

А.Б. Абен¹, Н. Әбдіжаппар²

¹ оқытушы магистр, Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті
(Қазақстан, Түркістан қ.), e-mail: arypzhan.aben@ayu.edu.kz,

² студент, Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті
(Қазақстан, Түркістан қ.), e-mail: axmet4047@gmail.com

ӨКПЕНІҢ ҚАТЕРЛІ ІСІГІН АНЫҚТАУДА МАШИНАЛЫҚ ОҚЫТУДЫ ҚОЛДАНУ
APPLICATION OF MACHINE LEARNING IN LUNG CANCER DETECTION
ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ РАКА ЛЕГКИХ

Аңдатпа. Қазіргі таңда жасанды интеллект көптеген салаларда кеңінен қолданылуда. Медицина саласында да жасанды интеллект кеңінен қолданыс тапқан. Бұл зерттеуде өкпенің қатерлі ісік ауруларын алдын ала болжау үшін машиналық оқыту модельдерін пайдаланылды. Зерттеуде өкпенің қатерлі ісігі ауруына шалдыққан науқастардың деректерінен тұратын деректер жиыны пайдаланылды. Өкпенің қатерлі ісігін ерте анықтауда машиналық оқыту модельдерінен K-Neighbor Classifier, Support Vector Machine (SVM), Logistic Regression, Random Forest Classifier, Gradient Boosting Classifier, LGBM Classifier, және Support Vector Classifier қолданылды. Бұл модельдердің болжам дәлдіктері алынды, нәтижесінде SVM және Support Vector Classifier модельдері 99% жоғары көрсеткіш көрсетсе, K-Neighbor Classifier және Random Forest сияқты басқа да модельдер 87.39% - 94.96% аралығында жоғарғы нәтижелерге алынды. Нәтижелер көрсеткендей өкпенің қатерлі ісігін анықтауда машиналық оқыту модельдері жоғарғы нәтижелер көрсетіп, болашақта өкпенің қатерлі ісігін анықтауда және осы сала зерттеулеріне үлкен үлес қосты. Зерттеуде пайдаланылған деректер жиыны мен машиналық оқыту модельдері медицина саласында пайдалану барысында жоғарғы нәтижелер алынатыны дәлелденді. Зерттеуде қарастырылған машиналық оқыту модельдерімен қоса болашақта терең оқыту модельдері пайдалану ұсынылды.

Негізгі сөздер: Өкпе қатерлі ісігі, жасанды интеллект, машиналық оқыту, болжам жасау.

Abstract. Currently, artificial intelligence is widely used in many fields. Artificial intelligence is also widely used in the field of medicine. This study used machine learning models to predict lung cancer. The study used a dataset consisting of lung cancer patients. Among the machine learning models used in the early detection of lung cancer are K-Neighbor Classifier, Support Vector Machine (SVM), Logistic Regression, Random Forest Classifier, Gradient Boosting Classifier, LGBM Classifier, and Support Vector Classifier. The prediction accuracies of these models were obtained, as a result, SVM and Support Vector Classifier models showed high performance of 99%, while other models such as K-Neighbor Classifier and Random Forest achieved high results ranging from 87.39% to 94.96%. The results show that machine learning models in lung cancer detection have shown superior results and have made a great contribution to future lung cancer detection and research in this field. The datasets and machine learning models used in the study have been proven to yield superior results when used in the medical field. In addition to the machine learning models considered in the study, it is recommended to use deep learning models in the future.

Keywords: Lung cancer, artificial intelligence, machine learning, prediction.

Аннотация. В настоящее время искусственный интеллект широко используется во многих сферах. Искусственный интеллект также широко используется в сфере медицины. В этом исследовании использовались модели машинного обучения для прогнозирования рака легких. В исследовании использовался набор данных, состоящий из пациентов с раком легких. Среди моделей машинного обучения, используемых для раннего выявления рака легких, — классификатор K-соседей, машина опорных векторов (SVM), логистическая регрессия, классификатор случайного леса, классификатор повышения градиента, классификатор LGBM и классификатор опорных векторов. Были получены точности прогнозирования этих моделей, в результате модели SVM и классификатора опорных векторов показали высокую производительность 99%, в то время как другие модели, такие как классификатор K-соседей и случайный лес, достигли высоких результатов в диапазоне от 87,39% до 94,96%. Результаты показывают, что модели машинного обучения при обнаружении рака легких показали превосходные результаты и внесли большой вклад в будущее обнаружение рака легких и исследования в этой области. Доказано, что наборы данных и модели машинного обучения, использованные в исследовании, дают превосходные результаты при использовании в медицинской сфере. Помимо рассмотренных в исследовании моделей машинного обучения, в дальнейшем рекомендуется использовать модели глубокого обучения.

Ключевые слова: рак легких, искусственный интеллект, машинное обучение, прогнозирование.

Кіріспе

Өкпенің қатерлі ісігі Қазақстанда көп таралған және қауіпті онкологиялық ауру түрі болып табылады. Статистика көрсеткендей елімізде жыл сайын 3500-ге жуық адам осы аурумен ауырады. Алайда науқастардың 80%-ы медициналық көмекке аурудың асқынған кезеңінде ғана жүгінетіні анықталған. Қатерлі ісіктерді ерте анықтау үшін онкологтар скринингтік тексеруден өту міндетті болуы керек деп есептейді. Өкпенің қатерлі ісігінің ең көп тараған түрлері шағын жасушалы өкпенің қатерлі ісігі (СКЛК) және шағын жасушалы емес өкпенің қатерлі ісігі (ҰТҚК) болып табылады. Бұл жағдайда SCLC барлық өкпе ісіктерінің шамамен 10-15% құрайды. Ал қалған өкпенің қатерлі ісігімен ауыратын науқастардың 80-85% -ы NSCLC ретінде жіктеледі[1].

Дүние жүзінде өкпенің қатерлі ісігі өлім-жітім санының артуына жетекші себебі болып отыр. Америка Құрама Штаттарының деректеріне сәйкес, өкпе рагы ерлер мен әйелдер арасында кең таралған. Бұл онкологиялық өлімнің әрбір бесінші бөлігін құрайды. 2024 жылға арналған болжамдар өкпе ісігінің шамамен 234 580 жаңа жағдайына диагноз қойылады. Нәтижесінде науқастардың 125 070 өлім жағдайы орын алады. Өкпенің қатерлі ісігі көбінесе 65 жастан асқан адамдарда кездеседі[2]. Зерттеуде нысаны ретінде алынған аурумен байланысты өлім-жітім деңгейі жоғары болғандықтан, ерте анықтау қажеттілігі ерекше маңызды ие.

