

Научная статья

УДК 004.27+004.056

DOI:10.31854/1813-324X-2022-8-4-89-99



# Методы машинного обучения для прогнозирования трафика в многоуровневой облачной архитектуре для сервисов автономных транспортных средств

✉ **Малик Аль-Свейти**, aldonasmar@gmail.com

Санкт-Петербургский государственный университет телекоммуникаций им. М.А. Бонч-Бруевича,  
Санкт-Петербург, 193232, Российская Федерация

**Аннотация:** Автономные транспортные средства (AV, аббр. от англ. Autonomous Vehicle) являются одним из наиболее важных новых вариантов использования и перспективной технологией для сетей пятого поколения (5G) и следующего поколения в многочисленных приложениях. В настоящее время использование AV экспоненциально растет во всем мире, благодаря быстрому росту осведомленности и применению методов искусственного интеллекта в различных областях. Прогнозирование потоков данных необходимо для AV, чтобы улучшить передачу данных и уменьшить задержки за счет более эффективного использования соответствующих возможностей, мониторинга, управления и контроля дорожной системы. В данной работе предлагается подход глубокого обучения с двунаправленной моделью с долгой краткосрочной памятью (BI-LSTM, аббр. от англ. Bidirectional Long-Short-Term Memory) для прогнозирования сетевого трафика AV с многоуровневыми сервисами облачных вычислений. С точки зрения точности прогнозирования проводится сравнение между BI-LSTM и однонаправленной моделью с долгой краткосрочной памятью (LSTM) в зависимости от количества используемых пакетов. Точность предсказания рассчитывается с помощью среднеквадратичной ошибки, средней абсолютной процентной ошибки, коэффициента детерминации ( $R^2$ ) и времени обработки. Результаты показывают, что точность прогнозирования с помощью BI-LSTM превосходит модель LSTM. Кроме того, точность прогнозирования с использованием размера обучающей партии (BatchSize) равной 8, превосходит конкурентов и обеспечивает высокую производительность.

**Ключевые слова:** беспилотные автомобили, 5G, MEC, ML, DL, прогнозирование, BI-LSTM, LSTM

**Ссылка для цитирования:** Аль-Свейти М. Методы машинного обучения для прогнозирования трафика в многоуровневой облачной архитектуре для сервисов автономных транспортных средств // Труды учебных заведений связи. 2022. Т. 8. № 4. С. 89–99. DOI:10.31854/1813-324X-2022-8-4-89-99

## Deep Learning Approaches for Traffic Prediction Forecasting in Multi-Level Cloud Architecture for Autonomous Vehicle Services

✉ **Malik Alsweity**, aldonasmar@gmail.com

The Bonch-Bruevich Saint-Petersburg State University of Telecommunications,  
St. Petersburg, 193232, Russian Federation

**Abstract:** Autonomous vehicle (AV) is one of the most new use cases and a technology for fifth-generation (5G) and next-generation mobile networks in numerous applications., the use of AVs has exponentially worldwide due to the rapidly growing awareness and use of artificial intelligence (AI) methods in various fields. Predicting data flows is essential for AVs to improve data transmission and decrease delays through more efficient use of appropriate

capabilities, monitoring, management, and control of the traffic system. This paper proposes a deep learning approach (DL) with the bidirectional long-short-term memory model (BI-LSTM) for predicting the traffic rates of AVs with multi-cloud services. In terms of prediction accuracy, a comparison is conducted between the BI-LSTM and the unidirectional LSTM based on the number of batch sizes used. The prediction accuracy is computed using the root mean square error (RMSE), mean absolute percentage error (MAPE), coefficient of determination ( $R^2$ ), and processing time. The results show that the prediction accuracy with BI-LSTM outperforms the LSTM model. Besides, the prediction accuracy using 8 batch sizes outperforms the competitors and offers outstanding performance.

**Keywords:** autonomous vehicles, 5G, MEC, ML, DL, prediction, BI-LSTM, LSTM

**For citation:** Alsweity M. Deep Learning Approaches for Traffic Prediction Forecasting in Multi-Level Cloud Architecture for Autonomous Vehicle Services. *Proc. of Telecom. Universities*. 2022;8(4):89–99. (in Russ.) DOI:10.31854/1813-324X-2022-8-4-89-99

## ВВЕДЕНИЕ

Многоуровневая сетевая структура отличается от любой другой структуры, поскольку она охватывает, контролирует, управляет локальными облачными структурами и предлагает лучшие возможности, чем поставщики облачных услуг. Многоуровневая архитектура обеспечивает стабильную и повторяемую архитектуру на нескольких платформах: она является основной в производстве, а также ядром нынешнего и будущего публичного облака. С ростом использования Интернета вещей (IoT, аббр. от англ. Internet of Things) в различных сферах жизни, включая обмен данными в реальном времени между пользователями и машинами, приобрел актуальность тактильный интернет, который позволил системе IoT мгновенно и быстро взаимодействовать в промышленном и коммерческом секторах.

Ожидается, что технологии беспроводной связи следующего поколения, включая сети 5G и выше, станут важным фактором успеха беспилотных автомобилей в ближайшие годы. Они позволят внедрить новые услуги и сценарии применения в будущем интеллектуальном транспорте, благодаря возможности быстрого и надежного обмена данными между различными субъектами. Развитие коммуникационных технологий для интеллектуальных транспортных систем в сочетании с передовой обработкой данных и искусственным интеллектом (AI, аббр. от англ. Artificial Intelligence) означает, что такие важные функции как обеспечение крупномасштабных сетей, выбор технологий для гетерогенного беспроводного доступа, а также интеллектуальные и безопасные интегрированные услуги, пользуются большим спросом [1].

Полностью функциональная и эффективная сеть 5G невозможна без AI. Сочетание машинного обучения (ML аббр. от англ. Machine learning) и AI на границе сети может быть достигнуто путем распределения сетей 5G. Данное сочетание обеспечивает синхронное подключение нескольких устройств IoT и генерирует большое количество данных, которые могут быть обработаны с помощью ML и AI. Это позволяет таким сетям быть

предиктивными и проактивными, что необходимо для эффективности сети пятого поколения. Добавив ML к технологии 5G, интеллектуальные каналы могут принимать собственные решения. Мобильные устройства могут создавать гибкие, динамические коллекции в соответствии со считанными данными, что повышает эффективность, латентность и надежность сетевых приложений.

Как только ML и AI будут объединены с 5G и периферийными вычислениями для множественного доступа (MEC, аббр. от англ. Multi-Access Edge Computing), мобильные операторы смогут обеспечить [2, 3]:

- значительные вычислительные возможности за счет распределенной структуры ML и AI на границе сети;

- управление трафиком в сетях доступа на основе приложений;

- динамическую нарезку сети (слайсинг, от англ. slice), которая охватывает различные сценарии с различными условиями качества обслуживания (QoS, аббр. от англ. Quality of Service) [4].

