

# Səhiyyədə Big Data Erası

Rasim Əliquliyev<sup>1</sup>, Məkrufə Hacırahimova<sup>2</sup>

Azərbaycan Milli Elmlər Akademiyası İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

<sup>1</sup>rasim@science.az, <sup>2</sup>makrufa@science.az

**Xülasə**—Məqalədə informasiya cəmiyyətinin əsas hərəkətverici qüvvəsi hesab edilən Big data mövzusunda baxılır, Big data-nın mahiyyəti açıqlanır. Digər sahələrdə olduğu kimi, səhiyyədə big data-nın mənbələri, imkanları şərh olunur. Həmçinin big data-nın səhiyyədə yaratdığı bəzi problemlər analiz olunur.

**Açar sözlər**— Big data; səhiyyə; e-tibb; genom; biotibb; teletibb; elektron sağlamlıq kartları; fərdi tibbi məlumatlar.

## I. GİRİŞ

Müasir informasiya texnologiyaları bizim şəxsi həyatımızı, düşüncəmizi əhəmiyyətli dərəcədə dəyişmişdir. Real (fiziki) dünyada baş verən proseslər onlayn mühitə daşınmışdır: xidmət və malların əldə olunması, tanışlarla ünsiyyət, işə götürən və dövlət orqanları ilə qarşılıqlı əlaqə və s. Bu vəziyyət bir çox faktorların təsirindən yaranmışdır:

1) İnternetin gündəlik həyata daxil olması;

2) elektron kommersiyanın yaranması;

3) fərdlərin İnternet şəbəkədə davranışları haqqında geniş informasiyanın toplanması (təxmin edilən və kökündə biznes model dayanan axtarış servislərinin yaranması və inkişafı);

4) tək-cə fərdlər haqqında verilənləri deyil, həm də onların münasibətlərini aqrəqasiya edən sosial şəbəkələrin meydana gəlməsi;

5) hər yerdə daim onlayn olmağa, istifadəçilərin hərəkət marşrutunu izləməyə, hər an məlumat mübadiləsinə imkan verən smartfon və planşetlərin yayılması.

Bunun nəticəsi olaraq, insan fəaliyyətinin əsas hissəsi İnternetə daşınır və fərdin istənilən hərəkəti müəyyən izlər qoyur ki, bunların ümumi cəmi rəqəmsal informasiyanın sürətlə artmasına səbəb olur. İnformasiyanın belə “partlayışı” dünyada “Big data” adlandırılan ciddi bir problemə çevrilmişdir. Özünü bütün fəaliyyət sahələrində göstərən bu problem səhiyyə sferası üçün də xarakterikdir [1–3].

Belə ki, səhiyyə sferasında fəaliyyətin ənənəvi (kağız) formadan elektron formaya keçirilməsi nəticəsində səhiyyə informasiya sistemlərində və klinikalarda qərarların qəbul edilməsini dəstəkləyən sistemlərdə (ing. *clinical decision support systems*—CDSS) müxtəlif növ böyük həcmdə verilənlər toplanmış və hazırda bu proses dinamik olaraq davam etməkdədir. Bu verilənlər toplusu (*pasiyentlərin elektron tibbi yazıları, dərman vasitələri, elektron reyestr sistemləri və s.*) bir-birindən həcmi, sürəti və müxtəlifliyi ilə fərqlənirlər [4, 5].

Tibbdə böyük verilənlər ənənəvi və İnternet, smartfonlar, sosial və kütləvi informasiya vasitələri və ya qeyd olunduğu kimi, xəstəxana və klinikaların informasiya sistemləri

vasitələri ilə yaradılır. Böyük həcmdə informasiyanı necə saxlamaq, necə idarə etmək, ondan faydalı informasiyanı necə əldə etmək çox ciddi məsələdir. Çünki tibbi verilənlər diaqnoz qoymaq üçün real vaxtda əməl tələb edir. Problem müxtəlif mənbələrdən və müxtəlif formatda (*strukturlaşdırılmış və strukturlaşdırılmamış*) fasiləsiz olaraq generasiya olunan verilənlərin ənənəvi üsul və alətlərin köməyi ilə əməl edilə bilməməsindədir [6]. Bunun üçün daha mükəmməl texnologiyalara ehtiyac vardır. Bu işdə məqsəd oxucuları və tədqiqatçıları səhiyyə sistemində big data texnologiyalarının imkanları və problemləri ilə tanış etmək, ekspertlər (*həkimlər*) tərəfindən əsaslandırılmış operativ qərarların qəbul edilməsində onun əsas şərt və komponent olduğunu göstərməkdir.

## II. BIG DATA PARADİQMASI

Veb, sosial şəbəkələr, mobil qurğular, kredit kartları vasitəsilə edilən tranzaksiyalar, müəssisələrin informasiya sistemləri, kameralar, elektron sağlamlıq kartları məlumatları və s. rəqəmsal verilənlər axınının artmasına gətirib çıxarmış, informasiya bolluğu yaranmış, dünya sanki informasiya ilə doldurulmuşdur [6].

XXI əsrin əvvəllərindən başlayaraq rəqəmsal verilənlər hər il həndəsi silsilə ilə artmışdır. Bunu informasiya texnologiyaları sferasında analitika üzrə ixtisaslaşmış IDC (International Data Corporation), Gartner və s. kimi analitik şirkətlərin hesabatları da təsdiq edir [7–9]. 2013-cü ildə dünyada informasiyanın həcmi 4,4 zettabayt olmuşdur və bunun ancaq 2%-dən azını qeyri-rəqəmsal informasiya təşkil etmişdir. IDC şirkətinin proqnozlarına görə, informasiyanın həcmi ildə 40% artır. 2020-ci ildə dünyadakı rəqəmsal informasiyanın həcmi 44 zettabayta çatacaqdır [7].

