

МРНТИ 20.51.19
УДК 631.362.36:57.087.3

<https://doi.org/10.51889/2021-3.1728-7901.16>

М.Е. Мансурова¹, В.Б. Барахнин², Г.А. Тюлепбердинова^{1*}, Ф.Р. Гусманова¹, А.А. Нұраханова¹
¹Ал-Фараби атындағы Қазақ Ұлттық Университеті, Алматы, Қазақстан
²Новосибирск мемлекеттік университеті, Новосибирск, Ресей Федерациясы
*e-mail: tyulepberdinova@mail.ru

ДЕНСАУЛЫҚ JAҒДАЙЫНЫҢ ЖІКТЕМЕСІН ЖАСАНДЫ ИНТЕЛЛЕКТ ӘДІСТЕРІМЕН САЛЫСТЫРМАЛЫ ТАЛДАУ

Аңдатпа

Ұсынылып отырған мақалада адам денсаулығына қатысты маңыздылығын жоғалтпайтын мәселелердің бірі қарастырылады. Бүгінде ағзалардың әртүрлі функцияларының жұмысын бақылайтын, халықаралық тұрғыдағы жіктемеге сәйкес бағдарламалардың заман талабына сай келмеуі, цифрлық денсаулыққа қатысты жасанды интеллект алгоритмдерінің негізіндегі профиль жасаудың маңызды екені баршаға мәлім. Мақалада халықаралық жіктемемен жұмыс істеудің және оны қолдануға мүмкіндік беретін жасанды интеллект алгоритмдері салыстырылады. Бұл мақалада жіктеме көрсеткіштерін кластарға бөлуге ыңғайлы классификациялық моделдерді дайындау мақсатында Adaboost алгоритмі, XGбуст алгоритмі және SVM алгоритмдеріне салыстырулар жүргізілді.

Мақалада деректерді өңдеудің негізгі әдістерін зерттеу, қарастырып отырған мәліметтер базасынан адам денсаулығының жағдайына байланысты статистика және машиналық оқытудың, ақпаратты іздеу әдістерінің жиынтығы ұсынылған.

Мәліметтерді талдау мәселелері мен алгоритмдері зерттеу және оларды адам денсаулығының жағдайы мәселелерін шешуге қолдану қарастылды.

Түйін сөздер: жасанды интеллект, машиналық оқыту, ақпаратты іздеу, сандық денсаулық профилі, жұмыс істеудің халықаралық жіктемесі, Тензорфло, Байес классификаторы, шешімдер ағашының әдіснамасы, к-жақын көршілер классификациясы, Adaboost алгоритмі, XGбуст алгоритмі, SVM алгоритмі.

Аннотация

М.Е. Мансурова¹, В.Б. Барахнин², Г.А. Тюлепбердинова^{1*}, Ф.Р. Гусманова¹, А.А. Нұраханова¹
¹Казахский национальный университет имени Аль-Фараби, Алматы, Қазақстан
²Новосибирский государственный университет, г. Новосибирск, Российская Федерация

СРАВНИТЕЛЬНЫЙ АНАЛИЗ КЛАССИФИКАЦИИ СОСТОЯНИЯ ЗДОРОВЬЯ МЕТОДАМИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В данной статье рассматривается одна из проблем, которая не теряет своей значимости для характеристики состояния здоровья человека. То, что в настоящее время еще не разработано, приложение для работы с алгоритмами искусственного интеллекта с помощью определителей международной классификации функционирования нескольких функций организма, говорит о важности этого проекта по созданию цифрового профиля здоровья. В статье мы изучим Международную классификацию функционирования и алгоритмы искусственного интеллекта, позволяющие ее использовать. Будем рассматривать работу по классификации полученных данных по классам по определителям международной классификации функционирования с использованием алгоритмов искусственного интеллекта и сравнение моделей прогнозов и дальнейшую оптимизацию для создания наиболее подходящей классификационной модели.

В статье представлен комплекс методов исследования основных методов обработки данных, статистики и машинного обучения, поиска информации по состоянию здоровья человека из рассматриваемой базы данных.

Рассмотрены проблемы и алгоритмы анализа данных и их применение для решения проблем состояния здоровья человека.

Ключевые слова: искусственный интеллект, машинное обучение, поиск информации, цифровой профиль здоровья, международная классификация функционирования, Тензорфло, Байесовский классификатор, методология дерева решений, К-классификация ближайших соседей, алгоритм Adaboost, алгоритм Xgбуста, алгоритм SVM.

Abstract

M. E. Mansurova¹, V.B.Barakhnin², G. A. Tyulepberdinova¹, F. R. Gusmanova¹, A. A. Nurakhanova¹

¹al-Farabi Kazakh National university, Almaty, Kazakhstan

²Novosibirsk State University, Novosibirsk, Russian federation

**COMPARATIVE ANALYSIS OF THE CLASSIFICATION OF HEALTH STATUS
BY ARTIFICIAL INTELLIGENCE METHODS**

This article considers one of the problems that does not lose its significance for the characteristics of the state of human health. The fact that an application for working with artificial intelligence algorithms using the determinants of the international classification of the functioning of several body functions has not yet been developed indicates the importance of this project to create a digital health profile. In this article, we will study the International Classification of Functioning and artificial intelligence algorithms that allow it to be used. We will consider the work on classifying the obtained data by classes according to the determinants of the international classification of functioning using artificial intelligence algorithms, comparing forecast models and further optimization to create the most suitable classification model. The article presents the study of the main methods of data processing, statistics related to the state of human health from the database under consideration, and a set of machine learning methods, information search methods. It is planned to study the problems and algorithms of data analysis and their application to solve problems of the state of human health.

