



ӘОЖ 004.08.528

FTAXP 20.23.15

DOI 10.37238/2960-1371.2960-138X.2026.102(2).35

Е.А.Тулегенов, Н.О.Мекебаев, Ж.А. Орынтаева

Қазақ ұлттық қыздар педагогикалық университеті, Алматы, Қазақстан

E-mail: nurbara80@gmail.com

## ГЕОГРАФИЯЛЫҚ НҮКТЕЛЕРІНІҢ БИІКТІГІН БОЛЖАУДА МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН САЛЫСТЫРУ

**Аңдатпа.** Бұл мақалада жер бедерлерінің морфометриялық (биіктігі) көрсеткіштерін болжау барысында қазіргі таңдағы өзекті мәселелері жүйелі түрде талданады. Онда машиналық оқыту алгоритмдерімен географиялық ақпараттық жүйелер интеграциялау негізделеді. Қазіргі уақытта ҚР аумағының дерекқорын құруда рельефтің сандық модельін әзірлеу маңызды мәселелердің бірі. Осы мәселе Қазақстан аумағының геоморфологиялық жағдайының сан алуандығымен және күрделілігімен тікелей тәуелді. Тянь – Шань таулары мен Алтай тауының биік таулы белдеулерінен бастап Каспий маңы ойпатының депрессиялық алаптарының ҚР геоморфологиялық әртүрлілігі кез-келген математикалық модельдің тиімділігін тексеру верификациялық міндет жүктейді.

Зерттеуіміздің нәтижелері дәстүрлі әдістерге қарағанда жасанды интеллект әдістері (машиналық оқыту) 25-36 % жоғары дәлдік көрсетті. Біз жүргізген зерттеуімізде модельдерді салыстыруда Random Forest әдісі нәтиже көрсетті. Ол деректерді өңдеуде тұрақтылық көрсеткіштеріне ие болды. Осы модельдің ерекшелігі жер бедерінің өзгерістерін сипаттап, олардың олардың ауытқуларын бейнелеуге мүмкіндік берді.

**Кілт сөздер:** Географиялық нүктелер, машиналық оқыту, кеңістік, алгоритм, жер бедері, нейрондық желі.

### *Kipicne*

Геоинформатика саласының бағыттарының бірі, қазіргі кезеңдегі геоғылымдар жүйесі мен аумақтық мәліметтерді құрылымдық тұрғыдан зерттеу.

Қазіргі кезде жер бедерінің сандық модельін құру, жер бедерінің динамикасын және биіктік айырмашылықтарын дәл болжау картаға түсіру картография міндеттерімен шектелмейді. Аталған үрдіс табиғат байлықтарды ұтымды пайдалану агроөнеркәсіпті сандық жүйеге көшіруге гидрологиялық жағдайларды зерттеу, инженерлік жолар үшін маңызды. Дегенмен ақпараттық технологиялардың дамуына қарамастан жер бедерін сандық үлгіде көрсету әлі шешелімеген мәселе болып тұр. Сондықтан бұл зерттеу тақырыбы ғалымдар арасында сынға ұшыраған, өзекті болып тұр. Осы аталған әдістер мәліметтердің сызықтық тиімділікке ие. Жер бедерінің геологиялық құрылым, климаттың өзгеруі, топырақ құрамы, ылғалдылық деңгейі және эрозия процестері сияқты көптеген факторлардың өзара әрекеттесуінен қалыптасқан күрделі жүйе. Біздің пікіріміше, рельефтің кепөлшемді сипатын тек математикалық интерполяция арқылы сипаттау жеткіліксіз. Осыған байланысты, күрделі сызықтық емес тәуелділіктерді модельдеу үшін машиналық оқытудың аналитикалық принципін қолдану шұғыл қажеттілік туындайды.