Həm də 2012-2020-ci illər arasında verilənlərin əksəriyyəti insanlar tərəfindən deyil, müxtəlif tip qurğuların bir-biri ilə qarşılıqlı əlaqəsi zamanı, radiotezlikli identifikasiya qurğuları (RFID), naviqasiya sistemləri (GPS) vasitəsilə generasiya olunacaqdır. İnternet şəbəkəsinin növbəti inkişaf mərhələsi hesab olunan “Əşyaların İnterneti” konsepsiyası əsasında müxtəlif tip qurğuların İnternet vasitəsilə qarşılıqlı əlaqəsi zamanı maşınlar yalnız informasiya istehsalçısı deyil, həm də onun istehlakçısına çevrilirlər. Bütün bu deyilənlərin məntiqi nəticəsi olaraq verilənlərin emalı, saxlanması və istifadəsində yeni eramı əks etdirən “böyük verilənlər” (ing. *Big data*) fenomeni meydana çıxmışdır [2, 3, 6]. Ümumiyyətlə, “big data” termini mürəkkəb strukturlu, müxtəlif tip, yüksək sürətli, çox böyük həcmli (*1 terabaytdan başlamış zettabaytlara qədər*) verilənlər toplusuna aiddir. Təəssüflər olsun ki, bu günə kimi böyük verilənlərin (BV) vahid tərifinin olmaması “big data” fenomeninin anlaşılmasında qeyri-müəyyənlik yaradır. Elm, sənaye və biznes kimi maraqlı tərəflərin müxtəlif

yanaşmaları mövcuddur. Tətbiq olunduğu sahələrdən asılı olmayaraq big data-nı təsvir etmək üçün ümumi xarakteristikalar - həcm (ing. *volume*), sürət (ing. *velocity*) və müxtəliflik (ing. *variety*) mövcuddur.

McKinsey Global İnstitutunun hesabatında bu termin informasiyanın toplanması, saxlanması, idarə edilməsi, axtarışı və analizinin tipik verilənlər bazasının imkanları xaricində olan verilənlər dəsti kimi xarakterizə olunur [8]. Daha bir populyar tərif isə Gartner şirkəti tərəfindən təklif olunmuş “3V” modelidir [9]. Bu model big data texnologiyalarının əsas konsepsiyasını – çox böyük sürətlə və müxtəlif mənbələrdən toplanan çox böyük həcmdə verilənləri daha səmərəli istifadə etmək, saxlamaq, analiz edərək ondan daha qiymətli informasiyanı əldə etmək ideyasını özündə əks etdirir. Strateji və idarəetmə baxımından mütəxəssislər bu modelin bir qədər təkmilləşdirilməsinə ehtiyacın olduğunu qeyd edərək, big data-nin həqiqilik (*veracity*), dəyər (*value*) kimi xarakteristikalarla ölçülməsini də vacib sayırlar [10, 11].

#### **A. Big data-nın potensialını formalaşdıran trendlər**

Müasir informasiya cəmiyyəti ancaq informasiyanın həcmünün artması ilə deyil, həm də müxtəlif fəaliyyət sahələrində informasiyanın rolunun artması ilə xarakterizə olunur. İnformasiya “yeni neft” adlandırılan yeni istehsal faktoru statusu əldə edir, informasiya cəmiyyətinin hərəkətverici qüvvəsinə çevrilir. Necə ki, sənaye cəmiyyəti dövründə ənənəvi neft əsas resurs idi. Ekspertlərin fikrincə, verilənlərin toplanması, inteqrasiyası və analizi daha biznes üçün xərc deyil, səmərəliliyə və gəlirə çatmaq üçün açaqdır.

Big data-nın analizi və onlardan biliklərin və faydalı informasiyanın əldə olunması idarəetmə və biznesdə inqilabi dəyişikliklər, müəssisələrdə yüksək mənfəət əldə etmək, bir çox sahələrdə elmi ideyaların inkişaf etdirilməsi və reallaşdırılması, yeni-yeni elmi kəşflərin edilməsində, təşkilatlarda əsaslandırılmış qərarların qəbul edilməsində, milli təhlükəsizlik, səhiyyə sahəsi və s. üçün böyük potensial vardır [6, 12]. Big data-nın bu potensialı istər, dövlət qurumları, istər elm, istərsə də biznes cəmiyyətləri tərəfindən dərk edilməkdədir. Bunu sübut edən aşağıdakı faktları qeyd etmək olar:

- Big data ABŞ və bir sıra qərb dövlətlərində elmi ictimaiyyət, biznes dairələri, hökumət strukturları tərəfindən neft qədər strateji resurs kimi dəyərləndirilir.

- Big data-nın müxtəlif aspektlərinə həsr olunmuş beynəlxalq konfranslar, seminarlar, forumlar keçirilməkdədir, elmi və populyar jurnalların xüsusi nömrələri bu mövzuya həsr olunmuşdur. 2014-cü ildən başlayaraq, “Big Data” adı ilə yeni akademik jurnalların nəşrinə başlanmışdır.

- Dünyanın tanınmış elmi mərkəzlərində BV-nin toplanması və emalı, saxlanması, arxitekturası, analitikası, təhlükəsizliyi, vizuallaşdırılması və s. istiqamətlərdə fundamental elmi-tədqiqat işləri aparılmaqdadır [3].

- 2013-cü ildən başlayaraq dünyanın bir çox aparıcı universitetlərində: “verilənlər haqqında elm” (*data science*) akademik fənn kimi bakalavr, magistr və doktorluq təhsil pilləsində tədris olunmağa başlanmışdır [6].

Heç də təsadüfi deyil ki, IDC, McKinsey Global İnstitutu, Gartner və s. kimi şirkətlər big data-nı əsas texnoloji istiqamət və informasiya-kommunikasiya texnologiyaları sahəsinin lokomotivi adlandırırlar [2–4].