Keywords: artificial intelligence, machine learning, information retrieval, digital health profile, international classification of functioning, Tensorflow, Bayes classifier, decision tree methodology, K-close neighborhood classification, Adaboost algorithm, Xgbust algorithm, SVM algorithm.

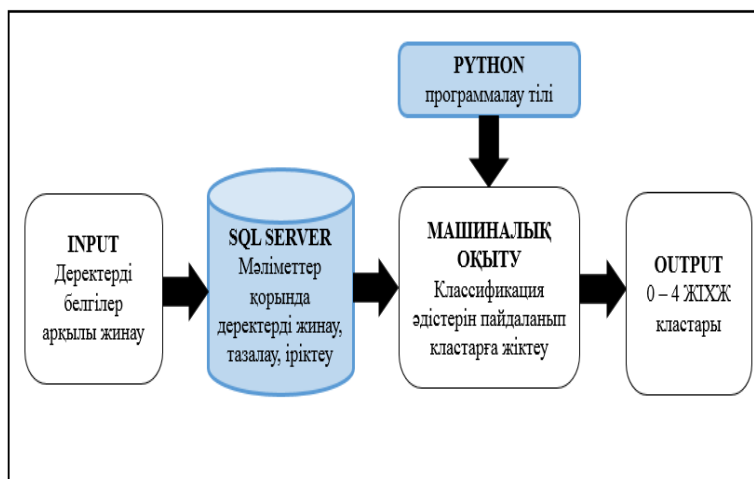
Кіріспе

Бүгінде қаншама ағза функцияларының жұмыс жасауындағы халықаралық тұрғыдағы жіктемесі анықтауыштар бойынша жасанды интеллектің алгоритмдері арқылы жұмыс істейтін бағдарламалар немесе бағдарламалық қамтамалар осы күнге дейін жасалынбаған, ал ол цифрлық денсаулыққа қатысты профиль жасауда маңызды мәселелердің бірі. Ғылыми жұмыстарда адамның денсаулығына қатысты жіктеменің толықтай нұсқасын жасау кезінде халықаралық тұрғыдағы жіктемесін таңдап алдық. Халықаралық тұрғыдағы жіктеме санаттары мүмкіндігімен денсаулықтың жағдайлары бойынша толық ақпарат, яғни жіктемелерге қол жеткізе аламыз.

Адам өзінің денсаулығының нашарлағанына көңіл бөлмеуіне байланысты, ағзасында белгісіз аурулардың асқынып кету мүмкіндігі бәрімізге мәлім. Әркім денсаулығы жайындағы ақпаратқа, қандай дәрежеде немесе емдік іс-әрекеттердің қандай дәрежеде әсер етіп жатқандығы бойынша мәліметтермен таныс болып отыруы медицина бойынша болып жататын әлемдегі қиындықтарды аз да болсын азайтуға көмегін тигізетіндігіне сенімдіміз. Онымен қоса денсаулыққа байланысты автоматты жіктеу дәрігерлерге де, науқастарға да оңтайлы және тиімді болады. Өйткені дәрігердің адамның денсаулығын аз уақытта жіктеуіне, саралай алуына және оған қатысты диагнозды қойып, алдымен әртүрлі іс-шараларды уақытымен бастап кетуіне болар еді. Ауырған адамдар өзінің денсаулық жағдайлары жайында жылдам мәліметтерді қажет кезінде ала отырып, денсаулығының жағдайы бойынша науқас жаңадан пайда болған кезінде дер кезінде іс-әрекет жасауына болар еді.

Зерттеудің әдістемелері

Бұл мақаланың басты тапсырмасы халықаралық тұрғыдағы жіктемесі бойынша зерттеу, жасанды интеллект әдістері бойынша жіктеулерді қарастыру болмақ. Жасанды интеллект бойынша мұндай бөлулер әртүрлі әдістермен ұйымдастырылады. Мысалы, машиналық оқыту - ол оқуға мүмкін алгоритмдерді құру тәсілдеріне зерттеу жасайтын жасанды интеллектің жан-жақты қамтылған бір бөлімі десек болады. Мұндағы негізгі қағида, есептің қағаз бетіне шығарылған нұсқасына бағдарлама құру ғана емес, сонымен қатар «ақпараттық үйрену» әдісін пайдалану болып табылады. 1-сурет бойынша қарасаңыз, енгізілетін ақпарат пен шығатын ақпарат арасындағы байланысты көресіз. Деректерді алдымен студенттердің өздері жинақтайды, халықаралық тұрғыдағы жіктемесімен әртүрлі функцияға қажетті белгілері бойынша сұрыпталады, содан кейін әртүрлі белгілері бойынша сандық түрге ие болады. Осыдан соң ақпараттар деректер қорына жіберіледі. Мұнда дұрыстау, тазалау, бөлу сияқты жұмыстар жүзеге асырылған болатын. Белгілі бір түрге келген мәліметтер машиналық оқыту бойынша іріктелінген модельге сүйеніп кластар бойынша жіктеу жұмысы іске асты. Кейіннен халықаралық тұрғыдағы жіктемесімен жіктелген 0-4 кластары, бес класс алынғанын көресіз.