Машиналық оқыту әдістері кеңістіктік деректер жиынтықтарында адам көзіне немесе дәстүрлі статистикаға көрінбейтін жасырын заңдылықтарды анықтай алады. Жақсы жалпылау қабілетіне ие машиналық оқыту модельдері биіктік көрсеткіштерін



болжау әдістерінен айтарлықтай жоғары нәтижелер көрсетуді бастаған. Осы мақалада біз Қазақстанның ерекше ландшафттық жағдайларына осы инновациялық алгоритмдерді бейімдеу арқылы географиялық нүктелердің биіктігін дәл болжау мәселесін зерттейміз.

Қазақстан территориясын зерттеу нысаны ретінде тандау кездейсоқ емес. Еліміз 2,7 миллион км үлкен аумақты алып жатыр, ал оның рельефі ерекше біркелкі сипатталады. Бір жағынан, бұл Батыс және Солтүстік Қазақстанның далалары мен теңіз деңгейінен төмен орналасқан Каспий ойпаты болса, екінші жағынан - оңтүстік пен шығыстағы Тянь-Шань, Семірічье Алатау, Тарбағатай және Алтай сияқты биік таулы жүйелер. Орталықтағы Саырарка мен жоғары емес төбелер, Маңғыстау мен Үстірттің кальцитті платолары, сондай-ақ Балқаш-Алакөл ойпаттарының морфологиялық құрылысы - барлығы бірыңғай цифрлық модель жасауға үлкен ғылыми кедергілер туғызады. Мұндай геоморфологиялық контраст жағдайында дәстүрлі сплайн немесе кригинг аппроксимациясының әдістері биіктік градиенттерінің кенет өзгерістерін ескере алмайды және нәтижесінде үлкен қателерге (RMSE) әкеледі [1],[4].

Қазақстандық ғылыми қоғамдастықта ГАЗ технологиялары соңғы онжылдықта ауыл шаруашылықты жерлерін бақылау, су ресурстарын қорғау және жер кадастрын реттеу сияқты салаларда кеңінен қолданылады. Дегенмен, машиналық оқыту негізінде сандық рельеф моделін автоматты түрде жасау және болжау мәселесі әлі де жүйелі зерттеуді қажет ететін болжау немесе климат карталарын жасау бойынша кейбір зерттеулер бар болса да, машиналық оқыту негізіндегі және елдің барлық ландшафттық спектрін қамтитын рельеф модельдері жоқ.

Халықаралық деңгейде жүргізілген фундаменталды зерттеулер кеңістіктік деректерді талдауда машиналық оқытудың артықшылығын ұзақ уақыттан бері дәлелдеді. Мысалы, 2014 жылы Blaschke, ГАЗ-да көппарадигмалық модельдеудің тұжырымдамасын ұсынды және күрделі деректерді өңдеуде Random Forest және Gradient Boosting әдістерінің тиімділігін көрсетті [5]. 2003 жылы Mark географиялық ақпараттық ғылымдағы жаңа мүмкіндіктерді машиналық оқытумен байланыстырады [6], ал 2019 жылы Lee осы әдістерді жоспарлау мен қоршаған ортаны бақылауға интеграциялаудың оң нәтижелерін ұсынды [7]. Әлемдік статистикаға сәйкес, машиналық оқыту алгоритмдерін қолдану гидрологиялық және геоморфологиялық болжамдардың дәлдігін 20% - 35% арттыруға мүмкіндік береді [8-10].

Соңғы жылдары Random Forest, Gradient Boosting және жасанды нейрондық желілер (ANN) сияқты құралдар үлкен деректерді өңдеуде орны толмас шешімдерге айналды. Random Forest әдісінің басты артықшылығы оның деректердегі артық шуды автоматты түрде сүзгілеу мүмкіндігі мен биіктікке әсер ететін ең маңызды факторларды анықтай алуы болып табылды [11], [12].

Gradient Boosting моделі жер бедерінің өзгерістерінің осал тұстарын дәл көрсетуге икемділік танытты. Ал нейрондық желілер таулы жерлердегі аумақтардағы геоморфологиялық құрылымын анықтауда, тануда, талдауда жоағыр тиімділік көрсетті [13], [14].

Зерттеудің жұмысында Қазақстанның табиғи – географиялық ерекшеліктері ескерілген. Сол ерекшеліктеріне бейімделген әртүрлі алгоритмдердің артықшылықтарын ұштастыратын болжау әдістерін ұсындық. Ұсынылған әдіс бағалау дәлдігін жақсартуға мүмкіндік береді.