### **III. SƏHIYYƏDƏ BIG DATA–MƏNBƏLƏRİ, İMKANLARI VƏ PROBLEMLƏRİ**

Səhiyyə rəqəmsal dünyanın sürətlə yüksələn seqmenti olmaqla, onun böyük hissəsini özündə əks etdirir. Bunu IDC-nin son illərdəki hesabatında verilmiş statistik rəqəmlərdə də görmək olar [7]. Hesabatdan da görüldüyü kimi səhiyyə sahəsində generasiya olunan verilənlərin həcmi son on ildə yüksələn xətt üzrə artmaqdadır. 2013-cü ildə səhiyyə sahəsindəki verilənlərin ölçüsü 153 ekzabayt idisə, 2020-ci ildə 2,314 ekzabayt olacağı proqnozlaşdırılır. Artımın səbəbi: yeni texnologiyaların yaranması; verilənlərin toplanması qurğuları, sensorlar, mobil tibbi tətbiqlərin geniş yayılması; genetik informasiyanın toplanmasının ucuzlaşması; pasiyentlərin rəqəmsal formada kommunikasiyalarının artması; daha çox tibbi biliyin toplanmasıdır və s. [5].

#### **A. Səhiyyədə Big data – mənbələri**

Ümumiyyətlə, səhiyyədə İT tətbiqləri iki kateqoriyaya bölünür [4]:

- **Klinika sistemləri** (ing. *clinical systems*). Bu kateqoriyaya düşən bir çox sistemlər vardır: Məsələn, təsvirlərin arxivləşdirilməsi və kommunikasiya sistemləri (ing. *Picture Archiving and Communications Systems – PACS*), xəstəxana informasiya sistemləri (ing. *Hospital Information System – HIS*), elektron sağlamlıq yazıları (ing. *Electronic Health Records – EHR*), radiologiya informasiya sistemləri (ing. *Radiology Information System – RIS*), təşkilatların kontentlərinin idarəedilməsi (ing. *Enterprise Content Management – ECM*), laboratoriya informasiya sistemləri (ing. *Laboratory Information System – LIS*) və s.

EHR pasiyentin tibbi məlumatlarının elektron versiyası olmaqla, onun demoqrafik verilənlərini, tibbi yazıları, dərmanları, həyati-vacib əlamətləri, keçdiyi tibbi tarixcə, immun, laboratoriya, radiologiya verilənləri və digər hesabatları özündə əks etdirir.

- **Administrasiya sistemləri** (ing. *administration systems*), mətnlərin işlənməsi vasitələri, elektron cədvəllər, təqdimatlar, hətta mühasibat proqram təminatı və s.

Bu sistemlərin yaratdığı verilənlər müxtəlif tipdə olur ki, bunları da strukturlaşdırılmış, strukturlaşdırılmamış, qismən strukturlaşdırılmış kateqoriyalara bölmək olar:

- strukturlaşdırılmış elektron sağlamlıq yazıları (*EHR*);
- strukturlaşdırılmamış klinika qeydləri (ing. *Unstructured Clinical Notes*);
- tibbi təsvir verilənləri (ing. *Medical Imaging Data*);
- genetik verilənlər (ing. *Genetic Data*);
- digər verilənlər (ing. *Epidemioloji və davranış*).

Bu təsnifat deməyə əsas verir ki, səhiyyədə toplanan verilənlər də bir qayda olaraq “5V” ilə xarakterizə olunur [3]:

- *Həcm (ing. volume):* pasiyentlərin elektron tibbi yazıları, dərman vasitələri, elektron reyestr sistemləri, tibbi elektron qurğular vasitəsilə yaradılan və ötürülən verilənlər və s. [4]-də qeyd olunduğu kimi, səhiyyə rəqəmsal dünyanın sürətlə yüksələn seqmenti olmaqla, onun 30% hissəsini özündə əks etdirir.

- *Müxtəliflik (ing. variety):* strukturlaşdırılmış EHR yazıları, strukturlaşdırılmamış kliniki qeydlər, video, audio və mətn kimi əlavə struktursuz kontentlər və s. International Healthcare Data məlumatında qeyd edilir ki, səhiyyədə verilənlərin 60-80%-ni təsvir, video və elektron poçt yazılarından ibarət strukturlaşdırılmamış verilənlər təşkil edir [5, 13].

1980-ci illərin ortalarından başlayaraq rəqəmsal tibbi qurğuların (*xüsusilə radiologiya, kardiologiya, oftomologiya və s. kimi sahələrdə*) meydana gəlməsi tibbi vizuallaşmanın simasını kökündən dəyişdi. Alınmış rəqəmsal təsvir verilənləri PACS sistemlərə ötürməklə, kompüterlərdə elektron şəkildə emal olunur və pasiyentləri yaxşı öyrənmək üçün onun demoqrafik, həyati vacib verilənləri ilə birgə saxlanılır. Nəticədə çox böyük ölçüdə məlumatlar toplanır.

- *Sürət (ing. velocity):* Yeni verilənlər yüksək tempolə axın şəklində yığılır və bəzi tibbi situasiyalarda verilənlər həyat və ölüm məsələsinə çevrilir. Real vaxt rejimində verilənləri analiz etmək (*travmaların monitorinqi, ürək monitorları, anesteziya və s.*), müqayisələr aparmaq və qərar qəbul etmək üçün də yüksək sürət lazımdır. Buraya, həmçinin gündəlik diabetik ölçmələr, arterial təzyiğin ölçülməsi, EKG (*ing. electrocardiography*) daxildir.

