

Ιατρικά δεδομένα στην εποχή των big data: Διαχείριση και Διάχυση

ΓΕΩΡΓΙΟΣ ΧΑΛΚΙΟΠΟΥΛΟΣ

ΠΜΣ ΔΗΜΟΣΙΑ ΥΓΕΙΑ ΙΑΤΡΙΚΗΣ ΠΑΤΡΩΝ

ΕΠΙΒΛΕΠΟΥΣΑ ΚΑΘΗΓΗΤΡΙΑ: ΕΛΕΝΗ ΓΕΛΑΣΤΟΠΟΥΛΟΥ – ΜΕΛΗ
ΤΡΙΜΕΛΟΥΣ ΕΠΙΤΡΟΠΗΣ: ΞΕΝΟΦΩΝ ΣΙΝΩΠΙΔΗΣ, ΚΑΡΟΛΙΝΑ
ΑΚΙΝΟΣΟΓΛΟΥ

“Data is the new oil.”

(Clive Humby, Mathematician)

“Τα δεδομένα είναι το νέο πετρέλαιο.”

(Clive Humby, Μαθηματικός)

Περιεχόμενα

Περιεχόμενα.....	2
Πρόλογος.....	4
Κεφάλαιο 1ο – Εισαγωγή στην έννοια των Big Data.....	10
1.1 Η έννοια.....	10
1.2 Η παραγωγή των μεγάλων δεδομένων.....	12
1.3 Οι τύποι των Big Data.....	14
Κεφάλαιο 2ο Τα ιατρικά δεδομένα και η σημασία τους.....	17
2.1 Εισαγωγή στην έννοια και τη σημασία της.....	17
2.1.1 Για το άτομο	17
2.1.2 Για την κοινωνία	18
2.1.3 Για την έρευνα.....	19
2.2 Προκλήσεις.....	21
2.3 Παραδείγματα δημιουργίας ιατρικών δεδομένων.....	22
2.3.1 Ηλεκτρονικά αρχεία υγείας.....	22
2.3.2 Ιατρική απεικόνιση.....	24
2.3.3 Γονιδιωματική αλληλουχία	24
2.3.4 Φορητές συσκευές και αισθητήρες	26
2.3.5 Κλινικές δοκιμές και ερευνητικές μελέτες	27
2.3.6 Επιτήρηση της δημόσιας υγείας	28
Κεφάλαιο 3ο - Διεθνείς οργανισμοί.....	29
3.1 Εισαγωγή	29
3.2 Health Level Seven International (HL7)	30
3.3 International Medical Informatics Association (IMIA)	31
3.4 European Institute for Health Records (EuroRec)	33
3.5 Health Data Research UK (HDR UK).....	34
3.6 Institute for Health Metrics and Evaluation (IHME)	36
Κεφάλαιο 4ο - Η διαχείριση μεγάλων δεδομένων στην ιατρική	39
4.1 Εισαγωγή	39
4.2 Απόκτηση και ενσωμάτωση	39
4.3 Αποθήκευση και υποδομές.....	42
4.4 Ανάλυση δεδομένων	45
4.5 Ποιότητα δεδομένων και προεπεξεργασία	48
4.6 Ασφάλεια δεδομένων	50

Κεφάλαιο 5° – Διάχυση των ιατρικών δεδομένων	56
5.1 Πλατφόρμες ανταλλαγής δεδομένων	56
5.1.1 Τύποι.....	56
5.1.2 Παραδείγματα	59
5.2 Πρότυπα δεδομένων	60
5.2.1 To Health Level Seven (HL7)	62
5.2.2 Η International Classification of Diseases (ICD).....	63
5.2.3 Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT)...	65
5.2.4 Logical Observation Identifiers Names and Codes (LOINC).....	66
5.3 Η ανάγκη εκπαίδευσης στον χώρο των ιατρικών δεδομένων	67
Επίλογος	70
Βιβλιογραφία	73

Πίνακας εικόνων

Εικόνα 1. Οι διαστάσεις των μεγάλων δεδομένων (Ristevski and Chen, 2018)	12
Εικόνα 2 . Η αγορά μεγάλων δεδομένων μέχρι το 2025 στην υγεία (BIS Research, 2023)	28
Εικόνα 3. Πρόβλεψη επενδύσεων στον χώρο των μεγάλων δεδομένων στην υγεία (Capitalist and Publicis, 2023)	55

Περίληψη

Η ταχεία πρόοδος των ψηφιακών τεχνολογιών και η εμφάνιση των μεγάλων δεδομένων έχουν φέρει επανάσταση στον κλάδο της υγειονομικής περίθαλψης, προσφέροντας τεράστιες ευκαιρίες για τη βελτίωση της περίθαλψης των ασθενών, την ενίσχυση της ιατρικής έρευνας και τη βελτιστοποίηση των πρωτοβουλιών για τη δημόσια υγεία. Ωστόσο, η αποτελεσματική διαχείριση και διάδοση των δεδομένων υγείας σε αυτή την εποχή των μεγάλων δεδομένων παρουσιάζει πολυάριθμες προκλήσεις που πρέπει να αντιμετωπιστούν για να διασφαλιστεί η πλήρης αξιοποίηση του δυναμικού τους. Στόχος της παρούσας διατριβής είναι να διερευνήσει τη διαχείριση και τη διάδοση των δεδομένων υγείας στην εποχή των μεγάλων δεδομένων.

Η έρευνα περιλαμβάνει εκτεταμένης ανασκόπησης του σημερινού τοπίου των πρακτικών διαχείρισης δεδομένων υγείας, εστιάζοντας στις προκλήσεις που θέτει η εκθετική αύξηση του όγκου, της ποικιλίας, της ταχύτητας και της αληθοφάνειας των δεδομένων. Διερευνά τα υφιστάμενα πλαίσια διακυβέρνησης δεδομένων, τις ανησυχίες για την προστασία της ιδιωτικής ζωής και τις κανονιστικές απαιτήσεις για τον εντοπισμό κενών και τομέων που απαιτούν βελτίωση. Υπογραμμίζει τη σημασία της λήψης αποφάσεων βάσει δεδομένων στην παροχή υγειονομικής περίθαλψης, στον προγραμματισμό της δημόσιας υγείας και στην ιατρική έρευνα, ενώ υπογραμμίζει την ανάγκη για κατανόηση των δεδομένων από διάφορους ενδιαφερόμενους. Επιπλέον, η διατριβή εξετάζει δεοντολογικά ζητήματα και νομικές πτυχές που σχετίζονται με τη διαχείριση και τη διάδοση δεδομένων υγείας. Εξετάζει τη σημασία της συγκατάθεσης μετά από ενημέρωση, των τεχνικών αυτοματοποίησης και της ανωνυμοποίησης δεδομένων για την προστασία της ιδιωτικής ζωής των ασθενών και τη συμμόρφωση με τα κανονιστικά πλαίσια.

Τα αποτελέσματα αυτής της έρευνας θα συμβάλουν στην ανάπτυξη βέλτιστων πρακτικών στη διαχείριση και τη διάδοση δεδομένων υγείας, ενημερώνοντας τους υπεύθυνους χάραξης πολιτικής, τους οργανισμούς υγειονομικής περίθαλψης και τους ερευνητές σχετικά με τις στρατηγικές και τα εργαλεία που απαιτούνται για την αξιοποίηση της δυναμικής των μεγάλων δεδομένων, αντιμετωπίζοντας παράλληλα

τις ανησυχίες για την προστασία της ιδιωτικής ζωής και τους ηθικούς προβληματισμούς.

Πρόλογος

Στην εποχή των μεγάλων δεδομένων (Big Data)¹, ο τομέας της ιατρικής έχει υποστεί βαθύτατο μετασχηματισμό με την έλευση προηγμένων τεχνολογιών και την εκθετική αύξηση των ιατρικών δεδομένων. Τα ιατρικά δεδομένα, τα οποία περιλαμβάνουν ένα τεράστιο φάσμα πληροφοριών που σχετίζονται με την υγεία των ασθενών, τη διάγνωση, τις θεραπείες και τα αποτελέσματα εξετάσεων, έχουν καταστεί πολύτιμος πόρος με τεράστιες δυνατότητες για τη βελτίωση της παροχής υγειονομικής περίθαλψης, την προώθηση της ιατρικής έρευνας και τελικά τη βελτίωση των αποτελεσμάτων των ασθενών. Η σημασία των ιατρικών δεδομένων στην εποχή των μεγάλων δεδομένων είναι σημαντική, καθώς παρέχουν πρωτοφανείς ευκαιρίες για την κατανόηση των ασθενειών, την προσαρμογή των θεραπειών και τη βελτιστοποίηση των πρακτικών υγειονομικής περίθαλψης.

Ένας από τους κύριους λόγους για την αξιοποίηση και την ανάλυση ιατρικών δεδομένων έγκειται στην υπόσχεση της εξατομικευμένης ιατρικής (Viceconti, Hunter and Hose, 2015). Κάθε ασθενής είναι μοναδικός και η ανταπόκρισή του στη θεραπεία μπορεί να διαφέρει με βάση πλήθος παραγόντων, όπως η γενετική, ο τρόπος ζωής και οι περιβαλλοντικές επιδράσεις. Αξιοποιώντας τη δύναμη της ανάλυσης μεγάλων δεδομένων, οι επαγγελματίες της υγειονομικής περίθαλψης μπορούν να εμβαθύνουν σε μεγάλους όγκους ιατρικών δεδομένων για τον εντοπισμό μοτίβων, συσχετίσεων και προγνωστικών μοντέλων που τους επιτρέπουν να προσαρμόζουν τις θεραπείες και τις παρεμβάσεις σε μεμονωμένους ασθενείς. Αυτή η μετάβαση από μια προσέγγιση που ταιριάζει σε όλους σε ένα παράδειγμα εξατομικευμένης ιατρικής και ιατρικής ακριβείας έχει τη δυνατότητα να φέρει επανάσταση στην υγειονομική περίθαλψη, βελτιώνοντας την αποτελεσματικότητα της θεραπείας, μειώνοντας τα ανεπιθύμητα συμβάντα και μεγιστοποιώντας τα αποτελέσματα των ασθενών (Viceconti, Hunter and Hose, 2015).

Η διπλωματική εργασία με τίτλο «Ιατρικά δεδομένα στην εποχή των Big data» έχει ως στόχο να εξετάσει τη διαχείριση, ανάλυση και διάχυση ιατρικών δεδομένων στο

¹ Έχει καθιερωθεί ο αγγλικός όρος ακόμα και στον ελληνικό χώρο.

πλαίσιο της εποχής των Big data. Η εργασία επιδιώκει να εξετάσει τον τρόπο που οι τεχνολογίες Big data επηρεάζουν την ιατρική, προσφέροντας νέες ευκαιρίες για την κατανόηση των ασθενειών, τη βελτίωση της θεραπείας και τη βελτίωση των αποτελεσμάτων των ασθενών.

Οι συγκεκριμένοι στόχοι της διπλωματικής εργασίας περιλαμβάνουν:

1. Ανάλυση της σημασίας των ιατρικών δεδομένων και των τεχνολογιών Big data: Η εργασία θα εξετάσει τη σημασία των ιατρικών δεδομένων στο πλαίσιο της εποχής των Big data και τον τρόπο με τον οποίο επηρεάζουν την ιατρική έρευνα, την πρόληψη των ασθενειών και την παροχή υγειονομικής περίθαλψης. Ταυτόχρονα θα εξετάσει τις τεχνολογίες και τα εργαλεία που χρησιμοποιούνται για τη συλλογή, αποθήκευση, διαχείριση και ανάλυσή τους. Θα εστιάσει στην εξέλιξη των τεχνολογιών Big data και την εφαρμογή τους στην ιατρική πρακτική.
2. Ανάλυση των προκλήσεων της διαχείρισης ιατρικών δεδομένων στην εποχή των Big data: Η εργασία θα εξετάσει τις προκλήσεις που σχετίζονται με τη διαχείριση, ανάλυση και διάχυση ιατρικών δεδομένων. Θα εστιάσει στα ζητήματα ασφάλειας, προστασίας προσωπικών δεδομένων, νομοθεσίας και ηθικής που πρέπει να ληφθούν υπόψη για την αποτελεσματική διαχείριση των ιατρικών δεδομένων.
3. Εξερεύνηση των εφαρμογών των ιατρικών δεδομένων στην εποχή των Big data και των προκλήσεων που παρουσιάζουν: Η εργασία θα εξετάσει τις διάφορες εφαρμογές των ιατρικών δεδομένων στην ιατρική πρακτική, την έρευνα και τη διαχείριση των υπηρεσιών υγείας. Θα εστιάσει στην ανίχνευση πρόβλεψης και πρόληψης ασθενειών, τη βελτίωση της θεραπείας, την υποστήριξη αποφάσεων, την ανάλυση πληθωρικών δεδομένων και την αποτελεσματικότερη διαχείριση των υποδομών υγείας. Ταυτόχρονα θα προσφέρει μια κριτική αξιολόγηση των προκλήσεων και των πιθανοτήτων που συνδέονται με τη χρήση των ιατρικών δεδομένων στην εποχή των Big data. Θα εξετάσει την επιπτώσεις της ανάλυσης των ιατρικών δεδομένων στην

ιατρική πρακτική, την ποιότητα της περίθαλψης, την προστασία της ιδιωτικότητας και την αποδοτικότητα του συστήματος υγείας.

Το πεδίο εφαρμογής της διπλωματικής εργασίας είναι η ιατρική και η υγειονομική περίθαλψη. Η εργασία απευθύνεται σε ερευνητές, επαγγελματίες υγείας, διαχειριστές υγείας, πολιτικούς και άλλους ενδιαφερόμενους φορείς που ασχολούνται με τη διαχείριση και αξιοποίηση των ιατρικών δεδομένων στο πλαίσιο των Big data. Η έρευνα και οι αναλύσεις που θα πραγματοποιηθούν στη διπλωματική εργασία θα βοηθήσουν στην προώθηση της κατανόησης των πλεονεκτημάτων, των προκλήσεων και των καλύτερων πρακτικών για τη διαχείριση και αξιοποίηση των ιατρικών δεδομένων σε αυτήν την εποχή των Big data.

Κεφάλαιο 1ο – Εισαγωγή στην έννοια των Big Data

1.1 Η έννοια

Η έννοια των μεγάλων δεδομένων αναφέρεται στον τεράστιο όγκο δομημένων και μη δομημένων δεδομένων που παράγονται με μεγάλη ταχύτητα από διάφορες πηγές (Sagiroglu and Sinanc, 2013). Οι πηγές αυτές περιλαμβάνουν ενδεικτικά τις πλατφόρμες κοινωνικής δικτύωσης, τα δίκτυα αισθητήρων, τα συστήματα υγειονομικής περίθαλψης, τις χρηματοπιστωτικές συναλλαγές, τις διαδικτυακές δραστηριότητες και την επιστημονική έρευνα. Τα μεγάλα δεδομένα χαρακτηρίζονται από τον όγκο, την ταχύτητα, την ποικιλία, την ειλικρίνειά τους, την μεταβλητότητα και την ακρίβεια, γνωστά ως τα 6V (Volume, Velocity, Variety και Veracity). Ας μελετήσουμε διεξοδικά αυτές τις διαστάσεις (Ristevski and Chen, 2018):

1. Όγκος: Τα μεγάλα δεδομένα χαρακτηρίζονται από τον απόλυτο όγκο τους. Τα παραδοσιακά συστήματα επεξεργασίας δεδομένων συχνά δυσκολεύονται να διαχειριστούν τις τεράστιες ποσότητες δεδομένων που παράγονται καθημερινά. Με την έλευση νέων τεχνολογιών και την ευρεία χρήση ψηφιακών συσκευών, τα δεδομένα παράγονται σε πρωτοφανή κλίμακα. Αυτό περιλαμβάνει δεδομένα της τάξης των terabytes, των petabytes, των exabytes και πέραν αυτών.
2. Ταχύτητα: Η ταχύτητα των μεγάλων δεδομένων αναφέρεται στην ταχύτητα με την οποία παράγονται τα δεδομένα και πρέπει να υποβάλλονται σε επεξεργασία σε πραγματικό χρόνο ή σχεδόν σε πραγματικό χρόνο. Τα δεδομένα ρέουν συνεχώς από διάφορες πηγές, όπως ενημερώσεις στα μέσα κοινωνικής δικτύωσης, διαδικτυακές συναλλαγές και δεδομένα αισθητήρων. Η ικανότητα καταγραφής, επεξεργασίας και ανάλυσης δεδομένων σε πραγματικό χρόνο παρέχει στους οργανισμούς πολύτιμες πληροφορίες και τη δυνατότητα να λαμβάνουν γρήγορα τεκμηριωμένες αποφάσεις.
3. Ποικιλία: Τα μεγάλα δεδομένα περιλαμβάνουν μεγάλη ποικιλία τύπων δεδομένων, συμπεριλαμβανομένων δομημένων, ημιδομημένων και

αδόμητων δεδομένων. Τα δομημένα δεδομένα αναφέρονται σε καλά οργανωμένα δεδομένα που είναι αποθηκευμένα σε παραδοσιακές βάσεις δεδομένων, ενώ τα αδόμητα δεδομένα αναφέρονται σε δεδομένα που δεν έχουν προκαθορισμένη δομή, όπως μηνύματα ηλεκτρονικού ταχυδρομείου, βίντεο, εικόνες, αναρτήσεις στα μέσα κοινωνικής δικτύωσης και δεδομένα αισθητήρων. Τα ημιδομημένα δεδομένα βρίσκονται κάπου στο ενδιάμεσο, καθώς περιέχουν κάποια οργανωτικά στοιχεία αλλά δεν συμμορφώνονται με ένα αυστηρό σχήμα. Η ικανότητα αξιοποίησης και ανάλυσης διαφορετικών τύπων δεδομένων αποτελεί βασική πτυχή της ανάλυσης μεγάλων δεδομένων.

4. Αλήθεια: Η ειλικρίνεια αναφέρεται στην ποιότητα και την αξιοπιστία των δεδομένων. Με τον πολλαπλασιασμό των πηγών δεδομένων και την πιθανότητα σφαλμάτων, ασυνέπειας και μεροληψίας, η διασφάλιση της ειλικρίνειας των μεγάλων δεδομένων καθίσταται ζωτικής σημασίας. Τα δεδομένα ενδέχεται να περιέχουν ανακρίβειες, ελλιπείς τιμές ή θόρυβο, γεγονός που μπορεί να επηρεάσει τις διαδικασίες ανάλυσης και λήψης αποφάσεων. Οι τεχνικές καθαρισμού, επικύρωσης και διασφάλισης ποιότητας δεδομένων είναι απαραίτητες για την αντιμετώπιση των προκλήσεων αληθοφάνειας που σχετίζονται με τα μεγάλα δεδομένα.
5. Μεταβλητότητα: Η μεταβλητότητα αφορά στη συνέπεια που διακρίνει τα δεδομένα στο πέρασμα του χρόνου.
6. Ακρίβεια: Η ποιότητα των δεδομένων είναι σημαντική



Εικόνα 1. Οι διαστάσεις των μεγάλων δεδομένων (Ristevski and Chen, 2018)

1.2 Η παραγωγή των μεγάλων δεδομένων

Όμως πως παράγονται τα δεδομένα αυτά; Γιατί σήμερα ζούμε στην εποχή τις τεράστιας παραγωγής τους; Στη συγκεκριμένη ενότητα θα μελετήσουμε τον τρόπο παραγωγής τους, προκειμένου να κατανοήσουμε την ιδιαιτερότητα της εποχής. Τα μεγάλα δεδομένα παράγονται με διάφορα μέσα. Καταρχάς τα δεδομένα παράγονται από τις διαδικτυακές αλληλεπιδράσεις (Chen, Mao, Zhang, Leung, Chen, *et al.*, 2014). Πλατφόρμες μέσω κοινωνικής δικτύωσης, ιστότοποι και διαδικτυακές υπηρεσίες παράγουν τεράστιες ποσότητες δεδομένων μέσω των αλληλεπιδράσεων των χρηστών. Αυτό περιλαμβάνει αναρτήσεις, σχόλια, likes/ dislikes, κοινοποιήσεις, κλικ και συμπεριφορές περιήγησης, οι οποίες είναι πολύτιμες για την κατανόηση των προτιμήσεων των χρηστών, των τάσεων και της ανάλυσης συναισθήματος. Σε δεύτερο επίπεδο τα δεδομένα παράγονται από συσκευές Internet of Things (IoT)

(Chen, Mao, Zhang, Leung and others, 2014). Ο πολλαπλασιασμός των συσκευών IoT, όπως αισθητήρες, wearables και έξυπνες συσκευές, παράγει τεράστιες ποσότητες δεδομένων. Αυτές οι συσκευές συλλέγουν και μεταδίδουν δεδομένα σχετικά με τις περιβαλλοντικές συνθήκες, τις μετρήσεις υγείας, τη χρήση ενέργειας και πολλά άλλα. Τα δεδομένα που παράγονται από το IoT παρέχουν πληροφορίες για τη βελτίωση της αποδοτικότητας, τη βελτιστοποίηση των διαδικασιών και τη δυνατότητα προληπτικής συντήρησης.

Παράλληλα παρατηρείται μεγάλη χρήση αισθητήρων. Τα αυτοματοποιημένα συστήματα, όπως οι βιομηχανικοί αισθητήρες, οι κάμερες παρακολούθησης και τα αρχεία καταγραφής διακομιστών, παράγουν συνεχώς δεδομένα (Lyko, Nitzschke and Ngonga Ngomo, 2016). Αυτές οι πηγές παράγουν μεγάλους όγκους δεδομένων που απαιτούν παρακολούθηση, ανάλυση και ανίχνευση ανωμαλιών σε πραγματικό χρόνο. Η μεγάλη χρήση του διαδικτύου και ειδικά μετά την πανδημία παρατηρείται αλματώδης αύξηση των δεδομένων συναλλαγών. Οι χρηματοοικονομικές συναλλαγές, οι δραστηριότητες ηλεκτρονικού εμπορίου και οι αλληλεπιδράσεις με τους πελάτες παράγουν τεράστιους όγκους δεδομένων. Αυτά περιλαμβάνουν ιστορικό αγορών, αρχεία συναλλαγών, προφίλ πελατών και δεδομένα ροής κλικ. Η ανάλυση των δεδομένων συναλλαγών βοηθά τις επιχειρήσεις να κατανοήσουν τη συμπεριφορά των καταναλωτών, να ανιχνεύσουν την απάτη και να εξατομικεύσουν τις εμπειρίες των πελατών (Viceconti, Hunter and Hose, 2015).

Τέλος δεν πρέπει να ξεχνάμε και την επιστημονική έρευνα, η οποία λόγω πίεσης των ακαδημαϊκών ιδρυμάτων έχει αυξηθεί εκθετικά (Chen, Mao, Zhang, Leung, Chen, *et al.*, 2014). Τα επιστημονικά πειράματα, οι προσομοιώσεις και οι παρατηρήσεις παράγουν σημαντικές ποσότητες δεδομένων σε τομείς όπως η γονιδιωματική, η αστρονομία, η κλιματική έρευνα και η σωματιδιακή φυσική. Αυτά τα σύνολα δεδομένων απαιτούν συχνά προηγμένες τεχνικές ανάλυσης και υπολογισμού για την εξαγωγή ουσιαστικών γνώσεων και την υποστήριξη επιστημονικών ανακαλύψεων.

Όπως έγινε κατανοητό, τα μεγάλα δεδομένα παράγονται από ποικίλες πηγές και απαιτούν προηγμένες τεχνολογίες και μεθοδολογίες για την αποτελεσματική καταγραφή, αποθήκευση, επεξεργασία και ανάλυση των δεδομένων.

1.3 Οι τύποι των Big Data

Τα μεγάλα δεδομένα είναι ένας όρος που περιλαμβάνει ένα ευρύ φάσμα τύπων δεδομένων, καθένα από τα οποία έχει τα δικά του μοναδικά χαρακτηριστικά και συνέπειες για την ανάλυση. Για να αναλύσουμε διεξοδικά τους τύπους μεγάλων δεδομένων, ας εμβαθύνουμε σε κάθε κατηγορία λεπτομερώς. Καταρχάς υπάρχουν τα δομημένα δεδομένα. Τα δομημένα δεδομένα αναφέρονται σε δεδομένα που έχουν μια σαφώς καθορισμένη μορφή και εντάσσονται σε ένα προκαθορισμένο σχήμα. Συνήθως είναι οργανωμένα και αποθηκεύονται σε σχεσιακές βάσεις δεδομένων, καθιστώντας τα εύκολα προσβάσιμα και αναλύσιμα. Τα δομημένα δεδομένα χαρακτηρίζονται από την ομοιομορφία τους, με σταθερά πεδία και προκαθορισμένους τύπους δεδομένων. Αυτός ο τύπος δεδομένων προσφέρεται για παραδοσιακές τεχνικές επεξεργασίας δεδομένων, όπως ερωτήματα SQL και σχεσιακές πράξεις. Παραδείγματα δομημένων δεδομένων περιλαμβάνουν αρχεία συναλλαγών, πληροφορίες πελατών, οικονομικά δεδομένα και τυποποιημένες μετρήσεις αισθητήρων (Sagiroglu and Sinanc, 2013).

Σε δεύτερο επίπεδο έχουμε τα ημιδομημένα δεδομένα. Τα ημιδομημένα δεδομένα γεφυρώνουν το χάσμα μεταξύ δομημένων και μη δομημένων δεδομένων. Διαθέτουν κάποια οργανωτικά στοιχεία, αλλά δεν τηρούν αυστηρά ένα προκαθορισμένο σχήμα (Hamouda and Zainol, 2019). Τα ημιδομημένα δεδομένα χαρακτηρίζονται από την ευελιξία και την αυτοπεριγραφική τους φύση. Συχνά χρησιμοποιούν ετικέτες, ετικέτες ή χαρακτηριστικά για να παρέχουν ένα επίπεδο οργάνωσης, ενώ παράλληλα επιτρέπουν τη διαφοροποίηση της δομής των δεδομένων. Αυτός ο τύπος δεδομένων μπορεί να αναπαρασταθεί σε μορφές όπως XML, JSON ή CSV. Οι πηγές ημιδομημένων δεδομένων περιλαμβάνουν αρχεία καταγραφής, αναρτήσεις στα μέσα κοινωνικής δικτύωσης, μηνύματα ηλεκτρονικού ταχυδρομείου, δεδομένα αισθητήρων με ποικίλες μορφές και άλλα δεδομένα που μπορεί να έχουν κάποια δομή αλλά δεν διαθέτουν πλήρη τυποποίηση.