Беспилотные автомобили с поддержкой многоуровневых облачных вычислений растут во всем мире по экспоненте, поскольку репутация методов искусственного интеллекта и их использование в многочисленных приложениях резко возросли. Прогнозирование транспортных потоков необходимо для того, чтобы беспилотные автомобили определяли маршрут движения и принимали адаптивные решения относительно окружающих их объектов. Связь между транспортными средствами (V2V, аббр. от англ. Vehicle-to-Vehicle) позволяет автомобилям общаться друг с другом, передавая данные о скорости и местоположении. Когда система транспортных средств обменивается данными с придорожной инфраструктурой (V2I, аббр. от англ. Vehicle-to-Infrastructure), придорожное устройство (RSU, аббр. от англ. Roadside Unit) берет на себя управление движением транспортных средств и отправляет широковещательное сообщение всем ближайшим автомобилям [5, 6].

Существуют специфические проблемы, связанные с маршрутизацией в автомобильных сетях

(VANET) – топология сети, различная плотность трафика в разное время и быстрые изменения прибывающих и убывающих транспортных средств и пр. Все эти проблемы требуют системного решения на базе современных технических решений, в качестве которого предлагается построение некой системы для сервисов AV с многоуровневой облачной архитектурой поддержки вычислений с использованием методов искусственного интеллекта.

Предлагаемая архитектура системы, показанная на рисунке 1, состоит из четырех уровней. Уровень 1 – это физическое оборудование, включающее в себя бортовые устройства (OBU, аббр. от англ. On-Board Unit), которые могут быть подключены к распределенным RSU. RSU может представлять собой интеллектуальный мультисервисный узел доступа, используемый вместо небольших сетевых узлов при расширении традиционных сетей и создании новых сетей с большим количеством удаленных объектов.

Уровень 2 содержит гетерогенные, облачные, граничные серверы, расположенные вдоль границы радиосети (RAN, аббр. от англ. Radio Access Network). Предлагаемая система состоит из двух серверов MEC: миниоблака и микрооблачного граничного сервера. Каждый RSU связан с мобильными базовыми станциями и граничным сервером микрооблака. Граничный сервер микрооблака работает с ограниченными вычислительными и энергетическими возможностями. Каждый набор серверов микрооблака связан с пограничным сервером миниоблака с мощными вычислительными и энергетическими ресурсами.

Уровень 3 – это программно-определяемая сеть (SDN, аббр. от англ. Software-Defined Network), распределенная по основной сети. Этот уровень состоит из двух основных подуровней: управления и уровня данных. На уровне данных находятся распределенные виртуальные коммутаторы OpenFlow, которые обслуживают протокол OpenFlow.

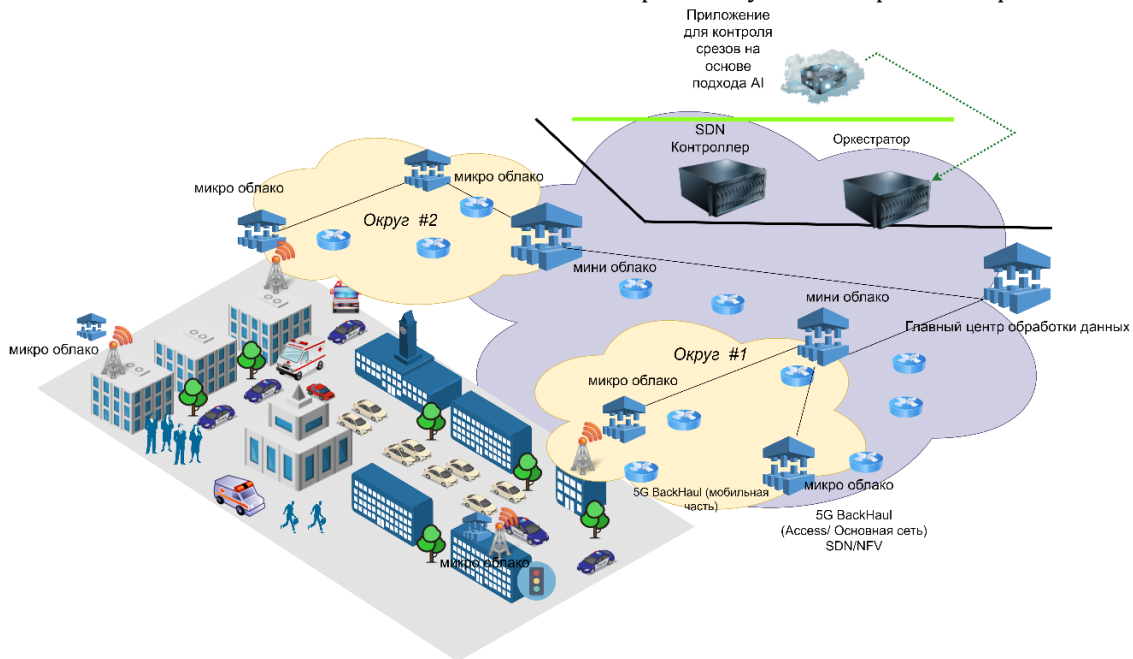


Рис. 1. Предлагаемая многоуровневая облачная система

Fig. 1. The proposed Multi-Level Cloud System

Уровень управления (Уровень 4) включает распределенные SDN для управления и контроля всей сети. Интерфейс прикладных программ (API, аббр. от англ. Application Programming Interface) управляет SDN и основными сетевыми интерфейсами к Уровню 1 и серверу приложений. Предлагаемая система имеет два прямых интерфейса на Уровне 1. Первичным интерфейсом является транспортное средство, а последующим – так называемая, «внутренняя инфраструктура» транспортного средства. Это может быть достигнуто через соответствующую сеть IEEE 802.11, различные стандарты связи для беспроводной локальной сети связи (WLAN,

аббр. от англ. Wireless Local Area Network) на различных частотах.

На Уровне 2 распределенные граничные серверы связаны через быструю беспроводную сеть. Каждый RSU связан с сервером MEC для обеспечения вычислительной мощности на границе сети доступа. Это снижает трафик на ведущую сеть, уменьшая ее перегрузку или потерю пакетов.

Концепции глубокого обучения (DL, аббр. от англ. Deep Learning) привлекают внимание многих исследователей благодаря своей способности решать проблемы распознавания, понимания естественного языка, сокращения размерности, иден-

тификации объектов, моделирования движения, прогнозирования трафика и т. д. DL использует концепции многослойных нейронных сетей для вывода глубинных свойств данных от самого низкого до самого высокого уровня. Они находят в данных большое количество структур, которые позволяют их представить и сделать логические выводы. Разработка беспилотного автомобиля, который может сделать транспортные сети гораздо более экономичными и снизить риск для жизни людей, также является предметом изучения разработчиков и исследователей интеллектуальных транспортных систем (ITS, аббр. от англ. *Intelligent transportation system*). Концепция DL также имеет такое существенное преимущество, как экономия времени, что особенно важно для обеспечения безопасного автоматизированного вождения – информация от систем помощи водителю, автономных транспортных средств и систем распознавания дорожных знаков [1, 7] должна поступать своевременно, без критической во величине задержки.

