

# Обзор методов искусственного интеллекта для прогнозирования трафика в сетях мобильной СВЯЗИ

Д.В. Кузьмин

Научный руководитель: А.В. Осовский

*Астраханский государственный технический университет, Астрахань, Россия*

**Аннотация** – В статье анализируются современные подходы к прогнозированию трафика в беспроводных сетях связи с использованием методов искусственного интеллекта. Рассмотрены статистические методы и традиционные алгоритмы машинного обучения, включая авторегрессионную интегрированную модель скользящего среднего (ARIMA) с её усовершенствованными вариантами. Среди нейросетевых архитектур представлены рекуррентные нейронные сети (RNN), сети с долгой краткосрочной памятью (LSTM), рекуррентные блоки с затворами (GRU), свёрточные нейронные сети (CNN), глубокая модель N-BEATS для прогнозирования временных рядов, а также генеративная предобученная модель TimeGPT. Особое внимание уделено комбинированным решениям, интегрирующим выбор признаков и оптимизацию гиперпараметров. Проанализированы аспекты вычислительной нагрузки, энергоэффективности и адаптации моделей к среде периферийных вычислений (edge computing). Цель исследования – систематизировать современные тенденции для практического внедрения методик прогнозирования в телекоммуникационную инфраструктуру.

**Ключевые слова** – прогнозирование трафика сотовой связи, искусственный интеллект, ИИ, глубокое обучение, рекуррентные нейронные сети, LSTM, GRU, гибридные модели, TimeGPT, N-BEATS, системная инженерия нейросетевых архитектур, временные ряды, ARIMA, AutoML, графовые нейронные сети, мобильные сети 5G, IoT-трафик.

## ВВЕДЕНИЕ

Переход на технологии 5G и начало проектирования сетей 6G сопровождаются экспоненциальным ростом объёмов пользовательского трафика. По данным аналитического отчёта Cisco [1], уже к 2025 году ежемесячный объём данных, передаваемых мобильными пользователями, превысит 100 эксабайт (ЭБ). Основные причины этого роста – массовое внедрение «Интернета вещей» (IoT), увеличение числа

подключённых устройств и ужесточение требований к качеству обслуживания (QoS).

Оптимальное использование сетевых ресурсов возможно при наличии точных прогнозов трафика. Традиционные статистические методы, к примеру, модели авторегрессии скользящего среднего интегрированного ряда (ARIMA) и её варианты, не всегда эффективны при обработке нелинейных и нестационарных временных рядов, часто встречающихся в телекоммуникационных системах [2–6].

В последние несколько лет стало ясно, что методы глубокого обучения, основанные на использовании нейросетей, показывают хорошие результаты при работе с временными рядами [7–10].

Данная статья посвящена анализу современных методов прогнозирования сетевого трафика с применением ИИ. Основные задачи исследования: 1) классификация существующих подходов; 2) выявление их сильных и слабых сторон; 3) определение перспективных направлений для внедрения в телекоммуникационные сети.

## I. ИСТОРИЧЕСКИЙ ОБЗОР: ОТ СТАТИСТИКИ К НЕЙРОСЕТЯМ

История прогнозирования сетевого трафика (Рис. 1) началась в середине XX века с развитием теории массового обслуживания [11]. В 1970–1980-е годы главным направлением стали авторегрессионные методы, в числе которых выделяется модель ARIMA, разработанная Джорджем Боксом и Гвилимом Дженкинсом [12]. Её популярность объяснялась понятной интерпретацией и приемлемой вычислительной сложностью. Сложность сетевых конструкций в 1990–2000-е годы и поведение пользователей усложнились, что вызвало потребность в переходе на нелинейные методы. В то время получили развитие методы машинного обучения, в числе которых стоит отметить метод опорных векторов

(SVM) и алгоритмы деревьев решений [13]. Значительный прогресс в обработке временных рядов связан с использованием рекуррентных нейронных сетей, в частности с архитектурой LSTM, введённой Хохрайтером и Шмидхубером в 1997 году [14]. Это исследование, до сих пор лежащее в основе современных разработок [15, 16], заложило основы эффективного моделирования временных зависимостей. К 2010-м годам акцент сместился на гибридные решения, сочетающие статистические методы и нейросетевые технологии [17]. Параллельно развивались системы автоматической оптимизации гиперпараметров (AutoML), сокращающие потребность в ручной настройке [18].

В настоящее время (2020–2025 гг.) лидирующими являются трансформерные архитектуры, графовые нейросети (GNN) для анализа пространственно-временных корреляций и генеративные модели для решения проблемы дефицита данных [19–21].

Наиболее важными критериями стали энергоэффективность и интерпретируемость моделей, что определяет их практическую ценность [22].

Эволюция методов прогнозирования трафика наглядно представлена на Рис. 1, который иллюстрирует переход от классических статистических методов к современным гибридным нейросетевым архитектурам.

## II. КЛАССИФИКАЦИЯ МЕТОДОВ ПРОГНОЗИРОВАНИЯ ТРАФИКА В СОТОВЫХ СЕТЯХ

Современные подходы к прогнозированию трафика могут быть классифицированы по нескольким критериям (Табл. I).

