθετα ερευνητικά άρθρα σε σχέση με την πανδημία COVID-19 από την προαναφερθείσα βάση δεδομένων που διατηρεί ο Π.Ο.Υ.. Μαζί με το σύνολο δεδομένων του κάθε άρθρου, παρέχεται και ένα αρχείο μεταδεδομένων που περιλαμβάνει το αναγνωριστικό δημοσίευσης (Digital Object Identifier - DOI) του άρθρου καθώς και τον εκδότη του, μεταξύ άλλων πληροφοριών. Πιθανές διπλοεγγραφές αφαιρούνται με βάση το αναγνωριστικό δημοσίευσης DOI. Στο σχεδιασμό του CORD-19 ελήφθησαν υπόψη διάφορες προδιαγραφές, όπως το να είναι το κείμενο σε μορφή αναγνώσιμη από υπολογιστή, διάφοροι περιορισμοί πνευματικών δικαιωμάτων και ύπαρξη αξιόπιστων μεταδεδομένων.

Το CORD-19 είναι πολύ δημοφιλές στην ερευνητική κοινότητα με περισσότερες από 3,35 εκατομμύρια προβολές και περισσότερες από 115 χιλιάδες λήψεις (όπως αποτυπώθηκαν στις 31-01-2021). Αξίζει επίσης να σημειωθεί ότι η Kaggle, μια online πλατφόρμα διαγωνισμών στον τομέα της επιστήμης των δεδομένων, έχει δημιουργήσει ένα διαγωνισμό (challenge) δημιουργίας εφαρμογών με χρήση αυτών των δεδομένων, που ονομάζεται COVID-19 Open Research Dataset Challenge [177].

Ακόμα μια αξιόλογη πρωτοβουλία συλλογής επιστημονικής βιβλιογραφίας είναι το LitCovid. Πρόκειται για έναν κόμβο συλλογής επιστημονικών άρθρων που έχουν αξιολογηθεί από ειδικούς και στοχεύουν στην άντληση ενημερωμένων επιστημονικών πληροφοριών για την πανδημία COVID-19 [178]. Το LitCovid περιέχει συνολικά 93.831 άρθρα (όπως αποτυπώθηκαν στις 31-01-2021) από την online βιβλιοθήκη επιστημών υγείας και βιοϊατρικής PubMed. Καθημερινά, ένα σύνολο άρθρων ανακτάται από το PubMed χρησιμοποιώντας συγκεκριμένο ερώτημα αναζήτησης[12] και στη συνέχεια, τα αποτελέσματα που επιστρέφονται εξετάζονται από ειδικούς, οι οποίοι απομονώνουν τα σχετικά άρθρα και απορρίπτουν τα μη-σχετικά, με τη βοήθεια και ενός αλγόριθμου μηχανικής μάθησης για ταξινόμηση κειμένου. Οι μέθοδοι μηχανικής μάθησης αναπτύχθηκαν για να βοηθήσουν τους επιστήμονες να ανταπεξέλθουν στον τεράστιο όγκο της βιβλιογραφίας που απαιτείται να ταξινομήσουν καθημερινά.

Τα σχετικά άρθρα στο LitCovid κατατάσσονται από τους επιστήμονες σε οκτώ κατηγορίες. Οι κατηγορίες είναι επί του παρόντος οι εξής: γενικές πληροφορίες, μηχανισμός, μετάδοση, διάγνωση, θεραπεία, πρόληψη, αναφορά κρουσμάτων και επιδημιολογική πρόβλεψη. Το LitCovid διαφέρει από το CORD-19, καθώς το πρώτο δεν περιλαμβάνει προ-δημοσιεύσεις (pre-prints). Επίσης, το LitCovid επικεντρώνεται στην παρακολούθηση εξειδικευμένων δημοσιεύσεων για την πανδημία COVID-19, ενώ το CORD-19 περιλαμβάνει ευρύτερη βιβλιογραφία σε σχέση με κορωνοϊούς (π.χ. περιλαμβάνει και άρθρα για τους SARS-CoV και MERS-CoV), καθώς και άρθρα μιας ευρύτερης χρονικής περιόδου, συμπεριλαμβανομένων και δημοσιεύσεων πριν από την εμφάνιση του νέου κορονωϊού SARS-CoV-2.

Άλλη μία ενδιαφέρουσα προσπάθεια είναι αυτή του εκδοτικού οίκου Elsevier, ο οποίος μέσω του νέου ιστοτόπου «Novel Coronavirus Information Center», έχει καταστήσει δημόσια διαθέσιμα επιστημονικά άρθρα σε αρχικό στάδιο έρευνας (early-stage research) καθώς και δημοσιεύσεις με

[12] "coronavirus"[All Fields] OR "ncov"[All Fields] OR "cov"[All Fields] OR "2019-nCoV"[All Fields] OR "COVID-19"[All Fields] OR "SARS-CoV-2"[All Fields]

αξιολόγηση από ομότιμους (peer-reviewed articles), σε σχέση με τη νέα πανδημία COVID-19 [179]. Ο πιο πάνω ιστότοπος περιλαμβάνει, επίσης, δεκάδες χιλιάδες άρθρα από την online βιβλιοθήκη ScienceDirect του ίδιου εκδοτικού οίκου.

Τέλος, αξίζει να αναφερθεί και η πρωτοβουλία του οργανισμού Lens, ο οποίος έχει επίσης δημοσιεύσει όλα τα δεδομένα του σε σχέση με διπλώματα ευρεσιτεχνίας (patents) σε ένα νέο ιστότοπο με την ονομασία «Human Coronavirus Innovation Landscape Patent and Research Works Open Datasets». Στόχος της πρωτοβουλίας αυτής είναι η υποστήριξη της ανάπτυξης βέλτιστων πρακτικών και καινοτόμων προϊόντων για τη διάγνωση, την ανακάλυψη θεραπειών και εμβολίων, την πρόληψη και άλλες πιθανές περιοχές εφαρμογών [180].

6.6 Δεδομένα Μέσων Κοινωνικής Δικτύωσης

Οι δημοφιλείς πλατφόρμες μέσων κοινωνικής δικτύωσης, όπως για παράδειγμα το Twitter και το Facebook, περιέχουν πληθώρα δεδομένων που μπορούν να χρησιμοποιηθούν τόσο για ερευνητικούς σκοπούς όσο και για την ανάπτυξη καινοτόμων εφαρμογών. Κατά τη διάρκεια της τελευταίας δεκαετίας, τα μέσα κοινωνικής δικτύωσης έχουν αποτελέσει ένα πολύτιμο κανάλι επικοινωνίας κατά τη διάρκεια καταστροφών ή άλλων έκτακτων καταστάσεων και κρίσεων. Τέτοιου είδους δεδομένα μπορούν να χρησιμοποιηθούν από διάφορους τύπους εφαρμογών τεχνητής νοημοσύνης στη μάχη κατά της πανδημίας COVID-19, όπως παρουσιάστηκαν σε προηγούμενα Κεφάλαια, και περιλαμβάνουν την έγκαιρη προειδοποίηση, την κατανόηση των συμπτωμάτων της νέας ασθένειας, τον υπολογισμό του κινδύνου, τη μοντελοποίηση της εξάπλωσης, την ανάλυση συναισθημάτων και τον έλεγχο για παραπληροφόρηση και ψευδείς ειδήσεις.

Ένα πολύ μεγάλης κλίμακας σύνολο δεδομένων αυτής της κατηγορίας δημιουργήθηκε από ερευνητές στο Panacea Lab του Georgia State University των ΗΠΑ και περιλαμβάνει δεδομένα από το Twitter [181]. Οι πρώτες 9 εβδομάδες δεδομένων (από την 1/1/2020 έως την 11/3/2020) περιέχουν πολύ λίγα tweets, καθώς τα δεδομένα φιλτράρονταν για άλλους ερευνητικούς σκοπούς. Παρά ταύτα, παρατηρήθηκε μια δραματική αύξηση στον αριθμό των tweets καθώς ο κόσμος ευαισθητοποιούνταν περισσότερο για το νέο ιό. Η πλήρης συλλογή δεδομένων από το Twitter ξεκίνησε στις 12 Μαρτίου και απέφερε πάνω από 4 εκατομμύρια tweets ημερησίως.

Τα δεδομένα που συλλέγονται στο πιο πάνω σύνολο περιλαμβάνουν tweets σε όλες τις γλώσσες, με το υψηλότερο ποσοστό να αντιστοιχεί στα αγγλικά, ισπανικά και γαλλικά. Όλα τα tweets και retweets περιλαμβάνονται στο πλήρες σύνολο δεδομένων, ενώ υπάρχει και μια φιλτραρισμένη έκδοση που δεν περιλαμβάνει τα retweets. Υπάρχουν διάφοροι πρακτικοί λόγοι για να εκμεταλλευθεί κάποιος τα retweets, όπως για παράδειγμα ο εντοπισμών των πιο σημαντικών / δημοφιλών tweets με βάση το πλήθος των retweets τους. Για να διευκολύνουν τις εφαρμογές επεξεργασίας φυσικής γλώσσας, οι δημιουργοί του συνόλου δεδομένων παρέχουν τις 1000 πιο συχνά εμφανιζόμενες λέξεις, καθώς και τις 1000 πιο συχνές φράσεις δύο και τριών λέξεων. Επίσης, περιλαμβάνουν ορισμένα γενικά στατιστικά στοιχεία ανά ημέρα. Το σύνολο δεδομένων ενημερώνεται κάθε δύο ημέρες στο αντίστοιχο αποθετήριο GitHub. Ξεκινώντας από την έκδοση 20 του συνόλου δεδομένων και έπειτα, οι δημιουργοί του έχουν συμπεριλάβει τη γλώσσα και την τοποθεσία ως μεταδεδομένα για όλα τα tweets.

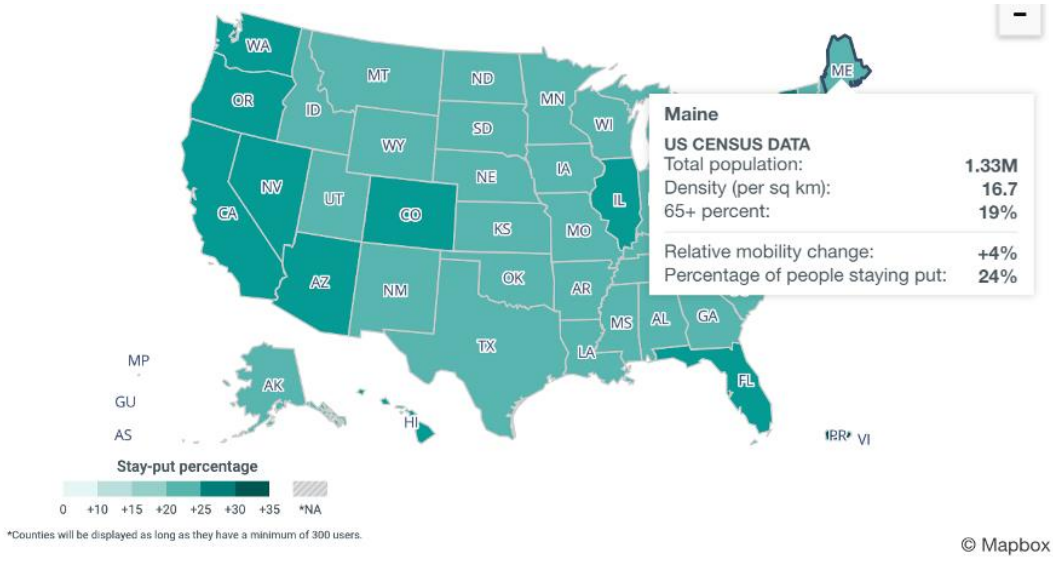
Ένα άλλο σύνολο δεδομένων από το Twitter που ενημερώνεται εβδομαδιαίως δημιουργήθηκε και συντηρείται από ερευνητές του Πανεπιστημίου της Νότιας Καλιφόρνιας [182]. Πρόκειται για ένα πολύγλωσσο σύνολο tweets σχετικών με την πανδημία COVID-19 που διατίθεται στην ερευνητική κοινότητα μέσω ενός αποθετηρίου GitHub. Η συλλογή δεδομένων ξεκίνησε στις 28 Ιανουαρίου 2020, αξιοποιώντας την προγραμματιστική διεπαφή (API) του Twitter καθώς και τη βιβλιοθήκη εργαλείων

Tweepy για να ακολουθήσει (follow) ορισμένες λέξεις-κλειδιά και λογαριασμούς χρηστών που ήταν δημοφιλείς εκείνη την περίοδο. Κατά τη συγγραφή της εργασίας αυτής (31-01-2021), περισσότερα από 1 δισεκατομμύριο tweets περιλαμβάνονται στο συγκεκριμένο σύνολο δεδομένων, με πάνω από το 65% αυτών να είναι στην αγγλική γλώσσα.

Εκτός από τις αναρτήσεις ατόμων (ιδιωτών) στα μέσα κοινωνικής δικτύωσης, υπάρχουν επίσης σημαντικές πληροφορίες που δημοσιεύονται σε αυτά από επίσημες πηγές, όπως για παράδειγμα κυβερνήσεις και οργανισμούς, καθώς και από ειδησεογραφικά πρακτορεία, σε μια προσπάθεια για συνεχή κάλυψη των θεμάτων της πανδημίας. Ένα τέτοιο παράδειγμα είναι το «Institutional, News and Media Tweet Dataset» [183], το οποίο συγκεντρώνει tweets με βάση μια λίστα περίπου 70 πηγών και μπορεί να χρησιμοποιηθεί για την παρακολούθηση των μηνυμάτων που εκπέμπουν επίσημες και θεσμικές πηγές σε διάφορες χώρες σε σχέση με την πανδημία COVID-19.

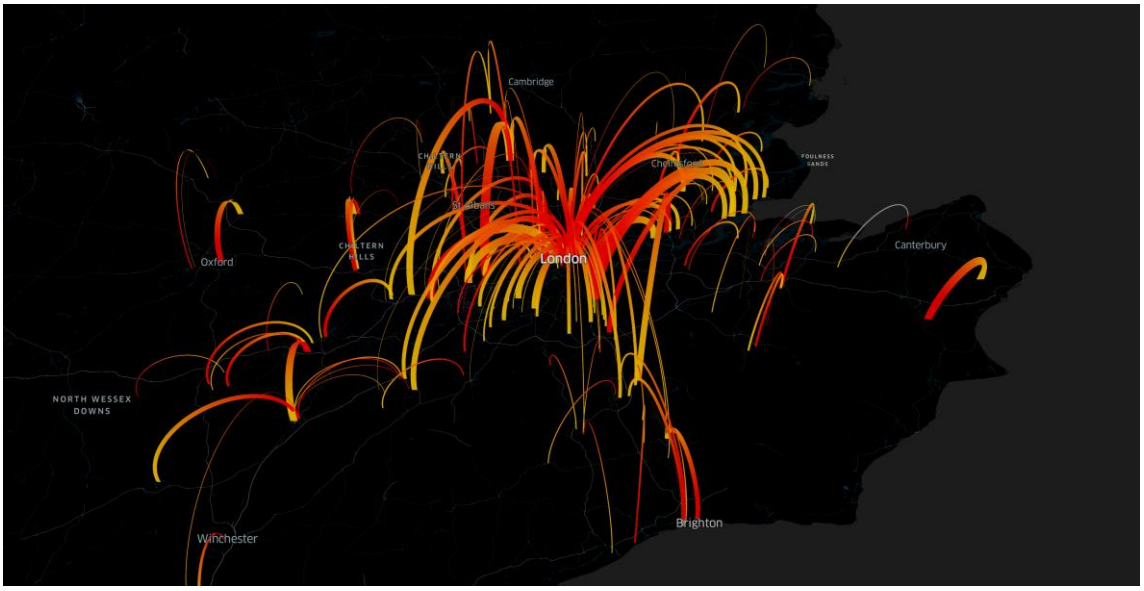
Εκτός από το Twitter, υπάρχουν και άλλες πλατφόρμες μέσων κοινωνικής δικτύωσης που περιέχουν χρήσιμα δεδομένα για την καλύτερη κατανόηση της πανδημίας και την ανάπτυξη εφαρμογών για την αντιμετώπιση διαφόρων πτυχών της. Η πρωτοβουλία «Data for Good» του Facebook [184] αποτελεί ένα τέτοιο παράδειγμα, μέσω της παροχής χρήσιμων πληροφοριών, είτε με τη μορφή ανοικτών συνόλων δεδομένων είτε με την παροχή δεδομένων κατόπιν αιτήματος ερευνητών και μη κερδοσκοπικών οργανισμών, με στόχο τη χρήση τους για τη δημιουργία εφαρμογών τεχνητής νοημοσύνης στα πλαίσια της πανδημίας COVID-19.