- *verilənlərin keyfiyyəti (ing. veracity – data quality):*

Verilənlərin keyfiyyəti səhiyyə sahəsində iki səbəbdən xüsusi diqqət tələb edir:

1. Bu həyat və ya ölüm məsələsidir ki, informasiyanın qaydada olmasından asılıdır.

2. Tibbi məlumatların keyfiyyəti, xüsusilə strukturlaşdırılmamış verilənlər (*məsələn, əl ilə yazılmış reseptlər*) çox vaxt düz olmur. Səhiyyədə pasiyent, xəstəxana, ödənişin miqdarı, ödəniş edən və s. verilənlərin həqiqiliyi üçün əsas parametrlərdir. Həqiqiliyin səhiyyə üçün digər unikal məsələləri isə diaqnozlar, reseptlər, müalicələr, prosedurlardır. Bir sözlə, tibbi xidmətlərin yaxşılaşdırılması, səhvlərin təkrarlanmaması, diaqnozların düzgünlüyü, dərman vasitələrinin effektivliyi, xərclərin azaldılması verilənlərin keyfiyyətindən asılıdır.

- *Dəyər (ing. value):*

- real vaxt rejimində əsaslandırılmış qərarın verilməsinə dəstək;
- düzgün diaqnoz;
- xərclərin azaldılması;
- xəstəlik risklərinin minimallaşdırılması;
- verilənlərin elmi tədqiqatlarda istifadəsi və s.

## B. Bio-tibbi verilənlər

Yüksək məhsuldarlıqlı biotexnoloji ölçmə texnologiyaları 21-ci əsrdə işlənmişdir və bio-tibb sahəsindəki tədqiqatlar da Big data erasına daxildir. Bio-tibbdə ağıllı, effektiv və dəqiq analitik modellərin, nəzəri sistemlərin yaradılması ilə mürəkkəb bioloji hadisənin arxasında mühüm idarə edən mexanizm aşkar edilə bilər. Bu yalnız biotibbin gələcək inkişafını təmin etmir, həmçinin milli təhlükəsizlik, tibbi xidmət kimi vacib təbiiqlər, yeni dərman istehsalı ilə əlaqəli olunan vacib strateji sənayenin inkişafında qəbul edilə bilər [14, 15].

İnsan Genomu Layihəsi (*ing. Human Genome Project – HGP*) və sonrakı inkişaf texnologiyasının uzunmüddətli inkişafı bu sahədə Big data-nın geniş təbiiqlərinə gətirir. Gen ardıcılığı (*ing. sequencing*) ilə əmələ gələn böyük məlumatların müxtəlif təbiiqlə tələblərə müvafiq olaraq ixtisaslaşdırılmış analizdən keçirməklə və onu klinik gen diaqnozu ilə birləşdirərək erkən diaqnoz və xəstəliyin fərdiləşdirilmiş müalicəsi üçün qiymətli informasiyanı vermək olar. İnsan geninin bir ardıcılığı 100 – 600 GB xam məlumatı əmələ gətirə bilər. Shenzhen-də Çin Milli Genombankında, izlənilə bilən 1.15 milyon insan nümunəsi və 150 000 heyvan, bitki və mikroorqanizm nümunələri daxil olmaqla, 1.3 milyon nümunə var.

Bu əvvəlcədən proqnozlaşdırıla bilər ki, biotibb texnologiyalarının inkişafı ilə, genin deşifrlənməsi daha sürətli və daha rahat olacaq və beləliklə, biotibbin böyük məlumatı fasiləsiz olaraq artır. Eyni zamanda kliniki tibbi yardım və tibbi tədqiqatlardan alınan məlumatlar sürətlə artır. Məsələn, ABŞ-ın Pitsburq Universiteti Tibb Mərkəzində (*ing. University of Pittsburgh Medical Center – UPMC*) belə 2TB məlumatı saxlayır. Practice Fusion Amerika şirkəti təxminən 200 000 xəstənin elektron tibbi yazılarını idarə edir. Explorys Amerika şirkəti isə klinik məlumatları, əməliyyat və texniki məlumatları, hətta maliyyə məlumatlarını təşkil etmək üçün platformalarla təmin edir.

## C. Teletibb

Pasiyentlər haqqında müxtəlif mənbələrdən alınmış informasiyanın birgə analizi ilə daha yaxşı xidmət təmin etmək mümkündür. Teletibbin əsas üstünlüyü ondadır ki, pasiyent xəstəxanaya getmədən smart qurğular vasitəsilə alınmış verilənlər əsasında həkimin ona real vaxt rejimində baxa, xroniki xəstəliyi monitorinq edə və müalicə təyin edə, dərmanların dozasını yaxşılaşdırma bilər. Digər bir cəhət isə xərclərin azalmasıdır. Teletibb hazırda geniş yayılmış və xəstəxanaların, şəxsi həkimlərin, səhiyyə agentliklərinin gündəlik fəaliyyəti ilə inteqrasiya olunmuşdur [16.]. Teletibbin əsas istiqamətləri isə aşağıdakılardır: onlayn və oflayn telekonsultasiyalar, cərrahi əməliyyatların translyasiyası, biomonitorinqlər, evdə teletibb.

#### IV. SƏHIYYƏDƏ BIG DATA İMKANLARI VƏ PROBLEMLƏRİ

##### A. Səhiyyədə Big Data imkanları

BV-in analitikası səhiyyə sistemində inqilabi dəyişikliklərə imkan verir. Əməliyyatların səmərəliliyinin yaxşılaşdırılması, xəstəlik epidemiyalarını öncədən söyləməyə, səhiyyə sahəsinə xərcləri optimallaşdırmağa, kliniki sınaqların monitorinqinin keyfiyyətini yaxşılaşdırmağa kömək edə bilər [1-3,6,17,18].

Səhiyyədə böyük verilənlərin ümumi hədəfləri aşağıdakılardır:

- böyük həcmdə tibbi məlumatlardan yararlanaraq pasiyentə lazım olan anda və lazım olan düzgün müdaxiləni təmin etmək.
- Pasiyentə fərdiləşdirilən qayğı göstərmək.
- Səhiyyə sisteminin bütün komponentlərinə (*provayder, ödəyən, xəstə və menecment*) potensial olaraq xeyir verməkdir.