### *Зерттеу материалдары мен әдістері*

Зерттеудің енгізгі міндеті Қазақстанның әртүрлі аймақтарының жер бедерінің биіктігін болжауға қабілетті алгоритм әзірлеп, мәліметтер жиынтығына ерекше мән берілді. Машиналық оқыту алгоритмдерінің жалпылау қабілетін жоғарлату мақсатында рельефтік ерекшеліктері бір-бірінен айқын ажыратылған бірнеше ландшафтық белдеулер қамтылды. Тәжірибелік аймақтар ретінде Солтүстік Тянь – Шаньның таулы жоталары, Ерейментау ласа таулы жүйесі, Жетісу – Алатау тау жоталары, өзіндік геоморфологиялық сипатқа ие Сарыарқа аумағы мен батыс өңіріндегі Маңғыстау үстіртінің тегіс аймақтары бар.

Зерттеу барысында Қазақстан Республикасының картография және геодезия қызметі ұсынған DEM – KZ жер бедерінің сандық моделі негізге алынды. Мәліметтер GeoTIFF форматында ұсынылған. Мәліметтер кеңістіктік ажыратымдылығы 30 метр. Жер бедерінің морфологиялық ерекшеліктерін жеткілікті дәлдікпен сипаттауға мүмкіндік береді [15].

Деректердің сенімділігін арттыру және олардың толықтығын қамтамасыз ету үшін спутниктік деректердің қосымша көздері қолданылды. Атап айтқанда NASA – ның SRTM деректері мен Sentinel-1 радиолокациялық суреттерді пайдаландық. Нәтижелерді тексеру үшін бақылау нүктелері ретінде 1:25000 масштабты топографиялық карталарды қолдандық. Python программалауда барлық баспатқы деректер өңделді. Бірыңғай дерекқорға біріктірдік. Нәтижесінде 5263 астам репрезентативтік нүктеден тұратын мәліметтер жиынтығы құрылды. Оның ішінде 4188 нүктесі (80%) – ы модельді оқытуға, ал қалған 1047 нүктесі (20%) – ы тестілеуге берілді.

#### *Мәліметтерді алдын ала өңдеу*

Машиналық оқыту алгоритмдерінің болжау дәлдігі мен тиімділігі алғашқы мәліметтердің сапасына байланысты болды. Мәліметтерді алдын ала өңдеудің бірнеше кезеңі қарастырылады. Онда мәліметтерді алдынала өңдеу жүргізіледі. Жерсеріктік түсірілімдерде көлеңке әсерінен атмосфералық жағдайлардан және өлшеу қателіктерінен туындаған ақау мәндердер анықталды. Ақау мәндері үзгілеу әдістері арқылы жойылды.

Оларға экспозиция, көлбеу бұрышы, жергілікті биіктік айырмашылықтары және көрші нүктелерге дейінгі орташа қашықтық жатады. Кеңістіктік қатынастарды сақтау үшін Делон триангуляциясы да қолданылды. Scipy модулімен жасалған 8шбұрыш торы `spatial.Delaunay()`, модельге нүктелердің геометриялық орналасуын түсіндіруге көмектесті.

#### *Модельдерді оқыту*

Зерттеу барысында біз машиналық оқытудың үш түрлі парадигмасын салыстырдық:

1. *Random Forest Regressor*: бұл модель 300 шешім ағашынан тұратын ансамбльдік әдіс (`n_estimators=300`). Оның басты ерекшелігі-жүктеу әдісі арқылы ішкі деректер үлгілерін жасау арқылы биіктікке әсер ететін факторларды автоматты түрде ажыратады. Біздің тәжірибемізде бұл модель ең кіші MAE (орташа абсолютті қате) көрсетті.

2. *Gradient Boosting Regressor*: градиентті күшейту (`learning_rate=0,05`) қателерді біртіндеп түзету арқылы тіпті ең кішкентай рельеф элементтерін ескеруге тырысады. Ол үлкен есептеу ресурстарын қажет етсе де, микрорельефті болжауда жоғары сезімталдықты көрсетті.