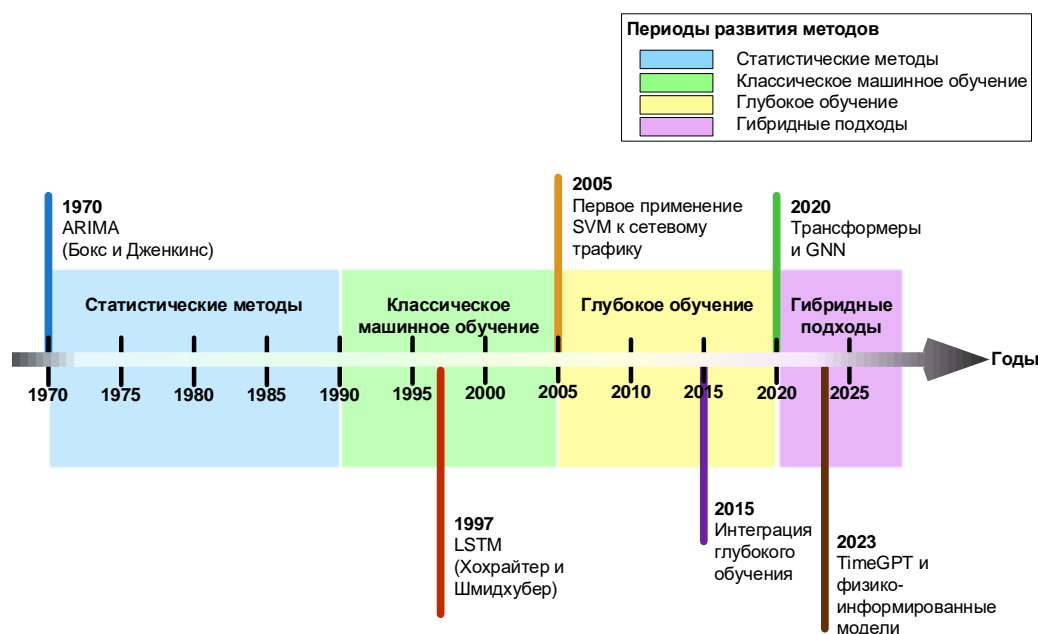


Рис. 1. Эволюция методов прогнозирования трафика в сотовых сетях

ТАБЛИЦА I  
КЛАССИФИКАЦИЯ МЕТОДОВ ПРОГНОЗИРОВАНИЯ ТРАФИКА

Класс метода	Представители	Преимущества	Ограничения	Область применения
Статистические	ARIMA, SARIMA, экспоненциальное сглаживание	Интерпретируемость, низкая вычислительная сложность	Требуют стационарности, плохо моделируют нелинейные зависимости	Краткосрочное прогнозирование в стабильных сетях
Машинное обучение	SVM, Random Forest, Gradient Boosting	Способность к обработке нелинейных зависимостей	Не учитывают последовательную природу временных рядов	Прогнозирование с использованием внешних признаков
Глубокое обучение	LSTM, GRU, CNN, Transformers	Автоматическое извлечение признаков, высокая точность	Высокая вычислительная сложность, требуют больших объемов данных	Краткосрочное и среднесрочное прогнозирование в динамичных сетях
Гибридные подходы	Комбинации статистических и нейросетевых моделей	Сочетание точности и устойчивости	Сложность разработки и настройки	Операторские сети с мультисервисной нагрузкой

### III. СИСТЕМНАЯ ИНЖЕНЕРИЯ МЕТОДИК ПРОГНОЗИРОВАНИЯ

Разработка эффективных методик прогнозирования трафика требует системного подхода, охватывающего все этапы от сбора данных до внедрения модели в эксплуатацию. Выделяются следующие ключевые аспекты.

#### A. Предварительная обработка данных

Надёжность и точность прогноза напрямую зависят от того, насколько качественно подготовлены исходные данные. В рамках предобработки решаются три основные задачи. Первая – обнаружение аномалий (выбросов): используются гибридные схемы, где статистические методы вроде Z-score и IQR дополняются алгоритмами машинного обучения [23,24]. Вторая – заполнение пропусков: эффективно работают генеративно-сопоставительные сети (GAN) и LSTM, которые выполняют задачу восполнения пропущенных значений во временных последовательностях [25]. Третья – нормализация: для устойчивого обучения нейронных сетей применяют масштабирование, адаптированное к специфике сетевого трафика [26].

#### B. Отбор признаков

Для снижения размерности признакового пространства применяются несколько подходов. Информационные критерии (взаимная информация, энтропия) позволяют отбирать наиболее значимые признаки [27]. Нейросетевые автоэнкодеры эффективно уменьшают размерность, сохраняя информативность данных [28]. Дополнительно используется мультикорреляционный анализ, который учитывает не только связь признаков с целевой переменной, но и взаимозависимости между самими признаками [29].

#### C. Выбор архитектуры и настройка гиперпараметров

Правильный выбор архитектуры модели и её гиперпараметров критически важен для достижения оптимального баланса между точностью и вычислительной сложностью:

- **Горизонт прогноза:** один из ключевых факторов при выборе архитектуры. Для краткосрочного прогнозирования (до 6 часов) достаточно лёгких рекуррентных сетей (LSTM или GRU) – они дают хорошую точность при умеренных вычислительных затратах. Для долгосрочного прогноза (сутки и более) лучше подходят трансформеры с механизмами внимания или гибридные модели, способные улавливать отдалённые временные зависимости [30].

- **Пространственно-временное моделирование:** для 5G/6G особое внимание уделяется методам,

учитывающим взаимодействие между ячейками сети с помощью графовых нейронных сетей [31].