Tibbdə big data pasiyentlərin xəstəlik əlamətlərinin səbəb və fəsadlarının aşkarlanmasında, xəstəliyin təhlükəlilik dərəcəsini və ya təkrar baş verməsini proqnozlaşdırmaqda, həmçinin ilkin tibbi-sanitar köməyin göstərilməsində vacib rol oynaya bilər. Böyük verilənlər tibbdə xəstəlik diaqnozunun qoyulmasında, xəstəliyin profilaktikasının proqnozunda, fərdi sağlamlığın və xəstəliyin idarə edilməsində fərdi proqramların yaradılmasında istifadə oluna bilər.

Big data texnologiyaları tibbi fəaliyyətdə də mühüm rol oynaya bilər. Bu sahədə big data texnologiyaları pasiyentin müalicəsində fərdi yanaşmanı təmin etməyə imkan verir, big data texnologiyaları istifadə edən qurğu həkimin fərdi köməkçisi qismində çıxış edir. Qabaqcıl elmi tədqiqatlara əsaslanan verilənlərin analizi, uyğun xəstəliyin müalicəsi təcrübəsi, ayrı-ayrı müalicə preparatlarının kliniki tədqiqatı pasiyent üçün orqanizmin xüsusiyyətlərini nəzərə alan fərdi müalicə planı işləmək olar. Bu istiqamətdə bir çox şirkətlər aktiv işlər aparırlar. Məsələn, IBM səhiyyə sahəsində insan genlərinin analizini və çox qısa zamanda pasiyentin düzgün müalicəsinin seçilməsini təmin edən sistemlərin işlənməsi üçün bir çox təşkilatlarla əməkdaşlıq edir. Bu sistemlər genlər haqqında informasiya toplayır, təyin olunmuş müalicəyə pasiyentin reaksiyasını öyrənir və insanın DNT-sinin xüsusiyyəti nəzərə alınmaqla variantlar təklif edir [18, 19].

##### B. Səhiyyədə Big Data problemləri

Səhiyyədə sürətlə böyüyən verilənlər geniş imkanlarla yanaşı bir sıra problemlər də gətirir. Problemləri üç qrupa bölmək olar: texniki-texnoloji, hüquqi və elmi problemlər [2,3].

Səhiyyə sistemlərində Big data platformaları üçün proqram təminatları, onların inteqrasiyası, istifadəçi interfeysləri və s. texnoloji problemlər sırasındadır.

Səhiyyə sistemində, xüsusən də e-tibbdə məlumatların əksəriyyətini pasiyentlərin fərdi məlumatları təşkil edir. Bu baxımdan big data analizi texnologiyalarının tətbiqində vətəndaşların şəxsi həyatlarının rəqəmsal mühitdə qorunması məsələsi gündəmə gəlir. Burada əsas problem big data analizi texnologiyalarının təbiəti ilə fərdi məlumatlar haqqında qanunvericiliklə uyğunsuzluqların, ziddiyyətlərin meydana çıxmasıdır. Əvvəlcədən müəyyən edilmiş məqsədlər üçün

emal və ya təkrar istifadə olunan fərdi məlumatların həcmində olan məhdudiyətlər Big data-nın fəlsəfəsinə ziddir və bu texnologiyaları malik olduğu imkanlardan məhrum edir [20,21]. Bu problemə sonrakı bölmədə bir qədər geniş baxacağıq.

Avropa İttifaqı səhiyyə sahəsində big data texnologiyalarının tətbiqindəki əsas problemləri aşağıdakı kimi təqdim edir: verilənlərin konfidensiallığı, təhlükəsizliyi, həqiqiliyi, uyğunluğu və idarə edilməsi [4].

Pasiyentlər onların tibbi məlumatlarının, xüsusən də genetik verilənlərinin mənimlənməsindən çox ehtiyatlanırlar. Verilənlərə əlyətərlik və konfidensiallıq riskləri bir-birilə əlaqəlidir. Avropa Komissiyasının Open Data siyasəti ilə bu təhlükənin aradan qaldırılması üçün qanunvericilikdə islahatlara ehtiyac yaranır.

Bütün sahələrdə olduğu kimi, böyük həcmdə verilənlər saxlama, emal və analiz sahəsində ciddi problemlər yaradırlar [22–24]:

- heterogen mənbələrdən bilik çıxarmaq;
- yazılarda pasiyent/məlumat korrelyasiyaları istifadə etmək;
- strukturlaşdırılmamış klinika qeydlərinin düzgün kontekstdə anlanması;
- böyük həcmli tibbi təsvirlərdən ibarət verilənlərin effektiv emalı və potensial lazımı informasiyanın və biomarkerlərin çıxarılması;
- genetik verilənlərin analizi hesablama məsələsidir və standart kliniki verilənlərlə birləşdirildikdə əlavə mürəkkəblik yaradır.
- pasiyentlərin sensorlar, sosial qarşılıqlı əlaqələr vasitəsilə davranışları haqqındakı məlumatların toplanması.

Bu istiqamətdə problemlərin həllində daha mükəmməl intellektual analiz üsulları, vizuallaşdırma üsulları və s. lazımdır.

##### C. Səhiyyədə Big Data tətbiqləri

ABŞ-da 2009-cu ildə “Health Information Technology for Economic and Clinical Health” (HITECH) qəbul edildikdən sonra EHR sistemlərinin tətbiqinə başlanmışdır. EHR-in əsas üstünlüyü tibb işçilərini tam və dəqiq tibbi informasiya ilə təmin edə bilməsidir. Hazırda ondan ABŞ-ın əksər (50%) xəstəxanalarında istifadə olunur.

