



SANA LOGOS – Республикалық ғылыми журнал., Том 1, №3-2026

ӘОЖ: 57:004.8

<https://doi.org/10.5281/zenodo.23477058>**Биологиялық зерттеулердегі жасанды интеллект және цифрлық әдістер**Ж.Ж. Рысбекова^{1*} , М.Н. Мәулен¹ 

¹Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан, Қазақстан
e-mail: rysbekova.zhanel2026@ayu.edu.kz

Қабылданған күні: 18.07.2026 | Қабылданды: 16.09.2026 | Жарияланды: 30.09.2026

Аңдатпа. Биологияда жоғары өнімді секвенирлеу, біржасушалық омика, сандық микроскопия, қашықтан зондтау және автоматтандырылған сенсорлар зерттеу деректерінің көлемі мен күрделілігін күрт арттырды. Осы әдеби шолудың мақсаты – жасанды интеллект пен цифрлық әдістердің биологиялық зерттеулердегі негізгі қолданылу бағыттарын, олардың ғылыми мүмкіндіктерін және сенімді қолдану шарттарын жүйелеу. Рецензияланған еңбектерді тақырыптық синтездеу нәтижесінде жасанды интеллект геномдық варианттарды анықтау, жасушаларды сегменттеу, ақуыз құрылымын болжау, біржасушалық деректерді интеграциялау, жаңа молекулаларды іріктеу, биоалуантүрлілікті мониторингтеу және зертханалық циклдерді автоматтандыруда елеулі нәтиже беретіні көрсетілді. Сонымен қатар жоғары болжамдық көрсеткіш биологиялық механизмнің дәлелденгенін білдірмейді: дерек ағуы, таңдаманың ығысуы, домендер арасындағы айырмашылық, түсіндірмеліліктің шектеулілігі және код пен модельдердің жабықтығы нәтижелерді жалпылауға кедергі келтіреді. Әдебиеттер синтезі жасанды интеллектіні биологті алмастыратын құрал емес, дерек сапасы, сыртқы валидация, пәндік сараптама және қайталанымдылық стандарттарымен басқарылатын зерттеу инфрақұрылымы ретінде қарастыру қажет екенін айқындайды.

Түйін сөздер: жасанды интеллект, машиналық оқыту, терең оқыту, биоинформатика, геномика, сандық микроскопия, биоалуантүрлілік мониторингі, зертханалық автоматтандыру

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ЦИФРОВЫЕ МЕТОДЫ В БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Аннотация. Высокопроизводительное секвенирование, одноклеточная омика, цифровая микроскопия, дистанционное зондирование и автоматизированные сенсоры привели к резкому увеличению объёма и сложности исследовательских данных в биологии. Цель данного обзора литературы – систематизировать основные направления применения искусственного интеллекта и цифровых методов в биологических исследованиях, их научные возможности и условия надёжного использования. В результате тематического синтеза рецензируемых научных публикаций показано, что искусственный интеллект обеспечивает значимые результаты при выявлении геномных вариантов, сегментации клеток, прогнозировании структуры белков, интеграции одноклеточных данных, скрининге новых молекул, мониторинге биоразнообразия и автоматизации лабораторных циклов. Вместе с тем высокие показатели прогнозирования не означают, что биологический механизм доказан: утечка данных, смещение выборки, различия между доменами, ограниченная интерпретируемость, а также закрытость кода и моделей препятствуют обобщению результатов. Синтез литературы показывает, что искусственный интеллект следует рассматривать не как инструмент замены биолога, а как исследовательскую инфраструктуру,



SANA LOGOS – Республикалық ғылыми журнал., Том 1, №3-2026

функционирующую на основе контроля качества данных, внешней валидации, экспертной оценки в предметной области и соблюдения стандартов воспроизводимости.

Ключевые слова: искусственный интеллект, машинное обучение, глубокое обучение, биоинформатика, геномика, цифровая микроскопия, мониторинг биоразнообразия, лабораторная автоматизация.

ARTIFICIAL INTELLIGENCE AND DIGITAL METHODS IN BIOLOGICAL RESEARCH

Abstract. High-throughput sequencing, single-cell omics, digital microscopy, remote sensing, and automated sensors have dramatically increased the volume and complexity of data generated in biological research. This literature review aims to systematize the main applications of artificial intelligence and digital methods in biological research, their scientific potential, and the conditions required for their reliable use. A thematic synthesis of peer-reviewed studies demonstrates that artificial intelligence delivers significant results in genomic variant detection, cell segmentation, protein structure prediction, single-cell data integration, screening of novel molecules, biodiversity monitoring, and laboratory workflow automation. However, high predictive performance does not establish that a biological mechanism has been proven: data leakage, sampling bias, domain shifts, limited interpretability, and the lack of transparency of code and models hinder the generalizability of results. The literature synthesis indicates that artificial intelligence should be regarded not as a tool for replacing biologists, but as a research infrastructure governed by data quality control, external validation, domain expertise, and reproducibility standards.

Keywords: artificial intelligence, machine learning, deep learning, bioinformatics, genomics, digital microscopy, biodiversity monitoring, laboratory automation.

Қысқашы

Қазіргі биологиялық зерттеу тек зертханалық бақылаумен шектелмейді. Геномдық секвенирлеу платформалары миллиардтаған нуклеотидтік оқылымдарды, микроскоптар мыңдаған кескіндерді, ал камера-тұзақтар, дрондар және акустикалық сенсорлар үздіксіз экологиялық деректерді өндіреді. Мұндай массивтерді дәстүрлі қолмен талдау баяу әрі субъективті болуы мүмкін. Сондықтан цифрлық деректерді сақтау, өңдеу және бөлісу технологиялары мен жасанды интеллект әдістері биологияның зерттеу логикасына біртіндеп енді. Терең оқыту күрделі бейне, реттілік және сигнал деректерінен пайдалы белгілерді автоматты түрде үйренуге мүмкіндік берсе, классикалық машиналық оқыту салыстырмалы түрде шағын, құрылымдалған деректерде әлі де маңызды орын алады (LeCun et al., 2015; Greener et al., 2022).

Жасанды интеллект ұғымы бұл шолуда машиналық оқыту, терең нейрондық желілер, генеративті модельдер және экспериментті жоспарлауға көмектесетін

алгоритмдерді қамтиды. Ал цифрлық әдістерге секвенирлеу, биоинформатикалық құбырлар, сандық бейнелеу, геоақпараттық жүйелер, қашықтан зондтау, сенсорлық мониторинг, бұлттық есептеу, электрондық зертханалық жазбалар және роботтандырылған эксперименттер жатады. Осы екі топ бір-бірінен бөлек емес: цифрлық құралдар стандартталған дерек ағынын жасайды, ал жасанды интеллект сол деректерден жіктеу, болжау, сегменттеу немесе оңтайландыру нәтижесін алады.

Мақаланың мақсаты – биологиялық зерттеулердегі жасанды интеллект пен цифрлық әдістердің жетістіктерін тізбектеу ғана емес, әр қолданбаның ғылыми дәлелге айналу шарттарын салыстырмалы түрде талдау. Негізгі сұрақтар: бұл технологиялар қандай биологиялық міндеттерді тиімді шешеді; қай жерде олардың қорытындысы шектеулі; және сенімді, қайталанатын зерттеу үшін қандай әдіснамалық қағидалар қажет?

Әдебиеттерді іріктеу тәсілі

Бұл жұмыс жүйелі шолу мәртебесіне үміткер емес, тақырыптық әдеби шолу



SANA LOGOS – Республикалық ғылыми журнал., Том 1, №3-2026

ретінде дайындалды. 2015–2026 жылдардағы еңбектерге басымдық беріліп, бағыттың қалыптасуына ықпал еткен ертерек жұмыс қосылды. Іздеу ұғымдары қазақша және ағылшынша баламалармен құрылды: artificial intelligence in biology, machine learning genomics, protein structure prediction, single-cell foundation model, deep learning microscopy, AI biodiversity monitoring, bioacoustics, self-driving laboratory, data leakage және reproducibility. Google Scholar-да индекстелетін рецензияланған журнал мақалалары негізге алынды; атау, автор, жыл, журнал және DOI мәліметтері журналдардың ресми беттері мен библиографиялық жазбалар арқылы қайта тексерілді.

