

Метод диагностики неисправностей подшипников на основе многомасштабной сети слияния внимания с использованием архитектуры с двумя ветвями

В. В. Потехин, Ч. Го✉

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Россия
✉ chenxiguo.academic@gmail.com

Аннотация. Для решения задач диагностики неисправностей качения подшипников – включая извлечение признаков с единой шкалы, высокую вычислительную сложность при моделировании дальнедействующих временных зависимостей и недостаточное слияние мультимодальных признаков – в данном исследовании предлагается метод диагностики неисправностей на основе сети мультишкального внимания и слияния (MultiScale Attention and Fusion Network, MSAFN). Этот подход использует двухветвевую архитектуру для совместной обработки временно-частотных карт и статистических векторных признаков. Во-первых, разрабатывается динамический многомасштабный сверточный модуль, использующий параллельные многомерные сверточные ядра для адаптивной настройки приемного поля, захватывая признаки неисправностей на различных уровнях гранулярности. Во-вторых, вводится остаточный механизм двойного внимания и модуль селективного пространства состояний (Selective State Space Module, SSM). Это улучшает восприятие критических компонентов частотной области, одновременно достигая эффективного моделирования дальнедействующих временных зависимостей в вибрационных сигналах с линейной вычислительной сложностью. На этой основе предлагается иерархическая прогрессивная стратегия слияния. Эта стратегия использует механизмы кросс-модального управления для обеспечения глубокого взаимодействия между разнородными признаками на различных уровнях абстракции. В сочетании с рамочной моделью многозадачного совместного обучения она оптимизирует задачи классификации, реконструкции и контрастного обучения для повышения обобщенной устойчивости модели. Экспериментальные результаты на наборах данных университета Кейс Вестерн Резерв (Case Western Reserve University, CWRU) и университета Падерборна (Paderborn University, PU) демонстрируют точности диагностики 99.67 и 96.0 % соответственно, превосходя основные модели глубокого обучения и обладая высокой интерпретируемостью.

Ключевые слова: качающиеся подшипники, диагностика неисправностей, многомасштабное внимание, мультимодальное слияние, многозадачное обучение, анализ во временно-частотной области

Для цитирования: Потехин В. В., Го Ч. Метод диагностики неисправностей подшипников на основе многомасштабной сети слияния внимания с использованием архитектуры с двумя ветвями // Изв. СПбГЭТУ «ЛЭТИ». 2026. Т. 19, № 7. С. 72–93. doi: 10.32603/2071-8985-2026-19-7-72-93.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Финансирование: Исследование финансировалось Китайским советом по стипендиям (CSC.202309810002).

Original article

A Method for Fault Diagnosis of Bearings Based on a Multi-Scale Attention Fusion Network Using Dual-Branch Architecture

V. V. Potekhin, Ch. Guo✉

Peter the Great Saint Petersburg Polytechnic University, Saint Petersburg, Russia
✉ chenxiguo.academic@gmail.com

Abstract. To solve the problems of diagnosing bearing rolling faults – including extracting features from a single scale, high computational complexity in modeling long-range temporal dependencies and insufficient fusion of

multimodal features – this study proposes a fault diagnosis method based on the Multi-Scale Attention and Fusion Network (MSAFN). This approach uses a two-branch architecture for joint processing of time-frequency maps and statistical vector features. First, a dynamic multi-scale convolution module is developed that uses parallel multidimensional convolution kernels for adaptive reception field tuning, capturing fault features at different levels of granularity. Second, a residual dual attention mechanism and a Selective State Space Module (SSM) are introduced. This improves the perception of critical components in the frequency domain while achieving effective modeling of long-range temporal dependencies in vibration signals with linear computational complexity. Based on this, a hierarchical progressive fusion strategy is proposed. This strategy utilizes cross-modal control mechanisms to ensure deep interaction between heterogeneous features at different levels of abstraction. Combined with a multi-task collaborative learning framework, it jointly optimizes classification, reconstruction, and contrastive learning tasks to improve the generalized robustness of the model. Experimental results on the Case Western Reserve University (CWRU) and Paderborn University (PU) datasets demonstrate diagnostic accuracies of 99.67 and 96.0 %, respectively, outperforming baseline deep learning models and exhibiting high interpretability.

Keywords: rolling bearings, fault diagnosis, multi-scale attention, multimodal fusion, multitasking learning, time-frequency domain analysis

For citation: Potekhin V. V., Guo Ch. A Method for Fault Diagnosis of Bearings Based on a Multi-Scale Attention Fusion Network Using Dual-Branch Architecture // LETI Transactions on Electrical Engineering & Computer Science. 2026. Vol. 19, no. 7. P. 72–93. doi: 10.32603/2071-8985-2026-19-7-72-93.

Conflict of interests. The authors declare no conflicts of interests.

Funding: The research was funded by the China Scholarship Council (CSC.202309810002).

Введение. Рабочее состояние качающихся подшипников, как наиболее широко используемых компонентов во вращающемся оборудовании, напрямую влияет на надежность и безопасность всей системы оборудования [1]. Однако несвоевременное обнаружение и устранение неисправностей подшипников может привести к незапланированным простоям оборудования, что в лучшем случае вызовет экономические потери, а в худшем – приведет к катастрофическим авариям, угрожающим безопасности персонала [2]. С развитием современной промышленности в сторону интеллектуальных и беспилотных операций модели обслуживания оборудования постепенно переходят от традиционного коррективного и планового к предсказательному обслуживанию. На этом фоне методы диагностики неисправностей подшипников, основанные на вибрационных сигналах, привлекают значительное внимание как в академической среде, так и в промышленности, становясь исследовательской горячей темой в области мониторинга состояния машин и управления их здоровьем [3].

Традиционные методы диагностики неисправностей подшипников в основном опираются на комбинированную стратегию ручного извлечения признаков и поверхностных классификаторов. Признаки, чувствительные к неисправностям, извлекаются из вибрационных сигналов с использованием таких методов, как статистика во

временной области, спектральный анализ или преобразования в ней, после чего проводится классификация неисправностей с помощью таких алгоритмов машинного обучения, как машины опорных векторов или случайные леса [4], [5]. Однако производительность подобных методов сильно зависит от опыта эксперта в области. Разработанные признаки плохо адаптируются к изменяющимся условиям эксплуатации, что ограничивает их возможности для обобщения в сложных и динамичных промышленных сценариях [6]. В последние годы быстрый прогресс технологий глубокого обучения привел к появлению новых исследовательских парадигм в области диагностики неисправностей [7]. Сверточные нейронные сети (CNN), рекуррентные нейронные сети (RNN) и их вариации широко применяются для автоматического обучения признаков из вибрационных сигналов, демонстрируя превосходную диагностическую производительность по сравнению с традиционными методами [8], [9]. Тем не менее, существующие подходы глубокого обучения все еще сталкиваются с рядом проблем: стандартные механизмы внимания недостаточно моделируют характеристики частотной области спативременных признаков; традиционные стратегии слияния не могут эффективно интегрировать дополнительную информацию из спативременных представлений и статистических признаков; а одноуровневое обучение с учителем под-

вержено переобучению при ограниченном числе размеченных образцов.

Для решения вышеупомянутых проблем в данной статье предлагается сеть мульти-шкального внимания и слияния (MSAFN) для диагностики неисправностей качающихся подшипников. Этот метод использует двухветвевую архитектуру для совместной обработки временно-частотных карт и статистических векторных признаков. С помощью тщательно разработанных механизмов многомасштабного извлечения признаков, усиления внимания и кросс-модального слияния, достигается комплексное представление и точная идентификация информации о неисправности. Основные вклады данной работы можно кратко изложить следующим образом:

1. Введение динамического многомасштабного сверточного модуля. С помощью параллельного использования сверточных ядер разных масштабов и их комбинирования с механизмом управления для адаптивного обучения весов важности для каждого масштаба, этот модуль достигает динамического восприятия паттернов неисправностей на различных уровнях гранулярности, преодолевая ограничения единого приемного поля в захвате многомасштабных признаков.

2. Разработка остаточного модуля двойного внимания, который последовательно интегрирует механизмы внимания канала, внимания пространства и внимания к координатам. Это всесторонне улучшает дискриминационную способность признаков по трем измерениям: важность канала, пространственная выразительность и чувствительность к позициям. Кроме того, введен модуль селективного состояния для моделирования долгосрочных временных зависимостей с линейной сложностью.

3. Сконструирован иерархический прогрессивный модуль слияния, который одновременно вводит кросс-модальные взаимодействия между векторными признаками и несколькими уровнями абстракции в ветви кодирования спектрограммы. Механизм управления динамически регулирует коэффициент вклада обоих типов признаков, эффективно извлекая дополнительную информацию из мультимодальных данных и преодолевая узкое место интеграции информации в традиционных одноуровневых стратегиях слияния.

4. Введение стратегии многозадачного совместного обучения. Помимо основной задачи классификации, она совместно оптимизирует две вспомогательные задачи: реконструкцию призна-

ков и контролируемое контрастивное обучение. Ограничивая процесс обучения в пространстве признаков через многомерные управляющие сигналы, повышают способность модели к обобщению при ограниченном числе размеченных образцов.

Связанные работы. Развитие методов диагностики неисправностей подшипников в основном прошло через два этапа: традиционная диагностика на основе обработки сигналов и интеллектуальная диагностика на основе глубокого обучения.

Традиционные подходы в основном опираются на теорию обработки сигналов для разработки признаков [10]. Авторы [11] предложили метод извлечения слабых признаков неисправностей качающихся подшипников, основанный на мультисенсорном оптимизационном алгоритме (MVO) – оптимизированном алгоритме минимальной энтропии деконволюции (MOMEDA). Этот метод повышает точность идентификации неисправностей в условиях сильных помех благодаря новым метрикам многозадачной оптимизации, глобальной поисковой способности MVO и улучшенной обработке сигналов. Авторы [12] представили адаптивный метод извлечения вариационной моды (AVME-MDMO), основанный на многозадачной многопараметрической оптимизации. Автоматически оптимизируя параметры VME с использованием многозадачного оптимизатора серого волка (MOGWO), метод улучшает возможности извлечения признаков неисправности качающихся подшипников, устраняя необходимость в ручной настройке параметров. Авторы [13] предложили метод симплектической геометрической декомпозиции мод (SGMD), улучшенный динамическим временем выравнивания (DTW), с использованием графовой схожести для выбора эффективных компонентов. Применяя SGMD для подавления шума и ввода графовую схожесть для выбора эффективных компонентов, они повысили точность традиционного DTW в диагностике неисправностей качающихся подшипников. Однако такие методы сильно зависят от ручного извлечения признаков, качество которых часто ограничено опытом эксперта, что ограничивает их способность к обобщению при сложных эксплуатационных условиях.

С развитием глубокого обучения диагностика неисправностей на основе интеллектуальных методов из конца в конец стала основным направлением. Сверточные нейронные сети (CNN) стали пионерами в этой области благодаря своим мощным возможностям извлечения локальных при-

знаков. Авторы [14] предложили метод проектирования сверточной нейронной сети, ориентированный на физику (PGCNN). Анализируя физические характеристики сигналов неисправностей подшипников, они оптимизировали размер входных данных и размер сверточных ядер в CNN, что позволило улучшить производительность модели в диагностике неисправностей. Впоследствии архитектура остаточной сети (ResNet), предложенная авторами [15], была широко адаптирована для диагностики неисправностей. Авторы [16] представили метод диагностики неисправностей качающихся подшипников, основанный на вейвлет-преобразовании (WT) и улучшенной остаточной нейронной сети (IResNet) (WT-IResNet). Этот подход включает новаторские слои пулинга и адаптивную стратегию глобального сингулярного разложения (SVD) для извлечения признаков, а также новую функцию потерь для решения задачи диагностики неисправностей с шумными метками.

Одновременно, используя временные характеристики вибрационных сигналов, рекуррентные нейронные сети (RNN) и их вариант – сети долгосрочной и краткосрочной памяти (LSTM) – были использованы для захвата временных зависимостей. Авторы [17] предложили новый метод диагностики неисправностей, объединяющий обработку сдвигающегося окна с моделью CNN-LSTM. Извлекая признаки через слои CNN и захватывая информацию о временных задержках неисправностей через слои LSTM, этот подход значительно улучшает точность диагностики и устойчивость к шуму. В последние годы архитектура Transformer приобрела популярность в машинном обучении благодаря преимуществам механизма самовнимания для моделирования глобальных зависимостей [18]. Авторы [19] предложили структуру диагностики неисправностей, основанную на генеративных состязательных сетях (GAN) с многозадачной преобразовательной трансформацией признаков (MDFT-GAN). Объединив отображение во временно-частотной области с преобразователями признаков Transformer, они улучшили качество образцов, диагностическую производительность и интерпретируемость модели. Авторы [20] представили модель Time-Channel Attention Convolutional Transformer (TCAC-Transformer) для эффективного извлечения признаков и реального времени диагностики неисправностей.

