

Костанайский региональный университет имени Ахмет Байтұрсынұлы

УДК 004.891.3

На правах рукописи

РЫСПАЕВА МАРЬЯ КУАНЫШЕВНА

**Искусственный интеллект для обнаружения рака груди с использованием
глубоких сверточных нейронных сетей**

8D06102 – Информационные технологии и робототехника

Диссертация на соискание степени
доктора философии (PhD)

Научный консультант
доктор PhD,
ассоциированный профессор
О.С. Салыкова

Зарубежный научный консультант
доктор PhD,
Kristoffer Johansen
(Норвегия)

Республика Казахстан
Костанай, 2026

СОДЕРЖАНИЕ

НОРМАТИВНЫЕ ССЫЛКИ.....	4
ОПРЕДЕЛЕНИЯ.....	5
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ.....	7
ВВЕДЕНИЕ.....	9
1 СОВРЕМЕННОЕ СОСТОЯНИЕ ИССЛЕДОВАНИЙ ПО ОБНАРУЖЕНИЮ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ.....	19
1.1 Медицинские аспекты ранней диагностики рака молочной железы....	19
1.2 Сверточные нейронные сети в задачах медицинской диагностики и обнаружения рака молочной железы.....	21
1.3 Проблема ограниченности и дисбаланса медицинских данных.....	22
1.4 Генеративные модели для синтеза медицинских изображений.....	24
1.5 Интеграция генеративных и классификационных подходов: современные решения и их ограничения.....	26
Выводы по первому разделу.....	28
2 МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССА И РАЗРАБОТКА МЕТОДА ОБНАРУЖЕНИЯ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ.....	30
2.1 Архитектуры сверточных нейронных сетей и их адаптация к классификации рака молочной железы.....	30
2.2 Transfer learning: использование предобученных весов.....	33
2.3 Математическая модель генеративных сетей: от классического GAN к WGAN и WGAN-GP.....	36
2.4 Датасеты и экспериментальные данные.....	40
2.5 Интеграция генеративной модели и feature extractor в критике (двухветвевая архитектура DGAN-WP-TL).....	43
2.6 Классификация рака молочной железы с использованием сверточных нейронных сетей и синтетически расширенного набора данных.....	47
2.7 Метрики оценки качества синтетических изображений и классификации.....	50
Выводы по второму разделу.....	53
3 ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ И РЕЗУЛЬТАТЫ..	55
3.1 Экспериментальные исследования генеративной модели DGAN-WP- TL.....	55
3.1.1 Результаты DGAN-WP-TL на датасете BUSI.....	56
3.1.2 Результаты DGAN-WP-TL на датасете CBIS-DDSM.....	58
3.1.3 Дополнительная апробация метода DGAN-WP-TL на датасете Alzheimer MRI.....	60
3.1.4 Анализ распределений синтетических и реальных изображений с использованием PCA и t-SNE.....	62
3.1.5 Сравнение с baseline-архитектурой StyleGAN2.....	65
3.2 Влияние предобученных весов ImageNet и RadImageNet на генерацию	67
3.2.1 Количественный анализ качества генерации: распределительные, перцептивные и структурные метрики.....	68

3.2.2 Качественные результаты и анализ признакового пространства.....	70
3.3 Абляционное исследование архитектуры DGAN-WP-TL.....	72
3.3.1 Влияние архитектуры критика и функции потерь.....	72
3.3.2 Анализ устойчивости и вычислительной эффективности DGAN-WP-TL при малом объеме данных.....	74
3.4 Классификация рака молочной железы на расширенной выборке.....	76
3.4.1 Оценка влияния синтетических данных на классификацию.....	76
3.4.2 Классификация BUSI с использованием моделей на предобученных весах ImageNet.....	79
3.4.3 Классификация CBIS-DDSM на основе синтетических изображений, сгенерированных DGAN-WP-TL (VGG19 и DenseNet121 в качестве перцептивных бэкбонов с ImageNet весами).....	83
3.4.4 Классификация CBIS-DDSM с использованием DGAN-WP-TL + RadImageNet-весов (DenseNet121 и InceptionV3 бэкбоны).....	86
Выводы по третьему разделу.....	89
4 АНАЛИЗ И ОБОБЩЕНИЕ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ....	91
4.1 Устойчивость и качество генерации: роль архитектуры, функции потерь и режима обучения.....	91
4.2 Сравнительный анализ: StyleGAN2 и влияние предобучения и перцептивного бэкбона.....	93
4.3 Сравнение предложенного метода DGAN-WP-TL с существующими подходами.....	94
4.4 Анализ влияния синтетических данных на качество классификации опухолей.....	97
4.5 Сравнительный анализ подходов к классификации РМЖ с использованием GAN-аугментации.....	98
4.6 Ограничения предложенного метода, практические рекомендации и направления дальнейших исследований.....	103
Выводы по четвертому разделу.....	105
ЗАКЛЮЧЕНИЕ.....	107
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	109
ПРИЛОЖЕНИЕ А – Акт внедрения.....	117
ПРИЛОЖЕНИЕ Б – Авторские свидетельства.....	120
ПРИЛОЖЕНИЕ В – Обзор генеративных моделей и их применения в диагностике рака молочной железы.....	123

НОРМАТИВНЫЕ ССЫЛКИ

В настоящей диссертации использованы ссылки на следующие стандарты: Инструкция по оформлению диссертации и автореферата, ВАК МОН РК, №377-3ж.

ГОСТ 7.32-2017. Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления.

СТ РК 34.005-2002. Информационная технология. Основные термины и определения.

СТ РК 34.006-2002. Информационная технология. Базы данных. Основные термины и определения.

СТ РК 34.014-2002. Информационная технология. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Термины и определения.

СТ РК 34.019-2005. (ISO/IEC 12207:1995, MOD) Информационная технология. Процессы жизненного цикла программных средств.

Указ Президента Республики Казахстан. О Государственной программе «Информационный Казахстан-2020» и внесении дополнения в Указ Президента Республики Казахстан от 19 марта 2010 года, №957 «Об утверждении Перечня государственных программ»: утв. 8 января 2013 года, №464.

Закон Республики Казахстан. Об информатизации: принят 24 ноября 2015 года, №418.

Постановление Республики Казахстан. Об утверждении Государственной программы «Цифровой Казахстан»: утв. 12 декабря 2017 года, №827.

Закон Республики Казахстан. О науке: принят 18 февраля 2011 года, №407-IV ЗРК.

ГОСТ 7.1-2003. Библиографическая запись. Библиографическое описание. Общие требования и правила составления.

ОПРЕДЕЛЕНИЯ

В настоящей диссертации применяются следующие термины с соответствующими определениями:

Абляционное исследование — метод анализа, заключающийся в последовательном исключении или изменении отдельных компонентов модели для оценки их вклада в итоговый результат.

Генератор — нейронная сеть в составе генеративно-состязательной архитектуры, предназначенная для синтеза изображений, стремящаяся формировать данные, неотличимые от реальных.

Градиентный штраф — метод регуляризации, применяемый в генеративных моделях WGAN-GP, заключающийся в добавлении к функции потерь слагаемого, ограничивающего норму градиента дискриминатора и обеспечивающего выполнение условия Липшица.

Дисбаланс данных — ситуация, при которой классы в наборе данных представлены неравномерно, что приводит к снижению качества распознавания редких классов.

Дискриминатор (критик) — нейронная сеть в составе генеративно-состязательной архитектуры, задача которой заключается в различении реальных изображений и изображений, сгенерированных генератором.

Искусственный интеллект — это область науки и техники, занимающаяся теорией и практикой разработки систем, которые способны имитировать интеллектуальную деятельность человека. В рамках данной работы под ИИ понимаются методы глубокого обучения, обеспечивающие автоматизированный анализ медицинских изображений и выявление злокачественных случаев.

Медицинское изображение — цифровое изображение, полученное методами медицинской визуализации (ультразвуковое исследование, маммография, магнитно-резонансная томография и др.), используемое для диагностики заболеваний.

Метод DGAN-WP-TL — предложенный в диссертации метод генеративного синтеза данных, основанный на принципах Wasserstein GAN с градиентным штрафом, использовании перцептивных признаков и обучении с transfer learning.

Миноритарный класс (minority class) — класс в наборе данных, представленный существенно меньшим количеством примеров по сравнению с другими классами (например, злокачественные опухоли).

Модель — это формализованное представление объекта, процесса или системы, реализованное в виде математической и/или вычислительной конструкции, предназначенной для анализа, аппроксимации и воспроизведения их свойств и закономерностей.

Перцептивные признаки — признаки изображений, извлекаемые предобученными сверточными нейронными сетями и используемые для оценки структурного и текстурного сходства изображений.

Предобученные веса – параметры нейронной сети, полученные в результате обучения на крупном наборе данных и используемые для инициализации модели при решении новой задачи.

Рак молочной железы (РМЖ) – злокачественное новообразование, развивающееся из ткани молочной железы.

Сверточная нейронная сеть – искусственная нейронная сеть, использующая свёрточные операции для автоматического извлечения признаков из изображений и решения задач классификации и/или обнаружения (детектирования) патологий.

Синтетические данные – искусственно созданные изображения, сгенерированные нейронной сетью, используемые для увеличения и балансировки обучающих выборок.

Чувствительность (Recall, полнота, Sensitivity) – доля корректно выявленных злокачественных случаев среди всех истинно злокачественных образцов; ключевая метрика в задачах обнаружения рака молочной железы.

Downstream-классификация – задача классификации, выполняемая с использованием данных или признаков, сформированных на предыдущем этапе обработки; в данной работе – после синтетического расширения обучающей выборки.

Generative Adversarial Network (GAN) – архитектура нейронной сети, включающая генератор и дискриминатор, обучающиеся в противостоянии для синтеза реалистичных данных.

ImageNet – крупный набор изображений общего назначения, широко используемый для предобучения сверточных нейронных сетей.

RadImageNet – специализированный набор медицинских изображений, используемый для предобучения моделей в задачах анализа медицинских данных.

Transfer learning – методика обучения, при которой модель использует предварительно обученные веса для ускорения и повышения качества обучения на новых данных.

