

УДК 614.4

Гречушкина Н.В., Димитриев С.В.

**ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЦИФРОВОЙ
ЭПИДЕМИОЛОГИИ**Рязанский государственный медицинский университет
имени академика И.П. Павлова, г. Рязань

Цифровая эпидемиология – отрасль эпидемиологии, в которой применяются цифровые инструменты обработки данных, представленных в цифровом формате. Технологическую основу этих инструментов в значительной мере составляют технологии искусственного интеллекта, а именно, системы прогнозирования эпидемий на основе искусственных нейронных сетей. В работе рассмотрены особенности применения искусственных нейронных сетей для прогнозирования эпидемий.

Ключевые слова: цифровая эпидемиология, электронная эпидемиология, прогнозирование эпидемий, искусственный интеллект, искусственные нейронные сети, машинное обучение.

Внедрение цифровых технологий в практику эпидемиологии сформировало новое направление – цифровую эпидемиологию, характеризующуюся использованием цифровых данных и инструментов их анализа, изменением существующих методов и средств обработки данных и построения моделей. Анализ успешных кейсов применения систем на основе искусственных нейронных сетей для прогнозирования эпидемий является актуальной задачей эпидемиологии и представляет исследовательский интерес, поскольку дополняет теоретическую основу для расширения практик применения технологий искусственного интеллекта в указанном контексте.

Цель работы

Проанализировать особенности применения технологий искусственного интеллекта в цифровой эпидемиологии.

Материал и методы

Материалами исследования являются научные публикации в отечественных и зарубежных изданиях, отобранные с использованием поисковых систем по ключевым словам «цифровая эпидемиология», «искусственный интеллект», «искусственные нейронные сети», «прогнозирование эпидемий». В работе использованы логико-теоретические методы исследования: анализ, сравнение, группировка.

Результаты и обсуждение

Цифровая (электронная) эпидемиология (Digital Epidemiology, E-Epidemiology), отмечают С. Таркома и его коллеги (S. Tarkomaa et al.), рассматривается исследователями в широком и в узком смыслах [11]. В первом случае цифровая эпидемиология понимается как направление эпидемиологии, технологическую основу которого составляет цифровой инструментарий, а данные, подлежащие обработке и анализу, представлены в цифровом формате [1, 4]. Во втором случае цифровая эпидемиология понимается как направление

эпидемиологии, основанное на аналитике цифровых данных, сгенерированных и полученных с основной целью, отличной от проведения эпидемиологических исследований [11]. И в той, и в другой трактовке цифровая эпидемиология оперирует цифровыми данными, ресурсами, методами и средствами их сбора, накопления, обработки и анализа, что приводит к изменениям характеристик эпидемиологических исследований, таким как:

разработка и применение новых методов и средств раннего выявления эпидемий и предиктивной аналитики данных, на основе технологий искусственного интеллекта, машинного обучения, искусственных нейронных сетей, больших данных;

использование в эпидемиологических исследованиях новых источников данных, таких как электронные медицинские карты, данные удаленного мониторинга и данные, получаемые от носимых и интеллектуальных устройств;

увеличение объема данных, подвергаемых анализу, и усложнение применяемых многофакторных моделей, для построения и обработки которых используются высокопроизводительная вычислительная техника и программное обеспечение;

использование в эпидемиологических исследованиях не только специальным образом отобранных и подготовленных медицинских данных, но и других, поступающих из широкого спектра источников, таких как онлайн ресурсы (базы данных, электронные библиотеки, онлайн издания СМИ, национальные и международные онлайн-системы эпидемиологического мониторинга) и цифровых следов пользователей;

изучение новых аспектов инфекционных заболеваний, хронических состояний, а также поведенческих и социальных явлений на основе анализа цифровых следов пользователей;

проведение распределенных (удаленных) эпидемиологических исследований с использованием средств цифровой коммуникации, а также инструментов для хранения и совместной работы с данными и ресурсами с помощью облачных технологий и технологий интернета вещей [4, 5, 7].