Ограничения существующих методов. Во-первых, механизмы извлечения признаков и вни-

мания не обладают специфичностью. Традиционные сверточные сети используют фиксированные приемные поля, что затрудняет адаптивное захватывание многомасштабных признаков, которые объединяют детализированные воздействия и крупномасштабные периодичности. В то же время, существующие механизмы внимания опираются в основном на глобальный средний пулинг, игнорируя разреженные характеристики распределения сигналов неисправности подшипников в частотной области. Это препятствует моделям точно локализовывать критические компоненты частот неисправностей и приводит к недостаточному кодированию пространственного распределения признаков.

Во-вторых, существуют узкие места эффективности при моделировании дальнедействующих зависимостей и мультимодального слияния. Ограниченные локальными индуктивными смещениями, CNN испытывают трудности с эффективным захватом долгосрочных временных корреляций в эволюции неисправностей. Хотя архитектуры Transformer обладают глобальными моделирующими способностями, их квадратичная вычислительная сложность ограничивает производительность в реальном времени для промышленного применения. Более того, существующие стратегии мультимодального слияния часто полагаются на поверхностную конкатенацию или простое взвешенное суммирование, пренебрегая гетерогенной природой и требованиями глубокого взаимодействия между различными модальностями, что не позволяет полностью использовать дополнительную информацию.

Наконец, обучение с одним сигналом от учителя накладывает ограничения на обобщение. Большинство моделей полагаются исключительно на функцию потерь кросс-энтропии от одной задачи классификации для оптимизации, что делает их подверженными переобучению при ограниченном числе образцов или дисбалансе классов. Этот единственный сигнал учителя не позволяет обеспечить компактность внутри классов и разделимость между классами в пространстве признаков, тем самым ослабляя обобщенную устойчивость модели при изменяющихся эксплуатационных условиях.

Для устранения этих ограничений предложенный метод MSAFN достигает адаптивного выбора приемного поля через динамические многомасштабные свертки. Он включает в себя внимание к каналам в частотной области для улучшения восприятия критических частотных ком-

понентов, использует остаточное двойное внимание и модуль селективного состояния (SSM) для балансировки важности признаков с эффективным моделированием дальнедействующих зависимостей и систематически строит устойчивую и эффективную структуру диагностики неисправностей через иерархическое прогрессивное слияние и стратегии многозадачного совместного обучения.

Предложенная модель. Предложенная в данной статье сеть мультишкального внимания и слияния (MSAFN) использует двухветвевую архитектуру для совместной обработки гетерогенных признаков. Общая структура модели представлена на рис. 1, а, а детальная структура блока MSAFN – на рис. 1, б.

Условные обозначения на рис. 1: X_s – трехканальная временно-частотная карта; x_v – статистический вектор признаков; блоки 1–4 – модули MSAFN с 64, 128, 256, 512 каналами соответственно; GAP – глобальный средний пулинг; FC – полностью связанный слой; BN – нормализация по батчам; HPF – иерархический прогрессивный модуль слияния; DMS – динамическая многомасштабная свертка; RDAB – остаточный блок двойного внимания; SSM – модуль селективного пространства состояний; $\mathcal{L}_{cls}$, $\mathcal{L}_{rec}$, $\mathcal{L}_{con}$ – потери классификации, реконструкции и контрастивные потери соответственно.

Ветвь обработки спектральных карт служит основной сети, извлекая многоуровневые глубокие признаки из трехканальных временно-частотных карт через четырехступенчатый каскадированный

модуль MSAFN. Каждый этап включает три подмодуля: динамическую многомасштабную свертку, остаточное двойное внимание и селективное пространство состояний, последовательно расширяя количество каналов от 3 до 64, 128, 256 и 512, при этом постепенно уменьшая пространственное разрешение с помощью операции максимального пулинга. Операция глобального среднего пулинга в конечном итоге генерирует компактное глобальное представление. Ветвь обработки векторных признаков использует двухслойную полностью связную сеть для отображения 109-мерных статистических признаков в высокомерное семантическое пространство. Иерархический прогрессивный модуль слияния извлекает проекционные спектральные признаки с первых трех уровней, динамически взаимодействуя с векторными признаками через механизмы кросс-модального управления. Результаты слияния с каждого уровня взвешенно конкатенируются и интегрируются с глобальными спектральными признаками для генерации окончательного слияния. Многоуровневая задача вывода включает две вспомогательные задачи – реконструкцию признаков и контрастивное обучение – помимо основной задачи классификации. Эти задачи совместно оптимизируются с использованием потерь кросс-энтропии, потерь по среднеквадратичной ошибке и потерь для контролируемого контрастивного обучения. Вся сеть проходит обучение от начала до конца с использованием оптимизатора Adam с косинусным планированием ослабления.

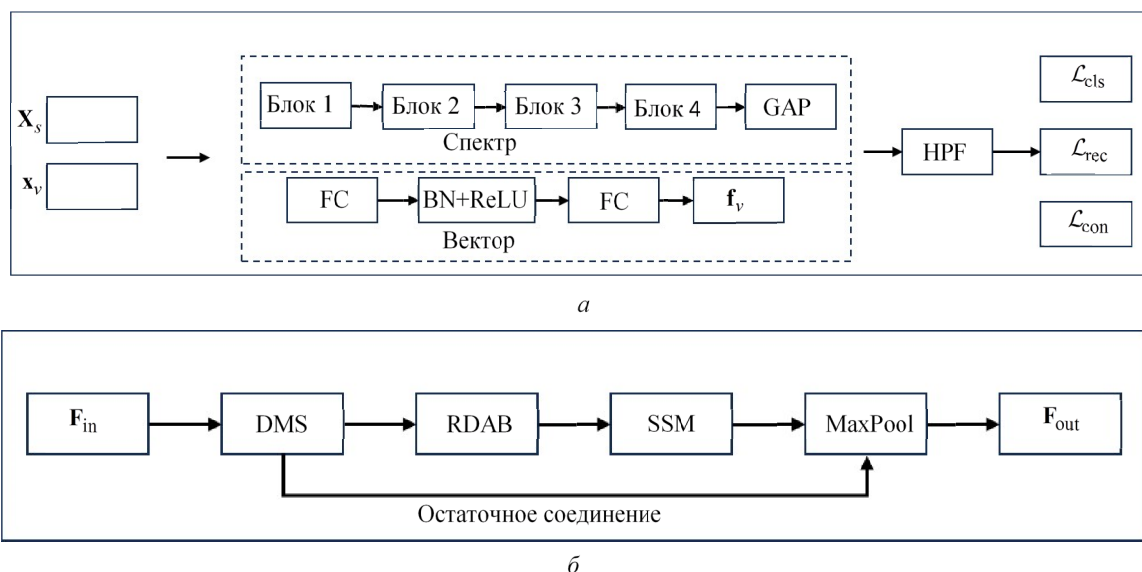


Рис. 1. Структура модели MSAFN: а – общая архитектура сети; б – структура блока MSAFN
Fig. 1. MSAFN model structure: а – overall network architecture; б – MSAFN block structure

Двухветвевой энкодер. В задачах диагностики неисправностей качающихся подшипников, представления во временно-частотной области и статистические признаки характеризуют условия работы оборудования с различных точек зрения: первые сохраняют локальную временно-частотную структуру и текстурные паттерны сигнала в двумерном формате изображения, что подходит для извлечения иерархических пространственных признаков через сверточные нейронные сети; вторые конденсируют глобальные статистические свойства сигнала в одномерный вектор, предлагая вычислительную эффективность и сильную физическую интерпретируемость. Однако однотипные представления признаков не могут всесторонне характеризовать признаки неисправностей при сложных эксплуатационных условиях. Для решения этой проблемы MSAFN разрабатывает архитектуру двухветвевой энкодера. Независимые пути извлечения признаков обрабатывают два типа гетерогенных данных отдельно, сохраняя их представительные преимущества и закладывая основу для последующего глубокого слияния.

Ветвь кодирования спектрограммы принимает трехканальную временно-частотную карту $\mathbf{X}_s \in \mathbb{R}^{3 \times H \times W}$ на вход, где H и W обозначают высоту и ширину временно-частотной карты соответственно, а три канала соответствуют спектру по преобразованию короткосрочного Фурье (STFT), временно-частотной карте по непрерывному вейвлет-преобразованию (CWT) и спектру огибающего. Преимущества этой многовидовой входной конструкции следующие: спектр STFT обеспечивает равномерное временно-частотное разрешение, подходящее для захвата частотных компонентов в устойчивом состоянии; временно-частотная карта CWT обладает многоуровневыми свойствами, позволяя адаптивно подстраиваться под кратковременные воздействия на разных масштабах; спектр огибающего выделяет модуляционные признаки и чувствителен к частотам характеристик подшипников. Синергетический ввод этих трех источников позволяет сети воспринимать информацию о неисправности с различных дополнительных точек зрения. Эта ветвь использует четырехступенчатую каскадную структуру, где процесс извлечения признаков на i -м этапе может быть выражен формулой

$$F_i, F_i^{\text{pre}} = \text{MSAFNBlock}_i(F_{i-1}),$$

где $i = 1, 2, 3, 4$, при этом $F_0 = \mathbf{X}_s$ – представляет входную спектрограмму; $F_i \in \mathbb{R}^{C_i \times H_i \times W_i}$ обозна-

чает подвыборку выходных признаков на уровне i ; F_i^{pre} – промежуточные признаки до подвыборки, используемые для последующего слияния слоев. Количество каналов на каждом уровне устанавливается последовательно как $C_1 = 64$, $C_2 = 128$, $C_3 = 256$, $C_4 = 512$ с прогрессивным удвоением, чтобы постепенно расширить семантическую емкость признаков. Первые три слоя используют максимальный пулинг с шагом 2 для уменьшения пространственного разрешения в два раза, что снижает вычислительные затраты, одновременно расширяя приемное поле. Эта иерархическая конструкция следует принципу абстракции признаков от мелких к более глубоким: поверхностные слои сохраняют низкоуровневые визуальные признаки, такие как края и текстуры, в то время как глубокие слои кодируют высокоуровневую семантическую информацию, связанную с типами неисправностей. Наконец, глобальные спектральные признаки извлекаются с помощью глобального среднего пулинга согласно формуле

$$\mathbf{f}_s^{\text{gl}} = \text{GAP}(F_4) \in \mathbb{R}^{512}.$$

Глобальный средний пулинг сжимает пространственное измерение в один вектор, тем самым избегая проблемы взрыва параметров, вызванной полностью связанными слоями, и улучшая инвариантность признаков к пространственным сдвигам. Более того, чтобы полностью использовать многомасштабную информацию из промежуточных слоев, признаки с первых трех слоев отображаются в единую размерность слияния d_f через независимый проекционный слой

$$\mathbf{f}_s^{(i)} = \mathbf{W}_i^{\text{proj}} \cdot \text{GAP}(\mathbf{F}_i^{\text{pre}}) + \mathbf{b}_i^{\text{proj}} \in \mathbb{R}^{d_f}, \quad i = 1, 2, 3,$$

где $\mathbf{W}_i^{\text{proj}} \in \mathbb{R}^{d_f \times C_i}$ и $\mathbf{b}_i^{\text{proj}} \in \mathbb{R}^{d_f}$ представляют собой матрицу весов и вектор смещения i -го проекционного слоя соответственно. Эта конструкция позволяет признакам с разных уровней абстракции взаимодействовать с векторными признаками в едином пространстве, эффективно устраняя потерю тонкой информации, присущую традиционным методам, которые полагаются исключительно на признаки последнего слоя.

Ветвь кодирования векторных признаков принимает статистический вектор признаков $\mathbf{x}_v \in \mathbb{R}^{d_v}$ размерностью d_v на вход, охватывающий традиционные ручные признаки – такие, как статистика во временной области, признаки в частотной

области и нелинейные динамические индикаторы. Несмотря на то, что методы глубокого обучения превосходят в автоматическом извлечении признаков, эти классические признаки были проверены в долгосрочной инженерной практике и демонстрируют четкие физические механизмы отклика для конкретных типов отказов. Их интеграция предоставляет сети основу предварительных знаний. Эта ветвь отображает низкоразмерные входы в высокоразмерное семантическое пространство через двухслойную полностью связанную сеть в соответствии с формулами

$$\mathbf{h}_1 = \text{Dropout}(\text{ReLU}(\text{BN}(\mathbf{W}_1 \mathbf{x}_v + \mathbf{b}_1)));$$

$$\mathbf{f}_v = \mathbf{W}_2 \mathbf{h}_1 + \mathbf{b}_2 \in \mathbb{R}^{d_f},$$

где $\mathbf{W}_1 \in \mathbb{R}^{d_h \times d_v}$ и $\mathbf{W}_2 \in \mathbb{R}^{d_f \times d_h}$ – матрицы весов; d_h – размерность скрытого слоя. Нормализация по батчам устраняет различия в размерностях между признаками и ускоряет сходимость. Функция активации ReLU вводит нелинейную выразительную способность, в то время как регуляризация Dropout случайным образом маскирует нейроны с вероятностью $p = 0.3$ для предотвращения переобучения. В отличие от наивного подхода, заключающегося в прямой конкатенации сырых признаков, этот процесс кодирования встраивает статистические признаки в пространство представлений, согласованное с семантикой глубоких признаков через нелинейные преобразования. Это позволяет последующим модулям слияния выполнять кросс-модальные взаимодействия на сопоставимом уровне абстракции, тем самым улучшая эффективность слияния.

