

УДК 629.4.018

## Многоклассовая классификация технического состояния дизельного двигателя тепловоза по вейвлет-признакам вибросигналов методами машинного обучения

**А. А. Беляев**

Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

**Для цитирования:** Беляев А. А. Многоклассовая классификация технического состояния дизельного двигателя тепловоза по вейвлет-признакам вибросигналов методами машинного обучения // Бюллетень результатов научных исследований. 2026. Вып. 1. С. 142–153. DOI: 10.20295/2223-9987-2026-1-142-153

### Аннотация

**Цель:** разработка и экспериментальная апробация алгоритма многоклассовой классификации технического состояния дизельного двигателя тепловоза на основе вейвлет-признаков вибросигналов с применением методов машинного обучения для обеспечения надежности и безопасности железнодорожного транспорта. **Методы:** экспериментальные исследования проводились на дизельном двигателе Д50 в Тепловозной лаборатории имени профессора Я. М. Гаккеля Петербургского государственного университета путей сообщения. Зарегистрирован 491 трехосевой вибросигнал в семи диагностических состояниях (исправное состояние и шесть вариантов имитации неисправности топливоподачи путем отключения подачи топлива в отдельные цилиндры). Вейвлет-пакетная декомпозиция до 11-го уровня с вейвлетом Добеши 4-го порядка позволила извлечь 45 984 диагностических признака из узлов декомпозиции. Для снижения размерности признакового пространства применен метод главных компонент (PCA) с порогом сохранения 95 % объясненной дисперсии. Проведено сравнительное исследование восьми методов машинного обучения с использованием стратифицированной пятикратной кросс-валидации для объективной оценки обобщающей способности моделей. **Результаты:** логистическая регрессия с L1-регуляризацией достигла наивысшей точности классификации 99,32 % (Accuracy = 0,9932, F1-macro = 0,9921). Стратифицированная кросс-валидация подтвердила стабильность и воспроизводимость результатов: F1-macro = 99,56 % ± 0,55 %. Метод продемонстрировал минимальную дисперсию среди всех исследованных алгоритмов, что свидетельствует о высокой устойчивости к вариациям в данных. **Практическая значимость:** экспериментально подтверждена возможность автоматической классификации дизельного двигателя по диагностическим классам на основе вейвлет-признаков вибросигналов. Логистическая регрессия выбрана как оптимальный метод благодаря сочетанию высокой точности, стабильности результатов и интерпретируемости модели. Разработанный алгоритм может быть использован в качестве основы для создания бортовых систем диагностики тепловозов в режиме реального времени.

**Ключевые слова:** машинное обучение, классификация, логистическая регрессия, L1-регуляризация, вейвлет-анализ, вибродиагностика, дизельный двигатель, тепловоз, анализ главных компонент, кросс-валидация

### Введение

Развитие эффективных методов диагностики технического состояния дизельных двигателей тепловозов представляет собой одну из ключевых задач обеспечения надежности и безопасности железнодорожного транспорта.

Возрастающие требования к надежности эксплуатации тепловозного парка обуславливают необходимость создания высокоточных систем автоматического диагностирования, способных обнаруживать зарождающиеся дефекты на ранних стадиях их развития [1, 2].

Вибрационная диагностика является одним из наиболее информативных методов оценки технического состояния дизельных двигателей [3, 4]. Виброакустические сигналы содержат информацию о динамических процессах в узлах силовой установки, позволяя выявлять нарушения в работе цилиндра-поршневой группы, кривошипно-шатунного механизма, систем газораспределения и топливоподачи.

Вейвлет-преобразование обеспечивает одновременную локализацию сигнала во временной и частотной областях, что делает его оптимальным инструментом для анализа нестационарных вибросигналов [5, 6]. Многомасштабная структура вейвлет-декомпозиции позволяет одновременно анализировать сигнал на различных уровнях детализации.

Применение методов машинного обучения для классификации технического состояния механических систем по вибрационным сигналам активно развивается в последние годы [7, 8]. Особую актуальность представляют методы, обеспечивающие одновременно высокую точность классификации и интерпретируемость результатов, что критически важно для практического применения.

