

**А. Д. КУБЕГЕНОВА^{1*}, А. И. ТАКУАДИНА², О. И. КРИВОРОТЬКО³,
Ж. Т. НУРУШЕВА⁴**

¹«Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті»
КеАҚ, Орал қ.

²«Қарағанды медициналық университеті» КеАҚ, Қарағанды қ.

³«Новосібір мемлекеттік университеті», Ресей, Новосібір қ.

⁴БҚО әкімдігі денсаулық сақтау басқармасының «ЖИТС-тің алды алу және оған қарсы күрес жөніндегі облыстық орталық» ШЖҚ МКК, Орал қ.
E-mail: aigul-03@mail.ru, alyoka.01@mail.ru, krivorotko.olya@mail.ru,
zkoaid.info@gmail.com

БАТЫС ҚАЗАҚСТАН ОБЛЫСЫНДАҒЫ АИТВ-ИНФЕКЦИЯСЫНЫҢ ЭПИДЕМИОЛОГИЯЛЫҚ ЖАҒДАЙЫН БОЛЖАУДАҒЫ ИНТЕЛЛЕКТУАЛДЫ ТАЛДАУ ТЕХНОЛОГИЯСЫ

Мақалада адамның иммун тапшылығы вирусын (АИТВ инфекциясы) бақылау мен таралуын математикалық модельдеуді дамытудың негізгі бағыттары, эпидемиялық процесінің динамикалық және құрылымдық классификациялары қолданылған.

Әлемдегі және Қазақстан Республикасы бойынша қол жетімді мәліметтер мен міндеттерге байланысты вирустың таралуы сипатталған, Батыс Қазақстан облысы әкімдігінің денсаулық сақтау басқармасының ЖИТС-тің алдын алу және оған қарсы күрес жөніндегі облыстық орталықтан, Батыс Қазақстан бойынша 2010-2020 жылғы статистикалық мәліметтер алынып АИТВ жұқтырғандардың ретроспективті, кластерлік талдаулар жүргізілді.

Сырқаттанушының кестесінің сипаттамасы берілген, тәуекелділігі анықталған, сырқаттанушылықтың статистикалық болжамдары тексерілді.

Ғылыми және практикалық зерттеулерде эпидемияның критерийлері мен бақылауының таралуы Ақпараттық технологиялар тарапынан қарастырылды. Болжау жүйесін және математикалық модельдің қолданыстағы зияткерлік жүйелерін қолдана отырып мысалдар келтірілді.

Маңызды бағыттардың бірі АИТВ індетінің салдарын болжау, бір организмнің ішіндегі иммунологиялық динамиканы зерттеу, халық деңгейінде эпидемиялық жағдайдың дамуын болжау мәселесі экономика мен әлеуметтік саласындағы жағдайлар қарастырылды.

АИТВ жұқтырғандарды өңдеу және жас ерекшеліктері бойынша жағдайды талдау үшін Data Mining классификация әдістері пайдаланылды.

Батыс Қазақстан облысының жас ерекшеліктері бойынша АИТВ жұқтырғандар туралы 10 жылдық кезеңінің сырқаттанушылығын болжау Statistica қолданбалы бағдарламалар пакетінің көмегімен жүргізіледі.

Түйін сөздер: ретроспективті талдау, кластерлік талдау, евклид арақашықтығы, Data Mining, Statistica advanced.

Кіріспе. Қоғамдық денсаулық сақтаудағы өзекті мәселелердің бірі АҚТҚ-жұқпасы проблемасы болып табылады, ол жедел шараларды талап ететін кең ауқымды әлеуметтік, медициналық, экономикалық салдарға әкеп соғады. АИТВ – ең алдымен жас ұрпақты, репродуктивті және еңбекке қабілетті жастағы адамдарды зақымдайтын қауіпті инфекциялық ауру.

* E-mail корреспондирующего автора: aigul-03@mail.ru

ЮНЭЙДС сарапшыларының бағалауы бойынша (2021), әлемде АИТВ жұқтырған 37,6 [30,2 - 45,0] млн. адам тұрады, 770 мың адам ЖИТС-пен байланысты аурулардан қайтыс болды. БҰҰ-ның АИТВ/ЖИТС бойынша біріккен бағдарламасының (ЮНЭЙДС) мәліметтері бойынша, қазіргі уақытта АИТВ-инфекциясының пандемиясы тұрақтанды, бірақ өте жоғары деңгейде.

2019-2020 жылдардағы жаңа коронавирустық инфекцияның пандемиясы адамзатқа барлық мемлекеттердің бұрын-соңды болмаған шараларымен жеңілетін міндеттер қойды, АИТВ инфекцияның таралу проблемасы өзекті болып қала береді, өйткені емдеу саласындағы табыстарға қарамастан әлі шешілмеген.

Қазақстанда ЖИТС-тың алдын алу және оған қарсы күрес жөніндегі республикалық орталықтың деректері бойынша 30.09.2020 ж. Өсу қорытындысымен АИТВ-инфекциясының 27 100 жағдайы тіркелді, оның ішінде ерлер – 16344, әйелдер – 10756, балалар – 146. Сондай-ақ, елімізде АИТВ-сы бар әйелдерден туған 4464 бала тіркелген[1].

Қаралып отырған мәселенің өзектілігі БҰҰ-ның 2016-2021 жылдарға арналған АИТВ/ЖИТС жөніндегі біріккен бағдарламасының (ЮНЭЙДС) жаңа стратегиясы: 2030 жылға қарай әлемде ЖҚТБ індетін тоқтату міндеттемесін өзіне алғандығына байланысты. Бұл Қазақстан Республикасының Денсаулық сақтау жүйесін дамытудың 2016-2019 жылдарға арналған «Денсаулық» мемлекеттік бағдарламасында көрініс тапты, бұл Қазақстанда АИТВ-инфекциясы эпидемиясының сипатын, әсіресе жұқтыру қаупі жоғары халықтың негізгі топтарында зерделеу қажеттілігін тудырды.