Іріктеуде үш өлшем қолданылды: биологиялық зерттеу міндетіне тікелей қатыстылық; әдістің немесе проблеманың ғылыми қауымдастық үшін маңыздылығы; нәтижені тексеруге мүмкіндік беретін тұрақты DOI не ресми баспа бетінің болуы. Зерттеу мақалалары нақты қолданбаларды көрсету үшін, ал шолу және ұсыным сипатындағы еңбектер салалық тенденциялар мен әдіснамалық шектеулерді синтездеу үшін пайдаланылды. Тек аннотациясы қолжетімді еңбектерден егжей-тегжейлі сандық нәтиже шығарылмады. Сондықтан төмендегі тұжырымдар «барлық әдебиетті толық қамтиды» дегенді емес, іріктелген сенімді еңбектердің тақырыптық синтезін білдіреді.

Биологиялық деректердің цифрлық трансформациясы

Цифрландыру биологиялық нысанды өлшеу тәсілін өзгертті. ДНҚ мен РНҚ реттіліктері символдық тізбекке, тіп кескіндері пиксельдер жиынына, жануар дауысы уақыт–жиілік спектріне, ал экожүйе координаталармен байланысқан кеңістіктік қабаттарға айналады. Мұндай репрезентациялар компьютерлік өндеуді жеңілдетеді, бірақ биологиялық шындықтың толық көшірмесі емес. Өлшеу құралы, үлгі дайындау хаттамасы, сенсордың сезімталдығы және деректі аннотациялау тәсілі кейінгі модельге жүйелі қате енгізуі мүмкін. Сондықтан цифрлық құбырдың сапасы тек алгоритм дәлдігімен емес, үлгі

алудан бастап метадеректерді сақтауға дейінгі бүкіл тізбекпен анықталады.

Машиналық оқытуда мақсат көбіне белгісіз үлгілерге болжам жасау, ал классикалық биостатистикада параметрді бағалау мен гипотезаны тексеру болады. Бұл айырмашылық практикалық мәнге ие. Модель түрді кескіннен жоғары дәлдікпен таныса да, популяция санының өзгеру себебін түсіндірмеуі мүмкін. Сол сияқты ген экспрессиясынан фенотипті болжау геннің себептік рөлін автоматты түрде дәлелдемейді. Pichler және Hartig (2023) экологияда болжау мен ғылыми қорытынды жасаудың мақсаттарын ажырату қажет екенін атап көрсетеді; бұл қағида биологияның өзге салаларына да ортақ.

Геномика және біржасушалық биология

Геномика жасанды интеллектіні ерте әрі кең қабылдаған бағыттардың бірі. Терең желілер ДНҚ тізбегіндегі функционалдық элементтерді тану, варианттарды шақыру, реттеуші белсенділікті болжау және көпомикалы деректерді біріктіру міндеттерінде қолданылады. Eraslan және әріптестері (2019) терең оқытудың үлкен әрі күрделі геномдық деректерден бейсызық тәуелділіктерді үйрене алатынын, бірақ модель интерпретациясы мен дерек қажеттілігі негізгі қиындық болып қалатынын көрсетті. Жалпы әдістемелік шолулар да биолог үшін модель таңдау міндеттен, дерек көлемінен және түсіндіру талабынан басталуы тиіс екенін айтады (Greener et al., 2022).

DeepVariant геномдық оқылымдардың локалдық туралануын кескінге ұқсас репрезентацияға түрлендіріп, бір нуклеотидтік варианттар мен шағын инделдерді терең желі арқылы жіктейді. Porlin және әріптестері (2018) әдістің әртүрлі секвенирлеу технологиялары мен эксперименттік дизайндарға бейімделе алатынын көрсетті. Бұл нәтиже биоинформатикалық ережелерді толық қолмен жазудың орнына, модельдің шынайы деректен пайдалы белгілерді үйрену мүмкіндігін танытты. Дегенмен өнімділік



SANA LOGOS – Республикалық ғылыми журнал., Том 1, №3-2026

референстік геномға, қамту тереңдігіне, платформаға және этникалық репрезентативтілікке тәуелді; клиникалық немесе популяциялық қорытынды алдында тәуелсіз дерекпен валидация қажет.

Біржасушалық RNA-seq әр жасушаның мыңдаған ген бойынша экспрессиясын өлшейді, бірақ сиректік, техникалық шу және партиялық әсер талдауды қиындатады. scVI ықтималдық генеративті модель арқылы төмен өлшемді жасырын кеңістік құрып, партиялық әсерді есепке алу мен жасушалық вариацияны талдауға мүмкіндік берді (Lopez et al., 2018). Кейін scGPT 33 миллионнан астам жасуша профилінде алдын ала оқытылған трансформерді ұсынып, жасуша типін аннотациялау, гендік желілерді зерттеу және көпомикалы міндеттерге бейімдеу мүмкіндігін көрсетті (Cui et al., 2024). Мұндай іргелі модельдер қайта пайдаланылатын репрезентация ұсынады, бірақ олардың көлемі биологиялық мағыналылыққа тең емес: оқу корпусындағы тіндер, аурулар және популяциялар тең ұсынылмаса, сирек жағдайларға қатысты қорытынды әлсіз болуы мүмкін.

Құрылымдық биология және дәрі іздеу

Ақуыз құрылымын аминқышқылдар тізбегінен болжау жасанды интеллектінің биологиядағы ең танымал жетістігіне айналды. AlphaFold2 CASP14 бағалауында көптеген нысандар үшін эксперименттік құрылымдарға жақын дәлдік көрсетті (Jumper et al., 2021). Кейін әдіс адам протеомының басым бөлігіне қолданылып, құрылымдық қамтуды айтарлықтай кеңейтті (Tunyasuvunakool et al., 2021). AlphaFold Protein Structure Database болжамдарды кең қауымдастыққа ашық ұсыну арқылы цифрлық инфрақұрылымның ғылыми ықпалын күшейтті (Varadi et al., 2022).

Сонымен бірге болжанған құрылым мен эксперименттік дәлелді теңестіруге болмайды. Модельдің сенімділік көрсеткіші төмен аймақтар, ретсіз домендер, балама конформациялар, лигандтар, посттрансляциялық модификациялар және

жасушалық орта толық сипатталмауы мүмкін. AlphaFold гипотеза қалыптастыруды, мутация орнын таңдауды немесе экспериментті басымдыққа қоюды жеделдетеді, бірақ функцияны, динамиканы және өзара әрекеттесуді растау үшін криоэлектрондық микроскопия, рентгендік кристаллография, ЯМР және биохимиялық сынақтар қажет болып қала береді.

Дәрі-дәрмек іздеуде машиналық оқыту нысана таңдаудан бастап молекула қасиеттерін болжауға, виртуалды скринингке және токсикологияға дейін қолданылады. Vamathevan және әріптестері (2019) мүмкіндіктің мол екенін, бірақ пайдалы нәтиже нақты қойылған сұрақ пен сапалы дерекке тәуелді екенін көрсетеді. Stokes және әріптестері (2020) терең оқыту арқылы құрылымдық жағынан бұрынғы антибиотиктерден өзгеше молекулаларды басымдыққа қоюға болатындығын көрсетті. Бұл модельдер зертханалық скрининг кеңістігін тарылтады; олар молекуланың тиімділігі мен қауіпсіздігін өздігінен дәлелдемейді. Сондықтан есептік болжам міндетті түрде *in vitro*, кейін қажет болса *in vivo* тексерумен жалғасуы тиіс.

Сандық микроскопия және биобейнелер

Сандық микроскопияда жасушаларды, ядроларды немесе ұлпа құрылымдарын қолмен белгілеу көп уақыт алады және зерттеушілер арасында айырмашылық туғызады. Терең оқыту кескін жіктеу, объектіні сегменттеу, траекторияны қадағалау және кескін сапасын қалпына келтіру міндеттерін автоматтандырды (Moen et al., 2019). Cellpose әртүрлі кескін типтерінде жасушаларды сегменттеуге арналған жалпылама модель ретінде ұсынылып, алдын ала қайта оқытусыз да кең қолдануға бейімделді (Stringer et al., 2021). Cellpose 2.0 адам қатысатын қайта оқыту арқылы жаңа бейне доменіне салыстырмалы аз аннотациямен бейімделуді жеңілдетті (Pachitariu & Stringer, 2022).

Биобейне модельдерінің осал тұсы – домен ауысуы. Басқа микроскоп, бояу, үлгі дайындау немесе жасуша морфологиясы



SANA LOGOS – Республикалық ғылыми журнал., Том 1, №3-2026

үйретілген модельдің қатесін арттыра алады. Сегменттеу қатесі кейінгі жасуша санына, морфометрияға және фенотиптік жіктеуге тізбекті түрде таралады. Сондықтан сапаны тек орташа дәлдікпен емес, биологиялық маңызды сирек объектілердегі қателермен, сыртқы зертхананың дерегімен және сарапшының визуалды тексеруімен бағалау қажет. Жасанды интеллект бұл жерде бақылаушыны толық алып тастамай, қайталанатын жұмысты азайтып, күмәнді мысалдарды адамға қайта бағыттайтын жүйе ретінде тиімді.