Yazılarda rəqəmsallaşdırılmasına beynəlxalq diqqət ABŞ EHR texnologiyalarının tətbiqində tək deyildir. Kanadada tibbi yardım və tibbi xidmətlərin yaxşılaşdırılması, ümumilikdə səhiyyə xidmətinin effektivliyini yüksəltmək məqsədi ilə sağlamlıq üçün EHR kimi rəqəmsal həllərin sürətləndirilməsi üçün “Canada Health Infoway” layihəsi qəbul edilmişdir. Kanada hökuməti tərəfindən 2,15 milyard Kanada dolları [25], Avropada isə EHR texnologiyalarının inkişafı üçün 3 milyard dollar vəsait ayrılmışdır [22].

Səhiyyə sahəsində big data tətbiqinin digər sinifinə isə MedicalBody Area Networks (MBAN) daxildir. Naqilsiz MBAN sensorları ürək döyüntülərini, elektrokardiogramı, bədən temperaturunu, tənəffüs tezliyini, döş qəfəsinin səslərini, qan təzyiqini və s. ötürməklə xəstənin vəziyyətini fasiləsiz monitorinq etməyə imkan verir. MBAN, həm də xəstənin sağlamlığına tarixi və real vaxt rejimində monitorinqi, infeksiya nəzarəti, pasiyentin izlənməsini həyata keçirməyə imkan verir. Buna baxmayaraq, yüz minlərlə xəstələrdən bu cür böyük MBAN massivlərini real vaxtda analiz etmək üçün səhiyyə

provayderləri intellektual və yüksək səviyyəli təhlükəsiz İKT infrastrukturuna malik olmalıdırlar.

#### D. Səhiyyənin gələcək modeli

Gələcək səhiyyənin modeli:

- səhiyyənin bütün sahələrində effektivliyin artırılması;
- daha çox əməkdaşlıq/özünə qayğı (ing. *collaborative/self-care*);
- əsas diqqətin əhalinin sağlamlığının idarə olunmasına yönəltməyi tələb edəcəkdir.

Verilənlər biliyə əsaslanan qərarlara imkan verəcək və bu işdə İT aparıcı faktor olacaqdır.

*İndi nə etmək lazımdır?* İlk növbədə aşağıdakılar tövsiyə olunur:

- işçi proseslər və tətbiqlər çərçivəsində İT imkanları və hüquqları yaxşılaşdırmaq və genişləndirmək;
- uyğun analitik texnologiyalar və mühit seçmək;
- təhlükəsizliyin daha effektiv idarə edilməsini təmin etmək;
- daha güclü mobil strategiyalar qəbul etmək;
- koordinasiya və komanda əsasında sağlamlıq üçün bulud tətbiqləri;
- tibbi ixtisaslar üzrə e-tibb məsələlərinin, tibb üzrə tədqiqatların aparılmasında verilənlər alimlərinin (ing. *data scientist*) hazırlanması və s.

#### V. FƏRDİ TİBBİ MƏLUMATLAR

Ümumiyyətlə, fərdi verilənlərin avtomatlaşdırılmış emalı problemlərinə həsr olunmuş xüsusi müddəalar ilk olaraq Avropada meydana çıxmış və sonralar bütün dünyaya yayılmışdır. 2012-ci ilin əvvəlinə olan məlumata görə, fərdi məlumatlar haqqında qanun dünyanın 89 ölkəsində qəbul edilmişdir. Bu sahədəki əsas hüquqi akt 1981-ci il 28 yanvar tarixində qəbul edilmiş Avropa Şurasının fərdi məlumatların avtomatlaşdırılmış emalı zamanı fiziki şəxslərin müdafiəsi Konvensiyasıdır. Bu Konvensiyanın müddəaları əsasında Avropa ölkələri milli səviyyədə fərdi verilənlərin tənzimlənməsinə həsr olunmuş qanunlar qəbul etmişlər [22].

Səhiyyə sferasının elektronlaşdırılması nəticəsində böyük həcmdə fərdi məlumatlar informasiya sistemlərinə daxil olur. Tədqiqatlar göstərir ki, əksər tibbi verilənlər kifayət qədər qorunmur və bu verilənlərin bir qismi – 7%-i açıq olmaqla mühafizə olunmağa ehtiyac duymur. Digər qismi (93%) isə gizli verilənləri təşkil edir və onların mühafizəsinə ehtiyac duyulur. Məsələn, tibbi yazılar, tarixçələr. Pasiyentlərin tibbi məlumatlarının 57% isə qismən qorunur, 43% isə adekvat olaraq qorunmur [7].

Tibb və səhiyyə verilənləri ən intensiv mübadilə edən sahələrdir. Big data erasında İKT, kompüter elmləri və tədqiqatlar arasında qarşılıqlı əlaqə səhiyyə sahəsində daha geniş yayılmışdır. Bu gün tibb müəssisələrinin həll edəcəyi ən aktual problem konfedensial fərdi tibbi verilənlərin mühafizəsidir.

Qanunvericilikdə fərdi məlumat, subyektin kimliyini birbaşa və ya dolayısı ilə müəyyənləşdirməyə imkan verən istənilən məlumatdır [26]. Bu sənədləşdirilmiş informasiyanın

onun sahibinin razılığı olmadan üçüncü şəxsə ötürülməsi qanunla qadağandır. Tibbi sirr daşıyan informasiyanın sahibi, məhz pasiyentdir və ya pasiyentin razılığını almış qanuni nümayəndəsidir. Fərdi məlumatların emalı dedikdə, onlar üzərindəki istənilən əməliyyat (*toplanması, sistemləşdirilməsi, saxlanması, yeniləndirilməsi, dəyişdirilməsi, istifadəsi, paylaşılması, məhv edilməsi*) nəzərdə tutulur. Tibbi müəssisələrdə fərdi məlumatların emalı informasiyanın konfidensiallığını təmin edən qaydalara riayət etməklə və icazəsiz müdaxilələrdən mühafizə olunmaqla həyata keçirilməlidir.

