

УДК 551.578.48
МРНТИ 37.29.25

В.В. Жданов

АО «Институт географии и водной безопасности» МНВО РК, Алматы, Казахстан

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ В ИЗУЧЕНИИ ПРИРОДНЫХ ОПАСНОСТЕЙ

Жданов Виталий Владимирович, к.т.н., старший научный сотрудник АО «Институт географии и водной безопасности» МНВО РК, Алматы, Казахстан.

E-mail: Zhdanovitaliy@yandex.kz; <https://orcid.org/000-0001-7678-1374>

Аннотация. Статья посвящена применению искусственного интеллекта и машинного обучения при прогнозировании снежных лавин и селевых потоков. В течение многих лет в Институте географии и водной безопасности проводятся работы по оценке и прогнозу природного риска. Для этого был изучен передовой мировой опыт в этой области и принято решение применить технологии интеллектуального анализа данных. В работе использовано программное обеспечение от мировых лидеров – «StatSoft». Симулятор нейронной сети был обучен на архивных данных о погодных условиях за период 1950-2023 гг. Качество нового метода было протестировано на практике и составило 85-95 %, что является хорошим показателем на уровне мировых прогностических центров. Разработанная автоматическая экспертная система упрощает работу инженера прогнозиста и улучшает качество прогнозов стихийных явлений. А это очень важно для обеспечения безопасности населения и объектов в горах. Была разработана концепция региональной системы оценки и прогноза снежных лавин и селевых потоков. Полученный нами опыт может быть полезен в других областях науки, где необходимо обрабатывать большие массивы входящей информации.

Ключевые слова: искусственный интеллект, машинное обучение, оценка, риска прогноз, природные опасности, снежные лавины, селевые потоки

В.В. Жданов

ҚР ҰҚМ «География және су қауіпсіздігі институты» АҚ, Алматы, Қазақстан

ТАБИҒИ ҚАУІПТЕРДІ ЗЕРТТЕУДЕГІ ЖАСАНДЫ ИНТЕЛЛЕКТ ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ

Жданов Виталий Владимирович, т.ғ.к., аға ғылыми қызметкер, ҚР ҰҚМ «География және су қауіпсіздігі институты» АҚ, Алматы, Қазақстан.

E-mail: Zhdanovitaliy@yandex.kz; <https://orcid.org/000-0001-7678-1374>

Аннотация. Мақалада жасанды интеллект пен машиналық оқытуды қар көшкіні мен сел тасқындарын болжауда қолдану мәселесі қарастырылған. Көп жылдар бойы География және су қауіпсіздігі институтында табиғи қатерлерді бағалау және болжау бойынша зерттеулер жүргізіліп келеді. Осы мақсатта аталған саладағы озық әлемдік тәжірибе зерттеліп, деректерді интеллектуалды талдау технологияларын қолдану туралы шешім қабылданды. Жұмыста әлемдік көшбасшылардың бағдарламалық қамтамасыз етуі – «StatSoft» пайдаланылды. Нейрондық желі симуляторы 1950–2023 жж. аралығындағы ауа райы жағдайларының архивтік деректері негізінде оқытылды. Жаңа әдістің сапасы тәжірибеде сыналып, 85–95% аралығында нәтиже көрсетті, бұл әлемдік болжамдық орталықтар деңгейіндегі жақсы көрсеткіш болып табылады. Дамытылған автоматтандырылған эксперттік жүйе инженер-болжамшының жұмысын жеңілдетіп, стихиялық құбылыстарды болжау сапасын арттырады. Бұл өз кезегінде таулы аймақтардағы халық пен нысандардың қауіпсіздігін қамтамасыз ету үшін аса маңызды. Аймақтық қар көшкіні мен сел тасқындарын бағалау және болжаудың тұжырымдамасы әзірленді. Алынған тәжірибеміз ауқымды ақпарат ағындарын өндеуді қажет ететін ғылымның басқа да салаларында пайдалы болуы мүмкін.

Түйін сөздер: жасанды интеллект, машиналық оқыту, бағалау, қатер болжамы, табиғи қатерлер, қар көшкіні, сел тасқындары.

V.Zhdanov

JSC Institute of Geography and Water Security, Almaty, Kazakhstan

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN THE STUDY OF NATURAL HAZARDS

Zhdanov Vitaly Vladimirovich, PhD., Senior Researcher, JSC Institute of Geography and Water Security of the Ministry of National Defense of the Republic of Kazakhstan, Almaty, Kazakhstan.

E-mail: Zhdanovitaliy@yandex.kz; <https://orcid.org/000-0001-7678-1374>

Abstract. The article is devoted to the application of artificial intelligence and machine learning in forecasting snow avalanches and mudflows. For many years, the Institute of Geography and Water Security has been working on assessing and forecasting natural risks. For this purpose, the best world practices in this field were studied and a decision was made to apply data mining technologies. The work used software from world leaders - StatSoft. The neural network simulator was trained on archived weather data for the period 1950-2023. The quality of the new method was tested in practice and amounted to 85-95%, which is a good indicator at the level of world forecasting centers. The developed automatic expert system simplifies the work of the forecast engineer and improves the quality of natural phenomena forecasts. And this is very important for ensuring the safety of the population and objects in the mountains. The concept of a regional system for assessing and forecasting snow avalanches and mudflows was developed. The experience we have gained can be useful in other areas of science where it is necessary to process large amounts of incoming information.