Қоғамда АҚТҚ жұқтырғандар мен ЖҚТБ-мен ауыратындар санының көп және жыл сайын ұлғайып отыруының болуы кез келген мемлекет үшін демографиялық, экономикалық және әлеуметтік проблема болып табылады [2-3]. Бағалау мәліметтері бойынша, әлемде АҚТҚ/ЖҚТБ-мен өмір сүретін адамдардың саны 46 миллионнан асады. Бұл аурудың жаһандық індеті қазірдің өзінде 30 миллионнан астам адамның өмірін қиды. Жыл сайын әлемде 5 миллион адам жұқтырады, 3 миллионнан астам адам ересектер мен балалар өмірден кетеді. Жастар АИТВ инфекциясына аса осал болып келеді: дамушы елдерде АИТВ жұқтырудың 65% фактісі анықталды 15-тен 24 жасқа дейінгі жастарға келеді.[4-6]

Бұл мәселенің өзектілігі Қазақстанда, әсіресе жұқтыру қаупі жоғары халық топтарында АИТВ-инфекциясы эпидемиясының сипатын зерттеу қажеттілігінен туындады. Эпидемиологиялық аурудың дамуын болдырмау үшін популяциядағы ауруды алдынала анықтауға мүмкіндік беретін терең талдау әдістері қолданылып, статистикалық өңдеуді енгізу қажеттілігі туындап отыр [7-8]. Осыған байланысты, медициналық мәліметтерге байланысты зерттеулер жүргізу, адамдардың бүкіл тобындағы аурудың сипатын зерттеу әдістердің интеграциясы және әмбебап тәсіл арқылы анықталуы керек.

Компьютерлік және ақпараттық технологиялардың, сондай-ақ сақтау технологиясының қарқынды дамуымен көптеген деректерді сақтауға болады. Деректерді іздеу технологиясы көптеген мәліметтерден ықтимал, құнды білімді іздеп және шығара алады. Деректер қоры технологиясы – бұл мәліметтер базасын басқаратын бағдарламалық жасақтама туралы ғылым [9-11]. Деректерді іздеу дегеніміз – мәліметтер шаблонын іздеу процесі, яғни көптеген толық емес, анық

емес, кездейсоқ мәліметтерден алынған мәліметтермен жұмыс жасау [12-13]. Деректерді іздеу – бұл мәліметтер базасы мен жасанды интеллект саласындағы өте белсенді зерттеу саласы. Статистикалық және оңтайландыру алгоритмдерінің жиынтығы 30% салыстырмалы дәлдікпен параметрлерді сәйкестендіруді көрсетеді [14-15].

Нәтижелерді Денсаулық сақтау ұйымдары модельдеу деректерін тарихи деректермен салыстыру арқылы белгілі бір аймақтағы жұқпалы аурулардың эпидемиясын болжау үшін қолдана алады.

Зерттеу мақсаты. Медициналық ақпаратты талдау үшін статистика әдістерін қолдану қазіргі уақытта Қазақстанда жеткілікті кең таралмаған, сондықтан жүргізілген зерттеулердің мақсаты Data Mining технологиясы бойынша эпидемиологиялық жағдайды талдау, болжау және алдын ала анықтау болды.

Қазақстан Республикасы және Батыс Қазақстан облысының жас ерекшеліктері бойынша АИТВ жұқтырғандар туралы 10 жылдық кезеңінің ретроспективті, кластерлік талдау жүргізу.

Материал және әдістер. Зерттеу объектісі ретінде Батыс Қазақстан облысының жас ерекшеліктері бойынша АИТВ жұқтырғандар туралы 10 жылдық кезеңінің (2010-2020 жж.) деректері таңдалды.

Халықтың аурушандығы бойынша деректерді жіктеу Data Mining технологиясының үлкен деректерін талдау арқылы жүргізілді. Деректерді талдау құралы ретінде Statistica қолданбалы бағдарламалар пакеті қолданылды: StatisticaBase, StatisticaAdvanced, Data Mining деректерін өндіруге арналған құралдар. Кластерлеудің жаңа әдістері толық байланыс әдісіне негізделген графикалық формалардың көмегімен талдау жасауға мүмкіндік берді. Графикалық формаларды қолдану арқылы деректерді кластерлеу талдау уақытын қысқартуға, сондай-ақ сықаттанушылықты болжау алгоритмін әзірлеуге мүмкіндік берді. Кластерлік талдауды деректерге қолданудың практикалық маңыздылығы мен өзектілігі күмән тудырмайды, өйткені қазіргі ақпараттық қоғамда деректер мен оларды талдау нәтижелері болған сайын үлкен рөл атқарады, ал кластерлеу бұл деректерді жақсы түсінуге мүмкіндік береді.

Эпидемиологиялық, статистикалық әдістерді пайдалана отырып, еліміздегі ЖҚТБ індетінің таралуын тоқтату жөніндегі жаһандық мақсаттарға қол жеткізу қажеттілігін ескере отырып, Батыс-Қазақстан облысындағы АИТВ-инфекциясы бойынша эпидемиологиялық жағдайды бағалау, АИТВ жұқтырғандардың жыныстық тиістілігі бойынша ретроспективті, кластерлік талдау жасау.

Нәтижелер мен талқылау. Эксперименттік мәліметтерді өңдеу компьютерде статистикалық пакеттерде жүргізілді.

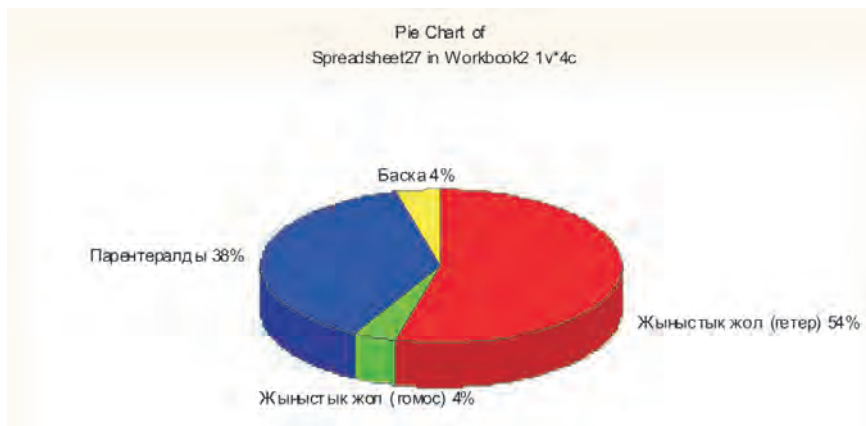
Қазақстан халқының негізгі топтары арасында АИТВ-инфекциясының жағдайларына жүргізілген мониторингтің 2010-2020 жылдарының деректері, АИТВ-ның таралуын эпидемиологиялық қадағалау нәтижелері елде індетті тиімді бақылау үшін ғана емес, сондай-ақ 2030 жылға қарай халықтың денсаулығына қатер ретінде АИТВ-инфекциясы эпидемиясының өсу қарқынын айтарлықтай қысқарту үшін барлық алғышарттардың бар екенін көрсетті.

2020 жылғы қыркүйекке арналған «Қазақстанда АИТВ» есебіне сәйкес АИТВ берілуінің негізгі жолдары (1-сурет) [1]:

– Ең жиі кездесетіні – вирустың жыныстық жолмен берілуі: ол 57,9% құрайды. Бұл жағдайда гетеросексуалдық жолмен 54,3%, ал гомосексуалдық жолмен тек 3,6% беріледі.

