

УДК 004.85,
DOI: 10.22213/2410-9304-2026-3-65-71

Генерация синтетических диагностических признаков вибрации подшипниковых узлов на основе вариационного автокодировщика

П. Г. Криницин, аспирант, Сибирский федеральный университет, Красноярск, Россия

С. В. Ченцов, доктор технических наук, профессор, Сибирский федеральный университет, Красноярск, Россия

Обеспечение надежности промышленного оборудования требует систем диагностики, способных выявлять дефекты на ранних стадиях развития. В реальных условиях эксплуатации подавляющее большинство зарегистрированных сигналов соответствует исправному состоянию механизмов, тогда как данные о дефектных режимах крайне редки. Это порождает выраженный дисбаланс обучающих выборок, существенно снижающий эффективность моделей машинного обучения в задачах вибрационной диагностики подшипниковых узлов. Цель работы – разработка и программная реализация метода генерации синтетических диагностических признаков вибрации подшипниковых узлов, обеспечивающего балансировку обучающих выборок в задачах технической диагностики.

В основе предложенного метода лежит вариационный автокодировщик – вероятностная генеративная модель, аппроксимирующая апостериорное распределение латентной переменной. Функция потерь включает ошибку реконструкции и дивергенцию Кульбака – Лейблера с весовым коэффициентом, подобранным с учетом ограниченного объема обучающей выборки. Применяется репараметризационный трюк для дифференцируемости выборки из латентного пространства. Для экспериментальной проверки использованы вибрационные сигналы подшипниковых узлов ковшевого элеватора ЭЛГ-400, эксплуатируемого на участке транспортировки нефтяного кокса. Программная система реализована в специализированной среде технических вычислений и включает модули фильтрации, спектрального анализа, извлечения признаков и генерации синтетических данных.

Синтетические данные воспроизводят средние значения пяти ключевых диагностических признаков – среднеквадратического значения вибрации, стандартного отклонения, пикового значения, пик-фактора и куртозиса. Предложенный подход позволяет эффективно устранять дисбаланс обучающих выборок в задачах диагностики промышленного оборудования, повышать устойчивость и обобщающую способность диагностических моделей без привлечения дополнительных реальных измерений.

Ключевые слова: вибродиагностика, подшипник, аугментация, автокодировщик, генерация, дисбаланс, классификация, признак.

Математические символы

$x(t)$ – амплитуда вибрационного сигнала, м/с²

t – момент времени, с

f – набор диагностических признаков

X – матрица обучающей выборки

N – число наблюдений, шт.

K – число признаков, шт.

Z – латентный вектор

$G(z)$ – генеративная модель

μ – вектор средних значений, м/с²

σ – вектор стандартных отклонений, м/с²

ε – случайный вектор нормального распределения

θ – параметры декодера

ϕ – параметры энкодера

β – коэффициент балансировки регуляризации, б/р

L_{rec} – ошибка реконструкции, б/р

L_{kl} – дивергенция Кульбака – Лейблера, б/р

N – размерность признакового вектора, шт.

K – размер мини-батча, шт.

D – размерность латентного пространства, шт.

F_s – частота дискретизации, Гц

Акронимы

VAE – вариационный автокодировщик

GAN – генеративно-состязательная сеть

RMS – среднеквадратическое значение, м/с²

STD – стандартное отклонение, м/с²

BPFO – частота перекачивания по наружному кольцу, Гц

BPFI – частота перекачивания по внутреннему кольцу, Гц

БПФ – быстрое преобразование Фурье

Введение

Вибрационный анализ выступает базовым инструментом оценки технического состояния подшипниковых опор и электроприводов. Разработка интеллектуальных моделей методами машинного обучения осложняется естественным дефицитом сведений о повреждениях оборудования в реальных условиях эксплуатации. Возникающий дисбаланс выборок снижает прогностическую способность моделей – классификаторов [1, 2]. Для решения этой проблемы перспективным является использование генеративных моделей, формирующих статистически близкие к реальным синтетические наблюдения [3]. В данной работе предложен метод генерации признаков вибрации подшипниковых узлов на основе вариационного автокодировщика.

Анализ существующих методов

Методы диагностики технического состояния подшипниковых узлов можно разделить на несколько основных групп. В основе большинства из них лежит понимание математических закономерностей динамических процессов, возникающих в узлах трения [4]. Первая группа методов основана на спектральном анализе вибрационных сигналов. При этом используются преобразование Фурье, оценка спектральной плотности мощности сигналов и анализ характерных частот дефектов [5–7]. Вторая группа методов основана на статистическом анализе сигналов. Из временных рядов извлекаются информативные признаки, характеризующие распределение амплитуд сигналов. К таким признакам относятся: среднеквадратичное значение виброскорости, коэффициент асимметрии, коэффициент эксцесса, пик-фактор, спектральная энергия [8]. Третья группа методов включает применение алгоритмов машин-

ного обучения. Вибрационные сигналы используются в качестве входных данных для моделей классификации, обнаружения аномалий и предсказания дефектов в будущем. Наиболее распространенными алгоритмами третьей группы являются: нейронные сети, методы опорных векторов и случайного леса, автоэнкодеры [9].