Экология және биоалуантүрлілікті цифрлық мониторингтеу

Экологияда камера-тұзақ, дрон, спутник, GPS және пассивті акустикалық құрылғылар кеңістіктік-уақыттық деректердің жаңа ағындарын қалыптастырады. Терең оқыту кескіннен түрді анықтау, жануар санын есептеу, мінез-құлықты жіктеу және дыбыстан вокализацияны тану міндеттеріне енгізілді (Christin et al., 2019). Norouzzadeh және әріптестері (2018) Snapshot Serengeti жиынтығындағы 3,2 миллион кескін бойынша 48 түрді анықтау, санын есептеу және сипаттау мүмкіндігін көрсетті. Бұл камера-тұзақ мониторингін масштабтауға жол ашты, бірақ модельдің бір аймақтағы жоғары нәтижесі басқа ландшафт, маусым немесе камера параметріне автоматты түрде көшпейді.

Өсімдік түрлерін кескін арқылы анықтау азаматтық ғылым мен далалық жұмысты қолдай алады. Wäldchen және Mäder (2018) автоматтандырылған өсімдік идентификациясының дамуын қорытындылап, жапырақтан бөлек гүл, жеміс, сабақ және экологиялық контекст сияқты белгілердің маңызын көрсетті. Практикалық жүйе пайдаланушыға бір атауды ғана емес, ықтимал нұсқаларды, сенімділік деңгейін және шешімге әсер еткен белгілерді ұсынуы керек. Сирек және жергілікті түрлердің оқу жиынында аз болуы қорғау басымдығы жоғары таксондарда қате тәуекелін күшейтеді.

Биоакустика құстарды, жарқанаттарды, жәндіктерді және теңіз

жануарларын ұзақ мерзім, жанаспай бақылауға мүмкіндік береді. Stowell (2022) терең оқытудың дыбыстық оқиғаларды табу мен жіктеуді ілгерілеткенін, алайда фондық шу, әлсіз белгі, таксондар арасындағы теңсіздік және басқа орынға тасымалдау мәселелері шешілмегенін атап өтеді. Экология мен эволюциядағы жүздеген зерттеуді қамтыған синтез терең оқытудың әсіресе кескіндік жіктеуде тез таралғанын көрсетті, бірақ механистік түсіндіру мен болжамдық өнімділікті ажырату керектігін ескертеді (Borowiec et al., 2022). Демек цифрлық мониторингтің ең сенімді үлгісі – сенсор, алгоритм, далалық тексеру және экологиялық интерпретацияның біріккен циклі.

Зертханалық автоматтандыру және тұйық циклді эксперимент

Зертханалық автоматтандырудың бастапқы мақсаты сұйықтық беру, инкубациялау және өлшеу сияқты қайталанатын әрекеттерді стандарттау болды. Жасанды интеллект қосылғанда жүйе алдыңғы нәтижеге сүйеніп келесі экспериментті таңдауға қабілетті тұйық циклге өтеді. King және әріптестері (2009) ұсынған Adam робот-ғалымы ашытқы функционалдық геномикасында гипотеза құрып, оны автоматты эксперимент арқылы тексерудің принциптік мүмкіндігін көрсетті. Бұл жұмыс ғылыми процестің кейбір формалды бөліктерін машинаға беруге болатынын дәлелдеді, бірақ ғылыми сұрақтың маңызын бағалау мен нәтижені кең теорияға енгізу адам сараптамасын талап етеді.

Қазіргі өздігінен басқарылатын зертханалар роботтық өндірісті, жоғары өнімді өлшеуді және белсенді оқыту немесе Байес оңтайландыруын біріктіреді. Протеин инженериясында алгоритм перспективалы варианттарды таңдап, робот оларды жасап өлшейді, ал жаңа дерек келесі циклді жаңартады. Мұндай тәсіл эксперимент санын азайтып, үлкен іздеу кеңістігінде пайдалы аймақты жылдам табуы мүмкін. Алайда жабдық сәйкестігі, калибрлеу, ақауларды анықтау, қауіпсіздік және дерек стандарттары



SANA LOGOS – Республикалық ғылыми журнал., Том 1, №3-2026

ескерілмесе, автоматтандыру қатені де міндеттің күрделілігіне, тәуекеліне және жоғары жылдамдықпен қайталайды. өлшеудің сенімділігіне сай болуы керек. Сондықтан автономия деңгейі биологиялық

1-кесте. Биологиялық зерттеулердегі негізгі цифрлық деректер, әдістер және валидация талаптары

Дерек түрі	Негізгі әдіс	Зерттеуге қосатын құндылық	Басты тексеру талабы
ДНК, РНК, омика	CNN, трансформер, вариациялық автоэнкодер	Вариант шақыру, жасуша типін ажырату, дерек интеграциясы	Тәуелсіз когорты, партиялық әсер, туыстық пен қайталанған үлгілерді бөлу
Ақуыз және молекула	Құрылымдық және генеративті модельдер	Құрылым болжамы, нысана мен кандидаттарды басымдыққа қою	Эксперименттік құрылым, биохимиялық белсенділік және қауіпсіздік сынағы
Микроскопиялық кескін	Сегменттеу және объектіні қадағалау	Жасушаларды автоматты санау, морфология мен динамиканы өлшеу	Басқа микроскоп пен хаттамада сыртқы валидация, сирек объектілерді тексеру
Экологиялық кескін мен дыбыс	Компьютерлік көру және акустикалық жіктеу	Түрді анықтау, санақ, мінез-құлық және фенология мониторингі	Жаңа орын, маусым және сенсорда тексеру; жалған оң нәтижені есепке алу
Роботтық эксперимент	Белсенді оқыту, Байес оңтайландыруы	Келесі тәжірибені бейімдеп таңдау, цикл уақытын қысқарту	Калибрлеу, бақылау үлгілері, қауіпсіз тоқтату және толық журналдау



Сенімділік, түсіндірмелілік және этика

Жасанды интеллект қолданылған биологиялық зерттеудің ең жиі қатесі – модельдің тест нәтижесін шынайы ғылыми жарамдылықпен теңестіру. Дерек ағуы кезінде бір дараға, бір отбасыға, бір зертханаға немесе бір үлгі сериясына қатысты ақпарат оқу және тест жиындарына қатар түседі. Сонда модель биологиялық заңдылықтың орнына техникалық ұқсастықты жаттап, нәтижені жасанды түрде көтереді. Bernett және әріптестері (2024) биологиялық машиналық оқытуда ағуды анықтауға арналған жеті бағыттаушы сұрақ ұсынады. Практикада бөлу бірлігі биологиялық тәуелсіздікке сай болуы керек: мысалы, кескінді емес, пациентті; дыбыс үзіндісін емес, жазба орнын; жасушаны емес, донорды бөлу.

Ашық дерекқорлар белгілі түрлерді, организмдерді, популяцияларды және зертханаларды шамадан тыс көрсетуі мүмкін. Модель осы таралымды үйреніп, сирек түрлерде немесе аз зерттелген популяцияларда нашарлайды. Бұл тек техникалық проблема емес: дерек жинау басымдығы ғылыми назарды қайта өндіреді. Сондықтан өнімділікті жалпы орташа көрсеткішпен қатар, маңызды ішкі топтарда, жаңа географиялық аймақта және басқа құрылғыда бағалау қажет.

Түсіндірмелілік пен себептілік меселесі белгі маңыздылығы немесе назар картасы модельдің қандай сигнал пайдаланғанын шамалап көрсетеді, бірақ биологиялық механизмнің дәлелі болмайды. Болжамнан кейін нокаут, мутагенез, тежегіш эксперименті немесе тәуелсіз далалық бақылау секілді тексеру қажет. Walsh және әріптестері (2021) ұсынған DOME ұсынымдары дерек, оңтайландыру, модель және бағалау туралы толық есеп беруге шақырады. Neil және әріптестері (2021) дерек, код, үйретілген салмақтар, бағдарламалық орта және автоматтандырылған жұмыс ағынын

жариялаудың қайталанымдылық үшін маңызын көрсетеді.