Σε τρίτο επίπεδο παράγονται αδόμητα δεδομένα. Τα αδόμητα δεδομένα αντιπροσωπεύουν τη μεγαλύτερη και πιο σύνθετη κατηγορία στο πλαίσιο των μεγάλων δεδομένων (Kitchin and McArdle, 2016). Αναφέρονται σε δεδομένα που δεν

διαθέτουν συγκεκριμένη δομή ή μορφή και δεν χωρούν εύκολα σε παραδοσιακές βάσεις δεδομένων. Τα μη δομημένα δεδομένα είναι συνήθως βαριά σε κείμενο, αλλά μπορεί επίσης να περιλαμβάνουν περιεχόμενο πολυμέσων, όπως εικόνες, αρχεία ήχου και αρχεία βίντεο. Η ανάλυση μη δομημένων δεδομένων παρουσιάζει σημαντικές προκλήσεις λόγω της έλλειψης οργάνωσης. Προηγμένες τεχνικές, όπως η επεξεργασία φυσικής γλώσσας, η μηχανική μάθηση και η όραση υπολογιστών, είναι απαραίτητες για την εξαγωγή ουσιαστικών πληροφοριών. Παραδείγματα μη δομημένων δεδομένων είναι τα έγγραφα κειμένου, το περιεχόμενο των μέσων κοινωνικής δικτύωσης, οι εικόνες, οι ηχογραφήσεις, το υλικό βίντεο, οι ιστοσελίδες και τα δεδομένα αισθητήρων χωρίς συγκεκριμένη μορφοποίηση.

Μία άλλη κατηγορία μεγάλων δεδομένων είναι τα δεδομένα χρονοσειρών. Τα δεδομένα χρονοσειρών χαρακτηρίζονται από τη διαδοχική τους φύση και καταγράφονται με την πάροδο του χρόνου σε τακτά χρονικά διαστήματα (Rakthanmanon *et al.*, 2013). Καταγράφουν μετρήσεις ή τιμές σε διαδοχικά χρονικά σημεία, επιτρέποντας την ανάλυση τάσεων, προτύπων και εποχικότητας. Τα δεδομένα χρονοσειρών χρησιμοποιούνται συχνά σε συστήματα πρόβλεψης, ανάλυσης τάσεων και παρακολούθησης. Συνδέονται με μια χρονοσφραγίδα για κάθε σημείο δεδομένων, διευκολύνοντας τη διαχρονική ανάλυση. Παραδείγματα δεδομένων χρονοσειρών περιλαμβάνουν δεδομένα χρηματιστηριακών αγορών, δεδομένα καιρού, δεδομένα αισθητήρων IoT (π.χ. θερμοκρασία, υγρασία, πίεση), δεδομένα επισκεψιμότητας ιστοτόπων και δεδομένα κατανάλωσης ενέργειας.

Η ανάπτυξη των γεωγραφικών συστημάτων οδήγησε παράλληλα στην ανάπτυξη γεωχωρικών δεδομένων (Goodchild, 2016). Τα γεωχωρικά δεδομένα ενσωματώνουν μια χωρική συνιστώσα και αντιπροσωπεύουν δεδομένα που σχετίζονται με συγκεκριμένες τοποθεσίες ή γεωγραφικές περιοχές. Συνήθως αναπαρίστανται με τη χρήση συντεταγμένων ή γεωγραφικών συστημάτων πληροφοριών (GIS). Τα γεωχωρικά δεδομένα επιτρέπουν την ανάλυση και την οπτικοποίηση προτύπων και σχέσεων δεδομένων σε ένα χωρικό πλαίσιο. Χρησιμοποιούνται σε διάφορους τομείς, όπως ο αστικός σχεδιασμός, η περιβαλλοντική παρακολούθηση, οι μεταφορές και η εφοδιαστική. Παραδείγματα γεωχωρικών δεδομένων είναι τα δεδομένα GPS, οι

δορυφορικές εικόνες, οι χάρτες, τα δεδομένα γεωγραφικών ερευνών και τα δεδομένα κοινωνικής δικτύωσης που βασίζονται στην τοποθεσία.

Στην καθημερινότητά μας ερχόμαστε συχνά σε επαφή με τα κοινωνικά δίκτυα. Τα δεδομένα των μέσων κοινωνικής δικτύωσης περιλαμβάνουν τον τεράστιο όγκο περιεχομένου που δημιουργείται από χρήστες σε πλατφόρμες κοινωνικής δικτύωσης (Bukovina, 2016). Περιλαμβάνει κείμενο, εικόνες, βίντεο και αλληλεπιδράσεις χρηστών. Τα δεδομένα των μέσων κοινωνικής δικτύωσης χαρακτηρίζονται από υψηλό όγκο, ταχύτητα και ποικιλία. Παρουσιάζουν μοναδικές προκλήσεις λόγω της μη δομημένης ή ημιδομημένης φύσης τους, καθώς και της ανάγκης εξαγωγής συναισθήματος, εντοπισμού τάσεων και ανάλυσης δομών κοινωνικών δικτύων. Παραδείγματα δεδομένων κοινωνικών μέσων περιλαμβάνουν tweets, αναρτήσεις στο Facebook, φωτογραφίες στο Instagram, βίντεο στο YouTube, σχόλια, συμπάθειες, κοινοποιήσεις και άλλες μορφές περιεχομένου που δημιουργείται από χρήστες.

Η κατανόηση των διαφόρων τύπων μεγάλων δεδομένων είναι ζωτικής σημασίας για την αποτελεσματική διαχείριση, επεξεργασία και ανάλυση των τεράστιων όγκων και της ποικίλης φύσης των δεδομένων που παράγονται. Κάθε τύπος δεδομένων απαιτεί συγκεκριμένες τεχνικές, εργαλεία και προσεγγίσεις για την εξαγωγή πολύτιμων πληροφοριών και τη βαθύτερη κατανόηση των υποκείμενων μοτίβων και σχέσεων. Με την αξιοποίηση των πλεονεκτημάτων κάθε τύπου δεδομένων, οι οργανισμοί μπορούν να αξιοποιήσουν τη δύναμη των μεγάλων δεδομένων για να προωθήσουν την καινοτομία, να λάβουν αποφάσεις βάσει δεδομένων και να αποκτήσουν ανταγωνιστικό πλεονέκτημα στον σημερινό κόσμο των δεδομένων.

Κεφάλαιο 2^ο Τα ιατρικά δεδομένα και η σημασία τους

2.1 Εισαγωγή στην έννοια και τη σημασία της

Τα ιατρικά δεδομένα αφορούν πληροφορίες και αρχεία που σχετίζονται με την υγεία, το ιατρικό ιστορικό και τις υπηρεσίες υγειονομικής περίθαλψης ενός ατόμου. Περιλαμβάνουν ένα ευρύ φάσμα σημείων δεδομένων, συμπεριλαμβανομένων προσωπικών στοιχείων, όπως όνομα, ηλικία και στοιχεία επικοινωνίας, καθώς και ειδικές ιατρικές πληροφορίες, όπως διαγνώσεις, θεραπείες, φάρμακα, αποτελέσματα εργαστηριακών εξετάσεων, εκθέσεις απεικόνισης και χειρουργικές επεμβάσεις. Τα ιατρικά δεδομένα μπορούν να συλλεχθούν από διάφορες πηγές, συμπεριλαμβανομένων παρόχων υγειονομικής περίθαλψης, νοσοκομείων, κλινικών, εργαστηρίων, φαρμακείων, ιατρικών συσκευών και ηλεκτρονικών αρχείων υγείας (Ristevski and Chen, 2018). Διαδραματίζουν κρίσιμο ρόλο στη φροντίδα των ασθενών, την κλινική έρευνα, τη διοίκηση της υγειονομικής περίθαλψης και τη δημόσια υγεία, παρέχοντας πληροφορίες για τη λήψη τεκμηριωμένων αποφάσεων, τη διάγνωση, τον προγραμματισμό της θεραπείας και τη βελτίωση της συνολικής παροχής υγειονομικής περίθαλψης. Στις υποενότητες που ακολουθούν παρουσιάζονται οι διαστάσεις της σημαντικότητας των ιατρικών δεδομένων.

2.1.1 Για το άτομο

Η σημασία των ιατρικών δεδομένων είναι κατά βάση σημαντική για το ίδιο το άτομο που αφορούν τα δεδομένα. Σε πρώτο επίπεδο τα ιατρικά δεδομένα επιτρέπουν στους επαγγελματίες του τομέα της υγειονομικής περίθαλψης να έχουν μια ολοκληρωμένη κατανόηση της κατάστασης της υγείας ενός ατόμου. Αυτό επιτρέπει εξατομικευμένα και στοχευμένα σχέδια θεραπείας, με αποτέλεσμα τη βελτίωση των αποτελεσμάτων των ασθενών και της ποιότητας της περίθαλψης (Viceconti, Hunter and Hose, 2015). Ταυτόχρονα παρέχουν βασικές πληροφορίες για ακριβείς διαγνώσεις. Βοηθούν τους

γιατρούς στον εντοπισμό μοτίβων, τάσεων και πιθανών παραγόντων κινδύνου, οδηγώντας σε έγκαιρο και ακριβή εντοπισμό των ιατρικών καταστάσεων (Senthilkumar *et al.*, 2018). Επίσης διασφαλίζουν τη συνέχεια της περίθαλψης, ιδίως όταν τα άτομα ζητούν θεραπεία από διαφορετικούς παρόχους υγειονομικής περίθαλψης ή ιδρύματα και επιτρέπουν την απρόσκοπτη ανταλλαγή πληροφοριών, μειώνοντας την πιθανότητα σφαλμάτων και βελτιώνοντας τον συντονισμό μεταξύ των επαγγελματιών υγείας. Τέλος, δεν πρέπει να υποτιμήσουμε τη σημασία της προληπτικής φροντίδας. Τα ιατρικά δεδομένα βοηθούν στον προσδιορισμό προληπτικών μέτρων και πρωτοκόλλων ελέγχου με βάση το ιατρικό ιστορικό ενός ατόμου. Αυτό διευκολύνει την έγκαιρη ανίχνευση ασθενειών, επιτρέποντας την έγκαιρη παρέμβαση και τη βελτίωση των στρατηγικών προληπτικής φροντίδας.

2.1.2 Για την κοινωνία

Ανάλογη είναι η σημασία των δεδομένων και για το κοινωνικό σύνολο, αφού μέσα από αυτά καθορίζει πολιτικές και σχεδιασμούς. Συγκεντρωτικά και ανώνυμα ιατρικά δεδομένα διαδραματίζουν ζωτικό ρόλο στην επιτήρηση της δημόσιας υγείας, ενώ ταυτόχρονα βοηθούν στην παρακολούθηση των εστιών ασθενειών, στον εντοπισμό τάσεων και στην εφαρμογή στοχευμένων παρεμβάσεων σε επίπεδο πληθυσμού, αποτρέποντας έτσι την εξάπλωση ασθενειών και προάγοντας τη συνολική δημόσια υγεία (Viceconti, Hunter and Hose, 2015). Όπως προαναφέρθηκε, τα δεδομένα είναι σημαντικά για την διαμόρφωση πολιτικής και στον σχεδιασμό. Τα ιατρικά δεδομένα χρησιμεύουν ως πολύτιμος πόρος για τους υπεύθυνους χάραξης πολιτικής και τους διαχειριστές της υγειονομικής περίθαλψης. Βοηθούν στην αξιολόγηση των αναγκών υγειονομικής περίθαλψης, στον εντοπισμό των κενών στην παροχή υπηρεσιών και στη διαμόρφωση τεκμηριωμένων πολιτικών για την αποτελεσματική αντιμετώπιση των προκλήσεων της δημόσιας υγείας.

Σημαντική είναι η συμβολή τους στην έρευνα και την καινοτομία. Είναι απαραίτητα για την ιατρική έρευνα και την καινοτομία. Οι ερευνητές μπορούν να αναλύσουν μεγάλα σύνολα δεδομένων για να αποκτήσουν γνώσεις σχετικά με τους μηχανισμούς των ασθενειών, την αποτελεσματικότητα της θεραπείας και την υγεία του

πληθυσμού. Οι γνώσεις αυτές τροφοδοτούν την ανάπτυξη νέων θεραπειών, παρεμβάσεων και ιατρικών τεχνολογιών, προωθώντας τον τομέα της υγειονομικής περίθαλψης (Baro *et al.*, 2015). Παράλληλα συμβάλουν στην ορθολογική κατανομή πόρων. Με την ανάλυση δεδομένων σχετικά με τον επιπολασμό των ασθενειών, τη χρήση της υγειονομικής περίθαλψης και τα αποτελέσματα της θεραπείας, οι υπεύθυνοι χάραξης πολιτικής μπορούν να κατανέμουν αποτελεσματικά τους πόρους, εξασφαλίζοντας ισότιμη πρόσβαση στις υπηρεσίες υγειονομικής περίθαλψης και μειώνοντας τις ανισότητες στην υγειονομική περίθαλψη.

2.1.3 Για την έρευνα

Τα δεδομένα είναι και η ζωή της επιστήμης. Είναι λογικό τα ιατρικά δεδομένα να έχουν βαθιά επίδραση στην έρευνα. Τα ιατρικά δεδομένα χρησιμεύουν ως πλούσια πηγή πληροφοριών για τους ερευνητές. Με την ανάλυση μεγάλων συνόλων δεδομένων, οι ερευνητές μπορούν να εντοπίσουν μοτίβα, τάσεις και συσχετίσεις, οδηγώντας σε βαθύτερη κατανόηση των ασθενειών, της αποτελεσματικότητας της θεραπείας και των αποτελεσμάτων της υγείας. Η γνώση αυτή συμβάλλει στην πρόοδο της επιστημονικής γνώσης και αποτελεί τη βάση για περαιτέρω ερευνητικές μελέτες. Ταυτόχρονα επιτρέπουν στους ερευνητές να εντοπίζουν παράγοντες κινδύνου για διάφορες ασθένειες και καταστάσεις. Αναλύοντας δεδομένα από άτομα με συγκεκριμένες παθήσεις, οι ερευνητές μπορούν να αναπτύξουν μοντέλα πρόβλεψης που βοηθούν στον εντοπισμό ατόμων που διατρέχουν κίνδυνο να αναπτύξουν ορισμένες ασθένειες. Αυτό επιτρέπει έγκαιρες παρεμβάσεις, προληπτικά μέτρα και εξατομικευμένες στρατηγικές υγειονομικής περίθαλψης (Senthilkumar *et al.*, 2018).

Μέσα από την εκμετάλλευση των ιατρικών δεδομένων μπορεί να επιτευχθεί η αξιολόγηση της αποτελεσματικότητας της θεραπείας. Συγκρίνοντας τα αποτελέσματα των θεραπειών σε διάφορους πληθυσμούς ασθενών, οι ερευνητές μπορούν να καθορίσουν ποιες παρεμβάσεις αποδίδουν τα καλύτερα αποτελέσματα, οδηγώντας σε τεκμηριωμένες κατευθυντήριες γραμμές και βελτιωμένη φροντίδα των ασθενών. Πλέον στην ιατρική χρησιμοποιείται και η έννοια της ιατρικής ακρίβειας. Τα ιατρικά δεδομένα υποστηρίζουν τον τομέα της ιατρικής ακριβείας, ο οποίος

αποσκοπεί στην προσαρμογή των θεραπειών και των παρεμβάσεων σε μεμονωμένους ασθενείς με βάση τα μοναδικά χαρακτηριστικά τους. Με την ανάλυση ιατρικών δεδομένων σε ατομικό επίπεδο, οι ερευνητές μπορούν να εντοπίσουν γενετικούς δείκτες, βιοδείκτες και άλλους παράγοντες που επηρεάζουν την ανταπόκριση στη θεραπεία. Αυτό επιτρέπει την ανάπτυξη εξατομικευμένων σχεδίων θεραπείας, με αποτέλεσμα καλύτερα αποτελέσματα για τους ασθενείς και μειωμένες ανεπιθύμητες ενέργειες (Senthilkumar *et al.*, 2018).

Παράλληλα παίζουν καθοριστικό ρόλο στο σχεδιασμό, την υλοποίηση και την ανάλυση των κλινικών δοκιμών. Οι ερευνητές χρησιμοποιούν ιατρικά δεδομένα για τον εντοπισμό κατάλληλων πληθυσμών ασθενών, την παρακολούθηση των αποτελεσμάτων της θεραπείας και την αξιολόγηση των αποτελεσμάτων ασφάλειας και αποτελεσματικότητας. Επιπλέον, τα ιατρικά δεδομένα συμβάλλουν στην ανάπτυξη νέων φαρμάκων και θεραπειών παρέχοντας πληροφορίες για τους μηχανισμούς της νόσου, τις αλληλεπιδράσεις των φαρμάκων και τις αντιδράσεις των ασθενών. Με την ανάλυση δεδομένων σε επίπεδο πληθυσμού, οι ερευνητές μπορούν να αξιολογήσουν την αποτελεσματικότητα των παρεμβάσεων δημόσιας υγείας, να εντοπίσουν τις ανισότητες στην υγειονομική περίθαλψη και να εκτιμήσουν τον αντίκτυπο των αλλαγών πολιτικής. Αυτή η έρευνα με βάση τα δεδομένα συμβάλλει στη διαμόρφωση στρατηγικών δημόσιας υγείας, στην κατανομή των πόρων και στο σχεδιασμό της υγειονομικής περίθαλψης (Athey, 2017).

Ωστόσο, είναι σημαντικό να διασφαλιστεί η δεοντολογική χρήση και προστασία των ιατρικών δεδομένων στην έρευνα. Η διασφάλιση της ιδιωτικής ζωής των ασθενών, η λήψη συγκατάθεσης μετά από ενημέρωση και η τήρηση των κανονιστικών κατευθυντήριων γραμμών, όπως ο HIPAA και ο GDPR, είναι υψίστης σημασίας για τη διατήρηση της εμπιστοσύνης και της εμπιστευτικότητας των πληροφοριών υγείας των ατόμων.

Συμπερασματικά, τα ιατρικά δεδομένα έχουν σημαντική σημασία για την έρευνα, διευκολύνοντας την πρόοδο στην επιστημονική γνώση, την ιατρική ακριβείας, την αξιολόγηση της θεραπείας και την πολιτική υγείας. Αξιοποιώντας τη δύναμη των ιατρικών δεδομένων με υπευθυνότητα και δεοντολογία, οι ερευνητές μπορούν να προωθήσουν την καινοτομία, να βελτιώσουν τα αποτελέσματα των ασθενών και να

συμβάλουν ουσιαστικά στον τομέα της έρευνας στον τομέα της υγειονομικής περίθαλψης. Η αξιοποίησή τους οδηγεί σε εξατομικευμένη φροντίδα, ακριβείς διαγνώσεις, συνέχεια της φροντίδας και προληπτική υγειονομική περίθαλψη για τα άτομα. Σε κοινωνικό επίπεδο, τα ιατρικά δεδομένα βοηθούν στην επιτήρηση της δημόσιας υγείας, στη διαμόρφωση πολιτικών υγείας, στην έρευνα και την καινοτομία και στη βέλτιστη κατανομή των πόρων. Αξιοποιώντας αποτελεσματικά και δεοντολογικά τη δύναμη των ιατρικών δεδομένων, μπορούμε να ενισχύσουμε τα αποτελέσματα της υγειονομικής περίθαλψης και να βελτιώσουμε την ευημερία τόσο των ατόμων όσο και των κοινοτήτων.

2.2 Προκλήσεις

Οι ιατρικά δεδομένα έχουν ιδιαιτερότητες και προκλήσεις λόγω της ευαίσθητης φύσης τους και των απαιτήσεων που συνδέονται με την επεξεργασία και την ανάλυσή τους. Καταρχάς αυτού του είδους τα δεδομένα περιέχουν ευαίσθητες πληροφορίες για την υγεία των ατόμων και απαιτούν αυστηρές προστατευτικές μέτρα για τη διατήρηση της ιδιωτικότητας και την πρόληψη της ανεπιθύμητης πρόσβασης ή διαρροής τους. Η εφαρμογή αποδοτικών μέτρων ασφάλειας και η συμμόρφωση με τις νομοθετικές διατάξεις περί προστασίας δεδομένων, όπως οι κανονισμοί GDPR και HIPAA, είναι απαραίτητες για την προστασία των ιατρικών δεδομένων (Shilo, Rossman and Segal, 2020). Ταυτόχρονα τα ιατρικά δεδομένα προέρχονται από ποικίλες πηγές, συμπεριλαμβανομένων των ιατρικών εγγραφών, των εργαστηριακών δοκιμών, των κλινικών δοκιμών και των φορέων υγείας. Οι προκλήσεις συνίστανται στον συντονισμό και τη συγχώνευση των δεδομένων από διάφορες πηγές, την αναγνώριση και επίλυση πιθανών αντικρουόμενων πληροφοριών και τη δημιουργία συμβατών συστημάτων για την ανταλλαγή δεδομένων μεταξύ των πηγών.

Δεν πρέπει να παραγνωρίζουμε την πολυπλοκότητα και τον όγκο των δεδομένων. Αυτά είναι πολυδιάστατα και πολύπλοκα, συχνά περιλαμβάνοντας πολλαπλές παραμέτρους και μεγάλο όγκο πληροφοριών (Asri *et al.*, 2015). Η αποτελεσματική επεξεργασία, ανάλυση και αξιοποίηση αυτών των δεδομένων απαιτεί προηγμένες

τεχνικές και εργαλεία, καθώς και αποδοτικούς υπολογιστικούς πόρους. Σημαντική πρόκληση είναι και η ανωνυμοποίηση δεδομένων. Για ερευνητικούς σκοπούς, μπορεί να απαιτείται η ανωνυμοποίηση των ιατρικών δεδομένων για τη διαφύλαξη της ιδιωτικότητας των ατόμων. Αυτό μπορεί να επιφέρει προκλήσεις, καθώς οι δεδομένοι πρέπει να είναι ανώνυμοι αλλά ταυτόχρονα να διατηρούν την αξιοπιστία και την αξία τους για την έρευνα. Τέλος σημασία έχει η κατηγοριοποίηση και η διαλειτουργικότητα των δεδομένων (Shilo, Rossman and Segal, 2020). Η σωστή ετικετοποίηση και κατηγοριοποίηση των ιατρικών δεδομένων απαιτεί την επιλογή των σωστών χαρακτηριστικών και την αποτελεσματική ανάθεση ετικετών σε αυτά. Επίσης, η διατήρηση της ποιότητας των δεδομένων και η αντιμετώπιση τυχόν λανθασμένων ή ασυνήθιστων τιμών είναι σημαντικές προκλήσεις. Τέλος η ανάγκη ανταλλαγής και ολοκλήρωσης ιατρικών δεδομένων μεταξύ διαφορετικών συστημάτων και πηγών μπορεί να προκαλέσει προκλήσεις στη διαλειτουργικότητα και στην ομαλή αλληλεπίδραση μεταξύ αυτών των συστημάτων (Asri *et al.*, 2015).

Η αντιμετώπιση αυτών των ιδιαιτεροτήτων και προκλήσεων απαιτεί την υιοθέτηση κατάλληλων τεχνολογικών λύσεων, την εφαρμογή προηγμένων μεθόδων ανάλυσης δεδομένων και την τήρηση αυστηρών κανονισμών απορρήτου και ασφάλειας. Με τη σωστή αντιμετώπιση αυτών των προκλήσεων, τα ιατρικά δεδομένα μπορούν να χρησιμοποιηθούν αποτελεσματικά για την προώθηση της έρευνας και τη βελτίωση των πρακτικών φροντίδας υγείας.

2.3 Παραδείγματα δημιουργίας ιατρικών δεδομένων

Όπως προαναφέρθηκε και σε προηγούμενο υποκεφάλαιο υπάρχουν πολλές εφαρμογές και συσκευές που ωθούν στη δημιουργία δεδομένων. Στο παρόν κεφάλαιο αναλύονται περιπτώσεις που δημιουργούν δεδομένα στον χώρο της υγείας.

2.3.1 Ηλεκτρονικά αρχεία υγείας

Όταν οι ασθενείς λαμβάνουν ιατρική περίθαλψη, οι πληροφορίες για την υγεία τους, συμπεριλαμβανομένου του ιατρικού ιστορικού, των διαγνώσεων, των θεραπειών, των φαρμάκων και των εργαστηριακών αποτελεσμάτων, καταγράφονται σε

ηλεκτρονικούς φακέλους υγείας (Ross, Wei and Ohno-Machado, 2014). Τα δεδομένα αυτά δημιουργούνται και ενημερώνονται κατά τη διάρκεια κάθε συνάντησης με τον ασθενή. Πιο συγκεκριμένα:

- Πληροφορίες για τον ασθενή: όπως το όνομα, η ημερομηνία γέννησης, τα στοιχεία επικοινωνίας και τα στοιχεία ασφάλισης. Περιέχουν επίσης πιο ολοκληρωμένες πληροφορίες, όπως ιατρικό ιστορικό, οικογενειακό ιστορικό, αλλεργίες, τρέχοντα και προηγούμενα φάρμακα, αρχεία εμβολιασμών και χειρουργικές επεμβάσεις.
- Κλινικές σημειώσεις και τεκμηρίωση: Αυτές επιτρέπουν στους παρόχους υγειονομικής περίθαλψης να τεκμηριώνουν τις συναντήσεις με τους ασθενείς, τις παρατηρήσεις, τις διαγνώσεις και τα σχέδια θεραπείας. Οι κλινικές σημειώσεις μπορεί να περιλαμβάνουν σημειώσεις προόδου, σημειώσεις διαβούλευσης, περιλήψεις εξόδου και άλλες σχετικές πληροφορίες. Αυτές οι σημειώσεις παρέχουν μια χρονολογική καταγραφή της φροντίδας ενός ασθενούς και διευκολύνουν τη συνέχεια και τον συντονισμό μεταξύ των παρόχων υγειονομικής περίθαλψης.
- Αποτελέσματα εξετάσεων και εκθέσεις απεικόνισης: Ενσωματώνουν και εμφανίζουν αποτελέσματα από διάφορες διαγνωστικές εξετάσεις, εργαστηριακές εκθέσεις και μελέτες ιατρικής απεικόνισης. Αυτό περιλαμβάνει εξετάσεις αίματος, παθολογικές εκθέσεις, ακτινολογικές εκθέσεις, ηλεκτροκαρδιογραφήματα (ΗΚΓ) και άλλες διαγνωστικές διαδικασίες. Η ύπαρξη αυτών των αποτελεσμάτων εντός του ΗΦΥ επιτρέπει την ευκολότερη πρόσβαση, ανασκόπηση και σύγκριση.
- Διαχείριση φαρμάκων: Καταγράφουν πληροφορίες σχετικά με τα συνταγογραφούμενα φάρμακα, συμπεριλαμβανομένης της δοσολογίας, της συχνότητας, της διάρκειας και τυχόν ειδικών οδηγιών. Παρέχουν επίσης ειδοποιήσεις για αλληλεπιδράσεις φαρμάκων, αλλεργίες και πιθανές αντενδείξεις για την ενίσχυση της ασφάλειας των ασθενών. Οι ΗΦΥ μπορούν να παρακολουθούν το ιστορικό της φαρμακευτικής αγωγής, συμπεριλαμβανομένων των προηγούμενων συνταγών και των πληροφοριών αναπλήρωσης.