Сурет 1. Машиналық оқыту бойынша Халықаралық тұрғыдағы жіктемесін жіктеу алгоритмін ұйымдастыру суреті

Ақпаратты жинау бөлімі машиналық оқыту жұмыстарындағы үлкен әрі маңызды істің бірі. Өйткені, 80% жұмыс модель құруға кетпейді, ол ақпаратты жинау және тазалау жұмыстарына кетеді. Бізге қажетті ақпаратты табу, біз білетін ақпаратты толықтыру, қажет ақпаратты табу, деректер жүйесімен интеграция, ақпаратты жіберу, керек түрге келтіру сияқты іс әрекеттер орындалады. Мәліметтердің қол жетімді, көп өлшемді, үйлесімді, толық, дәл, дәйекті және нақты онымен қоса өзекті болуы қажет. Жарамсыз ақпараттың сапасы нашар немесе пайдасы болмай кетуі ықтимал және ол бір жағынан жарамды болғанына қарамастан, қате тұжырымдардың туындауына әкеліп соғуы мүмкін. Зерттелетін жұмыстарда халықаралық тұрғыдағы жіктемесі бойынша функциялар ағзаның бөлімдері санаттарының белгілері бойынша жиналған. Жиналған ақпарат медицина бойынша негізделе келе, қолмен жинақталынған болатын. Алғашқы классификация бойынша: ағзаның функция бөлігі → жүрек және тамыр, иммундық, қан, тыныс алу сияқты жүйелердің функцияларының бөлімі → иммундық жүйе мен қан жүйесінің функциялары бойынша → қан жүйесі бойынша функциялар санаты.

Қан жүйесі бойынша функциялар класын анықтау үшін қажетті белгілерді зерханалық деректерден табуға болады. Халықаралық тұрғыдағы жіктеме бойынша алынатын функциялар қан жүйесі бойынша алынатын функциялармен байланысты. Өйткені, болжам бойынша қан құрамында болатын элементтер әртүрлі басқа жұмыс жасайтын функцияларға әсерін тигізеді. Мысал келтіретін болсақ, ол элементтер тыныс алу сияқты функциялармен байланыста болуы әбден мүмкін. Жіктеу бойынша алгоритмдерді жасап болғаннан кейін корреляцияның бар екендігін немесе оның мүлдем болмайтындығын көруге болады.

Әртүрлі белгілері бойынша қан жүйесі функциясы кластарын анықтайтын анағұрлым көбірек ақпарат алуға мүмкіндігіміз болады. Төмендегі кесте бойынша қалыпты облыстан ауытқып кеткен Халықаралық тұрғыдағы жіктеменің пайыз бойынша мөлшері есептелінді, ол 0 және 4 арасы пайызда қате кетуді жоқ деп айтуға келетін мөлшері, ал 5 және 24 пайыз арасы ол жеңіл тұрғыда қате кету, ал 25 және 49 пайыз ол орташа қате кетуді байқау мөлшері, ал 50 және 95 пайыз арасы ол ауыр қате кету мөлшерін көрсетеді, онымен қоса 96 және 100 пайыздық мөлшер абсолютті қате кетуді сипаттайды және осы белгілер бойынша қорытындылайтын болсақ 0 және 4 класс деңгей мөлшері қойылған еді. Мұндай белгілер 830 сынамалы объектке алынады, 800 алгоритм нысанда үйретуге жіберілетін болса, 30 алгоритм нысаны дұрыс жұмыс жасайтын бағалау мақсаты бойынша тест жасау үшін жинақталынған. Халықаралық тұрғыдағы жіктеме бойынша қан жүйесін жіктеу адамның денсаулығын бағалауда және инфекция немесе анемия, дәл сол сияқты лейкемия ауруларын тауып алдын алу шараларын ұйымдастыру үшін қажет болады. Ал келесі классификация бойынша: ағзаның функция бөлігі → урогениталды немесе репродуктивті функцияларының бөлімі → функция ол зәр шығару бойынша → b610 зәр шығару функциясы бойынша санат.

Ол зәр шығарудың функциялары класстарын анықтайтын қажетті белгілер сияқты зертханалық мәліметтерді қарастыруға болады.

Ал зәр шығарудың функциялары оң жұмыс жасауын қадағалау үшін неше түрлі несеп талдауларын көрсеткіштерімен қоса жүгіну қажет [1].

Зәр шығарудың функциялары жұмыстарын бағалау үшін неше түрлі зертханалық мәліметтерді қан жүйесі бойынша жұмыс жасау функциясын тексергенде 0 және 4 арасында класқа бөліп қарастырдық. Жалпы алғанда 830 сынама әртүрлі нысан бойынша алынды, ал олардың 800-дейі алгоритмді оған үйрету бойынша жіберілді, ал 30 нысаны алгоритмдердің оң жұмыс жасауын бақылау мақсатында тест жасауға жинақталынған болатын.

Келесі үшінші классификация бөлімі үшін: ағзаның функциялары бөлімінің → жүрек және тамыр, иммундық және қан, онымен қоса тыныс алудың жүйелері функцияларының бөлімі → тыныс алудың функциясы → ол *тыныс алудың функциялары бойынша* санат.

Ал тыныс алу функцияларының класстарын анықтау үшін қажетті белгілер, эксперимент нәтижесінде белгілі болады.