Ένα αντιπροσωπευτικό παράδειγμα των δεδομένων που διαθέτει το Facebook είναι οι Χάρτες Εύρους Μετακίνησης (Movement Range Maps) που μπορεί να φανούν χρήσιμοι σε ερευνητές και ειδικούς δημόσιας υγείας, καθώς αποκαλύπτουν τους τρόπους με τους οποίους οι διάφοροι πληθυσμοί ανταποκρίνονται στα μέτρα κοινωνικής αποστασιοποίησης (social distancing). Το συγκεκριμένο σύνολο δεδομένων παρέχει δύο διαφορετικούς δείκτες: το δείκτη Αλλαγής στην Κίνηση (Change in Movement) και το δείκτη Παραμονής (Stay Put). Η τιμή του δείκτη Change in Movement ουσιαστικά αντιπροσωπεύει ένα μέτρο της μετακίνησης των ανθρώπων ως προς μια περίοδο αναφοράς που συνήθως προηγείται των περισσότερων μέτρων κοινωνικής αποστασιοποίησης. Ο δείκτης Stay Put αντιστοιχεί με τη σειρά του στο ποσοστό των ανθρώπων που παραμένουν εντός μιας μικρής περιοχής γύρω από το σπίτι τους κατά τη διάρκεια μιας ολόκληρης μέρας. Ένα παράδειγμα ενός τέτοιου χάρτη φαίνεται στο Σχήμα 47.



Σχήμα 47 *- Παράδειγμα ενός χάρτη εύρους μετακίνησης (movement range map) του Facebook* [184]

Μια άλλη κατηγορία δεδομένων του Facebook είναι οι Χάρτες Πρόληψης Ασθενειών (Disease Prevention Maps) [185]. Οι χάρτες αυτοί σχεδιάστηκαν για να βοηθήσουν τους οργανισμούς δημόσιας υγείας των χωρών να κατανοήσουν καλύτερα τον τρόπο με τον οποίο μετακινούνται οι άνθρωποι, προκειμένου να βελτιωθεί η αποτελεσματικότητα των μέτρων αντιμετώπισης επιδημιών. Η λειτουργία των χαρτών αυτών ξεκίνησε το Μάιο του 2019 (δηλαδή πριν από την εμφάνιση της πανδημίας COVID-19) και έχουν χρησιμοποιηθεί και για άλλες υγειονομικές κρίσεις, όπως για την επιδημία χολέρας στη Μοζαμβίκη. Υπάρχουν διάφορες υποκατηγορίες χαρτών πρόληψης ασθενειών. Η πρώτη υποκατηγορία είναι οι χάρτες κίνησης, όπως το παράδειγμα που εμφανίζεται στο Σχήμα 48. Οι χάρτες κίνησης συλλέγουν πληροφορίες από χρήστες της εφαρμογής κινητών τηλεφώνων του Facebook, οι οποίοι έχουν ενεργοποιημένη τη λειτουργία ιστορικού τοποθεσίας (location history). Οι χάρτες αυτοί απεικονίζουν την κίνηση μεταξύ ζευγών σημείων στο χάρτη.

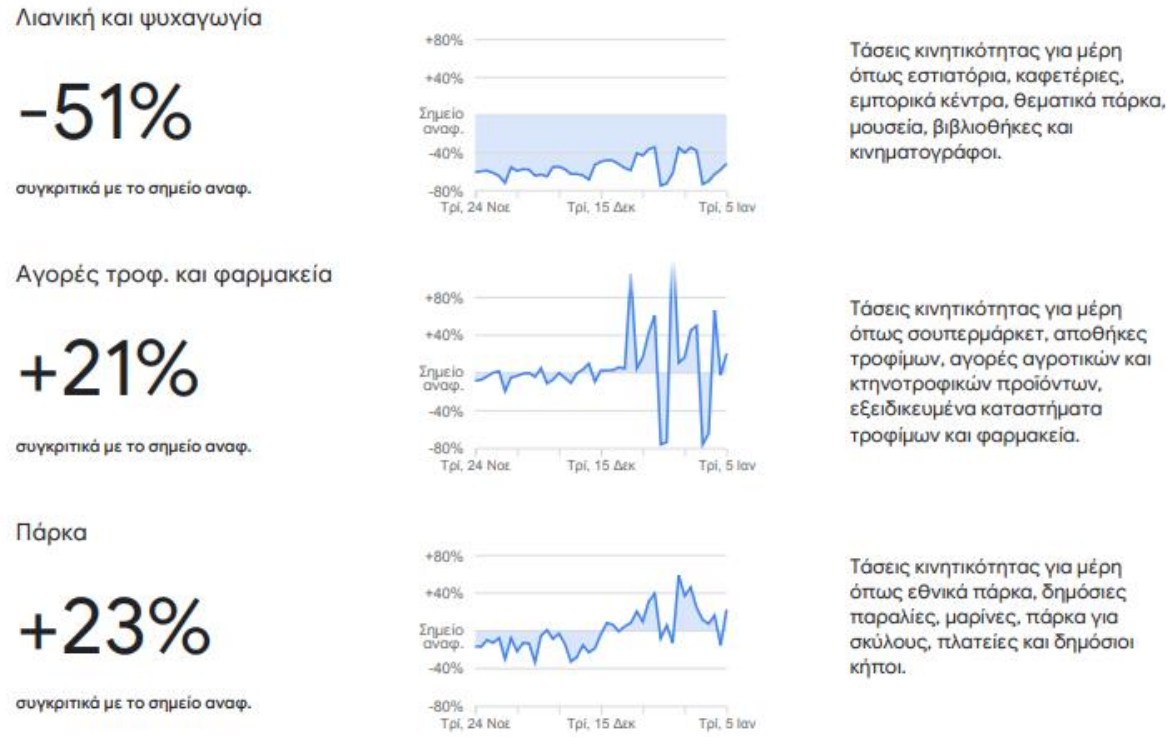


Σχήμα 48 *- Χάρτης κίνησης του Facebook που δείχνει την ημερήσια κίνηση ανθρώπων μεταξύ του κεντρικού Λονδίνου και των γύρω περιοχών* [184]

Η δεύτερη κατηγορία χαρτών κίνησης είναι οι Χάρτες Μοτίβων Κίνησης (Travel Patterns Maps), οι οποίοι απεικονίζουν συγκρίσεις του αριθμού των χρηστών του Facebook που μετακινούνται σε μεγάλες αποστάσεις, για παράδειγμα μέσω αεροπλάνου ή τρένου. Και αυτά τα δεδομένα βασίζονται μόνο σε χρήστες που έχουν επιλέξει να μοιράζονται ακριβή δεδομένα τοποθεσίας από την κινητή τους συσκευή μέσω της εφαρμογής Facebook. Επί του παρόντος, στους χάρτες αυτούς περιλαμβάνονται μόνο διεθνείς μετακινήσεις και οι αντίστοιχες μετρήσεις ενημερώνονται καθημερινά. Τέλος, η τρίτη κατηγορία χαρτών κίνησης ονομάζεται Colocation Maps. Οι χάρτες αυτοί δημιουργήθηκαν για την πανδημία COVID-19 και μετρούν την πιθανότητα οι άνθρωποι σε μια περιοχή να έρθουν σε επαφή με ανθρώπους από μια άλλη, υποδεικνύοντας έτσι τις επόμενες υποψήφιες περιοχές όπου μπορεί να εμφανιστούν νέα κρούσματα COVID-19.

Ένα παρόμοιο σύνολο δεδομένων διατίθεται από την Google μέσω των Αναφορών Κινητικότητας Κοινότητας (Community Mobility Reports) [186], με στόχο να χρησιμοποιηθεί από τους υπευθύνους στους τομείς της υγείας αλλά και από άλλους ενδιαφερόμενους στη λήψη κρίσιμων αποφάσεων για την καταπολέμηση της πανδημίας COVID-19. Αυτές οι αναφορές δημιουργούνται μέσω συγκέντρωσης ανώνυμων δεδομένων χρηστών που έχουν ενεργοποιημένη τη λειτουργία ιστορικού τοποθεσίας. Αποτυπώνουν τις τάσεις μετακίνησης των ανθρώπων με την πάροδο του χρόνου, ανά

γεωγραφική περιοχή, σε διάφορες κατηγορίες τοποθεσιών, όπως για παράδειγμα καταστήματα λιανικού εμπορίου, χώρους αναψυχής, σουπερμάρκετ, φαρμακεία, πάρκα, χώρους εργασίας και κατοικίες. Τα δεδομένα ενημερώνονται καθημερινά και είναι διαθέσιμα για λήψη σε μορφή PDF και CSV. Οι αναφορές δημοσιεύονται και για την Ελλάδα και περιέχουν διαφορετικές ενότητες, μία για κάθε γεωγραφική περιφέρεια. Ένα παράδειγμα αναφοράς για την περιφέρεια Αττικής εμφανίζεται στο Σχήμα 49.



Σχήμα 49 *- Στιγμιότυπο από τις Αναφορές Κινητικότητας της Google για την Ελλάδα και συγκεκριμένα για την περιφέρεια Αττικής* [186]

ΚΕΦΑΛΑΙΟ 7: Προκλήσεις, Προτάσεις και Συμπεράσματα

7.1 Εισαγωγή

Αρχικά, στο Κεφάλαιο αυτό θα εξετάσουμε τις διάφορες προκλήσεις στην ανάπτυξη εφαρμογών τεχνητής νοημοσύνης για την αντιμετώπιση της πανδημίας COVID-19, οι οποίες ενδέχεται να σταθούν εμπόδιο στην ποιότητα, την αποτελεσματικότητα και την ορθή χρήση των εφαρμογών αυτών. Αν και οι προκλήσεις αυτές είναι – σε μικρότερο ή μεγαλύτερο βαθμό – παρούσες σε όλους τους τύπους εφαρμογών που παρουσιάστηκαν στα προηγούμενα Κεφάλαια, θα αναφέρουμε συγκεκριμένα παραδείγματα στα οποία μπορεί να προκύψουν.

Το γεγονός ότι οι λύσεις τεχνητής νοημοσύνης που μελετήθηκαν σε αυτή την εργασία στοχεύουν στην καταπολέμηση μιας πανδημίας, καθιστά ορισμένες από τις παραδοσιακές προκλήσεις που υπάρχουν σε όλο το φάσμα των εφαρμογών τεχνητής νοημοσύνης ακόμα εντονότερες. Αφού πρώτα περιγράψουμε τις προκλήσεις αυτές, θα παρουσιάσουμε στη συνέχεια τις προτάσεις μας με τη μορφή επτά κατευθυντήριων γραμμών που αποσκοπούν στην αντιμετώπισή τους, μέσω της κατάλληλης σχεδίασης και υλοποίησης εύρωστων (robust) εφαρμογών τεχνητής νοημοσύνης. Τέλος, ολοκληρώνουμε την εργασία αυτή παρουσιάζοντας τα συμπεράσματά μας.

7.2 Προκλήσεις

7.2.1 Ποσότητα Δεδομένων, Ποιότητα Δεδομένων και Μεροληψία (Bias)

Η μηχανική μάθηση και ιδιαίτερα τα συστήματα βαθιάς μάθησης εκπαιδεύονται και βελτιώνονται με την πάροδο του χρόνου καθώς τροφοδοτούνται με περισσότερα δεδομένα. Στην ιδανική περίπτωση, τα δεδομένα θα πρέπει να είναι υψηλής ποιότητας και μεγάλου όγκου. Δυστυχώς, για πολλές από τις εφαρμογές που περιγράφονται στα προηγούμενα Κεφάλαια, δεν υπάρχουν ακόμη εκτεταμένα σύνολα δεδομένων. Για παράδειγμα, αν και υπάρχουν κάποια διαθέσιμα σύνολα δεδομένων με ιατρικές εικόνες, αυτά εξακολουθούν να είναι σχετικά μικρά συγκρινόμενα με τις απαιτήσεις των μοντέλων βαθιάς μάθησης.

Σε ορισμένες περιπτώσεις, η μικρή ποσότητα των δεδομένων οφείλεται στον κατανεμημένο χαρακτήρα πολλών πηγών, γεγονός που καθιστά δύσκολη τη συλλογή και τη συγκέντρωση / ενοποίησή τους. Παραδείγματος χάριν, τα αρχεία ηλεκτρονικού φακέλου ασθενούς είναι συχνά απομονωμένα σε επίπεδο χώρας, σε περιφερειακό επίπεδο ή ακόμα και σε επίπεδο μεμονωμένου νοσοκομείου. Επίσης, η περιορισμένη διαθεσιμότητα των δεδομένων οφείλεται πολλές φορές στο γεγονός ότι κάποιες πηγές δε μοιράζονται τα δεδομένα τους, είτε για λόγους σχετικούς με την προστασία των προσωπικών δεδομένων είτε για άλλους λόγους εμπορικής φύσεως.

Πέρα από το ζήτημα της ποσότητας, τα δεδομένα για την εκπαίδευση των συστημάτων μηχανικής μάθησης πρέπει να τηρούν και κατάλληλα επίπεδα ποιότητας. Η μία διάσταση της ποιότητας έχει να κάνει με τη μορφή (format) των δεδομένων, τα οποία σε πολλές περιπτώσεις μπορεί να διαφέρουν μεταξύ νοσοκομείων, περιφερειών και χωρών. Για παράδειγμα, στις εφαρμογές επιδημιολογικής μοντελοποίησης με τη χρήση μεθόδων τεχνητής νοημοσύνης, όπως αυτές που περιγράφονται στο Κεφάλαιο 2, παρατηρήσαμε ότι τα δεδομένα που συλλέγονται από διαφορετικές χώρες είναι πολλές φορές δύσκολο να συγκριθούν λόγω πολλαπλών παραγόντων, που περιλαμβάνουν διαφορετική στρατηγική και συχνότητα κλινικών δοκιμών (τεστ) ανά χώρα, καθώς και διαφορετικές μεθόδους αναφοράς και παρακολούθησης των κρουσμάτων. Επιπλέον, η εφαρμογή επιδημιολογικών μοντέλων εκπαιδευμένων σε συγκεκριμένο περιβάλλον για την πρόβλεψη της εξάπλωσης σε ένα

άλλο περιβάλλον, εγείρει ανησυχίες σε σχέση με την ικανότητα των μοντέλων αυτών να αποτυπώσουν πτυχές όπως διαφορετικά κοινωνικά, γεωγραφικά ή άλλα χαρακτηριστικά που μπορεί να επηρεάζουν την ταχύτητα εξάπλωσης του ιού.

Εκτός όμως από τη μορφή και την τυποποίηση, μια σημαντικότερη πτυχή των δεδομένων που χρησιμοποιούνται για την εκπαίδευση τέτοιων μοντέλων μηχανικής μάθησης έχει να κάνει με το πόσο αντιπροσωπευτικά είναι τα δεδομένα αυτά σε σχέση με τον πληθυσμό-στόχο, στον οποίο τελικά το μοντέλο θα κληθεί να εφαρμοστεί για να εξάγει προβλέψεις. Αυτό είναι ένα πρόβλημα που ονομάζεται αλγοριθμική μεροληψία (algorithmic bias) στην ορολογία της μηχανικής μάθησης. Έχει αναγνωριστεί ότι η έλλειψη ποικιλομορφίας στα δεδομένα και η υποεκπροσώπηση ή υπερεκπροσώπηση συγκεκριμένων ομάδων και πληθυσμών μπορεί να οδηγήσουν τα συστήματα, που έχουν εκπαιδευτεί με βάση αυτά τα δεδομένα, να δρουν μεροληπτικά και να παράγουν άνισα αποτελέσματα για ορισμένους υποπληθυσμούς.

Γενικότερα, τα δεδομένα από διαφορετικά κλινικά περιβάλλοντα μπορεί να περιέχουν διάφορους τύπους μεροληψίας και θορύβου, γεγονός που μπορεί, για παράδειγμα, να σημαίνει ότι ένα μοντέλο που έχει εκπαιδευτεί με βάση τα δεδομένα ενός νοσοκομείου, να αποτύχει να γενικεύσει τα συμπεράσματά του σε ένα δεύτερο νοσοκομείο. Όπως προκύπτει από τα παραπάνω, οι αλγόριθμοι τεχνητής νοημοσύνης υπόκεινται στο γνωστό κανόνα «garbage-in-garbage-out». Από τη φύση τους, οι αλγόριθμοι μηχανικής μάθησης έχουν σχεδιαστεί για να εντοπίζουν κρυμμένα μοτίβα (patterns) των δεδομένων και να δημιουργούν προβλέψεις εξόδου με βάση αυτά που έχουν μάθει στο παρελθόν. Καθώς πολλά σύνολα δεδομένων εισόδου περιέχουν χαρακτηριστικά ή μεροληπτούν υπέρ συγκεκριμένων χαρακτηριστικών, τα μοντέλα που εκπαιδεύτηκαν από τα δεδομένα αυτά «κουβαλούν» αυτά τα χαρακτηριστικά και τις ανισότητες και τείνουν να τα ενισχύουν.

