

**ВИКОРИСТАННЯ ПРИНЦИПІВ СИСТЕМНОГО АНАЛІЗУ
ПРИ РОЗРОБЦІ ІНТЕГРОВАНОГО СЕРЕДОВИЩА
ПРОГНОЗУВАННЯ ЗАБРУДНЕНОСТІ ПОВІТРЯ**

Анотація. У даній роботі прогнозування забрудненості повітря досліджується в контексті менш дослідженого забруднювача повітря, яким є діоксид азоту, з врахуванням впливу на його концентрацію автомобільного трафіку на основі прогнозованих значень, що не розглядається в інших існуючих дослідженнях. Мета роботи полягала в дослідженні впливу використання системного підходу на точність прогнозування забрудненості повітря діоксидом азоту. Проведене експериментальне дослідження вказує на те, що використання підходу, що базується на принципах системного аналізу, дозволяє підвищити точність прогнозування забрудненості повітря діоксидом азоту за рахунок використання в якості вхідних ознак моделі прогнозованих значень автомобільного трафіку замість значень за минулі періоди.

Ключові слова: прогнозування забрудненості повітря, діоксид азоту, концентрація забруднювача, машинне навчання, системний аналіз, автомобільний трафік, довга короткочасна пам'ять, середня абсолютна похибка.

Постановка проблеми. Проблеми екології так чи інакше впливають на життя кожної сучасної людини у світі. До таких проблем зокрема відноситься забрудненість повітря. Кожен з забруднювачів має свій вплив на стан здоров'я людини, при чому критичний вплив може досягатися не тільки, якщо він довгостроковий: людина не обов'язково повинна перебувати протягом всього року в екологічних умовах з великим середньорічним вмістом відповідної речовини.

Забруднення повітря має критичний вплив на здоров'я людей, адже підвищує смертність через респіраторні, серцево-судинні захворювання, ускладнює протікання захворювань і лікування від інфекцій нижніх дихальних шляхів, випадків передчасних пологів, що відповідно є основними причинами смертності або ускладнень у глобальному вираженні [1].

Згідно з рекомендаціями Всесвітньої організації здоров'я [1] вміст у пові-

трі діоксиду азоту, який розглядається в цій роботі, протягом періоду в 24 години має становити 25 мкг/м³, але якщо рівень смертності, що відповідає цьому значенню, встановлювати рівним 100, то у випадку, якщо вміст за такий період досягне 50 мкг/м³, рівень смертності оцінюється у 102, а за досягнення 120 мкг/м³ – 107. Тож коливання вмісту діоксиду азоту навіть у короткостроковому періоді важливо прогнозувати якомога точніше: фактично кожна наступна година може вказувати на виявлення такого тривалого періоду підвищеного вмісту забруднювача, визначаючи необхідність прийняття рішень для зменшення впливу на здоров'я людини.

Аналіз останніх досліджень і публікацій. На даний момент існує ряд робіт, зокрема [2-6], в яких запропоновано рішення щодо прогнозування вмісту забруднювачів у повітрі. У роботах [2-4] використовуються моделі на основі графових згорткових нейронних мереж з гібридним використанням інших архітектур, але для врахування просторових співвідношень, заради чого такі моделі створюються, необхідно мати розгалужену систему станцій вимірювання, яка не завжди наявна. Відповідно в умовах обмеженості даних та ресурсів, що розглядаються в даній роботі, застосовувати такі моделі недоцільно або неможливо. Також слід виокремити, що при цьому описані моделі прогнозування не стосуються всіх забруднювачів одночасно, тобто зазвичай не є універсальними. Зокрема одним з забруднювачів, який менш активно розглядається при розв'язанні задачі прогнозування, є діоксид азоту. Можна виділити роботи [4-6], в яких моделі прогнозування запропоновані саме для діоксиду азоту. У цих роботах застосовуються різні моделі машинного навчання, зокрема в роботі [4] досліджено щоденне прогнозування на основі моделей глибоких нейронних мереж з архітектурами трансформер, довга короткочасна пам'ять або long short-term memory (LSTM), InceptionTime, ResNet, XceptionTime, MiniRocket, у роботі [5] прогнозування відбувається погодинно, а в основі використовуються штучні нейронні мережі, у роботі [6] прогнозування виконується на наступні 6 годин на основі двонаправлених згорткових LSTM. Переважно такі моделі базуються на вхідних ознаках, до яких належать концентрація забруднювача в повітрі в попередні моменти. Додатково у роботах [5-6] використано метеорологічні дані та дані трафіку, проте трафік подається на вхід за минулі години, а не за години, на які відбувається прогнозування, що є недоліком, зважаючи на те, що саме трафік у цей час впливає на забруднення повітря на виході моделі. На усунення цього недоліку (врахування трафіку за наступні години як результату прогнозування) направлено дане дослідження.