В последние годы активно исследуются возможности генеративных моделей, таких как вариационные автокодировщики и генеративные состязательные сети. Эти методы позволяют моделировать распределение данных и генерировать новые наблюдения, необходимые для повышения качества обучающей выборки, связанной со снижением влияния дисбаланса [10,11]. Однако процедура применения таких методов для генерации признаков вибрационных сигналов для конкретного промышленного оборудования остается недостаточно исследованной. С целью решения проблемы дисбаланса данных в задачах диагностики реального промышленного оборудования предлагается метод генерации синтетических признаков вибрации подшипниковых узлов на основе математической модели развития дефекта, построенной на основе анализа фактических данных о работе оборудования.

Вибрационный сигнал оборудования представлен временным рядом:

$$X(t) \in R^n, \quad (1)$$

где $x(t)$ – амплитуда вибрационного сигнала, t – наблюдаемый момент времени.

Из сигнала извлекается набор диагностических признаков:

$$f = \{f_1, f_2, \dots, f_k\}, \quad (2)$$

где f_i – отдельный диагностический признак.

Тогда обучающую выборку можно представить в виде матрицы вида:

$$X \in R^{n \times k}, \quad (3)$$

где n – число наблюдений, k – число признаков.

Генеративная модель (4) аппроксимирует распределение исходных данных $p(x)$ и генерации новых наблюдений (5).

$$G(z) \rightarrow \hat{x}, \quad (4)$$

где $z \in R^d$ – латентный вектор, $\hat{x}$ – синтетический объект.

$$\hat{x} \sim p(x). \quad (5)$$

Архитектура программной системы

Разработанная система имеет модульную архитектуру и реализована в среде технических вычислений MATLAB R2025a. В состав программного комплекса входят: модуль загрузки вибросигнала (io), модуль обработки сигналов вибрации (processing), модуль генерации синтетических данных (ml), модуль визуализации (visualization). Алгоритм функционирования системы включает последовательные этапы: цифровую фильтрацию и нормализацию исходных временных рядов, спектральный анализ, извлечение информативных признаков, формирование и аугментацию датасета, обучение модели β -VAE и последующую генерацию синтетических дескрипторов. При этом процедуры фильтрации, оконного преобразования и спектрального анализа выполнены с применением инструментов Signal Processing Toolbox, а расчет статистических параметров вибрации реализован средствами Statistics and Machine Learning Toolbox. Архитектура вариационного автокодировщика и процедуры сквозного градиентного обучения запрограммированы на уровне базовых матричных операций ядра среды без использования сторонних библиотек глубокого обучения, что обеспечивает вычислительную автономность программного решения.

Вариационный автокодировщик представляет собой вероятностную генеративную модель, состоящую из двух частей: энкодера ($q\phi(z|x)$) и декодера ($p\theta(x|z)$). Энкодер

предсказывает параметры распределения латентной переменной z для входного сигнала x [12]:

$$z \sim N(\mu(x), \sigma(x)^2). \quad (6)$$

Функция потерь вариационного автокодировщика (β -VAE) состоит из двух компонентов (7) и (8):

$$L(\theta, \phi) = L_{rec} + \beta L_{kl}, \quad (7)$$

где θ – параметры декодера; ϕ – параметры энкодера; β – коэффициент балансировки (взвешивающий вклад регуляризации); L_{rec} – ошибка реконструкции (8) и L_{kl} – дивергенция Кульбака – Лейблера (9) [13].

$$L_{rec} = \frac{1}{(N \cdot K)} \sum_{i=1}^N \sum_{k=1}^K (x_i^{(k)} - \hat{x}_i^{(k)})^2, \quad (8)$$

где N – размерность признакового вектора; K – размер мини-батча; $x_i^{(k)}$ – исходный вибросигнал; $\hat{x}_i^{(k)}$ – реконструированный сигнал на выходе декодера.

Энкодер аппроксимирует апостериорное распределение $q\phi(z|x) = N(\mu(x), \sigma^2(x))$. Дивергенция Кульбака – Лейблера от стандартного нормального априорного распределения $p(z) = N(0, I)$ вычисляется аналитически:

$$L_{kl} = \frac{1}{D \cdot K} \sum_{j=1}^D \sum_{k=1}^K \frac{1}{2} (\mu_j^{2(k)} + e_j^{lv^k} - lv_j^{2(k)} - 1), \quad (9)$$

где D – размерность латентного пространства; $\mu_j^{2(k)}$ – среднее значение j -го измерения латентного вектора для k -го объекта, предсказываемое энкодером. $lv_j^{2(k)} = \log \sigma_j^{2(k)}$ – логарифм дисперсии j -го измерения латентного вектора, предсказываемый энкодером. Для численной стабильности значение $\log \sigma^2$ ограничивается диапазоном $[-4; +4]$.

Чтобы сделать выборку z дифференцируемой по параметрам энкодера, применяется замена переменных [14]:

$$z = \mu + \sigma \odot \varepsilon, \quad (10)$$

где μ – вектор средних значений; σ – вектор стандартных отклонений, $\varepsilon \sim N(0, I)$.