Μια πρόσφατη μελέτη πραγματοποίησε ανασκόπηση της υπάρχουσας επιστημονικής έρευνας στον τομέα των εφαρμογών τεχνητής νοημοσύνης για την αντιμετώπιση της πανδημίας COVID-19 και συγκεκριμένα στις περιοχές της επιδημιολογικής πρόβλεψης, της διάγνωσης και της πρόγνωσης [187]. Η μελέτη αυτή αποκάλυψε ότι τα μοντέλα, στις περισσότερες από τις επιστημονικές εργασίες που εξετάστηκαν, διέτρεχαν υψηλό κίνδυνο μεροληψίας, γεγονός που υποδηλώνει ότι η προγνωστική τους απόδοση, όταν εφαρμοστούν σε πραγματικές συνθήκες, θα είναι πιθανώς χαμηλότερη από αυτήν που αναφέρουν οι δημιουργοί τους. Το συμπέρασμα αυτό προέκυψε από το γεγονός ότι οι συμμετέχοντες, των οποίων τα δεδομένα χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων, φάνηκε να μην αποτελούν αντιπροσωπευτικό δείγμα του πραγματικού πληθυσμού, καθώς οι περισσότερες από τις μελέτες χρησιμοποίησαν μικρά, ενδεχομένως μεροληπτικά (biased) δείγματα κυρίως από την Κίνα. Άλλες ελλείψεις που εντοπίστηκαν ήταν η ασαφής αναφορά σε σχέση με τα κριτήρια ένταξης των συμμετεχόντων, καθώς και η έλλειψη πληροφοριών σε σχέση με τα στάδια προεπεξεργασίας των δεδομένων (π.χ. των ιατρικών εικόνων). Το μικρό μέγεθος των δειγμάτων είναι επίσης ένας παράγοντας που μπορεί να οδηγήσει σε αυξημένο κίνδυνο υπερβολικής προσαρμογής (overfitting) του μοντέλου[13].

Τέλος, μια άλλη σημαντική πτυχή της ποιότητας των δεδομένων, κυρίως στην επιτηρούμενη μηχανική μάθηση, είναι η ακριβής αξιολόγηση και επισήμανση των δεδομένων εκπαίδευσης που χρησιμοποιούνται ως αντικειμενική αλήθεια (ground truth). Η ακρίβεια των προβλέψεων ενός

[13] Η υπερβολική προσαρμογή (overfitting) συμβαίνει όταν ένα μοντέλο μαθαίνει πολύ καλά τις λεπτομέρειες και το θόρυβο στα δεδομένα εκπαίδευσης, σε βαθμό που επιδρά αρνητικά στην απόδοσή του σε νέα δεδομένα. Αυτό ουσιαστικά σημαίνει ότι ο θόρυβος ή οι τυχαίες διακυμάνσεις στα δεδομένα της εκπαίδευσης συλλέγονται και «μαθαίνονται» ως σημαντική πληροφορία από το μοντέλο. Το πρόβλημα είναι ότι αυτή η πληροφορία δε συναντάται σε νέα δεδομένα και το μοντέλο αποτυγχάνει να γενικεύσει τα συμπεράσματά του.

μοντέλου επιτηρούμενης μάθησης βασίζεται σε μεγάλο βαθμό στην ακρίβεια και την αξιοπιστία των δεδομένων εκπαίδευσης που εισάγονται στον αλγόριθμο. Τα δεδομένα με κακή ή λανθασμένη επισήμανση (labelling) θα αποφέρουν ανεπαρκή αποτελέσματα. Από τα πιο πάνω, συμπεραίνουμε ότι είναι πρωταρχικής σημασίας για τη διασφάλιση της ακρίβειας ενός μοντέλου να υπάρχει διαφάνεια στη διαδικασία της επισήμανσης των δεδομένων εκπαίδευσης.

7.2.2 Μοντέλα «Μαύρου Κουτιού»

Αν και τα μοντέλα μηχανικής μάθησης μπορούν – σε πολλές περιπτώσεις – να επιτύχουν καλύτερη απόδοση από τους ανθρώπους σε διάφορα προβλήματα, εντούτοις πολλές φορές δε μπορούμε να κατανοήσουμε ή να ερμηνεύσουμε πώς ένα δεδομένο μοντέλο κατέληξε σε μια συγκεκριμένη απόφαση, διάγνωση ή πρόβλεψη. Αυτά τα μοντέλα χαρακτηρίζονται πολλές φορές ως «μαύρα κουτιά» (black boxes), για να περιγράψουν μια διαδικασία η οποία καταλήγει σε κάποιο αποτέλεσμα πρόβλεψης χωρίς εμείς να μπορούμε να εξηγήσουμε με ποιο τρόπο κατέληξε σε αυτό [188]. Είναι προφανές ότι χωρίς διαφάνεια και «επεξηγησιμότητα» (explainability) είναι πολύ δύσκολο έως αδύνατο να εντοπιστούν οι αδυναμίες ενός μοντέλου έτσι ώστε να μπορούν να διορθωθούν.

Χαρακτηριστικά παραδείγματα που ενδέχεται να υπόκεινται σε αυτή την κατηγορία αφορούν στις εφαρμογές διάγνωσης με χρήση ιατρικής εικόνας, όπως μελετήθηκαν στην ενότητα 3.2. Αν και τα περισσότερα από τα μοντέλα αυτά παρουσιάζουν φαινομενικά μεγάλη ακρίβεια στη διάκριση ασθένειας COVID-19 από άλλους τύπους ιογενούς πνευμονίας μέσω χρήσης ακτινολογικών εικόνων, τα περισσότερα από αυτά στερούνται διαφάνειας και επεξηγησιμότητας. Αυτό σημαίνει ουσιαστικά ότι, στις περισσότερες από αυτές τις μελέτες, οι ερευνητές δεν κατάφεραν να εξηγήσουν ποια χαρακτηριστικά της απεικόνισης συνεισέφεραν περισσότερο στις αποφάσεις του μοντέλου. Μια ακόμα συνέπεια της ύπαρξης μοντέλων με τη μορφή μαύρων κουτιών είναι ότι πολύ δύσκολα θα τα εμπιστευθούν οι κλινικοί ιατροί για να βασίσουν σε αυτά αποφάσεις που μπορεί να είναι καθοριστικές για την περαιτέρω πορεία της εξέλιξης ενός ασθενούς.

Ένας άλλος λόγος για τον οποίο η διαφάνεια και η επεξηγησιμότητα ενός μοντέλου είναι σημαντικές οφείλεται στο γεγονός ότι οι τεχνολογίες τεχνητής νοημοσύνης υποφέρουν πολλές φορές από αλγοριθμική μεροληψία, όπως περιγράψαμε στην προηγούμενη ενότητα. Ακόμα και αν τα δεδομένα εκπαίδευσης είναι διαφανή και υψηλής ποιότητας, ένα μοντέλο τύπου μαύρου κουτιού δε θα επέτρεπε την εξέταση και την κατανόηση πιθανής μεροληψίας, ούτε την ικανότητα λήψης των κατάλληλων διορθωτικών μέτρων για να καταστεί το μοντέλο λιγότερο μεροληπτικό.

Αξίζει να σημειωθεί ότι η διαφάνεια και η επεξηγησιμότητα είναι κάποιες φορές δύσκολο να επιτευχθούν εάν οι εταιρείες που παράγουν μοντέλα μηχανικής μάθησης καθιστούν σκόπιμα τους αλγορίθμους τους αδιαφανείς για λόγους προστασίας της πνευματικής τους ιδιοκτησίας, εμπορικούς ή άλλους λόγους. Για παράδειγμα, ορισμένες από τις εμπορικά διαθέσιμες λύσεις που μελετήσαμε στα προηγούμενα Κεφάλαια δεν αποκαλύπτουν τις λεπτομέρειες των αλγορίθμων τους.

7.2.3 Έκτακτες Συνθήκες και Μεταβαλλόμενες Ανάγκες

Οι προσπάθειες για τη διαχείριση και την αντιμετώπιση της πανδημίας COVID-19 έχουν περάσει από διάφορα στάδια από την εμφάνιση της νέας ασθένειας μέχρι και σήμερα. Κάθε στάδιο διαφέρει ως προς το μετρούμενο και πραγματικό βαθμό εξάπλωσης της νόσου, που επηρεάζει άμεσα τη διαθεσιμότητα κλινικών πόρων, η οποία με τη σειρά της μπορεί σε διάστημα μερικών εβδομάδων να αλλάξει τις ανάγκες και τις αντίστοιχες προτεραιότητες. Ανάλογα με τη φάση της πανδημίας, οι προτεραιότητες μπορεί να μεταβάλλονται ανάμεσα στην κατανόηση του τρόπου εξάπλωσης, στην παροχή αξιόπιστων διαγνωστικών μεθόδων, στον έλεγχο και τον περιορισμό της καμπύλης

μετάδοσης, καθώς και στη διασφάλιση επαρκούς διαθεσιμότητας υγειονομικών πόρων. Αυτές οι γρήγορες αλλαγές, που συμβαίνουν ταυτόχρονα με τη σταδιακά αυξανόμενη γνώση της συμπεριφοράς του νέου ιού και την διαρκώς αυξανόμενη διαθεσιμότητα των αντίστοιχων δεδομένων, έχουν ως αποτέλεσμα την ανάγκη για συνεχή εξέλιξη και προσαρμογή των μοντέλων τεχνητής νοημοσύνης.

Είναι επίσης φανερό ότι οι εφαρμογές και τα μοντέλα τεχνητής νοημοσύνης που κρίνονται ως πιο επείγοντα και πιο αναγκαία διαφέρουν σε κάθε στάδιο της πανδημίας. Αυτό δημιουργεί πολλές δυσκολίες στη δημιουργία τέτοιων μοντέλων, καθώς τα δεδομένα που απαιτούνται για την εκπαίδευσή τους είναι συνήθως σε έλλειψη. Ας πάρουμε το παράδειγμα των διαγνωστικών εφαρμογών. Στα πρώτα στάδια της πανδημίας, ένα εξαιρετικά ευαίσθητο και ακριβές μοντέλο διάγνωσης με βάση ιατρικές εικόνες θα ήταν πολύ χρήσιμο, δεδομένης και της μικρής διαθεσιμότητας εργαστηριακών kit δοκιμών. Ταυτόχρονα όμως, η ελλιπής ποσότητα διαθέσιμων δεδομένων ιατρικής εικόνας από ασθενείς στη συγκεκριμένη φάση καθιστούσε ιδιαίτερα δύσκολη τη δημιουργία αξιόπιστων νέων μοντέλων διάγνωσης.

Ακόμα, η επείγουσα κατάσταση που δημιουργεί η κρίση μιας πανδημίας όπως η πανδημία COVID-19, απαιτεί άμεσες λύσεις. Αυτό σημαίνει πολλές φορές ότι ορισμένα άλλα επιθυμητά χαρακτηριστικά μιας εφαρμογής τεχνητής νοημοσύνης μπορεί να τεθούν σε δεύτερη μοίρα, λόγω έλλειψης χρόνου. Ωστόσο, εάν μια προτεινόμενη λύση δεν πληροί ορισμένα βασικά χαρακτηριστικά, μια πρόωρη διάθεσή της μπορεί να έχει τα αντίθετα αποτελέσματα από αυτά που προσδοκούν οι δημιουργοί της, με αποτέλεσμα σύντομα να χάσει την εμπιστοσύνη από τους δυνητικούς χρήστες της.

7.2.4 Προσωπικά Δεδομένα

Σε πολλές από τις λύσεις που περιγράφονται σε αυτή την εργασία, προσωπικά δεδομένα ασθενών ή τελικών χρηστών σχετικών εφαρμογών, για παράδειγμα δεδομένα που αφορούν την κινητικότητα, την τοποθεσία, κλινικές εξετάσεις και δημοσιεύσεις σε μέσα κοινωνικής δικτύωσης, χρησιμοποιούνται για την εκπαίδευση μοντέλων μηχανικής μάθησης. Λόγω της ευαίσθητης φύσης πολλών τέτοιων πληροφοριών, η προστασία της ιδιωτικής ζωής και των προσωπικών δεδομένων καθίσταται σημαντικό ζήτημα. Σε ορισμένες περιπτώσεις, οι ανησυχίες για την προστασία των προσωπικών δεδομένων μπορούν να επικρατήσουν έναντι του πιθανού οφέλους για τη δημόσια υγεία και τελικά, να οδηγήσουν σε περιορισμένη συλλογή και διάθεση δεδομένων για επιστημονικούς σκοπούς. Επιπλέον, υπάρχει ο φόβος, από ένα μεγάλο μέρος των ανθρώπων σε πολλές χώρες του κόσμου, ότι οι κυβερνήσεις θα εκμεταλλευθούν την ευρεία συλλογή προσωπικών δεδομένων που επιβλήθηκε για σκοπούς διασφάλισης της δημόσιας υγείας, συνεχίζοντας την παρακολούθηση των πολιτών τους για άλλους σκοπούς.

Οι εφαρμογές αυτοματοποιημένης ιχνηλάτησης επαφών, οι οποίες υιοθετήθηκαν από αρκετά κράτη, είναι μία από τις περιπτώσεις εφαρμογών που απαιτούν αναλυτική εξέταση των ζητημάτων προστασίας προσωπικών δεδομένων. Όπως μελετήσαμε στην ενότητα 2.3, οι εφαρμογές αυτές, οι οποίες εγκαθίστανται στα κινητά τηλέφωνα των χρηστών, παρακολουθούν τις αλληλεπιδράσεις μεταξύ χρηστών μέσω του πρωτοκόλλου επικοινωνίας Bluetooth. Ο πληθυσμός των χρηστών που διατρέχουν κίνδυνο μόλυνσης προσδιορίζεται μέσω των πρόσφατων επαφών τους με χρήστες που έχουν διαγνωσθεί θετικοί στην ασθένεια COVID-19. Εάν δε σχεδιαστούν σωστά, τέτοιες εφαρμογές ιχνηλάτησης ενδέχεται να χρησιμοποιηθούν για γενικότερη παρακολούθηση των πολιτών, ιδίως σε περιπτώσεις ορισμένων κεντρικών αρχιτεκτονικών, στις οποίες όλα τα δεδομένα των χρηστών και των επαφών τους βρίσκονται σε μια κεντρική βάση δεδομένων.

Εκτός από την περίπτωση των εφαρμογών ιχνηλάτησης επαφών, υπάρχουν και άλλες εφαρμογές που συλλέγουν δεδομένα κινητικότητας χρηστών, όπως για παράδειγμα αρχεία κλήσεων μέσω δικτύων κινητής τηλεφωνίας καθώς και τοποθεσίες χρηστών από διάφορα μέσα κοινωνικής δικτύωσης. Αν η συλλογή και η επεξεργασία των εν λόγω δεδομένων δεν προορίζεται μόνο για τον αρχικό τους σκοπό, ενδέχεται να δημιουργηθούν σημαντικά ζητήματα προστασίας των προσωπικών δεδομένων.

7.2.5 Ζητήματα Ηθικής

Ο ρόλος της ηθικής στην τεχνητή νοημοσύνη έχει απασχολήσει και εξακολουθεί να απασχολεί έντονα την ερευνητική κοινότητα. Σε ένα πολύ ενδιαφέρον άρθρο [189], οι ερευνητές προτείνουν τις εξής πέντε βασικές αρχές για τη θεμελίωση της ηθικής στην τεχνητή νοημοσύνη: beneficence (να είναι επωφελής), non-maleficence (να μη βλάπτει), autonomy (να είναι αυτόνομη και να μη χειραγωγείται), justice (να είναι δίκαιη) και explicability (να μπορεί να εξηγηθεί ο τρόπος αποφάσεών της και να είναι υπεύθυνη για τις αποφάσεις της)[14]. Όλες αυτές οι γενικές αρχές θα πρέπει προφανώς να ισχύουν και στην περίπτωση των εφαρμογών που αναπτύσσονται στα πλαίσια της αντιμετώπισης της πανδημίας COVID-19.

