

# Оперативное прогнозирование рисков возникновения дорожно-транспортных происшествий на основе нейросетевого анализа больших данных

О.К. Головнин<sup>1</sup>, Е.В. Сидорова<sup>1</sup>

<sup>1</sup>Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

**Аннотация.** В статье предлагается подход к оперативному прогнозированию дорожно-транспортных происшествий с разделением по видам на основе многослойного перцептрона Румельхарта. Используются большие данные, поступающие в онлайн режиме из внешних разнородных источников данных. Учитываются погодные, дорожные, организационные факторы и параметры транспортного потока. Программная реализация подхода использует фреймворк TensorFlow и библиотеку Keras. Проведенные эксперименты показали, что подход обеспечивает точность распознавания ситуаций в 90%. Программная реализация предназначена для функционирования в составе систем предотвращения аварий и инцидентов.

## 1. Введение

Ощутимой проблемой для транспортного комплекса современных городских агломераций остаются дорожно-транспортные происшествия (ДТП), которые наносят ущерб участникам дорожного движения, транспортным средствам и объектам транспортной инфраструктуры, что, в свою очередь, приводит к экономическим и социальным издержкам. По данным Госавтоинспекции на федеральных дорогах Российской Федерации только за октябрь 2019 г. произошло 17.0 тыс. ДТП, в которых погибшими числятся 3.9 тыс. человек, а пострадавшими – 25.6 тыс. человек [1]. Цифровизация процессов управления, развитие интеллектуальных технологий и методов обработки больших данных привели к появлению новых решений, которые могут быть использованы в задаче оперативного прогнозирования возникновения ДТП для принятия превентивных контрмер по предотвращению ДТП [2].

К актуальным проблемам современного дорожного движения, которые могут быть детектированы или предсказаны до возникновения ДТП, можно отнести [3]:

- несоответствующие скоростные режимы движения транспортных средств;
- экстремальные погодные условия;
- повреждения дорожного полотна и технических средств организации дорожного движения;
- опасное поведение: агрессивное вождение, препятствование обгону, несоблюдение безопасной дистанции между транспортными средствами, резкое торможение, выход пешеходов на проезжую часть;
- нарушение правил дорожного движения.

Развитие активных и пассивных средств обеспечения безопасности дорожного движения за последние годы позволило существенно снизить количество ДТП и тяжесть их последствий, однако в некоторых ситуациях внедряемые меры являются недостаточными, например, при заносе транспортного средства на скользкой дороге или в случае невнимательности водителя [4]. Обеспечение безопасности дорожного движения и снижение ущерба от предсказанного ДТП может быть достигнуто за счет директивного и косвенного воздействия на поведение на дороге путем активного управления светофорами и дорожными знаками переменной информации, оперативного уведомления спецслужб, а также информирования участников дорожного движения.

В данной статье предлагается подход, позволяющий на основе интеллектуальных технологий анализа больших данных в режиме онлайн (оперативно) прогнозировать возможность возникновения ДТП.

## 2. Анализ текущего состояния исследований

В рамках исследования под ДТП будем понимать такой дорожный инцидент, который возник с участием хотя бы одного транспортного средства в процессе его движения по улично-дорожной сети, при котором пострадали (погибли или ранены) люди или причинен ущерб транспортным средствам, перевозимому грузу, объектам транспортной инфраструктуры [5].

В настоящий момент ведется активная разработка методов и инструментальных средств, позволяющих обнаруживать [6], предсказывать [7], производить информирование [8] и предотвращать ДТП [9]. Существует ряд решений, основанных на измерительных датчиках [10, 11]. В [12] описан подход на основе инфракрасных датчиков, обеспечивающий работу в двухфазном режиме: обнаружение ДТП, профилактика ДТП. Реализация подхода оперирует показателями загруженности дороги, но не учитывает другие факторы, которые могут оказать влияние на моделирование опасной ситуации [13].