Прогнозирование сетевого трафика является важным инструментом для беспроводных сетей 5G и других сетей, позволяющим управлять перегрузкой и ресурсами сети, а также обеспечивать безопасность, надежность и устойчивость (связность) выбранных сетей. Использование статистических временных рядов или подходов машинного обучения позволяет прогнозировать будущие потоки сетевого трафика на основе предыдущих данных. Эффективное прогнозирование может улучшить QoS и снизить эксплуатационные расходы для поставщиков сетевых услуг. Существующие источники показывают, что DL-модели более эффективно подходят для моделирования сетевого трафика и более точно предсказывают будущий трафик, чем традиционные модели [2, 3, 8].

### Анализ существующих работ

В последнее время многие исследователи сосредоточились на прогнозировании трафика беспроводных сетей с использованием методов DL и технологии 5G. Данная статья посвящена прогнозированию трафика в автономных транспортных средствах с многоуровневыми облачными вычислениями с использованием DL-подхода на основе BI-LSTM (аббр. от англ. *Bidirectional Long-Short-Term Memoгу* – двунаправленная модель с долгой краткосрочной памятью) и LSTM (однонаправленная модель). Ниже представлены основные работы, имеющие отношение к предметной области.

В работе [9] рассматривается прогнозирование сетевого трафика с использованием глубоких нейронных сетей (DNN, аббр. от англ. *Deep Neural Network*). Исследуется то, как DNN могут заранее предсказать приблизительную скорость сетевого взаимодействия. В работе [10] было выполнено

прогнозирование задержки IoT с помощью многопараметрического метода DNN. В [11, 12] изучалось долгосрочное и краткосрочное прогнозирование трафика беспроводной сети с помощью LSTM с поддержкой DL. В [13] авторами представлены решения на основе DL для сетей 5G и интернета транспортных средств (IoV, аббр. от англ. *Internet of Vehicles*). В статье [14] рассматриваются последние исследования беспроводных сетей следующего поколения и роль методов машинного обучения (ML, аббр. от англ. *Machine Learning*) в этих будущих беспроводных системах. Прогнозирование трафика по временным рядам с использованием DL и LSTM-сети для граничных вычислений с поддержкой IoT выполнено в работе [15].

В работе [16] представлены исследования проблемы прогнозирования трафика IoT с помощью ML, DL, статистических методов прогнозирования на основе временных рядов, таких как LSTM, ARIMA, VARMA и нейронных сетей прямого распространения (FFNNs, аббр. от англ. *Feed Forward Neural Networks*). В статье [17] был представлен обширный обзор современных ML-подходов к прогнозированию сотового трафика в сетях 5G. В работе [17] проведены исследования прогнозирования маршрута транспортного средства на основе ML с использованием V2V-взаимодействия и бортовых датчиков. Представлен метод прогнозирования маршрута с использованием алгоритма случайного леса и архитектуры кодер-декодер LSTM.

Проанализированные релевантные работы позволили сделать некоторые предварительные выводы относительно способа решения задачи прогнозирования трафика в беспроводной коммуникационной среде AV с многоуровневой облачной архитектурой поддержки вычислений с использованием методов искусственного интеллекта. Представляется, что гипотетически перспективным является DL-подход на основе моделей рекуррентных нейронных сетей типа LSTM и их возможных модификаций, адаптированных под предметную область.

### Рекуррентные нейронные сети для решения задач последовательного

#### A. Сеть долгой краткосрочной памяти (LSTM)

Одной из самых популярных и часто используемых модельных структур в искусственном интеллекте является LSTM. Сеть LSTM – это тип рекуррентной нейронной сети (RNN, аббр. от англ. *Recurrent Neural Network*), наиболее подходящих для долгосрочного хранения данных и обычно используемых в области DL. Поскольку большие объемы данных могут повлиять на производительность модели, она является естественным выбором для построения, обучения и развертывания сети. Преимущество LSTM в том, что она помогает обес-



печить меньшее влияние на независимые связи. Кроме того, сеть должна быть обучена определять, какой фрагмент информации подходит для сети, и иметь точное представление о том, что нужно хранить. В случае DL, LSTM применяется многократно.

Стандартная структура ячеек LSTM показана на рисунке 2. Ячейка в LSTM имеет три основных узла, называемых гейтами (gate), управляющих ее состоянием: входной (input gate), забывания (forget gate) и выходной (output gate), а фактическая ячейка обратной связи – это скрытое состояние.

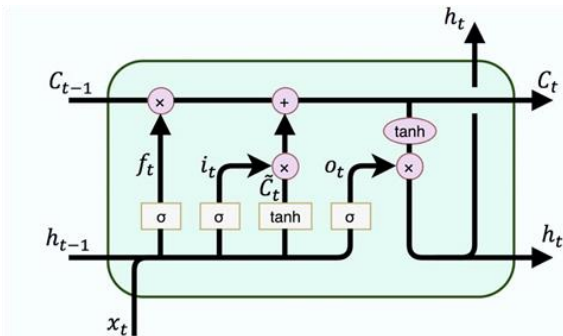


Рис. 2. Структура сети LSTM  
Fig. 2. Structure of LSTM Network

Гейты LSTM представляют собой сигмоидальную передаточную функцию (1):

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{1}$$

Следует оценить гейт забывания, как процентный вклад входа; входной гейт выбирает текущую информацию для хранения в ячейке памяти. Текущий вектор значений-кандидатов может быть использован для состояния слоя активации гиперболического тангенса (tanh-слой – один из типов слоев, используемых в нейронных сетях, который включает свои собственные весовые матрицы). Входной гейт подключается к tanh-слою для получения изменения состояния. Выход функции активации tanh соответствует значению между -1 и 1 для определения его значимости.

Оценка использует функцию tanh(x) следующим образом:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{2}$$

Другие связи увеличивают связность модели. Предположим, что  $x_t$  и  $h_t$  – это векторы входного и скрытого состояния по времени  $t$ . В этом случае  $W_i$  и  $W_h$  представляют собой матрицы весовых коэффициентов, которые применяются к входному пути и обратной связи, соответственно, а  $b$  – вектор смещения.

Учитывая следующий вход  $x_t$  для скрытого состояния  $f$  с предыдущего шага  $h_{t-1}$  и текущее состояние ячейки  $C_{t-1}$ , вычисляются зависимости для

входного гейта  $i_t$ , ворот забывания  $f_t$  и выходных ворот  $o_t$  в LSTM:

$$i_t = \sigma(w_i x_t + g_i h_{t-1} + b_i), \tag{3}$$

$$f_t = \sigma(w_f x_t + g_f h_{t-1} + b_f), \tag{4}$$

$$o_t = \sigma(w_o x_t + g_o h_{t-1} + b_o), \tag{5}$$

где  $i_t, f_t, o_t$  – функции весовых матриц  $w_i, g_i, w_f, g_f, w_o, g_o$  и векторов смещения  $b_i, b_f, b_o$ .

