

Нейросетевая модель классификации изображений динамограмм штанговых глубинных насосных установок

А.М. Вульфин¹, К.Ф. Тагирова¹, А.Д. Кириллова¹, А.Е. Сулавко², П.С. Ложников²

¹ Уфимский университет науки и технологий, 450076, Россия, г. Уфа, ул. Заки Валиди, д. 32;

² Омский государственный технический университет, 644050, Российская Федерация, г. Омск, пр-т Мира, д. 11

Аннотация

Работа посвящена проблеме повышения эффективности распознавания характерных признаков комплексных неисправностей на изображениях динамограмм работы погружного насосного оборудования в условиях ограниченной экспертной разметки, выраженного дисбаланса классов и наличия шумов измерений. Создан набор из 20164 практических динамограмм, распределенных по 10 классам. Выполнено систематическое сравнение шести архитектур нейронных сетей (ResNet-18, ResNet-50, EfficientNet-B0, Vision Transformer, ConvNeXt) и четырех способов представления динамограмм в виде изображений (контурное представление, скалограмма непрерывного вейвлет-преобразования, знаковое преобразование расстояний до контура и их комбинирование). В качестве базового использован подход на основе ручного конструирования признаков (цепной код Фримена, аппроксимация геометрическими примитивами) с классификацией методом градиентного бустинга CatBoost. Обучение и оценка моделей проводились по протоколу пятикратной стратифицированной кросс-валидации с компенсацией дисбаланса посредством взвешивания функции потерь. Показано, что представления с высокой долей информативных пикселей (скалограмма, знаковая карта расстояний) обеспечивают более стабильное обучение и превосходят разреженное контурное представление, особенно для архитектур с механизмами внимания. Наилучшее качество достигнуто комбинацией модели ResNet-18 и скалограммы (F1-масго = 0,831 при Ассигасу = 0,899). Научная новизна заключается в количественной оценке влияния типа представления динамограммы на эффективность различных семейств моделей при едином протоколе обучения и выраженном дисбалансе классов. Практическая значимость заключается в повышении точности распознавания изображений динамограмм с комплексными неисправностями, снижении трудоемкости ручной разметки.

Ключевые слова: динамограмма, классификация изображений, сверточные нейронные сети, вейвлет-преобразование, дисбаланс классов.

Цитирование: Вульфин, А.М. Нейросетевая модель классификации изображений динамограмм штанговых глубинных насосных установок / А.М. Вульфин, К.Ф. Тагирова, А.Д. Кириллова, А.Е. Сулавко, П.С. Ложников // Компьютерная оптика. – 2026. – Т. 50, № 4. – 1940. – doi: 10.18287/COJ1940.

Citation: Vulfin AM, Tagirova KF, Kirillova AD, Sulavko AE, Lozhnikov PS. Neural network model for classifying dynamometer cards of sucker-rod pumping units. Computer Optics 2026; 50(4): 1940. doi: 10.18287/COJ1940.

Введение

Штанговые глубинные насосные установки (ШГН) являются наиболее распространенным способом механизированной добычи нефти, обеспечивая эксплуатацию значительной доли малodeбитного фонда скважин. Своевременная диагностика технического состояния погружного оборудования является критически важной задачей, определяющей эффективность и безопасность нефтедобычи. Основным инструментом диагностики является динамограмма – зависимость нагрузки на полированном штоке от его положения за один полный цикл работы ШГН. Динамограмма представляет собой замкнутую параметрическую кривую на плоскости «положение штока – нагрузка», задаваемую парой функций $\{pos(t), load(t)\}_{t=1}^T$, где T – число отсчетов за один полный цикл работы насоса. Замкнутость кривой обусловлена периодичностью возвратно-поступательного движения: начальная и конечная точки цикла совпадают. Форма динамограммы при нормальной работе насоса приближается к параллелограмму: верхняя сторона соответствует ходу плунжера вверх под нагрузкой столба жидкости, нижняя – ходу вниз. Отклонения формы от эталонной параллелограммной конфигурации являются диагностическими признаками различных неисправностей и нарушений режима работы. Традиционно интерпретация динамограмм выполняется экспертами вручную, что требует значительных временных затрат и специализированных знаний. Автоматизация процесса с применением методов машинного обучения представляет собой актуальную научно-практическую задачу.

Целью исследования является определение оптимальной комбинации архитектуры глубокой нейронной сети и метода представления динамограммы в виде изображения в задаче классификации.

Для достижения поставленной цели решались следующие задачи:

- 1) Разработка и реализация методов преобразования временных рядов динамограмм в различные представления, пригодные для обработки сверточными нейронными сетями (CNN).
- 2) Адаптация современных архитектур глубокого обучения, включая трансферное обучение предобученных моделей.
- 3) Проведение систематических экспериментов с применением стратифицированной кросс-валидации.
- 4) Реализация базового метода на основе ручного конструирования признаков (цепной код Фримена, геометрические примитивы) с классификацией методом градиентного бустинга для количественного сопоставления с подходами глубокого обучения.

1. Анализ подходов к классификации изображений динамограмм

Предлагаемая классификация методов распознавания динамограмм, основанная на форме представления динамограмм и семействе моделей, приведена в табл. 1.

Табл. 1. Классификация методов распознавания динамограмм ШГН

Ветвь	Представление данных	Типичные преобразования / признаки	Модели	Достоинства	Ограничения и риски
А	Сигнал (траектория нагрузка-ход)	Нормализация масштаба; производные; контурные описатели (цепные коды); моменты; спектральные (Фурье) и время-частотные (вейвлет) признаки [1 – 3].	SVM/PSO-SVM, RF, kNN;	Интерпретируемость; работоспособность на малых данных; контроль инвариантностей.	Ограниченная переносимость между месторождениями; чувствительность к предобработке; трудоемкость дизайна признаков.
В	2D-изображение карты	Приведение к единому формату; аугментация; перенос обучения; балансировка классов [4, 5].	CNN (ResNet/EfficientNet), ViT/Swin [4, 5]	Автоматическое извлечение признаков; высокая точность при наличии разметки и корректной нормализации.	Риск переобучения; зависимость от протокола разбиения выборки; потребность в больших объемах данных.
С	Последовательность динамограмм во времени	Формирование окон/последовательностей; добавление временных признаков; многомоментные карты [6].	CNN+RNN (LSTM/GRU), временные трансформеры	Улавливание деградации и переходных состояний; повышение полноты по «медленно формирующимся» нарушениям.	Сложность протокола; риск утечки по времени; требования к полноте временных рядов.
Д	Гибридные модальности представления динамограмм	Слияние ручных признаков (Фурье/вейвлет коэффициенты и контур) и визуальных векторов вложений; ансамбли моделей; генеративные аугментации [3, 7, 8].	Ансамбли, мультиисточниковые признаки, генеративные модели	Повышение устойчивости к дисбалансу и вариативности формы; снижение чувствительности к отдельным ошибкам представления	Рост сложности и стоимости; сложнее обеспечить воспроизводимость и сравнение.

