

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Київський національний університет імені Тараса Шевченка
Фізичний факультет
Кафедра загальної фізики

На правах рукопису

Параметризація рухливості носіїв заряду в кремнії з використанням машинного навчання

Галузь знань: 10 Природничі науки
Спеціальність: 104 Фізика та астрономія
Освітня програма: Фізика наносистем

Кваліфікаційна робота магістра
студента 2 курсу
Кущ Іван

Науковий керівник:
доктор фізико-математичних наук,
професор, завідувач кафедри загальної фізики
Олег ОЛІХ

Робота заслухана на засіданні кафедри загальної фізики
та рекомендована до захисту на ЕК, протокол №_____ від «____»
_____ 2025р.

Завідувач кафедри загальної фізики _____ проф. **Олег ОЛІХ**

Київ – 2025

ВИТЯГ

з протоколу №_____
засідання Екзаменаційної комісії

Визнати, що студент _____ виконав та захистив
кваліфікаційну роботу магістра з оцінкою _____.

Голова ЕК _____ «____» ____ 2025 р.

АНОТАЦІЯ

Кущ Іван Параметризація рухливості носіїв заряду в кремнії з використанням машинного навчання. Кваліфікаційна робота магістра за спеціальністю 104 Фізика та астрономія, освітня програма «Фізика наносистем». – Київський національний університет імені Тараса Шевченка, фізичний факультет, кафедра загальної фізики. – Київ – 2025.

Науковий керівник: доктор фізико-математичних наук, професор, завідувач кафедри загальної фізики Олег Оліх.

У роботі наведено результати розрахунків рухливості носіїв заряду в монокристалічному кремнії з використанням методів машинного навчання: випадкового лісу, градієнтного бустингу, опорних векторів і нейронних мереж. За допомогою символьної регресії отримано аналітичні вирази для електронів і дірок у межах температур 200–500 К та концентрацій 10^{13} – 10^{19} см $^{-3}$. Моделі порівняно між собою та з відомими підходами.

Ключові слова: кремній, рухливість носіїв заряду, машинне навчання, символна регресія, моделі розсіяння.

SUMMERY

Kush Ivan Parameterization of charge carrier mobility in silicon using machine learning. Master's qualification work in specialty 104 Physics and Astronomy, educational program "Physics of Nanosystems". – Taras Shevchenko National University of Kyiv, Faculty of Physics, Department of General Physics. – Kyiv – 2025.

Research supervisor: – Doctor of Physical and Mathematical Sciences, Professor, Head of the Department of General Physics Oleh Olikh. The study presents charge carrier mobility calculations in monocrystalline silicon using machine learning methods, including random forest, gradient boosting, support vector machines, and neural networks. Symbolic regression was used to derive analytical expressions for electron and hole mobility within the 200–500 K temperature range and dopant concentrations of 10^{13} – 10^{19} cm $^{-3}$. The models were compared with each other and with known approaches.

Keywords: silicon, charge carrier mobility, machine learning, symbolic regression, scattering models.

Зміст

Вступ	4
1 Огляд літератури	5
1.1 Моделі для опису рухливості носіїв заряду у кремнії	5
1.2 Алгоритми машинного навчання	9
1.2.1 Загальні характеристики випадкового лісу	9
1.2.2 Основні властивості символної регресії	10
1.3 Застосування символної регресії у різних фізичних задачах	13
2 Методика обчислень	16
2.1 Створення тренувального та тестового наборів	16
2.2 Налаштування Symbolic Regression	17
2.3 Класичні регресійні моделі	17
2.4 Метрики оцінювання	19
3 Отримані результати	21
3.1 Символьна регресія	21
3.2 Класичні регресійні моделі	23
3.3 Порівняння точності різних алгоритмів машинного навчання	28
Висновки	32
Список використаних джерел	34

Вступ

Рухливість носіїв заряду є однією з фундаментальних фізичних характеристик напівпровідниківих матеріалів. Вона визначає швидкість переміщення електронів і дірок під впливом електричного поля, а отже, безпосередньо впливає на електропровідність, частотну характеристику, теплову стабільність та загальну ефективність функціонування електронних і оптоелектронних пристрій. Зокрема, в сучасних кремнієвих транзисторах, фотодіодах, сонячних елементах та сенсорах точне розуміння і контроль за рухливістю дозволяють досягти кращої продуктивності, енергоефективності та надійності.

У реальних умовах рухливість залежить від низки параметрів: температури, рівня легування, типу носіїв (електрони або дірки), механізмів розсіяння (на фононах, домішках, дефектах тощо), а також від внутрішньої структури матеріалу. Саме тому впродовж десятиліття науковці розробляли емпіричні та напівемпіричні моделі для кількісного опису цих залежностей. Серед найбільш відомих підходів варто відзначити моделі Agora, Masetti, Fletcher та Klaassen, кожна з яких має свої переваги і недоліки залежно від області застосування.

Хоча зазначені моделі залишаються актуальними, вони часто мають обмеження у вигляді великої кількості емпіричних коефіцієнтів, складних формул чи недостатньої точності в окремих діапазонах температур і концентрацій. Це створює потребу в розробці нових інструментів, які здатні враховувати більше факторів та адаптуватися до різних умов без надмірного ускладнення математичного апарату.

З огляду на бурхливий розвиток інтелектуальних алгоритмів обробки даних, машинне навчання (ML) пропонує нову парадигму моделювання фізичних властивостей матеріалів. Ці методи дають змогу автоматично виявляти нелінійні закономірності в багатовимірних наборах даних, створюючи моделі з високою точністю прогнозування. Крім класичних методів (таких як дерева рішень, Random Forest, градієнтний бустинг, підтримувальні векторні машини, нейронні мережі), особливу увагу привертає символьна регресія — підхід, який дозволяє отримувати компактні й аналітично зрозумілі формули, що легко інтерпретуються.

Символьна регресія поєднує точність з інтерпретованістю, що особливо цінно для фізичних задач, де критично важливо розуміти внутрішню структуру моделі. Такий підхід дає змогу не лише замінити традиційні моделі, але й побудувати більш універсальні й адаптивні формулі на основі сучасних даних і знань.

Метою даної роботи є побудова моделей для прогнозування рухливості носіїв заряду в монокристалічному кремнії з використанням алгоритмів машинного навчання. У роботі розглядаються переваги та обмеження класичних підходів, формуються навчальні та тестові вибірки на основі моделі Klaassen, реалізується декілька типів ML-моделей, а також досліджується ефективність символної регресії для створення інтерпретованих фізичних виразів. Застосування таких підходів відкриває нові можливості для оптимізації матеріалів і процесів у мікроелектроніці та суміжних галузях.

1. Огляд літератури

1.1. Моделі для опису рухливості носіїв заряду у кремнії

Перш ніж застосовувати машинне навчання для побудови моделей рухливості, необхідно мати фізично обґрунтовані теоретичні оцінки, які слугують базовими орієнтирами або джерелами навчальних даних. У цій роботі було розглянуто три основні аналітичні методи — Классена, Аори та Флетчера, кожен із яких має свої переваги, обмеження та сферу застосування.

Задача оцінки рухливості електронів μ_n та дірок μ_p у напівпровіднику за певних умов є достатньо поширеною у різноманітних фізичних дослідженнях. Один з варіантів її вирішення полягає у використанні загального підходу, згідно з яким

$$\mu = \frac{e\tau_p}{m_\sigma}, \quad (1.1)$$

де e — елементарний заряд, τ_p — середній час вільного пробігу носія заряду, m_σ — ефективна маса електропровідності.

Час вільного пробігу обмежується розсіянням носіїв заряду, яке може бути викликане декількома причинами, пов'язаними з порушеннями періодичності потенціалу. Зокрема виділяють розсіяння на коливаннях ґратки (акустичних та оптических фононах), заряджених та нейтральних домішках, дислокаціях, границях зерен та інших неоднорідностях структури, поверхнях та межах розділу, інших носіях. Кожен із цих механізмів має свою залежність від температури, рівня легування та розміру напівпровідникової структури і може бути визначальним для величини рухливості за певних умов. Проте найчастіше необхідно враховувати декілька можливих шляхів розсіяння носіїв заряду. В такому випадку для оцінки рухливості використовується правило Маттісена:

$$\mu^{-1} = \sum_i \mu_i^{-1}, \quad (1.2)$$

де сумування відбувається за механізмами розсіяння, μ_i — рухливість носіїв за наявності лише i -го механізму розсіяння. Для оцінки μ_i можна використовувати вирази, аналогічні формулі (1.1), розрахувавши відповідний час вільного пробігу.

Для переважної більшості механізмів розсіяння вирази для оцінки рухливості відомі. Так, при розсіянні на іонізованих домішках нерідко використовується вираз Brooks & Herring [1]:

$$\mu_I = \frac{3,68 \cdot 10^{20} \text{ см}^{-3}}{N_I Z^2} \left(\frac{\varepsilon}{16} \right)^2 \left(\frac{T}{100 \text{ K}} \right)^{1,5} \frac{(m_0/m)^{0,5}}{\left[\lg(1 + \beta_{BH}^2) - \frac{0,434 \beta_{BH}^2}{1 + \beta_{BH}^2} \right]}, \quad (1.3)$$

де N_I — концентрація домішок із зарядом Ze , ε — діелектрична проникність напівпровідника, T — температура, m — ефективна маса носія, m_0 — маса вільного електрону, а величина β_{BH} має вигляд

$$\beta_{BH} = \left(\frac{\varepsilon}{16} \right)^{0,5} \frac{T}{100 \text{ K}} \left(\frac{m}{m_0} \right)^{0,5} \left(\frac{2,08 \cdot 10^{18} \text{ см}^{-3}}{n_c} \right)^{0,5}, \quad (1.4)$$

n_c — концентрація носіїв заряду. Іншим наближенням для такого випадку є формула Conwell & Weisskopf [1]:

$$\begin{aligned} \mu_I &= \frac{3,68 \cdot 10^{20} \text{ см}^{-3}}{N_I Z^2} \left(\frac{\varepsilon}{16} \right)^2 \left(\frac{T}{100 \text{ K}} \right)^{1,5} \frac{(m_0/m)^{0,5}}{\lg(1 + \beta_{CW}^2)}, \\ \beta_{CW} &= \frac{1}{Z} \frac{\varepsilon}{16} \frac{T}{100 \text{ K}} \left(\frac{2,35 \cdot 10^{19} \text{ см}^{-3}}{N_I} \right)^{1/3}. \end{aligned} \quad (1.5)$$

Для іншого механізму, розсіяння носій-носій, застосовується підхід, розвинутий Fletcher [2]:

$$\mu_{cc} = \frac{\left(\frac{T}{T_{ref}^{Fl}} \right)^{3/2} F_1}{(np)^{1/2} \ln \left[1 + \left(\frac{T}{T_{ref}^{Fl}} \right)^2 (np)^{-1/3} F_2 \right]}, \quad (1.6)$$

де n та p — концентрації електронів та дірок, відповідно; T_{ref}^{Fl} , F_1 та F_2 — певні константи, які залежать від матеріалу. Як показано в роботах [3, 4], для кремнію доцільно застосовувати $T_{ref}^{Fl} = 300 \text{ K}$, $F_1 = 1,04 \cdot 10^{21} \text{ см}^{-1} \text{ В}^{-1} \text{ с}^{-1}$, $F_2 = 7,45 \cdot 10^{12} \text{ см}^2$ і тоді вираз (1.6) перетворюється на наступний

$$\mu_{cc} = \frac{2 \cdot 10^{17} T^{3/2}}{\sqrt{np} \ln \left[1 + 8,28 \cdot 10^8 T^2 (np)^{-1/3} \right]}. \quad (1.7)$$

Табл. 1. Коефіцієнти для розрахунку рухливості відповідно до моделі Caughey–Thomas (1.8)

Тип носіїв	Параметр			
	μ_{\max} , $\text{см}^2/(\text{B} \cdot \text{c})$	μ_{\min} , $\text{см}^2/(\text{B} \cdot \text{c})$	α	N_{ref} , см^{-3}
Електрони	1330	65	0,72	$8,5 \cdot 10^{16}$
Дірки	495	47,7	0,76	$6,3 \cdot 10^{16}$

Табл. 2. Коефіцієнти для розрахунку рухливості відповідно до моделі Masetti (1.9)

Параметр	Легант	
	Фосфор	Бор
μ_0 , $\text{см}^2/(\text{B} \cdot \text{c})$	68,5	44,9
μ_{\max} , $\text{см}^2/(\text{B} \cdot \text{c})$	1414	470,5
μ_0 , $\text{см}^2/(\text{B} \cdot \text{c})$	56,1	29,0
C_r , см^{-3}	$9,2 \cdot 10^{16}$	$2,23 \cdot 10^{17}$
C_s , см^{-3}	$3,41 \cdot 10^{20}$	$6,10 \cdot 10^{20}$
a	0,711	0,719
b	1,98	2,00
p_c , см^{-3}	—	$9,23 \cdot 10^{16}$

Проте в більшості випадків для більш-менш точного опису рухливості реального матеріалу необхідно враховувати значну кількість механізмів. Як наслідок, більше поширення отримав підхід оцінки величини μ з використанням апроксимаційної функції, яка часто базується на результатах експериментальних вимірювань. Як правило, для кожного матеріалу вигляд функції або наявні в ній коефіцієнти відрізняються.

Розглянемо декілька подібних підходів до опису рухливості носіїв заряду у монокристалічному кремнії, обмежуючись випадками об'ємного напівпровідника (без врахування впливу поверхні) та слабких полів. Одним з перших подібних наближень був вираз, запропонований Caughey & Thomas [5]:

$$\mu^{\text{CT}} = \mu_{\min} + \frac{\mu_{\max} - \mu_{\min}}{1 + (N/N_{ref})^\alpha}, \quad (1.8)$$

де N — концентрація легантів, а значення констант для випадків, коли розглядаються рухливості електронів та дірок, наведені у Табл. 1. Вираз насамперед призначений для оцінки залежності рухливості основних носіїв від концентрації легуючої домішки поблизу 300 К. З іншого боку, формула (1.8) не дозволяє оцінити рухливості неосновних носіїв та не враховує температурну залежність μ .

Модифікований варіант такого підходу був запропонований Masetti [6]:

$$\begin{aligned} \mu_n^{\text{M}} &= \mu_0 + \frac{\mu_{\max} - \mu_0}{1 + \left(\frac{n}{C_r}\right)^a} - \frac{\mu_1}{1 + \left(\frac{C_s}{n}\right)^b}, \\ \mu_p^{\text{M}} &= \mu_0 \exp\left(-\frac{p_c}{p}\right) + \frac{\mu_{\max}}{1 + \left(\frac{p}{C_r}\right)^a} - \frac{\mu_1}{1 + \left(\frac{C_s}{p}\right)^b}. \end{aligned} \quad (1.9)$$

Як видно з виразу (1.9), в такому випадку передбачено застосування різних виразів для опису рухливості електронів та дірок. Більше того, коефіцієнти мають залежати від того, які атоми були використання для легування — у Табл. 2 приведені їхні значення для бору та фосфору, які є найбільш типовими домішками для отримання кремнію з дірковою та електронною провідностями, відповідно. Теорія Masetti краще описує експериментальні дані, ніж підхід Caughey–Thomas, проте так само не дозволяє отримати інформацію для неосновних носіїв та температур, відмінних від кімнатної.

