

Санкт–Петербургский государственный университет

КОВАЛЕНКО Лев Алексеевич

Выпускная квалификационная работа
*Методы детектирования дефектов
нефтепроводов на основе данных внутритрубной
диагностики*

Уровень образования: магистратура

Направление 02.04.02 «фундаментальная информатика и
информационные технологии»

Основная образовательная программа ВМ.5503.2020 «Прикладная
математика, фундаментальная информатика и программирование»

Научный руководитель:

доцент, кафедра технологии программирования,
к.т.н. Блеканов Иван Станиславович

Рецензент:

начальник отдела, отдел стратегических инициатив,
ООО «Газпромнефть-ЦР»,
Кононов Ярослав Сергеевич

Санкт-Петербург

2022 г.

Содержание

Введение	4
Актуальность	4
Практическая значимость работы	5
Постановка задачи.	6
Цель работы.	6
Задачи работы.	6
Обзор литературы	7
Глава 1. Введение в предметную область	11
1.1. Процедура внутритрубной диагностики	11
1.2. Методы визуального контроля дефектов	13
Глава 2. Разработка методов детектирования дефектов	15
2.1. Структура и особенности данных	15
2.2. Аугментация датасета	21
2.3. Исследование дефектов	22
2.4. Поиск дефектов по паттерну	24
2.5. Детектирование дефектов методами классического машинного обучения	25
2.6. Классификация дефектов методами классического машинного обучения	26
2.7. Классификация дефектов методами на основе нейронных сетей	27
Глава 3. Проведение оценки качества алгоритмов	32
3.1. Процедура оценки качества	32
3.2. Результаты поиска дефекта по паттерну	32
3.3. Результаты детектирования методами классического машинного обучения	34
3.4. Результаты классификации методами классического машинного обучения	35
3.5. Результаты нейросетевого подхода	37
3.6. Выводы	40

Заключение	41
Результаты работы	41
Перспективы развития	41
Список литературы	43

Введение

Актуальность

Развитие современных информационных технологий затрагивает все области жизни общества. Не является исключением и сфера промышленности. В последнее время появилось огромное количество технологий, инструментов и программных комплексов, которые позволяют автоматизировать и оптимизировать производственные процессы. В настоящее время происходит повсеместная информатизация промышленности. Процесс информатизации производств подталкивает компании к проведению исследований в конкретной промышленной области и внедрению результатов исследований в технологические процессы.

Благодаря этому многие области промышленности получили большой толчок в развитии, например, автомобильная промышленность или нефтегазовая промышленность. В качестве примера можно привести то, что за последние годы в группе компаний «Газпром Нефть» были внедрены новейшие решения для оптимизации технологических и производственных процессов [1]:

- Проект «Когнитивный геолог», автоматизирующий предварительную обработку данных, полученных в результате геологоразведочных работ.
- Технология «КиберГРП» применяется для моделирования гидроразрыва пласта.
- Экспертная система «ProBA» накапливает информацию и формирует базу знаний по «Большой Ачимовке». На основе различных данных она помогает подобрать оптимальное технологическое решения для бурения новой скважины.

На текущий момент «Газпром Нефть» проводит огромную исследовательскую программу по созданию информационных и технологических решений для проведения геологоразведочных работ и дальнейшей разработки трудноизвлекаемых запасов нефти [2].

Практическая значимость работы

На балансе нефтедобывающей компании находится порядка 12 тыс. км промысловых трубопроводов. Одной из важных задач является оценка состояния эксплуатируемых трубопроводов [3]. На данный момент эта процедура требует большого количества ручного труда высококвалифицированных специалистов при визуальном контроле результатов внутритрубной диагностики (ВТД). Эксперты тратят до одного месяца на интерпретацию данных внутритрубной диагностики одного километра трубы. Автоматизация данного процесса позволит сократить время между проведением ВТД и интерпретацией результатов, а также уменьшить влияние человеческого фактора. В результате уменьшится количество прорывов нефтепроводов и утечек нефти. В компании разрабатывается программный комплекс для детектирования дефектов на основе экспертных правил. Использование методов машинного обучения может повысить качество детектирования дефектов.

Практическая значимость работы заключается в том, что предложенные методы детектирования дефектов позволят повысить качество процесса интерпретации данных ВТД и могут быть внедрены в разрабатываемый программный комплекс.

Постановка задачи.

Цель работы.

Цель данной работы — разработка методов и инструментов искусственного интеллекта для эффективного обнаружения дефектов в рамках внутритрубной диагностики.

Задачи работы.

Для достижения цели были поставлены следующие задачи:

- провести обзор и анализ подходов и методов машинного обучения для детектирования аномалий и дефектов;
- сбор и предобработка данных внутритрубной диагностики для обучения и оценки качества моделей;
- реализация методов интеллектуального анализа для обнаружения дефектов;
- тестирование и оценку качества разработанных алгоритмов на данных ВТД;
- экспертный анализ полученных результатов и рекомендации по улучшению качества моделей.

Обзор литературы

В настоящее время существует огромное количество алгоритмов для детектирования аномалий во временных рядах. Эти алгоритмы также могут быть применены и в задаче детектирования дефектов, так как данные ВТД являются набором временных рядов. В обзорной статье [4] методы детектирования аномалий разделяют на три класса: статистические методы, методы на основе классического обучения и подходы с использованием нейронных сетей.

Авторы статьи отнесли к статистическим методам обнаружения аномалий во временных рядах такие подходы, как ARIMA [5], Exponential Smoothing [6], Prediction Confidence Interval [7] и т.д. Эти методы основаны на предположении, что дефектом является статистическое отклонение во временном ряду. Для обнаружения дефектов они используют информацию о статистическом распределении значений и отклонений между ними.

В задаче детектирования аномалий временных рядов применяются следующие методы классического машинного обучения.

- Алгоритм k-means [8] предполагается использовать для кластеризации подпоследовательностей временного ряда. Таким образом, аномалии будут выделены в отдельный кластер или будут являться объектами, находящимися дальше всего от центроиды кластера.
- DBSCAN [9, 10] метод основан на иерархической кластеризации, он также работает с отрезками временного ряда и разделяет их на три категории: основные объекты, граничные объекты и аномалии.
- Алгоритм Isolation Forest [11] основан на предположении, что аномалии будут изолированными (удаленными) точками в данных. Он представляет собой ансамбль деревьев изоляции, которые отсекают наиболее изолированные объекты в начале. Этот подход также требует использования подпоследовательностей временного ряда в качестве входных объектов.
- Для детектирования аномалий с помощью SVM [12] используется его

модификация ОС-SVM [13], которая на вход принимает отрезки временного ряда с нормальными данными (без аномалий). В ходе обучения происходит выявление опорных объектов. При использовании ОС-SVM определяет объект либо как нормальный, либо как аномалию.

- XGboost [14] выступает предиктором временного ряда без аномалий. Аномалия во временном ряду вызывает сильное отклонение предсказанного значения XGboost от реального. То есть аномалия детектируется на основе ошибки предсказания модели.

Кроме методов, описанных ранее, можно использовать методы бинарной классификации, так как задачу детектирования можно свести к задаче классификации при наличии разметки временного ряда на нормальные и аномальные значения.

Обзор методов глубокого обучения стоит начать с фундаментальной нейронной сети — многослойного перцептрона (Multiple Layer Perceptron, MLP) [15]. Он может применяться для предсказания временных рядов [16], поэтому детектирование аномалий временного ряда происходит на основе ошибки предсказания MLP. Также для решения задачи детектирования аномалий могут быть использованы сверточные нейронные сети (convolutional neural networks CNNs). В области компьютерного зрения они решают задачи классификации [17], сегментации [18] и детекции [19]. В отличие от работы с изображениями, в задаче детекции аномалий используются одномерные сверточные сети. Одномерные сверточные сети также предсказывают следующие значения ряда, и аномалия детектируется на основе ошибки предсказания [20]. Кроме обычных CNNs могут использоваться различные модификации, например, Residual Neural Network (Resnet) [21], которая решает проблему исчезающих градиентов CNN. В 2016 году была разработана сеть wavenet [22], которая использовалась для генерации звуковых волн. Это также CNN сеть, но с нерегулярными, а расширенными слоями свертки. Это позволяет ей рассматривать не только локальные паттерны как CNN, но и обобщать более длинные последовательности. Далее рассмотрим рекуррентные нейронные сети на примере Long Short Term Memory

(LSTM) [23, 24] network и Gated recurrent unit (GRU) [25, 26]. Архитектуры этих сетей изначально проектировались для предсказания временных рядов, поэтому можно использовать ошибку предсказания для детектирования аномалии. Кроме этого способа сети позволяют предсказывать не одно, а несколько значений временного ряда, в этом случае используется метод правдоподобия для определения аномалии относительно предсказанной части временного ряда. Один из методов детектирования аномалии — это понижение размерности данных — создание проекции данных в пространстве меньшей размерности. В глубоком обучении для этого используются Autoencoders — сети, состоящие из энкодера и декодера. Autoencoders уже использовались для детекции аномалий на изображениях [27, 28]. В статье [29] их использовали для детекции аномалий временных рядов. Autoencoder был обучен кодировать участки временного ряда без аномалий в самих себя. Для детекции производилось сравнение между собой векторов в скрытом пространстве (результатов работы энкодера), так как аномалии приводили к сильным отклонениям в скрытом пространстве модели.

Для детектирования дефектов на поверхностях объектов часто используются методы на основе компьютерного зрения. В статье [30] проводилось построение сверточной сети для определения дефектов (загрязнений, царапин, заусенцев и следов износа на поверхностных деталях продукта). Авторы проводят ряд экспериментов, в которых демонстрируют качество работы предложенного решения. Авторы статьи [31] используют Faster-RCNN для определения дефектов нанесения химического волокна на высокоскоростных прядильных и намоточных машинах. По словам авторов, полученный алгоритм имеет лучшие результаты относительно используемых подходов в данный момент, точность в экспериментах достигала 98%. В статье [32] рассматривается задача поиска дефектов на изображениях керамической трубки с помощью гибридной модели SVM-QPSO, эксперименты проводились на выборке, включающей в себя 500 изображений с дефектами.

