Information about the authors

Mikhail Sergeyevich Murushkin – Senior GSD Coordinator and Business Expert in Digital Services, Airbus Kazakhstan LLP, Astana, Republic of Kazakhstan; e-mail: mikhail.murushkin@gmail.com. ORCID: https://orcid.org/0009-0006-9917-3842.

Kazbek Suleimenovich Baktybekov – Doctor of Physical and Mathematical Sciences, Chief Researcher, Ghalam LLP, Astana, Republic of Kazakhstan; e-mail: k.baktybekov@ghalam.kz. ORCID: https://orcid.org/0000-0002-6401-8053.

Berik Rakhimbekovich Zhumazhanov – Head of the Payload and Scientific Research Department, LLP «Ghalam»; e-mail: b.zhumazhanov@ghalam.kz. ORCID: http://orcid.org/0000-0001-5926-9619.

Aigul Yergaliyevna Kulakayeva* – PhD, Associate Professor of the Department of «Radio Engineering, Electronics and Telecommunications», International Information Technologies University, Almaty, Republic of Kazakhstan; e-mail: a.kulakayeva@iitu.edu.kz. ORCID: https://orcid.org/0000-0002-0143-085X.

Поступила в редакцию 05.01.2026
Поступила после доработки 04.02.2026
Принята к публикации 06.02.2026

https://doi.org/10.53360/2788-7995-2026-1(21)-6

МРНТИ: 50.43.15

Д.Т. Алдашева[1]* , О.С. Салыкова[1], М.Ю. Зарубин[2], Т.А. Жуаспаев[2], М.Д. Мусина[2]
[1]Костанайский региональный университет имени Ахмет Байтұрсынұлы,
110000, Республика Казахстан, г. Костанай, ул. А. Байтурсынова 47
[2]Костанайский инженерно-экономический университет имени М. Дулатова,
110000, Республика Казахстан, г. Костанай, ул. Чернышевского 59
*e-mail: aldasheva.dinara@mail.ru

EXPLAINABLE AI ДЛЯ ИНТЕРПРЕТАЦИИ ПРОГНОЗОВ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ ПРЕДИКТИВНОГО ОБСЛУЖИВАНИЯ

Аннотация: *В статье рассматривается применение методов Explainable Artificial Intelligence (XAI) для интерпретации прогнозов нейросетевых моделей в системах предиктивного обслуживания промышленного оборудования. Цель исследования – повышение прозрачности, точности и доверия к результатам машинного обучения при диагностике и прогнозировании технических неисправностей. В работе использованы методы SHAP, LIME и Grad-CAM, обеспечивающие визуализацию вклада признаков в предсказания моделей. На примере предприятий Костанайского автомобильного кластера показано, что интеграция Explainable AI позволяет повысить достоверность диагностики на 12-15%, сократить внеплановые простои оборудования на 13% и уменьшить время реагирования инженеров на 18%. Исследование демонстрирует, что интерпретируемые модели машинного обучения формируют основу для прозрачных решений в рамках цифровизации и внедрения промышленного интернета вещей (IIoT) в Казахстане. Предложенный подход соответствует стратегическим направлениям программ «Цифровой Казахстан» и «Концепция развития искусственного интеллекта до 2029 года», способствуя повышению эффективности и надёжности отечественных систем технического обслуживания.*

Экспериментальные результаты получены на обезличенных данных промышленного предприятия автомобильного машиностроения Республики Казахстан, что подтверждает применимость предложенного подхода в реальных производственных условиях.

Ключевые слова: *Explainable AI, предиктивное обслуживание, интерпретируемость моделей, машинное обучение, нейронные сети, XAI-подход.*

Введение

Современная промышленность в Казахстане, как и во многих странах, всё активнее переходит к стратегии предиктивного обслуживания (Predictive Maintenance, PdM), ориентированной на минимизацию простоев, продление срока службы оборудования и снижение операционных затрат. Это особенно важно для таких отраслей, как добыча,

переработка, машиностроение и энергетика, где сбои ведут к значительным экономическим потерям. Основой PdM служат методы машинного обучения и нейросетевые модели, анализирующие данные с датчиков, журналов, телеметрии, чтобы выявлять ранние признаки деградации. Однако высокая сложность моделей, особенно глубоких нейронных сетей, порождает проблему «чёрного ящика»: отсутствие прозрачности в том, как и почему принимается то или иное решение.

Современные исследования подчёркивают быстрое развитие интерпретируемых моделей для систем предиктивного обслуживания в промышленности [1]. Moosavi et al. (2024) представили систематический обзор методов Explainable AI в промышленности и кибер-физических производственных системах, показав, что интеграция XAI обеспечивает повышение доверия к ИИ-алгоритмам, улучшение контроля производственных процессов и возможность выявления аномалий в сложных инженерных системах [1]. Авторы подчёркивают, что именно интерпретируемость моделей становится ключевым фактором для их внедрения в промышленное производство.