Wasserstein GAN with Gradient Penalty (WGAN-GP) – модификация WGAN, использующая градиентный штраф для обеспечения устойчивости обучения и предотвращения коллапса мод.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

PMЖ	– рак молочной железы
AUC	– Area Under the Curve/площадь под ROC-кривой, интегральная метрика качества бинарной классификации
BUSI	– Breast Ultrasound Images Dataset/набор данных ультразвуковых изображений молочной железы
CBIS-DDSM	– Curated Breast Imaging Subset of Digital Database for Screening Mammography/курированный набор маммографических изображений с аннотациями опухолей молочной железы
CNN/CHC	– Convolutional Neural Network/сверточная нейронная сеть
DGAN-WP-TL	– предложенный в диссертации метод генеративного синтеза данных, основанный на принципах WGAN-GP и обучения с переносом (transfer learning)
FID	– Fréchet Inception Distance/метрика качества генерации изображений, основанная на сравнении распределений признаков реальных и синтетических изображений
F1-score	– гармоническое среднее точности (Precision) и полноты (Recall, чувствительность) классификации
GAN	– Generative Adversarial Network/генеративно-сопоставительная сеть
GP	– Gradient Penalty /градиентный штраф, регуляризатор для критика в модели WGAN-GP
KID	– Kernel Inception Distance/метрика качества генерации изображений, основанная на непараметрическом сравнении распределений признаков
LPIPS	– Learned Perceptual Image Patch Similarity/метрика перцептивного сходства изображений
MRI	– Magnetic Resonance Imaging/магнитно-резонансная томография
MS-SSIM	– Multi-Scale Structural Similarity Index Measure /мульти-масштабный индекс структурного сходства изображений
PCA	– Principal Component Analysis/метод главных компонент, используемый для анализа распределений признаков
ROC	– Receiver Operating Characteristic/кривая рабочих характеристик классификатора, отображающая зависимость между чувствительностью (True Positive Rate) и долей ложноположительных срабатываний (False Positive Rate) при изменении порога принятия решения
SOTA	– State-of-the-art /лучшие на текущий момент методы и результаты в научной области
t-SNE	– t-distributed Stochastic Neighbor Embedding/метод нелинейного снижения размерности для визуализации признаков

TL	– Transfer Learning/обучение с переносом
WGAN	– Wasserstein GAN /генеративно-сопязательная сеть с функцией потерь на основе расстояния Вассерштейна
WGAN-GP	– Wasserstein GAN with Gradient Penalty/модификация WGAN с использованием градиентного штрафа для обеспечения устойчивости обучения
λ	– коэффициент градиентного штрафа WGAN-GP

ВВЕДЕНИЕ

Проблема и актуальность темы исследования. Рак молочной железы (РМЖ) является одной из наиболее значимых онкологических проблем современного здравоохранения и занимает первое место среди злокачественных новообразований у женщин. По данным международной статистики ежегодно выявляется свыше двух миллионов новых случаев РМЖ, а число смертей в 2022 году превысило 700 тысяч [1-4]. В Республике Казахстан ежегодно регистрируется более 5 тысяч новых случаев РМЖ, при этом сохраняется проблема поздней диагностики: по оценкам, в 10-15% случаев заболевание выявляется на поздних стадиях, что ухудшает прогноз и повышает стоимость лечения [5-11].

Ключевая роль в раннем выявлении РМЖ принадлежит методам медицинской визуализации, включая ультразвуковое исследование (УЗИ) и маммографию. Однако интерпретация изображений зависит от человеческого фактора и качества протоколов, что стимулирует развитие методов искусственного интеллекта (ИИ), в частности сверточных нейронных сетей (СНС, CNN), показавших высокую эффективность в задачах автоматизированной классификации медицинских изображений и обнаружения патологий [12-14].

В практических сценариях применение CNN в диагностике РМЖ существенно ограничивается дефицитом и дисбалансом клинических данных, при котором злокачественные случаи представлены значительно реже, а формирование качественной разметки требует значительных экспертных ресурсов.

Традиционные методы аугментации (повороты, отражения, масштабирование) лишь частично компенсируют дисбаланс и не обеспечивают требуемого клинического разнообразия. Поэтому практической задачей является повышение чувствительности CNN-классификаторов к злокачественному (malignant) классу. В качестве вспомогательного инструмента рассматриваются генеративные модели (GAN), позволяющие формировать реалистичные синтетические медицинские изображения и восполнять дефицит миноритарного класса [12; 13; 14, p. 101552; 15-17].

В рамках настоящей диссертации генеративный синтез рассматривается не как самостоятельная цель, а как механизм повышения качества автоматизированного обнаружения РМЖ с использованием CNN. Для медицинских данных критично применение устойчивых схем обучения (WGAN-GP) и анализ влияния TL и источников предобучения (ImageNet, RadImageNet) на устойчивость синтеза и итоговые показатели downstream-классификации [12; 13; 14, p. 101552; 15, p. 609-616; 16; 17, p. 1-10]. Таким образом, актуальность исследования определяется необходимостью разработки и экспериментальной валидации подхода, повышающего эффективность обнаружения РМЖ за счёт синтетического расширения обучающей выборки и роста чувствительности CNN к злокачественному классу.

Степень изученности и научной разработанности темы исследования. Проблема ранней диагностики РМЖ с использованием методов ИИ активно исследуется в последние годы. Наибольшее распространение получили CNN, которые позволяют автоматически извлекать информативные признаки и выполнять классификацию изображений без необходимости ручного выделения признаков.

Основы применения глубоких нейронных сетей в медицинской диагностике заложены в ряде фундаментальных исследований, включая работы McKinney et al. [18], Yala et al. [19], Ehteshami Bejnordi et al. [20, 21], Janowczyk et al. [22], Geras et al. [23], а также классические архитектурные решения Krizhevsky et al. [24], Simonyan et al. [25], Szegedy et al. [26], He et al. [27] и др. Эти исследования заложили основу современного подхода к анализу медицинских изображений с помощью CNN, продемонстрировав, что такие модели способны достигать уровня квалифицированных радиологов при классификации опухолей молочной железы.

Несмотря на прогресс, практическая применимость CNN в медицинской диагностике ограничивается проблемой дисбаланса классов и дефицита данных. В связи с этим развивается направление использования GAN для генерации синтетических медицинских изображений, особенно в условиях малых выборок. Базовые принципы GAN предложены Goodfellow et al., а дальнейшее развитие устойчивых схем обучения представлено в работах Arjovsky et al. [28] и Gulrajani et al. (WGAN-GP). Показано, что устойчивые модификации GAN могут быть полезны как инструмент синтетического пополнения данных и повышения эффективности downstream-классификации в медицинских задачах [14, p. 101552; 26, p. 1-8; 27, p. 770-777; 28].

Множество исследований таких авторов, как Yi et al. [14], Wang et al. [15], Shin et al. [17], Chen et al. [29], Liu et al. [30], Li et al. [31] используют GAN в качестве инструмента для устранения дисбаланса данных, а также отмечают улучшение в результатах классификации на медицинских наборах данных. В области РМЖ исследования проводились такими учеными, как Liu et al. [30], Alruily et al. [32], Luo et al. [33], Jimenez-Gaona et al. [34, 35], Garuccho et al. [36], Park et al. [37], Oyelade et al. [38], Na et al. [39], Prodan et al. [39], Joseph et al. [41], Xing et al. [42], которые применяют генеративные модели также для балансировки датасетов с целью улучшения диагностики заболеваний рака груди. Однако в перечисленных работах присутствуют пробелы в исследованиях, такие как отсутствие комплексного анализа метрик генерации и классификации, информации о доле синтетического расширения медицинских датасетов и ее влиянии на итоговую задачу исследования. Кроме того, улучшение результатов достигается также за счет применения трансферного обучения с интеграцией в CNN, что позволяет повысить устойчивость методов и подтверждается исследованиями Raghu et al. [43], Mei et al. [44], Frid-Adar et al. [45] и Yang et al. [46] с применением предобученных моделей. Однако остаются недостаточно изученными вопросы архитектурной интеграции предобученных CNN в генеративные модели и условия, при которых синтетические данные

действительно улучшают качество классификации, а не приводят к переобучению или снижению обобщающей способности моделей.

В странах СНГ, включая Республику Казахстан, также наблюдается рост интереса к применению методов ИИ в медицинской визуализации, прежде всего в задачах автоматизированной классификации и сегментации изображений. Вместе с тем большинство отечественных исследований сосредоточено на непосредственном построении и оптимизации классификаторов, тогда как проблема ограниченности и дисбаланса медицинских данных зачастую рассматривается лишь косвенно. Подходы, основанные на использовании генеративных моделей, как вспомогательного инструмента для синтетического расширения обучающих выборок и повышения чувствительности классификационных моделей к редким патологическим классам, в отечественной научной литературе представлены ограниченно. Это подчёркивает актуальность и научную новизну данной работы, ориентированной на РМЖ с применением CNN в условиях дефицита и дисбаланса данных.

Проведённый анализ литературы позволил выявить ряд пробелов:

- ограниченная репрезентативность и дисбаланс клинических выборок, ведущие к снижению чувствительности CNN к злокачественному классу;
- необходимость устойчивой генерации синтетических изображений в условиях малых медицинских датасетов;
- недостаточная систематизация условий, при которых синтетическое расширение выборки действительно повышает качество классификации, а не приводит к переобучению или снижению обобщающей способности моделей.

Объект диссертационного исследования – системы автоматизированной диагностики рака молочной железы на основе глубоких сверточных нейронных сетей.

Предмет диссертационного исследования – закономерности повышения диагностической эффективности CNN-классификаторов рака молочной железы при обучении на ограниченных и несбалансированных медицинских данных в условиях их генеративного синтетического расширения.

Целью данной диссертации является разработка и экспериментальная валидация метода автоматизированного обнаружения РМЖ на медицинских изображениях с использованием CNN, устойчивого к дефициту и дисбалансу данных за счёт синтетического расширения обучающей выборки с применением генеративной модели DGAN-WP-TL.

Задачи диссертационного исследования, необходимые для достижения поставленной цели:

1. Провести анализ современных методов применения сверточных нейронных сетей и генеративных моделей в задачах обнаружения рака молочной железы в условиях дефицита и дисбаланса данных.
2. Обосновать применение генеративных моделей для синтетического расширения обучающих медицинских данных.
3. Разработать генеративную модель DGAN-WP-TL с двухветвевой архитектурой критика для генерации синтетических изображений миноритарного класса.

4. Разработать метод обучения CNN-классификаторов, основанный на управляемом формировании обучающей выборки за счёт варьирования доли синтетических данных.