В работе [33] авторы решают задачу обнаружения дефектов на поверхности дороги. В рамках статьи они сравнивают различные подходы:

Gradient Boosting, CNN, SVM и Random Forest. Для методов классического машинного обучения они собирают набор признаков на основе дескрипторов изображений. Авторы вводят несколько метрик для сравнения результатов алгоритмов: *Accuracy*, *Recall*, *Precision*, *F₁ – measure*. Эти метрики основываются на матрице ошибок (таблица 0.0.1).

Таблица 0.0.1: Матрица ошибок

	Дефекты по разметке	Нормальный объект по разметке
Предсказание класса дефекта	True positive	False positive
Предсказание класса нормального объекта	False negative	True negative

На основе представленной матрицы можно подсчитать выбранные метрики по следующим формулам:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$

$$Recall = \frac{TP}{TP + FN},$$

$$Precision = \frac{TP}{TP + FP},$$

$$F_1 - measure = \frac{2 * Recall * Precision}{Recall + Precision}.$$

Глава 1. Введение в предметную область

1.1 Процедура внутритрубной диагностики

Существует большое количество предложений по проведению магнитной и ультразвуковой внутритрубной диагностики нефтепровода для дальнейшей оценки состояния и определения плана ремонтных мероприятий. Для проведения ВТД используются автономные зонды (внутритрубные роботы) (рис. 1.1.1.), оборудованные профилометрами и магнитными или ультразвуковыми дефектоскопами.



Рис. 1.1.1: Зонд для проведения внутритрубной диагностики.

Внутритрубный робот имеет цилиндрическую форму, и по его окружности через равные угловые промежутки установлены измерительные блоки из 4 датчиков. Таких измерительных блоков может быть 48 или 64, в зависимости от диаметра робота. В процессе проведения диагностики автономный зонд загружается в нефтепровод, далее он перемещается внутри под давлением нефтяного потока, замераю изменение магнитного поля у стенок трубы (рис. 1.1.2.).

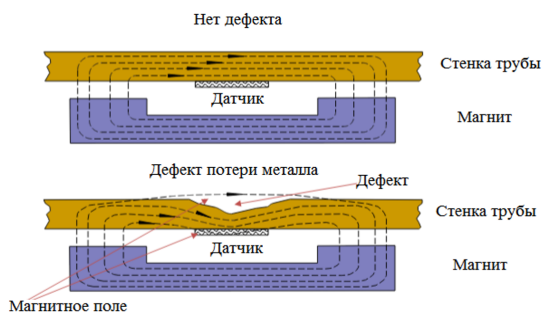


Рис. 1.1.2: Изменение магнитного поля при наличии дефекта.

Запись данных о текущих показателях магнитного поля происходит через равные промежутки времени одновременно со всех датчиков. Для ультразвуковой диагностики трубопровода зонд оснащается другими измерительными блоками с датчиками, использующими ультразвук для замера толщины стенок трубы.

Сервисные компании, такие как «Baker Hughes» [34] и «Интрон плюс» [35], предлагают услуги по проведению магнитной и ультразвуковой внутритрубной дефектоскопии. Компании используют собственные приборы и проприетарное программное обеспечение для расшифровки магнитограмм. Поиск дефектов производится посредством визуального контроля.

Для промышленных компаний, деятельность которых связана с трубопроводами, создан и включен в ГОСТ перечень единых стандартов. Существует ГОСТ [36] для единой категоризации дефектов трубопроводов:

- трещины и зоны трещин, включая стресс-коррозионные трещины;
- коррозия;
- механические повреждения;
- металлургические дефекты;
- дефекты геометрии трубы.

ВТД с помощью магнитных дефектоскопов позволяет выявлять коррозионные дефекты. Механические повреждения и дефекты геометрии трубы могут нанести повреждения внутритрубному зонду (рис. 1.1.3) и вывести из строя несколько датчиков или несколько измерительных блоков. Данные, снятые с поврежденных зондов, могут содержать недостоверные значения или же пропуски в данных. В таких случаях необходима повторная процедура ВТД.



Рис. 1.1.3: Повреждения внутритрубного дефектоскопа.

1.2 Методы визуального контроля дефектов

Для проведения визуального контроля в сервисных компаниях используется проприетарное программное обеспечение, позволяющее визуализировать данные дефектоскопов. Данные представлены в виде цветной полосы heatmap, где градациями цветов обозначены значения датчиков. Таким образом, пиковые значения относительно значений общей поверхности трубы резко выделяются. Пример магнитограммы с дефектом представлен на рисунке 1.2.1.

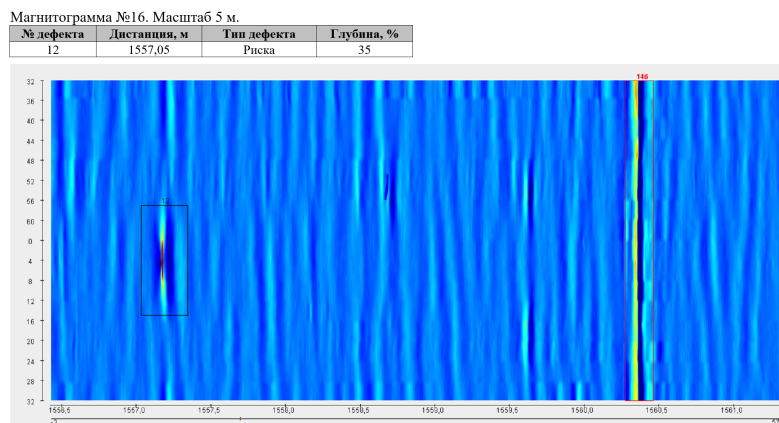


Рис. 1.2.1: Пример визуализации магнитограммы.

Для определения дефектов типа коррозия существуют способы классификации (таблица 1.2.1) на основе физических параметров дефекта [37].

Таблица 1.2.1: Классы дефектов

Класс	Определение размеров
Общая коррозия	$[W \geq 3A]$ и $[L \geq 3A]$
Питтинг	$([1A \leq W < 6A]$ и $[1A \leq L < 6A]$ и $[0.5 < L/W < 2])$ и не $([W \geq 3A]$ и $[L \geq 3A])$
Продольная канавка	$[1A \leq W < 3A]$ и $[L/W \geq 2]$
Поперечная канавка	$[L/W \leq 0.5]$ и $[1A \leq L < 3A]$
Продольная риска	$[0 < W < 1A]$ и $[L \leq 1A]$
Поперечная риска	$[W \leq 1A]$ и $[0 < L < 1A]$

L — длина дефекта, т.е. максимальный размер в продольном направлении (вдоль оси трубы); W — ширина дефекта, т.е. максимальный размер в поперечном (окружном) направлении; $A = 10$, если $\delta < 10$, $A = \delta$, если $\delta > 10$, где δ — номинальная толщина стенки трубы.

Глава 2. Разработка методов детектирования дефектов

2.1 Структура и особенности данных

Для проведения исследований по разработке методов детектирования дефектов была предоставлена выгрузка из 33 магнитограмм. В ее состав для каждой магнитограммы входили:

- исходный файл магнитограммы в бинарном виде;
- экспертный отчет в «.pdf» или «.doc» формате;
- реперные точки.

В исходном файле показания датчиков записаны в бинарном виде ради экономии памяти внутрирубного дефектоскопа. Для расшифровки данных был написан парсер с использованием инструмента для чтения структурированных бинарных данных kaitaistruct [38]. После расшифровки данных и приведения к табличному виду данные были сохранены в PostgreSQL [39] в целях консистентного хранения информации и возможности индексации по дистанции магнитограммы. Список параметров, полученных после расшифровки магнитограмм, приведен в таблице 2.1.1.

Таблица 2.1.1: Параметры магнитограмм

Параметр	Описание	Ед. измерения
Tag	Начало новой записи данных (технический параметр)	-
Size	Размер поля для записи данных (технический параметр)	-
Time	Момент записи	мкс
Dist	Значение дистанции	10 мкм
Index	Порядковый номер измерения	-
Flags	Флаги сканирования (технический параметр)	-

Status X	Статус для блока из четырех датчиков, описывающий возможные ошибки в них: 0 – отсутствие ошибок; X – номер блока	-
X.Y	Номер датчика Y из блока X, например, «1.3»	-

Понятие «Технический параметр» вводится в целях обозначения параметров, использующихся либо для расшифровки, либо для конверсии из исходного бинарного файла. Из приведенного выше списка параметров для исследований наиболее значимыми факторами являются Dist и X.Y, так как они указывают на какой дистанции от начала магнитограммы находится дефект и какой кадр измерений ему соответствует.

Экспертный отчет содержал информацию об условиях съемки, местоположении и типе конструктивных элементов, а также типе, местоположении и критичности дефектов трубопровода. Разметка конструктивных элементов и дефектов тоже была перенесена из отчета в табличный вид и сохранена в базе данных. Параметры, использующиеся в разметке, перечислены в таблицах 2.1.2 и 2.1.3.

Таблица 2.1.2: Параметры конструктивных элементов

Параметр	Описание	Ед. измерения
id	Уникальный идентификатор конструктивного элемента – строка	-
constructive type	Тип конструктивного элемента на основе отчетов	-
start dist	Начало конструктивного элемента	10 мкм
end dist	Конец конструктивного элемента	10 мкм
wall width	Толщина стенки трубы на основе отчета	мм

Таблица 2.1.3: Параметры дефектов

Параметр	Описание	Единица измерения
id	Уникальный идентификатор дефекта – строка	-
defect type	Тип дефекта на основе отчетов	-
defect place	Местоположение дефекта: «Стенка трубы» или «Сварной шов»	-
defect depth	Глубина дефекта на основе отчета	%
start dist	Начало дефекта	10 мкм
end dist	Конец дефекта	10 мкм

Для сопоставления дистанции из отчета и данных с датчиков зонда использовались реперные точки, которые представляли собой сопоставление координат, полученных от дефектоскопа, и реальных координат дефекта. На основе этого была пересчитана дистанция для каждого измерения датчиков.