Геномдық дерек тұлғаны қайта сәйкестендіруге мүмкіндік беруі мүмкін, ал биологиялық дизайн модельдері пайдалы молекулалармен қатар қауіпті сипаттамаларды да оңтайландыра алады. Сондықтан қолжетімділік деңгейі, деректі анонимдеу, аудит журналы, рөлдік рұқсат және тәуекелге сай адам бақылауы алдын ала жобалануы тиіс. FAIR қағидалары деректің табылатын, қолжетімді, үйлесімді және қайта пайдаланылатын болуын қолдайды, бірақ «ашық» пен «жауапты қолжетімділік» әрдайым бір нәрсе емес (Wilkinson et al., 2016).

Талқылау

Қарастырылған әдебиеттер жасанды интеллектінің биологиядағы әсері біркелкі емес екенін көрсетеді. Ең айқын жетістіктер үлкен стандартталған дерек, нақты мақсат және объективті эталон бар міндеттерде байқалады: ақуыз құрылымын болжау, нуклеотид варианттарын шақыру, жасуша сегменттеу немесе алдын ала белгіленген түрлерді кескіннен жіктеу. Ал себептік механизм, сирек жағдай, жаңа орта немесе көпдеңгейлі жүйе қажет болғанда белгісіздік артады. Сондықтан «ең күрделі модель ең жақсы модель» деген қағида ғылыми тұрғыдан дұрыс емес; қарапайым базалық модель, анық валидация және түсінікті қате талдауы кейде көбірек дәлел береді.

Цифрлық әдістердің басты құндылығы жекелеген алгоритмнен гөрі зерттеу циклін байланыстыруында. Сенсор немесе секвенатор дерек жинайды; стандартталған құбыр сапаны тексереді; модель ықтимал нысандарды басымдыққа қояды; эксперимент немесе далалық тексеру болжамды сынайды; нәтиже дерекқор мен модельді жаңартады. Осы циклде адам зерттеу сұрағын, өлшемнің жарамдылығын, этикалық шекті және биологиялық интерпретацияны анықтайды. Яғни болашақтағы тиімді үлгі – толық



автономия емес, адам мен есептік жүйенің міндеттері дәл бөлінген серіктестік.

Қазақстандық зерттеу ұйымдары үшін де бұл қорытынды өзекті. Қымбат үлкен модельді нөлден үйретуден бұрын ашық құралды жергілікті дерекпен тексеру, сапалы метадерекпен электрондық хаттама қалыптастыру, есептеу ортасын құжаттау және пәндік маман мен деректер маманының бірлескен жұмысын күшейту пайдалы. Биоалуантүрлілік саласында жергілікті түрлердің сурет және дыбыс коллекциялары, ал биомедицинада популяциялық репрезентативті деректер ерекше маңызға ие. Мұндай дерекқорларсыз сырттан алынған модельдің қателігі жасырын қалуы мүмкін.

Қорытынды

Жасанды интеллект және цифрлық әдістер биологиялық зерттеудің дерек жинау, талдау, гипотеза құру және экспериментті жоспарлау сатыларын өзгертіп отыр. Геномикада олар күрделі реттілік деректерін өңдейді; құрылымдық биологияда ақуыз

пішінін болжауды кеңейтеді; микроскопияда сандық фенотиптеуді жеделдетеді; экологияда үлкен аумақта үздіксіз мониторингке мүмкіндік береді; автоматтандырылған зертханада тәжірибені бейімдеп таңдауға көмектеседі. Бірақ бұл артықшылықтар деректің сапасы мен репрезентативтілігіне, дұрыс бөлінген оқу және тест жиындарына, сыртқы валидацияға, толық есеп беруге және биологиялық экспериментпен растауға тәуелді.

Әдебиеттердің ортақ тұжырымы мынадай: жасанды интеллект ғылыми әдісті алмастырмайды, оның кейбір кезеңдерін күшейтеді. Сенімді қолдану үшін әр болжамның дерек көзі, белгісіздігі, қолданылу шекарасы және тексеру жолы көрсетілуі керек. Болашақ зерттеулер мультимодальды деректерді біріктіруге, аз ұсынылған организмдер мен популяцияларды қамтуға, сыртқы валидацияға, түсіндірілетін модельдерге және ашық, қайталанатын жұмыс ағындарына бағытталғаны жөн.



Әдебиеттер тізімі

1. Bennett, J., Blumenthal, D. B., Grimm, D. G., Haselbeck, F., Joeres, R., Kalinina, O. V., & List, M. (2024). Guiding questions to avoid data leakage in biological machine learning applications. *Nature Methods*, 21, 1444–1453. <https://doi.org/10.1038/s41592-024-02362-y>
2. Borowiec, M. L., Dikow, R. B., Frandsen, P. B., McKeeken, A., Valentini, G., & White, A. E. (2022). Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*, 13(8), 1640–1660. <https://doi.org/10.1111/2041-210X.13901>
3. Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10), 1632–1644. <https://doi.org/10.1111/2041-210X.13256>
4. Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., & Wang, B. (2024). scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods*, 21, 1470–1480. <https://doi.org/10.1038/s41592-024-02201-0>
5. Eraslan, G., Avsec, Ž., Gagneur, J., & Theis, F. J. (2019). Deep learning: New computational modelling techniques for genomics. *Nature Reviews Genetics*, 20, 389–403. <https://doi.org/10.1038/s41576-019-0122-6>
6. Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology*, 23, 40–55. <https://doi.org/10.1038/s41580-021-00407-0>
7. Heil, B. J., Hoffman, M. M., Markowetz, F., Lee, S.-I., Greene, C. S., & Hicks, S. C. (2021). Reproducibility standards for machine learning in the life sciences. *Nature Methods*, 18, 1132–1135. <https://doi.org/10.1038/s41592-021-01256-7>
8. Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
9. King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., et al. (2009). The automation of science. *Science*, 324(5923), 85–89. <https://doi.org/10.1126/science.1165620>
10. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>
11. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., & Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15, 1053–1058. <https://doi.org/10.1038/s41592-018-0229-2>
12. Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., & Van Valen, D. (2019). Deep learning for cellular image analysis. *Nature Methods*, 16, 1233–1246. <https://doi.org/10.1038/s41592-019-0403-1>
13. Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 115(25), E5716–E5725. <https://doi.org/10.1073/pnas.1719367115>
14. Pachitariu, M., & Stringer, C. (2022). Cellpose 2.0: How to train your own model. *Nature Methods*, 19, 1634–1641. <https://doi.org/10.1038/s41592-022-01663-4>
15. Pichler, M., & Hartig, F. (2023). Machine learning and deep learning— A review for ecologists. *Methods in Ecology and Evolution*, 14(4), 994–1016. <https://doi.org/10.1111/2041-210X.14061>



16. Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36, 983–987. <https://doi.org/10.1038/nbt.4235>
17. Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., et al. (2020). A deep learning approach to antibiotic discovery. *Cell*, 180(4), 688–702.e13. <https://doi.org/10.1016/j.cell.2020.01.021>
18. Stowell, D. (2022). Computational bioacoustics with deep learning: A review and roadmap. *PeerJ*, 10, e13152. <https://doi.org/10.7717/peerj.13152>
19. Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2021). Cellpose: A generalist algorithm for cellular segmentation. *Nature Methods*, 18, 100–106. <https://doi.org/10.1038/s41592-020-01018-x>
20. Tunyasuvunakool, K., Adler, J., Wu, Z., et al. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, 596, 590–596. <https://doi.org/10.1038/s41586-021-03828-1>
21. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., et al. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18, 463–477. <https://doi.org/10.1038/s41573-019-0024-5>
22. Varadi, M., Anyango, S., Deshpande, M., et al. (2022). AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1), D439–D444. <https://doi.org/10.1093/nar/gkab1061>
23. Wäldchen, J., & Mäder, P. (2018). Machine learning for image based species identification. *Methods in Ecology and Evolution*, 9(11), 2216–2225. <https://doi.org/10.1111/2041-210X.13075>
24. Walsh, I., Fishman, D., Garcia-Gasulla, D., et al. (2021). DOME: Recommendations for supervised machine learning validation in biology. *Nature Methods*, 18, 1122–1127. <https://doi.org/10.1038/s41592-021-01205-4>
25. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>



SANA LOGOS – Республикалық ғылыми журнал., Том 1, №3-2026

ӘОЖ: 58:004.8

<https://doi.org/10.5281/zenodo.23477056>

Өсімдіктер систематикасындағы жасанды интеллект цифрлық гербарийлер түрлерді анықтау және таксономиялық зерттеулердің жаңа мүмкіндіктері (әдеби шолу)

А.А. Құдайберген^{1*} ,¹әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстанe-mail: aiauzhankudaibergen@gmail.com**Қабылданған күні: 29.07.2026 | Қабылданды: 23.09.2026 | Жарияланды: 30.09.2026**