5. Разработать методику оценки синтетических данных и провести экспериментальную апробацию предложенного подхода с анализом влияния синтетического расширения на качество классификации.

Материалы исследования

Экспериментальные исследования проведены на открытых наборах медицинских изображений, соответствующих задачам классификации опухолей молочной железы в различных модальностях:

1. BUSI (Breast Ultrasound Images Dataset) – набор ультразвуковых изображений молочной железы с выраженным дисбалансом между benign и malignant классами. [47].

2. CBIS-DDSM (Curated Breast Imaging Subset of Digital Database for Screening Mammography) – набор маммографических изображений с аннотациями для доброкачественных и злокачественных образований. [48].

Для того чтобы проверить устойчивость предложенного метода DGAN-WP-TL, эксперименты были дополнительно проведены на Alzheimer MRI [49] датасете для междоменной апробации на малых медицинских выборках за пределами РМЖ.

Методы исследования

В основе диссертационной работы лежат CNN и GAN, относящиеся к методам ИИ и глубокого обучения с дополнительной интеграцией трансферного обучения в CNN и GAN, которые были использованы для классификации и генерации медицинских изображений. Трансферное обучение применялось на базе предобученных весов ImageNet и RadImageNet. Для разработки метода DGAN-WP-TL применен метод оптимизации, а именно Wasserstein GAN (WGAN-GP), что позволило обеспечить стабильное обучение генерации и отсутствие коллапса мод. Генеративная модель в рамках работы использовалась как вспомогательный инструмент синтетического расширения обучающей выборки для последующего повышения эффективности CNN-классификаторов.

Для объективной оценки качества синтетических изображений применялись метрики статистического и перцептивного сходства, включая Fréchet Inception Distance (FID) [50] и Kernel Inception Distance (KID) [51], а также Learned Perceptual Image Patch Similarity (LPIPS) [52] и Multi-Scale Structural Similarity Index (MS-SSIM) [53].

Для анализа пространственного распределения признаков и визуализации взаимного расположения реальных и синтетических изображений в признаковом пространстве использовались методы снижения размерности, включая Principal Component Analysis (PCA) [54] и t-distributed Stochastic Neighbor Embedding (t-SNE) [55]. Эти методы позволили исследовать степень согласованности распределений признаков и подтвердить статистическую состоятельность синтетических данных.

Для оценки эффективности автоматизированного обнаружения опухолей молочной железы использовались CNN, обучаемые на исходных и синтетически

расширенных обучающих выборках. Классификация выполнялась в постановке бинарного разделения изображений на доброкачественные и злокачественные, с последующим сравнительным анализом влияния синтетических данных на качество работы моделей. Эффективность классификации оценивалась по клинически значимым метрикам Recall, F1-score и AUC, позволяющим количественно охарактеризовать чувствительность моделей к злокачественному классу и общую разделимость классов [56-58].

Экспериментальные исследования проводились на медицинских наборах данных, относящихся к задачам визуальной диагностики РМЖ, включая BUSI (УЗИ) и CBIS-DDSM (маммографические изображения). Дополнительно использовался набор Alzheimer MRI для междоменной апробации метода и оценки его устойчивости в условиях малых медицинских выборок другого типа.

Программная реализация экспериментальных исследований выполнена в средах TensorFlow и PyTorch, что обеспечило воспроизводимость результатов, гибкую настройку гиперпараметров и возможность проведения сравнительных экспериментов.

Научная новизна диссертационного исследования

Научная новизна диссертационного исследования заключается в разработке комплексного подхода к автоматизированному обнаружению рака молочной железы, включающего генерацию, использование и оценку синтетических медицинских данных. Предложена генеративная модель DGAN-WP-TL с двухветвевой архитектурой критика, в которой за счёт интеграции предобученного и доменно-адаптированного признаков пространств обеспечивается более точная аппроксимация распределения медицинских изображений и устойчивость обучения генеративной модели. Разработан метод обучения CNN-классификаторов, в котором реализовано управляемое формирование обучающей выборки за счёт варьирования доли синтетических данных миноритарного класса, что позволяет целенаправленно изменять распределение данных и повышать диагностическую чувствительность модели. Разработана методика оценки синтетических данных, основанная на сочетании распределительных, перцептивных и структурных метрик, а также анализе их влияния на результаты классификации, что позволяет выявлять коллапс мод и оценивать диагностическую значимость синтетических изображений.

Основные научные положения, выносимые на защиту и обладающие признаками научной новизны:

1. Генеративная модель DGAN-WP-TL с двухветвевой структурой критика, в которой объединяются предобученные и доменно-адаптированные признаки, что обеспечивает более точную оценку реалистичности синтетических изображений и повышает качество их генерации

2. Метод обучения классификатора сверточной нейронной сети для обнаружения злокачественных новообразований, в котором реализовано управляемое формирование обучающей выборки за счёт варьирования доли синтетических данных, что обеспечивает контролируемое изменение распределения обучающей выборки и приводит к повышению диагностической чувствительности (Recall) модели.

3. Методика оценки синтетических данных, основанная на сочетании распределительных, перцептивных и структурных метрик, а также анализе их влияния на результаты классификации, что позволяет оценивать их диагностическую значимость.

Теоретическая значимость результатов диссертационного исследования.

Теоретическая значимость диссертационной работы заключается в развитии научных представлений о применении методов ИИ, в частности CNN и генеративных моделей, для решения задач автоматизированного обнаружения РМЖ на медицинских изображениях. В работе предложен и теоретически обоснован подход, в котором генеративная модель используется не как самостоятельный объект исследования, а как вспомогательный инструмент повышения эффективности классификации в условиях дефицита и дисбаланса данных.

Результаты исследования расширяют теоретические основы использования генеративно-сопоставительных сетей в медицинской визуализации за счёт обоснования их роли в формировании синтетически расширенных обучающих выборок, ориентированных на повышение чувствительности CNN-классификаторов к редкому злокачественному классу. В рамках единого экспериментального исследования системно проанализирована взаимосвязь между параметрами обучения генеративной модели, источниками предобученных весов (ImageNet и RadImageNet) и распределительными и перцептивными метриками качества синтетических изображений (FID, KID, LPIPS, MS-SSIM), а также их влиянием на показатели downstream-классификации.

Отдельное теоретическое значение имеет исследование применения unconditional-подхода к генерации медицинских изображений в задачах балансировки обучающих выборок. В работе показана практическая состоятельность данного подхода при генерации синтетических malignant-изображений без явного задания класса, что позволяет снизить зависимость метода от объёма размеченных данных и рассматривать unconditional-генерацию как допустимый инструмент в задачах медицинской классификации при ограниченной разметке.

Таким образом, диссертационное исследование вносит вклад в развитие теории интеграции генеративных моделей и CNN в медицинской визуализации, формируя методологические основания для оценки качества генерации через её вклад в повышение эффективности автоматизированного обнаружения патологий, а не исключительно по визуальным или распределительным критериям.

Практическая ценность результатов исследования.

Практическая значимость диссертационного исследования заключается в возможности применения разработанного метода DGAN-WP-TL для повышения эффективности автоматизированного анализа медицинских изображений в задачах обнаружения РМЖ. Использование синтетических данных, сформированных с помощью предложенного метода, позволяет снизить влияние

дисбаланса обучающих выборок, повысить чувствительность CNN к редким злокачественным паттернам и улучшить надёжность классификации опухолей молочной железы.

Результаты исследования могут быть использованы при разработке и совершенствовании методов компьютерной поддержки принятия врачебных решений (Computer-Aided Detection, CAD), а также в научно-исследовательских и образовательных проектах в области медицины и машинного обучения. Предложенный подход ориентирован на использование генеративных моделей как вспомогательного инструмента синтетического расширения обучающих данных и может быть интегрирован в существующие конвейеры обучения CNN-классификаторов для медицинской визуализации.

Практическая реализация результатов исследования способствует повышению воспроизводимости экспериментальных исследований, снижению зависимости качества классификации от объёма размеченных данных и расширению возможностей применения методов ИИ в условиях реальных клинических и скрининговых сценариев.

Социальная значимость исследования заключается в повышении эффективности раннего выявления РМЖ за счёт применения методов искусственного интеллекта для анализа медицинских изображений.

Апробация результатов диссертационного исследования.

Основные результаты диссертационной работы были представлены на международных научных конференциях, включённых в базы данных Scopus и IEEE Xplore, что подтверждает высокий уровень научной апробации и соответствие работы современным мировым тенденциям развития искусственного интеллекта в медицине.

Результаты исследования опубликованы и доложены в следующих работах:

1. Generative Adversarial Network as Data Balance and Augmentation Tool in Histopathology of Breast Cancer. // Proceedings of the 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST) (Astana, 2023. – P. 99-104) [59].

2. Enhancing Grayscale Image Synthesis with Deep Conditional GAN and Transfer Learning. // Proceedings of the 2024 IEEE AITU: Digital Generation Conference (Astana, 2024. – P. 122-127) [60].

3. Effect of Data Balancing Methods on MRI Alzheimer's Classification. // Proceedings of the 2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST) (Astana, 2025. – P. 1-7) [61].

Результаты исследования также обсуждались на кафедральных и университетских научных семинарах Костанайского регионального университета имени Ахмет Байтұрсынұлы, где получили положительные отзывы и рекомендации к практическому внедрению (Приложение А).

Кроме того, результаты диссертационной работы внедрены в учебный процесс кафедры программного обеспечения НАО «Костанайский региональный университет имени Ахмет Байтұрсынұлы». Это подтверждается Актом о практическом применении результатов диссертационного исследования

Рыспаевой М.К. на тему «Искусственный интеллект для обнаружения рака груди с использованием глубоких свёрточных нейронных сетей» от 03.03.2026 г. (Приложение А).

Практическая значимость полученных результатов также подтверждается наличием охраняемых объектов интеллектуальной собственности – программ для ЭВМ, реализующих разработанные в рамках диссертационного исследования методы генерации и балансировки медицинских изображений, а также архитектуру генеративной модели DGAN-WP-TL. Указанные программные реализации зарегистрированы в государственном реестре прав на объекты, охраняемые авторским правом, что подтверждается свидетельствами №29428 от 14.10.2022 г., №57670 от 06.05.2025 г. и №67033 от 04.02.2026, и свидетельствует о доведении научных результатов до уровня практической реализации и их готовности к использованию в образовательной и научно-исследовательской деятельности (Приложение Б).

