

УДК 004.9

НЕЙРОСЕТЕВЫЕ МЕТОДЫ ОЦЕНКИ БЕЗОТКАЗНОСТИ И ОСТАТОЧНОГО РЕСУРСА НАСОСНОГО АГРЕГАТА НА ОСНОВЕ DATASETA PUMP SENSOR DATA

Воронов Р. Е., аспирант гр. МРа-241, II курс
Научный руководитель: Чичерин И. В., к.т.н., доцент
Кузбасский государственный технический университет
имени Т. Ф. Горбачева, г. Кемерово

В современных промышленных системах насосные агрегаты выполняют критически важные функции, обеспечивая транспортировку жидкостей, водоотведение и устойчивость технологических процессов. Нарушение их работоспособности ведет к простоям оборудования, увеличению затрат на ремонт, снижению безопасности эксплуатации и ухудшению производственных показателей.

Традиционные подходы к оценке надежности насосного оборудования основаны на статистике отказов, регламентном обслуживании и периодическом контроле диагностических параметров. Однако при наличии непрерывного сенсорного мониторинга возникает возможность перейти к более точной модели управления техническим состоянием, основанной на выявлении ранних признаков деградации и прогнозе отказа в реальном времени. Как отмечается в литературе, «выбор подхода к прогнозированию отказов оборудования зависит от специфики задач, доступных ресурсов и требований к точности прогнозов» [1].

Особую ценность в этом контексте представляют открытые наборы данных, позволяющие проверять методы машинного обучения и нейросетевого анализа на стандартизированной эмпирической базе. Одним из таких наборов является pump sensor data, опубликованный на платформе Kaggle и содержащий многомерные временные ряды, сформированные по показаниям сенсоров насосного агрегата, а также целевую метку состояния оборудования.

По описанию дата сета, в его структуру входят временная метка, 52 сенсорных переменных и колонка статуса оборудования. Сами данные организованы как последовательность наблюдений с минутным шагом. Кроме того, набор содержит три класса состояния оборудования – normal, recovering, broken, что делает его пригодным для исследования задач раннего обнаружения отказов и классификации технического состояния насоса.

Таблица 1 – Основные характеристики датасета

Параметр	Характеристика
Объект наблюдения	Насосная установка
Тип данных	Многомерные временные ряды сенсорных измерений

Временная структура	Последовательность наблюдений с минутным шагом
Состав признаков	Временная метка, 52 сенсорных переменных, статус оборудования
Целевая переменная	Machine status
Класс состояния	Normal, recovering, broken



Рисунок 1 – Изменение состояния насосного агрегата во времени

Как видно из рисунка №1, наблюдения формируют непрерывный временной ряд, в котором доминирует состояние NORMAL, а переходы в RECOVERING и особенно в BROKEN встречаются существенно реже. Это подтверждает целесообразность рассмотрения набора данных как основы для прогнозирования технического состояния во времени.