Аңдатпа. Бұл әдеби шолудың мақсаты – цифрлық гербарийлер мен жасанды интеллект әдістерінің өсімдіктер систематикасындағы орнын, түрлерді автоматтандырылған анықтау мүмкіндіктерін және таксономиялық зерттеулерге ықпалын кешенді бағалау. Әдебиеттер Google Scholar арқылы табуға болатын рецензияланған жарияланымдар арасынан тақырыптық сәйкестік бойынша іріктеліп, DOI және баспа беттерімен салыстырыла тексерілді. Негізгі іздеу кезеңі 2015–2026 жылдарды қамтыды, ал тұжырымдамалық маңызы бар ертерек еңбектер қосымша енгізілді. Нәтижелер цифрлық гербарийлердің сурет пен жапсырма метадеректерін біріктіретін ауқымды зерттеу инфрақұрылымына айналғанын көрсетті. Терең оқыту үлгілері ірі және сапалы таңбаланған деректер жиындарында түрлерді жоғары дәлдікпен ажырата алады; сонымен бірге өсімдік мүшелерін сегментациялау, морфологиялық белгілерді өлшеу, фенологиялық фазаларды тану, жапсырма мәтінін шығару және ықтимал қате анықтамаларды белгілеу мүмкіндіктері кеңеюде. Алайда дәлдік көрсеткіштері деректердің теңгерімсіздігіне, географиялық және коллекциялық ығысуға, сирек таксондардың аз ұсынылуына, таксономиялық атаулардың өзгеруіне және модельдің бұрын көрмеген түрлерді тану қабілетіне тәуелді. Сондықтан жасанды интеллект таксономисті алмастыратын құрал ретінде емес, сарапшы басқаратын, шешім қабылдауды қолдайтын жүйе ретінде қарастырылуы тиіс. Болашақ бағыттар мультимодальды деректерді, түсіндірілетін модельдерді, ашық классты тануды және халықаралық дерек стандарттарын біріктіруге негізделеді.

Түйін сөздер: жасанды интеллект, өсімдіктер систематикасы, цифрлық гербарий, компьютерлік көру, түрлерді анықтау, машиналық оқыту, интегративті таксономия.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В СИСТЕМАТИКЕ РАСТЕНИЙ: ЦИФРОВЫЕ ГЕРБАРИИ, ИДЕНТИФИКАЦИЯ ВИДОВ И НОВЫЕ ВОЗМОЖНОСТИ ТАКСОНОМИЧЕСКИХ ИССЛЕДОВАНИЙ (ЛИТЕРАТУРНЫЙ ОБЗОР)

Аннотация. Цель данного литературного обзора – комплексно оценить роль цифровых гербариев и методов искусственного интеллекта в систематике растений, возможности автоматизированной идентификации видов и влияние этих технологий на таксономические



исследования. Литература отбиралась по тематическому соответствию среди рецензируемых публикаций, доступных для поиска через Google Scholar, с последующей перекрёстной проверкой DOI и страниц издательств. Основной период поиска охватывал 2015–2026 годы, при этом дополнительно включались более ранние работы, имеющие концептуальное значение. Результаты показали, что цифровые гербарии превратились в масштабную исследовательскую инфраструктуру, объединяющую изображения образцов и метаданные с этикеток. Модели глубокого обучения способны с высокой точностью различать виды на крупных и качественно аннотированных наборах данных. Кроме того, расширяются возможности сегментации органов растений, измерения морфологических признаков, распознавания фенологических фаз, извлечения текста с этикеток и выявления потенциально ошибочных определений. Однако показатели точности зависят от несбалансированности данных, географических и коллекционных смещений, недостаточной представленности редких таксонов, изменений таксономической номенклатуры и способности моделей распознавать ранее не встречавшиеся виды. Поэтому искусственный интеллект следует рассматривать не как инструмент замены таксономиста, а как систему поддержки принятия решений под контролем эксперта. Перспективные направления дальнейших исследований связаны с интеграцией мультимодальных данных, разработкой интерпретируемых моделей, применением методов распознавания открытого класса и внедрением международных стандартов данных.

Ключевые слова: искусственный интеллект, систематика растений, цифровой гербарий, компьютерное зрение, идентификация видов, машинное обучение, интегративная таксономия.

ARTIFICIAL INTELLIGENCE IN PLANT SYSTEMATICS: DIGITAL HERBARIA, SPECIES IDENTIFICATION, AND NEW OPPORTUNITIES FOR TAXONOMIC RESEARCH (A LITERATURE REVIEW)

Abstract. The aim of this literature review is to comprehensively assess the role of digital herbaria and artificial intelligence methods in plant systematics, the potential of automated species identification, and the impact of these technologies on taxonomic research. The literature was selected from peer-reviewed publications searchable through Google Scholar based on thematic relevance, with DOI records and publisher websites cross-checked for verification. The main search period covered 2015–2026, while earlier studies of conceptual significance were also included. The findings indicate that digital herbaria have evolved into large-scale research infrastructures integrating specimen images with label metadata. Deep learning models can distinguish species with high accuracy when trained on large, high-quality annotated datasets. In addition, capabilities for plant organ segmentation, morphological trait measurement, phenological stage recognition, label text extraction, and the flagging of potentially erroneous identifications are expanding. However, accuracy metrics depend on data imbalance, geographic and collection biases, the underrepresentation of rare taxa, changes in taxonomic nomenclature, and the ability of models to recognize previously unseen species. Therefore,



artificial intelligence should be regarded not as a replacement for taxonomists but as an expert-supervised decision-support system. Future research directions include the integration of multimodal data, interpretable models, open-set recognition, and international data standards.

Keywords: *artificial intelligence, plant systematics, digital herbarium, computer vision, species identification, machine learning, integrative taxonomy.*

Кіріспе

Өсімдіктер систематикасы өсімдік алуантүрлілігін сипаттау, атау, жіктеу және эволюциялық байланыстарын түсіндіру міндеттерін атқарады. Бұл міндеттердің эмпирикалық негізі – түрдің морфологиялық белгілерін, таралуын, жиналған орнын және таксономиялық анықтамаларын сақтайтын гербарий үлгілері. Гербарий парағы тек кептірілген өсімдік емес: ол қайта тексеруге болатын заттай куәлік, жапсырмадағы кеңістік-уақыттық дерек және кейінгі ревизиялардың тарихы. Сондықтан гербарий коллекцияларын цифрландыру систематикадағы деректердің қолжетімділігі мен салыстырмалылығын түбегейлі өзгертті (Soltis, 2017; Meineke et al., 2019).

Жаппай цифрландыру нәтижесінде миллиондаған жоғары ажыратымдылықтағы кескіндер GBIF, iDigBio және институционалдық виртуалды гербарийлер арқылы таратылады. Бұл қорлар зерттеушіге бір коллекциямен шектелмей, әр аймақ пен тарихи кезеңдегі үлгілерді қатар талдауға мүмкіндік береді. Бірақ цифрлық дерек көлемінің өсуі жаңа кедергі туғызды:

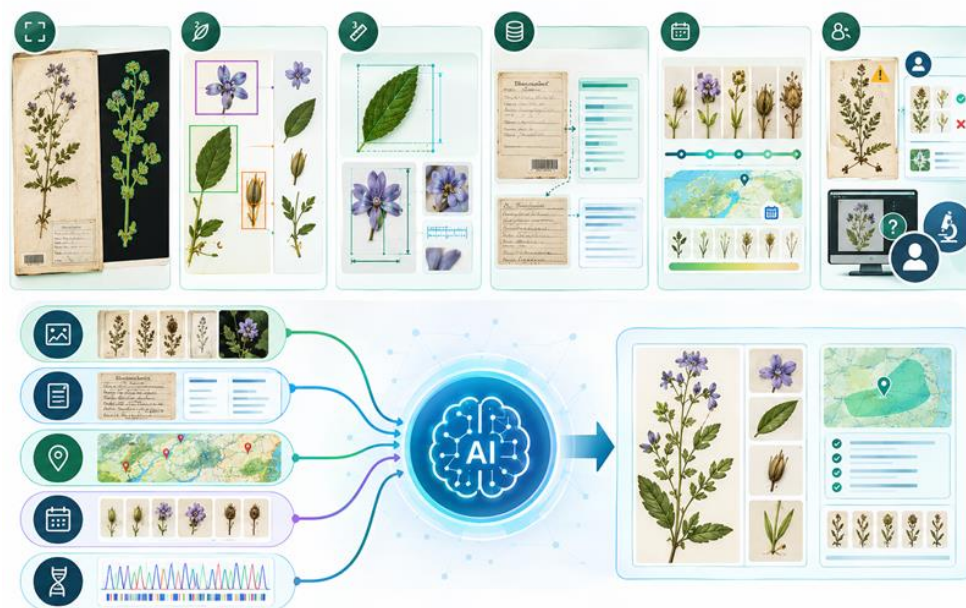
суреттерді қолмен анықтау, жапсырмаларды транскрипциялау, мүшелерді белгілеу және қате жазбаларды тексеру қарқынды дерек ағынына ілесе алмайды. Осы тұста компьютерлік көру, табиғи тілді өңдеу және машиналық оқыту жүйелері қайталанатын операцияларды жеделдететін құрал ретінде маңызға ие болды (Pearson et al., 2020; Hussein et al., 2022).