Li & Li (2025) выполнили сравнительный анализ архитектур глубокого обучения – CNN, LSTM и гибридных моделей – для задач предиктивной диагностики оборудования на основе сенсорных данных. Результаты показали, что гибридная архитектура LSTM + CNN обеспечивает точность свыше 96% при диагностике деградации механических агрегатов и вибрационного мониторинга [2]. Это подтверждает целесообразность использования рекуррентных моделей, аналогичных применённым в настоящем исследовании.

Систематический обзор Cummins et al. (2023) показал, что XAI-технологии становятся критически значимыми для промышленной эксплуатации ИИ-моделей, так как позволяют инженерам интерпретировать предсказания, анализировать вклад отдельных параметров и принимать обоснованные решения при планировании технического обслуживания [3]. Аналогичные выводы представлены в исследовании Nor et al. (2021), где авторы отмечают, что включение XAI-модулей в PHM-системы (Prognostics and Health Management) повышает доверие персонала, сокращает риски неверных решений и предоставляет прозрачную аргументационную базу для выводов предиктивных моделей [4]. Кроме того, исследователи подчеркивают необходимость разработки единых метрик оценки объяснимости и стандартизированных подходов к внедрению XAI в промышленности.

Таким образом, международный опыт свидетельствует о высокой перспективности развития Explainable AI для PdM-задач и подтверждает актуальность использования интерпретируемых нейросетевых моделей в промышленном секторе Казахстана.

На основе анализа мировых и отечественных исследований сформирована собственная методология, сочетающая элементы Explainable AI и предиктивной аналитики для промышленного оборудования Казахстана. Далее изложены материалы и методы, отражающие архитектуру модели, особенности данных и алгоритмы интерпретации прогнозов.

В рамках исследования впервые предложена адаптация методов Explainable AI (SHAP и LIME) для интерпретации прогнозов рекуррентной нейронной сети LSTM в условиях реального производственного оборудования казахстанского машиностроительного сектора [5-7]. Новизна заключается в выявлении устойчивых закономерностей влияния вибрации, температуры и времени наработки на вероятность отказа агрегатов с возможностью объяснения причин формирования прогноза. Разработанная модель обеспечивает повышение прозрачности вычислительных решений и формирует методическую основу для внедрения интерпретируемых систем предиктивного обслуживания на отечественных предприятиях.

Условия и методы исследования

В качестве базового объекта исследования выбрано ТОО «СарыаркаАвтоПром», являющееся ядром Костанайского автомобильного кластера и одним из ведущих предприятий Республики Казахстан по выпуску легковых и коммерческих автомобилей.

Исследование проводилось на станочном оборудовании механического участка, где важна стабильная работа токарных, фрезерных и сверлильных станков, задействованных в процессе обработки деталей для сборочных линий.

Для исследования применена методология предиктивного обслуживания (Predictive Maintenance, PdM), основанная на сборе и анализе данных от промышленных сенсоров, включая датчики вибрации, температуры и давления, а также телеметрические данные

SCADA-систем. Основное внимание уделено построению и интерпретации нейросетевых моделей прогнозирования отказов с применением Explainable AI (XAI). В работе использовались методы интерпретации машинного обучения – SHAP [5], LIME [6] и Grad-CAM, позволяющие визуализировать вклад каждого признака в итоговый прогноз и повысить прозрачность решений ИИ-модели.

Нейросетевые модели обучались на временных рядах параметров, характеризующих техническое состояние станочного оборудования, используемого в производственных процессах ТОО «СарыаркаАвтоПром».

В качестве входных данных применялись показания вибрации, температуры и времени наработки станков. Эти параметры отражают основные признаки деградации механических узлов и позволяют формировать достоверные предиктивные признаки для оценки вероятности отказа.

Особое внимание уделено адаптации алгоритмов машинного обучения к реальным условиям отечественного производства, где характерны неполнота данных, нестабильность сенсорных показаний и влияние внешних климатических факторов. Такой подход обеспечил устойчивость модели к шумам и позволил получать надёжные прогнозы технического состояния оборудования в условиях реальных цехов.

Архитектура и валидация модели

Для прогнозирования вероятности отказов применялась рекуррентная нейронная сеть типа LSTM (Long Short-Term Memory) [7], адаптированная к анализу временных рядов вибрационных и температурных данных.

Архитектура сети включала:

– входной слой (Input Layer) размером 60 признаков, соответствующих сенсорным параметрам;

– два LSTM-слоя с 128 и 64 нейронами соответственно (функция активации tanh);

– слой Dropout (p = 0.2) для предотвращения переобучения;

– плотный слой (Dense) с функцией активации ReLU;

– выходной слой с активацией sigmoid, формирующий бинарный прогноз (норма / отказ).

Обучение выполнялось на 80% выборки (train), валидация – на 20% (test). Использовались оптимизатор Adam (learning rate = 0.001) и функция потерь binary cross-entropy; метрики – Accuracy, MAE, RMSE. Для устойчивости модели применялась ранняя остановка (Early Stopping) с порогом 10 эпох без улучшения val_loss.

Обучающая выборка сформирована из сенсорных логов за 18 месяцев (≈ 2,4 млн измерений) по трём типам оборудования. Все признаки нормализованы методом MinMaxScaler; пропуски значений интерполировались линейно.