Модуль MSAFN. Служит основным строительным блоком для кодирования спектральных графов, отвечая за извлечение высокодискриминирующих признаков неисправностей из временно-частотных графов. Учитывая многомасштабные характеристики и сложные временно-частотные структуры сигналов неисправностей качающихся подшипников, этот модуль включает три подмодуля: динамическую многомасштабную свертку, остаточное двойное внимание и селективное пространство состояний. Эти подмодули усиливают возможности представления признаков в трех измерениях: адаптивный выбор поля восприятия, улучшение значимости признаков и моделирование дальнедействующих зависимостей.

Качающиеся подшипники демонстрируют заметно различающиеся временно-частотные паттерны в зависимости от типа неисправности, уровня повреждения и условий эксплуатации: локализованные дефекты типа ям образуют периодические удары, проявляясь как узкополосные кратковременные отклики в временно-частотных графиках; износные неисправности вызывают сдвиги в распределении энергии с широким диапазоном, в то время как комплексные неисправности часто содержат многомасштабные компоненты признаков одновременно. Традиционные одноуровневые сверточные ядра не могут эффективно захватывать эти разнообразные паттерны. Для решения этой проблемы подмодуль динамической многомасштабной свертки одновременно использует сверточные ядра трех масштабов и применяет механизм управления для адаптивного слияния. Конкретно, входной признак $\mathbf{F}_{in} \in \mathbb{R}^{C_{in} \times H \times W}$, где C_{in} обозначает количество входных каналов, обрабатывается через сверточные ядра 3×3 , 5×5 и 7×7 , что описывается формулой

$$\mathbf{F}^{(k)} = \text{Conv}_{k \times k}(\mathbf{F}_{in}), \quad k \in \{3, 5, 7\},$$

где $\mathbf{F}^{(k)} \in \mathbb{R}^{C_{out} \times H \times W}$ – выход свертки на соответствующем масштабе. Приемные поля на трех масштабах настроены для захвата тонких текстур, средних структур и крупномасштабных глобальных паттернов соответственно. Для динамической настройки вклада каждого масштаба был введен механизм масштабного управления на основе внимания канала, определяемый формулой

$$\alpha = \text{Softmax}(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \text{GAP}(\mathbf{F}_{in}))) \in \mathbb{R}^3,$$

где $\mathbf{W}_1 \in \mathbb{R}^{(C_{in}/r) \times C_{in}}$ и $\mathbf{W}_2 \in \mathbb{R}^{3 \times (C_{in}/r)}$ – матрицы проекции для уменьшения и увеличения размерности соответственно; r – коэффициент сжатия. Окончательный вывод многомасштабного слияния вычисляется по формуле

$$\mathbf{F}_{ms} = \alpha_1 \mathbf{F}^{(3)} + \alpha_2 \mathbf{F}^{(5)} + \alpha_3 \mathbf{F}^{(7)}.$$

Эта конструкция позволяет сети адаптивно выбирать оптимальную комбинацию приемных полей в зависимости от характеристик входных образов, предлагая большую гибкость представления по сравнению с архитектурами с фиксированными весами в многоветвевых структурах. Слитые признаки проходят нормализацию по батчам, а затем активируются после добавления к входу через остаточные соединения согласно формуле

$$\mathbf{F}_{dms} = \text{ReLU}(\text{BN}(\mathbf{F}_{ms}) + \mathbf{F}_{res}),$$

где $\mathbf{F}_{\text{res}}$ представляет собой тождественное отображение входных признаков или проекцию с использованием свертки 1×1 (когда количество каналов на входе и выходе различается). Остаточные соединения обеспечивают плавную обратную передачу градиентов, эффективно устраняя проблемы деградации в глубоких сетях.

Признаки, обработанные через динамические многомасштабные свертки, содержат богатую многомасштабную информацию, однако не все каналы признаков и пространственные расположения имеют одинаковую важность для идентификации неисправностей. Подмодуль остаточного двойного внимания адаптивно усиливает дискриминирующие признаки и подавляет избыточные отклики, сочетая три механизма: внимание канала, внимание пространства и внимание к координатам. Внимание канала моделирует взаимозависимости между каналами карты признаков, обучая веса каналов после сжатия пространственной информации через глобальный пулинг, как представлено в формуле

$$s_c = \sigma(\mathbf{W}_c^2 \cdot \text{ReLU}(\mathbf{W}_c^1 \cdot (\text{GAP}(\mathbf{F}) + \text{GMP}(\mathbf{F})))),$$

где $\text{GAP}(\cdot)$ и $\text{GMP}(\cdot)$ – глобальные средний и максимальный пулинги соответственно; $\sigma(\cdot)$ – функция активации Sigmoid; индекс c обозначает размерность канала. Дизайн с двойным пулингом одновременно захватывает среднюю интенсивность отклика и пик важности по каналам, улучшая устойчивость оценки внимания. Пространственное внимание фокусируется на распределении пространственной значимости карт признаков, выявляя критические области, связанные с неисправностями. Этот процесс описывается формулой

$$s_s = \sigma(\text{Conv}_{7 \times 7}([\text{AvgPool}_c(\mathbf{F}); \text{MaxPool}_c(\mathbf{F})])),$$

где $[\cdot; \cdot]$ обозначает конкатенацию вдоль размерности канала; AvgPool_c и MaxPool_c представляют операции среднего и максимального пулинга вдоль размерности канала соответственно. Сверточное ядро размером 7×7 расширяет диапазон восприятия пространственного контекста, помогая захватывать глобальное пространственное распределение режимов отказа. Более того, механизм координатного внимания явно кодирует позиционную информацию в веса внимания, компенсируя ограничения традиционных механизмов внимания канала и пространства в точном восприятии местоположений, что выражается как

$$\mathbf{z}_h = \frac{1}{W} \sum_{j=1}^W \mathbf{F}(:, :, j), \quad \mathbf{z}_w = \frac{1}{H} \sum_{i=1}^H \mathbf{F}(:, i, :),$$

$$\mathbf{g} = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}([\mathbf{z}_h; \mathbf{z}_w])),$$

$$\mathbf{a}_h = \sigma(\text{Conv}_{1 \times 1}(\mathbf{g}_h)), \quad \mathbf{a}_w = \sigma(\text{Conv}_{1 \times 1}(\mathbf{g}_w)),$$

где $\mathbf{z}_h \in \mathbb{R}^{C \times H \times 1}$ и $\mathbf{z}_w \in \mathbb{R}^{C \times 1 \times W}$ агрегируют пространственную информацию вдоль горизонтального и вертикального направлений соответственно; индексы h и w обозначают пространственные размерности высоты и ширины; $\mathbf{g}$ – результат совместного кодирования конкатенированных признаков через сверточный слой 1×1 с нормализацией по батчам и функцией активации ReLU; $\mathbf{a}_h$ и $\mathbf{a}_w$ – веса внимания для горизонтального и вертикального направлений соответственно, полученные разделением закодированных признаков и применением функции активации Sigmoid. После конкатенации и кодирования они разделяются для генерации весов внимания для обоих направлений. В конечном итоге три механизма внимания последовательно применяются к карте признаков, сохраняя исходную информацию через остаточные соединения согласно формуле

$$\mathbf{F}_{\text{att}} = \mathbf{F}_{\text{dms}} + (\mathbf{F}_{\text{dms}} \odot \mathbf{s}_c \odot \mathbf{s}_s \odot \mathbf{a}_h \odot \mathbf{a}_w),$$

где $\odot$ – поэлементное умножение. Эта каскадная конструкция многократного внимания всесторонне улучшает дискриминационные способности признаков по трем измерениям: важность канала, пространственная выразительность и чувствительность к позициям.

Поскольку сигналы вибрации качающихся подшипников проявляют несомненную периодичность и дальнедействующие временные зависимости, циклическое повторение воздействий неисправностей и их модуляционная связь с частотой вращения содержат важную диагностическую информацию. Традиционные сверточные нейронные сети, ограниченные локальными приемными полями, испытывают трудности с эффективным моделированием таких глобальных зависимостей. Подмодуль селективного состояния заимствует концепции из моделей состояния для захвата дальнедействующих зависимостей с линейной вычислительной сложностью. Сначала двумерная карта признаков была преобразована в последовательное представление:

$$\mathbf{X}_{\text{seq}} = \text{Reshape}(\mathbf{F}_{\text{att}}) \in \mathbb{R}^{H \times W \times C}.$$

Затем была выполнена нормализация по слоям и через линейную проекцию были созданы две ветви, как показано в формуле

$$\mathbf{X}_{\text{inner}}, \mathbf{Z} = \text{Split}(\mathbf{W}_{\text{in}} \cdot \text{LN}(\mathbf{X}_{\text{seq}})),$$

где $\mathbf{X}_{\text{inner}}, \mathbf{Z} \in \mathbb{R}^{L \times D}$, при этом D – расширенная размерность скрытого слоя. Локальная ветвь захватывает краткосрочные локальные паттерны с помощью глубоко разделяемых одномерных сверток согласно формуле

$$\mathbf{Y}_{\text{loc}} = \text{SiLU}(\text{BN}(\text{DWConv}_{1D}(\mathbf{X}_{\text{inner}}))).$$

Глобальная ветвь моделирует дальнедействующий контекст с помощью глобального пулинга $\mathbf{w}_g$ и механизмов управления $\mathbf{Y}_{\text{gl}}$, определяемых формулами

$$\begin{aligned} \mathbf{w}_g &= \sigma(\mathbf{W}_g^2 \cdot \text{SiLU}(\mathbf{W}_g^1 \cdot \text{GAP}(\mathbf{X}_{\text{inner}}))), \\ \mathbf{Y}_{\text{gl}} &= \mathbf{X}_{\text{inner}} \odot \mathbf{w}_g. \end{aligned}$$

Обе ветви выбираются для слияния через обучаемое управление в соответствии с формулами

$$\begin{aligned} \mathbf{G} &= \sigma(\mathbf{W}_s \cdot \mathbf{X}_{\text{inner}}), \\ \mathbf{Y} &= \mathbf{G} \odot \mathbf{Y}_{\text{loc}} + (1 - \mathbf{G}) \odot \mathbf{Y}_{\text{gl}}, \end{aligned}$$

где $\mathbf{W}_s$ – матрица проекции для управления слиянием ветвей.

Этот выборочный механизм позволяет сети динамически балансировать вклады локальных деталей и глобального контекста в зависимости от входного контента. Окончательный выход восстанавливается в двумерную карту признаков через управляемую модуляцию и линейную проекцию:

$$\mathbf{F}_{\text{out}} = \mathbf{F}_{\text{att}} + \gamma \cdot \text{Reshape}(\mathbf{W}_{\text{out}}(\mathbf{Y} \odot \text{SiLU}(\mathbf{Z}))),$$

где γ – это обучаемый параметр масштабирования, инициализируемый малым значением для стабилизации потока градиентов на ранних этапах обучения. По сравнению с вычислительной сложностью $O(L^2)$ механизмов самовнимания, этот модуль требует только $O(L)$ вычислительных затрат. Это значительно снижает вычислительную нагрузку при сохранении способностей к моделированию дальнедействующих зависимостей, что делает его подходящим для эффективной обработки высокоразреженных спatioвременных графов.

Иерархический прогрессивный модуль слияния признаков. Слияние признаков – важный этап в мультимодальной диагностике неисправностей, направленный на эффективную интеграцию дополнительной информации из различных источников данных для улучшения диагностической производительности. Традиционные стратегии слияния обычно используют простые операции – конкатенацию признаков или поэле-

ментное сложение, интегрируя признаки только на последнем этапе сети. Однако этот одноуровневый подход слияния имеет два ограничения: во-первых, хотя признаки глубоких слоев содержат богатую семантическую информацию, тонкие локальные паттерны постепенно теряются в процессе абстракции на каждом слое, что затрудняет сохранение целостности многомасштабных признаков в одном слое; во-вторых, простая конкатенация или суммирование предполагают равную важность различных типов признаков, игнорируя различные вклады разных модальностей в зависимости от конкретных типов неисправностей. Для преодоления этих ограничений в данной статье предлагается иерархический прогрессивный модуль слияния, который достигает глубокой интеграции гетерогенных признаков через многократное взаимодействие между уровнями абстракции и динамические механизмы управления.

Основная идея иерархического прогрессивного модуля слияния заключается в одновременном введении взаимодействий с векторными признаками на нескольких уровнях абстракции в ветви кодирования спектрограммы, полностью используя многомасштабную информацию от поверхностных до глубоких слоев. Как описано в разделе «Двухветвевой энкодер», верхние три выхода ветви спектрограммы проецируются для получения трех наборов иерархических признаков $\{\mathbf{f}_s^{(1)}, \mathbf{f}_s^{(2)}, \mathbf{f}_s^{(3)}\}$. Признаки поверхностного уровня сохраняют низкоуровневую визуальную информацию – такую, как края и текстуры, в то время как признаки глубокого уровня кодируют высокоуровневую семантику, связанную с типами неисправностей. Этот многослойный дизайн позволяет процессу слияния одновременно обращаться к представлениям на разных уровнях абстракции, компенсируя потерю информации, которая происходит при использовании только признаков последнего слоя.