Keywords. Artificial intelligence, machine learning, assessment, risk forecast, natural hazards, snow avalanches, mudflows

Введение. Проблема снежных лавин и селевых потоков актуальна во всём мире, поскольку данные природные явления способны вызывать значительный материальный ущерб и приводить к многочисленным человеческим жертвам. В связи с этим защита от них имеет важное государственное и социальное значение. На основе своевременных предупреждений организуется работа спасательных служб, осуществляется эвакуация населения и принимаются меры по минимизации последствий чрезвычайных ситуаций (Баймолдаев, 2007:284).

В последние годы разработка новейших методов прогнозирования селевых потоков и лавин активно ведётся в ведущих научно-исследовательских центрах мира. Лидирующие позиции занимают государства, где горные районы занимают значительные территории – США, Китай, Индия, Россия, Швейцария, Япония и Южная Корея (Yang, 2024:623; Tang, 2021:51-67; Zhao, 2024:397-412; Ponzani, 2023:143-162). Применяются два подхода к прогнозу – детерминистский и стохастический. В первом используют расчетные формулы, учитывая физические и гидравлические законы возникновения потоков. Во втором применяют статистические зависимости между опасными явлениями погоды и возникновением селевого потока.

В прогнозировании лавинного и селевого риска существуют несколько проблем – редкость мониторинговой сети в горах, влияние малоизученных погодных факторов, короткий период архивных данных. Поэтому некоторые прогнозные модели имеют невысокую точность. В среднем точность предупреждений 80 % для редкого опасного процесса является успешным (Черников, 1990:198). Большинство мировых методов прогноза являются сверхкраткосрочными, относятся к классу «Наукастинг» прогнозов. В них учитывается статистическая зависимость между текущими гидрометеорологическими условиями и прохождением катастрофического потока (Таланов, 2001: 410). Для увеличения заблаговременности предупреждений повсеместно применяются данные численных моделей погоды. Однако с увеличением заблаговременности падает общая точность модели. Часто для моделирования потока применяются формулы гидравлики и GIS-технологии. Подобные методы позволяют оценить территории, подверженные селевому потоку, и оценить время добегания паводочной волны.

В Казахстане снежные лавины и селевые потоки широко распространены в предгорных районах на востоке, юго-востоке и юге республики. Масштабы явления могут меняться от локальных явлений до гигантских с многочисленными жертвами. Их объемы могут достигать миллионов кубометров. Крупнейшие селевые катастрофы в 1921, 1963, 1973 г. приводили к огромным жертвам и ущербу в миллиарды долларов (Виноградов, 1976:60-72). Крупнейшая лавинная катастрофа произошла в марте 1966 г., когда был нанесен огромный ущерб и разрушены старые здания альпинистских и

туристических баз, автодорог и линий электропередач в ущельях рек Улкен и Киши Алматы, Талгар и Турген (Северский, 2006:184). Опасность этих природных явлений вызывает необходимость в разработке разных методов защиты. Долговременная постоянная защита в виде дамб и отводящих каналов стоит очень дорого, но значительно уменьшает риск для населения и объектов. Прогнозирование опасности прохождения селя или лавины стоит значительно меньше и позволяет организовать предупреждение для населения. Поэтому оно очень широко применяется в стране (Степанов, 2023:113-123).

Цель и задачи исследований. Целью выполненной работы было внедрение технологий машинного обучения в методы оценки и прогноза природного риска. Поскольку качественные прогнозы снежных лавин и селевых потоков являются важным элементом государственной системы предупреждения населения, то совершенствование методов оценки очень важны для обеспечения безопасности населения. В ходе работы выполнялись следующие задачи: изучение передового мирового опыта, сбор и анализ архивных данных, поиск подходящего программного обеспечения, оценка качества работы полученной модели.

Искусственный интеллект (ИИ) – это область современной информатики и программирования, в рамках которой разрабатываются алгоритмы, имитирующие свойства человеческого мозга, включая способность к обучению (Бишоп, 2020:960). Часто в этом контексте используется синоним – (машинное обучение). ИИ находит широкое применение в различных областях науки и техники, выполняя задачи, связанные с математическими расчётами, статистической обработкой данных, анализом текстов, изображений и других видов информации. Благодаря своей универсальности и эффективности, технологии искусственного интеллекта активно внедряются в научные исследования по всему миру.

Одним из ключевых направлений машинного обучения является «аналитика данных» (также известная как «наука о данных» или «добыча данных»). Это современный инструмент обработки и анализа больших объёмов информации, особенно актуальный в таких сферах, как климатология, археология и банковское дело, где исследователи работают с масштабными статистическими архивами (Бурова, 2020:60-72; Lutsenko, 2022:96-177).