Компактные формы для уравнений состояния клетки-кандидата, состояния ячейки-кандидата и вектора скрытого состояния (выходного вектора) следующие:

$$C'_t = \tanh(w_c x_t + g_c h_{t-1} + b_c), \tag{6}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot C'_t, \tag{7}$$

$$h_t = o_t \odot \tanh(C'_t), \tag{8}$$

Состояние ячейки-кандидата  $\tilde{C}_t$  определяются как функция весовых матриц  $w_c$  и  $g_c$ ; оператор  $\odot$  обозначает произведение Hadamard (произведение по элементам).

### В. Двухнаправленная LSTM (BI-LSTM)

Двухнаправленные LSTM основаны на блоке LSTM, где два LSTM подаются на входной сигнал. Одна подается на входной ряд (путь прямого распространения). На втором этапе обратная форма входного ряда подается на модель LSTM (путь обратного распространения). Сеть LSTM анализирует только влияние прошлой серии на текущую и игнорирует более позднюю информацию, поэтому оптимальная производительность не достигается. В отличие от этого BI-LSTM связывает выход блока LSTM двухнаправленно. Она улавливает двухнаправленные зависимости сигналов, что повышает общую производительность модели. BI-LSTM – это последовательная вычислительная модель, состоящая из двух LSTM. Каждая нейронная сеть имеет ряд информации в обратном пути (от приходящей к предыдущей) и прямом пути (от последней к приходящей). В двухнаправленной сети входные данные поступают в двух направлениях, что отличает BI-LSTM от обычной LSTM. В обычной LSTM входной поток может идти в одном направлении: либо назад, либо вперед. В двухнаправленной LSTM можно заставить входные данные поступать в обоих направлениях, чтобы получить информацию из будущего и прошлого.

BI-LSTM эффективно увеличивают количество информации, доступной сети, улучшая содержание, доступное алгоритму. Прямое и обратное направления BI-LSTM одновременно передаются на выходной модуль. Таким образом, могут быть получены прошлые и предстоящие данные, как показано на рисунке 3. В каждый момент времени  $t$  вход подается на прямой и обратный путь сети LSTM.

Выход BI-LSTM может быть представлен следующим образом:

$$\vec{h}_t = H(W_1 x_t + W_2 \vec{h}_{t-1} + \vec{b}), \quad (9)$$

$$\vec{h}_t = H(W_3 x_t + W_5 \vec{h}_{t-1} + \vec{b}), \quad (10)$$

$$y_t = W_4 \vec{h}_t + W_6 \vec{h}_t + b_y, \quad (11)$$

где  $\vec{h}_t$ ,  $\vec{h}_t$ , и  $y_t$  – векторы для прямого, обратного и выходного слоев;  $W$  – весовые коэффициенты;  $\vec{b}$ ,  $\vec{b}$  и  $b_y$  – соответствующие векторы смещения.

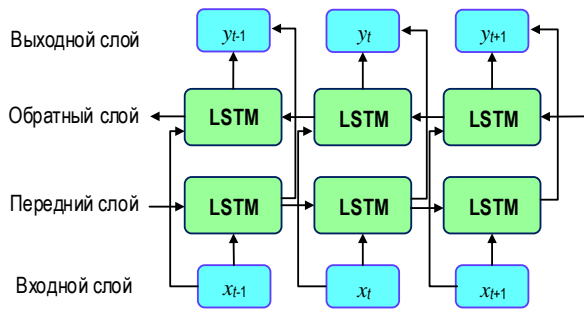


Рис. 3 Структура нейронной сети Bi-LSTM

Fig. 3. Structure of Bi-LSTM Neural Network

На рисунке 3 можно видеть поток информации в обратном и прямом слоях. BI-LSTM обычно используется там, где требуется последовательность задач. Этот тип сети может использоваться в классификации текстов, распознавании речи и прогнозических моделях.

### Постановка задачи и модель системы

Для прогнозирования сетевого трафика использовалось множество методов машинного обучения, таких как Random Forest [18], ARIMA [16, 19], VARMA [16], LSTM [2, 11, 12, 15], NARX [3, 8, 20, 21] и др. В данной статье предложены двунаправленная BI-LSTM и однонаправленная LSTM с DNN для прогнозирования трафика в автономном транспортном средстве на основе многоуровневых облачных вычислений.

Для проведения исследования было выполнено построение DL на основе блоков LSTM и BI-LSTM для прогнозирования трафика. Для получения обучающего набора данных была создана система AV с помощью симулятора AnyLogic. После сбора, исследования и обработки набора данных они использовались в качестве входных данных для DL и для метода прогнозирования. Затем набор данных применялся к сети в двух подмножествах: в столбцах входных данных ( $I$ ) и выводных ( $O$ ), – а затем разбивался на подмножества для обучения и тестирования. Нормирование входных данных должно было находиться в интервале  $[-1, 1]$ , что соответствует фактическим максимальным или минимальным значениям.

Далее выполнялись этапы обучения, тестирования и внедрения. Следует отметить, что в данном

исследовании не использовался проверочный набор для сети и не проводилась оптимизация статических гиперпараметров сети, таких как размер слоя LSTM. Модель DL была обучена в среде VANET с использованием различных размеров порций для BI-LSTM и LSTM. Цель этапа состояла в том, чтобы найти оптимальный размер обучающей партии, который обеспечит наилучшую точность прогнозирования.

### А. Среда VANET

Пусть трафик представляет собой пуассоновский поток, а экспоненциальное распределение описывает время обслуживания. Модель очереди в целом совместима с методом  $M/M/1/k$  с ограниченным буфером.

Тогда вероятность того, что входящая информация не попадет в систему (вероятность потери) определяется выражением:

$$p = \frac{1 - \rho}{1 - \rho^{n+2}} \rho^{n+1}. \quad (12)$$

В этом уравнении  $\rho = a/\mu$ , где  $a$  – интенсивность входящих сообщений;  $\mu$  – интенсивность входящего трафика; в этом случае производительность сервера равна  $ti$ .