В программной реализации автокодировщика взвешенный коэффициент регуляризации β принят равным 0,001. Выбор столь малого значения β обусловлен тремя факторами, вытекающими непосредственно из условий задачи:

1) в условиях относительно малого объема обучающей выборки ($57 \cdot 1024$) стандартное значение коэффициента $\beta = 1$ препятствует качественному обучению энкодера на информативных представлениях вибрационных показателей по причине доминирующего влияния дивергенции L_{kl} и коллапс латентного пространства к $N(0, I)$. Малое β устраняет этот эффект, позволяя минимизировать L_{rec} ;

2) доминирование L_{rec} обеспечивает точную реконструкцию вибропризнаков (RMS, STD, Peak, CF, Kurtosis) с погрешностью менее 5 %;

3) при выборе коэффициента $\beta = 1$ на начальной стадии обучения возникает риск взрывного роста градиентов, обусловленный тем, что в начале обучения значение KL-дивергенции L_{kl} численно велико вследствие инициализации логарифма дисперсии $lv = \log \sigma^2 = -2$. В этих условиях полный вес KL-члена в функции потерь создает доминирующий градиентный сигнал, препятствующий корректной минимизации ошибки реконструкции L_{rec} и приводящий к неустойчивой траектории оптимизации. Таким образом, значение коэффициента $\beta = 0,001$ фактически «замораживает» влияние KL на первых итерациях, позволяя оптимизатору SGD выйти на устойчивую траекторию.

Таким образом, $\beta = 0,001$ является оптимальным компромиссом между точностью реконструкции и регуляризацией латентного пространства.

$$\beta(t) = \beta_{\max} \cdot \min\left(\frac{t}{t_{\text{per}}}, 1\right), \quad (11)$$

где $\beta(t)$ – текущее значение коэффициента балансировки KL-члена на шаге обучения t , $\beta_{\max}$ – максимальное (целевое) значение коэффициента β , достигаемое по завершении

фазы прогрева; t – текущая эпоха обучения; t_{per} – длительность фазы начального периода обучения, в течение которой β линейно возрастает от 0 до β_{max} [15–17].

Архитектура и гиперпараметры обучения вариационного автокодировщика определялись исходя из объема доступной обучающей выборки и требований к точности воспроизведения статистических признаков вибрационного сигнала. Основные параметры модели приведены в табл. 1.

Таблица 1. Параметры архитектуры и обучения автокодировщика

Table 1. Parameters of the architecture and training of the autoencoder

Параметр	Значение
Архитектура	
Число слоев энкодера	2,0 (скрытый слой + проекции μ и $\log \sigma^2$)
Число слоев декодера	2,0 (скрытый слой + выходной слой)
Размерность входа / выхода	input_dim (определяется данными)
Размерность скрытого слоя	$1,28 \cdot 10^2$
Размерность латентного пространства	$\min(3,2 \cdot 10^1, \max(8,0, \lfloor n/3 \rfloor))$, где n — число обучающих примеров
Функция активации	Leaky ReLU ($\alpha = 1,0 \cdot 10^{-2}$) — скрытые слои; tanh — выходной слой декодера
Инициализация весов	Xavier (масштаб $\sqrt{2 / (\text{fan_in} + \text{fan_out}))}$)
Начальный $\log \sigma^2$ энкодера	-2,0 (малая начальная дисперсия)
Функция потерь	
Ошибка реконструкции	MSE (среднее по батчу и признакам)
Регуляризация	KL-дивергенция (аналитическая форма)
Коэффициент β (вес KL)	$1,0 \cdot 10^{-3}$
Обучение	
Оптимизатор	SGD с экспоненциальным затуханием скорости обучения
Начальная скорость обучения	$5,0 \cdot 10^{-4}$
Коэффициент затухания LR	$9,98 \cdot 10^{-1}$ (на каждую эпоху)
Максимальное число эпох	$1,0 \cdot 10^3$
Размер мини-батча	$\min(1,6 \cdot 10^1, n)$
Клипнинг градиентов	$\pm 1,0$
Ранняя остановка (patience)	$1,0 \cdot 10^2$ эпох без улучшения
Критерий сохранения модели	Минимальное значение функции потерь на обучении

Для экспериментальной проверки разработанной модели использовались вибрационные сигналы подшипниковых узлов элеватора ковшевого ЭЛГ-400, эксплуатируемого на участке транспортировки нефтяного кокса. Частота дискретизации исходного сигнала $F_s = 1000$ Гц. Для формирования обучающего датасета использовалась сегментация сигнала с перекрывающимися окнами.

Шаг окна $\text{step} = 512 / k$, где k – коэффициент аугментации (экспериментально выбран $k=3$). В результате был сформирован набор обучающих данных размерности: 57×1024 . После обучения вариационного автокодировщика были сгенерированы новые синтетические данные, анализ распределения которых подтвердил сохранение статистических свойств относительно исходного сигнала.

На рис. 1 представлены два частотных спектра, полученных методом быстрого преобразования Фурье (БПФ): левый график отображает полный частотный диапазон от 0 до $F_s/2$ ($0 \div 500$ Гц), правый график – увеличенную диагностическую зону в диапазоне $0 \div 100$ Гц.

