

DOI: [https://doi.org/10.25140/2411-5363-2026-1\(43\)-45-57](https://doi.org/10.25140/2411-5363-2026-1(43)-45-57)

УДК 004.852:665.63

Сергій Михайлович Ткачук¹, Вадим Валерійович Бондар²

¹аспірант кафедри технічних та програмних засобів автоматизації
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (Київ, Україна)

E-mail: tkachuksergiy8@gmail.com. ORCID: <https://orcid.org/0009-0002-4013-8875>

²аспірант кафедри технічних та програмних засобів автоматизації
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (Київ, Україна)

E-mail: v.v.bondar@kpi.ua. ORCID: <https://orcid.org/0009-0000-3320-4550>

ІНТЕЛЕКТУАЛЬНА СИСТЕМА ОЦІНЮВАННЯ ЯКОСТІ НАФТОПРОДУКТІВ НА ОСНОВІ АНСАМБЛЕВИХ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Розроблено інтелектуальну систему оцінювання якості продукції каталітичного риформінгу в реальному часі на основі ансамблевих методів машинного навчання. Систему побудовано із застосуванням методу *Random Forest*, що забезпечує прогнозування октанового числа риформату із середньою абсолютною похибкою 0,41 пункту та коефіцієнтом детермінації $R^2 = 0,974$. Для підвищення адекватності порівняння додатково залучено глибоку нейронну мережу *LSTM*. Реалізовано механізм інкрементальної онлайн-адаптації моделі та квантифікації невизначеності прогнозів для детекції аномальних режимів роботи. Інтеграція розробленої системи в контур нелінійного предиктивного управління забезпечує зниження варіації октанового числа на 57 % та скорочення часу перехідних процесів у 2,8 рази порівняно з традиційним *PID*-управлінням.

Ключові слова: *soft*-сенсор; оцінювання якості; *Random Forest*; нейронні мережі; онлайн-адаптація; каталітичний риформінг; предиктивне управління; ансамблеві методи; *LSTM*.

Рис.: 3. Табл.: 5. Бібл.: 12.

Актуальність теми дослідження. Каталітичний риформінг є одним із ключових процесів нафтопереробної промисловості, що забезпечує виробництво високооктанових компонентів автомобільного бензину та водню для процесів гідроочищення [1]. В умовах четвертої промислової революції (*Industry 4.0*) цифровізація нафтопереробних виробництв та впровадження концепції цифрових двійників ставлять нові вимоги до оперативності та точності вимірювання якості продукції. Установки каталітичного риформінгу з безперервною регенерацією каталізатора (*Continuous Catalytic Reforming, CCR*) є найбільш прогресивним типом установок, що дозволяють підтримувати високу активність каталізатора та забезпечувати стабільну якість продукції протягом тривалого періоду експлуатації [2].

Ефективне управління якістю продукції на установках *CCR* потребує оперативної інформації про октанове число риформату та вихід рідкого продукту. Проте ці показники визначаються лабораторними методами з періодичністю 2–8 годин, що створює значну затримку в контурі управління [3]. Поточні аналізатори на основі інфрачервоної спектроскопії або хроматографії потребують регулярного калібрування та мають обмежену надійність в умовах промислової експлуатації. Це суттєво обмежує можливості реалізації стратегій оптимального управління в реальному часі та призводить до підвищеної варіабельності якості продукції, перевитрат енергоресурсів та неоптимального використання каталізатора.

Перспективним вирішенням цієї проблеми є застосування *soft*-сенсорів (програмних сенсорів, віртуальних аналізаторів) — математичних моделей, що забезпечують оцінку важковимірюваних показників якості на основі легкодоступних вимірювань технологічних параметрів [4]. На відміну від апаратних аналізаторів, *soft*-сенсори не потребують спеціального обладнання, проте їх точність залежить від якості математичної моделі, а їх деградація з часом внаслідок дрейфу характеристик процесу є серйозною проблемою для промислової експлуатації [5]. За даними досліджень, значна частка впроваджених *soft*-сенсорів втрачає прийнятну точність протягом першого року без регулярного перенавчання, що зумовлює актуальність розробки адаптивних механізмів.

Постановка проблеми. Процес каталітичного риформінгу є складним нелінійним багатовимірним об'єктом управління з багатомасштабною динамікою: постійні часу контурів якості становлять 25–45 хвилин, транспортні затримки — 30–60 хвилин, а динаміка дезактивації каталізатора розвивається протягом сотень годин [2]. Нелінійність процесу обумовлена експоненційною залежністю швидкостей реакцій від температури та складними кінетичними механізмами каталітичних перетворень.