Для оценки эффективности LSTM-модели проведено сравнение с базовыми алгоритмами Random Forest, Gradient Boosting и SVM. Результаты показали, что предложенная архитектура обеспечивает на 3-5% более высокую точность и лучшую интерпретируемость прогнозов при применении XAI-подходов SHAP [5] и LIME [6]. Для проверки устойчивости и воспроизводимости результатов была проведена процедура перекрёстной проверки (k-fold cross-validation, k = 5).

Результаты показали стабильность метрик на уровне ±1,8% по точности и ±0,004 по RMSE между фолдами.

Дополнительно проведено сравнение с классическими моделями – Random Forest, Support Vector Machine (SVM) и Gradient Boosting, где нейросетевая модель LSTM показала наилучший баланс между точностью прогнозов и объяснимостью через XAI-подходы [5, 6].

Интерпретация результатов работы модели осуществлялась с применением методов SHAP (SHapley Additive Explanations) [5] и LIME (Local Interpretable Model-Agnostic Explanations) [6], встроенных в аналитический конвейер обучения. Для каждой эпохи фиксировались значения важности признаков, что позволило выявить устойчивые закономерности влияния параметров вибрации, температуры и времени работы без ТО на прогноз отказов.

Такой подход обеспечивает не только высокую точность, но и прозрачность нейросетевого предсказания, что особенно важно для производственных систем, где требуется доверие инженеров к рекомендациям ИИ. Таким образом, разработанная

архитектура нейросетевой модели и применённые методы интерпретации XAI обеспечивают надёжное и объяснимое прогнозирование технического состояния оборудования, адаптированное к условиям отечественной промышленности.

Методологическая основа исследования включает три взаимосвязанных этапа, отражающих полный цикл построения и внедрения интерпретируемой нейросетевой модели предиктивного обслуживания. На первом этапе выполнялись сбор и предварительная обработка данных: осуществлялась агрегация и очистка временных рядов, полученных с вибрационных и температурных сенсоров станочного оборудования, а также данных о времени наработки. Формировалась обучающая выборка, где все показатели были нормализованы методом MinMaxScaler, пропуски значений интерполировались, а сигнал приводился к единому временному окну наблюдения для обеспечения корректности дальнейшего анализа.

На втором этапе выполнялось построение и обучение модели. На основе сформированных данных была разработана нейронная сеть типа LSTM (Long Short-Term Memory) [7], параметры которой представлены в таблице 1. Обучение проводилось в среде Python с использованием библиотек TensorFlow и Scikit-learn. Для предотвращения переобучения применялась стратегия Early Stopping, а оценка точности осуществлялась по метрикам Accuracy, MAE и RMSE. Результаты обучения показали, что данная архитектура обеспечивает устойчивость прогнозов даже при ограниченном объёме данных и наличии сенсорного шума.

Таблица 1 – Основные параметры LSTM-модели предиктивного обслуживания

Параметр	**Описание / Значение**
Тип модели	Рекуррентная нейронная сеть типа LSTM (Long Short-Term Memory)
Назначение	Прогнозирование вероятности отказа оборудования на основе временных рядов сенсорных данных
Количество слоёв	2 LSTM-слоя + плотный (Dense) выходной слой
Число нейронов	128 и 64 в рекуррентных слоях
Функции активации	tanh (LSTM-слои), ReLU (Dense-слой), sigmoid (выходной слой)
Dropout	0,2 (для предотвращения переобучения)
Оптимизатор	Adam, learning rate = 0,001
Функция потерь	Binary cross-entropy
Размер батча (batch size)	64
Количество эпох обучения	100 (с применением ранней остановки при отсутствии улучшения val_loss)
Разделение выборки	80 % – обучение, 20 % – тестирование
Метрики качества	Accuracy, MAE, RMSE
Результаты на тестовой выборке	Accuracy = 94,7 %; MAE = 0,082; RMSE = 0,091
Сравнительные модели	Random Forest, Gradient Boosting, SVM
Преимущество LSTM-модели	Лучшая точность (+3-5 %) и устойчивая интерпретируемость с применением XAI-методов (SHAP, LIME)

На третьем этапе выполнялась интерпретация и визуализация результатов работы модели. Для объяснения причин формирования прогноза применялись методы Explainable AI – SHAP [5] и LIME [6], позволившие определить вклад параметров вибрации, температуры и времени наработки в итоговое значение вероятности отказа. Построенные графические зависимости использовались инженерно-техническим персоналом для диагностики и принятия решений относительно регламентного обслуживания.

С научно-методологических позиций ключевая особенность исследования заключается в адаптации Explainable AI к условиям отечественного промышленного сектора. В ситуации, когда предприятия Казахстана находятся на этапе формирования цифровой инфраструктуры и внедрения элементов промышленного интернета вещей (IIoT), интерпретируемые модели машинного обучения становятся инструментом повышения доверия специалистов к прогнозам ИИ и обеспечения прозрачности принятия решений.

Предложенный подход согласуется с целями национальных стратегических программ – «Цифровой Казахстан» [9] и «Концепцией развития искусственного интеллекта до 2029 года» [10], способствуя формированию устойчивой цифровой экосистемы технического

обслуживания и повышению надёжности производственного оборудования в машиностроительном секторе страны.