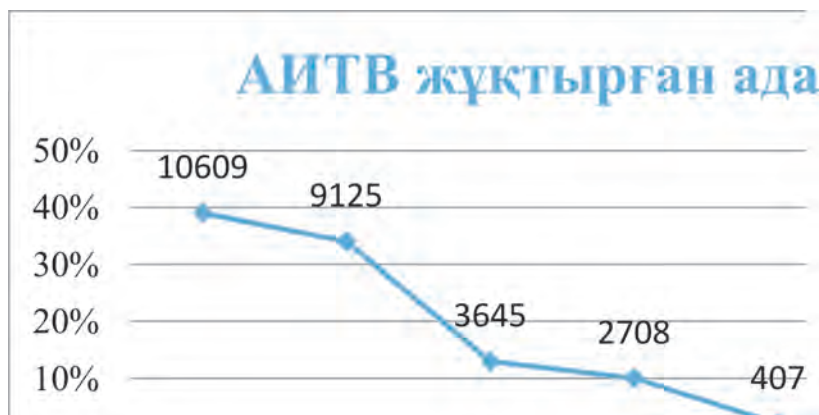
– АИТВ-ның екінші негізгі берілу жолы – парентеральды (яғни, инъекциялық есірткіні қолдану кезінде инелер мен шприцтер арқылы): 38,1% жағдай.

– Басқа және белгіленбеген беру жолдары – 4%



Сурет 1 – Қазақстан Республикасында АИТВ-инфекциясының берілу жолдары.

АИТВ жұқтырған адамдардың әлеуметтік мәртебесінің мәліметі бойынша жұмыс жасайтын АИТВ жұқтырған адам – 33,4% ,жұмыссыздар – 48,8 %, сотталғандар және тергеумен қамалғандар 12,6 % құрайды. АИТВ жұқтырған адамдардың жас ерекшелігі бойынша талдайтын болсақ ең көп кездесіп тұрғаны 30-39 жас аралығында 39% (10 609 адам), 40-49 жас аралығында 34% (9 125 адам), 50-59 жас аралығында 13% (3 645 адам), 20-29 жас аралығында 10% (2 708 адам), 60 жастан жоғары 2% (407 адам), 0-14 жас аралығында 1% (334 адам), 15-19 жас аралығында 1% (272 адам) 2 суретте көрсетілген.



Сурет 2 – Қазақстан Республикасының жас ерекшелігі бойынша АИТВ жұқтырған адамдар

2020 жылғы қыркүйекке арналған «Қазақстанда АИТВ» есебіне сәйкес АИТВ берілуінің негізгі жолдары.

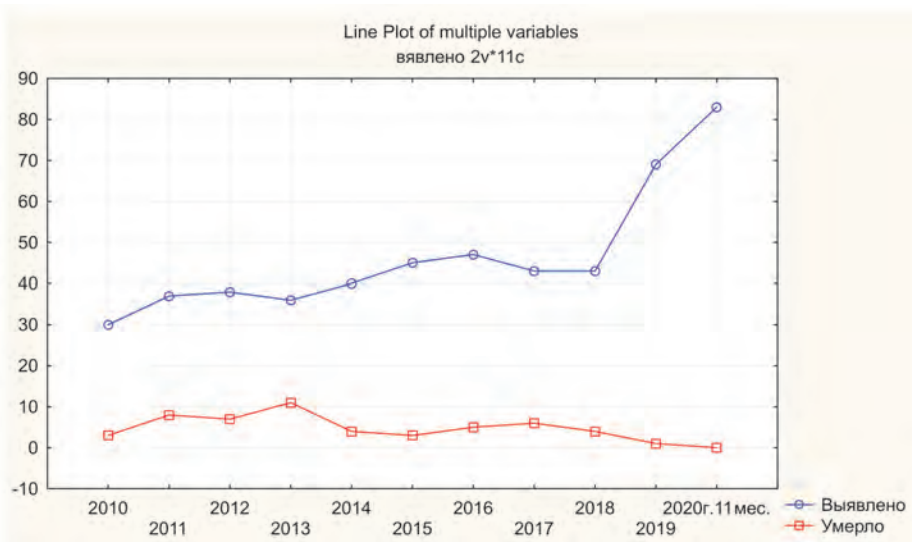
Ақпараттық технологиялар тұрғысынан эпидемияның пайда болуын, таралуын және бақылау критерийлерін зерттеу маңызды ғылыми және практикалық зерттеулердің бірі болып табылады. АИТВ індетінің алдын алу мәселесін шешу үшін, эпидемияның өршуінің уақытында болжау ошағы үлкен рөл атқарады. Болжау сөзі – грек сөзінен алынған болашақта объектінің ықтималдығын, жай-күйі туралы және (немесе) оларды жүзеге асырудың балама жолдары мен мерзімдері туралы ғылыми негізделген пайымдауды білдіреді [3].

Болжаудың негізгі әдістеріне мыналар жатады:

- модельдеу әдістері;
- сараптамалық бағалау;
- статистикалық әдістер;
- интуитивті (маман бұрын болжаудың белгілі бір түріне ғылыми әдістерді қолданған және техникалық құралдарды қолданбай орындай алатын кезде) [4].

Осылайша, болжау жүйесін құру немесе математикалық модельдерді қолдануға негізделген қолданыстағы интеллектуалды жүйелерді қолдану маңызды болып келеді. Алынған мәліметтерді және оларды талдауды ескере отырып, нақты аймақтағы нақты жағдайды неғұрлым дәл сипаттайтын үлгіні таңдау.

2010-2020 жылдар аралығында Батыс Қазақстан облысы бойынша АИТВ ауруының анықталғаны жәнеде қайтыс болған туралы сызықтық графикте көрсетілген (3 сурет).



Сурет 3 – АИТВ ауруының анықталғаны жәнеде қайтыс болған туралы сызықтық график

Аурудың сызықтық кестесі (сурет 3.) АҚТҚ – жұқпасы (анықталған және қайтыс болғандар саны) Батыс Қазақстан халқы бойынша 10 жылдық кезеңнің

жиынтық деректерін ескере отырып құрылған (2010-2020 жж.). Абсцисса осі бойынша АИТВ жұқтырғандарды зерттеу жылдары, координаталар осі бойынша – АИТВ жұқтырғандардың абсолюттік саны қойылған. Бұл сызықтық график 2011-2013 жылдар аралығында анықталған тұрақты үрдісті көрсетеді. 2014 жылдан бастап анықталған аурулардың нәтижесі екі есеге көбеюіне байланысты нашарлайды. 2019-2020 жылдары зерттеудің алғашқы жылдарымен салыстырғанда тұрғындардың АИТВ ауруы бірнеше есе артып, өзіндік шыңына жетеді. Бұл көрсеткіштің жоғарылау ересек жастағы тұрғындар және осал топтағы-есірткі тұтынушылар, сотталғандар арасында АИТВ-инфекциясының таралуымен және АИТ -инфекциясының жыныстық трансмиссиясының орын алуымен түсіндіріледі.

