

УДК 004.8:630

## Нейросетевая модель оценки и прогнозирования экологического риска изменения состояния лесного фонда\*

*В.В. Бурлуцкий, Г.А. Кочергин, А.Л. Царегородцев, А.В. Якимчук*

Югорский НИИ информационных технологий (Ханты-Мансийск, Россия)

## Neural Network Model of Assessment and Forecasting of Environmental Risks of Forest Fund Changes

*V.V. Burlutsky, G.A. Kochergin, A.L. Tsaregorodtsev, A.V. Yakimchuk*

Ugra Research Institute of Information Technologies (Khanty-Mansiysk, Russia)

Представлены результаты разработки и вычислительные аспекты нейросетевой модели оценки и прогнозирования экологических рисков изменения состояния лесных массивов на территории Ханты-Мансийского автономного округа — Югры с использованием разнородных многомерных данных о произошедших в 2011–2017 гг. авариях на трубопроводах. В основе разработанной модели лежат методы машинного обучения, позволяющие определить степень риска возникновения повторных аварийных ситуаций на территории лицензионных участков, и методы геоинформационного анализа, которые использовались для построения прогнозной карты рисков негативного воздействия на земли лесного фонда в 2018 г. Достоверность обсуждаемых результатов подтверждена проведенными численными экспериментами. Показано, что результаты моделирования удовлетворительно согласуются с данными дистанционного космического и наземного мониторинга, а также данными аэрофотосъемки. Главным результатом работы является вывод о возможности использования разработанной модели при осуществлении контрольно-надзорной деятельности с целью выявления территорий, наиболее подверженных рискам загрязнения нефтью и нефтепродуктами.

**Ключевые слова:** нейронные сети, машинное обучение, градиентный спуск, искусственный интеллект, геоинформационный анализ, экологические риски.

DOI 10.14258/izvasu(2019)1-10

### Введение

В последние годы Правительством Российской Федерации ведется активная работа над внедрением дифференцированного, или риск-ориентированного подхода [1] при организации и осуществлении контрольно-надзорных мероприятий, что позволит су-

The paper considers the development and computational aspects of the neural network model for assessing and forecasting of environmental risks of changes in Khanty-Mansi Autonomous Okrug based on the data on pipeline accidents in 2011–2017. The model is based on the machine learning methods that allow defining the degree of risk of repeated accidents on the territory of licensed blocks, and methods of geoinformation analysis are used to create a forecast risk map of the negative impact on the forest fund land in 2018. The fidelity of the results is confirmed by numerical experiments. It is shown that modeling results are accorded with the data of remote space and ground monitoring, as well as with the data of aerial photography. The main result of the work is the conclusion about the possibility of using the developed model to identify areas most exposed to the risks of oil pollution.

**Key words:** neural networks, machine learning, gradient descent, artificial intelligence, geoinformation analysis, environmental risks.

щественно повысить эффективность расходования ресурсов на функционирование контрольно-надзорных органов путем сосредоточения усилий инспекторского состава государственного надзора на наиболее значимых направлениях. Так, в Постановлении Правительства РФ №1029 от 28.09.2015 [2] объекты

\*Работа выполнена при финансовой поддержке гранта РФФИ № 18-45-860003.

хозяйственной деятельности, осуществляющие добычу сырой нефти и природного газа, включая переработку природного газа, отнесены к первой категории объектов, оказывающих значительное негативное воздействие на окружающую среду. В Постановлении Правительства РФ №1410 от 22.11.2017 [3] приведены критерии отнесения производственных объектов, используемых юридическими лицами и индивидуальными предпринимателями, оказывающих негативное воздействие на окружающую среду, к определенной категории риска для регионального государственного экологического надзора.

Параллельно с законотворческими процессами велась активная работа научного сообщества по обобщению накопленного опыта управления рисками и разработке различных подходов к решению задач мониторинга [4, 5], оценки и прогнозирования экологических рисков. Здесь можно выделить работы [6, 7], в которых систематизированы основные понятия риск-ориентированного подхода к контрольно-надзорной деятельности, представлены различные методы оценки рисков и предложена целевая модель данного подхода при осуществлении государственного контроля. Н.В. Акининой и соавторами [8] предлагается новый подход к организации процесса анализа и мониторинга экологических рисков в условиях нечеткости, в основе которого лежит качественный анализ экологических рисков, выполняемый экспертными методами с использованием ГИС-технологий.