В [14] представлена модель краткосрочного прогнозирования транспортного потока, учитывающая пространственное и временное канализирование. Реализация модели выполнена с помощью фреймворка Apache Spark на основе модели распределённых вычислений MapReduce, тем самым достигнута высокая скорость работы, достаточная для прогнозирования в режиме онлайн, однако не реализован функциональный блок по анализу возможности ДТП.

Интеллектуальный подход на базе нейронной сети, позволяющий автоматически обнаруживать уже случившееся ДТП по косвенным дорожным данным, представлен в [15]. В основу подхода положено предположение о закономерности изменения средней скорости движения транспортного потока в случае ДТП. Предложенный подход не позволяет прогнозировать возможность ДТП. В [16] приведены геоинформационные модели для управления транспортными потоками в случае выявления ДТП, однако получение достоверных данных о факте ДТП не выполняется.

В [17] проведен анализ нескольких контролируемых методов обучения для классификации степени ущерба в результате ДТП: со смертельным исходом, тяжелые травмы, легкие травмы и автомобильное столкновение. Предложенное в рассматриваемой статье решение не может использоваться для мониторинга обстановки на участке улично-дорожной сети и, соответственно, не может использоваться как составная часть системы предотвращения ДТП.

В [18] предложен метод определения временных характеристик ДТП на основе скоростной термограммы, однако он обеспечивает низкие показатели качества. В [19] предложено использовать вейвлет-спектрограммы для оценки характеристик транспортных потоков, но определение факторов, приводящих к ДТП, возможно только по косвенным признакам с недостаточной временной точностью привязки события.

В настоящем исследовании предлагается подход к прогнозированию ДТП с разделением по видам ДТП, использующий многослойный перцептрон Румельхарта [20] применительно к большим данным, поступающим в онлайн режиме из внешних разнородных источников данных, предоставляющих погодные условия, информация об участке улично-дорожной сети и событиях на нем, информации о транспортных средствах и др.

### 3. Модель нейронной сети

#### 3.1. Структура входных данных

Для обучения и тестирования нейронной сети применяется модель обучения с учителем, поэтому потребуются  $n$ -мерные вектора, описывающие погодные, дорожные и организационные факторы.

Данные об участке улично-дорожной сети включают в себя:

- тип устройства контроля движения, TRAFDEV  $\in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ ,  $x_1$ ;
- состояние устройства контроля движения, TRAFFUNCT  $\in \{0, 1, 2, 9\}$ ,  $x_2$ ;
- ограничение скорости, SPEEDLIMIT  $\in \{0, 24, 25, \dots, 119, 120, 121, 999\}$ ,  $x_3$ ;
- тип дороги, RELTOJUNCT  $\in \{0, 1, 2, 3, 4, 5, 9\}$ ,  $x_4$ ;
- тип дорожного покрытия, SURFTYPE  $\in \{1, 2, 3, 4, 5, 8, 9\}$ ,  $x_5$ ;
- количество полос, RDLANES  $\in \{1, 2, 3, 4, 5, 6, 7, 9\}$ ,  $x_6$ ;
- тип разделяющей полосы справа, LINERIGHT  $\in \{0, 1, 2, 3, 4, 5, 9\}$ ,  $x_7$ ;
- тип разделяющей полосы слева, LINELEFT  $\in \{0, 1, 2, 3, 4, 5, 9\}$ ,  $x_8$ .

Данные о погодных условиях:

- погодные условия их набора, WEATHER  $\in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 98, 99\}$ ,  $x_9$ ;
- состояние покрытия дорожной части, SURFCOND  $\in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 98, 99\}$ ,  $x_{10}$ ;
- условия освещения, LIGHTCOND  $\in \{1, 2, 3, 4, 5, 9\}$ ,  $x_{11}$ .

Данные о дате и времени:

- время, CRASHTIME  $\in$  Time,  $x_{12}$ ;
- день недели, DAYOFWEEK  $\in \{1, 2, 3, 4, 5, 6, 7\}$ ,  $x_{13}$ ;
- месяц, CRASHMONTH  $\in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ ,  $x_{14}$ .