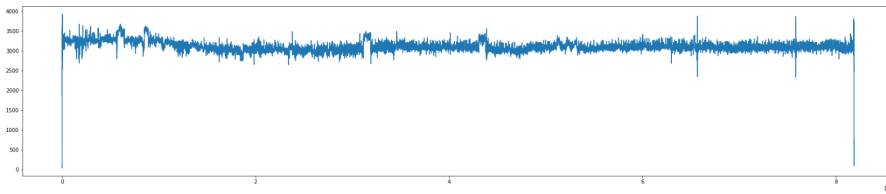
Важно заметить, что число кадров (строк с показаниями датчиков) в магнитограмме зависит от времени прохождения зондом трубопровода: для прохождения трубопроводов длиной 40 км требуется до нескольких дней. В представленном ниже датасете проходимое расстояние составляло от 1.5 км до 40 км. Участки свыше 10 км были исключены из рассмотрения в связи с объемом ресурсов, требуемых на их обработку и анализ.

Итоговый датасет состоял из 18 магнитограмм для анализа дефектов (магнитограммы 1–4 не содержали дефектов согласно отчетам). Конечный список магнитограмм для построения математической модели приведен в Таблице 2.1.4.

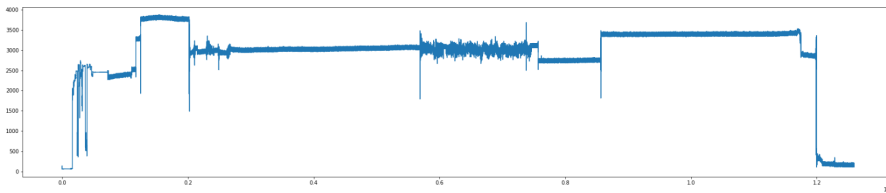
Таблица 2.1.4: Итоговый список магнитограмм

Магнитограмма	Количество записей
Магнитограмма №5	3761227
Магнитограмма №6	1072643
Магнитограмма №7	449613
Магнитограмма №8	711720
Магнитограмма №9	412681
Магнитограмма №10	1165283
Магнитограмма №11	3047114
Магнитограмма №12	575441
Магнитограмма №13	1498927
Магнитограмма №14	683275
Магнитограмма №15	1108527
Магнитограмма №16	1630936
Магнитограмма №17	338127
Магнитограмма №18	233018
Магнитограмма №19	291116
Магнитограмма №20	222167
Магнитограмма №21	860107
Магнитограмма №22	1772941

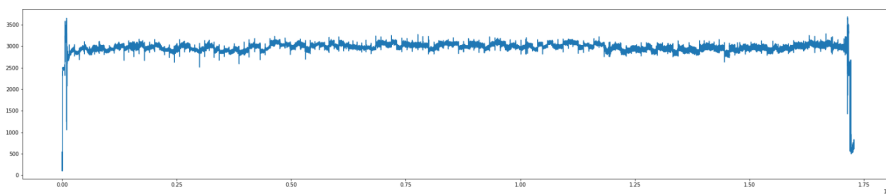
Для проверки данных были визуализированы сигналы датчиков всех магнитограмм. Так как данные были представлены сразу 64 значениями в один момент времени, поскольку записывались всеми датчиками, то для отрисовки графиков использовалось среднее значение по всем датчикам в определенный момент времени.



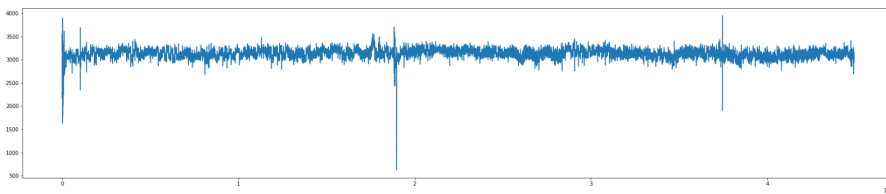
Магнитограмма №1



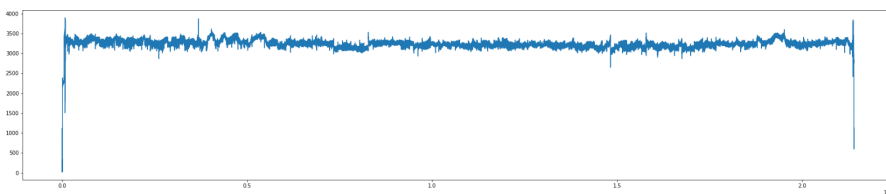
Магнитограмма №3



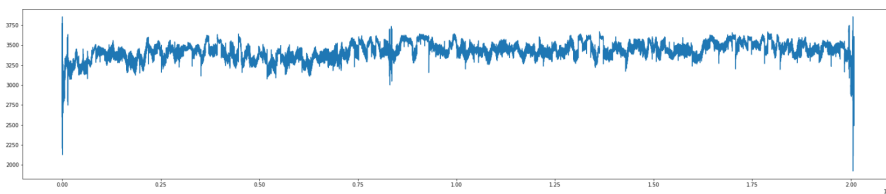
Магнитограмма №9



Магнитограмма №10

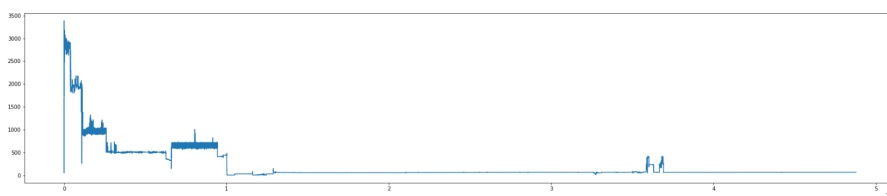


Магнитограмма №12

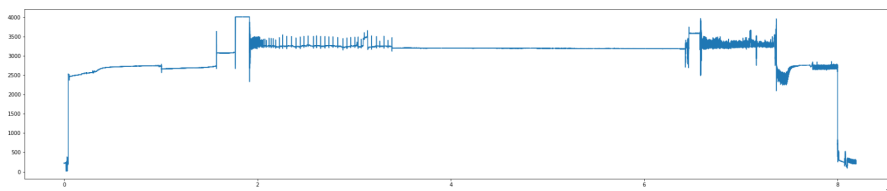


Магнитограмма №14

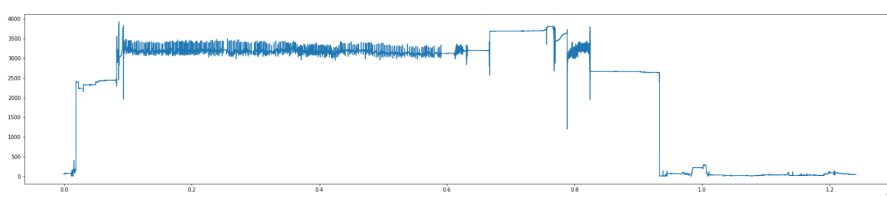
Рис. 2.1.1: Магнитограммы с сильными шумами.



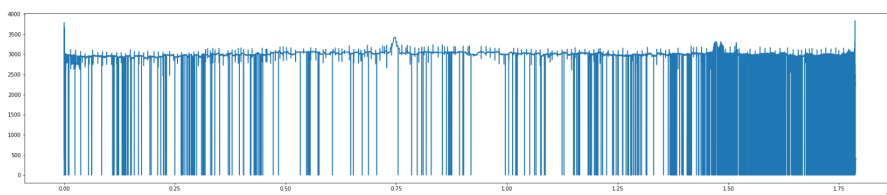
Магнитограмма №2



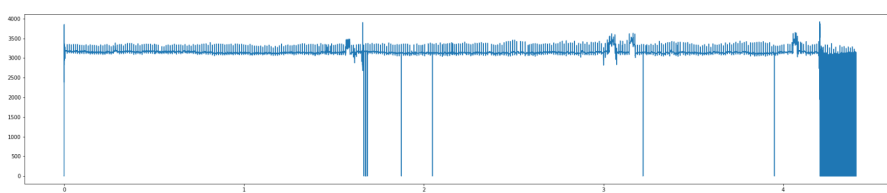
Магнитограмма №4



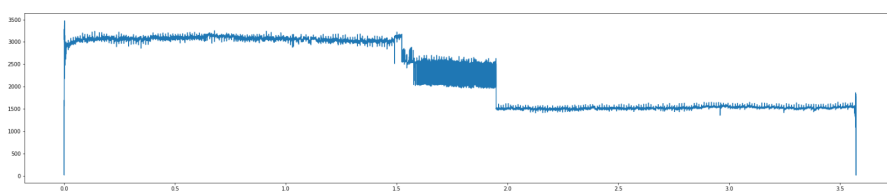
Магнитограмма №5



Магнитограмма №7



Магнитограмма №15



Магнитограмма №16

Рис. 2.1.2: Магнитограммы с битами или недостоверными данными.

На рисунках 2.1.1 и 2.1.2 представлены зашумленные магнитограм-

мы или «битые» магнитограммы, которые также использовались для проведения экспериментов, но с заведомым ожиданием ухудшения качества работы алгоритмов на них.

При проверке статусов блоков датчиков у «битых» магнитограмм было выявлено, что большая часть блоков датчиков выдавала сообщения об ошибках: «снижение напряжение питания», «ошибки приемника данных (шум, перегрузка, ошибка кадра) UART БЭ» или «неверный заголовок данных БД». Было принято решение восстановить эти данные на основе верных значений. С этой целью было рассчитано среднее значение по всей длине магнитограммы для достоверных данных. Если недостоверные значения имели резкие отклонения — более тысячи условных единиц от соседних показаний — то они заменялись на среднее значение магнитограммы. После проведения такой предобработки появилась возможность использовать весь датасет.

2.2 Аугментация датасета

Для увеличения обучающей выборки и повышения стабильности работы модели рассматривалась задача аугментации датасета.

Самым простым методом аугментации являлся циклический сдвиг показаний датчиков, так как магнитограмма представляет собой плоскую развертку цилиндрической трубы. Для каждого кадра записи показаний показание первого датчика становилось вторым, показание второго датчика — третьим, и так далее до показания шестьдесят четвертого датчика, оно становилось первым. Эту операцию можно было повторять многократно, до 64 раз. Это позволяло сильно увеличить объем данных.

Также был рассмотрен метод аугментации, основанный на том, что при движении внутри трубы автономный зонд имеет некую угловую скорость и потому проворачивается вокруг своей оси. Для аугментации проводилось моделирование такого движения. Как только моделируемый зонд менял свое положение на 5,65 градуса вокруг своей оси, то производился циклический сдвиг всех последующих кадров записи. Таким образом, зонд мог выполнить несколько оборотов вокруг своей оси во время сканирова-

ния трубопровода.