2.3.2 Ιατρική απεικόνιση

Οι τεχνικές ιατρικής απεικόνισης, όπως οι ακτίνες Χ, οι αξονικές τομογραφίες, οι μαγνητικές τομογραφίες, οι υπέρηχοι και οι τομογραφίες PET, παράγουν μεγάλες ποσότητες οπτικών δεδομένων. Οι εικόνες αυτές βοηθούν στη διάγνωση και παρακολούθηση διαφόρων ιατρικών καταστάσεων, όπως κατάγματα, όγκοι, καρδιαγγειακές παθήσεις και νευρολογικές διαταραχές. Υπάρχουν διάφοροι κοινοί τύποι τεχνικών ιατρικής απεικόνισης, όπως οι Ακτίνες Χ (Aiello *et al.*, 2019). Οι ακτίνες Χ χρησιμοποιούν ηλεκτρομαγνητική ακτινοβολία για την παραγωγή εικόνων των οστών και ορισμένων μαλακών ιστών. Χρησιμοποιούνται συνήθως για τον εντοπισμό καταγμάτων, πνευμονικών παθήσεων, οδοντιατρικών ζητημάτων και την αξιολόγηση της θέσης ιατρικών συσκευών όπως οι βηματοδότες. Μία άλλη μορφή απεικόνισης είναι η αξονική τομογραφία. Οι αξονικές τομογραφίες παρέχουν λεπτομερέστερες πληροφορίες για τα οστά, τα όργανα, τα αιμοφόρα αγγεία και τους ιστούς. Είναι ιδιαίτερα χρήσιμες στη διάγνωση καταστάσεων όπως όγκοι, τραυματισμοί και καρδιαγγειακές παθήσεις (Κουανου *et al.*, 2018). Μία άλλη διαδεδομένη μορφή είναι η μαγνητική τομογραφία. Η μαγνητική τομογραφία χρησιμοποιεί ισχυρό μαγνητικό πεδίο και ραδιοκύματα για τη δημιουργία λεπτομερών εικόνων των οργάνων, των μαλακών ιστών και των δομών στο εσωτερικό του σώματος. Είναι ιδιαίτερα αποτελεσματική στην απεικόνιση του εγκεφάλου, του νωτιαίου μυελού, των αρθρώσεων και των εσωτερικών οργάνων. Η μαγνητική τομογραφία χρησιμοποιείται συχνά για τη διάγνωση καταστάσεων όπως όγκοι, νευρολογικές διαταραχές, μυοσκελετικές κακώσεις και καρδιαγγειακές παθήσεις. Τέλος έχουμε του υπέρηχους, την τομογραφία εκπομπής ποζιτρονίων και την πυρηνική ιατρική.

2.3.3 Γονιδιωματική αλληλουχία

Με την πρόοδο της γονιδιωματικής, οι τεχνολογίες αλληλούχισης του DNA παράγουν τεράστιες ποσότητες γενετικών δεδομένων (Καο *et al.*, 2014). Η γονιδιωματική αλληλούχιση παρέχει πληροφορίες για τη γενετική σύνθεση ενός ατόμου, εντοπίζοντας γενετικές παραλλαγές, προδιάθεση σε ασθένειες και πιθανές αντιδράσεις σε φάρμακα. Η γονιδιωματική αλληλουχία, επίσης γνωστή ως αλληλουχία DNA ή αλληλουχία γονιδιώματος, είναι μια εργαστηριακή διαδικασία που προσδιορίζει την πλήρη αλληλουχία του DNA ενός ατόμου, το οποίο είναι το γενετικό υλικό που φέρει τις οδηγίες για την ανάπτυξη, τη λειτουργία και τα

χαρακτηριστικά ενός οργανισμού. Περιλαμβάνει την ανάγνωση της σειράς των νουκλεοτιδικών βάσεων (αδενίνη, κυτοσίνη, γουανίνη και θυμίνη) στο γονιδίωμα ενός ατόμου, παρέχοντας μια ολοκληρωμένη εικόνα της γενετικής του σύνθεσης.

Υπάρχουν διαφορετικές μέθοδοι και τεχνολογίες που χρησιμοποιούνται για τη γονιδιωματική αλληλούχιση, αλλά η πιο συχνά χρησιμοποιούμενη τεχνική ονομάζεται αλληλούχιση επόμενης γενιάς (NGS) (Kao *et al.*, 2014). Η NGS επιτρέπει την ταχεία και οικονομικά αποδοτική αλληλούχιση μεγάλων ποσοτήτων DNA. Η γονιδιωματική αλληλουχία παρέχει πολύτιμες πληροφορίες σχετικά με τον γενετικό κώδικα ενός ατόμου και μπορεί να χρησιμοποιηθεί για διάφορους σκοπούς, όπως η διάγνωση ασθενειών, η φαρμακογονιδιωματική και ο οικογενειακός προγραμματισμός. Αναφορικά με το πρώτο, η γενετική αλληλουχία μπορεί να εντοπίσει γενετικές παραλλαγές που σχετίζονται με κληρονομικές διαταραχές ή ασθένειες. Βοηθά στη διάγνωση γενετικών καταστάσεων, στον προσδιορισμό του κινδύνου εμφάνισης ορισμένων ασθενειών και στην πρόβλεψη του τρόπου με τον οποίο ένα άτομο μπορεί να ανταποκριθεί σε ορισμένες θεραπείες. Επίσης οι ερευνητές χρησιμοποιούν τα δεδομένα αλληλούχισης για να μελετήσουν τη γενετική βάση των ασθενειών, να εντοπίσουν νέα γονίδια, να διερευνήσουν τη γενετική του πληθυσμού και να αποκαλύψουν συσχετίσεις μεταξύ γενετικών παραλλαγών και συγκεκριμένων χαρακτηριστικών ή καταστάσεων.

Αναφορικά με την φαρμακογονιδιωματική θα πρέπει να αναφέρουμε ότι αναλύοντας τις γενετικές παραλλαγές ενός ατόμου, η γονιδιωματική αλληλούχιση μπορεί να προσφέρει πληροφορίες σχετικά με τον τρόπο με τον οποίο ένα άτομο μπορεί να μεταβολίζει ορισμένα φάρμακα. Οι πληροφορίες αυτές βοηθούν στη βελτιστοποίηση της επιλογής φαρμάκων, της δοσολογίας και στη μείωση του κινδύνου ανεπιθύμητων αντιδράσεων φαρμάκων. Τέλος η γονιδιωματική αλληλουχία μπορεί να βοηθήσει τα άτομα και τα ζευγάρια στη λήψη τεκμηριωμένων αποφάσεων σχετικά με τον οικογενειακό προγραμματισμό, τον προγεννητικό έλεγχο και την αναπαραγωγική υγεία. Μπορεί να προσδιορίσει την κατάσταση φορέα για ορισμένες γενετικές παθήσεις και να παράσχει πληροφορίες σχετικά με τον κίνδυνο μεταβίβασης γενετικών διαταραχών στους απογόνους.

2.3.4 Φορητές συσκευές και αισθητήρες

Οι φορητές συσκευές, όπως οι συσκευές παρακολούθησης φυσικής κατάστασης, τα έξυπνα ρολόγια και οι ιατρικές συσκευές, μπορούν να παρακολουθούν συνεχώς διάφορες παραμέτρους υγείας, όπως ο καρδιακός ρυθμός, τα πρότυπα ύπνου, η σωματική δραστηριότητα και τα επίπεδα γλυκόζης. Αυτές οι συσκευές παράγουν δεδομένα σε πραγματικό χρόνο που βοηθούν τα άτομα και τους παρόχους υγειονομικής περίθαλψης να παρακολουθούν και να διαχειρίζονται την υγεία τους (Purswani *et al.*, 2019). Τύποι συσκευών είναι οι παρακάτω:

- Συνεχείς μετρητές γλυκόζης (Continuous Glucose Monitors - CGM): Είναι μικροί αισθητήρες που παρακολουθούν τα επίπεδα γλυκόζης σε άτομα με διαβήτη. Φοριούνται συνήθως στο σώμα και παρέχουν μετρήσεις γλυκόζης σε πραγματικό χρόνο, επιτρέποντας στα άτομα να παρακολουθούν και να διαχειρίζονται αποτελεσματικότερα τα επίπεδα σακχάρου στο αίμα τους.
- Οθόνες μέτρησης της αρτηριακής πίεσης: Τα φορητά μόνιτορ αρτηριακής πίεσης χρησιμοποιούν αισθητήρες για τη συνεχή μέτρηση της αρτηριακής πίεσης καθ' όλη τη διάρκεια της ημέρας. Αυτές οι συσκευές είναι βολικές για άτομα που χρειάζονται τακτική παρακολούθηση της αρτηριακής πίεσης, ιδίως για άτομα με υπέρταση ή καρδιαγγειακές παθήσεις.
- Παλμικά οξύμετρα: Τα παλμικά οξύμετρα είναι φορητές συσκευές ή αισθητήρες που μετρούν τα επίπεδα κορεσμού οξυγόνου στο αίμα. Συνήθως κουμπώνουν σε ένα δάχτυλο ή μπορούν να φορεθούν στον καρπό. Τα παλμικά οξύμετρα είναι χρήσιμα για άτομα με αναπνευστικές παθήσεις ή για όσους παρακολουθούν τα επίπεδα οξυγόνου τους κατά τη διάρκεια σωματικών δραστηριοτήτων ή σε μεγάλα υψόμετρα.
- Οθόνες: Τα μόνιτορ ΗΚΓ (ηλεκτροκαρδιογράφημα) ή ΕΚΓ (ηλεκτροκαρδιογράφημα) καταγράφουν την ηλεκτρική δραστηριότητα της καρδιάς. Τα φορητά μόνιτορ ΗΚΓ, που συνήθως φοριούνται ως ιμάντες ή επιθέματα στο στήθος, παρέχουν συνεχή παρακολούθηση του καρδιακού ρυθμού. Χρησιμοποιούνται για την ανίχνευση ακανόνιστων καρδιακών

ρυθμών, την παρακολούθηση της υγείας της καρδιάς και τη βοήθεια στη διάγνωση καρδιακών παθήσεων.

- Αισθητήρες ανίχνευσης πτώσης: Οι αισθητήρες ανίχνευσης πτώσης έχουν σχεδιαστεί για να ανιχνεύουν πτώσεις ή ξαφνικές κινήσεις. Αυτοί οι αισθητήρες μπορούν να ενσωματωθούν σε φορητές συσκευές, όπως βραχιολάκια ή μενταγιόν, και είναι ιδιαίτερα χρήσιμοι για ηλικιωμένους ή άτομα που διατρέχουν κίνδυνο πτώσεων. Μπορούν να ενεργοποιούν αυτόματα ειδοποιήσεις ή ειδοποιήσεις σε περίπτωση πτώσης, επιτρέποντας την έγκαιρη παροχή βοήθειας.
- Αισθητήρες υπεριώδους ακτινοβολίας: Οι αισθητήρες υπεριώδους ακτινοβολίας μετρούν την ένταση της έκθεσης στην υπεριώδη ακτινοβολία (UV) από τον ήλιο. Οι φορητοί αισθητήρες υπεριώδους ακτινοβολίας, συχνά με τη μορφή βραχιολιών ή επιθεμάτων, βοηθούν τα άτομα να παρακολουθούν την έκθεσή τους στην υπεριώδη ακτινοβολία και να λαμβάνουν τις απαραίτητες προφυλάξεις για την προστασία του δέρματός τους από την ηλιακή βλάβη και την ελαχιστοποίηση του κινδύνου ηλιακού εγκαύματος ή καρκίνου του δέρματος.
- Αισθητήρες ιδρώτα: Οι αισθητήρες ιδρώτα είναι φορητές συσκευές που αναλύουν τη σύνθεση του ιδρώτα για να παρέχουν πληροφορίες σχετικά με τα επίπεδα ενυδάτωσης, την ισορροπία ηλεκτρολυτών και τους μεταβολικούς δείκτες. Αυτοί οι αισθητήρες, που συχνά ενσωματώνονται σε βραχιολάκια ή επιθέματα, μπορούν να χρησιμοποιηθούν από αθλητές, λάτρεις της γυμναστικής ή άτομα που παρακολουθούν τη συνολική κατάσταση της υγείας και της ενυδάτωσής τους.

2.3.5 Κλινικές δοκιμές και ερευνητικές μελέτες

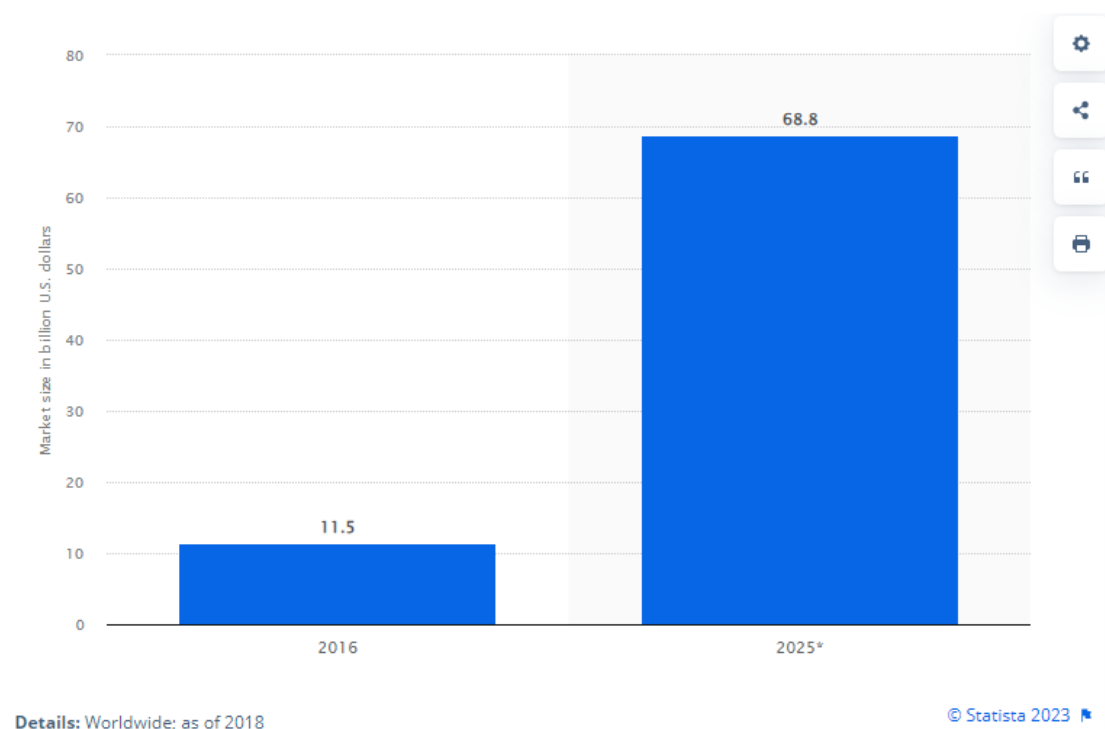
Τα ιατρικά δεδομένα παράγονται μέσω κλινικών δοκιμών και ερευνητικών μελετών, όπου καταγράφονται και αναλύονται οι πληροφορίες υγείας των συμμετεχόντων, οι αντιδράσεις στη θεραπεία και τα αποτελέσματα για την αξιολόγηση της αποτελεσματικότητας και της ασφάλειας νέων θεραπειών και παρεμβάσεων.

2.3.6 Επιτήρηση της δημόσιας υγείας

Οι οργανισμοί δημόσιας υγείας συλλέγουν και αναλύουν δεδομένα για την παρακολούθηση των εστιών ασθενειών, την παρακολούθηση της εξάπλωσης μολυσματικών ασθενειών και την αξιολόγηση των τάσεων της υγείας του πληθυσμού. Τα δεδομένα αυτά περιλαμβάνουν πληροφορίες σχετικά με τη συχνότητα εμφάνισης των ασθενειών, τον επιπολασμό, τους δημογραφικούς παράγοντες και τους παράγοντες κινδύνου.

Τα παραδείγματα αυτά καταδεικνύουν τον τρόπο με τον οποίο παράγονται ιατρικά δεδομένα από διάφορες πηγές και τεχνολογίες, παρέχοντας πολύτιμες πληροφορίες για τους παρόχους υγειονομικής περίθαλψης, τους ερευνητές και τους φορείς δημόσιας υγείας για τη βελτίωση της περίθαλψης των ασθενών, την προαγωγή της ιατρικής γνώσης και την ενίσχυση των παρεμβάσεων δημόσιας υγείας.

Στην εικόνα που ακολουθεί παρουσιάζεται το μέγεθος της παγκόσμιας αγοράς μεγάλων δεδομένων που σχετίζονται με την υγειονομική περίθαλψη το 2016 και μια πρόβλεψη για το 2025. Εκτιμάται ότι κατά τη διάρκεια αυτής της περιόδου η αγορά θα αυξηθεί από περίπου 11,5 δισεκατομμύρια σε σχεδόν 70 δισεκατομμύρια δολάρια ΗΠΑ.



Εικόνα 2 . Η αγορά μεγάλων δεδομένων μέχρι το 2025 στην υγεία (BIS Research, 2023)

Κεφάλαιο 3^ο- Διεθνείς οργανισμοί

3.1 Εισαγωγή

Είναι προφανές από τα παραπάνω ότι ο χώρος των ιατρικών δεδομένων και στη μεγέθυνσή του, ο χώρος των Μεγάλων Δεδομένων είναι ένας πολυσύνθετος χώρος που η οποιαδήποτε τοπική ή ιδρυματική προσέγγιση θα ήταν περιορισμένη και μονολιθική. Στον χώρο των μεγάλων δεδομένων στην υγεία είναι επιβεβλημένη η επιτήρηση και διαμόρφωση του χώρου από διεθνείς οργανισμούς που θα διαμορφώσουν τις εξελίξεις κάτω από ένα ολιστικό πλαίσιο. Σε πρώτο επίπεδο θα πρέπει η πληροφορία να τυποποιηθεί. Οι διεθνείς οργανισμοί μπορούν να συμβάλουν στη θέσπιση κοινών προτύπων και κατευθυντήριων γραμμών για τη συλλογή, την ανταλλαγή, την προστασία της ιδιωτικής ζωής και την ασφάλεια των δεδομένων. Η τυποποίηση διασφαλίζει τη διαλειτουργικότητα και την εναρμόνιση των δεδομένων σε διαφορετικά συστήματα υγειονομικής περίθαλψης, επιτρέποντας την απρόσκοπτη συνεργασία και έρευνα. Ταυτόχρονα θα πρέπει να ενισχυθεί διεθνής παρακολούθηση και η έρευνα, οι οποίες μπορούν να προωθήσουν την ποιότητα και την ακεραιότητα των δεδομένων. Θέτοντας σημεία αναφοράς και διενεργώντας ελέγχους, οι διεθνείς φορείς μπορούν να διασφαλίσουν ότι τα δεδομένα που συλλέγονται είναι ακριβή, αξιόπιστα και συνεπή.

Οι διεθνείς οργανισμοί μπορούν να συμβάλουν στην διαφύλαξη της ιδιωτικότητας και της δεοντολογίας, αλλά και στη διασυννοριακή συνεργασία. Η υγειονομική περίθαλψη αποτελεί παγκόσμιο ζήτημα και πολλά ζητήματα υγείας υπερβαίνουν τα εθνικά σύνορα. Η διεθνής παρακολούθηση και έρευνα διευκολύνουν τη διασυννοριακή συνεργασία, επιτρέποντας την ανταλλαγή δεδομένων, γνώσεων και βέλτιστων πρακτικών. Με τη συνεργασία σε μεγάλα δεδομένα στον τομέα της υγειονομικής περίθαλψης, οι χώρες μπορούν να επωφεληθούν από συλλογικές γνώσεις, να εντοπίσουν παγκόσμιες τάσεις στον τομέα της υγείας και να αναπτύξουν αποτελεσματικότερες στρατηγικές για την πρόληψη, τη διάγνωση και τη θεραπεία ασθενειών. Τέλος είναι αυτοί που θα αναπτύξουν πολιτικές και θα προωθήσουν την έρευνα. Μπορούν να αξιολογήσουν τις ηθικές, νομικές και κοινωνικές επιπτώσεις της

χρήσης μεγάλων δεδομένων, να εντοπίσουν πιθανούς κινδύνους και να προτείνουν κατάλληλα πλαίσια και κατευθυντήριες γραμμές για την υπεύθυνη και ασφαλή χρήση τους. Τέτοιες πολιτικές και κανονισμοί συμβάλλουν στην οικοδόμηση της εμπιστοσύνης του κοινού, στη διασφάλιση της λογοδοσίας και στον μετριασμό της πιθανής κατάχρησης των δεδομένων. Τα μεγάλα δεδομένα έχουν τη δυνατότητα να φέρουν επανάσταση στην έρευνα και την καινοτομία στον τομέα της υγειονομικής περίθαλψης. Οι διεθνείς προσπάθειες παρακολούθησης και έρευνας μπορούν να προωθήσουν τη συνεργασία μεταξύ ερευνητών, επιστημόνων και ιδρυμάτων παγκοσμίως. Αυτό το συνεργατικό περιβάλλον επιταχύνει την ανακάλυψη νέων θεραπειών, βελτιώνει την επιτήρηση ασθενειών, εντοπίζει μοτίβα και τάσεις και βελτιώνει τα αποτελέσματα της υγειονομικής περίθαλψης σε παγκόσμιο επίπεδο.

3.2 Health Level Seven International (HL7)

Το Health Level Seven International (HL7) είναι ένας μη κερδοσκοπικός οργανισμός που αναπτύσσει διεθνή πρότυπα για την ανταλλαγή, ολοκλήρωση, διαμοιρασμό και ανάκτηση ηλεκτρονικών πληροφοριών υγείας (Health Level Seven International, 2023). Ιδρύθηκε το 1987 και έχει καταστεί μια από τις κορυφαίες αρχές στη διαλειτουργικότητα των δεδομένων υγειονομικής περίθαλψης. Πρωταρχικός στόχος του HL7 είναι να καταστήσει δυνατή την απρόσκοπτη επικοινωνία και διαλειτουργικότητα μεταξύ συστημάτων υγειονομικής περίθαλψης, ιατρικών συσκευών και άλλων εφαρμογών τεχνολογίας πληροφοριών υγείας (HIT). Ο οργανισμός αναπτύσσει και διατηρεί ένα σύνολο προτύπων, πρωτοκόλλων και μεθοδολογιών που διευκολύνουν την ανταλλαγή δεδομένων υγείας μεταξύ διαφορετικών συστημάτων και οργανισμών. Τα πρότυπα HL7 αναπτύσσονται μέσω μιας διαδικασίας συνεργασίας και συναίνεσης στην οποία συμμετέχει ένα ευρύ φάσμα ενδιαφερόμενων φορέων, συμπεριλαμβανομένων παρόχων υγειονομικής περίθαλψης, προμηθευτών, κυβερνητικών φορέων και άλλων οργανισμών ανάπτυξης προτύπων. Τα πρότυπα αυτά καθορίζουν τη δομή, το περιεχόμενο και τη σημασιολογία των κλινικών και διοικητικών δεδομένων, διασφαλίζοντας ότι οι πληροφορίες μπορούν να ανταλλάσσονται με ακρίβεια και νόημα.

Τα πρότυπα HL7 χρησιμοποιούνται ευρέως σε παγκόσμιο επίπεδο και έχουν υιοθετηθεί από πολυάριθμους οργανισμούς υγειονομικής περίθαλψης, κυβερνήσεις και προμηθευτές. Το πιο γνωστό πρότυπο που αναπτύχθηκε από το HL7 είναι το πρότυπο ανταλλαγής μηνυμάτων HL7 Version 2.x, το οποίο ορίζει μια ευρέως χρησιμοποιούμενη μορφή για την ανταλλαγή δεδομένων υγειονομικής περίθαλψης μεταξύ συστημάτων. Ο HL7 έχει επίσης αναπτύξει το πρότυπο HL7 Clinical Document Architecture (CDA) και το πρότυπο HL7 Fast Healthcare Interoperability Resources (FHIR®), το οποίο κερδίζει σημαντική απήχηση στον κλάδο της υγειονομικής περίθαλψης λόγω της ευελιξίας και της σύγχρονης προσέγγισής του στην ανταλλαγή δεδομένων (Health Level Seven International, 2023). Εκτός από την ανάπτυξη προτύπων, ο HL7 παρέχει εκπαιδευτικούς πόρους, προγράμματα πιστοποίησης και υποστηρίζει ευκαιρίες συνεργασίας και δικτύωσης για επαγγελματίες που εργάζονται στην πληροφορική της υγειονομικής περίθαλψης και την ανταλλαγή δεδομένων. Διοργανώνει εκδηλώσεις, συνέδρια και ομάδες εργασίας όπου οι ειδικοί συναντώνται για να αντιμετωπίσουν τις τρέχουσες προκλήσεις και να προωθήσουν τον τομέα της διαλειτουργικότητας των δεδομένων υγείας.

Συνολικά, το HL7 διαδραματίζει κρίσιμο ρόλο στην προώθηση της διαλειτουργικότητας και της ανταλλαγής δεδομένων στον κλάδο της υγειονομικής περίθαλψης, επιτρέποντας την απρόσκοπτη επικοινωνία και την ενσωμάτωση των πληροφοριών υγείας για τη βελτίωση της περίθαλψης των ασθενών, της έρευνας και των πρωτοβουλιών για τη δημόσια υγεία.