Основна проблема полягає у відсутності комплексного підходу до побудови інтелектуальної системи оцінювання якості, що поєднує високу точність прогнозування з можливістю автоматичної адаптації до змін процесу та оцінкою надійності прогнозів. Існуючі рішення на основі лінійних методів не забезпечують необхідної точності при суттєвих змінах режиму, а нейромережеві soft-сенсори потребують складної процедури перенавчання при дрейфі характеристик.

Аналіз останніх досліджень і публікацій. Проблемі розробки soft-сенсорів для промислових процесів присвячено значну кількість наукових праць. Комплексний огляд soft-сенсорів для моніторингу, управління та оптимізації наведено в [3], де систематизовано основні підходи: від лінійних методів до глибокого навчання. Лінійні методи – метод головних компонент (PCA), проєкція на латентні структури (PLS) – залишаються базовим інструментом завдяки простоті інтерпретації, проте мають обмежену точність для нелінійних процесів. Роль штучного інтелекту в розвитку soft-сенсорів для сталого промислового виробництва проаналізовано в [4], де показано переваги data-driven підходів над фізичними моделями для складних процесів.

Ансамблеві методи — Random Forest (RF), Gradient Boosting (XGBoost, LightGBM) — поєднують прогнози кількох базових моделей для підвищення точності та робастності. Сучасні модифікації RF, зокрема з покращеним відбором ознак та адаптивними стратегіями, продемонстровано в [7]. Підходи на основі графових нейронних мереж (GCN) для побудови soft-сенсорів запропоновано в [6], що дозволяє моделювати складні кореляції між змінними. Гібридні моделі з адаптивними механізмами набувають все більшого поширення: у [8] розроблено глибоку гаусівську змішану адаптивну мережу з механізмом калібрування для компенсації дрейфу процесу.

Проблема адаптації soft-сенсорів до дрейфу характеристик процесу (concept drift) досліджується у [9], де систематизовано методи адаптації в рамках глибокого навчання. Проблему трансферного навчання та переносимості soft-сенсорів між різними установками розглянуто в [5]. Квантифікація невизначеності прогнозів ML-моделей, зокрема через ансамблювання та бассівські підходи, проаналізовано у [10]. Питанням інтеграції ML-моделей з нелінійним предиктивним управлінням для хімічних процесів присвячено роботу [12].

Виділення недосліджених частин загальної проблеми. Проведений аналіз показав, що існуючі дослідження мають такі обмеження. По-перше, більшість soft-сенсорів для процесу риформінгу базуються на лінійних методах (PLS, PCR) або простих нейронних мережах, що не забезпечують достатньої точності при суттєвих змінах режиму або складу сировини. По-друге, механізми онлайн-адаптації здебільшого потребують повного перенавчання моделі. По-третє, відсутній комплексний підхід, що поєднує ансамблеві методи з вбудованою квантифікацією невизначеності та інкрементальним оновленням. По-четверте, недостатньо досліджено питання інтеграції адаптивних soft-сенсорів з системами предиктивного управління для процесу CCR-риформінгу. По-п'яте, порівняння ансамблевих методів із сучасними архітектурами глибокого навчання (LSTM) для задачі оцінювання якості CCR-процесу залишається недостатньо висвітленим у літературі.

Метою статті є розробка інтелектуальної системи оцінювання якості продукції каталітичного риформінгу в реальному часі на основі ансамблевих методів машинного навчання з механізмами онлайн-адаптації та квантифікації невизначеності прогнозів, порівняння з методами глибокого навчання, а також верифікація ефективності розробленої системи в контурі нелінійного предиктивного управління.

Виклад основного матеріалу.

1. Архітектура інформаційної системи оцінювання якості. Задача розробки soft-сенсора формулюється таким чином. Дано: вектор доступних вимірювань $x(t)$, що включає температури, тиски, витрати, показники густини та заломлення потоків. Потрібно: побудувати модель $f(x; \theta)$, що забезпечує оцінку цільової змінної $\hat{y} = f(x; \theta)$ з мінімальною похибкою відносно лабораторних значень y . Вхідний вектор soft-сенсора для оцінки октанового числа включає 10 параметрів:

$$x = [T_1, T_2, T_3, T_4, T_{out}^t, P, F_{sir}, \rho_{sir}, \rho_{r1}^f, n^D]^T \tag{1}$$

де T_1 – T_4 – температури входу в реактори, °С; T_{out}^t – температура виходу з останнього реактора; P – тиск, МПа; F_{sir} – витрата сировини, м³/год; ρ_{sir}, ρ_{r1}^f – густини сировини та риформату, кг/м³; n^D – показник заломлення риформату.

Вимоги до soft-сенсора визначаються потребами системи управління: точність прогнозування RON не гірше $\pm 1,0$ пункту; час розрахунку не більше 100 мс для забезпечення роботи в реальному часі; можливість онлайн-адаптації при надходженні нових лабораторних даних; оцінка невизначеності прогнозу для детекції аномальних станів.