В рамках оценки качества для расширения обучающей выборке применялась процедура циклического сдвига с шагом равному количеству датчиков в блоке, то есть 4. Размер выборки был увеличен в 16 раз и составил 288 магнитограмм. Это позволило расширить вариацию положения дефектов на магнитограмме, но разнообразить типы дефектов не является возможным.

2.3 Исследование дефектов

На основе экспертной разметки были получены средние размеры дефекта: протяженность 15 тиков, ширина – показания четырёх датчиков. Для наглядности были выделены паттерны дефектов – двумерный (матричный) и одномерный (усреднённый по датчикам). На рисунке 2.3.1 представлен матричный паттерн дефекта, на рисунке 2.3.2 – одномерный паттерн дефекта.

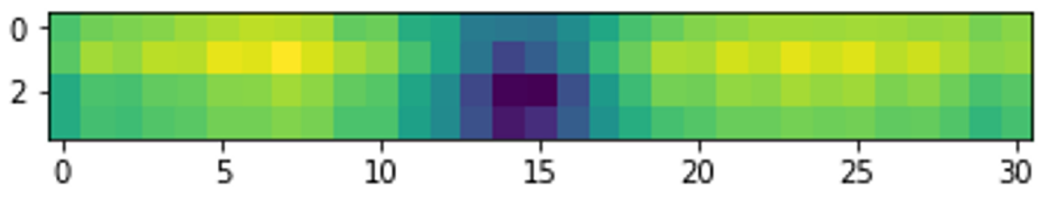


Рис. 2.3.1: Матричный паттерн дефекта.

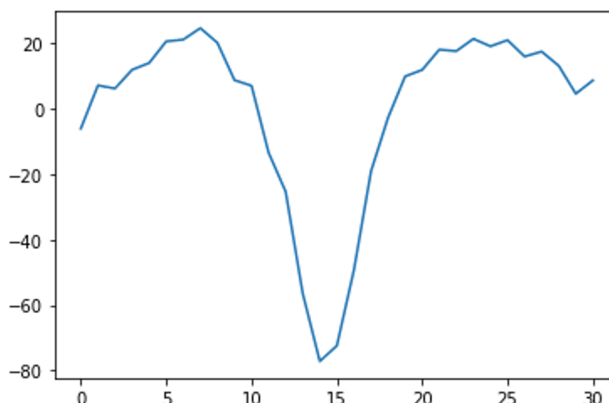


Рис. 2.3.2: Одномерный паттерн дефекта.

Также было проведено построение одномерных паттернов для каждой группы дефектов согласно классификации 1.2.1.

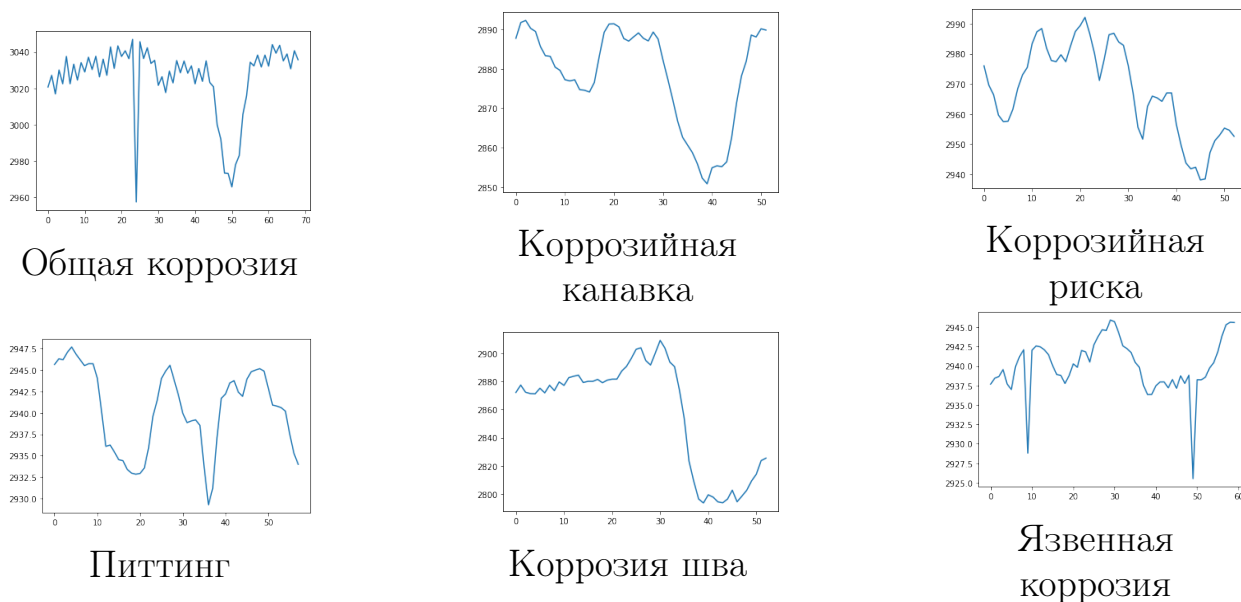


Рис. 2.3.3: Паттерны классов дефектов.

На рисунке 2.3.3 приведены примеры паттернов дефектов. Можно заметить, что дефекты довольно сильно отличаются между собой по форме и далеко не все подходят под приведённый ранее паттерн дефекта. Единственное обобщение, которое можно вынести из приведенного примера, заключается в том, что дефект является областью вокруг локального минимума.

Для признакового описания дефекта могла быть использована область вокруг дефекта (матричное представление), но так как большинство алгоритмов классического машинного обучения не умеют выделять пространственные признаки, был составлен набор дескрипторов (признаков) для описания дефекта на основе матричного представления:

- среднее значения по всем элементам матрицы;
- стандартное отклонение значений по всем элементам матрицы;
- длина пика вдоль магнитограммы;
- количество датчиков, зафиксировавших пик;

- глубина пика относительно показаний одного датчика;
- глубина пика относительно показаний датчиков в этом кадре;
- глубина пика относительно среднего значения матрицы;
- минимальное значение среди элементов матрицы;
- L_1 — расстояние от одномерного представления пика до паттерна;
- L_2 — расстояние от одномерного представления пика до паттерна;
- L_2^N — расстояние от одномерного представления пика до паттерна;
- косинусное расстояние от одномерного представления пика до паттерна;
- значение корреляции между пиком и паттерном;
- дистанция, посчитанная между одномерным представлением пика и паттерном по алгоритму Dynamic Time Warping (DTW) [40].

2.4 Поиск дефектов по паттерну

После проведения исследований дефектов было принято решение рассматривать в качестве одного из подходов детектирования дефектов поиск по паттерну. Для реализации этого подхода использовался метод `correlate` из библиотеки `scipy` [41]. Кроме данного метода рассматривались различные меры для определения близости возможного дефекта к паттерну:

$$L_1(x, y) = \sum_{i=0}^n |x_i - y_i| \text{ — манхэттенское расстояние,}$$

$$L_2(x, y) = \sum_{i=0}^n (x_i - y_i)^2 \text{ — евклидово расстояние,}$$

$$L_2^N(x, y) = \sum_{i=0}^n \frac{(x_i - y_i)^2}{1 + (x_i - y_i)^2} \text{ — нормализованная мера } L_2,$$

$$\text{cos_dist}(x, y) = \frac{x \cdot y}{\|x\| \|y\|} - \text{косинусное расстояние.}$$

Применение данного подхода для всей магнитограммы затруднительно, так как дефекты — это редкие одиночные события, большое количество расчетов будет для нормальных областей. В связи с этим было принято решение производить поиск областей интереса — участков временного ряда, имеющих пик (локальный минимум) с определенными параметрами. Такая оптимизация позволила в разы сократить время обработки магнитограммы. Для расчета меры близости брался паттерн дефекта и фрагмент временного ряда схожей длины из области интереса. Проводился расчет близости между фрагментом магнитограммы и паттерном, далее на основе порогового значения определялось, является ли фрагмент дефектом или нет. Пороговое значение позволяло находить компромисс между метриками качества Recall и Precision. Увеличение порогового значения позволяло повысить полноту детекции дефектов. Уменьшение порогового значения давало повышение точности детекции.

2.5 Детектирование дефектов методами классического машинного обучения

Среди методов детектирования аномалий был рассмотрен ОС-SVM [13]. Для применения метода использовалась имплементация, реализованная в библиотеке PyOD [42]. Методы, основанные на ошибке предсказания, не рассматривались, так как ряд показаний датчиков магнитограммы имеет сложную структуру и может быть зашумлен, в связи с чем возникало бы большое количество ложноположительных предсказаний. Также на отказ от рассмотрения этих методов повлияло время, требуемое на предсказание всего ряда показаний датчика. От использования методов кластеризации также было решено отказаться, так как они требовали точной настройки для каждой отдельной магнитограммы, соответственно, их использование на новых магнитограммах невозможно.

Для обучения выбранного метода требовалось собрать выборку нормальных участков магнитограмм, не содержащих дефектов или конструк-

тивных элементов. В качестве входных данных для обучения алгоритма могли использоваться матричные представления дефектов, но так как алгоритм не учитывал пространственные признаки, а матричное представление содержало избыточное количество информации (например, шумы сигнала), был использован набор признаков, описанный в разделе 2.3. В обучающую выборку вошли участки магнитограмм, не имеющие дефектов. Также для увеличения точности детекции в обучающую выборку включались конструктивные элементы и зашумленные участки магнитограммы.

Для детектирования дефектов на магнитограмме использовалась оптимизация, примененная при поиске дефектов по паттерну. Аналитическим способом проводился поиск областей интереса, для каждой области подсчитывались дескрипторы. Алгоритм принимал на вход признаковое описание и относил их либо к нормальным данным, либо к дефектам.

2.6 Классификация дефектов методами классического машинного обучения

Как уже говорилось ранее, задачу детектирования аномалий временного ряда можно свести к задаче классификации участков временного ряда. При такой постановке задачи были рассмотрены следующие методы классического машинного обучения: Random Forest [43], Extra Trees [44] и CatBoost [45]. Их реализации были взяты из библиотек Scikit-Learn [46] и CatBoost [47].