Адам тамырының соғуы жылдамдығы 60 соққыны көрсетсе 1/мин, онда тыныс алудың жиілік мөлшері сәйкес мән бойынша, демін шығарған кезінен кейін автоматты үзіліс түріне 4 максималды кідіріске 120 с болатын болса онда науқаста өкпесінің альвеолаларында көмірқышқылының газы мөлшерлемесі 6,5 пайызға жетеді, ал оны тыныс алу функциялары дұрыс жұмыс жасауының көрінісі ретінде қарастыруға болады [2]. Мұндай көрсеткіш көбінесе су спорт түрлерімен айналысқан адамдарда және жеңіл атлетикамен айналысатын адамдарда кездесуі ықтимал. Мұндай көрсеткіштерді көрсету үшін көбінесе уақыт бойынша бақылап іс-әрекеттер жүргізеді [3]. Төмендегі 1-кестеде тыныс алу функциясының белгілері оның бірлігі және қалыпты аумақтарымен берілген. Қаншалықты Халықаралық тұрғыдағы жіктеме бойынша қате кетуі, кластар бойынша анықтау мүмкін. 2-кестеде тыныс алудың функцияларының белгілері бойынша 10 нысанның кіріс және шығыс мәндерінің үлгісі келтірілген.

Кесте 1. Тыныс алу функцияларының белгілері

№	Белгілер	Бірлігі	Қалыпты ауқымы
1	Жынысы	1 – ер 0 – әйел	
2	Тыныс алу жиілігі	Минутына тыныс алу саны	Ер – 12 – 15 Әйел – 14 – 19
3	Пульс жиілігі	соққы / минут	Ер – 60 – 70 Әйел – 65 – 80
4	Автоматты үзіліс	Саны	4
5	Бақылау кідіруі	Секунд	40 – 60
6	Ерікті кідіру	Секунд	40 – 60
7	Максималды кідіру	Секунд	80 – 120
8	Тыныс алу тереңдігі	•	1 – 1,5
9	Ерік индексі	•	1

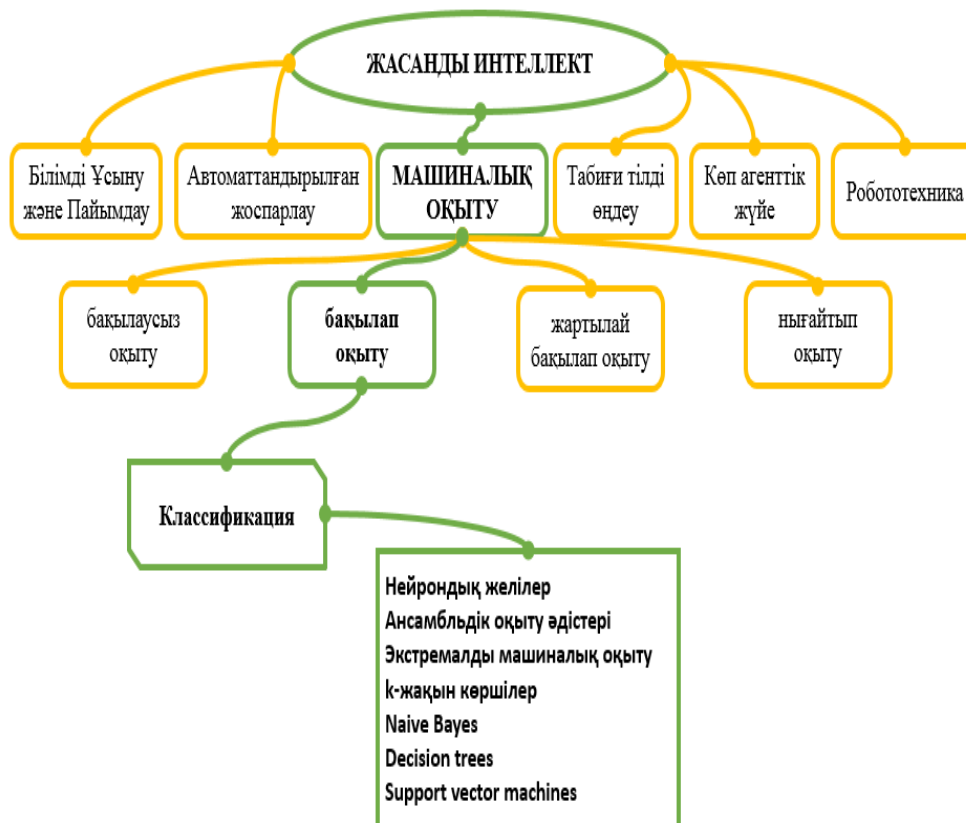
Кесте 2. Тыныс алу функцияларының белгілері бойынша кіріс және шығыс мәндерінің үлгісі

Кіріс мәндері									Шығыс мәндері
Жынысы	Тыныс алу жиілігі	Пульс жиілігі	Автоматты үзіліс	Бақылау кідіруі	Ерікті кідіру	Максималды кідіру	Тыныс алу тереңдігі	Ерік индексі	
0	18	67	4	40	32	72	1,5	0,8	1
0	18	75	4	48	48	96	1,25	1	0
0	16	71	4	50	49	99	1,2	0,98	0
1	14	65	4	58	59	117	1,034	1,017	0
1	13	64	4	63	64	127	0,952	1,015	0
1	27	91	2	15	12	27	4	0,8	4
0	22	75	3	52	38	90	1,153	0,730	2
0	27	101	2	24	11	35	2,5	0,458	3
0	29	98	2	26	11	37	2,30	0,423	3
1	13	60	4	45	38	83	1,33	0,844	1

Тыныс алу функцияларының жұмысын бағалауда әртүрлі эксперименттік көрсеткіштері жоғары. Ал келесі төртінші классификация бойынша тоқталатын болсақ: ағзада функция бөлігі → ас қорыту, метаболика, эндокриндік және басқа функциялар бөлімі → ас қорыту жүйесі бойынша функциялар → салмақты сақтау бойынша функциялардың санаты.