Осылайша, Батыс Қазақстанда АИТВ-мен сырқаттанушылықтың көпжылдық серпінін бағалау кезінде 2013-2020 жылдар кезеңінде жылдам көтерілу және 2012-2013 жылдар, 2016-2018 жылдар уақыт аралығында құлдырау анықталады.

Вертикалды жолының көрсеткіші 1,3%-ға дейін төмендеуі бұл тенденция абсолютті дегенді білдірмейді, өйткені нәтижелердің жақсаруынан кейін біртіндеп нашарлайды. АИТВ жұқтырғандардың сырқаттанушылық деңгейі бойынша желілік кестені талдау негізінде жыл бойынша үш топқа бөліп көрсетуге болады:

- 1) (2011-2013) – орташа көтерілу жылдары;
- 2) (2010,2014-2020) – жоғары көтерілу жылдары;
- 3) (2012-2013,2016-2018) құлдырау жылдары және аралық жылдар (2013, 2016, 2017).

Сызықтық графикада БҚО бойынша АИТВ-дан қайтыс болған деректердің нәтижелері 2010 - 2013 жылдар аралығындағы өлім-жітімнің жылдам көтерілуін көрсетеді.

2013-2015 жылдар аралығында қайтыс болғандардың саны төмендегені, осы жылдары анықталған науқастардың саны жоғарылауы байқалады себебі олар есепке алынып ем алғандар болып келеді.

2016-2017 жылдары деректер тұрақтанбады, инъекциялық есірткіні тұтынушылар арасында АИТВ-инфекциясының таралуы, сондай-ақ жыныстық жолмен жұғатындар көбейіп келеді. Алайда, 2019-2020 жылдарға қарай көрсеткіштің аздап төмендегенін байқаймыз.Бұл көрсеткіш бір жағынан жаңа короновирустық инфекция пандемиясы кезеңінде ақпарат жинау жүйесінің нашарлауымен, ал екінші жағынан оның өлім түріндегі салдарымен түсіндіріледі.

Variable	Descriptive Statistics (выявлено)									
	Valid N	% Valid obs.	Mean	Median	Mode	Frequency of Mode	Sum	Minimum	Maximum	Std.Dev.
Выявлено	11	100,0000	46,45455	43,00000	43,00000	2	511,0000	30,00000	83,00000	15,66119
Умерло	11	100,0000	4,72727	4,00000	Multiple	2	52,0000	0,00000	11,00000	3,16515

Сурет 4 – АИТВ жұқтырғандар нәтижелерінің сипаттамалық статистикасы бар электрондық кесте

Бақыланатын айнымалының іріктемелі орташа мәні формула бойынша анықталады(1):

$$\bar{x} = \frac{i \sum_{i=1}^n x_i}{n} \quad (1)$$

мұндағы n -үлгінің көлемі (айнымалы x бақылауларының нақты саны).

Медиана-тапсырыс берілген мәндер қатарын жоғарыда да, төменде де осы мәндердің тең санымен жартысына бөледі.

Мода – бұл мәліметтер жиынтығында жиі кездесетін мән.

Іріктемелі дисперсия айнымалының өзгергіштігін сипаттайды, формула бойынша есептеледі(2):

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (2)$$

мұндағы \bar{x} – үлгінің орташа мәнін білдіреді.

Дисперсия 0-ден шексіздікке дейін өзгереді. Тұрақты айнымалылар – 0-дің экстремалды мәні өзгергіштіктің болмауын білдіреді.

Бастапқы деректер файлында Батыс Қазақстан обылысының жас ерекшеліктері бойынша АИТВ жұқтырғандар туралы ақпарат бар. Осы кластерлік талдаудың мақсаты кластерлерге бөлу және тәуекел топтарын анықтау үшін тиісті кластерді анықтау болып табылады. Бұл мәселені шешу үшін кластерлік талдауды қолдану негізгі тиімді және жаппай қолданылатын әдістердің бірі болып саналады.

Жақындық шарасы ретінде Евклид арақашықтығын (Евклидово расстояние) пайдалана отырып, кластерлік талдаудың иерархиялық рәсімінің көмегімен өңірдің жас ерекшеліктері бойынша жіктеуді жүргіземіз, ал кластерлерді біріктіру үшін – толық байланыс әдісін (метод полной связи) әдістерін қолданамыз. Осы әдістердің көмегімен екі кластерді бір-бірімен байланыстыруға болады.

Кез-келген екі кластер бір-біріне жақындағанда, олар бір-біріне жақындай бастайды және байланыс қашықтығы мен ерекшеленеді. Тиісінше, біріктірілген кластерлер кездейсоқ қалған бөліктерден бөлек элементтерге айналады. Бұл құбылыс нысандарды бір-бірімен байланыстырады және кластерлерді құрайды. Алынған кластерлер ұзын тізбектермен ұсынылған. Кластерлердің табиғи санын анықтау аймақтарды кластерлерге біріктіру арқылы жүргізілді.

Жас ерекшеліктері бойынша АИТВ ауруын жұқтырған адамдарды кластерлерге біріктіру тәртібі иерархиялық ағашта 5 суретте көрсетілген.

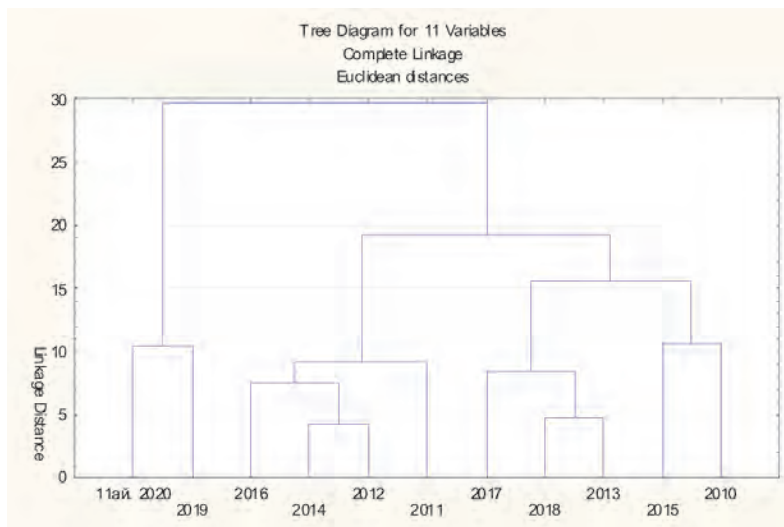
Бөлуді бағалау үшін төмендегі формула бойынша қарастыруға болады:

$$\pi = \frac{a_i + a_j}{2b_{ij}} \quad (3) \quad S = \frac{1}{k} \sum_{i=1}^g \max \pi_{ij} \quad (4)$$

мұндағы a_i, a_j – кластардың орта кластерлік арақашықтығы; $i: j; b_{ij}$ – сол кластерлік арасындағы орташа кластерлік қашықтық болып келеді.