Как видно из табл. 1, подходы ветви А, основанные на ручном конструировании признаков из сигнала динамограммы, обеспечивают высокую интерпретируемость и работоспособность при малых объемах данных, однако характеризуются трудоемкостью проектирования признаков и ограниченной переносимостью между различными месторождениями. Подходы ветви В, использующие 2D-изображения и глубокое обучение, позволяют автоматически извлекать признаки, достигающей высокой точности, но требуют значительных объемов размеченных данных и корректного протокола разбиения выборки. Подходы ветви С, учитывающие временную динамику последовательностей динамограмм, перспективны для выявления медленно формирующихся нарушений, однако предъявляют повышенные требования к полноте временных рядов. Гибридные подходы ветви Д комбинируют преимущества нескольких представлений, повышая устойчивость к дисбалансу и вариативности формы, но сопряжены с ростом вычислительной сложности.

Табл. 2 обобщает ключевые исследования по распознаванию динамограмм ШГН с группировкой по использованным наборам данных. В качестве базовых показателей качества приводятся метрики, заявленные авторами работ. При отсутствии раскрытия протокола или состава данных это учитывается при интерпретации сопоставимости результатов.

Анализ табл. 2 показывает, что прямое сравнение результатов различных исследований затруднено по ряду причин: существенные различия в объемах и составе наборов данных (от 500 до 19510 карт), неоднородность номенклатуры классов (от 5 до 24), различающиеся протоколы разбиения выборки, а также неполнота раскрытия метрик по отдельным классам. Большинство работ приводят только общую точность, которая при выраженном

дисбаланс классов не является информативной метрикой для оценки качества распознавания редких неисправностей. Данные обстоятельства подчеркивают необходимость систематического сравнительного исследования с единым протоколом оценки, что и является одной из задач настоящей работы.

Табл. 2. Сводный обзор исследований по классификации динамограмм ШГН с группировкой по наборам данных

Набор данных (объем; классы)	Работа	Ветвь	Модель/ архитектура	Результаты	Примечания
РФ, 500 реальных карт; 8 классов	Mikhailov et al., 2018 [9]	B	CNN (LeNet-подобная), MLP; сравнение с SVM	Точность 0,75-0,80 (CNN) и 0,60 (SVM, оценочно)	Малая выборка; метрики по редким классам не приведены (частичная сопоставимость)
РФ (ПАО «Татнефть»), 3000 динамограмм; отклонения от нормы	Volkov et al., 2021 [10]	D	Ансамбли ML (Random Forest, MLP и др.) и CNN	Точность до 0,90	Ограниченная спецификация классов и протокола (низкая сопоставимость)
Китай (Чанцинское месторождение), 19510 карт с 1226 скважинами; 24 класса (20 одиночных и 4 комбинированных)	Tan et al., 2022 [6]	B/D	ResNet18; мульти-нормализация с формированием нескольких вариантов карты	Точность 0,958; полнота 0,931; сообщается рост полноты по редким классам на 35 %	Реальные данные; отчетливо описаны данные и сравнения (сопоставимость внутри набора данных)
Сербия, несколько тысяч карт; 11 классов; 2 года наблюдений	Martinović et al., 2023 [11]	B	ResNet/DenseNet/ ViT (набор архитектур)	Лучшая модель ResNet34: Точность 0,945; ViT 0,93; DenseNet 0,92	Указано разбиение 70/15/15; метрики по редким классам не акцентированы (частичная сопоставимость)
Китай, 11941 карта; 10-12 классов	Sun et al., 2024 [12]	B	EfficientViT; сравнение с ResNet, MobileNet, EfficientNet, ViT и др.	Точность 0,9573 (EfficientViT)	Детали классов представлены неполно (частичная сопоставимость)
Actual dynagraph datasets (объем не раскрыт); 7 классов	Dong et al., 2024 [5]	B	Swin-T; сравнение с ViT, ResNet50, DenseNet121, LeNet	Точность 0,960; точность по положительному классу 0,954; полнота 0,952; дополнительно приводится F1 0,952	Часть характеристик набора данных не раскрыта; сравнение архитектур информативно (частичная сопоставимость)
Практические данные месторождения (объем не указан); 5 классов	Cheng et al., 2020 [4]	B	AlexNet (перенос обучения) как извлекатель признаков и SVM	Заявлено повышение точности и снижение доли ручных признаков (численные метрики в обзоре не указаны)	Применимо при малых выборках; сопоставимость по метрикам ограничена (низкая сопоставимость)
Синтетические динамограммы, 11000 симулированных карт; 13 классов	Gumerova, 2023 [13]	B	Сравнение MLP, CNN, Siamese CNN	Точность: MLP 0,081; CNN 0,963; Siamese CNN 0,954	Синтетическая природа данных ограничивает переносимость
Последовательность и (3 динамограммы); 5 классов	Danilov, 2019 [14]	C	CNN и LSTM	Без количественных оценок	Показана релевантность учета динамики; сопоставимость по метрикам ограничена
12 типов динамограмм	Sharaf, 2019 [15]	B	Сравнение нескольких CNN (AlexNet, ResNet50, ResNeXt и др.)	Сообщается точность выше 0,98	Отсутствуют характеристики данных и протокола; использование как вспомогательного источника тенденций
—	Wang et al., 2021 [16]	B/D	Формирование входа путем совмещения двух динамограмм в одном изображении	Сообщается об улучшении различимости близких форм. Без количественных оценок	Прием усиления отличий; сопоставимость по метрикам ограничена

Сводные результаты показывают, что прямое сравнение качества классификации изображений динамограмм затруднено из-за разнородных наборов классов, разных протоколов разбиения и неполноты раскрытия данных. Во многих публикациях основной акцент сделан на общей точности, тогда как сведения о качестве по редким классам представлены ограниченно.

Перспективным направлением анализа динамограмм является поэтапное предобучение вложений (embeddings) с последующей адаптацией под целевую номенклатуру классов и формализованными процедурами оценки робастности (вариации предобработки, тестирование чувствительности к шумам, контроль переносимости).

2. Материалы и методы

Экспериментальные исследования проводились на размеченном наборе данных практических динамограмм, полученных с действующих нефтедобывающих скважин. После предварительной фильтрации по комплексному критерию качества (ошибки формы, пропуски данных и т.п.) итоговый набор данных включал 20 164 динамограммы, распределенные по 10 классам режимов работы погружного насосного оборудования (табл. 3).

Табл. 3. Распределение образцов по классам

№	Класс	Количество	Доля, %
1	Нормальная работа	12245	60,7
2	Низкий коэффициент наполнения	4099	20,3
3	АСПО, эмульсия	2236	11,1
4	Высокая посадка плунжера	626	3,1
5	Обрыв штанг	410	2,0
6	Утечки в приемной части насоса	134	0,7
7	Выход плунжера из цилиндра	129	0,6
8	Утечки в нагнетательной части	118	0,6
9	Низкая посадка плунжера	98	0,5
10	Подклинивание плунжера	69	0,3
	Итого	20164	100,0

Как видно из табл. 3, набор данных характеризуется выраженным дисбалансом классов (177 к 1). Каждая динамограмма (рис. 1) представляет собой пару временных рядов $\{pos(t), load(t)\}_{t=1}^T$, где $T=384$ – число отсчетов после передискретизации, $pos(t)$ – положение штока (мм), $load(t)$ – нагрузка на полированном штоке (кН).