Студенттердің дамуы бойынша деректер салмақ өзгерісі кластарын анықтау үшін қажет белгілер ретінде қарастырылған.

Біздің машиналық оқытудың келесі процесі бөлігі – ол классификация моделдерін оңтайлы анықтау. Халықаралық тұрғыдағы жіктеме бойынша анықтауыштарын сәйкесінше анықтағаннан соң көп класты моделді таңдау болып табылады. Көп класты жіктеу төмендегі 2-суретте келтірілген моделдер біреусін қолдана отырып ұйымдастыруға арналған.



Сурет 2. Моделдерді жіктеу

Тиімді классификация моделін жасауда оларды салыстыру мен кейіннен тиімдісін таңдау үшін ерекшеліктерін қарастыра отырып іс әрекетке көшу керек. Қажетті моделді таңдағанда көп кедергілер де болады. Моделдер есептеу кезінде уақыттың көп кетуі және дәлдігінің аздығы немесе ауытқу мөлшерінің көбеюі ықтимал деп біз де гипотеза жасадық [4].

Тензорфло дегеніміз ол google бойынша терең оқу үшін қосымшалар қажеттігі үшін жасалынған ашық түрдегі кітапхана. Ал ол онымен қоса машиналық оқытуды да қолдайды. Тензорфло ол мәліметтерді тензор атымен айтатын үлкен көлемді өлшемі бар массив деп қарастыратын болады. Үлкен көлемді массив деректерді түрлендіру кезінде қолайлы. Ағза функцияларына әсер ететіндіктен және оның жұмысын бағалауда кейбір ерекшеліктері қарастырылатындықтан мәліметтердің үлкен көлемді массивтері пайдаланылады. Байес классификаторы дегеніміз постериоримен мүмкіндігінің барынша жоғары түріне жоспарланған жіктеудің алгоритмдерінің үлкен сыныбы. Жіктелінген әрбір объектке сыныптың мүмкіндігі жоғары функциялары таңдалынады және оны сыныптардың келесі мүмкіндіктерімен қарастыратын болады.

Белгілеп отырған нысана постериоридің мүмкіндігі жоғары болатын сыныбына қатысты. Шешімдер ағашы әдіснамасы дегеніміз, ол бірнеше ковариаттарға арналған классификация жүйелерін жасауға және мақсатпен белгісіздерді болжау алгоритмдерін жасауға негізделген

мәліметтерді енгізудің көп қолданылып жүрген тәсілі. Мұндай тәсілде популяция тамырдың түйіні бойынша және ішкі түйіндер арқылы жапырақтың түйіндері төңкеріліп ағаштың бұтақ тәрізді етіп жасайтын сегменттеріне жіктеледі. Ал алгоритм ол параметрлік емес немесе күрделі, параметрлік тәрізді құрылымдарды таңдамай-ақ үлкен немесе күрделі ақпараттар жиынын оңтайлы жүргізе алады. k-жақын көршілер әдісінің классификациясы анағұрлы қарапайым және әртүрлі белгілері бойынша олардың жақын көршілер сыныбына қатысты жіктеліп отырады. Көп жағдайда жан-жағындағы көршілерін санап отыру пайдалы, себебі мұндай тәсіл көп жағдайда жақын көршілердің сұраныс жіктелуі деп те аталады.

Мұндағы сыныпты анықтау жағдайында жақын көршілер әдісін пайдаланамыз. Мұндай классификациялар екі кезектен, біріншісі ол жақын көршілер анықтамасы, ал екіншісі ол - көршілерді қолданатын сыныптың анықтамасы [5]. SVM дегеніміз ол регрессия мен жіктеулер үшін пайдаланылатын бақылауда оқытылатын тәсілдерінің жиынтығы. Бұлар жалпы алғанда сызықты классификацияның тобына жатады. SVM қасиеттерінің ерекшелігі ол бір мезгілде эмпирикалық классификация қателіктерін азайтып отырады, ал геометриялық шекті ол барынша жоғарғы мәнге жеткізеді. SVM-ның енгізілетін векторын барынша ең үлкен бөлетін гиперпланның салынылған максималды өлшемі кеңістікте бейнеленеді. Гиперплан жан жағында деректерді бөлуге екі параллель болатын гиперпландар салынған. Ал бөлінетін гиперплан ол екі параллель болатын гиперпландар арасында қашықтығы барынша үлкен мәнге жететін гиперплан. Параллель гиперпландар шекарасын және қашықтық бойынша қаншалықты көп болса, онда ол жіктеуіштің жалпы ерекшелігі соншалықты максималды болмақ [6, 7].