Внедрение результатов способствует повышению качества образовательного процесса, развитию компетенций обучающихся в области современных методов искусственного интеллекта и формированию практических навыков работы с синтетическими данными и нейронными сетями.

Личный вклад автора

Все научные результаты и положения, выносимые на защиту, получены автором лично под руководством научных консультантов. Автором самостоятельно выполнены формулировка замысла и цели исследования, постановка задач, а также выбор методологического подхода к автоматизированному обнаружению РМЖ на основе CNN с использованием методов синтетического расширения данных.

В рамках диссертационного исследования автором проведён анализ и экспериментальное исследование различных архитектур CNN, применяемых для классификации медицинских изображений РМЖ, включая настройку архитектурных параметров, выбор функций потерь и стратегий обучения, а также оценку их эффективности по клинически значимым метрикам классификации. Выполнено построение и обучение CNN-классификаторов для маммографических, ультразвуковых и МРТ-изображений, а также анализ чувствительности моделей к дисбалансу классов.

Для повышения эффективности автоматизированного обнаружения РМЖ автором разработан и теоретически обоснован метод DGAN-WP-TL, основанный на архитектуре Wasserstein GAN с градиентным штрафом (WGAN-GP) и механизмах обучения с переносом (TL), предназначенный для синтетического расширения обучающих выборок и повышения качества классификации CNN. Автором выполнено проектирование архитектур генератора и дискриминатора, выбор и адаптация предобученных весов (ImageNet, RadImageNet), а также программная реализация модели в средах TensorFlow и PyTorch.

Автором осуществлена подготовка и верификация экспериментальных данных, включая сбор, стандартизацию и разметку изображений для наборов BUSI и CBIS-DDSM, а также проведение дополнительной междоменной апробации предложенного подхода на наборе Alzheimer MRI. Проведены

абляционные исследования влияния параметров генеративной модели, в том числе коэффициента градиентного штрафа λ , на качество синтетических изображений и последующее влияние синтетических данных на показатели качества CNN-классификаторов. Выполнено вычисление и анализ метрик качества генерации (FID, KID, LPIPS, MS-SSIM), а также клинически значимых метрик классификации, включая Recall, F1-score и AUC. Проведён сравнительный анализ эффективности классификации при использовании исходных и синтетически расширенных обучающих выборок.

В рамках работы автором разработан воспроизводимый конвейер обучения и экспериментальной оценки моделей, включающий единообразную настройку гиперпараметров, стандартизированную процедуру обучения и тестирования моделей, а также визуализацию распределений признаков реальных и синтетических изображений с использованием методов PCA и t-SNE.

Автором подготовлены иллюстративные материалы, таблицы и тексты разделов диссертационной работы, оформлен список использованных источников и приложения. Публикации по теме диссертационного исследования подготовлены и поданы автором лично; основные результаты работы докладывались и обсуждались на научных конференциях и научных семинарах.

Публикации по теме диссертационного исследования

По теме диссертационной работы опубликованы 9 работ, в том числе 1 статья в журнале, входящем в базу данных Scopus (процентиль по Cite Score равный 54%), 5 статей в изданиях, рекомендованных Комитетом по обеспечению качества в сфере науки и высшего образования МОН РК, 3 статьи в трудах международных конференций, индексируемых в Scopus/IEEE Xplore. Имеется 3 свидетельства о государственной регистрации программы для ЭВМ (Приложение Б).

Структура и объем диссертационной работы

Диссертационная работа состоит из введения, четырёх разделов, заключения, списка использованных источников и приложений. Общий объем работы составляет 116 страниц машинописного текста, содержит 30 рисунков, 41 таблицу и 116 использованных источников, и 3 приложения.

Во введении обоснована актуальность темы диссертационного исследования, сформулированы цель и задачи работы, определены объект и предмет исследования, а также раскрыты научная новизна, теоретическая и практическая значимость полученных результатов.

В первом разделе представлен аналитический обзор современного состояния исследований по обнаружению РМЖ с использованием методов ИИ. Рассмотрены медицинские аспекты ранней диагностики РМЖ, применение CNN в задачах медицинской визуализации, а также проблема ограниченности и дисбаланса клинических данных. Отдельное внимание уделено генеративным моделям для синтеза медицинских изображений и существующим подходам в интеграции генеративных и классификационных методов.

Во втором разделе изложены методологические и математические основы разработки метода обнаружения РМЖ. Рассмотрены архитектуры CNN и их

адаптация к задаче классификации РМЖ, использование обучения с TL и предобученных весов ImageNet и RadImageNet, а также математическая модель генеративных сетей от классического GAN к WGAN и WGAN-GP. Описаны используемые датасеты и экспериментальные данные, представлена двухветвевая архитектура DGAN-WP-TL с интеграцией генеративной модели и feature extractor в дискриминаторе, а также сформулирован подход к классификации опухолей на основе синтетически расширенной обучающей выборки.

В третьем разделе представлены экспериментальные исследования и результаты, направленные на оценку эффективности автоматизированного обнаружения РМЖ с использованием CNN в условиях дефицита и дисбаланса данных. Ключевым элементом раздела является анализ влияния синтетически расширенной обучающей выборки на качество downstream-классификации опухолей, включая оценку изменений клинически значимых метрик Recall, F1-score и AUC. Для обеспечения условий повышения качества классификации выполнена экспериментальная апробация предложенного метода DGAN-WP-TL, используемого в качестве инструмента синтетического расширения данных. Приведены результаты генерации синтетических изображений для датасетов BUSI и CBIS-DDSM, а также выполнена дополнительная методическая апробация метода на данных Alzheimer MRI. Проведён анализ распределений реальных и синтетических изображений с использованием методов PCA и t-SNE, выполнено сравнение с baseline-архитектурой StyleGAN2, а также исследовано влияние предобученных весов ImageNet и RadImageNet и архитектурных параметров DGAN-WP-TL на устойчивость и качество генерации при малом объёме данных.

В четвёртом разделе выполнен аналитический разбор и обобщение результатов исследования, включая интерпретацию экспериментальных данных, обоснование архитектурных решений, сравнительный анализ с существующими подходами и baseline-архитектурами, а также формулирование ограничений, практических рекомендаций и направлений дальнейших исследований.

В заключении подведены основные итоги диссертационного исследования, сформулированы выводы по результатам выполненной работы и рекомендации, а также обозначены перспективы дальнейшего развития предложенного подхода.

В приложениях представлены вспомогательные материалы, включая акт внедрения и документы, подтверждающие практическое использование результатов диссертационного исследования.

1 СОВРЕМЕННОЕ СОСТОЯНИЕ ИССЛЕДОВАНИЙ ПО ОБНАРУЖЕНИЮ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ

1.1 Медицинские аспекты ранней диагностики рака молочной железы

Рак молочной железы (РМЖ) остаётся одной из наиболее актуальных медико-социальных проблем современной онкологии, занимая ведущее место среди злокачественных новообразований у женщин и являясь одной из основных причин онкологической смертности [1, р. 1154-1161; 2, р. 103851; 3, р. 229-262; 4, р. 334]. По данным эпидемиологических исследований, ежегодно в мире регистрируется около двух миллионов новых случаев РМЖ, а число летальных исходов в 2022 году превысило 700 тысяч [1, р. 1154-1161; 3, р. 229-262; 4, р. 334]. Ряд ученых, а именно Kim J. et al., Liao L. и Bray F. et al., прогнозируют, что заболеваемость и смертность РМЖ увеличатся к 2050 году. В связи с этим вопрос автоматизированных методов диагностики РМЖ остается актуальным, особенно на ранних этапах скрининга, позволяя повысить точность обнаружения РМЖ с использованием ИИ.

Казахстанские ученые [5, р. 2807-2816; 6, р. 3361-3370; 7, р. 128-136; 8, р. 1-9; 9, р. 9-15; 10, р. 20536; 11, р. 1340] отмечают общемировую тенденцию роста РМЖ за счет снижения среднего возраста женщин, например, заболевание все чаще диагностируется у женщин моложе 40 лет [5, р. 2807-2816; 6, р. 3361-3370]. Ученые связывают данное наблюдение с рядом изменений в современном образе жизни, а именно поздний возраст первых родов, рост ожирения, низкая физическая активность. Кроме того, национальные скрининговые программы позволяют охватить молодое население современными методами визуальной диагностики (УЗИ, маммография) [6, р. 3361-3370; 7, р. 128-136]. Однако, несмотря на широкое использование методов лучевой диагностики и их эффективность, отмечаются такие основные проблемы, как человеческий фактор в интерпретации результатов и увеличение нагрузки на врачей-радиологов. В исследованиях отмечается устойчивая тенденция к снижению среднего возраста выявления РМЖ и увеличению доли случаев заболевания у женщин младше 40 лет [5, р. 2807-2816; 6, р. 3361-3370].

Рисунок 1 позволяет отобразить эпидемиологическую ситуацию по РМЖ в Республике Казахстан в общем виде. Так по данным Dunenova G.A. et al. [7, р. 128-136], за 2012-2021 гг. заболеваемость растет и достигает своего пика в 2019 году с 52.3 случая на 100 тысяч женщин с последующим частичным снижением к 2020 году. [7, р. 128-136] связывает снижение с пандемией COVID-19, что привело к коронавирусным ограничениям проведения плановых скринингов. В последующие годы наблюдалось восстановление уровня выявляемости заболевания. При этом за аналогичный период зарегистрировано устойчивое снижение смертности от РМЖ с 16.3 до 11.2 случая на 100 тысяч женщин [7, р. 128-136].

Схожие эпидемиологические тенденции представлены в работах Midlenko A. et al. [8, р. 1-8] и Aidarov D. et al. [9, р. 9-15], которые также связывают рост заболеваемости РМЖ с распространением скринингов среди женщин. Дополнительно казахстанские ученые Toguzbayeva A. et al., Igissinov N. et al.

также отмечают, что увеличилось число случаев I-II стадий, которые ранее пропускались и повышали смертность. За счет ранней диагностики I-II стадий РМЖ снизилась летальность заболевания, что напрямую связано с высоким уровнем урбанизации и доступной медициной [10, р. 20536; 11, р. 1340].

