

## УПРАВЛІННЯ В ТЕХНІЧНИХ СИСТЕМАХ

## УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

## CONTROL IN TECHNICAL SYSTEMS

УДК 004.852

DOI: 10.20998/2079-0023.2021.02.03

*Р. А. БАБУДЖАН, К. О. ИСАЕНКОВ, Д. М. КРАСИЙ, О. О. ВОДКА, І. В. ЗАДОРЖНИЙ, М. В. ЮЩУК***ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ БІНАРНОЇ КЛАСИФІКАЦІЇ РОБОЧОГО СТАНУ ПІДШИПНИКІВ ЗА СИГНАЛАМИ ЇХ ВІБРОПРИСКОРЕННЯ**

В роботі досліджується зв'язок між віброприскоренням підшипників з їх робочим станом. Для визначення цих залежностей було побудовано випробувальний стенд та проведено 112 експериментів з різними підшипниками: 100 підшипників, у яких під час експлуатації розвинувся внутрішній дефект та 12 підшипників без дефекту. З отриманих записів було сформовано набір даних, який використовувався для побудови класифікатору та знаходиться у вільному доступі. Був запропонований метод для класифікації нових та використаних підшипників, що полягає у пошуку залежностей та закономірностей сигналу за допомогою описових функцій: статистичних, ентропій, фрактальних розмірностей та інших. Окрім обробки самого сигналу, також використовувалося частотне представлення сигналу роботи підшипників для доповнення простору ознак. У роботі було перевірено можливість узагальнення класифікації для її застосування на тих сигналах, які не були отримані під час лабораторних експериментів. Сторонній набір даних було знайдено у вільному доступі. Цей набір даних був використаний для того, щоб визначити, наскільки точним буде класифікатор, який навчався та тестувався на істотно різних сигналах. Навчання та валідація проводилась методом бутстрапування для викоринення ефекту випадковості з огляду на малий об'єм наявних даних для навчання. Для оцінки якості класифікаторів було використано  $F1$ -міру, як основну метрику, через незбалансованість наборів даних. В якості моделей класифікатору були обрані наступні алгоритми машинного навчання з вчителем: логістична регресія, метод опорних векторів, випадковий ліс та метод найближчих сусідів. Результати представлені в вигляді графіків густини розподілу та діаграм.

**Ключові слова:** машинне навчання, вібродіагностика дефектів підшипників кочення, обробка даних, вилучення ознак сигналів, швидке перетворення Фур'є, класифікація незбалансованих наборів даних, метод Монте-Карло, бутстрап.

*Р. А. БАБУДЖАН, К. А. ИСАЕНКОВ, Д. М. КРАСИЙ, А. А. ВОДКА, И. В. ЗАДОРЖНИЙ, М. В. ЮЩУК***ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ БИНАРНОЙ КЛАССИФИКАЦИИ РАБОЧЕГО СОСТОЯНИЯ ПОДШИПНИКОВ ПО СИГНАЛАМ ИХ ВИБРОУСКОРЕНИЯ**

В работе исследуется связь между виброускорением подшипников с их эксплуатационным состоянием. Для определения этих зависимостей был построен испытательный стенд и проведено 112 экспериментов с разными подшипниками: 100 подшипников, у которых во время эксплуатации развился внутренний дефект и 12 подшипников без дефекта. Из полученных записей был сформирован набор данных, который использовался для построения классификаторов и находится в свободном доступе. Был предложен метод классификации новых и использованных подшипников, заключающийся в поиске зависимостей и закономерностей сигнала с помощью описательных функций: статистических, энтропий, фрактальных размерностей и других. Кроме обработки самого сигнала также использовалось частотное представление сигнала работы подшипников для дополнения пространства признаков. В работе была рассмотрена возможность обобщения классификации для ее применения на тех сигналах, которые не были получены в ходе лабораторных экспериментов. Посторонний набор данных был найден в свободном доступе. Этот набор данных был использован для того, чтобы определить, насколько точен классификатор, который учился и тестировался на существенно разных сигналах. Обучение и валідація проводилась методом бутстрапирования для искоренения эффекта случайности, учитывая малый объем имеющихся обучающих данных. Для оценки качества классификаторов была использована  $F1$ -мера как основная метрика из-за несбалансированности наборов данных. В качестве моделей классификатора были выбраны следующие методы машинного обучения с учителем: логистическая регрессия, метод опорных векторов, случайный лес и метод ближайших соседей. Результаты представлены в виде графиков плотности распределения и диаграмм.

**Ключевые слова:** машинное обучение, вибродиагностика дефектов подшипников качения, обработка данных, извлечение признаков сигналлов, быстрое преобразование Фурье, классификация несбалансированных наборов данных, метод Монте-Карло, бутстрап.

*Р. А. BABUDZHAN, К. О. ISAIENKOV, Д. М. KRASII, О. О. VODKA, І. V. ZADOROZHNIY, М. В. YUSHCHUK***THE USE OF MACHINE LEARNING METHODS FOR BINARY CLASSIFICATION OF THE WORKING CONDITION OF BEARINGS USING THE SIGNALS OF VIBRATION ACCELERATION**

The paper investigates the relationship between vibration acceleration of bearings with their operational state. To determine these dependencies, a test bench was built and 112 experiments were carried out with different bearings: 100 bearings that developed an internal defect during operation and 12 bearings without a defect. From the obtained records, a dataset was formed, which was used to build classifiers. Dataset is freely available. A method for classifying new and used bearings was proposed, which consists in searching for dependencies and regularities of the signal using descriptive

© Р. А. Бабуджан, К. О. Исаенков, Д. М. Красий, О. О. Водка, І. В. Задоржний, М. В. Ющук, 2021

functions: statistical, entropy, fractal dimensions and others. In addition to processing the signal itself, the frequency domain of the bearing operation signal was also used to complement the feature space. The paper considered the possibility of generalizing the classification for its application on those signals that were not obtained in the course of laboratory experiments. An extraneous dataset was found in the public domain. This dataset was used to determine how accurate a classifier was when it was trained and tested on significantly different signals. Training and validation were carried out using the bootstrapping method to eradicate the effect of randomness, given the small amount of training data available. To estimate the quality of the classifiers, the *F1*-measure was used as the main metric due to the imbalance of the data sets. The following supervised machine learning methods were chosen as classifier models: logistic regression, support vector machine, random forest, and *K* nearest neighbors. The results are presented in the form of plots of density distribution and diagrams.