Медицинадағы соңғы жылдарда жасанды интеллект (AI) және машиналық оқыту алгоритмдері салаларындағы өкпе ісігін ерте анықтау үшін жаңа мүмкіндіктер ашып отыр. Машиналық оқыту үлкен деректер жиынтығын өңдеу мүмкіндігі арқылы дәстүрлі әдістермен салыстырғанда ауруды диагностикалауда үлкен дәлдік пен жылдамдықты ұсына алады. Бұл алгоритмдер өкпе ісігінің ерте кезеңдерінде анықтау қиын болатын үлгілер мен ауытқуларды анықтау үшін үлкен көлемдегі радиологиялық кескіндер мен медициналық деректерді талдай алады. Мұндай технологияларды қолдану диагностиканың сапасын арттырады. Бұл технологиялар сондай ақ емдеуді уақтылы бастауды, пациенттердің өмір сүру сапасын жақсартады.

Бұл зерттеуде машиналық оқыту алгоритмдерін пайдалана отырып аурудың ерте кезеңнен анықтаудың тәсілдерін ұсынады. Қазіргі уақытта өкпенің қатерлі ісігін диагностикалауда әртүрлі машиналық оқыту әдістері қолданылады. Бұл салада зерттеулерде терең оқыту және жасанды нейрондық желілер сияқты технологиялар мен модельдер жиі қолданылып жүр. Бұл әдістердің артықшылығы үлкен деректер жиынтығын жылдам өңдеу және ауруды анықтау дәлдігін жақсарту қабілетінде ие екені болып отыр.

Бұл зерттеудің мақсаты машиналық оқыту әдістерін қолдану арқылы өкпенің қатерлі ісігін ерте диагностикалаудың тиімділігін зерттеу. Сонымен қатар осы саладағы соңғы жетістіктерді талқылау болып табылады. Зерттеу әртүрлі машиналық оқыту алгоритмдерінің ерекшеліктерін және олардың клиникалық қолдану мүмкіндігін зерттейді. Мақалада сонымен қатар диагностикалық әдістердің артықшылықтары мен шектеулері талданады. Зерттеуде болашақ зерттеулерге ұсыныстар атап өтілетін болады.

Ерте анықтау және биомаркерді табу

Гулд және т.б. (2021) әдеттегі клиникалық және зертханалық деректерді пайдалана отырып, өкпенің қатерлі ісігін ерте анықтау үшін ML модельдерінің пайдалылығын зерттеді[3]. Олардың нәтижелері логистикалық регрессия және кездейсоқ орман сияқты алгоритмдер ерте диагностиканы айтарлықтай жақсарта алады, өкпенің қатерлі ісігінің қаупін жоғарылататын дәлдікпен болжайды. Хіе және т.б. (2021) ML әдістерін қолдана отырып, ерте диагностикалық биомаркерлерді ашуға бағытталған [4]. Функцияларды таңдау әдістерін қосу арқылы олар диагностикалық процесті жақсарта алатын тиісті биомаркерлерді анықтады. Дегенмен, Nageswaran және т.б. Өкпенің қатерлі ісігін ML және кескінді өңдеу арқылы жіктеуді мақсат еткен (2022) кейінірек алынып тасталды, бұл зерттеудегі әдістемелік қатаңдықтың маңыздылығын атап өтті[5].

Болжау және жіктеу үшін машиналық оқыту алгоритмдері

Өкпенің қатерлі ісігін болжау саласында Патра (2020) әртүрлі ML моделдерін, соның ішінде шешім ағаштарын, қолдау векторлық машиналарын (SVM) және k-ең жақын көршілерді (KNN) пайдаланған болатын [6]. Бұл зерттеу модель дәлдігі мен сыныптандыру тиімділігін арттырудағы алгоритмді таңдаудың маңызды рөліне атқарды. Сол сияқты Чатурведи және т.б. (2021) кездейсоқ орман алгоритмі тексерілген әдістердің арасында ең жоғары дәлдікке қол жеткізгенін анықтады. Өкпенің қатерлі ісігін болжау және жіктеу үшін бірнеше ML әдістерін енгізді [7]. Mukherjee and Bohra (2020) диагностикалық нәтижелерді жақсартуда деректерді алдын ала өңдеудің және мүмкіндіктерді таңдаудың маңыздылығын көрсетті [8].

Болжамдық модельдер және тәуекелді бағалау

Абдулла және т.б. (2021) ML әдістерімен біріктірілген корреляциялық таңдау әдістеріне негізделген өкпе обырын болжау және жіктеу моделін ұсынды [9]. Олардың нәтижелері мүмкіндікті тиімді таңдау болжау дәлдігін айтарлықтай арттыра алатынын көрсетті. Дрицас пен Тригга (2022) пациенттердің денсаулығының әртүрлі айнымалыларын біріктіретін өкпе ісігі қаупін болжау үшін ML үлгілерін әзірледі. Зерттеушілер ансамбльдік әдістер тәуекелді жоғары бағалауды қамтамасыз ететінін анықтады [10]. Сонымен қатар, Тункал және т.б. (2020) әртүрлі ML алгоритмдері арқылы өкпенің қатерлі ісігімен сырқаттанушылықты болжауды зерттеген болатын. Болжау дәлдігі градиентті күшейтетін машиналар ең жақсы нәтиже береді деген қорытындыға келді.

Жетілдірілген әдістер және түсіндіру

Алсинглави және т.б. (2022) өкпе обыры бар науқастардың ауруханада болу ұзақтығын болжауға арналған түсіндірілетін ML құрылымын енгізді [12]. Бұл модель болжаулардың денсаулық сақтау провайдерлері үшін түсінікті болуын үшін интерпретациялау әдістерін қолданды. ML қолданбаларында ашықтық қажеттілігін қарастырады. Осы сияқты зерттеулердің бірі Чаунзва және т.б. (2021) КТ кескіндері арқылы өкпе ісігі гистологиясын жіктеу үшін терең оқыту әдістерін қолданды. Зерттеуде терең нейрондық желілер гистологиялық ішкі типтерді жоғары дәлдікпен жіктей алатындығын көрсетті. Осылайша радиологиялық диагностика үшін құнды құралды қамтамасыз етеді [13].