3. *Жасанды нейрондық желі (ANN):* жасанды нейрондық желі үш жасырын қабаттан (64-32-16 нейрон) жасалған. ReLU белсендіру функциясы мен Adam оптимизаторы қолданылды. Жасанды нейрондық желілер таулы аймақтардағы күрделі морфологияны тануда жақсы жұмыс істегенімен, олар деректері аз жазық жерлерде "қайта оқыту" белгілерін көрсетті. Төмендегі 1-ші кестеде машиналық оқытудың үш түрлі парадигмасын салыстырулары көрсетілген.

1 – Кесте – Машиналық оқытудың үш түрлі парадигмасын салыстыру

| Модельдер         | Параметрлер          |
|-------------------|----------------------|
| Random Forest     | n_estimators=300     |
| Gradient Boosting | learning_rate=0,05   |
| ANN               | 64-32-16, ReLU, Adam |

*Бағдарлама ортадағы бағалау көрсеткіштері*

Мәліметтер train – test split әдісі арқылы кездейсоқ 80-де 20 қатынасында тестілеуге, оқытуға бердік. Модельдерді болжаудың дәлдігін бағалау үшін жалпы көрсеткіштер қолданылды. Атап айтқанда, орташа абсолютті қате (MAE), орташа квадрат қате (RMSE) және анықтау коэффициенті (R2) есептелді. Бұл көрсеткіштер модельдердің нақты мәндерге қаншалықты жақын екенін және олардың түсіндіру қабілеті қандай екенін жан- жақты бағалауға мүмкіндік берді.

Деректерді есептеу мен талдау Python ортасында жүргізілді. Деректерді өңдеу үшін NumPy, Pandas кітапханалы қоладынылды. Машиналық оқыту модельдерін құру және бағалау үшін Scikit – learn пакетін пайдаландық. Нәтижелерді графиктік түрде көрсету үшін Matplotlib кітапханасын қолдандық. 1 суретте деректерді оқыту, тестілеу жиындары көрсетілген.

```
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.datasets import make_regression

# Үлгі деректерін жасау (шынайы қолданбаларда сіз өз деректеріңізді жүктейсіз)
X, y = make_regression(n_samples=100, n_features=10, random_state=42)

# Деректерді оқыту және тестілеу жиындарына бөлу
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

print(f"X_train niwini: {X_train.shape}")
print(f"y_train niwini: {y_train.shape}")
print(f"X_test niwini: {X_test.shape}")
print(f"y_test niwini: {y_test.shape}")

... X_train niwini: (80, 10)
y_train niwini: (80,)
X_test niwini: (20, 10)
y_test niwini: (20,)
```

1 – Сурет – Деректердің оқыту және тестілеу жиындары

Бұл рельефтің практикалық модельдерін жасау үшін машиналық оқыту алгоритмдерін тікелей қолдануға мүмкіндік береді.

*Зерттеу нәтижелері*

Зерттеу барысында алынған эксперименттік деректер машиналық оқыту алгоритмдері Қазақстанның өте күрделі және қарама-қарсы физика-географиялық жағдайларындағы биіктік мәндерін болжауға толық сәйкес келетінін көрсетті. Random Forest, Gradient Boosting және жасанды нейрондық желілер (ANN) модельдерінің нәтижелерін салыстырмалы талдау рельефтің морфологиялық құрылымына байланысты айтарлықтай айырмашылықтар мен заңдылықтарды анықтады. Біздің ойымызша, бұл айырмашылықтар модельдердің математикалық архитектурасымен және кеңістіктік деректерді өңдеу ерекшеліктерімен тікелей байланысты.