Для преобразования исходных временных рядов в формат, пригодный для обработки сверточными нейронными сетями, были реализованы четыре метода представления динамограмм в виде изображений (табл. 4, рис. 2).

1) Контурное представление (Contour) формируется путем растеризации динамограммы на двумерной плоскости. Координаты нормализуются в диапазон $[0, S-1]$, где $S=128$ – размер изображения.

Последовательные точки соединяются отрезками с использованием алгоритма Брезенхема, формируя бинарное изображение $I \in \{0,1\}^{S \times S}$. Существенным недостатком данного представления является его разреженность: доля ненулевых пикселей составляет лишь 2,93 %.

2) Непрерывное вейвлет-преобразование (CWT) обеспечивает время-частотное представление сигнала. Для сигнала нагрузки $s(t) = load(t)$ вычисляется скалограмма (1):

$$W_s(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \cdot \psi^* \left(\frac{t-b}{a} \right) dt, \quad (1)$$

где $\psi(t) = e^{i\omega_0 t} e^{-t^2/2}$ – комплексный вейвлет Морле с центральной частотой $\omega_0 = 6$, $a \in [1, S]$ – параметр масштаба, $b \in [0, T]$ – параметр временного сдвига. Модуль коэффициентов $|W_s(a, b)|$ формирует плотное изображение размером $S \times S$ с долей ненулевых пикселей 99,99 %.

3) Преобразование расстояний с учетом принадлежности точки объекту (signed distance transform, SDT), формирующее знаковую карту расстояний до ближайшей точки контура. Алгоритм построения изображения включает следующие этапы:

- построение бинарного контура C динамограммы методом Брезенхема;
- заполнение внутренней области контура C_{filled} ;
- вычисление евклидова расстояния до контура для внешних (2) и внутренних (3) точек:

$$D_{out}(x, y) = \min_{(x', y') \in C} \sqrt{(x - x')^2 + (y - y')^2}, \quad (2)$$

$$D_{in}(x, y) = \min_{(x', y') \notin C_{filled}} \sqrt{(x - x')^2 + (y - y')^2}. \quad (3)$$

- формирование знакового поля: $SDT = D_{out} - D_{in}$, нормализованного в диапазон $[-1, 1]$.

Результирующее изображение является полностью плотным и сохраняет информацию о геометрии контура динамограммы в неявном виде.

4) Мультимодальная нейронная сеть на основе позднего слияния признаков CWT и SDT (CWT+SDT). Представления CWT и SDT характеризуют динамограмму в принципиально различных пространствах: CWT является частотно-временным разложением сигнала нагрузки, тогда как SDT отражает пространственно-геометрическую структуру контура на плоскости «положение-нагрузка». Для совместного использования

CWT и SDT реализована мультимодальная нейросетевая архитектура с отдельными ветвями обработки. Каждая ветвь представляет собой независимый экземпляр базовой архитектуры (EfficientNet-B0), принимающий на вход одноканальное изображение соответствующего представления. Обе ветви инициализируются одинаковыми предобученными весами, но дообучаются независимо. Интеграция информации осуществляется на уровне высокоуровневых векторов признаков путем их конкатенации и последующей обработки общим полносвязным классификатором (4):

$$\mathbf{f}_{CWT} = \text{Encoder}_{CWT}(I_{CWT}), \quad \mathbf{f}_{SDT} = \text{Encoder}_{SDT}(I_{SDT}), \quad \hat{y} = \text{softmax}(W[\mathbf{f}_{CWT}; \mathbf{f}_{SDT}] + \mathbf{b}), \quad (4)$$

где $\mathbf{f}_{CWT} \in \mathbf{R}^d$ и $\mathbf{f}_{SDT} \in \mathbf{R}^d$ – векторы признаков, извлеченные из глобального пулинга соответствующих ветвей; $[\cdot; \cdot]$ – операция конкатенации; $W \in \mathbf{R}^{C \times 2d}$, $\mathbf{b} \in \mathbf{R}^C$ – параметры классификатора; C – число классов.

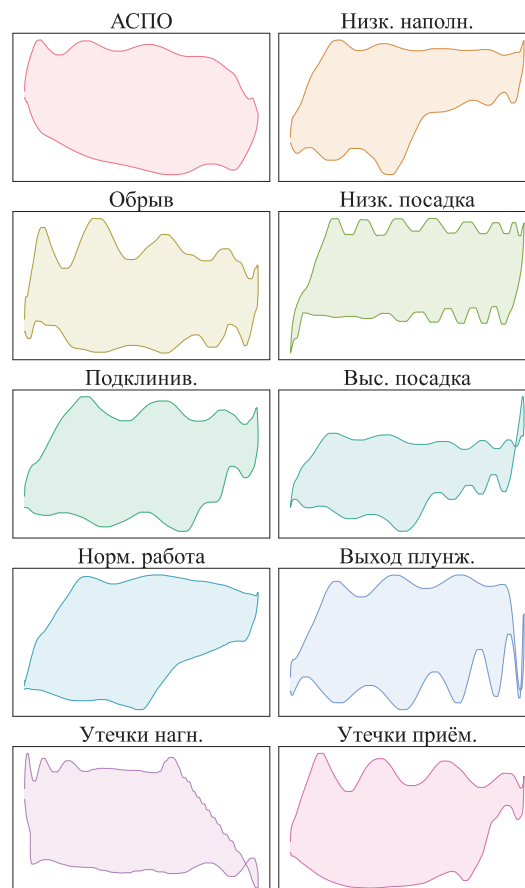


Рис. 1. Примеры динамограмм по классам состояний погружного насосного оборудования

Такой подход обеспечивает отдельное извлечение частотно-временных и геометрических признаков в пространствах, адекватных природе каждого представления, и их интеграцию только на уровне абстрактных дискриминативных признаков, где различие физических пространств уже не играет роли.

Следует отметить, что CWT и SDT являются функционально зависимыми представлениями одного и того же объекта: из SDT может быть реконструирован контур, из которого, в свою очередь, восстанавливается исходный сигнал и его CWT. Таким образом, мультимодальное объединение не увеличивает информационное содержание входных данных, а лишь изменяет форму их представления, потенциально упрощая задачу извлечения дискриминативных признаков для нейронной сети за счет явного предоставления двух взаимодополняющих представлений динамограммы (табл. 4, рис. 2).

Далее представления, в которых доля ненулевых (информативных) пикселей близка к 100 % (CWT, SDT, CWT+SDT), будем называть плотными, в отличие от разреженного контурного представления с долей ненулевых пикселей около 3 %.

Другой вариант формирования признаков на основе анализа изображений динамограмм представляет интерес с точки зрения интерпретируемости и позволяет оценить информативность различных типов геометрических признаков. Рассмотрены два метода.

1) Цепной код Фримена (Freeman chain code) – классический метод кодирования контуров, широко применяемый в задачах распознавания образов. Код описывает последовательность направлений движения вдоль границы объекта, обеспечивая компактное и инвариантное к переносу представление формы.