Кешенді тәсілдер мен шолулар

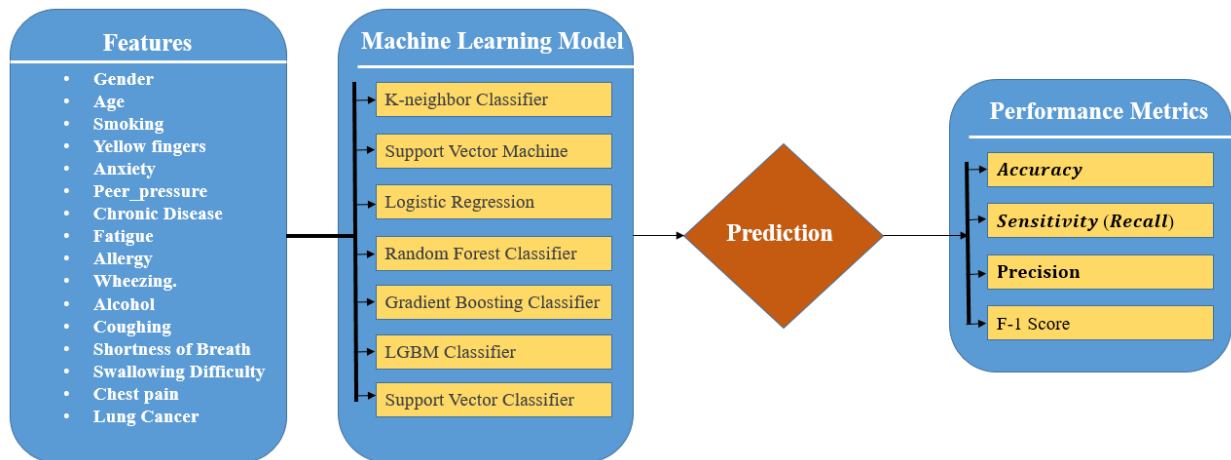
Ли және т.б. (2022) әртүрлі әдістерді, соның ішінде бақыланатын оқытуды, терең оқытуды және нығайтуды оқытуды талқылады. Өкпенің қатерлі ісігін диагностикалауда, емдеуде және болжауда ML қосымшаларына мұқият шолу жасады [14]. Рауф және т.б. (2020) дәлдік пен дәлдік [15] сияқты өнімділік көрсеткіштеріне негізделген бірнеше модельдерді бағалады. ML көмегімен өкпе обырын болжауға кешенді тәсілді ұсынды. Бұл өкпенің қатерлі ісігінің диагностикасын тиімдірек іздеуде зерттелетін әртүрлі әдістемелерді көрсетеді.

Өкпенің қатерлі ісігін диагностикалауда және болжауда машиналық оқытуды қолдану дәстүрлі алгоритмдерден күрделі тәсілдерге бөлді. Соның ішінде терең оқыту мен ансамбльдік әдістерге ауыса отырып, айтарлықтай дамыды. Қаралған зерттеулер диагностикалық дәлдік пен клиникалық қолдану мүмкіндігін арттыру үшін ерекшеліктерді таңдаудың дәйектілік көрсетті. Биомаркерді ерте анықтаудың және модельді түсіндірудің маңыздылығын дәйекті түрде атап көрсетеді. Алгоритмді таңдау және әдістемелік қатандық сияқты қиындықтар өкпенің қатерлі ісігін анықтау үшін ML шешімдерін тиімді қолдануда айтарлықтай кедергілер тудыруда. Осы қиындықтарды шешу машиналық оқыту технологияларын клиникалық тәжірибеге интеграциялауды ілгерілету маңызды.

Материалдар мен тәсілдер

Бұл зерттеуде өкпенің қатерлі ісігінің қаупін болжау үшін арнайы әзірленген деректер жиынтығы қолданылады. Зерттеуде пайдаланылған деректер жиыны науқастардың

деректерін пайдалана отырып, қатерлі ісікті дәл болжауға өз үлесін қосады. Зертеу барысында машиналық оқыту алгоритімдерін салсыра отырып жасалынған модель ұсынылады. Зерттеуде пайдаланылған архитектура төменде 1-суретте көрсетілген.



Сурет 1. Зертеуде қолданылған модель архитектурасы.

Деректер жиынына шолу

Зертеуден пайдаланылатын деректер жиыны өкпенің қатерлі ісігімен ауырған науқастар деректерімен жасалынған. Деректер жиыны науқастардың денсаулықтарына байланысты әр түрлі факторларды ескере отырып және қатерлі ісікті бағалауға арналған 284 инстанция мен 16 бағанан тұрады. Деректер қкпенің қатерлі ісігін анықтауда маңызды рөл атқарады. Төменде осы деретер жиынын сипаттама берілген:

- Жыныс: науқастың жынысын білдіретін категория (М = Ер, F = Әйел).
- Жасы: науқастың жасын көрсететін баған атауы.
- Темекі шегу: науқастың темекі шегіп, шекпегендігі туралы ақпарат беретін катигория (ИӘ = 2, ЖОҚ = 1).
- Сары саусақтар: науқастың сары саусақтардың бар немесе жоқ екенін көрсететін категория (ИӘ = 2, ЖОҚ = 1).
- Мазасыздық: мазасыздану деңгейін көрсететін категориялық айнымалы (ИӘ = 2, ЖОҚ = 1).
- Бірлескен қысым: тең дәрежедегі қысым әсерін көрсететін категориялық айнымалы (ИӘ = 2, ЖОҚ = 1).
- Созылмалы ауру: созылмалы аурулардың болуын көрсететін категориялық айнымалы (ИӘ = 2, ЖОҚ = 1).
- Шаршау: Шаршау деңгейін көрсететін категориялық айнымалы (ИӘ = 2, ЖОҚ = 1).
- Аллергия: Аллергия күйін көрсететін категориялық айнымалы (ИӘ = 2, ЖОҚ = 1).
- Қырылдау: сырылдар белгілерін көрсететін категориялық айнымалы (ИӘ = 2, ЖОҚ = 1).
- Алкоголь: алкогольді тұтынуды көрсететін категориялық айнымалы (ИӘ = 2, ЖОҚ = 1).
- Жөтел: жөтел симптомдарын көрсететін категориялық айнымалы (ИӘ = 2, ЖОҚ = 1).
- Тыныс алудың қысқа болуы: еңтігуді көрсететін категориялық айнымалы (ИӘ = 2, ЖОҚ = 1).
- Жұтыну қиындығы: жұтылу қиындықтарын көрсететін категориялық айнымалы мән (ИӘ = 2, ЖОҚ = 1).
- Кеудедегі ауырсыну: Кеудедегі ауырсынуды көрсететін категориялық айнымалы (ИӘ = 2, ЖОҚ = 1).