Цифровая эпидемиология, используя расширенный спектр методов и привлекая новые источники данных для проведения исследований, сформировала и собственный стек технологий для работы с ними, который включает в себя облачные технологии, искусственный интеллект, большие данные, интернет вещей, технологии беспроводной связи, технологии семантического связывания данных, технологии Web 2.0 и др. Важное место здесь отводится технологиям искусственного интеллекта, использование которых для решения задач эпидемиологии насчитывает значительное количество кейсов в реальной практике.

Анализ научной литературы [1, 3, 4, 8, 11] показал, что применение систем на основе искусственного интеллекта в цифровой эпидемиологии направлено на решение четырех ключевых задач: распознавание рисков – раннее выявление вспышек распространения

заболеваний, оценка (анализ) рисков – прогнозирование эпидемий, их масштабов, продолжительности, скорости распространения, показателей заболеваемости и смертности и др., предоставление оперативной информации об эпидемиях – оперативный анализ развития ситуации и обеспечение поддержки принятия решений на основе анализа рисков, моделирования и симуляций различных сценариев принятия мер по предотвращению эпидемий и борьбы с ними.

Успешное построение и эффективное функционирование системы прогнозирования эпидемий на основе искусственного интеллекта требует оперативного (предпочтительно – в режиме реального времени) осуществления обработки на естественном языке (NLP) большого количества многомерных данных из открытых источников для раннего выявления вспышек эпидемий, выявления местных и региональных закономерностей в обнаруженных сигналах, построения моделей распространения эпидемий при разных условиях, быстрого выявления дезинформации и сокрытия релевантных данных.

Известными системами сбора и анализа данных об эпидемиях с использованием искусственного интеллекта являются Epidemic Intelligence from Open Sources (EIOS), BlueDot, Metabiota Epidemic Tracker, EPIWATCH и др.

Примерами успешного применения искусственного интеллекта для прогнозирования эпидемий конкретного заболевания являются такие проекты как AIME (Artificial Intelligence in Medical Epidemiology) и ODL (ObservatoriiodaDengue Lab), а также проект BlueDotInsights, которые применяются при прогнозировании вспышек эпидемии тропической лихорадки денге [3, 6, 9]. Денге является оптимальным примером для иллюстрации применения искусственного интеллекта для прогнозирования эпидемий, поскольку имеет региональный, а не планетарный характер (распространена в странах Южной и Юго-Восточной Азии, Африки, Океании и Карибского бассейна и Латинской Америки), демонстрирует тенденцию к расширению региона распространения (в последние годы случаи заболеваний регистрируются и в континентальной Европе (Франция, Хорватия и Португалия)), имеет повторяющийся характер (заболеваемость составляет более 50 млн человек ежегодно).

Технологическую основу указанных проектов составляют искусственные нейронные сети, машинное обучение и анализ данных. Для анализа данных в проекте AIME используется метод опорных векторов (SVM), в проекте ODL – обобщенная аддитивная модель (GAM), BlueDotInsights – модель деревьев регрессии с градиентным усилением (GBRT). Эти методы машинного обучения используются для геопространственно-временного прогнозирования и позволяют с высокой степенью точности предсказывать возникновение и сроки региональных вспышек, а также моделировать тенденции изменения данных во времени с учетом задержки между принимаемыми мерами и эффектом от их применения. Это обеспечивает основу для

ранней готовности и реагирования национальных систем здравоохранения и поддержку принятия решений соответствующими органами управления. В моделях геопространственно-временного прогнозирования используются региональные данные о прошлых вспышках, экологические факторы, данные о поездках, социальные факторы, распределение переносчиков и спутниковые метеорологические данные. Это позволяет определить метеорологические, эпидемиологические, геосоциальные и демографические параметры, выступающие предикторами для многофакторного анализа.