Указанные выше нормативно-правовые акты не содержат рекомендаций по оценке рисков несоблюдения требований лесного законодательства, а предлагаемые подходы оценки рисков не учитывают особенностей пространственного расположения объектов контроля и надзора, что в условиях Ханты-Мансийского автономного округа является одним из важнейших факторов для принятия решения. В связи с этим можно сделать вывод о том, что задача разработки модели оценки рисков изменения состояния лесного фонда является востребованной для региона.

#### **Данные и методы исследования**

В качестве исходной информации об экологических нарушениях использовался накопленный за 2011–2017 гг. массив данных об авариях, выявленных на внутрипромысловых трубопроводах, предоставленный службой по контролю и надзору в сфере охраны окружающей среды, объектов животного мира и лесных отношений Ханты-Мансийского автономного округа — Югры (Природнадзор Югры). Массив данных представляет собой сложно структурированный набор пространственной и атрибутивной информации с мест аварий. Накопление информации осуществляется в государственной информационной системе «Эконадзор», используемой инспекторским составом Природнадзора Югры при осуществлении

контрольно-надзорных функций. Общее количество инцидентов, хранящихся в информационной системе, составляет более 25 тысяч записей. В качестве дополнительной информации использовались векторные пространственные слои о лесоустройстве автономного округа, инфраструктуре топливно-энергетического комплекса, расположении лицензионных участков, населенных пунктов и дорожной сети.

Анализ исходной информации позволил выделить основные критерии, оказывающие наибольшее влияние на риски возникновения экологических нарушений в границах лицензионных участков, расположенных на территории лесного фонда. Это наличие следующего:

- нефтезагрязненных участков;
- ранее произошедших инцидентов;
- техногенных объектов;
- нарушений правил санитарной безопасности;
- объектов размещения отходов;
- лесопользователя.

А также приближенность к транспортной инфраструктуре и населенным пунктам.

Для комплексного анализа имеющейся разнородной информации о территории исследования и выявленных на ней фактах экологических нарушений нами предлагается использование методов машинного обучения в совокупности с методами пространственного анализа. Имеющийся у коллектива опыт разработки сложных систем подобного класса [9–12] позволяет сделать вывод о целесообразности использования методов машинного обучения для оценки степени негативного воздействия на земли лесного фонда, методов пространственного анализа для учета информации о фактическом расположении мест аварий и объектов инфраструктуры, а также представления результатов анализа и построения прогнозной тепловой карты рисков. Комплексное использование упомянутых технологий позволит создать модель оценки и прогноза экологических рисков изменения состояния лесного фонда на территории автономного округа в границах лицензионных участков.

#### **Нейросетевая модель прогнозирования экологических рисков**

В соответствии с Постановлением Правительства РФ №1410 от 22.11.2017 [3] принято выделять следующие пять категорий риска для регионального государственного экологического надзора: высокий, значительный, средний, умеренный и низкий, для каждого из которых устанавливается своя очередность плановых проверок. С учетом данной категоризации рисков создаваемая нами модель будет относить все лицензионные участки, расположенные на территории автономного округа, к пяти категориям.

Для решения поставленной задачи нами предлагается следующий алгоритм обработки и анализа разнородной информации:

- 1) предварительная обработка данных;
- 2) пространственный анализ данных;
- 3) анализ данных с использованием методов машинного обучения;
- 4) представление результатов анализа средствами ГИС.

Рассмотрим каждый из этапов более подробно. На этапе предварительной обработки были выполнены операции по преобразованию некорректных координат и восстановлению пространственных характеристик исследуемых объектов, фильтрация и очистка от ошибочных данных для обеспечения полноты набора данных. Далее числовые признаки из массива унифицированных записей были нормализованы по максимальному значению, а текстовые признаки переведены в числовые, что позволило привести все признаки набора данных к числовому виду. Объем полученного в результате предварительной обработки массива унифицированных данных составил более 20 тысяч записей, что позволяет говорить о корректности применения методов машинного обучения для решения поставленной задачи.

На этапе пространственного анализа данных средствами геоинформационной системы QGIS каждому лицензионному участку были поставлены в соответствие записи из массива данных, полученных на предыдущем этапе, а также дополнительная информация о годовой аварийности, расстояниях до населенных пунктов, транспортной инфраструктуры и других техногенных объектов. Таким образом, общее количество признаков, оказывающих влияние на риски изменения состояния лесного фонда для каждого лицензионного участка, составило 83 единицы.