Егер объектілер арасындағы барлық айырмашылықтар бір-біріне тең болса, онда бұл жағдайда олар 1-ге тең болады. Жоғарыда сипатталған алгоритмнің көмегімен

алынған бөлімдер 1-ге тең немесе артық емес болып келеді. Нәтижесінде барлық нысандар бір кластерге біріктірілген кезде олар 1-ге тең болады деп айта аламыз.



Сурет 5 – 2010-2020 жыл аралығындағы, жас ерекшелігі бойынша АИТВ ауруын жұқтырған адамдардың классификациясы.

Толық байланыс әдісі. Толық байланыс әдісі (ағылш. Complete-linkage clustering) – иерархиялық кластерлеу алгоритмдерінің бірі. Бастапқыда үлгінің әр элементі жеке кластер болып саналады. Содан кейін кластерлер барлық элементтер бір кластерге енгенше дәйекті түрде біріктіріледі. Алгоритмнің әр қадамында екі кластер біріктіріледі, олардың арасындағы қашықтық минималды болып келеді.

Объектілер жиынтығы бөлінген кластерлердің табиғи санын анықтау үшін иерархиялық кластерлеудің әр деңгейінде жиынтықты осы класстың санына бөлінді. Кластерлердің әр жұбымен олардың бір-бірімен ішкі байланысы бағаланды. Және де әр кластер үшін орташа кластерлік қашықтық есептелді.

Байланыстылықты бағалау кезінде орташа кластерішілік қашықтықтың кластераралық қашықтыққа қатынасы қолданылады:

Дендрограммада объектілер кластерлерге біріктірілетін қашықтықтар (шартты бірліктерде) көлденеңінен белгіленеді.

Көлденең ось бақылауларды білдіреді, тік-біріктіру қашықтығын білдіреді.

Жас ерекшелігі бойынша АИТВ ауруын жұқтырған адамдардың классификациясының алғашқы қадамы келесі кластерлерден тұрады:

- 1) 2019 – 2020 (11 ай) жылдардың көрсеткішінен;
- 2) 2016, 2014, 2012, 2011 жылдардың көрсеткішінен;
- 3) 2017, 2018, 2013, 2015, 2010 жылдардың көрсеткішінен құралды.

Бұдан әрі бір кластерге (2016, 2014, 2012, 2011) жылдардың көрсеткіштерімен (2017, 2018, 2013, 2015, 2010) жылдардың көрсеткіштері кластерлер біріктіріледі. Процесс барлық нысандарды бір кластерге біріктірумен аяқталады. Сонымен, дендограмм бойынша, бұл жағдайда үш кластерге бөлуге болады.

Case ID	Cluster Means (Возраст)		
	Cluster No. 1	Cluster No. 2	Cluster No. 3
0-14 жас	0,50000	0,40000	1,00000
15-19 жас	1,50000	0,60000	0,00000
20-29 жас	17,50000	7,80000	17,50000
30-39 жас	34,50000	18,00000	12,50000
40-49 жас	15,50000	6,00000	6,75000
50 жас+	6,50000	6,60000	2,75000

Сурет 6 – Cluster Means электрондық кестесі

Алынған кластерлердің ерекшеліктерін талдап, жас ерекшелігі бойынша АИТВ ауруын жұқтырған адамдардың көрсеткіштерінің орташа мәнін салыстыра отырып, біз келесі нәтижелерге қол жеткіздік:

Variable	Euclidean distances (Возраст)										
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	11ай. 2020
2010	0,0	11,4	7,5	6,0	7,7	10,6	12,8	10,9	9,6	23,1	28,3
2011	11,4	0,0	9,2	16,1	7,1	15,0	8,4	19,2	17,7	23,2	29,8
2012	7,5	9,2	0,0	11,6	4,2	12,0	7,5	13,7	13,7	21,7	26,1
2013	6,0	16,1	11,6	0,0	11,4	10,0	16,0	7,4	4,8	21,9	26,0
2014	7,7	7,1	4,2	11,4	0,0	12,4	5,4	13,2	13,1	21,5	26,1
2015	10,6	15,0	12,0	10,0	12,4	0,0	16,6	15,6	8,8	13,4	20,0
2016	12,8	8,4	7,5	16,0	5,4	16,6	0,0	15,9	17,3	23,2	27,0
2017	10,9	19,2	13,7	7,4	13,2	15,6	15,9	0,0	8,5	24,8	26,5
2018	9,6	17,7	13,7	4,8	13,1	8,8	17,3	8,5	0,0	18,5	22,3
2019	23,1	23,2	21,7	21,9	21,5	13,4	23,2	24,8	18,5	0,0	10,5
11ай. 2020	28,3	29,8	26,1	26,0	26,1	20,0	27,0	26,5	22,3	10,5	0,0

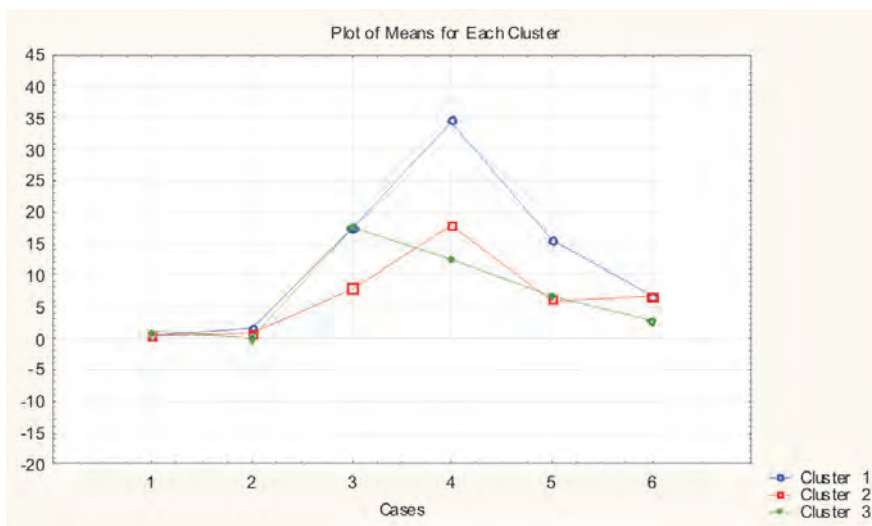
Сурет 7 – Евклидово арақашықтығы электрондық кестесі