Цель настоящего исследования — разработка и экспериментальная апробация алгоритма многоклассовой классификации технического состояния дизельного двигателя Д50 на основе вейвлет-признаков вибросигналов с использованием методов машинного обучения.

## **Материалы и методы**

### *Объект исследования и экспериментальная установка*

Экспериментальные исследования проводились в Тепловозной лаборатории имени профессора Я. М. Гаккеля Петербургского государственного университета путей сообщения Императора Александра I (ПГУПС). Источником экспериментальных данных служил четырехтактный дизельный двигатель Д50 производства Коломенского завода со следующими характеристиками: номинальная мощность — 735 кВт, частота вращения коленчатого вала — 750 об/мин, шесть цилиндров рядного расположения.

Для регистрации вибрационных процессов использовался цифровой трехосевой виброметр ZET 7152-N (производство ООО «ЭТМС», Россия) с преобразователем интерфейса ZET 7174 (CAN 2.0 ↔ USB). Виброметр обеспечивает измерение виброускорения в диапазоне от 0,01 до 50 м/с<sup>2</sup> в полосе частот от 3 до 200 Гц по трем взаимно перпендикулярным осям [3, 4]. Датчик установлен на верхней плоскости блока цилиндров между крышками 3-го и 4-го цилиндров с использованием магнитного крепления. Частота дискретизации составляла 200 Гц.

Контролируемое моделирование неисправностей осуществлялось путем программного отключения подачи топлива в отдельные цилиндры двигателя. Экспериментальная программа включала регистрацию вибросигналов для семи диагностических классов: класс 0 — исправное состояние (все цилиндры работают); классы 1–6 — отключение подачи топлива в цилиндры № 1–6 соответственно.

В результате проведения экспериментальных исследований был зарегистрирован 491 вибросигнал. Распределение данных по диагностическим классам представлено в табл. 1.

ТАБЛИЦА 1. Распределение экспериментальных данных по диагностическим классам

Класс	Техническое состояние	Количество сигналов	Доля от общего объема, %
0	Исправное состояние	114	23,2
1	Отключение цилиндра № 1	62	12,6
2	Отключение цилиндра № 2	63	12,8
3	Отключение цилиндра № 3	60	12,2
4	Отключение цилиндра № 4	60	12,2
5	Отключение цилиндра № 5	61	12,4
6	Отключение цилиндра № 6	71	14,5
Всего	—	491	100,0

### Извлечение диагностических признаков

Для декомпозиции вибросигналов применялось вейвлет-пакетное преобразование с использованием вейвлета Добеши 4-го порядка (db4) [5, 6]. Максимальный уровень декомпозиции определялся длиной анализируемых сигналов [5]:

$$L_{\max} = \log_2(N), \quad (1)$$

где  $N$  — длина сигнала в отсчетах.

Для вибросигналов длиной  $N = 2100$  отсчетов максимальный уровень составил  $L_{\max} = 11$ , что обеспечивало разбиение частотного диапазона на  $2^{11} = 2048$  равных частотных полос.

Из каждого узла вейвлет-декомпозиции ( $j, k$ ) извлекался комплексный набор из 30 диагностических характеристик, сгруппированных в семь категорий [7]: базовые статистические моменты (среднее значение, стандартное отклонение, асимметрия, эксцесс); амплитудные характеристики (среднеквадратичное значе-

ние, пиковое значение, пик-фактор, форм-фактор, коэффициент импульсности, коэффициент зазора); перцентили и меры разброса (10-й, 25-й, 50-й, 75-й, 90-й перцентили, межквартильный размах, коэффициент вариации); спектральные характеристики (спектральный центроид, спектральный разброс, спектральная асимметрия, спектральный эксцесс, спектральная энтропия); энтропийные характеристики (энтропия, временная энтропия); автокорреляционные признаки (автокорреляция с лагами 1, 5, 10, значение следующего пика); энергетические характеристики (энергия узла, нормализованная энергия).