Табл. 4. Сравнительные характеристики методов представления динамограмм в виде изображений

Метод	Размерность	Плотность (доля ненулевых пикселей), %	Тип информации	Время генерации, мс
Contour	128×128×1	2,93	Геометрические признаки	0,5
CWT	128×128×1	99,99	Время-частота	15,0
SDT	128×128×1	100,00	Геометрические признаки и расстояние до контура	3,0
CWT+SDT	2 × 128×128×1	100,00	Частотно-временная и геометрическая; раздельная обработка, интеграция на уровне признаков	18,0

Для последовательности точек контура $\{(x_i, y_i)\}_{i=0}^{N-1}$ вычисляется направление перехода от точки i к точке $i + 1$. В 8-связной схеме используются восемь направлений, соответствующих углам $0^\circ, 45^\circ, 90^\circ, \dots, 315^\circ$ (5):

$$d_i = \left\lfloor \frac{\text{atan2}(\Delta y_i, \Delta x_i) \cdot 4}{\pi} + 0,5 \right\rfloor \bmod 8, \quad (5)$$

где $\Delta x_i = x_{i+1} - x_i$, $\Delta y_i = y_{i+1} - y_i$.

Из последовательности кодов $\{d_i\}_{i=0}^{N-2}$ формируется нормализованная гистограмма (6):

$$h_k = \frac{1}{N-1} \sum_{i=0}^{N-2} 1[d_i = k], \quad k \in \{0, 1, \dots, 7\}. \quad (6)$$

Гистограмма инвариантна к масштабу контура и содержит информацию о распределении направлений движения. Для динамограмм характерно преобладание горизонтальных ($k = 0, 4$) и вертикальных ($k = 2, 6$) направлений, отражающих прямоугольную структуру контура.

2) Для извлечения высокоуровневых геометрических характеристик (табл. 5) выполняется аппроксимация контура динамограммы четырьмя типами примитивов. Для каждого примитива решается задача нелинейной оптимизации, минимизирующая сумму расстояний от точек контура до границы примитива.

Табл. 5. Структура признаков геометрических примитивов

Примитив	Параметры	Семантика	Используемые признаки, шт.
Прямоугольник	$x_0, y_0, w, h, \theta, \varepsilon$	Центр, размеры, поворот, ошибка	6
Параллелограмм	$x_c, y_c, a, b, c, d, \varepsilon$	Центр, образующие векторы, ошибка	7
Параллелограмм с вырезом	$v_{1x}, v_{1y}, v_{2x}, v_{2y}, \alpha, \beta, h, t_x, t_y, \varepsilon$	Образующие, параметры выреза, смещение, ошибка	10
Параллелограмм	$x_0, y_0, a, b, \phi, \varepsilon$	Центр, полуоси, ориентация, ошибка	6

Для каждого из четырех типов примитивов вектор признаков формируется из всех оптимизируемых параметров и значения ошибки аппроксимации ε . Ошибка аппроксимации ε при этом является важным дискриминативным признаком, характеризующим степень соответствия формы динамограммы данному типу примитива. Итого вектор геометрических признаков содержит 29 компонентов; с добавлением нормализованной площади фигуры итоговый вектор составляет 30 признаков.

$$d_i(\theta) = \min_{(x', y') \in \Delta P(\theta)} \sqrt{(x_i - x')^2 + (y_i - y')^2}. \quad (7)$$

Задача оптимизации формулируется как минимизация суммы расстояний по всем N точкам контура (8):

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^N d_i(\theta). \quad (8)$$

Для прямоугольника и эллипса оптимизация выполняется методом SLSQP с ограничениями неотрицательности размеров; для параллелограммов – методом Левенберга–Марквардта. Начальное приближение получается из описанного прямоугольника (для прямоугольника и параллелограмма) или прямой подгонки эллипса (для эллипса):

1) Прямоугольник. Модель параметризуется центром (x_0, y_0) , размерами (w, h) и углом поворота θ (9):

$$\min_{x_0, y_0, w, h, \theta} \sum_{i=1}^N d_i^{rect}, \quad (9)$$

где d_i^{rect} – расстояние от точки контура до границы прямоугольника в локальной системе координат. Оптимизация выполняется методом SLSQP с ограничениями $w, h > 0$.

2) Параллелограмм. Задается центром (x_c, y_c) и двумя образующими векторами $\mathbf{u} = (a, b)$, $\mathbf{v} = (c, d)$ (10):

$$V_1 = C + \mathbf{u} + \mathbf{v}, \quad V_2 = C + \mathbf{u} - \mathbf{v}, \quad V_3 = C - \mathbf{u} - \mathbf{v}, \quad V_4 = C - \mathbf{u} + \mathbf{v}. \quad (10)$$

Ошибка вычисляется как сумма минимальных расстояний до четырех сторон. Используется метод наименьших квадратов (Levenberg-Marquardt).

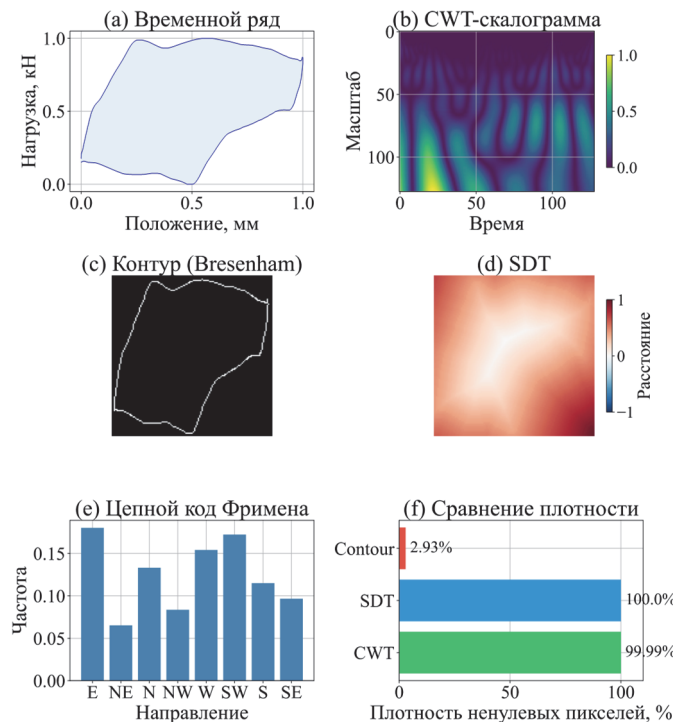


Рис. 2. Используемые методы представления динамограмм в виде изображений и сравнение плотности представления. Для каждого примитива решается задача нелинейной оптимизации. Пусть контур динамограммы задан упорядоченным множеством из N точек $\{(x_i, y_i)\}_{i=1}^N$, полученных rasterизацией алгоритмом Брезенхема. Для заданного примитива с вектором параметров θ , граница которого определяет множество точек $\Delta P(\theta)$, расстояние от i -й точки контура до границы примитива вычисляется как евклидово расстояние до ближайшей точки границы (7)

3) Параллелограмм с вырезом. Специализированный примитив для динамограмм с характерной выемкой в нижней части. Задается 9 параметрами (11):

$$\text{params} = (v_{1x}, v_{1y}, v_{2x}, v_{2y}, \alpha, \beta, h, t_x, t_y), \quad (11)$$

где v_1, v_2 – образующие векторы, $\alpha, \beta \in [0, 1]$ – относительные позиции выреза на сторонах, h – глубина выреза, (t_x, t_y) – смещение. Примитив формирует шестиугольник ABCDEF с характерной конфигурацией. Ограничение $\beta \geq \alpha$ обеспечивает корректную геометрию выреза.