Применение ИИ особенно важно в ситуациях, когда традиционные методы анализа дают слабые результаты из-за слабовыраженных статистических зависимостей. В таких случаях машинное обучение позволяет выявлять скрытые закономерности и делать более точные прогнозы. Кроме того, технологии искусственного интеллекта применяются для прогнозирования и оценки опасных природных процессов, таких как снежные лавины, селевые потоки, оползни и обвалы.

Существуют два способа внедрения технологий ИИ в работу ученых, анализирующих природные явления. Первый из них – это написание собственной программы нейросимулятора. Самый распространенный язык программирования – PYTHON с открытым кодом и использованием бесплатных библиотек. Но этот способ требует соответствующей квалификации программиста и больших затрат времени на разработку собственной нейронной сети (ИНС). Другой способ создания ИНС – это использование готового программного обеспечения от известных производителей, таких как “StatSoft”, “MathCad”, “SPSS” (Ada, 2018: 237-263; Akgum, 2012: 23-34).

Мы применили современные ИИ-методы обработки данных для изучения природных опасностей. Это позволило выявить новые зависимости и разработать методы оценки лавинной и селевой опасности. Накопленный опыт может использоваться и в других географических исследованиях, требующих анализа больших массивов данных.

Материалы и методы исследований. Аналитика и добыча данных – это современное направление математической статистики, позволяющее решать задачи классификации, кластеризации, регрессии, анализа временных рядов и прогнозирования. Оно упрощает работу с большими массивами информации и автоматизирует вычисления статистических коэффициентов. Машинное обучение снижает необходимость ручных расчетов, однако его точность зависит от качества обучающей выборки. При отсутствии надежных статистических зависимостей или при больших погрешностях модель может выдавать ложные результаты, поэтому необходима калибровка на реальных данных. Поскольку все природные явления подчиняются законам распределения, важным этапом исследования является выбор подходящего закона, от которого зависят вычисления средних характеристик и ошибок (Чернышова, 2009:88; Аргучинцева, 2007:105).

Параметрическая статистика и непараметрическая статистика. В точных науках и фундаментальных исследованиях переменные подчиняются закону нормального распределение (Гаусса, Стьюдента). При этом среднее значение переменной (математическое ожидание) и

стандартное отклонение описывают большинство случаев. При этом отсутствуют редкие крупные события (выбросы), которые искажают величину математического ожидания. «Золотое сечение» или «Правило трех сигм» означает, что в пределах трех стандартных отклонений (σ) расположено 99 % случаев; двух (σ) – 95 % случаев. Это называется доверительным интервалом. Соответственно в точных науках используют стандартные 1 и 5 % уровни значимости – вероятности ошибок.

В природе, экономике и социологии встречаются непараметрические распределения случайных величин (Парето, Экспоненциальное, Вейбулла). Степенное распределение характеризуется резким уменьшением вероятности больших значений переменной (например, количества больших звезд или миллионеров). Величина математического ожидания сильно искажается из-за одного редкого, но крупного события (выброса). Также характерно «Золотое сечение» 80×20. Это означает, что 20 % миллионеров имеют 80 % богатств; 20 % крупных звезд имеют 80 % массы вещества галактики. Для оценки социологических и природных явлений используют непараметрические статистические величины – медиану (центральную квартиль) и межквартильный размах. Параметры, показывающие срединное значение и процентное соотношение – 25, 50, 75 % от всей выборки. Виды распределений показаны на рисунке 1.

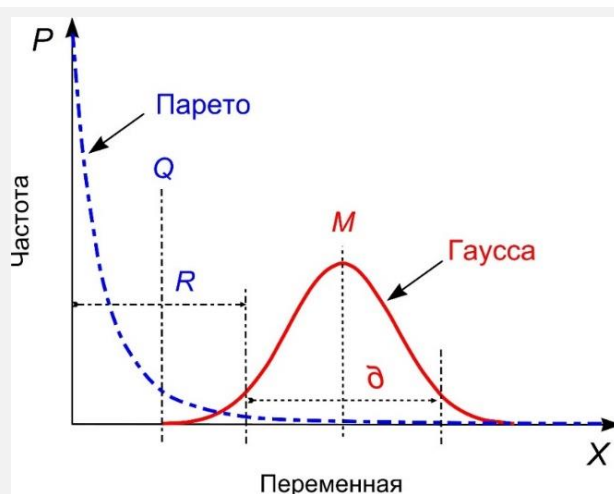


Рисунок 1 – Виды распределений случайной величины и их основные параметры.
X – изучаемая переменная; P – вероятность ее появления; M – математическое ожидание;
 σ – стандартное отклонение; Q – квартиль; R – квартильный размах

Figure 2 – Types of distributions of random variables and their main parameters.

X – variable under study; P – probability of its occurrence; M – mathematical expectation;
□ – standard deviation; Q – quartile; R – quartile range

Нейросетевое программное обеспечение. Для практического применения искусственного интеллекта в математической статистике были разработаны нейросимуляторы от различных производителей программного обеспечения (Черепанов, 2013; Боровиков, 2018:354). Приведем несколько из них:

1. Надстройки для программ прикладных вычислений.