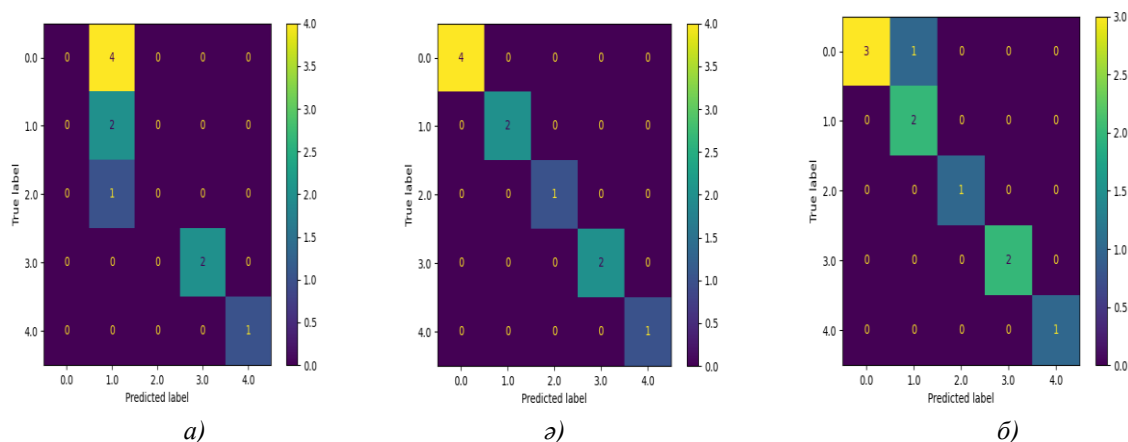
Келесі алгоритм ол – Adaboost, оқу мәліметтері негізінде салмақ жиынтығын өзгертпей және әрбір әлсіз оқудағы циклдан соң бұларды түрге келтіріп жөндеулер арқылы нашар үйретілген жиынтығын жасап отырады.

Зерттеу нәтижелері

Халықаралық тұрғыдағы жіктемесімен топтастырылған деректер ағзадағы функциялар негізінде айрықша классификация жасалынады. Халықаралық тұрғыдағы жіктемесінің өзіндік ерекшелігі бойынша функция басқаша жіктелгеннен кейін, олардың арасында байланыс болуын анықтау керек. Мысал келтіретін болсақ, қан жүйесі функциясының тыныс алу функцияларына қандай дәрежеде әсер етуі мүмкін болатынын анықтау керек. Түрлі болжамдар бойынша гемоглобин көрсеткіштері аз нысандар тыныс алу функцияларының жұмыс жасауына кедергілер жасап жеңіл және орташа қате кетулер болуы мүмкін.

Жинақталған мәліметтерді жоғары жақта көрсетіп өткеніміздей тізімде келтірілген жіктеулер бойынша тексереміз. 1-ші және 2-ші кестелерде көрсетілген деректер бойынша жіктеу алгоритмдері жүргізілді. Басқа да функциялар үшін жіктеу алгоритмдерін таңдау алгоритмдері жүргізілді.

Confusion матрицалары - тыныс алу функциялары жұмыстарын бақылаудағы мәндерін төмендегі 3-ші және 4-ші суреттерде бейнеленген салыстырулардан көруге болады.

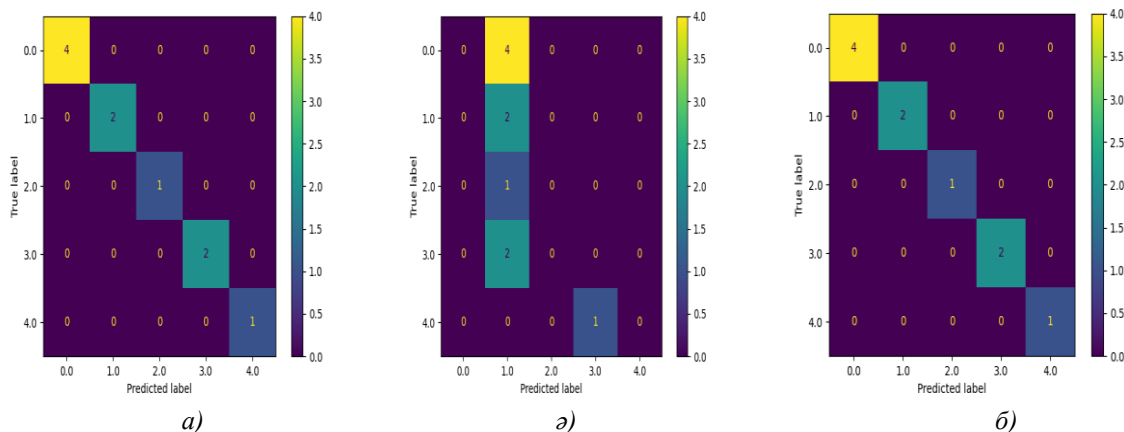


Сурет 3. Confusion матрицалары, мұндағы а) тыныс алу функциясының жіктелуі бойынша алынған шешімдер ағашы бойынша методологиясы, б) бұл k - жақын көршілер, в) бұл SVM алгоритмі

3-суреттегі а) нұсқасы бұл тестілеу кестесіндегі шешімдер ағашы бойынша методологияның жіктелуі қаншалықты оң болжамдарды көрсететіндігінің дәлелі болмақ: болжамдардың көпшілік бөлігі 1 сыныпқа жататын, дегенмен негізгі шығатын сыныптарына сәйкес келмейтін болады.