Для каждого уровня в данной статье разрабатывается механизм кросс-модального управления, чтобы обеспечить динамическое взаимодействие между спектральными и векторными признаками. Мотивация для этого механизма исходит из наблюдения, что относительная важность этих двух типов признаков значительно различается в зависимости от типа неисправности и условий эксплуатации. Например, для дефектов ударного типа, демонстрирующих ярко выраженные визуальные паттерны на временно-частотных графиках, спектральные признаки могут доминировать.

Напротив, для ранних, тонких неисправностей, которые легче различить с помощью статистических метрик, векторные признаки становятся более важными. Таким образом, процесс слияния должен адаптивно регулировать веса вклада обоих типов признаков в зависимости от характеристик входного образца. Процесс кросс-модального слияния на i -м уровне формализуется формулами:

$$\begin{aligned} \mathbf{h}_s^{(i)} &= \mathbf{W}_s^{(i)} \mathbf{f}_s^{(i)} + \mathbf{b}_s^{(i)}, \\ \mathbf{h}_v^{(i)} &= \mathbf{W}_v^{(i)} \mathbf{f}_v^{(i)} + \mathbf{b}_v^{(i)}, \end{aligned}$$

где $\mathbf{W}_s^{(i)}, \mathbf{W}_v^{(i)} \in \mathbb{R}^{d_f \times d_f}$ – матрицы проекции, которые отображают два типа признаков в единое пространство взаимодействия. Веса управления генерируются конкатенацией признаков и применением нелинейного преобразования согласно формуле

$$\mathbf{g}^{(i)} = \sigma\left(\mathbf{W}_g^{(i)} \left[\mathbf{h}_s^{(i)}; \mathbf{h}_v^{(i)}\right] + \mathbf{b}_g^{(i)}\right),$$

где $\mathbf{W}_g^{(i)} \in \mathbb{R}^{d_f \times 2d_f}$ – матрица проекции управления, которая ограничивает вывод интервалом $[0, 1]$. Этот управляемый вектор контролирует относительный вклад двух типов признаков поэлементно, что выражается формулой

$$\mathbf{f}_{\text{fuse}}^{(i)} = \mathbf{W}_o^{(i)} \left(\mathbf{g}^{(i)} \odot \mathbf{h}_s^{(i)} + (1 - \mathbf{g}^{(i)}) \odot \mathbf{h}_v^{(i)} \right) + \mathbf{b}_o^{(i)},$$

где $\mathbf{W}_o^{(i)} \in \mathbb{R}^{d_f \times d_f}$ – матрица проекции для вывода. По сравнению с фиксированным слиянием весов этот механизм управления демонстрирует более высокую адаптивность, позволяя тонко настраивать вклад признаков по измерениям в зависимости от входного контента.

После получения признаков слияния с каждого уровня их необходимо интегрировать в единую представление слияния. Учитывая, что признаки с разных уровней могут по-разному способствовать финальной диагностике, в данной статье предлагаются обучаемые иерархические веса для адаптивной агрегации по формуле

$$\beta = \text{Softmax}([\beta_1, \beta_2, \beta_3]),$$

где $\beta_1, \beta_2, \beta_3$ – обучаемые параметры, которые после нормализации через Softmax служат агрегированными весами для каждого слоя. Этот параметризованный дизайн позволяет сети автоматически обучаться оптимальным коэффициентам вклада для каждого слоя в процессе обучения, исключая необходимость в ручной настройке

параметров. Затем признаки с каждого слоя конкатенируются и проходят глубокую интеграцию через два полностью связанных слоя, как показано в формулах

$$\begin{aligned} \mathbf{f}_{\text{conc}} &= [\mathbf{f}_{\text{fuse}}^{(1)}; \mathbf{f}_{\text{fuse}}^{(2)}; \mathbf{f}_{\text{fuse}}^{(3)}] \in \mathbb{R}^{3d_f}, \\ \mathbf{f}_{\text{hier}} &= \mathbf{W}_{f2} (\text{Dropout}(\text{ReLU}(\mathbf{W}_{f1} \mathbf{f}_{\text{conc}} + \\ &\quad + \mathbf{b}_{f1}))) + \mathbf{b}_{f2}, \end{aligned}$$

где $\mathbf{W}_{f1} \in \mathbb{R}^{d_f \times 3d_f}$ и $\mathbf{W}_{f2} \in \mathbb{R}^{d_f \times d_f}$ представляют собой матрицы уменьшения размерности и проекции вывода соответственно; $\mathbf{f}_{\text{conc}}$ – взвешенная конкатенация слитых признаков со всех трех уровней. Регуляризация Dropout случайным образом маскирует нейроны с вероятностью $p = 0.3$, чтобы предотвратить избыточную зависимость сети от признаков, исходящих от определенных слоев.

Наконец, слитые иерархические признаки комбинируются с глобальными спектральными признаками для генерации полного слитого представления для последующих задач согласно

$$\mathbf{f}_{\text{fin}} = \text{Dropout}(\text{ReLU}(\mathbf{W}_{\text{int}} [\mathbf{f}_{\text{hier}}; \mathbf{f}_s^{\text{gl}}] + \mathbf{b}_{\text{int}})),$$

где $\mathbf{f}_s^{\text{gl}} \in \mathbb{R}^{512}$ представляет глобальные признаки пулинга с последнего слоя ветви кодирования спектрограммы; $\mathbf{W}_{\text{int}} \in \mathbb{R}^{d_f \times (d_f + 512)}$ – матрица проекции для интеграции. Этот дизайн сочетает многоуровневые кросс-модальные слияния признаков с глубокими глобальными семантическими признаками, сохраняя тонкие многомасштабные данные, одновременно обеспечивая целостность высокоуровневой семантической информации.

Многозадачный выходной модуль. Модели глубокого обучения часто сталкиваются с проблемами в задачах диагностики неисправностей, такими как ограниченное количество размеченных образцов и дрейф распределений, вызванный изменениями в эксплуатации. Сигналы управления от одной задачи классификации могут быть недостаточными для того, чтобы направить сеть на обучение представлений признаков с сильными возможностями обобщения. Многозадачное обучение эффективно снижает риски переобучения и улучшает устойчивость признаков, одновременно оптимизируя несколько связанных задач и используя общую информацию между задачами для предоставления дополнительных индуктивных смещений. Исходя из этого принципа, в данной статье разрабатывается многозадачный

выходной модуль, который вводит два вспомогательных задания – реконструкцию признаков и контрастивное обучение – наряду с основной задачей классификации. Совместная оптимизация этих задач улучшает диагностическую производительность модели. Все три головки задач используют одни и те же слитые признаки $\mathbf{f}_{\text{fin}}$ в качестве входных данных, совместно оптимизируя всю сеть через обратное распространение.

Головка классификации служит основным модулем задачи, ответственным за отображение слитых признаков в пространство категорий неисправностей. Этот модуль использует двухслойную полностью связанную сеть в соответствии с формулами

$$\mathbf{h}_{\text{cls}} = \text{Dropout}\left(\text{ReLU}\left(\mathbf{W}_{\text{cls}}^1 \mathbf{f}_{\text{fin}} + \mathbf{b}_{\text{cls}}^1\right)\right),$$

$$\mathbf{o}_{\text{cls}} = \mathbf{W}_{\text{cls}}^2 \mathbf{h}_{\text{cls}} + \mathbf{b}_{\text{cls}}^2 \in \mathbb{R}^{N_c},$$

где $\mathbf{W}_{\text{cls}}^1 \in \mathbb{R}^{(d_f/2) \times d_f}$ и $\mathbf{W}_{\text{cls}}^2 \in \mathbb{R}^{N_c \times (d_f/2)}$ – матрицы для уменьшения размерности и классификации соответственно, при этом N_c обозначает общее количество категорий неисправностей, а d_f – размерность слитых признаков. Вектор $\mathbf{o}_{\text{cls}}$ представляет выходной вектор головки классификации до нормализации Softmax, где каждый элемент соответствует ненормализованному значению (логиту) для соответствующей категории неисправности. Регуляризация Dropout в промежуточном слое случайным образом маскирует нейроны с вероятностью $p = 0.5$, эффективно подавляя избыточную зависимость головки классификации от определенных признаков. Во время вывода выходной вектор проходит нормализацию через Softmax для получения предсказанного распределения вероятностей по категориям, как определено формулой

$$p(y = k | \mathbf{x}) = \frac{\exp(o_{\text{cls}}^k)}{\sum_{j=1}^{N_c} \exp(o_{\text{cls}}^j)}, \quad k = 1, 2, \dots, N_c. \quad (1)$$

В то же время, головка реконструкции служит первым вспомогательным модулем задачи, ограничивая структурную целостность пространства признаков через самоконтролируемые сигналы. Она имеет следующую логику проектирования: если слитые признаки могут эффективно восстанавливать высокоразмерное представление до сжатия, то эти признаки сохраняют критическую информацию из входных данных, а не просто

учат короткие признаки, связанные с метками классификации. Данный модуль использует двухслойную полностью связанную сеть, структурированную как автокодировщик, согласно

$$\mathbf{h}_{\text{rec}} = \text{ReLU}\left(\mathbf{W}_{\text{rec}}^1 \mathbf{f}_{\text{fin}} + \mathbf{b}_{\text{rec}}^1\right),$$

$$\hat{\mathbf{f}} = \mathbf{W}_{\text{rec}}^2 \mathbf{h}_{\text{rec}} + \mathbf{b}_{\text{rec}}^2 \in \mathbb{R}^{d_r},$$

где $\mathbf{W}_{\text{rec}}^1 \in \mathbb{R}^{d_f \times d_f}$ и $\mathbf{W}_{\text{rec}}^2 \in \mathbb{R}^{d_r \times d_f}$ – матрицы весов сети реконструкции; d_r – размерность цели реконструкции; $\hat{\mathbf{f}}$ – выход головки реконструкции, используемый для восстановления сжатого представления признаков.

Кроме того, головка контрастивной проекции служит вторым вспомогательным модулем задачи, способствуя внутриклассовой когезии и межклассовому разделению через контролируемое контрастивное обучение. Основная идея контрастивного обучения заключается в том, чтобы приблизить схожие образцы и отдалить несхожие в пространстве встраивания, тем самым обучая дискриминирующие представления признаков. Этот модуль проецирует слитые признаки в низкоразмерное нормализованное пространство по формулам

$$\mathbf{h}_{\text{proj}} = \text{ReLU}\left(\mathbf{W}_{\text{proj}}^1 \mathbf{f}_{\text{fin}} + \mathbf{b}_{\text{proj}}^1\right),$$

$$\mathbf{z} = \frac{\mathbf{W}_{\text{proj}}^2 \mathbf{h}_{\text{proj}} + \mathbf{b}_{\text{proj}}^2}{\|\mathbf{W}_{\text{proj}}^2 \mathbf{h}_{\text{proj}} + \mathbf{b}_{\text{proj}}^2\|_2} \in \mathbb{R}^{d_p},$$

где $\mathbf{W}_{\text{proj}}^1 \in \mathbb{R}^{d_f \times d_f}$ и $\mathbf{W}_{\text{proj}}^2 \in \mathbb{R}^{d_p \times d_f}$ – матрицы весов проекционных сетей; d_p – размерность проекционного пространства; $\|\cdot\|_2$ – L_2 -норма;

$\mathbf{z} \in \mathbb{R}^{d_p}$ – нормализованный проекционный вектор признаков, используемый для вычисления контрастивных потерь. Проецированные векторы признаков ограничиваются на единичной гиперсфере через L_2 -нормализацию, что позволяет напрямую измерять сходство между образцами через скалярные произведения. По сравнению с исходным высокоразмерным пространством признаков, контрастивное обучение в низкоразмерном проекционном пространстве более эффективно оптимизирует геометрические отношения между образцами, одновременно снижая вычислительные затраты.

Проектирование функции потерь. Для трех ветвей задач многозадачного выходного модуля в данной статье разработаны соответствующие

функции потерь, которые совместно оптимизируются через взвешенное суммирование. Общая функция потерь определяется как

$$L_{\text{tot}} = L_{\text{cls}} + \lambda_1 L_{\text{rec}} + \lambda_2 L_{\text{con}},$$

где λ_1 и λ_2 – гиперпараметры, балансирующие вклады вспомогательных задач.

Потери классификации используют функцию кросс-энтропии с размягчением меток. Размягчение меток смягчает целевое распределение, равномерно перераспределяя вероятность из истинного класса в другие:

$$y_{\text{smth}}^k = \begin{cases} 1 - \varepsilon + \frac{\varepsilon}{N_c}, & k = y_{\text{true}}, \\ \frac{\varepsilon}{N_c}, & k \neq y_{\text{true}}, \end{cases} \quad (2)$$

где ε – коэффициент сглаживания; y_{true} – индекс истинного класса образца. Потери классификации определяются как кросс-энтропия между предсказанным распределением и сглаженным распределением меток по формуле

$$L_{\text{cls}} = - \sum_{k=1}^{N_c} y_{\text{smth}}^k \log p(y = k | \mathbf{x}),$$

где y_{smth}^k – сглаженное значение метки для класса k , определяемое (2); $p(y = k | \mathbf{x})$ – предсказанная вероятность для класса k , определяемая (1).