– Matlab_Neural_Network – набор нейросетевых расширений для пакета прикладных вычислений Matlab.

– Statistica_Neural_Networks – набор нейросетевых расширений для пакета прикладной статистики Statistica.

– Excel_Neural_Package – набор библиотек и скриптов для электронных таблиц Excel, реализуют некоторые возможности нейросетевой обработки данных.

2. Универсальные нейросетевые пакеты.

– NeuroSolutions – нейропакет предназначен для моделирования широкого круга искусственных нейронных сетей.

– NeuroPro – менеджер обучаемых искусственных нейронных сетей.

– NeuralWorks – нейропакет, в котором основной упор сделан на применение стандартных нейронных парадигм и алгоритмов обучения.

3. Специализированные.

–Neuroshell Trader – одна из наиболее известных программ создания нейронных сетей для анализа рынков.

–Глаз – используется для обработки аэрокосмической информации.

Нейросимулятор Ясницкого и Черепанова (версия 1,2,3,4). Был разработан авторами Ясницким Л.Н. и Черепановым Ф.М. из Пермской школы искусственного интеллекта при Пермском Государственном Гуманитарно-педагогическом университете в учебных целях и распространяется совершенно бесплатно. Однако он включает в себя современные алгоритмы обучения ИНС. Симулятор способен обучаться на вложенных архивных данных, которые можно загрузить из электронных таблиц или вбивать вручную. Также включена функция проверки полученной сети на тестовой выборке. Это делает программу универсальной и пригодной для научно-исследовательской работы. С применением нейросимулятора (версия 1,2,3,4) проводились научные исследования и опубликованы серии статей, что доказывает его практическую пользу.

Нейросетевое приложение “Statistica neural networks” от компании “StatSoft”. Пакет профессиональной статистики “StatSoft” от компании “Dell”. Русскоязычная версия адаптирована для рынка СНГ компанией “StatSoft Russia” и широко применяется многими научными организациями (Сайт StatSoft, 2022). Пакет представляет электронные таблицы данных и блоки статистического анализа. Программа включает интеллектуальный анализ данных использующих различные алгоритмы контролируемого и неконтролируемого обучения (Боровиков, 2018:354). Присутствуют режимы обучения, тестирования и прогнозирования с помощью полученной ИНС.

В симуляторе нейронной сети “StatSoft” реализованы следующие параметры:

– Тип сети: многослойный персептрон (MPL). Математическая модель восприятия информации, подражающая живому мозгу. Логический элемент электроники – включает серию входных, решающих и выходных элементов. При поступлении сигнала со входного элемента происходит обработка сигнала в решающем элементе. При превышении пороговых значений он активируется и передает сигнал на выходные элементы.

– Алгоритм обучения: итерационный метод численной оптимизации (BFGS), разновидность метода обратного распространения ошибки. Метод обучения персептронов. Многократное повторение для получения правильной реакции персептрона на входные сигналы. Метод BFGS названный в честь авторов в настоящее время считается очень эффективным и используется во многих симуляторах ИНС. Относится к классу методов контролируемого обучения с известным входным сигналом.

– Функция активации: Тожественная, Гиперболическая, Логистическая, Софтмакс. Функция активации или передаточная функция – это способность модели нейрона реагировать на изменение входных сигналов. Функция Софтмакс – частный случай логистической функции, эффективный для анализа многомерных входных данных.

– Эпохи обучения. Полный цикл обработки всего обучающего массива. Реализация от 1 до нескольких тысяч.

– Функция ошибки обучения: сумма квадратов и кросс-энтропия. Математический метод выражения суммарной ошибки расчетов. Сумма квадратов – известный статистический метод учета средней квадратической ошибки. Кросс-энтропия – новый метод идентификации различий для двух распределений вероятностей. Распространен в современном программировании.

– Режим хранения и прогнозирования. Позволяет сохранить готовую нейронную сеть в формате PMML и использовать для прогнозирования. Язык разметки прогнозного моделирования. Формат, используемый в совместимых программах «добычи данных» и «аналитики данных».

Данный интеллектуальный пакет программ позволяет решать большинство задач современной математической статистики и анализа больших массивов данных. Это подходит для создания собственной модели оценки и прогноза опасных природных явлений.

Результаты исследований. В течение многих лет в лаборатории природных опасностей проводились исследования, направленные на разработку и внедрение новых экспериментальных методов оценки и прогнозирования снежных лавин и селевых потоков. В ходе работы были достигнуты следующие результаты:

– Разработаны и протестированы экспериментальные методы оценки лавиноопасности и селеопасности на основе анализа метеорологических и геоморфологических факторов.

– Созданы алгоритмы прогнозирования, использующие технологии искусственного интеллекта и машинного обучения для обработки больших массивов данных.

– Разработана автоматизированная экспертная система, позволяющая классифицировать текущую ситуацию и повышающая точность прогнозов.

– Сформирована концепция региональной системы мониторинга и прогноза, обеспечивающая комплексный подход к снижению природных рисков.