- **Оптимизация параметров:** ставится акцент на автоматизированные системы (AutoML, байесовская оптимизация) для настройки гиперпараметров моделей [32].



Рис. 2. Системная архитектура методики прогнозирования трафика

Комплексный подход к разработке методик прогнозирования иллюстрируется на Рис. 2, где представлена системная архитектура, охватывающая все этапы жизненного цикла модели.

#### IV. ТИПОЛОГИЯ ТРАФИКА И СПЕЦИФИКА ПРОГНОЗИРОВАНИЯ

Одной из ключевых проблем в разработке универсальных методик прогнозирования является гетерогенность современного сетевого трафика. Можно

выделить несколько основных типов трафика, каждый из которых обладает уникальными характеристиками и предъявляет специфические требования к методам прогнозирования. На Рис. 3 представлена визуализация распределения типов трафика по ключевым параметрам.

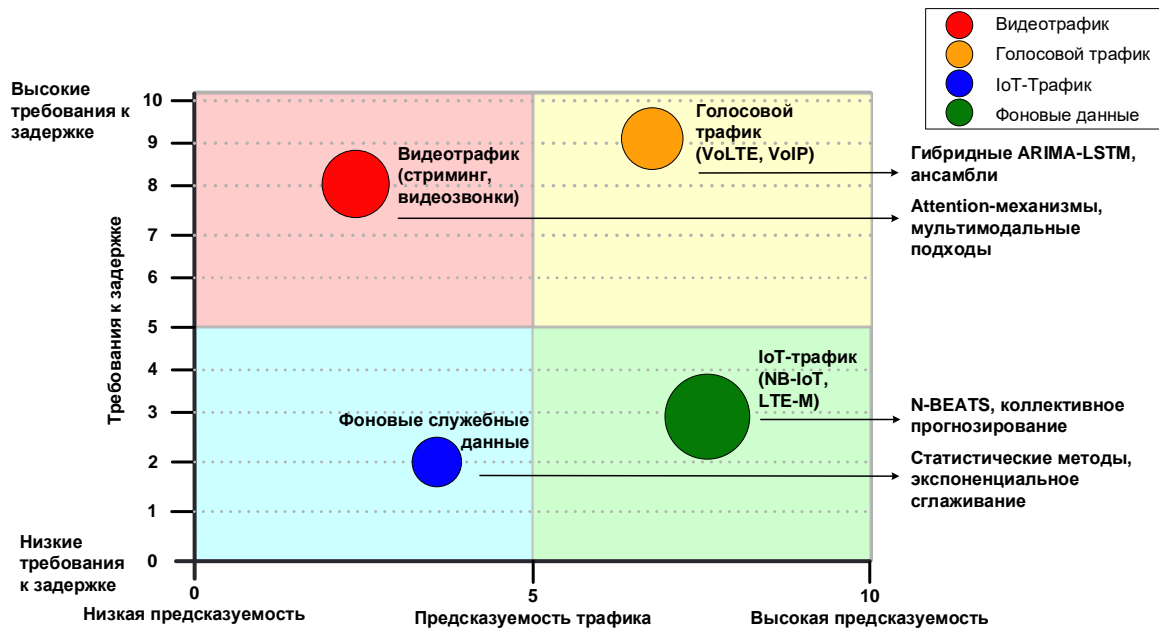


Рис. 3. Характеристики различных типов сетевого трафика

##### A. Видеотрафик (стриминг, видеозвонки)

Видеотрафик отличается высокой интенсивностью, нелинейными паттернами активности и значительной зависимостью от внешних факторов, таких как время суток и происходящие события. Для прогнозирования этого типа трафика наиболее эффективными являются:

- *Механизмы внимания (attention mechanisms)*: в моделях LSTM и трансформеров, которые позволяют выделять наиболее значимые временные интервалы [33].
- *Мультимодальные подходы*, объединяющие данные о календарных событиях и активности в социальных сетях для повышения точности прогнозов [34].

##### B. IoT трафик

Трафик от устройств IoT имеет свои особенности: низкую интенсивность, высокую периодичность и предсказуемость. Для его прогнозирования предлагаются:

- *Модель N-BEATS* с адаптированной архитектурой, специально разработанной для обработки периодических сигналов [35].
- *Коллективные методы прогнозирования*, которые позволяют обрабатывать данные от множества

устройств одновременно с учётом их взаимосвязей [36].

##### C. Голосовой трафик и служебные данные

Голосовой трафик и служебные данные, такие как управление сетью и синхронизация, характеризуются стабильной интенсивностью и высокими требованиями к надежности прогнозов. Для этих типов трафика предпочтительны:

- *Гибридные статистико-нейросетевые модели*, сочетающие интерпретируемость ARIMA и точность LSTM [37].
- *Ансамблевые подходы*, повышающие устойчивость прогноза за счёт комбинирования результатов нескольких моделей [38].

#### V. ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ РАЗВИТИЯ МЕТОДИК ПРОГНОЗИРОВАНИЯ ТРАФИКА

Анализ современных исследований выявляет несколько ключевых проблем и перспективных направлений развития методик прогнозирования трафика в сотовых сетях. На Рис.4 представлена визуализация интеграции этих направлений в единую методологию.