Данные о транспортных средствах и детектируемых событиях:

- тип транспортного средства, BODYCAT  $\in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 99\}$ ,  $x_{15}$ ;
- событие на дороге, CRITCAT  $\in \{1, 2, 3, 4, 5, 6, 8, 9\}$ ,  $x_{16}$ ;
- скорость транспортного средства, DVEST  $\in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ ,  $x_{17}$ .

#### 3.2. Структура выходных данных

В качестве выходных данных используется вектор со следующими возможными значениями:

- нет ДТП,  $y_1$ ;
- лобовое столкновение,  $y_2$ ;
- боковое столкновение,  $y_3$ ;
- столкновение сзади,  $y_4$ ;
- опрокидывание,  $y_5$ ;
- наезд на объект вне дороги,  $y_6$ ;
- наезд на объект на дороге,  $y_7$ ;
- иной вид ДТП,  $y_8$ .

Так как вид ДТП кодируется как целое число, то для решения задачи многоклассовой классификации к полученным категориям используется one-hot кодирование:  $0 \rightarrow [1, 0, 0, 0, 0, 0, 0, 0]$ ,  $1 \rightarrow [0, 1, 0, 0, 0, 0, 0, 0]$ , ...,  $7 \rightarrow [0, 0, 0, 0, 0, 0, 0, 1]$ .

#### 3.3. Топология нейронной сети

Прогнозирование ДТП с точки зрения классического машинного обучения относится к задаче множественной классификации. Таким образом, в соответствии с заранее определенными входными и выходными данными, определяется количество входных и выходных нейронов: на вход нейросети поступает вектор из 17 значений, на выходе – вектор из 8 значений. Нейронная сеть выполнена на основе перцептрона Румельхарта с 1 скрытым слоем. Число нейронов скрытого слоя определено по правилу геометрической пирамиды:

$$k = \sqrt{nm},$$

где  $k$  – число нейронов в скрытом слое,  $n$  – число нейронов во входном слое;  $m$  – число нейронов в выходном слое.

Таким образом, число нейронов в скрытом слое – 12. Топология используемой искусственной нейронной сети представлена на рисунке 1.

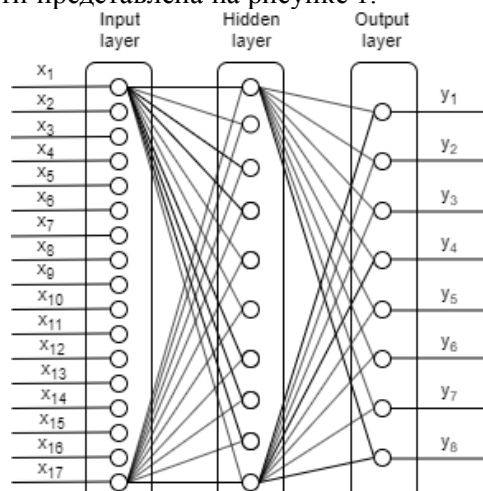


Рисунок 1. Используемая нейронная сеть.

#### 4. Программная реализация

Предлагаемый нейросетевой подход для прогнозирования ДТП реализован в виде программной подсистемы профайлинга, предназначенной для функционирования в составе системы предотвращения ДТП (рисунок 2). Поток данных, поступающие в подсистему профайлинга, логически объединяются в источники данных. Подсистема профайлинга реализована на языке Python в среде PyCharm; для непосредственного использования нейронных сетей применяется фреймворк TensorFlow и библиотека Keras.