- Өкпенің қатерлі ісігі: Өкпенің қатерлі ісігінің бар екенін көрсететін екілік мақсатты категория(ИӘ, ЖОҚ).

Машиналық оқыту үлгілері

Бұл зерттеуде берілген деректер жиынтығына негізделген өкпе ісігі қаупін болжау үшін машиналық оқыту модельдерін қолданылды. Әрбір модельдің бірегей күшті және әлсіз жақтары бар. Бұл оларды жіктеу тапсырмасының әртүрлі ерекшелікпен орындауы үшін қолайлы етеді.

К-Ең жақын көршілер классификаторы (KNN)

К-Ең жақын көршілер (KNN) данаға негізделген оқыту алгоритмі болып табылады. Ол Евклидтік қашықтық сияқты қашықтық өлшемдерімен өлшенеді. k-ең жақын көршілерінің көпшілік класына негізделген деректер нүктелерін жіктейді.

Қолдау векторлық машинасы (SVM)

Қолдау векторлық машинасы (SVM) класстарды бөлу үшін жоғары өлшемді кеңістікте гипержазықтықтарды құрастырады. Мақсат - әртүрлі сыныптар арасындағы маржаны барынша арттыратын гипержазықтықты табу.

Логистикалық регрессия

Логистикалық регрессия – логистикалық функцияны қолдану арқылы бір немесе бірнеше болжаушы айнымалылар негізінде екілік нәтиженің ықтималдығын бағалайтын және екілік жіктеу үшін қолданылатын статистикалық әдіс.

Кездейсоқ орман классификаторы

Кездейсоқ орман классификаторы – жаттығу кезінде бірнеше шешім ағаштарын құрастыратын және дәлдікті жақсарту қолданылады. Сонымен қатар артық орнатуды бақылау үшін олардың болжамдарын біріктіретін ансамбльді оқыту әдісі. Ол әрбір ағашты деректердің кездейсоқ жиынына үйрету үшін қаптауды пайдаланады.

Градиентті күшейтетін классификатор

Gradient Boosting классификаторы үлгілерді дәйекті түрде құрастырады. Мұнда әрбір жаңа үлгі алдыңғысының қателерін түзетуге әрекет жасайды. Ол жоғалту функциясын азайту үшін градиентті түсіруді қолдана отырып, әлсіз оқушыларды күшті оқушыға біріктіреді.

LightGBM классификаторы (LGBM)

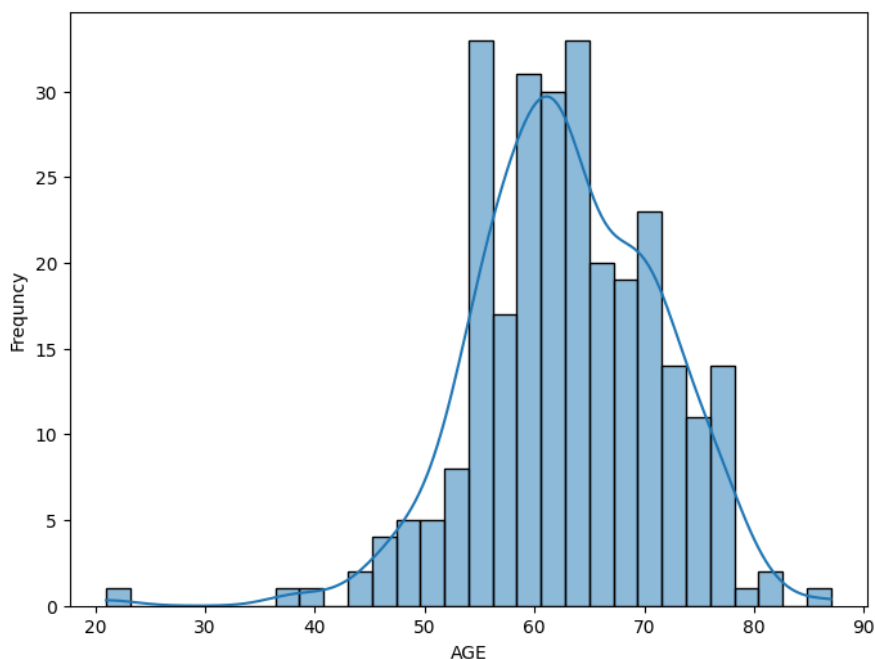
LightGBM (Light Gradient Boosting Machine) – жылдамдық пен тиімділікке арналған. Жадты аз қолданумен үлкен деректер жиынын өңдеуге қабілетті ағаш негізделген. LightGBM классификаторы оқыту алгоритмдерін пайдаланатын градиентті күшейтетін әдіс.

Бұл модельдер өкпе ісігі қаупін болжаудың әртүрлі тәсілдерін ұсынады. Модельдердің әрқайсысы дәл бағалау үшін ең тиімді шешімді анықтау үшін деректер жиынтығын кешенді бағалауға ықпал етеді.