**Keywords:** machine learning, vibration diagnostics of rolling bearing defects, data processing, signal feature mining, fast Fourier transform, classification of unbalanced datasets, Monte Carlo method, bootstrap.

**Вступ.** Обробка та аналіз сигналів широко поширений у завданнях радіоелектроніки, аналізі сейсмічної активності, розпізнаванні мови, а також вібродіагностики промислових конструкцій. З погляду різноманітності задач та застосувань обробки сигналів вібродіагностика конструкцій представляє великий інтерес. Розробка нових методів аналізу вібрацій установок активно проводиться сьогодні. Велику популярність отримали задачі аналізу підшипникових вузлів як одного з найчастіших місць поломки конструкцій, що обертаються. Близько 50 % таких конструкцій виходять з ладу саме через дефекти підшипникових вузлів [1].

Поширені зараз емпіричні підходи до вібродіагностики дають порівняно якісні результати та активно застосовуються у промисловості і сьогодні. Але розвиток обчислювальної техніки дає можливість створювати якісно нові підходи, які допоможуть покращити якість діагностики, автоматизувати та уніфікувати діагностичні процеси, зменшити витрати на діагностичне обладнання. Так, метою цієї роботи є огляд новітніх методів розробки математичних моделей діагностики, побудованих на кумулятивних характеристиках вібросигналів підшипників, та аналіз можливості уніфікації цих характеристик для різних конструкцій.

Мета цього дослідження – розробити підхід до аналізу сигналів вібрації установок, який б надав змогу класифікувати стан підшипників незважаючи на геометричні параметри установки та умови експлуатації.

**Огляд проблеми.** Задача діагностики підшипникових вузлів має широкий спектр підходів, що застосовуються у галузі протягом останніх років. Класичні підходи являють собою аналіз спектрів частот сигналів за допомогою емпіричних індикаторів, таких як аналіз зміни амплітуд частот, визваних різними конструкційними елементами підшипникових вузлів. Так, у статтях [2, 3, 4] розглядається класичний підхід до оцінки якості підшипника на підставі огляду спектру сигналу підшипнику. Такий підхід є виправданим та широко використовуваним у галузі, але потребує використання інформації щодо геометричних та експлуатаційних параметрів вузлів та установки, що унеможливило уніфікацію методу.

Дещо інший підхід пропонується у статті [5]. Автор пропонує використовувати статистичний аналіз метрики PSNR (пікового відношення сигналу до шуму) над коефіцієнтами розкладання сигналу в вейвлет-скейлограму. Цей підхід є більш уніфікованим, але для перевірки за статистичним критерієм необхідно мати сигнали підшипника, який перевіря-

ється під час діагностики. Для вирішення цієї проблеми можна використати методологію спектрального аналізу оминаючої за [6]. Цей потужний підхід дає спеціалістам з діагностики можливість використовувати перелік емпіричних індикаторів для детекції несправностей підшипникових вузлів, однак передбачає ручний вибір ширини фільтру та аналіз спектрів сигналу, що робить цей підхід затратним до людських ресурсів. Інший підхід до аналізу оминаючої пропонується у роботі [7]. Автори статті використовують синтезований набір даних сигналів прискорень підшипників для розрахунку ізоліній та побудови діагностичних карт.

У наш час проводяться дослідження з діагностики стану підшипників, використовуючи новітні методи побудови математичних моделей. Так, у статті [8] розглядаються дискримінантні класифікатори та машина опорних векторів для діагностики дефектів в підшипниках. Ці моделі мають високу точність розпізнавання, але у роботі не представлено застосування даних моделей для класифікації сигналів з інших установок обертання. Автори статті [9] використовують штучну нейронну мережу з архітектурою довгої короткочасної пам'яті для багатокласової класифікації аудіосигналів. За результатами статті можна побачити, що якість класифікації значним чином погіршується при використанні іншого експериментального обладнання. Таким чином питання уніфікації підходів для класифікації сигналів на наш час все ще є відкритим.

**Збір даних.** У вільному доступі було знайдено декілька наборів даних роботи підшипників, які можна було б використати для досліджень, але було вирішено сконструювати власний стенд для випробування. Власний експеримент проводився для збільшення кількості даних, та для виходу за обмеження, продиктованими параметрами експериментів, що були обрані іншими дослідниками. На рис. 1 зображений зовнішній вигляд пристрою. Характеристики пристрою та використаних підшипників знаходяться в табл. 1.

Таблиця 1 – Характеристики установки

Параметр	Значення	
Частота запису	3 кГц	
Швидкість ротору	0–1500–200 об/хв	
Вага валу	35 Н	
Тип підшипнику	6204	6202
Внутрішній діаметр	20 мм	15 мм
Зовнішній діаметр	47 мм	35 мм
Довжина	14 мм	11 мм

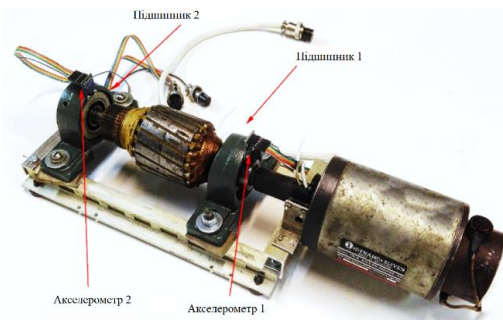


Рис. 1. Вигляд експериментальної установки

Для тестування різних методик використовується набір даних за вільним доступом [10]. Конструкція установки для проведення експериментів є однаковою для всіх наборів даних. На одному боці стенду встановлений підшипник без дефекту, він є незмінним для усіх експериментів. На іншому боці встановлений підшипник з дефектом, сигнал якого повинен бути проаналізований.