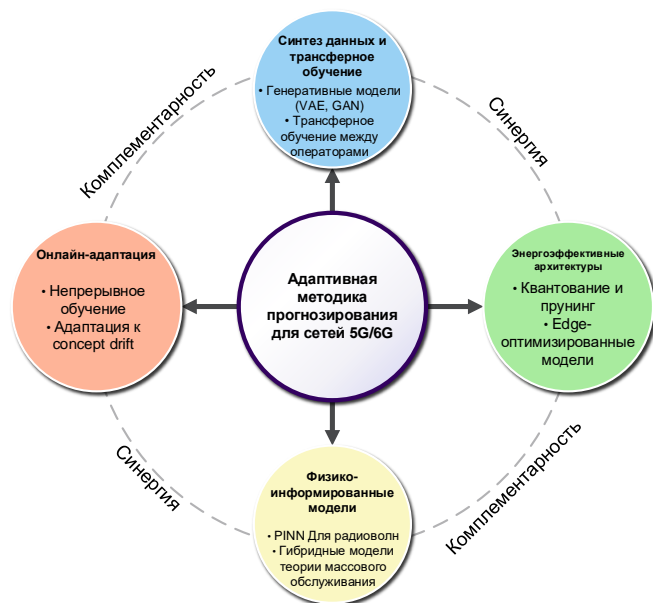


Рис. 4. Перспективные направления развития методик прогнозирования

#### А. Ограниченность данных

Проблема ограниченного доступа к реальным данным операторов связи, вызванная соображениями безопасности и коммерческой конфиденциальности, требует разработки альтернативных методов. Для преодоления этой проблемы предлагаются следующие подходы:

- **Синтез данных.** Один из путей решения проблемы дефицита реальных данных – их синтез с помощью генеративных моделей: вариационных автоэнкодеров (VAE), генеративно-состязательных сетей (GAN) и диффузионных моделей. Такие методы создают искусственные временные ряды, которые по статистическим свойствам (распределение, автокорреляция, сезонность) практически неотличимы от реального трафика и могут использоваться для обучения моделей [39,40].

- **Transfer learning (Трансферное обучение),** которое позволяет переносить знания, полученные на открытых или синтетических данных, на специфические условия конкретного оператора. Этот подход обеспечивает возможность адаптации моделей к уникальным характеристикам реальных сетей [41].

#### В. Энергоэффективность и вычислительная сложность

Требования к энергоэффективности и низкой задержке критичны для развертывания моделей прогнозирования в условиях edge-вычислений. Для решения этих задач проводятся исследования в следующих направлениях:

- **Квантование и оптимизация нейронных сетей.** Для работы на периферийных устройствах (edge computing) критичны энергопотребление и задержки.

Эффективный приём – квантование нейронных сетей (переход от 32-битных чисел к 8-битным или даже бинарным весам). Это резко снижает объём вычислений и памяти при незначительной (1–2%) потере точности. В сочетании с оптимизацией архитектуры такие методы делают модели пригодными для развертывания прямо на базовых станциях или устройствах пользователей [42].

- **Архитектурные оптимизации,** включая использование разделенных свёрток и рекуррентных блоков с адаптивной глубиной, направлены на улучшение производительности моделей. Такие подходы способствуют созданию более эффективных и экономичных решений [43].

#### С. Физико-информированные модели

Интеграция физических законов и доменных знаний в архитектуру нейронных сетей открывает новые перспективы для повышения точности и интерпретируемости прогнозов. В этом контексте выделяются:

- **Перспективное направление – физико-информированные нейронные сети (PINN).** В отличие от data-driven («управляемые данными») / «основанные исключительно на данных») моделей, PINN включают в функцию потерь не только ошибку на данных, но и остатки дифференциальных уравнений (например, описывающих затухание радиоволн или распределение нагрузки в соте). Благодаря этому модель лучше обобщает на нетипичные ситуации и требует меньше обучающих данных. Для сотовых сетей PINN позволяют прогнозировать трафик с учётом реального расположения базовых станций, рельефа и помех [44].

- **Гибридные модели,** сочетающие машинное обучение с традиционными методами теории массового обслуживания. Такой подход обеспечивает баланс между гибкостью машинного обучения и структурированностью традиционных методов, что способствует созданию более надежных и точных моделей [45].

## ВЫВОДЫ И ЗАКЛЮЧЕНИЕ

Проведённый систематический анализ исследований в области прогнозирования трафика [46] в сотовых сетях выявил ключевые тенденции и направления развития. Наблюдается переход от традиционных статистических методов к более сложным гибридным архитектурам, объединяющим преимущества машинного обучения, глубокого обучения и методов обработки сигналов. Эффективность прогнозирования существенно зависит от типа трафика (видеопотоки, данные IoT, голосовые сервисы, служебные сообщения), что требует разработки специализированных методик. Успешное применение методик определяется не только выбором модели, но и качеством предобработки данных, отбором признаков

и оптимизацией гиперпараметров – эти аспекты играют критическую роль. Современные исследования всё чаще акцентируют внимание на энергоэффективности моделей, их вычислительной сложности и возможности работы в условиях edge-вычислений, что особенно важно для мобильных устройств и сетей с ограниченными ресурсами.

Перспективные направления развития включают интеграцию генеративных моделей для решения проблем ограниченности данных, разработку энергоэффективных архитектур для edge-устройств, применение физико-информированных моделей для повышения интерпретируемости и развитие методов онлайн-адаптации к изменениям паттернов трафика.

Дальнейшие исследования следует направить на создание комплексных методик, сочетающих высокую точность с практической применимостью, с учётом специфики стандартов 6G и возможностей квантовых коммуникаций.