Общее количество извлеченных признаков составило  $p = 45984$ , что при  $n = 491$  наблюдении формирует сверхвысокоразмерное признаковое пространство с соотношением  $p/n \approx 94$ .

### *Снижение размерности методом главных компонент*

В условиях сверхвысокой размерности ( $p \gg n$ ) применение классификаторов напрямую к исходным признакам приводит к проблеме переобучения [2]. Для ее решения использован метод главных компонент (Principal Component Analysis, PCA) [10].

Пусть  $\bar{X}$  — центрированная матрица наблюдений размером  $n \times p$ . Метод главных компонент выполняет ортогональное линейное преобразование:

$$Y = \bar{X}W, \quad (2)$$

где  $Y$  — матрица главных компонент размером  $n \times p$ ;

$W$  — ортогональная матрица собственных векторов ковариационной матрицы.

Собственные векторы находятся из решения задачи [10]:

$$Sw_i = \lambda_i w_i, \quad (3)$$

где  $S = (1/(n-1))\bar{X}^T\bar{X}$  — выборочная ковариационная матрица;

$\lambda_i$  —  $i$ -е собственное значение;

$w_i$  — соответствующий собственный вектор.

Доля дисперсии, объясняемая  $i$ -й главной компонентой:

$$\rho_i = \frac{\lambda_i}{\sum_{j=1}^r \lambda_j}, \quad (4)$$

где  $r = \min(n-1, p)$  — ранг матрицы данных.

Количество сохраняемых компонент  $d$  определялось из условия достижения порога кумулятивной объясненной дисперсии 95%:

$$\rho_{\text{cum}}(d) = \frac{\sum_{i=1}^d \lambda_i}{\sum_{j=1}^r \lambda_j} \geq 0,95. \quad (5)$$

### Логистическая регрессия с L1-регуляризацией

Для многоклассовой классификации с  $K = 7$  диагностическими классами применена логистическая регрессия со стратегией «один против всех» (One-vs-Rest, OvR) [2]. Для каждого класса  $k$  строится бинарный классификатор, оценивающий вероятность принадлежности объекта  $x$  к данному классу:

$$P(y = k | \mathbf{x}) = \sigma(\mathbf{w}_k^T \mathbf{x} + b_k), \quad (6)$$

где  $\mathbf{w}_k^T$  — вектор весов для класса  $k$ ;

$b_k$  — смещение;

$\sigma(\cdot)$  — логистическая (сигмоидная) функция:

$$\sigma(z) = \frac{1}{1 + e^{-z}}. \quad (7)$$

Для предотвращения переобучения и обеспечения разреженности модели применена L1-регуляризация (Lasso). Оптимизируемая функция потерь имеет вид:

$$\mathcal{L}(\mathbf{w}) = -\sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] + C^{-1} \sum_{j=1}^d |w_j|, \quad (8)$$

где  $y_i$  — истинная метка  $i$ -го наблюдения;  $\hat{y}_i$  — предсказанная вероятность;

$C$  — параметр регуляризации (в исследовании  $C = 1,0$ );

$d$  — количество признаков (главных компонент).

L1-регуляризация обеспечивает обнуление малозначимых коэффициентов, что повышает интерпретируемость модели и снижает риск переобучения.

Итоговое предсказание класса определяется как:

$$\hat{y} = \arg \max_{k \in \{0,1,\dots,K-1\}} P(y = k | \mathbf{x}). \quad (9)$$

### Методология оценки качества

Для объективной оценки обобщающей способности моделей использовалась стратифицированная 5-кратная кросс-валидация. Данные разбивались на пять непересекающихся подмножеств (фолдов) с сохранением пропорций классов в каждом подмножестве. На каждой итерации четыре фолда использовались для

обучения модели, один — для тестирования. Процедура повторялась пять раз, каждый фолд однократно выступал в роли тестового.

Важной особенностью методологии является применение PCA внутри каждого фолда кросс-валидации. Параметры преобразования (средние значения признаков, собственные векторы ковариационной матрицы) вычислялись исключительно на обучающей части фолда и затем применялись к тестовой части. Такой подход исключает утечку информации из тестовой выборки в обучающую и обеспечивает корректную оценку обобщающей способности модели.