Архітектура розробленої інформаційної системи включає чотири функціональні блоки: попередня обробка вхідних даних, ансамблева модель Random Forest, квантифікація невизначеності та детекція аномалій, онлайн-адаптація при зміні характеристик процесу. Система інтегрується з контуром нелінійного предиктивного управління (NMPC), замикаючи контур управління якістю. Структуру системи представлено на рис. 1.

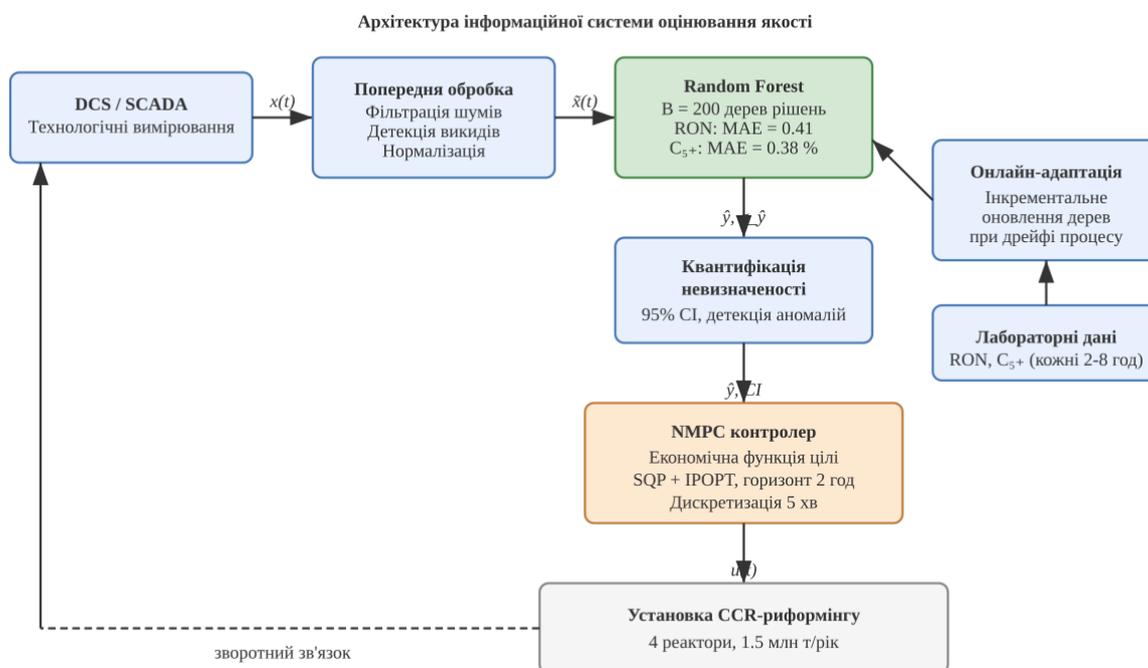


Рис. 1. Архітектура інформаційної системи оцінювання якості
Джерело: розроблено авторами.

2. Побудова soft-сенсорів на основі методу Random Forest. Для побудови soft-сенсора обрано метод випадкового лісу (Random Forest, RF) — ансамблевий метод машинного навчання, що базується на агрегації прогнозів множини дерев рішень [7]. Вибір RF обумовлено: високою точністю без ретельного налаштування; природною оцінкою невизначеності через дисперсію прогнозів; робастністю до шуму; інтерпретованістю через оцінку важливості ознак.

Алгоритм Random Forest будує ансамбль з B дерев рішень, кожне з яких навчається на бутстреп-вибірці з навчальних даних. При побудові кожного вузла розглядається випадкова підмножина ознак розміром $m < p$, де p — загальна кількість ознак. Прогноз ансамблю для задачі регресії визначається як середнє прогнозів окремих дерев:

$$\hat{y} = (1/B) \cdot \sum^{b_{oi}} T_o(x), \tag{2}$$

де $T_o(x)$ — прогноз b -го дерева для входу x ; B — кількість дерев в ансамблі.

Оцінка невизначеності прогнозу обчислюється як стандартне відхилення прогнозів окремих дерев:

$$\sigma_{\hat{y}} = \sqrt{[(1/(B-1)) \cdot \sum^{b_{oi}} (T_o(x) - \hat{y})^2]}, \tag{3}$$

де висока дисперсія прогнозів $\sigma_{\hat{y}}$ сигналізує про ненадійність прогнозу.

Важливість j -ої ознаки оцінюється за зниженням якості моделі при випадковій перестановці значень цієї ознаки (permutation importance):

$$I_j = (1/B) \cdot \sum^{b_{oi}} (MSE^{bperm(j)} - MSE^b) \tag{4}$$

де MSE_o — похибка b -го дерева на OOB-вибірці; $MSE_o^{perm(j)}$ — похибка після перестановки j -ої ознаки.