Обробка даних. Використовуючи стенд з попереднього пункту, було створено набір даних з 10265700 записів, в середньому по 91600 записів на кожен експеримент. Отримані дані знаходяться у вільному доступі: їх було завантажено на платформу Kaggle [11]. Набір даних складається з 6 колонок, які відповідають прискоренню кожного з двох підшипників уздовж кожної з трьох вісей, а також супутньої інформації про швидкість обертання ротору, час виміру від початку експерименту, та потужність двигуна. Для навчання класифікаторів використовуються миттєві прискорення підшипників у трьох осях: X, Y, Z.

На рис. 2 приведені миттєві прискорення деяких підшипників уздовж осі Y.

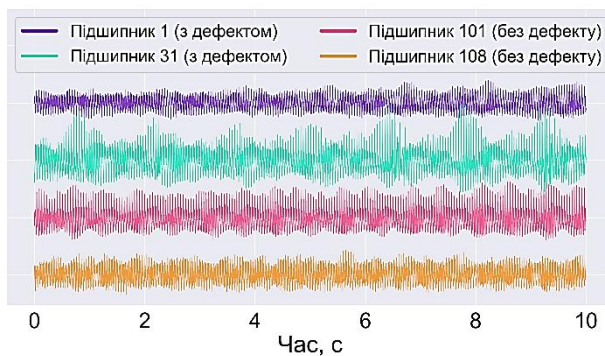


Рис. 2. Сигнали підшипників

Оскільки порівняння значень прискорення в конкретний момент часу між двома різними сигналами не є інформативним через велику кількість шумів та фазовий зсув, необхідно створити простір кумулятивних ознак, які б характеризували сигнал. Таким чином, для створення простору ознак використовується наступний підхід:

Початковий сигнал представляється чотирма способами:

- Чистий сигнал;
- Масштабування Z-значенням;

- Масштабування сигналу від  $-1$  до  $1$ ;
- Робастне масштабування [12];

Отриманий з попереднього пункту сигнал представляється методами описової статистики, різними визначеннями ентропії, фрактальних розмірностей та іншими, що будуть названі далі.

Сигнал представляється в частотному просторі за допомогою перетворення Фур'є (рис. 3), після чого використовуються методи з попереднього пункту.

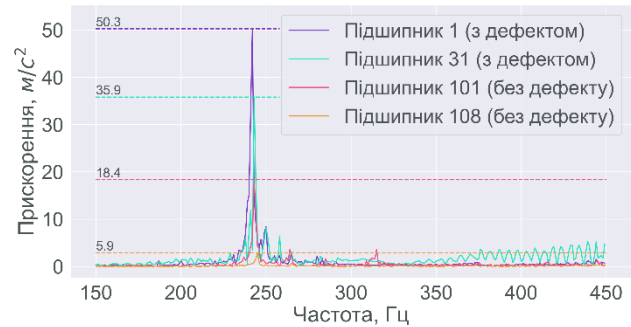


Рис. 3. Спектри сигналів

Набір використаних методів опису сигналу виглядає наступним чином:

- Коефіцієнт варіації;
- Розмах;
- Інтерквартильний розмах (IQR);
- Асиметрія;
- Екссес;
- Ентропія:
  - Ентропія за Шенном;
  - Приблизна ентропія;
  - Зразкова ентропія;
  - Ентропія перестановки;
  - Ентропія сингулярного розкладу матриці.
- Енергія;
- Параметри Хьорта:
  - Активність;
  - Мобільність.
- Показник Херста;
- Фрактальні розміри:
  - За Петросяном;
  - За Хігучі;
  - За Кацом.
- Кількість переходів через нуль;
- Пік-фактор.

Детальний розгляд даних ознак був приведений у статті [13].

**Моделі машинного навчання.** У даній роботі розглядаються 4 алгоритми машинного навчання:

- Логістична регресія
- Машина опорних векторів
- Випадковий ліс
- K-найближчих сусідів

Логістична регресія аналогічна множинній лінійній регресії за одним винятком – результат є бінарним. Перший крок полягає в тому, щоб представити змінну

результату не як бінарну мітку, а як ймовірність  $p$ , що мітка дорівнює "1". Для цього використовується функція логістичного відгуку:

$$p = \frac{1}{1 + e^{-(b_0 + b_1x_1 + b_2x_2 + \dots)}} \quad (1)$$

Для того, щоб отримати експоненційний вираз із знаменника, замість ймовірностей ми розглядаємо шанси. Шанси – це ймовірність події, поділена на ймовірність, що подія не станеться.

$$\text{Odds}(Y = 1) = \frac{p}{1 - p} \quad (2)$$

Підставивши  $p$  у функцію шансів і взявши логарифм від отриманого виразу, ми отримаємо так звану логіт-функцію  $\text{logit}(p)$ . Невідомі вагові коефіцієнти  $b_i$  обчислюються методом максимальної правдоподібності.

Метод опорних векторів (support vector machine або SVM) – одна з найпопулярніших методологій навчання за прецедентами. У процесі навчання SVM обчислює важливість кожної точки навчальних даних з погляду визначення вирішальної межі між двома класами. Зазвичай лише частина точок навчального набору важлива для визначення межі прийняття рішення: точки, які лежать на межі між класами. Вони називаються опорними векторами (support vectors) і дали назву машині опорних векторів.