3.3 International Medical Informatics Association (IMIA)

Η Διεθνής Ένωση Ιατρικής Πληροφορικής (IMIA) διαδραματίζει σημαντικό ρόλο στον τομέα των ιατρικών δεδομένων (Association, 2023). Η IMIA προωθεί την ανάπτυξη και την υιοθέτηση προτύπων δεδομένων που διευκολύνουν τη διαλειτουργικότητα και την απρόσκοπτη ανταλλαγή ιατρικών δεδομένων μεταξύ διαφορετικών συστημάτων υγειονομικής περίθαλψης. Μέσω των ομάδων εργασίας της, η IMIA συμμετέχει ενεργά στην ανάπτυξη και προώθηση διεθνών προτύπων, όπως το Health Level Seven (HL7) και το Fast Healthcare Interoperability Resources (FHIR®),

διασφαλίζοντας ότι τα ιατρικά δεδομένα μπορούν να διαμοιραστούν και να αξιοποιηθούν αποτελεσματικά. Ταυτόχρονα τονίζει τη σημασία της διακυβέρνησης δεδομένων και της δεοντολογίας στην πληροφορική της υγείας. Ο οργανισμός προωθεί την υπεύθυνη και ηθική χρήση των ιατρικών δεδομένων, συμπεριλαμβανομένων των θεμάτων που σχετίζονται με την προστασία της ιδιωτικής ζωής των δεδομένων, την ασφάλεια και τη συγκατάθεση. Η IMIA παρέχει κατευθυντήριες γραμμές και συστάσεις για να διασφαλίσει ότι τα ιατρικά δεδομένα χειρίζονται με ασφαλή και δεοντολογικό τρόπο, προστατεύοντας το απόρρητο και την ιδιωτική ζωή των ασθενών (Association, 2023).

Επιπλέον, η IMIA υποστηρίζει την έρευνα και την καινοτομία στην ανάλυση ιατρικών δεδομένων. Ο οργανισμός ενθαρρύνει τη χρήση προσεγγίσεων και τεχνολογιών με γνώμονα τα δεδομένα για την εξαγωγή συμπερασμάτων από σύνολα ιατρικών δεδομένων μεγάλης κλίμακας. Οι ομάδες εργασίας και τα συνέδρια της IMIA παρέχουν πλατφόρμες στους ερευνητές και τους επαγγελματίες να μοιράζονται τα ευρήματά τους και τις προόδους τους στον τομέα της ανάλυσης ιατρικών δεδομένων. Η IMIA διευκολύνει την παγκόσμια συνεργασία και δικτύωση μεταξύ επαγγελματιών και οργανισμών που ασχολούνται με την ιατρική πληροφορική. Συνδέοντας εμπειρογνώμονες από διαφορετικές περιοχές και τομείς, η IMIA προωθεί την ανταλλαγή γνώσεων, εμπειριών και βέλτιστων πρακτικών που σχετίζονται με την κοινή χρήση, τη διαχείριση και την ανάλυση ιατρικών δεδομένων.

Τέλος, η IMIA προωθεί την εκπαίδευση και την κατάρτιση στην πληροφορική της υγείας, συμπεριλαμβανομένης της διαχείρισης και της ανάλυσης ιατρικών δεδομένων. Ο οργανισμός υποστηρίζει την ανάπτυξη εκπαιδευτικών προγραμμάτων, μαθημάτων και πιστοποιήσεων που εφοδιάζουν τους επαγγελματίες του τομέα της υγείας με τις απαραίτητες δεξιότητες για τον αποτελεσματικό χειρισμό ιατρικών δεδομένων και την αξιοποίησή τους για την έρευνα, τη λήψη αποφάσεων και τη βελτίωση της υγειονομικής περίθαλψης.

Συνολικά, η IMIA διαδραματίζει καθοριστικό ρόλο στην προώθηση του τομέα της διαχείρισης και αξιοποίησης ιατρικών δεδομένων. Υποστηρίζοντας τα πρότυπα δεδομένων, τη δεοντολογία και τη διαλειτουργικότητα, προωθώντας την έρευνα και την καινοτομία και παρέχοντας εκπαιδευτικούς πόρους, η IMIA συμβάλλει στην

αποτελεσματική και υπεύθυνη χρήση των ιατρικών δεδομένων προς όφελος των ασθενών, των παρόχων υγειονομικής περίθαλψης και της κοινωνίας στο σύνολό της.

3.4 European Institute for Health Records (EuroRec)

Το Ευρωπαϊκό Ινστιτούτο για τα αρχεία υγείας (EuroRec) είναι ένας οργανισμός που επικεντρώνεται στην ανάπτυξη και προώθηση ηλεκτρονικών αρχείων υγείας (ΗΦΥ) υψηλής ποιότητας στην Ευρώπη (European Institute for Health Records, 2023). Το EuroRec διαδραματίζει καθοριστικό ρόλο στο πλαίσιο των ιατρικών δεδομένων. Η EuroRec παρέχει προγράμματα πιστοποίησης και πλαίσια αξιολόγησης της ποιότητας για συστήματα ΗΦΥ (Ηλεκτρονικών φακέλων υγείας). Τα προγράμματα αυτά διασφαλίζουν ότι οι ΗΦΥ πληρούν συγκεκριμένα κριτήρια και πρότυπα, συμπεριλαμβανομένης της ποιότητας των δεδομένων, της ασφάλειας, της διαλειτουργικότητας και της συμμόρφωσης με τις νομικές και ηθικές απαιτήσεις. Η διαδικασία πιστοποίησης της EuroRec βοηθά τους οργανισμούς υγειονομικής περίθαλψης και τους πωλητές να προσδιορίσουν τα συστήματα EHR που είναι αξιόπιστα και τηρούν τις καθιερωμένες πρακτικές διαχείρισης δεδομένων. Ταυτόχρονα προωθεί τη χρήση διεθνών προτύπων για τους ΗΦΥ και υποστηρίζει τη διαλειτουργικότητα μεταξύ διαφορετικών συστημάτων ΗΦΥ. Ο οργανισμός συνεργάζεται στενά με φορείς τυποποίησης, όπως ο Health Level Seven (HL7) και ο Διεθνής Οργανισμός Τυποποίησης (ISO), για την ανάπτυξη και εφαρμογή προτύπων δεδομένων που επιτρέπουν την απρόσκοπτη ανταλλαγή και κοινή χρήση ιατρικών δεδομένων μεταξύ οργανισμών και συστημάτων υγειονομικής περίθαλψης (European Institute for Health Records, 2023).

Η EuroRec τονίζει τη σημασία της ποιότητας των δεδομένων στους ΗΦΥ. Ο οργανισμός παρέχει κατευθυντήριες γραμμές και μεθοδολογίες για τη βελτίωση της ποιότητας των δεδομένων και της διακυβέρνησης, διασφαλίζοντας ότι τα ιατρικά δεδομένα που καταγράφονται στους ΗΦΥ είναι ακριβή, πλήρη και αξιόπιστα. Οι πρωτοβουλίες της EuroRec επικεντρώνονται στην επικύρωση δεδομένων, την ακεραιότητα δεδομένων και την εφαρμογή στρατηγικών διακυβέρνησης δεδομένων για τη διατήρηση ιατρικών δεδομένων υψηλής ποιότητας στα συστήματα ΗΦΥ.

Ταυτόχρονα η EuroRec διευκολύνει την ανταλλαγή βέλτιστων πρακτικών, εμπειριών και γνώσεων που σχετίζονται με την εφαρμογή HAY και τη διαχείριση ιατρικών δεδομένων. Ο οργανισμός διοργανώνει συνέδρια, εργαστήρια και συνεργατικά έργα όπου επαγγελματίες του τομέα της υγείας, ερευνητές και υπεύθυνοι χάραξης πολιτικής μπορούν να ανταλλάσσουν ιδέες, καινοτομίες και διδάγματα στον τομέα των ΗΦΥ και των ιατρικών δεδομένων (European Institute for Health Records, 2023).

Τέλος η EuroRec υποστηρίζει δραστηριότητες έρευνας και καινοτομίας που σχετίζονται με τους ΗΦΥ και τα ιατρικά δεδομένα. Ο οργανισμός συνεργάζεται με ερευνητικά ιδρύματα, βιομηχανικούς εταίρους και οργανισμούς υγειονομικής περίθαλψης για τη διερεύνηση αναδυόμενων τεχνολογιών, αναλύσεων δεδομένων και συστημάτων υποστήριξης αποφάσεων που μπορούν να αξιοποιήσουν τα ιατρικά δεδομένα για τη βελτίωση της περίθαλψης των ασθενών, της έρευνας και της διαχείρισης της υγείας του πληθυσμού.

Μέσω των πρωτοβουλιών της, η EuroRec διαδραματίζει σημαντικό ρόλο στην προώθηση της ποιότητας, της διαλειτουργικότητας και της διακυβέρνησης των ιατρικών δεδομένων εντός των συστημάτων EHR. Με την προώθηση της πιστοποίησης, της υιοθέτησης προτύπων, της βελτίωσης της ποιότητας των δεδομένων και της ανταλλαγής γνώσεων, η EuroRec συμβάλλει στην αποτελεσματική χρήση των ιατρικών δεδομένων για καλύτερα αποτελέσματα της υγειονομικής περίθαλψης και την ασφάλεια των ασθενών στην Ευρώπη.

3.5 Health Data Research UK (HDR UK)

Το Health Data Research UK (HDR UK) είναι ένα εθνικό ινστιτούτο του Ηνωμένου Βασιλείου αφιερωμένο στην αξιοποίηση της δύναμης των δεδομένων υγείας για την έρευνα και την καινοτομία (Health Data Research UK, 2023). Το HDR UK διαδραματίζει ζωτικό ρόλο στο πλαίσιο των μεγάλων δεδομένων στην υγειονομική περίθαλψη. Το HDR UK εργάζεται προς την κατεύθυνση της ενοποίησης και της επιμέλειας συνόλων δεδομένων υγείας μεγάλης κλίμακας από διάφορες πηγές, συμπεριλαμβανομένων των ηλεκτρονικών αρχείων υγείας, της γονιδιωματικής, των κλινικών δοκιμών και των δεδομένων για την υγεία του πληθυσμού. Με την

ενοποίηση και την εναρμόνιση αυτών των διαφορετικών συνόλων δεδομένων, το HDR UK δίνει τη δυνατότητα στους ερευνητές και τους καινοτόμους να έχουν πρόσβαση σε ολοκληρωμένα και αντιπροσωπευτικά δεδομένα υγείας για ανάλυση και ανακάλυψη. Παράλληλα προωθεί την έρευνα στον τομέα της επιστήμης των δεδομένων και της ανάλυσης για την άντληση πληροφοριών από μεγάλα δεδομένα υγείας. Το ινστιτούτο υποστηρίζει την ανάπτυξη και εφαρμογή προηγμένων αναλυτικών μεθόδων, μηχανικής μάθησης, τεχνητής νοημοσύνης και τεχνικών οπτικοποίησης δεδομένων για την αποκάλυψη μοτίβων, τάσεων και συσχετίσεων σε δεδομένα υγείας. Με την προώθηση καινοτόμων προσεγγίσεων ανάλυσης δεδομένων, το HDR UK ενισχύει την κατανόηση των ασθενειών, των θεραπευτικών αποτελεσμάτων και της δυναμικής της υγείας του πληθυσμού (Health Data Research UK, 2023).

Το HDR UK διασφαλίζει τη δημιουργία ασφαλούς και αξιόπιστης υποδομής δεδομένων για μεγάλα δεδομένα υγείας. Το ινστιτούτο τηρεί αυστηρές αρχές προστασίας και διακυβέρνησης δεδομένων, εφαρμόζοντας ισχυρά μέτρα ασφαλείας για τη διασφάλιση της ιδιωτικής ζωής και της εμπιστευτικότητας των ασθενών. Το HDR UK συνεργάζεται με ρυθμιστικούς φορείς και οργανισμούς για τη δημιουργία πλαισίων πρόσβασης σε δεδομένα που επιτρέπουν την υπεύθυνη και ηθική χρήση των δεδομένων υγείας. Ταυτόχρονα προωθεί τη συνεργασία μεταξύ ερευνητών, οργανισμών υγειονομικής περίθαλψης, βιομηχανικών εταίρων και φορέων χάραξης πολιτικής. Το ινστιτούτο προωθεί τη διεπιστημονική συνεργασία για την αντιμετώπιση σύνθετων προκλήσεων στον τομέα της υγείας με τη χρήση μεγάλων δεδομένων. Μέσω συνεργασιών με ακαδημαϊκά ιδρύματα, οργανισμούς NHS και βιομηχανικούς φορείς, το HDR UK διευκολύνει την ανταλλαγή δεδομένων, την ανταλλαγή γνώσεων και κοινά ερευνητικά έργα, επιτρέποντας τη διατομεακή συνεργασία στον τομέα των μεγάλων δεδομένων στην υγειονομική περίθαλψη. Τέλος υποστηρίζει τη μετάφραση των ερευνητικών ευρημάτων σε πρακτικές εφαρμογές και παρεμβάσεις. Το ινστιτούτο ενθαρρύνει την καινοτομία και την επιχειρηματικότητα γεφυρώνοντας το χάσμα μεταξύ έρευνας και βιομηχανίας. Το HDR UK διευκολύνει την ανάπτυξη εργαλείων, τεχνολογιών και υπηρεσιών που βασίζονται στα δεδομένα και μπορούν να επηρεάσουν άμεσα την παροχή

υγειονομικής περίθαλψης, την ιατρική ακριβείας και τις στρατηγικές δημόσιας υγείας (Health Data Research UK, 2023).

Το HDR UK διαδραματίζει καθοριστικό ρόλο στην αξιοποίηση των μεγάλων δεδομένων για την προώθηση της έρευνας, της καινοτομίας και των βελτιώσεων στην υγειονομική περίθαλψη. Με την ενσωμάτωση διαφορετικών συνόλων δεδομένων, την προώθηση της επιστήμης των δεδομένων και της ανάλυσης, τη διασφάλιση της ασφάλειας των δεδομένων, την προώθηση της συνεργασίας και την προώθηση της μεταφραστικής έρευνας, το HDR UK επιτρέπει την αποτελεσματική και υπεύθυνη χρήση των μεγάλων δεδομένων υγείας προς όφελος των ασθενών, των παρόχων υγειονομικής περίθαλψης και της κοινωνίας στο σύνολό της (Health Data Research UK, 2023).

3.6 Institute for Health Metrics and Evaluation (IHME)

Το Institute for Health Metrics and Evaluation (IHME) είναι ένας ανεξάρτητος ερευνητικός οργανισμός για την παγκόσμια υγεία που επικεντρώνεται στη μέτρηση και την αξιολόγηση των αποτελεσμάτων της υγείας παγκοσμίως (Evaluation, 2023). Το IHME διαδραματίζει σημαντικό ρόλο στη δημιουργία και διάδοση δεδομένων, ερευνών και μετρήσεων που σχετίζονται με την υγεία. Το IHME συλλέγει, αναλύει και συνθέτει τεράστιες ποσότητες δεδομένων που σχετίζονται με την υγεία από διάφορες πηγές, συμπεριλαμβανομένων ερευνών, απογραφών, μητρώων και διοικητικών αρχείων. Ο οργανισμός χρησιμοποιεί προηγμένες στατιστικές μεθόδους και τεχνικές μοντελοποίησης για την εκτίμηση και την προβολή δεικτών υγείας, όπως τα ποσοστά θνησιμότητας, ο επιπολασμός ασθενειών, οι παράγοντες κινδύνου και τα έτη ζωής προσαρμοσμένα στην αναπηρία (DALYs). Η προσέγγιση του IHME που βασίζεται στα δεδομένα παρέχει μια ολοκληρωμένη και λεπτομερή κατανόηση των παγκόσμιων τάσεων και προτύπων στον τομέα της υγείας. Επίσης αναπτύσσει και διατηρεί ένα ευρύ φάσμα μετρήσεων υγείας που χρησιμεύουν ως πολύτιμα εργαλεία για τους υπεύθυνους χάραξης πολιτικής, τους ερευνητές και τους επαγγελματίες της δημόσιας υγείας. Οι μετρήσεις αυτές περιλαμβάνουν δείκτες όπως το προσδόκιμο ζωής, οι αιτίες θανάτου, η επιβάρυνση από ασθένειες, οι παράγοντες κινδύνου και

οι επιδόσεις του συστήματος υγείας. Οι μετρικές του IHME επιτρέπουν συγκρίσεις μεταξύ χωρών και περιφερειών, διευκολύνοντας τη λήψη αποφάσεων βάσει στοιχείων και την κατανομή πόρων στον τομέα της παγκόσμιας υγείας (Evaluation, 2023).

Το IHME διενεργεί εμπειριστατωμένες αξιολογήσεις της επιβάρυνσης από ασθένειες και τραυματισμούς, συμπεριλαμβανομένων τόσο των μεταδοτικών όσο και των μη μεταδοτικών ασθενειών. Με την ποσοτικοποίηση του αντίκτυπου συγκεκριμένων συνθηκών υγείας στην υγεία του πληθυσμού, το IHME παρέχει πολύτιμες πληροφορίες για τους τομείς προτεραιότητας για παρέμβαση, κατανομή πόρων και ανάπτυξη πολιτικής. Επίσης το IHME αναπτύσσει μοντέλα πρόβλεψης για την πρόβλεψη μελλοντικών τάσεων και αποτελεσμάτων στον τομέα της υγείας. Λαμβάνοντας υπόψη διάφορους παράγοντες, όπως δημογραφικά στοιχεία, παράγοντες κινδύνου και παρεμβάσεις υγειονομικής περίθαλψης, τα μοντέλα πρόβλεψης του IHME βοηθούν στην πρόβλεψη των μελλοντικών επιβαρύνσεων από ασθένειες, των αναγκών υγειονομικής περίθαλψης και του δυνητικού αντίκτυπου των παρεμβάσεων. Οι προβλέψεις αυτές βοηθούν τους υπεύθυνους χάραξης πολιτικής στο σχεδιασμό και την εφαρμογή αποτελεσματικών στρατηγικών για τη βελτίωση της υγείας του πληθυσμού (Evaluation, 2023).

Τέλος το IHME χρησιμοποιεί την έρευνά του και τα δεδομένα του για να υποστηρίξει πολιτικές και παρεμβάσεις βασισμένες σε στοιχεία που βελτιώνουν τα αποτελέσματα της παγκόσμιας υγείας. Ο οργανισμός συνεργάζεται με φορείς χάραξης πολιτικής, διεθνείς οργανισμούς και ενδιαφερόμενους φορείς για να επηρεάσει τις πολιτικές υγείας και να υποστηρίξει τη λήψη αποφάσεων με βάση αυστηρά δεδομένα και ερευνητικά ευρήματα. Το έργο του IHME στοχεύει να έχει άμεσο αντίκτυπο στα συστήματα υγείας, στην κατανομή των πόρων και στις παρεμβάσεις για τη μείωση των παγκόσμιων ανισοτήτων στην υγεία. Επίσης η IHME θέτει ως προτεραιότητα την αποτελεσματική οπτικοποίηση και επικοινωνία των δεδομένων, ώστε οι πολύπλοκες πληροφορίες για την υγεία να είναι προσιτές και κατανοητές. Μέσω διαδραστικών εργαλείων, οπτικοποιήσεων και εκθέσεων, το IHME παρουσιάζει τα αποτελέσματα της έρευνάς του με φιλικό προς το χρήστη τρόπο, επιτρέποντας στους υπεύθυνους

χάραξης πολιτικής, τους ερευνητές και το κοινό να διερευνήσουν και να κατανοήσουν τις παγκόσμιες τάσεις της υγείας.

Ο ρόλος του ΙΗΜΕ ως ερευνητικού ινστιτούτου για την παγκόσμια υγεία έγκειται στην ικανότητά του να παράγει ακριβή, αξιόπιστα και ολοκληρωμένα δεδομένα, μετρήσεις και ερευνητικά ευρήματα για την υγεία. Παρέχοντας κρίσιμες γνώσεις σχετικά με την υγεία του πληθυσμού, την επιβάρυνση από ασθένειες και την απόδοση του συστήματος υγείας, το ΙΗΜΕ ενημερώνει για τις αποφάσεις πολιτικής, καθοδηγεί τα ερευνητικά προγράμματα και υποστηρίζει τις προσπάθειες για τη βελτίωση των αποτελεσμάτων της υγείας παγκοσμίως.

Κεφάλαιο 4ο - Η διαχείριση μεγάλων δεδομένων στην ιατρική

4.1 Εισαγωγή

Η αποτελεσματική διαχείριση δεδομένων επιτρέπει στους επαγγελματίες υγείας να λαμβάνουν τεκμηριωμένες αποφάσεις. Οργανώνοντας, αποθηκεύοντας και αναλύοντας τα δεδομένα υγείας, μπορούν να αποκτήσουν γνώσεις σχετικά με τις συνθήκες των ασθενών, τα αποτελέσματα της θεραπείας και τις τάσεις της δημόσιας υγείας. Αυτό τους επιτρέπει να λαμβάνουν αποφάσεις βασισμένες σε στοιχεία που βελτιώνουν τη φροντίδα και τα αποτελέσματα των ασθενών.

Η διαχείριση δεδομένων στον τομέα της υγείας διαδραματίζει ζωτικό ρόλο στην ενίσχυση της περίθαλψης των ασθενών, στη διευκόλυνση της έρευνας και της καινοτομίας, στη βελτίωση της υγείας του πληθυσμού, στην προώθηση της βελτίωσης της ποιότητας και στη διατήρηση της ασφάλειας και της ιδιωτικότητας των δεδομένων. Αξιοποιώντας τη δύναμη των δεδομένων, οι οργανισμοί υγειονομικής περίθαλψης μπορούν να επιτύχουν καλύτερα αποτελέσματα στην υγεία και να παρέχουν πιο αποτελεσματική και εξατομικευμένη φροντίδα.

4.2 Απόκτηση και ενσωμάτωση

Η απόκτηση δεδομένων και η ενσωμάτωση δεδομένων αποτελούν θεμελιώδη στοιχεία για την αποτελεσματική διαχείριση και αξιοποίηση των μεγάλων δεδομένων. Στην παρούσα υποενότητα, θα παρουσιαστεί τι συνεπάγεται η απόκτηση δεδομένων και η ολοκλήρωση δεδομένων στο πλαίσιο των μεγάλων δεδομένων και πώς συμβάλλουν στη συνολική διαδικασία εξαγωγής πολύτιμων πληροφοριών από μεγάλα και ποικίλα σύνολα δεδομένων.

Η απόκτηση δεδομένων αναφέρεται στη διαδικασία συλλογής δεδομένων από διάφορες πηγές και συστήματα. Στο πλαίσιο των μεγάλων δεδομένων, οι πηγές αυτές μπορεί να περιλαμβάνουν δομημένα δεδομένα από σχεσιακές βάσεις δεδομένων, μη δομημένα δεδομένα από έγγραφα κειμένου ή αναρτήσεις στα μέσα κοινωνικής δικτύωσης, ημιδομημένα δεδομένα από αρχεία XML ή JSON, δεδομένα που

παράγονται από μηχανές από αισθητήρες ή συσκευές του Διαδικτύου των Πραγμάτων (IoT) και άλλα (Stone *et al.*, 2007). Ο πρωταρχικός στόχος της απόκτησης δεδομένων είναι η σύλληψη και η εισαγωγή δεδομένων σε ένα κεντρικό αποθετήριο, όπου μπορούν να υποστούν περαιτέρω επεξεργασία και ανάλυση.

Σε περιβάλλοντα μεγάλων δεδομένων, η απόκτηση δεδομένων θέτει διάφορες προκλήσεις (Henkel and Stieltjes, 2020). Ο όγκος των δεδομένων μπορεί να είναι τεράστιος, συχνά κυμαίνεται από terabytes έως petabytes και πέραν αυτού. Ως εκ τούτου, είναι ζωτικής σημασίας η ύπαρξη ισχυρών μηχανισμών για την αποτελεσματική καταγραφή και αποθήκευση τέτοιων τεράστιων όγκων δεδομένων. Κλιμακούμενα πλαίσια και τεχνολογίες εισαγωγής δεδομένων, όπως το Apache Kafka ή το Apache Flume, χρησιμοποιούνται συνήθως για να χειρίζονται ροές δεδομένων μεγάλου όγκου και να εξασφαλίζουν αξιόπιστη καταγραφή δεδομένων (Akhtar, Khattak and Lee, 2016).

Επιπλέον, η ταχύτητα με την οποία παράγονται τα δεδομένα και η ανάγκη για αναλύσεις σε πραγματικό ή σχεδόν πραγματικό χρόνο απαιτούν μηχανισμούς συλλογής δεδομένων χαμηλής καθυστέρησης. Πλατφόρμες επεξεργασίας ροών, όπως το Apache Spark Streaming ή το Apache Flink, χρησιμοποιούνται συχνά για την εισαγωγή και επεξεργασία δεδομένων σε πραγματικό χρόνο, επιτρέποντας στους οργανισμούς να λαμβάνουν έγκαιρα και βάσει δεδομένων αποφάσεις (Akhtar, Khattak and Lee, 2016).

Εκτός από τον όγκο και την ταχύτητα, τα μεγάλα δεδομένα χαρακτηρίζονται επίσης από την ποικιλία τους. Τα δεδομένα μπορεί να υπάρχουν σε πολλαπλές μορφές και δομές, γεγονός που καθιστά απαραίτητη τη διαχείριση διαφορετικών πηγών δεδομένων κατά τη φάση της απόκτησης. Αυτό απαιτεί ευέλικτους συνδέσμους και προσαρμογείς δεδομένων που μπορούν να εξάγουν δεδομένα από διαφορετικά συστήματα και να τα μετατρέπουν σε μια κοινή μορφή κατάλληλη για περαιτέρω ανάλυση. Επιπλέον, η απόκτηση δεδομένων συχνά περιλαμβάνει διαδικασίες καθαρισμού και επικύρωσης δεδομένων για να διασφαλιστεί η ποιότητα και η συνοχή των δεδομένων πριν από την ενσωμάτωση (Wang *et al.*, 2018).

Μόλις αποκτηθούν τα δεδομένα, το επόμενο βήμα είναι η ολοκλήρωση των δεδομένων. Η ολοκλήρωση δεδομένων περιλαμβάνει τον συνδυασμό δεδομένων από διάφορες πηγές και συστήματα για τη δημιουργία μιας ενιαίας και ολοκληρωμένης άποψης των δεδομένων. Στόχος της είναι να ξεπεραστούν τα "σιλό" δεδομένων και να καταστεί δυνατή η διαλειτουργική ανάλυση με τη συνένωση διαφορετικών συνόλων δεδομένων (Jaleel *et al.*, 2020). Η ενοποίηση δεδομένων είναι ζωτικής σημασίας σε σενάρια μεγάλων δεδομένων, επειδή επιτρέπει στους οργανισμούς να αποκτήσουν πληροφορίες που μπορεί να μην είναι δυνατές μόνο με μεμονωμένα σύνολα δεδομένων.