Ωστόσο, η περίπτωση χρήσης τεχνολογιών που βασίζονται στην τεχνητή νοημοσύνη για τη διαχείριση πανδημιών, όπως η πανδημία COVID-19, αποτελεί χαρακτηριστικό παράδειγμα του τρόπου με τον οποίο αυτές οι θεμελιώδεις αρχές ενδέχεται να έρθουν σε σύγκρουση. Υπάρχουν, δηλαδή, αρκετές ερωτήσεις ή διλήμματα που ο σχεδιαστής των αντίστοιχων εφαρμογών ή ακόμα και ο χρήστης τους (π.χ. ένας ιατρός που συμβουλεύεται ένα σύστημα τεχνητής νοημοσύνης για μια απόφασή του) θα πρέπει να απαντήσει. Παραδείγματα τέτοιων ερωτημάτων / διλημμάτων είναι τα εξής:

- Προκρίνουμε τη δημόσια υγεία (beneficence) έναντι της προστασίας προσωπικών δεδομένων (non-maleficence);
- Πώς αντιμετωπίζει το σύστημα τεχνητής νοημοσύνης τη λήψη αποφάσεων για ζητήματα όπως, «ποιος θα μπει σε μια μονάδα εντατικής θεραπείας;», δεδομένων των περιορισμένων κρεβατιών και λαμβάνοντας παράλληλα υπόψη τόσο την ανάγκη για δικαιοσύνη (justice) όσο και το να μη βλάψει άλλους ασθενείς (non-maleficence);
- Σε περίπτωση έκτακτων συνθηκών για τη δημιουργία εφαρμογών αντιμετώπισης της πανδημίας με χρήση της τεχνητής νοημοσύνης, πόσο σημαντική θεωρούμε την έννοια της επεξηγησιμότητας (explainability);

Αυτά είναι μερικά από τα πολύ δύσκολα ερωτήματα που πρέπει να εξεταστούν παράλληλα με τη συνεχή ανάπτυξη εφαρμογών και λύσεων τεχνητής νοημοσύνης που θα χρησιμοποιηθούν σε κρίσεις πανδημιών, όπως αυτή που ζούμε σήμερα. Και όπως προκύπτει από τα παραπάνω, υπάρχουν σημαντικά διλήμματα όταν πρέπει να αποφασίσουμε για ενδεχόμενους συμβιβασμούς μεταξύ αυτών των αρχών.

[14] Οι συντάκτες του άρθρου περιγράφουν τον όρο «explicability» ως μια σύνθετη έννοια που συνδυάζει την επιστημολογική έννοια του «intelligibility» (ως απάντηση στο ερώτημα «πώς λειτουργεί;») και την ηθική διάσταση του «accountability» (ως απάντηση στο ερώτημα «ποιος είναι υπεύθυνος για τον τρόπο με τον οποίο λειτουργεί;»).

7.3 Προτάσεις

7.3.1 Κοινή Χρήση Δεδομένων και Ανοικτά Δεδομένα

Όπως περιγράψαμε στην ενότητα 7.2.1, μία από τις κρίσιμες προκλήσεις για την ανάπτυξη αξιόπιστων λύσεων τεχνητής νοημοσύνης έναντι της πανδημίας COVID-19 είναι η έλλειψη τυποποιημένων και υψηλής ποιότητας δεδομένων. Για να αντιμετωπιστεί αυτή η πρόκληση, κυβερνήσεις, οργανισμοί και εταιρείες από όλο τον κόσμο έχουν ξεκινήσει διάφορες μορφές συνεργασίας που στοχεύουν στη δημιουργία και τη συντήρηση ανοικτών συνόλων δεδομένων διαφορετικών ειδών, όπως είδαμε στο Κεφάλαιο 6. Η δημόσια διάθεση και η ελεύθερη πρόσβαση σε δεδομένα επιτρέπει σε διάφορα είδη χρηστών, όπως ερευνητές, επιστήμονες, προγραμματιστές εφαρμογών, επιχειρηματίες και άλλους ενδιαφερομένους, να εκμεταλλευθούν τα δεδομένα με νέους και καινοτόμους τρόπους για την υλοποίηση νέων ιδεών μέσω των αντίστοιχων εφαρμογών.

Θεωρούμε ότι τέτοιες πρωτοβουλίες για την ανταλλαγή δεδομένων μέσω της δημιουργίας, φιλοξενίας, συντήρησης και αξιολόγησης ανοικτών αποθετηρίων (repositories) πρέπει να πληθύνουν και να ενισχυθούν περαιτέρω. Με αυτό τον τρόπο, οι ερευνητές, οι επιστήμονες και γενικότερα οι χρήστες τέτοιων δεδομένων θα είναι σε θέση να τα εκμεταλλευθούν και να συνεισφέρουν με τη σειρά τους πρόσθετα δεδομένα και εφαρμογές. Πιστεύουμε ότι η καλλιέργεια ενός οικοσυστήματος ανοικτών δεδομένων δεν αφορά απλά τη δημοσίευση των δεδομένων στο διαδίκτυο. Χρειάζονται κατάλληλες πολιτικές και κίνητρα για την κινητοποίηση και την ενεργό συμμετοχή και συνεισφορά όλων των διαφορετικών εμπλεκόμενων μερών.

Ένα παράδειγμα ενός χρήσιμου συνόλου ανοικτών δεδομένων, που έχει ήδη αναφερθεί, είναι τα παγκόσμια αποθετήρια με ανώνυμα κλινικά δεδομένα, τα οποία περιλαμβάνουν δεδομένα ιατρικής απεικόνισης καθώς και άλλα στοιχεία ιστορικού ασθενών. Τα δεδομένα αυτά παρουσιάζουν μεγάλο ενδιαφέρον σε ό,τι αφορά την ανταλλαγή γνώσης μεταξύ διαφορετικών κλινικών ιδρυμάτων (π.χ. νοσοκομείων). Για να διευκολυνθεί η ανταλλαγή τέτοιων δεδομένων, θα πρέπει να σχεδιαστούν κλινικά πρωτόκολλα και αρχιτεκτονικές διαμοιρασμού δεδομένων, καθώς επίσης και να εφαρμοστούν πλαίσια διακυβέρνησης των δεδομένων αυτών (data governance). Αξίζει να σημειωθεί ότι η κοινή χρήση κλινικών βάσεων δεδομένων απαιτεί σημαντική προσοχή για τη σωστή διαχείριση ρυθμιστικών θεμάτων καθώς και θεμάτων προστασίας προσωπικών δεδομένων, όπως θα εξηγήσουμε και στη συνέχεια. Η ταχεία και αποτελεσματική επίλυση αυτών των ζητημάτων μπορεί να είναι ιδιαίτερα δύσκολη, ιδίως κατά την περίοδο που πολλοί δημόσιοι οργανισμοί βρίσκονται υπό την πίεση άμεσων λύσεων.

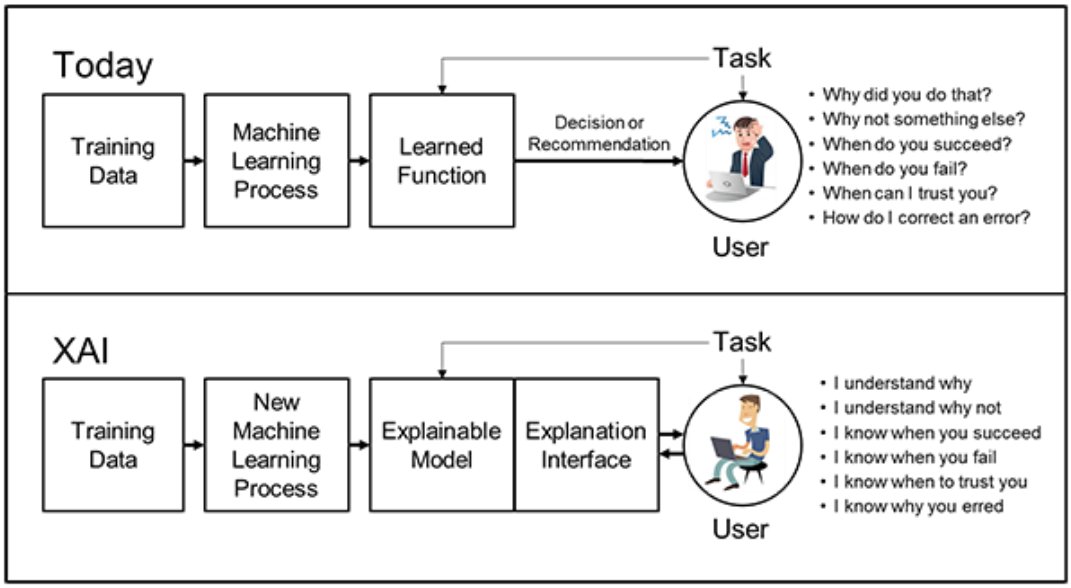
7.3.2 Επεξηγησιμότητα και Δυνατότητα Ερμηνείας των Μοντέλων Τεχνητής Νοημοσύνης

Όπως αναφέρθηκε και προηγουμένως, η έννοια της επεξηγησιμότητας (explainability) είναι εκείνο το χαρακτηριστικό ενός συστήματος τεχνητής νοημοσύνης που επιτρέπει σε ένα άτομο να κατανοήσει το γιατί και με ποιον τρόπο ένα συγκεκριμένο μοντέλο τεχνητής νοημοσύνης κατέληξε σε συγκεκριμένες προβλέψεις. Μέσω επεξηγήσιμων μοντέλων, οι εμπλεκόμενοι χρήστες θα μπορέσουν να κατανοήσουν τα πιθανά οφέλη, τους περιορισμούς ενός μοντέλου και – στο μέτρο του δυνατού – τον τρόπο λειτουργίας του, δηλαδή ποια είναι εκείνα τα χαρακτηριστικά των δεδομένων εισόδου που συμβάλλουν περισσότερο ή λιγότερο στην έξοδο του μοντέλου [190].

Ας πάρουμε ως παράδειγμα τη διάγνωση ασθενών με COVID-19, χρησιμοποιώντας ιατρική απεικόνιση και βαθιά νευρωνικά δίκτυα. Θεωρούμε ότι απαιτείται επιπλέον έρευνα σε σχέση με μεθόδους επεξηγήσιμης τεχνητής νοημοσύνης, που θα έχουν στόχο να εξηγήσουν την απόδοση των μοντέλων βαθιάς μάθησης. Συγκεκριμένα, οι μέθοδοι αυτές θα πρέπει να αναδείξουν τα

χαρακτηριστικά των εικόνων που συμβάλλουν περισσότερο στη διάκριση μεταξύ ασθενών με COVID-19, ασθενών με άλλους τύπους πνευμονίας ή υγιών ασθενών. Αυτό θα βοηθήσει ακτινολόγους και κλινικούς γιατρούς να αποκτήσουν καλύτερη γνώση σχετικά με τον ιό, να εμπιστευθούν περισσότερο τα μοντέλα μηχανικής μάθησης και τελικά να μπορέσουν να διαγνώσουν πιθανή ασθένεια πολύ πιο αποτελεσματικά.

Ένας τρόπος για την επίτευξη επεξηγησιμότητας στα συστήματα τεχνητής νοημοσύνης είναι η χρήση αλγορίθμων μηχανικής μάθησης που είναι εγγενώς επεξηγήσιμοι. Για παράδειγμα, κάποιες σχετικά απλές μορφές μηχανικής μάθησης, όπως τα δέντρα αποφάσεων (decision trees), οι bayesian ταξινομητές (classifiers) και άλλοι αλγόριθμοι που έχουν συγκεκριμένες ιδιότητες ιχνηλασιμότητας και διαφάνειας στη λήψη των αποφάσεών τους, μπορούν να παρέχουν τη διαφάνεια και τη δυνατότητα εποπτείας που απαιτείται, χωρίς να υστερούν υπερβολικά σε απόδοση ή ακρίβεια. Ωστόσο, πιο περίπλοκοι αλλά και δυνητικά πιο ισχυροί αλγόριθμοι, όπως τα βαθιά νευρωνικά δίκτυα, θυσιάζουν συχνά τη διαφάνεια και την επεξηγησιμότητα με αντάλλαγμα μεγαλύτερη απόδοση και ακρίβεια. Σημαντική έρευνα που βρίσκεται σήμερα σε εξέλιξη εστιάζει σε νέες μεθόδους προσθήκης επεξηγησιμότητας σε νευρωνικά δίκτυα και μοντέλα βαθιάς μάθησης [191]. Οι προκλήσεις ενός παραδοσιακού συστήματος τεχνητής νοημοσύνης σε αντιπαραβολή με τα οφέλη της Επεξηγήσιμης Τεχνητής Νοημοσύνης (Explainable Artificial Intelligence - XAI) παρουσιάζονται στο Σχήμα 50.



Σχήμα 50 *- Η έννοια της Επεξηγήσιμης Τεχνητής Νοημοσύνης (XAI)* [192]

Συνειδητοποιώντας την ανάγκη για βελτίωση της επεξηγησιμότητας σε μοντέλα βαθιάς μάθησης, καθώς και σε άλλες σύνθετες αλγοριθμικές προσεγγίσεις, η αμερικανική υπηρεσία DARPA (Defense Advanced Research Projects Agency) χρηματοδοτεί πολλές ερευνητικές προσπάθειες και αντίστοιχες πρωτοβουλίες [192].

7.3.3 Ηθική με Διαφάνεια και Αίσθηση του Επείγοντος

Όπως περιγράψαμε στις ενότητες 7.2.4 και 7.2.5, υπάρχουν διάφορες δεοντολογικές πτυχές και ζητήματα προστασίας προσωπικών δεδομένων, καθώς και κρίσιμα ερωτήματα και διλήμματα που σχετίζονται με την υιοθέτηση εφαρμογών τεχνητής νοημοσύνης για την καταπολέμηση της πανδημίας COVID-19. Σημαντικό κομμάτι έρευνας έχει επικεντρωθεί στην εύρεση βέλτιστων τρόπων

αντιμετώπισης αυτών των θεμάτων. Επίσης, στο χώρο της Ευρωπαϊκής Ένωσης, το Ευρωκοινοβούλιο συνέστησε να επιλέγονται – όπου είναι εφικτό – οι «εναλλακτικές λύσεις που παραβιάζουν λιγότερο την ελευθερία» (least liberty-infringing alternatives), προκειμένου να εξισορροπηθούν όσο το δυνατόν καλύτερα οι απαιτήσεις των αντικρουόμενων ηθικών αρχών που σχετίζονται με την τεχνητή νοημοσύνη [193].

Σε μια πολύ ενδιαφέρουσα μελέτη για την ηθική των εφαρμογών τεχνητής νοημοσύνης στην εποχή της πανδημίας COVID-19 [194], οι ερευνητές προτείνουν τρόπους για την υιοθέτηση ηθικής σε συνθήκες έκτακτων καταστάσεων μέσω της ύπαρξης τριών θεμελιωδών αρχών: (α) εισαγωγή της ηθικής από τη φάση της σχεδίασης (ethics by design), (β) εκτέλεση αντιπροσωπευτικών δοκιμών για τη διασφάλιση της ορθής λειτουργίας και της ασφάλειας των συστημάτων και (γ) οικοδόμηση εμπιστοσύνης των χρηστών μέσω ανεξάρτητης εποπτείας και αξιολόγησης.

Η πρώτη αρχή (ethics by design) περιλαμβάνει την ανάλυση πιθανών ηθικών ζητημάτων και κινδύνων όσο το δυνατόν πιο διεξοδικά κατά τη φάση της σχεδίασης, δηλαδή πριν από την ανάπτυξη και την παραγωγική λειτουργία συστημάτων σε πραγματικές συνθήκες. Για παράδειγμα, ερωτήματα όπως «ποια δεδομένα χρειαζόμαστε και ποια ζητήματα μπορεί να εγείρει η συλλογή τους;» πρέπει να απαντηθούν όσο το δυνατόν νωρίτερα. Αυτό σημαίνει ότι ειδικοί σε θέματα ηθικής και αξιολόγησης κινδύνων πρέπει να αποτελούν μέρος των ομάδων που σχεδιάζουν λύσεις τεχνητής νοημοσύνης από την αρχή, και να θέτουν πολύ σαφείς κατευθυντήριες γραμμές που θα ακολουθηθούν από τους μηχανικούς και τους προγραμματιστές κατά τη φάση της ανάπτυξης.