Щоб отримати прогноз для нової точки, вимірюється відстань до кожного опорного вектору. Класифікаційне рішення приймається виходячи з відстаней до опорних векторів, а також важливості опорних векторів, отриманих у процесі навчання. Вимірювання відстані між точками зазвичай виконується за допомогою гаусовського ядра, але ним не обмежено.

Випадковий ліс (англ. Random Forest) – це метод ансамблювання (агрегації) простих моделей дерева прийняття рішень [14]. Випадковий ліс навчає сукупність дерев рішень на різних підвбірках набору даних і використовує усереднення для підвищення точності прогнозування та контролю перенавчання. Дерево рішень [14] – це рекурсивний алгоритм, який прогнозує значення цільової змінної, вивчаючи прості правила прийняття рішень, виведені з ознак даних. На кожній ітерації дерево розбиває набір даних так, щоб максимізувати функцію інформативності розбиття за якимось критерієм, і створює вузол дерева з оптимальним правилом розбиття. У нашому випадку ми використовуємо критерій Джині. Оскільки дерево є жадібним алгоритмом – воно оптимізує критерій інформативності локально у кожному вузлі, набагато краще використовувати ансамбль з дерев для збільшення точності. У методі випадкового лісу використовується беггінг та метод випадкових підпросторів (RSM). Використовуючи перший метод, ми будемо навчати кожне дерево на підвбірці, взятій з поверненням із початкової вибірки, а використовуючи RSM ми будемо навчати кожне дерево лише на

частині ознак з набору даних. Остаточне рішення відбувається голосуванням по всіх деревах.

$K$ -найближчих сусідів (англ.  $K$ -nearest neighbors, або KNN) – метричний метод, сутність якого є у знаходженні "сусідів" елемента, найближчих до нього за певною метрикою [14]. Значення цільової величини, яке найчастіше зустрічається серед сусідніх елементів, стає значенням цього елемента. У даній роботі використовувалась метрика Міньковського – узагальнення евклідовської відстані з довільним значенням показнику.

Для кожної моделі у цій роботі була зроблена оптимізація гіперпараметрів методом повного перебору всіх комбінацій. Оптимізація гіперпараметрів виконувалась як максимізація якості розпізнавання на валідаційних вибірках.

**Оцінки якості.** Якість алгоритмів оцінювалась за допомогою  $F1$ -міри [15]. Ця метрика використовується для оцінки якості алгоритму класифікації, який побудований на наборі даних зі значним дисбалансом класів. Вимірюється від 0 до 1, де 1 – найкращий результат.  $F1$ -міра розраховується за формулою 3:

$$F1 = \frac{2 \times \text{влучність} \times \text{повнота}}{\text{влучність} + \text{повнота}} \quad (3)$$

Так, обрана метрика є гармонійним середнім між двома іншими метриками – влучністю – впевненістю у тому, що модель не робить багато хибно-позитивних помилок, та повнотою – впевненістю у тому, що модель не робить багато хибно-негативних помилок.  $F1$ -міра буде дорівнювати 0 якщо хоча б одна з цих складових дорівнюватиме 0, та 1 якщо обидві складові дорівнюють 1. Влучність та повноту можна розрахувати за формулами 4, 5:

$$\text{влучність} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{повнота} = \frac{TP}{TP + FN} \quad (5)$$

де  $TP$  – правильно розпізнані позитивні результати (підшипники без дефекту);

$FP$  – кількість хибно-позитивних помилок;

$FN$  – кількість хибно-негативних помилок.

У роботі також розраховувались супутні метрики для більш тонкого контролю математичних моделей – точність, або відсоток правильних відповідей; складові  $F1$ -міри; TNR – метрика, які вказує, наскільки якісно ми розпізнаємо негативні приклади.

Оскільки ми маємо занадто малу кількість прикладів для навчання та тестування моделі, конкретні показники якості будуть залежати від того, як ми розіб'ємо вибірку на навчальну та тестову підвбірку. Тому ми використовували валідацію методом бутстрапу [16]. Цей метод полягає в багатократному стохастичному відборі з поверненням елементів для навчання з початкової вибірки. Так, для 100 ітерацій відбору ми будемо мати 100 різних пар підвбірок, де навчальна вибірка буде містити 63 %

елементів з початкової вибірки, а тестова – відповідно, 37 %. Цей метод розрахунку називається "Out-of-bag error" [17]. Цей метод також застосовується у методи машинного навчання "випадковий ліс". Необхідно зазначити, що ми окремо розбиваємо позитивні і негативні приклади, щоб зберегти розподіл класів та не отримати елементи лише одного класу у навчальній або тестовій підвибірках.

Якщо для кожної ітерації бутстрапованого відбору ми можемо навчити математичну модель на 63 % даних та обчислити метрики на відкладених 37 %, ми отримаємо розподіл значень метрики. Так, маючи середнє значення з розподілу, ми можемо зробити стійкий кількісний висновок про якість моделі.

Для того, щоб зробити якісний висновок про роботу моделі – чи є її прогноз кращим за випадкове вгадування, ми провели експеримент Монте-Карло [18]. Так, ми  $M$  разів створили випадкові прогнози для кожної з наших 100 пар підвибірок. Потім ми обчислили  $F1$ -міри для кожного прогнозу та знайшли середні значення метрики за усіма підвибірками. Так, ми маємо вибірку об'ємом  $M$  середніх значень  $F1$ -міри. За центральною граничною теоремою, розподіл середніх можна описати нормальним законом. Порівнюючи значення метрики для певної моделі з 99 % довірчим інтервалом із розподілу середніх і побачивши, що наше значення метрики більше ніж права границя, ми можемо зробити статистично значимий висновок, що наша модель навчається на даних, а не лише запам'ятовує навчальну вибірку.