Η ολοκλήρωση δεδομένων σε περιβάλλοντα μεγάλων δεδομένων συχνά περιλαμβάνει την ενσωμάτωση τόσο δομημένων όσο και μη δομημένων δεδομένων από διαφορετικές πηγές, συμπεριλαμβανομένων των παραδοσιακών βάσεων δεδομένων, του κατακευκτισμένου συστήματος αρχείων Hadoop (HDFS), των πλατφορμών αποθήκευσης cloud και των εξωτερικών παρόχων δεδομένων (Bahga and Madisetti, 2015). Η διαδικασία ολοκλήρωσης μπορεί να περιλαμβάνει μετασχηματισμό δεδομένων, όπου τα δεδομένα τυποποιούνται, καθαρίζονται και εμπλουτίζονται για να διασφαλιστεί η συνέπεια και η συμβατότητα μεταξύ διαφορετικών συνόλων δεδομένων. Αυτό το βήμα μπορεί να περιλαμβάνει τη χρήση εργαλείων και τεχνολογιών ολοκλήρωσης δεδομένων, όπως το Apache Nifi ή το Talend, για την αυτοματοποίηση και τον εξορθολογισμό της διαδικασίας (Bahga and Madisetti, 2015).

Επιπλέον, η ολοκλήρωση δεδομένων στα μεγάλα δεδομένα απαιτεί συχνά το χειρισμό δεδομένων σε πραγματικό ή σχεδόν πραγματικό χρόνο. Πλαίσια ολοκλήρωσης δεδομένων ροής όπως το Apache Kafka, το Apache NiFi ή το Apache Flink χρησιμοποιούνται για τη συνεχή επεξεργασία και ολοκλήρωση δεδομένων ροής από πολλαπλές πηγές. Αυτές οι τεχνολογίες επιτρέπουν την εισαγωγή, μετατροπή και ολοκλήρωση δεδομένων σε πραγματικό χρόνο, επιτρέποντας στους οργανισμούς να αντλούν πληροφορίες από ενημερωμένες πληροφορίες. Η ολοκλήρωση δεδομένων περιλαμβάνει επίσης την επίλυση ασυνεπειών δεδομένων, όπως ο χειρισμός αντικρουόμενων αναπαραστάσεων δεδομένων, η επίλυση διπλών εγγραφών ή η διαχείριση της εξέλιξης του σχήματος δεδομένων. Αυτές οι προκλήσεις προκύπτουν

λόγω της αποκεντρωμένης φύσης των πηγών δεδομένων και της πιθανής έλλειψης προτύπων διακυβέρνησης δεδομένων. Η αντιμετώπιση αυτών των ζητημάτων απαιτεί την εφαρμογή κατάλληλων τεχνικών και στρατηγικών ολοκλήρωσης δεδομένων, συμπεριλαμβανομένης της αντιστοίχισης δεδομένων, του αποδιπλασιασμού, της επίλυσης οντοτήτων και της διαχείρισης της εξέλιξης του σχήματος (Brazhnik and Jones, 2007).

Επιπλέον, η ολοκλήρωση δεδομένων μπορεί να περιλαμβάνει τη χρήση τεχνικών εικονικοποίησης δεδομένων, όπου τα δεδομένα ενσωματώνονται εικονικά χωρίς φυσική μετακίνηση ή αντιγραφή τους. Η προσέγγιση αυτή παρέχει μια ενοποιημένη άποψη των δεδομένων, διατηρώντας παράλληλα τα δεδομένα στις αρχικές τους πηγές. Η εικονικοποίηση δεδομένων μπορεί να είναι ιδιαίτερα επωφελής σε σενάρια μεγάλων δεδομένων, όπου η μετακίνηση και η αντιγραφή τεράστιων όγκων δεδομένων μπορεί να είναι ανέφικτη ή να απαιτεί πολλούς πόρους (Brazhnik and Jones, 2007).

Συνολικά, η απόκτηση δεδομένων και η ενσωμάτωση δεδομένων είναι ζωτικής σημασίας βήματα για την αποτελεσματική αξιοποίηση των μεγάλων δεδομένων. Η απόκτηση δεδομένων επικεντρώνεται στη σύλληψη και τη λήψη ποικίλων δεδομένων από διάφορες πηγές, εξασφαλίζοντας επεκτασιμότητα, χαμηλή καθυστέρηση και υψηλή ποιότητα δεδομένων. Η ολοκλήρωση δεδομένων, από την άλλη πλευρά, περιλαμβάνει τον συνδυασμό και την εναρμόνιση δεδομένων από διαφορετικές πηγές για τη δημιουργία μιας ενοποιημένης προβολής που επιτρέπει την ολοκληρωμένη ανάλυση και την εξαγωγή πολύτιμων συμπερασμάτων. Μαζί, αυτές οι διαδικασίες θέτουν τα θεμέλια για προηγμένη ανάλυση, μηχανική μάθηση και λήψη αποφάσεων βάσει δεδομένων στο βασίλειο των μεγάλων δεδομένων.

4.3 Αποθήκευση και υποδομές

Η αποθήκευση και η υποδομή των δεδομένων υγείας στο πλαίσιο των μεγάλων δεδομένων διαδραματίζουν κρίσιμο ρόλο στη διαχείριση και ανάλυση τεράστιων ποσοτήτων πληροφοριών που σχετίζονται με την υγειονομική περίθαλψη. Στην παρούσα υποενότητα θα παρουσιαστούν οι βασικές πτυχές της αποθήκευσης

δεδομένων και της υποδομής για τα δεδομένα υγείας στο πεδίο των μεγάλων δεδομένων, εστιάζοντας στις μοναδικές προκλήσεις, τεχνολογίες και εκτιμήσεις που αφορούν ειδικά τον κλάδο της υγειονομικής περίθαλψης.

Μία από τις πρωταρχικές προκλήσεις στην αποθήκευση δεδομένων υγείας σε περιβάλλοντα μεγάλων δεδομένων είναι ο τεράστιος όγκος των πληροφοριών που παράγονται. Οι οργανισμοί υγειονομικής περίθαλψης παράγουν καθημερινά τεράστιο όγκο δεδομένων και τα δεδομένα αυτά πρέπει να αποθηκεύονται με ασφάλεια και αποτελεσματικότητα. Για την αντιμετώπιση αυτής της πρόκλησης χρησιμοποιούνται συνήθως κλιμακούμενα και κατανεμημένα συστήματα αποθήκευσης. Τεχνολογίες όπως το Apache Hadoop Distributed File System (HDFS) και λύσεις αποθήκευσης αντικειμένων όπως το Amazon S3 ή το Google Cloud Storage παρέχουν κλιμακούμενη και ανεκτική σε σφάλματα αποθήκευση για αποθετήρια δεδομένων υγείας μεγάλης κλίμακας (Nazari, Shahriari and Tabesh, 2019).

Ένα άλλο κρίσιμο ζήτημα στην αποθήκευση δεδομένων υγείας είναι η ιδιωτικότητα και η ασφάλεια των δεδομένων. Τα δεδομένα υγείας είναι ιδιαίτερα ευαίσθητα και υπόκεινται σε αυστηρούς κανονισμούς προστασίας της ιδιωτικής ζωής, όπως ο νόμος περί φορητότητας και λογοδοσίας της ασφάλισης υγείας (HIPAA) στις Ηνωμένες Πολιτείες. Ως εκ τούτου, είναι σημαντικό να διασφαλιστεί ότι τα συστήματα αποθήκευσης συμμορφώνονται με τους εν λόγω κανονισμούς και ενσωματώνουν ισχυρά μέτρα ασφαλείας για την προστασία της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας των δεδομένων. Η κρυπτογράφηση, οι έλεγχοι πρόσβασης, τα αρχεία καταγραφής ελέγχου και οι τεχνικές απόκρυψης δεδομένων είναι μερικά από τα μέτρα ασφαλείας που χρησιμοποιούνται συνήθως στην αποθήκευση δεδομένων υγειονομικής περίθαλψης.

Εκτός από τον όγκο και την ασφάλεια, η ποικιλία και η ετερογένεια των δεδομένων υγείας δημιουργούν σημαντικές προκλήσεις στην αποθήκευση και την υποδομή. Τα δεδομένα υγείας διατίθενται σε διάφορες μορφές και δομές, συμπεριλαμβανομένων δομημένων δεδομένων (π.χ. δημογραφικά στοιχεία ασθενών, εργαστηριακά αποτελέσματα), μη δομημένων δεδομένων (π.χ. κλινικές σημειώσεις, ιατρικές απεικονίσεις) και ημιδομημένων δεδομένων (π.χ. μηνύματα HL7, αρχεία υγείας βασισμένα σε XML). Τα συστήματα αποθήκευσης πρέπει να προσαρμόζονται σε αυτή

την ποικιλία και να παρέχουν μηχανισμούς για την αποτελεσματική αποθήκευση, οργάνωση και ανάκτηση διαφορετικών τύπων δεδομένων υγείας (Manogaran et al., 2017).

Για να αντιμετωπιστούν οι ποικίλες ανάγκες αποθήκευσης δεδομένων υγείας, χρησιμοποιούνται συχνά αποθήκες δεδομένων σε περιβάλλοντα μεγάλων δεδομένων. Αυτές αποτελούν ένα κεντρικό αποθετήριο που αποθηκεύει ακατέργαστα, μη επεξεργασμένα δεδομένα από διάφορες πηγές, επιτρέποντας την ευέλικτη ανάλυση και εξερεύνηση (Wade, 2014). Οι λίμνες δεδομένων επιτρέπουν στους οργανισμούς υγειονομικής περίθαλψης να αποθηκεύουν και να διαχειρίζονται τεράστιους όγκους δεδομένων υγείας, διατηρώντας παράλληλα την αρχική τους μορφή. Από την άλλη πλευρά, οι αποθήκες δεδομένων παρέχουν μια δομημένη και οργανωμένη προβολή των δεδομένων, συχνά μέσω προσεγγίσεων μοντελοποίησης δεδομένων και *schema-on-read*. Οι αποθήκες δεδομένων επιτρέπουν βελτιστοποιημένες δυνατότητες αναζήτησης και υποβολής εκθέσεων, καθιστώντας τις κατάλληλες για αναλυτικές περιπτώσεις χρήσης (Wade, 2014).

Στο πλαίσιο των δεδομένων υγείας, οι εκτιμήσεις για την υποδομή υπερβαίνουν την αποθήκευση και μόνο. Οι οργανισμοί υγειονομικής περίθαλψης απαιτούν μια στιβαρή και κλιμακούμενη υποδομή για την υποστήριξη της αποθήκευσης, της επεξεργασίας και της ανάλυσης μεγάλων δεδομένων. Αυτή η υποδομή περιλαμβάνει συνήθως έναν συνδυασμό πόρων στις εγκαταστάσεις, πλατφόρμες που βασίζονται στο νέφος ή υβριδικά περιβάλλοντα (Rahimi et al., 2022).

Το υπολογιστικό νέφος έχει αποκτήσει σημαντική έλξη στην υγειονομική περίθαλψη λόγω της επεκτασιμότητας, της ελαστικότητας και της σχέσης κόστους-αποτελεσματικότητας. Οι πάροχοι υπολογιστικού νέφους προσφέρουν διαχειριζόμενες υπηρεσίες αποθήκευσης, όπως το Amazon S3 ή το Azure Blob Storage, οι οποίες επιτρέπουν στους οργανισμούς υγειονομικής περίθαλψης να αποθηκεύουν και να έχουν πρόσβαση σε μεγάλους όγκους δεδομένων υγείας χωρίς την ανάγκη αρχικών επενδύσεων σε υποδομές. Οι υποδομές που βασίζονται στο νέφος παρέχουν επίσης την ευελιξία να αυξάνουν ή να μειώνουν τους πόρους ανάλογα με τη ζήτηση, καθιστώντας τις κατάλληλες για τον χειρισμό κυμαινόμενων φόρτων εργασίας (Rahimi et al., 2022).

Εκτός από την αποθήκευση, η υποδομή για τα δεδομένα υγείας σε περιβάλλοντα μεγάλων δεδομένων θα πρέπει να υποστηρίζει αποτελεσματική επεξεργασία και ανάλυση δεδομένων. Τεχνολογίες όπως το Apache Hadoop, το Apache Spark ή πλατφόρμες ανάλυσης που βασίζονται στο cloud, όπως το Google BigQuery ή το Amazon Redshift, προσφέρουν δυνατότητες κατανεμημένης επεξεργασίας για τη διαχείριση εργασιών επεξεργασίας και ανάλυσης δεδομένων μεγάλης κλίμακας. Αυτές οι τεχνολογίες επιτρέπουν στους οργανισμούς υγειονομικής περίθαλψης να εκτελούν προηγμένες αναλύσεις, μηχανική εκμάθηση και διορατικότητα με βάση την τεχνητή νοημοσύνη σε δεδομένα υγείας (Nazari, Shahriari and Tabesh, 2019).

Επιπλέον, η υποδομή θα πρέπει να παρέχει τα απαραίτητα εργαλεία και πλαίσια για την ολοκλήρωση δεδομένων, τον καθαρισμό δεδομένων και τη διακυβέρνηση δεδομένων. Πλαίσια ολοκλήρωσης δεδομένων όπως το Apache Nifi ή το Talend μπορούν να χρησιμοποιηθούν για την εξαγωγή, τον μετασχηματισμό και τη φόρτωση δεδομένων υγείας από διάφορες πηγές στα συστήματα αποθήκευσης. Οι διαδικασίες εκκαθάρισης και διακυβέρνησης δεδομένων είναι ζωτικής σημασίας για τη διασφάλιση της ποιότητας των δεδομένων, της τυποποίησης και της συμμόρφωσης με τις κανονιστικές απαιτήσεις (Nazari, Shahriari and Tabesh, 2019).

Η αποθήκευση και η υποδομή δεδομένων υγείας σε περιβάλλοντα μεγάλων δεδομένων απαιτούν κλιμακούμενες, ασφαλείς και ευέλικτες λύσεις. Οι οργανισμοί υγειονομικής περίθαλψης πρέπει να λαμβάνουν υπόψη τους τον όγκο, την προστασία της ιδιωτικής ζωής και την ετερογένεια των δεδομένων υγείας, ενώ εφαρμόζουν συστήματα αποθήκευσης και υποδομές που μπορούν να αντεπεξέλθουν στις μοναδικές απαιτήσεις του κλάδου της υγειονομικής περίθαλψης. Με τα κατάλληλα συστήματα αποθήκευσης, υποδομές και αναλυτικά εργαλεία, οι οργανισμοί υγειονομικής περίθαλψης μπορούν να αξιοποιήσουν τα μεγάλα δεδομένα για να βελτιώσουν τα αποτελέσματα των ασθενών, να προωθήσουν την έρευνα και να βελτιώσουν την επιχειρησιακή αποτελεσματικότητα.

4.4 Ανάλυση δεδομένων

Η ανάλυση δεδομένων υγείας στο πλαίσιο των μεγάλων δεδομένων περιλαμβάνει την εξαγωγή, μετατροπή και ανάλυση τεράστιου όγκου πληροφοριών που σχετίζονται με την υγειονομική περίθαλψη για την εξαγωγή πολύτιμων

συμπερασμάτων και την τεκμηριωμένη λήψη αποφάσεων. Η ανάλυση δεδομένων υγείας είναι ζωτικής σημασίας για τους οργανισμούς υγειονομικής περίθαλψης προκειμένου να αποκτήσουν αξιοποιήσιμες γνώσεις, να βελτιώσουν τα αποτελέσματα των ασθενών, να βελτιστοποιήσουν τις λειτουργίες και να καταστήσουν δυνατή τη λήψη αποφάσεων βάσει τεκμηρίων. Αξιοποιώντας τη δύναμη των μεγάλων δεδομένων, οι πάροχοι υγειονομικής περίθαλψης μπορούν να αποκαλύψουν μοτίβα, τάσεις, συσχετίσεις και προγνωστικά μοντέλα που μπορούν να οδηγήσουν σε εξελίξεις στην εξατομικευμένη ιατρική, τη διαχείριση της υγείας του πληθυσμού, την πρόληψη ασθενειών και την αποτελεσματικότητα της παροχής υγειονομικής περίθαλψης (Kankanhalli *et al.*, 2016).

Μία από τις πρωταρχικές προκλήσεις στην ανάλυση δεδομένων υγείας είναι ο μεγάλος όγκος και η ποικιλία των δεδομένων υγειονομικής περίθαλψης. Τα μεγάλα δεδομένα στην υγειονομική περίθαλψη περιλαμβάνουν δομημένα δεδομένα (π.χ. ηλεκτρονικά αρχεία υγείας, εργαστηριακά αποτελέσματα), μη δομημένα δεδομένα (π.χ. κλινικές σημειώσεις, ιατρική απεικόνιση), δεδομένα χρονοσειρών (π.χ. ζωτικά σημεία, wearables) και γονιδιωματικά δεδομένα, μεταξύ άλλων. Η ανάλυση και η ενσωμάτωση αυτών των ποικίλων πηγών δεδομένων απαιτούν προηγμένες τεχνικές ανάλυσης και κλιμακούμενη υποδομή ικανή να χειρίζεται επεξεργασία δεδομένων μεγάλης κλίμακας (Saranya and Asha, 2019).

Για την αντιμετώπιση αυτών των προκλήσεων, οι οργανισμοί υγειονομικής περίθαλψης χρησιμοποιούν διάφορες τεχνικές ανάλυσης δεδομένων στο πεδίο των μεγάλων δεδομένων (Elgendy and Elragal, 2014):

- Περιγραφική ανάλυση: Η περιγραφική ανάλυση περιλαμβάνει την εξερεύνηση και τη σύνοψη ιστορικών δεδομένων υγείας για την κατανόηση μοτίβων, τάσεων και πληροφοριών. Παρέχει μια αναδρομική άποψη των λειτουργιών υγειονομικής περίθαλψης και των αποτελεσμάτων των ασθενών, βοηθώντας στον εντοπισμό τομέων για βελτίωση και συγκριτική αξιολόγηση των επιδόσεων.
- Διαγνωστική ανάλυση: Η διαγνωστική ανάλυση αποσκοπεί στην ανακάλυψη των βαθύτερων αιτιών παρελθόντων γεγονότων ή αποτελεσμάτων

εξετάζοντας δεδομένα υγειονομικής περίθαλψης. Περιλαμβάνει την ανάλυση συσχετίσεων, την εκτέλεση αναλύσεων κοόρτης και τη διεξαγωγή αναδρομικών μελετών για την αποκάλυψη παραγόντων που επηρεάζουν συγκεκριμένες συνθήκες υγείας ή αποτελέσματα.

- Προγνωστική ανάλυση (Predictive Analytics): Η προγνωστική ανάλυση αξιοποιεί τη στατιστική μοντελοποίηση και τους αλγόριθμους μηχανικής μάθησης για την πρόβλεψη μελλοντικών συμβάντων ή αποτελεσμάτων. Χρησιμοποιεί ιστορικά δεδομένα για τον εντοπισμό μοτίβων και τη δημιουργία προγνωστικών μοντέλων που μπορούν να εκτιμήσουν την πιθανότητα συγκεκριμένων καταστάσεων υγείας, επανεισαγωγών ή ανεπιθύμητων συμβάντων. Η προγνωστική ανάλυση βοηθά στη διαστρωμάτωση του κινδύνου, την έγκαιρη ανίχνευση και τον εξατομικευμένο σχεδιασμό της θεραπείας (Poornima and Pushpalatha, 2018).
- Προδιαγραφική ανάλυση: Τα Prescriptive analytics συνδυάζουν ιστορικά δεδομένα με τεχνικές βελτιστοποίησης και μοντέλα λήψης αποφάσεων για την παροχή αξιοποιήσιμων συστάσεων. Βοηθά τους παρόχους υγειονομικής περίθαλψης να βελτιστοποιήσουν την κατανομή των πόρων, τα σχέδια θεραπείας, τον προγραμματισμό και τη διαχείριση της ροής εργασίας, προσομοιώνοντας διαφορετικά σενάρια και προσδιορίζοντας την καλύτερη πορεία δράσης.
- Ανάλυση κειμένου: Οι τεχνικές ανάλυσης κειμένου επιτρέπουν την εξαγωγή πολύτιμων πληροφοριών από αδόμητες κλινικές σημειώσεις, ερευνητικά άρθρα και άλλες πηγές κειμένου. Οι αλγόριθμοι επεξεργασίας φυσικής γλώσσας (NLP) και οι τεχνικές εξόρυξης κειμένου χρησιμοποιούνται για την εξαγωγή πληροφοριών, τον εντοπισμό σχέσεων και την ταξινόμηση δεδομένων κειμένου για την υποστήριξη της κλινικής λήψης αποφάσεων και της έρευνας.
- Μηχανική μάθηση: Οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούνται ευρέως στην ανάλυση δεδομένων υγείας για την ανακάλυψη μοτίβων, την ταξινόμηση δεδομένων, την πραγματοποίηση προβλέψεων και την αυτοματοποίηση διαδικασιών λήψης αποφάσεων. Οι αλγόριθμοι μάθησης με

επίβλεψη, όπως η λογιστική παλινδρόμηση ή οι μηχανές διανυσμάτων υποστήριξης, μπορούν να εφαρμοστούν σε δεδομένα υγειονομικής περίθαλψης για την ανάπτυξη μοντέλων πρόβλεψης. Οι τεχνικές μάθησης χωρίς επίβλεψη, όπως η ομαδοποίηση ή η μείωση διαστάσεων, βοηθούν στον εντοπισμό υποομάδων ασθενών ή στην επιλογή χαρακτηριστικών (Beam and Kohane, 2018).

4.5 Ποιότητα δεδομένων και προεπεξεργασία

Η ποιότητα των δεδομένων και η προεπεξεργασία είναι κρίσιμα βήματα στην ανάλυση των δεδομένων υγείας σε περιβάλλοντα μεγάλων δεδομένων. Η υψηλή ποιότητα των δεδομένων είναι απαραίτητη για την ακριβή και αξιόπιστη ανάλυση, ενώ οι τεχνικές προεπεξεργασίας συμβάλλουν στην προετοιμασία των δεδομένων για τη μετέπειτα ανάλυση. Η ποιότητα δεδομένων αναφέρεται στην ακρίβεια, την πληρότητα, τη συνέπεια και την αξιοπιστία των δεδομένων. Στο πλαίσιο των δεδομένων υγείας, η διατήρηση της ποιότητας των δεδομένων είναι ζωτικής σημασίας για τη διασφάλιση της εγκυρότητας και της αξιοπιστίας των αποτελεσμάτων της ανάλυσης. Η ποιότητα των δεδομένων έχει συγκεκριμένες διαστάσεις που θα πρέπει να καλυφθούν (Rao, Gudivada and Raghavan, 2015). Καταρχάς η ακρίβεια. Τα δεδομένα δεν πρέπει να περιέχουν λάθη, ασυνέπειες ή αποκλίσεις. Θα πρέπει να αντικατοπτρίζουν τις πραγματικές τιμές και τα χαρακτηριστικά των οντοτήτων που σχετίζονται με την υγεία και μετρούνται. Επίσης η πληρότητα. Τα δεδομένα θα πρέπει να είναι πλήρη και να περιλαμβάνουν όλα τα σχετικά χαρακτηριστικά και μεταβλητές που απαιτούνται για την ανάλυση. Τα δεδομένα που λείπουν μπορεί να εισάγουν μεροληψίες και να παρεμποδίζουν την αποτελεσματικότητα της ανάλυσης. Ακολούθως είναι η συνέπεια. Τα δεδομένα θα πρέπει να είναι εσωτερικά συνεπή, δηλαδή να μην περιέχουν αντικρουόμενες ή αντιφατικές πληροφορίες. Η συνέπεια εξασφαλίζει ότι τα δεδομένα είναι συνεκτικά και ευθυγραμμίζονται με προκαθορισμένους κανόνες ή πρότυπα. Τέλος είναι η αξιοπιστία. Τα δεδομένα πρέπει να είναι αξιόπιστα και να προέρχονται από αξιόπιστες πηγές. Θα πρέπει να συλλέγονται με τη χρήση επικυρωμένων οργάνων μέτρησης ή καθιερωμένων διαδικασιών συλλογής δεδομένων (Rao, Gudivada and Raghavan, 2015).

Από την άλλη πλευρά προεπεξεργασία δεδομένων περιλαμβάνει μια σειρά βημάτων για τη μετατροπή των ακατέργαστων δεδομένων υγείας σε μορφή κατάλληλη για ανάλυση. Οι τεχνικές προεπεξεργασίας βοηθούν στην αντιμετώπιση ζητημάτων όπως τα ελλιπή δεδομένα, οι ακραίες τιμές, ο θόρυβος και η ενσωμάτωση δεδομένων. Ακολουθούν ορισμένες κοινές τεχνικές προεπεξεργασίας που χρησιμοποιούνται στην ανάλυση δεδομένων υγείας (Taleb, Dssouli and Serhani, 2015; Hariharakrishnan *et al.*, 2017):

- Καθαρισμός δεδομένων: Αυτό το βήμα περιλαμβάνει τον χειρισμό των δεδομένων που λείπουν, των ακραίων τιμών και των ασυνεπειών. Τα ελλείποντα δεδομένα μπορούν να αντιμετωπιστούν μέσω τεχνικών όπως ο υπολογισμός, όπου οι ελλείπουσες τιμές εκτιμώνται με βάση τις διαθέσιμες πληροφορίες. Οι ακραίες τιμές μπορούν να εντοπιστούν και να αντιμετωπιστούν βάσει στατιστικών μεθόδων ή γνώσεων του τομέα. Τα ασυνεπή δεδομένα μπορούν να επιλυθούν με επικύρωση βάσει προκαθορισμένων κανόνων ή προτύπων.
- Ολοκλήρωση δεδομένων: Τα δεδομένα υγείας συχνά προέρχονται από διάφορες πηγές και μπορεί να είναι αποθηκευμένα σε διαφορετικές μορφές ή δομές. Η ενοποίηση δεδομένων περιλαμβάνει το συνδυασμό και τη συγχώνευση δεδομένων από διαφορετικές πηγές, εξασφαλίζοντας τη συνοχή και τη συμβατότητα. Αυτό μπορεί να περιλαμβάνει την αντιστοίχιση των χαρακτηριστικών των δεδομένων, την επίλυση συγκρούσεων και τη μετατροπή των δεδομένων σε ενιαία μορφή.
- Επιλογή χαρακτηριστικών: Στην ανάλυση δεδομένων υγείας, είναι σύνηθες να αντιμετωπίζεται μεγάλος αριθμός μεταβλητών ή χαρακτηριστικών. Οι τεχνικές επιλογής χαρακτηριστικών βοηθούν στον εντοπισμό των πιο σχετικών και κατατοπιστικών χαρακτηριστικών για ανάλυση. Αυτό μειώνει την υπολογιστική πολυπλοκότητα, ελαχιστοποιεί τον θόρυβο και βελτιώνει την ερμηνευσιμότητα των αποτελεσμάτων της ανάλυσης.
- Μείωση διαστάσεων: Τα δεδομένα υγείας μπορεί να είναι υψηλής διάστασης, δηλαδή περιέχουν μεγάλο αριθμό μεταβλητών. Οι τεχνικές μείωσης της διαστατικότητας, όπως η ανάλυση κύριων συνιστωσών (PCA) ή η

t-SNE, μειώνουν τον αριθμό των μεταβλητών, διατηρώντας παράλληλα σημαντικές πληροφορίες. Αυτό απλοποιεί τη διαδικασία ανάλυσης και οπτικοποίησης.