Источниками данных для обучения нейросети и составления прогноза являются данные официальной статистики, как размещенные в открытом доступе, так и предоставленные профильными ведомствами и министерствами. Помимо этих данных используются геопространственные, временные, экологические и метеорологические спутниковые данные. Также используются большие массивы неструктурированных данных из открытых источников, таких как социальные сети и новостные сообщения, регулярный сбор которых позволяет выявить ранние сигналы об эпидемиях до их обнаружения органами здравоохранения. Так, например, указанные проекты использовали такие общедоступные данные, как: публикации в СМИ и сообщения врачей, метеосводки (AIME), посты в социальных сетях (ODL), сведения о перемещениях людей коммерческими рейсами, климатические данные со спутников (BlueDotInsights) и др.

Количество прогностических переменных в моделях составляло от 13 до 20 (AIME: общее количество зарегистрированных случаев денге, количество дней с начала лихорадки, направление и скорость ветра, количество осадков, случаи грозы, плотность населения и др.); ODL: численность населения, ВВП на душу населения, средний индекс развития человеческого потенциала, индекс развития человеческого потенциала в отношении доходов, продолжительности жизни, в области образования, обеспеченность домов персональными компьютерами, охват домов доступом в Интернет, общее число случаев заболевания, общая заболеваемость на 100 000 жителей, общее число случаев заболевания в логарифмической шкале, общее количество постов в Twitter, показатель активности на 100 000 жителей, общее количество постов в Twitter в логарифмической шкале и др.; BlueDotInsights: численность населения, общее число случаев заболевания, количество и направление ежегодных перемещений людей коммерческими рейсами, данные о популяции животных и насекомых, направление и скорость ветра, сведения об осадках и др.).

Точность прогнозов AIME составляет 89%, а дальность прогнозирования – 3 месяца. Точность прогнозов ODL составляет 88% при дальности прогнозирования 2 месяца и 94% при дальности прогнозирования 7 дней. Относительно точности прогнозов BlueDotInsights и

дальности прогнозирования эпидемий лихорадки денге достоверные сведения не удалось получить из открытых источников.

Указанные методы машинного обучения широко применяются для решения различных задач, связанных с геопространственно-временным прогнозированием в эпидемиологии и других областях. Поскольку данные, полученные из открытых научных источников, оказались неполными, был проведен анализ рассмотренных методов машинного обучения с помощью 10 подготовленных наборов данных для обучения и тестирования нейросети (источник – БД UCI) для каждой из задач (классификации и регрессии). Для каждого набора данных были сгенерированы 10 обучающих наборов по 100 примеров каждый и 10 тестовых наборов по 1000 примеров каждый. Показателем производительности стали среднеквадратичная ошибка для регрессии и ошибка классификации для классификации. Для каждой комбинации метода и набора данных были рассчитаны среднее, медиана и межквартильный размах частоты ошибок тестового набора с использованием средних значений результатов перекрестной проверки. Все методы показали высокую производительность, однако метод SVM незначительно уступает по показателям другим рассмотренным методам. Полученные результаты в целом соотносятся с данными научных публикаций [10] и позволяют сделать вывод о том, что выбор метода машинного обучения должен основываться не только на показателях дальности и точности прогнозирования, которые исследователь желает получить, но также и на других факторах, таких как тип решаемой задачи, состав и объем данных для анализа, доступная вычислительная мощность. Тем не менее, окончательное решение о целесообразности применения того или иного метода для решения задач эпидемиологии определяется не возможной точностью алгоритма и параметрами процесса обучения ИНС, а требуемыми показателями дальности и точности прогнозирования [2].