Для обучения моделей была составлена выборка из областей интереса, определенная аналитическим подходом (раздел 2.4). Для каждой области присваивалась метка нормального участка, конструктивного элемента или дефекта. После этого области были описаны с помощью выбранных дескрипторов (раздел 2.3). Затем была проведена балансировка выборки с использованием методов SMOTE и NeighbourhoodCleaningRule, реализованных в библиотеке imbalanced-learn [48]. Сначала было проведено увеличение выборки дефектов алгоритмом SMOTE, затем с помощью NeighbourhoodCleaningRule были уменьшены выборки конструктивных элементов и нормальных участков до количества дефектов.

На подготовленной выборке проводилось обучение моделей и настройка гиперпараметров. Подбор гиперпараметров производился по сетке, в качестве метрики выступала F_1 measure, как комбинация recall и precision. Для ее расчета по трем классам применялся weighted macro-average подход. Для результатов F_1 категории дефектов вес был в два раза больше, чем по остальным классам при усреднении. Для замера точности при процедуре оценки подсчитывались *Recall*, *Precision* и F_1 measure для класса дефекта относительно всех остальных. По сути классы нормального участка и конструктивного элемента объединялись.

Для детектирования дефектов на магнитограмме формировался набор областей интереса. Он описывался с помощью дескрипторов и подавался на вход модели. Модель на основании этих данных делала вывод о том, к какой категории отнести область интереса.

2.7 Классификация дефектов методами на основе нейронных сетей

Для обучения нейронных сетей был сформирован специальный датасет. Для этого были сделаны следующие шаги.

- Показания датчиков нормализованы.
- Было взято окно в 100, 50, 30 (разные датасеты) тиков (записей/строк в магнитограмме). Начиная с 200 тика на магнитограмме был совершён проход окна с шагом 50 тиков и 25 тиков для создания двух датасетов. Все данные с одинаковым шагом окна были объединены. Окно считается атомарным элементом магнитограммы. Если в окне был участок с длиной больше 5 тиков с отметкой, что там есть дефект, такое окно получало флаг, что содержит дефект. Для конструктивных элементов аналогично, но участки были длиной 10 тиков.
- Данные были не сбалансированы, было много окон без дефектов. Для балансировки были взяты все окна, содержащие дефект, и подсчитано их количество. В таком же количестве были взяты окна без дефектов и конструктивных элементов и окна с конструктивными элементами.

Используя встроенные средства tensorflow [49], балансировка проводилась на лету. Таким образом было удовлетворено требование, что все типы без дефектов и конструктивных элементов, с дефектом, с конструктивным элементом были в одинаковом количестве для обучения. Тут же легко решалась проблема с тем, что есть пересечения, когда в одном окне есть и дефект, и конструктивный элемент.

- После были созданы обучающие и тестовые выборки. Сначала выборка просто разбилась по стратегии 80% — для обучения, 20% — для тестов и для проверки и первичной отладки моделей. После того, как стало ясно, что модель сходится и получаются ожидаемые результаты, стратегия разбиения изменилась. Использовалась кроссвалидация, т.е. брались все объекты (X) магнитограммы за исключением одной отдельной магнитограммы (x). Данные по взятым объектам X использовались как обучающая выборка. Валидация и тестирование проходили на данных объектах.

При моделировании использовались достаточно простые модели для проверки возможности и потенциала использования нейронных сетей. В дальнейшем следует провести поиск по лучшим архитектурам и параметрам. К тому же, были использованы параметры, делающие модель легче (для каждой архитектуры свои), для того, чтобы общее количество оцениваемых параметров было меньше и можно было бы уйти от проблемы переобучения.

Для построения моделей была выбрана следующая обобщенная архитектура:

- слой для приема входных данных;
- слои модели в части их особенностей, для CNN - сверточные слои, для LSTM - двунаправленные слои LSTM;
- несколько полносвязанных слоев;
- 3 выхода под каждый тип (без конструктивных элементов и дефектов, с конструктивным элементом, с дефектом). При этом выход для

дефектов вносит в 2 раза больший вклад в ошибку. В таком случае модель в общем показывает результаты лучше.

Выбор такой архитектуры модели обоснован тем, что для бинарной классификации — есть конструктивный элемент/дефект или нет — изменение архитектуры отличалось бы лишь количеством выходов. Возможно потребовалось бы увеличивать датасет или уменьшать количество параметров в модели для того, чтобы не наступало переобучение. Но если есть несколько выходов, то модель учится сначала смотреть на данные обобщенно и также потом делать предсказание на одном и том же представлении. Поэтому возможно использование меньшего датасета и большего количества параметров. Такой подход позволяет сократить общее время разработки и обучения.

Выбранные архитектуры моделей представлены на рисунках 2.7.1–2.7.3.

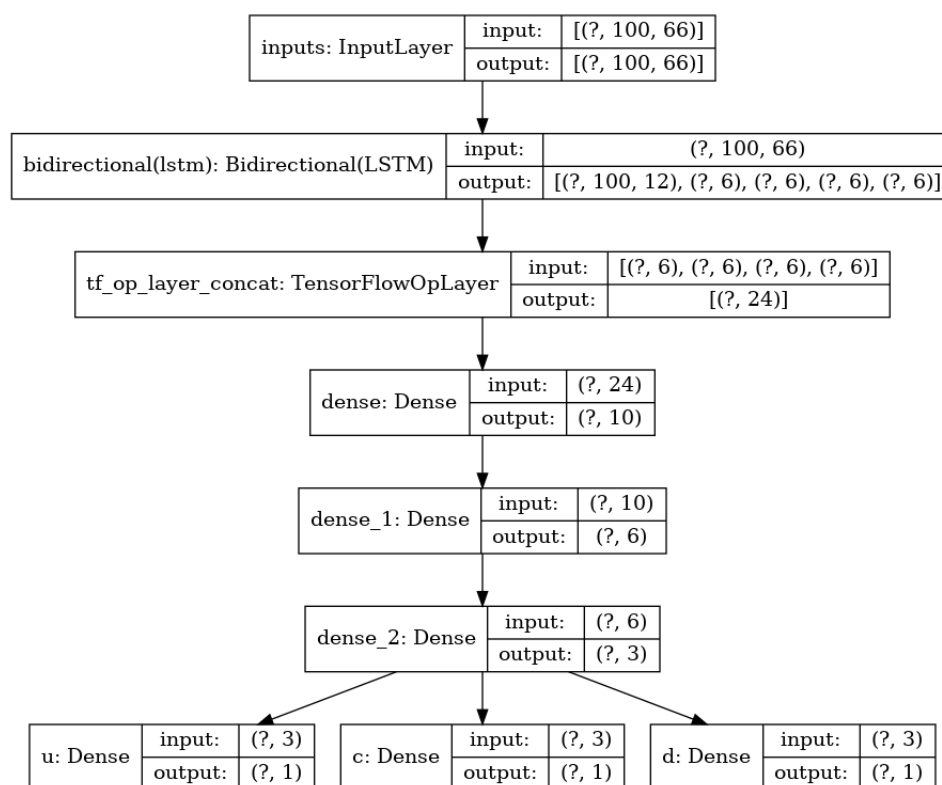


Рис. 2.7.1: Архитектура LSTM, количество параметров: 3,853.

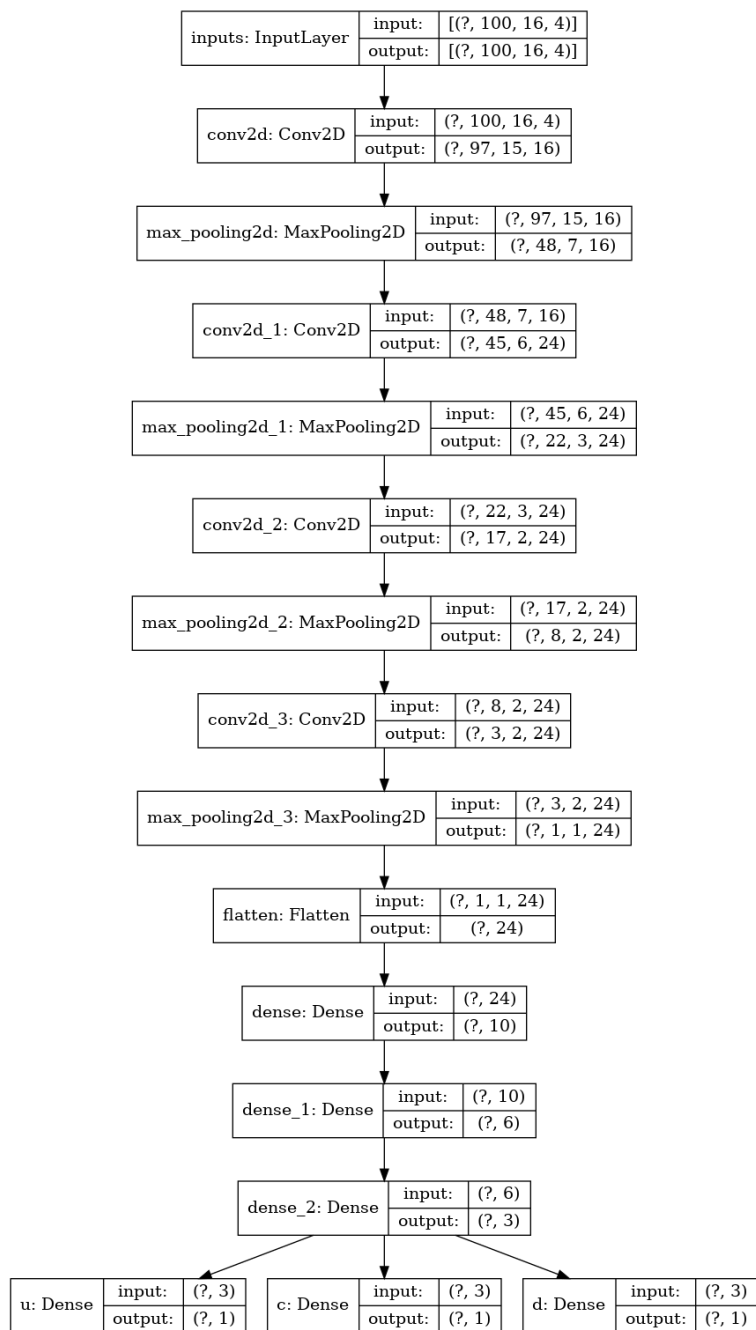


Рис. 2.7.2: Архитектура CNN, количество параметров: 14,389.