- Κανονικοποίηση και τυποποίηση: Οι τεχνικές κανονικοποίησης ή τυποποίησης δεδομένων εφαρμόζονται για να φέρουν τα δεδομένα σε μια κοινή κλίμακα ή κατανομή. Αυτό εξασφαλίζει ότι διαφορετικές μεταβλητές με διαφορετικές μονάδες μέτρησης ή εύρος τιμών είναι συγκρίσιμες και δεν εισάγουν μεροληψία στην ανάλυση.
- Επεξεργασία κειμένου: Τα μη δομημένα δεδομένα κειμένου, όπως κλινικές σημειώσεις ή ερευνητικά άρθρα, μπορούν να υποβληθούν σε επεξεργασία με τη χρήση τεχνικών επεξεργασίας φυσικής γλώσσας (NLP). Οι αλγόριθμοι NLP εξάγουν σχετικές πληροφορίες, εκτελούν αναγνώριση οντοτήτων, ανάλυση συναισθήματος ή μοντελοποίηση θεμάτων, επιτρέποντας τη δομημένη ανάλυση δεδομένων κειμένου.

Για να διασφαλιστεί η ποιότητα των δεδομένων και η ακεραιότητα της προεπεξεργασίας, είναι ζωτικής σημασίας η καθιέρωση διαδικασιών διασφάλισης ποιότητας και η τεκμηρίωση όλων των βημάτων που εκτελούνται κατά την προεπεξεργασία. Αυτό περιλαμβάνει τη διατήρηση μιας διαδρομής ελέγχου των μετασχηματισμών δεδομένων, την καταγραφή των αποφάσεων που λαμβάνονται κατά τον καθαρισμό ή τον υπολογισμό των δεδομένων και την τεκμηρίωση των παραδοχών και των μεθοδολογιών που χρησιμοποιήθηκαν κατά την προεπεξεργασία. Η σωστή τεκμηρίωση εξασφαλίζει την αναπαραγωγικότητα και τη διαφάνεια στη διαδικασία ανάλυσης. Συνοπτικά, η ποιότητα των δεδομένων και η προεπεξεργασία αποτελούν βασικές πτυχές της ανάλυσης δεδομένων υγείας σε περιβάλλοντα μεγάλων δεδομένων. Η διασφάλιση της ποιότητας των δεδομένων μέσω της ακρίβειας, της πληρότητας, της συνέπειας και της αξιοπιστίας ενισχύει την εγκυρότητα.

4.6 Ασφάλεια δεδομένων

Η ασφάλεια των δεδομένων υγείας στα μεγάλα δεδομένα αποτελεί κρίσιμο ζήτημα λόγω της ευαίσθητης και προσωπικής φύσης των σχετικών πληροφοριών. Καθώς ο

κλάδος της υγειονομικής περίθαλψης υιοθετεί ολοένα και περισσότερο ψηφιακά συστήματα και αναλύσεις μεγάλων δεδομένων, η διασφάλιση της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας των δεδομένων υγείας καθίσταται υψίστης σημασίας. Σε πρώτο επίπεδο η κρυπτογράφηση δεδομένων (Benjelloun and Lahcen, 2019). Η κρυπτογράφηση είναι μια θεμελιώδης τεχνική που χρησιμοποιείται για τη διασφάλιση των δεδομένων υγείας στα μεγάλα δεδομένα. Περιλαμβάνει την κωδικοποίηση των πληροφοριών με τη χρήση κρυπτογραφικών αλγορίθμων ώστε να καταστούν μη αναγνώσιμες από μη εξουσιοδοτημένα άτομα. Συνήθως χρησιμοποιούνται δύο πρωταρχικοί τύποι κρυπτογράφησης: σε κατάσταση ηρεμίας και κατά τη μεταφορά. Η κρυπτογράφηση σε κατάσταση ηρεμίας διασφαλίζει τα δεδομένα όταν αποθηκεύονται σε βάσεις δεδομένων ή σε φυσικά μέσα, ενώ η κρυπτογράφηση κατά τη μεταφορά προστατεύει τα δεδομένα καθώς μεταδίδονται μεταξύ συστημάτων ή μέσω δικτύων. Αλγόριθμοι κρυπτογράφησης, όπως το Advanced Encryption Standard (AES), χρησιμοποιούνται συνήθως για την προστασία των δεδομένων υγείας.

Ταυτόχρονα απαιτείται έλεγχος πρόσβασης (Benjelloun and Lahcen, 2019). Ο έλεγχος της πρόσβασης στα δεδομένα υγείας είναι ζωτικής σημασίας για τη διασφάλιση της ασφάλειας. Εφαρμόζονται μηχανισμοί ελέγχου πρόσβασης για τον περιορισμό της πρόσβασης στα δεδομένα μόνο σε εξουσιοδοτημένο προσωπικό. Συχνά χρησιμοποιείται έλεγχος πρόσβασης βάσει ρόλων (RBAC), όπου τα δικαιώματα πρόσβασης εκχωρούνται βάσει συγκεκριμένων ρόλων και αρμοδιοτήτων εντός του οργανισμού υγειονομικής περίθαλψης. Η προσέγγιση αυτή διασφαλίζει ότι τα άτομα μπορούν να έχουν πρόσβαση μόνο στα δεδομένα που είναι απαραίτητα για τα εργασιακά τους καθήκοντα. Επιπλέον, μπορεί να εφαρμοστεί έλεγχος ταυτότητας πολλαπλών παραγόντων (MFA) για την ενίσχυση των ελέγχων πρόσβασης, απαιτώντας από τους χρήστες να παρέχουν πολλαπλές μορφές ταυτοποίησης (π.χ. κωδικούς πρόσβασης, βιομετρικά στοιχεία) για να αποκτήσουν πρόσβαση σε δεδομένα υγείας.

Τι γίνεται όμως με την εξουσιοδότηση των χρηστών; Η σωστή ταυτοποίηση, αυθεντικοποίηση και εξουσιοδότηση των χρηστών είναι απαραίτητη για την προστασία των δεδομένων υγείας. Συνήθως εφαρμόζονται ισχυρές πολιτικές

κωδικών πρόσβασης, συμπεριλαμβανομένων απαιτήσεων για σύνθετους κωδικούς πρόσβασης και τακτικές αλλαγές κωδικών πρόσβασης. Οι μηχανισμοί ελέγχου ταυτότητας χρηστών, όπως συνδυασμοί ονόματος χρήστη και κωδικού πρόσβασης ή βιομετρικός έλεγχος ταυτότητας (π.χ. σάρωση δακτυλικών αποτυπωμάτων ή ίριδας), βοηθούν στην επαλήθευση της ταυτότητας των ατόμων που έχουν πρόσβαση στα δεδομένα. Οι διαδικασίες εξουσιοδότησης καθορίζουν τις ενέργειες και τα προνόμια που μπορούν να εκτελέσουν οι πιστοποιημένοι χρήστες στα δεδομένα υγείας, διασφαλίζοντας ότι τα άτομα έχουν πρόσβαση μόνο στις απαραίτητες πληροφορίες (Patil and Seshadri, 2014).

Για την περαιτέρω προστασία των δεδομένων υγείας μπορούν να χρησιμοποιηθούν τεχνικές όπως η ανωνυμοποίηση και η απο-ταυτοποίηση. Η ανωνυμοποίηση περιλαμβάνει την αφαίρεση ή την τροποποίηση των προσωπικών πληροφοριών (PII) από το σύνολο δεδομένων, καθιστώντας αδύνατη τη σύνδεση των δεδομένων με ένα άτομο. Οι τεχνικές αποταυτοποίησης, όπως η αφαίρεση άμεσων αναγνωριστικών στοιχείων (π.χ. όνομα, διεύθυνση) ή η εφαρμογή μάσκας δεδομένων, μετασχηματίζουν τα δεδομένα ώστε να ελαχιστοποιηθεί ο κίνδυνος επαναταυτοποίησης (Abouelmehdi *et al.*, 2017). Αυτές οι μέθοδοι επιτρέπουν στους οργανισμούς να χρησιμοποιούν τα δεδομένα υγείας για έρευνα και ανάλυση, διατηρώντας παράλληλα την ιδιωτικότητα. Για την οχύρωση των ιατρικών δεδομένων στην εποχή των μεγάλων δεδομένων εφαρμόζονται και άλλα προληπτικά μέτρα. Εφαρμόζονται μέτρα πρόληψης απώλειας δεδομένων για την αποτροπή μη εξουσιοδοτημένης αποκάλυψης ή τυχαίας απώλειας δεδομένων υγείας (Abouelmehdi *et al.*, 2017). Οι στρατηγικές DLP περιλαμβάνουν την παρακολούθηση των ροών δεδομένων και την εφαρμογή πολιτικών για τον εντοπισμό και την πρόληψη παραβιάσεων δεδομένων. Αυτό περιλαμβάνει την εφαρμογή τειχών προστασίας, συστημάτων ανίχνευσης και πρόληψης εισβολών (IDPS) και λογισμικού πρόληψης απώλειας δεδομένων για την παρακολούθηση και τον έλεγχο της πρόσβασης και της κίνησης δεδομένων. Οι προηγμένες λύσεις DLP χρησιμοποιούν αλγορίθμους μηχανικής μάθησης για τον εντοπισμό μοτίβων ενδεικτικών πιθανών παραβιάσεων δεδομένων, εξασφαλίζοντας ταχεία αντίδραση και μετριασμό.

Επίσης οι συνεχείς έλεγχοι ασφαλείας και οι αξιολογήσεις κινδύνου είναι ζωτικής σημασίας για τον εντοπισμό τρωτών σημείων και αδυναμιών στην υποδομή ασφαλείας. Οι οργανισμοί θα πρέπει να διενεργούν τακτικές αξιολογήσεις για την αξιολόγηση της αποτελεσματικότητας των ελέγχων ασφαλείας, τον εντοπισμό πιθανών απειλών και τρωτών σημείων και την εφαρμογή μέτρων αποκατάστασης. Οι αξιολογήσεις αυτές συμβάλλουν στη διασφάλιση ότι τα μέτρα ασφαλείας είναι επικαιροποιημένα, ευθυγραμμισμένα με τις βέλτιστες πρακτικές του κλάδου και ικανά να προστατεύουν τα δεδομένα υγείας από τις εξελισσόμενες απειλές. Ταυτόχρονα όταν τα δεδομένα υγείας μεταφέρονται μεταξύ διαφορετικών συστημάτων ή μοιράζονται με εξωτερικές οντότητες, θα πρέπει να χρησιμοποιούνται ασφαλή πρωτόκολλα μετάδοσης. Τα πρωτόκολλα Secure Socket Layer (SSL) ή Transport Layer Security (TLS) παρέχουν ασφαλή κανάλια επικοινωνίας, κρυπτογραφώντας τα δεδομένα κατά τη μεταφορά και προστατεύοντάς τα από υποκλοπή ή παραποίηση.

Κλασσικό παράδειγμα/ κίνηση διασφάλισης των δεδομένων είναι η δημιουργία αντιγράφων ασφαλείας δεδομένων και αποκατάσταση καταστροφών. Οι αξιόπιστες στρατηγικές δημιουργίας αντιγράφων ασφαλείας δεδομένων και ανάκτησης από καταστροφές είναι απαραίτητες για τη διασφάλιση της διαθεσιμότητας και της ακεραιότητας των δεδομένων υγείας. Πραγματοποιούνται τακτικά αντίγραφα ασφαλείας για τη δημιουργία αντιγράφων δεδομένων που μπορούν να αποκατασταθούν σε περίπτωση απώλειας δεδομένων ή βλάβης του συστήματος. Τα εφεδρικά δεδομένα αποθηκεύονται συχνά σε ασφαλείς εξωτερικές τοποθεσίες για τον μετριασμό των κινδύνων που συνδέονται με φυσικές ζημιές ή φυσικές καταστροφές. Επιπλέον, τα σχέδια ανάκαμψης από καταστροφές περιγράφουν διαδικασίες για την ταχεία αποκατάσταση δεδομένων και συστημάτων σε περίπτωση συμβάντος για την ελαχιστοποίηση του χρόνου διακοπής λειτουργίας και της απώλειας δεδομένων (Marx, 2013).

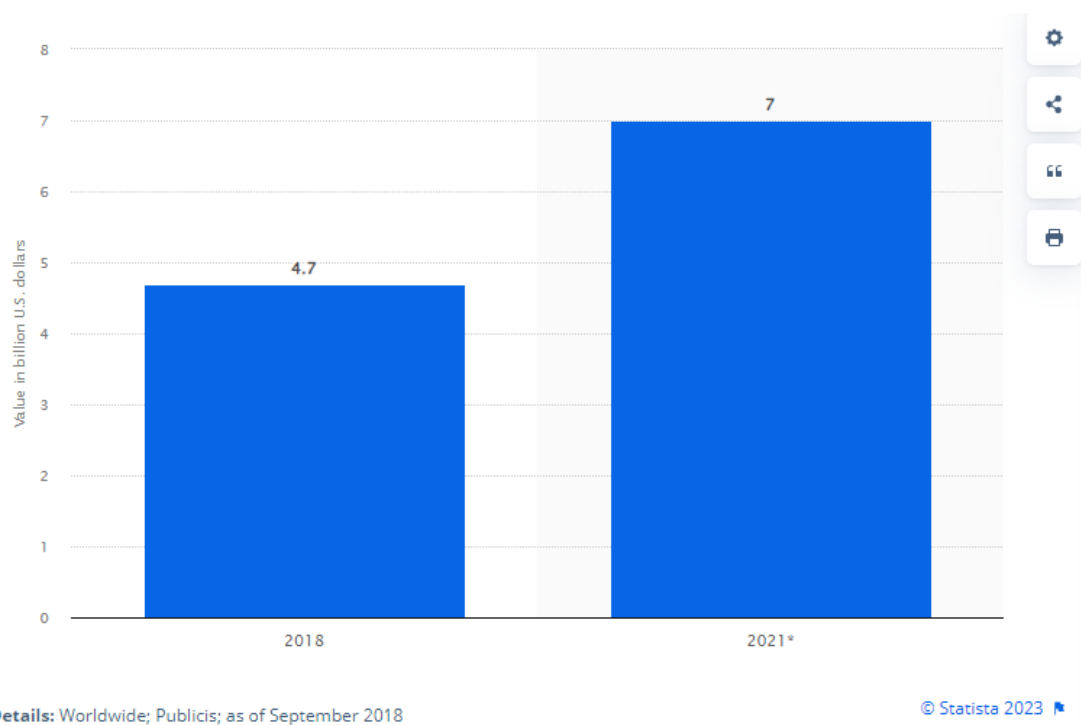
Σημαντικοί κίνδυνοι για την ασφάλεια των δεδομένων υγείας είναι το ανθρώπινο λάθος και οι εσωτερικές απειλές. Είναι ζωτικής σημασίας η παροχή ολοκληρωμένων προγραμμάτων κατάρτισης και ευαισθητοποίησης για την εκπαίδευση των μελών του προσωπικού σχετικά με τις βέλτιστες πρακτικές ασφάλειας, τις διαδικασίες

χειρισμού δεδομένων και τη σημασία της διατήρησης της εμπιστευτικότητας των δεδομένων υγείας. Οι τακτικές εκπαιδευτικές συνεδρίες μπορούν να βοηθήσουν τους υπαλλήλους να ενημερώνονται για τις αναδυόμενες απειλές, να ενισχύουν τα πρωτόκολλα ασφαλείας και να καλλιεργούν μια κουλτούρα ευαισθητοποίησης σε θέματα ασφάλειας εντός του οργανισμού (Marx, 2013).

Θα πρέπει να έχουμε στο νου μας ότι θα πρέπει να υπάρξει κανονιστική συμμόρφωση. Η συμμόρφωση με τους σχετικούς κανονισμούς για την προστασία των δεδομένων και της ιδιωτικής ζωής είναι απαραίτητη στον κλάδο της υγειονομικής περίθαλψης. Νόμοι όπως ο νόμος περί φορητότητας και λογοδοσίας για την ασφάλιση υγείας (HIPAA) στις Ηνωμένες Πολιτείες και ο γενικός κανονισμός για την προστασία δεδομένων (GDPR) στην Ευρωπαϊκή Ένωση επιβάλλουν αυστηρές απαιτήσεις για την ασφάλεια και την προστασία της ιδιωτικής ζωής των δεδομένων υγείας. Οι οργανισμοί που χειρίζονται δεδομένα υγείας πρέπει να διασφαλίζουν ότι συμμορφώνονται με τους εν λόγω κανονισμούς, να εφαρμόζουν τις απαραίτητες διασφαλίσεις και να είναι προετοιμασμένοι για ελέγχους και πιθανές κυρώσεις σε περίπτωση μη συμμόρφωσης (Starkbaum and Felt, 2019).

Συμπερασματικά, η ασφάλεια των δεδομένων υγείας στα μεγάλα δεδομένα περιλαμβάνει μια ολοκληρωμένη προσέγγιση που συνδυάζει κρυπτογράφηση, έλεγχο πρόσβασης, έλεγχο ταυτότητας, ανωνυμοποίηση, πρόληψη απώλειας δεδομένων, τακτικούς ελέγχους, ασφαλή διαβίβαση, δημιουργία αντιγράφων ασφαλείας και ανάκτηση, εκπαίδευση του προσωπικού και συμμόρφωση με τους κανονισμούς. Με την εφαρμογή αυτών των μέτρων, οι οργανισμοί μπορούν να ελαχιστοποιήσουν τους κινδύνους που σχετίζονται με την αποθήκευση, την επεξεργασία και την κοινή χρήση ευαίσθητων δεδομένων υγείας, διασφαλίζοντας την εμπιστευτικότητα, την ακεραιότητα και τη διαθεσιμότητά τους στην εποχή της ανάλυσης μεγάλων δεδομένων.

Στην εικόνα που ακολουθεί παρουσιάζονται οι επενδύσεις που έχουν γίνει στον χώρο των μεγάλων δεδομένων στον τομέα της υγείας (Capitalist and Publicis, 2023). Από το διάγραμμα είναι προφανές ότι οι επενδύσεις έχουν ανοδική τάση, όπως είναι και η τάση των δεδομένων που παράγονται.



Εικόνα 3. Πρόβλεψη επενδύσεων στον χώρο των μεγάλων δεδομένων στην υγεία (Capitalist and Publicis, 2023)

Κεφάλαιο 5^ο – Διάχυση των ιατρικών δεδομένων

5.1 Πλατφόρμες ανταλλαγής δεδομένων

Οι πλατφόρμες ανταλλαγής δεδομένων για δεδομένα υγείας διαδραματίζουν κρίσιμο ρόλο στη διευκόλυνση της ανταλλαγής πληροφοριών μεταξύ παρόχων υγειονομικής περίθαλψης, ερευνητών και άλλων ενδιαφερόμενων φορέων. Οι πλατφόρμες αυτές αποσκοπούν στην προώθηση της συνεργασίας, στη βελτίωση των αποτελεσμάτων της υγειονομικής περίθαλψης και στην προώθηση των εξελίξεων της ιατρικής έρευνας. Μέσα από αυτούς τους χώρους μπορεί η ιατρική κοινότητα να εξελιχθεί και να λάβει αποφάσεις σε επίπεδο πολιτικής, αλλά και σε επίπεδο καθημερινής άσκησης του λειτουργήματος της ιατρικής. Ακολουθούν διάφορες εξέχουσες πλατφόρμες ανταλλαγής δεδομένων που χρησιμοποιούνται για δεδομένα υγείας.

5.1.1 Τύποι

Στην παρούσα ενότητα παρουσιάζονται οι τύποι πλατφορμών που απασχολούν την ιατρική κοινότητα. Αξίζει να σημειωθεί ότι αυτές οι πλατφόρμες θέτουν ως προτεραιότητα την προστασία της ιδιωτικής ζωής και την ασφάλεια των δεδομένων, ώστε να διασφαλίζεται η συμμόρφωση με τους ισχύοντες κανονισμούς και να προστατεύονται οι ευαίσθητες πληροφορίες υγείας. Συνήθως εφαρμόζονται πλαίσια διακυβέρνησης δεδομένων, διαχείριση συναίνεσης, τεχνικές ανωνυμοποίησης και αυστηροί έλεγχοι πρόσβασης για τη διασφάλιση της εμπιστευτικότητας των δεδομένων υγείας, επιτρέποντας παράλληλα την υπεύθυνη κοινή χρήση και αξιοποίησή τους για την έρευνα, την κλινική περίθαλψη και τις πρωτοβουλίες δημόσιας υγείας. Πιο συγκεκριμένα:

- Ανταλλαγές πληροφοριών υγείας (Health Information Exchanges - HIEs): Είναι περιφερειακές ή κρατικές πλατφόρμες που επιτρέπουν την ασφαλή ανταλλαγή πληροφοριών υγείας μεταξύ οργανισμών υγειονομικής περίθαλψης, συμπεριλαμβανομένων νοσοκομείων, κλινικών, εργαστηρίων

και φαρμακείων (Rudin *et al.*, 2014). Παρέχουν ένα τυποποιημένο πλαίσιο για την ανταλλαγή δεδομένων ασθενών, όπως ιατρικοί φάκελοι, εργαστηριακά αποτελέσματα και ιστορικό φαρμάκων, μεταξύ διαφορετικών φορέων υγειονομικής περίθαλψης εντός μιας συγκεκριμένης γεωγραφικής περιοχής. Οι HIE ενισχύουν το συντονισμό της περίθαλψης, μειώνουν τις διπλές εξετάσεις και διαδικασίες και υποστηρίζουν την ταχύτερη και πιο τεκμηριωμένη λήψη κλινικών αποφάσεων.

- Δίκτυα ερευνητικών δεδομένων: Τα δίκτυα ερευνητικών δεδομένων είναι πλατφόρμες που διευκολύνουν την ανταλλαγή δεδομένων υγείας για ερευνητικούς σκοπούς. Τα δίκτυα αυτά φέρνουν σε επαφή διάφορα ιδρύματα, όπως πανεπιστήμια, ερευνητικούς οργανισμούς και παρόχους υγειονομικής περίθαλψης, για τη συγκέντρωση ανώνυμων ή αποσταυρωμένων δεδομένων υγείας από διάφορες πηγές (Maro *et al.*, 2009). Επιτρέπουν στους ερευνητές να έχουν πρόσβαση σε σύνολα δεδομένων μεγάλης κλίμακας, να διεξάγουν μελέτες και να αποκτούν γνώσεις σχετικά με την υγεία του πληθυσμού, τα πρότυπα ασθενειών, την αποτελεσματικότητα της θεραπείας και τις τάσεις της υγειονομικής περίθαλψης. Παραδείγματα δικτύων ερευνητικών δεδομένων περιλαμβάνουν το PCORnet (Patient-Centered Outcomes Research Network) και το OMOP (Observational Medical Outcomes Partnership).
- Πλατφόρμες ανταλλαγής δεδομένων για συγκεκριμένες ασθένειες: Ορισμένες πλατφόρμες ανταλλαγής δεδομένων επικεντρώνονται σε συγκεκριμένες ασθένειες ή ιατρικές καταστάσεις. Αυτές οι πλατφόρμες έχουν σχεδιαστεί για τη συλλογή και την ανταλλαγή δεδομένων που σχετίζονται με μια συγκεκριμένη ασθένεια, επιτρέποντας στους ερευνητές, τους κλινικούς ιατρούς και τους ασθενείς να συνεργάζονται για την κατανόηση της ασθένειας, την ανάπτυξη νέων θεραπειών και τη βελτίωση των αποτελεσμάτων των ασθενών (Kong *et al.*, 2022). Παραδείγματα περιλαμβάνουν το Global Alzheimer's Association Interactive Network (GAAIN) για την έρευνα της νόσου του Alzheimer και το Multiple Sclerosis Data Alliance (MSDA) για την έρευνα της σκλήρυνσης κατά πλάκας.

- Πλατφόρμες ανταλλαγής γονιδιωματικών δεδομένων: Οι πλατφόρμες ανταλλαγής γονιδιωματικών δεδομένων επικεντρώνονται στην ανταλλαγή και την ανάλυση γονιδιωματικών πληροφοριών, όπως αλληλουχίες DNA και γενετικές παραλλαγές, για την προώθηση της εξατομικευμένης ιατρικής και της γονιδιωματικής έρευνας (Schatz *et al.*, 2022). Αυτές οι πλατφόρμες επιτρέπουν στους ερευνητές να έχουν πρόσβαση σε μεγάλα σύνολα γονιδιωματικών δεδομένων, να συνεργάζονται σε γονιδιωματικές μελέτες και να επιταχύνουν τις ανακαλύψεις που σχετίζονται με γενετικές καταστάσεις, αντιδράσεις σε φάρμακα και προδιάθεση για ασθένειες. Η Παγκόσμια Συμμαχία για τη Γονιδιωματική και την Υγεία (GA4GH) και η Βάση Δεδομένων Γονότυπων και Φαινοτύπων (dbGaP) των Εθνικών Ινστιτούτων Υγείας αποτελούν παραδείγματα πλατφορμών ανταλλαγής γονιδιωματικών δεδομένων.
- Πλατφόρμες ανοικτών δεδομένων: Οι πλατφόρμες ανοικτών δεδομένων παρέχουν δημόσια πρόσβαση σε σύνολα δεδομένων που σχετίζονται με την υγεία, προωθώντας τη διαφάνεια, την καινοτομία και τη συνεργασία. Αυτές οι πλατφόρμες καθιστούν ανώνυμα ή απο-ταυτοποιημένα δεδομένα υγείας διαθέσιμα σε ερευνητές, προγραμματιστές και το ευρύ κοινό, επιτρέποντάς τους να εξερευνήσουν και να αναλύσουν τα δεδομένα για διάφορους σκοπούς, συμπεριλαμβανομένης της έρευνας, της ανάπτυξης εφαρμογών και της χάραξης πολιτικής (Navar *et al.*, 2016). Τα ανοικτά δεδομένα των Κέντρων Ελέγχου και Πρόληψης Νοσημάτων (CDC) και το Παγκόσμιο Παρατηρητήριο Υγείας του Παγκόσμιου Οργανισμού Υγείας (ΠΟΥ) αποτελούν παραδείγματα πλατφορμών ανοικτών δεδομένων.
- Δίκτυα συνεργατικής έρευνας: Τα συνεργατικά ερευνητικά δίκτυα φέρνουν σε επαφή πολλαπλά ενδιαφερόμενα μέρη, όπως ερευνητές, κλινικούς ιατρούς, ασθενείς και βιομηχανικούς εταίρους, για να συνεργαστούν σε ερευνητικά έργα και να μοιραστούν δεδομένα υγείας (Robinson *et al.*, 2015). Τα δίκτυα αυτά δημιουργούν ένα συνεργατικό περιβάλλον όπου οι συμμετέχοντες μπορούν να συνεισφέρουν την εμπειρογνωμοσύνη τους, να μοιράζονται δεδομένα και να συνεργάζονται για την αντιμετώπιση σύνθετων προκλήσεων στον τομέα της υγειονομικής περίθαλψης.