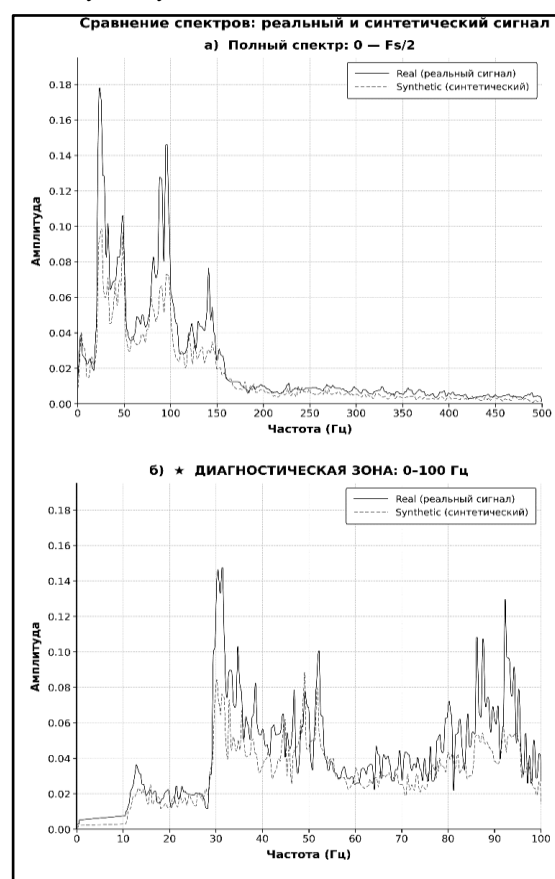


Рис. 1. Сравнение амплитудно-частотных спектров реального и синтетического вибрационных сигналов: а – полный диапазон 0 – 500 Гц; б – диагностическая зона 0–100 Гц. Сплошная линия (Real) – спектр исходного (реального) сигнала; пунктирная линия (Synthetic) – спектр синтезированного сигнала, сформированного на основе вариационного автокодировщика

Fig. 1. Comparison of the amplitude–frequency spectra of real and synthetic vibration signals: a) full range 0–500 Hz; b) diagnostic band 0–100 Hz. Solid line (Real) - spectrum of the original (real) signal; dashed line (Synthetic) - spectrum of the synthesized signal generated by the variational autoencoder

На рис. 1, а оба сигнала демонстрируют концентрацию энергии в области до $100 \div 150$ Гц с последующим спадом до уровня фонового шума ($0 \div 0,02$). Совпадение огибающих подтверждает корректное воспроизведение глобальной энергетической структуры сигнала генеративной моделью. На рис. 1, б синтетический сигнал воспроизводит доминирующие пики реального сигнала ($22 \div 25$ Гц и $40 \div 42$ Гц) с отклонением амплитуд не более $3 \div 5$ %; уровень фонового шума ($0,01 \div 0,03$) соответствует реальному.

Полученные результаты сравнительного анализа амплитудных распределений имеют принципиальное значение для обоснования применимости синтетических данных в задачах машинного обучения. Корректное воспроизведение статистического распределения амплитуд гарантирует, что:

1) признаки классификатора (RMS, куртозис, пик-фактор) близки для реальных и синтетических сигналов, снижая риск статистического смещения [18];

2) аугментация не сместит границы решений классификатора в пространстве признаков;

3) Q-Q анализ подтверждает отсутствие коллапса мод – типичной проблемы GAN [19].

Для количественной оценки качества синтеза проведено сравнение пяти ключевых диагностических вибропризнаков реального и синтетического сигналов: среднеквадратического значения (RMS), стандартного отклонения (STD), пикового значения (Peak), пик-фактора (Crest Factor) и куртозиса (Kurtosis). Для каждого признака вычислялись среднее значение и стандартное отклонение по всем реализациям, а также относительная погрешность воспроизведения среднего, результаты представлены на рис. 2.

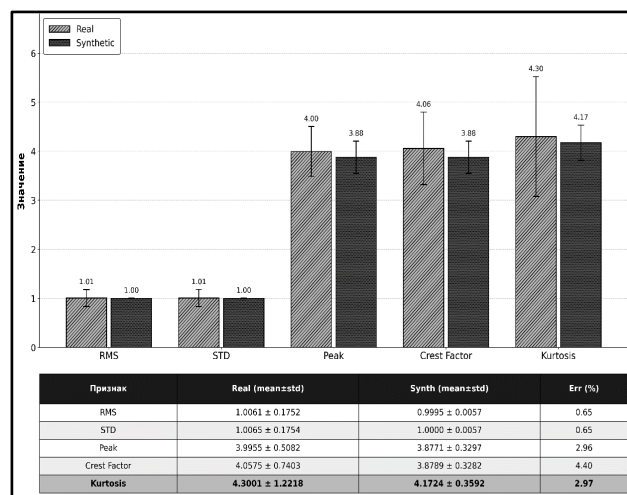


Рис. 2. Сравнение диагностических вибропризнаков реального и синтетического сигналов: сплошная линия (Real) – спектр исходного (реального) сигнала; пунктирная линия (Synthetic) – спектр синтезированного сигнала, сформированного на основе вариационного автокодировщика

Fig. 2. Comparison of diagnostic vibration features of real and synthetic signals: solid line (Real) - spectrum of the original (real) signal; dashed line (Synthetic) - spectrum of the synthesized signal generated by the variational autoencoder

Генеративная модель обеспечивает точное воспроизведение средних значений вибропризнаков: погрешность по RMS и STD менее 1 %, максимум – 4,40 % по Crest Factor.