– Полученные результаты рекомендованы для практического применения в горных районах, а также могут быть использованы в смежных областях науки и техники.

Создание обучающей выборки. Для оценки селевой опасности в горах Иле Алатау используются многолетние архивы Казахстанской метеорологической службы «Казгидромет». Данные наблюдений включены в международный фонд Всемирной метеорологической организации (ВМО) и находятся в свободном доступе в стандартном коде КН-01 – SYNOP (сайт погода и климат в мире 2024; архив Казгидромета, 2024). Это старейшие пункты наблюдений, которые являются репрезентативными для горных районов всего Юго-Восточного Казахстана:

– метеостанция «Мынжылки», высота 3017 м н.у.м., индекс 36889 (N43°05'06,9"; E77°04'37,5");

– метеостанция «озеро Улкен Алматы», высота 2502 м н.у.м., индекс 36789 (N43°03'33,3"; E76°58'55,0");

– метеостанция Шымбулак, высота 2200 м н.у.м., индекс станции 36873 (N43°07'57,3"; E77°04'39,51").

На этих станциях проводится полный комплекс метеорологических наблюдений по стандарту ВМО. По данным метеорологических станций была создана база данных со следующими характеристиками:

– период наблюдений: 1950-2024 гг.,

– дискретность наблюдений: суточные данные метеоэлементов за селеопасный период с 1 мая по 30 сентября;

– объем выборки: 16 рядов переменных по 11100 значений в каждом;

Данные о снежных лавинах и селевых потоках собираются в многолетних архивах наблюдений Национальной метеорологической службы, государственного селезащитного управления «Казселезащита» (селевой бюллетень, 2024; сайт МЧС, 2024). Анализировались опасные явления на реках северного склона Иле Алатау: Узынкаргалы, Узынагаш, Шамолган, Каскелен, Каргалы, Аксай, Улкен и Киши Алматы, Талгар, Есик, Турген. Данные хранятся в форматах «Microsoft Excel» и «Statistica Spreadsheet».

Все специалисты по машинному обучению отмечают важный факт – качество работы сети больше всего зависит от полноты и достоверности обучающей выборки. Были рассчитаны статистические характеристики этих рядов наблюдений за период 1950-2023 гг. Они оказались однородными и достоверными. Стандартная статистическая ошибка не превышает 10 % от среднего значения, а критерий Стьюдента является статистически значимым. Данные показаны в таблице 1.

Таблица 1 – Образец электронной базы данных в таблице Excel, для обучения искусственной нейронной сети (фрагмент)

Table 1 – Sample of an electronic database in an Excel table for training an artificial neural network (fragment)

Дата	Степень лавинной опасности	Количество осадков, мм	Высота старого снега, см	Прирост снега за снегопад, см	Максимальная температура воздуха, °C	Коэффициент устойчивости снега
10.04.2017	умеренная	0,0	65	0	10,5	0,47
11.04.2017	умеренная	25,0	64	0	10,1	0,36
12.04.2017	значительная	26,6	64	5	5,6	0,28
13.04.2017	значительная	44,0	68	0	2	0,26
14.04.2017	экстремальная	45,7	65	37	6,2	0,26
15.04.2017	высокая	45,7	75	37	15,5	0,28
16.04.2017	высокая	45,7	63	37	17	0,24
17.04.2017	умеренная	16,0	55	0	3,5	0,83
18.04.2017	значительная	0,0	45	0	9,7	1,13
19.04.2017	умеренная	0,0	43	0	12,1	0,68

Обучение компьютерной программы. Симулятор искусственного интеллекта был обучен распознаванию лавиноопасных и селеопасных ситуаций разного уровня в Иле Алатау. Эксперимент показал, что ИИ способен распознавать текущую лавиноопасную и селеопасную ситуацию с точностью 70-90 %. Наибольшая ошибка распознавания отмечалась при использовании учебного нейросимулятора Ясницкого и Черепанова при короткой обучающей выборке (бассейн Улкен Алматы за 2002-2022 гг.). Лучшие результаты показал нейросимулятор компании StatSoft при использовании длинной обучающей выборки (бассейн Киши Алматы за 1950-2022 гг.). Точность обучения зависит от совершенства алгоритмов. В учебном нейросимуляторе используется более простой и быстрый алгоритм обратного распространения ошибки, а в платной версии более совершенный, но медленный алгоритм численной оптимизации BFGS. Скорость работы сильно зависит от системных требований ЭВМ (Сайт StatSoft, 2022).

В процессе эксперимента проводилось обучение нейронных сетей с различными наборами параметров и алгоритмов обучения. В результате были выбраны варианты с наименьшими ошибками обучения. Они приведены в таблице 2. Отрабатывались два варианта статистических задач – регрессия и классификация. При обучении закладывались механизмы активации нейронов, по умолчанию рекомендованные производителем программ. В режиме классификации при одинаковой обучающей выборке ошибки оказались ниже. Для дальнейшей работы будет использоваться нейросеть, которая дает наименьшие ошибки при обучении. На ее основе создается экспертная система оценки уровня селевой опасности.