Для анализа данных методами машинного обучения на следующем этапе была реализована нейронная сеть [13, 14], в которой в качестве метки для обучения используется признак произошедших в 2017 г. аварий. Всего было выявлено 180 лицензионных участков, на территории которых значение метки для обучения отлично от нуля, что говорит о наличии на данном участке аварий в указанном году. Для более точного обучения нейросетевой модели было отобрано такое же количество лицензионных участков, где аварий в 2017 г. не было зафиксировано. Таким образом, набор данных для обучения модели составил 360 записей, содержащих значения признаков за 2014–2016 гг.

Разработанная нейросетевая модель имеет 83 нейрона на входном слое (по количеству признаков для каждого лицензионного участка), три скрытых слоя, в каждом из которых в качестве функции активации используется сигмоид, и выходной слой, возвращающий вещественное число в интервале от 0 до 1. Это число, характеризующее степень негативного воз-

действия на лесной фонд в границах лицензионного участка, будем трактовать как оценку экологического риска изменения состояния лесного фонда по аналогии с отнесением производственных объектов, оказывающих негативное воздействие на окружающую среду, к определенной категории риска [3]. Зависимость между входными и выходными данными описывается следующим отношением:

$$Y = F\left(\sum_{i=1}^m F\left(\sum_{i=1}^n X_i * W_{ij}\right) * V_j\right), \quad (1)$$

где  $X$  является вектором входных данных;  $w$  — весовые коэффициенты между входным и обрабатываемым слоем;  $v$  — весовые коэффициенты между обрабатываемым и выходным слоем;  $F$  — функция активации;  $m$  — количество нейронов во входном слое;  $n$  — количество нейронов в обрабатываемом слое;  $y$  — выходное значение нейрона.

Для оптимизации показателей нейронной сети использовался метод градиентного спуска, который обеспечивает подстройку весовых коэффициентов по следующему правилу:

$$\Delta w_{ij}^{(q)} = \eta \frac{\partial E}{\partial w_{ij}}, \quad (2)$$

где  $\Delta w_{ij}^{(q)}$  — величина изменения веса связи, соединяющей  $i$ -й нейрон  $(q-1)$  слоя с  $j$ -м нейроном слоя  $q$ ;  $\eta$  — коэффициент скорости обучения.

На этапе представления результатов анализа средствами ГИС проводится категорирование степени негативного воздействия на лесной фонд на пять категорий риска в зависимости от полученного на предыдущем этапе значения оценки экологического риска по следующему правилу:

$$\left\{ \begin{array}{l} y = 0 - \text{низкий риск,} \\ 0 < y \leq 0,19 - \text{умеренный риск,} \\ 0,19 < y \leq 0,6 - \text{средний риск,} \\ 0,6 < y \leq 0,85 - \text{значительный риск,} \\ 0,85 < y \leq 1 - \text{высокий риск.} \end{array} \right. \quad (3)$$

Все лицензионные участки в зависимости от рассчитанной категории риска отображаются на цифровой карте с использованием уникального условного знака, что позволяет построить тепловую карту рисков.

#### Проверка качества модели

Для оценки качества реализованной нейросетевой модели были проведены численные эксперименты на основе метода кросс-валидации по  $k$ -блокам ( $k$ -fold cross-validation) [15, 16]:

$$CV(\mu, X^L) = \frac{1}{N}. \quad (4)$$

Выборка  $X^L$  разбивается  $N$  различными способами на две непересекающиеся подвыборки:

$$X^L = X_n^m \cap X_n^k, \quad (5)$$

где  $X_n^m$  — обучающая подвыборка длины  $m$ ;  $X_n^k$  — контрольная подвыборка длины  $k = L - m$ ,  $n = 1, \dots, N$  — номер разбиения.

Набор данных для обучения был разбит на десять одинаковых по размеру блоков. Всего было проведено

десять численных экспериментов, в каждом из которых один блок из десяти использовался для тестирования модели, а оставшиеся девять блоков применялись в качестве тренировочного набора. По результатам численных экспериментов с различными комбинациями блоков набора данных были рассчитаны промежуточные оценки качества модели (табл.).