Техника размягчения меток накладывает имплицитную энтропийную регуляризацию на предсказания сети, побуждая модели выводить более гладкие распределения вероятностей. Это улучшает устойчивость к шумным меткам и образцам за пределами распределения.

Потери реконструкции используют функцию среднеквадратичной ошибки для измерения расхождения между восстановленными и целевыми признаками, что выражается как

$$L_{\text{rec}} = \frac{1}{d_r} \|\hat{\mathbf{f}} - \mathbf{f}_{\text{targ}}\|_2^2,$$

где $\hat{\mathbf{f}}$ представляет выход головки реконструкции; $\mathbf{f}_{\text{targ}} = \text{sg}(\mathbf{f}_{\text{fin}})$ – слитые признаки после останковки градиента. Этот компонент потерь служит самоконтролируемым сигналом, ограничивающим структурную целостность пространства признаков, побуждая сеть обучать сжатое представление, которое полностью сохраняет входную информацию.

Контрастивные потери используют задачу контролируемого контрастивного обучения. Пусть в батче содержится B образцов, чьи нормализованные проекционные признаки обозначаются как $\{\mathbf{z}_i\}_{i=1}^B$, с соответствующими метками $\{y_i\}_{i=1}^B$. Для анкерного образца i положительное множество образцов определяется, как другие образцы в батче с той же меткой: $P_i = \{j : y_j = y_i, j \neq i\}$. Контрастивные потери с контролируемым обучением определяются как

$$L_{\text{con}} = -\frac{1}{B} \sum_{i=1}^B \frac{1}{|P_i|} \sum_{j \in P_i} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_j / \tau)}{\sum_{k \neq i} \exp(\mathbf{z}_i \cdot \mathbf{z}_k / \tau)},$$

где τ – это гиперпараметр температуры. Эта функция потерь явно оптимизирует геометрическую структуру пространства встраивания, приближая схожие образцы и отдаляя несхожие, тем самым способствуя внутриклассовой кластеризации и межклассовому разделению.

Вкратце, потери классификации направляют оптимизацию сети на обеспечение диагностической точности, потери реконструкции усиливают целостность признаков через ограничения на сохранение информации, а контрастивные потери улучшают разделимость классов через структурную оптимизацию. Эти три компонента синергетически ограничивают процесс обучения признаков по трем измерениям – дискриминируемости, целостности и структуры – предоставляя надежную цель для оптимизации диагностики неисправностей качающихся подшипников.

Экспериментальная настройка. В данной статье используются два общедоступных набора данных о неисправностях подшипников для проверки эффективности предложенного метода. Платформа сбора данных представлена на рис. 2. Набор данных CWRU (Case Western Reserve University) [21], предоставленный Центром данных о подшипниках CWRU, служит стандартным эталонным набором данных для диагностики неисправностей подшипников. Этот набор данных содержит вибрационные сигналы от подшипников на приводной стороне мотора при различных типах неисправностей, местах и уровнях повреждений, с частотой дискретизации 12 кГц. Были выбраны десять рабочих условий для формирования задачи классификации на 10 классов, включая одно нормальное состояние и девять состояний неисправности (по три разных диаметра

повреждений для неисправностей внутреннего кольца, наружного кольца и катящегося элемента). Набор данных PU (Paderborn University) [22], собранный в Лаборатории интеллектуальных систем обслуживания Университета Падерборна, включает более сложные условия эксплуатации и данные о процессе деградации с частотой дискретизации 64 кГц. В данной статье были выбраны пять типичных режимов отказа для формирова-

ния задачи классификации на пять классов, охватывающих здоровое состояние, неисправность внутреннего кольца, неисправность наружного кольца и смешанные неисправности.

Исходный вибрационный сигнал делится на сегменты по 512 выборок с использованием метода скользящего окна с 50 %-м перекрытием окна для увеличения разнообразия данных. Шаги предварительной обработки данных показаны на

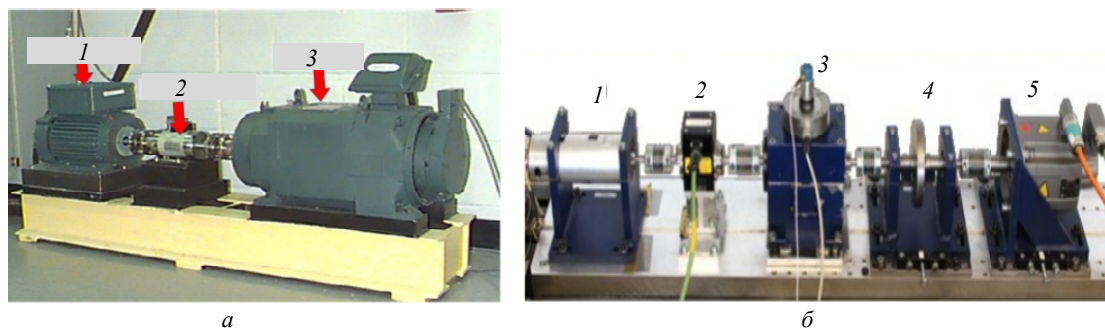


Рис. 2. Платформа сбора данных: набор данных CWRU (а): 1 – электродвигатель; 2 – подшипник приводного конца; 3 – динамометр; набор данных PU (б): 1 – электродвигатель; 2 – валы для измерения крутящего момента; 3 – модуль испытания подшипников качения; 4 – маховик; 5 – нагрузочный двигатель
Fig. 2. Data collection platform: CWRU dataset (a): 1 – electric motor; 2 – drive-end bearing; 3 – dynamometer; PU dataset (b): 1 – electric motor; 2 – torque measurement shafts; 3 – rolling bearing test module; 4 – flywheel; 5 – load motor

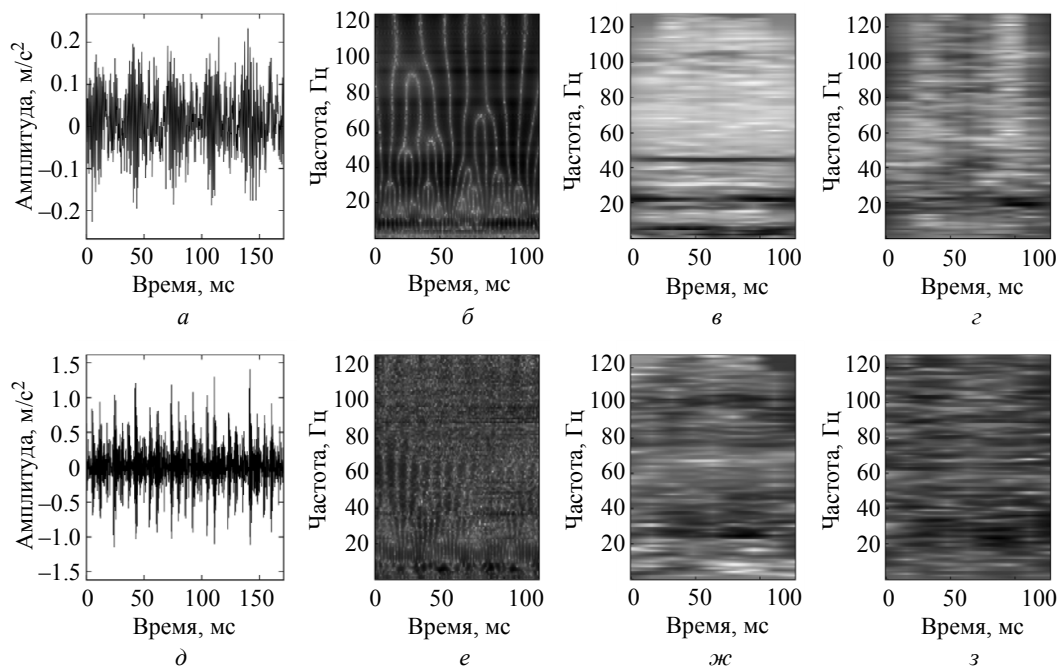


Рис. 3. Процесс предварительной обработки данных: а – исходный вибрационный сигнал в нормальном состоянии; б – спектрограмма непрерывного вейвлет-преобразования (CWT) в нормальном состоянии; в – спектрограмма короткосрочного преобразования Фурье (STFT) в нормальном состоянии; г – спектр огибающей в нормальном состоянии; д – исходный вибрационный сигнал при неисправности внутреннего кольца; е – спектрограмма CWT при неисправности внутреннего кольца; ж – спектрограмма STFT при неисправности внутреннего кольца; з – спектр огибающей при неисправности внутреннего кольца

Fig. 3. Data preprocessing process: а – raw vibration signal under normal condition; б – CWT spectrogram under normal condition; в – STFT spectrogram under normal condition; г – envelope spectrum under normal condition; д – raw vibration signal under inner race fault; е – CWT spectrogram under inner race fault; ж – STFT spectrogram under inner race fault; з – envelope spectrum under inner race fault

рис. 3. Каждый сегмент генерирует трехканальные временно-частотные графики в качестве входов для спектральной ветви: канал 1: спектр короткосрочного преобразования Фурье (STFT) с использованием окна Ханна длиной 256, перекрытием длиной 192 и 512 точками FFT; канал 2: временно-частотный график непрерывного вейвлет-преобразования (CWT) на основе вейвлета Морле с масштабом 128; третий канал – спектр огибающего, полученный выполнением STFT на огибающей сигнала, извлеченной через преобразование Хильберта. Все три временно-частотных канала подвергаются логарифмическому сжатию для улучшения динамического диапазона и масштабируются до разрешения 128×128 . Одновременно были извлечены 109-мерные ручные векторные признаки, включая: 20-мерные статистические признаки временной области (среднее, стандартное отклонение, амплитуда пиков, среднеквадратичное значение, фактор формы, фактор пика, импульсный фактор, фактор края, асимметрия, эксцесс, коэффициент пересечений с нулем, энтропия Шеннона и др.); 44-мерные признаки частотной области (спектральная статистика, центральная частота, среднеквадратичная частота, спектральная энтропия, спектральная плоскостность, точка спектрального отката и подполосные признаки в 6 частотных полосах); 35 признаков вейвлет-пакетов (32 нормализованные энергии подполос и статистика энергии на основе 5-уровневой вейвлет-декомпозиции db4); 10 признаков огибающего (статистика огибающего, статистика мгновенной частоты и спектральные признаки огибающего). Данные были разделены на тренировочную, валидационную и тестовую выборки в соотношении 7 : 1.5 : 1.5 и стандартизированы с

использованием Z-оценок. Для обеспечения баланса классов было выбрано 400 образцов на класс для формирования итогового набора данных.

В данной статье комплексно оценивается производительность модели с использованием четырех широко применяемых метрик: точности, полноты, отзывчивости и $F1$ -меры. Кроме того, для всесторонней оценки преимуществ предложенного метода были выбраны пять представительных моделей глубокого обучения в качестве сравнительных алгоритмов: MLP Mixer [23] использует архитектуру чистого многослойного перцептрона для обработки пространственной и каналовой информации; EfficientNet [24] оптимизирует структуры сверточных сетей через комплексную стратегию масштабирования; ViT [25] использует механизмы самовнимания для моделирования глобальных зависимостей в изображениях; Swin Transformer [26] вводит механизм скользящего окна для эффективного извлечения многомасштабных признаков; ConvNeXt [27] модернизирует сверточные сети, принимая принципы дизайна Transformer.

Вся разработка модели и экспериментальное тестирование в данной работе проводились с использованием фреймворка глубокого обучения PyTorch и Python 3.11. Среда разработки – PyCharm 2023. Экспериментальное оборудование включало процессор Intel Xeon(R) Platinum 8352 и графический процессор NVIDIA GeForce RTX 4090.

Анализ случаев. В табл. 1 представлена детальная матрица ошибок MSAFN на наборе данных CWRU: 1 – нормальное состояние; 2 – неисправность внутреннего кольца 0.18 мм; 3 – неисправность внутреннего кольца 0.36 мм; 4 – неисправность внутреннего кольца 0.53 мм; 5 – неисправность внутреннего кольца 0.53 мм; 6 – неисправность тела качения 0.18 мм; 6 – неисправ-

Табл. 1. Матрица ошибок MSAFN для набора данных CWRU

Tab. 1. MSAFN confusion matrix for the CWRU dataset

Истинный класс	Предсказанный класс									
	1	2	3	4	5	6	7	8	9	10
1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
10	0.0	0.0	0.03	0.0	0.0	0.0	0.0	0.0	0.0	0.97

ность тела качения 0.36 мм; 7 – неисправность тела качения 0.53 мм; 8 – неисправность наружного кольца 0.18 мм; 9 – неисправность наружного кольца 0.36 мм; 10 – неисправность наружного кольца 0.53 мм. Черным выделены верные классификации, светло-серым – ошибки ≤ 0.03 , темно-серым – ошибки > 0.03 . MSAFN достигает почти идеальной классификационной производительности по всем категориям благодаря механизму многомасштабного слияния внимания, который эффективно захватывает различные признаки неисправностей. Примечательно, что существует незначительное смещение в классе 10, где 3 % образцов ошибочно классифицируются как класс 3. Это объясняется схожими вибрационными характеристиками при данных уровнях повреждений.