Мета роботи полягала в дослідженні впливу використання системного підходу на точність прогнозування забрудненості повітря діоксидом азоту.

Викладення основного матеріалу дослідження. Прогнозування забрудненості повітря діоксидом азоту є задачею прогнозування часового ряду. У класичній постановці цієї задачі необхідно за значеннями концентрації діоксиду азоту в повітрі на станції вимірювання D за минулий період часу виконати прогнозування концентрації діоксиду азоту в повітрі в наступний період часу. У такому випадку вхідні ознаки представлені набором послідовних значень концентрації діоксиду азоту в повітрі протягом H^F годин, які є послідовними. Результатом прогнозування є концентрація діоксиду азоту в повітрі протягом наступних H^{FP} годин. Тобто значення вхідних і вихідних ознак у підсумку формують часовий ряд, представлений значеннями, що відповідають $H^F + H^{FP}$ послідовним годинам. У випадку, коли кількість вхідних ознак відповідає кількості вихідних ознак моделі, тобто часові горизонти співпадають ($H^F = H^{FP}$), виходить, що $H^F + H^{FP} = 2H^F$.

Двигуни внутрішнього згорання є одним з головних джерел діоксиду азоту [7]. Тому логічною є необхідність врахування даного впливу в моделі прогнозування забрудненості повітря діоксидом азоту. Таке врахування зазвичай виконується на основі визначення автомобільного трафіку, який долає певну ділянку, безпосередньо пов'язану зі станцією D . Проте важливим є те, що при прогнозуванні концентрації діоксиду азоту в наступні $H^F + 1, H^F + 2, \dots, H^F + H^F$ години основний вплив на таку концентрацію здійснює обсяг трафіку відповідно в $H^F + 1, H^F + 2, \dots, H^F + H^F$ години, а не в $1, 2, \dots, H^F$ години відповідно. Звичайно, що значення трафіку в $H^F + 1, H^F + 2, \dots, H^F + H^F$ години невідомі в момент прогнозування забрудненості повітря, тому для цього необхідно попередньо виконати прогнозування автомобільного трафіку. При системному розгляді пов'язаних проблем з даною можуть бути створені або вже наявні моделі прогнозування, зокрема прогнозування автомобільного трафіку, адже ця сукупність моделей зокрема відноситься до задач керування, планування діяльності, прийняття рішень у місті. Наявність сукупності таких моделей дозволяє розглядати відповідну інформаційну технологію як цілісну в складі інтегрованого середовища прогнозування забрудненості повітря. Запропоновано для прогнозування значень трафіку використовувати метод прогнозування автомобільного трафіку в умовах обмеженості даних та ресурсів [8].

Для побудови моделей прогнозування забрудненості повітря діоксидом азоту можуть використовуватися різні архітектури, проте в даній роботі розг-

лядається випадок обмеженості даних та ресурсів, тому базова модель побудована на основі LSTM. Такий вибір обумовлений зокрема і тим, що мета роботи пов'язана з дослідженням впливу використання системного підходу на точність прогнозування. Тому достатньо використання LSTM як базової моделі, але дослідження різних варіантів вхідних ознак, пов'язаних з трафіком.