Результаты оценки качества нейросетевой модели

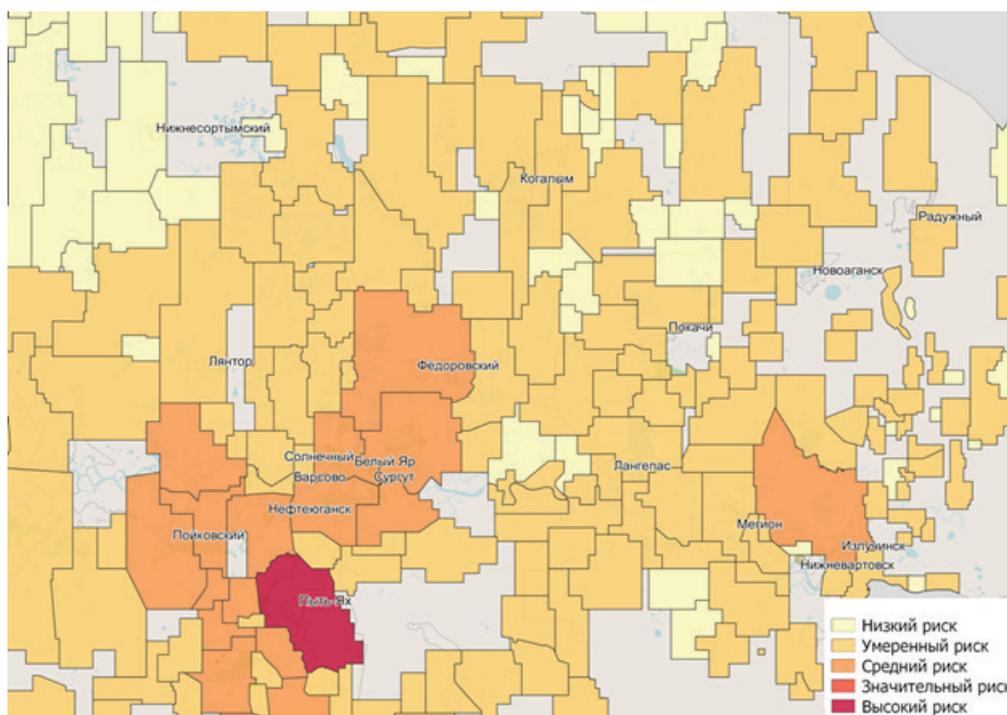
Номер блока данных	1	2	3	4	5	6	7	8	9	10
Рассчитанная оценка, %	93,7	94,1	94,3	93,5	95,6	94,6	94,8	94,7	93,9	94,2

Итоговая оценка качества реализованной нейросетевой модели, полученная как среднее от всех промежуточных оценок, составила 94,3%, что говорит о ее достаточно высоком качестве и практической применимости для решения данного класса задач.

#### Использование нейросетевой модели

С использованием разработанной нейросетевой модели был проведен анализ имеющейся информации с целью построения прогноза рисков негативного

воздействия на лесной фонд в 2018 г. По результатам анализа была построена тепловая карта лицензионных участков Ханты-Мансийского автономного округа — Югры (рис.), отражающая риски изменения состояния лесного фонда в соответствии с правилом (3). Анализ полученного результата показал, что Мамонтовский лицензионный участок, расположенный в Нефтеюганском районе, единственный в автономном округе относится к категории высокого риска.



Тепловая карта рисков негативного воздействия

По результатам анализа, проведенного с использованием нейросетевой модели, был осуществлен комплекс заверочных работ. На первом этапе было проведено дешифрирование космического снимка на территорию Мамонтовского лицензионного участка, площадь которого составляет почти 1300 кв. км, с целью выявления вероятных мест возникновения

аварийных разливов нефти в 2018 г. По результатам дешифрирования космического снимка был выделен участок площадью 25 кв. км с наибольшей концентрацией вероятных мест нефтезагрязнений. Далее, сотрудниками Природнадзора Югры была осуществлена съемка этого участка с использованием беспилотного летательного аппарата. Дешифрирование

данных аэрофотосъемки подтвердило наличие 11 новых разливов нефти общей площадью 22 га.

Положительные результаты использования нейросетевой модели позволяют говорить о новом подходе к проведению контрольно-надзорных мероприятий по выявлению экологических нарушений на территории региона интенсивной нефтедобычи. Этот подход включает комплекс мероприятий по прогнозированию рисков с использованием разработанной модели для выявления участков со значительным и высоким риском, дешифрированию космических снимков и других пространственных данных для сужения области поиска новых разливов нефти и нефтепродуктов, применению беспилотных летательных аппаратов для заверки результатов и расчета площадей загрязненных земель.