Бірінші кластер тұтастай алғанда 20-29, 30-39, 40-49 жас аралығындағы ерекшелігі бойынша АИТВ ауруын жұқтырған адамдар ересек тұрғындар мен халықтың осал топтары-есірткі тұтынушылар, сотталғандар арасында АИТВ инфекциясының таралуының орташа деңгейімен ерекшеленеді және АИТВ инфекциясының жыныстық трансмиссиясының үлес салмағын алады;

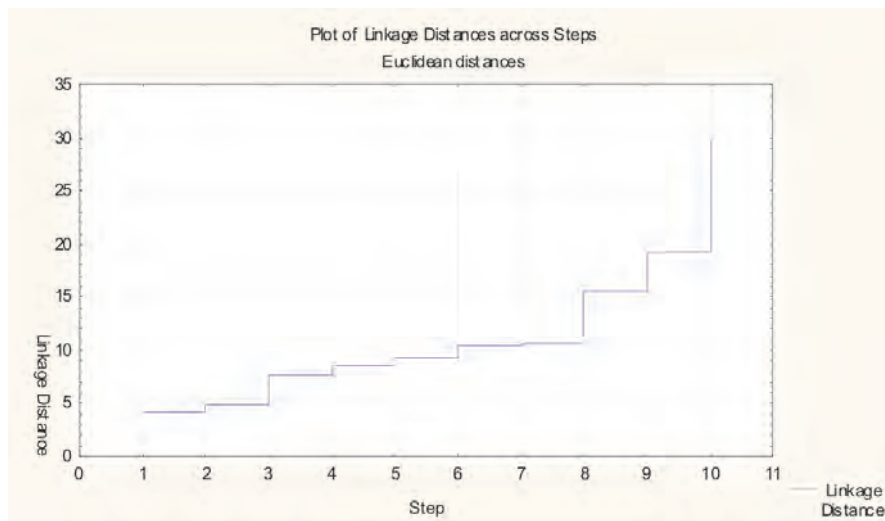
Екінші кластер үшін 1 кластермен салыстырғанда АИТВ-жұқпасының төмен көрсеткіші тән. Екінші кластерге 20-29, 30-39, жас ерекшелігі бойынша АИТВ ауруын жұқтырған адамдар тәуекелі жоғары болып келеді, алайда алкоголизмнен, нашақорлықтан және басқа да әлеуметтік аурулардан зардап шегетін адамдар тобы кіреді;

Үшінші кластерде 0-14,15-19 жас аралығында және 50 жастан жоғары жас ерекшелігі бойынша АИТВ ауруын жұқтырған адамдар инъекциялық есірткіні тұтынушылар арасында, сондай-ақ ауру анандан ұрыққа жыныстық жолмен және құрсаққа жұғатын АИТВ-инфекциясының таралуы мен берілу жолы көрсетілген.

8-суретте жас ерекшелігі бойынша АИТВ ауруын жұқтырған адамдардың үш тобына қатысты айтарлықтай айырмашылықтарды атап өтуге болады.



Сурет 8 – Әр кластер үшін орташа мәндер графигі.



Сурет 9 – Қадам бойынша біріктіру схемасының графигі.
(Евклидово расстояние)

Ағаш тәрізді (древовидный) кластерлеу нәтижелері. Диаграммада көлденең ось бойынша қадамдар, тігінен – қашықтықты бейнелейді (Сурет 9.) Барлық нысандарды бір кластерге біріктіру үшін алгоритмге 11 қадам қажет болды. Бастапқы популяция үшін үшінші қадамда байқалды, ал максималды секіру сегізінші қадамнан кейін байқалады, осылайша 3 кластерге бөлінуі анықталды.

Жас ерекшелігі бойынша АИТВ ауруын жұқтырған адамдардың классификация саңы бойынша кластерге біріктіріліп жәнеде кластерлердін жоғары өсуін анықтады.

Жас ерекшелігі бойынша біртекті топтарға біріктіру шешу жолымен алынған статистикалық болжамдау нәтижелері инъекциялық есірткіні тұтынушылар (ИЕТ) сырқаттанушылықтың болжаушылары болып табылатынын көрсетті. Data mining көмегімен өңдеу халықтың осы тобы АИТВ-инфекциясы індетінің өсуін ынталандыруды жалғастырып отырғанын көрсетті. Жыныстық жолмен берілетін инфекциялар құрылымында маңызды рөл атқаратын коинфекциялар үлесінің артуы үлкен алаңдаушылық тудырады.

Қорытынды. Қазақстанда 2030 жылға қарай ЮНЕЙДС-тің «90-90-90» жеделдетілген стратегиясының мақсаттарына қол жеткізу үшін елде АИТВ инфекциясының эпидемиялық процесінің дамуындағы қазіргі үрдістерді, берілу жолдары мен инфекцияның ықтимал таралуын айқындайтын, қауіп факторларын (эпидемиологиялық, әлеуметтік-экономикалық, мәдени, мінез-құлық және басқалар) ескеру мақсатында АИТВ-инфекциясының эпидемиялық процестері, динамикалық және құрылымдық модельдер қарастырылды.

Эпидемиологиялық, статистикалық әдістерді пайдалана отырып, еліміздегі ЖҚТБ індетінің таралуын тоқтату мақсатында Батыс Қазақстан облысындағы АИТВ-инфекциясы бойынша эпидемиологиялық жағдайдың 2010-2020 жылғы статистикалық мәліметтер алынып АИТВ жұқтырғандардың жыныстық тиістілігі бойынша ретроспективті, кластерлік талдау жасалды. Талдауда нәтижесі АИТВ жұқтырғандар ең көп және жыл сайын саңы өсіп келген және сол жастағы өлім-жітімнің жоғары деңгейі көрсетіліп, ерлер арасында 30-39 жас аралығындағы адамдар байқалды. Қазақстан есебіне сәйкес, АИТВ-инфекциясының берілуінің негізгі жолдары, өңірлерінде АИТВ-мен өмір сүретін адамдардың саны графикалық түрде қарастырылды.

Қолданыстағы интеллектуалды жүйелерді қолдану маңызды болғандықтан алынған мәліметтерді талдап, болжау жүйесі құрылып, жақындық шарасы ретінде Евклидово арақашықтығын (Евклидово расстояние) пайдалана отырып, кластерлік талдаудың иерархиялық рәсімінің көмегінің қажеттілігі анықталды.