## ЛИТЕРАТУРА

- [1] Cisco. Cisco Annual Internet Report (2018–2023) White Paper. San Jose: Cisco Systems, 2020.
- [2] Alawe I., Ksentini A., Hadjadj-Aoul Y., Bertin P. Improving Traffic Forecasting for 5G Core Network Scalability: A Machine Learning Approach // *IEEE Network*. – 2018. – Vol. 32. – No. 6. – P. 42–49.
- [3] Zhang C., Patras P. Long-Term Mobile Traffic Forecasting Using Deep Spatio-Temporal Neural Networks // *Proceedings of the 19th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*. – 2018. – P. 231–240.
- [4] Анализ и прогнозирование трафика современных телекоммуникационных систем на основе методов искусственного интеллекта / Д. В. Кутузов, А. В. Осовский, Д. В. Старов [и др.] // *Вестник Астраханского государственного технического университета. Сер.: Управление, вычислительная техника и информатика*. – 2024. – № 1. – С. 73–87. – DOI 10.24143/2072-9502-2024-1-73-87.
- [5] Свойства трафика видеоконференций 5G и его прогнозирование методами искусственного интеллекта / Д. В. Кутузов, А. В. Осовский, Н. С. Мальцева, С. В. Мартынов // *Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика*. – 2025. – № 1. – С. 103–116. – DOI 10.24143/2072-9502-2025-1-103-116.
- [6] Управление трафиком системы обработки данных на основе NoC для информационно-измерительной сети IoT / Д. В. Кутузов, А. В. Осовский, Д. В. Старов, С. В. Мартынов // *Датчики и системы*. – 2025. – № 5(283). – С. 8–14. – DOI 10.24412/1992-7185-2025-5-8-14.
- [7] Zhang C., Patras P., Haddadi H. Deep Learning in Mobile and Wireless Networking: A Survey // *IEEE Communications Surveys & Tutorials*. – 2019. – Vol. 21. – No. 3. – P. 2224–2287.
- [8] Lopez-Martin M. Novel applications of Machine Learning to Network Traffic Analysis and Prediction: PhD Thesis. – Valladolid: University of Valladolid, 2019.
- [9] Hou Y., Zheng X., Han C., Wei W., Scherer R., Polap D. Deep Learning Methods in Short-Term Traffic Prediction: A Survey // *Information Technology and Control*. – 2022. – Vol. 51. – No. 1. – P. 139–157.
- [10] Trinh H. D., Giupponi L., Dini P. Mobile Traffic Prediction from Raw Data Using LSTM Networks // *IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. – 2018. – P. 1827–1832.
- [11] Zukerman M. Introduction to Queueing Theory and Stochastic Teletraffic Models // *arXiv preprint arXiv:1307.2968*. – 2013.
- [12] Box G., Jenkins G., Reinsel G., Ljung G. *Time Series Analysis: Forecasting and Control*. 5th ed. Hoboken: Wiley, 2015.
- [13] Ferreira G. O., Ravazzi C., Dabbene F., Calafiore G. C., Fiore M. Forecasting Network Traffic: A Survey and Tutorial With Open-Source Comparative Evaluation // *IEEE Access*. – 2023. – Vol. 11. – P. 9691–9723.
- [14] Hochreiter S., Schmidhuber J. Long short-term memory // *Neural Computation*. – 1997. – Vol. 9. – No. 8. – P. 1735–1780.
- [15] Greff K., Srivastava R. K., Koutník J., Steunebrink B. R., Schmidhuber J. LSTM: A search space odyssey // *IEEE Transactions on Neural Networks and Learning Systems*. – 2017. – Vol. 28. – No. 10. – P. 2222–2232.
- [16] Goodfellow I., Bengio Y., Courville A. *Deep Learning*. Cambridge: MIT Press, – 2016.
- [17] Feng J., Chen X., Gao R., Zeng M., Li Y. DeepTP: An End-to-End Neural Network for Mobile Cellular Traffic Prediction // *IEEE Network*. – 2018. – Vol. 32. – No. 6. – P. 108–115.
- [18] He X., Zhao K., Chu X. AutoML: A survey of the state-of-the-art // *Knowledge-Based Systems*. – 2021. – Vol. 212. – Article ID 106622.
- [19] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention is all you need // *Advances in Neural Information Processing Systems (NIPS)*. – 2017. – Vol. 30. – P. 5998–6008.
- [20] Velickovic P., Cucurull G., Casanova A., Romero A., Lio P., Bengio Y. Graph attention networks // *International Conference on Learning Representations (ICLR)*. – 2018.
- [21] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Networks // *Advances in Neural Information Processing Systems (NIPS)*. – 2014. – Vol. 27. – P. 2672–2680.
- [22] Strubell E., Ganesh A., McCallum A. Energy and policy considerations for deep learning in NLP // *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*. – 2019. – P. 3645–3650.
- [23] Wu Y., Dai H. N., Tang H. Graph Neural Networks for Anomaly Detection in Industrial Internet of Things // *IEEE Internet of Things Journal*. – 2022. – Vol. 9. – No. 12. – P. 9214–9231.
- [24] Zhao Z., Chen W., Wu X., Chen P. C. Y., Liu J. LSTM network: a deep learning approach for short-term traffic forecast // *IET Intelligent Transport Systems*. – 2017. – Vol. 11. – No. 2. – P. 68–75.
- [25] Che Z., Purushotham S., Cho K., Sontag D., Liu Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values // *Scientific Reports*. – 2018. – Vol. 8. – Article ID 6085.
- [26] Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift // *ICML*. – 2015. – P. 448–456.
- [27] Chandrashekar G., Sahin F. A survey on feature selection methods // *Computers & Electrical Engineering*. – 2014. – Vol. 40. – No. 1. – P. 16–28.
- [28] Hinton G. E., Salakhutdinov R. R. Reducing the dimensionality of data with neural networks // *Science*. – 2006. – Vol. 313. – №. 5786. – P. 504–507.
- [29] Zhang J., Zheng Y., Qi D. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction // *AAAI*. – 2017. – P. 1655–1661.
- [30] Lai G., Chang W. C., Yang Y., Liu H. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks // *SIGIR*. – 2018. – P. 95–104.
- [31] Мартынов С., Осовский А., Кутузов Д. Стандарты сетей связи 5G/5G-Advanced для цифровой трансформации железных дорог и развития интеллектуальных транспортных систем // *Системная инженерия и инфокоммуникации*. – 2026. – №. 1. – С. 40–51.
- [32] Bergstra J., Bengio Y. Random search for hyper-parameter optimization // *Journal of Machine Learning Research*. – 2012. – Vol. 13. – P. 281–305.
- [33] Стукач О., Ершов И., Кутузов Д. LSTM-модель потребления тепловой энергии в многоквартирном жилом здании // *Системная инженерия и инфокоммуникации*. – 2025. – №. 4. – С. 11–14.
- [34] Zhang J., Zheng Y., Qi D., Li R., Yi X. DNN-Based Prediction Model for Spatio-Temporal Data // *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. – 2016. – Article No. 92.