Разбиение данных на обучающую (70%) и тестовую (30%) выборки для финальной оценки выполнялось с учетом идентификаторов сигналов, что предотвращает попадание фрагментов одного сигнала в разные выборки.

Качество классификации оценивалось с помощью макроусредненной F1-меры (F1-macro), которая вычисляется как среднее арифметическое F1-мер по всем классам. Данная метрика учитывает качество распознавания каждого класса равнозначно, что важно при несбалансированном распределении классов в выборке.

По результатам кросс-валидации вычислялись два показателя: среднее значение F1-macro по пяти фолдам, характеризующее типичную точность модели, и стандартное отклонение (Std), характеризующее стабильность результатов при различных разбиениях данных. Низкое значение стандартного отклонения свидетельствует о робастности модели и воспроизводимости результатов.

## Результаты исследования

### *Сравнительный анализ методов классификации*

Проведено сравнение семи методов машинного обучения. Результаты представлены в табл. 2.

ТАБЛИЦА 2. Сравнение методов классификации на тестовой выборке

Метод	Ассурасу (точность)	Взвешенная F1-мера	F1 macro	ROC AUC	Время, с
Logistic Regression	0,9932	0,9932	0,9921	0,9905	0,10
Extra Trees	0,9932	0,9931	0,9939	1,0000	0,19
SGD Classifier	0,9932	0,9932	0,9921	0,9922	0,04
Naive Bayes	0,9864	0,9863	0,9867	0,9913	0,02
KNN (k=5)	0,9864	0,9863	0,9867	0,9999	0,22
Random Forest	0,9728	0,9729	0,9743	0,9995	0,25
Hist Gradient Boosting	0,9660	0,9659	0,9649	0,9953	2,48

Три метода достигли одинаковой точности Accuracy = 0.9932: логистическая регрессия, Extra Trees и SGD Classifier.

### Результаты кросс-валидации

Стратифицированная 5-кратная кросс-валидация подтвердила стабильность результатов (табл. 3).

ТАБЛИЦА 3. Результаты стратифицированной 5-кратной кросс-валидации

Метод	F1-macro (CV)	Std
Logistic Regression	99,56 %	± 0,55 %
Extra Trees	99,31 %	± 0,64 %
Hist Gradient Boosting	99,22 %	± 0,40 %
SGD Classifier	99,02 %	± 1,07 %
Random Forest	98,93 %	± 0,88 %
Naive Bayes	98,79 %	± 1,17 %
KNN (k=5)	98,47 %	± 1,17 %

### Обоснование выбора логистической регрессии

После применения PCA данные становятся практически линейно разделимыми. Логистическая регрессия оптимально использует линейную структуру редуцированного признакового пространства. L1-регуляризация обеспечивает разреженность модели, минимизируя риск переобучения. Низкая дисперсия результатов кросс-валидации критически важна для практического применения и воспроизводимости.

Логистическая регрессия выбрана как оптимальный метод по совокупности критериев. Она демонстрирует максимальное среднее значение F1-macro (99,56%) и минимальную дисперсию ( $\pm 0,55\%$ ), что свидетельствует о высокой стабильности и воспроизводимости результатов.

### Визуализация результатов

Проекция данных на плоскость первых двух главных компонент (рис. 1) демонстрирует структуру разделения диагностических классов.

На рис. 1 видно четкое пространственное разделение класса 0 (исправное состояние) от всех классов неисправностей. Классы неисправностей образуют характерную структуру: классы 1–2 группируются в области отрицательных значений PC1, классы 3–6 располагаются в области положительных значений. Такое распределение может отражать конструктивные особенности шестицилиндрового дизельного двигателя Д50.

Матрица ошибок классификации (рис. 2) показывает распределение предсказаний по классам.