Жасанды интеллектінің өсімдік систематикасындағы қолданылуы тек «суреттен түр атауын табумен» шектелмейді. Қазіргі жүйелер гербарий парағындағы өсімдік пен фон элементтерін бөледі, жапырақ, гүл, жеміс сияқты мүшелерді табады, морфометриялық белгілерді өлшейді, жапсырма мәтінін құрылымдандырады, фенологиялық күйді бағалайды және коллекциядағы күмәнді анықтамаларды сарапшыға ұсынады (1-сурет). Соңғы еңбектер мұны бірнеше дерек түрін – кескін, мәтін, география, фенология және генетикалық ақпаратты – біріктіретін интегративті таксономиялық AI бағытына ұластырады (Guo et al., 2025; Takano et al., 2026).



SANA LOGOS

SANA LOGOS – Республикалық ғылыми журнал., Том 1, №3-2026



1-сурет. Жасанды интеллекттің өсімдік систематикасындағы қолданылуы: интегративті таксономиялық талдау

Мақаланың мақсаты – цифрлық гербарийлерге негізделген жасанды интеллект әдістерінің түрлерді анықтау мен таксономиялық зерттеулерге қосатын үлесін әдеби деректер негізінде жүйелеу, олардың сенімділік шектерін көрсету және өсімдіктер систематикасындағы болашақ зерттеу бағыттарын айқындау.

Материалдар мен әдістер

Жұмыс эмпирикалық эксперимент емес, құрылымдалған нарративтік әдеби шолу ретінде орындалды. РСС логикасы қолданылды: зерттеу нысаны – гербарий үлгілері мен өсімдік таксондары; негізгі ұғым – жасанды интеллект, машиналық оқыту және компьютерлік көру; контекст – цифрлық гербарийлер, түрлерді анықтау және таксономиялық зерттеу. Үш сұрақ қойылды: 1) цифрлық гербарий деректері AI үшін қандай инфрақұрылым қалыптастырады; 2) AI өсімдік

түрін анықтау және гербарийді өңдеу міндеттерінде қандай нәтижелер көрсетеді; 3) бұл әдістерді таксономиялық шешімге айналдыруға қандай шектеулер кедергі жасайды? Ақпарат іздеу стратегиясы Google Scholar-да табылатын ғылыми еңбектерді қамту мақсатымен ағылшын тіліндегі мына ұғымдар жеке және біріктіріліп пайдаланылды: “artificial intelligence”, “machine learning”, “deep learning”; “plant taxonomy”, “plant systematics”, “species identification”, “species delimitation”; “digital herbarium”, “digitized herbarium specimen”; “computer vision”; “label extraction”; “trait extraction”; “open-set recognition”. Нәтижелердің библиографиялық дұрыстығы журналдың ресми беті, DOI жазбасы, PubMed Central немесе ұйымдық репозиторий арқылы салыстырыла тексерілді. Тікелей тақырыптық еңбектердің әдебиеттер тізімдері бойынша кері



дәйексөздік іздеу де жүргізілді. Google Scholar библиографиялық іздеу құралы ретінде пайдаланылғанымен, басылымның сенімділігі оның Scholar-да көрінуімен ғана емес, рецензиялануы, баспасы, әдісінің ашықтығы және дерекке сәйкестігі арқылы бағаланды.

Деректерді шығару және синтездеу

Әр еңбек бойынша авторы, жылы, дерек көзі, зерттеу мақсаты, дерек жиынтығының сипаты, модель немесе аналитикалық тәсіл, негізгі нәтиже, авторлар көрсеткен шектеулер және систематикаға қолданылуы тіркелді. Дәлдік, үлгі саны және таксон саны тек бастапқы мақалада немесе ресми толық мәтінде нақты көрсетілген жағдайда ғана пайдаланылды. Әртүрлі міндеттердің метрикалары бір-біріне тікелей теңестірілмеді: классификация дәлдігі, объектіні анықтау көрсеткіші және белгіні өлшеу сапасы бөлек түсіндірілді. Зерттеулер әдістемелік гетерогенді болғандықтан мета-талдау жасалмады; нәтижелер тақырыптық синтез арқылы цифрлық инфрақұрылым, автоматты идентификация, дерек шығару, таксономиялық интерпретация және тәуекелдер топтарына бөлінді. Бұл жұмыс толық жүйелі шолу немесе PRISMA есебі ретінде ұсынылмайды, себебі барлық дерекқорлар бойынша тәуелсіз қосарланған скрининг жүргізілген жоқ.

Нәтижелер

Цифрлық гербарийлер AI үшін дерек инфрақұрылымы ретінде

Гербарийлерді цифрландыру физикалық коллекцияны әлемдік деңгейде іздеуге болатын дерек жүйесіне айналдырды. Цифрлық жазба әдетте үлгі кескінін, таксон атауын, жинаушыны, күнді, географиялық

орынды, коллекция нөмірін және кейінгі анықтау белгілерін қамтиды. Осындай құрылым модельді оқытуға қажетті кескін-таңба жұптарын береді. Soltis (2017) цифрландырудың тарихи үлгілерді жаңа зерттеу сұрақтарына қайта пайдалануға жол ашқанын көрсетті; Meineke et al. (2019) табиғи-тарихи коллекцияларды антропоцендегі биоалуантүрлілік өзгерістерін түсіндіретін ұзақ мерзімді дерек ретінде бағалады.

AI үшін гербарий кескіндерінің маңызды артықшылығы – көпшілігінің мамандар берген таксономиялық атаумен байланысуы. Сонымен қатар бұл атаулар толық мінсіз «ақиқат белгі» емес: кейбір жазбалар ескірген номенклатураны сақтайды, кейбір үлгілер қате анықталған, ал бір таксон әр коллекцияда әртүрлі атпен берілуі мүмкін. Цифрландыру қарқыны да біркелкі емес; ірі мекемелер мен кең таралған түрлер басым, шағын гербарийлер, сирек түрлер және географиялық тұрғыдан аз зерттелген аймақтар жеткіліксіз көрінеді. Демек, цифрлық гербарий – дайын және бейтарап оқу жиынтығы емес, кураторлық тарихы бар дерек көзі.

Түрлерді автоматтандырылған анықтау

Carranza-Rojas et al. (2017) гербарий кескіндерін терең нейрондық желілермен жіктеудің жүзеге асатынын ірі масштабта көрсетті. Бұл еңбек кейінгі гербарийлік компьютерлік көру зерттеулеріне негіз болды. Little et al. (2020) ұйымдастырған Herbarium 2019 байқауында Melastomataceae тұқымдасының 46 469 үлгісі мен 683 түрі қолданылып, үздік төрт модель 88%-дан жоғары дәлдікке жетті. Мұндай нәтиже алгоритмдердің маман жұмысын жылдамдата



алатынын көрсеткенімен, деректер бір жаңа, гибридті, толық емес немесе таксономиялық топпен шектелді және анықталмаған материалмен кездеседі. түрлердің оқу мысалдары тең бөлінбеді.

Жапония флорасының 500 554 кескіні мен 2171 таксонында жасалған жүйе 96,4% дәлдік көрсетті және модель назар аударған кескін бөліктерін талдау арқылы қате анықталуы ықтимал үлгілерді белгілеуге мүмкіндік берді (Takano et al., 2022). Бұл нәтиже жоғары сапалы, аймақтық тұрғыдан үйлесімді коллекцияда AI өте тиімді болатынын дәлелдейді. Алайда көрсеткішті басқа құрлықтарға, басқа гербарий стандарттарына немесе оқу жиынтығында жоқ түрлерге тікелей көшіруге болмайды. Жабық классты сынақта барлық тест түрлері алдын ала белгілі болады; нақты гербарий жұмысында жүйе

Ірі деректердегі өкілдік оқыту да маңызды бағыт болып отыр. Walker et al. (2022) бес гербарийден алынған 2,25 миллионнан астам кескін мен 64 500 түрді пайдаланып, кескіндердің жалпыға тасымалданатын сандық сипаттамаларын үйренді. Үштік желі арқылы алынған кеңістік бірнеше кейінгі міндетке жақсырақ ауысқан. Мұндай морфологиялық кеңістік ұқсас үлгілерді топтауға, атипиялық жазбаларды табуға және сарапшыға салыстыру үшін жақын үлгілерді ұсынуға мүмкіндік береді.