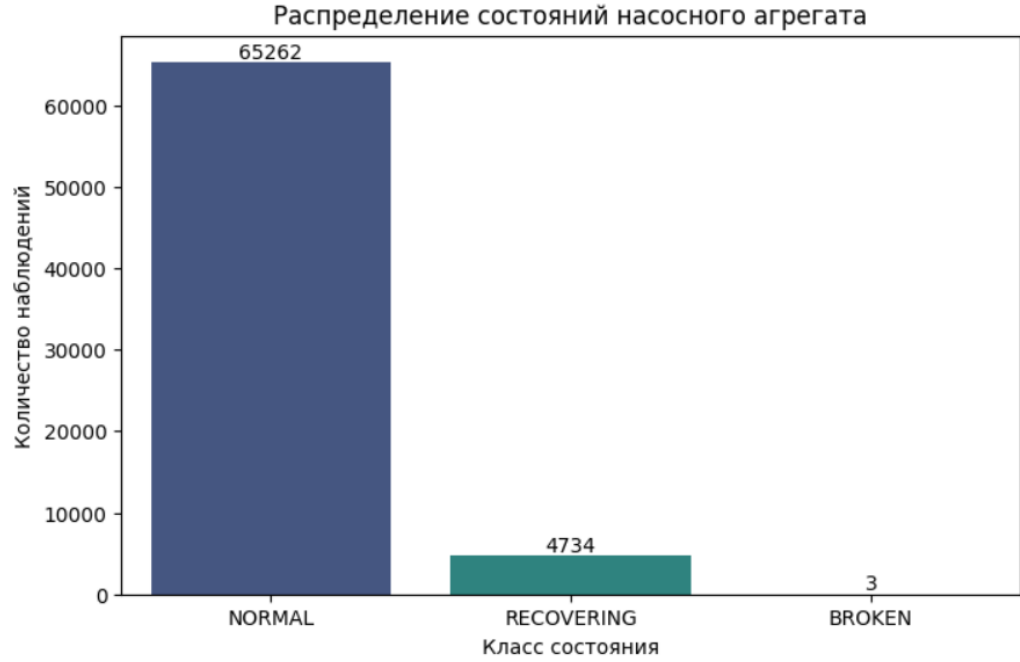


Рисунок 2 – распределение наблюдений по классам состояния

Следует отметить, что рассматриваемый датасет не содержит готовой колонки остаточного ресурса в явном виде. Поэтому для решения задачи прогнозирования «remaining useful life» необходимо построение производной целевой переменной, отражающей временное расстояние от текущего наблюдения до ближайшего отказа, зафиксированного в ряду состояний – machine_status. При

этом необходимо учитывать, что «чем выше качество данных, тем более точными и надежными будет прогноз» [2].

В литературе по прогнозированию остаточного ресурса многокомпонентных систем показано, что нейросетевые архитектуры, в частности CNN-LSTM, обладает высокой эффективностью в задачах обработки временных рядов деградации. В работе по прогнозированию RUL на данных NASA Turbofan гибридная модель CNN-LSTM продемонстрировала лучшие результаты по сравнению с рядом классических методов машинного обучения, что делает подобный подход методически релевантным и для насосных систем. Одновременно исследования по насосному оборудованию подтверждают, что нейросетевые методы применимы и к оценке вероятности безотказной работы. В частности, в работе по вероятностному анализу надежности установлено, что нейросетевой прогноз позволяет уточнить техническое состояние насосной установки, а повышение температурных параметров работы электродвигателя сопровождается снижением вероятности безотказной работы.

Доступные характеристики датасета, даже без полного вычислительного эксперимента, позволяют сделать ряд методически важных выводов. Во-первых, наличие большого числа сенсорных каналов создает предпосылки для выявления скрытых взаимосвязей между признаками и техническим состоянием насоса, что делает нейросетевые методы более предпочтительными по сравнению с простыми пороговыми схемами. Во-вторых, наличие трех классов состояния оборудования показывает, что деградация представлена не только как бинарный переход «исправлен/неисправлен», но и как промежуточная стадия «recovering». Это важно для практики предупреждения, технического обслуживания, поскольку позволяет строить модели раннего предупреждения, ориентированные не только на отказ, но и на предаварийные переходные режимы. В-третьих, использование производной величины RUL представляется методически оправданным компромиссом между содержанием датасета и задачами статьи. Поскольку исходный набор не содержит инженерно измеренного остаточного ресурса, расчет времени до ближайшего «broken» позволяет сформировать воспроизводимую цель для регрессионного нейросетевого моделирования.

При этом следует учитывать ряд ограничений. Открытые описания датасета не заменяют полного технического паспорта измерительной системы, а целевая колонка состояния сама по себе не раскрывает физический механизм деградации отдельных узлов агрегата. Поэтому результаты моделирования следует интерпретировать как оценку поведенческого остаточного ресурса.

Еще одним важным ограничением является риск переобучения. Материалы Google for Developers подчеркивают, что переобучение возникает тогда, когда модель хорошо запоминает обучающие примеры, но теряет способность корректно работать на новых данных, для временных рядов промышленного мониторинга это особенно вероятно при неправильном разбиении выборки или избыточной сложности сети.

Следовательно, при построении нейросетевой модели необходимо обеспечивать временно корректную валидацию, контроль дисбаланса классов и сопоставление метрик качества с реальной задачей эксплуатации. Только при соблюдении этих условий результаты могут быть интерпретированы как основание для разработки системы прогнозного обслуживания насосного оборудования.

В ходе вычислительного эксперимента была реализована LSTM – модель для классификации технического состояния насосного агрегата по многоканальным сенсорным временным рядам. При подготовке данных использовались удаление неинформативных признаков, заполнение пропусков.