4) Эллипс. Модель параметризуется центром (x_0, y_0) , полуосями (a, b) и углом ориентации ϕ . Начальное приближение получается методом прямой подгонки эллипса, затем уточняется методом SLSQP.

2.1. Архитектуры нейронных сетей для классификации изображений динамограмм

В рамках исследования была проведена сравнительная оценка шести архитектур нейронных сетей, относящихся к различным семействам.

2.1.1. ResNet-18 и ResNet-50

Остаточные сети (Residual Networks) (табл. 6) используют блоки с остаточными связями (skip connections), позволяющие эффективно обучать более глубокие модели. Базовый блок ResNet реализует преобразование (12):

$$y = F(x, \{W_i\}) + x, \quad (12)$$

где $F(x, \{W_i\})$ – остаточное отображение, реализуемое двумя сверточными слоями 3×3 с пакетной нормализацией.

Табл. 6. Сравнение параметров ResNet-18 и ResNet-50

Параметр	ResNet-18	ResNet-50
Число параметров, млн	11,7	25,6
Число слоев	18	50
Предобучение	ImageNet	ImageNet

Для адаптации предобученных на ImageNet моделей к одноканальным входным изображениям применялось расширение каналов путем дублирования. Первые два блока (Layer1, Layer2) замораживались, последние два блока (Layer3, Layer4) и полносвязный классификатор дообучались.

2.1.2. EfficientNet-B0

Архитектура EfficientNet основана на принципе составного масштабирования (compound scaling), где глубина, ширина и разрешение сети масштабируются совместно: $d = \alpha^\phi$, $w = \beta^\phi$, $r = \gamma^\phi$ при ограничении $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$ и $\phi = 1$ для варианта B0.

Основным блоком является MBConv (Mobile Inverted Bottleneck Convolution) с механизмом Squeeze-and-Excitation. Модель содержит 5,3 млн параметров.

2.1.3. Vision Transformer (ViT-Tiny)

Vision Transformer применяет архитектуру Transformer к изображениям путем разбиения на непересекающиеся «патчи» размером 16×16 пикселей. Каждый «патч» проецируется в пространство вложений (embeddings) размерности $d = 192$. Механизм многоголового самовнимания (Multi-Head Self-Attention) определяется выражением (13), в котором матрицы запросов Q , ключей K и значений V являются линейными проекциями одного и того же набора входных вложений Z : $Q = ZW^Q$, $K = ZW^K$, $V = ZW^V$, что отличает механизм самовнимания от перекрестного внимания, где Q и $\{K, V\}$ формируются из различных источников:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (13)$$

Для входного изображения 224×224 формируется 196 патчей, к которым добавляется специальный токен [CLS] для классификации. Модель ViT-Tiny содержит 12 слоев Transformer и 5,7 млн параметров.

2.1.4. ConvNeXt-Tiny

ConvNeXt представляет собой модернизированную сверточную архитектуру, заимствующую ряд идей из Transformer: глубокие и широкие свертки 7×7 , нормализация слоев (LayerNorm), активационная функция GELU и стохастическая глубина распространения сигнала (DropPath). Модель содержит 28,6 млн параметров.

2.2. Подход для классификации изображений динамограмм с помощью ручного конструирования признаков

В качестве базового метода, не использующего глубокое обучение, был реализован подход на основе ручного конструирования признаков с последующей классификацией алгоритмом градиентного бустинга CatBoost [17]. Данный подход представляет интерес с точки зрения интерпретируемости и позволяет оценить информативность различных типов геометрических признаков.

Были протестированы три варианта формирования вектора признаков: нормализованная гистограмма цепного кода Фримена (8 признаков, соответствующих 8 направлениям, формула (6)); параметры аппроксимации геометрическими примитивами (30 признаков, табл. 5); комбинация гистограммы и геометрических признаков (38 признаков).

2.3. Протокол обучения моделей классификации изображений динамограмм и оценки их эффективности

Все эксперименты проводились с использованием 5-кратной стратифицированной кросс-валидации. Гиперпараметры обучения были унифицированы для всех моделей (табл. 7).

Табл. 7. Гиперпараметры обучения моделей классификации изображений динамограмм

Гиперпараметры	Значение
Размер мини-пакета	32
Максимальное число эпох	100
Начальная скорость обучения	10^{-3}
Регуляризация весов	10^{-4}
Планировщик изменения скорости обучения	Косинусное затухание с прогревом (5 эпох)
Ранняя остановка	20 эпох
Смешанная точность	FP16

Для компенсации дисбаланса классов применялось взвешивание функции потерь (14):

$$\mathcal{L} = -\sum_{c=1}^C w_c \cdot y_c \cdot \log(\hat{y}_c), \quad (14)$$

где $w_c = 1/\sqrt{N_c}$ – вес класса c , обратно пропорциональный корню из числа образцов N_c .

В качестве основной метрики качества использовалась макроусредненная F1-мера (15):

$$\text{F1-макро} = \frac{1}{C} \sum_{c=1}^C \frac{2 \cdot P_c \cdot R_c}{P_c + R_c}, \quad (15)$$

где P_c и R_c – точность и полнота для класса c соответственно.

Все эксперименты проводились на рабочей станции со следующей конфигурацией: процессор AMD Ryzen 7 5700X (8 ядер/16 потоков), 128 ГБ оперативной памяти, GPU NVIDIA GeForce RTX 3060 (12 ГБ VRAM). Обучение проводилось в режиме смешанной точности (mixed precision fp16) для ускорения вычислений и снижения потребления видеопамати. Программная реализация выполнена на языке Python 3.10 с использованием фреймворка PyTorch 2.8.0 (CUDA 12.8) и надстройки PyTorch Lightning 2.5.5. Предобученные модели ResNet-18, ResNet-50, EfficientNet-B0, ViT-Tiny и ConvNeXt-Tiny загружались из библиотеки timm 1.0.24 (PyTorch Image Models) с весами, обученными на ImageNet. Вычисление непрерывного вейвлет-преобразования (CWT-скалограмм на базе вейвлета Морле) выполнялось с использованием библиотеки PyWavelets 1.8.0. Классификация типа аппроксимирующего геометрического примитива методом градиентного бустинга реализована с помощью библиотеки CatBoost 1.2.8. Воспроизводимость экспериментов обеспечивалась фиксацией случайного начального значения (random seed = 42) для всех компонентов конвейера, включая детерминистические алгоритмы CUDA (torch.backends.cudnn.deterministic = True).

3. Результаты экспериментов

Результаты экспериментов для всех комбинаций архитектур и представлений приведены в табл. 8. Модели ранжированы по значению F1-макро.