### Заключение

В статье предложен новый подход к осуществлению контрольно-надзорной деятельности при осуществлении регионального экологического надзора, основанный на разработанной нейросетевой модели, позволяющей прогнозировать возникновение экологических рисков загрязнения земель лесного фонда нефтью и нефтепродуктами. Путем анализа разнородных данных была построена тепловая карта рисков негативного воздействия на территории Ханты-Мансийского автономного округа, выявлен лицензионный участок, относящийся к наивысшей категории риска, проведен комплекс заверочных работ, который подтвердил высокое качество разработанной модели и позволил выявить 11 новых аварийных разливов нефти.

## Библиографический список

1. План мероприятий по совершенствованию контрольно-надзорной деятельности в Российской Федерации на 2016–2017 годы : распоряжение Правительства Российской Федерации от 1 апреля 2016 г. № 559-р // Электронный фонд правовой и нормативно-технической документации.
2. Об утверждении критериев отнесения объектов, оказывающих негативное воздействие на окружающую среду, к объектам I, II, III и IV категорий : постановление Правительства Российской Федерации от 28 сентября 2015 г. № 1029 // Электронный фонд правовой и нормативно-технической документации.
3. О критериях отнесения производственных объектов, используемых юридическими лицами и индивидуальными предпринимателями, оказывающих негативное воздействие на окружающую среду, к определенной категории риска для регионального государственного экологического надзора и об особенностях осуществления указанного надзора : постановление Правительства Российской Федерации от 22 ноября 2017 г. № 1410 // Электронный фонд правовой и нормативно-технической документации.
4. Волков Н.В., Донцов А.А., Лагутин А.А. Разработка геопортальной системы для решения задач регионального космического мониторинга // Известия Алтайского гос. ун-та. 2013. № 1/2 (77).
5. Донцов А.А., Волков Н.В., Лагутин А.А. Разработка технологии организации каталогов спутниковых данных // Известия Алтайского гос. ун-та. 2014. № 1/2 (81).
6. Чаплинский А.В., Плаксин С.М. Управление рисками при осуществлении государственного контроля в России // Вопросы государственного и муниципального управления. 2016. № 2.
7. Авдийский В.И., Безденежных В.М., Катаева Е.Г. Управление рисками как ключевой элемент обеспечения реализации риск-ориентированного подхода в деятельности хозяйствующих субъектов // Экономика. Налоги. Право. 2017. №6.
8. Акинина Н.В., Псоянц В.Г., Колесенков А.Н., Таганов А.И. Теория и практика применения нечетких сетей Петри для мониторинга экологических рисков // Вестник Томского гос. ун-та. 2017. №41.
9. Полищук Ю.М., Кокорина Н.В., Кочергин Г.А., Перемитина Т.О., Токарева О.С. Методология оценки экологического риска воздействия точечных источников атмосферного загрязнения на основе данных биоиндикации // Проблемы анализа риска. 2011. Т. 8, №4.
10. Карташев Е.А., Царегородцев А.Л. Автоматизированная информационная система поиска и анализа информации в сети Интернет // Фундаментальные исследования. 2016. № 10, ч. 2.
11. Волошин С.В., Царегородцев А.Л., Карташев Е.А., Славский В.В. Анализ качества бинарной классификации веб-страниц методом опорных векторов // Известия Алтайского гос. ун-та. 2017. № 4 (96).
12. Кочергин Г.А., Зыков А.С., Царегородцев А.Л., Горлов Н.В. Автоматизированная информационная система мониторинга изменений на основе космических снимков // Известия Алтайского гос. ун-та. 2017. № 1 (93).
13. Галушкин А. И. Нейронные сети: основы теории. М., 2010.
14. Riedmiller M., Braun H. A Direct Adaptive Method for Faster Backpropagation Learning: the RPROP Algorithm // International Conference of Neural Networks. San Francisco, 1993. Vol. 1.
15. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning // Springer, 2014.
16. Refaeilzadeh P., Tang L., Liu H. "Cross Validation." Encyclopedia of Database Systems. Eds M. Tamer A-Zsu, and Ling Liu. New York, 2009.