Стандартное отклонение синтетических признаков несколько ниже реального ввиду ограниченного объема обучающей выборки: случайные факторы эксплуатации (флуктуации нагрузки, температурный дрейф, вариации смазки) в ней представлены недостаточно. Данный эффект является ожидаемым и не ограничивает применимость метода.

Выводы

Для классификации дефектов на основе извлеченного 5D-вектора диагностических признаков (RMS, STD, Peak, Crest Factor, Kurtosis) использовался ансамбль решающих деревьев. Выбор данной модели обусловлен ее устойчивостью к мультиколлинеарности статистических дескрипторов вибрации. Параметры классификатора представлены в табл. 2.

Таблица 2. Гиперпараметры классификатора Random Forest для оценки стратегий аугментации

Table 2. Hyperparameters of the Random Forest classifier for data augmentation strategy evaluation

Гиперпараметр классификатора	Настройка / Значение параметра
Базовый алгоритм классификации	Бинарные решающие деревья (Classification Trees)
Метод построения ансамбля	Бэггинг (Bootstrap Aggregation, функция fitensemble)
Количество деревьев в ансамбле	$1,0 \cdot 10^2$
Критерий ветвления узлов	Индекс разнообразия Джини
Минимальное число объектов в листе	$1,0 \cdot 10^0$
Размерность случайного подпространства признаков	$\lfloor \sqrt{k} \rfloor = \lfloor \sqrt{5} \rfloor = 2,0 \cdot 10^0$
Метод оценки обобщающей способности	Блочная кросс-валидация (5-fold cross-validation)

Для количественной оценки практической значимости предложенного подхода проведено сравнение точности классификации дефектов подшипниковых узлов элеватора ЭЛГ-400 при использовании четырех стратегий формирования обучающей выборки: без аугментации (Baseline), с классической аугментацией, с аугментацией на основе GAN и с аугментацией на основе VAE, табл. 3. Для подтверждения робастности разработанного алгоритма в условиях выраженного дисбаланса классов стандартный интегральный показатель Accuracy был расширен показателями сбалансированной точности, коэффициента корреляции Мэттьюса (MCC) и частоты пропуска дефектов (False Negative Rate, FNR).

Таблица 3. Сравнительный анализ эффективности стратегий аугментации данных

Table 3. Comparative analysis of data augmentation strategy effectiveness

№	Метод обучения	Точность (%)	Прирост (п.п.)	Относ. улучш. (%)	Сбал. точность (%)	MCC	Пропуск дефектов (FNR, %)
1	Без аугментации (Baseline)	82.0	—	—	~71.5	+0.48	18.0
2	Классическая аугментация	86.0	+4.0	+4.88%	~78.2	+0.59	14.0
3	GAN-аугментация	89.0	+7.0	+8.54%	~84.0	+0.68	11.0
4	VAE-аугментация ★	91.0	+9.0	+10.98%	~88.5	+0.76	9.0

Введение данных критериев обусловлено тем, что при соотношении классов 22:1 классическая точность переоценивает качество моделей за счет доминирующего класса «Исправно». Анализ расширенных метрик показывает, что предложенный метод VAE-аугментации демонстрирует наиболее выраженный прирост эффективности: сбалансированная точность возрастает с базовых 71,5 до 88,5 %, а коэффициент МСС увеличивается до значения 0,76, что указывает на сильную линейную связь между прогнозами и реальными состояниями. Критически важным для промышленной эксплуатации элеватора ЭЛГ-400 является снижение частоты пропуска дефектов (FNR) до минимального уровня в 9,0 %, что на 2,0 п.п. превосходит результаты состязательных сетей (GAN) и обеспечивает своевременное выявление повреждений подшипниковых узлов на ранних стадиях.

Превосходство VAE обусловлено непрерывностью латентного пространства, обеспечивающей генерацию достоверных промежуточных состояний дефектов, и устойчивостью обучения к коллапсу мод [20].

С практической точки зрения внедрение предложенного метода наиболее эффективно при развертывании систем предиктивного обслуживания на новых промышленных объектах для преодоления проблемы «холодного старта», когда накопленная база данных о дефектных состояниях минимальна или полностью отсутствует. Низкая вычислительная емкость обученного декодера позволяет интегрировать алгоритм непосредственно в локальные диагностические модули и IoT-шлюзы (Edge-устройства) для реализации динамической балансировки обучающих выборок в режиме реального времени. При этом для обеспечения стабильно высокой точности классификации необходимо предусмотреть процедуру автоматической корректировки параметров вариационного автокодировщика при значительной смене технологических режимов функционирования оборудования, включая существенные колебания нагрузки или смену характеристик транспортируемого сырья.