Так, для нашого набору даних ми маємо 12 позитивних прикладів та 100 негативних. Тестова вибірка містить 37 % прикладів, тобто у кожній підвибірці ми маємо 41 приклад. На рис. 4, *a* представлений розподіл середніх та 99 % довірчі інтервали для нашого набору даних.

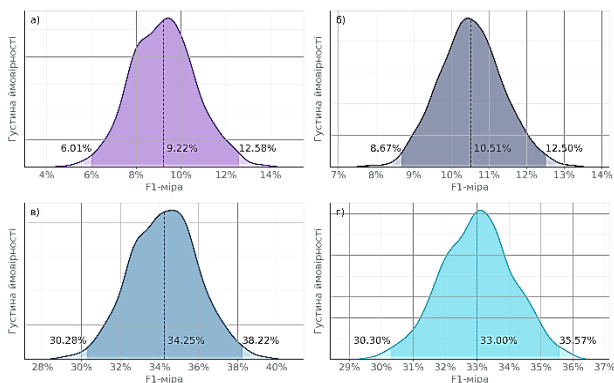


Рис. 4. Середнє значення  $F1$ -міри при випадковому вгадуванні стану підшипника: *a* – наш набір даних, 37 % набору; *б* – наш набір даних, 100 % набору; наш набір даних, *в* – сторонній набір даних, 37 % від набору; *г* – сторонній набір даних, 37 % від набору

Якщо брати весь наш набір даних за тестову підвибірку, то будемо мати рис. 4, *б*. Сторонні набори даних містять 33 % позитивних результатів, або 15 прикладів. Розподіл середніх для цього набору даних при тестовій підвибірці з 37 % початкової вибірки

проілюстрований на рис. 4, *в*. Для усіх 100 % (45 прикладів) – на рис. 4, *г*. Якщо ми розглядаємо навчання і тестування на одному й тому ж наборі даних, нас цікавить порівняння метрики з рис. 4, *a* та рис. 4, *в*. Якщо ж ми розглядаємо узагальнення математичних моделей на інші набори даних, нас цікавлять довірчі інтервали на рис. 4, *б* та рис. 4, *г*. У даній роботі значення  $M$  дорівнює 1000.

**Результати.** Першим етапом цієї роботи є створення математичних моделей для розпізнавання підшипників з дефектом у рамках одного експерименту. На рис. 5 зображені розподіли  $F1$ -міри для 100 бутстрапованих підвибірок з нашого набору даних. На рис. 6 зображені середні значення усіх метрик, які були обчислені у даній роботі. На рис. 7 надані середні значення метрик для іншого набору даних.

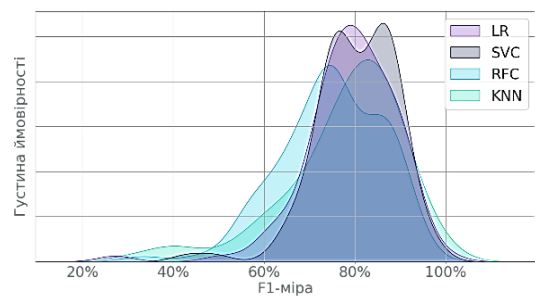


Рис. 5. Розподіл  $F1$ -міри при навчанні та тестуванні моделей на нашому наборі. Якість на валідаційній вибірці.

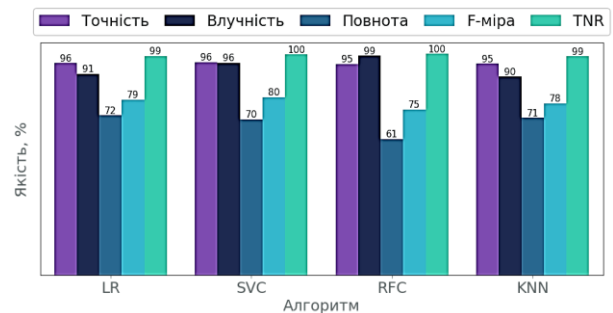


Рис. 6. Середні значення метрик при навчанні та тестуванні моделей на нашому наборі. Якість на валідаційній вибірці.

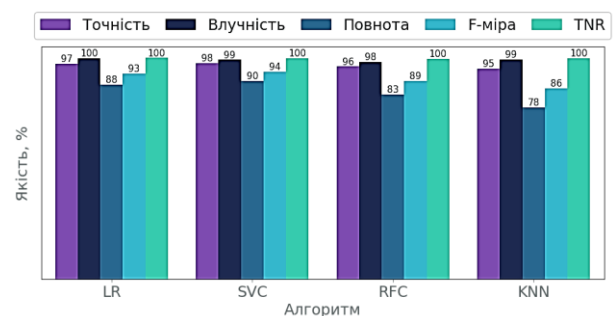


Рис. 7. Середні значення метрик при навчанні та тестуванні моделей на іншому наборі. Якість на відкладених вибірках.

Як можна бачити з наведених рисунків, метод опорних векторів має найкращий результат. Порівнюючи середні значення з рис. 4, *a*, *в*, ми можемо бути впевнені, що всі моделі придатні для детекції підшипників з дефектом у рамках одного експерименту.

Логістична регресія на рисунках позначається як LR, метод опорних векторів – SVC, випадковий ліс – RF, К найближчих сусідів – KNN. Усі метрики розраховувались для валідаційних вибірок, тобто на вибірках даних, які не приймали участі у навчанні моделей, а використовувались при оптимізації гіперпараметрів.