Налаштування гіперпараметрів виконано методом крос-валідації (5-fold) з використанням RandomizedSearchCV. Результати оптимізації наведено в таблиці 1.

Таблиця 1 – Оптимальні гіперпараметри Random Forest

Гіперпараметр	Діапазон	Оптимум
Кількість дерев (n_estimators)	[50, 500]	200
Максимальна глибина (max_depth)	[5, 30]	15
Мін. зразків для розщеплення	[2, 20]	5
Кількість ознак (max_features)	[sqrt, log2, all]	sqrt
Мін. зразків у листі	[1, 10]	2

Джерело: розроблено авторами.

Результати порівняння soft-сенсора з базовими методами, включаючи модель глибокого навчання LSTM, наведено в таблиці 2.

Таблиця 2 – Порівняння методів для soft-сенсора RON

Метод	MAE	RMSE	R ²	Час, мс	Адапт.	Невизначеність
Лінійна регресія	0,95	1,21	0,872	<1	проста	ні
PLS (5 комп.)	0,82	1,05	0,903	<1	проста	ні
Дерево рішень	0,68	0,89	0,931	<1	середня	ні
Random Forest	0,41	0,54	0,974	12	середня	вбудована
Gradient Boosting	0,39	0,51	0,977	8	складна	потребує калібр.
LSTM (64 нейрони)	0,37	0,49	0,979	45	складна	потребує калібр.

Джерело: розроблено авторами.

Random Forest забезпечує $MAE = 0,41$ пункту RON ($R^2 = 0,974$), що суттєво перевищує вимогу $\pm 1,0$ пункту. Gradient Boosting та LSTM показують дещо кращі результати за метриками точності ($MAE = 0,39$ та $0,37$ відповідно), проте RF обрано як основний метод з таких причин: по-перше, RF забезпечує вбудовану квантифікацію невизначеності через дисперсію прогнозів дерев без додаткових процедур калібрування [10]; по-друге, інкрементальне оновлення ансамблю (заміна окремих дерев) є значно простішим та швидшим, ніж дообучення градієнтного бустингу чи LSTM; по-третє, час розрахунку RF (12 мс) є прийнятним для реального часу, тоді як LSTM потребує 45 мс. Слід зазначити, що Gradient Boosting (зокрема LightGBM) теоретично також підтримує інкрементальне навчання, проте його квантифікація невизначеності потребує додаткових обчислювальних витрат (наприклад, через NGBBoost або конформні інтервали), що нівелює перевагу у швидкодії.

Аналіз важливості ознак (рис. 2) підтвердив відповідність моделі фізичним закономірностям: найбільший внесок мають температура виходу з останнього реактора ($I = 0,285$), різниця густин ($I = 0,195$) та середня температура ($I = 0,148$). За аналогічною методологією розроблено soft-сенсор для виходу C_{5+} : $MAE = 0,38$ % мас., $R^2 = 0,968$.

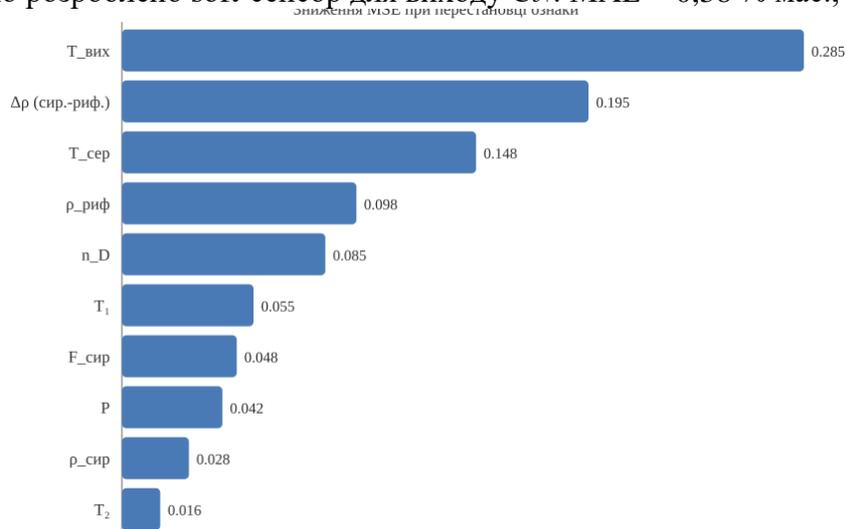


Рис. 2. Важливість ознак для soft-сенсора RON (permutation importance)

Джерело: розроблено авторами

3. Механізм онлайн-адаптації та квантифікації невизначеності. Промислові процеси характеризуються поступовою зміною характеристик з часом (concept drift), що призводить до деградації точності soft-сенсорів [9]. Для підтримки точності реалізовано механізм інкрементального оновлення soft-сенсора при надходженні нових лабораторних вимірювань.