5.1.2 Παραδείγματα

Τα τελευταία χρόνια έχουν εμφανιστεί διάφορες δημοφιλείς πλατφόρμες κοινής χρήσης δεδομένων για δεδομένα υγείας, λόγω της αυξανόμενης ανάγκης για συνεργασία, έρευνα και καινοτομία στον κλάδο της υγειονομικής περίθαλψης. Αν και η δημοτικότητα μπορεί να ποικίλλει ανάλογα με τη γεωγραφική τοποθεσία και τους συγκεκριμένους τομείς ενδιαφέροντος, παραθέτουμε ορισμένες ευρέως αναγνωρισμένες και ευρέως χρησιμοποιούμενες πλατφόρμες κοινής χρήσης δεδομένων για δεδομένα υγείας. Πιο συγκεκριμένα:

1. Το Εθνικό Κέντρο Βιοτεχνολογικών Πληροφοριών (NCBI): NCBI φιλοξενεί μια σειρά βιοϊατρικών και γονιδιωματικών βάσεων δεδομένων, συμπεριλαμβανομένων των PubMed, GenBank, dbGaP και ClinVar. Η PubMed είναι μια ολοκληρωμένη βάση δεδομένων επιστημονικών δημοσιεύσεων στον τομέα της βιοϊατρικής, ενώ η GenBank αποθηκεύει δεδομένα γενετικών ακολουθιών (National Center for Biotechnology Information, 2023). Η dbGaP είναι ένα αποθετήριο δεδομένων γονότυπων και φαινοτύπων που σχετίζονται με την ανθρώπινη γενετική και η ClinVar παρέχει πληροφορίες σχετικά με τις γενετικές παραλλαγές και την κλινική τους σημασία.
2. The Cancer Genome Atlas (TCGA): Το TCGA είναι ένα ολοκληρωμένο πρόγραμμα γονιδιωματικής του καρκίνου που συλλέγει και αναλύει γονιδιωματικά δεδομένα από διάφορους τύπους καρκίνου (Cancer Genome Atlas, 2023). Παρέχει μια πλούσια πηγή για τους ερευνητές που μελετούν τη γενετική βάση του καρκίνου και διερευνούν πιθανές στρατηγικές θεραπείας.
3. Το Παγκόσμιο Παρατηρητήριο Υγείας (GHO) του Παγκόσμιου Οργανισμού Υγείας (ΠΟΥ): Το GHO είναι ένα εκτεταμένο αποθετήριο δεδομένων που συντηρείται από τον ΠΟΥ και περιέχει στατιστικά στοιχεία και δείκτες που σχετίζονται με την υγεία από χώρες σε όλο τον κόσμο (WHO, 2023). Παρέχει δεδομένα για διάφορα θέματα υγείας, όπως η θνησιμότητα, η επιβάρυνση από ασθένειες, οι υποδομές υγειονομικής περίθαλψης και οι επιδόσεις των συστημάτων υγείας.
4. Τα Κέντρα Ελέγχου και Πρόληψης Νοσημάτων (CDC) Wonder: Το CDC Wonder είναι μια διαδικτυακή πύλη που παρέχει πρόσβαση σε ένα ευρύ φάσμα δεδομένων και πληροφοριών για τη δημόσια υγεία (CDC, 2023). Προσφέρει

πρόσβαση σε σύνολα δεδομένων, εκθέσεις και άλλους πόρους που σχετίζονται με μολυσματικές ασθένειες, χρόνιες ασθένειες, περιβαλλοντική υγεία και υγεία του πληθυσμού.

5. Το ClinicalTrials.gov: Το ClinicalTrials.gov είναι ένα μητρώο και μια βάση δεδομένων αποτελεσμάτων κλινικών δοκιμών που διεξάγονται παγκοσμίως (ClinicalTrials.gov, 2023). Παρέχει πληροφορίες σχετικά με τρέχουσες και ολοκληρωμένες κλινικές δοκιμές, συμπεριλαμβανομένου του σχεδιασμού, των στόχων, των αποτελεσμάτων και των κριτηρίων επιλεξιμότητας των συμμετεχόντων. Αυτό το αποθετήριο χρησιμεύει ως πολύτιμη πηγή για ερευνητές, κλινικούς ιατρούς και ασθενείς που αναζητούν πληροφορίες σχετικά με κλινικές δοκιμές.
6. Το OpenFEMA: Το OpenFEMA είναι ένα αποθετήριο δεδομένων που συντηρείται από την Ομοσπονδιακή Υπηρεσία Διαχείρισης Εκτάκτων Αναγκών (FEMA) στις Ηνωμένες Πολιτείες. Παρέχει πρόσβαση σε ένα ευρύ φάσμα δεδομένων που σχετίζονται με καταστροφές, συμπεριλαμβανομένων πληροφοριών σχετικά με δηλώσεις καταστροφών, προγράμματα βοήθειας και προσπάθειες μετριασμού των κινδύνων (OpenFEMA, 2023).

5.2 Πρότυπα δεδομένων

Στον κλάδο της υγειονομικής περίθαλψης, τα πρότυπα δεδομένων και η διαλειτουργικότητα διαδραματίζουν κρίσιμο ρόλο στη διευκόλυνση της απρόσκοπτης ανταλλαγής και ενσωμάτωσης δεδομένων υγείας σε διαφορετικά συστήματα, πλατφόρμες και οργανισμούς (Khan *et al.*, 2014). Αυτά τα πρότυπα και οι πρακτικές διασφαλίζουν ότι οι πληροφορίες υγείας μπορούν να διαμοιράζονται, να κατανοούνται και να χρησιμοποιούνται αποτελεσματικά για τη βελτίωση της περίθαλψης των ασθενών, τη βελτίωση των αποτελεσμάτων της υγειονομικής περίθαλψης και την προώθηση της έρευνας και της καινοτομίας. Τα πρότυπα δεδομένων αναφέρονται σε ένα σύνολο κανόνων, μορφοτύπων και ορισμών που επιτρέπουν τη συνεπή και δομημένη αναπαράσταση των πληροφοριών. Στο πλαίσιο των δεδομένων υγείας, τα πρότυπα έχουν σχεδιαστεί για να διασφαλίζουν ότι τα στοιχεία δεδομένων, όπως τα δημογραφικά στοιχεία των ασθενών, οι κλινικές παρατηρήσεις, οι διαγνώσεις, οι διαδικασίες, τα φάρμακα και άλλα, καταγράφονται,

αποθηκεύονται και διαβιβάζονται με τυποποιημένο τρόπο. Καθιερώνουν κοινά λεξιλόγια, μοντέλα δεδομένων και τεχνικές προδιαγραφές, επιτρέποντας σε διαφορετικά συστήματα υγειονομικής περίθαλψης να μιλούν την ίδια γλώσσα και να ερμηνεύουν τα δεδομένα με συνέπεια.

Η υιοθέτηση προτύπων δεδομένων και η επίτευξη διαλειτουργικότητας στα δεδομένα υγείας προσφέρουν πολυάριθμα οφέλη στους ασθενείς, τους παρόχους υγειονομικής περίθαλψης, τους ερευνητές και το οικοσύστημα υγειονομικής περίθαλψης στο σύνολό του (Jaleel *et al.*, 2020). Για παράδειγμα, τα πρότυπα δεδομένων και η διαλειτουργικότητα επιτρέπουν μια ολιστική εικόνα των αρχείων υγείας των ασθενών, παρέχοντας στους επαγγελματίες του τομέα της υγειονομικής περίθαλψης ολοκληρωμένες και επικαιροποιημένες πληροφορίες. Αυτό διευκολύνει τη λήψη καλύτερων κλινικών αποφάσεων, τον καλύτερο συντονισμό της περίθαλψης, τη μείωση των ιατρικών σφαλμάτων και την ενίσχυση της ασφάλειας των ασθενών. Ταυτόχρονα επιτρέπει την απρόσκοπτη ροή των δεδομένων υγείας μεταξύ διαφορετικών συστημάτων και οργανισμών υγειονομικής περίθαλψης. Αυτό διευκολύνει την ανταλλαγή πληροφοριών μεταξύ νοσοκομείων, κλινικών, εργαστηρίων και άλλων φορέων υγειονομικής περίθαλψης, επιτρέποντας την ομαλότερη μετάβαση στη φροντίδα και την αποφυγή της περιττής καταχώρησης δεδομένων.

Τα τυποποιημένα δεδομένα υγείας μπορούν να συγκεντρωθούν και να αναλυθούν για ερευνητικούς σκοπούς, επιτρέποντας τη διαχείριση της υγείας του πληθυσμού, τις επιδημιολογικές μελέτες και την ανακάλυψη νέων θεραπειών και παρεμβάσεων (Pezoulas *et al.*, 2019). Η διαλειτουργικότητα διευκολύνει την ενσωμάτωση δεδομένων από πολλαπλές πηγές, υποστηρίζοντας μελέτες μεγάλης κλίμακας και τη δημιουργία στοιχείων από τον πραγματικό κόσμο. Επίσης αξίζει να αναφερθεί ότι προωθούν την ενσωμάτωση των συστημάτων υγειονομικής περίθαλψης με αναδυόμενες τεχνολογίες, όπως η τηλεϊατρική, οι φορητές συσκευές και οι εφαρμογές υγείας. Αυτό επιτρέπει την απρόσκοπτη ενσωμάτωση δεδομένων από διάφορες πηγές, προωθώντας την καινοτομία και διευρύνοντας τις δυνατότητες για εξατομικευμένη και ασθενοκεντρική φροντίδα .

Τέλος, σε μία κοινωνία στην οποία η έννοια του κόστους λαμβάνει σημαντικές διαστάσεις θα πρέπει να αναφέρουμε ότι η διαλειτουργικότητα μειώνει την ανάγκη για χειροκίνητη καταχώρηση δεδομένων, διπλές εξετάσεις και περιττές διαδικασίες, οδηγώντας σε εξοικονόμηση κόστους για τους οργανισμούς υγειονομικής περίθαλψης. Εξορθολογίζει τις διοικητικές διαδικασίες, επιτρέπει την αποτελεσματική διαχείριση της τιμολόγησης και των απαιτήσεων και υποστηρίζει πρωτοβουλίες φροντίδας βάσει αξίας.

5.2.1 To Health Level Seven (HL7)

To Health Level Seven (HL7) είναι ένα σύνολο διεθνών προτύπων για την ανταλλαγή, ολοκλήρωση, διαμοιρασμό και ανάκτηση ηλεκτρονικών πληροφοριών υγείας. Παρέχει ένα πλαίσιο και προδιαγραφές για τη διαλειτουργική ανταλλαγή δεδομένων υγειονομικής περίθαλψης μεταξύ διαφορετικών συστημάτων και οργανισμών. Τα πρότυπα HL7 καθορίζουν τη δομή, το περιεχόμενο και τη μορφή των μηνυμάτων που διευκολύνουν την επικοινωνία και τη μεταφορά δεδομένων υγείας (Hettinger and Brazile, 1994). Η πιο διαδεδομένη έκδοση του HL7 είναι η έκδοση 2.x του HL7, η οποία χρησιμοποιεί μια μορφή μηνυμάτων που βασίζεται σε τμήματα, πεδία και συστατικά. Τα μηνύματα αυτά μπορούν να περιέχουν διάφορους τύπους πληροφοριών, συμπεριλαμβανομένων δημογραφικών στοιχείων του ασθενούς, κλινικών παρατηρήσεων, διαγνώσεων, διαδικασιών, φαρμάκων και διοικητικών δεδομένων.

To HL7 περιλαμβάνει επίσης και άλλα πρότυπα πέραν της ανταλλαγής μηνυμάτων, όπως το HL7 Clinical Document Architecture (CDA) και το Fast Healthcare Interoperability Resources (FHIR). Το CDA είναι ένα πρότυπο βασισμένο σε XML για την ανταλλαγή κλινικών εγγράφων, όπως περιλήψεις εξιτηρίου, σημειώσεις προόδου και διαγνωστικές εκθέσεις. Καθορίζει μια δομημένη μορφή για την καταγραφή και την ανταλλαγή κλινικών πληροφοριών με τρόπο αναγνώσιμο από τον άνθρωπο και επεξεργάσιμο από μηχανήματα (Hettinger and Brazile, 1994). Από την άλλη πλευρά, το FHIR είναι ένα σύγχρονο και διαδικτυακό πρότυπο για την ανταλλαγή δεδομένων υγειονομικής περίθαλψης. Χρησιμοποιεί μια προσέγγιση RESTful API και αξιοποιεί κοινές τεχνολογίες ιστού όπως HTTP, XML, JSON και OAuth για την ανταλλαγή δεδομένων. Το FHIR επιτρέπει τη λεπτομερή ανταλλαγή διακριτών στοιχείων

δεδομένων που ονομάζονται πόροι, τα οποία μπορούν να αντιπροσωπεύουν ένα ευρύ φάσμα κλινικών και διοικητικών εννοιών (Hettinger and Brazile, 1994).

Τα πρότυπα HL7 προωθούν τη διαλειτουργικότητα παρέχοντας ένα κοινό πλαίσιο και λεξιλόγιο για την επικοινωνία και την ανταλλαγή δεδομένων μεταξύ των συστημάτων υγειονομικής περίθαλψης. Παρέχουν τη δυνατότητα στους οργανισμούς υγειονομικής περίθαλψης να μοιράζονται πληροφορίες απρόσκοπτα, να ενσωματώνουν διαφορετικά συστήματα και να υποστηρίζουν τη συνεργατική φροντίδα σε διαφορετικές συνθήκες περίθαλψης (Jorgia *et al.*, 2018).

Τα πρότυπα HL7 διευκολύνουν την αποτελεσματική και ακριβή ανταλλαγή δεδομένων υγείας μεταξύ συστημάτων, μειώνοντας τη χειροκίνητη εισαγωγή δεδομένων, ελαχιστοποιώντας τα σφάλματα και ενισχύοντας την ακεραιότητα των δεδομένων. Επιτρέποντας την ανταλλαγή ολοκληρωμένων και επικαιροποιημένων πληροφοριών για τον ασθενή, το HL7 υποστηρίζει τον καλύτερο συντονισμό της περίθαλψης μεταξύ των παρόχων υγειονομικής περίθαλψης. Αυτό οδηγεί σε βελτιωμένη συνέχεια της περίθαλψης, μειωμένη επανάληψη των εξετάσεων και καλύτερα τεκμηριωμένη λήψη κλινικών αποφάσεων. Ταυτόχρονα τα πρότυπα HL7 επιτρέπουν στους οργανισμούς υγειονομικής περίθαλψης να ενσωματώνουν συστήματα συμβατά με το HL7 με την υπάρχουσα υποδομή τους, όπως συστήματα ηλεκτρονικού φακέλου υγείας (EHR), εργαστηριακά συστήματα και συστήματα χρέωσης. Τέλος τα συμβατά με το HL7 δεδομένα μπορούν να συγκεντρωθούν και να αναλυθούν για ερευνητικούς σκοπούς, για τη διαχείριση της υγείας του πληθυσμού και για πληροφορίες βάσει δεδομένων. Οι ερευνητές μπορούν να έχουν πρόσβαση σε τυποποιημένα δεδομένα από πολλαπλές πηγές, διευκολύνοντας μελέτες μεγάλης κλίμακας και αναλύσεις δεδομένων (Jorgia *et al.*, 2018).

5.2.2 Η International Classification of Diseases (ICD)

Η Διεθνής Ταξινόμηση Ασθενειών (ICD) είναι ένα τυποποιημένο σύστημα ταξινόμησης και κωδικοποίησης ασθενειών, τραυματισμών και αιτιών θανάτου. Παρέχει μια καθολική γλώσσα για την περιγραφή και την κατηγοριοποίηση των καταστάσεων υγείας, η οποία διευκολύνει τη σύγκριση, την ανάλυση και την ανταλλαγή πληροφοριών για την υγεία σε παγκόσμια κλίμακα (Hong and Zeng, 2023). Ο Παγκόσμιος Οργανισμός Υγείας (ΠΟΥ) είναι υπεύθυνος για τη διατήρηση και την

ενημέρωση του ICD. Η τρέχουσα έκδοση είναι το ICD-10, το οποίο εισήχθη το 1994. Ωστόσο, μια επικαιροποιημένη έκδοση, το ICD-11, κυκλοφόρησε από τον ΠΟΥ το 2018 και υιοθετείται σταδιακά από τις χώρες και τα συστήματα υγειονομικής περίθαλψης. Τα βασικά χαρακτηριστικά και στοιχεία του ICD περιλαμβάνουν (Hong and Zeng, 2023):

- Σύστημα κωδικοποίησης: Το ICD χρησιμοποιεί αλφαριθμητικούς κωδικούς για την αναπαράσταση διαφόρων ασθενειών, τραυματισμών και καταστάσεων υγείας. Οι κωδικοί αυτοί παρέχουν έναν τυποποιημένο τρόπο για τον εντοπισμό και την ταξινόμηση συγκεκριμένων διαγνώσεων και αιτιών θανάτου. Κάθε κωδικός αποτελείται από έναν συνδυασμό γραμμάτων και αριθμών, επιτρέποντας την ακριβή ταξινόμηση και διαφοροποίηση.
- Ιεραρχική δομή: Το ICD διαθέτει ιεραρχική δομή που οργανώνει τις ασθένειες και τις καταστάσεις σε κεφάλαια, τμήματα και κατηγορίες. Ακολουθεί ένα πολυαξονικό πλαίσιο, με τους κωδικούς να διατάσσονται με βάση την ανατομική θέση, την αιτιολογία (αιτία) και την κλινική σοβαρότητα. Αυτή η δομή επιτρέπει τη συστηματική ομαδοποίηση και ταξινόμηση των συναφών καταστάσεων υγείας.
- Κλινικές τροποποιήσεις: Η ICD ενημερώνεται περιοδικά για να προσαρμόζεται στις εξελίξεις των ιατρικών γνώσεων και στις αλλαγές στις πρακτικές υγειονομικής περίθαλψης. Διαφορετικές χώρες και περιοχές μπορούν να αναπτύξουν τις δικές τους κλινικές τροποποιήσεις του ICD για να καλύψουν συγκεκριμένες ανάγκες κωδικοποίησης και περιφερειακές διαφορές στα πρότυπα ασθενειών.
- Κωδικοποίηση αιτίας θανάτου: Το ICD περιλαμβάνει ένα ειδικό τμήμα για την κωδικοποίηση των αιτιών θανάτου, γνωστό ως ICD Mortality Coding. Παρέχει μια τυποποιημένη προσέγγιση για τον εντοπισμό και την καταγραφή της υποκείμενης αιτίας θανάτου, των παραγόντων που συμβάλλουν και των συναφών καταστάσεων. Οι πληροφορίες αυτές είναι απαραίτητες για τις ζωτικές στατιστικές, την επιτήρηση της δημόσιας υγείας και την έρευνα σχετικά με τις τάσεις της θνησιμότητας.

5.2.3 Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT)

Η Συστηματοποιημένη Ονοματολογία της Ιατρικής - Κλινικοί Όροι (SNOMED CT) είναι ένα ολοκληρωμένο σύστημα κλινικής ορολογίας και κωδικοποίησης που χρησιμοποιείται ευρέως στον κλάδο της υγειονομικής περίθαλψης. Παρέχει ένα τυποποιημένο και διεθνώς αναγνωρισμένο λεξιλόγιο για την περιγραφή των κλινικών εννοιών, της ιατρικής ορολογίας και της τεκμηρίωσης της υγειονομικής περίθαλψης. Το SNOMED CT βασίζεται σε ένα μοντέλο προσανατολισμένο στις έννοιες, όπου σε κάθε κλινική έννοια αποδίδεται ένα μοναδικό αναγνωριστικό. Οι έννοιες στο SNOMED CT αντιπροσωπεύουν κλινικές ιδέες, όπως ασθένειες, συμπτώματα, διαδικασίες, ανατομικές δομές και φάρμακα. Το μοντέλο αυτό επιτρέπει την ακριβή αναπαράσταση και ταξινόμηση των κλινικών πληροφοριών (Peterson, 2020).

Την ίδια στιγμή, οργανώνει τις έννοιες σε μια ιεραρχική δομή, με ευρύτερες και στενότερες σχέσεις μεταξύ τους. Αυτή η ιεραρχία επιτρέπει την ταξινόμηση των εννοιών με βάση τις σχέσεις τους και παρέχει έναν δομημένο τρόπο πλοήγησης και ανάκτησης συγκεκριμένων κλινικών όρων. Επιτρέπει επίσης την καταγραφή λεπτομερών κλινικών πληροφοριών, όπως η θέση, η σοβαρότητα και άλλα χαρακτηριστικά. Το SNOMED CT καλύπτει εξίσου, ένα ευρύ φάσμα κλινικών τομέων, συμπεριλαμβανομένης της γενικής ιατρικής, των ειδικοτήτων, των διαδικασιών, των εργαστηριακών ευρημάτων, της ακτινολογίας, των φαρμάκων και της νοσηλείας. Παρέχει ένα πλούσιο λεξιλόγιο που μπορεί να αποτυπώσει την πολυπλοκότητα και την ποικιλομορφία των κλινικών πληροφοριών σε διάφορους κλάδους της υγειονομικής περίθαλψης (Peterson, 2020).

Τέλος, αξίζει να αναφερθεί ότι έχει σχεδιαστεί για να υποστηρίζει πολλές γλώσσες, επιτρέποντας την τυποποιημένη αναπαράσταση κλινικών εννοιών σε διάφορα γλωσσικά πλαίσια. Αυτό το χαρακτηριστικό ενισχύει τη διεθνή συνεργασία και επιτρέπει τη χρήση του SNOMED CT σε διαφορετικές χώρες και περιοχές. Το SNOMED CT ενημερώνεται και συντηρείται συνεχώς από τον Διεθνή Οργανισμό Ανάπτυξης Προτύπων Ορολογίας Υγείας (IHTSDO). Οι τακτικές ενημερώσεις διασφαλίζουν ότι το SNOMED CT παραμένει επίκαιρο, αντικατοπτρίζοντας τις εξελίξεις στην ιατρική γνώση, τους νέους κλινικούς όρους και τις εξελισσόμενες πρακτικές υγειονομικής περίθαλψης.

5.2.4 Logical Observation Identifiers Names and Codes (LOINC)

Τα ονόματα και οι κωδικοί αναγνώρισης λογικών παρατηρήσεων (LOINC) είναι ένα ευρέως χρησιμοποιούμενο τυποποιημένο σύστημα ορολογίας για τον προσδιορισμό και την αναφορά ιατρικών εργαστηριακών παρατηρήσεων και κλινικών μετρήσεων. Παρέχει ένα καθολικό σύνολο κωδικών και ονομάτων για εργαστηριακές εξετάσεις, κλινικές μετρήσεις και άλλες παρατηρήσεις, διευκολύνοντας την ανταλλαγή και την κοινή χρήση εργαστηριακών δεδομένων μεταξύ διαφορετικών συστημάτων και οργανισμών υγειονομικής περίθαλψης (Yeh *et al.*, 2021). Το LOINC αποσκοπεί στην τυποποίηση των ονομάτων και των κωδικών που χρησιμοποιούνται για τον προσδιορισμό εργαστηριακών δοκιμών και παρατηρήσεων. Αποδίδει ένα μοναδικό αναγνωριστικό σε κάθε δοκιμή, επιτρέποντας την ακριβή ταυτοποίηση και ερμηνεία των εργαστηριακών αποτελεσμάτων. Οι κωδικοί LOINC καλύπτουν ένα ευρύ φάσμα εργαστηριακών τομέων, συμπεριλαμβανομένης της χημείας, της αιματολογίας, της μικροβιολογίας, της μοριακής διάγνωσης και της παθολογίας. Το LOINC ακολουθεί μια ιεραρχική δομή, οργανώνοντας τις εργαστηριακές εξετάσεις σε κατηγορίες και υποκατηγορίες. Αυτή η δομή παρέχει μια λογική ομαδοποίηση των σχετικών δοκιμών, επιτρέποντας την αποτελεσματική πλοήγηση και ανάκτηση συγκεκριμένων εργαστηριακών παρατηρήσεων. Υποστηρίζει τη συνεπή οργάνωση και παρουσίαση των εργαστηριακών δεδομένων (Yeh *et al.*, 2021).

Επίσης θα πρέπει να αναφερθεί ότι υποστηρίζει πολλές γλώσσες, επιτρέποντας την αναπαράσταση εργαστηριακών δοκιμών και παρατηρήσεων σε διαφορετικά γλωσσικά πλαίσια. Αυτό το χαρακτηριστικό ενισχύει τη διεθνή συνεργασία και επιτρέπει τη χρήση του LOINC σε διαφορετικά περιβάλλοντα υγειονομικής περίθαλψης. Ταυτόχρονα περιλαμβάνει αντιστοιχίσεις και διασταυρούμενες παραπομπές σε άλλα συστήματα ορολογίας, όπως το SNOMED CT και το ICD-10, διευκολύνοντας τη διαλειτουργικότητα και την ολοκλήρωση των δεδομένων. Αυτές οι αντιστοιχίσεις επιτρέπουν τη μετάφραση των εργαστηριακών παρατηρήσεων σε ευρύτερα κλινικά πλαίσια και υποστηρίζουν την ανταλλαγή δεδομένων σε διάφορους τομείς της υγειονομικής περίθαλψης (Yeh *et al.*, 2021). Τέλος το LOINC ενημερώνεται και συντηρείται τακτικά από το Ινστιτούτο Regenstrief, διασφαλίζοντας ότι η ορολογία παραμένει επίκαιρη και ευθυγραμμίζεται με τις εξελισσόμενες εργαστηριακές πρακτικές και τις αναδυόμενες τεχνολογίες. Οι

ενημερώσεις περιλαμβάνουν την προσθήκη νέων εργαστηριακών δοκιμών, αναθεωρήσεις των υφιστάμενων κωδικών και βελτιώσεις στην ιεραρχία του LOINC.