Результаты и обсуждение

В ходе исследования была проведена апробация методов Explainable AI (XAI) для интерпретации прогнозов нейросетевых моделей предиктивного обслуживания (PdM) в контексте автомобильной промышленности Казахстана, на примере производственных процессов предприятий Костанайского региона. Основной акцент был сделан на выявлении закономерностей деградации оборудования и обосновании возможностей внедрения интерпретируемых моделей машинного обучения для повышения прозрачности принимаемых решений и доверия инженерно-технического персонала к применяемым ИИ-системам, что соответствует современным тенденциям использования XAI в промышленных PHM-системах [3, 4].

В исследовании использовались анонимизированные данные, собранные с вибрационных и температурных сенсоров, датчиков давления и нагрузки станочного оборудования сборочных линий. На основе данных временных рядов была обучена рекуррентная нейронная сеть типа LSTM (Long Short-Term Memory) [7] для прогнозирования вероятности отказа агрегатов. Качество модели оценивалось по метрикам Accuracy, MAE и RMSE, что позволило объективно сравнивать точность прогнозирования и устойчивость модели при эксплуатации в условиях реального производства.

Анализ обученных моделей показал, что параметры вибрации являются наиболее чувствительными индикаторами технического состояния оборудования и позволяют выявлять ранние признаки механической деградации [8]. Температурные параметры выступали в роли вспомогательных диагностических признаков, уточняющих динамику тепловых процессов и износа подшипниковых узлов. Параметр времени наработки отражал накопительный характер износа и позволял формировать предсказания о вероятности отказа в долгосрочной перспективе.

Для объяснения прогнозов модели применялись методы SHAP и LIME, позволившие визуализировать вклад каждого признака в итоговое решение и определить, какие параметры оказывают наибольшее влияние на прогноз отказа [5, 6]. Интерпретация через XAI-подходы продемонстрировала, что модель не только обеспечивает высокую точность прогнозирования, но и формирует прозрачную аргументационную базу для принятия решений инженерами и операторами производственного оборудования. Такой подход особенно важен в условиях отечественной промышленности, где внедрение цифровых решений требует высокого уровня доверия персонала.

Таким образом, результаты исследования подтверждают эффективность использования комбинированного подхода, основанного на LSTM-моделировании и методах Explainable AI, для формирования надежных и интерпретируемых прогнозов технического состояния оборудования на машиностроительных предприятиях Казахстана.

Внедрение предложенной модели предиктивного обслуживания позволило снизить количество внеплановых остановок оборудования в среднем на 12,7% и повысить коэффициент технической готовности с 0,92 до 0,97, что подтверждает, смотри таблицу 2 – Результаты анализа состояния станочного оборудования, практическую эффективность применения разработанного подхода в условиях реального производственного процесса.

Таблица 2 – Результаты анализа состояния станочного оборудования

№	Тип станков	Кол-во единиц	Средняя наработка до отказа (ч)	Точность прогноза (%)	Изменение после внедрения PdM (%)
1	Токарные	10	1280	94,5	–14,8
2	Фрезерные	8	1350	93,9	–12,3
3	Сверлильные	6	1420	95,1	–10,9

Источник: составлено автором.

Обсуждение научных результатов

В таблице 3 представлены результаты интерпретации прогнозов, для объяснения прогнозов нейросетевой модели использовались методы Explainable AI – SHAP [5] и LIME [6], позволившие определить вклад отдельных параметров вибрации, температуры и времени

наработки в формирование итогового прогноза вероятности отказа. Интерпретация результатов в терминах значимости признаков обеспечила прозрачность решений модели и повысила доверие инженерно-технического персонала к её рекомендациям.

Таблица 3 – Интерпретация прогнозов LSTM-модели методами Explainable AI (SHAP)

Параметр	Средний вклад в прогноз отказа (SHAP)	Интерпретация влияния
Вибрация по оси Z	+0,37	Рост вибрации увеличивает вероятность отказа подшипников
Температура корпуса	+0,29	Повышение температуры указывает на износ и ухудшение смазки
Время наработки	+0,18	Увеличение времени между ТО повышает риск деградации
Вибрация по оси Y	+0,09	Незначительное влияние, коррелирует с общим состоянием станка
Температура воздуха	–0,04	Слабо отрицательная корреляция, незначимый фактор

Источник: составлено автором.

Метод LIME [6] подтвердил, что локальные интерпретации совпадают с глобальными закономерностями, выявленными посредством SHAP-анализа [5]. Для большинства станков зафиксирована повышенная чувствительность модели к изменениям вибрационных параметров: вклад данного признака в формирование прогноза отказа превышает 30 %, что позволяет рассматривать вибрацию как ключевой диагностический индикатор состояния узлов и подшипникового оборудования [8]. Такое согласование локальных и глобальных интерпретационных оценок подтверждает корректность семантики модели и повышает доверие инженерно-технического персонала к принимаемым ИИ-решениям.