Нәтижелер

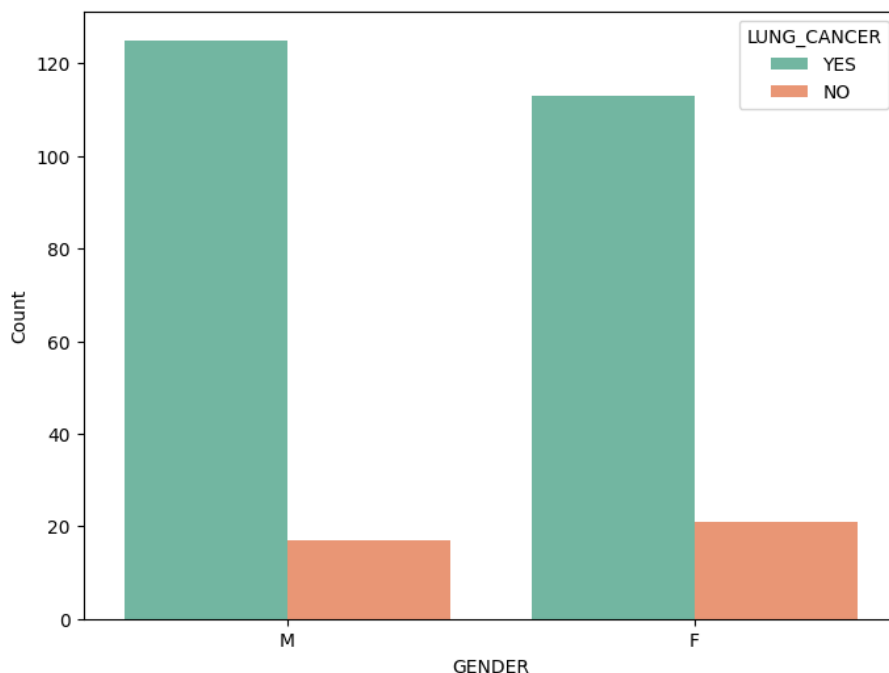
Бұл зерттеуде науқастардың жасына, жынысына, темекі шегу әдеттеріне, алкогольді тұтынуға және өкпе ісігі жағдайларының жалпы таралуына назар аударылды. Зерттеу барысында өкпе ісігі деректер жинағын талдау үшін әртүрлі визуализациялар жасалды. Келесі бөлімдер орындалған көрнекі талдаулардан алынған нәтижелерді егжей-тегжейлі сипаттайды.

Төменде Сурет 2-де науқастардың жасының таралуын бағалау үшін гистограмма көрсетілген. Бұл гистограмма өкпенің қатерлі ісігімен ауыратын науқастардың жастық бөлінуін нақты көрсету үшін 30 қалтаны пайдаланды. Гистограммадағы тығыздық қисығы (KDE) өкпенің қатерлі ісігінен ең көп зардап шеккен жас топтарын қосымша көрсетеді.



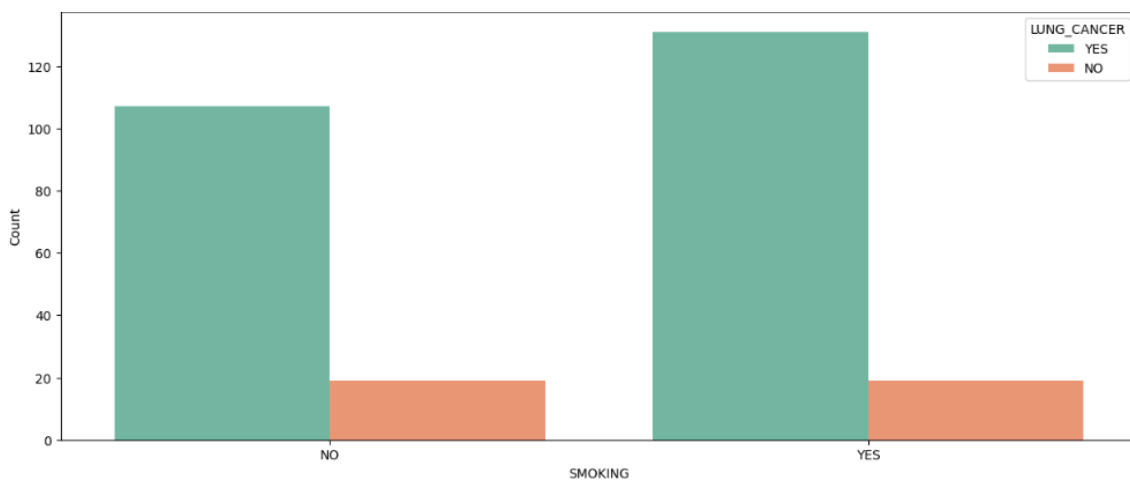
Сурет 2. Өкпенің қатерлі ісігі үшін жастарының таралуы диаграммасы.

Төменде Сурет 3-те өкпенің қатерлі ісігімен сырқаттанушылыққа қатысты жыныстың таралуы көрсетілген. Көрсетілген диаграммада ерлер мен әйелдердің арасында аурудың таралуын анықтауда маңызды рөл атқарады.



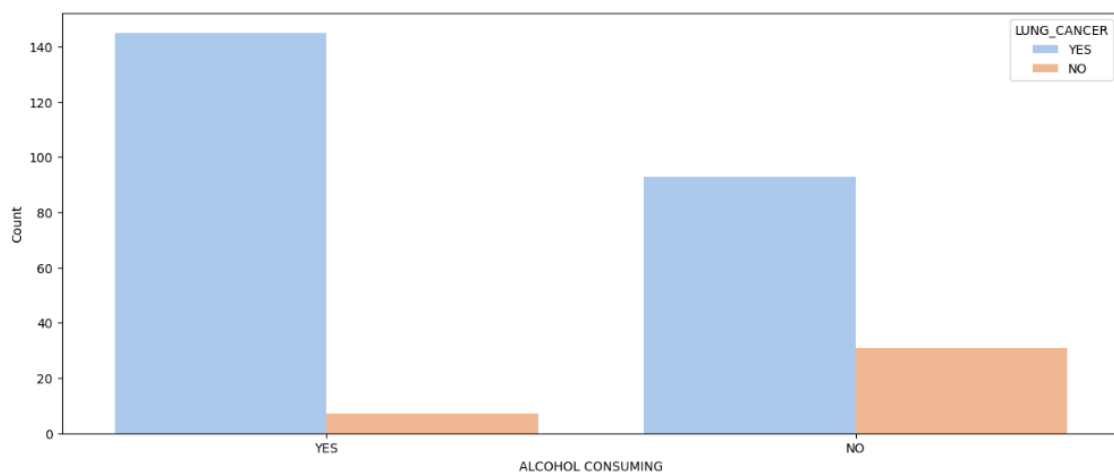
Сурет 3. Өкпе ісігі бойынша жыныстың таралуы.