5.3 Η ανάγκη εκπαίδευσης στον χώρο των ιατρικών δεδομένων

Από τα παραπάνω κρίνεται δόκιμο να γίνει αναφορά στη σημασία της εκπαίδευσης όλων των εμπλεκομένων στο χώρο της υγείας, για τις διαδικασίες παραγωγής, επεξεργασίας και διάχυσης των δεδομένων προκειμένου να μπορούν να εντοπίσουν και να τα επεξεργαστούν. Η εκπαίδευση και η κατάρτιση είναι ουσιαστικής σημασίας για την προώθηση του αλφαριθμητισμού στα δεδομένα υγείας και τη διασφάλιση ότι τα άτομα διαθέτουν τις απαραίτητες γνώσεις και δεξιότητες για την αποτελεσματική κατανόηση, ερμηνεία και αξιοποίηση των δεδομένων υγείας (Bleakley, 2005).

Τα δεδομένα υγείας μπορεί να είναι πολύπλοκα, περιλαμβάνοντας διάφορους τύπους πληροφοριών, μορφότυπους και ορολογία. Η εκπαίδευση και η κατάρτιση βοηθούν τα άτομα να κατανοήσουν τις θεμελιώδεις έννοιες των δεδομένων, όπως οι πηγές δεδομένων, οι μέθοδοι συλλογής δεδομένων, η ποιότητα των δεδομένων και οι τεχνικές ανάλυσης δεδομένων. Η κατανόηση αυτή επιτρέπει στα άτομα να αξιολογούν και να ερμηνεύουν κριτικά τα δεδομένα υγείας, εξασφαλίζοντας την ακριβή κατανόηση των συνεπειών και των περιορισμών τους. Ο αλφαριθμητισμός στα δεδομένα υγείας δίνει τη δυνατότητα στα άτομα να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με την υγεία τους και την υγειονομική περίθαλψη. Μέσω της εκπαίδευσης και της κατάρτισης, τα άτομα μπορούν να μάθουν να έχουν πρόσβαση και να αναλύουν δεδομένα υγείας που σχετίζονται με τις παθήσεις τους, τις θεραπευτικές επιλογές και τα αποτελέσματά τους. Μπορούν να χρησιμοποιήσουν αυτές τις γνώσεις για να συμμετάσχουν ενεργά στη λήψη κοινών αποφάσεων με τους παρόχους υγειονομικής περίθαλψης, οδηγώντας τελικά σε καλύτερα αποτελέσματα για την υγεία και την ικανοποίηση των ασθενών.

Ταυτόχρονα, η εκπαίδευση και η κατάρτιση στον αλφαριθμητισμό στα δεδομένα υγείας εφοδιάζουν τους ερευνητές, τους επαγγελματίες υγείας και τους υπεύθυνους χάραξης πολιτικής με τις δεξιότητες που απαιτούνται για την πρόσβαση, την αξιολόγηση και τη χρήση ερευνητικών ευρημάτων που βασίζονται σε δεδομένα υγείας (Kuhn *et al.*, 2018). Μπορούν να αξιολογούν κριτικά τις μελέτες, να κατανοούν τις στατιστικές έννοιες και να εφαρμόζουν πρακτικές βασισμένες σε αποδείξεις σε

κλινικά περιβάλλοντα ή στην ανάπτυξη πολιτικής. Αυτό βελτιώνει την ποιότητα και την αποτελεσματικότητα της έρευνας και της λήψης αποφάσεων στον τομέα της υγειονομικής περίθαλψης. Παράλληλα, επιτρέπει στους ερευνητές να αξιοποιούν τον τεράστιο όγκο δεδομένων υγείας που είναι διαθέσιμα για ερευνητικούς σκοπούς. Με την κατανόηση των μεθοδολογιών συλλογής δεδομένων, των τεχνικών ανάλυσης δεδομένων και των δεοντολογικών εκτιμήσεων, οι ερευνητές μπορούν να σχεδιάζουν αυστηρές μελέτες, να εξάγουν σημαντικές γνώσεις από τα δεδομένα και να αναπτύσσουν καινοτόμες λύσεις στις προκλήσεις της υγειονομικής περίθαλψης. Η εκπαίδευση και η κατάρτιση προάγουν μια νοοτροπία με γνώμονα τα δεδομένα, προωθώντας τη δημιουργικότητα και την καινοτομία στις ερευνητικές προσπάθειες.

Η εκπαίδευση και η κατάρτιση σχετικά με τον αλφαριθμητισμό στα δεδομένα υγείας περιλαμβάνουν επίσης την κατανόηση της σημασίας της ιδιωτικής ζωής και της ασφάλειας (Kuhn *et al.*, 2018). Τα άτομα πρέπει να γνωρίζουν τους πιθανούς κινδύνους και τα τρωτά σημεία που σχετίζονται με τα δεδομένα υγείας, καθώς και τα μέτρα που εφαρμόζονται για την προστασία των προσωπικών τους πληροφοριών. Η εκπαίδευση βοηθά τα άτομα να κατανοήσουν τα δικαιώματά τους, τις ευθύνες τους και τη σημασία της συγκατάθεσης μετά από ενημέρωση όταν πρόκειται για την κοινή χρήση δεδομένων υγείας και την προστασία της ιδιωτικής ζωής. Με την κατανόηση των εννοιών και της ορολογίας των δεδομένων, άτομα με διαφορετικό υπόβαθρο μπορούν να επικοινωνούν με μεγαλύτερη σαφήνεια, να μοιράζονται ιδέες και να συνεργάζονται για την επίτευξη κοινών στόχων. Αυτή η συνεργασία ενισχύει τη διεπιστημονική έρευνα, την παροχή υγειονομικής περίθαλψης και την ανάπτυξη πολιτικής.

Στην εποχή της υπερφόρτωσης με πληροφορίες, η εκπαίδευση και η κατάρτιση στον αλφαριθμητισμό των δεδομένων υγείας επιτρέπει στα άτομα να αξιολογούν κριτικά τις πληροφορίες υγείας από διάφορες πηγές, συμπεριλαμβανομένων των διαδικτυακών πλατφορμών, των επιστημονικών άρθρων και των εκθέσεων (Kuhn *et al.*, 2018). Μπορούν να διακρίνουν μεταξύ αξιόπιστων και αναξιόπιστων πηγών, να αξιολογούν την αξιοπιστία των δεδομένων και να λαμβάνουν τεκμηριωμένες αποφάσεις με βάση ακριβείς πληροφορίες. Με την προώθηση του αλφαριθμητισμού των δεδομένων υγείας μέσω της εκπαίδευσης και της κατάρτισης, τα άτομα και οι επαγγελματίες είναι

καλύτερα εξοπλισμένοι για να περιηγηθούν και να χρησιμοποιήσουν αποτελεσματικά τα δεδομένα υγείας, συμβάλλοντας στη βελτίωση των αποτελεσμάτων της υγειονομικής περίθαλψης, στη λήψη αποφάσεων βάσει στοιχείων και στην πρόοδο της έρευνας και της καινοτομίας.

Επίλογος

Η ανάλυση των ιατρικών δεδομένων στην εποχή των μεγάλων δεδομένων μπορεί να προωθήσει την ιατρική έρευνα και την καινοτομία. Οι ερευνητές και οι επιστήμονες μπορούν να αξιοποιήσουν μεγάλα σύνολα δεδομένων για να αποκαλύψουν νέες γνώσεις, να ανακαλύψουν νέους βιοδείκτες, να εντοπίσουν παράγοντες κινδύνου ασθενειών και να αναπτύξουν καινοτόμες θεραπευτικές προσεγγίσεις. Αναλύοντας ποικίλες πηγές ιατρικών δεδομένων, συμπεριλαμβανομένων των ηλεκτρονικών αρχείων υγείας, της γονιδιωματικής, της ιατρικής απεικόνισης, των φορητών συσκευών και των αποτελεσμάτων που αναφέρουν οι ασθενείς, οι ερευνητές μπορούν να αποκτήσουν μια ολιστική κατανόηση των ασθενειών, της εξέλιξής τους και του αντίκτυπου των διαφόρων θεραπειών. Η γνώση αυτή μπορεί να δώσει πληροφορίες για την ανάπτυξη νέων φαρμάκων, θεραπειών και παρεμβάσεων, προωθώντας έτσι την ιατρική επιστήμη και προσφέροντας νέες δυνατότητες πρόληψης, έγκαιρης ανίχνευσης και αποτελεσματικής διαχείρισης ασθενειών.

Εκτός από τα δυνητικά οφέλη, είναι ζωτικής σημασίας να αναγνωριστεί η σημασία της προστασίας των ιατρικών δεδομένων στην εποχή των μεγάλων δεδομένων. Τα ιατρικά δεδομένα περιλαμβάνουν ευαίσθητες και ιδιωτικές πληροφορίες, συμπεριλαμβανομένων των προσωπικών αρχείων υγείας, των γενετικών προφίλ και των δημογραφικών στοιχείων των ασθενών. Η αυξημένη ψηφιοποίηση και η ανταλλαγή ιατρικών δεδομένων παρουσιάζουν τόσο ευκαιρίες όσο και προκλήσεις όσον αφορά την ασφάλεια των δεδομένων και την προστασία της ιδιωτικής ζωής. Οι οργανισμοί υγειονομικής περίθαλψης και οι ερευνητές πρέπει να υιοθετήσουν ισχυρά μέτρα για τη διασφάλιση της ιδιωτικής ζωής των ασθενών, τη διατήρηση της ακεραιότητας των δεδομένων και τη συμμόρφωση με τους ηθικούς και νομικούς κανονισμούς. Η κρυπτογράφηση δεδομένων, οι τεχνικές ανωνυμοποίησης, οι έλεγχοι πρόσβασης και τα πρωτόκολλα ασφαλούς ανταλλαγής δεδομένων είναι μερικές από τις στρατηγικές που χρησιμοποιούνται για τη διασφάλιση της εμπιστευτικότητας και της ακεραιότητας των ιατρικών δεδομένων. Η προστασία των ιατρικών δεδομένων δεν αποτελεί μόνο νομική και δεοντολογική επιταγή, αλλά είναι επίσης ζωτικής σημασίας για την ενίσχυση της εμπιστοσύνης του κοινού, την ενθάρρυνση της

συμμετοχής των ασθενών στην έρευνα και τη διευκόλυνση της συνεργασίας για την ανταλλαγή δεδομένων μεταξύ ιδρυμάτων

Επιπλέον, η διασυνδεδεμένη φύση των ιατρικών δεδομένων στην εποχή των μεγάλων δεδομένων αναδεικνύει τη σημασία της διαλειτουργικότητας και της ολοκλήρωσης των δεδομένων. Τα ιατρικά δεδομένα παράγονται και αποθηκεύονται σε διάφορα συστήματα, συμπεριλαμβανομένων των ηλεκτρονικών φακέλων υγείας, των εργαστηριακών πληροφοριακών συστημάτων, των πλατφορμών απεικόνισης και των φορητών συσκευών. Η ενοποίηση αυτών των διαφορετικών πηγών δεδομένων και η καθιέρωση διαλειτουργικών πλαισίων επιτρέπουν στους παρόχους υγειονομικής περίθαλψης να έχουν πρόσβαση σε ένα ολοκληρωμένο προφίλ ασθενούς, να διευκολύνουν την απρόσκοπτη ανταλλαγή πληροφοριών και να εξάγουν χρήσιμες πληροφορίες. Η διαλειτουργικότητα προωθεί τη λήψη αποφάσεων βάσει δεδομένων, μειώνει την επανάληψη των προσπαθειών και ενισχύει το συντονισμό της περίθαλψης μεταξύ των διάφορων παρόχων υγειονομικής περίθαλψης. Επιπλέον, επιτρέπει τη διαχείριση της υγείας του πληθυσμού, τις επιδημιολογικές μελέτες και τον εντοπισμό τάσεων και παρεμβάσεων στον τομέα της δημόσιας υγείας.

Παρά τις τεράστιες δυνατότητες και ευκαιρίες, πρέπει να αντιμετωπιστούν αρκετές προκλήσεις στην εποχή των μεγάλων δεδομένων στην ιατρική. Ο τεράστιος όγκος, η ταχύτητα και η ποικιλία των ιατρικών δεδομένων θέτουν σημαντικές προκλήσεις σε επίπεδο υπολογισμού, αποθήκευσης και ανάλυσης. Οι οργανισμοί υγειονομικής περίθαλψης πρέπει να επενδύσουν σε ισχυρές υποδομές, δυνατότητες αποθήκευσης δεδομένων και κλιμακούμενους υπολογιστικούς πόρους για την αποτελεσματική διαχείριση και επεξεργασία μεγάλων ιατρικών δεδομένων. Επιπλέον, υπάρχει ανάγκη για εξειδικευμένους επιστήμονες δεδομένων, βιοπληροφορικούς και επαγγελματίες του τομέα της υγειονομικής περίθαλψης που διαθέτουν την τεχνογνωσία για την ανάλυση και την ερμηνεία σύνθετων ιατρικών δεδομένων. Τα εκπαιδευτικά προγράμματα, οι συνεργασίες μεταξύ ακαδημαϊκών και βιομηχανικών φορέων και οι πρωτοβουλίες συνεχούς επαγγελματικής ανάπτυξης μπορούν να συμβάλουν στη γεφύρωση αυτού του χάσματος δεξιοτήτων και στην προώθηση ενός

εργατικού δυναμικού εξοπλισμένου για την αξιοποίηση της δύναμης των μεγάλων δεδομένων στην ιατρική.

Τα ιατρικά δεδομένα έχουν τεράστια σημασία στην εποχή των μεγάλων δεδομένων, προσφέροντας μετασχηματιστικές ευκαιρίες για εξατομικευμένη ιατρική, ιατρική έρευνα και βελτιστοποίηση της υγειονομικής περίθαλψης. Η ανάλυση των ιατρικών δεδομένων επιτρέπει την προσαρμογή των θεραπειών, την προώθηση της ιατρικής γνώσης και τη βελτίωση των αποτελεσμάτων των ασθενών. Ωστόσο, είναι ζωτικής σημασίας η διασφάλιση της προστασίας και της ιδιωτικής ζωής των ιατρικών δεδομένων, η τήρηση των δεοντολογικών και νομικών πλαισίων και η αντιμετώπιση των προκλήσεων που σχετίζονται με τη διαλειτουργικότητα και την ανάλυση των δεδομένων. Αγκαλιάζοντας το δυναμικό των ιατρικών δεδομένων στην εποχή των μεγάλων δεδομένων, μπορούμε να ξεκλειδώσουμε νέα σύνορα στην υγειονομική περίθαλψη, να προωθήσουμε την καινοτομία και τελικά να βελτιώσουμε την ευημερία των ατόμων και των πληθυσμών. Στην παρούσα εργασία θα μελετηθεί αναλυτικά η σημασία των ιατρικών δεδομένων και των δυνατοτήτων που προσφέρει η τεχνολογία και η επιστήμη της πληροφορίας για την ανάδειξη και προστασία των παραπάνω.

Βιβλιογραφία

Abouelmehdi, K. *et al.* (2017) 'Big data security and privacy in healthcare: A Review', *Procedia Computer Science*, 113, pp. 73–80.

Aiello, M. *et al.* (2019) 'The challenges of diagnostic imaging in the era of big data', *Journal of clinical medicine*, 8(3), p. 316.

Akhtar, U., Khattak, A.M. and Lee, S. (2016) 'Challenges in managing real-time data in health information system (HIS)', in *Inclusive Smart Cities and Digital Health: 14th International Conference on Smart Homes and Health Telematics, ICOST 2016, Wuhan, China, May 25-27, 2016. Proceedings 14*, pp. 305–313.

Asri, H. *et al.* (2015) 'Big data in healthcare: challenges and opportunities', in *2015 International Conference on Cloud Technologies and Applications (CloudTech)*, pp. 1–7.

Association, I.M.I. (2023) *International Medical Informatics Association*,. Available at: <https://imia-medinfo.org/wp/>.

Athey, S. (2017) 'Beyond prediction: Using big data for policy problems', *Science*, 355(6324), pp. 483–485.

Bahga, A. and Madiseti, V.K. (2015) 'Healthcare data integration and informatics in the cloud', *Computer*, 48(2), pp. 50–57.

Baro, E. *et al.* (2015) 'Toward a literature-driven definition of big data in healthcare', *BioMed research international*, 2015.

Beam, A.L. and Kohane, I.S. (2018) 'Big data and machine learning in health care', *Jama*, 319(13), pp. 1317–1318.

Benjelloun, F.-Z. and Lahcen, A.A. (2019) 'Big data security: challenges, recommendations and solutions', in *Web Services: Concepts, Methodologies, Tools, and Applications*. IGI Global, pp. 25–38.

BIS Research (2023) *Global healthcare big data market size in 2016 and a forecast for 2025*, *Statista*. Available at: <https://www.statista.com/statistics/909654/global-big->

data-in-healthcare-market-size/.

Bleakley, A. (2005) 'Stories as data, data as stories: making sense of narrative inquiry in clinical education', *Medical education*, 39(5), pp. 534–540.

Brazhnik, O. and Jones, J.F. (2007) 'Anatomy of data integration', *Journal of biomedical informatics*, 40(3), pp. 252–269.

Bukovina, J. (2016) 'Social media big data and capital markets—An overview', *Journal of Behavioral and Experimental Finance*, 11, pp. 18–26.

Cancer Genome Atlas (2023) *The Cancer Genome Atlas Program (TCGA) - NCI*. Available at: <https://www.cancer.gov/ccg/research/genome-sequencing/tcga> (Accessed: 26 June 2023).

Capitalist, V. and Publicis (2023) *Projected value of big data investments in healthcare globally in 2018 and 2021*, Statista. Available at: <https://www.statista.com/statistics/1038084/big-data-investments-in-healthcare-value/>.

CDC (2023) *CDC WONDER*. Available at: <https://wonder.cdc.gov/> (Accessed: 26 June 2023).

Chen, M., Mao, S., Zhang, Y., Leung, V.C.M. and others (2014) *Big data: related technologies, challenges and future prospects*. Springer.

Chen, M., Mao, S., Zhang, Y., Leung, V.C.M., Chen, M., *et al.* (2014) 'Big data generation and acquisition', *Big Data: Related Technologies, Challenges and Future Prospects*, pp. 19–32.

ClinicalTrials.gov (2023) *Beta ClinicalTrials.gov*. Available at: <https://clinicaltrials.gov/> (Accessed: 26 June 2023).

Elgendy, N. and Elragal, A. (2014) 'Big data analytics: a literature review paper', in *Advances in Data Mining. Applications and Theoretical Aspects: 14th Industrial Conference, ICDM 2014, St. Petersburg, Russia, July 16-20, 2014. Proceedings 14*, pp. 214–227.

European Institute for Health Records (2023) *European Institute for Health Records*.

Available at: <https://www.eurorec.org/>.

Evaluation, I. for H.M. and (2023) *Institute for Health Metrics and Evaluation*.

Available at: <https://www.healthdata.org/>.

Goodchild, M.F. (2016) 'GIS in the Era of Big Data', *Cybergeo: European journal of geography* [Preprint].

Hamouda, S. and Zainol, Z. (2019) 'Semi-Structured Data Model for Big Data (SS-DMBD).', in *DATA*, pp. 348–356.

Hariharakrishnan, J. *et al.* (2017) 'Survey of pre-processing techniques for mining big data', in *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)*, pp. 1–5.

Health Data Research UK (2023) *Health Data Research UK*. Available at: <https://www.hdruk.ac.uk/>.

Health Level Seven International (2023) *Health Level Seven International - Homepage / HL7 International*. Available at: <https://www.hl7.org/> (Accessed: 20 June 2023).

Henkel, M. and Stieltjes, B. (2020) 'Structured data acquisition in oncology', *Oncology*, 98(6), pp. 423–429.

Hettinger, B.J. and Brazile, R.P. (1994) 'Health Level Seven (HL7): standard for healthcare electronic data transmissions', *Computers in nursing*, 12(1), pp. 13–16.

Hong, Y. and Zeng, M.L. (2023) 'International classification of diseases (ICD)', *KNOWLEDGE ORGANIZATION*, 49(7), pp. 496–528.

Jaleel, A. *et al.* (2020) 'Towards medical data interoperability through collaboration of healthcare devices', *IEEE Access*, 8, pp. 132302–132319.

Joyia, G.J. *et al.* (2018) 'Evolution of health level-7: A survey', in *Proceedings of the 2018 International Conference on Software Engineering and Information Management*, pp. 118–123.

Kankanhalli, A. *et al.* (2016) 'Big data and analytics in healthcare: Introduction to the special section', *Information Systems Frontiers*, 18, pp. 233–235.

- Kao, R.R. *et al.* (2014) 'Supersize me: how whole-genome sequencing and big data are transforming epidemiology', *Trends in microbiology*, 22(5), pp. 282–291.
- Khan, W.A. *et al.* (2014) 'An adaptive semantic based mediation system for data interoperability among health information systems', *Journal of medical systems*, 38, pp. 1–18.
- Kitchin, R. and McArdle, G. (2016) 'What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets', *Big Data & Society*, 3(1), p. 2053951716631130.
- Kong, X. *et al.* (2022) 'Disease-specific data processing: An intelligent digital platform for diabetes based on model prediction and data analysis utilizing big data technology', *Frontiers in Public Health*, 10, p. 1053269.
- Kouanou, A.T. *et al.* (2018) 'An optimal big data workflow for biomedical image analysis', *Informatics in Medicine Unlocked*, 11, pp. 68–74.
- Kuhn, S. *et al.* (2018) 'Data literacy in medicine: What competences does a physician need?', *Der Onkologe*, 24, pp. 368–377.
- Lyko, K., Nitzschke, M. and Ngonga Ngomo, A.-C. (2016) 'Big data acquisition', *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*, pp. 39–61.
- Manogaran, G. *et al.* (2017) 'Big data knowledge system in healthcare', *Internet of things and big data technologies for next generation healthcare*, pp. 133–157.
- Maro, J.C. *et al.* (2009) 'Design of a national distributed health data network', *Annals of internal medicine*, 151(5), pp. 341–344.
- Marx, V. (2013) 'The big challenges of big data', *Nature*, 498(7453), pp. 255–260.
- National Center for Biotechnology Information (2023) *National Center for Biotechnology Information*. Available at: <https://www.ncbi.nlm.nih.gov/> (Accessed: 26 June 2023).
- Navar, A.M. *et al.* (2016) 'Use of open access platforms for clinical trial data', *Jama*, 315(12), pp. 1283–1284.

Nazari, E., Shahriari, M.H. and Tabesh, H. (2019) 'BigData analysis in healthcare: apache hadoop, apache spark and apache flink', *Frontiers in Health Informatics*, 8(1), p. 14.

OpenFEMA (2023) *OpenFEMA / FEMA.gov*. Available at: <https://www.fema.gov/about/reports-and-data/openfema> (Accessed: 26 June 2023).

Patil, H.K. and Seshadri, R. (2014) 'Big data security and privacy issues in healthcare', in *2014 IEEE international congress on big data*, pp. 762–765.

Peterson, K. (2020) *A Corpus-Driven Standardization Framework for Encoding Clinical Problems with SNOMED CT Expressions and HL7 FHIR*. University of Minnesota.

Pezoulas, V.C. *et al.* (2019) 'Medical data quality assessment: On the development of an automated framework for medical data curation', *Computers in biology and medicine*, 107, pp. 270–283.

Poornima, S. and Pushpalatha, M. (2018) 'A survey of predictive analytics using big data with data mining', *International journal of bioinformatics research and applications*, 14(3), pp. 269–282.

Purswani, J.M. *et al.* (2019) 'Big data from small devices: the future of smartphones in oncology', in *Seminars in radiation oncology*, pp. 338–347.

Rahimi, M. *et al.* (2022) 'Cloud healthcare services: A comprehensive and systematic literature review', *Transactions on Emerging Telecommunications Technologies*, 33(7), p. e4473.

Rakthanmanon, T. *et al.* (2013) 'Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping', *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 7(3), pp. 1–31.

Rao, D., Gudivada, V.N. and Raghavan, V. V (2015) 'Data quality issues in big data', in *2015 IEEE International Conference on Big Data (Big Data)*, pp. 2654–2660.

Ristevski, B. and Chen, M. (2018) 'Big Data Analytics in Medicine and Healthcare', *Journal of Integrative Bioinformatics*, 15(3). Available at: <https://doi.org/10.1515/JIB->

2017-0030.

Robinson, T.E. *et al.* (2015) 'Collaborative research networks in health: a pragmatic scoping study for the development of an imaging network', *Health Research Policy and Systems*, 13(1), pp. 1–7.

Ross, M.K., Wei, W. and Ohno-Machado, L. (2014) "'Big data" and the electronic health record', *Yearbook of medical informatics*, 23(01), pp. 97–104.

Rudin, R.S. *et al.* (2014) 'Usage and effect of health information exchange: a systematic review', *Annals of internal medicine*, 161(11), pp. 803–811.

Sagiroglu, S. and Sinanc, D. (2013) 'Big data: A review', in *2013 international conference on collaboration technologies and systems (CTS)*, pp. 42–47.

Saranya, P. and Asha, P. (2019) 'Survey on big data analytics in health care', in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 46–51.

Schatz, M.C. *et al.* (2022) 'Inverting the model of genomics data sharing with the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space', *Cell Genomics*, 2(1).

Senthilkumar, S.A. *et al.* (2018) 'Big data in healthcare management: a review of literature', *American Journal of Theoretical and Applied Business*, 4(2), pp. 57–69.

Shilo, S., Rossman, H. and Segal, E. (2020) 'Axes of a revolution: challenges and promises of big data in healthcare', *Nature medicine*, 26(1), pp. 29–38.

Starkbaum, J. and Felt, U. (2019) 'Negotiating the reuse of health-data: Research, big data, and the European general data protection regulation', *Big Data & Society*, 6(2), p. 2053951719862594.

Stone, A. *et al.* (2007) *The science of real-time data capture: Self-reports in health research*. Oxford University Press.

Taleb, I., Dssouli, R. and Serhani, M.A. (2015) 'Big data pre-processing: A quality framework', in *2015 IEEE international congress on big data*, pp. 191–198.

Viceconti, M., Hunter, P. and Hose, R. (2015) 'Big data, big knowledge: big data for personalized healthcare', *IEEE journal of biomedical and health informatics*, 19(4), pp. 1209–1215.

Wade, T.D. (2014) 'Traits and types of health data repositories', *Health information science and systems*, 2, pp. 1–8.

Wang, W. *et al.* (2018) 'Experimental and numerical validation of guided wave phased arrays integrated within standard data acquisition systems for structural health monitoring', *Structural Control and Health Monitoring*, 25(6), p. e2171.

WHO (2023) *Global Health Observatory*. Available at: <https://www.who.int/data/gho> (Accessed: 26 June 2023).

Yeh, C.-Y. *et al.* (2021) 'Logical observation identifiers names and codes (Loinc®) applied to microbiology: A national laboratory mapping experience in Taiwan', *Diagnostics*, 11(9), p. 1564.