Спроба взяти до уваги температурні залежності була зроблена N. Arora зі співавторами [7]. При цьому розглядалися розсіяння на фононах та іонізованих домішках і було запропоновано використання узагальненого виразу Caughey–Thomas

$$\mu^{\text{A}} = \mu_{\min} + \frac{\mu_0}{1 + \left(\frac{N_I}{N_0}\right)^\alpha}, \quad (1.10)$$

де $N_I = N_a^- + N_d^+$ — загальна концентрація іонізованих легантів (N_a^- та N_d^+ — кількості іонізованих акцепторів та донорів у одиниці об'єму, відповідно), а решта коефіцієнтів у виразі (1.10) залежать від температури

Табл. 3. Коефіцієнти для розрахунку рухливості відповідно до моделі Arora (1.10)-(1.11)

Параметр	Тип носіїв			
	електрони		дірки	
	основні [7]	неосновні [8]	основні [7]	неосновні [9]
T_{ref} , К	300	295	300	292
μ_{min}^{ref} , $\text{см}^2/(\text{B} \cdot \text{с})$	88	232	54.3	130
μ_0^{ref} , $\text{см}^2/(\text{B} \cdot \text{с})$	1252	1180	407	370
N_0^{ref} , см^{-3}	$1,26 \cdot 10^{17}$	$8 \cdot 10^{16}$	$2,35 \cdot 10^{17}$	$8 \cdot 10^{17}$
α^{ref}	0,88	0,9	0,88	1,25
β_1		-0,57		
β_2		-2,33		-2,23
β_3			2,4	
β_4			-0,146	

Табл. 4. Коефіцієнти для розрахунку рухливості відповідно до моделі Klaassen, формули (1.13)-(1.14)

Тип носіїв	Параметр			
	μ_{max} , $\text{см}^2/(\text{B} \cdot \text{с})$	μ_{min} , $\text{см}^2/(\text{B} \cdot \text{с})$	α	N_{ref} , см^{-3}
Електрони	1414	68,5	0,711	$9,2 \cdot 10^{16}$
Дірки	495	44,9	0,719	$2,23 \cdot 10^{17}$

схожим чином:

$$\begin{aligned} \mu_{min} &= \mu_{min}^{ref} \left(\frac{T}{T_{ref}} \right)^{\beta_1}, \\ \mu_0 &= \mu_0^{ref} \left(\frac{T}{T_{ref}} \right)^{\beta_2}, \\ N_0 &= N_0^{ref} \left(\frac{T}{T_{ref}} \right)^{\beta_3}, \\ \alpha &= \alpha^{ref} \left(\frac{T}{T_{ref}} \right)^{\beta_4}. \end{aligned} \quad (1.11)$$

Тобто, відбулося повернення до ситуації, коли вираз для опису рухливості не залежить від типу носіїв, проте різними є константи, що мають бути використані під час розрахунків. Зауважимо, що у оригінальній роботі [7] вирази отримані (та визначені значення констант) на основі аналізу експериментальних даних щодо рухливості основних носіїв при різних значеннях температури та ступеню легування. Проте нерідко вирази (1.10)-(1.11) поширюють і для неосновних носіїв, використовуючи значення констант, отримані в роботах [8] (для неосновних електронів) та [9] (для неосновних дірок). Проте у вказаних роботах приведені дані лише для однієї температури і тому поширення виразів на широкий температурний діапазон дещо самовпевнено. Значення констант для всіх випадків наведено у Табл. 3.

Проте найбільш точним з точки зору збіжності з експериментальними даними вважається підхід до оцінки рухливості, розвинутий Klaassen [10] (на який в літературі також посилаються як на модель Philips). В цьому випадку вважається, що визначальними для рухливості є процеси розсіяння на коливаннях гратки, іонізованих домішках та інших носіях, а температурну та концентраційну залежності μ як для основних, так і неосновних носіїв заряду можна описати наступними виразами:

$$\mu^K = \frac{\mu_L \mu_{DA}}{\mu_L + \mu_{DA}}, \quad (1.12)$$

де

$$\mu_L = \mu_{max} \left(\frac{300}{T} \right)^{2,25} \quad (1.13)$$

$$\mu_{DA} = \frac{\mu_{max}^2}{\mu_{max} - \mu_{min}} \cdot \frac{N_{sc}}{N_{eff}} \cdot \left(\frac{N_{ref}}{N_{sc}} \right)^\alpha \cdot \left(\frac{T}{300} \right)^{3\alpha-1,5} + \frac{\mu_{max} \mu_{min}}{\mu_{max} - \mu_{min}} \cdot \frac{n+p}{N_{eff}} \cdot \left(\frac{300}{T} \right)^{0,5} \quad (1.14)$$

а константи, що входять до рівнянь (1.13)-(1.14) наведені у Табл. 4. Величини N_{sc} та N_{eff} є функціями, вигляд яких залежить від типу носіїв:

Табл. 5. Коефіцієнти для розрахунку рухливості відповідно до моделі Klaassen, формули (1.17) та (1.18)

Параметр	S_1	S_2	S_3	S_4	S_5	S_6	S_7
Значення	0,89233	0,41372	0,19778	0,28227	0,005978	1,80618	0,72169
Параметр	r_1	r_2	r_3	r_4	r_5	r_6	
Значення	0,7643	2,2999	6,5502	2,3670	-0,01552	0,6478	

$$\text{електрони: } N_{sc} = N_d^+ + N_a^- + p, \quad (1.15a)$$

$$\text{дірки: } N_{sc} = N_a^- + N_d^+ + n, \quad (1.15b)$$

$$\text{електрони: } N_{\text{eff}} = N_d^+ + N_a^- \cdot G_n(P_n, T) + \frac{p}{F_n(P_n, T)}, \quad (1.16a)$$

$$\text{дірки: } N_{\text{eff}} = N_a^- + N_d^+ \cdot G_p(P_p, T) + \frac{n}{F_p(P_p, T)}. \quad (1.16b)$$

В свою чергу, для функцій $G_{n(p)}$, $F_{n(p)}$ та $P_{n(p)}$ справедливі наступні вирази:

$$G_n = 1 - \frac{S_1}{\left[S_2 + \left(\frac{m_0}{m_e} \frac{T}{300} \right)^{S_4} P_n \right]^{S_3}} + \frac{S_5}{\left[\left(\frac{m_e}{m_0} \frac{300}{T} \right)^{S_7} P_n \right]^{S_6}}, \quad (1.17a)$$

$$G_p = 1 - \frac{S_1}{\left[S_2 + \left(\frac{m_0}{m_h} \frac{T}{300} \right)^{S_4} P_p \right]^{S_3}} + \frac{S_5}{\left[\left(\frac{m_h}{m_0} \frac{300}{T} \right)^{S_7} P_p \right]^{S_6}}, \quad (1.17b)$$

$$F_n = \frac{r_1 P_n^{r_6} + r_2 + r_3 \frac{m_e}{m_h}}{P_n^{r_6} + r_4 + r_5 \frac{m_e}{m_h}}; \quad F_p = \frac{r_1 P_p^{r_6} + r_2 + r_3 \frac{m_h}{m_e}}{P_p^{r_6} + r_4 + r_5 \frac{m_h}{m_e}}, \quad (1.18)$$

$$P_n = (P_{bn} + P_{cn})^{-1}; \quad P_p = (P_{bp} + P_{cp})^{-1}, \quad (1.19)$$

де m_e та m_h — ефективні маси електрона та дірки, відповідно, а константи, що входять до виразів (1.17) та (1.18), наведені в Табл. 5. У випадку, коли мова йде про неосновні носії, то $P_{cn} = 0$ ($P_{cp} = 0$). При розрахунках для основних носіїв

$$P_{cn} = \frac{2,46}{3,97 \cdot 10^{19} \left[\left(\frac{T}{300 Z_n} \right)^3 \frac{1}{N_d^+} \right]^{2/3}}; \quad P_{cp} = \frac{2,46}{3,97 \cdot 10^{19} \left[\left(\frac{T}{300 Z_p} \right)^3 \frac{1}{N_a^-} \right]^{2/3}}, \quad (1.20a)$$

$$Z_n = 1 + \frac{1}{0,21 + \left(\frac{4 \cdot 10^{20}}{N_d^+} \right)^2}; \quad Z_p = 1 + \frac{1}{0,5 + \left(\frac{7,2 \cdot 10^{20}}{N_a^-} \right)^2}, \quad (1.20b)$$

$$P_{bn} = \frac{3,83}{\frac{1,36 \cdot 10^{26}}{N_d^+ + p} \left(\frac{T}{300} \right)^2 \frac{m_e}{m_0}}; \quad P_{bp} = \frac{3,83}{\frac{1,36 \cdot 10^{26}}{N_a^- + n} \left(\frac{T}{300} \right)^2 \frac{m_h}{m_0}}. \quad (1.20b)$$

Як можна бачити, підхід Klaassen (1.12)-(1.20) передбачає застосування достатньо комплексних виразів та більше двадцяти констант. Це, звичайно, не створює непереборних складностей, але призводить до того, що на практиці для оцінки рухливості носіїв заряду в кремнії нерідко використовується більш простий підхід Arora (1.10)-(1.11), незважаючи на його меншу точність.

Зазначимо, що питання опису рухливості продовжує привертати увагу і періодично з'являються нові моделі. Наприклад, пропонуються вирази, які дозволяють врахувати додаткові механізми розсіяння за рахунок модифікації моделі Masetti [11], або використовують апроксимаційні поліноми [12]. Якщо узагальнити все вищесказане, то можна виділити три основні моделі, які використовуються для оцінки рухливості носіїв заряду у кремнії:

- **Классен:** найбільш повна модель, підходить для симуляцій на всьому діапазоні температур та легування; складна у реалізації, потребує великої кількості емпіричних параметрів.

- **Арора:** компроміс між точністю та простотою, добре працює при середніх температурах та концентраціях.

- **Флетчер:** підходить для попередніх оцінок або швидких розрахунків у системах з відомими температурними профілями.

Таким чином, використання одразу кількох теоретичних моделей дозволяє гнучко адаптувати обчислення під конкретні умови, а також створює якісну основу для побудови та навчання моделей машинного навчання.

1.2. Алгоритми машинного навчання

Машинне навчання та штучний інтелект – це не просто інструменти, а нова філософія роботи з інформацією. Вони змінюють підхід до наукових досліджень, роблячи їх швидшими, точнішими та масштабнішими. У майбутньому ці технології можуть привести до відкриттів, які ми сьогодні навіть не можемо уявити. Машинне навчання (ML) ділиться на три основні категорії: навчання з учителем (Supervised Learning, SL), навчання без учителя (Unsupervised Learning, UL) і навчання з підкріпленням (Reinforcement Learning, RL).

У методі навчання з учителем (Supervised Learning, SL) [13], модель навчається на розмічених даних, тобто кожен входний об'єкт супроводжується відповідним вихідним значенням. Мета - знайти функцію f , яка апроксимує залежність між входами X і виходами Y :

$$Y = f(X) + \epsilon$$

де ϵ – випадкова помилка.

Навчання без учителя (Unsupervised Learning) Цей метод використовується, коли дані не мають розмічених міток. Алгоритм шукає приховані структури в даних. Прикладом такого методу є метод кластеризації об'єднані схожі об'єкти.

В навчання з підкріпленням (Reinforcement Learning) агент взаємодіє зі середовищем та отримує винагороду за дії. Мета - максимізувати функцію винагороди:

$$Q(s, a) = r + \gamma \max_{a'} Q(s', a')$$

де $Q(s, a)$ - цінність дії a в стані s , r - винагорода, γ - коефіцієнт дисконтування.

1.2.1 Загальні характеристики випадкового лісу

У сучасному світі величезний потік інформації вимагає ефективних методів її аналізу. Одним із найбільш популярних і потужних алгоритмів машинного навчання є випадковий ліс (Random Forest, RF). Він поєднує простоту дерева рішень з ансамблевим підходом, що дозволяє підвищити точність і зменшити ризик переобучення. Випадковий ліс є ансамблевим методом, який базується на методі бутстреп-агрегування (bagging). Основна ідея полягає в тому, щоб створити кілька незалежних дерев рішень і об'єднати їхні результати. У кожному вузлі дерева випадковим чином вибирається тільки частина ознак, що зменшує кореляцію між деревами. Дерево росте до максимальної глибини або досягнення порогу розбиття. Для класифікації використовується голосування більшості: клас, що зустрічається найчастіше серед дерев, обирається як остаточний. Нехай маємо датасет:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

де x_i – вектор ознак, а y_i – мітка класу (у класифікації) або числове значення (у регресії).

Остаточний клас \hat{y} визначається більшістю голосів:

$$\hat{y} = \arg \max_c \sum_{i=1}^m \mathbb{I}(T_i(x) = c)$$

де:

- m – кількість дерев у лісі, - $T_i(x)$ – прогноз i -го дерева, - c – один із класів, - $\mathbb{I}(\cdot)$ – індикаторна функція, що дорівнює 1, якщо прогноз дерева дорівнює класу c , і 0 в іншому випадку.

Ансамблевий підхід дозволяє випадковому лісу перевершувати класичні моделі, такі як лінійна регресія або поодинокі дерева рішень, особливо у випадку складних нелінійних залежностей.

Для кожного дерева використовується випадкова підмножина ознак, що зменшує кореляцію між деревами та підвищує стійкість алгоритму.

Формально, помилка узгодженого ансамблю (ensemble error) визначається як:

$$E = \rho E_T + \frac{1 - \rho}{m} E_I$$

де:

- E – загальна помилка ансамблю,
- ρ – середня кореляція між деревами,
- E_T – помилка окремого дерева,
- E_I – незалежна помилка.

Якщо $m \rightarrow \infty$ і ρ мале, тоді $E \rightarrow E_I$, тобто помилка зменшується!

Оскільки алгоритм використовує бутстреп-вибірки, окрім аномальні точки не впливають на всі дерева, що робить модель стійкою до викидів (outliers).

Формально, випадковий ліс може зменшувати вплив викидів на оцінку середнього значення в задачах регресії:

$$\hat{y}_{RF} = \frac{1}{m} \sum_{i=1}^m T_i(x) \approx \text{median}(T_i(x))$$

Медіана є більш стійкою до викидів, ніж середнє значення!

Випадковий ліс дозволяє визначити, які ознаки мають найбільший вплив на прогноз.

Метрика важливості ознаки (Gini Importance) обчислюється як:

$$I(X_j) = \sum_{t \in T} p(t)(1 - p(t)) - \sum_{k \in \{\text{left}, \text{right}\}} p(t_k)(1 - p(t_k))$$

де:

- $I(X_j)$ — важливість ознаки X_j ,
- $p(t)$ — частка вибірки, яка потрапляє у вузол t ,
- $p(t_k)$ — частки після розбиття вузла на дочірні вузли.

Це допомагає проводити відбір ознак, спрощуючи модель!

Головний недолік випадкового лісу — це час навчання та використання, оскільки він створює багато дерев і робить численні обчислення. [13]

Оптимальне налаштування:

$$m = 100 - 500 \text{ дерев}, \quad k = \frac{p}{\sqrt{m}}, \quad \text{де } p \text{ — загальна кількість ознак.}$$

1.2.2 Основні властивості символальної регресії

Випадковий ліс, демонструючи вражуючу ефективність у задачах прогнозування, страждає від проблеми "чорної скриньки". Хоча модель здатна генерувати якісні результати, її внутрішня структура залишається непрозорою, що ускладнює інтерпретацію отриманих висновків та розуміння факторів, які впливають на результат. Для більш якіснох інтерпретацій наших даних та їх передбаченню був сворений метод — символна регресія (Symbolic Regression, SR). Символьна регресія (Symbolic Regression, SR) — це потужний метод машинного навчання, який дозволяє автоматично знаходити математичні функції, що найкраще описують дані, без необхідності заздалегідь визначати функціональну форму. [14] Цей метод є особливо корисним у випадках, коли немає можливості вивести функцію з перших принципів, або коли форма розподілу даних є довільною та складною для опису традиційними методами. Однією з ключових властивостей символальної регресії є її здатність автоматично шукати функції, які найкраще описують дані. [14] На відміну від традиційних методів регресії, таких як поліноміальна регресія, які вимагають заздалегідь визначені функціональні форми, SR дозволяє моделі самостійно знаходити оптимальну функцію. Це досягається за рахунок використання генетичного програмування, де функції представлені у вигляді дерев виразів, які еволюціонують протягом процесу навчання.