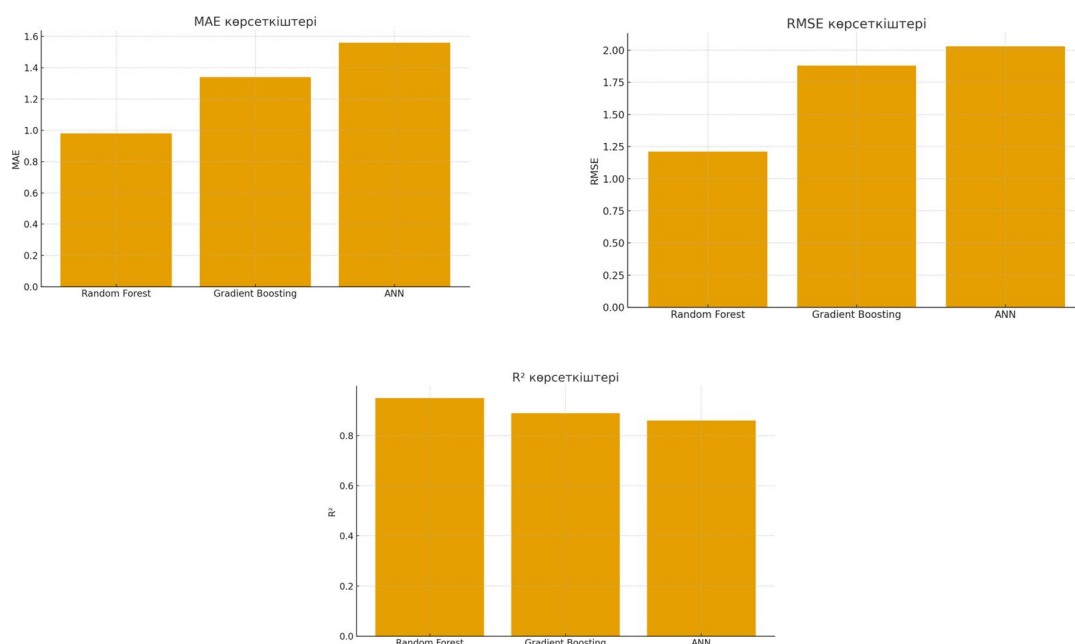
### Модельдің өнімділік көрсеткіштерін бағалау

Жүргізілген зерттеу жұмысы аясында үш модельдің тиімділігі жалпы қабылданған халықаралық регрессия көрсеткіштері - MAE, RMSE және  $R^2$  негізге алынды. Есептеу нәтижесі көрсеткендей, Random Forest моделі күрделі кеңістік сипаттамаларды жоғары деңгейде қабылдап, барлық сынақ айтмақтарында нақты және тұрақты көрсеткіштерді көрсетті. Бұл модельдің басты басымдылығы оның ансамбльдік жүйесі, яғни көптеген шешім ағақтарының орташаландыру арқылы рельефтің күтпеген өзгеруіне төзімділік көрсете алады. Модельдердің дәлдік деңгейін көрсететін салыстырмалы нәтижелері 2-ші кестеде көрсетілген.

2 – Кесте – Модельдердің болжамдық дәлдігінің салыстырмалы көрсеткіштері

| Модель               | MAE (м) | RMSE (м) | $R^2$ |
|----------------------|---------|----------|-------|
| Random Forest        | 0.98    | 1.21     | 0.95  |
| Gradient Boosting    | 1.34    | 1.88     | 0.89  |
| ANN (Нейрондық желі) | 1.56    | 2.03     | 0.86  |

Random Forest моделі барлық негізгі көрсеткіштер бойынша ең жоғары дәлдікті көрсетті. 2-ші суретте MAE, RMSE және  $R^2$  салыстырмалы нәтижелері көрсетілген.



2 – Сурет– MAE, RMSE және  $R^2$  көрсеткіштерінің салыстырмалы нәтижелері

Кестеден көріп отырғандарыңыздай, Random Forest моделі 0,98 метр орташа абсолютті қателікпен (MAE) алдыңғы қатарда тұр. Бұл көрсеткіш модельдің Қазақстан рельефінің микродеңгейлі вариацияларын жоғары дәлдікпен түсіре алатынын дәлелдейді. 1,21 М деңгейіндегі RMSE көрсеткіші модельдің үлкен қателіктерге (шығарындыларға) төзімділігін көрсетеді, ал значение  $R^2 = 0,95$  значение мәні оның биіктіктегі өзгерістердің 95% дәл болжай алатынын көрсетеді.