Verilənlərin əksəriyyəti fərdi məlumat olduğundan, big data texnologiyaları şəxsi həyatın toxunulmazlığı baxımından ciddi problemlər yaradır. Ona görə də bu verilənlər xüsusi mühafizə olmalıdır.

[11]-də qeyd olunduğu kimi, müasir izləmə texnologiyaları və big data analizi əsasında insanın “doğulduğu gündən ölənə qədər” izlənməsi imkanının yaranması onun şəxsi həyatının toxunulmazlığını sual altında qoyur. İnsanların xəbəri olmadan onların haqqındakı verilənlərin analiz olunması etik və hüquqi cəhətdən yolverilməzdir.

#### NƏTİCƏ

Mütəxəssislər haqlı olaraq 21-ci əsr insan həyatının bütün aspektlərini, o cümlədən tibbi, əhatə edən böyük həcmli verilənlər – big data dövrü adlandırırlar. Tibbi verilənlərin həcmi və müxtəlifliyinin artması bir daha göstərir ki, big data dövrü məhz səhiyyə üçün yaranmışdır. Bunun ancaq informasiya sistemlərinin həyata keçirilməsi və qiymətləndirilməsi nöqtəyi-nəzərindən deyil, həm də informatika sahəsində tədqiqatçıların və peşəkarların hazırlanmasında böyük təsiri vardır. Yəni tibbi mütəxəssislər informatika sahəsində nə etməlidirlər – big data və analitika sahəsində işləyənlər nəyi bilməlidirlər.

Ənənəvi tibbi kağız yazılardan elektron tibbi kart sisteminə keçilməsi verilənlərin eksponensial artmasına səbəb olmuşdur. Nəticədə böyük verilənlər həkimlər və səhiyyə sahəsində ekspertlərə verilənlər əsasında idarəetmə qərarlarının qəbulu üçün çox yaxşı imkanlar verir və tibbi tədqiqatlar üçün geniş poliqon. Səhiyyədəki aparıcı tendensiyalar verilənlərin həcmi artıracaq və verilənlərdən asılılıq yaradacaqdır. Belə bir mühitdə fərdi tibbi verilənlər xüsusi mühafizə olunmalıdırlar.

#### ƏDƏBİYYAT

- [1] Sh. Zhang, “Big Data for Healthcare in China: a Review of the State-of-art”, Cambridge Journal of China Studies, vol.10, no.1, pp.32-40., 2015.
- [2] S. V. Sebastian, A. Haimann, E. Mossialos, “Big Data and Health Care: Challenges and Opportunities for Coordinated Policy Development in the EU”, Health Systems & Reform, vol. 1, no. 4, pp.2 85–300, 2015.
- [3] E. W. Susan, “A review of big data in health care: challenges and opportunities”, Open Access Bioinformatics,;no.6, pp. 13–18, 2014.
- [4] Whitepaper:Healthcare Disaster Recovery. www.bridgeheadssoftware.com
- [5] R. Wyber, et.al.,Big data in global health: improving health in low- and middle-income countries. www.who.int/bulletin/volumes/93/3/14-139022.pdf
- [6] R. M. Əliquliyev, M. Ş. Hacırəhimova, "Big Data" fenomeni: problemlər və imkanlar, İnformasiya texnologiyaları problemləri, 2014, №2, səh. 3-16.
- [7] The Digital Universe in driving data growth in healthcare. 2014. www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-healthcare.pdf

- [8] Big data: The next frontier for innovation, competition, and productivity. Analyst report, McKinsey Global Institute, May 2011. [www.mckinsey.com/](http://www.mckinsey.com/)
- [9] D. Laney, “3D Data Management: Controlling Data Volume, Velocity and Variety.” Technical report, META Group, Inc (now Gartner, Inc.), February 2001. <http://blogs.gartner.com/>
- [10] M.Ş. Hacırahimova, “Big Data” texnologiyaları və informasiya təhlükəsizliyi problemləri, *İnformasiya texnologiyaları problemləri*, №2, səh. 49-56, 2014.
- [11] Y.N. İmamverdiyev, “Big data texnologiyalarının böyük perspektivləri və problemləri”, *İnformasiya cəmiyyəti problemləri*, №1, səh.23–34, 2016.
- [12] В. Майер-Шенбергер, К. Кукьер, Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. М. 2014. С. 17.
- [13] <http://medcitynews.com/2013/03/the-body-in-bytes-medical-images-as-a-source-of-healthcare-big-data-infographic>
- [14] A.W. Toga, I. D. Dinov, Sharing big biomedical data, *Journal of Big Data*, vol. 2, no. 7, pp. 1-12, 2015.
- [15] О. П. Трифонова, В. А. Ильин, Е. В. Колкер, А. В. Лисица, Большие данные» в биологии и медицине, *Actanaturae*, том 5, № 3 (18) , стр.14-17, 2013.
- [16] M. Coakley, G. Crocetti, P. Dressner, W. Kellum Transforming Telemedicine Through Big Data Analytics, Cornell University Library, 2015
- [17] K. Jee, G.-H. Kim, “Potentiality of Big Data in the Medical Sector: Focus on How to Reshape the Healthcare System”, *Healthcare Informatics Research*, vol. 19, no. 2, pp.79-85, 2013.
- [18] R. Wullianallur, R Viju, “Big Data Analytics in Healthcare: Promise and Potential”, *Health Information Science and Systems*, no. 2. vol. 3. 2014.
- [19] Big data at the speed of business. 2015. <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>.
- [20] O. Tene, J.Polonetsky, Privacy in the age of big data: A time for big decisions, *Stanford Law Review Online*, 2012. <http://www.stanfordlawreview.org/online/privacy-paradox/big-data>
- [21] G. Greenleaf, “Global Data Privacy Laws: 89 Countries, and Accelerating”, *Privacy Laws & Business International Report*, N 115, February 2012; Queen Mary School of Law Legal Studies Research Paper No. 98/2012. <http://ssrn.com/abstract=2000034>
- [22] The Use of Big Data in Public Health Policy and Research, 2014. [http://ec.europa.eu/research/health/pdf/public-health-research\\_en.pdf](http://ec.europa.eu/research/health/pdf/public-health-research_en.pdf)
- [23] M. Herland, T M. Khoshgoftar, R. Wald, “A review of data mining using big data in health informatics”, *Journal of Big Data*, vol. 1, no. 2, pp.1-35, 2014.
- [24] A. Alyass, M. Turcotte, D. Meyre, From big data analysis to personalized medicine for all: challenges and opportunities, *BMC Medical Genomics*, vol. 8, no. 33, pp.1-12, 2015.
- [25] Finance, Government of Canada, Department of. "Budget 2016: Chapter 5 - An Inclusive and Fair Canada". [www.budget.gc.ca](http://www.budget.gc.ca). Retrieved 2016-05-02.
- [26] Fərdi məlumatlar haqqında Azərbaycan Respublikasının Qanunu, 11 may 2010-cu il, [www.president.az](http://www.president.az).