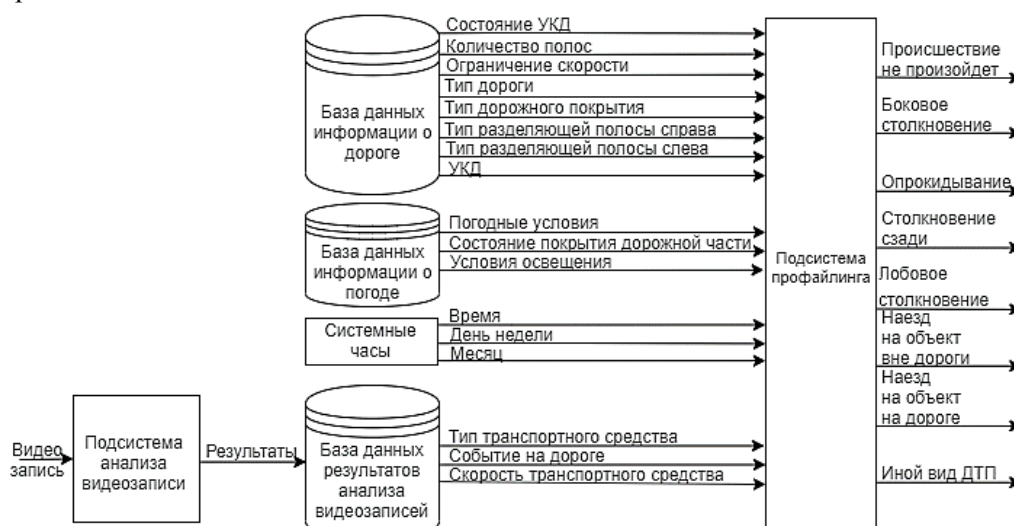


Рисунок 2. Поток данных.

Получаемая в результате работы подсистемы профайлинга аналитика по ДТП сохраняется в базе данных. Для доступа и управления данными использована библиотека `psycopg2` и система управления базами данных PostgreSQL, обеспечивающая пространственно-временную привязку ДТП. Поступающая сопроводительная информация о дороге, погоде, транспортных средствах также логируется в привязке к ДТП (рисунок 3).

Подсистема профайлинга не обладает графическим интерфейсом для конечного пользователя, обеспечивая, помимо записи в базу данных, API для подключения внешних подсистем.

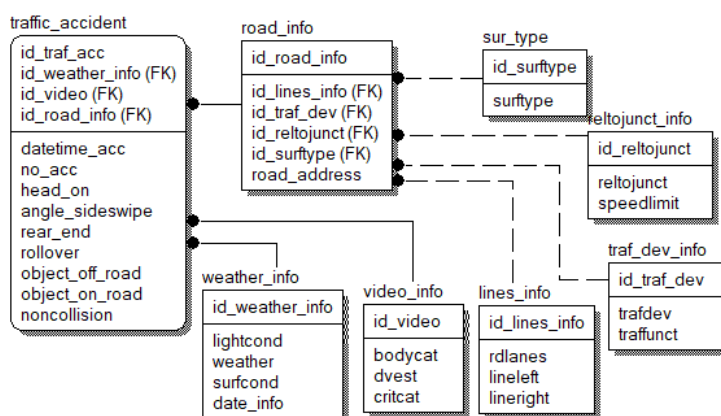


Рисунок 3. Модель данных.

Отдельно разработан графический интерфейс, позволяющий через API подсистемы профайлинга осуществлять обучение и настройку нейросети, выполнять принудительный запуск прогнозирования по указанным наборам данных, осуществлять просмотр уведомлений о возможном возникновении ДТП в режиме ситуационного центра.

## 5. Результаты

Для обучения нейронной сети использовались достоверные данные специального формата, размещенные в свободном доступе на сервере data.gov.uk под лицензией OGL (Open Government License) [21].

Для оценки полученных результатов применялась графическая интерпретация результатов и метрика `roc_auc_score` пакета `sklearn`. Для повышения точности обучения применялось прореживание в 20%. Наилучшие результаты показало количество обучающих примеров за одно обучение – 7.

На рисунках 4 и 5 показаны графики зависимости показателей точности и ошибки для каждой эпохи с конечной 300 эпохой при использовании 7 примеров для обучения одновременно.