Η δεύτερη αρχή αφορά στην εξεύρεση τρόπων για τη διεξαγωγή ρεαλιστικών δοκιμών και την επαλήθευση των συστημάτων, με στόχο να διασφαλιστεί ότι συμπεριφέρονται κατά τρόπο που συνάδει με τις αρχές σχεδιασμού, οι οποίες έχουν τεθεί κατά τη φάση της σχεδίασης. Τέλος, η τρίτη αρχή, που αφορά στην οικοδόμηση της εμπιστοσύνης των χρηστών, μπορεί να επιτευχθεί αν οι εφαρμογές που δημιουργούνται είναι διαφανείς και ανοικτές σε δημόσιο έλεγχο και κριτική. Ο στόχος αυτός μπορεί να υποβοηθηθεί από ανεξάρτητα εποπτικά όργανα, τα οποία θα είναι αρμόδια για την επανεξέταση τυχόν κινδύνων και ηθικών ζητημάτων που προκύπτουν από τη χρήση των νέων αυτών λύσεων, καθώς και για την κατάρτιση δημόσια διαθέσιμων εκθέσεων αξιολόγησης.

Εξετάζοντας το παράδειγμα της κατανομής περιορισμένων υγειονομικών πόρων κατά τη διάρκεια μιας κρίσης πανδημίας, το οποίο παρουσιάστηκε στην ενότητα 3.4.2, παρατηρούμε ότι τίθενται σημαντικά διλήμματα ηθικής σε σχέση με τη δίκαιη κατανομή τους στους ασθενείς, ανεξαρτήτως αν οι αποφάσεις λαμβάνονται από ανθρώπους ή προτείνονται από εφαρμογές τεχνητής νοημοσύνης. Όπως περιγράφεται στις κατευθυντήριες γραμμές του Π.Ο.Υ., οι αρχές της χρησιμότητας (utility) και της ισότητας (equality) πρέπει να ληφθούν υπόψη και να εξισορροπηθούν για να μεγιστοποιηθούν τα συνολικά οφέλη, διατηρώντας παράλληλα το αίσθημα της δικαιοσύνης (justice) [195].

Υπάρχουν πάντα συμβιβασμοί που θα πρέπει να γίνουν, αλλά αυτοί οι συμβιβασμοί πρέπει να είναι πλήρως διαφανείς. Επίσης, είναι ιδιαίτερα σημαντικό οι αποφάσεις σχετικά με τη μεταχείριση και την κατανομή των πόρων που βασίζονται στην τεχνητή νοημοσύνη να έχουν κάποια μορφή επεξήγησης, προκειμένου να δημιουργηθεί εμπιστοσύνη στα αντίστοιχα εργαλεία και να μπορούν να τροποποιηθούν αν κριθεί απαραίτητο.

7.3.4 Αρχιτεκτονικές που Παρέχουν Ασφάλεια και Προστασία των Προσωπικών Δεδομένων

Η κοινή χρήση ιατρικών και άλλων προσωπικών δεδομένων με ασφαλή τρόπο και σεβασμό στην προστασία της ιδιωτικότητας για σκοπούς ιατρικής ανάλυσης είναι ένα ζήτημα που προϋπήρχε της πανδημίας COVID-19. Ένα παράδειγμα σχετικής μελέτης στο χώρο της κλινικής ιατρικής είναι

το [196], στο οποίο οι ερευνητές προτείνουν τη χρήση ενός συστήματος με την ονομασία MedCo. Το σύστημα αυτό επιτρέπει σε μια ομάδα εγκαταστάσεων κλινικής υγείας (clinical sites), για παράδειγμα νοσοκομεία, να διασυνδεθούν με τρόπο που σέβεται την ασφάλεια και την προστασία των ευαίσθητων δεδομένων, έτσι ώστε τα δεδομένα αυτά να μπορούν να διαμοιραστούν σε εξωτερικούς ερευνητές. Το σύστημα MedCo χρησιμοποιεί μια ειδική μορφή κρυπτογράφησης, που ονομάζεται συλλογική ομομορφική κρυπτογράφηση (collective homomorphic encryption), για την παροχή αποκεντρωμένου μηχανισμού εμπιστοσύνης και την προστασία της εμπιστευτικότητας από άκρο σε άκρο, καθώς και τεχνικές απόκρυψης των προσωπικών δεδομένων.

Άλλες προσεγγίσεις για τη σχεδίαση αρχιτεκτονικών που μπορούν να αντιμετωπίσουν τα ζητήματα της ασφάλειας, της ιδιωτικότητας και της προστασίας των προσωπικών δεδομένων, περιλαμβάνουν τη χρήση τεχνολογιών, όπως το Blockchain και η Ομόσπονδη Μάθηση (Federated Learning). Το Blockchain είναι ουσιαστικά μια κατανεμημένη και δημόσια βάση δεδομένων, της οποίας το ιστορικό δεν μπορεί να μεταβληθεί (immutable), και επίσης κάθε «συναλλαγή» (εγγραφή, τροποποίηση, διαγραφή δεδομένων) επαληθεύεται από όλους τους κόμβους του δικτύου μέσω αλγορίθμων συναίνεσης. Η τεχνολογία Blockchain έχει ήδη χρησιμοποιηθεί από διάφορες εφαρμογές στο χώρο της υγείας και θα μπορούσε να υιοθετηθεί και για σκοπούς ασφαλούς ανταλλαγής δεδομένων κατά τη διάρκεια της πανδημίας COVID-19. Παράδειγμα τέτοιας λύσης είναι το σύστημα με την ονομασία MiPasa, το οποίο συνδυάζει δύο σύγχρονες τεχνολογίες από την εταιρεία IBM: το Blockchain και το cloud computing [197]. Σκοπός αυτής της λύσης είναι να παρέχει αξιόπιστα, εύκολα προσβάσιμα και υψηλής ποιότητας δεδομένα σε επιστημονικές και άλλες κοινότητες για την ανάπτυξη εφαρμογών που μπορούν να βοηθήσουν στην καταπολέμηση της πανδημίας.

Η Ομόσπονδη Μάθηση (Federated Learning) είναι μια ακόμα προσέγγιση αρχιτεκτονικής μηχανικής μάθησης με έμφαση στην προστασία των ευαίσθητων δεδομένων. Στα παραδοσιακά συστήματα μηχανικής μάθησης, τα δεδομένα συλλέγονται και αποθηκεύονται κεντρικά, προκειμένου να μπορούν να χρησιμοποιηθούν για την εκπαίδευση των μοντέλων μηχανικής μάθησης. Η ομόσπονδη μάθηση αποτελεί μια εναλλακτική προσέγγιση στην οποία τα περισσότερα προσωπικά δεδομένα δεν απαιτείται να διαβιβάζονται σε έναν κεντρικό διακομιστή [198]. Λαμβάνοντας ως παράδειγμα εφαρμογές τεχνητής νοημοσύνης σε κινητά τηλέφωνα για την αρχική διάγνωση COVID-19 (βλ. ενότητα 3.2.3), στη λογική της ομόσπονδης μάθησης κάθε τηλέφωνο μπορεί να εκπαιδεύσει το δικό του μοντέλο βαθιάς μάθησης, χρησιμοποιώντας τοπικά τα δεδομένα του χρήστη. Κατόπιν, τα κινητά τηλέφωνα στέλνουν τα εκπαιδευμένα μοντέλα τους σε έναν κεντρικό διακομιστή, ο οποίος είναι υπεύθυνος για τη δημιουργία ενός οικουμενικού μοντέλου (global model), που στη συνέχεια αποστέλλεται σε όλα τα κινητά τηλέφωνα. Να σημειωθεί ότι η πιο πάνω αρχιτεκτονική μπορεί να γενικευθεί. Συγκεκριμένα, στη θέση του κινητού τηλεφώνου μπορεί να βρίσκεται οποιαδήποτε περιφερειακή συσκευή, όπως για παράδειγμα ένας διακομιστής ενός νοσοκομείου, ενώ ο κεντρικός διακομιστής μπορεί να βρίσκεται στο cloud.

7.3.5 Επαλήθευση Μοντέλων σε Πραγματικές Συνθήκες

Οι περισσότερες από τις εφαρμογές που αφορούν στην κλινική ιατρική και οι οποίες εξετάστηκαν στο Κεφάλαιο 3, αναπτύχθηκαν και δοκιμάστηκαν από τους ερευνητές χρησιμοποιώντας σύνολα δεδομένων δοκιμών στα εργαστήριά τους. Θεωρούμε ότι οι εφαρμογές τεχνητής νοημοσύνης στο χώρο της υγείας θα πρέπει να αποδεικνύουν την αποτελεσματικότητα, την αποδοτικότητα και την ασφάλειά τους όχι μόνο σε εργαστηριακά σύνολα δεδομένων και απομονωμένα περιβάλλοντα, αλλά και σε πραγματικές συνθήκες, μέσω της ενσωμάτωσης των εν λόγω συστημάτων σε πραγματικές κλινικές ροές εργασίας. Μόνο τότε μπορούμε να αξιολογήσουμε τις πραγματικές τους δυνατότητες και να είμαστε βέβαιοι ότι είναι έτοιμα για παραγωγική λειτουργία.

Πριν από την ενσωμάτωση των υποψήφιων εφαρμογών / συστημάτων στην κλινική ροή, είναι ζωτικής σημασίας να δοκιμαστούν σε μια λεγόμενη «σιωπηλή» λειτουργία (silent mode). Κατά τη διάρκεια της λειτουργίας αυτής, το σύστημα παράγει κανονικά προβλέψεις σε πραγματικό χρόνο που παρουσιάζονται σε μια ομάδα εμπειρογνωμόνων, οι οποίες δε χρησιμοποιούνται για τη λήψη αποφάσεων, παρά μόνο για την αξιολόγηση της επίδοσης του συστήματος. Αυτή η μορφή ελέγχου και επαλήθευσης επιτρέπει στους κλινικούς ιατρούς να εντοπίζουν και να εξετάζουν πιθανά σφάλματα του μοντέλου σε πραγματικό χρόνο. Μετά τη θετική έκβαση των δοκιμών αυτών, μπορεί να αξιολογηθεί περαιτέρω η αποτελεσματικότητα του μοντέλου σε επιπλέον κλινικές μελέτες αν αυτό απαιτηθεί [199].

Εκτός από την εκτέλεση ρεαλιστικών δοκιμών σε πραγματικές συνθήκες, η κατανόηση του τρόπου με τον οποίο θα ενσωματωθεί καλύτερα ένα σύστημα τεχνητής νοημοσύνης εντός μιας ευρύτερης ροής εργασίας είναι εξίσου σημαντική. Για παράδειγμα, θα πρέπει να αποφασιστεί πώς και σε ποιον θα πρέπει να παρουσιάζονται τα αποτελέσματα πρόβλεψης του μοντέλου, καθώς και ποιος θα έχει την τελική ευθύνη για τη λήψη αποφάσεων. Ακόμα και προηγμένα μοντέλα εξαιρετικής ακρίβειας μπορεί να μην έχουν τον κατάλληλο αντίκτυπο εάν δεν ενσωματωθούν με σαφή και προσεγμένο τρόπο σε μια ροή εργασίας.

7.3.6 Διεπιστημονικές Συνεργασίες

Η πολυπλοκότητα και ο διεπιστημονικός χαρακτήρας της έρευνας που απαιτείται για τη δημιουργία και τη λειτουργία συστημάτων τεχνητής νοημοσύνης στη μάχη για την καταπολέμηση της πανδημίας COVID-19 απαιτούν τη δημιουργία ετερογενών ομάδων που απαρτίζονται από ανθρώπους διαφορετικών ειδικοτήτων, με συμπληρωματικές οπτικές, καθώς και τη σύναψη μακροπρόθεσμων συνεργασιών μεταξύ ακαδημαϊκών ιδρυμάτων, εταιρειών και δημόσιων υπηρεσιών. Μηχανικοί στον τομέα της τεχνητής νοημοσύνης, επιστήμονες δεδομένων, ερευνητές ιατρικής και άλλων βιολογικών επιστημών, καθώς και επαγγελματίες υγείας θα πρέπει να συνεργαστούν με έναν κοινό στόχο. Στόχος των ομάδων αυτών θα είναι να αναπτύξουν με προτεραιότητα εκείνες τις εφαρμογές που θα αντιμετωπίσουν τις κρίσιμες ανάγκες στον αντίστοιχο τομέα, όπως για παράδειγμα στην επιδημιολογία, την κλινική ιατρική και τη φαρμακολογία που παρουσιάστηκαν σε αυτή την εργασία.

Χρησιμοποιώντας και πάλι ένα παράδειγμα που αναφέρθηκε προηγουμένως, η χρήση μοντέλων τύπου μαύρου κουτιού στη διάγνωση για την ασθένεια COVID-19 μέσω ιατρικής εικόνας, θα μπορούσε φαινομενικά να αποτελέσει μια πρακτική λύση, που όμως είναι σχεδόν άχρηστη χωρίς την κατάλληλη ερμηνεία και αξιολόγηση από ειδικούς ιατρούς. Συνεπώς, η δημιουργία ομάδων από ανθρώπους με αλληλοσυμπληρούμενη εμπειρία είναι σημαντική σε όλα τα στάδια του κύκλου ζωής των εφαρμογών τεχνητής νοημοσύνης: στον ολοκληρωμένο σχεδιασμό, την ανάπτυξη, τη συλλογή των κατάλληλων δεδομένων, την εκτέλεση ρεαλιστικών δοκιμών, την παραγωγική λειτουργία καθώς και τη συνεχή προσαρμογή και βελτίωσή τους.

Είναι προφανές ότι ο ρόλος κάθε ειδικότητας στις ομάδες αυτές είναι μοναδικός και σημαντικός. Οι επιστήμονες δεδομένων και οι μηχανικοί υπολογιστών, οι οποίοι αναπτύσσουν τα συστήματα τεχνητής νοημοσύνης, γνωρίζουν τι μπορεί να κάνει η τεχνητή νοημοσύνη αλλά και τι δεν μπορεί να κάνει. Οι ειδικοί ανά περιοχή (domain experts), όπως για παράδειγμα επιδημιολόγοι, κλινικοί ιατροί και βιολόγοι, κατανοούν τη φύση του προβλήματος, θέτουν τα κατάλληλα ερωτήματα που πρέπει να απαντηθούν και μπορούν να προσδιορίσουν, για παράδειγμα, αν τα δεδομένα της αρχικής εκπαίδευσης ενός μοντέλου εξακολουθούν να είναι σχετικά και επαρκή ή αν το μοντέλο πρέπει να επανεκπαιδευθεί. Τέλος, ανεξαρτήτως ειδικότητας και ρόλου, οι άνθρωποι που συμμετέχουν σε αυτές τις ομάδες θα πρέπει να επιδεικνύουν ανοιχτή σκέψη που μπορεί να οδηγήσει σε νέες ιδέες και καινοτόμες εφαρμογές.

7.3.7 Διεθνής Συνεργασία και Λύσεις για Αναπτυσσόμενες Χώρες

Θεωρούμε ότι η διεθνής συνεργασία μπορεί να διαδραματίσει σημαντικότατο ρόλο σε αυτή την πανδημία που δε γνωρίζει σύνορα. Λύσεις και εφαρμογές τεχνητής νοημοσύνης που έχουν δοκιμαστεί σε κάποια χώρα του κόσμου μπορούν να διαμοιραστούν σε παγκόσμιο επίπεδο και να προσαρμοστούν σε άλλα περιβάλλοντα, δίνοντας προτεραιότητα στην κάλυψη τοπικών αναγκών. Πολλοί διεθνείς οργανισμοί, εταιρείες του ιδιωτικού τομέα καθώς και συνεργασίες στον τομέα της έρευνας λειτουργούν διασυνοριακά και είναι σε θέση να διευκολύνουν τη διάδοση της γνώσης και την ανάπτυξη ικανοτήτων σε τοπικό επίπεδο. Με αυτό τον τρόπο, χώρες με λιγότερους πόρους και τεχνολογικές υποδομές μπορούν να επωφεληθούν από την παγκόσμια συνεργασία και να επικεντρώσουν τις προσπάθειές τους στις σημαντικότερες τοπικές προκλήσεις.