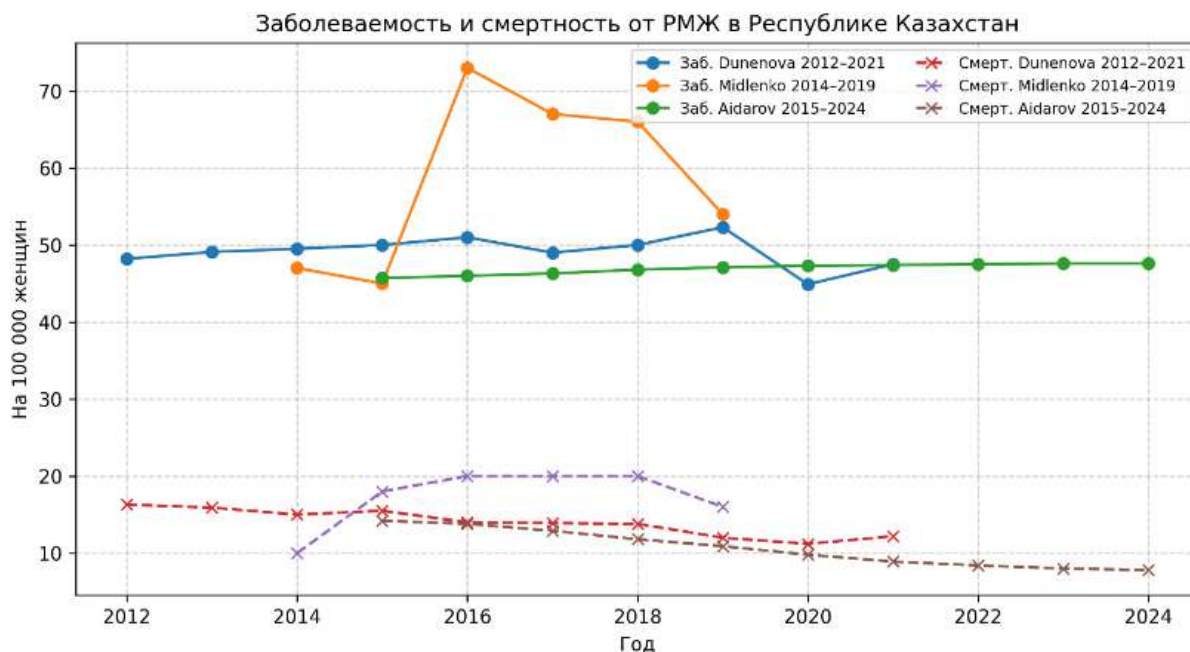


Рисунок 1 – Заболеваемость и смертность РМЖ в Казахстане по данным

Примечание – Составлено по источникам [7, р. 129; 8, р. 2; 9, р. 10]

Таким образом, распространение национальных программ и доступности визуальной диагностики вызывает рост выявляемости случаев РМЖ [5, р. 2807-2816; 6, р. 3361-3370; 7, р. 128-136; 8, р. 1-9; 9, р. 9-15; 10, р. 20536; 11, р. 1340]. Однако наблюдаются региональные различия, так в сельских районах, например, сохраняются проблемы диагностики и интерпретации медицинских изображений, что вызывает рост диагностических ошибок, дополнительной нагрузки на врачей для подтверждения диагноза и сказывается на стадиях выявления заболевания.

Среди методов ранней диагностики маммография остается «золотым стандартом» скрининга, обнаруживая микрокальцинаты и малые опухоли на доклинической стадии [62]. В связи с увеличением количества УЗИ и маммографий согласно цифровизации систем здравоохранения, нагрузка на радиологов возросла многократно, что вызывает усталость врача и риск диагностической ошибки. Становится очевидным, что внедрение ИИ в качестве CAD-систем позволит снизить нагрузку за счет автоматизации анализа медицинских изображений как первичное мнение с дальнейшей обязательной интерпретацией врача-эксперта и постановкой диагноза [63]. Одним из таких автоматизированных методов являются сверточные нейронные сети (CNN), которые показали высокую эффективность в задачах автоматизированного анализа и классификации маммографических и УЗИ изображений РМЖ,

повышая тем самым чувствительность диагностики [18, р. 89-93; 19, р. 60-65; 34, р. 2737-2755].

К сожалению, эффективность CNN снижается из-за недостаточного количества данных, а именно несбалансированных наборов данных. Данная проблема приводит к смещению модели ИИ к преобладающему классу, например, к доброкачественным новообразованиям. Из-за этого, снижается выявляемость миноритарного класса (злокачественных опухолей) и рост ложноотрицательных результатов [47, р. 104863; 48; 64]. Для компенсации дисбаланса классов и дефицита данных в литературе обсуждаются различные стратегии расширения выборок, включая методы генерации и синтетической аугментации данных [12; 13; 14, р. 101552].

1.2 Сверточные нейронные сети в задачах медицинской диагностики и обнаружения рака молочной железы

В основе большинства CAD-систем лежат архитектуры глубокого обучения CNN [65, 66]. Главная идея CNN заключается в обнаружении опухолей и других объектов с помощью автоматизированного извлечения признаков (границы, микрокальцинаты, тканевая неоднородность), вместо ручного метода, используемого в машинном обучении. Это позволяет автоматизировать процесс и снизить влияние человеческого фактора и улучшить дальнейшую интерпретацию результатов.

Однако на результаты CNN остро влияет распределение данных внутри клинических наборов, а именно дисбаланс классов. Даже при использовании современных методов, архитектурные улучшения не обеспечивают роста чувствительности модели к миноритарному классу без соответствующей аугментации. Особенно в задачах медицинской диагностики РМЖ, важно выявлять злокачественные новообразования, которые представлены реже чем доброкачественные [59, р. 99-103; 67].

CNN, в свою очередь, базируются на иерархических представлениях, где низкоуровневые слои автоматически выделяют базовые признаки (границы, текстуры), а высокоуровневые слои обнаруживают анатомические объекты (опухоли, микрокальцинаты) [65, р. 541-550; 66, р. 2278-2323].

Впервые архитектурные принципы CNN были представлены в модели неокогнитрона Фукусимы, продемонстрировавшей поэтапное распознавание образов [68]. Далее LeCun Y. et al. [65, р. 541-550; 66, р. 2278-2323] интегрировал обучение по градиенту, позволяя CNN автоматически формировать признаки без ручного извлечения, как принято в машинном обучении. Существенный прогресс был достигнут с появлением глубоких архитектур, таких как AlexNet [24], VGG [25], Inception [26, р. 1-8], ResNet [27, р. 770-777] и EfficientNet [69], ставшие базой для задач медицинской визуализации.

Практическая эффективность CNN подтверждается крупными валидационными исследованиями. В международной работе McKinney S. et al. [18, р. 89-93] система на основе CNN снизила частоту ложноположительных и ложноотрицательных результатов при маммографическом скрининге и сохранила высокую точность на мультицентровых данных, что демонстрирует

потенциал моделей как «второго читающего врача». Подход Yala A. et al. [19, p. 60-65] расширил парадигму от простой классификации к оценке индивидуального риска на основе истории маммографий и клинических факторов. Аналогичные подходы успешно применяются и в других областях медицинской визуализации, включая нейровизуализацию, пульмонологию и задачи обработки низкодозовых КТ, что подтверждает универсальность CNN для анализа медицинских изображений [23; 46, p. 1348-1356].

Вместе с тем клиническое внедрение CNN ограничивается рядом факторов: дефицитом и дисбалансом данных, межцентровой гетерогенностью и ограниченной объяснимостью решений. «Чёрный ящик» остаётся барьером доверия врача: методы интерпретации позволяют визуализировать вклад фрагментов изображения в решение, однако не устраняют всех вопросов воспроизводимости в реальных потоках [18, p. 89-93; 19, p. 60-65; 20, p. 2199-2209; 21, p. 044504; 22, p. 29; 23].

Одним из ключевых подходов к смягчению ограничений, связанных с дефицитом данных, является перенос обучения (transfer learning, TL), позволяющий инициализировать модели предобученными весами и повысить устойчивость обучения при малых выборках. Чаще всего используются такие веса как ImageNet [70, 71], которые, несмотря на различие доменов, часто обеспечивают стабильную сходимость и конкурентоспособные результаты за счёт универсальности низкоуровневых признаков. Параллельно развиваются доменно-специфичные источники, включая RadImageNet, ориентированный на КТ, МРТ, УЗИ, однако требуют эмпирической проверки на медицинских датасетах [44, p. E210315; 45, p. 321-330].

Таким образом, на практике устойчивость CNN при дефиците данных обеспечивается сочетанием TL и методов расширения выборок; далее рассматриваются причины дефицита (1.3) и генеративные подходы, применяемые в схемах «генерация – классификация» (1.4)-(1.5).

1.3 Проблема ограниченности и дисбаланса медицинских данных

Медицинские данные подвержены особому вниманию при обработке изображений, которые требуют экспертной разметки с привлечением классифицированных специалистов именно врачей-радиологов. Единая стандартизированная разметка является трудоемким, дорогостоящим процессом, что ограничивает количество обработанных изображений. Кроме того, медицинские изображения требуют полной анонимизации и сохранения конфиденциальности. Данные ограничения влияют на доступность данных, их распространение и использование за пределами медицинских учреждений.

В Казахстане, как и во многих других странах, отсутствие единого национального репозитория медицинских изображений и различия в форматах хранения между учреждениями препятствуют формированию открытых и сбалансированных датасетов, что сужает возможности исследований и замедляет внедрение интеллектуальных диагностических систем.

Как было указано в подразделе 1.2, одним из важных ограничений является неравномерность распределения изображений между классами. В клинических

архивах и открытых базах данных количество нормальных или доброкачественных случаев значительно превышает число злокачественных. Это связано с тем, что такие методы медицинской диагностики, как УЗИ используются для первичного скрининга, маммография – для профилактических обследований, и содержат преобладающее большинство изображений без патологий.

Таблица 1 отображает соотношение распределения доброкачественных и злокачественных опухолей в двух датасетах, а именно BUSI [47, p. 104863] и CBIS-DDSM [48]. Преобладание мажоритарного класса приводит к тому, что CNN «привыкает» к преобладающему классу и модель оптимизирует функцию потерь в пользу доминирующего класса. Такое «привыкание» приводит к снижению метрик Recall, F1-score и AUC к редким злокачественным случаям [14, p. 101552; 29, p. 493-496; 72].