### Аймақтық ландшафтты талдау және модельді бейімдеу

Модельдердің тиімділігі географиялық аймақтардың морфологиялық күрделілігіне байланысты әртүрлі динамиканы көрсетті. Біз Қазақстан ландшафтының үш негізгі түрін қарастырдық:



Жазық аймақтар (Солтүстік және Батыс Қазақстан): бұл аймақтардағы рельеф салыстырмалы түрде біркелкі болғандықтан, биіктік айырмашылықтары аз болғандықтан, барлық модельдер қанағаттанарлық нәтиже берді. Дегенмен, нейрондық желілер (ANN) осындай біртекті деректерде “қайта оқыту” белгілерін көрсетті. Талдау нәтижелері күрделі архитектурасы бар модельдер оқу деректерінде қайта оқытуға бейім екенін көрсетті. Бұл жағдайда рельефтің кездейсоқ өзгеруі пайдалы үлгі ретінде қабылданды. Тестілеу жиынындағы болжамдардың сапасының төмендеуіне әкеледі. Сол себепті Random forest алгоритмі тұрақты нәтиже көрсетті. Тегіс жерде ең төменгі қате мәні ( $MAE \gg 0,7$  м).

### *Қорытынды*

Бұл жұмыста Қазақстанның кең және геоморфологиялық алуан түрлі аумағындағы географиялық нүктелердің биіктігін модельдеу үшін машиналық оқыту алгоритмдерінің мүмкіндіктерін жан-жақты бағалау ұсынылған. Алгоритмдерді тексеруден басқа, зерттеу сонымен қатар эмпирикалық цифрлық деректерге негізделген дәстүрлі геостатистикалық интерполяция әдістерінен деректерге бағытталған тәсілдердің артықшылығын көрсетті.

Салыстырмалы талдау кездейсоқ орман моделі барлық көрсеткіштер бойынша ең жоғары дәлдікке ие екенін көрсетті ( $MAE = 0,98$  м;  $RMSE = 1,21$  м;  $R^2 = 0,95$ ). Бұл модельдің ансамбльдік сипаты Қазақстанның әртүрлі ландшафттарынан алынған “шулы” деректермен жұмыс істеу кезінде тиімді нәтижелерге қол жеткізуге мүмкіндік берді. Оның көп факторлы кеңістіктік қатынастарды анықтау қабілеті оның сандық рельефтік модельдерді (DEM) құрудың ең сенімді әдісі екенін дәлелдеді.

### **ӘДЕБИЕТ**

- [1] Blaschke T., Merschdorf H. Geographic information science as a multidisciplinary and multiparadigmatic field // Cartography and Geographic Information Science. – 2014. – Vol. 41. – №3. – P. 196–213.
- [2] Lee D., Oh K. The Green Infrastructure Assessment System (GIAS) and its applications for urban development and management // Sustainability. – 2019. – Vol. 11. – №14. – Article 3798.
- [3] Scull P., Franklin J., Chadwick O.A. The application of classification tree analysis to soil type prediction in a mountainous area // Geoderma. – 2005. – Vol. 124. – №3–4. – P. 323–338.
- [4] Hengl T., Heuvelink G.B.M., Rossiter D.G. About regression-kriging: From equations to case studies // Computers & Geosciences. – 2007. – Vol. 33. – №10. – P. 1301–1315.
- [5] Breiman L. Random Forests // Machine Learning. – 2001. – Vol. 45. – P. 5–32.
- [6] Friedman J.H. Greedy function approximation: A gradient boosting machine // Annals of Statistics. – 2001. – Vol. 29. – №5. – P. 1189–1232.
- [7] LeCun Y., Bengio Y., Hinton G. Deep learning // Nature. – 2015. – Vol. 521. – P. 436–444.
- [8] Schmidt A., Gopal S., Shepherd J. Geographic random forest models for spatial prediction // GIScience & Remote Sensing. – 2020. – Vol. 57. – №1. – P. 1–17.
- [9] Farr T.G. және т.б. The Shuttle Radar Topography Mission // Reviews of Geophysics. – 2007. – Vol. 45. – RG2004.
- [10] Ковин Р.В., Марков Н.Г. Геоинформационные системы: учебное пособие. – Томск: Изд-во ТПУ, 2008. – 180 с.