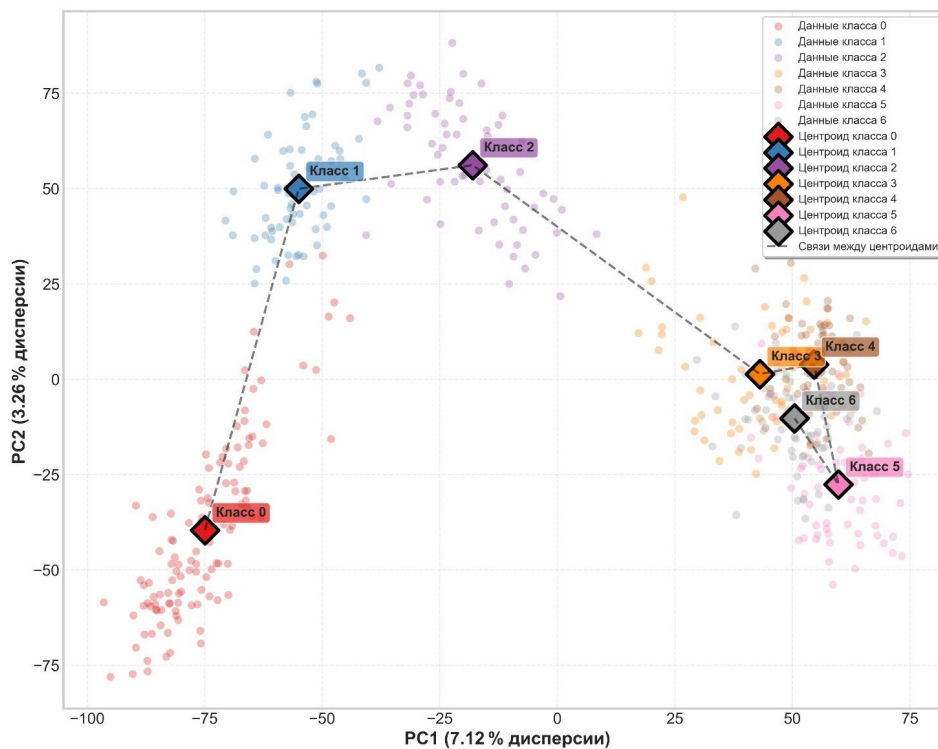


Рис. 1. Проекция данных на плоскость первых двух главных компонент с центриодами классов