Қолдану бағыты	Негізгі дерек	AI нәтижесі	Таксономиялық мәні
Түрді жіктеу	Гербарий парағының кескіні	Таксонға ықтималдық береді	Алдын ала анықтау және ревизияны басымдықтау
Сегментация	Өсімдік пен парақ элементтері	Өсімдікті фоннан бөледі	Шу әсерін азайтып, өлшеуді тұрақтандырады
Мүше тану	Жапырақ, гүл, жеміс, сабақ	Мүшелерді табады және санайды	Диагностикалық белгілер мен фенологияны шығарады
Мәтін шығару	Жапсырма және аннотация	OCR және NER арқылы өрістерді бөледі	Локалитет, күн, жинаушы және анықтамаларды құрылымдайды
Морфологиялық кеңістік	Көп үлгінің үйренілген белгілері	Ұқсастықтар мен аномалияларды көрсетеді	Түр шекарасын тексеруге гипотеза ұсынады

1-кесте. Цифрлық гербарийлердегі AI міндеттері және олардың систематикадағы қызметі

Морфологиялық белгілер мен LeafMachine және оның кейінгі нұсқалары фенологиялық ақпаратты автоматты шығару жапырақтарды табу, өлшеу және Таксономия үшін тек соңғы түр атауы морфологиялық белгілерді көп мөлшерде емес, оны негіздейтін белгілер маңызды. шығару жұмысын автоматтандырады (Weaver et al., 2020). GinJinn сияқты объектіні анықтау құбырлары Younis et al. (2020) жапырақ пен гербарий кескіндеріндегі құрылымдарды сабақтың жақсы танылғанын, ал ұсақ немесе автоматты табуға арналған (Ott et al., 2020). қабаттасқан гүл мүшелерінің қиынырақ



болғанын көрсетті. Бұл айырмашылық таксономиялық белгілердің визуалды қолжетімділігі модель нәтижесіне тікелей әсер ететінін білдіреді.

Фенологиялық кезеңдерді анықтау гүлдеу және жеміс беру уақытын тарихи қатарлар бойынша зерттеуге мүмкіндік береді. Pearson et al. (2020) гербарий кескіндерінен фенологиялық дерек алуға арналған модульдік машиналық оқыту жұмыс үдерісін сипаттады. Мұнда модель адам еңбегін түгел жоймайды: сарапшы оқу дерегін белгілейді, модель көп кескінді өңдейді, ал қателері бақылау іріктемесінде бағаланады. Осы схема таксономиялық зерттеуге де сәйкес келеді, себебі морфологиялық белгіні өлшеу мен оның таксономиялық мағынасын түсіндіру екі бөлек кезең болып қала береді.

Жапсырма мәтіні мен коллекциялық метадеректерді өңдеу

Гербарий парағындағы жапсырма түр атауымен қатар үлгінің қай жерде, қашан және кім жинағанын көрсетеді. OCR мәтінді таниды, ал атаулы мәндерді тану алгоритмдері сол мәтіннен ел, елді мекен, координата, жинаушы, күн және коллекция нөмірі секілді өрістерді бөледі. Соңғы жүйелер объектіні анықтау, OCR және табиғи тілді өңдеуді бір құбырға біріктіріп, қолжазба, типографиялық әркелкілік және көптілді жапсырмалар жағдайында жұмыс істеуге ұмтылады. Мұндай автоматтандыру үлгілерді картаға түсіруді, синонимдерді салыстыруды және тарихи анықтамаларды іздеуді жеделдетеді, бірақ нәтижелерді Darwin Core сияқты стандарттарға сәйкестендіріп, адамның сапа бақылауынан өткізу қажет.

Түр шекаралары және интегративті таксономия

Классификатор белгілі атаулардың бірін таңдайды, ал түрді делимитациялау деректердегі шекаралардың өзін тексереді. Сондықтан автоматты идентификация мен жаңа түрді сипаттау бір міндет емес. Бақылаусыз оқыту, кластерлеу және өлшемділікті азайту морфологиялық немесе генетикалық кеңістіктегі табиғи топтарды көрсетуге көмектеседі, бірақ кластер өздігінен биологиялық түр болып саналмайды. Species delimitation 4.0 тұжырымдамасы AI көмегімен морфология, геномика, экология және географияны бірге талдап, субъективтілікті азайтуды ұсынады; түпкі шешім бірыңғай түр тұжырымдамасы мен таксономиялық сараптамаға сүйенуі керек.

Takano et al. (2026) ұсынған интегративті таксономиялық AI тәсілінде жеке кескінге бір атау беру емес, көптеген үлгіні ортақ үйренілген морфологиялық кеңістікте салыстыру басым. Бұл кеңістік түрішілік өзгергіштікті, ықтимал гибридтерді, шеткі үлгілерді, қате анықтамаларды және қазіргі түр шекарасына сәйкес келмейтін топтарды көрсетуі мүмкін. AI мұнда жаңа таксономиялық гипотеза түзеді; гипотеза типтік материал, диагноз, номенклатура, география, молекулалық филогения және сарапшының морфологиялық бағалауымен тексеріледі.

Талқылау

Ғылыми және практикалық артықшылықтар

Әдебиеттер AI-дың ең күшті әсері ауқым мен басымдықта екенін көрсетеді. Біріншіден, модель миллиондаған кескін арасынан белгілі



бір таксонға, мүшеге немесе фенологиялық күйге сәйкес жазбаларды жылдам іріктейді. Екіншіден, күмәнді немесе атипиялық үлгілерді белгілеп, сарапшы уақытын ең күрделі жағдайларға бағыттайды. Үшіншіден, бірдей алгоритмдік ереже көптеген коллекцияға қолданылғанда өлшеу қайталанымдылығы жоғарылайды. Төртіншіден, цифрлық кескіндер физикалық үлгіні қайта-қайта ұстау қажеттігін азайтып, сирек және типтік материалды қашықтан салыстыруға жағдай жасайды.

Жасанды интеллект әсіресе «адам–алгоритм» бірлескен үдерісінде тиімді. Модель алғашқы болжамды және сенімділік мәнін береді; таксономист диагностикалық белгілерді, атаулардың қолданылуын және типтік үлгілерді тексереді; расталған түзетулер қайтадан оқу дерегіне қосылады. Осындай белсенді оқыту циклі сирек таксондарға арналған таңбаланған деректі біртіндеп жақсартады және модель қатесін жасырмай, кураторлық шешімге айналдырады.

Сенімділік шектері және әдіснамалық тәуекелдер

Бірнеше кең таралған түрдің мыңдаған кескіні бар, ал сирек түрлердің бірен-саран үлгісі ғана кездеседі. Жалпы дәлдік жоғары болғанымен, сирек немесе қорғауды қажет ететін таксондағы қате үлесі жасырын қалуы мүмкін. Сондықтан macro-F1, әр таксон бойынша сезімталдық, top-k дәлдігі және калибрленген сенімділік секілді көрсеткіштер қажет. Модель өсімдік морфологиясымен қатар гербарий қағазының түсін, масштаб сызғышын, штрихкодты, мекеме мөрін немесе кескін түсіру стилін үйренуі мүмкін. Бір

коллекцияда жақсы нәтиже көрсеткен жүйе басқа гербарийде әлсірейді.

Көп модель тест кескінін оқу кезінде көрген түрлердің біріне мәжбүрлеп жатқызады. Нақты таксономиялық жұмыста анықталмаған, жаңа, гибриді немесе оқу жиынтығынан тыс таксондар бар. Ашық классты тану жүйесі мұндай үлгіні «белгісіз» деп тоқтатып, сарапшыға жіберуі тиіс. Жылу картасы модель назар аударған аймақты көрсете алады, бірақ ол ботаникалық диагностикалық белгіні пайдаланғанын толық дәлелдемейді. Сондықтан түсіндірме морфологиялық негізбен және қарсы мысалдармен тексерілуі керек (Guo et al., 2025).

Гербарий қорлары жинау экспедицияларының, қолжетімді жолдардың, саяси шекаралардың және ғылыми басымдықтардың ізін сақтайды. AI осы үлгіні қайталайды: деректе аз аймақ пен аз таксон модельде де әлсіз болады. Бұған қоса, қабылданған атаулар мен синонимдер уақыт өте өзгереді. Оқу жиынтығын тұрақты таксономиялық анықтамалықпен нұсқалау, бастапқы verbatim identification өрісін сақтау және модель шығарған атаудың қай номенклатуралық нұсқаға сәйкес екенін тіркеу қажет.