Другий етап – застосування моделей для класифікації сигналів, отриманих з інших експериментів. Тут перевірялося, чи можуть дані математичні моделі правильно класифікувати підшипники інших розмірів, які були встановлені на інших машинах обертання. На цьому етапі було розглянуто два випадки – навчання на нашому наборі даних і тестування на іншому наборі, та навчання на іншому і тестування на нашому. У першому випадку серед усіх варіантів масштабування сигналу та відбору статистик не було знайдено моделі, яка б перевершила 99 % довірчий інтервал випадкового вгадування з рис. 4, з. Але у другому випадку – навчанні на інших даних – ми отримали результат 18 % при використанні алгоритму "випадкового лісу", що на 5.5 % випереджає відповідний 99 % довірчий інтервал з рис. 6. На рис. 8 можна побачити розподіл  $F_1$ -міри, а на рис. 9 – середні значення усіх метрик.

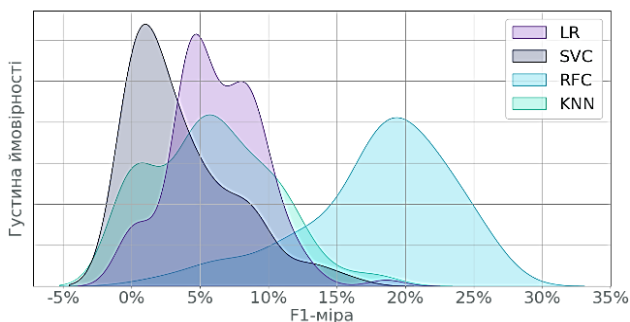


Рис. 8. Розподіл  $F_1$ -міри при навчанні моделей на іншому наборі та тестуванні на нашому.

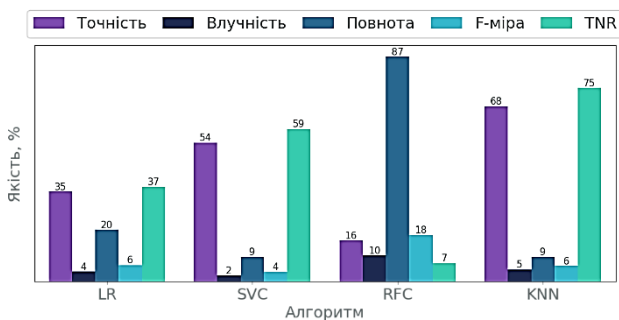


Рис. 9. Середні значення метрик при навчанні моделей на іншому наборі та тестуванні на нашому.

**Висновки.** У даній роботі розглядався підхід до аналізу дефектів підшипників кочення на основі застосування алгоритмів машинного навчання до оброблених сигналів миттєвого прискорення підшипників. Для отримання набору даних був створений експериментальний стенд, на якому було розглянуто 112 підшипників двох типорозмірів. У роботі також використовувався сторонній набір даних за відкритим доступом. У роботі запропоновані методи розширення

набору даних шляхом побачевого роздіблення початкових сигналів.

Для застосування алгоритмів машинного навчання був запропонований підхід для створення простору ознак сигналів за допомогою використання різних статистик сигналу. Простір ознак також містить статистику частотного представлення вхідного сигналу.

Для стійкої кількісної оцінки якості алгоритмів був використаний бутстрапований відбір прикладів для навчання моделей. Цей метод дав змогу побудувати розподіл цільової метрики для оцінки якості моделі. Для якісної оцінки методів використовувалося порівняння середнього значення цільової метрики з довірчими інтервалами з розподілу, отриманому за допомогою методу Монте-Карло для середніх значень при випадковому вгадуванні. Цей підхід дав змогу оцінити можливість застосування методів машинного навчання у поєднанні з використанням статистик сигналу як простору ознак для бінарної класифікації підшипників на підшипники з дефектом та без.

#### Список літератури

1. Nguyen P., Kang M., Kim J., Ahn B., Ha J., Choi B. Robust condition monitoring of rolling element bearings using de-noising and envelope analysis with signal decomposition techniques. *Expert systems with applications*. 2015. Vol. 42, no. 22. P. 9024–9032. doi: 10.1016/j.eswa.2015.07.064.
2. Graney B., Starry K. Rolling element bearing analysis. *Engineering. Materials evaluation*. 2012. Т. 70. С. 78–85.
3. Науменко А. П., Кудрявцева И. С. *Диагностирование дефектов и неисправностей машин и механизмов: указания к лабораторным работам*. Омск: Изд-во ОмГТУ, 2017. 60 с.
4. Таджикибаев А. И., Драганчев Х., Яровой В. С., Гаюров Л., Коркин Л. М. Прогнозирование технического состояния подшипников качения на основе вибрационного анализа. *Материаловедение. Энергетика*. 2021. Т. 27, № 3. С. 77–85. doi: 10.18721/JEST.27307.
5. Акутин М. В. Метод контроля состояния подшипников качения на основе сравнения вейвлет скейлограмм. *Известия вузов. Проблемы энергетики*. 2009. № 1–2. С. 145–149.
6. Кан Ш. Ч., Микулович А. В., Микулович В. И. Спектральный анализ огибающей высокочастотных составляющих сложных сигналов на основе эмпирической декомпозиции мод и преобразования Гильберта. *Информатика*. 2010. Т. 4, № 28. С. 16–26.
7. Soto-Ocampo R., Cano-Moreno J., Mera J., Maroto J. Bearing severity fault evaluation using contour maps-case study. *Applied sciences*. 2021. Vol. 11, no. 14. P. 2–30. doi: 10.3390/app11146452.
8. Cococcioni M., Forte P., Sacchi C., Manconi S. A machine learning approach to fault diagnosis of rolling bearings. *2008 IEEE international conference on computational cybernetics (ICCC), Stara Lesna, Slovakia, 27–29 November 2008*. 2008. P. 209–214. doi: 10.1109/icccyb.2008.4721407.
9. Nishizaki H., Makino K. Signal classification using deep learning. *2019 IEEE international conference on sensors and nanotechnology (SENSORS & NANO), Penang, Malaysia, 24–25 July 2019*. 2019. P. 1–4. doi: 10.1109/sensorsnano44414.2019.8940077.
10. Soto-Ocampo C., Cano-Moreno J., Mera J., Maroto J. *Bearing Database*. URL: <https://zenodo.org/record/3898942#.YbZtrr1ByUm> (дата звернення: 01.12.2021).
11. Isaienkov K. *Bearing classification*. URL: <https://www.kaggle.com/isaienkov/bearing-classification> (дата звернення: 22.09.2021).
12. Pedregosa F., Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011. Vol. 12. P. 2825–2830.
13. Babudzhani R., Isaienkov K., Krasii D., Melkonian R., Vodka O., Zadorozhnyi I., Yushchuk M. Collection and processing of bearing vibration data for their technical condition classification by machine learning methods. *Modeling Control and Information Technologies*. 2021. Vol. 5. P. 10–15. doi: 10.31713/mcit.2021.02.