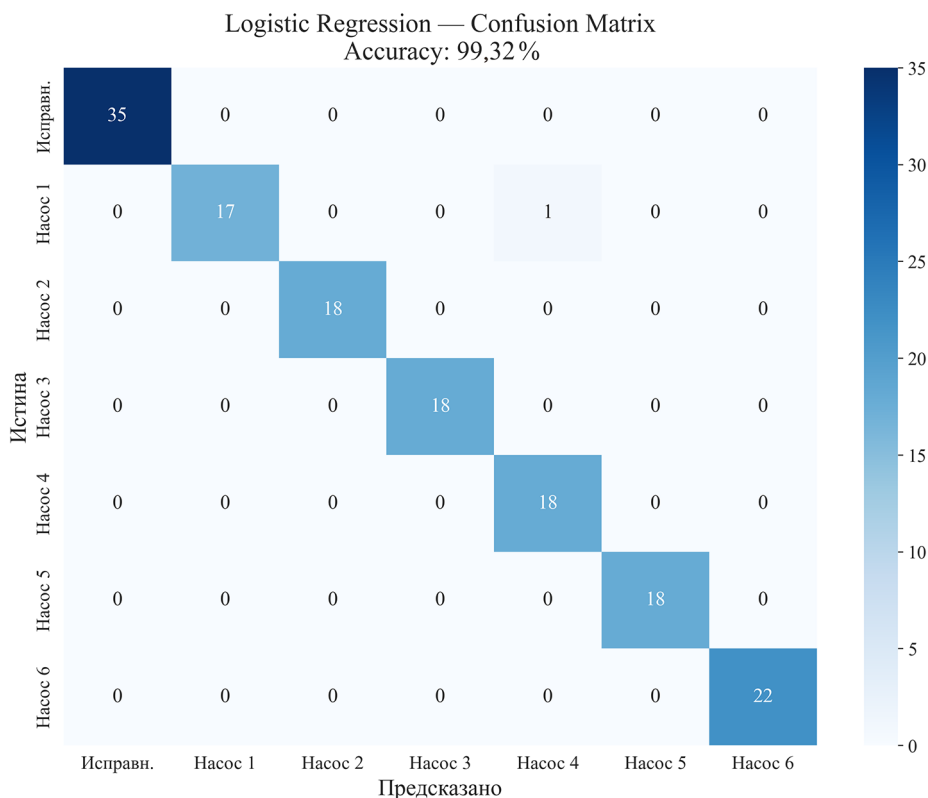
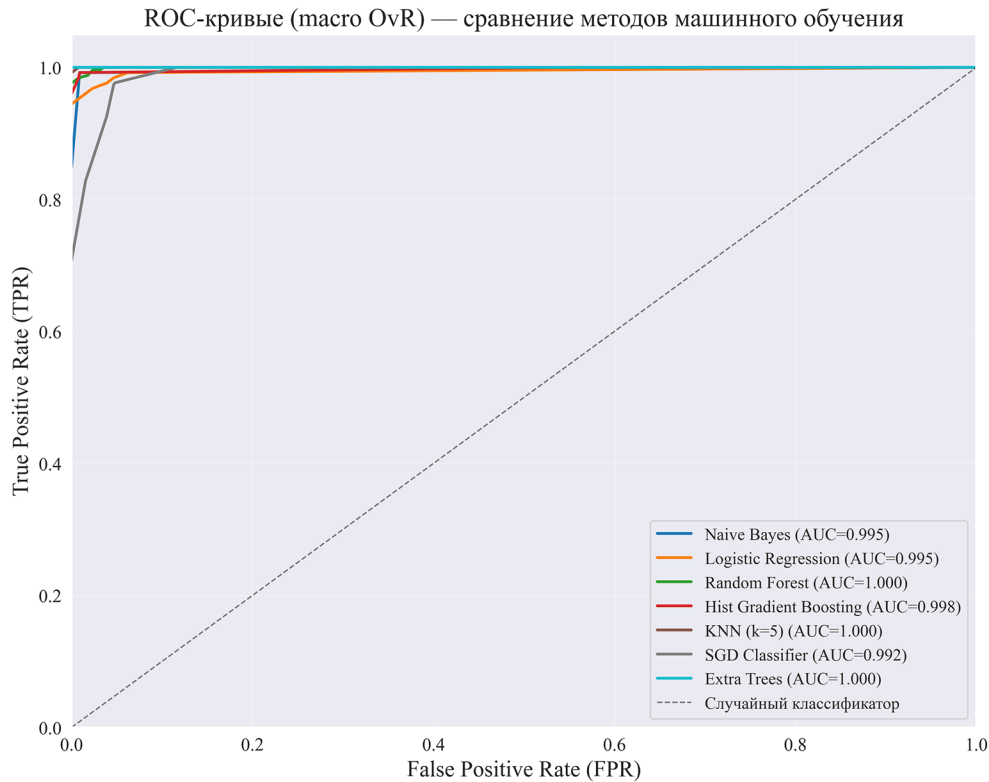


Рис. 2. Матрица ошибок классификации методом логистической регрессии (Accuracy = 99,32%)

Анализ матрицы ошибок выявил единственную ошибку классификации: один образец класса «Насос 1» (отключение 1-го цилиндра) был ошибочно классифицирован как «Насос 4» (отключение 4-го цилиндра).

ROC-кривые для всех исследованных методов представлены на рис. 3.



**Рис. 3.** ROC-кривые (масштаб OvR) для сравнительного анализа методов машинного обучения

Все методы демонстрируют высокие значения площади под ROC-кривой ( $AUC > 0.99$ ), что свидетельствует об отличной разделимости классов в редуцированном признаковом пространстве после PCA.

### Обсуждение и заключение

Проведенное исследование экспериментально подтвердило возможность автоматической классификации технического состояния дизельного двигателя Д50 по диагностическим классам на основе вейвлет-признаков вибросигналов.