Окрім того випадок обмеженості даних та ресурсів характеризується також і тим, що станції вимірювання трафіку та забрудненості повітря не охоплюють всі потрібні локації, а відповідно можуть не охоплювати і одні ті самі ділянки, тобто на одній ділянці вимірюється забрудненість повітря діоксидом азоту, але не вимірюється автомобільний трафік. Тоді розглянуте вище питання вибору вхідних ознак набуває потреби додаткового врегулювання. Для цього було використано метод вибору вхідних ознак на основі Random Forest. Для цього при створенні кожної моделі прогнозування забрудненості повітря діоксидом азоту для відповідної станції D відбувався вибір максимум 2 станцій вимірювання автомобільного трафіку на основі методу Random Forest з формуванням підмножин B^{SP} . Таким чином визначалися станції B^{SP} , обсяг автомобільного трафіку за якими мав найбільший вплив на забрудненість повітря діоксидом азоту за станцією D . Для визначення такого впливу використовувались або дані автомобільного трафіку за $1, 2, \dots, H^F$ годин для моделей з додатковими ознаками на основі минулих значень, або фактично зафіксовані значення обсягу автомобільного трафіку за $H^F + 1, H^F + 2, \dots, H^F + H^F$ годин.

Для перевірки ефективності запропонованого рішення було виконано експериментальне дослідження, для якого використано дані про автомобільний трафік [9] та забрудненість повітря діоксидом азоту [10] в місті Мадрид (Іспанія), які доступні через Портал відкритих даних Мадридської міської ради [11]. З цих даних виділено записи, які стосуються періоду з 1 січня 2019 року до 30 вересня 2022 року. Ці дані було розподілено на навчальні, валідаційні та тестові у співвідношенні відповідно 60 %, 20 %, 20 %. На основі цих даних було створено і навчено 3 типи моделей для кожної станції вимірювання забрудненості повітря діоксидом азоту:

– моделі прогнозування концентрації діоксиду азоту на наступні 6 годин на основі LSTM з використанням значень концентрації діоксиду азоту за минулі 6 годин (LSTM-base);

– моделі прогнозування концентрації діоксиду азоту на наступні 6 годин на основі LSTM з використанням значень концентрації діоксиду азоту та обсягу автомобільного трафіку за минулі 6 годин за станціями вимірювання, відібрані

ними для кожної моделі окремо (LSTM-traffic);

– моделі прогнозування концентрації діоксиду азоту на наступні 6 годин на основі LSTM з використанням значень концентрації діоксиду азоту за минулі 6 годин та значень обсягу автомобільного трафіку за наступні 6 годин, отриманих шляхом прогнозування за методом [8], за станціями вимірювання з підмножин B^{SP} , відібраними для кожної моделі окремо (LSTM-predicted-traffic).

Результати експериментального дослідження приведені в таблиці 1. Для оцінювання моделей використано показник середньої абсолютної похибки або mean absolute error (MAE).

Отримані результати продемонстрували, що середня абсолютна похибка для моделі без додаткових вхідних ознак (LSTM-base) була покращена на 3 % за використання як додаткових вхідних ознак значень обсягу автомобільного трафіку за минулі 6 годин (LSTM-traffic).

За використання прогнозованих значень автомобільного трафіку (LSTM-predicted-traffic) середня абсолютна похибка була покращена на 5 %, а в найкращому випадку – на 10 %. Порівнюючи використання значень трафіку за минулі години та прогнозованих значень, варто також вказати на те, що окрім зменшення середньої абсолютної похибки відбулось збільшення кількості станцій вимірювання, для яких було покращено значення середньої абсолютної похибки, яке виявилось меншим за базову модель тільки у 2 випадках. Усе це дозволяє стверджувати про доцільність створення інтегрованого середовища прогнозування забрудненості повітря замість побудови окремих незв'язаних моделей. При цьому отриманий результат прогнозування залежить також і від точності прогнозування автомобільного трафіку. Слід розуміти, що в даній постановці задачі розглядається саме випадок обмеженості даних та ресурсів, тому моделі, які використовуються для прогнозування автомобільного трафіку визначаються в умовах, коли складність таких моделей має бути обмеженою, а дані про трафік наявні на обмеженому просторі, що має певний вплив і на точність моделей прогнозування автомобільного трафіку. З іншого боку загальний автомобільний трафік тільки частково визначає забрудненість повітря діоксидом азоту. Відповідно дані експериментальні дослідження вказують на те, що саме в умовах обмеженості даних та ресурсів було досягнуто покращення отриманих результатів прогнозування забрудненості повітря діоксидом азоту за використання принципів системного аналізу та врахування наявних зв'язків між моделями прогнозування забрудненості повітря та автомобільного трафіку.