Алгоритм онлайн-адаптації включає три етапи. На першому етапі відбувається накопичення нових даних: при надходженні лабораторного аналізу формується пара (x, y) , де x — вектор параметрів на момент відбору проби з урахуванням транспортної затримки, y — лабораторне значення RON. На другому етапі оцінюється необхідність адаптації: якщо похибка $|\hat{y} - y|$ перевищує поріг 2σ від середньої похибки на валідації, точка позначається як значуща. Вибір порогу 2σ обумовлено балансом між чутливістю до реального дрейфу та стійкістю до випадкових викидів: за результатами аналізу чутливості, збільшення порогу до 3σ призводило до пропуску 18 % випадків дрейфу, а зменшення до $1,5\sigma$ — до 25 % хибних спрацювань. На третьому етапі при накопиченні 10–20 нових точок виконується оновлення ансамблю:

$$RF_{ne}^W = \{T_{k+1}, \dots, T^b\} \cup \{T^{b+1}, \dots, T^{b+k}\}, \quad (5)$$

де видаляються k найстаріших дерев $\{T_1, \dots, T_k\}$ та додаються k нових дерев $\{T^{b+1}, \dots, T^{b+k}\}$, навчених на оновленій вибірці. Альтернативною стратегією є зважене оновлення, при якому кожному дереву присвоюється вага пропорційно його точності на нових даних. Порівняння стратегій показало, що просте заміщення забезпечує робастніший результат при меншій обчислювальній складності.

Інтервал довіри для прогнозу обчислюється на основі квантилів розподілу прогнозів дерев:

$$CI = [Q_{\alpha/2}, Q_{1-\alpha/2}], \quad (6)$$

де $Q_{\alpha/2}$ та $Q_{1-\alpha/2}$ — квантилі відповідного рівня розподілу прогнозів дерев. Для $\alpha = 0,05$ отримуємо 95 % інтервал довіри.

Для детекції аномальних станів використовується комбінований критерій:

$$Anomaly = (Q_{0,975} - Q_{0,025} > 2 \cdot CI_{av}^g) \vee (\hat{y} \notin [RON_{min}, RON_{max}]), \quad (7)$$

де CI_{av}^g — середня ширина інтервалу довіри на валідаційній вибірці; RON_{min} , RON_{max} — межі історичного діапазону. При детекції аномалії система управління отримує попередження, і прогноз не використовується для прийняття рішень до з'ясування причини.

Таблиця 3 – Зведені характеристики soft-сенсорів

Показник	Значення
MAE для RON, пункт	0,41
R ² для RON	0,974
MAE для виходу C ₅₊ , % мас.	0,38
R ² для виходу C ₅₊	0,968
Час розрахунку RON / C ₅₊ , мс	12 / 10
Середня ширина 95 % CI (RON), пункт	0,92
Середня ширина 95 % CI (C ₅₊), % мас.	0,85

Джерело: розроблено авторами

4. Верифікація системи в контурі предиктивного управління. Для верифікації створено програмний комплекс імітаційного моделювання установки CCR-риформінгу потужністю 1,5 млн т/рік на Python 3.10 з бібліотеками NumPy, SciPy, scikit-learn, CasADi, IPOPT [11]. Слід зазначити, що верифікація виконана на симуляційній моделі, а не на реальних промислових даних, що є типовим обмеженням для досліджень даного типу. Soft-сенсори інтегровано в контур NMPC з економічним критерієм оптимізації [12]. Функція цілі має вигляд:

$$J = \sum_{01}^{N} [\alpha_1 \cdot RON(k) + \alpha_2 \cdot YC_{5+}(k) - \alpha_3 \cdot E(k) - \alpha_4 \cdot \Phi_{деакт}(k)] \rightarrow \max, \quad (8)$$

де $RON(k)$ — октанове число; $YC_{5+}(k)$ — вихід рідкого продукту; $E(k)$ — енерговитрати; $\Phi_{деакт}(k)$ — штраф за інтенсивну дезактивацію каталізатора; α_1 – α_4 — вагові коефіцієнти; N — горизонт прогнозування (24 кроки = 2 год).

Задача оптимізації розв'язується методом послідовного квадратичного програмування (SQP) з використанням солвера IPOPT [11] з урахуванням обмежень:

$$u_{min} \leq u(k) \leq u_{max}, |Du(k)| \leq \Delta u_{max}, y_{min} \leq y(k) \leq y_{max} \quad (9)$$

де $u(k)$ — вектор керуючих впливів (температури входу в реактори, витрата сировини); $y(k)$ — вектор виходів (RON, C₅₊).