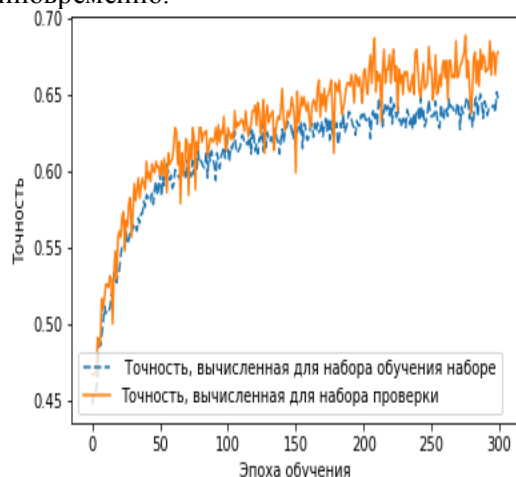


Рисунок 4. График зависимости количества правильных классификаций нейронной сети от эпохи обучения. Конечная эпоха 300.

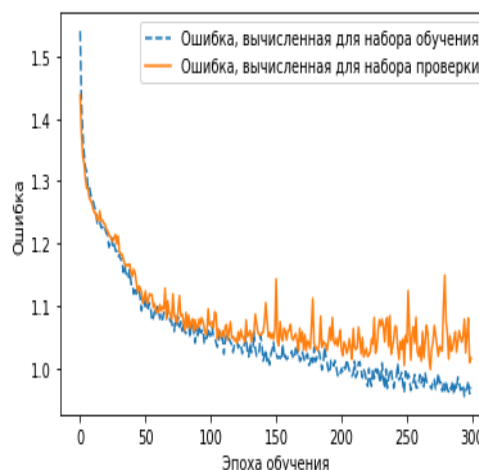
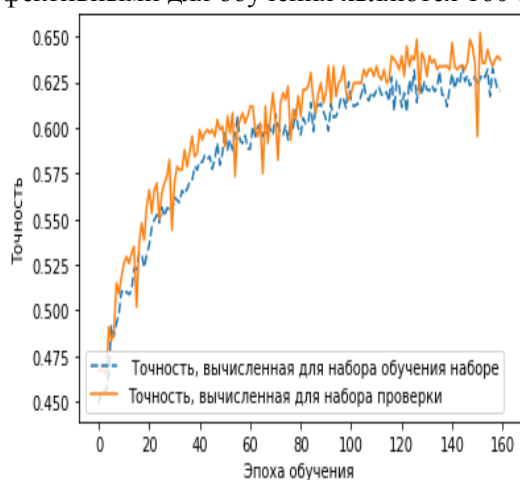


Рисунок 5. График зависимости количества ошибочных классификаций от эпохи обучения. Конечная эпоха 300.

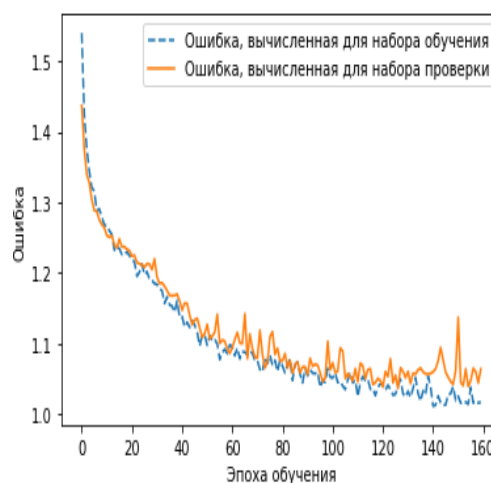
Начиная с 200 эпохи показатели остаются примерно в одних и тех же пределах, при этом метрика `roc_auc_score` показывает результат 0.90.

В целях недопущения переобучения нейросети было снижено количество эпох. Графики на рисунках 6 и 7 показывают показатели при обучении на 160 эпохах; метрика `roc_auc_score`

показывает результат 0.91. Таким образом, при таких же параметрах обучения и топологии сети эффективными для обучения являются 160 эпох.



**Рисунок 6.** График зависимости количества правильных классификаций нейронной сети от эпохи обучения. Конечная эпоха 160.