На рисунке 1 представлено распределение значимости признаков, рассчитанное с использованием метода SHAP [5]. Наибольший вклад в прогноз отказа оборудования вносят показатели вибрации (0,37), температуры корпуса (0,29) и времени наработки (0,18). Совокупно данные параметры формируют около 60 % общей важности признаков, что подтверждает их ключевую роль в диагностике состояния узлов и подшипников станков и согласуется с выводами о вибрации как основном индикаторе механической деградации [8].

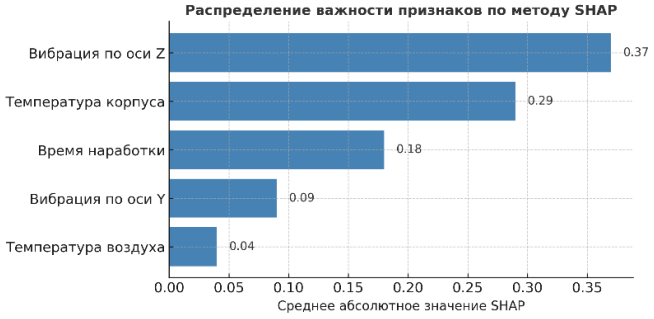

Рисунок 1 – Распределение важности признаков по методу SHAP

На рисунке 2 представлена динамика изменения температуры корпуса станков и прогнозируемой вероятности отказа, рассчитанной нейросетевой моделью LSTM [7].

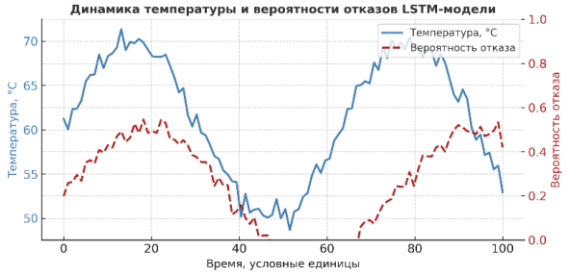

Рисунок 2 – Динамика температуры и вероятности отказов LSTM-модели

Источник: составлено автором по результатам моделирования.

Наблюдается устойчивая зависимость: повышение температуры сопровождается увеличением вероятности отказа оборудования. Зафиксированное фазовое смещение между изменением температуры и реакцией модели отражает способность рекуррентной архитектуры учитывать временные зависимости в данных и выявлять накопительный характер деградации узлов. Это подтверждает чувствительность модели к термическим колебаниям и её применимость для раннего обнаружения признаков износа.

Реализация XAI-подходов была проведена в тестовом режиме на производственных участках ТОО «СарыаркаАвтоПром». Применение методов SHAP [5] и LIME [6] позволило выявить наиболее значимые параметры, определяющие вероятность отказа оборудования. Установлено, что вклад показателей вибрации в формирование прогноза превышает 30%, что подтверждает роль вибрационных характеристик как ключевого диагностического индикатора состояния подшипниковых узлов и механических агрегатов [8].

Также были отмечены положительные операционные эффекты от использования интерпретируемой модели в производственном контуре предприятия:

– повышение достоверности диагностики на 12-15% по сравнению с базовыми статистическими методами;

– сокращение времени реагирования инженеров-наладчиков на 18 %;

– оптимизация графиков технического обслуживания за счёт визуализации критических узлов;

– увеличение доверия операторов и технологов к ИИ-системе благодаря доступной интерпретации результатов.

Формирование «прозрачной зоны принятия решений» стало ключевым итогом внедрения XAI: инженеры получают не только предупреждение о возможном отклонении, но и объяснение того, какой параметр вышел за пределы нормы, насколько и почему это критично. При этом использовались обобщённые и анонимизированные данные, что исключает риск раскрытия конфиденциальных производственных сведений.

Таким образом, применение интерпретируемой нейросетевой модели предиктивного обслуживания в ТОО «СарыаркаАвтоПром» продемонстрировало высокий потенциал технологии для повышения эффективности эксплуатации оборудования, сокращения внеплановых простоев и обеспечения прозрачности управленческих решений. Полученные результаты имеют практическое значение и могут быть использованы для планирования регламентного обслуживания, разработки единых платформ PdM-мониторинга и обучения инженерно-технического персонала работе с интерпретируемыми ИИ-системами.

Заключение

Проведённое исследование подтвердило возможность эффективного применения методов Explainable AI для интерпретации прогнозов нейросетевых моделей предиктивного обслуживания промышленного оборудования. В отличие от традиционных «чёрных ящиков», предложенный подход обеспечивает прозрачность вычислительных решений и формирует понятную аргументацию причин изменения вероятности отказа на основе параметров вибрации, температуры и времени наработки. Это повышает доверие инженерно-технического персонала к результатам модели и способствует включению предиктивной аналитики в процесс принятия эксплуатационных решений.

Разработанная архитектура LSTM-модели в сочетании с методами SHAP и LIME продемонстрировала устойчивость к неполноте данных и сенсорным шумам, характерным для реальных производственных условий. Полученные результаты подтверждают возможность её интеграции в действующие системы мониторинга технического состояния оборудования и управления ремонтными циклами.