3- суреттегі ә) нұсқасы тестілеу кестесіндегі k - жақын көрші әдісімен жіктеу алгоритмінің: көпшілік болжамдар бойынша өз сыныбына сәйкес келетін болады.

3- суреттегі б) нұсқасы тестілеудің кестесіндегі SVM жіктеуінің алгоритмін: 0 және 1 сыныптарына болжау кезінде қателік барын көрсетеді.



Сурет 4. тыныс алу функциялары, оның ішінде а) Naive Bayes, ә) Adaboost, б) Xgboost алгоритмдерінің confusion матрицалары бойынша жіктелуі

4 суреттің а) нұсқасында тестілеу кестесі бойынша Naive Bayes жіктеуінің алгоритмі оң және теріс болжамдардың ара-қатынастарын көрсетеді: көпшілік көрсетілген болжамдар негізгі шығыс сыныптарына ұқсас келеді.

4 суреттің ә) нұсқасында Adaboost жіктеуінің алгоритмі қаншалықты оң болжамдар көрсететінінің дәлелін береді: 9 нысанды әртүрлі белгілеріне сүйеніп тестілеу көрсеткіштерін 1 сыныбына жатқызсақ, дегенмен суреттен көргеніміздей екеуі ғана бұл болжамдарға сәйкес екенін көреміз.

4 суреттің б) нұсқасында тестілеу кестесін Xgboost алгоритмі бойынша жіктеуде оң және теріс болжамдардың қатынасын көрсетеді: көпшілік байқайтын болжамдар негізгі шығыс сыныптарына жақындайды.

Дискуссия

Машиналық оқытуда k - жақын көршілер әдісі Naive Bayes ансамблдік оқыту тәсіліне кіретін Xgboost жіктеуі бойынша модельдері арасында тыныс алу функцияларының жұмысын бағалау үшін таңдау болды және бұл жерде үшінші моделге таңдау жасалынды. Енгізілген деректер Naive Bayes моделіне сәйкес өзара тәуелсіз болуы шарт, дегенмен зерттеу жұмыстарында таңдау жасалынған белгілерден өзара тәуелді екені анықталады. Сондықтан бұл моделге таңдау жасау кейінге шегерілді. Әртүрлі тексерулер негізінде он нысанға бақылау көрсеткіштерін жасап, тексерілді. Таңдаулар саны жеткілікті болатын кезде, сауалнаманың мәліметтерін тексеру, оқыту жиынтығына бөліп қарастыруға болады [8].

Шешімдерге ағаш моделі құрылуы үшін, жаттығу жиынтықтары мен тиімді шеткі моделге жету мақсатында қажет болатын ағаш өлшемдеріне таңдау жасау мақсатында тексеру бойынша ақпараттар жиынтығының қажеті айтарлықтай. Басқа аналитика тәсілдері тәрізді қолданушыларға ескертетін шешім ағашы әдісі бойынша шектеулер бар. Кемшілігі – ол аз-маз ақпаратты пайдалған кезде, артық камтуға және айтарлықтай ескермеуге болады.

Мұндай кемшілік нәтижесі бойынша алынған модельдер сенімділігі мен жалпылануына шектеулер болуы мүмкін. Келесі болжанатын кемшілік, ол әртүрлі мүмкін кіретін айнымалылардың арасында болатын үлкен корреляция моделдерінің статистикасын жоғарылататын, дегенмен қызығушылық нәтижесі себептік байланысқа емес айнымалылардың таңдауына әкеліп соғуы ықтимал.

Сонымен шешім ағашы моделдері бойынша түсінік және сол моделдер нәтижелері, себептік гипотезалар жасау кезінде қажет болады [9, 10]. Болашақта сандық денсаулық профилін құрастыру үшін жұмыстар жалғасатын болады.

Қорытынды

Функциялар іс-әрекетін бағалау кезінде машиналық оқыту, ансамбльдік оқыту тәсілдеріне жататын Adaboost алгоритмінің жіктеу моделі таңдалынды. Өйткені басқа моделдермен салыстырғанда оның жоғары дәлдікті көрсететіні анықталды. Зерттеулер барысында он нысан бақылауға алынып, әртүрлі көрсеткіш мәндері енгізіліп және талдаулар жасалынды.

Моделдер классификациясы бойынша дәл өлшеу, машиналық оқыту негізінде бағалайтын моделдер үшін маңызды. Келтірілген суреттер нәтижесі бойынша Adaboost, SVM, XGбуст жоғары дәлдік көрсеткіштеріне ие екендігін көреміз. Бұл мақаладағы бес жіктеу бойынша осы үш моделге таңдау жасалынды. Халықаралық тұрғыдағы жіктеме санаттары мен оны қолдайтын жасанды интеллект алгоритмдерін жан - жақты қарастыра келе біз,

Халықаралық тұрғыдағы жіктеме санаттары жіктеулерінің қолданылуы аймағы өте кең және әртүрлі екеніне тұжырымдама жасадық. Жасанды интеллект негізінде қарастырылған жіктеулер әдістері бойынша, Халықаралық тұрғыдағы жіктеме санаттарын жіктеуде қолданылған зерттеулер аз болса да 0 және 1 жіктеулеріне пайдаланылғаны анықталды.