Результати експериментального дослідження прогнозування забрудненості повітря діоксидом азоту

Номер станції вимірювання	MAE для LSTM-base	MAE для LSTM-traffic	MAE для LSTM-predicted-traffic
1	0,03156	0,03412	0,03642
2	0,03585	0,03535	0,03322
3	0,04641	0,04366	0,04436
4	0,03634	0,03607	0,03747
5	0,06193	0,05997	0,05825
6	0,05002	0,04921	0,0477
7	0,04449	0,04165	0,04231
8	0,05408	0,05167	0,05069
9	0,05137	0,05165	0,04783
10	0,04794	0,04431	0,04452
11	0,04491	0,04368	0,04322
12	0,05863	0,05724	0,05355
13	0,05451	0,05547	0,05379
14	0,06211	0,05703	0,05697
15	0,05283	0,04967	0,04889
16	0,0454	0,04532	0,04488
17	0,05254	0,05334	0,05086
18	0,03527	0,03854	0,0342
19	0,04348	0,0417	0,04087
20	0,04331	0,0421	0,04152
21	0,05669	0,05356	0,05351
22	0,04317	0,03836	0,03886
23	0,05001	0,04672	0,04567
24	0,05213	0,04989	0,04746

Висновки. Проведений аналіз проблеми прогнозування забрудненості повітря діоксидом азоту в контексті останніх проведених досліджень дозволив встановити, що тільки в деяких з запропонованих моделей трафік враховується як вхідні ознаки, але використовуються виключно значення, зафіксовані в минулі періоди часу. Тоді навіть якщо паралельно існують моделі прогнозування

автомобільного трафіку і забрудненості повітря діоксидом азоту, існуючі зв'язки між результатами використання цих моделей не враховуються системно.

Враховуючи, що на забрудненість повітря в певну годину впливає автомобільний трафік не за минулі години, а за поточну, у роботі було досліджено, як впливає використання прогнозованих значень автомобільного трафіку на прогнозування забрудненості повітря діоксидом азоту порівняно з використанням значень трафіку за минулі години. Для цього було створено моделі на основі LSTM тільки на основі концентрації забруднювача за минулі години, з додатковим використанням значень автомобільного трафіку за минулі години та отриманих шляхом прогнозування. Останній з перелічених варіантів моделей представляв системний підхід, оскільки безпосередньо застосовував інші існуючі в середовищі прогнозування моделі. Створені моделі були порівняні в процесі експериментального дослідження на основі даних, зібраних у місті Мадрид.

Використання підходу, що базується на принципах системного аналізу, дозволяє підвищити точність прогнозування забрудненості повітря діоксидом азоту, створюючи для цього інтегроване середовище прогнозування забрудненості повітря, яке об'єднує в собі інформаційні технології прогнозування автомобільного трафіку та прогнозування забрудненості повітря.

ЛІТЕРАТУРА / REFERENCES

1. World Health Organization, WHO global air quality guidelines. Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: World Health Organization, 2021.
2. P. Muthukumar, K. Nagrecha, D. Comer et al., "PM_{2.5} Air Pollution Prediction through Deep Learning Using Multisource Meteorological, Wildfire, and Heat Data", *Atmosphere*, vol. 13, no. 5, article 822, 2022. DOI: 10.3390/atmos13050822.
3. B. Liu, M. Wang M., and H. W. Guesgen, "A hybrid model for spatial-temporal prediction of PM_{2.5} based on a time division method", *International Journal of Environmental Science and Technology*, vol. 20, no. 11, pp. 12195 – 12206, 2023. DOI: 10.1007/s13762-023-04813-z.
4. A. AlShehhi, and R. Welsch, "Artificial intelligence for improving Nitrogen Dioxide forecasting of Abu Dhabi environment agency ground-based stations", *Journal of Big Data*, vol. 10, art. 92, 2023. DOI: 10.1186/s40537-023-00754-z.
5. M. I. Rodriguez-Garcia, M. C. Ribeiro Rodrigues, J. Gonzalez-Enrique, J. J. Ruiz-Aguilar, and I. J. Turias, "Forecasting air pollutants using classification models: a case