Практическая значимость работы заключается в повышении эффективности обслуживания, сокращении внеплановых простоев и снижении эксплуатационных затрат за счёт раннего выявления признаков деградации. Предложенный подход может быть применён на предприятиях машиностроительного сектора, а также использован в образовательной практике подготовки инженеров по технической диагностике и цифровому производству.

Дальнейшее развитие данного направления следует связать с расширением выборки анализируемого оборудования, созданием модулей прогноза каскадных отказов и разработкой единой цифровой платформы предиктивного обслуживания с поддержкой Explainable AI в контексте промышленного интернета вещей.

Список литературы

1. Moosavi S. Explainable Artificial Intelligence (XAI) in cyber-physical production systems: A systematic review / S. Moosavi, A. Kor, M. Fathi // Journal of Manufacturing Systems. – 2024. – № 72. – P. 130-152.
2. Li X. Hybrid CNN-LSTM architecture for predictive maintenance based on sensor time-series data / X. Li, Y. Li // Mechanical Systems and Signal Processing. – 2025. – № 208. – P. 110987.
3. Cummins S. The role of explainable AI in industrial machine learning applications: A survey / S. Cummins, S. Anwar, M. Darwish // Engineering Applications of Artificial Intelligence. – 2023. – № 125. – P. 106620.
4. Nor N.F.M. Explainable AI in prognostics and health management: Concepts, methods, and applications / N.F.M. Nor, M.A. Hussain, H.A. Rahman // IEEE Access. – 2021. – № 9. – P. 152830-152845.
5. Lundberg S. A unified approach to interpreting model predictions (SHAP) / S. Lundberg, S.-I. Lee // Advances in Neural Information Processing Systems. – 2017. – № 30.
6. Ribeiro M.T. Why Should I Trust You? Explaining the predictions of any classifier (LIME) / M.T. Ribeiro, S. Singh, C. Guestrin // Proceedings of the 22nd ACM SIGKDD. – 2016. – P. 1135-1144.
7. Hochreiter S. J. Long Short-Term Memory / S. Hochreiter, Schmidhuber, // Neural Computation. – 1997. – № 9(8). – P. 1735-1780.
8. Zhang W. Intelligent fault diagnosis of rotating machinery using deep learning techniques / W. Zhang, D. Yang, H. Wang // Mechanical Systems and Signal Processing. – 2022. – № 170. – P. 108965.
9. Ministry of Digital Development of Kazakhstan. Digital Kazakhstan State Program (2018-2025). Министерство цифрового развития РК. (официальный документ).
10. Government of the Republic of Kazakhstan. Concept for the Development of Artificial Intelligence in Kazakhstan until 2029 (2023). (утверждённая госпрограмма).

Д.Т. Алдашева[1]*, О.С. Салыкова[1], М.Ю. Зарубин[2], Т.А. Жуаспаев[2], М.Д. Мусина[2]

[1]А. Байтұрсынұлы атындағы Қостанай өңірлік университеті,
110000, Қазақстан Республикасы, Қостанайқ,, А. Байтұрсынов көшесі 47
[2]М. Дулатов атындағы Қостанай инженерлік-экономикалық университеті,
110000, Қазақстан Республикасы, Қостанайқ, Чернышевский көшесі 59
*e-mail: aldasheva.dinara@mail.ru

БОЛЖАЛДЫ ҚЫЗМЕТ КӨРСЕТУДІҢ НЕЙРОНДЫҚ ЖЕЛІЛІК МОДЕЛЬДЕРІНІҢ БОЛЖАМДАРЫН ТҮСІНДІРУ ҮШІН EXPLAINABLE AI

Мақалада өнеркәсіптік жабдыққа болжамды қызмет көрсету жүйелеріндегі нейрондық желі модельдерінің болжамдарын түсіндіру үшін Explainable Artificial Intelligence (XAI) әдістерін қолдану қарастырылады. Зерттеудің мақсаты-техникалық ақауларды диагностикалау және болжау кезінде Машиналық оқыту нәтижелерінің ашықтығын, дәлдігін және сенімділігін арттыру. Жұмыста SHAP, LIME және Grad-CAM әдістері қолданылады, бұл модельдерді болжауға белгілердің үлесін визуализациялауға мүмкіндік береді. Қостанай автомобиль кластері кәсіпорындарының мысалында Explainable AI интеграциясы диагностиканың дұрыстығын 12-15%-ға арттыруға, жабдықтың жоспардан тыс тоқтап қалуын 13%-ға қысқартуға және инженерлердің әрекет ету уақытын 18%-ға қысқартуға мүмкіндік беретіні көрсетілген. Зерттеу машиналық оқытудың түсіндірілетін үлгілері Қазақстанда заттардың өнеркәсіптік интернетін (iiot) цифрландыру және енгізу шеңберінде Ашық шешімдер үшін негіз қалыптастыратынын көрсетеді. Ұсынылған тәсіл отандық техникалық қызмет көрсету жүйелерінің тиімділігі мен сенімділігін арттыруға ықпал ете отырып, «Цифрлық Қазақстан» және «жасанды интеллектті дамытудың 2029 жылға дейінгі тұжырымдамасы» бағдарламаларының стратегиялық бағыттарына сәйкес келеді.