**Рисунок 7.** График зависимости количества ошибочных классификаций от эпохи обучения. Конечная эпоха 160.

Поскольку прогнозирование осуществляется с учетом показателей распознавания событий, то учитывая приведенные выше результаты, можно выполнить оценку временного промежутка при решении задачи прогнозирования. При стационарном режиме работы производится постоянный мониторинг обстановки на дороге и учитываются показатели, изменяющиеся каждую секунду, что приводит к максимальному значению временного интервала в 1 мин. Однако, ввиду того, что учитываются и относительно постоянные показатели, такие как погодные условия и состояние участка улично-дорожной сети, то результаты прогнозирования могут быть полезны в течении часа с момента их расчёта, например, в виде ориентировочных данных для водителей.

## 6. Заключение и обсуждение

Повышение безопасности на дорогах с помощью новейших достижений науки и техники – очевидный путь для развитого общества по сокращению количества инцидентов и ДТП. Непрерывно совершенствуются интеллектуальные транспортные системы, системы для «умного» и «безопасного» города, передовые технические средства обеспечения пассивной и активной безопасности. Внедрение технологий обработки больших данных и машинного обучения видится эффективным во многих сферах транспортной отрасли, в том числе и для прогнозирования возможности возникновения ДТП.

Представленный в работе подход позволяет повысить точность определения возможности возникновения ДТП за счет анализа классов параметров, охватывающих такие важные факторы как: погодные условия, вид и состояние дороги и управляющих устройств, сезонные колебания трафика, скорость транспортного потока и отдельного транспортного средства. Классификация по видам ДТП позволяет принимать наиболее эффективные меры, направленные на недопущение ДТП конкретного вида. Проведенные эксперименты позволили выявить наиболее эффективные параметры нейросети и добиться точности распознавания ситуаций в 90%. Программная реализация предлагаемого подхода позволит в интеграции с системами предотвращения ДТП достичь уменьшения аварийности, снизить тяжесть последствий ДТП и повысить информированность участников движения.

## 7. Литература

- [1] Официальный сайт Госавтоинспекции [Электронный ресурс]. – Режим доступа: <http://stat.gibdd.ru> (01.12.2019).