Следующее уравнение позволяет рассчитать средний размер очереди при  $i = 1$ :

$$\bar{L} = \frac{\rho(1 - (n + 1)\rho^n + k\rho^{n+1})}{(1 - \rho)(1 - \rho^{n+1})}. \quad (13)$$

Среднее время отклика первого слоя представлено следующим образом:

$$T = (\bar{L} + 1)\bar{t}. \quad (14)$$

Сравнивая конструкцию с ограниченным и неограниченным размером очереди, вычисляется задержка системы  $M/M/1$ :

$$\bar{T}_m = \frac{\bar{t}}{1 - \rho}. \quad (15)$$

Для метода  $M/M/1$  нагрузка на сервер достаточна для обеспечения трафика; для метода  $M/M/1/k$  нагрузка на сервер может быть вычислена следующим образом:

$$u_i = 1 - P_0, \quad (16)$$

$$P_0 = \begin{cases} \frac{1}{n + 1} & \rho_i = 1 \\ \frac{1 - \rho_i}{1 - \rho_i^{n+1}} & \rho_i \neq 1 \end{cases}, \quad (17)$$

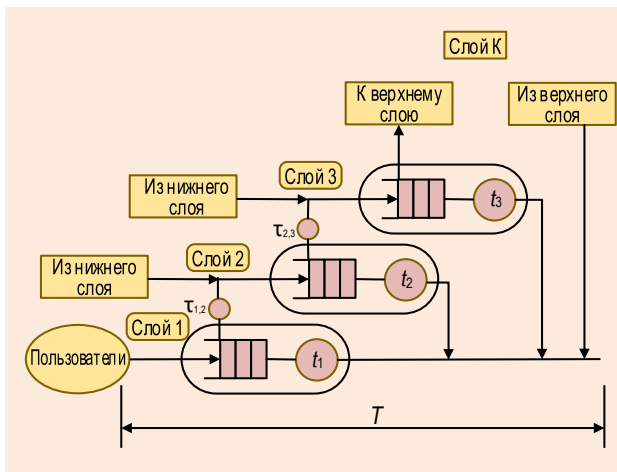
где  $P_0$  – вероятность того, что сервер не работает в это время.

Предлагается модель реализации для МЕС с учетом вышеприведенного предположения о модели очереди. Чтобы увеличить использование метода

МЕС (микро- или миниоблака), следует выбрать режим повышенного использования облака; однако это приведет к ухудшению QoS. Предлагается использовать несколько уровней, включая облака, предназначенные для обслуживания избыточного трафика, поступающего с соседних уровней. Такой метод работает по следующей схеме: коммуникации с регионом VANET, ограниченным регионом EDC, поступают в облако первого уровня, описываемое системой очередей  $M/M/1/k$  (первый уровень). Предполагается, что производительность облака  $\mu_1$  определяется временем обслуживания  $t_1$ . И выбираем значение  $t$  на основе времени доставки  $T_0$  и требований к нагрузке сервера.

**В. Имитационная модель**

Как показано на рисунке 4, структура моделируемой сети представляет собой некую реализацию метода очередей.



**Рис. 4. Структура сети**  
Fig. 4. Network Structure

Каждый слой модели состоит из одного сервера и одной очереди, и каждая очередь имеет два выхода и только один вход. Вход очереди принимает входящую информацию от этого слоя. Выходом является верхняя часть очереди, подключенная к серверу. Если некоторые запросы не были удовлетворены на данном уровне (потерянные рекомендации), они выходят из очереди и продолжают поступать на следующий уровень через этот выход. Существуют разные причины, по которым запросы не были обслужены от одной среды к другой.

Для среды VANET использовалась следующая структура модели: каждый слой описывает стационарную ячейку базовой станции и потерянные запросы из-за чрезмерного трафика в первом случае; во втором случае потери вызваны слишком большой задержкой, превышающей  $T_m$ .

Обычно в этих двух случаях можно использовать несколько приоритетных запросов. Тем не

менее, предполагалось, что все пакеты (сообщения) в среде VANET однородны, и из-за этого использовались неприоритетные очереди.

**Результаты моделирования**

В этой статье осуществляется прогнозирование трафика на основе DL-подхода с использованием BI-LSTM и LSTM сетей для AV с многоуровневыми облачными вычислениями. Точность прогнозирования измерялась с помощью RMSE, MAPE,  $R^2$  и времени обработки, с использованием функции потерь MSE и скорости обучения 0,1. Было проведено сравнение между BI-LSTM и однонаправленной LSTM с использованием RMSE, MAPE,  $R^2$  и времени обработки при различных размерах для среды VANET. Наборы данных для обучения были получены из многоуровневых облачных вычислений с поддержкой AV. Данная работа была выполнена с использованием MATLAB, так как набор данных был собран и обработан, он был разделен на 70 и 30 % для обучения, а также 30 % для тестирования.

Для расчета точности прогнозирования применялись математические формы для RMSE, MAPE и  $R^2$ , как описано в уравнениях (9) и (10):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \tag{28}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{x_t} \right|, \tag{29}$$

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)}{\sum_i (y_i - \bar{y}_i)}, \tag{30}$$

где  $N$  – общее число наблюдений;  $y_i$  – фактическое значение;  $\hat{y}_i$  – предсказанное значение;  $\bar{y}_i$  – среднее значение всех значений;  $SS_{Total}$  – сумма общих квадратичных ошибок;  $SS_{Regression}$  – сумма квадратичных ошибок регрессии.

В таблице 1 показана точность предсказания скорости трафика в случае использования среды VANET при трех различных размерах обучающей партии BatchSizes с точки зрения использования двух различных BI-LSTM и LSTM-моделей с учетом функции потерь MSE в качестве меры производительности и скорости обучения 0,1. Чтобы исследовать модель прогнозирования, которая обеспечивает парадигмальную точность и максимальное среднее улучшение, масштабировалась точность прогнозирования с помощью среднеквадратичной ошибки (RMSE, аббр. от англ. Root Mean Square Error), средней абсолютной процентной ошибки (MAPE, аббр. от англ. Mean Absolute Percentage Error) и коэффициента детерминации  $R^2$ .

**ТАБЛИЦА 1. Точность прогнозирования трафика в среде VANET**  
 TABLE 1. The Prediction Accuracy for the Traffic Rate in the VANET Environment