У генетичному програмуванні функції генеруються шляхом мутацій (зміни вузлів дерева) та кросоверу (обмін піддеревами між різними кандидатами). Цей процес дозволяє SR шукати функції в широкому просторі можливих математичних виразів, не обмежуючись попередньо заданими формами. Таким чином, SR усуває необхідність емпіричного підбору функцій, що є трудомістким і часто неточним процесом.

SR дозволяє знаходити функції, які точно описують такі розподіли, навіть якщо вони не мають явного аналітичного виразу. Це досягається за рахунок того, що SR не обмежується попередньо заданими функціональними формами, а шукає їх у широкому просторі можливих математичних виразів. Сума квадратів похибок (SSE) є широко використовуваною метрикою оцінки якості моделі в алгоритмах символальної регресії. Вона базується на обчисленні різниці між прогнозованими та фактичними значеннями цільової змінної та визначається наступним чином

$$\text{SSE} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{1.21}$$

де y_i — фактичне значення, \hat{y}_i — прогнозоване значення, n — кількість спостережень.

В процесі символальної регресії (SR) початкова популяція розв'язків формується випадковим чином з нескінченного пулу операторів та терміналів. Розмір популяції задається користувачем. Кожен окремий розв'язок може характеризуватися різною мірою придатності, впливаючи на загальну ефективність. Алгоритм SR постійно оцінює якість кожного розв'язку, відираючи для подальшої еволюції найбільш перспективні варіанти, що демонструють мінімальну похибку, та відкидаючи неефективні. В алгоритмі символальної лічної регресії (SR) еволюційний процес характеризується ітеративним зменшенням середньої помилки за рахунок елімінації розв'язків з низькою ефективністю. Завершення алгоритму SR відбувається при виконанні однієї з наступних умов: досягнення максимальної кількості ітерацій, заданої користувачем; досягнення

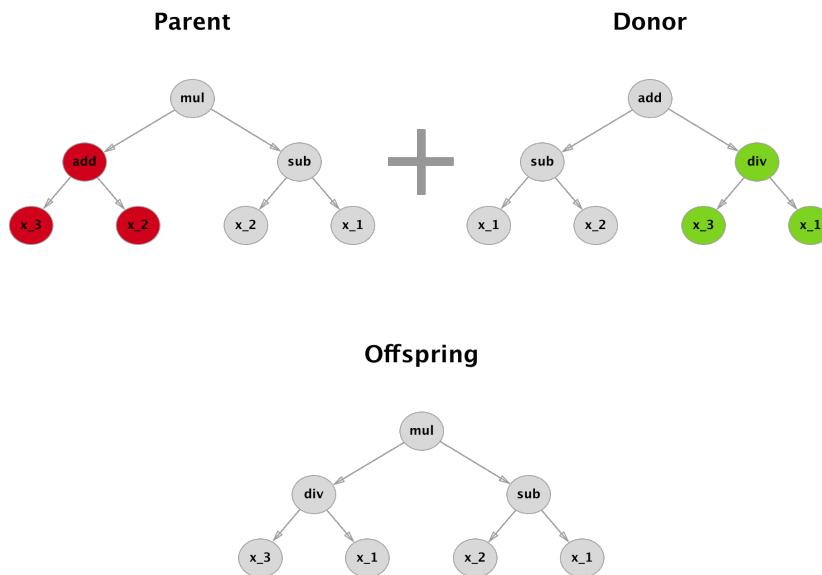


Рис. 1.1. Підхід генетичного програмування до символної регресії. Функції представлені деревами виразів. Нові функції генеруються за допомогою мутації деревних вузлів.

оптимального розв'язку з нульовою помилкою; досягнення прийнятного рівня помилки, встановленого користувачем (наприклад, 0,01). Зазвичай використовується комбінація цих критеріїв. Важливо зазначити, що SR може генерувати не тільки окреме рівняння, але й набір рішень, що відрізняються за складністю. Символьна регресія є ефективним методом, який дозволяє значно скоротити час і зусилля, необхідні для моделювання даних. Завдяки автоматизації процесу пошуку функцій, SR усуває необхідність ручного підбору функціональних форм, що є трудомістким і часто неточним процесом. Це особливо важливо в таких галузях, як фізика високих енергій, де дані можуть мати складні форми, і традиційні методи моделювання можуть бути недостатньо ефективними.

В ході багатьох спроб підбору найкращих параметрів, було прийнято метод оцінки моделі "model selection = "best" — цей метод оцінки моделі оптимізує фінальний результат навчання, між складністю моделі та середньою квадратичною помилкою. Тобто не дає "дереву" розроститися до неінтерпритованих варіантів і при цьому шукає максимально якісну апроксимацію даних. "populations = 3 * cpi count()" — Кількість окремих популяцій у стилі "island model"— кожна популяція еволюціонує незалежно, іноді обмінюючись найкращими формулами. Параметр покращує різноманіття рішень, зменшує ймовірність застрягнути в локальному мінімумі. Проте дуже залежить від числа процесорів, бо можна паралеліти. "population size = 1000" — Кількість формул в одній популяції. Працює за доволі простим принципом: більше = більше варіантів, але й довше обчислення. В описі цієї бібліотеки автор рекомендує значення від 100 до 1000, залежно від задачі й ресурсів. Параметр "ncycles per iteration = 500" це скільки еволюційних циклів робиться всерединіожної ітерації (до оновлення результату). Також модель можна контролювати, а саме скільки максимальних ітерацій всього процесу буде проведено, можна обмежувати модель за кількістю ітерацій, за цей метод відповідає параметр "niterations", проте в нього є великий недолік, що модель може застягти в локальному мінімумі і процес просто заступориться, проте не закінчиться до кінця ітерацій. Звісно, можна поставити обмеження по похибці, проте дуже важко в реальному часі спрогнозувати, як поведе модель до моменту поки не побачиш результати. Тому було вирішено обмежувати модель по часу, тоді ми не ризикуємо витрачати час та ресурси комп'ютера в "холосту". Так як під час навчання автоматично створюється файл, за допомогою якого можна запустити модель на місці, на якому вона закінчила навчання по закінченню часу, за цей метод відповідає параметр "timeout in seconds". В ході роботи було виявлено, що на один пакет даних, модель знаходить найкращі формули за 5-6 годин. "warm start" — Якщо вже був попередній запуск і є збережені тимчасові дані, алгоритм продовжить з того місця. Також ми можемо задати максимальний розмір формули, що генерується, задавши параметр "maxsize". Розмір формули рахується за кількістю всіх елементів у рівнянні (оператори + змінні + константи). Параметри "complexity of constants", "complexity of variables" та "complexity of operators" відповідають як раз за ціну одного елементу в рівнянні: константи, змінні та оператори (повноцінні функції як "sin" та "cos") відповідно. Також, аби фінальна формула не використовувала лише один з якихось операторів, використовується метод "constraints" — він обмежує можливу кількість користування операторами. "parsimony" — ще один з операторів, що допомагає уникати

```

model = PySRRegressor(
    model_selection="best",
    # model_selection="accuracy",
    # elementwise_loss = "L1DistLoss()", 
    # loss="L1DistLoss()", 
    populations = 3*cpu_count(),
    population_size = 1000,
    ncycles_per_iteration=500,
    # population_size = 300,
    # ncycles_per_iteration=3000,
    # niterations = 100,
    niterations = 500000,
    timeout_in_seconds=2 * 60 * 60,
    warm_start=True,
    early_stop_condition=("stop_if(loss, complexity) = loss < 1e-8 && complexity < 10"), # Stop early if we find
    maxsize = 30,
    # maxdepth= 7,
    # binary_operators=["*", "+", "/", "pow", "-",
    #                   "Zved(x,y)=x*y/(x+y)" ]
    #                   ],
    # unary_operators=["exp", "log", "tanh", "sin", "neg",
    #                   "PowInv(x)=1.0f0/(1.0f0+x)" ],
    complexity_of_constants = 1,
    complexity_of_variables = 1,
    complexity_of_operators={"tanh": 3, "sin":3},
    constraints={
        "pow": (-1, 5),
        "exp": 4,
    },
    # extra_sympy_mappings={
    #   "PowInv": lambda x: 1/(1+x),
    #   "Zved": lambda x,y: x*y/(x+y)
    # },
    # nested_constraints={
    #   "sin": {"sin": 0},
    #   "tanh": {"tanh": 0},
    #   "PowInv": {"PowInv": 0, "tanh": 0},
    #   "Zved": {"Zved": 0, "tanh": 0},
    # },
    parsimony=1e-5 ,
    # weight_optimize = 0.001,
    adaptive_parsimony_scaling = 100 ,
    # warmup_maxsize_by = 0.1,
    verbosity = 1,
    print_precision = 3,
    # -----
    # batching = True,
    # batch_size = 8,
    # -----
    # temp_equation_file = True,
    # temp_equation_file = False,
    # tempdir = "temp"
    # delete_tempfiles = False,
    # delete_tempfiles = True,
    # -----
    # procs= cpu_count()
)

```

Рис. 1.2. Основний код моделі

занадто складних формул, навіть якщо вони точні, балансує між точністю й простотою. Проте є можливість більш продвинуту впливати на параметр "parsimony" прямо під час роботи програми, це метод "adaptive parsimony scaling" — він автоматично масштабує вплив parsimony залежно від значень втрат (loss). Це особливо корисно, коли у нас loss змінюється в широкому діапазоні — підтримує баланс між loss і complexity. "verbosity" — це параметр, що відповідає за рівень виводу в консоль, 1 = базова інформація (поточний лідер, ітерації тощо) чи 0 = тихий режим. "procs" — Скільки процесів запускати паралельно. В ході налаштування моделі для оптимізації важливими параметрами стали Binary_operators та Unary_operators. Перші визначають, які бінарні операції можуть бути використані в процесі формування моделей. Це дозволяє регулювати складність рівнянь, які можна отримати, а також зберігати контроль над точністю та інтерпретованістю кінцевої формули. Операторів може бути кілька, і їх використання в кожному окремому випадку має критичне значення для балансу між ефективністю та складністю.

Unary_operators, в свою чергу, обмежують можливості роботи з унарними операціями, що дає можливість точніше контролювати математичні трансформації і спрощувати процес знаходження оптимальних рішень. Вони допомагають моделі уникати надмірної складності в процесі побудови формул.

Параметр Extra_sympy_mappings додає додаткові відображення в процес розв'язування задачі, що може бути корисним при роботі з складними виразами, де є потреба в інтеграції нестандартних математичних функцій, що не входять до стандартного набору.

Nested_constraints служать для введення додаткових обмежень на внутрішню структуру моделей. Цей параметр дозволяє розширювати можливості моделі, не порушуючи загальних правил побудови, таким чином, забезпечуючи більшу гнучкість і можливість оптимізації для складних систем, де потрібен контроль над кожним окремим рівнянням і його складовими. Метод Extra_sympy_mappings дозволяє створювати унікальні оператори, які можна використовувати під час побудови математичних моделей. Цей параметр дає можливість додавати кастомізовані функції, що можуть бути зручними для вирішення конкретних задач, де стандартні оператори не дають достатньо гнучкості або точності.

1.3. Застосування символної регресії у різних фізичних задачах

Вивчення потоків розрідженої газу у циліндричних трубах є ключовим для низки інженерних застосувань — від мікрофлюїдних пристрій до вакуумних систем і термоядерних реакторів. У таких задачах класичні методи, що базуються на кінетичній теорії, забезпечують високу точність, проте вимагають значних обчислювальних ресурсів. Тому зростає інтерес до методів машинного навчання, здатних швидко та точно відтворити поведінку системи на основі вже обчислених даних.

У статті [13] було продемонстровано ефективність використання методів випадкового лісу та символної регресії для апроксимації зменшеної витрати маси газу W як функції трьох безрозмірних параметрів: відношення тисків $p = P_2/P_1$, відносної довжини трубы $l = L/R$ та параметра розрідженності δ , що пов'язаний із числом Кнудсена.

У рамках дослідження, на основі згенерованого кінетичного набору даних було побудовано кілька аналітичних формул для $W(p, l, \delta)$ за допомогою символної регресії. Зокрема, було отримано загальну апроксимацію, що охоплює весь діапазон параметрів, з максимальною абсолютною похибкою менше 17%. Для підвищення точності дані були розділені на три підгрупи відповідно до значень δ , для яких побудовано окремі формули, що досягли похибки не більше 9%.

Особливістю методу є можливість отримання саме аналітичного вигляду функцій, що, на відміну від «чорних скринь» типу нейронних мереж, дозволяє не тільки прогнозувати, а й аналізувати фізичний зміст процесів. При цьому, формули не потребують великої кількості параметрів або ітеративних процедур, як це часто буває у напівемпіричних підходах.

У випадку $\delta \rightarrow 0$ та $l \rightarrow 0$, тобто у вільномуолекулярному режимі, всі моделі, побудовані методом симвосимвольної лінійної регресії, коректно наближаються до теоретично відомої залежності $W = 1 - p$. Це підтверджує фізичну обґрунтованість отриманих виразів.

Таким чином, використання символної регресії для задач цього типу дозволяє суттєво знижити обчислювальні витрати, зберігаючи при цьому точність і фізичну інтерпретованість моделі. Подібний підхід може бути легко адаптований для інших задач кінетичної теорії, таких як моделювання взаємодії частинок у газових сумішах або вивчення ефектів на границях розділу.

Символьна регресія також знайшла своє застосування у вивчені акустичних метаструктур, зокрема, сонічних кристалів. Ці структури здатні контролювати поширення звукових хвиль завдяки своїй періодичній геометрії та матеріальним властивостям. Вони мають широкий спектр застосувань, від шумозаглушення до створення нових акустичних пристрій.

Сонічні кристали — це періодичні акустичні структури, здатні змінювати характеристики поширення звукових хвиль завдяки своєму геометричному або матеріальному профілю. Незважаючи на тридцятичну історію вивчення, кількість точних аналітичних виразів для таких структур обмежена найпростішими випадками, що ускладнює проектування систем із заданими резонансними або забороненими зонами.

Основою задачі є рівняння Вебстера — модифікація хвильового рівняння, що описує поширення квазіплощинних хвиль у хвилеводах із змінним поперечним перерізом. Для періодичних структур застосовується теорія Флоке-Блоха, яка дозволяє обчислити дисперсійну залежність — ключовий параметр для виявлення зон заборонених частот (bandgaps), де хвилі не поширюються.

У дослідженні [15] запропоновано новий підхід до побудови компактних аналітичних формул, що пов'язують геометрію періодичної комірки з характеристиками дисперсії. Замість прямого чисельного розв'язку рівнянь було згенеровано великий набір даних, які проаналізовано за допомогою методів фізично-обґрунтованого машинного навчання, зокрема символної регресії.

Геометрія одиничної комірки хвилеводу параметризована кубічним сплайном із чотирма керуючими точками, що забезпечує гнучкість при збереженні гладкості та симетрії. Основними змінними стали чотири безрозмірні параметри: зміщення і рознесення по координаті x , а також зміщення і різниця по радіусу r .