Είναι γενικώς αποδεκτό ότι οι συνέπειες και οι προκλήσεις της πανδημίας COVID-19 είναι εντονότερες στους πληθυσμούς εκείνους που έχουν περιορισμένη πρόσβαση σε αξιοπρεπή υγειονομική περίθαλψη, όπως για παράδειγμα οι πληθυσμοί των φτωχότερων χωρών. Ταυτόχρονα, οι πολίτες τέτοιων χωρών επηρεάζονται αρνητικά και από την περιορισμένη πρόσβαση στην πληροφορία. Συνεπώς, μια βασική πρόκληση είναι η ανάπτυξη τεχνολογικών λύσεων που σχεδιάζονται με τρόπο ώστε να μην προκαλούν αποκλεισμούς. Αυτό απαιτεί να εξεταστεί, για παράδειγμα, ο τρόπος με τον οποίο μπορούν να εφαρμοστούν αυτές οι λύσεις σε αγροτικές και οικονομικά φτωχότερες περιοχές, οι οποίες θέτουν πρόσθετες πρακτικές δυσκολίες που θα πρέπει να ληφθούν υπόψη. Για παράδειγμα, μια εφαρμογή για κινητές συσκευές για την αυτοματοποιημένη ιχνηλάτηση επαφών θα πρέπει να σχεδιαστεί για περιορισμένους υπολογιστικούς πόρους, θα πρέπει να μπορεί να λειτουργήσει υπό περιορισμένη δικτυακή συνδεσιμότητα, θα πρέπει να υποστηρίζει πολλές γλώσσες και θα πρέπει να είναι εύχρηστη ακόμα και από αναλφάβητους χρήστες. Γενικότερα, η ευρεία προσβασιμότητα των τεχνολογικών λύσεων τεχνητής νοημοσύνης θα τις καταστήσει εφαρμόσιμες σε μεγαλύτερη κλίμακα, κάτι που μπορεί να είναι ζωτικής σημασίας για την επιτυχή αντιμετώπιση αυτής της παγκόσμιας κρίσης.

7.4 Συμπεράσματα

Η πανδημία COVID-19 έχει επηρεάσει σημαντικά τις ζωές των ανθρώπων σε όλο τον κόσμο ενώ παράλληλα ο αριθμός των θανάτων από την αντίστοιχη ασθένεια εξακολουθεί να αυξάνεται συνεχώς. Η έκτακτη αυτή κατάσταση έχει προκαλέσει σημαντική πίεση στα συστήματα υγείας των περισσοτέρων χωρών και έχει δημιουργήσει πολλές οικονομικές και κοινωνικές προκλήσεις στην ανθρωπότητα.

Είναι γενικώς αποδεκτό ότι η τεχνητή νοημοσύνη διαδραματίζει πλέον ένα σημαντικό ρόλο στην καθημερινότητά μας και έχει ενσωματωθεί σε πολλούς διαφορετικούς τομείς των επιστημών και της βιομηχανίας, συμπεριλαμβανομένου του τομέα της ιατρικής και της υγείας. Η δύναμη της τεχνητής νοημοσύνης πηγάζει από τη δυνατότητά της να αναλύσει και να συνθέσει έναν τεράστιο όγκο διαφορετικών δεδομένων, με ταχύτητα και αποτελεσματικότητα, που πρακτικά θα ήταν αδύνατο να τύχουν επεξεργασίας από τον άνθρωπο με τη χρήση παραδοσιακών υπολογιστικών μεθόδων. Από την εμφάνιση του νέου ιού μέχρι και σήμερα, πάρα πολλοί ερευνητές, οργανισμοί και εταιρείες ανά τον κόσμο έχουν μπει στη μάχη για την αξιοποίηση της τεχνητής νοημοσύνης και της μηχανικής μάθησης σε μια προσπάθεια να συνεισφέρουν στην αντιμετώπιση διαφορετικών πτυχών της πανδημίας.

Στη Διπλωματική αυτή Εργασία έγινε μια προσπάθεια ανασκόπησης εφαρμογών και τεχνολογιών τεχνητής νοημοσύνης για την αντιμετώπιση της πανδημίας COVID-19 σε τέσσερις βασικές περιοχές: στην Επιδημιολογία, την Κλινική Ιατρική, τη Φαρμακολογία και την Επιδημία της Πληροφορίας

(Infodemic). Παρουσιάστηκαν εφαρμογές και λύσεις προερχόμενες τόσο από την ακαδημαϊκή έρευνα όσο και από εταιρείες και οργανισμούς που δραστηριοποιούνται σε τομείς σχετιζόμενους με την τεχνητή νοημοσύνη. Στόχος ήταν να δοθεί μια εικόνα του εύρους των πιθανών εφαρμογών, των τεχνολογιών που εκμεταλλεύονται, των δεδομένων που χρησιμοποιούν και των προβλημάτων που επιχειρούν να επιλύσουν.

Μέσω της ανασκόπησης αυτής συμπεράναμε ότι πράγματι η τεχνητή νοημοσύνη φαίνεται να μπορεί να δώσει λύσεις σε πολλά και δύσκολα προβλήματα με τα οποία έρχεται αντιμέτωπη η παγκόσμια κοινότητα. Κάποιες από τις εφαρμογές που εξετάστηκαν βρίσκονται ακόμα σε πρώιμο, ερευνητικό στάδιο ενώ κάποιες άλλες έχουν ήδη τεθεί σε παραγωγική λειτουργία και χρησιμοποιούνται για να βοηθήσουν τους ανθρώπους στην εξαγωγή χρήσιμων συμπερασμάτων και προβλέψεων.

Αξίζει να τονίσουμε ότι η τεχνητή νοημοσύνη και οι εφαρμογές της στην αντιμετώπιση της πανδημίας COVID-19 δεν αποτελούν πανάκεια και θα πρέπει να τυγχάνουν κριτικής αξιολόγησης και συνεχούς βελτίωσης. Ανεξαρτήτως περιοχής, παρατηρήσαμε ότι πολλές από τις εν λόγω εφαρμογές έχουν να αντιμετωπίσουν μια σειρά κοινών προκλήσεων σε σχέση με το πλήθος και την ποιότητα των δεδομένων, την προστασία ευαίσθητων και προσωπικών πληροφοριών, την ασφάλεια, την επεξηγησιμότητα, τη μεροληψία και άλλα ζητήματα ηθικής. Αναγνωρίζοντας τις προκλήσεις αυτές, προτείναμε μια σειρά από κατευθύνσεις για την αντιμετώπισή τους, στοχεύοντας παράλληλα στη μεγιστοποίηση του οφέλους από τη χρήση τέτοιων εφαρμογών.

Αναμφισβήτητα, η ανθρωπότητα θα μπορούσε να είναι καλύτερα προετοιμασμένη για μια πανδημία όπως η σημερινή, έτσι ώστε να αντιμετώπιζε με αποτελεσματικότερο τρόπο την έκτακτη αυτή κατάσταση. Εντούτοις, η ευκαιρία που δημιουργήθηκε για διεθνείς και διεπιστημονικές συνεργασίες σε διάφορους τομείς, που περιλαμβάνουν και την τεχνητή νοημοσύνη, είναι μοναδική. Ακόμα και αν κάποιες από τις εφαρμογές που δημιουργήθηκαν στα πλαίσια της τρέχουσας πανδημίας δεν έφτασαν στο επιθυμητό στάδιο παραγωγικής λειτουργίας, αξίζει να βελτιωθούν και να γενικευθούν, έτσι ώστε να μπορούμε να τις εκμεταλλευθούμε έγκαιρα και αποδοτικά σε αντίστοιχες μελλοντικές κρίσεις.

Βιβλιογραφικές Αναφορές

[1] CDC, “COVID-19 FAQs,” 2020. https://www.cdc.gov/coronavirus/2019-ncov/faq.html (accessed Nov. 14, 2020).

[2] ECDC, “Coronaviruses,” 2020. https://www.ecdc.europa.eu/en/covid-19/latest-evidence/coronaviruses (accessed Nov. 14, 2020).

[3] M. Cevik, K. Kuppalli, J. Kindrachuk, and M. Peiris, “Virology, transmission, and pathogenesis of SARS-CoV-2,” *BMJ (Clinical research ed.)*, vol. 371, p. m3862, 2020, doi: 10.1136/bmj.m3862.

[4] A. A. T. Naqvi *et al.*, “Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach,” *Biochimica et Biophysica Acta - Molecular Basis of Disease*, vol. 1866, no. 10. Elsevier B.V., Oct. 01, 2020, doi: 10.1016/j.bbadis.2020.165878.

[5] E. K. Stokes *et al.*, “Coronavirus Disease 2019 Case Surveillance — United States, January 22–May 30, 2020,” *MMWR. Morbidity and Mortality Weekly Report*, vol. 69, no. 24, pp. 759–765, 2020, doi: 10.15585/mmwr.mm6924e2.

[6] WHO, “Coronavirus disease (COVID-19) pandemic.” https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed Nov. 15, 2020).

[7] S. Neilson, A. Woodward, and D. Mosher, “A comprehensive timeline of the coronavirus pandemic at 9 months, from China’s first case to the present,” Sep. 2020.

[8] C. Hoffmann and B. S. Kamps, *COVID Reference*, 5th editio. 2020.

[9] N. Zhu *et al.*, “A Novel Coronavirus from Patients with Pneumonia in China, 2019,” *New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, 2020, doi: 10.1056/nejmoa2001017.

[10] C. Huang *et al.*, “Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China,” *The Lancet*, vol. 395, no. 10223, pp. 497–506, 2020, doi: 10.1016/S0140-6736(20)30183-5.

[11] John Hopkins, “Coronavirus Resource Center,” 2020. https://coronavirus.jhu.edu/map.html (accessed Feb. 28, 2021).

[12] WHO, “Estimating Mortality Rates from COVID-19.” https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19 (accessed Nov. 15, 2020).

[13] P. Ferrara and L. Albano, “COVID-19 and healthcare systems: What should we do next?,” *Public Health*, vol. 185. pp. 1–2, 2020, doi: 10.1016/j.puhe.2020.05.014.

[14] S. J. Russel and P. Norvig, *Artificial Intelligence, A Modern Approach*, 3rd ed. Prentice Hall.

[15] V. Gadepally *et al.*, “AI enabling technologies: A survey,” *arXiv*, vol. 7014, no. Feb 2014, 2019.

[16] A. Mezic, “Why Unsupervised Machine Learning is the Future of Cybersecurity.” https://www.technative.io/why-unsupervised-machine-learning-is-the-future-of-cybersecurity/ (accessed Feb. 28, 2021).

[17] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[18] K. H. Yu, A. L. Beam, and I. S. Kohane, “Artificial intelligence in healthcare,” *Nature Biomedical Engineering*, vol. 2, no. 10, pp. 719–731, 2018, doi: 10.1038/s41551-018-0305-z.

[19] R. C. Deo, “Machine learning in medicine,” *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015, doi: 10.1161/CIRCULATIONAHA.115.001593.

[20] T. B. Murdoch and A. S. Detsky, “The inevitable application of big data to health care,” *JAMA - Journal of the American Medical Association*, vol. 309, no. 13, pp. 1351–1352, 2013, doi: 10.1001/jama.2013.393.

[21] B. Ristevski and M. Chen, “Big Data Analytics in Medicine and Healthcare,” *Journal of integrative bioinformatics*, vol. 15, no. 3, pp. 1–5, 2018, doi: 10.1515/jib-2017-0030.

[22] Ι. Ντζούφρας, “Σημειώσεις Μαθήματος Βιοστατιστικής - Ειδικές Ιατρικές και Επιδημιολογικές Μελέτες.” [Online]. Available: http://www2.stat-athens.aueb.gr/~jbn/courses/biostats2_uoa/pdf/01_medical_epidemiologic_studies_intro.pdf.

[23] WHO, “Emergencies Preparedness, Response.” https://www.who.int/csr/don/12-january-%0A2020-novel-coronavirus-china/en/ (accessed Nov. 28, 2020).

[24] BlueDot, “Anticipate outbreaks. Mitigate risk. Build resilience.” https://bluedot.global/ (accessed Nov. 28, 2020).

[25] I. I. Bogoch, A. Watts, A. Thomas-Bachli, C. Huber, M. U. G. Kraemer, and K. Khan, “Potential for global spread of a novel coronavirus from China,” *Journal of Travel Medicine*, vol. 27, no. 2, pp. 1–3, 2020, doi: 10.1093/jtm/taaa011.

[26] I. I. Bogoch, A. Watts, A. Thomas-Bachli, C. Huber, M. U. G. Kraemer, and K. Khan, “Pneumonia of unknown aetiology in Wuhan, China: Potential for international spread via commercial air travel,” *Journal of Travel Medicine*, vol. 27, no. 2, pp. 1–3, 2020, doi: 10.1093/jtm/taaa008.

[27] Z. Allam, G. Dey, and D. S. Jones, “Artificial Intelligence (AI) Provided Early Detection of the Coronavirus (COVID-19) in China and Will Influence Future Urban Health Policy Internationally,” *Ai*, vol. 1, no. 2, pp. 156–165, 2020, doi: 10.3390/ai1020009.

[28] Metabiota, “Making the world more resilient to epidemics.” https://metabiota.com/ (accessed Nov. 29, 2020).

[29] Z. Allam, “The Rise of Machine Intelligence in the COVID-19 Pandemic and Its Impact on Health Policy,” *Surveying the Covid-19 Pandemic and its Implications*, no. January, pp. 89–96, 2020, doi: 10.1016/b978-0-12-824313-8.00006-1.

[30] W. Douglas Heaven, “AI could help with the next pandemic - but not with this one,” *MIT Technology Review*, pp. 1–14, 2020, [Online]. Available: https://www.technologyreview.com/s/615351/ai-could-help-with-the-next-pandemicbut-not-with-this-one/.

[31] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, “Detecting influenza epidemics using search engine query data,” *Nature*, vol. 457, no. 7232, pp. 1012–1014, 2009, doi: 10.1038/nature07634.

[32] Y. Teng *et al.*, “Dynamic forecasting of zika epidemics using google trends,” *PLoS ONE*, vol. 12, no. 1, pp. 1–10, 2017, doi: 10.1371/journal.pone.0165085.

[33] S. Lim, C. S. Tucker, and S. Kumara, “An unsupervised machine learning model for discovering latent infectious diseases using social media data,” *Journal of Biomedical Informatics*, vol. 66, no. February, pp. 82–94, 2017, doi: 10.1016/j.jbi.2016.12.007.

[34] L. Ferretti *et al.*, “Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing,” *Science*, vol. 368, no. 6491, pp. 0–7, 2020, doi: 10.1126/science.abb6936.

[35] E. Yoneki and J. Crowcroft, “EpiMap: Towards quantifying contact networks for understanding epidemiology in developing countries,” *Ad Hoc Networks*, vol. 13, no. PART A, pp. 83–93, 2014, doi: 10.1016/j.adhoc.2012.06.003.

[36] N. Ahmed *et al.*, “A Survey of COVID-19 Contact Tracing Apps,” *IEEE Access*, vol. 8, pp. 134577–134601, 2020, doi: 10.1109/ACCESS.2020.3010226.

[37] NIST, “Pilot TC4TL Challenge.” https://tc4tlchallenge.nist.gov/ (accessed Dec. 06, 2020).

[38] J. Hsu, “Can AI Make Bluetooth Contact Tracing Better?,” *IEEE Spectrum*, p. 8th September, 2020, [Online]. Available: https://spectrum.ieee.org/the-human-os/artificial-intelligence/machine-learning/ai-bluetooth-contact-tracing.

[39] F. Sattler *et al.*, “Risk estimation of sars-cov-2 transmission from bluetooth low energy measurements,” *arXiv*, 2020.

[40] P. Policy, C. Policy, P. Health, and D. Muoio, “Machine learning algorithm helps Bluetooth contact tracing systems discern per-event risk of COVID- 19 transmission,” 2020.

[41] M. Akhtar, M. U. G. Kraemer, and L. M. Gardner, “A dynamic neural network model for predicting risk of Zika in real-time,” *bioRxiv*, pp. 1–16, 2018, doi: 10.1101/466581.

[42] K. Hao, “This is how the CDC is trying to forecast coronavirus’s spread - MIT Technology Review,” *MIT Technology Report*, 2020, [Online]. Available: https://www.technologyreview.com/2020/03/13/905313/cdc-cmu-forecasts-coronavirus-spread/?truid=b43522c9ab0b0cde77b8d8b698991108&utm_source=the_algorithm&utm_medium=email&utm_campaign=the_algorithm.unpaid.engagement&utm_content=11-01-2020%0Ahttps://www.tec.

[43] S. L. Chang, N. Harding, C. Zachreson, O. M. Cliff, and M. Prokopenko, “Modelling transmission and control of the COVID-19 pandemic in Australia,” *Nature Communications*, vol. 11, no. 1, pp. 1–13, 2020, doi: 10.1038/s41467-020-19393-6.

[44] C. J. Huang, Y. H. Chen, Y. Ma, and P. H. Kuo, “Multiple-Input Deep Convolutional Neural Network Model for COVID-19 Forecasting in China,” *medRxiv*, pp. 1–16, 2020, doi: 10.1101/2020.03.23.20041608.