Заключение и выводы

Возможности применения технологий искусственного интеллекта в цифровой эпидемиологии включают в себя раннее выявление эпидемий, анализ рисков возникновения эпидемий, оперативный анализ ситуации во время эпидемии, поддержку принятия решений для определения органами здравоохранения мер по предотвращению распространения эпидемий. Несмотря на значительные преимущества применения систем эпидмониторинга на основе искусственного интеллекта, его внедрение в практику эпидемиологии существенно уступает клинической медицине, где искусственный интеллект применяется как технологическая основа различных медицинских ИТ-решений [1, 4, 8]. Изучение и совершенствование практик применения ИТ-решений на основе искусственного интеллекта может повысить эффективность эпиднадзора и потому является актуальным направлением развития цифровой эпидемиологии. Решение о целесообразности применяемого метода

должно основываться не на его технических показателях (скорость, производительность, точность алгоритма и др.), а на тех параметрах результата, которые необходимо получить (точность и дальность прогнозирования) и на специфике данных, которые подлежат обработке.

ЛИТЕРАТУРА

1. Денисов Н.С. Тренды популяционных исследований: молекулярная и цифровая эпидемиология (обзор) / Н.С. Денисов, Е.М. Каменских, О.С. Федорова // *Современные технологии медицины*. – 2022. – Т. 14. – №4. – С. 60-72.
2. Романюта, М. А. Системы на основе нейронных сетей в медицине / М. А. Романюта, Н. В. Гречушкина // *Наука и образование: актуальные вопросы теории и практики : Материалы III Международной научно-методической конференции, посвященной 50-летию Самарского государственного университета путей сообщения, Самара, 21–22 марта 2023 года*. – Оренбург, 2023. – С. 161-164.
3. Datilo P.M. A Review of Epidemic Forecasting Using Artificial Neural Networks / P.M. Datilo, Z. Ismail, J. Dare // *International Journal of Epidemiologic Research*. – 2019. – Vol. 6(3). – P. 132-143.
4. Engelmann L. Digital epidemiology, deep phenotyping and the enduring fantasy of pathological omniscience / L. Engelmann // *Big Data & Society*. – 2022. – Vol. 9. – No. 1. – Article ID: 205395172110664.
5. Gagalova K.K. What you need to know before implementing a clinical research data warehouse: comparative review of integrated data repositories in health care institutions / K.K. Gagalova, M.A. Leon Elizalde, E. Portales Casamar, M. Görges // *JMIR Form Res*. – 2020. – Vol. 4. – No. 8. – Article: e17687.
6. Ho C.C. Using Public Open Data to Predict Dengue Epidemic: Assessment of Weather Variability, Population Density, and Land Use As Predictor Variables For Dengue Outbreak Prediction using Support Vector Machine / C.C. Ho, T.C. Yee, D.B. Raja // *Indian Journal of Science and Technology*. – 2018. – Vol. 11(4). – P. 1-8.
7. Kuzmenkov A.Y. AMRmap: an interactive web platform for analysis of antimicrobial resistance surveillance data in Russia. / A.Y. Kuzmenkov, I.V. Trushin, A.G. Vinogradova, A.A. Avramenko [et al.] // *Frontiers in Microbiology*. – 2021. – Vol. 12. – Article 620002.
8. MacIntyre C.R. Artificial intelligence in public health: the potential of epidemic early warning systems / C.R. MacIntyre, X. Chen, M. Kunasekaran [et al.] // *Journal of International Medical Research*. – 2023. – Vol. 51(3). – Article 3000605231159335. – P. 1-18.
9. Marques-Toledo C.dA. Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting Dengue at country and city level / C.dA. Marques-Toledo, C.M. Degener, L. Vinhal, G. Coelho [et al.] // *PLoS Neglected Tropical Diseases*. – 2017. – Vol. 11. – Issue 7. – Article e0005729.
10. Meyer D. The support vector machine under test / D.Meyer, F.Leisch, K.Hornik // *Neurocomputing*. – 2003. – Vol. 55. – Iss. 1-2. – P.169-186.
11. Tarkomaa S. Fighting pandemics with digital epidemiology / S. Tarkomaa, S. Alghnamb, M.D. Howell // *EClinicalMedicine*. – 2020. – Vol. 26. – Article 10051.