Отримано формули, які з високою точністю апроксимують центри та ширини перших трьох зон заборони, а також повну дисперсійну залежність. Наприклад, центр першої зони заборони описується виразом:

$$m_1 = \pi + 0.38 \cdot \max(0.17, x_d r_d),$$

а ширина цієї зони:

$$w_1 = 2.61 \cdot [\max(r_m, r_d) - \min(0, r_m)].$$

Ці формули зручні для швидкої оцінки параметрів системи та демонструють узгодженість із фізичними принципами, наприклад, залежність ширини першої зони переважно від об'єму одиничної комірки. Такий підхід спрощує проектування акустичних метаструктур без постійного використання чисельних симуляцій.

Також отримано узагальнене рівняння для передбачення повної дисперсійної залежності:

$$\cos \xi = b_0(\kappa) + b_1(\kappa)c_1 + b_2(\kappa)c_2 + b_3(\kappa)c_3,$$

де $b_i(\kappa)$ — функції хвильового числа, а c_i — функції геометричних параметрів. Така декомпозиція дозволяє відокремити залежність від частоти та геометрії, підвищуючи гнучкість моделі.

Усі отримані формули продемонстрували високу точність на навчальному та тестовому наборах даних. Важливо, що побудовані залежності мають зрозумілу фізичну інтерпретацію. Наприклад, у випадку хвилеводу з постійним радіусом модель повертається до стандартної залежності $\cos \xi = \cos \kappa$, що відповідає неперевному середовищу.

Таким чином, використання символної регресії в задачах акустики дозволяє будувати прозорі й ефективні моделі для складних періодичних структур. Це відкриває перспективи для аналогічних задач у фотоніці, механіці, електродинаміці та інших галузях, де важливими є дисперсійні властивості середовища.

Також символна регресія може бути використана для побудови аналітичних виразів для різних фізичних величин, таких як енергія, імпульс, частота тощо. Це дозволяє не тільки спростити обчислення, але й отримати глибше розуміння фізичних процесів, що відбуваються в системі. У фізиці важких іонних зіткнень струмені (джети), що виникають унаслідок жорсткого розсіювання кварків і глюонів, є чутливими зондами до властивостей кварк-глюонної плазми. Проте точність таких вимірювань обмежується флуктуаційним фоном, утвореним м'якими частинками з гідродинамічним походженням. Традиційні методи віднімання фону, зокрема area-based підхід, демонструють високу варіативність результатів, особливо в області малих імпульсів p_T .

У дослідженні [16] вперше застосовано інтерпретоване машинне навчання для реконструкції струменевого імпульсу, зокрема поєднання глибокої нейронної мережі та символної регресії. Після навчання нейромережі на даних PYTHIA + TENNGEN, автори використали PySR для отримання аналітичного виразу, що апроксимує поведінку нейромережі:

$$p_{T,\text{corr}} = p_{T,\text{tot}} - C_1 \cdot (N_{\text{tot}} - C_2),$$

де $p_{T,\text{tot}}$ — повний імпульс джета, N_{tot} — кількість частинок у струмені, а C_1, C_2 — коефіцієнти, підібрані PySR. Отримана формула тісно пов'язана з **мультиплікативним методом**, де корекція імпульсу базується на числі частинок, а не площині струменя.

На відміну від area-методу:

$$p_{T,\text{corr}} = p_{T,\text{tot}} - \rho A,$$

де ρ — густина імпульсу, A — площа струменя, новий підхід значно знижує вплив флуктуацій, спричинених гідродинамічними ефектами. Формула стандартного відхилення залишки імпульсу δp_T показує, що останні два члени усуваються при множинному підході:

$$\sigma_{\delta p_T} = \sqrt{N \cdot \sigma_{p_T}^2 + \text{флуктуації від } v_n},$$

де N — кількість частинок, v_n — коефіцієнти азимутальної анізотропії.

Застосування символної регресії дозволило зменшити модельну залежність від нейромережі та надати отриманим виразам фізичну інтерпретацію. Коефіцієнти C_1 та C_2 узгоджуються з параметрами ρ_{Mult} та N_{signal} з мультиплікативного методу, особливо для колізій Au+Au на RHIC.

Цей підхід, що поєднує глибокого навчання із прозорістю символного моделювання, демонструє, як машинне навчання може розширити класичні методи фізичного аналізу, зберігаючи інтерпретованість результатів.

У роботі було продемонстровано, що символна регресія може бути ефективно використана для побудови аналітичних виразів у задачах, пов'язаних із струменями в важких іонних зіткненнях. Це відкриває нові можливості для аналізу даних та розуміння фізичних процесів, що відбуваються в кварк-глюонній плазмі.

Вивчення зв'язку між масою надмасивної чорної діри (SMBH) та властивостями її галактики-хазяїна залишається ключовим напрямком сучасної астрофізики. Протягом останніх десятиліть науковці аналізують так звані *масштабні співвідношення* — емпіричні залежності, які дозволяють оцінити масу чорної діри на основі доступних параметрів, таких як світність або швидкість обертання галактики. Проте більшість таких залежностей базується на еліптичних галактиках, що обмежує точність екстраполяції до спіральних галактик, де очікується наявність проміжних за масою чорних дір (IMBHs).

У роботі Davis & Jin (2023) [17] запропоновано нову трипараметричну залежність маси чорної діри від кутового параметра спіралі ϕ (pitch angle) та максимальної швидкості обертання диска v_{\max} . Автори використали ці параметри для побудови *фундаментальної площини* — узагальненого співвідношення, яке значно зменшує статистичне розсіювання.

Залежність була побудована за допомогою символної регресії (пакет PySR), яка автоматично шукала математичний вираз, що найкраще описує дані. Для уточнення коефіцієнтів застосовано метод Hyper-Fit. Остаточна формула для логарифма маси чорної діри M_{\bullet} (у сонячних масах) має вигляд:

$$M_{\bullet} = \alpha \cdot (\tan |\phi| - 0.24) + \beta \cdot \log \left(\frac{v_{\max}}{211 \text{ км/с}} \right) + \gamma,$$

де $\alpha = -5.58 \pm 0.06$, $\beta = 3.96 \pm 0.06$, $\gamma = 7.33 \pm 0.05$, а внутрішнє розсіювання $\sigma = 0.22 \pm 0.06 \text{ dex}$ — одне з найменших серед відомих масштабних співвідношень.

Фізичне тло цієї залежності узгоджується з попередніми спостереженнями: масивні чорні діри трапляються в галактиках із щільно закрученими спіральними рукавами (малий ϕ) та високою швидкістю обертання, тоді як легкі — у галактиках із розгорнутими рукавами та повільним обертанням. Це підкреслює взаємодію між SMBH та глобальною структурою галактики, на противагу традиційним підходам, які акцентувалися на центральних властивостях.

Модель дозволяє визначити умови, за яких маса чорної діри ймовірно менша за $10^5 M_\odot$: $|\phi| > 26.8^\circ$ та $v_{\max} < 130 \text{ км/с}$. Це відкриває можливості для пошуку IMBH на основі архівних даних.

Окрім цього, нове співвідношення є практичним: pitch angle легко оцінити навіть з немаркованих зображень, а дані про v_{\max} доступні в численних базах радіоспостережень. Це робить модель універсальним інструментом для великих оглядів, зокрема на високих червоних зміщеннях.

Отримане тривимірне співвідношення між M_\bullet , ϕ та v_{\max} значно випереджає попередні моделі за точністю та інтерпретованістю. Воно відкриває нові можливості в дослідженні еволюції галактик, статистики чорних дір і пошуку IMBH.

2. Методика обчислень

2.1. Створення тренувального та тестового наборів

Метою цієї роботи було створення моделей для оцінки рухливості носіїв заряду у кремнії при певних температурі T та рівні легування N_D з використанням алгоритмів машинного навчання (МН). Іншими словами, з точки зору МН, ми вирішували задачу у якій набір ознак (дескрипторів) складався з двох величин (T, N_D), а цільовим результатом була величина μ . Зрозуміло, що для досягнення поставленої мети насамперед необхідні дані і для їх отримання ми використовували модель Klaassen. Тобто, передбачення моделі Klaassen вважалися істинними і ми намагалися створити регресійні моделі, передбачення яких максимально узгоджуються з підходом, запропонованим у роботі [10].

У роботі ми розглядали наступні діапазони температур та рівнів легування:

$$T \in [200, 500] \text{ K}; \quad N_D \in [10^{13}, 10^{19}] \text{ см}^{-3}. \quad (2.1)$$

Під час створення тренувального набору даних випадковим чином вибиралися 500 пар значень (T, N_D) з діапазону (2.1) і обчислювалися значення μ , використовуючи вирази (1.12)-(1.20) та дані таблиць 4 та 5. Тестовий набір, який використовувався для оцінки якості передбачень моделей, складався з 2550 зразків, що відповідали всім комбінаціям 51 значень температури (від 200 К до 500 К з кроком 6 К) та 50 значень рівня легування, рівномірно розподіленим в логарифмічному масштабі від 10^{13} см^{-3} до 10^{19} см^{-3} . У роботі розроблялися моделі, здатні передбачати рухливість як основних так і неосновних електронів та дірок у кремнії і для кожного з цих чотирьох випадків створювалися окремі тренувальний та тестові набори. Надалі рухливість електронів у n-Si позначатиметься $\mu_{n,n}$, у p-Si — $\mu_{n,p}$. Аналогічно, $\mu_{p,p}$ та $\mu_{p,n}$ позначатимуть рухливості дірок, коли вони є основними та неосновними носіями заряду, відповідно.

Під час розрахунків величини рухливості вважалося, що: 1) у кристалі наявні легуючі домішки лише одного типу (випадок компенсованого напівпровідника не розглядався); 2) концентрація нерівноважних (інжеекторованих) носіїв набагато менша концентрації електронів у n-Si та дірок у p-Si.

У випадку електронного напівпровідника для обчислення концентрацій носіїв заряду використовувався вираз

$$n = \frac{N_d^+}{2} + \sqrt{\left(\frac{N_d^+}{2}\right)^2 + n_i^2}, \quad (2.2)$$

$$p = \frac{n_i^2}{n}, \quad (2.3)$$

де n_i — концентрація електронів у власному напівпровіднику, для обчислення температурної залежності якої, використовувався підхід з роботи [18]. В свою чергу, кількість іонізованих донорів у одиниці об'єму обчислювалася з урахуванням статистики Фермі–Дірака:

$$N_d^+ = N_d \left(1 - \frac{1}{1 + 0,5 \exp \left[\frac{E_F - E_d}{kT} \right]} \right), \quad (2.4)$$

де E_F — положення рівня Фермі, E_d — енергія розташування донорного рівня. Оцінка розташування рівня Фермі здійснювалася з використанням умови електронейтральності

$$n = p + N_d^+, \quad (2.5)$$

при цьому вважалося, що

$$n = N_C F_{1/2} \left(-\frac{E_F}{kT} \right), \quad (2.6)$$

$$p = N_V F_{1/2} \left(-\frac{E_G - E_F}{kT} \right), \quad (2.7)$$

де N_C та N_V — ефективні густини станів біля дна зони провідності та вершини валентної зони, обчислювалися відповідно до [18]; E_G — ширина забороненої зони, яка розраховувалася згідно з наближенням Passler [19]; $F_{1/2}$ — інтеграл Фермі–Дірака ступеня 1/2, для обчислення якого використовувався підхід [20].

У випадку p-Si розрахунки проводилися аналогічно.

І для електронного, і для діркового напівпровідника вважалося, що енергія активації легуючої домішки складає 45 меВ. Ця величина відповідає енергіям іонізації атомів бору та фосфору, які є основними легуючими домішками для кремнію.

Результати розрахунків істинних значень рухливості представлені на Рис. 2.1.

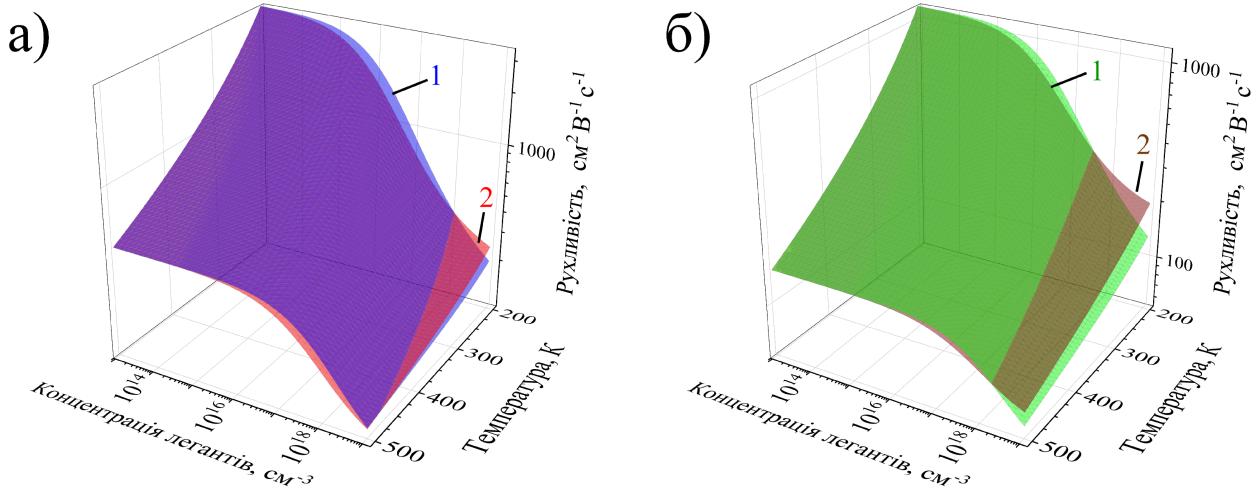


Рис. 2.1. Залежності рухливості електронів (а) та дірок (б) у випадку, коли вони є основними (поверхні 1) та неосновними (поверхні 2) носіями заряду від температури та концентрації легуючої домішки. Розрахунки виконані відповідно до теорії Klaassen.

2.2. Налаштування Symbolic Regression

Символьна регресія (Symbolic Regression, SR) — це сучасний алгоритм машинного навчання, який апроксимує залежності між вхідними та вихідними параметрами, дозволяючи отримати аналітичний зв'язок у вигляді математичної формули [21]. SR використовує еволюційні алгоритми для побудови функціональної залежності та пошуку числових параметрів. Цей метод поєднує дані, моделі машинного навчання та наукові теорії, забезпечуючи інтерпретованість, обчислювальну ефективність і легку інтеграцію у вирішення фізичних задач.

Для реалізації SR використовувався open source пакет PySR, орієнтований на Python. Для побудови рівнянь застосовувалися стандартні математичні функції ($+$, $-$, \times , \div , \exp , \ln , x^y , \tanh , \sin) та дійсні константи. Крім того, були додані спеціально сконструйовані оператори:

- PowInv: функція, що обчислює обернену залежність $1/(1+x)$.
 - Zved: функція, що обчислює середнє гармонійне двох чисел x і y , тобто $\frac{x \cdot y}{x+y}$.
- Щоб уникнути складних і важко інтерпретованих виразів, застосовувалися обмеження:
- Заборона вкладених операцій $\sin(\sin(\bullet))$, PowInv, Zved, tanh, а також вкладеності tanh в PowInv та Zved.
 - Заборона вкладеності \ln і x^y в \exp .
 - Обмеження складності показника експоненти до 3 та показника функції x^y до 5.

Попередня обробка даних полягала у нормалізації, тобто на вхід SR моделі подавалися нормовані величини T_n і N_{Dn} , пов'язані з реальними температурою T та концентрацією легуючої домішки N_D наступним чином:

$$T_n = \frac{T}{300\text{ K}}, \quad (2.8a)$$

$$N_{Dn} = \frac{N_D}{10^{17} \text{ cm}^{-3}}. \quad (2.8b)$$

Цей підхід забезпечує стабільність і точність моделі при роботі з широкими діапазонами значень.