Проведено комплекс обчислювальних експериментів за шістьма сценаріями (стаціонарна робота, зміна завдання RON, ступінчасте та поступове збурення складу сировини, дезактивація каталізатора, комбінований сценарій). Кожен сценарій повторено 10 разів. Результати наведено в таблиці 4.

Таблиця 4 – Зведені результати порівняння стратегій управління

Показник	PID	Лін. MPC	NMPC
MAE RON, пункт	0,42±0,08	0,28±0,05	0,18±0,03
Час перех. процесу, хв	185±25	95±12	65±8
Макс. ΔRON при збуренні	1,85±0,25	1,22±0,18	0,68±0,12
MAE при дезактивації	0,85	0,52	0,22
Покращення vs PID, %	—	37	63

Джерело: розроблено авторами

Ефективність механізму адаптації підтверджено при моделюванні дезактивації каталізатора (рис. 3): повна адаптація (bias update + оновлення дерев) забезпечує підтримку MAE RON на рівні 0,24 пункту — лише на 33 % вище початкового рівня, тоді як без адаптації деградація становить 430 %.

Ефективність механізму адаптації при дезактивації каталізатора

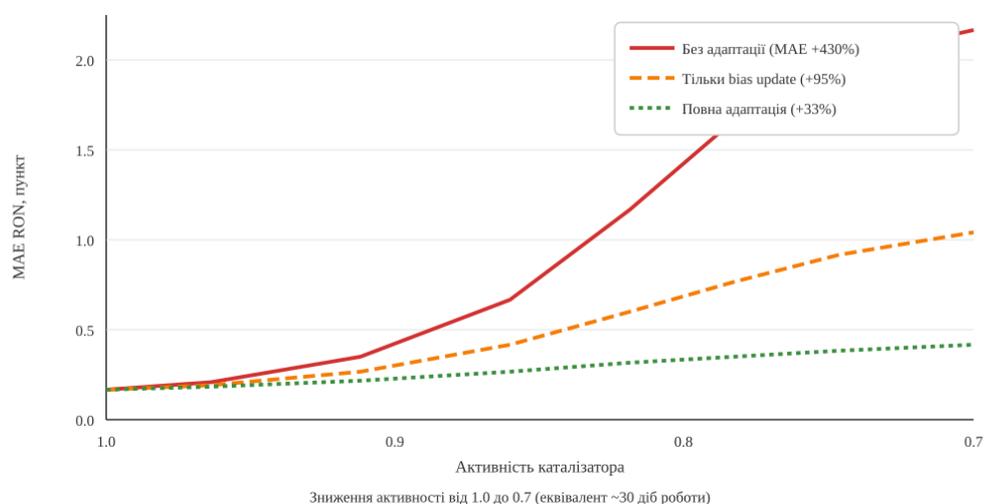


Рис. 3. Ефективність адаптації при дезактивації каталізатора

Джерело: розроблено авторами

Таблиця 5 – Економічні показники систем управління (30 діб)

Показник	PID	Лін. MPC	NMPC
Середнє RON	97,58	97,85	98,22
Середній вихід C ₅₊ , %	86,15	86,48	86,92
Час нижче специфікації, %	3,8	1,5	0,4
Приріст прибутку, тис. \$/міс	—	+104	+224

Джерело: розроблено авторами

Економічний ефект від впровадження NMPC з soft-сенсорами становить близько 2,7 млн \$ на рік. Основні складові ефекту: підвищення RON на 0,64 пункту, що збільшує маржу на 52 % від загального ефекту; збільшення виходу C₅₊ на 0,77 % мас. (35 %); зниження витрат палива на 6 % за рахунок оптимізації температурного профілю (8 %); зменшення втрат від невідповідності специфікації (5 %).

Висновки та обмеження. У роботі розроблено інтелектуальну систему оцінювання якості продукції каталітичного риформінгу в реальному часі на основі ансамблевих методів машинного навчання. Основні результати полягають у наступному.

1. Обґрунтовано вибір методу Random Forest для побудови soft-сенсорів якості риформату. Порівняльний аналіз шести методів, включаючи LSTM, показав, що RF забезпечує оптимальне поєднання точності (MAE = 0,41 пункту RON, $R^2 = 0,974$), вбудованої квантифікації невизначеності та можливості інкрементального оновлення. Хоча LSTM та Gradient Boosting показали дещо кращу точність, RF обрано завдяки простоті адаптації та вбудованій оцінці надійності прогнозів. Час розрахунку не перевищує 12 мс.

2. Реалізовано механізм онлайн-адаптації шляхом інкрементального оновлення ансамблю дерев (5), що забезпечує адаптацію без повного перенавчання. Механізм квантифікації невизначеності на основі розподілу прогнозів (3), (6) дозволяє детектувати аномальні режими за критерієм (7).