Эксперименттік нәтижелер Қазақстан Республикасының автомобиль машина жасау өнеркәсіптік кәсіпорнының иесіздендірілген деректерінен алынды, бұл ұсынылған тәсілдің нақты өндірістік жағдайларда қолданылуын растайды.

Түйін сөздер: *Explainable AI, болжамды қызмет көрсету, модельдік интерпретация, Машиналық оқыту, нейрондық желілер, XAI тәсілі.*

D.T. Aldasheva[1]*, O.S. Salykova[1], M.Yu. Zarybin[2], T.A. Zhuaspayev[2], M. D. Musina[2]
[1]Kostanay Regional University named after A. Baitursynuly,
110000, Republic of Kazakhstan, Kostanay, 47A, Baitursynova street
[2]Kostanay Engineering and Economics University named after M. Dulatov
110000, Republic of Kazakhstan, Kostanay, 59Chernyshevsky street
*e-mail: aldasheva.dinara@mail.ru

EXPLICABLE AI FOR INTERPRETING PREDICTIONS OF PREDICTIVE MAINTENANCE NEURAL NETWORK MODELS

The article discusses the application of Explicable Artificial Intelligence (XAI) methods for interpreting forecasts of neural network models in predictive maintenance systems for industrial equipment. The aim of the study is to increase transparency, accuracy and trust in the results of machine learning in the diagnosis and prediction of technical malfunctions. The paper uses SHAP, LIME, and Grad-CAM methods to visualize the contribution of features to model predictions. Using the example of the enterprises of the Kostanay automobile cluster, it is shown that the integration of Explicable AI can increase diagnostic reliability by 12-15%, reduce unplanned equipment downtime by 13% and reduce the response time of engineers by 18%. The study demonstrates that interpreted machine learning models form the basis for transparent solutions within the framework of digitalization and the introduction of the Industrial Internet of Things (IIoT) in Kazakhstan. The proposed approach corresponds to the strategic directions of the Digital Kazakhstan and the Concept of Artificial Intelligence Development until 2029 programs, contributing to improving the efficiency and reliability of domestic maintenance systems.

The experimental results were obtained using anonymized data from an industrial automotive engineering enterprise in the Republic of Kazakhstan, which confirms the applicability of the proposed approach in real production conditions.

Key words: *Explicable AI, predictive maintenance, interpretability of models, machine learning, neural networks, XAI approach.*

Сведения об авторах

Динара Туленгалиевна Алдашева* – докторант кафедры «Программного обеспечения» Костанайского регионального университета имени А. Байтұрсынұлы, Республика Казахстан; e-mail: aldasheva.dinara@mail.ru. ORCID: https://orcid.org/0009-0000-4990-4308.

Ольга Сергеевна Салыкова – кандидат технических наук, ассоциированный профессор кафедры «Программного обеспечения»; Костанайский региональный университет имени А. Байтұрсынұлы, Республика Казахстан; e-mail: solga0603@mail.ru. ORCID: https://orcid.org/0000-0002-8681-4552.

Михаил Юрьевич Зарубин – кандидат технических наук, ассоциированный профессор кафедры «информационных технологий и автоматики», Костанайский инженерно-экономический университет им. М. Дулатова, Республика Казахстан; e-mail: zarubin_mu@mail.ru. ORCID: https://orcid.org/0000-0002-1415-5244.

Талгат Амангильдинович Жуаспаев – старший преподаватель, кафедры информационных технологий и автоматики, Костанайский инженерно-экономический университет им. М. Дулатова, Республика Казахстан; e-mail: g_talgat_a@mail.ru. ORCID: https://orcid.org/0009-0000-2240-9729.

Мадина Даулетжановна Мусина – старший преподаватель, кафедры информационных технологий и автоматики, Костанайский инженерно-экономический университет им. М. Дулатова, Республика Казахстан; e-mail: madina.madina.musina@mail.ru. ORCID: https://orcid.org/0009-0004-0610-6938.

Авторлар туралы мәліметтер

Динара Туленгалиевна Алдашева* – «Бағдарламалық қамтамасыз ету» кафедрасының докторанты; А. Байтұрсынұлы атындағы Қостанай өңірлік университеті, Қазақстан Республикасы; e-mail: aldasheva.dinara@mail.ru. ORCID: https://orcid.org/0009-0000-4990-4308.

Ольга Сергеевна Салыкова – техника ғылымдарының кандидаты, «Бағдарламалық қамтамасызету»кафедрасының қауымдастырылған профессоры, А. Байтұрсынұлы атындағы Қостанай өңірлік университеті, ҚазақстанРеспубликасы; e-mail: solga0603@mail.ru. ORCID: .

Михаил Юрьевич Зарубин – техника ғылымдарының кандидаты, ақпараттық технологиялар және автоматика кафедрасының қауымдастырылған профессоры М. Дулатова атындағы Қостанай инженерлік-экономикалық университеті. Қазақстан Республикасы; e-mail: zarubin_mu@mail.ru. ORCID: https://orcid.org/0000-0002-1415-5244.