Таблица 2 – Параметры обучения нейронной сети на длинной обучающей выборке (1950-2023 гг.) для классификации потенциально селеопасных условий

Table 2 – Parameters for training a neural network on a long training sample (1950-2023) for classifying potentially mudflow-hazardous conditions

Режим классификации с применением ИНС	Архитектура ИНС	Производительность обучения	Контрольная производительность	Тестовая производительность	Количество эпох обучения	Функция ошибки	Функция активации скрытых нейронов	Функция активации выходных нейронов
Уровень селевой опасности	MLP 12-600-4	92,3	91,9	92,0	1000	Энтропия	Гиперболическая	Софтмакс
Факт прохождения потока	MLP 12-500-2	97,2	97,3	97,4	1000	Энтропия	Гиперболическая	Софтмакс
Высотная зона формирования селя	MLP 7-500-4	99,4	99,7	99,3	500	Энтропия	Гиперболическая	Софтмакс

Качество работы модели. Были проведены эксплуатационные испытания нового метода оценки и прогноза уровня лавинной и селевой опасности. Результаты приведены в таблице 3. Критерии для сравнения данных моделирования с реальными данными оказались статистически значимыми при стандартном уровне значимости 5 % и вероятности ошибки p-level менее 0,01. Для подтверждения точности модели мы использовали несколько оценок. Оправдываемость (эффективность) сверхкраткосрочного прогноза 91-94 % является хорошим показателем. Критерий Брайера для вероятностных прогнозов, также близок к 0, что является хорошим показателем (Cagnati, 1998:130-134; Schweizer, 2020:737-750).

Таблица 3 – Статистические критерии для оценки достоверности полученных моделей
Table 3 – Statistical criteria for assessing the reliability of the obtained models

	Процент точных оценок	Критерий Вилкоксона	Критерий Брайера	Корреляция Спирмена
Селевые потоки	93,6	2,2	0,026	0,74
Снежные лавины	91,2	3,65	0,09	0,70

Региональная концепция модернизации системы оценки и прогноза снежных лавин и селевых потоков должна включать следующие направления развития:

- Переход на вероятностные формы прогноза и применение 5 балльных шкал опасности с цветовым обозначением каждой градации. Это позволит оценить риск и принять соответствующие меры, адекватные прогнозируемой опасности.
- Внедрение программ интеллектуального анализа данных для оценки текущего и прогнозного уровня опасности стихийных природных явлений.
- Для получения более подробной информации применение данных автоматических метеостанций, дистанционного зондирования земли и международных погодных баз данных реанализов.
- Применение современных моделей численного прогноза погоды для увеличения заблаговременности экстренных предупреждений, что позволит дать время на принятие защитных мер.

Выводы. Методы искусственного интеллекта и машинного обучения являются хорошим инструментом при изучении опасных природных процессов. Эти технологии позволяют быстро и качественно обработать большие архивы и дают статистически значимые результаты. Современное направление математической статистики – интеллектуальный анализ данных начинает широко применяться в метеорологии, гидрологии и климатологии. Эти модели позволяют с высокой точностью классифицировать лавиноопасную и селеопасную обстановку.

Однако их эффективность зависит от алгоритмов обучения и качества выборки, поэтому очень важно оценивать качество входных архивных данных. Если многолетние архивы нерепрезентативны и недостоверны, то невозможно получить качественную модель. Поэтому достоверная обучающая выборка является важнейшим этапом при работе с симуляторами нейронной сети. Математические алгоритмы обработки данных без контрольной проверки выдают только преобразованный результат, не всегда достоверный.

Точность прогнозов опасных явлений напрямую связана с развитием численных моделей погоды, поэтому сочетание методов прогнозирования атмосферных процессов и интеллектуальной оценки природных опасностей является перспективным направлением гидрометеорологии и повышает уровень безопасности населения.

Литература

- 1 Ada M, San BT. (2018). Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey. *Nat Hazards*, 90, 237–263.
- 2 Akgun A, и др. (2012). An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Comput Geosci*, 38(1), 23–34.
- 3 Баймолдаев Т., Виноходов В.Н. (2007). *Казселезащита – оперативные меры до и после стихии*, Алматы: изд-во «Бастау», 284.
- 4 Бишоп Кристофер М. (2020). *Распознавание образов и машинное обучение*. Пер. с англ. Д. А. Ключина. Москва; Санкт-Петербург: Диалектика, 960.
- 5 Боровиков В.П. (2018). *Популярное введение в современный анализ данных и машинное обучение на Statistica*. Москва, Изд-во «Горячая линия – Телеком», 354.
- 6 Бурова О. А., Судакова А. А. 2020. Искусственный интеллект – помощник статистики будущего. *Наука о данных: Материалы международной научно-практической конференции*, Санкт-Петербург, 5-7 февраля 2020 года. Санкт-Петербург: Санкт-Петербургский гос. эконом. университет, 74-76.
- 7 Виноградов Ю.Б. и др. (1976). Селевой поток 15 июля 1973 г. на Малой Алматинке. *КазНИГМИ. Селевые потоки*, 1, 60–72.
- 8 Информация о погоде и климате по данным метеостанций. URL:<http://www.rp5.ru> [Электрон. ресурс]. (дата обращения 10.10.2024)
- 9 Ежедневный селевой бюллетень. URL:<https://www.kazhydromet.kz/ru/selevedenie-str/ezhednevnyy-byulleten-selevo-opasnosti-dozhdevogo-genezisa>. [Электронный ресурс]. (Дата обращения 30.09.2024).