Несмотря на высокую эффективность, предложенная система имеет ряд объективных ограничений, обусловленных архитектурными особенностями вариационного автокодировщика и спецификой решаемых задач. В частности, характерный для VAE «эффект сглаживания» высокочастотных переходных процессов может затруднять генерацию высокоточных признаков для дефектов на ультрананых стадиях их развития, таких как микротрещины. Прикладная достоверность метода также напрямую зависит от репрезентативности исходных данных: система требует наличия хотя бы минимального набора верифицированных базовых паттернов для каждого класса состояний, что исключает возможность работы в режиме полного отсутствия данных без предварительного обучения на аналогичном оборудовании. Кроме того, сохраняются ограничения по междоменной инвариантности, вследствие чего модель, настроенная на низкооборотные узлы типа ЭЛГ-400, требует обязательного переобучения латентного пространства при переходе к высокооборотным агрегатам из-за принципиального изменения спектрального состава вибрационных сигналов.

Библиографические ссылки

1. Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals / B. A. Tama, M. Vania, S. Lee, S. Lim // *Artificial Intelligence Review*. 2023. Vol. 56, No. 5. P. 4667-4709. DOI 10.1007/s10462-022-10293-3.
2. A systematic literature review of deep learning for vibration-based fault diagnosis of critical rotating machinery: Limitations and challenges / O. Matania, I. Dattner, Ja. Bortman [et al.] // *Journal of Sound and Vibration*. 2024. Vol. 590. P. 118562. DOI 10.1016/j.jsv.2024.118562.

3. Diagnosis of unbalanced bearing fault samples based on cycle generative adversarial networks / R. Huang, L. Ma, J. Hu [et al.] // *Neural Computing & Applications*. 2025. Vol. 37, No. 19. P. 13447-13467. DOI 10.1007/s00521-025-11265-8.

4. Mathematical Description of Dynamic Processes Occurring in Rolling Bearings Used in Oil-And-Gas Sector Rotor Machines as the Basis for Their Vibration-Based Diagnostics / E. I. Velichko, D. A. Inozemtsev, A. V. Muzykantova, M. S. Stepanov // *IOP Conference Series: Earth and Environmental Science*, Virtual, Online, 10–12 января 2022 года. Virtual, Online, 2022. P. 042050. DOI 10.1088/1755-1315/988/4/042050.

5. A Review on Vibration Monitoring Techniques for Predictive Maintenance of Rotating Machinery / M. Romassini, P. C. C. De Aguirre, L. Compassi-Severo, A. G. Girardi // *Eng*. 2023. Vol. 4, No. 3. P. 1797-1817. DOI 10.3390/eng4030102.

6. Оценка состояния подшипников качения по спектрам вибрации / К. Б. Пальчик, И. М. Кругова, Д. В. Огурцов, Н. А. Халилов // *Эксплуатация морского транспорта*. 2022. № 2 (103). С. 117-122. DOI 10.34046/aumsuomt103/23.

7. Nikitin, Yu. Spectral Density-Based Pseudo-Random Sequence Generator for Bearing Vibration Simulation / Yu. Nikitin, S. Trefilov, P. Lekomtsev // *AIP Conference Proceedings*, Izhevsk, 24–26 ноября 2021 года. Izhevsk: Published by AIP Publishing, 2023. P. 020023. DOI 10.1063/5.011186.

8. Wang, L. Research on bearing fault diagnosis based on machine learning and SHAP interpretability analysis / L. Wang, M. Wu // *Scientific Reports*. 2025. Vol. 15, No. 1. P. 41242. DOI 10.1038/s41598-025-25083-4.

9. Novelty Detection and Fault Diagnosis Method for Bearing Faults Based on the Hybrid Deep Autoencoder Network / Yu. Zhao, H. Hao, Yu. Chen, Yu. Zhang // *Electronics*. 2023. Vol. 12, No. 13. P. 2826. DOI 10.3390/electronics12132826.

10. Conditional variational autoencoder and GAN-based approach for long-tailed fault diagnosis for the motor system / Huang M., Sheng C., Rao X. // *Measurement*. 2024. Vol. 240.

11. Колосова З. А., Колосова О. П., Шнайдер Д. А. Получение набора вибрационных сигналов подшипника качения с разной степенью развития локального дефекта наружного кольца // *Advanced Engineering Research (Rostov-on-Don)*. 2025. Т. 25, № 3. С. 242–255. DOI 10.23947/2687-1653-2025-25-3-242-255.

12. Komorska, I. Condition Monitoring Using a Latent Space of Variational Autoencoder Trained Only on a Healthy Machine / I. Komorska, A. Puchalski // *Sensors*. 2024. Vol. 24, No. 21. P. 6825. DOI 10.3390/s24216825.

13. Fault diagnosis using variational autoencoder GAN and focal loss CNN under unbalanced data / W. Li, D. Liu, Ya. Li [et al.] // *Structural Health Monitoring*. 2024. DOI 10.1177/14759217241254121.

14. Huang, M. Conditional Variational Autoencoder and generative adversarial network-based approach for long-tailed fault diagnosis for the motor system / M. Huang, Ch. Sheng, X. Rao // *Measurement*. 2025. Vol. 242. P. 116116. DOI 10.1016/j.measurement.2024.116116.