3. Верифікація в контурі NMPC з економічним критерієм (8) підтвердила: зниження варіації RON на 57 % порівняно з PID, скорочення перехідних процесів у 2,8 рази, зниження максимального відхилення при збуреннях на 63 %. Економічний ефект ~2,7 млн \$/рік для установки 1,5 млн т/рік.

4. Обмеження дослідження. Верифікація виконана на імітаційній моделі, а не на реальних промислових даних, що є типовим для досліджень на етапі розробки, проте потребує подальшої валідації. Модель симулятора базується на спрощеній кінетиці, що може не повністю відображати всі режими роботи реальної установки.

5. Перспективи досліджень пов'язані з валідацією на промислових даних, врахуванням динаміки регенерації катализатора, застосуванням методів трансферного навчання для адаптації до різних установок [5], інтеграцією з цифровими двійниками та дослідженням можливостей навчання з підкріпленням для оптимізації управління.

Заява про використання генеративного ШІ та технологій на основі ШІ в процесі написання тексту статті.

Під час написання цього матеріалу автор(и) використовували Claude (Anthropic) для допомоги у структуруванні та оформленні тексту статті. Після використання цього інструменту автор(и) переглянув(ли) та відредагував(ли) зміст за потреби і взяв(ли) на себе повну відповідальність за зміст публікації.

Список використаних джерел

1. Ali, S., Theravalappil, R., Alzaid, A., Alshareef, A., Huang, H., & Hossain, M. (2024). Molecular-level kinetic modeling of catalytic naphtha reforming. *Industrial & Engineering Chemistry Research*, 63. <https://doi.org/10.1021/acs.iecr.4c00735>.
2. Dong, X., Shen, J., Ma, Z.-F., & He, Y.-J. (2022). Robust optimal operation of continuous catalytic reforming process under feedstock uncertainty. *International Journal of Hydrogen Energy*, 47. <https://doi.org/10.1016/j.ijhydene.2022.08.161>.
3. Jiang, Y., Yin, S., Dong, J., & Kaynak, O. (2020). A review on soft sensors for monitoring, control and optimization of industrial processes. *IEEE Sensors Journal*. <https://doi.org/10.1109/JSEN.2020.3033153>.
4. Hemavathy, S. (2025). Soft sensors. https://doi.org/10.1007/978-981-96-2407-2_148
5. Curreri, F., Patané, L., & Xibilia, M. G. (2021). Soft sensor transferability: A survey. *Applied Sciences*, 11(16), 7710. <https://doi.org/10.3390/app11167710>.
6. Jia, M., Xu, D., Yang, T., Liu, Y., & Yao, Y. (2023). Graph convolutional network soft sensor for process quality prediction. *Journal of Process Control*, 123, 12–25. <https://doi.org/10.1016/j.jprocont.2023.01.010>.
7. Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2023). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, 237, 121549. <https://doi.org/10.1016/j.eswa.2023.121549>.
8. Shao, W., Ge, Z., & Song, Z. (2018). Soft-sensor development for processes with multiple operating modes based on semisupervised Gaussian mixture regression. *IEEE Transactions on Control Systems Technology*. <https://doi.org/10.1109/TCST.2018.2856845>.

9. Xiang, Q., Zi, L., Cong, X., & Wang, Y. (2023). Concept drift adaptation methods under the deep learning framework: A literature review. *Applied Sciences*, 13(11), 6515. <https://doi.org/10.3390/app13116515>.
10. Heid, E., McGill, C., Vermeire, F., & Green, W. (2023). Characterizing uncertainty in machine learning for chemistry. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv-2023-00vcg>.
11. Andersson, J., Gillis, J., Horn, G., Rawlings, J., & Diehl, M. (2018). CasADi: A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1), 1–36. <https://doi.org/10.1007/s12532-018-0139-4>.
12. Bagla, G., Patwardhan, S., & Valluru, J. (2023). Intelligent state estimation for fault tolerant integrated frequent RTO and adaptive nonlinear MPC. *Journal of Process Control*, 131, 103092. <https://doi.org/10.1016/j.jprocont.2023.103092>.