Таблица 1 – Соотношение доброкачественных и злокачественных случаев в медицинских наборах данных по РМЖ

Набор данных	Модальность	Всего изображений	Benign	Malignant	Соотношение (benign : malignant)
BUSI	УЗИ	647	437	210	$\approx 2.1 : 1$
CBIS-DDSM	Маммография	1227	718	509	$\approx 1.4 : 1$
Примечание – Составлено по источникам [47, p. 104863; 48]					

Кроме вышеперечисленных ограничений, существует дополнительное при обучении моделей CNN, так называемая межцентровая гетерогенность. Отсутствие единых параметров оборудования и стандартов по обработке медицинских изображений приводит к тому, что модель, обученная на данных одной клиники, теряет точность при применении к другой. В мультицентровых исследованиях показано, что перенос моделей между клиническими центрами может приводить к заметному снижению предсказательной способности модели при переносе между медицинскими учреждениями [18, p. 89-93; 20, p. 2199-2209].

Для компенсации дисбаланса классов применяются различные подходы, включая oversampling и undersampling, статистическую генерацию признаков, аугментацию изображений и использование специальных функций потерь (Focal loss, Dice loss, Tversky loss). Несмотря на повышение устойчивости обучения, данные методы не устраняют ключевое ограничение – недостаток клинически достоверного и структурно разнообразного minority-класса, что ограничивает их эффективность в задачах ранней диагностики.

Таким образом, дефицит и дисбаланс данных в сочетании с межцентровой гетерогенностью и вариабельностью аннотаций приводят к снижению чувствительности CNN и требуют методов компенсации minority-класса [30, p. e19585; 36, p. 1044496]. В подразделе 1.4 рассматриваются генеративные модели, а в разделе 2 представлен подход DGAN-WP-TL, в котором синтез

данных используется как вспомогательный механизм повышения надёжности downstream-классификации.

1.4 Генеративные модели для синтеза медицинских изображений

Сегодня генеративные модели занимают ведущее место в ИИ и находятся под пристальным вниманием ученых. Это связано с тем, что в отличие от традиционных методов аугментации, генеративные модели создают новые распределения данных с высокой фотореалистичностью. Использование генеративных моделей в качестве аугментации позволяет решать проблему ограниченности и дисбаланса в медицинских датасетах РМЖ за счет создания похожих изображений. Таким образом, синтетические изображения близки к реальным за счет воспроизведения распределения именно реальных данных. Прикладная ценность генеративных моделей определяется, в свою очередь, не только визуальной правдоподобностью, но и их влиянием на downstream-метрики (Recall, F1-score, AUC) при обучении CNN-классификаторов [14, p. 101552; 67, p. 2556687; 72, p. 110094].

Поворотным моментом в развитии генеративных моделей стала работа Goodfellow I. et al. [12], в которой были предложены генеративные состязательные сети (GAN), основанные на совместном обучении генератора и дискриминатора. После того, как Goodfellow I. et al. [12] представили базовую архитектуру модели, были предложены различные модификации. Данные модификации направлены на улучшение качества синтетических изображений и устойчивость обучения генеративной модели. Одной из первых модификаций была условная генеративная модель (conditional GAN) [73]. Идея модификации заключалась в управлении генерацией класса в датасете, что вызывало дополнительную нагрузку на вычислительные ресурсы за счет одновременного обучения всех классов. При этом cGAN не решает проблему ограниченности данных, так как модель также «привыкает» к мажоритарному классу и ухудшает качество синтеза.

Последующей модификацией архитектуры была интеграция сверточных слоев с иерархическими представлениями, что позволило улучшить обучение модели и реалистичность изображений. Однако архитектура DCGAN (Deep Convolutional GAN) [74] критична к несбалансированным и малым выборкам и склонна к коллапсу мод (потеря разнообразия в генерации). Дальнейшая модификация была предложена Arjovsky et al. [28] под названием Wasserstein GAN (WGAN). Отличие заключается в применении другой функции потерь с использованием расстояния Вассерштейна, тогда как в DCGAN и в GAN используется классическая функция потерь с Jensen-Shannon Divergence. Однако жесткое ограничение весов (weight clipping) в WGAN снижает стабильность обучения. В результате Gulrajani et al. [13] улучшили WGAN добавлением градиентного штрафа (gradient penalty), и модификация называется WGAN-GP. Градиентный штраф позволяет получить более устойчивое обучение, качественный синтез даже в малых датасетах, что часто является характерным в медицинских выборках [13; 67, p. 255687; 75].

Дальнейшие модификации, например CycleGAN [53, p. 1398-1401], стали использоваться для узкоспециализированных задач, а именно переноса изображений между доменами. Предложенная StyleGAN архитектура демонстрирует высокую реалистичность изображений, однако имеет два ограничения [76-78]. Ограничения заключаются в том, что требуются значительные вычислительные ресурсы и большой объем данных. В связи с этим, StyleGAN ограничен в применении для генерации изображений в медицинской диагностике.

Последующие модификации GAN, такие как ACGAN, BEGAN, InfoGAN, были также адаптированы для специализированных задач и не получили широкого распространения за счет менее устойчивого обучения по сравнению с WGAN-GP [16; 79, 80].

В (Приложении В) дано сравнение архитектурных решений, а также указаны их ограничения и преимущества.

Среди представленных модификаций наибольшее распространение получили WGAN-GP и StyleGAN. Для малых медицинских датасетов лучше подходит WGAN-GP за счет компромисса между стабильностью обучения, количеством данных и вычислительных ресурсов. Однако, WGAN-GP необходимо дополнительно адаптировать под малые медицинские датасеты с помощью интеграции перцептивного анализа, который подразумевает внедрение предобученных моделей CNN. Данная интеграция относится к трансферному обучению, а именно применение предобученных моделей таких как VGG16, VGG19 [25], к примеру. Предобучение модели позволяет извлекать признаки из заранее обученной модели на базе ImageNet или RadImageNet датасетов. С применением такого подхода генеративная модель адаптируется под клинические особенности медицинских изображений, такие как опухоли, микрокальцинаты, плотность тканей и т.д. В данной диссертационной работе перцептивный анализ реализован с помощью интеграции предобученных моделей в архитектуру дискриминатора. В свою очередь, предобученные модели инициализируются ImageNet/RadImageNet весами [44, p. e210315; 70, p. 248-254; 71, p. 211-251]. Такое сочетание модулей в архитектуре, а именно WGAN-GP и трансферное обучение, делает процесс обучения и генерации более стабильным в малых медицинских выборках [14, p. 101552; 17, p. 1-10].

Генеративные модели рассматриваются как инструмент синтетического расширения редких патологических классов и повышения устойчивости обучения диагностических CNN. В настоящей работе акцент сделан на WGAN-GP и на перцептивном контроле качества синтеза через интеграцию предобученных признаков и перенос обучения (TL) [14, p. 101552; 67, p. 2556687]. Практическая полезность синтетических данных оценивается через downstream-метрики классификации, что детально рассматривается в разделе 1.5.

1.5 Интеграция генеративных и классификационных подходов: современные решения и их ограничения

В автоматизированных системах обнаружения РМЖ генеративные модели выступают вспомогательным инструментом в роли аугментации миноритарного класса, а классификационные CNN применяются для обнаружения опухолей. Подобная комбинация моделей позволяет обойти ограничение дисбаланса и нехватки медицинских изображений. Однако недостаточно только оценивать генеративную модель и качество синтезированных данных. Необходимо проводить оценку влияния синтетических данных на метрики downstream-классификации. Такие метрики как Recall, F1-score, AUC отображают, насколько сбалансированный класс с помощью генеративной аугментации обнаруживается классификационной моделью [18, p. 89-93; 19, p. 60-65; 35, p. 106255; 59, p. 99-103; 60, p. 122-126; 67, p. 2556687]. Дополнительно ряд работ проверяют качество синтетических данных с привлечением врачей-радиологов, которые проводят «ослепленную» оценку и интерпретацию изображений [17, p. 1-10; 18, p. 89-93].

Существует ряд исследований в области маммографии об интеграции генеративных и классификационных моделей. В (Приложении В) сравнивают подходы, где аугментация происходит за счет синтетических данных с дальнейшей классификацией в области маммографии. Сначала генеративная аугментация данных применялась для сегментации и доменной адаптации, например, image-to-image, использующейся для переноса между модальностями маммографических снимков [81-85]. Применение условных GAN для image-to-image задач также оставалось нестабильным и, наоборот, ухудшало классификационные метрики особенно локально в ROI-изображениях (Region of Interest) [38, p. 6166; 86, 87].

Дальнейшая адаптация более сложных архитектур, например, HRGAN и StyleGAN, отличается реалистичностью, ориентированностью на безусловную генерацию и устранение дисбаланса. Однако данные исследования подчеркивают, что перечисленные архитектуры отличаются высокой вычислительной сложностью и требуют большого количества данных. Оба перечисленных ограничения влияют на их широкое применение в медицинской диагностике [39, p. 723-732; 77, p. 8107-8115; 78, p. 852-862; 88, 89]. Работа [41, p. 105456] предлагает prior-guided cGAN для трех классов, представленных в выборке (normal, benign, malignant). Авторы [41, p. 105456] заявляют, что классификация с генеративной аугментацией улучшилась на малых и несбалансированных выборках.

Перечисленные исследования объединяет недостаток, который не раскрыт и отражается на результатах моделей, а именно, как и в каком процентном соотношении были добавлены синтетические изображения в обучающие выборки. Авторы не исследовали вопрос влияния синтетики на классификационную модель, а также не дана оценка насколько искажены результаты автоматизированных систем при использовании генеративной аугментации. Дополнительно проведенные исследования направлены на аугментацию всех классов, представленных в датасетах. Авторы исследований не изучают вопрос влияния синтетических данных на миноритарный класс и

снижения доли ложноотрицательных случаев. Подобные пробелы в исследованиях нуждаются в особом внимании, так как доля синтетики может существенно влиять на результаты downstream-классификации. Кроме того, также ставится под сомнение вопрос аугментации мажоритарных классов, которые вызывают смещение предсказательной способности CNN-классификатора [36, p. 1044496; 90].

В УЗИ также применяется интеграция генеративных и классификационных моделей для аугментации малых обучающих выборок. Исследования направлены на улучшение клинически значимых метрик, как и в области маммографии. Сравнительная характеристика представлена в (Приложении В).