Ці обмеження дозволяють моделі працювати більш ефективно, підтримуючи баланс між складністю та точністю, що є особливо важливим при роботі з великими обсягами даних або складними математичними моделями.

2.3. Класичні регресійні моделі

Окрім символьної регресії та випадкового лісу, в роботі для оцінки рухливості також були реалізовані декілька моделей, які базуються на використанні більш класичних алгоритмів МН. А саме, були використані: *Gradient Boosting* (GB) — об'єднане декілька слабких моделей, як правило, дерев рішень, з метою покращення ефективності прогнозування; кожна нова модель виправляє помилки попередньої, тим самим підвищуючи загальну точність; остаточний прогноз отримують шляхом агрегування прогнозів, зазвичай за допомогою зваженої суми [22];

Support Vector Regression(SVR) — передбачає знаходження функції, яка наближує залежність між вхідними змінними та цільовими значеннями з максимально допустимою похибкою, водночас забезпечуючи якнаймен-

Табл. 6. Пошуковий простір гіперпараметрів для RF моделей

Гіперпараметр	Можливі значення (діапазон пошуку)
# estimators	100 \div 700
max depth	10 \div 70
min samples leaf	1, 2, 3, 4, 5
min samples split	2, 3, 4, 5
bootstrap	True, False
max features	0,5, 0,8, 1

шу складність моделі; може використовувати різні функції ядра (наприклад, радіально-базисні чи поліноміальні), щоб працювати з нелінійними залежностями [23];

Deep Neural Network (DNN) — складається з декількох шарів взаємопов'язаних нейронів, які обробляють вхідні дані шляхом послідовних нелінійних перетворень [24].

Підкреслимо, що хоча подібні моделі також дозволяють передбачати певні величини (наприклад, рухливість носіїв, як це відбувалося в нашому випадку), проте вони є фактично чорною скринькою, що суттєво утруднює обґрунтування отриманих результатів і поступово перестає задовольняти дослідників.

Вказані моделі були імплементовані на мові Python, використовуючи пакети Keras (для DNN) та Scikit-learn (для RF, GB та SVR).

Розглядалися два варіанти кожного типу моделей, які відрізнялися попередньою обробкою даних. В одному випадку як вхідні ознаки використовувалися нормовані значення температури та концентрації легуючої домішки (визначені відповідно до (2.8)), а прогнози були орієнтовані на отримання величини μ . Відповідні моделі надалі позначатимуться шляхом додавання до абревіатури використаного алгоритму суфіксу “_N” (RF_N, DNN_N тощо) Проте ознаки достатньо сильно відрізняються за величиною між собою, а N_D ще й змінюється у межах шести порядків, що може бути перепоною для точних прогнозів. Тому для другого типу моделей проводилася попередня обробка вхідних ознак, яка передбачала нормалізацію (Z-стандартизацію, вже вибачте за усталений термін) як значень температури T , так і логарифмів ступеню легування $\lg N_D$ (здійснювалися лінійні перетворення, в результаті яких тренувальні набори кожної з ознак характеризувалися нульовим середнім і одиничним стандартним відхиленням). Моделі були орієнтовані на передбачення нормалізованих значень $\lg \mu$. Процес нормалізації був реалізований за допомогою функції StandardScaler з пакету Scikit-learn і позначення відповідних моделей міститиме суфікс “_S”.

Відомо, що налаштування гіперпараметрів є надзвичайно важливим для оптимізації продуктивності моделей [25]. Перелік гіперпараметрів, які налаштовувалися для кожної з моделей, а також діапазони пошуку вказані в Таблицях 6-9.

Табл. 7. Пошуковий простір гіперпараметрів для GB моделей

Гіперпараметр	Можливі значення (діапазон пошуку)
# estimators	200 \div 900
max depth	2 \div 40
min samples leaf	1, 2, 3, 4, 5, 6, 7
min samples split	2, 3, 4, 5, 6, 7
learning rate	$10^{-3} \div 1$
max features	0,5, 0,6, 0,7, 0,8, 0,9, 1

Табл. 8. Пошуковий простір гіперпараметрів для SVR моделей

Гіперпараметр	Можливі значення (діапазон пошуку)
kernel	linear, poly, rbf, sigmoid
degree*	2, 3, 4, 5, 6
C_0	0 – 8
Tolerance	$10^{-5} \div 10^{-2}$
C	$10^{-2} \div 15$
ε	$10^{-4} \div 0,5$

* липше для kernel=poly

Зазначимо, що у випадку DNN вхідний шар складався з двох вузлів, вихідний містив один вузол, а структура схованих повнозв'язних шарів наведена на Рис. 2.2. Як видно з рисунку, використовувалося шість схованих шарів, кількість вузлів у яких поступово зменшувалося: від шару до шару на 10% від кількості вузлів у першому схованому N_{nl}), яка була одним із гіперпараметрів, що налаштовувалися.

Табл. 9. Пошуковий простір гіперпараметрів для DNN моделей

Гіперпараметр	Можливі значення (діапазон пошуку)
# nodes for first hidden layer	5 ÷ 150
batch size	1, 4, 16
activation function	ReLU, sigmoid, tanh, SELU, ELU
optimizer	SGD, RMSprop, Adam, Adadelta, Adagrad, Adamax, Nadam, Ftrl
learning rate	$10^{-4} \div 10^{-2}$
weight initializer	Xavier Normal, Xavier Uniform, He Normal, He Uniform, Random Normal, Random Uniform

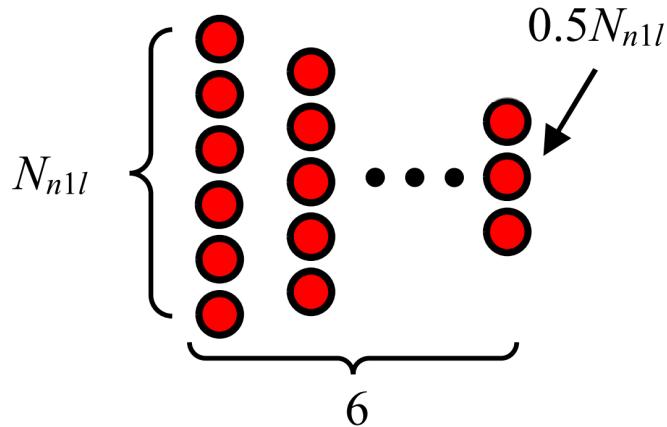


Рис. 2.2. Конфігурація схованих шарів для використаних DNN моделей.

Налаштування моделей здійснювалося за допомогою пакету Optuna [26] з використанням TPE самплера та Hyperband прюнера для ефективного вибору гіперпараметрів. Під час налаштування використовувалась 5-кратна крос-валідація, яка передбачала почергове використання кожних 20% тренувального набору як валідаційних даних для оцінки ефективності моделей, навчених з використанням решти 80%.

2.4. Метрики оцінювання

При побудові регресійних моделей надзвичайно важливо використовувати різні метрики, що дозволяють оцінити якість навчання та точність прогнозів. В роботі були використані декілька кількісних показників для оцінювання створених моделей. Одним з них була середня квадратична похибка (the mean squared error, MSE):

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2, \quad (2.9)$$

де \hat{y}_i — прогнозоване моделлю значення для i -го зразка (набору вхідних параметрів), y_i — відоме цільове значення для i -го зразка, N — кількість зразків у наборі даних. У нашому випадку для символної регресії та *_N моделей y_i збігалося зі величиною рухливості, для *_S моделей при обчисленні y_i проводилися логарифмування та стандартизація (див. розділ 2.3). MSE є однією з найбільш популярних метрик і в нашому випадку переважно використовувалося для оцінки якості навчання моделей, зокрема під час налаштування гіперпараметрів та в процесі безпосереднього навчання (середня квадратична похибка розглядалася як цільова функція для мінімізації під час навчання). Певним виключенням був лише випадок SR, коли під час навчання мінімізувалося середнє значення зваженої суми, причому вагові коефіцієнти мали вигляд

$$\omega_i = \frac{\log N_{D,i}}{\mu_i}. \quad (2.10)$$

Використання ω_i дозволяло врахувати, що рухливість у вибраних діапазонах температури та ступеня легування змінюються у достатньо широких межах, і перешкоджало тому, щоб остаточний аналітичний вираз орієнтувався переважно на відтворення великих значень рухливості.

Для оцінки якості прогнозів використовувалися декілька показників. Один з них — середня відносна похибка (mean absolute percentage error, MAPE):

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|\mu_i^{\text{PRED}} - \mu_i^{\text{TRUE}}|}{\mu_i^{\text{TRUE}}} \times 100\%, \quad (2.11)$$

де μ_i^{PRED} — передбачене значення рухливості (після застосування операцій, обернених до нормування та нормалізації), μ_i^{TRUE} — істинне значення рухливості (отримане в результаті розрахунків відповідно до теорії Klaassen). У випадку, коли частина прогнозів за своєю точністю суттєво відрізняється від інших, MAPE не дозволяє отримати повну картину щодо акуратності передбачень. Тому ми також використовували медіанне значення відносної похибки APE_{MED} , яке відображає величину помилки не більше якої мають 50% передбачень у наборі даних. Водночас для характеризації максимального відхилення від істини ми використовували максимальне значення відносної похибки APE_{MAX} ,

Нарешті, ми розглядали середнє значення абсолютної похибки (mean absolute error, MAE):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\mu_i^{\text{PRED}} - \mu_i^{\text{TRUE}}|. \quad (2.12)$$

Ця метрика дозволяє оцінити середню величину абсолютної похибки між прогнозованими та істинними значеннями рухливості. Вона є менш чутливою до викидів, ніж MAPE, і може бути корисною для оцінки точності моделі в умовах, коли дані містять аномальні значення. Всі вказані метрики використовувалися для оцінки якості навчання моделей, а також для порівняння між собою різних моделей.

3. Отримані результати

3.1. Символьна регресія

У випадку опису рухливості електронів, коли вони є основними носіями у монокристалічному кремнії в результаті застосування SR був отриманий наступний вираз:

$$\mu_{n,n} = \frac{N_{Dn} + 1413,2}{T_n^{2,2508} + \frac{a_{n,n} \cdot b_{n,n}}{a_{n,n} + b_{n,n}}}, \quad (3.1a)$$

$$a_{n,n} = \left(\frac{N_{Dn}}{0,873T_n + 0,0727} \right)^{0,7212}, \quad (3.1b)$$

$$b_{n,n} = 0,215N_{Dn}^{0,58} + 10,5T_n - 1,342. \quad (3.1b)$$

Нагадаємо, що нормовані величини температури та концентрації донорів описуються співвідношеннями (2.8). Звертає на себе увагу наявність складової, яка містить $T_n^{2,2508}$, що надзвичайно близько до температурної залежності рухливості, пов'язаної з розсіянням на фононах. Загалом, вираз містить всього 9 параметрів і є набагато простішим ніж той, який пропонується в теорії Klaasen. Розподіл похибки формули (3.1) порівняно з теорією Klaasen залежно від температури та ступеню легування представлений на Рис. 3.1,а. Як видно з представлених даних, найбільші відхилення спостерігаються при $T = 200$ К та N_d в околі 10^{18} см⁻³. Проте навіть в цьому випадку похибка не перевищує 1.2 %. Для більшості ж значень аргументів у тих діапазонах, що розглядалися абсолютна відносна похибка не перевищує 0,15 %.

Наступна формула (3.2) ілюструє отриманий аналітичний вираз для опису рухливості неосновних електронів у монокристалічному кремнії. Розподіл АРЕ як функції температури та концентрації акцепторів представлений на Рис. 3.1,б. В цьому випадку максимальні розбіжності між (3.2) та теорією (1.12)-(1.20) спостерігаються при концентрації легуючої домішки 10^{19} см⁻³ та температурі близько 500 К. Інші “проблемні” області це $T = 230 - 270$ К і $N_a = (3 - 6) \cdot 10^{17}$ см⁻³ і $T \simeq 200$ К і $N_a = 10^{19}$ см⁻³. Проте загалом помилки при оцінці рухливості електронів, коли вони є неосновними носіями заряду, дещо нижчі ніж коли вони є основними.

$$\mu_{n,p} = 26,3 \cdot \left(\frac{N_{Dn}}{N_{Dn} + 192} \right)^{0,62} + \frac{1412,3}{T_n^{2,25} + \frac{a_{n,p} \cdot b_{n,p}}{a_{n,p} + b_{n,p}}}, \quad (3.2a)$$

$$a_{n,p} = 1,92 \cdot \left(\frac{N_{Dn}}{T_n + 0,071} \right)^{0,717}, \quad (3.2b)$$

$$b_{n,p} = 5,7N_{Dn}^{0,103} \cdot T_n, \quad (3.2b)$$

Вираз для опису рухливості дірок у p-Si (формула (3.3)) містить найменшу кількість параметрів з усіх отриманих виразів. Проте і максимальні похибки в цьому випадку найбільші, хоча вони спостерігаються у достатньо вузькому діапазоні параметрів — див. Рис. 3.1,в.

$$\mu_{p,p} = N_{Dn}^{0,448} + \frac{470}{T_n^{2,2505} + \frac{a_{p,p} \cdot b_{p,p}}{a_{p,p} + b_{p,p}}}, \quad (3.3a)$$

$$a_{p,p} = 2,652N_{Dn}^{0,1669} \cdot T_n, \quad (3.3b)$$

$$b_{p,p} = 0,578N_{Dn}^{0,7587} \cdot T_n^{-0,755}, \quad (3.3b)$$

У випадку неосновних дірок вдалося досягти найбільшого збігу значень, що обчислюються відповідно до отриманої формули (3.4) та теорії Klaassen — див. Рис. 3.1,г. Зокрема з рисунку видно, що для переважної частини значень температури та рівня легування помилка менше 0,08%.