Табл. 8. Результаты классификации (5-кратная кросс-валидация)

№	Модель	Представление	F1-макро, %	Accurasy, %	Эпохи	Параметры, млн	Время, мин.
1	ResNet-18	CWT	83,07 ± 1,29	89,91 ± 0,50	35	11,7	35
2	ResNet-50	CWT	82,07 ± 0,30	89,16 ± 0,67	39	25,6	58
3	ResNet-18	SDT	81,63 ± 0,24	89,78 ± 1,11	39	11,7	40
4	EfficientNet-B0	CWT	80,31 ± 1,18	87,67 ± 0,58	38	5,3	40
5	ResNet-18	Contour	79,98 ± 1,27	88,11 ± 1,12	28	11,7	30
6	EfficientNet-B0	CWT+SDT	80,11 ± 1,76	87,68 ± 1,32	29	10,8	65
7	EfficientNet-B0	SDT	78,16 ± 0,96	86,51 ± 0,69	51	5,3	50
8	ResNet-50	Contour	76,67 ± 1,42	85,36 ± 1,63	26	25,6	40
9	EfficientNet-B0	Contour	75,29 ± 1,63	84,90 ± 0,78	44	5,3	45
10	CatBoost	Combined	68,04 ± 1,67	82,52 ± 0,96	245	–	8
11	CatBoost	Geom	63,68 ± 2,33	81,20 ± 0,24	216	–	6
12	CatBoost	ChainCode	48,83 ± 3,00	70,05 ± 0,33	155	–	5
13	ViT-Tiny-v3	Contour	36,86 ± 3,44	71,04 ± 2,55	25	5,7	100
14	ConvNeXt-Tiny	Contour	34,51 ± 33,00	70,45 ± 11,90	11	28,6	90

Наилучший результат достигнут комбинацией ResNet-18 с представлением CWT. Архитектура ResNet-50 с CWT уступает ResNet-18, что подтверждает гипотезу о переобучении более глубоких сетей на ограниченном объеме данных.

Мультимодальная модель CWT+SDT с отдельными ветвями обработки показала результат F1-макро = 80,11%, что ниже одномодальной модели EfficientNet-B0 с представлением CWT (80,31%). Данный результат объясняется функциональной зависимостью между представлениями CWT и SDT: оба являются производными одного и того же сигнала, и мультимодальная архитектура не получает дополнительной информации, но при этом удваивает число обучаемых параметров энкодера, что при ограниченном объеме выборки приводит к переобучению. Кроме того, CWT-представление уже неявно содержит геометрическую информацию о форме контура через структуру частотно-временного спектра, что делает SDT-ветвь в значительной степени избыточной.

Анализ влияния метода представления динамограммы в виде изображения проводился для двух архитектур: ResNet-18 и EfficientNet-B0. Представление CWT обеспечивает наибольший прирост качества для всех архитектур. Наиболее значительное улучшение наблюдается для ResNet-50, что свидетельствует о важности плотного представления для более глубоких сетей. Тем не менее даже с представлением CWT архитектура ResNet-50 уступает ResNet-18, что подтверждает наличие переобучения при увеличении глубины сети на данном существенно ограниченном объеме выборки.

При фиксированном представлении Contour сравнение архитектур показало следующие результаты (табл. 9).

Архитектуры, основанные на механизме внимания (ViT, ConvNeXt), показали неудовлетворительные результаты на контурном представлении. Причина заключается в высокой разреженности входных данных: при 2,93% ненулевых пикселей большинство патчей 16×16 оказываются полностью пустыми, что приводит к неинформативным ключам и значениям в механизме внимания.

Стоит отметить, что увеличение глубины сети (переход от ResNet-18 к ResNet-50) привело к снижению качества классификации, что скорее всего объясняется переобучением более емкой модели на относительно небольшом наборе доступных размеченных данных динамограмм.

Табл. 9. Сравнение архитектур (представление Contour)

Архитектура	F1-масро, %	Параметры, млн	Время обучения для одного прохода, мин.
ResNet-18	79,98	11,7	30
ResNet-50	76,67	25,6	40
EfficientNet-B0	75,29	5,3	45
ViT-Tiny	33,94	5,7	100
ConvNeXt-Tiny	34,51	28,6	90

Детальный анализ качества классификации для отдельных классов позволяет выявить «сложные» категории (табл. 10).

Табл. 10. F1-мера по классам для лучших моделей

Класс	ResNet-18 CWT	EfficientNet CWT	ResNet-18 Contour
Нормальная работа	78,9 %	73,0 %	74,2 %
Низкий коэффициент заполнения	76,7 %	69,3 %	65,7 %
АСПО, эмульсия	82,9 %	79,6 %	78,8 %
Высокая посадка	87,9 %	85,2 %	86,5 %
Обрыв штанг	92,7 %	91,2 %	91,6 %
Утечки приемной части	75,9 %	80,7 %	77,1 %
Выход плунжера	85,5 %	83,1 %	81,8 %
Утечки нагнет. части	76,9 %	71,9 %	74,5 %
Низкая посадка	91,4 %	86,5 %	84,6 %
Подклинивание	83,2 %	82,7 %	84,9 %

Наибольшие значения F1-меры (более 90 %) достигаются для классов «Обрыв штанг» и «Низкая посадка», несмотря на их малую представленность в выборке. Это объясняется характерной геометрией динамограмм данных классов, существенно отличающейся от остальных.

Наименьшие значения F1-меры наблюдаются для классов, связанных с утечками (71 – 80 %), что обусловлено как малым числом образцов, так и схожестью формы динамограмм при различных типах утечек.

Подход на основе градиентного бустинга (CatBoost) признаками, конструируемыми вручную, существенно уступает методам глубокого обучения. Вместе с тем интерпретируемость признакового подхода может быть востребована в задачах объяснимого искусственного интеллекта.

4. Анализ и обсуждение результатов

Ключевым фактором, определяющим качество классификации динамограмм, является доля информативных (ненулевых) пикселей во входном представлении. Разреженное контурное изображение с большим количеством нулевых пикселей создает неблагоприятные условия для архитектур, основанных на механизме внимания, где информативность зависит от содержательности патчей.

Представление CWT обеспечивает информативность каждого патча, что объясняет успешную работу как сверточных, так и потенциально трансформерных архитектур при его использовании.

Предобученные на ImageNet нейросетевые модели ResNet-18 и EfficientNet-B0 продемонстрировали высокую эффективность при адаптации к специализированной задаче, несмотря на существенные различия между естественными изображениями и скалограммами. Это свидетельствует об универсальности низкоуровневых признаков (границы, текстуры), извлекаемых первыми слоями сверточных сетей.

Увеличение емкости модели (переход от ResNet-18 к ResNet-50) привело к снижению качества даже при использовании наилучшего представления CWT. Данный результат согласуется с теоретическими представлениями: при объеме обучающей выборки 16131 образец и 25,6 млн параметров нейронной сети отношение «образцы/параметры» составляет лишь 0,63, что недостаточно для эффективного обучения. Примечательно, что CWT существенно сокращает разрыв между архитектурами, но не устраняет его полностью.

Отсутствие выигрыша от мультимодального объединения CWT и SDT подтверждает, что при функциональной зависимости между представлениями увеличение числа входных модальностей не компенсирует роста числа параметров модели.