Зерттеу инфрақұрылымына қойылатын талаптар

Сенімді жүйе құру үшін кескіннің тұрақты идентификаторы, коллекция коды, таксон атауының авторы, анықтаушы мен анықтау күні, географиялық дерек, лицензия және таксономиялық нұсқа сақталуы тиіс. Оқу, валидация және тест жиынтықтарын кездейсоқ кескін деңгейінде ғана бөлу бір үлгінің



көшірмелері немесе бір жинау оқиғасының сериясы арқылы дерек ағып кетуіне әкелуі мүмкін. Сондықтан бөлу үлгі идентификаторы, мекеме, география немесе уақыт бойынша жүргізіліп, сыртқы коллекцияда сынақ жасалғаны дұрыс. Код, модель салмақтары, дерек іріктеу ережелері және қате матрицалары ашық берілгенде ғана нәтиже қайталанып, басқа флораға бағаланады.

Қазақстан және Орталық Азия гербарийлері үшін басым қадам – бірден күрделі универсал модель жасау емес, сапалы цифрлық корпус қалыптастыру. Алдымен типтік және флористикалық маңызды үлгілерді жоғары сапада түсіру, жапсырмаларды екі немесе үш тілде құрылымдау, синонимдерді біріздендіру және сарапшы растаған шағын эталон жиынтықтарын құру қажет. Одан кейін аймақтық модельді сыртқы гербарий материалымен сынап, сирек түрлер бойынша қателерді бөлек көрсету керек. Бұл тәсіл жергілікті флораны халықаралық дерек желілеріне қосып, сонымен бірге жергілікті таксономиялық білімнің жоғалуына жол бермейді.

Бұл шолу барлық библиографиялық дерекқорды толық қамтыған жүйелі шолу емес. Негізгі назар ағылшын тіліндегі, онлайн тексеруге болатын рецензияланған әдебиетке аударылды; сондықтан жергілікті тілдердегі және индекстелуі әлсіз еңбектер толық көрінбеуі мүмкін. Өртүрлі зерттеу деректері, таксон саны және метрикалары біркелкі болмағандықтан, нәтижелер сандық түрде біріктірілмеді. Кейбір жаңа еңбектер 2025–2026 жылдары жарияланғандықтан, олардың

тәуелсіз қайталануы мен ұзақ мерзімді әсері әлі толық бағаланбаған.

Қорытынды

Цифрлық гербарийлер өсімдіктер систематикасына жоғары көлемді, тексерілетін және тарихи терең дерек берді; жасанды интеллект сол деректі іздеу, жіктеу және өлшеу мүмкіндігін кеңейтті. Әдебиеттер терең оқыту жүйелерінің түрлерді анықтау, мүшелерді табу, белгілерді шығару, фенологияны бағалау және жапсырма мәтінін құрылымдау міндеттерінде нақты нәтижелер бергенін көрсетеді. Ең перспективалы өзгеріс – жеке кескінге атау беруден көптеген үлгіні ортақ морфологиялық және мультимодальды кеңістікте салыстыруға көшу.

Сонымен бірге AI нәтижесі таксономиялық акт емес. Жоғары дәлдік деректер жиынтығының шегінде ғана мағыналы; жаңа түрді тану, түр шекарасын өзгерту немесе номенклатуралық шешім қабылдау үшін типтік материал, диагноз, әдебиет, генетикалық және экологиялық деректер, сондай-ақ маманның жауапкершілігі қажет. Сондықтан болашақтағы тиімді модель – түсіндірілетін, белгісіз таксонды қабылдайтын, бірнеше дерек модальдігін біріктіретін және адам бақылауын жүйенің міндетті бөлігі ретінде сақтайтын модель. Осындай жағдайда AI өсімдіктер систематикасын алмастырмайды, оның дәлелдік ауқымын және зерттеу жылдамдығын арттырады.



Әдебиеттер тізімі

1. Carranza-Rojas, J., Goëau, H., Bonnet, P., Mata-Montero, E., & Joly, A. (2017). Going deeper in the automated identification of herbarium specimens. *BMC Evolutionary Biology*, 17, 181. <https://doi.org/10.1186/s12862-017-1014-z>
2. Guo, Y.-Y., Cai, H., Bramley, G. L. C., Atkins, H. J., Li, B., & Theodossiades, S. (2025). A review of artificial intelligence in herbarium specimen image analysis. *Artificial Intelligence Review*, 58, 402. <https://doi.org/10.1007/s10462-025-11408-2>
3. Hussein, B. R., Malik, O. A., Ong, W.-H., & Slik, J. W. F. (2022). Applications of computer vision and machine learning techniques for digitized herbarium specimens: A systematic literature review. *Ecological Informatics*, 69, 101641. <https://doi.org/10.1016/j.ecoinf.2022.101641>
4. Little, D. P., Tulig, M., Tan, K. C., Liu, Y., Belongie, S., Kaeser-Chen, C., Michelangeli, F. A., Panesar, K., Guha, R. V., & Ambrose, B. A. (2020). An algorithm competition for automatic species identification from herbarium specimens. *Applications in Plant Sciences*, 8(6), e11365. <https://doi.org/10.1002/aps3.11365>
5. Meineke, E. K., Davies, T. J., Daru, B. H., & Davis, C. C. (2019). Biological collections for understanding biodiversity in the Anthropocene. *Philosophical Transactions of the Royal Society B*, 374, 20170386. <https://doi.org/10.1098/rstb.2017.0386>
6. Ott, T., Palm, C., Vogt, R., & Oberprieler, C. (2020). GinJinn: An object-detection pipeline for automated feature extraction from herbarium specimens. *Applications in Plant Sciences*, 8(6), e11351. <https://doi.org/10.1002/aps3.11351>
7. Pearson, K. D., Nelson, G., Aronson, M. F. J., Bonnet, P., Brenskelle, L., Davis, C. C., Denny, E. G., Ellwood, E. R., Goëau, H., Heberling, J. M., Joly, A., Lorieul, T., Mazer, S. J., Meineke, E. K., Stucky, B. J., Sweeney, P., & White, A. E. (2020). Machine learning using digitized herbarium specimens to advance phenological research. *BioScience*, 70(7), 610–620. <https://doi.org/10.1093/biosci/biaa044>
8. Soltis, P. S. (2017). Digitization of herbaria enables novel research. *American Journal of Botany*, 104(9), 1281–1284. <https://doi.org/10.3732/ajb.1700281>
9. Takano, A., Moon, H., Shirai, M., Tagane, S., Vasques, D. T., Lee, C.-K., Khan, A., Seo, J., Han, S., Park, C., Lee, J., Akihiro, T., & Kim, S. (2026). Herbarium specimens in the age of artificial intelligence: From herbarium image identification to Integrative Taxonomic AI. *Journal of Plant Research*. <https://doi.org/10.1007/s10265-026-01743-y>
10. Takano, A., et al. (2022). Development of a system for the automated identification of herbarium specimens with high accuracy. *Scientific Reports*, 12, 8066. <https://doi.org/10.1038/s41598-022-11450-y>
11. Walker, B. E., Tucker, A., & Nicolson, N. (2022). Harnessing large-scale herbarium image datasets through representation learning. *Frontiers in Plant Science*, 12, 806407. <https://doi.org/10.3389/fpls.2021.806407>
12. Wäldchen, J., & Mäder, P. (2018). Plant species identification using computer vision techniques: A systematic literature review. *Archives of Computational Methods in Engineering*, 25, 507–543. <https://doi.org/10.1007/s11831-016-9206-z>
13. Weaver, W. N., Ng, J., & Laport, R. G. (2020). LeafMachine: Using machine learning to automate leaf trait extraction from digitized herbarium specimens. *Applications in Plant Sciences*, 8(6), e11367. <https://doi.org/10.1002/aps3.11367>
14. Younis, S., Weiland, C., Hoehndorf, R., Dressler, S., Hickler, T., Seeger, B., & Schmidt, M. (2018). Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks. *Botany Letters*, 165(3–4), 377–383. <https://doi.org/10.1080/23818107.2018.1446357>
15. Younis, S., Schmidt, M., Weiland, C., Dressler, S., Seeger, B., & Hickler, T. (2020). Detection and annotation of plant organs from digitised herbarium scans using deep learning. *Biodiversity Data Journal*, 8, e57090. <https://doi.org/10.3897/BDJ.8.e57090>