[45] Z. Hu, Q. Ge, S. Li, L. Jin, and M. Xiong, “Artificial intelligence forecasting of covid-19 in China,” *arXiv*, no. April, 2020.

[46] M. A. A. Al-Qaness, A. A. Ewees, H. Fan, and M. A. el Aziz, “Optimization method for forecasting confirmed cases of COVID-19 in China,” *Applied Sciences*, vol. 9, no. 3, 2020, doi: 10.3390/JCM9030674.

[47] R. Dandekar and G. Barbastathis, “Neural network aided quarantine control model estimation of global Covid-19 spread,” *arXiv*, vol. 2020, no. Who, pp. 1–9, 2020.

[48] Z. Yang *et al.*, “Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions,” *Journal of Thoracic Disease*, vol. 12, no. 3, pp. 165–174, 2020, doi: 10.21037/jtd.2020.02.64.

[49] S. J. Fong, G. Li, N. Dey, R. G. Crespo, and E. Herrera-Viedma, “Finding an accurate early forecasting model from small dataset: A Case of 2019-nCoV novel coronavirus outbreak,” *arXiv*, no. February, 2020, doi: 10.9781/ijimai.2020.02.002.

[50] C. Poirier *et al.*, “Real-time forecasting of the COVID-19 outbreak in Chinese provinces: Machine learning approach using novel digital data and estimates from mechanistic models,” *Journal of Medical Internet Research*, vol. 22, no. 8, pp. 1–15, 2020, doi: 10.2196/20285.

[51] D. Balcan, B. Gonçalves, H. Hu, J. J. Ramasco, V. Colizza, and A. Vespignani, “Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model,” *Journal of Computational Science*, vol. 1, no. 3, pp. 132–145, 2010, doi: 10.1016/j.jocs.2010.07.002.

[52] V. Lampos *et al.*, “Tracking COVID-19 using online search,” 2020, [Online]. Available: http://arxiv.org/abs/2003.08086.

[53] L. Qin *et al.*, “Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 7, 2020, doi: 10.3390/ijerph17072365.

[54] R. Biele, “COVID-19 : What Is Hidden Behind the Official Numbers? Machine Learning Reveals Insights: A Time Series Cluster Analysis,” *Towards Data Science*, 2020, [Online]. Available: https://towardsdatascience.com/which-countries-are-affected-the-most-by-covid-19-4d4570852e31.

[55] R. M. Carrillo-Larco and M. Castillo-Cara, “Using country-level variables to classify countries according to the number of confirmed COVID-19 cases: An unsupervised machine learning approach,” *Wellcome Open Research*, vol. 5, pp. 1–22, 2020, doi: 10.12688/wellcomeopenres.15819.3.

[56] Z. Hu, Q. Ge, S. Li, L. Jin, and M. Xiong, “Artificial intelligence forecasting of covid-19 in China,” *arXiv*, pp. 1–20, 2020.

[57] P. Hartono, “Generating similarity map for COVID-19 transmission dynamics with topological autoencoder,” *arXiv*, 2020, doi: 10.1016/j.imu.2020.100386.

[58] R. Pal, A. A. Sekh, S. Kar, and D. K. Prasad, “Neural network based country wise risk prediction of COVID-19,” *Applied Sciences (Switzerland)*, vol. 10, no. 18, 2020, doi: 10.3390/APP10186448.

[59] G. Bruno, R. Wemole, G. B. Ronsivalle, L. Foresti, and G. Poledda, “A Prototype Model of Georeferencing the Inherent Risk of Contagion from COVID-19,” no. March, 2020, doi: 10.13140/RG.2.2.30077.72163.

[60] Y. Ye *et al.*, “α-Satellite: An AI-Driven System and Benchmark Datasets for Dynamic COVID-19 Risk Assessment in the United States,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2755–2764, 2020, doi: 10.1109/JBHI.2020.3009314.

[61] C. S. of E. (CSE), “α-Satellite: COVID-19 in the US.” https://covid-19.yes-lab.org/.

[62] W. Choi *et al.*, “Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT,” *Radiology*, vol. 1, pp. 1–13, 2020.

[63] L. Li *et al.*, “Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy,” *Radiology*, vol. 296, no. 2, pp. E65–E71, 2020, doi: 10.1148/radiol.2020200905.

[64] X. Xu *et al.*, “A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia,” *Engineering*, vol. 6, no. 10, pp. 1122–1129, 2020, doi: 10.1016/j.eng.2020.04.010.

[65] O. Gozes *et al.*, “Rapid AI Development Cycle for the Coronavirus (COVID-19) Pandemic: Initial Results for Automated Detection & Patient Monitoring using Deep Learning CT Image Analysis,” no. April, 2020, [Online]. Available: http://arxiv.org/abs/2003.05037.

[66] RADLogics, “AI-Based detection and the COVID-19 pandemic.” https://www.radlogics.com/covid-19/ (accessed Jan. 23, 2021).

[67] F. Shan+ *et al.*, *Lung Infection Quantification of COVID-19 in CT Images with Deep Learning*. 2020.

[68] L. Wang, Z. Q. Lin, and A. Wong, “COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images,” *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-76550-z.

[69] K. Murphy *et al.*, “COVID-19 on chest radiographs: A multireader evaluation of an artificial intelligence system,” *Radiology*, vol. 296, no. 3, pp. E166–E172, 2020, doi: 10.1148/radiol.2020201874.

[70] Lunit Inc, “Lunit Insight CXR for COVID-19,” 2020. https://insight.lunit.io/covid19/login (accessed Dec. 19, 2020).

[71] VUNO Med, “LungQuant COVID-19 Version,” 2020. https://covid19.vunomed.com/ (accessed Dec. 19, 2020).

[72] G. Battineni, N. Chintalapudi, and F. Amenta, “AI Chatbot Design during an Epidemic like the Novel Coronavirus,” *Healthcare*, vol. 8, no. 2, p. 154, 2020, doi: 10.3390/healthcare8020154.

[73] Symptoma, “Determine your risk of COVID-19.” https://www.symptoma.com/ (accessed Dec. 26, 2020).

[74] A. Martin *et al.*, “An artificial intelligence-based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot,” *Scientific Reports*, vol. 10, no. 1, pp. 1–7, 2020, doi: 10.1038/s41598-020-75912-x.

[75] I. Song, “Diagnosis of pneumonia from sounds collected using low cost cell phones,” in *2015 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2015, pp. 1–8, doi: 10.1109/IJCNN.2015.7280317.

[76] B. W. Schuller, D. M. Schuller, K. Qian, J. Liu, H. Zheng, and X. Li, “COVID-19 and computer audition: An overview on what speech & sound analysis could contribute in the SARS-COV-2 corona crisis,” *arXiv*, no. March, 2020.

[77] A. Imran *et al.*, "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app," *Informatics in Medicine Unlocked*, vol. 20, p. 100378, 2020, doi: 10.1016/j.imu.2020.100378.

[78] J. Laguarta, F. Hueto, and B. Subirana, "COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 275–281, 2020, doi: 10.1109/ojemb.2020.3026928.

[79] Amanda Morris, "Monitoring COVID-19 from hospital to home: First wearable device continuously tracks key symptoms," *NORTHWESTERN NOW*, 2020, [Online]. Available: https://news.northwestern.edu/stories/2020/04/monitoring-covid-19-from-hospital-to-home-first-wearable-device-continuously-tracks-key-symptoms/.

[80] H. S. Maghded, K. Z. Ghafoor, A. S. Sadiq, K. Curran, D. B. Rawat, and K. Rabie, "A Novel AI-enabled Framework to Diagnose Coronavirus COVID-19 using Smartphone Embedded Sensors: Design Study," *Proceedings - 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science, IRI 2020*, pp. 180–187, 2020, doi: 10.1109/IRI49571.2020.00033.

[81] Y. Wang, M. Hu, Q. Li, X. P. Zhang, G. Zhai, and N. Yao, "Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner," *arXiv*, 2020.

[82] A. A. Khorana, "Artificial intelligence for cancer-associated thrombosis risk assessment," *The Lancet Haematology*, vol. 5, no. 9, pp. e391–e392, 2018, doi: 10.1016/S2352-3026(18)30118-2.

[83] V. N. Dubey, J. M. Dave, J. Beavis, and D. v Coppini, "Predicting Diabetic Neuropathy Risk Level Using Artificial Neural Network and Clinical Parameters of Subjects With Diabetes.," *Journal of diabetes science and technology*, p. 1932296820965583, Oct. 2020, doi: 10.1177/1932296820965583.

[84] M. John and H. Shaiba, "Main factors influencing recovery in MERS Co-V patients using machine learning," *Journal of Infection and Public Health*, vol. 12, no. 5, pp. 700–704, 2019, doi: https://doi.org/10.1016/j.jiph.2019.03.020.

[85] A. Colubri, T. Silver, T. Fradet, K. Retzepi, B. Fry, and P. Sabeti, "Transforming Clinical Data into Actionable Prognosis Models: Machine-Learning Framework and Field-Deployable App to Predict Outcome of Ebola Patients," *PLoS Neglected Tropical Diseases*, vol. 10, no. 3, pp. 1–17, 2016, doi: 10.1371/journal.pntd.0004549.

[86] EPIC, "Epic AI Helps Clinicians Predict When COVID-19 Patients Might Need Intensive Care," 2020. https://www.epic.com/epic/post/epic-ai-helps-clinicians-predict-covid-19-patients-might-need-intensive-care (accessed Dec. 19, 2020).

[87] K. Singh *et al.*, "Evaluating a Widely Implemented Proprietary Deterioration Index Model Among Hospitalized COVID-19 Patients," *medRxiv*, 2020, doi: 10.1101/2020.04.24.20079012.

[88] A. L. Booth, E. Abels, and P. McCaffrey, "Development of a prognostic model for mortality in COVID-19 infection using machine learning," *Modern Pathology*, 2020, doi: 10.1038/s41379-020-00700-x.

[89] X. Jiang *et al.*, “Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity,” *Computers, Materials and Continua*, vol. 63, no. 1, pp. 537–551, 2020, doi: 10.32604/cmc.2020.010691.

[90] Z. Tang *et al.*, “Severity assessment of COVID-19 using CT image features and laboratory indices,” *Physics in Medicine & Biology*, 2020, [Online]. Available: http://iopscience.iop.org/article/10.1088/1361-6560/abbf9e.

[91] W. Shi *et al.*, “Deep Learning-Based Quantitative Computed Tomography Model in Predicting the Severity of COVID-19: A Retrospective Study in 196 Patients,” *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3546089.

[92] University of Pennsylvania Health System, “COVID-19 Hospital Impact Model for Epidemics (CHIME),” 2020. https://code-for-philly.gitbook.io/chime/ (accessed Jan. 08, 2020).

[93] Z. Qian, A. M. Alaa, and M. van der Schaar, “CPAS: the UK’s national machine learning-based hospital capacity planning system for COVID-19,” *Machine Learning*, no. 0123456789, 2020, doi: 10.1007/s10994-020-05921-4.

[94] M. van der Schaar, “Progress using COVID-19 patient data to train machine learning models for healthcare,” *University of Cambridge - Research News*. https://www.cam.ac.uk/research/news/progress-using-covid-19-patient-data-to-train-machine-learning-models-for-healthcare (accessed Dec. 20, 2020).

[95] M. van der Schaar *et al.*, “How artificial intelligence and machine learning can help healthcare systems respond to COVID-19,” *Machine Learning*, no. 0123456789, pp. 1–9, 2020, doi: 10.1007/s10994-020-05928-x.

[96] A. M. Alaa, “Bayesian Inference of Individualized Treatment Effects using Multi-task Gaussian Processes,” no. Nips, pp. 1–9, 2017.

[97] J. Yoon, J. Jordon, and M. van der Schaar, “Ganite: Estimation of individualized treatment effects using generative adversarial nets,” *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, no. 2010, pp. 1–22, 2018.

[98] Y. Zhang, A. Bellot, and M. van der Schaar, “Learning Overlapping Representations for the Estimation of Individualized Treatment Effects,” *arXiv*, vol. 108, 2020.

[99] Protein Structure Prediction Center, “Artificial intelligence solution to a 50-year-old science challenge could ‘revolutionise’ medical research,” *30-11-2020*, 2020. https://predictioncenter.org/casp14/doc/CASP14_press_release.html (accessed Dec. 27, 2020).

[100] DeepMind, “AlphaFold can accurately predict 3D models of protein structures and has the potential to accelerate research in every field of biology.,” [Online]. Available: https://deepmind.com/research/case-studies/alphafold.

[101] D. Hassabis. John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Kathryn Tunyasuvunakool, Olaf Ronneberger, Russ Bates, Augustin Žídek, Alex Bridgland, Clemens Meyer, Simon A A Kohl, Anna Potapenko, Andrew J Ballard, Andrew Cowie, Bernardino Ro, “High Accuracy Protein Structure Prediction Using Deep Learning,” in *Fourteenth Critical Assessment of Techniques for Protein Structure Prediction (Abstract Book)*, 2020.

[102] E. Sanz, “Anfinsen and the Architecture of Proteins,” *OpenMind BBVA*, 2016. .

[103] N. Parker, M. Schneegurt, A.-H. Thi Tu, P. Lister, and B. Forster, “Proteins,” in *Microbiology*, .

[104] M. Torrisi, G. Pollastri, and Q. Le, “Deep learning methods in protein structure prediction,” *Computational and Structural Biotechnology Journal*, vol. 18, pp. 1301–1310, 2020, doi: 10.1016/j.csbj.2019.12.011.

[105] Protein Structure Prediction Center, “Homepage,” 2020. https://predictioncenter.org/index.cgi (accessed Dec. 27, 2020).

[106] A. W. Senior *et al.*, “Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13),” *Proteins: Structure, Function and Bioinformatics*, vol. 87, no. 12, pp. 1141–1148, 2019, doi: 10.1002/prot.25834.

[107] A. W. Senior *et al.*, “Improved protein structure prediction using potentials from deep learning,” *Nature*, vol. 577, no. 7792, pp. 706–710, 2020, doi: 10.1038/s41586-019-1923-7.

[108] DeepMind, “Computational predictions of protein structures associated with COVID-19,” *DeepMind Research*, pp. 1–7, 2020, [Online]. Available: https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19.

[109] J. Yang, I. Anishchenko, H. Park, Z. Peng, S. Ovchinnikov, and D. Baker, “Improved protein structure prediction using predicted interresidue orientations,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 3, pp. 1496–1503, 2020, doi: 10.1073/pnas.1914677117.

[110] W. Zheng, Y. Li, C. Zhang, R. Pearce, S. M. Mortuza, and Y. Zhang, “Deep-learning contact-map guided protein structure prediction in CASP13,” *Proteins*, vol. 87, no. 12, pp. 1149–1164, Dec. 2019, doi: 10.1002/prot.25792.

[111] N. Fleming, “How artificial intelligence is changing drug discovery,” *Nature*, vol. 557, no. 7707, pp. S55–S57, 2018, doi: 10.1038/d41586-018-05267-x.

[112] K. K. Mak and M. R. Pichika, “Artificial intelligence in drug development: present status and future prospects,” *Drug Discovery Today*, vol. 24, no. 3, pp. 773–780, 2019, doi: 10.1016/j.drudis.2018.11.014.

[113] J. Avorn, “The $2.6 Billion Pill — Methodologic and Policy Considerations,” *New England Journal of Medicine*, vol. 372, no. 20, pp. 1877–1879, May 2015, doi: 10.1056/NEJMp1500848.

[114] Insider Intelligence, “Big pharma is using AI and machine learning in drug discovery and development to save lives,” *Business Insider*, 2020, [Online]. Available: https://www.businessinsider.com/ai-machine-learning-in-drug-discovery-development-2020.

[115] Y. Zhou, F. Wang, J. Tang, R. Nussinov, and F. Cheng, “Artificial intelligence in COVID-19 drug repurposing,” *The Lancet Digital Health*, vol. 2, no. 12, pp. e667–e676, 2020, doi: 10.1016/S2589-7500(20)30192-8.

[116] D. N. Nicholson and C. S. Greene, “Constructing knowledge graphs and their biomedical applications,” *Computational and Structural Biotechnology Journal*, vol. 18, pp. 1414–1428, 2020, doi: 10.1016/j.csbj.2020.05.017.

[117] Z. Zhu, M. Qu, S. Xu, and J. Tang, “GraphVite: A high-performance CPU-GPU hybrid system for node embedding,” *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, pp. 2494–2504, 2019, doi: 10.1145/3308558.3313508.