Сопоставление полученных результатов с исследованиями, представленными в табл. 2, позволяет сделать следующие наблюдения. Достигнутое значение F1-масро = 83,07 % при 10 классах и выраженном дисбалансе не может быть напрямую сопоставлено с показателями точности, заявленными в большинстве работ (0,90 – 0,96), поскольку эти метрики имеют принципиально различную чувствительность к миноритарным классам. Тем не менее ряд содержательных сравнений возможен. Результаты подтверждают выводы работы [6] о значимости метода формирования входного представления: авторы отмечали рост полноты по редким классам на 35 % при использовании мультиномализации, что согласуется с обнаруженным в настоящем исследовании приростом F1-масро на 3 – 5 % при переходе от контурного представления к скалограмме CWT. Результаты [11] о сопоставимой

эффективности ResNet и ViT на их наборе данных не нашли подтверждения в наших экспериментах: ViT-Tiny показал крайне низкие результаты ($F1\text{-macro} = 36,86\%$) на контурном представлении. Это расхождение, вероятно, объясняется различием в плотности входных представлений: авторы [11] использовали стандартные изображения динамограмм, тогда как в нашем исследовании разреженное контурное представление с 2,93 % ненулевых пикселей оказывается неблагоприятным для архитектур на основе внимания. Наблюдение о переобучении при увеличении глубины модели (ResNet-18 превосходит ResNet-50) согласуется с общей тенденцией, отмечаемой при обучении на ограниченных специализированных наборах данных, и подтверждает целесообразность использования более компактных архитектур при объемах обучающей выборки порядка 16 – 20 тысяч образцов.

Существенным ограничением сравнения с существующими исследованиями остается отсутствие в большинстве работ детализированных метрик по отдельным классам и единого протокола оценки, что дополнительно подчеркивает значимость проведенного систематического сравнения при едином экспериментальном протоколе.

Проведенное исследование имеет ряд ограничений:

1. Эксперименты выполнены на данных одного месторождения; обобщаемость результатов на другие условия эксплуатации требует дополнительной проверки.
2. Не исследовано влияние методов аугментации данных, которые могут улучшить результаты для миноритарных классов. Одна аугментация данных может вносить искажения в форму динамограммы и нарушать физическую интерпретируемость протекающих в погружном оборудовании процессов.

Заключение

В настоящей работе проведено систематическое сравнительное исследование методов глубокого обучения для задачи классификации изображений динамограмм штанговых глубинных насосных установок. Основные результаты:

1. Наилучшее качество классификации ($F1\text{-macro} = 83,07\%$) достигается при использовании архитектуры ResNet-18 с представлением данных на основе непрерывного вейвлет-преобразования.
2. Метод представления динамограммы в виде изображения оказывает существенное влияние на качество классификации: переход от контурного изображения к скалограмме обеспечивает прирост $F1\text{-macro}$ на 3 – 5 % для всех протестированных нейросетевых архитектур.
3. Архитектуры, основанные на механизме внимания (Vision Transformer, ConvNeXt), непригодны для работы с разреженными контурными изображениями в случае ограниченных по размеру выборок и наличии дисбаланса классов, но потенциально могут быть эффективны с плотными представлениями.
4. Методы на основе ручного конструирования признаков (CatBoost) значительно уступают глубокому обучению, однако могут быть востребованы в задачах с требованиями к интерпретируемости принимаемых диагностических решений.
5. Увеличение глубины нейронной сети приводит к снижению качества при ограниченном объеме данных, хотя плотное представление существенно сокращает этот разрыв по сравнению с контурным.

Полученные результаты создают основу для разработки промышленной системы автоматической диагностики насосного оборудования. Дальнейшие исследования будут направлены на изучение методов аугментации данных, ансамблевых подходов и механизмов объяснимости принимаемых решений.

Благодарности

Работа выполнена в ОмГТУ в рамках государственного задания Минобрнауки России на 2023–2025 годы № FSGF-2023-0004.

References

- [1] Li K, Gao X-W, Yang W-B, Dai Y-L, Tian Z-D. Multiple fault diagnosis of down-hole conditions of sucker-rod pumping wells based on Freeman chain code and DCA. *Pet. Sci* 2013; 10: 347-360. DOI: 10.1007/s12182-013-0283-4.
- [2] Li K, Gao X-W, Tian Z, Qiu Z. Using the curve moment and the PSO-SVM method to diagnose downhole conditions of a sucker rod pumping unit. *Pet. Sci* 2013; 10: 73-80. DOI: 10.1007/s12182-013-0252-y.
- [3] Abdalla R, El Ela MA, El-Banbi A. Identification of Downhole Conditions in Sucker Rod Pumped Wells Using Deep Neural Networks and Genetic Algorithms. *SPE Prod. Oper* 2020; 35: 435-447. DOI: 10.2118/200494-PA.
- [4] Cheng H, Yu H, Zeng P, Osipov E, Li S, Vyatkin V. Automatic Recognition of Sucker-Rod Pumping System Working Conditions Using Dynamometer Cards with Transfer Learning and SVM. *Sensors* 2020; 20: 5659. DOI: 10.3390/s20195659.
- [5] Dong G, Li W, Dong Z, Wang C, Qian S, Zhang T, Ma X, Zou L.; Lin, K.; Liu, Z. Enhancing Dynagraph Card Classification in Pumping Systems Using Transfer Learning and the Swin Transformer Model. *Appl. Sci.* 2024, 14, 1657. <https://doi.org/10.3390/app14041657>.
- [6] Tan C, Chen P, Feng Z, Ai X, Lu M, Zhou Q, Feng G. Multi-Scale Normalization Method Combined With a Deep CNN Diagnosis Model of Dynamometer Card in SRP Well. *Front. Earth Sci* 2022; 10: 852633. DOI: 10.3389/feart.2022.852633.
- [7] Yuan C, Wu W, Li X. Dynamometer card generation for pumping units based on CNN and electrical parameters. *Sci. Rep* 2024; 14: 18657. DOI: 10.1038/s41598-024-69516-y.

- [8] Tan X, Wang S, Wu H, Chen F. Enhancing pumping unit diagnosis with similarity splicing data augmentation and wavelet denoising. Sci. Rep 2025; 15: 37045. DOI: 10.1038/s41598-025-20819-8.
- [9] Mihajlov AG, Shubin SS, Alferov AV, Imashev RN, Yamaliev VU. Improvement of efficiency of diagnostics of rod pumps with use of deep neural networks. Neftyanoye khozyaystvo 2018; 9: 122-126. DOI: 10.24887/0028-2448-2018-9-122-126.
- [10] Volkov MG, Silnov DV, Topolnikov AS, Latypov BM, Katermin AV, Enikeev RM. Automated system for interpreting technical condition from dynamograms based on machine learning tools. Neftyanoye khozyaystvo 2021; 4: 102-105. DOI: 10.24887/0028-2448-2021-4-102-105.
- [11] Martinović A, Bijanić M, Danilović D, Petrović A, Delibašić B. Unveiling deep learning insights: a specialized analysis of sucker rod pump dynamographs, emphasizing visualizations and human insight. Mathematics 2023; 11: 4782. DOI: 10.3390/math11234782.
- [12] Sun Y, Wang H, Zhang X, Yang J, Wang C, Shao J, Li Y. Comparative Analysis and Suggestions for Sucker Rod Pump Working Condition Diagnosis Using Machine Learning Techniques. SPE Asia Pacific Oil and Gas Conf. and Exhib. SPE 2024; D021S012R008.
- [13] Gumerova VI, Korobkov GE. Assessment of the technical condition of oil industrial equipment using digital technologies. Obshchestvo 2023; 3(30): 12-16.
- [14] Danilov SO. Identification of complications and malfunctions of submersible equipment of sucker rod pumping units using neural networks [In Russian]. Molodoy uchenyy 2019; 15(253): 17-22.
- [15] Sharaf SA. Beam pump dynamometer card prediction using artificial neural networks. KnE Eng 2018; 198-212.
- [16] Wang X, He Y, Li F, Wang Z, Dou X, Xu H, Fu L. A Working Condition Diagnosis Model of Sucker Rod Pumping wells Based on Deep Learning. SPE Prod. Operations 2021; 36(2): 317-326. DOI:10.2118/205015-PA.
- [17] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulina A. CatBoost: unbiased boosting with categorical features. Advances in Neural Information Processing Systems 2018; 31: 6638-6648.