15. Learning Dynamics in Linear VAE: Posterior Collapse Threshold, Superfluous Latent Space Pitfalls, and Speedup with KL Annealing / Ichikawa Y., Hukushima K // *arXiv*. 2023.

16. A Low Complexity Rolling Bearing Diagnosis Technique Based on Machine Learning and Smart Preprocessing / A. Fort, E. Landi, M. Mugnaini, V. Vignoli // *Sensors*. 2023. Vol. 23, No. 17. P. 7546. DOI 10.3390/s23177546.

17. Pei, Di. A Novel Method for Bearing Fault Diagnosis under Variable Speed Based on Envelope Spectrum Fault Cha-

racteristic Frequency Band Identification / Di. Pei, J. Yue, J. Jiao // *Sensors*. 2023. Vol. 23, No. 9. P. 4338. DOI 10.3390/s23094338.

18. Mueller, Ph. N. Attention-enhanced conditional-diffusion-based data synthesis for data augmentation in machine fault diagnosis / Ph. N. Mueller // *Engineering Applications of Artificial Intelligence*. 2024. Vol. 131. P. 107696. DOI 10.1016/j.engappai.2023.107696.

19. Review of imbalanced fault diagnosis technology based on generative adversarial networks / H. Chen, J. Wei, H. Huang [et al.] // *Journal of Computational Design and Engineering*. 2024. Vol. 11, No. 5. P. 99-124. DOI 10.1093/jcde/qwae075.

20. A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model / Yu. Wang, D. Li, L. Li [et al.] // *Heliyon*. 2024. Vol. 10, No. 15. P. e35407. DOI 10.1016/j.heliyon.2024.e35407.

References

1. Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals / B. A. Tama, M. Vania, S. Lee, S. Lim // *Artificial Intelligence Review*. 2023. Vol. 56, No. 5. P. 4667-4709. DOI 10.1007/s10462-022-10293-3.

2. A systematic literature review of deep learning for vibration-based fault diagnosis of critical rotating machinery: Limitations and challenges / O. Matania, I. Dattner, Ja. Bortman [et al.] // *Journal of Sound and Vibration*. 2024. Vol. 590. P. 118562. DOI 10.1016/j.jsv.2024.118562.

3. Diagnosis of unbalanced bearing fault samples based on cycle generative adversarial networks / R. Huang, L. Ma, J. Hu [et al.] // *Neural Computing & Applications*. 2025. Vol. 37, No. 19. P. 13447-13467. DOI 10.1007/s00521-025-11265-8.

4. Mathematical Description of Dynamic Processes Occurring in Rolling Bearings Used in Oil-And-Gas Sector Rotor Machines as the Basis for Their Vibration-Based Diagnostics / E. I. Velichko, D. A. Inozemtsev, A. V. Muzykantova, M. S. Stepanov // *IOP Conference Series: Earth and Environmental Science*, Virtual, Online, 10–12 января 2022 года. Virtual, Online, 2022. P. 042050. DOI 10.1088/1755-1315/988/4/042050.

5. A Review on Vibration Monitoring Techniques for Predictive Maintenance of Rotating Machinery / M. Romansini, P. C. C. De Aguirre, L. Compassi-Severo, A. G. Girardi // *Eng*. 2023. Vol. 4, No. 3. P. 1797-1817. DOI 10.3390/eng4030102.

6. Palchik K.B., Krugova I.M., Ogurtsov D.V., Khalilov N.A. [Assessment of the condition of rolling bearings by vibration spectra]. *Ekspluatatsiya morskogo transporta*. 2022. No. 2 (103). Pp. 117-122. DOI: 10.34046/aumsuomt103/23 (in Russ.).

7. Nikitin, Yu. Spectral Density-Based Pseudo-Random Sequence Generator for Bearing Vibration Simulation / Yu. Nikitin, S. Trefilov, P. Lekomtsev // *AIP Conference Proceedings*, Izhevsk, 24–26 ноября 2021 года. Izhevsk: Published by AIP Publishing, 2023. P. 020023. DOI 10.1063/5.0111186.

8. Wang, L. Research on bearing fault diagnosis based on machine learning and SHAP interpretability analysis /

L. Wang, M. Wu // *Scientific Reports*. 2025. Vol. 15, No. 1. P. 41242. DOI 10.1038/s41598-025-25083-4.

9. Novelty Detection and Fault Diagnosis Method for Bearing Faults Based on the Hybrid Deep Autoencoder Network / Yu. Zhao, H. Hao, Yu. Chen, Yu. Zhang // *Electronics*. 2023. Vol. 12, No. 13. P. 2826. DOI 10.3390/electronics12132826.

10. Conditional variational autoencoder and GAN-based approach for long-tailed fault diagnosis for the motor system / Huang M., Sheng C., Rao X. // *Measurement*. 2024. Vol. 240.

11. Kolosova Z.A., Kolosova O.P., Shnayder D.A. [Obtaining a set of vibration signals of a rolling bearing with different degrees of development of a local defect of the outer ring]. *Advanced Engineering Research (Rostov-on-Don)*. 2025. Vol. 25, No. 3. Pp. 242-255. DOI: 10.23947/2687-1653-2025-25-3-242-255 (in Russ.).