- [2] Майоров, В.И. Управление риском в системе обеспечения безопасности дорожного движения // Вестник Уральского института экономики, управления и права. – 2018. – Т. 3, № 44. – С. 8-12.
- [3] Соколов, Д.В. Актуальные проблемы обеспечения безопасности дорожного движения в Российской Федерации // Д.В. Соколов, Е.В. Колузакова // Наука. Мысль: электронный периодический журнал, 2016.
- [4] Tarko, A.P. Surrogate measures of safety / A.P. Tarko // *Safe Mobility: Challenges, Methodology and Solutions*, 2018. – P. 383-405.
- [5] Федеральный закон от 10.12.1995 N 196-ФЗ (ред. от 27.12.2018) «О безопасности дорожного движения» (с изм. и доп., вступ. в силу с 01.11.2019).
- [6] Khalil, U. Automatic Road Accident Detection Using Ultrasonic Sensor / U. Khalil, A. Nasir, S.M. Khan, T. Javid, S.A. Raza, A. Siddiqui // *IEEE 21st International Multi-Topic Conference (INMIC)*, 2018. – P. 206-212.
- [7] Ming, Z. Traffic accident's severity prediction: A deep-learning approach-based CNN network / Z. Ming, L. Tong, Z.R.C. Jing, M. Zifei, T. Mingjing, C. Zhongqiang, W. Zhan // *IEEE Access*. – 2019. – Vol. 7. – P. 39897-39910.
- [8] Tai, W. RTAIS: Road Traffic Accident Information System / W. Tai, H. Wang, C. Chiang, C. Chien, K. Lai, T. Huang // *IEEE 4th International Conference on Data Science and Systems*. – 2018. – P. 1393-1397.
- [9] Lamr, M. Big Data and Its Usage in Systems of Early Warning of Traffic Accident Risks / M. Lamr // *IEEE*, 2018. – P. 154-157.
- [10] Pan, Y. Structural health monitoring and assessment using wavelet packet energy spectrum / Y. Pan, L. Zhang, X. Wu, K. Zhang, M.J. Skibniewski // *Safety Science*. – 2019. – Vol. 120. – P. 652-665.
- [11] Dehzangi, O. GSR-based distracted driving identification using discrete & continuous decomposition and wavelet packet transform / O. Dehzangi, V. Sahu, V. Rajendra, M. Taherisadr // *Smart Health*. – 2019. – № 100085.
- [12] Wadhahi, N.T.S.A. Accident detection and prevention system to reduce the risk of traffic using infrared sensors / N.T.S.A. Wadhahi, S.M. Hussain, K.M. Yosof, S.A. Hussain, A.V. Singh // *7th International Conference on Reliability, Information and Communication Technologies and Optimization (ICRITO)*, 2018. – P. 737-741.
- [13] Головнин, О.К. Управляемый данными анализ транспортных потоков в различных дорожных условиях / О.К. Головнин // V Международная конференция и молодёжная школа «Информационные технологии и нанотехнологии» (ИТНТ) – Самара: Новая техника. – 2019. – Т. 4. – С. 533-542.
- [14] Агафонов, А.А. Анализ больших данных в геоинформационной задаче краткосрочного прогнозирования параметров транспортного потока на базе метода k ближайших соседей // А.А. Агафонов, А.С. Юмаганов, В.В. Мясников // *Компьютерная оптика*. – 2018. – Т. 42, № 6. – С. 1101-1111.
- [15] Ki, Y. Method for Automatic Detection of Traffic Incidents Using Neural Networks and Traffic Data / Y. Ki, J. Kim, T. Kim, N. Heo, J. Choi, J. Jeong // *IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2018. – P. 184-188.
- [16] Mikheeva, T.I. GIS-based Models for Transport Network Emergency Management / T.I. Mikheeva, A.A. Osmushin, S.V. Mikheev, O.K. Golovnin // *Journal of Physics: Conference Series*. – 2019. – Vol. 1353(1). – P. 012009.
- [17] Labib, M.F. Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh / M.F. Labib, A.S. Rifat, M.M. Hossain, A.K. Das, F. Nawrine // *7th International Conference on Smart Computing & Communications (ICSCC)*, 2019. – P. 1-5.
- [18] Zhang, J. Prediction of urban expressway total traffic accident duration based on multiple linear regression and artificial neural network / J. Zhang, J. Wang, S. Fang // *5th International Conference on Transportation Information and Safety (ICTIS)*, 2019. – P. 503-510.

- [19] Головнин, О.К. Вейвлет-анализ как инструмент исследования характеристик дорожного движения для интеллектуальных транспортных систем в условиях недостающих данных / О.К. Головнин, А.А. Столбова // Труды СПИИРАН. – 2019. – № 2(18). – С. 326-353.
- [20] Самонина, О.А. Методы и проблемы обучения многослойной нейронной сети, используемой для оценки характеристик проектируемых информационных систем // Известия Петербургского университета путей сообщения. – 2008. – № 4. – С. 148-159.
- [21] Road Safety Data [Электронный ресурс]. – Режим доступа: <https://data.gov.uk/dataset/road-accidents-safety-data> (01.12.2019).

## Operational forecasting of road traffic accidents via neural network analysis of Big Data

О.К. Golovnin<sup>1</sup>, E.V. Sidorova<sup>1</sup>

<sup>1</sup>Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

**Abstract.** The paper proposes an approach to the operational forecasting of traffic accidents with the separation of accident types based on the multilayer Rumelhart perceptron. The approach is applied to analyze Big Data consists of data collected from external heterogeneous data sources. Weather, road, organizational factors and traffic flow parameters are taken into account. The software implementation of the approach uses the TensorFlow framework and the Keras library. The experiments showed that the approach provides a 90% accuracy in recognizing situations. The software implementation is intended to function as part of accident prevention systems.