14. Müller A., Guido S. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, 2016. 400 p.
15. Sulaiman M., Hossin M. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*. 2015. Vol. 5. P. 1–11. doi: 10.5121/ijdkp.2015.5201.
16. Janitza S., Hornung R. On the overestimation of random forest's out-of-bag error. *PLoS ONE*. 2018, Vol. 13. P. 2–19. doi: 10.1371/journal.pone.0201904.
17. Bruce P., Bruce A. *Practical Statistics for Data Scientists: 50 Essential Concepts*. O'Reilly Media, 2017. 318 p.
18. Fishman G. *Monte Carlo: concepts, algorithms, and applications*. Springer, 1996. 689 p.

## References (transliterated)

1. Nguyen P., Kang M., Kim J., Ahn B., Ha J., Choi B. Robust condition monitoring of rolling element bearings using de-noising and envelope analysis with signal decomposition techniques. *Expert systems with applications*. 2015, vol. 42, no. 22, pp. 9024–9032. doi: 10.1016/j.eswa.2015.07.064.
2. Graney B., Starry K. Rolling element bearing analysis. *Engineering. Materials evaluation*. 2012, vol. 70, pp. 78–85.
3. Naumenko A. P., Kudryavtseva Y. S. Dyagnostyrovanye defektov y neyspravnostey mashyn y mekhanyzmov: ukazaniya k laboratornim rabotam. [Diagnosis of defects and malfunctions of machines and mechanisms: instructions for laboratory work]. Omsk: OmHTU Publ., 2017. 60 p.
4. Tadzhibaev A. I., Draganchev H., Yarovoy V. S., Gayurov L., Korin L. M., Forecasting of the technical condition of rolling bearings based on vibration analysis. *Materials Science. Power Engineering*. 2021, pp. 77–85. doi: 10.18721/JEST.27307.
5. Akutin M. V. Method of monitoring the state of rolling bearings based on the comparison of wavelet scalegrams. *Izvestiya vuzov. Energy problems*. 2009, no. 1–2, pp. 145–149. URL: <https://cyberleninka.ru/article/n/metod-kontrolya-sostoyaniya-podshipnikov-kacheniya-na-osnove-sravneniya-veyvlet-skeylogramm> (date accessed 07.12.2021).
6. Kan Sh. Ch., Mykulovych A. V., Mykulovych V. Y. Spektral'nyy analiz ohybayushchey visokochastotnikh sostavlyayushchykh slozhnikh syhnalov na osnove empyrycheskoy dekompozitsyy mod y preobrazovaniya Hyl'berta. [Spectral analysis of the envelope of high-frequency components of complex signals based on empirical mode decomposition and Hilbert transform]. *Informatyka*. 2010, vol. 4, no. 28, pp. 16–26.
7. Soto-Ocampo R., Cano-Moreno J., Mera J., Maroto J. Bearing severity fault evaluation using contour maps-case study. *Applied sciences*. 2021, vol. 11, no. 14, pp. 2–30. doi: 10.3390/app11146452.
8. Cococcioni M., Forte P., Sacchi C., Manconi S. A machine learning approach to fault diagnosis of rolling bearings. *2008 IEEE international conference on computational cybernetics (ICCC), Stara Lesna, Slovakia, 27–29 November 2008*. 2008, pp. 209–214. doi: 10.1109/icccyb.2008.4721407.
9. Nishizaki H., Makino K. Signal classification using deep learning. *2019 IEEE international conference on sensors and nanotechnology (SENSORS & NANO), Penang, Malaysia, 24–25 July 2019*. 2019, pp. 1–4. doi: 10.1109/sensorsnano44414.2019.8940077.
10. Soto-Ocampo C., Cano-Moreno J., Mera J., Maroto J. *Bearing Database*. Available at: <https://zenodo.org/record/3898942#YbZtrr1ByUm> (accessed 01.12.2021).
11. Isaienkov K. *Bearing classification*. Available at: <https://www.kaggle.com/isaienkov/bearing-classification> (accessed 22.09.2021).
12. Pedregosa F., Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011, vol. 12, pp. 2825–2830.
13. Babudzhan, R., Isaienkov, K., Krasii, D., Melkonian, R., Vodka, O., Zadorozhnyi, I., & Yushchuk, M. Collection and processing of bearing vibration data for their technical condition classification by machine learning methods. *Modeling Control and Information Technologies*. 2021, vol. 5, pp. 10–15. doi: 10.31713/mcit.2021.02.
14. Müller A., Guido S. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, 2016. 400 p.
15. Sulaiman M., Hossin M. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*. Vol. 5, pp. 1–11. doi: 10.5121/ijdkp.2015.5201.
16. Janitza S., Hornung R. On the overestimation of random forest's out-of-bag error. *PLoS ONE*. 2018, vol. 13, pp. 2–19. doi: 10.1371/journal.pone.0201904.
17. Bruce P. and Bruce A. *Practical Statistics for Data Scientists: 50 Essential Concepts*. O'Reilly Media, 2017. 318 p.
18. Fishman G. *Monte Carlo: concepts, algorithms, and applications*. Springer, 1996. 689 p.