Табл. 2 представляет сравнение точности диагностики различных методов на наборе данных CWRU. Предложенный метод MSAFN демонстрирует наивысшую точность 99.67 %, превосходя все эталонные модели. ConvNeXt показывает относительно более слабую производительность с точностью 94.33 %, что указывает на ограничения единой сверточной архитектуры в обработке многоклассовой тонкой диагностики неисправностей. Swin Transformer достиг второго наивысшего показателя точности – 99.17 %.

Табл. 2. Сравнение точности диагностики различных моделей на наборе данных CWRU

Tab. 2. Comparison of diagnostic accuracy of different models on the CWRU dataset

Модель	Точность, %
MLPMixer	97.00
EfficientNet	98.83
ViT	97.17
Swin Transformer	99.17
ConvNeXt	94.33
MSAFN	99.67

В табл. 3 представлена детальная матрица ошибок MSAFN на наборе данных PU: 1 – нормальное состояние; 2 – неисправность наружного кольца; 3 – неисправность внутреннего кольца; 4 – смешанная неисправность; 5 – неисправность сепаратора. Черным выделены верные классификации, светло-серым – ошибки ≤ 0.03 , темно-серым – ошибки > 0.03 . MSAFN достиг идеальной классификации для класса 3, что демонстрирует синергетическое преимущество представления двухветвевое кодировщика как для временно-частотных, так и для статистических признаков. Однако класс 1, представляющий категорию

нормального состояния, продемонстрировал склонность к ошибочной классификации в категории неисправностей (точность 91 %). Это явление объясняется нечеткими границами признаков между нормальными состояниями и ранними незначительными неисправностями в наборе данных PU.

Табл. 3. Матрица ошибок MSAFN для набора данных PU

Tab. 3. MSAFN confusion matrix for the PU dataset

Истинный класс	Предсказанный класс				
	1	2	3	4	5
1	0.91	0.0	0.05	0.04	0.0
2	0.0	0.93	0.03	0.01	0.03
3	0.0	0.0	1.0	0.0	0.0
4	0.03	0.0	0.0	0.97	0.0
5	0.0	0.0	0.02	0.0	0.98

Табл. 4 представляет сравнение точности различных методов на наборе данных PU. Этот набор данных характеризуется более высокой частотой дискретизации и более сложными условиями эксплуатации, что представляет собой более серьезную задачу для обобщающих способностей модели. MSAFN достигает точности 96.0 %. EfficientNet показывает наибольшее снижение производительности на этом наборе данных, достигая только 88.67 % точности, что указывает на то, что архитектуры сетей, использующие однотипный вход, испытывают трудности с полным извлечением дискриминирующих признаков в сложных эксплуатационных условиях.

Табл. 4. Сравнение точности диагностики различных моделей на наборе данных PU

Tab. 4. Comparison of diagnostic accuracy of different models on the PU dataset

Модель	Точность, %
MLPMixer	97.0
EfficientNet	88.67
ViT	97.0
Swin Transformer	95.67
ConvNeXt	96.0
MSAFN	96.0

Суммируя результаты с обоих наборов данных, MSAFN сохраняет стабильную классификационную производительность на сложных образцах с нечеткими границами классов, интегрируя дополнительную информацию из трехканальных спatio-временных представлений и статистических векторных признаков. Это подтверждает положительное влияние стратегий мультимодального слияния на повышение устойчивости модели.

Анализ интерпретируемости. Для анализа механизма принятия решений моделью в данной статье применяется метод Grad-CAM [28], результаты визуализации распределения внимания представлены на рис. 4. На тепловой карте светлые области соответствуют высоким весам внимания, темные – низким. Результаты показывают, что MSAFN формирует различные паттерны внимания в зависимости от типа неисправности. Для нормального состояния внимание распределено равномерно без выраженной фокусировки. Для неисправностей наружного и внутреннего колец внимание сосредоточено на характерных частотах дефектов и их гармониках. Для составных неисправностей наблюдается активация в нескольких частотных диапазонах, а для неисправности сепаратора – концентрация в среднем и низком частотном диапазоне. Полученные результаты подтверждают физическую интерпретируемость модели MSAFN и эффективность извлеченных признаков.

При углубленном анализе процесса обучения MSAFN был использован метод t -SNE для визуализации представлений признаков по слоям сети (рис. 5). Рис. 5, *а* и *б* иллюстрируют эволюцию распределения признаков по четырем уровням сети для наборов данных CWRU и PU соответственно. На первом уровне образцы различных категорий демонстрируют значительное перекрытие, а четкие межклассовые границы еще не сформированы, что свидетельствует о преимущественном кодировании низкоуровневой информации без выраженной семантической дискриминации. По мере увеличения глубины сети распределение признаков постепенно проявляет тенденцию к внутриклассовой агрегации и межклассовому разделению. На втором уровне отдельные категории начинают формировать различимые центры кластеризации при сохранении существенного пересечения границ. На третьем

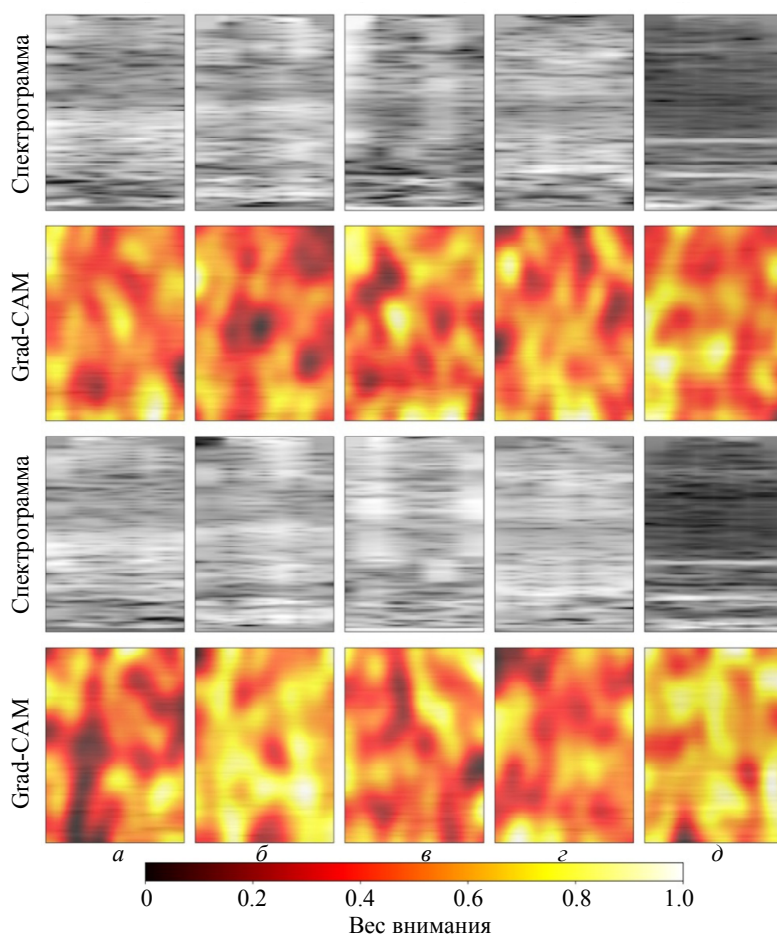


Рис. 4. Визуализация распределения внимания по различным типам неисправностей: *а* – неисправность сепаратора; *б* – смешанная неисправность; *в* – неисправность внутреннего кольца; *г* – неисправность наружного кольца; *д* – нормальное состояние

Fig. 4. Visualization of attention distribution across different fault types: *a* – cage fault; *б* – mixed fault; *в* – inner race fault; *г* – outer race fault; *д* – normal condition

уровне распределение признаков оптимизируется: большинство категорий образуют компактные кластеры, и лишь некоторые схожие типы неисправностей сохраняют близость. На четвертом уровне образцы различных категорий формируют независимые кластеры с четкими границами и значительно увеличенными межклассовыми расстояниями. Данная прогрессивная эволюция подтверждает эффективность иерархической архитектуры MSAFN: неглубокие модули извлекают многомасштабные локальные признаки, а глубокие модули интегрируют их в высокоуровневые семантические представления с выраженной дискриминационной способностью. Синергетическое взаимодействие механизма остаточного двойного внимания и модуля селективного состояния дополнительно усиливает межклассовую разделяемость признаков.

Исследования абляции. Для проверки эффективности каждого компонента модуля в MSAFN были спроектированы систематические исследования абляции. Поочередно удаляя или заменяя инновационные модули, количественно оценили их влияние на общую производительность. Табл. 5 представляет конфигурации модели для экспериментов абляции, включающие восемь сравнительных экспериментов. Полная модель обозначает предложенную MSAFN; DMS – динамическая многомасштабная свертка; FCA – внимание к каналам в частотной области; RDAB – остаточное двойное внимание; S3M – модуль селективного состояния; HPF – иерархическое прогрессивное слияние; MTL – удаление многозадачного обучения; VEC – векторная ветвь признаков. Эксперименты абляции были проведены с использованием более сложного набора

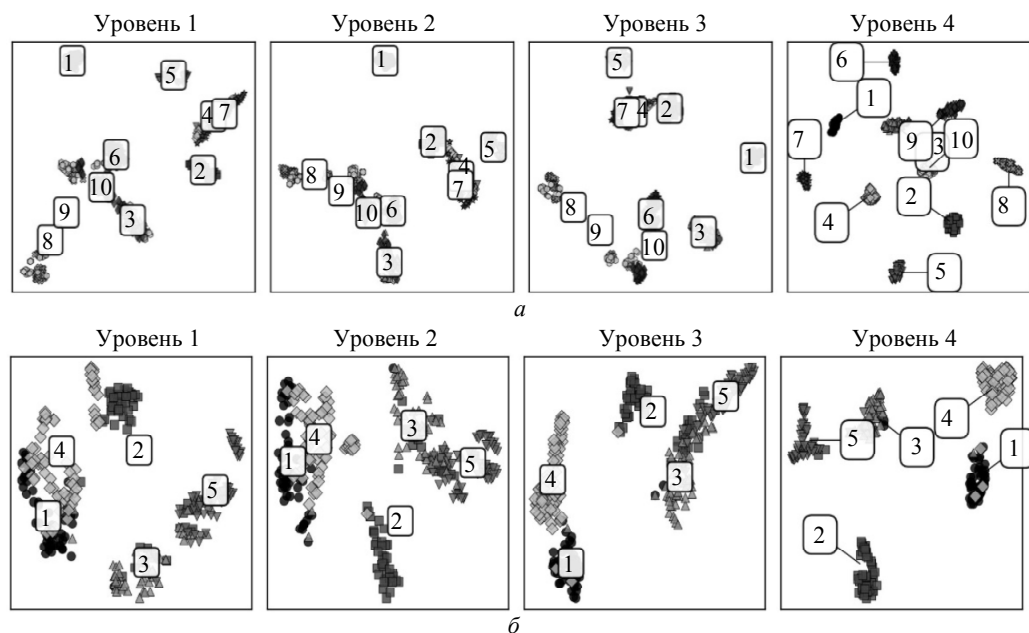


Рис. 5. Графики эволюции t -SNE для каждого слоя сети MSAFN: a – набор данных CWRU; b – набор данных PU. Цифры в рамках обозначают номера классов неисправностей
Fig. 5. t -SNE evolution plots for each layer of the MSAFN network: a – CWRU dataset; b – PU dataset. Numbers in boxes indicate fault class labels

Табл. 5. Конфигурации модели для экспериментов абляции
Tab. 5. Model configurations for ablation experiments

Метод	DMS	FCA	RDAB	S3M	HPF	MTL	VEC
Полная модель	✓	✓	✓	✓	✓	✓	✓
без DMS	×	✓	✓	✓	✓	✓	✓
без FCA	✓	×	✓	✓	✓	✓	✓
без RDAB	✓	✓	×	✓	✓	✓	✓
без S3M	✓	✓	✓	×	✓	✓	✓
без HPF	✓	✓	✓	✓	×	✓	✓
без MTL	✓	✓	✓	✓	✓	×	✓
без VEC	✓	✓	✓	✓	✓	✓	×

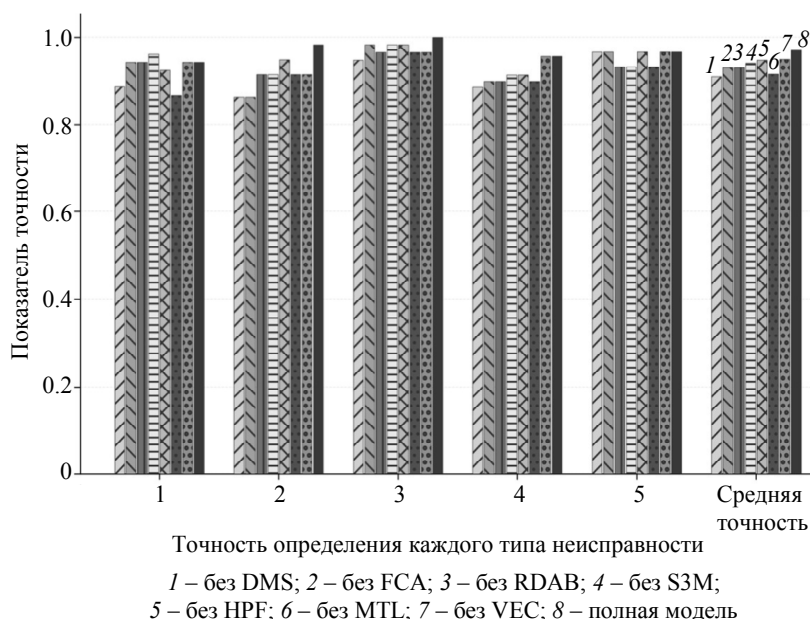


Рис. 6. Сравнение точности моделей абляции на наборе данных PU

Fig. 6. Comparison of ablation model accuracy on the PU dataset

данных PU. На рис. 6 представлены результаты экспериментов абляции на наборе данных PU.