Батыс-Қазақстан облысындағы АИТВ-инфекциясы бойынша эпидемиологиялық жағдайдың 10 жылдық кезеңде (2010-2020 жж.) АИТВ бойынша сырқаттанушылыққа жүргізілген талдау АИТВ жұқтырғандардың күрт өсуін және сырқаттанушылықтың тұрақты үрдісін анықтауға мүмкіндік берді;

Кластерлік жіктеу алгоритмі объектілер арасындағы ішкі байланысты анықтады және АИТВ эпидемиологиясының математикалық моделінің дұрыстығын көрсетті;

Data mining көмегімен өңдеу Батыс-Қазақстан облысындағы АИТВ-инфекциясы бойынша сырқаттанушылықтың тұрақты өсуін көрсетті;

Кластерлеу және алынған кластерлердің сипаты олардың жиынтығынан Батыс Қазақстан облысындағы АИТВ-инфекциясымен күресте нақты антиретровирустық препараттарды және терапевтік схемаларды модельдеу, оңтайландыру және таңдау бойынша арнайы деректерді қалыптастыруға мүмкіндік береді.

ӘДЕБИЕТ

- 1 Проект ОФ «Центр научно-практических инициатив» (30-сентябрь 2020г.)
<https://amanbol.kz/vich/nauka/tpost/ryp0vd2yh1-vich-v-kazahstane-aktualnie-dannie-na-ko>
- 2 2019 Глобальный мониторинг эпидемии СПИДа (2020). //– Руководство ЮНЭЙДС, 2020. – С.15-21.
- 3 И.С. Светульников, С.Г. Светульников. Методы социально-экономического прогнозирования. Том 1. Теория и методология. — Москва: Юрайт, 2015. — 351 с. — ISBN 978-5-9916-4903-2, 978-5-9916-4905-6.
- 4 А.И. Такуадина. Обзор существующих математических моделей как эффективных методов прогнозирования распространения эпидемии. V Международная научно-практическая конференция «Европа и тюркский мир: наука, техника и технологии» г. Анкара, Турция, 6-8 май 2020г. Том 1. С.108-115.
- 5 Романюха А.А., Носова Е.А. Модель распространения ВИЧ-инфекции в результате социальной дезадаптации // Управление большими системами, 2011. Выпуск 34. М.: ИПУ РАН. С. 227–253.
- 6 Карач А.С., Романюха А.А. Современные подходы к анализу и прогнозированию здоровья населения с помощью математических моделей. Математические методы прогнозирования. 2014, №1 с.38-47.
- 7 Xinyi Wang. The Role of Data Mining Technology in Advertising Marketing. J. Phys.: Conf. Ser. 1744 042202, 2021. DOI: 10.1088/1742-6596/1744/4/042202
- 8 Kabanikhin, S.I., Krivorotko, O.I. Mathematical Modeling of the Wuhan COVID-2019 Epidemic and Inverse Problems. Comput.Math.and Math. Phys. 60, 1889–1899 (2020).<https://doi.org/10.1134/S0965542520110068>
- 9 Kabanikhin S., Krivorotko O., Takuadina A., Andornaya D., Zhang. Sh. Geo-information system of tuberculosis spread based on inversion and prediction. Journal of Inverse and Ill-posed Problems, 2021, vol. 29, no. 1, pp. 65-79, doi.org/10.1515/jiip-2020-0022.
- 10 May R., Anderson R.M. Transmission dynamics of HIV infection. Nature, 1987, vol. 326, pp. 137–142 [doi:10.1038/326137A0](https://doi.org/10.1038/326137A0)
- 11 Roeger L.V., Feng Z., Castillo-Chaves.C. Modelling TB and HIV Co-infections. Mathematical biosciences and engineering, 2009, vol.6, no.4, pp.815-837, [doi:10.3934/mbe.2009.6.815](https://doi.org/10.3934/mbe.2009.6.815)
- 12 S. Kabanikhin, O. Krivorotko and D. Yermolenko, «A parameter identification problem for the mathematical model of HIV dynamics,» 2017 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Novosibirsk, 2017, pp. 82-86, [doi: 10.1109/SIBIRCON.2017.8109842](https://doi.org/10.1109/SIBIRCON.2017.8109842).
- 13 Banks H.Th., Kabanikhin. S.I., Krivorotko O.I, Yermolenko D.V., A numerical algorithm for constructing an individual mathematical model of HIV dynamics at cellular level. [Journal of Inverse and ILL-Posed Problems]. 2018, no.26(6). Pp.859-873.
- 14 Bouchnita A., Bocharov G., Meyerhans A., Volpert V. Towards a multiscale model of acute HIV infection. Computation, 2017, vol.5, no.1(6). pp.1-22, [doi:10.3390/5010006](https://doi.org/10.3390/5010006)
- 15 Omondi. E.O., A mathematical modelling study of HIV infection in two heterosexual age groups in Kenya. Infectious Disease Modelling, 2019, vol. 4, pp. 83-98, [doi:10.1016/j.idm.2019.04.003](https://doi.org/10.1016/j.idm.2019.04.003).

REFERENCES

- 1 Project OF (2020) «Center for Scientific and Practical Initiatives»
<https://amanbol.kz/vich/nauka/tpost/ryp0vd2yh1-vich-v-kazahstane-aktualnie-dannie-na-ko>
(accessed 30 September 2020)

2 2019 Global Monitoring of the AIDS Epidemic (2020). // – UNAIDS Handbook, 2020. – pp.15-21.

3 I.S.Svetunkov, S.G.Svetunkov. Methods of socio-economic forecasting. Volume 1. Theory and methodology. – Moscow: Yurayt, 2015. - 351 p. – ISBN 978-5-9916-4903-2, 978-5-9916-4905-6.

4 A.I.Takuadina. Review of existing mathematical models as effective methods for predicting the spread of the epidemic. V International Scientific and Practical Conference "Europe and the Turkic World: Science, technology and technology" Ankara, Turkey, 6-8 May 2020. Volume 1. pp.108-115.

5 Romaniukha A.A., Nosova E.A. A model of the spread of HIV infection as a result of social maladaptation // Management of large systems, 2011. Issue 34. Moscow: IPU RAS. pp. 227-253.

6 Karach A.S., Romaniukha A.A., Modern approaches to the analysis and prediction of public health using mathematical models. Mathematical methods of forecasting. 2014, No. 1, pp.38-47.