Надійшла (received) 08.11.2021

## Відомості про авторів / Сведения об авторах / About the Authors

**Бабуджан Руслан Андрійович** – Національний Технічний Університет «Харківський Політехнічний Інститут», студент кафедри динаміки та міцності машин; м. Харків, Україна; ORCID: <https://orcid.org/0000-0001-5765-9234>, e-mail: [ruslanbabudzhan@gmail.com](mailto:ruslanbabudzhan@gmail.com)

**Ісаєнков Константин Олександрович** – інженер, Quantum Inc.; м. Харків, Україна; ORCID: <https://orcid.org/0000-0001-5266-7040>; e-mail: [k.isaienkov@quantumobile.com](mailto:k.isaienkov@quantumobile.com)

**Красій Данило Максимович** – Національний технічний університет «Харківський політехнічний інститут», студент; м. Харків, Україна; ORCID: <https://orcid.org/0000-0003-1599-2295>, e-mail: [krasiydm@gmail.com](mailto:krasiydm@gmail.com)

**Водка Олексій Олександрович** – кандидат технічних наук, доцент, Національний технічний університет «Харківський політехнічний інститут»; м. Харків, Україна; ORCID: <https://orcid.org/0000-0002-4462-9869>; e-mail: [Oleksii.vodka@gmail.com](mailto:Oleksii.vodka@gmail.com)

**Задорожний Іван Вячеславович** – Національний Технічний Університет «Харківський Політехнічний Інститут», аспірант кафедри динаміки та міцності машин; м. Харків, Україна; ORCID: <https://orcid.org/0000-0002-6952-3024>, e-mail: [Ivan.Zadorozhnyi@khp.edu.ua](mailto:Ivan.Zadorozhnyi@khp.edu.ua)

**Ющук Михайло Вікторович** – інженер, Quantum Inc; м. Харків, Україна; ORCID: <https://orcid.org/0000-0002-2618-0893>, e-mail: [m.yushchuk@quantumobile.com](mailto:m.yushchuk@quantumobile.com)

**Бабуджан Руслан Андреевич** – Национальный технический университет «Харьковский политехнический институт», студент кафедры динамики и прочности машин; г. Харьков, Украина; ORCID: <https://orcid.org/0000-0001-5765-9234>, e-mail: [ruslanbabudzhan@gmail.com](mailto:ruslanbabudzhan@gmail.com)

**Исаенков Константин Александрович** – Quantum Inc., инженер; г. Харьков, Украина; ORCID: <https://orcid.org/0000-0001-5266-7040> e-mail: [k.isaienkov@quantumobile.com](mailto:k.isaienkov@quantumobile.com)

**Красий Данил Максимович** – Национальный технический университет «Харьковский политехнический институт», студент кафедры динамики и прочности машин; г. Харьков, Украина; ORCID: <https://orcid.org/0000-0003-1599-2295>, e-mail: [krasiydm@gmail.com](mailto:krasiydm@gmail.com)

**Водка Алексей Александрович** – кандидат технических наук, доцент, Национальный технический университет «Харьковский политехнический институт»; г. Харьков, Украина; ORCID: <https://orcid.org/0000-0002-4462-9869>; e-mail: [Oleksii.vodka@gmail.com](mailto:Oleksii.vodka@gmail.com)

**Задорожный Иван Вячеславович** – Национальный Технический Университет "Харьковский Политехнический Институт", аспирант кафедры динамики и прочности машин; г. Харьков, Украина; ORCID: <https://orcid.org/0000-0002-6952-3024>, e-mail: [Ivan.Zadorozhnyi@khpi.edu.ua](mailto:Ivan.Zadorozhnyi@khpi.edu.ua)

**Ющук Михаил Викторович** – Quantum Inc., инженер; г. Харьков, Украина; ORCID: <https://orcid.org/0000-0002-2618-0893>, e-mail: [m.yushchuk@quantumobile.com](mailto:m.yushchuk@quantumobile.com)

**Babudzhan Ruslan Andriovich** – National Technical University "Kharkiv Polytechnic Institute", student of the Department of Dynamics and Strength of Machines; Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0001-5765-9234>, e-mail: [ruslanbabudzhan@gmail.com](mailto:ruslanbabudzhan@gmail.com)

**Isaienkov Konstantyn Oleksandrovich** – Quantum Inc., engineer; Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0001-5266-7040>; e-mail: [k.isaienkov@quantumobile.com](mailto:k.isaienkov@quantumobile.com)

**Krasiy Danilo Maksimovich** – National Technical University "Kharkov Polytechnic Institute", student of the Department of Dynamics and Strength of Machines; Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0003-1599-2295>, e-mail: [krasiydm@gmail.com](mailto:krasiydm@gmail.com)

**Vodka Oleksii Oleksandrovich** – Candidate of Technical Sciences, Associate Professor, National Technical University "Kharkiv Polytechnic Institute"; Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0002-4462-9869>; e-mail: [Oleksii.vodka@gmail.com](mailto:Oleksii.vodka@gmail.com)

**Zadorozhny Ivan Vyacheslavovich** – National Technical University "Kharkov Polytechnic Institute", post-graduate student of the Department of Dynamics and Strength of Machines; Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0002-6952-3024>, e-mail: [Ivan.Zadorozhnyi@khpi.edu.ua](mailto:Ivan.Zadorozhnyi@khpi.edu.ua)

**Yushchuk Michael Viktorovich** – Quantum Inc., engineer; Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0002-2618-0893>, e-mail: [m.yushchuk@quantumobile.com](mailto:m.yushchuk@quantumobile.com)