[11] Яковенко Н.В., Марков Д.С., Туркина Е.П. ГИС-технологии в геоэкологических исследованиях [Электронный ресурс] // Современные проблемы науки и образования. – 2014. – №5. – Режим доступа: <https://science-education.ru/ru/article/view?id=15277>

[12] Plata I.T., Panganiban E.B. және т.б. Drone-based GIS mapping of cassava disease for precision agriculture // International Journal of Emerging Trends in Engineering and Advanced Technology (IJETAET). – 2022. – Vol. 12. – №2. – P. 1–9.

[13] Mambetqaziev E., Berdigozhin L. Qazaqstan tabigaty: oqu quraly. – Almaty: Qazaq universiteti, 2015. – 320 b.

[14] Qazaqstan Ulttyq Geodeziya zhane Kartografiya qyzmeti. Qazaqstannyn cifrlyq biiktik modeli (DEM–KZ) [Электрондық ресурс]. – 2024.

### REFERENCES

[1] Blaschke T., Merschdorf H. Geographic information science as a multidisciplinary and multiparadigmatic field // Cartography and Geographic Information Science. – 2014. – Vol. 41. – №3. – P. 196–213.

[2] Lee D., Oh K. The Green Infrastructure Assessment System (GIAS) and its applications for urban development and management // Sustainability. – 2019. – Vol. 11. – №14. – Article 3798.

[3] Scull P., Franklin J., Chadwick O.A. The application of classification tree analysis to soil type prediction in a mountainous area // Geoderma. – 2005. – Vol. 124. – №3–4. – P. 323–338.

[4] Hengl T., Heuvelink G.B.M., Rossiter D.G. About regression-kriging: From equations to case studies // Computers & Geosciences. – 2007. – Vol. 33. – №10. – P. 1301–1315.

[5] Breiman L. Random Forests // Machine Learning. – 2001. – Vol. 45. – P. 5–32.

[6] Friedman J.H. Greedy function approximation: A gradient boosting machine // Annals of Statistics. – 2001. – Vol. 29. – №5. – P. 1189–1232.

[7] LeCun Y., Bengio Y., Hinton G. Deep learning // Nature. – 2015. – Vol. 521. – P. 436–444.

[8] Schmidt A., Gopal S., Shepherd J. Geographic random forest models for spatial prediction // GIScience & Remote Sensing. – 2020. – Vol. 57. – №1. – P. 1–17.

[9] Farr T.G. және т.б. The Shuttle Radar Topography Mission // Reviews of Geophysics. – 2007. – Vol. 45. – RG2004.

[10] Ковин Р.В., Марков Н.Г. Геоинформационные системы: учебное пособие. – Томск: Изд-во ТПУ, 2008. – 180 с.

[11] Яковенко Н.В., Марков Д.С., Туркина Е.П. ГИС-технологии в геоэкологических исследованиях [Электронный ресурс] // Современные проблемы науки и образования. – 2014. – №5. – Режим доступа: <https://science-education.ru/ru/article/view?id=15277>

[12] Plata I.T., Panganiban E.B. және т.б. Drone-based GIS mapping of cassava disease for precision agriculture // International Journal of Emerging Trends in Engineering and Advanced Technology (IJETAET). – 2022. – Vol. 12. – №2. – P. 1–9.

[13] Mambetqaziev E., Berdigozhin L. Qazaqstan tabigaty: oqu quraly. – Almaty: Qazaq universiteti, 2015. – 320 b.

[14] Qazaqstan Ulttyq Geodeziya zhane Kartografiya qyzmeti. Qazaqstannyn cifrlyq biiktik modeli (DEM–KZ) [Электрондық ресурс]. – 2024.