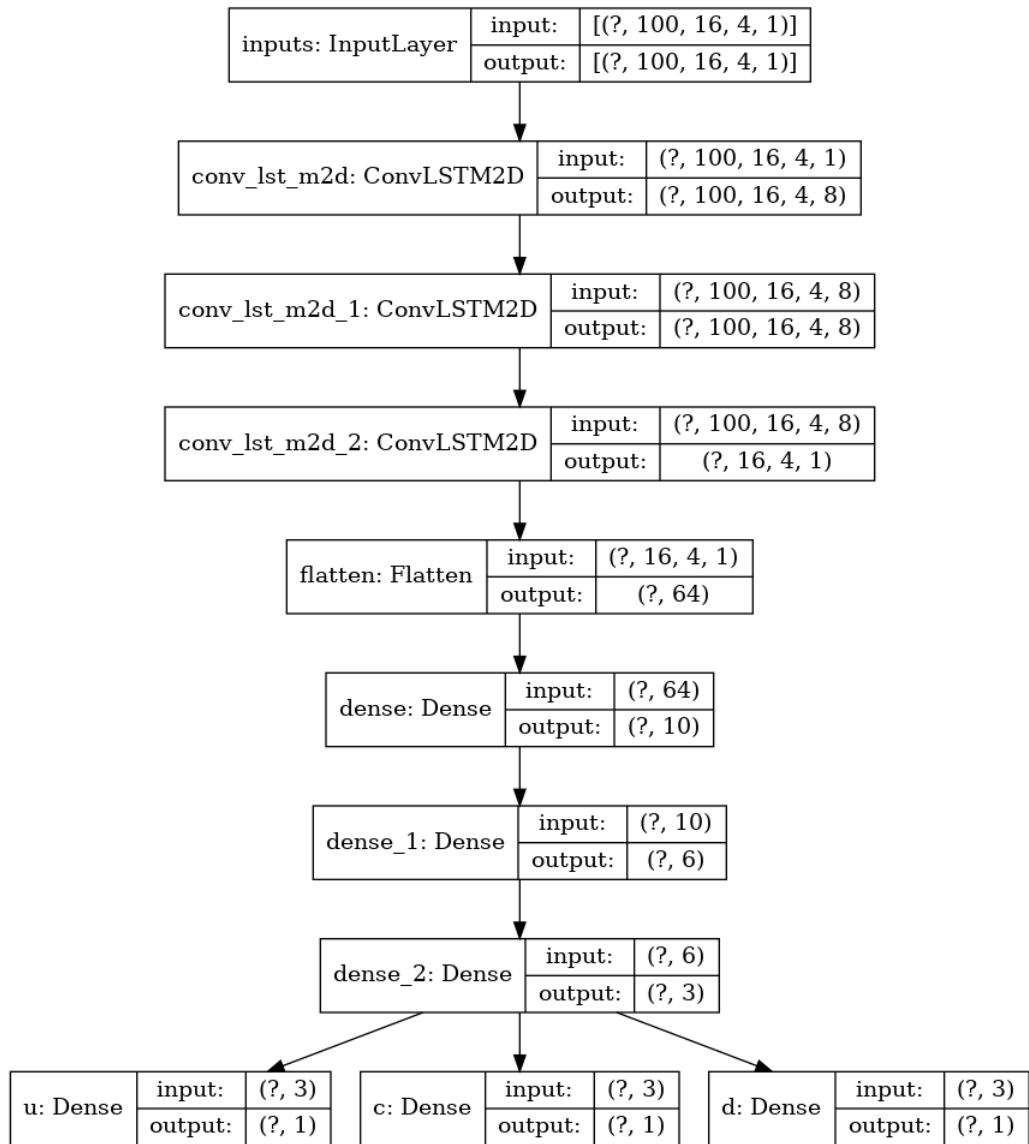


Рис. 2.7.3: Архитектура CNN+LSTM, количество параметров: 14,193.

Глава 3. Проведение оценки качества алгоритмов

3.1 Процедура оценки качества

Для оценки качества использовался весь набор магнитограмм. На его основе проводилась кроссвалидация — цикл оценки решения. На каждой итерации цикла выбиралась одна магнитограмма в качестве отложенной выборки, а на основе оставшихся формировалась тренировочная выборка и проводилось обучение модели. После получения детектирования дефектов результаты сопоставлялись с разметкой экспертов и рассчитывались метрики *Recall*, *Precision* и F_1 — *measure*.

В используемой выборке количество дефектов было распределено неравномерно, т.е. существуют магнитограммы, содержащие порядка 300 дефектов, и магнитограммы с 3 или 5 дефектами, что сильно влияет на результаты измерения качества, поэтому итоговый результат кроссвалидации считается как средневзвешенное по метрикам всех магнитограмм, где веса зависят от количества дефектов в разметке.

3.2 Результаты поиска дефекта по паттерну

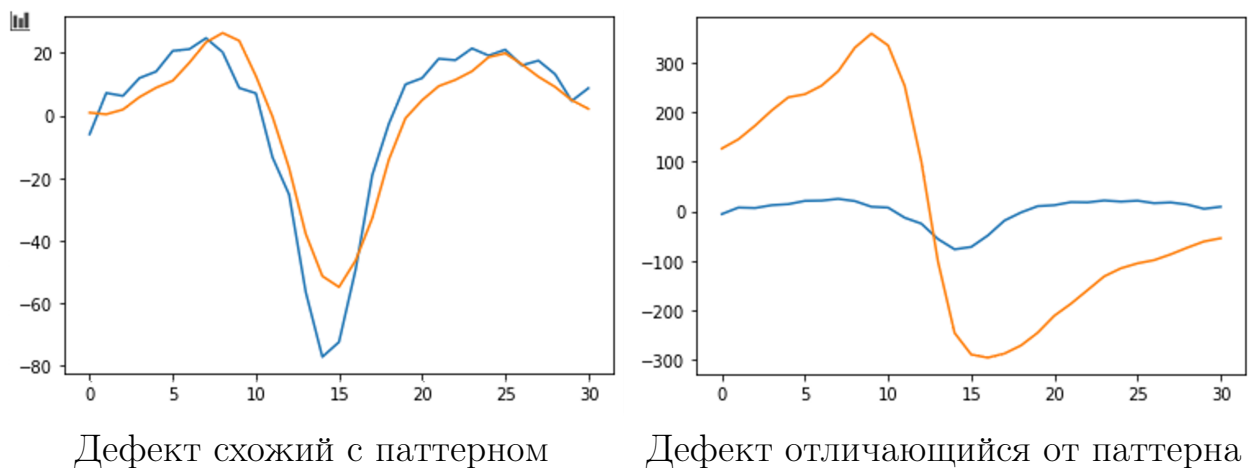
В таблице 3.2.1 представлены результаты кроссвалидации поиска дефекта по паттерну. По результатам проведенного эксперимента заметно, что у выбранного подхода очень низкая точность. Данная ситуация объясняется тем, что магнитограммы помимо дефектов содержат шумы и конструктивные элементы.

Таблица 3.2.1: Результаты поиска дефекта по паттерну

Магнитограмма	<i>Recall</i>	<i>Precision</i>	F_1 — <i>measure</i>
Магнитограмма №5	0,9375	0,0002	0,0004
Магнитограмма №6	0,4118	0,0003	0,0006
Магнитограмма №7	0,7273	0,0009	0,0018
Магнитограмма №8	0,8889	0,0006	0,0012
Магнитограмма №9	0,608	0,0097	0,0191

Магнитограмма №10	1	0,0001	0,0002
Магнитограмма №11	0,85	0,0013	0,0026
Магнитограмма №12	0,7857	0,001	0,002
Магнитограмма №13	0,8367	0,0014	0,0028
Магнитограмма №14	1	0,0015	0,003
Магнитограмма №15	1	0,0007	0,0014
Магнитограмма №16	0,9354	0,0009	0,0018
Магнитограмма №17	1	0,0032	0,0064
Магнитограмма №18	1	0,0011	0,0022
Магнитограмма №19	0,9523	0,0011	0,0022
Магнитограмма №20	1	0,002	0,004
Магнитограмма №21	0,8333	0,0011	0,0022
Магнитограмма №22	0,8981	0,0055	0,0109
Средневзвешенное	0,8298	0,0041	0,0082

На некоторых магнитограммах recall качества поиска меньше единицы, это связано с тем, что существуют дефекты, сильно отличающиеся от паттерна. Пример такого отличия представлен на рисунке 3.2.1.



Дефект схожий с паттерном

Дефект отличающийся от паттерна

Рис. 3.2.1: Сравнение дефектов с паттерном.

3.3 Результаты детектирования методами классического машинного обучения

Результаты детектирования дефектов с помощью ОС-SVM представлены в таблице 3.3.1. Алгоритм корректно детектирует дефекты, но при этом детектирует большое количество шумов и конструктивных элементов, поэтому точность детекции низкая. Для улучшения работы можно обогатить обучающую выборку большим количеством шумов, но есть риск уменьшить полноту детекции дефектов.