АЛҒЫС

Бұл мақала Қазақстан Республикасы Білім және ғылым министрлігінің гранттық қаржыландыруы есебінен № AP09260767 "Қазақстан студенттерінің денсаулық жағдайын бағалаудың зияткерлік ақпараттық-аналитикалық жүйесін әзірлеу" жобасы аясында орындалды.

Пайдаланылған дереккөздер тізімі:

- 1 Alanjary M., Cano-Prieto C., Gross H., Medema M.H. Computer-aided re-engineering of nonribosomal peptide and polyketide biosynthetic assembly lines. // *Nat., «Prod. Rep.»* - 2019.- №36. –pp. 1249–1261
- 2 Montalbán-López M., et al. New developments in RiPP discovery, enzymology and engineering // *Nat. Prod. Rep.*, 2020, №38, pp. 130–239.
- 3 Hannigan, G. D., et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. // *Nucleic Acids Res.* 2019. №47, pp. 110.
- 4 Frontino G., Meschi F., Bonfanti R., Rigamonti A., Battaglini R., Favalli V., Bonura C., Ferro G., Chiumello G. Future perspectives in glucose monitoring sensors.// *European Endocrinology.* 2020, №9(1). Pp 6- 11.
- 5 Villena Gonzales W., Mobashsher A. T., Abbosh A. The Progress of Glucose Monitoring-A Review of Invasive to Minimally and Non-Invasive Techniques, Devices and Sensors. // *Sensors (Basel, Switzerland).* 2020. №19(4), P.800. <https://doi.org/10.3390/s19040800>
- 6 Facchinetti A. Continuous Glucose Monitoring Sensors: Past, Present and Future Algorithmic Challenges. // *Sensors (Basel, Switzerland).* 2016. №16(12). P 2093. <https://doi.org/10.3390/s16122093>
- 7 Salam N. A., Saad Z., Manap F., Salehuddin "The Evolution of Non-invasive Blood Glucose Monitoring System for Personal Application." // *Journal of Telecommunication, Electronic and Computer Engineering.* 2016. №8. pp. 59-65.
- 8 Kumar P. M., Lokesh S., Varatharajan R., Babu G. C., Parthasarathy P. Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier. // *Future Generation Computer Systems.* 2016. №86. pp. 527-534.
- 9 Narkhede P., Dhalwar S., Karthikeyan B. NIR Based Non-Invasive Blood Glucose Measurement. // *Indian Journal of Science and Technology.* 2016. №9(41). <https://doi.org/10.17485/ijst/2016/v9i41/98996>
- 10 Жолдас Н.А., Мансурова М.Е. Информационные технологии для мониторинга юных пациентов с сахарным диабетом. // *Автоматика и программная инженерия.* 2021, №1(35). Стр. 11-20.

References

- 1 Alanjary M., Cano-Prieto C., Gross H., Medema M.H. (2019) Computer-aided re-engineering of nonribosomal peptide and polyketide biosynthetic assembly lines. // *Nat., «Prod. Rep.»* №36. pp. 1249–1261
- 2 Montalbán-López M.,(2020) et al. New developments in RiPP discovery, enzymology and engineering // *Nat. Prod. Rep.*, №38, pp. 130–239.
- 3 Hannigan, G. D., (2019) et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. // *Nucleic Acids Res.* №47, pp. 110.
- 4 Frontino G., Meschi F., Bonfanti R., Rigamonti A., Battaglini R., Favalli V., Bonura C., Ferro G., Chiumello G. (2020) Future perspectives in glucose monitoring sensors.// *European Endocrinology.* №9(1). pp 6- 11.

- 5 Villena Gonzales W., Mobashsher A. T., Abbosh A. (2020) *The Progress of Glucose Monitoring-A Review of Invasive to Minimally and Non-Invasive Techniques, Devices and Sensors.* // *Sensors (Basel, Switzerland).* №19(4), P.800. <https://doi.org/10.3390/s19040800>
- 6 Facchinetti A. (2016) *Continuous Glucose Monitoring Sensors: Past, Present and Future Algorithmic Challenges.* // *Sensors (Basel, Switzerland).* 2016. №16(12). P 2093. <https://doi.org/10.3390/s16122093>
- 7 Salam N. A., Saad Z., Manap F., Salehuddin (2016) "The Evolution of Non-invasive Blood Glucose Monitoring System for Personal Application." . *Journal of Telecommunication, Electronic and Computer Engineering.* 2016. №8. pp. 59-65.
- 8 Kumar P. M., Lokesh S., Varatharajan R., Babu G. C., (2016) Parthasarathy P. *Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier.* *FutureGeneration Computer Systems.* 2016. №86. pp. 527-534.
- 9 Narkhede P., Dhalwar S., Karthikeyan B. (2016) *NIR Based Non-Invasive Blood Glucose Measurement.* // *Indian Journal of Science and Technology.* 2016. №9(41). <https://doi.org/10.17485/ijst/2016/v9i41/98996>
- 10 Zholdas N.A., Mansurova M.E. (2021) *Informacionnye tehnologii dlja monitoringa junyh pacientov s saharnym diabetom [Information technologies for monitoring young patients with diabetes mellitus].* *Nauchno-tehnicheskij zhurnal «Avtomatika i programmaja inzhenerija» (Novosibirskij institut programmyh sistem. RINC),* №1(35), 11-20.