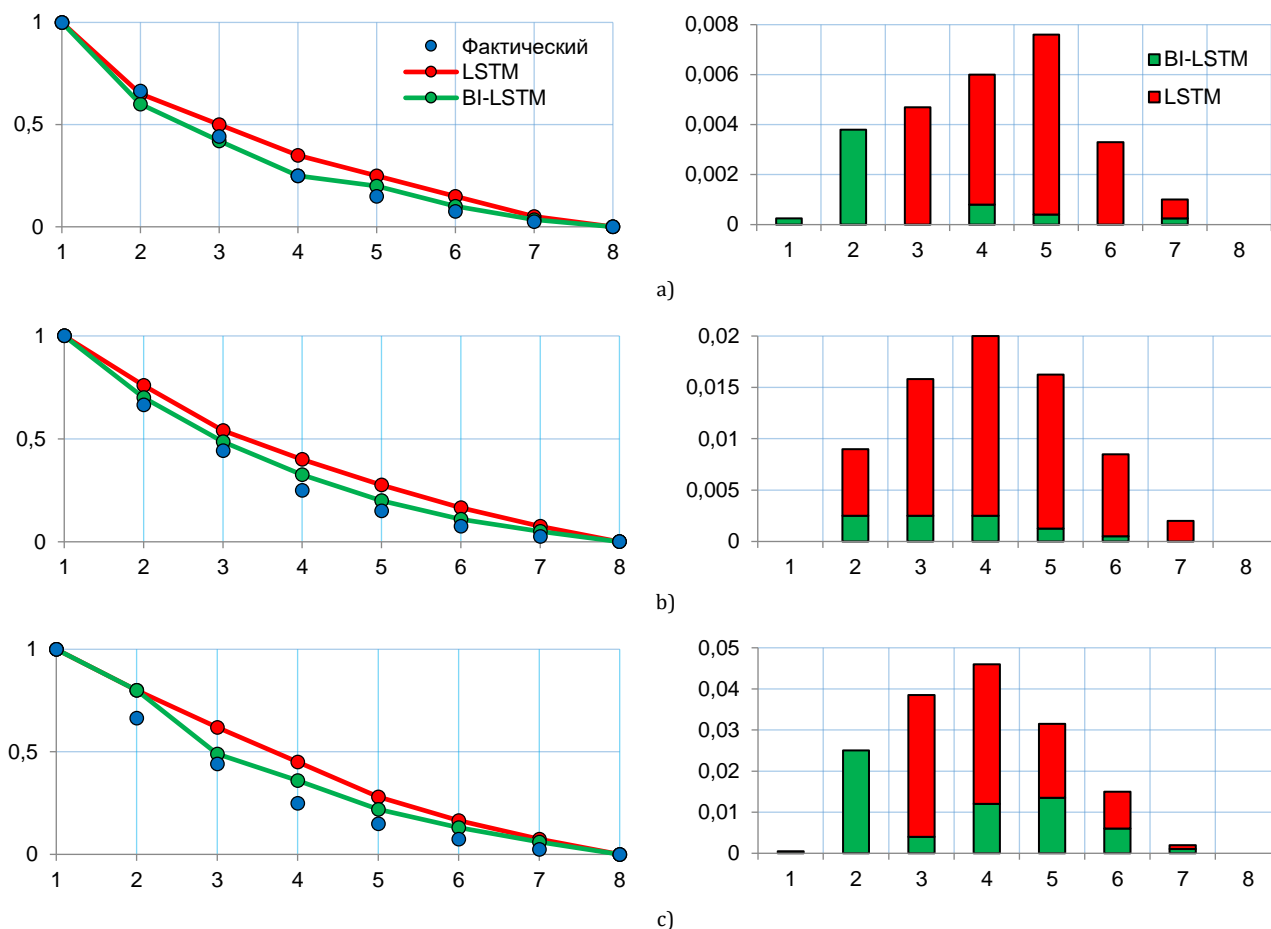
Batch Sizes	BI-LSTM / LSTM			
	RMSE	MAPE%	$R^2$	Время обработки
8	0,0241 / 0,0547	2,9902 / 5,1858	0,9959 / 0,9862	44,5000 / 27,7969
16	0,0350 / 0,0954	4,3937 / 10,3052	0,9958 / 0,9732	39,8750 / 28,6563
64	0,0943 / 0,1379	10,8848 / 15,3204	0,9721 / 0,9436	25,1094 / 36,1719

Как видно из результатов в таблице 1, при использовании BatchSizes = 8 точность предсказания превосходит конкурентов и имеет наилучшие показатели для BI-LSTM и LSTM сетей.

Соответственно, максимальное среднее улучшение в обоих случаях составляет 7.9 и 10.135 %. Однако это занимает больше всего времени обработки для BI-LSTM и быстрее всего для LSTM. Точность прогнозирования при BatchSizes = 16 примерно такая же, как и при BatchSizes = 8. В этом случае максимальное среднее улучшение составляет 6.5 и 5.12 % соответственно для BI-LSTM и LSTM. Однако, точность прогнозирования при использовании BatchSizes = 64 достаточно низкая по сравнению с другими, хотя имеет самое быстрое

время обработки для BI-LSTM, но самое длительное для LSTM. Кроме того, коэффициент детерминации  $R^2$  примерно одинаков в обоих случаях и близок к 1, что указывает на хорошее соответствие. С другой стороны, точность предсказания с помощью BI-LSTM выше, чем у LSTM во всех случаях, а максимальное среднее улучшение BI-LSTM по сравнению с LSTM во всех случаях: 2.2, 5.9 и 4.4 %, соответственно.

На рисунках 5–7 показаны результаты прогнозирования для различных размеров партии Batch Size в BI-LSTM и LSTM-моделях: кривые (слева) представляет собой вероятности ошибки предсказания скорости трафика, а гистограммы (справа) – вероятности потери с течением времени (в секундах).



**Рис. 5. Прогнозирование выходных паттернов в случае использования BatchSize, равной: а) 8, б) 16 и в) 64**  
 Fig. 5. Prediction of Output Patterns in Case of Using a) 8, b) 16 and c) 64 Batch Size



Модели прогнозирования BI-LSTM и LSTM для случаев, когда используются BatchSizes = 8, 16 и 64, как показывают кривые слева, дают примерно одинаковый результат: вероятности ошибки предсказания скорости трафика постепенно уменьшаются со временем, что достаточно похоже на реально наблюдаемый процесс (на графиках – «Фактический»). Гистограммы справа показывают, что модель на основе BI-LSTM имеет наименьшие потери со временем во всех случаях по сравнению с моделью на основе LSTM, а наибольшие потери составляют 5, 4 и 4 сек. при использовании BatchSizes = 8, 16 и 64, соответственно.

### Заключение

В данной статье предлагаются методы DL для прогнозирования трафика в автономных транспортных средствах с многоуровневыми облачными вычислениями. DL был обучен на примере среды VANET, используя BI-LSTM и LSTM сети для различных размеров партии, чтобы исследовать модель прогнозирования, которая обеспечивает парадигмальную точность, а также максимальное среднее улучшение. Точность прогнозирования рассчитана с использованием среднеквадратичной ошибки, средней абсолютной процентной ошибки, коэффициента детерминации  $R^2$  и времени обработки.

Результаты моделирования точности прогнозирования показывают, что BI-LSTM превосходит LSTM во всех случаях. Для BatchSize = 64 точность прогнозирования низкая по сравнению с другими, а при BatchSize = 8 – превосходит конкурентов. Коэффициент детерминации  $R^2$  примерно одинаков в обоих случаях и близок к 1, что означает, что сеть хорошо обучена. При использовании среды VANET время обработки при BatchSize = 64 самое короткое, а при BatchSize = 8 – самое длительное в сравнении с другими.

Полученные результаты исследования, без потери общности, позволяют масштабировать прогнозные модели для систем облачных и туманных вычислений в системе VANETs [21], которые обеспечивают значительные вычислительные возможности за счет DL на границах сети для различного количества облаков, в том числе, в условиях мигрирующего тумана.

Кроме этого, использования данного подхода позволяют перейти к решению оптимизационных задач при взаимодействии граничных устройств VANETs [22] с учетом ограничений полученных прогнозных моделей.

Указанные вопросы определяют будущее направления исследований по данной работе.