$$\mu_{p,n} = \frac{0,0226N_{Dn} \cdot \ln T_n}{\ln T_n + 0,63} + \frac{113,56}{\left[0,15617T_n^{2,94} + \left(\frac{a_{p,n} \cdot b_{p,n}}{1,151a_{p,n} + 4,21b_{p,n}} \right)^{1,177} \right]^{0,76547}}, \quad (3.4a)$$

$$a_{p,n} = N_{Dn}^{0,6085}, \quad (3.4b)$$

$$b_{p,n} = T_n^{1,143}. \quad (3.4b)$$

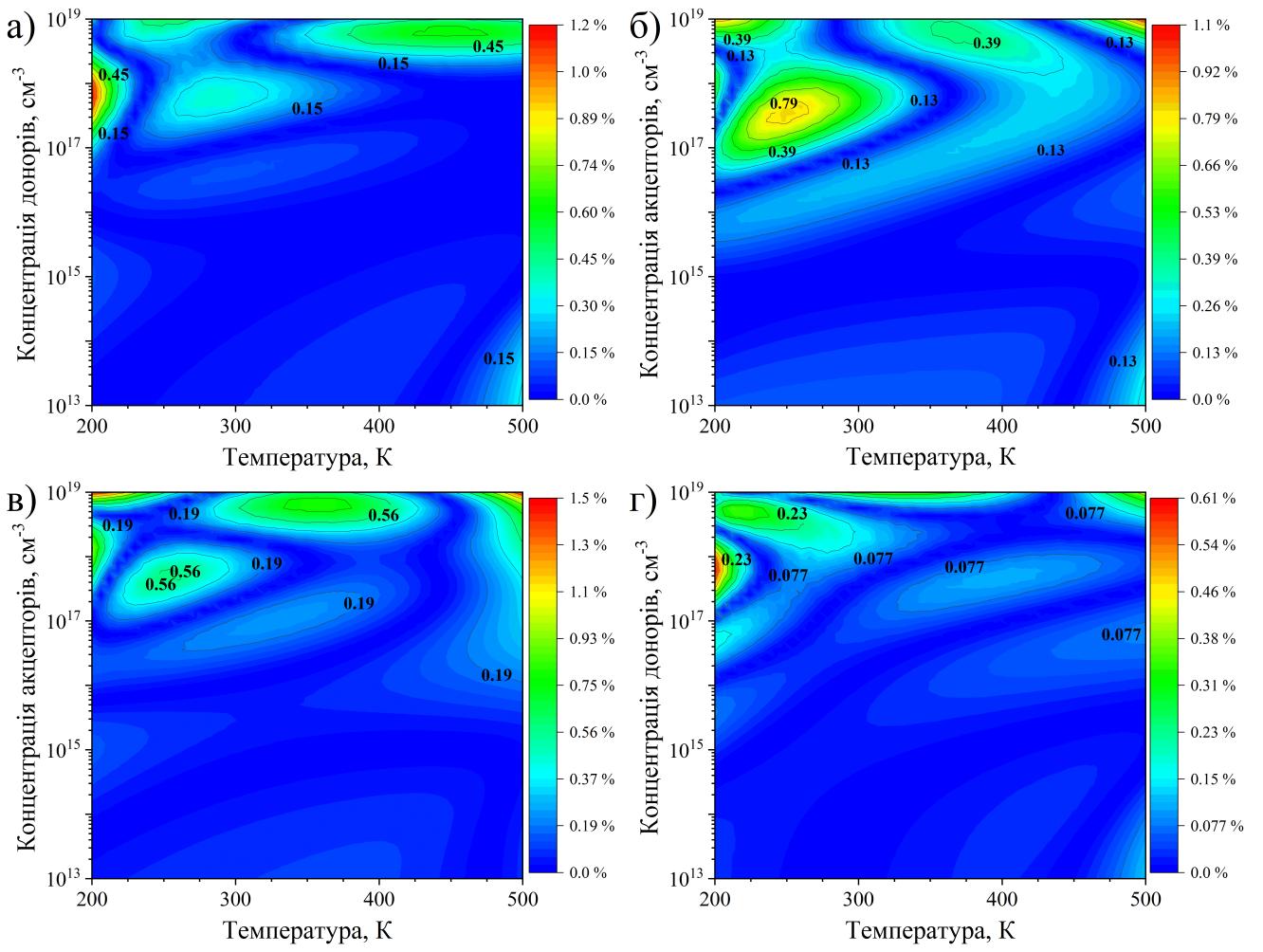


Рис. 3.1. Залежності величини відносної похибки при оцінці рухливості носіїв заряду у кремнії за допомогою виразів (3.1) – (3.4), отриманих за допомогою символьної регресії, від температури та рівня легування. Представлені випадки оцінки рухливості електронів (а, б) та дірок (в, г), коли вони є основними (а, в) та неосновними (б, г) носіями заряду.

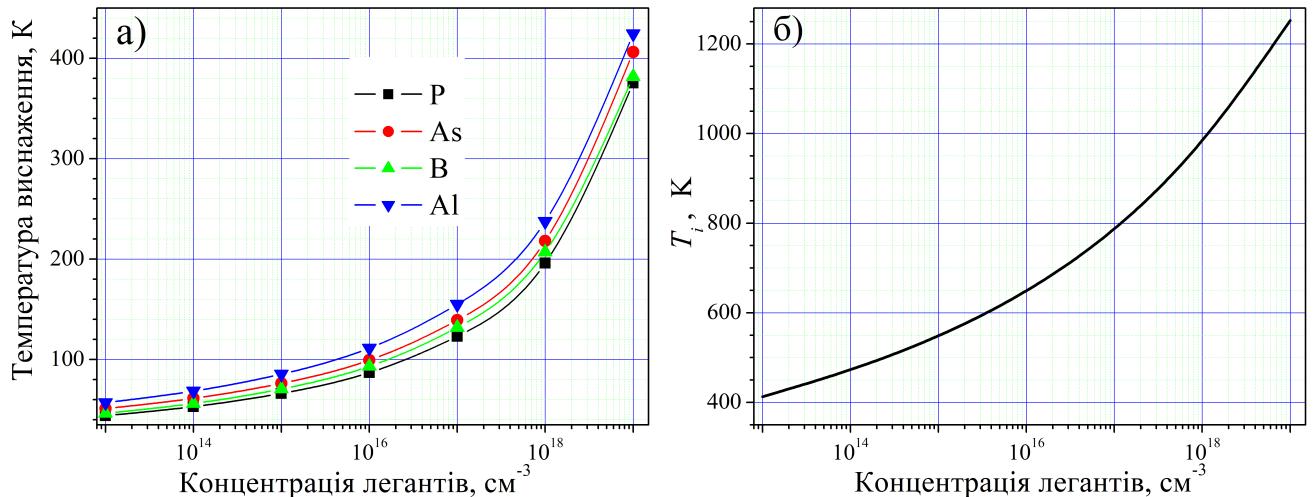


Рис. 3.2. Залежності виснаження (а) та температури початку власної провідності (б) від концентрації легуючих домішок. У частині а наведено результати розрахунків для декількох типів легуючих домішок.

Табл. 10. Використані значення гіперпараметрів для RF моделей

Модель	RF_N				RF_S			
	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$
Гіперпараметр	Значення							
# estimators	468	500	658	510	278	564	304	204
max depth	19	58	37	19	30	15	55	55
min samples leaf	1	1	1	1	1	1	1	1
min samples split	2	2	2	2	2	2	2	2
bootstrap	True							
max features	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0

Зазначимо, що порівняно з теорією Klaassen запропоновані вирази (3.1)–(3.4) зручніші для використання не лише через свою меншу ускладненість. У формулах (1.12)–(1.20) фігурують концентрації носіїв та іонізованих домішок, тоді як запропоновані в цій роботі вирази дозволяють використовувати концентрацію легуючої домішки, що є паспортною характеристикою матеріалу. Умови $n = N_d$, $N_d^+ = Nd$ (для n-кремнію та $p = N_a$, $N_a^- = Na$ для p-кремнію) виконуються лише для певного температурного діапазону. А саме T має бути достатньо високою, щоб виконувалась умова повної іонізації легуючої домішки, коли рівень Фермі знаходиться на декілька kT близьче до середини забороненої зони ніж енергетичний рівень домішкового атому. Зокрема температура при якій $E_F = E_d$ називається температурою виснаження і може бути визначена наступним чином

$$n - \text{тип} : T_s = \frac{E_d}{k \ln(N_C/N_d)}; \quad p - \text{тип} : T_s = \frac{E_d}{k \ln(N_V/N_a)}. \quad (3.5)$$

На Рис. 3.2,а представлено залежності T_i від рівня легування для різних типових легантів. Як видно з представлених даних, при низьких рівнях легування температура виснаження не потрапляє у діапазон температур, для якого знаходилися аналітичні вирази. Проте при $N_D > 10^{18} \text{ см}^{-3}$ умова повної іонізації виконується не завжди.

З іншого боку, температура має бути не занадто високою для того, щоб концентрацією власних носіїв заряду можна було знехтувати у порівнянні з кількістю носіїв, зобов'язаних своїм утворенням легуванню. Температура початку власної провідності оцінюється за допомогою наступного співвідношення:

$$T_i = \frac{E_G}{k \ln \left(\frac{N_C N_V}{N_D^2} \right)}. \quad (3.6)$$

Відповідна концентраційна залежність наведена на Рис. 3.2,б. В цьому випадку при $N_D < 5 \cdot 10^{13} \text{ см}^{-3}$ в діапазоні $200 \div 500 \text{ К}$ може реалізуватися ситуація, коли наявність вільних носіїв заряду не пов'язана лише з легантами.

При $T < T_s$ або $T > T_i$ у випадку застосування виразів (1.12)–(1.20) необхідно додатково враховувати відмінності n та N_d^+ від N_d (p та N_a^- від N_a). У формулах (3.1)–(3.4) подібні ефекти вже враховані. Звичайно, так як при розрахунках використовувалося значення енергії іонізації домішки 0,045 еВ, то отримані вирази мають певне обмеження до застосування лише у випадках саме такого розташування енергетичного рівня леганта відносно дозволених зон напівпровідника. Проте враховуючи, що саме таку енергетичну структуру мають найпопулярніші під час забезпечення домішкової провідності у кремнії атоми бору та фосфору, обмеження не є надто жорстким. Крім того, величина енергії іонізації відіграватиме роль лише у випадку неповної іонізації леганта. Тобто при інших температурах (а при $N_D < 10^{18} \text{ см}^{-3}$ практично завжди) запропоновані формули будуть справедливі і для інших варіантів легування.

3.2. Класичні регресійні моделі

Значення гіперпараметрів, отримані в результаті налаштування і в наступному використані під час навчання та тестування RF, GB, SVR та DNN моделей наведені в Таблицях 10-13.

Результати застосування натренованих моделей до тестового набору ілюстровано Рис. 3.3-3.6. Як правило, відхилення зростають при збільшенні величини рухливості, причому ця тенденція спостерігається незалежно від вибраного алгоритму МН, попередньої підготовки даних чи типу носіїв заряду. Нормалізація даних фактично не впливає на точність прогнозів у випадку використання Random Forest та Gradient Boosting (див. Рис. 3.3 та Рис. 3.4), що, загалом є типовим для деревоподібних алгоритмів. У випадку глибоких нейронних мереж додаткова попередня обробка має сенс, як це можна бачити при порівнянні DNN_S та DNN_N мереж на Рис. 3.6. При застосуванні алгоритму SVR стандартизація даних виглядає абсолютно необхідною — див. Рис. 3.5.

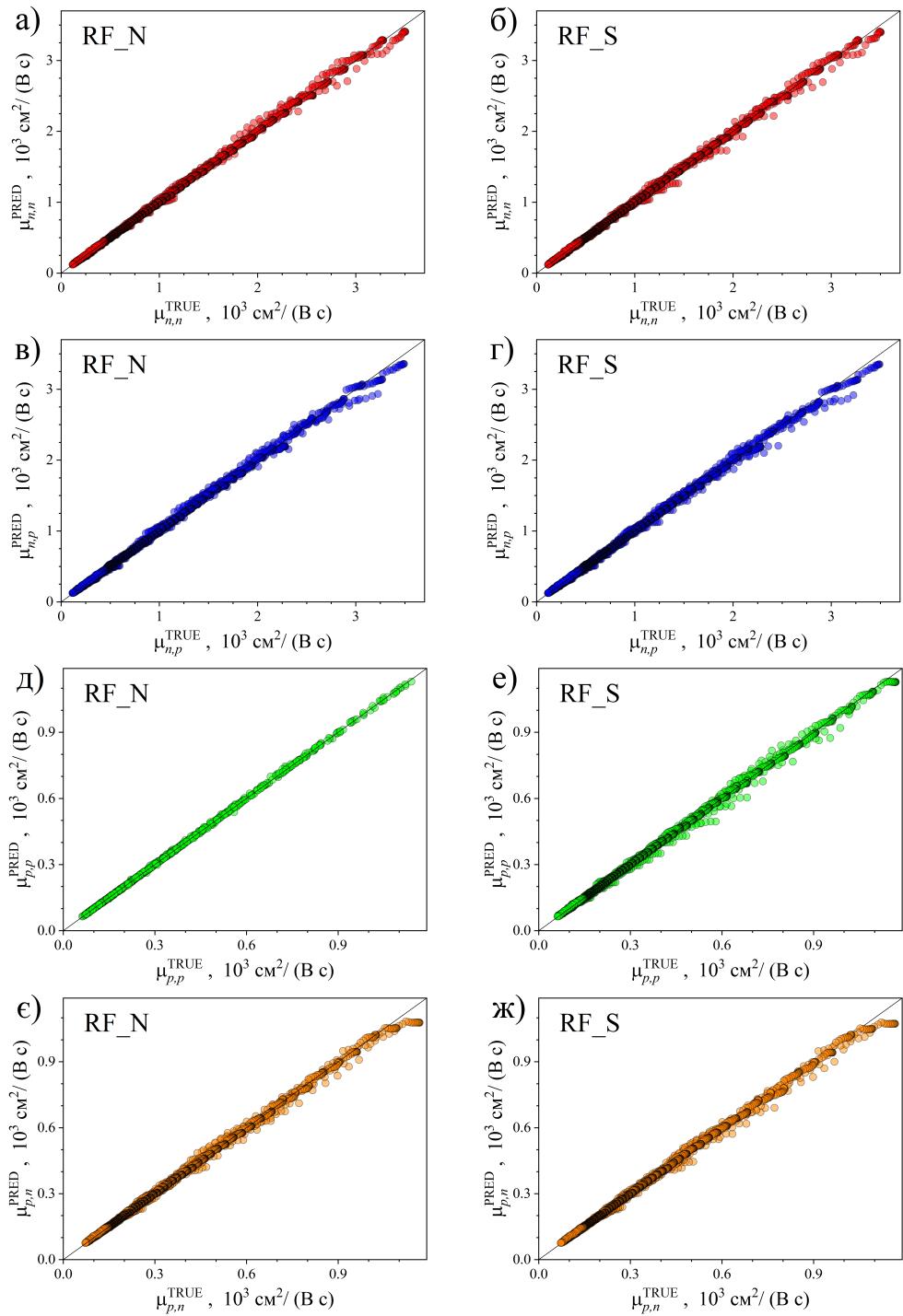


Рис. 3.3. Діаграми розсіювання, що порівнюють еталонні значення рухливості із значеннями, передбаченими RF моделями на тестовому наборі даних. Представлені випадки оцінки рухливості електронів (а–г) та дірок (д–ж), коли вони є основними (а, б, д, е) та неосновними (в, г, е, ж) носіями заряду. Попередня підготовка даних передбачала нормування (а, в, д, е) або нормалізацію (б, г, е, ж). Чорні прямі — лінії ідентичності для зручності порівняння.