Полная модель достигла оптимальной производительности со средней точностью 97.0 %, включая 100 %-ную точность распознавания для класса 3, что подтверждает необходимость синергетического взаимодействия между модулями. Наибольшее влияние оказало удаление динамической многомасштабной свертки, снизившее среднюю точность до 91.0 % с выраженным ухудшением распознавания классов 3 и 4. Это свидетельствует о ключевой роли механизма адаптивного выбора рецептивного поля в захвате паттернов неисправностей на различных масштабах. Существенное снижение производительности наблюдалось также при удалении иерархического прогрессивного слияния (точность 94.8 %) и модуля селективного состояния (точность 94.1 %), что подтверждает эффективность многоуровневых кросс-модальных взаимодействий и моделирования дальнедействующих зависимостей соответственно. Удаление внимания к каналам в частотной области, остаточного двойного внимания и многозадачного обучения в каждом случае приводило к снижению производительности приблизительно на 4 процентных пункта, отражая положительный вклад данных компонентов в общую точность модели. Исключение ветви статистических признаков снизило точность до 95.0 %, подтверждая дополнительную ценность статистического представления для временно-

частотного анализа. Таким образом, каждый модуль вносит значимый вклад в итоговую производительность, что обосновывает целесообразность предложенной архитектуры.

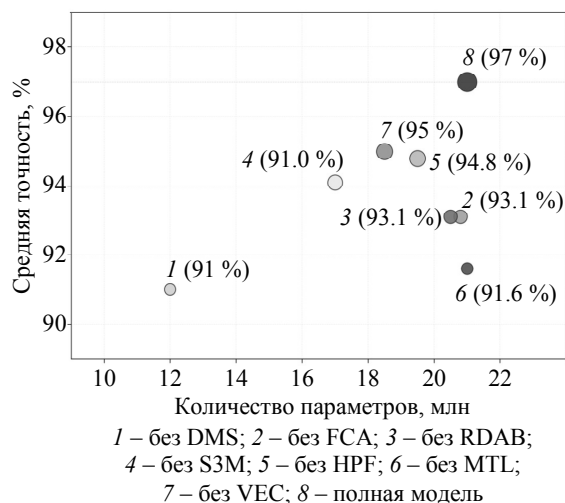


Рис. 7. Сравнение точности и количества параметров моделей абляции

Fig. 7. Comparison of accuracy and parameter count of ablation models

Рис. 7 визуально иллюстрирует зависимость между точностью и количеством параметров в моделях абляции с помощью диаграммы в виде пузырьков, где размер пузырька отражает относительную производительность модели. Полная модель достигает наивысшей точности 97.0 % с примерно 21 млн параметров, располагается в верхнем правом квадранте диаграммы, демонстрируя оптимальный баланс между производи-

тельностью и сложностью. Применительно к эффективности параметров, удаление динамического многомасштабного сверточного слоя значительно снижает их количество (до ≈ 12 млн,) но точность резко падает до 91.0 %. Это указывает на то, что, несмотря на увеличение количества параметров, этот модуль служит ключевым фактором повышения производительности. Удаление модуля селективного состояния снизило количество параметров на ≈ 4 млн, что привело к снижению точности до 94.1 %. Такое соотношение производительности и параметров указывает на то, что этот модуль достигает значительного улучшения способностей к моделированию дальнodelействующих зависимостей при умеренных вычислительных затратах. Удаление остаточного двойного внимания и многозадачного обучения привело к незначительным изменениям параметров, но точность снизилась до 93.1 и 91.6 % соответственно. Это демонстрирует, что данные два модуля обеспечивают значительный прирост производительности с минимальными затратами по параметрам, показывая высокую эффективность. Удаление ветви признаков вектора и иерархического прогрессивного слияния привело к точности 95.0 и 94.8 % соответственно, что показывает относительно умеренное снижение производительности, но при этом подтверждает эффективность двойной архитектуры и стратегии многоуровневого слияния. Удаление модуля внимания к каналам в частотной области снизило точность до 93.1 %, оказав незначительное влияние на количество параметров, что демонстрирует его рентабельность как компонента для легкого улучшения. В целом, полная модель достигает разумного баланса между эффективностью параметров и вкладом в производительность через ее модули, что в совокупности поддерживает превосходную производительность модели.

Заключение. В данной статье решаются проблемы диагностики неисправностей качающихся подшипников, включая трудности извлечения многомасштабных признаков, несовершенство механизмов внимания, ограниченные возможности моделирования дальнodelействующих зависимостей и подоптимальные эффекты мультимодального слияния. Авторы предлагают сеть мультимасштабного внимания и слияния (MSAFN). Этот подход использует двухветвевую архитектуру для одновременной обработки временно-частотных карт и статистических векторных при-

знаков. Она использует динамические многомасштабные сверточные модули для адаптивного выбора приемного поля, эффективно захватывая паттерны неисправностей на различных уровнях детализации. Внимание к каналам в частотной области улучшает восприятие критических частотных компонентов, в то время как остаточное двойное внимание комплексно усиливает дискриминирующие признаки по каналам, пространству и позициям. Модуль селективного состояния моделирует дальнodelействующие временные зависимости с линейной вычислительной сложностью, эффективно захватывая периодические характеристики вибрационных сигналов. Иерархический прогрессивный модуль слияния позволяет динамически взаимодействовать с несколькими уровнями абстракции, полностью используя дополнительную информацию из гетерогенных признаков. Стратегия многозадачного совместного обучения соединенными усилиями оптимизирует задачи классификации, реконструкции и контрастного обучения, улучшая обобщающую способность модели. Экспериментальные результаты на публичных наборах данных CWRU и PU показывают, что MSAFN достигает диагностической точности 99.67 и 96.0 % соответственно, превосходя основные методы глубокого обучения. Эксперименты абляции подтвердили положительный вклад каждого инновационного модуля в общую производительность. Визуализация с использованием *t*-SNE и анализ с тепловыми картами Grad-CAM показали, что признаки, изученные моделью, обладают хорошей разделяемостью классов, а процесс принятия решений с большой точностью соответствует механизму вибрации неисправностей подшипников, обеспечивая отличную интерпретируемость.

Дальнейшие исследования будут сосредоточены на следующих направлениях:

1. Изучение разработки легких моделей с использованием таких методов, как дистилляция знаний и обрезка сети, чтобы уменьшить вычислительную сложность и соответствовать требованиям реального времени на периферийных устройствах.
2. Введение механизмов адаптации к области для улучшения переносимости модели в разных эксплуатационных условиях и типах оборудования.
3. Расширение предложенного метода для диагностики неисправностей других вращающихся механических компонентов, чтобы подтвердить его обобщаемость и практическую ценность.

Список литературы

1. AGFCN: A bearing fault diagnosis method for high-speed train bogie under complex working conditions / D. He, J. Wu, Z. Jin, C. Huang, Z. Wei, C. Yi // *Reliability Engin. & Syst. Safety*. 2025. Vol. 258. P. 110907. doi: 10.1016/j.ress.2025.110907.
2. LLM-based framework for bearing fault diagnosis / L. Tao, H. Liu, G. Ning, W. Cao, B. Huang, Ch. Lu // *Mechan. Syst. and Signal Proc.* 2025. Vol. 224. P. 112127. doi: 10.1016/j.ymssp.2024.112127.
3. A triple domain adversarial neural network for bearing fault diagnosis / C. Chen, X. Li, J. Shi, D. Yue, C. Wang, L. Feng, H. Chen, A. A. Churakova, I. V. Alexandrov // *Mechan. Syst. and Signal Proc.* 2025. Vol. 238. P. 113202. doi: 10.1016/j.ymssp.2025.113202.
4. Interpretable convolutional neural network with multilayer wavelet for noise-robust machinery fault diagnosis / H. Wang, Z. Liu, D. Peng, M. J. Zuo // *Mechan. Syst. and Signal Proc.* 2023. Vol. 195. P. 110314. doi: 10.1016/j.ymssp.2023.110314.
5. Attention-based bilinear feature fusion method for bearing fault diagnosis / D. Wang, Y. Li, L. Jia, Y. Song, T. Wen // *IEEE/ASME Trans. on Mechatronics*. 2023. Vol. 28, no. 3. P. 1695–1705. doi: 10.1109/TMECH.2022.3223358.
6. Neural-transformer: A brain-inspired lightweight mechanical fault diagnosis method under noise / C. Wang, B. Tian, J. Yang, H. Jie, Y. Chang, Z. Zhao // *Reliability Engin. Syst. Safety*. 2024. Vol. 251. P. 110409. doi: 10.1016/j.ress.2024.110409.
7. Federated transfer learning for machinery fault diagnosis: A comprehensive review of technique and application / Q. Qian, B. Zhang, C. Li, Y. Mao, Y. Qin // *Mechan. Syst. and Signal Proc.* 2025. Vol. 223. P. 111837. doi: 10.1016/j.ymssp.2024.111837.
8. RTSMFFDE-HKRR: A fault diagnosis method for train bearing in noise environment / D. He, Z. Zhang, Z. Jin, F. Zhang, C. Yi, S. Liao // *Measurement*. 2025. Vol. 239. P. 115417. doi: 10.1016/j.measurement.2024.115417.
9. A roadmap to fault diagnosis of industrial machines via machine learning: A brief review / G. Vashishtha, S. Chauhan, M. Sehri, R. Zimroz, P. Dumond, R. Kumar, M. K. Gupta // *Measurement*. 2025. Vol. 242. P. 116216. doi: 10.1016/j.measurement.2024.116216.
10. A comprehensive survey on domain adaptation for intelligent fault diagnosis / C. Wang, Z. Wang, Q. Liu, H. Dong, W. Liu, X. Liu // *Knowledge-Based Syst.* 2025. Vol. 327. P. 114109. doi: 10.1016/j.knosys.2025.114109.
11. Weak fault feature extraction method of rolling bearings based on MVO-MOMEDA under strong noise interference / Z. Lv, L. Peng, Y. Cao, L. Yang, L. Li, C. Zhou // *IEEE Sensors J.* 2023. Vol. 23, no. 14. P. 15732–15740. doi: 10.1109/JSEN.2023.3277516.
12. An adaptive variational mode extraction method based on multi-domain and multi-objective optimization for bearing fault diagnosis / X. Yan, W. Yan, K. V. Yuen, Z. Yang, X. Wang // *Structural Health Monitoring*. 2023. Vol. 22, no. 4. P. 2708–2733. doi: 10.1177/14759217221133283.
13. Dynamic time warping using graph similarity guided symplectic geometry mode decomposition to detect bearing faults / J. Guo, Z. Si, Y. Liu, J. Li, Y. Li, J. Xiang // *Reliability Engin. & Syst. Safety*. 2022. Vol. 224. P. 108533. doi: 10.1016/j.ress.2022.108533.
14. CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis / D. Ruan, J. Wang, J. Yan, C. Gühmann // *Advanced Engin. Informatics*. 2023. Vol. 55. P. 101877. doi: 10.1016/j.aei.2023.101877.
15. Deep residual learning for image recognition / K. He, X. Zhang, Sh. Ren, J. Sun // 2016 IEEE Conf. on Comp. Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. P. 770–778. doi: 10.1109/CVPR.2016.90.
16. Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment / P. Liang, W. Wang, X. Yuan, S. Liu, L. Zhang, Y. Cheng // *Engin. Appl. of Artificial Intelligence*. 2022. Vol. 115. P. 105269. doi: 10.1016/j.engap-pai.2022.105269.
17. A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems / T. Huang, Q. Zhang, X. Tang, S. Zhao, X. Lu // *Artificial Intelligence Rev.* 2022. Vol. 55, no. 2. P. 1289–1315. doi: 10.1007/s10462-021-09993-z.
18. Attention is all you need / A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin. 2017. URL: <https://arxiv.org/abs/1706.03762> (дата обращения: 20.02.2026).
19. MDFT-GAN: A multi-domain feature transformer GAN for bearing fault diagnosis under limited and imbalanced data conditions / C. Guo, V. V. Potekhin, P. Li, E. A. Kovalchuk, J. Lian // *Appl. Scie.* 2025. Vol. 15, no. 11. P. 6225. doi: 10.3390/app15116225.
20. TCAC-transformer: A fast convolutional transformer with temporal-channel attention for efficient industrial fault diagnosis / W. Wu, N. Zhou, X. Liang, W. Gui, C. Yang, Y. Liu // *Expert Syst. with Appl.*, 2026. Vol. 297, part C. P. 129473. doi: 10.1016/j.eswa.2025.129473.
21. Robust bearing fault diagnosis via bio-inspired spectral enhancement and multimodal contrastive graph attention / Ch. Guo, V. V. Potekhin, J. Lian, P. Li // *IEEE Sensors J.* 2026. Vol. 26, no. 10. P. 15872–15898. doi: 10.1109/JSEN.2026.3680626.
22. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification / Ch. Lessmeier, J. K. Kimotho, D. Zimmer, W. Sextro // *Proc. of 3rd Europ. Conf. of the Prognostics and Health Management Soc. Bilbao, Spain: PHM Society*, 2016. Vol. 3, no. 1. P. 1–17.