### Список источников

1. Abdellah A.R., Alshahrani A., Muthanna A., Koucheryavy A. Performance Estimation in V2X Networks Using Deep Learning-Based M-Estimator Loss Functions in the Presence of Outliers // *Symmetry*. 2021. Vol. 13. Iss. 11. P. 2207. DOI:10.3390/sym13112207.
2. Chochliouros I.P., Spiliopoulou A.S., Lazaridis P., Dardamanis A., Zaharis Z., Kostopoulos A. Dynamic Network Slicing: Challenges and Opportunities // *Proceedings of the International Conference on Artificial Intelligence Applications and Innovations (AIAI 2020, Neos Marmaras, Greece, 5–7 June 2020)*. IFIP WG 12.5 International Workshops. IFIP Advances in Information and Communication Technology. Vol. 585. Cham: Springer, 2020. DOI:10.1007/978-3-030-49190-1\_5
3. Miglani A., Kumar N. Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges // *Vehicular Communications*. 2019. Vol. 20. P. 100184. DOI:10.1016/j.vehcom.2019.100184
4. Gillani M., Niaz H.A., Farooq M.U., Ullah A. Data collection protocols for VANETs: a survey // *Complex & Intelligent Systems*. 2022. Vol. 8. Iss. 3. PP. 2593–2622. DOI:10.1007/s40747-021-00629-x
5. Grigorescu S., Trasnea B., Cocias T., Macesanu G. A survey of deep learning techniques for autonomous driving // *Journal of Field Robotics*. 2020. Vol. 37. Iss. 3. PP. 362–386. DOI:10.1002/rob.21918
6. Pfülb B., Hardegen C., Gepperth A., Rieger S. A Study of Deep Learning for Network Traffic Data Forecasting // *Proceedings of the 28th International Conference on Artificial Neural Networks and Machine Learning (ICANN 2019, Munich, Germany, 17–19 September 2019)*. Lecture Notes in Computer Science. Vol. 11730. Cham: Springer, 2019. PP. 497–512. DOI:10.1007/978-3-030-30490-4\_40
7. Ateeq M., Ishmanov F., Afzal M.K., Naeem M. Predicting Delay in IoT Using Deep Learning: A Multiparametric Approach // *IEEE Access*. 2019. Vol. 7. PP. 62022–62032. DOI:10.1109/ACCESS.2019.2915958
8. Abdellah A.R., Koucheryavy A. Deep Learning with Long Short-Term Memory for IoT Traffic Prediction // *Proceedings of the 20th International Conference on Next Generation Networks and Systems, NEW2AN 2020, and 13th Conference on Internet of Things, Smart Spaces, ruSMART 2020 (St. Petersburg, Russia, 26–28 August 2020)*. Lecture Notes in Computer Science. Vol. 12525. Cham: Springer, 2020. PP. 267–280. DOI:10.1007/978-3-030-65726-0\_24
9. Almutairi M.S. Deep Learning-Based Solutions for 5G Network and 5G-Enabled Internet of Vehicles: Advances, Meta-Data Analysis, and Future Direction // *Mathematical Problems in Engineering*. 2022. Vol. 2022. P. 6855435. DOI:10.1155/2022/6855435
10. Kaur J., Khan M.A., Iftikhar M., Imran M., Haq Q.E.U. Machine Learning Techniques for 5G and Beyond // *IEEE Access*. 2021. Vol. 9. PP. 23472–23488. DOI:10.1109/ACCESS.2021.3051557
11. Abdellah A.R., Volkov A., Muthanna A., Gallyamov D., Koucheryavy A. Deep Learning for IoT Traffic Prediction Based on Edge Computing // *Proceedings of the 23rd International Conference on Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN 2020, Moscow, Russia, 14–18 September 2020)*. Com-

munications in Computer and Information Science. Vol. 1337. PP. 18–29. Cham: Springer, 2020. DOI:10.1007/978-3-030-66242-4\_2

12. Khedkar S.P., Canessane R.A., Najafi M.L. Prediction of Traffic Generated by IoT Devices Using Statistical Learning Time Series Algorithms // *Wireless Communications and Mobile Computing*. 2021. Vol. 2021. P. 5366222. DOI:10.1155/2021/5366222

13. Selvamanju E., Shalini V.B. Machine Learning Based Mobile Data Traffic Prediction in 5G Cellular Networks // *Proceedings of the 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA, Coimbatore, India, 02–04 December 2021)*. IEEE, 2021. PP. 1318–1324. DOI:10.1109/ICECA52323.2021.9675887

14. Choi D., Yim J., Baek M., Lee S. Machine Learning-Based Vehicle Trajectory Prediction Using V2V Communications and On-Board Sensors // *Electronics*. 2021. Vol. 10. Iss. 4. P. 420. DOI:10.3390/electronics10040420

15. Rasouli A., Tsotsos J.K. Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice // *IEEE Transactions on Intelligent Transportation Systems*. 2020. Vol. 21. Iss. 3. PP. 900–918. DOI:10.1109/TITS.2019.2901817

16. 19.Badicu A., Suciuc G., Balanescu M., Dobrea M., Birdici A., Orza O., Pasat A. PMs concentration forecasting using ARIMA algorithm // *Proceedings of the 91-st Vehicular Technology Conference (VTC2020-Spring, Antwerp, Belgium, 25–28 May 2020)*. IEEE, 2020. DOI:10.1109/VTC2020-Spring48590.2020.9129390

17. Shang Q., Tan D., Gao S., Feng L. A Hybrid Method for Traffic Incident Duration Prediction Using BOA-Optimized Random Forest Combined with Neighborhood Components Analysis // *Journal of Advanced Transportation*. 2019. Vol. 2019. P. 4202735. DOI:10.1155/2019/4202735

18. Кучерявый А.Е. Сети связи с ультрамалыми задержками // *Труды научно-исследовательского института радио*. 2019. № 1. С. 69–74.

19. Мутханна А.С. Интеллектуальная распределенная архитектура сети связи для поддержки беспилотных автомобилей // *Электросвязь*. 2020. № 7. С. 29–34. DOI:10.34832/ELSV.2020.8.7.004

20. Владыко А.Г., Мутханна А.С., Кучерявый А.Е. Метод выгрузки трафика в V2X/5G сетях на основе системы граничных вычислений // *Электросвязь*. 2020. № 8. С. 24–30. DOI:10.34832/ELSV.2020.9.8.004

21. Shrestha R., Vajracharya R., Nam S.Y. Challenges of Future VANET and Cloud-Based Approaches // *Wireless Communications and Mobile Computing*. 2018. Vol. 2018. P. 5603518. DOI:10.1155/2018/5603518

22. Плотников П.В., Владыко А.Г. Минимизация задержек при взаимодействии граничных устройств с использованием кластеризации в сетях VANETs // *Труды учебных заведений связи*. 2022. Т. 8. № 2. С. 6–13. DOI:10.31854/1813-324X-2022-8-2-6-13