Келесі Сурет 4-те темекі шегудің өкпенің қатерлі ісігі қаупіне әсері санау сызбасы арқылы көрсетілген. Бұл науқастардың темекі шегу мәртебесіне қарай жіктейді. Темекі шегу мен өкпе ісігінің пайда болуы арасындағы байланысты ашуға көмегін тигізеді.



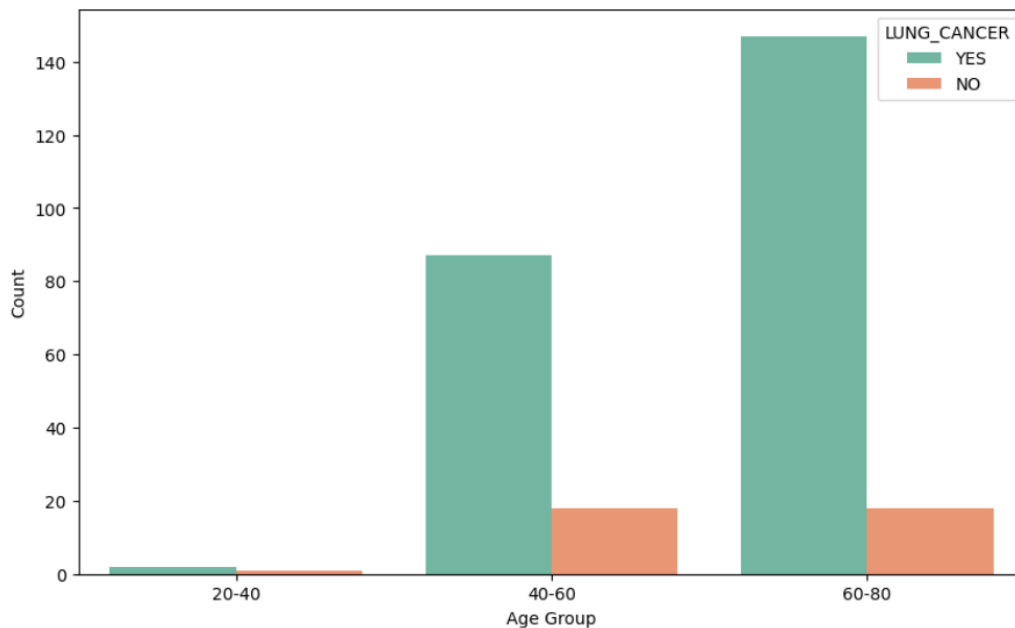
Сурет 4. Өкпенің қатерлі ісігі бойынша темекі шегудің таралуы.

Алкогольді тұтынудың өкпенің қатерлі ісігіне әсерін бағалау үшін басқа санау сызбасы салынды. Сурет 5-те алкогольді тұтыну әдеттері мен өкпе ісігінің жиілігі арасындағы байланысты көрсетеді.



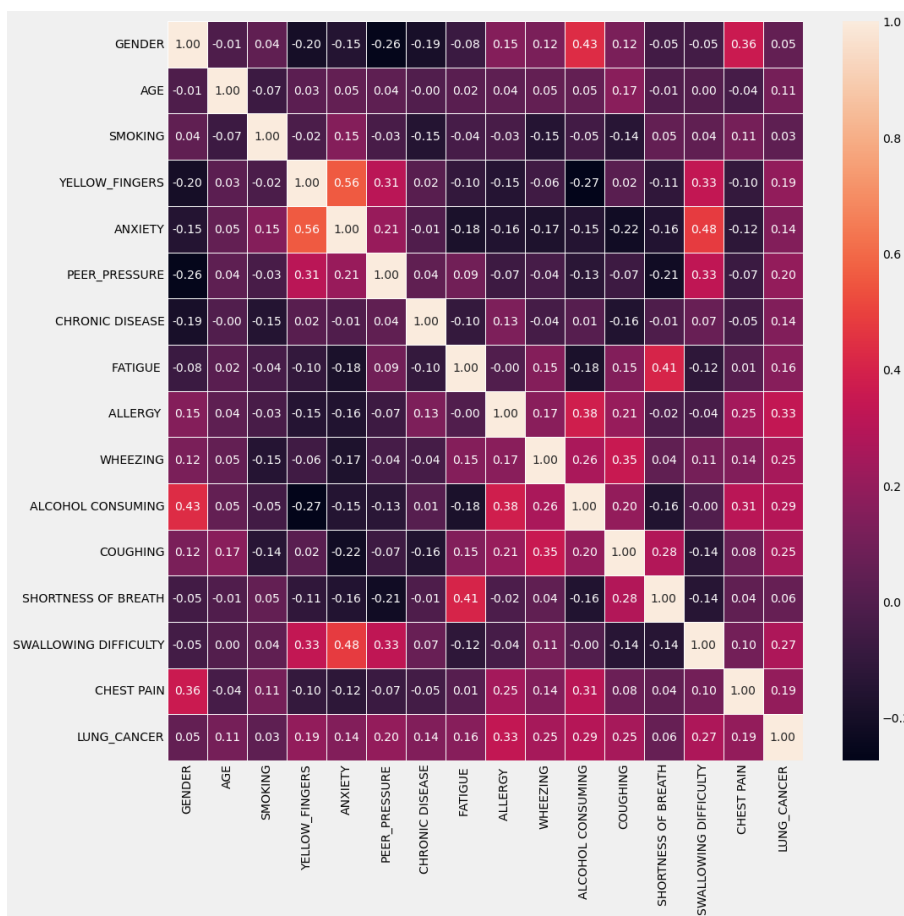
Сурет 5. Алкогольді тұтынуды өкпе ісігі бойынша бөлу.

Жасқа байланысты тенденцияларды әрі қарай талдау үшін пациенттердің жасы әртүрлі жас топтары бойынша анықтау Сурет 6-да көрсетілген. Суретте өкпе обырының таралуын зерттеуге мүмкіндік беретін жәшіктерге (AGE_BIN) жіктелді. Бұл жіктеу өкпенің қатерлі ісігімен ауыратын науқастарға байланысты жасқа байланысты қауіп факторларын жақсы түсінуге мүмкіндік береді.



Сурет 6. Жас топтары бойынша өкпе ісігінің таралуы.