study in the Bay of Algeciras (Spain)”, Stochastic Environmental Research and Risk Assessment, vol. 37, pp. 4359-4383, 2023. DOI: 10.1007/s00477-023-02512-2.

6. D. Iskandaryan, F. Ramos, and S. Trilles, “Bidirectional convolutional LSTM for the prediction of nitrogen dioxide in the city of Madrid”, PLoS One, vol. 17, no. 6, article e0269295, 2022. DOI: 10.1371/journal.pone.0269295.

7. European Environment Agency, Explaining road transport emissions: A non-technical guide. Luxembourg: Publications Office of the European Union, 2016. doi: 10.2800/71804.

8. V. M. Lovkin, S. A. Subbotin, and A. O. Oliinyk, “Method for Agent-Oriented Traffic Prediction under Data and Resource Constraints”, Radio Electronics, Computer Science, Control, vol. 4, pp. 99-110, 2023. DOI: 10.15588/1607-3274-2023-4-10.

9. Aforos de trafico en la ciudad de Madrid permanentes – Portal de datos abiertos del Ayuntamiento de Madrid, 2023. [Online]. Available: <https://datos.madrid.es/sites/v/index.jsp?vgnextoid=fabfb3e1de124610VgnVCM2000001f4a900aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD>.

10. Calidad del aire. Datos horarios desde 2001 – Portal de datos abiertos del Ayuntamiento de Madrid, 2023. [Online]. Available: <https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=f3c0f7d512273410VgnVCM2000000c205a0aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD>.

11. En portada – Portal de datos abiertos del Ayuntamiento de Madrid, 2023. [Online]. Available: <https://datos.madrid.es/portal/site/egob>.

Received 03.01.2024.
Accepted 08.01.2024.

Application of the principles of systems analysis for the development of environment for integrated air pollution forecasting

Air pollution forecasting is a well-known scientific problem, but nitrogen dioxide is not often considered as a pollutant in the problem. Traffic, presented by vehicle emissions, is one of the main sources of air pollution by nitrogen dioxide. Some existing investigations take traffic into account as input features for forecasting models but traffic is presented by values measured during previous hours. Therefore, separate models are created and trained for every air pollution measurement station. Trained forecasting models are not used systematically and models for forecasting other indicators are not applied. But air pollution over current hour is caused by traffic mostly at the same place over the same hour but not over previous hours. Therefore, it is important to investigate how usage of forecasted values of traffic impacts on air pollution by nitrogen dioxide in comparison with usage of traffic values measured over previous hours. This task is con-

sidered in the paper.

The objective of the study is to investigate impact of system approach on accuracy of forecasting air pollution by nitrogen dioxide. Three types of forecasting models based on long short-term memory were created to investigate such an impact. These types include models with concentration of nitrogen dioxide over previous hours as input features and two types of models with additional input features presented by values of traffic over previous hours and by forecasted values of traffic over output hours correspondingly. The last type of models presented application of systems approach, as it uses other models available in the integrated forecasting environment. Experimental investigation allowed to create and train models as well as to compare the results of the models on the datasets representing car traffic and air pollution in Madrid. Application of the approach based on the principles of systems analysis increases accuracy of forecasting air pollution by nitrogen dioxide. It enables to create integrated air pollution forecasting environment, which combines information technologies of traffic and air pollution forecasting.

Льовкін Валерій Миколайович – кандидат технічних наук, доцент, доцент кафедри програмних засобів, Національний університет «Запорізька політехніка», Запоріжжя, Україна.

Lovkin Valerii – Ph.D., Associate Professor, Associate Professor of the Department of Software Tools, National University «Zaporizhzhia Polytechnic», Zaporizhzhia, Ukraine.