Таблица 3.3.1: Результаты детектирования ОС-SVM

Магнитограмма	<i>Recall</i>	<i>Precision</i>	$F_1 - measure$
Магнитограмма №5	1	0,002	0,004
Магнитограмма №6	0,27	0,003	0,0059
Магнитограмма №7	0,74	0,0015	0,003
Магнитограмма №8	0,81	0,003	0,006
Магнитограмма №9	0,55	0,02	0,0386
Магнитограмма №10	1	0,0015	0,003
Магнитограмма №11	0,75	0,003	0,006
Магнитограмма №12	0,67	0,012	0,0236
Магнитограмма №13	0,73	0,0025	0,005
Магнитограмма №14	1	0,002	0,004
Магнитограмма №15	0.85	0,0015	0,003
Магнитограмма №16	0.78	0,003	0,006
Магнитограмма №17	1	0,01	0,0198
Магнитограмма №18	1	0,0011	0,0022
Магнитограмма №19	1	0,0007	0,0014
Магнитограмма №20	1	0,002	0,004
Магнитограмма №21	0,8333	0,01	0,0198
Магнитограмма №22	0,8981	0,0125	0,0247
Средневзвешенное	0,7858	0,0095	0,0188

3.4 Результаты классификации методами классического машинного обучения

В данном эксперименте были рассмотрены три ансамблевые модели на основе деревьев решений. Результаты кроссвалидации представлены в таблицах 3.4.1, 3.4.2 и 3.4.3. Наилучший результат показал градиентный бустинг CatBoost. Он показал высокую полноту и приемлемую точность (на один корректно найденный дефект 9 фантомных дефектов). С учетом, что дефектов редко бывает более 100 на магнитограмме, то провалидировать его работу вручную представляется вполне возможным. Random Forest и Extra Trees выдают схожие между собой результаты при кроссвалидации, это связано с тем, что эти модели основаны на одинаковом принципе ансамблирования — бэккинга. Их результаты практически в два раза проигрывают CatBoost по F_1 — *measure*. Возможно, для увеличения полноты детекции стоит пересмотреть правила с весами оценки моделей во время обучения.

Таблица 3.4.1: Результаты детектирования Extra Trees

Магнитограмма	<i>Recall</i>	<i>Precision</i>	F_1 — <i>measure</i>
Магнитограмма №5	0,24	0,03	0,0533
Магнитограмма №6	0,02	0	0
Магнитограмма №7	0,83	0,003	0,006
Магнитограмма №8	0,43	0,04	0,0732
Магнитограмма №9	0,41	0,2	0,2689
Магнитограмма №10	0,33	0,005	0,0099
Магнитограмма №11	0,6	0,018	0,035
Магнитограмма №12	0,21	0,014	0,0263
Магнитограмма №13	0,06	0,01	0,0171
Магнитограмма №14	0,86	0,011	0,0217
Магнитограмма №15	0,03	0	0
Магнитограмма №16	0,31	0,005	0,0098
Магнитограмма №17	0,25	0,04	0,069

Магнитограмма №18	0,8	0	0
Магнитограмма №19	0,79	0	0
Магнитограмма №20	0,11	0,01	0,0183
Магнитограмма №21	0,07	0,03	0,042
Магнитограмма №22	0,24	0,11	0,1509
Средневзвешенное	0,3378	0,0791	0,1282

Таблица 3.4.2: Результаты детектирования Random Forest

Магнитограмма	<i>Recall</i>	<i>Precision</i>	<i>F₁ – measure</i>
Магнитограмма №5	0,24	0,04	0,0686
Магнитограмма №6	0,04	0,02	0,0267
Магнитограмма №7	0,82	0,003	0,006
Магнитограмма №8	0,55	0,04	0,0746
Магнитограмма №9	0,41	0,2	0,2689
Магнитограмма №10	0,33	0,005	0,0099
Магнитограмма №11	0,7	0,018	0,0351
Магнитограмма №12	0,21	0,014	0,0263
Магнитограмма №13	0,04	0,015	0,0218
Магнитограмма №14	0,86	0,013	0,0256
Магнитограмма №15	0,066	0,005	0,0093
Магнитограмма №16	0,204	0,011	0,0209
Магнитограмма №17	0,23	0,06	0,0952
Магнитограмма №18	0,6	0,05	0,0923
Магнитограмма №19	0,71	0,05	0,0934
Магнитограмма №20	0,11	0,01	0,0183
Магнитограмма №21	0,08	0,03	0,0436
Магнитограмма №22	0,26	0,12	0,1642
Средневзвешенное	0,3479	0,0863	0,1383

Таблица 3.4.3: Результаты детектирования CatBoost

Магнитограмма	<i>Recall</i>	<i>Precision</i>	$F_1 - measure$
Магнитограмма №5	0,34	0,1	0,1545
Магнитограмма №6	0,1	0,08	0,0889
Магнитограмма №7	0,95	0,02	0,0392
Магнитограмма №8	0,67	0,12	0,2035
Магнитограмма №9	0,54	0,37	0,4391
Магнитограмма №10	0,39	0,03	0,0557
Магнитограмма №11	0,9	0,06	0,1125
Магнитограмма №12	0,27	0,03	0,054
Магнитограмма №13	0,11	0,04	0,0587
Магнитограмма №14	0,92	0,03	0,0581
Магнитограмма №15	0,15	0,02	0,0353
Магнитограмма №16	0,18	0,005	0,0097
Магнитограмма №17	0,31	0,13	0,1832
Магнитограмма №18	0,8	0,2	0,32
Магнитограмма №19	0,84	0,15	0,2545
Магнитограмма №20	0,23	0,09	0,1294
Магнитограмма №21	0,17	0,1	0,1259
Магнитограмма №22	0,44	0,12	0,1886
Средневзвешенное	0,4805	0,1398	0,2166

3.5 Результаты нейросетевого подхода

Была проведена оценка качества для нейронных сетей, результаты для используемых архитектур находятся в таблицах 3.5.1, 3.5.2 и 3.5.3. Сеть на основе CNN архитектуры показывают почти всегда лучшие результаты. Наибольшие ошибки возникают при наличии битых или зашумленных данных, как на 7 или 20 магнитограммах. Для нейронных сетей существует нехватка данных. Даже при аугментации датасета сети имеют свойство переобучаться уже на 10 эпохах.

При использовании нескольких отдельных сетей с бинарными выходами для определения каждого класса, переобучение наступает раньше, поэтому в дальнейшем стоит рассматривать предложенную архитектуру с несколькими выходами. Также необходимо изучить влияния весов ошибок для каждого из выходов, т.к. их настройка существенно может изменить результаты. Так, поставив более высокую важность для выхода, классифицирующего дефекты, можно получить лучшие результаты для поиска дефектов.

Таблица 3.5.1: Результаты детектирования CNN

Магнитограмма	<i>Recall</i>	<i>Precision</i>	$F_1 - measure$
Магнитограмма №5	0	0	0
Магнитограмма №6	0,01	0,06	0,0171
Магнитограмма №7	0	0	0
Магнитограмма №8	0,69	0,06	0,1104
Магнитограмма №9	0,02	0,31	0,0376
Магнитограмма №10	0,15	0,005	0,0097
Магнитограмма №11	0,015	0,12	0,0267
Магнитограмма №12	0,24	0,04	0,0686
Магнитограмма №13	0,02	0,005	0,008
Магнитограмма №14	0,41	0,05	0,0891
Магнитограмма №15	0,11	0,33	0,165
Магнитограмма №16	0,23	0,02	0,0368
Магнитограмма №17	0,17	0,05	0,0773
Магнитограмма №18	0,4	0,13	0,1962
Магнитограмма №19	0,11	0,13	0,1192
Магнитограмма №20	0	0	0
Магнитограмма №21	0,05	0,09	0,0643
Магнитограмма №22	0,07	0,23	0,1073
Средневзвешенное	0,0819	0,1663	0,1097

Таблица 3.5.2: Результаты детектирования LSTM

Магнитограмма	<i>Recall</i>	<i>Precision</i>	$F_1 - measure$
Магнитограмма №5	0,26	0,03	0,0538
Магнитограмма №6	0,03	0,03	0,03
Магнитограмма №7	0	0	0
Магнитограмма №8	0,09	0,05	0,0643
Магнитограмма №9	0,11	0,23	0,1488
Магнитограмма №10	0	0	0
Магнитограмма №11	0	0	0
Магнитограмма №12	0,07	0,08	0,0747
Магнитограмма №13	0,1	0,02	0,0333
Магнитограмма №14	0,02	0,02	0,02
Магнитограмма №15	0	0	0
Магнитограмма №16	0,12	0,03	0,048
Магнитограмма №17	0,21	0,06	0,0933
Магнитограмма №18	0,17	0,12	0,1407
Магнитограмма №19	0,27	0,15	0,1929
Магнитограмма №20	0,05	0,14	0,0737
Магнитограмма №21	0,1	0,17	0,1259
Магнитограмма №22	0,08	0,005	0,0094
Средневзвешенное	0,0881	0,0668	0,076

Таблица 3.5.3: Результаты детектирования CNN+LSTM

Магнитограмма	<i>Recall</i>	<i>Precision</i>	$F_1 - measure$
Магнитограмма №5	0,24	0,04	0,0686
Магнитограмма №6	0,04	0,02	0,0267
Магнитограмма №7	0	0	0
Магнитограмма №8	0,02	0,04	0,0267
Магнитограмма №9	0,12	0,25	0,1622

Магнитограмма №10	0	0	0
Магнитограмма №11	0,12	0,11	0,1148
Магнитограмма №12	0,13	0,09	0,1064
Магнитограмма №13	0,07	0,14	0,0933
Магнитограмма №14	0,01	0,02	0,0133
Магнитограмма №15	0	0	0
Магнитограмма №16	0,12	0,03	0,048
Магнитограмма №17	0,18	0,1	0,1286
Магнитограмма №18	0,02	0,1	0,0333
Магнитограмма №19	0,25	0,14	0,1795
Магнитограмма №20	0	0	0
Магнитограмма №21	0,03	0,03	0,03
Магнитограмма №22	0,07	0,11	0,0856
Средневзвешенное	0,0927	0,1187	0,1041

3.6 Выводы

По результатам проведенных экспериментов можно сказать о приемлемом качестве по детекции дефектов подхода с использованием CatBoost. Он показывает хорошие результаты на магнитограммах, не имеющих шумов и «битых» данных. Результаты на примерах, содержащих шумы или некорректные данные, значительно хуже.

Также по результатам экспериментов выдвинуто предположение, что текущая разметка не является полной, так как эксперты не смогли определить все дефекты на магнитограммах в силу зашумления, недостоверности или отсутствия данных.

В дальнейшем для улучшения результатов работы необходимо исключить магнитограммы, содержащие сильные шумы и «битые» данные, увеличить используемую выборку качественными магнитограммами. Такие шаги позволят улучшить качество детекции многих предложенных алгоритмов.