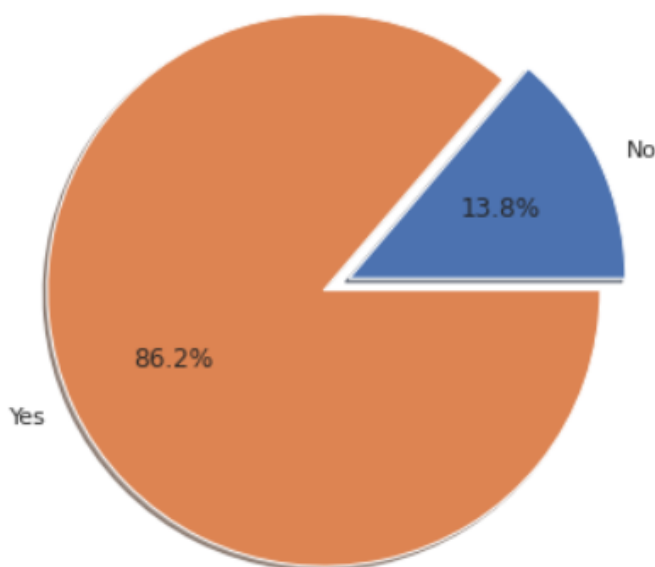
Деректер жиынындағы әртүрлі атрибуттар арасындағы корреляцияны зерттеу үшін жылу картасы көрсетілді. Бұл визуализация әртүрлі қауіп факторлары арасындағы қарым-қатынасты көрсетеді. Сурет 8-де науқастардың өкпенің қатерлі ісігіне ықтимал әсері туралы түсінік береді.



Сурет 7. Деректер жиынының корреляциялық жылу картасы.

Зерттеу барысында болжалды үлгілердің өнімділігін бағалау үшін ROC AUC талдауы жүргізілді. Бұл талдау өкпе ісігінің оң және теріс жағдайларын ажыратудағы әрбір модельдің тиімділігінің жан-жақты көрінісін қамтамасыз етеді. Жалған оң көрсеткішке қарсы шынайы оң көрсеткішті бағалайды.

Төменде сурет 9-да өкпенің қатерлі ісігі бар және онсыз науқастардың таралуын көрсету үшін дөңгелек диаграмма көрсетілген. Бұл визуализация өкпенің қатерлі ісігінің болуына қатысты деректер жиынтығының құрамының нақты шолуын қамтамасыз етеді.



Сурет 8. Өкпе ісігі жағдайларының таралуы.

Осы бөлімде ұсынылған суретте өкпе ісігі деректерінің жиынтық талдауын қамтамасыз етеді, өкпе ісігінің жиілігіне әсер ететін маңызды факторларды көрсетеді. Жастың, жыныстың, темекі шегудің, алкогольді тұтынудың және өкпе обырының жалпы таралуының таралуын зерттей отырып, бұл зерттеу осы аурумен байланысты қауіп факторлары туралы құнды түсініктерге ықпал етеді. Бұл нәтижелер біздің өкпе рагы динамикасын түсінуді жақсартып қана қоймайды, сонымен қатар одан әрі зерттеулер мен мақсатты араласу стратегияларына жол ашады.

Өкпенің қатерлі ісігін болжауда әртүрлі машиналық оқыту үлгілерінің өнімділігі Кесте 1-де жинақталған нәтижелермен бағаланды. Дәлдік көрсеткіштері деректер жиынына негізделген өкпе ісігі жағдайларын жіктеудегі әрбір модельдің тиімділігі көрсетілген.

Кесте 1. Өкпенің қатерлі ісігін болжау үшін машиналық оқыту үлгілерінің нәтижелері.

| | Модель | Дәлдік (%) |
|---|------------------------------|-------------------|
| 1 | K-Neighbor Classifier | 90.76 |
| 2 | Support Vector Machine (SVM) | 99.16 |
| 3 | Logistic Regression | 87.39 |
| 4 | Random Forest Classifier | 94.12 |
| 5 | Gradient Boosting Classifier | 94.12 |
| 6 | LGBM Classifier | 94.96 |
| 7 | Support Vector Classifier | 99.16 |

Кестеде берілгендей қолдау векторлық машинасы (SVM) және қолдау векторының классификаторы екеуі де ең жоғары дәлдікке қол жеткізді - 99,16%. Бұл ерекше өнімділік бұл модельдердің өкпенің қатерлі ісігі жағдайлары мен емес жағдайларын ажыратуда жоғары тиімді екенін көрсетеді. Модель науқастардағы қатерлі ісік болжау үшін баға жетпес құрал етеді. SVM жоғары өлшемді кеңістіктерді өңдеу және күрделі деректер жиынын жіктеу қабілеті оның табысқа жетуіне ықпал етеді.

K-Neighbor классификаторы 90,76% дәлдікке қол жеткізді. Модель сенімді өнімділікті көрсетті. Алайда SVM модельімен салыстырғанда өкпенің қатерлі ісігін анықтауда жеткіліксіз болып отыр. Ал логистикалық регрессия моделі 87,39% дәлдікке ие нәтижені көрсетті. Логистикалық регрессия басқа модельдерге қарағанда төмен нәтижеге ие болды. Логистикалық регрессия өзінің қарапайымдылығы мен түсіндірмелілігі үшін жиі қолданылады. Логистикалық регрессия денсаулық сақтау мамандарына әртүрлі болжаушылар мен өкпе ісігі қаупі арасындағы қарым-қатынасты түсінуге көмектеседі. Кездейсоқ орман классификаторы, градиентті күшейту классификаторы және LGBM классификаторы сәйкесінше 94,12%, 94,12% және 94,96% дәлдікке қол жеткізді. Бұл ансамбльдік үлгілердің салыстырмалы өнімділігі олардың деректер жиынының күрделілігін түсірудегі тиімділігін көрсетеді. Random Forest, Gradient Boosting және LGBM классификаторларының салыстырмалы түрде өкпе обырын болжау ерекшеліктерін өңдеуде сенімді екенін көрсетеді.

Нәтижелер өкпенің қатерлі ісігін болжаудағы SVM негізіндегі әдістердің артықшылығын көрсетті. Сонымен бірге Random Forest және Gradient Boosting сияқты ансамбльдік әдістердің тиімділігіне баса назар аударады. Нәтижелер олардың интерпретациялануын жақсарту және клиникалық жағдайларда қолданылуын қамтамасыз ету үшін осы үлгілерді одан әрі зерттеуді ынталандырады. Өкпе обырын диагностикалау мен емдеуде ерте анықтауды және науқастардың нәтижелерін жақсартуға мүмкіндік береді.