# Технологии Data Mining в Медицине

Рамиз Алыгулиев

Институт Информационных Технологий НАНА, Баку, Азербайджан

r.aliguliyev@gmail.com

**Аннотация**– В статье описаны цели, задачи и этапы обнаружения знаний в базах данных (KDD) и интеллектуального анализа данных (data mining). Исследована роль технологии data mining в анализе медицинских данных и перечислены ее этапы. Указаны проблемы, ограничения и тенденции применения технологии data mining к медицинским данным.

**Ключевые слова**– Knowledge Discovery in Database, data mining, medical data mining.

## I. ВВЕДЕНИЕ

Современные больницы хорошо оснащены информационно-коммуникационными технологиями и другими средствами по сбору данных, которые обеспечивают сравнительно недорогое средство для сбора и хранения данных в меж- и внутри- госпитальных информационных системах. Обширное количество данных, собранных в медицинских базах, требует специальных инструментов для хранения и доступа к ним, а также для анализа и эффективного использования. Такое увеличение объема данных вызывает большие трудности при извлечении полезной информации для поддержки принятия решений из-за того, что традиционные методы не в состоянии обрабатывать такой объем данных. В таком случае методы эффективного компьютерного анализа незаменимы. Чтобы удовлетворить эту потребность, медицинская информатика использует технологии KDD (Knowledge Discovery in Database – обнаружение знаний в базах данных), разработанные в новой междисциплинарной области, охватывающей статистику, распознавание образов, машинное обучение, а также инструменты визуализации для анализа данных и обнаружения закономерностей в сырых данных. Термин KDD впервые был введен в середине 1990-х годов с целью извлечения знаний из базы данных [3].

## II. ОБНАРУЖЕНИЕ ЗНАНИЙ В БАЗАХ ДАННЫХ

“KDD – это нетривиальный процесс выявления действительных, новых, потенциально полезных, в конечном счете, понятных закономерностей в данных”. KDD представляет собой автоматический, поисковый анализ и моделирование больших хранилищ данных.

Процесс KDD обычно состоит из следующих этапов [3, 9–11]:

1) **Понимание предметной области и постановка цели.** Этот шаг готовит почву для понимания того, что должно быть сделано с многочисленными решениями (о преобразовании, алгоритме, представлении и т.д.). Людям, которые отвечают за проект KDD, необходимо понять и

определить цели конечного пользователя и окружающей среды, в которой процесс обнаружения знаний будет происходить.

2) **Выбор и создание набора данных.** Имея определенную цель, должны быть определены данные, которые будут использованы для обнаружения знаний. Этот процесс включает в себя выбор способов поиска, выявление доступных данных и интегрирование всех данных в один набор для обнаружения знаний, включая также те атрибуты, которые будут рассмотрены.

3) **Очистка и предварительная обработка данных.** Этап предназначен для повышения надежности данных. Он включает в себя очистку данных, таких как обработка пропущенных (недостающих) значений, удаление шумов и выбросов.

4) **Преобразование данных.** Вычислительная эффективность является одной из проблем анализа данных. Исследователи находятся под влиянием принципа Оккама, который может быть интерпретирован как “чем проще, тем лучше”. Методы преобразования данных включают уменьшение размерности (например, выбор и извлечение признаков, выборка записей), а также преобразование атрибутов (например, дискретизация численных атрибутов и функциональное преобразование). Этот шаг часто имеет решающее значение для успеха всего процесса KDD, но, как правило, зависит от конкретного случая. Например, в медицинских обследованиях отношение атрибутов часто может быть более важным фактором, чем каждый из них сам по себе. Предложены четыре подхода для уменьшения размерности набора данных. Первый подход заключается в использовании соответствующего источника знаний, таких как онтология или терминология, например SNOMED. Второй подход является субъективным, где эксперты в предметной области выполняют функцию выбора признаков. Третий подход использует алгоритмы интеллектуального анализа данных для оценки прогностической силы каждого атрибута или их комбинации. Последний – четвертый подход использует информационные меры, например, прироста информации для выбора соответствующих признаков. Этот подход измеряет важность каждой переменной по отношению к конкретной цели, но он не может обнаружить эффект комбинации подмножества переменных.

## 5) Data mining

5.1) **Выбор подходящей задачи data mining.** Принимается решение, какую задачу data mining использовать - классификацию, регрессию или кластеризацию? Это во многом зависит от целей KDD, а также от предыдущих этапов. Data mining предназначена