7 Xinyi Wang. The Role of Data Mining Technology in Advertising Marketing. J. Phys.: Conf. Ser. 1744 042202, 2021. DOI: 10.1088/1742-6596/1744/4/042202

8 Kabanikhin, S.I., Krivorotko, O.I. Mathematical Modeling of the Wuhan COVID-2019 Epidemic and Inverse Problems. Comput.Math.and Math. Phys. 60, 1889–1899 (2020).<https://doi.org/10.1134/S0965542520110068>

9 Kabanikhin S., Krivorotko O., Takuadina A., Andornaya D., Zhang. Sh. Geo-information system of tuberculosis spread based on inversion and prediction. Journal of Inverse and Ill-posed Problems, 2021, vol. 29, no. 1, pp. 65-79, doi.org/10.1515/jiip-2020-0022.

10 May R., Anderson R.M. Transmission dynamics of HIV infection. Nature, 1987, vol. 326, pp. 137–142 doi:10.1038/326137A0

11 Roeger L.V., Feng Z., Castillo-Chaves.C. Modelling TB and HIV Co-infections. Mathematical biosciences and engineering, 2009, vol.6, no.4, pp.815-837, doi:10.3934/mbe.2009.6.815

12 S. Kabanikhin, O. Krivorotko and D. Yermolenko, «A parameter identification problem for the mathematical model of HIV dynamics.» 2017 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Novosibirsk, 2017, pp. 82-86, doi: 10.1109/SIBIRCON.2017.8109842.

13 Banks H.Th., Kabanikhin. S.I., Krivorotko O.I., Yermolenko D.V., A numerical algorithm for constructing an individual mathematical model of HIV dynamics at cellular level. [Journal of Inverse and ILL-Posed Problems]. 2018, no.26(6). Pp.859-873.

14 Bouchnita A., Bocharov G., Meyerhans A., Volpert V. Towards a multiscale model of acute HIV infection. Computation, 2017, vol.5, no.1(6). pp.1-22, doi:10.3390/5010006

15 Omondi E.O., A mathematical modelling study of HIV infection in two heterosexual age groups in Kenya. Infectious Disease Modelling, 2019, vol. 4, pp. 83-98, doi:10.1016/j.idm.2019.04.003.

***А. Д. КУБЕГЕНОВА¹, А. И. ТАКУАДИНА², О. И. КРИВОРОТЬКО³,
Ж. Т. НУРУШЕВА⁴***

*¹НАО «Западно-Казахстанский аграрно-технический университет
имени Жангир хана», Уральск, Казахстан*

²НАО «Медицинский Университет Караганды», Караганда, Казахстан

³«Новосибирский государственный университет», Новосибирск, Россия

*⁴ГКП на ПВХ «Областной центр по профилактике и борьбе со СПИД» Управления
здравоохранения акимата Западно-Казахстанской области, Уральск, Казахстан*

*E-mail: aigul-03@mail.ru, alyoka.01@mail.ru, krivorotko.olya@mail.ru
zkoaid.info@gmail.com*

ТЕХНОЛОГИЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА В ПРОГНОЗИРОВАНИИ ЭПИДЕМИОЛОГИЧЕСКОЙ СИТУАЦИИ ПО ВИЧ-ИНФЕКЦИИ В ЗАПАДНО-КАЗАХСТАНСКОЙ ОБЛАСТИ

Рассмотрены основные направления развития математического моделирования контроля и распространения вируса иммунодефицита человека (ВИЧ-инфекции), динамическая и структурная классификации эпидемического процесса. Из областного центра по профилактике и борьбе со СПИД Управления здравоохранения акимата Западно-Казakhstanской области взяты статистические данные по Западно-Казakhstanской области за период 2010-2020 годы, с описанием распространения вируса в зависимости от имеющихся данных и задач в мире и по Республике Казахстан, проведен ретроспективный, кластерный анализ ВИЧ-инфицированных.

Дана характеристика графика заболеваемости, выявлен риск, проверены статистические прогнозы заболеваемости.

В научных и практических исследованиях распространённость критериев и наблюдений за эпидемией рассматривалась со стороны информационных технологий. Приведены примеры с использованием системы прогнозирования и существующих интеллектуальных систем математической модели. Одним из важных направлений является прогнозирование последствий эпидемии ВИЧ, изучение иммунологической динамики внутри одного организма, проблема прогнозирования развития эпидемической ситуации на уровне населения.

Для обработки ВИЧ-инфицированных и анализа ситуации по возрастным особенностям были использованы методы классификации Data Mining.

Прогноз заболеваемости Западно-Казakhstanской области по возрастной группе ВИЧ-инфицированных за 10-летний период проводилось с помощью пакета прикладных программ Statistica.

Ключевые слова: ретроспективный анализ, кластерный анализ, евклидово расстояние, Data Mining, Statistica advanced.

**A. D. KUBEGENOVA¹, A. I. TAKUADINA², O. I. KRIVOROTKO³,
ZH. T. NURUSHEVA⁴**

¹Non-profit JSC «West Kazakhstan Agrarian and Technical University named after Zhangir Khan» Uralsk, Kazakhstan

²Non-profit JSC «Karaganda Medical University», Karaganda, Kazakhstan

³«Novosibirsk State University», Novosibirsk, Russia

⁴SME on the REM «Regional Center for the Prevention and Control of AIDS» of the Health Department of the Akimat of the West Kazakhstan region, Uralsk, Kazakhstan

E-mail: aigul-03@mail.ru, alyoka.01@mail.ru, krivorotko.olya@mail.ru
zkoaid.info@gmail.com

THE TECHNOLOGY OF INTELLIGENT ANALYSIS IN PREDICTING THE EPIDEMIOLOGICAL SITUATION OF HIV INFECTION IN THE WEST KAZAKHSTAN REGION

The article uses the main directions of development of mathematical modeling of control and spread of human immunodeficiency virus (HIV infection), dynamic and structural classification of the epidemic process. Statistical data on the West Kazakhstan region for the period 2010-2020 were taken from the regional center for the prevention and control of AIDS of the health department of the Akimat of the West Kazakhstan region, with a description of the spread of the virus, depending on the available data and tasks

in the world and in the Republic of Kazakhstan, a retrospective, cluster analysis of HIV-infected people was carried out.

The characteristics of the morbidity schedule are given, the risk is identified, statistical forecasts of morbidity are checked.

In scientific and practical studies, the prevalence of criteria and observations of the epidemic was considered from the information technology side. Examples are given using a forecasting system and existing intelligent mathematical model systems. One of the important directions is the prediction of the consequences of the HIV epidemic, the study of immunological dynamics within one organism, the problem of predicting the development of the epidemic situation at the population level.

Data Mining classification methods were used to process HIV-infected people and analyze the situation by age characteristics.

The prognosis of the morbidity of the West Kazakhstan region by the age group of HIV-infected people over a 10-year period was carried out using the Statistica application software package.

Key words: *retrospective analysis, cluster analysis, Euclidean distance, Data Mining, Statistica advanced.*