Қорытынды

Бұл зерттеу пациенттің сипаттамалары мен симптомдарына қатысты әртүрлі атрибуттарды қамтитын деректер жиынтығын пайдаланылды. Өкпе обырын болжау үшін әртүрлі машиналық оқыту модельдерінің тиімділігін зерттеді. Нәтижелер машиналық оқыту өкпенің қатерлі ісігін ерте анықтауды айтарлықтай жақсарту алатынын көрсетті.

Жұмысты бағалау қолдау векторлық машинасы (SVM) және қолдау векторлық классификаторы 99,16% деңгейінде ең жоғары дәлдікке қол жеткізгенін көрсетті. Бұл нәтижелер SVM негізіндегі модельдер өкпенің қатерлі ісігі жағдайлары мен емес жағдайларды ажырату үшін өте қолайлы екенін көрсетті. Модель нәтижелері клиникалық шешім қабылдаудағы үлкен пайдасын көрсетеді. Ал K-Neighbor классификаторы, логистикалық регрессия және кездейсоқ орман, градиентті күшейту және LGBM сияқты әдістер де 87,39%-дан 94,96%-ға дейінгі жоғарғы нәтижелер көрсете алды.

Бұл зерттеуде дәлелденгендей машиналық оқыту модельдері өкпе обырын анықтауда жоғарғы нәтижелер көрсете алды. Зерттеудің барысында қолданылған деректер жиыны нақты деректерге негізделген. Болашақта осы сала зерттеушілеріне зерттеу объектісі ретінде бір медициналық орталықты алыну ұсынылып отыр. Сонымен қатар зерттеудің маңыздылығын ескере отырып машиналық оқытудан басқада салалармен байласын жасау ұсынылып отыр. Атап айтатын болсақ жасанды интеллекттің терең оқыту саласымен медициналық суреттерге, дыбыстарға жүргізу жоғарғы нәтиже береді деп күтілуде.

Пайдаланылған әдебиеттер тізімі

1. Qazaqstan. (2023, July 14). High incidence of lung cancer in Almaty. *Qazaqstan National Television*. Retrieved October 20, 2024, from <https://qazaqstan.tv/news/178923/>

- American Cancer Society. (n.d.). Key statistics for lung cancer. In *Cancer.org*. Retrieved October 20, 2024, from <https://www.cancer.org/cancer/types/lung-cancer/about/key-statistics.html>
- Gould, M. K., Huang, B. Z., Tammemagi, M. C., Kinar, Y., & Shiff, R. (2021). Machine learning for early lung cancer identification using routine clinical and laboratory data. *American journal of respiratory and critical care medicine*, 204(4), 445-453.
- Xie, Y., Meng, W. Y., Li, R. Z., Wang, Y. W., Qian, X., Chan, C., ... & Leung, E. L. H. (2021). Early lung cancer diagnostic biomarker discovery by machine learning methods. *Translational oncology*, 14(1), 100907.
- Nageswaran, S., Arunkumar, G., Bisht, A. K., Mewada, S., Kumar, J. S., Jawarneh, M., & Asenso, E. (2022). [Retracted] Lung Cancer Classification and Prediction Using Machine Learning and Image Processing. *BioMed research international*, 2022(1), 1755460.
- Patra, R. (2020). Prediction of lung cancer using machine learning classifier. In *Computing Science, Communication and Security: First International Conference, COMS2 2020, Gujarat, India, March 26–27, 2020, Revised Selected Papers I*. 132-142. Springer Singapore.
- Chaturvedi, P., Jhamb, A., Vanani, M., & Nemade, V. (2021, March). Prediction and classification of lung cancer using machine learning techniques. In *IOP conference series: materials science and engineering* (Vol. 1099, No. 1, p. 012059). IOP Publishing.
- Mukherjee, S., & Bohra, S. U. (2020, December). Lung cancer disease diagnosis using machine learning approach. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 207-211). IEEE.
- Abdullah, D. M., Abdulazeez, A. M., & Sallow, A. B. (2021). Lung cancer prediction and classification based on correlation selection method using machine learning techniques. *Qubahan Academic Journal*, 1(2), 141-149.
- Dritsas, E., & Trigka, M. (2022). Lung cancer risk prediction with machine learning models. *Big Data and Cognitive Computing*, 6(4), 139.
- Tuncal, K., Sekeroglu, B., & Ozkan, C. (2020). Lung cancer incidence prediction using machine learning algorithms. *Journal of advances in information technology*, 11(2).
- Alsinglawi, B., Alshari, O., Alorjani, M., Mubin, O., Alnajjar, F., Novoa, M., & Darwish, O. (2022). An explainable machine learning framework for lung cancer hospital length of stay prediction. *Scientific reports*, 12(1), 607.
- Chaunzwa, T. L., Hosny, A., Xu, Y., Shafer, A., Diao, N., Lanuti, M., ... & Aerts, H. J. (2021). Deep learning classification of lung cancer histology using CT images. *Scientific reports*, 11(1), 1-12.
- Li, Y., Wu, X., Yang, P., Jiang, G., & Luo, Y. (2022). Machine learning for lung cancer diagnosis, treatment, and prognosis. *Genomics, Proteomics and Bioinformatics*, 20(5), 850-866.
- Raof, S. S., Jabbar, M. A., & Fathima, S. A. (2020, March). Lung Cancer prediction using machine learning: A comprehensive approach. In *2020 2nd International conference on innovative mechanisms for industry applications (ICIMIA)* (pp. 108-115). IEEE.

Сведение об авторах

Абен А.Б. – преподаватель магистр, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави (Казахстан, г. Туркестан), e-mail: arypzhan.aben@ayu.edu.kz

Әбдіжаппар Н. - студент, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави (Казахстан, г. Туркестан), e-mail: axmet4047@gmail.com

Information about authors

Aben A.B. - master lecturer, Khoja Akhmet Yassawi International Kazakh-Turkish University, (Kazakhstan, Turkestan), e-mail: arypzhan.aben@ayu.edu.kz

Abdizhappar N. - student, Khoja Akhmet Yassawi International Kazakh-Turkish University (Kazakhstan, Turkestan), e-mail: axmet4047@gmail.com