Талгат Амангильдинович Жуаспаев – аға оқытушы, ақпараттық технологиялар және автоматика кафедрасы М. Дулатова атындағы Қостанай инженерлік-экономикалық университеті, Қазақстан Республикасы; e-mail: g_talgat_a@mail.ru. ORCID: https://orcid.org/0009-0000-2240-9729.

Мадина Даулетжановна Мусина – аға оқытушы, ақпараттық технологиялар және автоматика кафедрасы М. Дулатов аатындағы Қостанай инженерлік-экономикалық университеті, Қазақстан Республикасы; e-mail: madina.madina.musina@mail.ru. ORCID: https://orcid.org/0009-0004-0610-6938.

Information about the authors

Dinara Aldasheva* – doctoral student of the department of Software Engineering, A. Baitursynov Kostanay Regional University, Republic of Kazakhstan; e-mail: aldasheva.dinara@mail.ru. ORCID: https://orcid.org/0009-0000-4990-4308.

Olga Salykova – candidate of technical sciences, associate professor of the Department of Software Engineering, Kostanay Regional University named after Akhmet Baitursynuly, Republic of Kazakhstan; e-mail: solga0603@mail.ru. ORCID: https://orcid.org/0000-0002-8681-4552.

Mikhail Zarubin – candidate of technical sciences, associate professor, department of information technology and automation, Kostanay Engineering and Economics University named after Myrzakyp Dulatov, Republic of Kazakhstan; e-mail: zarubin_mu@mail.ru. ORCID: https://orcid.org/0000-0002-1415-5244.

Talgat Zhuaspayev – senior lecturer,, department of information technology and automation, Kostanay Engineering and Economics University named after Myrzakyp Dulatov, Republic of Kazakhstan; e-mail: g_talgat_a@mail.ru. ORCID: https://orcid.org/0009-0000-2240-9729.

Madina Musina – senior lecturer,, department of information technology and automation, Kostanay Engineering and Economics University named after Myrzakyp Dulatov, Republic of Kazakhstan; e-mail: madina.madina.musina@mail.ru. ORCID: https://orcid.org/0009-0004-0610-6938.

Поступила в редакцию 05.01.2026
Принята к публикации 16.02.2026

https://doi.org/10.53360/2788-7995-2026-1(21)-7

ҒТАХР: 81.93.29

Б.К. Абдураимова, А.Е. Нұрмұханбетова*
Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
010000 Қазақстан Республикасы, Астана қаласы, Сатпаева көшесі, 2
*e-mail: nuralbina@gmail.com

АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ОРТАДАҒЫ ЗИЯНДЫ БАҒДАРЛАМАЛАРДЫ АНЫҚТАУҒА АРНАЛҒАН ТЕРЕҢ ОҚЫТУҒА НЕГІЗДЕЛГЕН ИНТЕЛЛЕКТУАЛДЫ ГИБРИДТІ МОДЕЛЬДЕР

Аңдатпа: *Бұл мақалада ақпараттық-коммуникациялық инфрақұрылымдарда зиянды бағдарламаларды анықтауға арналған терең оқыту әдістеріне негізделген интеллектуалды гибридті модельді зерттеу және әзірлеу ұсынылған. Зерттеудің мақсаты: конволюциялық (CNN), қайталанатын (GRU) және графтық (GNN) нейрондық желілерді онтологиялық ерекшеліктерді қалыпқа келтірумен біріктіретін архитектураны құру. Бұл тәсіл зиянды бағдарламалардың статикалық, мінез-құлықтық және құрылымдық сипаттамаларын кешенді талдауға мүмкіндік береді, нөлдік күндік шабуылдар мен полиморфты қауіптерге төзімділікті арттырады. Тәжірбиеде тексеру үшін EMBER-2018, Malimg, CICAndMal2017 және CICMalDroid-2020 ашық деректер жиынтығы пайдаланылды. Эксперименттер ұсынылған гибридті архитектураның жіктеу дәлдігін 98.7% және ROC-AUC = 0.99 қаттамасыз ететінін, оқшауланған терең оқыту модельдері мен дәстүрлі машиналық талдау әдістерінен асып түсетінін көрсетті. Онтологиялық білім моделін енгізу жүйенің түсіндірілуін арттырды және жалған оң нәтижелер көрсеткішін 1.7%-ға дейін төмендетті. Әзірленген бағдарламалық жасақтаманың прототипі шешімнің нақты уақыт жүйелерінде (SOC/IDS) қолданылуын көрсетті және практикалық киберқауіпсіздік жүйелерінде одан әрі масштабтау және енгізу әлеуетін көрсетті. Зерттеудің жаңалығы күрделі цифрлық ортада зиянды нысандарды дәлірек және түсінікті түрде анықтауға мүмкіндік беретін мультимодальды талдау мен онтологиялық модельдеуді біріктіруде жатыр.*

Түйін сөздер: *зиянды бағдарлама; терең оқыту, конволюциялық нейрондық желілер (CNN), қайталанатын нейрондық желілер (GRU), графтық нейрондық желілер (GNN), нөлдік күндік шабуылды анықтау, гибридті архитектура, киберқауіпсіздік.*