[118] D. N. Sosa, A. Derry, M. Guo, E. Wei, C. Brinton, and R. B. Altman, “A Literature-Based Knowledge Graph Embedding Method for Identifying Drug Repurposing Opportunities in Rare Diseases,” *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, vol. 25, pp. 463–474, 2020, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31797619.

[119] D. M. Gysi *et al.*, “Network medicine framework for identifying drug repurposing opportunities for COVID-19,” *arXiv*, 2020.

[120] BenevolentAI, “BenevolentAI Mission.” https://www.benevolent.com/about-us (accessed Jan. 03, 2021).

[121] P. Richardson *et al.*, “Baricitinib as potential treatment for 2019-nCoV acute respiratory disease,” *The Lancet*, vol. 395, no. 10223, pp. e30–e31, 2020, doi: 10.1016/S0140-6736(20)30304-4.

[122] A. C. Kalil *et al.*, “Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19.,” *The New England journal of medicine*, pp. 1–13, 2020, doi: 10.1056/NEJMoa2031994.

[123] P. R. U.S. Food and Drug Administration, “Coronavirus (COVID-19) Update: FDA Authorizes Drug Combination for Treatment of COVID-19.” https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-drug-combination-treatment-covid-19 (accessed Jan. 04, 2021).

[124] X. Zeng *et al.*, “Repurpose Open Data to Discover Therapeutics for COVID-19 Using Deep Learning,” *Journal of proteome research*, vol. 19, no. 11, pp. 4624–4636, 2020, doi: 10.1021/acs.jproteome.0c00316.

[125] Y. Ge *et al.*, “A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19,” pp. 1–62, 2020, doi: 10.1101/2020.03.11.986836.

[126] A. Zhavoronkov, Q. Vanhaelen, and T. I. Oprea, “Will Artificial Intelligence for Drug Discovery Impact Clinical Pharmacology?,” *Clinical Pharmacology and Therapeutics*, vol. 107, no. 4, pp. 780–785, 2020, doi: 10.1002/cpt.1795.

[127] A. Zhavoronkov *et al.*, “Potential 2019-nCoV 3C-like protease inhibitors designed using generative deep learning approaches,” *ChemRxiv*, 2020, doi: 10.26434/chemrxiv.11829102.v2.

[128] B. Tang, F. He, D. Liu, M. Fang, Z. Wu, and D. Xu, “AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2,” *bioRxiv : the preprint server for biology*, 2020, doi: 10.1101/2020.03.03.972133.

[129] A. Razzaq, “CogMol: Framework developed by IBM based on deep learning to accelerate therapeutic developments for COVID-19,” *Marktechpost*, 2020, [Online]. Available: https://www.marktechpost.com/2020/04/05/cogmol-framework-developed-by-ibm-based-on-deep-learning-generative-modeling-to-accelerate-therapeutic-developments-for-covid-19/.

[130] V. Chenthamarakshan *et al.*, “Target-Specific and selective drug design for covid-19 using deep generative models,” *arXiv*, pp. 1–29, 2020.

[131] IBM Research, “Explore novel drug candidates for COVID-19.” https://covid19-mol.mybluemix.net/ (accessed Jan. 25, 2021).

[132] N. Bung, S. R. Krishnan, G. Bulusu, and A. Roy, “De Novo Design of New Chemical Entities (NCEs) for SARS-CoV-2 Using Artificial Intelligence,” no. March, 2020, doi: 10.26434/chemrxiv.11998347.

[133] K. Kaushal and P. Sarma, “Emerging role of artificial intelligence in therapeutics for COVID-19 : a systematic review,” *Journal of Biomolecular Structure and Dynamics*, vol. 0, no. 0, pp. 1–16, 2020, doi: 10.1080/07391102.2020.1855250.

[134] C. J. Janeway, P. Travers, and M. Walport, “Immunobiology: The Immune System in Health and Disease,” in *Immunobiology: The Immune System in Health and Disease*, 5th ed., G. Science, Ed. .

[135] E. Fast, R. B. Altman, and B. Chen, “Potential T-cell and B-cell Epitopes of 2019-nCoV,” pp. 1–9, 2020.

[136] B. Malone *et al.*, “Artificial intelligence predicts the immunogenic landscape of SARS-CoV-2: toward universal blueprints for vaccine designs,” *Scientific Reports*, pp. 1–14, 2020, doi: 10.1101/2020.04.21.052084.

[137] E. Ong, M. U. Wong, A. Huffman, and Y. He, “COVID-19 Coronavirus Vaccine Design Using Reverse Vaccinology and Machine Learning,” *Frontiers in immunology*, vol. 11, no. July, p. 1581, 2020, doi: 10.3389/fimmu.2020.01581.

[138] J. W. Scannell, A. Blanckley, H. Boldon, and B. Warrington, “Diagnosing the decline in pharmaceutical R&D efficiency,” *Nature Reviews Drug Discovery*, vol. 11, no. 3, pp. 191–200, 2012, doi: 10.1038/nrd3681.

[139] C. H. Wong, K. W. Siah, and A. W. Lo, “Estimation of clinical trial success rates and related parameters,” *Biostatistics*, vol. 20, no. 2, pp. 273–286, Apr. 2019, doi: 10.1093/biostatistics/kxx069.

[140] S. Harrer, P. Shah, B. Antony, and J. Hu, “Artificial Intelligence for Clinical Trial Design,” *Trends in Pharmacological Sciences*, vol. 40, no. 8, pp. 577–591, 2019, doi: 10.1016/j.tips.2019.05.005.

[141] M. Molteni, “Meet the Company Trying to Democratize Clinical Trials With AI,” *Wired*, 2018, [Online]. Available: https://www.wired.com/story/meet-the-company-trying-to-democratize-clinical-trials-with-ai/.

[142] M. Robinson, “Deep6 AI - Why diversity is more important than ever for clinical trials.” https://deep6.ai/why-diversity-is-more-important-than-ever-for-clinical-trials/ (accessed Jan. 04, 2021).

[143] S. Buhr, “Mendel.ai nabs $2 million to match cancer patients with the latest clinical trials,” 2020.

[144] W. R. Zame *et al.*, “Machine learning for clinical trials in the era of COVID-19,” *Statistics in Biopharmaceutical Research*, vol. 12, no. 4, pp. 506–517, 2020, doi: 10.1080/19466315.2020.1797867.

[145] Merriam-Webster Dictionary, “Words We’re Watching: ‘Infodemic,’” *2020*. https://www.merriam-webster.com/words-at-play/words-were-watching-infodemic-meaning.

[146] B. Zimmer, “‘Infodemic’: When Unreliable Information Spreads Far and Wide,” *The Wall Street Journal*, pp. 2–5, 2020.

[147] J. Gómez-Gardeñes, L. Lotero, S. N. Taraskin, and F. J. Pérez-Reche, “Explosive Contagion in Networks,” *Scientific Reports*, vol. 6, no. November 2015, pp. 1–9, 2016, doi: 10.1038/srep19767.

[148] WHO, “Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation,” 2020. [Online]. Available: https://www.who.int/news-room/detail/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation.

[149] R. Schulman and D. Siman-Tov, “From Biological Weapons to Miracle Drugs: Fake News about the Coronavirus Pandemic,” *Institute for National Security Studies*, no. 1275, pp. 0–5, 2020, [Online]. Available: https://www.jstor.org/stable/resrep23529.

[150] R. Gallotti, F. Valle, N. Castaldo, P. Sacco, and M. de Domenico, “Assessing the risks of ‘infodemics’ in response to COVID-19 epidemics,” *Nature Human Behaviour*, vol. 4, no. December, 2020, doi: 10.1038/s41562-020-00994-6.

[151] CoMuNelab, “Covid19 Infodemics Observatory,” 2020. https://covid19obs.fbk.eu/ (accessed Jan. 06, 2021).

[152] M. Cinelli *et al.*, “The COVID-19 social media infodemic,” *Scientific Reports*, vol. 10, no. 1, pp. 1–10, 2020, doi: 10.1038/s41598-020-73510-5.

[153] T. Wang, K. Lu, K. P. Chow, and Q. Zhu, “COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model,” *IEEE Access*, vol. 8, pp. 138162–138169, 2020, doi: 10.1109/ACCESS.2020.3012595.

[154] L. Schild, C. Ling, J. Blackburn, G. Stringhini, Y. Zhang, and S. Zannettou, “‘Go eat a bat, Chang!’: An Early Look on the Emergence of Sinophobic Behavior on Web Communities in the Face of COVID-19,” *arXiv*, no. April, 2020.

[155] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, “The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge,” *Wspolczesna Onkologia*, vol. 1A, pp. A68–A77, 2015, doi: 10.5114/wo.2014.47136.

[156] C. Sudlow *et al.*, “UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age,” *PLoS Medicine*, vol. 12, no. 3, pp. 1–10, 2015, doi: 10.1371/journal.pmed.1001779.

[157] V. Ljosa, K. L. Sokolnicki, and A. E. Carpenter, “Annotated high-throughput microscopy image sets for validation,” *Nature methods*, vol. 9, no. 7, p. 637, Jun. 2012, doi: 10.1038/nmeth.2083.

[158] E. Williams *et al.*, “Image Data Resource: A bioimage data integration and publication platform,” *Nature Methods*, vol. 14, no. 8, pp. 775–781, 2017, doi: 10.1038/nmeth.4326.

[159] National Institutes of Health - Office of Data Science Strategy, “Open-Access Data and Computational Resources to Address COVID-19.” https://datascience.nih.gov/covid-19-open-access-resources (accessed Jan. 16, 2021).

[160] E. Dong, H. Du, and L. Gardner, “An interactive web-based dashboard to track COVID-19 in real time,” *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, 2020, doi: 10.1016/S1473-3099(20)30120-1.

[161] DXY-DX Doctor, “COVID-19 Global Pandemic Real-time Report.” https://ncov.dxy.cn/ncovh5/view/en_pneumonia (accessed Jan. 31, 2021).

[162] 1Point3Acres, “Global COVID-19 Tracker and Interactive Charts.” https://coronavirus.1point3acres.com/ (accessed Jan. 31, 2021).

[163] Worldometer, “COVID-19 Coronavirus Pandemic.” https://www.worldometers.info/coronavirus/ (accessed Jan. 31, 2021).

[164] The COVID Tracking Project, “The COVID Tracking Project collects and publishes the most complete data about COVID-19 in the US.” https://covidtracking.com/ (accessed Jan. 31, 2021).

[165] The New York Times, “We’re Sharing Coronavirus Case Data for Every U.S. County.”

[166] COVID-19 Response Greece, “Data Analytics.” https://www.covid19response.gr/project5_en.html (accessed Jan. 08, 2021).

[167] Hellenic National Public Health Organization, “COVID-19 Daily Updates.” https://eody.gov.gr/category/covid-19/ (accessed Jan. 08, 2021).

[168] J.P. Cohen et al, “covid-chestxray-dataset.” https://github.com/ieee8023/covid-chestxray-dataset (accessed Jan. 08, 2021).

[169] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, “COVID-19 Image Data Collection: Prospective predictions are the future,” *arXiv*, pp. 1–38, 2020.

[170] NHS, “National COVID-19 Chest Image Database (NCCID).” https://nhsx.github.io/covid-chest-imaging-database/ (accessed Jan. 08, 2021).

[171] covid19challenge.eu, “A platform for open research on large Covid-19 imaging datasets.” https://www.covid19challenge.eu/ (accessed Jan. 08, 2021).

[172] GISAID, “Enabling rapid and open access to epidemic and pandemic virus data.” https://www.gisaid.org/about-us/mission/ (accessed Jan. 08, 2021).

[173] RCSB Protein Data Bank (PDB), “Enabling Breakthroughs in Scientific and Biomedical Research and Education.” https://www.rcsb.org/pages/about-us/index (accessed Jan. 08, 2021).

[174] American Chemical Society CAS, “CAS Releases Open Access Dataset of Antiviral Chemical Compounds to Aid COVID-19 Discovery and Analysis,” 2020, [Online]. Available: https://www.cas.org/resources/press-releases/open-access-covid-19-dataset.

[175] WHO, “Global research on coronavirus disease (COVID-19).” https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov (accessed Jan. 08, 2021).

[176] L. L. Wang *et al.*, “CORD-19: The COVID-19 open research dataset,” *arXiv*, 2020.

[177] Kaggle, “COVID-19 Open Research Dataset Challenge (CORD-19).” https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge (accessed Jan. 09, 2021).

[178] Q. Chen, A. Allot, and Z. Lu, “Keep up with the latest coronavirus research.,” *Nature*, vol. 579, no. 7798. England, p. 193, Mar. 2020, doi: 10.1038/d41586-020-00694-1.

[179] Elsevier, “Novel Coronavirus Information Center.” https://www.elsevier.com/connect/coronavirus-information-center (accessed Jan. 09, 2021).

[180] The Lens, “COVID-19 Datasets.” https://about.lens.org/covid-19/ (accessed Jan. 09, 2021).

[181] J. M. Banda *et al.*, “A large-scale COVID-19 twitter chatter dataset for open scientific research - An international collaboration,” *arXiv*, 2020.

[182] E. Chen, K. Lerman, and E. Ferrara, “Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set,” *arXiv*, vol. 6, 2020, doi: 10.2196/19273.

[183] J. Yu, “Open access institutional and news media tweet dataset for COVID-19 social science research,” *arXiv*, pp. 1–8, 2020.

[184] Facebook, “Data for Good.” https://dataforgood.fb.com/ (accessed Jan. 09, 2021).

[185] Facebook, “Data for Good - Disease Prevention Maps.” https://dataforgood.fb.com/tools/disease-prevention-maps/ (accessed Jan. 09, 2021).

[186] Google, “COVID-19 Community Mobility Reports.” https://www.google.com/covid19/mobility/ (accessed Jan. 31, 2021).

[187] L. Wynants *et al.*, “Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal,” *The BMJ*, vol. 369, 2020, doi: 10.1136/bmj.m1328.

[188] Μ. Ανδριωτάκης, “Οι μεροληψίες και η ηθική της τεχνητής νοημοσύνης,” *Η Καθημερινή*, pp. 1–11, 2021, [Online]. Available: https://www.kathimerini.gr/culture/books/561254644/oi-merolipsies-kai-i-ithiki-tis-technitis-noimosynis-k/.

[189] L. Floridi *et al.*, “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations,” *Minds and Machines*, vol. 28, no. 4, pp. 689–707, 2018, doi: 10.1007/s11023-018-9482-5.

[190] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, “Explainability for artificial intelligence in healthcare: a multidisciplinary perspective,” *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1–9, 2020, doi: 10.1186/s12911-020-01332-6.

[191] G. Montavon, W. Samek, and K. R. Müller, “Methods for interpreting and understanding deep neural networks,” *Digital Signal Processing: A Review Journal*, vol. 73, pp. 1–15, 2018, doi: 10.1016/j.dsp.2017.10.011.

[192] Defence Advanced Research Projects Agency and M. Turek, “Explainable Artificial Intelligence (XAI).” https://www.darpa.mil/program/explainable-artificial-intelligence (accessed Jan. 16, 2021).

[193] M. Kritikos, “What if we could fight coronavirus with artificial intelligence?” [Online]. Available: https://www.europarl.europa.eu/RegData/etudes/ATAG/2020/641538/EPRS_ATA(2020)641538_EN.pdf.

[194] A. Tzachor, J. Whittlestone, L. Sundaram, and S. hÉigeartaigh, “Artificial intelligence in a crisis needs ethics with urgency,” *Nature Machine Intelligence*, vol. 2, no. 7, pp. 365–366, 2020, doi: 10.1038/s42256-020-0195-0.

[195] World Health Organization, “Guidance for Managing Ethical Issues in Infectious Disease Outbreaks,” 2016. [Online]. Available: https://apps.who.int/iris/handle/10665/250580.

[196] J. L. Raisaro *et al.*, “MEDCO: Enabling secure and privacy-preserving exploration of distributed clinical and genomic data,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 4, pp. 1328–1341, 2019, doi: 10.1109/TCBB.2018.2854776.

[197] G. Singh and J. Levi, “MiPasa project and IBM Blockchain team on open data platform to support Covid-19 response,” 2020.

[198] Q. Yang, Y. Liu, Tianjian Chen, and Yongxin Tong, “Federated Machine Learning: Concept and Applications,” *arXiv*, vol. 10, no. 2, pp. 1–19, 2019.

[199] J. Wiens *et al.*, “Do no harm: a roadmap for responsible machine learning for health care,” *Nature Medicine*, pp. 15–18, 2019, doi: 10.1038/s41591-019-0548-6.