Большинство исследований, как и в маммографии, направлены на генеративную аугментацию в задачах классификации и сегментации и демонстрируют рост метрик [91]. В УЗИ также генеративная аугментация рассматривается как вспомогательный модуль.

В дальнейшем в области УЗИ, в отличие от маммографии, задачи генеративной аугментации, сегментации и классификации стали объединяться в единую схему. Например, Attentional GAN [92] и двухстадийная архитектура 2S-BUSGAN [33, p. 8614] одновременно генерируют опухоли и контуры для сегментации. Вследствие этого, такая взаимосвязь улучшила сегментацию и в дальнейшем классификацию. Аналогичный подход реализован и в SPCGAN [42, p. 2555-2564] при ограниченных данных. Отдельное направление связано с расширением информативности входных данных для классификаторов за счёт image-to-image преобразований, например, в исследованиях [32, p. 8599; 93-95], которые подтверждают улучшение классификации за счет интеграции генеративной аугментации и сегментации.

В данной диссертационной работе генеративная аугментация в УЗИ и маммографии применяется как вспомогательный инструмент классификации, именно для компенсации дисбаланса злокачественных опухолей. Классификация при этом является основным этапом в обнаружении РМЖ с использованием глубоких сверточных нейронных сетей. Дополнительно подчеркивается выбор безусловной генерации, тогда как условная генерация увеличивает вычислительные ресурсы, приводит к смещению классификационной модели к мажоритарному классу и избыточности синтетических данных с доброкачественными новообразованиями. В этих условиях, принято решение применить безусловную генерацию с архитектурой WGAN-GP и интеграцией перцептивного анализа в роли трансферного обучения в дискриминатор для улучшения синтеза именно медицинских изображений [16; 33, p. 8614; 37, p. 2925; 39, p. 723-732; 77, p. 8107-8115; 78, p. 852-862; 79, p. 2642-2650; 82, p. 105012; 89, p. 868257; 94, p. 788].

Для оценки качества синтетических данных, используемых в составе интегрированных генеративно-классификационных систем, применяются распределительные и перцептивные метрики, включая FID, KID, LPIPS и MS-SSIM, которые дополняются downstream-оценкой влияния синтетики на Recall, F1-score и AUC. Такой подход отражает переход от субъективной оценки

визуального качества генерации к анализу её вклада в задачи автоматического обнаружения и классификации опухолей [33, р. 8614; 39, р. 723-732; 89, р. 868257; 93, р. 388].

С учётом указанных ограничений в данной работе разработана архитектура DGAN-WP-TL, ориентированная на практическую интеграцию генеративной и классификационной систем. Модель сочетает устойчивость WGAN-GP с перцептивным анализом признаков, реализованным через предобученную сверточную сеть с TL [13], что позволяет согласовать процесс синтеза с требованиями downstream-диагностики. Для комплексной оценки эффективности используются метрики качества генерации и показатели качества классификации, отражающие влияние синтетических данных на способность CNN-моделей обнаруживать злокачественные паттерны.

Выводы по первому разделу

1. РМЖ остаётся одной из актуальных онкопатологий; для Казахстана отмечается рост выявляемости на фоне развития скрининговых программ и снижение смертности, что подчёркивает значимость ранней диагностики и доступности лучевой визуализации [1, р. 1154-1161; 2, р. 103851; 3, р. 229-262; 4, р. 334; 5, р. 2807-2816; 6, р. 3361-3370; 7, р. 128-136; 8, р. 1-9; 9, р. 9-15; 10, р. 20536; 11, р. 1340; 12].

2. Маммография сохраняет статус «золотого стандарта» скрининга, а УЗИ широко применяется на этапах первичного обследования. Рост объёма исследований увеличивает нагрузку на врачей-радиологов и риск диагностических ошибок, что обосновывает необходимость САД-систем на основе ИИ как инструмента поддержки принятия решений [18, р. 89-93; 19, р. 60-65; 34, р. 2337-2755; 62, р. 2353-2357; 63, р. 1828-1836; 96].

3. CNN показали высокую эффективность в задачах анализа маммографических и УЗИ изображений, однако их клиническая устойчивость ограничивается дефицитом и дисбалансом данных, межцентровой гетерогенностью и ограниченной объяснимостью результатов [18, р. 89-93; 19, р. 60-65; 20, р. 2199-2209; 21, р. 044504; 22, р. 29; 23].

4. Дисбаланс выборок приводит к смещению CNN в сторону мажоритарного класса и снижению чувствительности к злокачественным случаям, увеличивая долю ложноотрицательных решений. Классические методы балансировки и аугментации не обеспечивают достаточного клинического разнообразия minority-класса [29, р. 493-496; 30, р. e19585; 36, р. 1044496; 47, р. 104863; 48; 64, р. 170177; 72, р. 110094].

5. Перенос обучения (TL) повышает устойчивость обучения CNN на малых выборках, однако сам по себе не устраняет проблему дефицита патологических примеров и требует сочетания с методами расширения данных [43; 44, р. e210315; 70, р. 248-254; 71, р. 211-251].

6. Генеративные модели рассматриваются как перспективный инструмент синтетического расширения данных для компенсации дефицита malignant-примеров. Для малых медицинских датасетов наиболее практичным подходом является WGAN-GP, а повышение клинической согласованности синтеза

достигается за счёт перцептивного анализа и интеграции предобученных CNN (TL) [13; 14, p. 101552; 17, p. 1-10; 67, p. 2556687; 75, p. 99-111].

7. Практическая ценность синтетических данных должна оцениваться не только по визуальному качеству и распределительным метрикам, но и по влиянию на downstream-метрики (Recall, F1-score, AUC) диагностических CNN. При этом в литературе недостаточно раскрываются доля синтетики в обучении и её влияние на миноритарный класс, что ограничивает воспроизводимость и интерпретацию результатов [18, p. 89-93; 19, p. 60-65; 34, p. 2737-2756; 35, p. 106255; 36, p. 1044496; 90, p. 7281].

2 МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССА И РАЗРАБОТКА МЕТОДА ОБНАРУЖЕНИЯ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ

2.1 Архитектуры сверточных нейронных сетей и их адаптация к классификации рака молочной железы

В подразделе 2.1 рассматривается интеграция генеративной и классификационной моделей для обнаружения РМЖ. В качестве CNN-классификаторов рассмотрены и адаптированы пять backbone-архитектур: ResNet18, VGG16, DenseNet121, MobileNetV2 и EfficientNet-V2-S. А в генеративной модели DGAN-WP-TL для реализации перцептивного анализа интегрированы такие модели как VGG19, DenseNet121, InceptionV3 в роли экстрактора признаков в дискриминаторе. Выбор CNN-архитектур осуществлялся на основе таких характеристик как распространенность, число параметров, вычислительные ресурсы и устойчивость обучения при малых выборках.

Ряд использованных CNN-архитектур применяется в данном исследовании в качестве классификатора, а часть CNN в качестве признакового экстрактора в дискриминаторе. Например, VGG19 [25] извлекает признаки из преобразованных одноканальных изображений в трехканальные. При этом веса модели фиксируются в дискриминаторе. В случае VGG16 [25], предобученная модель дообучается и решает задачу обнаружения РМЖ. Обе модели инициализированы весами ImageNet.

Из семейства ResNet [27, p. 770-777] в исследовании применяются две модификации: ResNet18 и ResNet50. ResNet18 используется в качестве классификатора, который является одной из базовой моделей валидации в генеративной аугментации в медицинских датасетах. Использование ResNet18 позволяет минимизировать влияние сложности классификатора и более корректно оценивать вклад синтетической балансировки в downstream-результаты. В свою очередь, ResNet50 адаптирована в роли экстрактора в дискриминаторе. Необходимо отметить, что ResNet50 инициализирована весами RadImageNet.

DenseNet121 [97] выполняет обе роли в исследовании и выступает в качестве классификатора и экстрактора признаков с весами ImageNet и RadImageNet в дискриминаторе. Использование архитектуры в генеративной и классификационной моделях позволяет сопоставить результаты и их интерпретацию в зависимости от выполняемой задачи на каждом этапе исследования.

Роль InceptionV3 [98] ограничена, и модель адаптирована только как экстрактор в DGAN-WP-TL и только с RadImageNet весами. Инициализация доменно-специфичными весами позволяет провести оценку влияния радиологических изображений на обнаружение РМЖ.

MobileNetV2 [99] и EfficientNet-V2-S [100] включены только в классификационные модели как примеры легких архитектур, что позволяет сравнить их с ёмкими backbone-архитектурами в условиях ограниченных данных и последующей синтетической балансировки.

В таблице 2 дана сравнительная характеристика использованных CNN в данной диссертационной работе, их особенности, ограничения и преимущества.

Все CNN, используемые в диссертации реализуют параметризованное отображение входного изображения в пространство классов [65, р. 541-550]:

$$f_{\theta} : X \mapsto Y, Y = (\hat{y}_0, \hat{y}_1), \hat{y}_0 + \hat{y}_1 = 1 \quad (1)$$

где X – входное медицинское изображение,

θ – совокупность обучаемых параметров сети.

На архитектурном уровне CNN f_{θ} представляет собой последовательность стандартных свёрточных блоков, которые поэтапно извлекают признаки из входного изображения и получают представления, которые затем используются для классификации. Базовой операцией является свёртка: небольшое ядро размера $M \times N$ последовательно проходит по локальным участкам изображения, формируя карты признаков [66, р. 2278-2323]. Базовые операции (свёртка, активации, pooling, регуляризация) используются в стандартной постановке и не рассматриваются детально, поскольку не определяют новизну предложенного метода.