References

1. Ali, S., Theravalappil, R., Alzaid, A., Alshareef, A., Huang, H., & Hossain, M. (2024). Molecular-level kinetic modeling of catalytic naphtha reforming. *Industrial & Engineering Chemistry Research*, 63. <https://doi.org/10.1021/acs.iecr.4c00735>.
2. Dong, X., Shen, J., Ma, Z.-F., & He, Y.-J. (2022). Robust optimal operation of continuous catalytic reforming process under feedstock uncertainty. *International Journal of Hydrogen Energy*, 47. <https://doi.org/10.1016/j.ijhydene.2022.08.161>.
3. Jiang, Y., Yin, S., Dong, J., & Kaynak, O. (2020). A review on soft sensors for monitoring, control and optimization of industrial processes. *IEEE Sensors Journal*. <https://doi.org/10.1109/JSEN.2020.3033153>.
4. Hemavathy, S. (2025). Soft sensors. https://doi.org/10.1007/978-981-96-2407-2_148
5. Curreri, F., Patané, L., & Xibilia, M. G. (2021). Soft sensor transferability: A survey. *Applied Sciences*, 11(16), 7710. <https://doi.org/10.3390/app11167710>.
6. Jia, M., Xu, D., Yang, T., Liu, Y., & Yao, Y. (2023). Graph convolutional network soft sensor for process quality prediction. *Journal of Process Control*, 123, 12–25. <https://doi.org/10.1016/j.jprocont.2023.01.010>.
7. Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2023). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, 237, 121549. <https://doi.org/10.1016/j.eswa.2023.121549>.
8. Shao, W., Ge, Z., & Song, Z. (2018). Soft-sensor development for processes with multiple operating modes based on semisupervised Gaussian mixture regression. *IEEE Transactions on Control Systems Technology*. <https://doi.org/10.1109/TCST.2018.2856845>.
9. Xiang, Q., Zi, L., Cong, X., & Wang, Y. (2023). Concept drift adaptation methods under the deep learning framework: A literature review. *Applied Sciences*, 13(11), 6515. <https://doi.org/10.3390/app13116515>.
10. Heid, E., McGill, C., Vermeire, F., & Green, W. (2023). Characterizing uncertainty in machine learning for chemistry. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv-2023-00vcg>.
11. Andersson, J., Gillis, J., Horn, G., Rawlings, J., & Diehl, M. (2018). CasADi: A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1), 1–36. <https://doi.org/10.1007/s12532-018-0139-4>.
12. Bagla, G., Patwardhan, S., & Valluru, J. (2023). Intelligent state estimation for fault tolerant integrated frequent RTO and adaptive nonlinear MPC. *Journal of Process Control*, 131, 103092. <https://doi.org/10.1016/j.jprocont.2023.103092>.

Дата першого надходження статті до видання: 13.01.2026

Дата прийняття статті до друку після рецензування: 02.01.2026

Serhiy Tkachuk¹, Vadym Valeriyovych Bondar²

PhD Student, Department of Automation of Chemical Production
National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute” (Kyiv, Ukraine)

PhD Student, Department of Automation of Chemical Production
National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute” (Kyiv, Ukraine)

INTELLIGENT SYSTEM FOR PETROLEUM PRODUCT QUALITY ESTIMATION BASED ON ENSEMBLE MACHINE LEARNING METHODS

The efficiency of catalytic reforming units with continuous catalyst regeneration (CCR) significantly depends on the availability of real-time product quality information. However, key quality indicators, including the research octane number (RON) and liquid product yield are determined by laboratory methods with a periodicity of 2 to 8 hours, creating a significant delay in the control loop. The problem addressed in this study is the lack of a comprehensive approach to building an intelligent quality estimation system that combines high prediction accuracy with automatic adaptation to process changes and reliability assessment of predictions. The purpose of this article is to develop an intelligent system for real-time quality estimation of catalytic reforming products based on ensemble machine learning methods with online adaptation and prediction uncertainty quantification mechanisms. The Random Forest method was selected for soft sensor construction based on a comparative analysis of six methods (linear regression, PLS, decision tree, RF, Gradient Boosting, and LSTM). The soft sensor for RON prediction achieves MAE of 0.41 points with $R^2 = 0.974$. Although LSTM and Gradient Boosting showed slightly better accuracy metrics, RF was chosen due to its built-in uncertainty quantification and simpler incremental adaptation mechanism. An incremental online adaptation mechanism was implemented through the addition of new decision trees trained on updated data with simultaneous removal of the oldest trees. The uncertainty quantification mechanism based on the distribution of tree predictions enables the detection of abnormal operating modes. Verification within a nonlinear model predictive control (NMPC) loop with an economic objective function confirmed the system effectiveness: RON variation reduction by 57% compared to PID control, settling time reduction by a factor of 2.8, and maximum deviation reduction under disturbances by 63%. The adaptation mechanism maintains MAE at only 33% above the initial level during catalyst deactivation, compared to 430% degradation without adaptation. The economic effect for a 1.5 million tons per year unit amounts to approximately 2.7 million USD per year.

Keywords: soft sensor; quality estimation; Random Forest; neural networks; online adaptation; catalytic reforming; predictive control; ensemble methods; LSTM.

Fig.: 3. Tables: 5. References: 12.