Заключение

Результаты работы

В рамках проделанной работы были выполнены следующие задачи:

- проведен обзор и анализ подходов и методов машинного обучения для детектирования аномалий и дефектов;
- собраны и предобработаны данных внутритрубной диагностики для обучения и оценки качества моделей, написан парсер для расшифровки бинарных файлов, а также проведена аугментация и балансировка датасетов;
- реализованы методы с использованием нейронных сетей и алгоритмов машинного обучения для обнаружения дефектов;
- проведено тестирование и оценка качества разработанных алгоритмов в процессе кроссвалидации на подготовленной выборке;
- полученные результаты проанализированы и сформированы рекомендации по улучшению разработанных подходов в будущем;
- полученное решение интегрировано в программный комплекс для информатизации процесса внутритрубной диагностики;
- получено свидетельство о государственной регистрации программы для ЭВМ [50].

Перспективы развития

На основании проведенных экспериментов для улучшения качества детекции дефектов следует предпринять следующие шаги:

- предложить решение, позволяющее определять состояние магнитогаммы;

- исключить из выборки некорректные магнитограммы;
- обогатить выборку хорошими магнитограммами;
- перепроверить текущую разметку магнитограмм с помощью экспертов;
- расширить признаковое описание дефектов на основе работы CNN сетей.
- усложнить архитектуру используемых сетей CNN и добавить Batch Normalization слои;
- рассмотреть возможность использования autoencoders в данной задаче.

Список литературы

- [1] Алексеев, А. Точки инновационного роста /А. Алексеев // «Сибирская нефть». — 2019. — №161. — С. 46–52.
- [2] Статья ТАСС «Ачимовка — нефть будущего»: сайт. — URL: <https://spec.tass.ru/achimovka/> (дата обращения: 05.05.2022). — Текст: электронный.
- [3] Шалай, В.В. Анализ технического состояния объектов линейной част магистральных Нефтепроводов, определение оптимальных способов поддержания объектов линейной части в нормативном состоянии / В.В. Шалай, М.М. Васильев, К.А. Шумаков // «Омский научный вестник»— Омск, 2004. — С. 196–199.
- [4] Braei, Mohammad & Wagner, Sebastian. Anomaly Detection in Univariate Time-series: A Survey on the State-of-the-Art. 2020.
- [5] A. H. Yaacob, I. K. T. Tan, S. F. Chien and H. K. Tan, "ARIMA Based Network Anomaly Detection,"2010 Second International Conference on Communication Software and Networks, 2010, pp. 205-209.
- [6] Jašek, R., Szmit, A., Szmit, M. (2013). Usage of Modern Exponential-Smoothing Models in Network Traffic Modelling. In: Zelinka, I., Chen, G., Rössler, O., Snasel, V., Abraham, A. (eds) Nostradamus 2013: Prediction, Modeling and Analysis of Complex Systems. Advances in Intelligent Systems and Computing, vol 210. Springer, Heidelberg.
- [7] Yufeng Yu, Yuelong Zhu, Shihua Li, and Dingsheng Wan. Time series outlier detection based on sliding window prediction. In Mathematical Problems in Engineering, 2014.
- [8] J. MacQueen. Some Methods for Classification and Analysis of Multivariate Observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, pages 281–297, Berkeley, Calif., 1967. University of California Press.

- [9] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96, pages 226–231. AAAI Press, 1996.
- [10] M. Çelik, F. Dadaşer-Çelik, and A. Ş. Dokuz. Anomaly detection in temperature data using dbscan algorithm. In 2011 International Symposium on Innovations in Intelligent Systems and Applications, pages 91–95, 06 2011.
- [11] Fei Tony Liu, Kai Ming Ting, and Zhi hua Zhou. Isolation forest. In ICDM '08: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining. IEEE Computer Society, pages 413–422, 2008.
- [12] V.N. Vapnik and A.Ya Chervonenkis. A class of algorithms for pattern recognition learning. *Avtomat. i Telemekh.*,25:937–945, 1964.
- [13] Bernhard Schölkopf, Robert Williamson, Alex Smola, John Shawe-Taylor, and John Platt. Support vector method for novelty detection. In Proceedings of the 12th International Conference on Neural Information Processing Systems, NIPS'99, pages 582–588, Cambridge, MA, USA, 1999. MIT Press.
- [14] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.
- [15] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2 edition, 2009.
- [16] R.J. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 2014.
- [17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C.

- J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR*, abs/1505.04597, 2015.
- [19] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object Detection with Deep Learning: A Review. *CoRR*, abs/1807.05511, 2018.
- [20] Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and J. Leon Zhao. Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks. In Feifei Li, Guoliang Li, Seung-won Hwang, Bin Yao, and Zhenjie Zhang, editors, *Web-Age Information Management*, pages 298–310, Cham, 2014. Springer International Publishing.
- [21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *CoRR*, abs/1512.03385, 2015.
- [22] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. WaveNet: A Generative Model for Raw Audio. *arXiv*, 9 2016.
- [23] S. Chauhan and L. Vig. Anomaly detection in ECG Time Signals via Deep Long Short-Term Memory Networks. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–7, 10 2015.
- [24] Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. Long Short Term Memory Networks for Anomaly Detection in Time Series. In *ESANN*, 2015.
- [25] Kyunghyun Cho, Bart van Merriënboer, Çağlar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *CoRR*, abs/1406.1078, 2014.

- [26] Wentai Wu, Ligang He, and Weiwei Lin. Local Trend Inconsistency: A Prediction-driven Approach to Unsupervised Anomaly Detection in Multi-seasonal Time Series. ArXiv, abs/1908.01146, 2019.
- [27] Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Deep Autoencoding Models for Unsupervised Anomaly Segmentation in Brain MR Images. CoRR, abs/1804.04488, 2018.
- [28] Chong Zhou and Randy C. Paffenroth. Anomaly Detection with Robust Deep Autoencoders. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17, pages 665–674, New York, NY, USA, 2017. ACM.
- [29] Mayu Sakurada and Takehisa Yairi. Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction. In Proceedings of the MLSDA 2014 2Nd Workshop on Machine Learning for Sensory Data Analysis, MLSDA'14, pages 4:4–4:11, New York, NY, USA, 2014. ACM.
- [30] Park, JK., Kwon, BK., Park, JH. et al. Machine learning-based imaging system for surface defect inspection. Int. J. of Precis. Eng. and Manuf.-Green Tech. 3, 303–310 (2016).
- [31] Y. Shi, Y. Li, X. Wei and Y. Zhou, "A Faster-RCNN Based Chemical Fiber Paper Tube Defect Detection Method,"2017 5th International Conference on Enterprise Systems (ES), 2017, pp. 173-177.
- [32] L. Xiao-Cong, "A Hybrid SVM-QPSO Model Based Ceramic Tube Surface Defect Detection Algorithm,"2014 Fifth International Conference on Intelligent Systems Design and Engineering Applications, 2014, pp. 28-31.
- [33] Chatterjee, Sromona & Saeedfar, Pouya & Tofangchi, Schahin & Kolbe, Lutz. (2018). INTELLIGENT ROAD MAINTENANCE: A MACHINE LEARNING APPROACH FOR SURFACE DEFECT DETECTION.

- [34] Официальный сайт «Baker Hughes»: сайт. — URL: <https://www.bakerhughes.com/> (дата обращения: 06.03.2020). — Текст: электронный.
- [35] Официальный сайт «Интрон плюс»: сайт. — URL: <https://www.intron.ru/> (дата обращения: 06.03.2020). — Текст: электронный.
- [36] ГОСТ Р 55999-2014 Внутритрубное техническое диагностирование газопроводов. Общие требования. — 2015.02.01 — с.7.
- [37] Методические указания по организации и исполнению программ диагностики промысловых трубопроводов компании ОАО «Газпром Нефть». — 2017. — с.43.
- [38] Официальный сайт документации Kaitai Struct: сайт. — URL: <https://kaitai.io/> (дата обращения: 10.05.2022). — Текст: электронный.
- [39] Официальный сайт документации PostgreSQL: сайт. — URL: <https://www.postgresql.org/> (дата обращения: 10.05.2022). — Текст: электронный.
- [40] Salvador, Stan & Chan, Philip. (2004). Toward Accurate Dynamic Time Warping in Linear Time and Space. *Intelligent Data Analysis*. 11. 70–80.
- [41] Официальный сайт документации Scipy: сайт. — URL: <https://www.scipy.org/> (дата обращения: 10.05.2022). — Текст: электронный.
- [42] Официальный сайт документации PyOD: сайт. — URL: <https://pyod.readthedocs.io/en/latest/> (дата обращения: 10.05.2022). — Текст: электронный.
- [43] Mariana Belgiu, Lucian Drăguț, Random forest in remote sensing: A review of applications and future directions, *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 114, 2016, Pages 24–31

- [44] Shafique, Rahman & Mehmood, Arif & Ullah, Dr. Saleem & Choi, Gyu Sang. (2019). Cardiovascular Disease Prediction System Using Extra Trees Classifier. 10.21203/rs.2.14454/v1.
- [45] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. 2018. CatBoost: unbiased boosting with categorical features. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18). Curran Associates Inc., Red Hook, NY, USA, 6639–6649.
- [46] Официальный сайт документации Scikit-Learn: сайт. — URL: <https://scikit-learn.org/> (дата обращения: 10.05.2022). — Текст: электронный.
- [47] Официальный сайт документации CatBoost: сайт. — URL: <https://catboost.ai/> (дата обращения: 10.05.2022). — Текст: электронный.
- [48] Официальный сайт документации Imbalanced-Learn: сайт. — URL: <https://imbalanced-learn.org/stable/> (дата обращения: 10.05.2022). — Текст: электронный.
- [49] Официальный сайт документации Tensorflow: сайт. — URL: <https://www.tensorflow.org/> (дата обращения: 10.05.2022). — Текст: электронный.
- [50] Свидетельство 2022619105. ИНТЕЛЛЕКТУАЛЬНАЯ РАСШИФРОВКА МАГНИТОГРАММ: программа для ЭВМ / Л. А. Коваленко, Д. В. Терещенко, А. А. Бабурин и др. (RU) ; ПУБЛИЧНОЕ АКЦИОНЕРНОЕ ОБЩЕСТВО «ГАЗПРОМ НЕФТЬ» (ПАО «ГАЗПРОМ НЕФТЬ») (RU) — № 2022618606 ; заявл. 12.05.2022 ; опубл. 19.05.2022. 5 Мб.