23. A deep learning approach based on MLP-mixer models for bearing fault diagnosis / V. Q. Nguyen, M. H. Vu, V. T. Pham, T. T. Tran // 2023 Intern. Conf. on Syst. Sci. and Engin. (ICSSE). Ho Chi Minh, Vietnam: IEEE, 2023. P. 16–21. doi: 10.1109/ICSSE58758.2023.10227218.

24. An attention EfficientNet-based strategy for bearing fault diagnosis under strong noise / B. Hu, J. Tang, J. Wu, J. Qing // Sensors. 2022. Vol. 22, no. 17. P. 6570. doi: 10.3390/s22176570.

25. Rolling bearing fault diagnosis method based on PE-DCM and ViT / Y. Hui, K. Xu, P. Chen, X. Zhao // Measurement Sci. and Technol. 2024. Vol. 35, no. 10. P. 105107. doi: 10.1088/1361-6501/ad5eab.

26. FEV-Swin: Multi-source heterogeneous information fusion under a variant swin transformer frame-

work for intelligent cross-domain fault diagnosis / K. Zhou, N. Lu, B. Jiang, Z. Ye // Knowledge-Based Syst., 2025. Vol. 310. P. 112982. doi: 10.1016/j.knsys.2025.112982.

27. Attention-based ConvNeXt with a parallel multiscale dilated convolution residual module for fault diagnosis of rotating machinery / B. Guo, Z. Qiao, N. Zhang, Y. Wang, F. Wu, Q. Peng // Expert Syst. with Appl. 2024. Vol. 249, part C. P. 123764. doi: 10.1016/j.eswa.2024.123764.

28. Grad-CAM: Visual explanations from deep networks via gradient-based localization / R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, Dh. Parikh, D. Batra // 2017 IEEE Intern. Conf. on Comp. Vision (ICCV). Venice, Italy: IEEE, 2017. P. 618–626. doi: 10.1109/ICCV.2017.74.

Информация об авторах

Потехин Вячеслав Витальевич – канд. техн. наук, доцент Высшей школы управления киберфизическими системами, Санкт-Петербургский политехнический университет Петра Великого, ул. Политехническая, 29, Санкт-Петербург, 195251, Россия.

E-mail: slava.potekhin@spbstu.ru

<https://orcid.org/0000-0001-9850-9558>

Го Чэньси – аспирант Высшей школы управления киберфизическими системами, Санкт-Петербургский политехнический университет Петра Великого, ул. Политехническая, 29, Санкт-Петербург, 195251, Россия.

E-mail: chenxiguu.academic@gmail.com

<https://orcid.org/0000-0002-1497-2893>

References

1. AGFCN: A bearing fault diagnosis method for high-speed train bogie under complex working conditions / D. He, J. Wu, Z. Jin, C. Huang, Z. Wei, C. Yi // Reliability Engin. & Syst. Safety. 2025. Vol. 258. P. 110907. doi: 10.1016/j.ress.2025.110907.

2. LLM-based framework for bearing fault diagnosis / L. Tao, H. Liu, G. Ning, W. Cao, B. Huang, Ch. Lu // Mechan. Syst. and Signal Proc. 2025. Vol. 224. P. 112127. doi: 10.1016/j.ymssp.2024.112127.

3. A triple domain adversarial neural network for bearing fault diagnosis / C. Chen, X. Li, J. Shi, D. Yue, C. Wang, L. Feng, H. Chen, A. A. Churakova, I. V. Alexandrov // Mechan. Syst. and Signal Proc. 2025. Vol. 238. P. 113202. doi: 10.1016/j.ymssp.2025.113202.

4. Interpretable convolutional neural network with multilayer wavelet for noise-robust machinery fault diagnosis / H. Wang, Z. Liu, D. Peng, M. J. Zuo // Mechan. Syst. and Signal Proc. 2023. Vol. 195. P. 110314. doi: 10.1016/j.ymssp.2023.110314.

5. Attention-based bilinear feature fusion method for bearing fault diagnosis / D. Wang, Y. Li, L. Jia, Y. Song, T. Wen // IEEE/ASME Trans. on Mechatronics. 2023. Vol. 28, no. 3. P. 1695–1705. doi: 10.1109/TMECH.2022.3223358.

6. Neural-transformer: A brain-inspired lightweight mechanical fault diagnosis method under noise / C. Wang, B. Tian, J. Yang, H. Jie, Y. Chang, Z. Zhao // Relia-

bility Engin. Syst. Safety. 2024. Vol. 251. P. 110409. doi: 10.1016/j.ress.2024.110409.

7. Federated transfer learning for machinery fault diagnosis: A comprehensive review of technique and application / Q. Qian, B. Zhang, C. Li, Y. Mao, Y. Qin // Mechan. Syst. and Signal Proc. 2025. Vol. 223. P. 111837. doi: 10.1016/j.ymssp.2024.111837.

8. RTSMFFDE-HKRR: A fault diagnosis method for train bearing in noise environment / D. He, Z. Zhang, Z. Jin, F. Zhang, C. Yi, S. Liao // Measurement. 2025. Vol. 239. P. 115417. doi: 10.1016/j.measurement.2024.115417.

9. A roadmap to fault diagnosis of industrial machines via machine learning: A brief review / G. Vashishtha, S. Chauhan, M. Sehri, R. Zimroz, P. Dumond, R. Kumar, M. K. Gupta // Measurement. 2025. Vol. 242. P. 116216. doi: 10.1016/j.measurement.2024.116216.

10. A comprehensive survey on domain adaptation for intelligent fault diagnosis / C. Wang, Z. Wang, Q. Liu, H. Dong, W. Liu, X. Liu // Knowledge-Based Syst. 2025. Vol. 327. P. 114109. doi: 10.1016/j.knsys.2025.114109.

11. Weak fault feature extraction method of rolling bearings based on MVO-MOMEDA under strong noise interference / Z. Lv, L. Peng, Y. Cao, L. Yang, L. Li, C. Zhou // IEEE Sensors J. 2023. Vol. 23, no. 14. P. 15732–15740. doi: 10.1109/JSEN.2023.3277516.

12. An adaptive variational mode extraction method based on multi-domain and multi-objective optimization for bearing fault diagnosis / X. Yan, W. Yan, K. V. Yuen, Z. Yang, X. Wang // *Structural Health Monitoring*. 2023. Vol. 22, no. 4. P. 2708–2733. doi: 10.1177/14759217221133283.
13. Dynamic time warping using graph similarity guided symplectic geometry mode decomposition to detect bearing faults / J. Guo, Z. Si, Y. Liu, J. Li, Y. Li, J. Xiang // *Reliability Engin. & Syst. Safety*. 2022. Vol. 224. P. 108533. doi: 10.1016/j.ress.2022.108533.
14. CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis / D. Ruan, J. Wang, J. Yan, C. Gühmann // *Advanced Engin. Informatics*. 2023. Vol. 55. P. 101877. doi: 10.1016/j.aei.2023.101877.
15. Deep residual learning for image recognition / K. He, X. Zhang, Sh. Ren, J. Sun // 2016 IEEE Conf. on Comp. Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. P. 770–778. doi: 10.1109/CVPR.2016.90.
16. Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment / P. Liang, W. Wang, X. Yuan, S. Liu, L. Zhang, Y. Cheng // *Engin. Appl. of Artificial Intelligence*. 2022. Vol. 115. P. 105269. doi: 10.1016/j.engap-pai.2022.105269.
17. A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems / T. Huang, Q. Zhang, X. Tang, S. Zhao, X. Lu // *Artificial Intelligence Rev.* 2022. Vol. 55, no. 2. P. 1289–1315. doi: 10.1007/s10462-021-09993-z.
18. Attention is all you need / A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin. 2017. URL: <https://arxiv.org/abs/1706.03762> (дата обращения: 20.02.2026).
19. MDFT-GAN: A multi-domain feature transformer GAN for bearing fault diagnosis under limited and imbalanced data conditions / C. Guo, V. V. Potekhin, P. Li, E. A. Kovalchuk, J. Lian // *Appl. Sci.* 2025. Vol. 15, no. 11. P. 6225. doi: 10.3390/app15116225.
20. TCAC-transformer: A fast convolutional transformer with temporal-channel attention for efficient industrial fault diagnosis / W. Wu, N. Zhou, X. Liang, W. Gui, C. Yang, Y. Liu // *Expert Syst. with Appl.*, 2026. Vol. 297, part C. P. 129473. doi: 10.1016/j.eswa.2025.129473.
21. Robust bearing fault diagnosis via bio-inspired spectral enhancement and multimodal contrastive graph attention / Ch. Guo, V. V. Potekhin, J. Lian, P. Li // *IEEE Sensors J.* 2026. Vol. 26, no. 10. P. 15872–15898. doi: 10.1109/JSEN.2026.3680626.
22. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification / Ch. Lessmeier, J. K. Kimotho, D. Zimmer, W. Sextro // *Proc. of 3rd Europ. Conf. of the Prognostics and Health Management Soc. Bilbao, Spain: PHM Society*, 2016. Vol. 3, no. 1. P. 1–17.
23. A deep learning approach based on MLP-mixer models for bearing fault diagnosis / V. Q. Nguyen, M. H. Vu, V. T. Pham, T. T. Tran // 2023 Intern. Conf. on Syst. Sci. and Engin. (ICSSE). Ho Chi Minh, Vietnam: IEEE, 2023. P. 16–21. doi: 10.1109/ICSSE58758.2023.10227218.
24. An attention EfficientNet-based strategy for bearing fault diagnosis under strong noise / B. Hu, J. Tang, J. Wu, J. Qing // *Sensors*. 2022. Vol. 22, no. 17. P. 6570. doi: 10.3390/s22176570.
25. Rolling bearing fault diagnosis method based on PE-DCM and ViT / Y. Hui, K. Xu, P. Chen, X. Zhao // *Measurement Sci. and Technol.* 2024. Vol. 35, no. 10. P. 105107. doi: 10.1088/1361-6501/ad5eab.
26. FEV-Swin: Multi-source heterogeneous information fusion under a variant swin transformer framework for intelligent cross-domain fault diagnosis / K. Zhou, N. Lu, B. Jiang, Z. Ye // *Knowledge-Based Syst.*, 2025. Vol. 310. P. 112982. doi: 10.1016/j.knsys.2025.112982.
27. Attention-based ConvNeXt with a parallel multiscale dilated convolution residual module for fault diagnosis of rotating machinery / B. Guo, Z. Qiao, N. Zhang, Y. Wang, F. Wu, Q. Peng // *Expert Syst. with Appl.* 2024. Vol. 249, part C. P. 123764. doi: 10.1016/j.eswa.2024.123764.
28. Grad-CAM: Visual explanations from deep networks via gradient-based localization / R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, Dh. Parikh, D. Batra // 2017 IEEE Intern. Conf. on Comp. Vision (ICCV). Venice, Italy: IEEE, 2017. P. 618–626. doi: 10.1109/ICCV.2017.74.

Information about the authors

Vyacheslav V. Potekhin – Cand. Sci. (Eng.), Associate Professor of the Graduate School of Cyber-Physical Systems Control, Peter the Great St. Petersburg Polytechnic University, 29 Polytechnicheskaya St., Saint Petersburg, 195251, Russia.

E-mail: slava.potekhin@spbstu.ru

<https://orcid.org/0000-0001-9850-9558>

Chenxi Guo – postgraduate student of the Graduate School of Cyber-Physical Systems Control, Peter the Great St. Petersburg Polytechnic University, 29 Polytechnicheskaya St., Saint Petersburg, 195251, Russia.

E-mail: chenxiguao.academic@gmail.com

<https://orcid.org/0000-0002-1497-2893>

Статья поступила в редакцию 19.01.2026; принята к публикации после рецензирования 21.05.2026; опубликована онлайн 28.09.2026.

Submitted 19.01.2026; accepted 21.05.2026; published online 28.09.2026.