Метод главных компонент с порогом 95 % объясненной дисперсии позволил эффективно снизить размерность признакового пространства с 45984 до нескольких сотен компонент, устранив мультиколлинеарность и создав оптимальные условия для работы линейных классификаторов.

Логистическая регрессия с L1-регуляризацией достигла точности классификации 99,32 % на тестовой выборке. Стратифицированная кросс-валидация

подтвердила стабильность результатов:  $F1\text{-macro} = 99,56\% \pm 0,55\%$ . Выбор логистической регрессии обоснован сочетанием максимальной точности, минимальной дисперсии результатов и высокой интерпретируемости модели.

Практическая значимость работы заключается в экспериментальной апробации полного цикла классификации: от извлечения вейвлет-признаков до сравнительного анализа методов машинного обучения с честной оценкой обобщающей способности. Достигнутые результаты создают основу для разработки систем диагностики реального времени.

Ограничения исследования включают проведение экспериментов в контролируемых лабораторных условиях с имитацией неисправностей. Для практического применения требуется валидация на реальных эксплуатационных дефектах.

Перспективы дальнейших исследований включают: расширение набора диагностических классов; валидацию в эксплуатационных условиях; разработку алгоритмов для встраиваемых систем реального времени; исследование применимости глубокого обучения.

## СПИСОК ИСТОЧНИКОВ

1. Jardine A. K. S., Lin D., Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance // *Mechanical Systems and Signal Processing*. 2006. Vol. 20 (7). Pp. 1483–1510.
2. Applications of machine learning to machine fault diagnosis: a review / Y. Lei [et al.] // *Mechanical Systems and Signal Processing*. 2020. Vol. 138.
3. Randall R. B. *Vibration-based condition monitoring: industrial, aerospace and automotive applications*, R. B. Randall. 2nd ed. // Chichester: John Wiley & Sons, 2021, 456 p.
4. Зигельман Е. Б., Лоцинин И. А., Скворцов Д. Ф. Исследование возможности вибродиагностики среднеоборотных дизель-генераторов // *Известия высших учебных заведений. Машиностроение*. 2013. № 6. С. 42–48.
5. Mallat S. *A wavelet tour of signal processing*. 3rd ed. San Diego: Academic Press, 2009. 832 p.
6. Yan R., Gao R. X., Chen X. Wavelets for fault diagnosis of rotary machines: a review with applications // *Signal Processing*. 2014. Vol. 96. Pp. 1–15.
7. Fault diagnosis of diesel engine valve clearance based on wavelet packet decomposition and neural networks / Y. Liu [et al.] // *Electronics*. 2023. Vol. 12 (2). P. 353.
8. Диагностирование газоздушного тракта тепловозного дизеля с использованием интеллектуального классификатора / В. В. Грачев [и др.] // *Бюллетень результатов научных исследований*. 2022. Вып. 2. С. 124–140.
9. Breiman L. Random forests // *Machine Learning*. 2001. Vol. 45 (1). Pp. 5–32.
10. Jolliffe I. T., Cadima J. Principal component analysis: a review and recent developments // *Philosophical Transactions of the Royal Society A*. 2016. Vol. 374 (2065). — 20150202.

Дата поступления: 09.01.2026

Решение о публикации: 11.02.2026

**Контактная информация:**

БЕЛЯЕВ Андрей Александрович — старший преподаватель; belyaevaa@list.ru

## **Multiple-Class Classification of a Diesel Locomotive Engine's Mechanical Condition Based on Wavelet Properties of Vibration Signals via Machine-Learning Techniques**

**A. A. Beliaev**

Emperor Alexander I Petersburg State Transport University, 9 Moskovsky pr., Saint Petersburg, 190031, Russia

**For citation:** *Beliaev A. A.* Multiple-Class Classification of a Diesel Locomotive Engine's Mechanical Condition Based on Wavelet Properties of Vibration Signals via Machine-Learning Techniques // Bulletin of Scientific Research Results, 2026, iss. 1. pp. 142–153. DOI: 10.20295/2223-9987-2026-1-142-153. (In Russian)