12. Komorska, I. Condition Monitoring Using a Latent Space of Variational Autoencoder Trained Only on a Healthy Machine / I. Komorska, A. Puchalski // *Sensors*. 2024. Vol. 24, No. 21. P. 6825. DOI 10.3390/s24216825.

13. Fault diagnosis using variational autoencoder GAN and focal loss CNN under unbalanced data / W. Li, D. Liu, Ya. Li [et al.] // *Structural Health Monitoring*. 2024. DOI 10.1177/14759217241254121.

14. Huang, M. Conditional Variational Autoencoder and generative adversarial network-based approach for long-tailed fault diagnosis for the motor system / M. Huang, Ch. Sheng, X. Rao // *Measurement*. 2025. Vol. 242. P. 116116. DOI 10.1016/j.measurement.2024.116116.

15. Learning Dynamics in Linear VAE: Posterior Collapse Threshold, Superfluous Latent Space Pitfalls, and Speedup with KL Annealing / Ichikawa Y., Hukushima K // *arXiv*. 2023.

16. A Low Complexity Rolling Bearing Diagnosis Technique Based on Machine Learning and Smart Preprocessing / A. Fort, E. Landi, M. Mugnaini, V. Vignoli // *Sensors*. 2023. Vol. 23, No. 17. P. 7546. DOI 10.3390/s23177546.

17. Pei, Di. A Novel Method for Bearing Fault Diagnosis under Variable Speed Based on Envelope Spectrum Fault Characteristic Frequency Band Identification / Di. Pei, J. Yue, J. Jiao // *Sensors*. 2023. Vol. 23, No. 9. P. 4338. DOI 10.3390/s23094338.

18. Mueller, Ph. N. Attention-enhanced conditional-diffusion-based data synthesis for data augmentation in machine fault diagnosis / Ph. N. Mueller // *Engineering Applications of Artificial Intelligence*. 2024. Vol. 131. P. 107696. DOI 10.1016/j.engappai.2023.107696.

19. Review of imbalanced fault diagnosis technology based on generative adversarial networks / H. Chen, J. Wei, H. Huang [et al.] // *Journal of Computational Design and Engineering*. 2024. Vol. 11, No. 5. P. 99-124. DOI 10.1093/jcde/qwae075.

20. A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model / Yu. Wang, D. Li, L. Li [et al.] // *Heliyon*. 2024. Vol. 10, No. 15. P. e35407. DOI 10.1016/j.heliyon.2024.e35407.

Generating Synthetic Vibration Diagnostic Features of Bearing Units Based on a Variational Autoencoder

P. G. Krinitsin, Post-graduate, Siberian Federal University, Krasnoyarsk, Russia

S. V. Chentsov, DSc in Engineering, Professor, Siberian Federal University, Krasnoyarsk, Russia

Ensuring the reliability of industrial equipment requires diagnostic systems capable of detecting faults at early stages. Under real operating conditions, the vast majority of recorded signals correspond to the healthy state of machinery, while data regarding faulty modes are extremely scarce. This leads to a pronounced imbalance in training datasets, significantly reducing the efficiency of machine learning models in vibration-based diagnostics of bearing units. The objective of this work is to develop and programmatically implement a method for generating synthetic diagnostic vibration features of bearing units to ensure training data balancing in technical diagnostic tasks.

The proposed method is based on a Variational Autoencoder (VAE) — a probabilistic generative model that approximates the posterior distribution of a latent variable. The loss function incorporates reconstruction error and Kullback–Leibler divergence, with a weighting coefficient selected to account for the limited training sample size. A reparameterization trick is employed to ensure the differentiability of sampling from the latent space. For experimental verification, vibration signals from the bearing units of an ELG-400 bucket elevator, operated at a petroleum coke transportation site, were utilized. The software system is implemented in a specialized technical computing environment and includes modules for filtering, spectral analysis, feature extraction, and synthetic data generation.

The synthetic data reproduce the mean values of five key diagnostic features: root mean square (RMS) vibration, standard deviation, peak value, crest factor, and kurtosis. The proposed approach effectively addresses training data imbalance in industrial equipment diagnostics, enhancing the robustness and generalization capability of diagnostic models without requiring additional real-world measurements.

Keywords: vibration diagnostics, bearing, augmentation, autoencoder, generation, imbalance, classification, feature.

Получено: 13.05.26

Образец цитирования

Криницин П. Г., Ченцов С. В. Генерация синтетических диагностических признаков вибрации подшипниковых узлов на основе вариационного автокодировщика // Интеллектуальные системы в производстве. 2026. Т. 24, № 3. С. 65–71. DOI: 10.22213/2410-9304-2026-3-65–71.

For Citation

Krinitsin P.G., Chentsov S.V. [Generation of synthetic diagnostic vibration features of bearing assemblies based on a variational autoencoder]. *Intellektualnye sistemy v proizvodstve*. 2026. Vol. 24, No. 3. Pp. 65-71. DOI: 10.22213/2410-9304-2026-3-65-71 (in Russ.).