- 10 Cagnati A., Valt M., Soratroi G., Gavalda J. and Sellés C. G. (1998). A field method for avalanche danger-level verification. *Ann. Glaciology*, 26, 130-134. doi: 10.3189/1998AoG26-1-343-346.
- 11 Lutsenko E. V. (2022). Automated system-cognitive analysis of the level of consistency of natural numbers as systems of prime factors. E. V. Lutsenko. Text: electronic. *Polythematic Online Scientific Journal of Kuban State Agrarian University*, 179, 96-177.
- 12 А. В. Аргучинцева. (2007). *Методы статистической обработки и анализа гидрометеорологических наблюдений: учеб. пособие*. Иркутск: Иркут. гос. ун-т, 105.
- 13 Метеорологическая база данных РГП «Казгидромет». [Электрон. ресурс]. https://meteo.kazhydromet.kz/database_meteo/ (дата обращения 10.10.2024)
- 14 Ponziani M. и др. (2023). The use of machine learning techniques for a predictive model of debris flows triggered by short intense rainfall. *Nat. Hazards*, 117, 143–162.
- 15 Черников Н.Г. (1990). *Руководство селестоковым станциям и гидрографическим партиям. РД 52.30.238–89. Вып.1. Организация и проведение работ по изучению селей*. Москва: Гидрометеоиздат., 198.
- 16 Schweizer J., Mitterer C., Techell F., Stoffell A., and Reuter B. (2020). On the relation between avalanche occurrence and avalanche danger level. *The Cryosphere*, 14, 737–750. <https://doi.org/10.5194/tc-14-737-2020>.
- 17 Сайт российского представительства фирмы StatSoft. URL: <http://www.StatSoft.ru> [Электрон. ресурс]. (дата обращения 10.10.2022 г.).
- 18 Сайт Министерства по чрезвычайным ситуациям. URL: <https://www.gov.kz/memleket/entities/emer/press/article/1?lang=en> [Электрон. ресурс] (дата обращения 10.10.2024 г.).
- 19 Степанов Б.С. (2023). Вклад национальной гидрометеорологической службы республики Казахстан в селевую безопасность страны. *Гидрометеорология и экология*, 4, 113-123. DOI: 10.54668/2789-6323-2023-111-4-113-123.
- 20 Северский И., Благовещенский В. и др. (2006). *Снежный покров и лавины Тянь-Таня*. Алматы, 184.
- 21 Таланов Е.А. (2001). Селевые явления и зоны, подверженные их воздействию в бассейне реки Б. Алматинка. *Современные проблемы геоэкологии и созоологии: Доклады Международной научно-практической конференции 22-23 января 2001*, Алматы. Под редакцией М.М. Буракова, Л.М. Павличенко, Е.А. Таланова, Алматы: изд-во Шартапа, 410.
- 22 Tang W. (2021). Artificial Neural Network-based prediction of glacial debris flows in the ParlungZangbo Basin, southeastern Tibetan Plateau, China. *J. Mt. Sci.*, 18, 51–67.
- 23 Yang L. и др. (2024). Machine-Learning-Based Prediction Modeling for Debris Flow Occurrence: A Meta-Analysis. *Water*, 16, 923-943. <https://doi.org/10.3390/w16070923>
- 24 Zhao H., Wei A., Ma F. et al. (2024). Comparison of debris flow susceptibility assessment methods: support vector machine, particle swarm optimization, and feature selection techniques. *J.Mt.Sci.*, 21, 397–412. <https://doi.org/10.1007/s11629-023-8395-9>
- 25 Черепанов Ф.М. (2013). Исследовательский симулятор нейронных сетей, обзор его приложений и возможности применения для создания системы диагностики заболеваний сердечно-сосудистой системы. *Современные проблемы науки и образования*, 1; URL: www.science-education.ru/107-8392 (in Russian)
- 26 Чернышева Л.С., Платонова В.А. (2009). *Расчёт и интерпретация основных климатических показателей отдельных метеорологических величин*. Уч.-метод. пособие. Владивосток: Изд-во Дальневост. ун-та, 88.