- [35] Oreshkin B.N., Carpow D., Chapados N., Bengio Y. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting // ICLR. 2020.
- [36] Lim W. Y. B., Luong N. C., Hoang D. T., Jiao Y., Liang Y. C., Yang Q., Niyato D., Miao C. Federated Learning in Mobile Edge Networks: A Comprehensive Survey // IEEE Communications Surveys & Tutorials. – 2020. – Vol. 22. – №. 3. – P. 2031–2063.
- [37] Oliveira J. F. L., Silva E. G., Mattos Neto P. S. G. A Hybrid System Based on Dynamic Selection for Time Series Forecasting // IEEE Transactions on Neural Networks and Learning Systems. – 2022. – Vol. 33. – №. 8. – P. 3251–3263.
- [38] Bandara K., Bergmeir C., Hewamalage H. LSTM-MSNet: leveraging forecasts on sets of related time series with multiple seasonal patterns // IEEE Transactions on Neural Networks and Learning Systems. – 2021. – Vol. 32. – №. 4. – P. 1586–1599.
- [39] Song J., Meng C., Ermon S. Denoising diffusion implicit models // ICLR. 2021.
- [40] Kingma D. P., Welling M. Auto-Encoding Variational Bayes // International Conference on Learning Representations (ICLR). 2014.
- [41] Pan S. J., Yang Q. A Survey on Transfer Learning // IEEE Transactions on Knowledge and Data Engineering. – 2010. – Vol. 22. – №. 10. – P. 1345–1359.
- [42] Jacob B., Kligys S., Chen B., Zhu M., Tang M., Howard A., Adam H., Kalenichenko D. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference // CVPR. – 2018. – P. 2704–2713.
- [43] Hasani R. M., Lechner M., Amini A., Rus D., Grosu R. Liquid time-constant networks // AAAI. – 2021. – Vol. 35. – No. 9. – P. 7657–7666.
- [44] Raissi M., Perdikaris P., Karniadakis G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations // Journal of Computational Physics. – 2019. – Vol. 378. – P. 686–707.
- [45] Fu M., Wang P., Wang Z., Li Z. Deep Learning for Network Traffic Prediction: An Overview // IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC). – 2023. – P. 665–671.
- [46] Осовский, А. В. Анализ и расчёт трафика в телекоммуникационных системах : учебное пособие / А. В. Осовский, Н. С. Мальцева, Д. В. Кутузов ; ФГБОУ ВО «Астраханский государственный технический университет». – Астрахань : Астраханский государственный технический университет, 2022. – 164 с. – ISBN 978-5-89154-739-1.

### Информация об авторах

Кузьмин Дмитрий Васильевич, магистрант направления подготовки «Инфокоммуникационные технологии и системы связи» Астраханского государственного технического университета, г. Астрахань, Россия, [diman\\_k30rus@mail.ru](mailto:diman_k30rus@mail.ru)

Осовский Алексей Викторович – к.т.н., доц., доцент кафедры «Связь» Астраханского государственного технического университета, г. Астрахань, Россия, [a\\_osovskiy@mail.ru](mailto:a_osovskiy@mail.ru), ORCID: 0000-0002-3174-9765

### A Review of Artificial Intelligence Methods for Traffic Forecasting in Mobile Networks

D.V. Kuzmin

Academic advisor: A.V. Osovsky

Astrakhan State Technical University, Astrakhan,  
Russia

**Abstract** – The article analyzes modern approaches to traffic forecasting in wireless communication networks using artificial intelligence methods. Statistical methods and traditional machine learning algorithms are reviewed, including the autoregressive integrated moving average (ARIMA) model with its improved variants. Neural network architectures presented include recurrent neural networks (RNN), long short-term memory networks (LSTM), gated recurrent units (GRU), convolutional neural networks (CNN), the deep N-BEATS model for time series forecasting, and the generative pre-trained TimeGPT model. Special attention is paid to combined solutions integrating feature selection and hyperparameter optimization. The aspects of computational load, energy efficiency, and model adaptation to edge computing are analyzed. The purpose of the study is to systematize current trends for the practical implementation of forecasting techniques in telecommunications infrastructure.