## References

1. Abdellah A.R., Alshahrani A., Muthanna A., Koucheryavy A. Performance Estimation in V2X Networks Using Deep Learning-Based M-Estimator Loss Functions in the Presence of Outliers. *Symmetry*. 2021;13(11);2207. DOI:10.3390/sym13112207

2. Chochliouros I.P., Spiliopoulou A.S., Lazaridis P., Dardamanis A., Zaharis Z., Kostopoulos A. Dynamic Network Slicing: Challenges and Opportunities. *Proceedings of the International Conference on Artificial Intelligence Applications and Innovations, AIAI 2020, 5–7 June 2020, Neos Marmaras, Greece. IFIP WG 12.5 International Workshops. IFIP Advances in Information and Communication Technology, vol.585*. Cham: Springer; 2020. DOI:10.1007/978-3-030-49190-1\_5

3. Miglani A., Kumar N. Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges. *Vehicular Communications*. 2019;20:100184. DOI:10.1016/j.vehcom.2019.100184

4. Gillani M., Niaz H.A., Farooq M.U., Ullah A. Data collection protocols for VANETs: a survey. *Complex & Intelligent Systems*. 2022;8(3):2593–2622. DOI:10.1007/s40747-021-00629-x

5. Grigorescu S., Trasnea B., Cocias T., Macesanu G. A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*. 2020;37(3):362–386. DOI:10.1002/rob.21918

6. Pfülb B., Hardegen C., Gepperth A., Rieger S. A Study of Deep Learning for Network Traffic Data Forecasting. *Proceedings of the 28th International Conference on Artificial Neural Networks and Machine Learning, ICANN 2019, 17–19 September 2019, Munich, Germany. Lecture Notes in Computer Science, vol.11730*. Cham: Springer; 2019. p.497–512. DOI:10.1007/978-3-030-30490-4\_40

7. Ateeq M., Ishmanov F., Afzal M.K., Naeem M. Predicting Delay in IoT Using Deep Learning: A Multiparametric Approach. *IEEE Access*. 2019;7:62022–62032. DOI:10.1109/ACCESS.2019.2915958

8. Abdellah A.R., Koucheryavy A. Deep Learning with Long Short-Term Memory for IoT Traffic Prediction. *Proceedings of the 20th International Conference on Next Generation Networks and Systems, NEW2AN 2020, and 13th Conference on Internet of Things, Smart Spaces, ruSMART 2020, 26–28 August 2020, St. Petersburg, Russia. Lecture Notes in Computer Science, vol.12525*. Cham: Springer; 2020. p.267–280. DOI:10.1007/978-3-030-65726-0\_24

9. Almutairi M.S. Deep Learning-Based Solutions for 5G Network and 5G-Enabled Internet of Vehicles: Advances, Meta-Data Analysis, and Future Direction. *Mathematical Problems in Engineering*. 2022;2022:6855435. DOI:10.1155/2022/6855435

10. Kaur J., Khan M.A., Iftikhar M., Imran M., Haq Q.E.U. Machine Learning Techniques for 5G and Beyond. *IEEE Access*. 2021;9:23472–23488. DOI:10.1109/ACCESS.2021.3051557

11. Abdellah A.R., Volkov A., Muthanna A., Gallyamov D., Koucheryavy A. Deep Learning for IoT Traffic Prediction Based on Edge Computing. *Proceedings of the 23rd International Conference on Distributed Computer and Communication Networks: Control, Computation, Communications, DCCN 2020, 14–18 September 2020, Moscow, Russia. Communications in Computer and Information Science, vol.1337*. p.18–29. Cham: Springer, 2020. DOI:10.1007/978-3-030-66242-4\_2

12. Khedkar S.P., Canessane R.A., Najafi M.L. Prediction of Traffic Generated by IoT Devices Using Statistical Learning Time Series Algorithms. *Wireless Communications and Mobile Computing*. 2021;2021:5366222. DOI:10.1155/2021/5366222


13. Selvamanju E., Shalini V.B. Machine Learning Based Mobile Data Traffic Prediction in 5G Cellular Networks. *Proceedings of the 5th International Conference on Electronics, Communication and Aerospace Technology, ICECA, 02–04 December 2021, Coimbatore, India*. IEEE; 2021. p.1318–1324. DOI:10.1109/ICECA52323.2021.9675887
14. Choi D., Yim J., Baek M., Lee S. Machine Learning-Based Vehicle Trajectory Prediction Using V2V Communications and On-Board Sensors. *Electronics*. 2021;10(4):420. DOI:10.3390/electronics10040420
15. Rasouli A., Tsotsos J.K. Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice. *IEEE Transactions on Intelligent Transportation Systems*. 2020;21(3):900–918. DOI:10.1109/TITS.2019.2901817
16. Badicu A., Suci G., Balanescu M., Dobrea M., Birdici A., Orza O., Pasat A. PMs concentration forecasting using ARIMA algorithm. *Proceedings of the 91-st Vehicular Technology Conference, VTC2020-Spring, 25–28 May 2020, Antwerp, Belgium*. IEEE; 2020. DOI:10.1109/VTC2020-Spring48590.2020.9129390
17. Shang Q., Tan D., Gao S., Feng L. A Hybrid Method for Traffic Incident Duration Prediction Using BOA-Optimized Random Forest Combined with Neighborhood Components Analysis. *Journal of Advanced Transportation*. 2019;2019:4202735. DOI:10.1155/2019/4202735
18. Kucheryavy A. Network Delay Communications Networks. *Trudy NIIR*. 2019;1:69–74. (in Russ.)
19. Muthanna A.S. Distributed Intelligent Communication Network Architecture for Unmanned Vehicles. *Electrosvyaz*. 2020;7:29–34. (in Russ.) DOI: 10.34 832/ELSV.2020.8.7.004
20. Vladyko A.G., Muthanna A.S., Kucheryavy A.Eu. Traffic Offloading Method for V2X/5G Networks Based on the Edge Computing System. *Electrosvyaz*. 2020;8:24–30. (in Russ.) DOI: 10.34 832/ELSV.2020.9.8.004
22. Shrestha R., Bajracharya R., Nam S.Y. Challenges of Future VANET and Cloud-Based Approaches. *Wireless Communications and Mobile Computing*. 2018;2018:5603518. DOI:10.1155/2018/5603518
23. Plotnikov P., Vladyko A. Minimizing Delays in the Interaction of Edge Devices Using Clustering in VANETs. *Proceedings of Telecommunication Universities*. 2022;8(2):6–13. (in Russ.) DOI:10.31854/1813-324X-2022-8-2-6-13

Статья поступила в редакцию 24.11.2022; одобрена после рецензирования 20.12.2022; принята к публикации 21.12.2022.

The article was submitted 24.11.2022; approved after reviewing 20.12.2022; accepted for publication 21.12.2022.

## Информация об авторе:

**АЛЬ-СВЕЙТИ  
Малик**

аспирант кафедры сети связи и передачи данных Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича  
 <https://orcid.org/0000-0002-6267-4727>