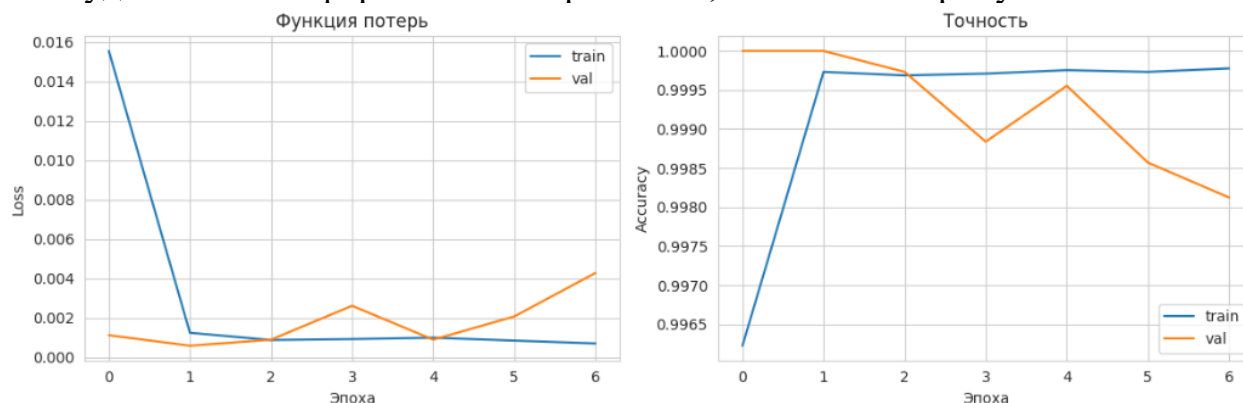


Рисунок 3 - Обучение модели

По результатам обучения модель продемонстрировала общую точность на тестовой выборке 0,96%.

	precision	recall	f1-score	support
BROKEN	0.00	0.00	0.00	1
NORMAL	0.96	1.00	0.98	13313
RECOVERING	1.00	0.10	0.18	680
accuracy			0.96	13994
macro avg	0.65	0.37	0.39	13994
weighted avg	0.96	0.96	0.94	13994

Рисунок 4 - Оценка качества модели

Таким образом, результаты эксперимента подтверждают применимость нейросетевого подхода для анализа сенсорных временных рядов насосного оборудования, но одновременно показывают необходимость дополнительных мер по компенсации дисбаланса классов. В качестве таких мер могут рассматриваться взвешивание классов, переразметка задачи, укрупнение классов или переход к прогнозированию остаточного ресурса как более устойчивой постановке задачи.

Открытый датасет pump sensor data представляет собой релевантную эмпирическую базу для исследования нейросетевых методов оценки технического состояния насосного агрегата. Наличие многоканальных сенсорных временных рядов и целевой переменной machine_status делает этот набор данных пригодным для решения задач классификации безотказности и построения приближенной оценки остаточного ресурса.

Показано, что наиболее корректной постановкой задачи для данного набора является сочетание двух направлений анализа: прогноза вероятности безотказной работы по классам состояния и расчета прокси-остаточного ресурса как времени до ближайшего отказа. Такая схема позволяет связать нейросетевую обработку сенсорных данных с задачами профилактического технического обслуживания.

Перспективными моделями для решения поставленной задачи являются рекуррентные и гибридные нейросетевые архитектуры, прежде всего LSTM, NARX и CNN-LSTM, показавшие высокую результативность в задачах анализа временных рядов деградации технических систем.

Таблица 2 – Сопоставление нейросетевых моделей

Модель	Сильные стороны	Ограничения	Предпочтительная задача
LSTM	Учитывает долгосрочные зависимости во временных рядах	Требует большой объем данных	Прогноз технического состояния и RUL
NARX	Моделирование динамики системы	Сложная настройка лагов	Прогнозирование поведения сенсорных сигналов
CNN-LSTM	Извлечение локальных паттернов	Сложна в настройке и чувствительна к данным	Совместный анализ сенсорных каналов и оценка RUL

Список литературы:

1. Воронов, Р. Е. Преимущества и недостатки различных подходов к прогнозированию отказов оборудования с помощью технологий искусственного интеллекта / Р. Е. Воронов // Современные тенденции развития и перспективы внедрения инновационных технологий в машиностроении, образовании и экономике. – 2025. – Т. 7, № 1. – С. 58-61. – EDN ZBQVIB.

2. Влияние качества данных на точность прогнозирования отказов оборудования с использованием ии // Экономика будущего: тренды, вызовы и возможности : материалы III Всероссийской научно-практической студенческой конференции с международным участием, Казань, 22–23 мая 2025 года. – Казань: ИП Сагиев А.Р., 2025. – С. 726-727. – EDN XLCPYQ.