**Keywords** – cellular traffic forecasting, artificial intelligence, AI, deep learning, recurrent neural networks, LSTM, GRU, hybrid models, TimeGPT, N-BEATS, system engineering of neural network architectures, time series, ARIMA, AutoML, graph neural networks, 5G mobile networks, IoT traffic.

### REFERENCES

- [1] Cisco. Cisco Annual Internet Report (2018–2023) White Paper. San Jose: Cisco Systems, 2020.
- [2] Alawe I., Ksentini A., Hadjadj-Aoul Y., Bertin P. Improving Traffic Forecasting for 5G Core Network Scalability: A Machine Learning Approach // IEEE Network. 2018. Vol. 32. No. 6. P. 42–49.
- [3] Zhang C., Patras P. Long-Term Mobile Traffic Forecasting Using Deep Spatio-Temporal Neural Networks // Proceedings of the 19th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc). 2018. P. 231–240.
- [4] Analiz i prognozovanie trafika sovremennyh telekommunikacionnyh sistem na osnove metodov iskusstvennogo intelekta / D. V. Kutuzov, A. V. Osovskij, D. V. Starov [i dr.] // Vestnik Astrahanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naya tekhnika i informatika. – 2024. – № 1. – S. 73-87. – DOI 10.24143/2072-9502-2024-1-73-87.
- [5] Svoystva trafika videokonferencij 5G i ego prognozovanie metodami iskusstvennogo intelekta / D. V. Kutuzov, A. V. Osovskij, N. S. Mal'ceva, S. V. Martynov // Vestnik Astrahanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naya tekhnika i informatika. – 2025. – № 1. – S. 103-116. – DOI 10.24143/2072-9502-2025-1-103-116.
- [6] Upravlenie trafikom sistemy obrabotki dannyh na osnove NoC dlya informacionno-izmeritel'noi seti IoT / D. V. Kutuzov, A. V. Osovskij, D. V. Starov, S. V. Martynov // Datchiki i sistemy. – 2025. – № 5(283). – S. 8-14. – DOI 10.24412/1992-7185-2025-5-8-14.
- [7] Zhang C., Patras P., Haddadi H. Deep Learning in Mobile and Wireless Networking: A Survey // IEEE Communications Surveys & Tutorials. 2019. Vol. 21. No. 3. P. 2224–2287.
- [8] Lopez-Martin M. Novel applications of Machine Learning to Network Traffic Analysis and Prediction: PhD Thesis. Valladolid: University of Valladolid, 2019.
- [9] Hou Y., Zheng X., Han C., Wei W., Scherer R., Polap D. Deep Learning Methods in Short-Term Traffic Prediction: A Survey // Information Technology and Control. 2022. Vol. 51. No. 1. P. 139–157.
- [10] Trinh H. D., Giupponi L., Dini P. Mobile Traffic Prediction from Raw Data Using LSTM Networks // IEEE 29th Annual International

- Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). 2018. P. 1827–1832.
- [11] Zukerman M. Introduction to Queueing Theory and Stochastic Teletraffic Models // arXiv preprint arXiv:1307.2968. 2013.
- [12] Box G., Jenkins G., Reinsel G., Ljung G. Time Series Analysis: Forecasting and Control. 5th ed. Hoboken: Wiley, 2015.
- [13] Ferreira G. O., Ravazzi C., Dabbene F., Calafiore G. C., Fiore M. Forecasting Network Traffic: A Survey and Tutorial With Open-Source Comparative Evaluation // IEEE Access. 2023. Vol. 11. P. 9691–9723.
- [14] Hochreiter S., Schmidhuber J. Long short-term memory // Neural Computation. 1997. Vol. 9. No. 8. P. 1735–1780.
- [15] Greff K., Srivastava R. K., Koutník J., Steunebrink B. R., Schmidhuber J. LSTM: A search space odyssey // IEEE Transactions on Neural Networks and Learning Systems. 2017. Vol. 28. No. 10. P. 2222–2232.
- [16] Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge: MIT Press, 2016.
- [17] Feng J., Chen X., Gao R., Zeng M., Li Y. DeepTP: An End-to-End Neural Network for Mobile Cellular Traffic Prediction // IEEE Network. 2018. Vol. 32. No. 6. P. 108–115.
- [18] He X., Zhao K., Chu X. AutoML: A survey of the state-of-the-art // Knowledge-Based Systems. 2021. Vol. 212. Article ID 106622.
- [19] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention is all you need // Advances in Neural Information Processing Systems (NIPS). 2017. Vol. 30. P. 5998–6008.
- [20] Velickovic P., Cucurull G., Casanova A., Romero A., Lio P., Bengio Y. Graph attention networks // International Conference on Learning Representations (ICLR). 2018.
- [21] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Networks // Advances in Neural Information Processing Systems (NIPS). 2014. Vol. 27. P. 2672–2680.
- [22] Strubell E., Ganesh A., McCallum A. Energy and policy considerations for deep learning in NLP // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL). 2019. P. 3645–3650.
- [23] Wu Y., Dai H. N., Tang H. Graph Neural Networks for Anomaly Detection in Industrial Internet of Things // IEEE Internet of Things Journal. 2022. Vol. 9. No. 12. P. 9214–9231.
- [24] Zhao Z., Chen W., Wu X., Chen P. C. Y., Liu J. LSTM network: a deep learning approach for short-term traffic forecast // IET Intelligent Transport Systems. 2017. Vol. 11. No. 2. P. 68–75.
- [25] Che Z., Purushotham S., Cho K., Sontag D., Liu Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values // Scientific Reports. 2018. Vol. 8. Article ID 6085.
- [26] Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift // ICML. 2015. P. 448–456.
- [27] Chandrashekar G., Sahin F. A survey on feature selection methods // Computers & Electrical Engineering. 2014. Vol. 40. No. 1. P. 16–28.
- [28] Hinton G. E., Salakhutdinov R. R. Reducing the dimensionality of data with neural networks // Science. 2006. Vol. 313. No. 5786. P. 504–507.
- [29] Zhang J., Zheng Y., Qi D. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction // AAAI. 2017. P. 1655–1661.
- [30] Lai G., Chang W. C., Yang Y., Liu H. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks // SIGIR. 2018. P. 95–104.
- [31] Martynov S., Osovskij A., Kutuzov D. Standarty setej svyazi 5G/5G-Advanced dlya cifrovoj transformacii zheleznyh dorog i razvitiya intellektual'nyh transportnyh sistem //Sistemnaya inzheneriya i infokommunikacii. – 2026. – №. 1. – S. 40-51.
- [32] Bergstra J., Bengio Y. Random search for hyper-parameter optimization // Journal of Machine Learning Research. 2012. Vol. 13. P. 281–305.
- [33] Stukach O., Ershov I., Kutuzov D. LSTM-model' potrebleniya teplovoj energii v mnogoetazhnom zhilom zdanii //Sistemnaya inzheneriya i infokommunikacii. – 2025. – №. 4. – S. 11-14.
- [34] Zhang J., Zheng Y., Qi D., Li R., Yi X. DNN-Based Prediction Model for Spatio-Temporal Data // Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2016. Article No. 92.
- [35] Oreshkin B. N., Carpow D., Chapados N., Bengio Y. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting // ICLR. 2020.
- [36] Lim W. Y. B., Luong N. C., Hoang D. T., Jiao Y., Liang Y. C., Yang Q., Niyato D., Miao C. Federated Learning in Mobile Edge Networks: A Comprehensive Survey // IEEE Communications Surveys & Tutorials. 2020. Vol. 22. No. 3. P. 2031–2063.
- [37] Oliveira J. F. L., Silva E. G., Mattos Neto P. S. G. A Hybrid System Based on Dynamic Selection for Time Series Forecasting // IEEE Transactions on Neural Networks and Learning Systems. 2022. Vol. 33. No. 8. P. 3251–3263.
- [38] Bandara K., Bergmeir C., Hewamalage H. LSTM-MSNet: leveraging forecasts on sets of related time series with multiple seasonal patterns // IEEE Transactions on Neural Networks and Learning Systems. 2021. Vol. 32. No. 4. P. 1586–1599.
- [39] Song J., Meng C., Ermon S. Denoising diffusion implicit models // ICLR. 2021.
- [40] Kingma D. P., Welling M. Auto-Encoding Variational Bayes // International Conference on Learning Representations (ICLR). 2014.
- [41] Pan S. J., Yang Q. A Survey on Transfer Learning // IEEE Transactions on Knowledge and Data Engineering. 2010. Vol. 22. No. 10. P. 1345–1359.
- [42] Jacob B., Kligys S., Chen B., Zhu M., Tang M., Howard A., Adam H., Kalenichenko D. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference // CVPR. 2018. P. 2704–2713.
- [43] Hasani R. M., Lechner M., Amini A., Rus D., Grosu R. Liquid time-constant networks // AAAI. 2021. Vol. 35. No. 9. P. 7657–7666.
- [44] Raissi M., Perdikaris P., Karniadakis G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations // Journal of Computational Physics. 2019. Vol. 378. P. 686–707.
- [45] Fu M., Wang P., Wang Z., Li Z. Deep Learning for Network Traffic Prediction: An Overview // 2023 IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC). 2023. P. 665–671.
- [46] Osovskij, A. V. Analiz i raschyot trafika v telekommunikacionnyh sistemah : uchebnoe posobie / A. V. Osovskij, N. S. Mal'ceva, D. V. Kutuzov ; FGBOU VO «Astrahanskij gosudarstvennyj tekhnicheskij universitet». – Astrahan' : Astrahanskij gosudarstvennyj tekhnicheskij universitet, 2022. – 164 s. – ISBN 978-5-89154-739-1.

#### Information about the authors

Kuzmin Dmitry Vasilyevich, Master's student in the Infocommunication Technologies and Communication Systems program at Astrakhan State Technical University, Astrakhan, Russia, [diman\\_k30rus@mail.ru](mailto:diman_k30rus@mail.ru)

Osovsky Alexey Viktorovich, Cand. of Techn. Sci., Associate Professor in the Communications Department at Astrakhan State Technical University, Astrakhan, Russia, [a\\_osovskiy@mail.ru](mailto:a_osovskiy@mail.ru), ORCID: 0000-0002-3174-9765