Сведения об авторах

Вульфин Алексей Михайлович, окончил Уфимский государственный нефтяной технический университет. В 2022 году защитил диссертацию на соискание ученой степени доктора технических наук. Работает профессором кафедры вычислительной техники и защиты информации Уфимского университета науки и технологий. Сфера научных интересов: интеллектуальные системы защиты информации, кибербезопасность и машинное обучение. E-mail: vulfin.am@ugatu.su

Тагирова Клара Фоатовна, окончила Уфимский авиационный институт. В 2008 году защитила диссертацию на соискание ученой степени доктора технических наук, профессор с 2010 года. Работает профессором кафедры технической кибернетики Уфимского университета науки и технологий. Сфера научных интересов: автоматизация управления технологическими процессами добычи нефти из малодобитных скважин на основе динамических моделей. E-mail: tagirovakf@yandex.ru

Кириллова Анастасия Дмитриевна, окончила Уфимский государственный авиационный технический университет, в 2023 году защитила диссертацию на соискание степени кандидата технических наук. Работает доцентом кафедры вычислительной техники и защиты информации Уфимского университета науки и технологий. Сфера научных интересов: кибербезопасность, управление рисками информационной безопасности, интеллектуальные системы защиты информации. E-mail: kirillova.ad@ugatu.su

Сулавко Алексей Евгеньевич, в 2009 году окончил Сибирскую государственную автомобильно-дорожную академию, в 2023 году защитил диссертацию на соискание ученой степени доктора технических наук, профессор кафедры комплексной защиты информации Омского государственного технического университета. Область научных интересов: биометрия, распознавание образов, машинное обучение, защита информации. E-mail: sulavich@mail.ru

Ложников Павел Сергеевич, окончил Омский государственный технический университет. В 2019 году защитил диссертацию на соискание ученой степени доктора технических наук. Заведующий кафедрой комплексной защиты информации Омского государственного технического университета, руководитель научной школы «Нейросетевые алгоритмы искусственного интеллекта в защищенном исполнении», особенностью которой является использование Байесовских нейронных сетей, способных анализировать данные любой размерности. E-mail: lozhnikov@mail.ru

Поступила в редакцию 16 января 2026 г.



Окончательный вариант – 12 апреля 2026 г.

Neural network model for classifying dynamometer cards of sucker-rod pumping units

A.M. Vulfin¹, K.F. Tagirova¹, A.D. Kirillova¹, A.E. Sulavko², P.S. Lozhnikov²

¹ Ufa University of Science and Technology, Zaki Validi Str. 32, Ufa, 450076, Russia;

² Omsk State Technical University, Prospekt Mira 11, Omsk, 644050, Russia

Abstract

The work is devoted to the problem of increasing the efficiency of recognizing characteristic signs of complex faults in images of dynamometer charts of submersible pumping equipment under conditions of limited expert marking, pronounced class imbalance and the presence of measurement noise. A set of 20164 practical dynamometer charts, divided into 10 classes, is created. A systematic comparison of six neural network architectures (ResNet-18, ResNet-50, EfficientNet-B0, Vision Transformer, ConvNeXt) and four methods of representing dynamometer charts as images (contour representation, continuous wavelet transform scalogram, signed transform of distances to the contour, and their combination) is performed. The basic approach used is based on manual feature construction (Freeman chain code, approximation by geometric primitives) with classification using the CatBoost gradient boosting method. Model training and evaluation are performed using a 5-fold stratified cross-validation protocol with imbalance compensation via loss function weighting. It is shown that representations with a high proportion of informative pixels (scalogram, signed distance map) provide more stable learning and outperform sparse contour representation, especially for architectures with attention mechanisms. The best quality is achieved by combining the ResNet-18 model and scalogram (F1-macro = 0.831 with Accuracy = 0.899). The scientific novelty lies in the quantitative assessment of the influence of the type of dynamometer chart representation on the efficiency of different families of models under a single training protocol and a pronounced class imbalance. The practical significance lies in increasing the recognition accuracy of dynamometer chart images with complex faults and reducing the labor intensity of manual marking.

Keywords: dynamometer card, image classification, convolutional neural networks, wavelet transform, class imbalance.

Citation: Vulfin AM, Tagirova KF, Kirillova AD, Sulavko AE, Lozhnikov PS. Neural network model for classifying dynamometer cards of sucker-rod pumping units. *Computer Optics* 2026; 50(4): 1940. doi: 10.18287/COJ1940.

Acknowledgements: The research was financially supported by the Ministry of Science and Higher Education of the Russian Federation under project FSGF-2023-0004.

About authors

Alexey Mikhailovich Vulfin graduated from the Ufa State Petroleum Technological University, Doctor of Technical Sciences 2022, Professor at the Ufa University of Science and Technology. Research: intelligent information security systems, cybersecurity, and machine learning. E-mail: vulfin.am@ugatu.su

Klara Foatovna Tagirova graduated from the Ufa Aviation Institute, Doctor of Technical Sciences 2008, Professor since 2010, Professor at the Ufa University of Science and Technology. Research: automation of process control for oil production from low-yield wells based on dynamic models. E-mail: tagirovakf@yandex.ru

Anastasia Dmitrievna Kirillova graduated from the Ufa State Aviation Technical University, Candidate of Technical Sciences 2023, Associate Professor at the Ufa University of Science and Technology. Research: cybersecurity, information security risk management, intelligent information security systems. E-mail: kirillova.ad@ugatu.su

Alexey Evgenievich Sulavko graduated from the Siberian State Automobile and Highway Academy, Doctor of Technical Sciences 2023, Professor at Omsk State Technical University. Research: biometrics, pattern recognition, machine learning, information security. E-mail: sulavich@mail.ru

Pavel Sergeevich Lozhnikov graduated from Omsk State Technical University, Doctor of Technical Sciences 2019, Head of Department and scientific school (Bayesian neural networks for any-dimension data). E-mail: lozhnikov@mail.ru

Received January 16, 2026. The final version – April 12, 2026.