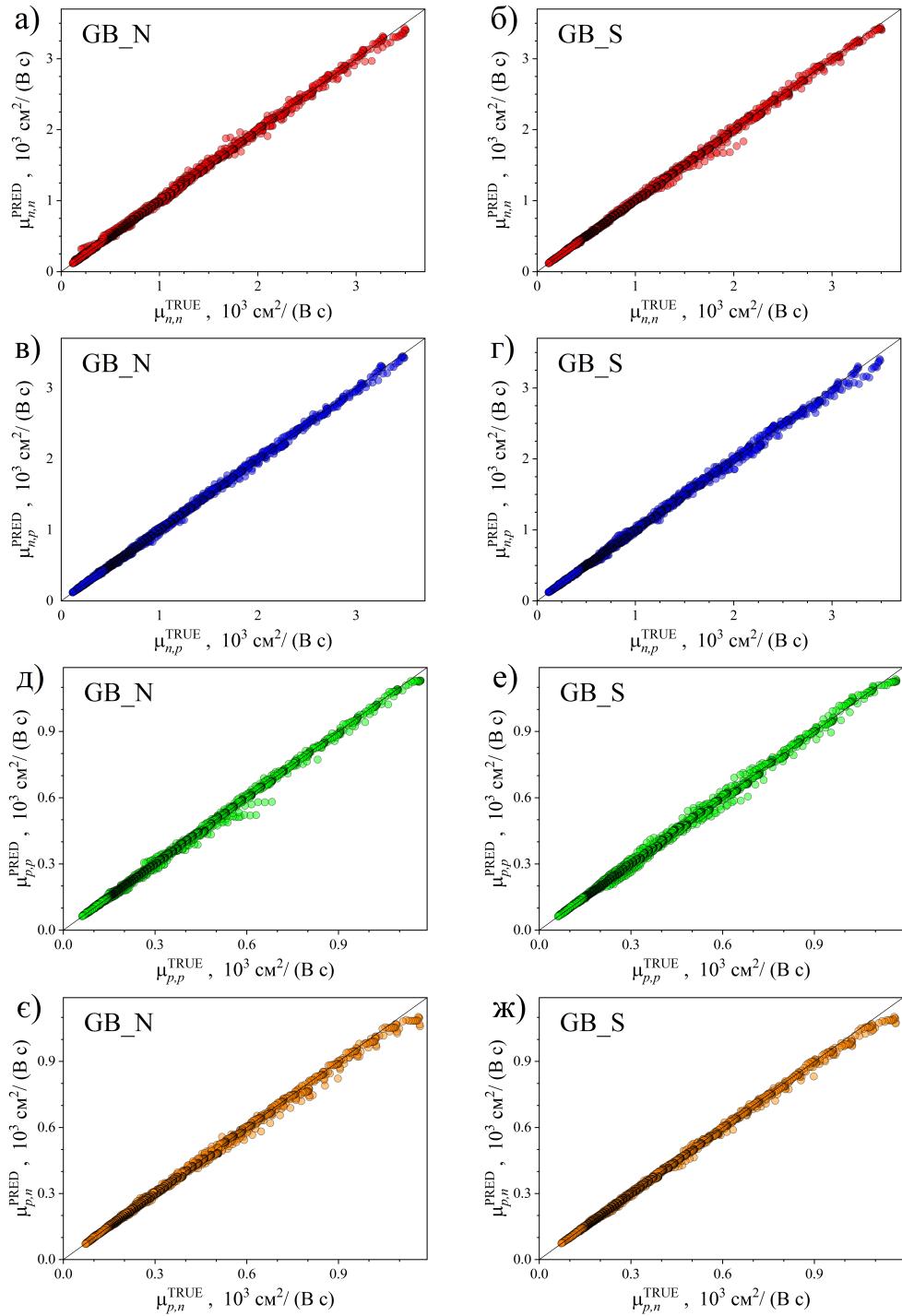


Рис. 3.4. Діаграми розсіювання, що порівнюють еталонні значення рухливості із значеннями, передбаченими GB моделями на тестовому наборі даних. Представлені випадки оцінки рухливості електронів (а–г) та дірок (д–ж), коли вони є основними (а, б, д, е) та неосновними (в, г, е, ж) носіями заряду. Попередня підготовка даних передбачала нормування (а, в, д, е) або нормалізацію (б, г, е, ж). Чорні прямі — лінії ідентичності для зручності порівняння.

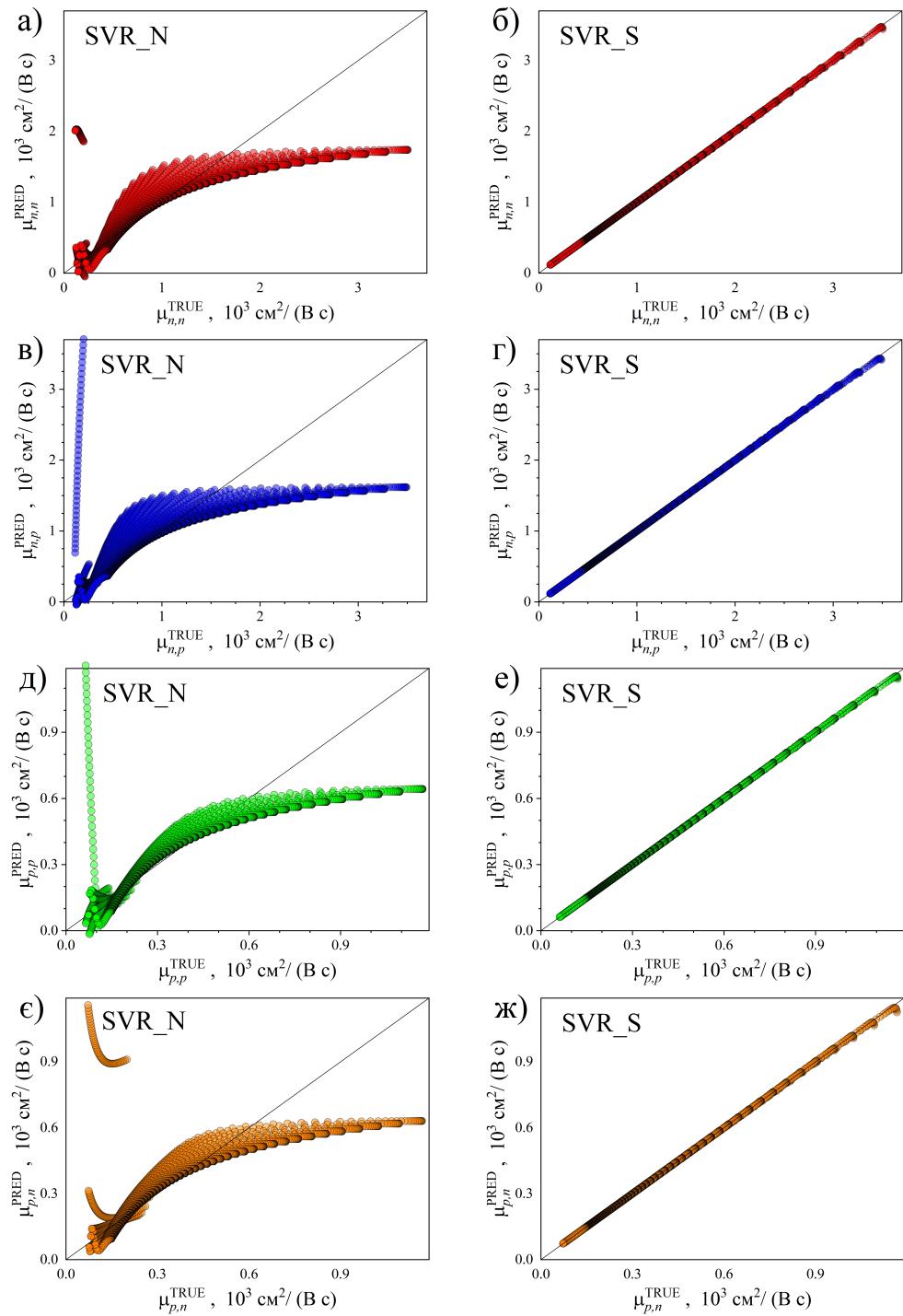


Рис. 3.5. Діаграми розсіювання, що порівнюють еталонні значення рухливості із значеннями, передбаченими SVR моделями на тестовому наборі даних. Представлені випадки оцінки рухливості електронів (а–г) та дірок (д–ж), коли вони є основними (а, б, д, е) та неосновними (в, г, е, ж) носіями заряду. Попередня підготовка даних передбачала нормування (а, в, д, е) або нормалізацію (б, г, е, ж). Чорні прямі — лінії ідентичності для зручності порівняння.

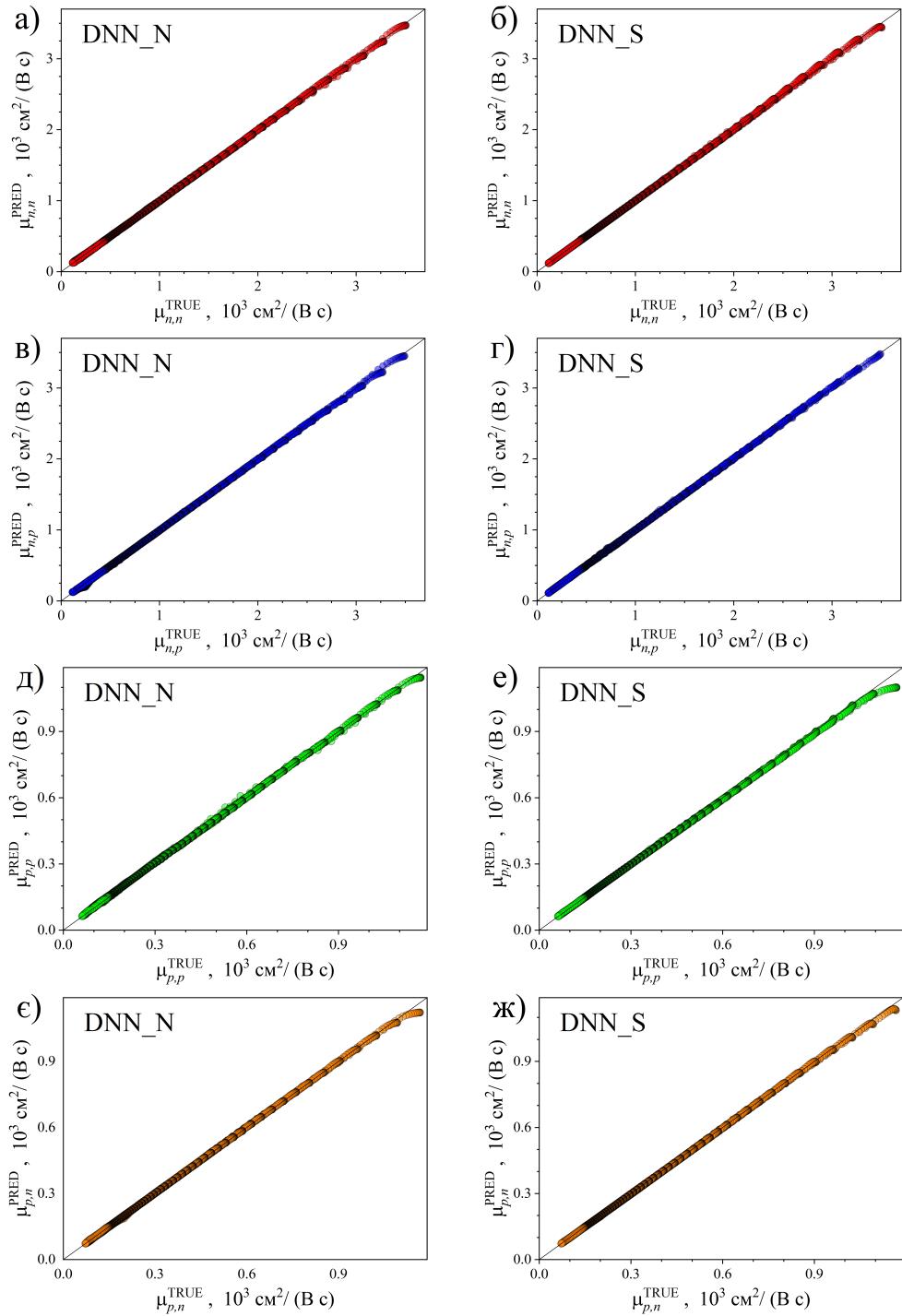


Рис. 3.6. Діаграми розсіювання, що порівнюють еталонні значення рухливості із значеннями, передбаченими DNN моделями на тестовому наборі даних. Представлені випадки оцінки рухливості електронів (а–г) та дірок (д–ж), коли вони є основними (а, б, д, е) та неосновними (в, г, є, ж) носіями заряду. Попередня підготовка даних передбачала нормування (а, в, д, є) або нормалізацію (б, г, е, ж). Чорні прямі — лінії ідентичності для зручності порівняння.

Табл. 11. Використані значення гіперпараметрів для GB моделей

Модель	GB_N				GB_S			
Цільова величина	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$
Гіперпараметр	Значення							
# estimators	700	800	800	570	840	770	810	850
max depth	4	4	4	4	4	4	2	4
min samples leaf	6	2	5	4	5	3	5	4
min samples split	2	2	3	3	5	5	5	6
learning rate, 10^{-2}	6,44	9,78	7,69	4,68	9,49	4,63	9,67	5,29
max features	0,8	1,0	1,0	0,6	1,0	0,7	1,0	0,8

Табл. 12. Використані значення гіперпараметрів для SVR моделей

Модель	SVR_N				SVP_S			
Цільова величина	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$
Гіперпараметр	Значення							
kernel	poly	poly	poly	poly	rbf	rbf	rbf	rbf
degree	6	6	6	6	—	—	—	—
C_0	7,52	7,54	7,08	7,02	4,92	2,73	3,88	6,78
Tolerance, 10^{-3}	5,64	2,22	0,60	0,015	0,75	0,32	0,083	0,068
C	9,79	7,04	3,00	4,19	15	15	15	15
$\varepsilon, 10^{-3}$	33,2	11,6	10,4	0,91	0,55	0,50	2,44	0,84

Табл. 13. Використані значення гіперпараметрів для DNN моделей

Модель	DNN_N				DNN_S			
Цільова величина	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$
Гіперпараметр	Значення							
# nodes for first hidden layer	25	100	25	75	25	75	100	25
batch size	16							
activation function	SELU	ReLU	SELU	sigmoid	ReLU	sigmoid		
optimizer	Nadam	Adamax			Nadam	Adamax		
learning rate, 10^{-3}	0,58	1,25	8,05	5,17	13,0	4,89	12,8	2,84
weight initializer	Xavier Normal	He Normal	Xavier Normal	He Normal			Xavier Normal	
# epoch	1000							

3.3. Порівняння точності різних алгоритмів машинного навчання

Окрім оцінки точності передбачень величини рухливості, що отримуються при використанні різних моделей, доцільно також порівняти результати теорії Klaassen з іншими, більш спрощеними, варіантами визначення μ . В роботі ми зосередили увагу на теорії Arora, як найбільш популярної теорії, що зустрічається в літературі, провівши розрахунки відповідно до виразів (1.10)–(1.11). На Рис. 3.7,а наведено порівняння $\mu_{n,n}$, обчислених за теорією Klaassen та Arora, а на Рис. 3.7,а — відносні похибки. Також для порівняння на рисунку представлені залежності, отримані при використанні аналітичного виразу, запропонованого за допомогою символичної регресії. Як видно, відхилення у випадку застосування (1.10)–(1.11) набагато більші. Приведені залежності є типовими, для $\mu_{n,p}$, $\mu_{p,p}$ та $\mu_{p,n}$ співвідношення схожі.

Відомо, що теорія Arora не враховує розсіяння носій–носій. Тому ми також провели розрахунки, вважаючи що

$$\frac{1}{\mu^{AD}} = \frac{1}{\mu^A} + \frac{1}{\mu_{cc}}, \quad (3.7)$$

де μ^A та μ_{cc} описуються виразами (1.10) та (1.7), відповідно. Виявилося, що за умов, які розглядалися внесок розсіяння носій–носій надзвичайно малий: максимальні відхилення значень, отриманих відповідно до (3.7) від теорії Arora не перевищують 0,059% для $\mu_{n,n}$, 0,068% для $\mu_{n,p}$, 0,023% для $\mu_{p,p}$ та 0,028% для $\mu_{p,n}$. Тобто, врахування додаткового механізму не дозволяє наблизити значення рухливості, що обчислюються за теорією Arora, до більш точних величин, які можуть бути отримані з використанням теорії Klaasen та запропонованих нами апроксимаційних наближень.