### **Abstract**

**Objective:** to develop and experimentally validate a multiclass classification algorithm for assessing the technical condition of a diesel locomotive engine based on wavelet properties derived from vibration signals, employing machine learning methods to ensure the reliability and safety of railway transport. **Methods:** experimental measurements were performed on a D50 diesel engine at the Diesel Locomotive Laboratory named after Professor Ya. M. Gakkel, Emperor Alexander I St. Petersburg State Transport University. A total of 491 triaxial vibration recordings were obtained across seven diagnostic states (one of which represented normal operation and six of which simulated fuel supply faults by cutting off fuel delivery to specific cylinders). Wavelet packet decomposition to the eleventh level, using a fourth-order Daubechies wavelet, yielded 45,984 diagnostic features from the decomposition nodes. Principal Component Analysis (PCA) was employed to minimize the dimensionality of the extracted features, retaining components that together explained 95 % of the variation. Stratified 5-fold cross-validation was used to compare eight machine-learning algorithms in order to ensure an unbiased evaluation of the models' generalization capabilities. **Results:** the highest classification performance of 99.32 % was obtained using logistic regression with L1 regularization (Accuracy = 0.9932, F1-macro = 0.9921). The stratified cross-validation confirmed the stability and reproducibility of this outcome, yielding an F1-macro score of 99.56 %  $\pm$  0.55 %. This method demonstrated the lowest variability across all tested algorithms, indicating strong robustness to data variability. **Practical significance:** the experimental findings have confirmed the feasibility of automatic classification of diesel engines into diagnostic categories based on wavelet properties of vibration signals. Logistic regression has been selected as the optimal method because it provides a favorable balance of predictive accuracy, stability of results, and model interpretability. The developed algorithm can serve as a foundation for real-time on-board diagnostic systems for diesel locomotives.

**Keywords:** machine learning, classification, logistic regression, L1 regularization, wavelet analysis, vibration diagnostics, diesel engine, locomotive, principal component analysis, cross-validation

## References

1. Jardine A. K. S., Lin D., Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing*, 2006, vol. 20 (7). Pp. 1483–1510.
2. Y. Lei [et al.]. Applications of machine learning to machine fault diagnosis: a review // *Mechanical Systems and Signal Processing*, 2020, vol. 138.
3. Randall R. B. *Vibration-based condition monitoring: industrial, aerospace and automotive applications*, R. B. Randall. 2nd ed. // Chichester: John Wiley & Sons, 2021, 456 p.
4. Zigel'man E. B., Loshchinin I. A., Skvortsov D. F. Issledovanie vozmozhnosti vibrodiagnostiki sredneoborotnykh dizel'-generatorov // *Izvestiya vysshikh uchebnykh zavedenij. Mashinostroenie*. 2013. № 6. S. 42–48. (In Russian)
5. Mallat S. *A Wavelet Tour of Signal Processing*. 3rd ed. San Diego: Academic Press, 2009, 832 p.
6. Yan R., Gao R. X., Chen X. Wavelets for fault diagnosis of rotary machines: a review with applications, *Signal Processing*, 2014, vol. 96. Pp. 1–15.
7. Y. Liu [et al.]. Fault diagnosis of diesel engine valve clearance based on wavelet packet decomposition and neural networks, *Electronics*, 2023, vol. 12 (2). P. 353.
8. Diagnostirovanie gazovozdushnogo trakta teplovoznogo dizelya s ispol'zovaniem intellektual'nogo klassifikatora / V. V. Grachev [i dr.] // *Byulleten' rezul'tatov nauchnykh issledovanij*. 2022. Vyp. 2. S. 124–140. (In Russian)
9. Breiman L. Random forests // *Machine Learning*, 2001, vol. 45 (1). Pp. 5–32.
10. Jolliffe I. T., Cadima J. Principal component analysis: a review and recent developments, *Philosophical Transactions of the Royal Society A*, 2016, vol. 374 (2065). 20150202.

Received: 09.01.2026

Accepted: 11.02.2026

### Author's information:

Andrei A. BELIAEV — Senior Lecturer; belyaevaa@list.ru