References

- 1 Ada M, San BT. (2018). Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey. *Nat Hazards*, 90, 237–263.
- 2 Akgun A, et al. (2012). An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Comput Geosci*, 38(1), 23–34.
- 3 Baymoldayev T., Vinokhodov V.N. (2007). [*Kazselezashchita – operational measures before and after the disaster*], Almaty: Bastau Publishing House, 284.
- 4 Bishop Christopher M. (2020). *Pattern recognition and machine learning*. Transl. from English. D. A. Klyushina. Moscow; Saint Petersburg: Dialectika, 960.
- 5 Borovikov V. P. (2018). [*Popular Introduction to Modern Data Analysis and Machine Learning on Statistica*]. Moscow, Goryachaya Liniya-Telecom Publishing House, 354.
- 6 Burova O. A., Sudakova A. A. 2020. Artificial Intelligence – Assistant to Statistics of the Future. *Data Science: Proceedings of the International Scientific and Practical Conference*, Saint Petersburg, February 5-7, 2020. Saint Petersburg: Saint Petersburg State University of Economics, 74-76.
- 7 Vinogradov Yu. B. et al. (1976). Mudflow on July 15, 1973 on Malaya Almatinka. [*KazNIGMI. Mudflows*], 1, 60–72.
- 8 Weather and Climate Information Based on Weather Station Data. URL: <http://www.rp5.ru> [Electronic resource]. (Accessed 10.10.2024)
- 9 Daily mudflow bulletin. URL: <https://www.kazhydromet.kz/ru/selevedenie-str/ezhednevnyy-byulleten-selevoy-opasnostidozhdevogo-genezisa>. [Electronic resource]. (Accessed 30.09.2024).
- 10 Cagnati A., Valt M., Soratroi G., Gavalda J. and Sellés C. G. (1998). A field method for avalanche danger-level verification. *Ann. Glaciology*, 26, 130-134. doi: 10.3189/1998AoG26-1-343-346.
- 11 Lutsenko E. V. (2022). Automated system-cognitive analysis of the level of consistency of natural numbers as systems of prime factors. E. V. Lutsenko. Text: electronic. [*Polythematic Online Scientific Journal of Kuban State Agrarian University*, 179, 96-177.
- 12 А. В. Аргучинцева. (2007). [*Methods of statistical processing and analysis of hydrometeorological observations: tutorial. manual*]. Irkutsk: Irkutsk state university, 105.
- 13 Meteorological database of RSE "Kazhydromet". [Electronic resource]. https://meteo.kazhydromet.kz/database_meteo/ (date of access 10.10.2024)
- 14 Ponziani M. et al. (2023). The use of machine learning techniques for a predictive model of debris flows triggered by short intense rainfall. *Nat. Hazards*, 117, 143–162.
- 15 Chernikov N. G. (1990). [*Mudflow Station and Hydrographic Party Guide. RD 52.30.238–89. Issue 1. Organization and Implementation of Mudflow Study Work*]. Moscow: Hydrometeoizdat, 198.

- 16 Schweizer J., Mitterer C., Techell F., Stoffel A., and Reuter B. (2020). On the Relation between Avalanche Occurrence and Avalanche Danger Level. *The Cryosphere*, 14, 737–750. <https://doi.org/10.5194/tc-14-737-2020>.
- 17 Website of the Russian Representative Office of StatSoft. URL: <http://www.StatSoft.ru> [Electronic resource]. (date of access 10.10.2022).
- 18 Website of the Ministry of Emergency Situations. URL: <https://www.gov.kz/memleket/entities/emer/press/article/1?lang=en> [Electronic resource] (date of access 10.10.2024).
- 19 Stepanov B.S. (2023). Contribution of the National Hydrometeorological Service of the Republic of Kazakhstan to the mudflow safety of the country. [*Hydrometeorology and Ecology*], 4, 113-123. DOI: 10.54668/2789-6323-2023-111-4-113-123.
- 20 Seversky I., Blagoveshchensky V. et al. (2006). [*Snow cover and avalanches of the Tien Tan*]. Almaty, 184.
- 21 Talanov E.A. (2001). Mudflow phenomena and zones affected by them in the B. Almatinka River basin. *Modern problems of geoecology and zoology: Reports of the International scientific-practical conference on January 22-23, 2001*, Almaty, Edited by M.M. Burakov, L.M. Pavlichenko, E.A. Talanov, Almaty: Shartarap Publishing House, 410.
- 22 Tang W. (2021). Artificial Neural Network-based prediction of glacial debris flows in the Parlung Zangbo Basin, southeastern Tibetan Plateau, China. *J. Mt. Sci.*, 18, 51–67.
- 23 Yang L. et al. (2024). Machine-Learning-Based Prediction Modeling for Debris Flow Occurrence: A Meta-Analysis. *Water*, 16, 923-943. <https://doi.org/10.3390/w16070923>
- 24 Zhao H., Wei A., Ma F. et al. (2024). Comparison of debris flow susceptibility assessment methods: support vector machine, particle swarm optimization, and feature selection techniques. *J. Mt. Sci.*, 21, 397–412. <https://doi.org/10.1007/s11629-023-8395-9>
- 25 Cherepanov F. M. (2013). Research simulator of neural networks, a review of its applications and the possibility of using it to create a diagnostic system for cardiovascular diseases. [*Modern problems of science and education*], 1; URL: www.science-education.ru/107-8392 (in Russian)
- 26 Chernysheva L. S., Platonova V. A. (2009). Calculation and interpretation of the main climatic indicators of individual meteorological quantities. [*Educational and methodological manual*]. Vladivostok: Publishing house of the Far Eastern University, 88.