Значення метрик, отримані при розрахунках рухливості різними методами узагальнені в Табл. 14. Загалом, числові дані підтверджують висновки, зроблені раніше з Рис. 3.3–3.6. А саме, доцільність застосування

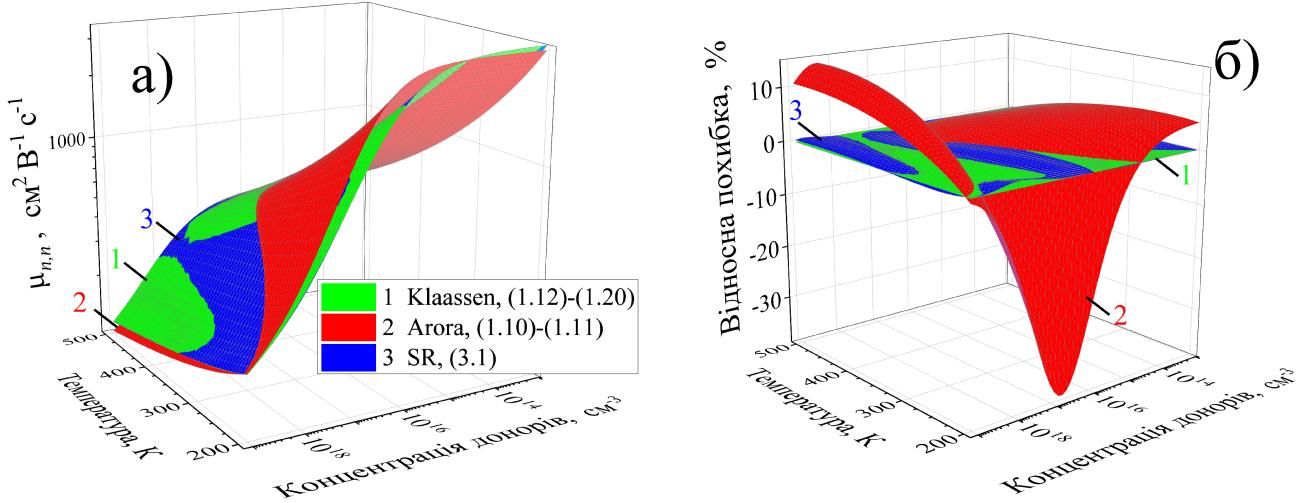


Рис. 3.7. Температурні та концентраційні залежності рухливості основних носіїв заряду у n-Si (а) та помилки розрахунків відносно теорії Klaassen. Обчислення проводилися відповідно до формул (1.12)-(1.20) (поверхня 1), (1.10)-(1.11) (поверхня 2) та (3.1) (поверхня 3).

попередньої стандартизації при використанні SVR та DNN (всі метрики для SVR_S та DNN_S моделей кращі, ніж для SVR_N та DNN_N, відповідно); несуттєва відмінність точностей RF_S і RF_N та GB_S і GB_N моделей.

Як видно з наведених даних, заміна теорії Klaassen більш простими виразами теорії Arora призводить до значно більших помилок ніж використання будь-якої з моделей машинного навчання (за виключенням SVR_N). Серед МН моделей найгірші показники спостерігаються при використанні Gradient Boosting. Трохи кращі кількісні характеристики у випадку використання алгоритму Random Forest, але відмінності від GB досить незначні. Далі у рейтингу точності розташовується модель DNN_N (за виключенням максимальної помилки, яка співрозмірна з показниками попередньо згаданих моделей). Для DNN_S практично всі показники приблизно у 1,5 рази кращі. Найкращі результати продемонстрували SVR_S та SR моделі, причому остання переважає лише за значеннями MAPE та APE_{MED} для кремнію з дірковою провідністю (обидва типи носіїв). Тобто SVR_S має найкращі показники у 4 випадках з 16, решта 12 метрик кращі у SR.

Табл. 14. Метрики відхилень розрахованих значень рухливості у кремнії від теорії Klaassen. Жирним виділено найкращі результати для кожної метрики при передбаченнях рухливості електронів та дірок, коли вони є основними та неосновними носіями заряду

Модель	MAPE, %				APE _{MAX} , %				APE _{MED} , %				MAE, см ² /(B · c)			
	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$	$\mu_{n,n}$	$\mu_{n,p}$	$\mu_{p,p}$	$\mu_{p,n}$
Arora [7]	5,35	20,3	5,62	37,8	38,2	70,7	17,6	111	3,67	16,3	5,01	32,6	45,7	103	16,0	82,6
RF_N	1,29	1,69	1,52	1,47	9,33	14,5	4,60	10,9	0,805	0,994	0,264	0,913	11,0	13,5	4,50	4,76
RF_S	1,42	1,77	1,46	1,41	12,3	14,4	14,5	12,6	0,852	1,04	0,744	0,872	11,6	14,1	4,40	4,76
SVR_N	46,5	51,5	32,6	33,4	1630	1750	2070	1510	14,9	17,4	13,1	12,9	220	244	68,8	73,6
SVR_S	0,108	0,087	0,098	0,082	1,71	2,16	2,26	3,56	0,088	0,064	0,069	0,050	1,20	1,03	0,394	0,410
GB_N	2,20	1,46	1,60	1,55	58,2	10,8	17,3	11,0	1,46	0,939	0,905	1,15	15,43	10,85	4,75	5,24
GB_S	1,32	1,60	1,82	1,23	1,42	11,7	15,0	8,48	0,841	1,104	1,19	0,874	11,3	13,4	5,92	4,19
DNN_N	0,925	0,942	2,01	0,529	9,39	20,0	8,07	6,02	0,621	0,702	1,893	0,383	5,23	5,37	4,22	1,39
DNN_S	0,374	0,623	0,655	0,671	2,23	3,75	5,96	2,97	0,323	0,392	0,556	0,631	3,58	3,49	2,88	1,97
SR	0,094	0,137	0,140	0,050	1,19	0,992	1,48	0,596	0,044	0,077	0,079	0,026	0,475	0,708	0,314	0,115

Наприкінці підсумуємо переваги використання виразів (3.1)–(3.4) порівняно з іншими варіантами оцінки рухливості носіїв заряду. Порівняно з теорією Arora вони дозволяють отримати наближення до теорії Klaassen (яка вважається найбільш точним варіантом опису рухливості) з набагато меншими відхиленнями. Формули (3.1)–(3.4) переважають за точністю регресійні моделі, що використовують алгоритми випадкового лісу, градієнтного бустингу та глибоких нейронних мереж. Порівняно з алгоритмом опорних векторів перевага в точності не настільки велика, але практичне використання SVR моделей (як і інших класичних підходів до вирішення регресійних задач) вимагає як самих натренованих моделей, так і використання відповідного програмного забезпечення. Такий процес не є занадто складним та критичним для реалізації, проте виглядає дешço складнішим за розрахунки за відомою формулою. Нарешті, якщо порівнювати обчислення з використанням формул (3.1)–(3.4) та (1.12)–(1.20), то варто наголосити наступне: 1) втрати точності при застосуванні (3.1)–(3.4) мінімальні: при $T = 200 \div 500$ К та $N_D = 10^{13} \div 10^{19}$ см⁻³ приблизно для 50% можливих комбінацій параметрів похибки менші 0,044%, 0,077%, 0,079% та 0,026% при обчисленні $\mu_{n,n}$, $\mu_{n,p}$, $\mu_{p,p}$ та $\mu_{p,n}$, відповідно; 2) запропоновані вирази містять більше ніж у двічі менше параметрів; 3) при обчисленнях за допомогою (3.1)–(3.4) не потрібно попередньо оцінювати ступінь іонізації легантів; 4) нарешті, вирази просто простіші, хоча це, звичайно, не є критичним при сучасному розвитку комп’ютерної техніки.

Таким чином, незважаючи на певні обмеження у застосуванні (об’ємний кристалічний нескомпенсований кремній легований бором або фосфором (хоча останнє обмеження є суттєвим лише у достатньо вузькому температурному діапазоні), температура $200 \div 500$ К, ступінь легування $10^{13} \div 10^{19}$ см⁻³) запропоновані вирази мають значні перспективи для практичного використання.

Висновки

1. Проведено розрахунки рухливості основних та неосновних носіїв заряду у монокристалічному кремнії відповідно до теорії Klaassen, теорії Agora та теорії Agora з врахуванням розсіяння носій-носій відповідно до теорії Dorkel та порівняно точність отриманих величин.
2. Розроблено та налаштовано моделі машинного навчання для оцінки рухливості електронів та дірок у n-Si та p-Si з використанням алгоритмів випадкового лісу, градієнтного бустинга, опорних векторів, глибоких нейронних мереж та символьної регресії; проведено порівняння точності прогнозів вказаних моделей у порівнянні з теорією Klaassen та показано доцільність застосування підходів Support Vector Regression та Deep Neural Network з попередньою стандартизацією вхідних даних, а також Symbolic Regression.
3. Запропоновано аналітичні вирази для опису рухливості електронів та дірок у монокристалічному кремнії для діапазону температур $200 - 500$ К та діапазону концентрації легуючої домішки (бор або фосфор) $10^{13} - 10^{19}$ см^{-3} ; показано, що середня відносна похибка запропонованих виразів від теорії Klaassen не перевищує 0,14%, а абсолютна — $0,71 \text{ см}^2 / (\text{B} \cdot \text{c})$; водночас вирази містять суттєво меншу кількість параметрів та більш зручні для використання.

Список використаних джерел

- [1] *Seeger, Karlheinz.* Semiconductor Physics. An Introduction / Karlheinz Seeger. Advanced Texts in Physics. — 9 edition. — Springer Berlin, Heidelberg, 2004. — Jun.
- [2] *Fletcher, Neville H.* The High Current Limit for Semiconductor Junction Devices / Neville H. Fletcher // *Proc. IRE.* — 1957. — Vol. 45, no. 6. — Pp. 862–872.
- [3] *Choo, Seok Cheow.* Theory of a forward-biased diffused-junction P-L-N rectifier—Part I: Exact numerical solutions / Seok Cheow Choo // *IEEE Trans. Electron Devices.* — 1972. — Vol. 19, no. 8. — Pp. 954–966.
- [4] *Dorkel, J.M.* Carrier mobilities in silicon semi-empirically related to temperature, doping and injection level / J.M. Dorkel, Ph. Leturcq // *Solid-State Electron.* — 1981. — Vol. 24, no. 9. — Pp. 821–825.
- [5] *Caughey, D.M.* Carrier mobilities in silicon empirically related to doping and field / D.M. Caughey, R.E. Thomas // *Proc. IEEE.* — 1967. — Vol. 55, no. 12. — Pp. 2192–2193.
- [6] *Masetti, G.* Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon / G. Masetti, M. Severi, S. Solmi // *IEEE Trans. Electron Devices.* — 1983. — Vol. 30, no. 7. — Pp. 764–769.
- [7] *Arora, N.D.* Electron and hole mobilities in silicon as a function of concentration and temperature / N.D. Arora, J.R. Hauser, D.J. Roulston // *IEEE Trans. Electron Devices.* — 1982. — Vol. 29, no. 2. — Pp. 292–295.
- [8] *Swirhun, S.E.* Measurement of electron lifetime, electron mobility and band-gap narrowing in heavily doped p-type silicon / S.E. Swirhun, Y.-H. Kwark, R.M. Swanson // 1986 International Electron Devices Meeting. — 1986. — Pp. 24–27.
- [9] *del Alamo, J.* Simultaneous measurement of hole lifetime, hole mobility and bandgap narrowing in heavily doped n-type silicon / J. del Alamo, S. Swirhun, R.M. Swanson // 1985 International Electron Devices Meeting. — 1985. — Pp. 290–293.
- [10] *Klaassen, D.B.M.* A unified mobility model for device simulation — I. Model equations and concentration dependence / D.B.M. Klaassen // *Solid-State Electron.* — 1992. — Jul. — Vol. 35, no. 7. — Pp. 953–959.
- [11] Mobility Modeling Considerations for Radiation Effects Simulations in Silicon / Daniel J. Cummings, Arthur F. Witulski, Hyunwoo Park et al. // *IEEE Trans. Nucl. Sci.* — 2010. — Vol. 57, no. 4. — Pp. 2318–2326.
- [12] Temperature, electric field, and doping dependent mobilities of electrons and holes in semiconductors / S. Noor Mohammad, Andrew V. Bemis, Ronald L. Carter, Robert B. Renbeck // *Solid-State Electron.* — 1993. — Vol. 36, no. 12. — Pp. 1677–1683.
- [13] Computation of flow rates in rarefied gas flow through circular tubes via machine learning techniques / F. Sofos, C. Dritselis, S. Misdanitis et al. // *Microfluidics and Nanofluidics.* — 2023.
- [14] *Cranmer, Miles.* PySR: Fast & Interpretable Symbolic Regression in Python and Julia. — <https://www.pysr.org>. — 2025. — April.
- [15] *Hruška, Viktor.* Analytical formulae for design of one-dimensional sonic crystals with smooth geometry based on symbolic regression / Viktor Hruška, Aneta Furmanová, Michal Bednářík // *Journal of Sound and Vibration.* — 2025. — Vol. 597.
- [16] Interpretable machine learning methods applied to jet background subtraction in heavy-ion collisions / Tanner Mengel, Patrick Steffanic, Charles Hughes, Antonio Carlos Oliveira da Silva // *Physical Review C.* — 2023.
- [17] *Davis, Benjamin L.* Discovery of a Planar Black Hole Mass Scaling Relation for Spiral Galaxies / Benjamin L. Davis, Zehao Jin // *The Astrophysical Journal Letters.* — 2023.
- [18] *Couderc, Romain.* Reassessment of the intrinsic carrier density temperature dependence in crystalline silicon / Romain Couderc, Mohamed Amara, Mustapha Lemiti // *J. Appl. Phys.* — 2014. — Mar. — Vol. 115, no. 9. — P. 093705.
- [19] *Pässler, Roland.* Dispersion-related description of temperature dependencies of band gaps in semiconductors / Roland Pässler // *Phys. Rev. B.* — 2002. — Aug. — Vol. 66. — P. 085201.

- [20] *Bednarczyk, D.* The approximation of the Fermi-Dirac integral F12 / D. Bednarczyk, J. Bednarczyk // *Phys. Lett. A.* — 1978. — Vol. 64, no. 4. — Pp. 409–410.
- [21] *Angelis, Dimitrios.* Artificial Intelligence in Physical Sciences: Symbolic Regression Trends and Perspectives / Dimitrios Angelis, Filippos Sofos, Theodoros E. Karakasidis // *Arch. Comput. Methods Eng.* — 2023. — Jul. — Vol. 30, no. 6. — Pp. 3845–3865.
- [22] *Natekin, Alexey.* Gradient boosting machines, a tutorial / Alexey Natekin, Alois Knoll // *Front. Neurorob.* — 2013. — Vol. 7.
- [23] *Cao, Weidong.* Parameter Optimization of Support Vector Regression Using Henry Gas Solubility Optimization Algorithm / Weidong Cao, Xin Liu, Jianjun Ni // *IEEE Access.* — 2020. — Vol. 8. — Pp. 88633–88642.
- [24] Machine-Learning Accelerating the Development of Perovskite Photovoltaics / Tiantian Liu, Sen Wang, Yingguang Shi et al. // *Sol. RRL.* — 2023. — Vol. 7, no. 23. — P. 2300650.
- [25] Enhancing Solar Forecasting Accuracy with Sequential Deep Artificial Neural Network and Hybrid Random Forest and Gradient Boosting Models across Varied Terrains / Muhammad Farhan Hanif, Muhammad Umar Siddique, Jicang Si et al. // *Advanced Theory and Simulations.* — 2024. — Vol. 7, no. 7. — P. 2301289.
- [26] Optuna: A Next-generation Hyperparameter Optimization Framework / Takuya Akiba, Shotaro Sano, Toshihiko Yanase et al. // Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. — KDD '19. — New York, NY, USA: Association for Computing Machinery, 2019. — Pp. 2623–2631.