**Е.А.Тулегенов, Н.О.Мекебаев, Ж.А.Орынтаева**  
**СРАВНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ПРИ ПРОГНОЗИРОВАНИИ**  
**ВЫСОТЫ ГЕОГРАФИЧЕСКИХ ТОЧЕК**

**Аннотация.** В данном исследовании всесторонне рассматриваются современные вопросы прогнозирования высотных показателей рельефа на основе интеграции географических информационных систем (ГИС) и алгоритмов машинного обучения. На сегодняшний день актуальность создания высокоточных цифровых моделей рельефа для территории Казахстана крайне высока. Это объясняется, прежде всего, уникальной геоморфологической сложностью страны: ландшафт Казахстана, где низменности Прикаспия соседствуют с высокогорными системами Тянь-Шаня и Алтая, представляет собой серьезный вызов для любой математической модели.

С практической точки зрения такая природная неоднородность значительно снижает эффективность классических методов геостатистической интерполяции таких как кригинг или метод обратных взвешенных расстояний (IDW). Стандартные подходы не способны в полной мере учитывать градиенты рельефа, что приводит к значительным погрешностям. Для устранения этих недостатков в ходе исследования было решено использовать современные подходы, ориентированные на данные (data-driven). Предметом исследования являются прогрессы регрессивного прогнозирования высотных точек путем внедрения пространственных данных из ГИС-платформ в модели машинного обучения.

Полученные результаты показали, что методы машинного обучения обеспечивают точность на 20-35% выше по сравнению с традиционными подходами. В ходе сравнительного анализа модель Random Forest продемонстрировала наивысшую стабильность и способность к обобщению при работе со сложными наборами данных.

В заключение следует отметить, что предложенная методика закладывает основу для вывода цифровой картографии Казахстана на качественно новый уровень. Результаты исследования рекомендуются к внедрению в геодезическое проектирование, гидрологическое моделирование, экологический мониторинг и государственные системы управления природными ресурсами.

**Ключевые слова:** Географические точки, машинное обучение, модель высот, рельеф, пространственная интерполяция, ГИС, Random Forest.

**E.Tulegenov, N.Mekebaev, Zh. Oryntaeva**  
**COMPARISON OF MACHINE LEARNING METHODS IN PREDICTING THE**  
**HEIGHT OF GEOGRAPHICAL POINTS**

**Annotation.** This study comprehensively examines contemporary issues in predicting topographic elevation through the integration of Geographic Information Systems (GIS) and Machine Learning algorithms. Currently, the development of high-precision Digital Elevation Models (DEM) for the territory of Kazakhstan is of paramount importance. This is primarily due to the unique geomorphological complexity of the country; the Kazakh landscape, ranging from the Caspian depressions to the high-altitude systems of the Tien Shan and Altai, poses a significantly challenge for any mathematical model.

From a practical standpoint, such natural heterogeneity significantly reduces the effectiveness of classical geostatistical interpolation methods, such as Kriging or Inverse Distance Weighting (IDW). Standard approaches often fail to account for sharp topographic gradients, leading to substantial errors. To address these shortcomings, this research employs modern data-driven approaches. The subject of the study is the process of regression forecasting of elevation points by integrating spatial data from GIS platforms into machine learning models.

The primary objective is to identify the most optimal model adapted to the specific characteristics of the local terrain. To this end, three different architectures were tested: Random Forest, Gradient Boosting, and Artificial Neural Networks (ANN). In the initial stage, GIS data



underwent preprocessing and normalizations, after which the predictive capacity of each model was compared against actual measurement points.

The results indicated that machine learning methods provide 20-35% higher accuracy compared to traditional techniques. During the comparative analysis, the Random Forest model datasets. This model allowed for a more accurate description of local terrain fluctuations.

In conclusion, the proposed methodology paves the way for elevating digital cartography in Kazakhstan to a qualitatively new level. It is recommended to implement these research findings in geodetic design, hydrological modeling, environmental monitoring, and state systems for transformation aimed at the efficient use of the country's land resources.

**Keywords:** geographical points, machine learning, elevation model, terrain, spatial interpolation, GIS, Random Forest.