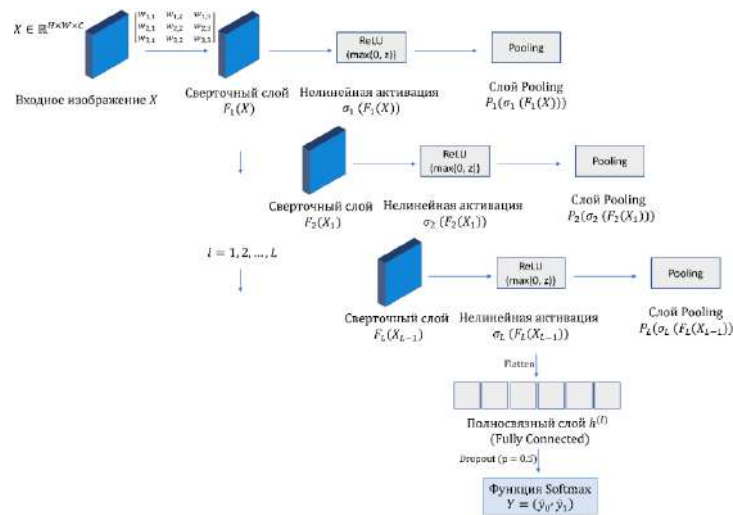


Рисунок 2 – Архитектура CNN, включающая последовательность свёрточных, активационных и pooling-слоёв, операцию Flatten, полносвязный слой, регуляризацию Dropout и финальную функцию Softmax

Рисунок 2 отображает обобщенную структуру CNN, а также ключевые этапы. Схема начинается с входного изображения и далее проходит свёртку, активацию, объединение, векторизацию признаков (Flatten), полносвязную классификацию и применение функции Softmax.

Таблица 2 – Сравнительные характеристики архитектур CNN, используемых в исследовании

Архитектура	Глубина / Параметры	Ключевая особенность	Преимущества	Ограничения	Применение в работе
VGG16	16 слоёв / ~138 млн.	Последовательные 3×3 свёртки	Простота архитектуры; стабильное обучение	Высокая вычислительная сложность	Классификация опухолей
VGG19	19 слоёв / ~143 млн.	Углублённая версия VGG	Высокая чувствительность к текстурам	Большой объём памяти	Экстрактор признаков в дискриминаторе (ImageNet)
ResNet18	18 слоёв / ~11 млн.	Остаточные связи (skip connections)	Устойчивая сходимост; хорошая обобщаемость	Ограниченная глубина признаков	Классификация опухолей
ResNet50	50 слоёв / ~25 млн.	Глубокие residual- блоки	Высокая выразительность признаков	Возможная потеря мелких деталей	Экстрактор признаков в дискриминаторе (RadImageNet)
DenseNet121	121 слой / ~8 млн.	Плотные связи между слоями	Эффективное переиспользо вание признаков; устойчи вость при малых данных	Повышенные требования к памяти	Классификация и экстрактор в дискриминаторе (ImageNet / RadImageNet)
InceptionV3	~48 слоёв / ~24 млн.	Многомасштабные свёртки	Анализ структур разных размеров	Сложность архитектуры	Экстрактор признаков в дискриминаторе (RadImageNet)
MobileNetV2	53 слоя / ~3,4 млн.	Глубинно- разделённые свёртки	Низкая вычислительная сложность	Ограниченная глубина признаков	Классификация опухолей
EfficientNet-V2-S	/ ~22 млн.	Fused-MBConv + MBConv	Быстрое и устойчивое обучение; хорошее соотно шение точности и сложности	Более сложная настройка	Классификация опухолей
Примечание – Составлено по источникам [25; 27, p.770-777; 97, p. 2261-2268; 98, p. 2818-2825; 99, p. 4510-4519; 100, p. 10096-10105]					

Таким образом, в разделе 2.1 определены backbone-архитектуры для классификации и фиксированные экстракторы признаков для дискриминатора DGAN-WP-TL, а также обоснованы их роли в экспериментальной схеме. В разделе 2.2 рассматриваются режимы переноса обучения и выбор предобученных весов (ImageNet/RadImageNet) как ключевой механизм повышения устойчивости при ограниченных данных.

2.2 Transfer learning: использование предобученных весов

Одной из ключевых стратегий повышения эффективности CNN при ограниченном объеме медицинских данных является transfer learning (TL) – перенос знаний, полученных при обучении на одной задаче, на другую, смежную по постановке. Данный подход основан на том, что нижние слои CNN извлекают универсальные визуальные признаки, полезные для широкого круга задач, что позволяет инициализировать модели предобученными весами вместо обучения с нуля.

Наиболее широко используемыми весами являются ImageNet, которые были получены на обучении датасета из 14 млн. изображений повседневных объектов [70, р. 248-254; 71, р. 211-251]. Однако медицинские изображения отличаются от повседневных контрастностью, текстурой и самими объектами (опухоли, микрокальцинаты и т.д.). В связи с этим ImageNet веса ограничены в адаптации к медицинским изображениям [43].

Веса ImageNet могут быть применены для извлечения низкоуровневых признаков, таких как контуры, текстуры, границы, которые характерны и для повседневных объектов. При этом высокоуровневые представления необходимо адаптировать для более специфичных объектов, изображенных на медицинских снимках. В итоге для обнаружения РМЖ применяется трансферное обучение, которое позволяет дообучать высокоуровневые признаки с радиологическими паттернами. Дополнительно применяются специализированные веса RadImageNet, которые были получены после обучения на медицинских снимках КТ, МРТ и УЗИ. По сравнению с ImageNet, количество медицинских изображений уступает в RadImageNet, а именно 1.3 млн. против 14 млн. в ImageNet [71, р. 211-251]. С другой стороны, интеграция доменно-специфичных весов RadImageNet требует валидации и проверки доменной близости к УЗИ и маммографическим снимкам [43].

Сравнительные характеристики наборов ImageNet и RadImageNet, а также их релевантность для медицинских задач, представлены в таблице 3.

Таблица 3 – Сравнительная характеристика предобученных наборов

Характеристика	ImageNet	RadImageNet
1	2	3
Тип данных	Изображения повседневных объектов	Медицинские изображения (рентген, КТ, МРТ, УЗИ)
Объем данных	≈14 млн изображений, 1000 классов	≈1,3 млн изображений, ≈ 165 категорий

Продолжение таблицы 3

1	2	3
Тип аннотаций	Категории объектов	Анатомические области и типы модальностей
Цветовая структура	RGB (разнообразные текстуры и цвета)	Преимущественно градации серого
Релевантность медицинским задачам	Низкая (обобщённые признаки)	Высокая (доменные признаки тканей и структур)
Преимущества	Универсальные низкоуровневые признаки; доступность	Доменно-специфическая структура; улучшенные результаты в медицине
Ограничения	Слабая переносимость на медицинские данные	Ограниченное разнообразие и меньший объем классов
Типичные применения	Общие задачи классификации и детекции	Диагностика, сегментация, генерация медицинских изображений

Эффективность модели $f_{\theta}(X)$ существенно зависит от объёма обучающей выборки, и в условиях ограниченных медицинских данных обучение с нуля часто приводит к переобучению и нестабильной сходимости [65, р. 541-550; 66, р. 2278-2323].

В настоящей работе перенос обучения используется таким образом, что параметры нижних свёрточных слоёв инициализируются предобученными весами, обеспечивающими извлечение универсальных и доменно-специфических признаков, тогда как верхние слои, формирующие вектор признаков, адаптируются к задаче классификации РМЖ.

Следует отметить, что в данной работе перенос обучения используется как в задачах классификации, так и при построении дискриминатора генеративной модели, где предобученные сети выполняют роль экстракторов признаков.

Таким образом, уже определённая ранее модель $f_{\theta}(X)$ уравнение (1) может быть представлена в виде композиции (2) [101, 102]:

$$f_{\theta}(X) = g_{\theta_c}(\phi_{\theta_f}(X)) \quad (2)$$

где ϕ_{θ_f} – сверточная часть, выполняющая роль feature extractor, включающая операции свёртки, активации и pooling;

g_{θ_c} – классификатор, формирующий выходное решение.

Параметры сети $\theta = \{\theta_f, \theta_c\}$ инициализируются предобученными весами, перенесёнными из исходной модели, тогда как классификационная голова и (при необходимости) верхние свёрточные слои дообучаются на целевой задаче.

В практике TL различают два основных подхода [43; 102, р. 1345-1358]:

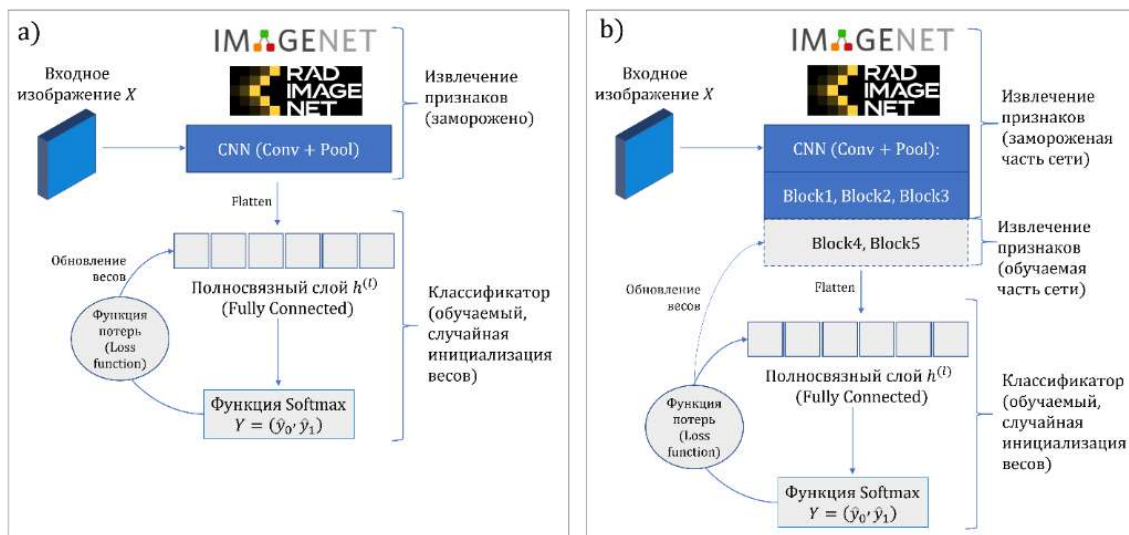
1. Замороженные сверточные слои (feature extraction).

В первом подходе веса предобученной модели фиксируются, а обучение выполняется только для классификатора g_{θ_c} . Такой режим называется feature

extraction, поскольку сеть использует уже готовые признаки, извлечённые из базовой модели, не изменяя их (рисунок 3а).

2. Тонкая настройка (fine-tuning).

Во втором подходе часть свёрточных слоёв (обычно верхние) остаётся обучаемой. За счет этого, верхние слои дообучаются под специфичные медицинские изображения, сохраняя низкоуровневые признаки, извлеченные из весов ImageNet/RadImageNet (рисунок 3b) [43; 103].



а – извлечение признаков с фиксированными весами (feature extraction); б – тонкая настройка (fine-tuning), при которой обучается верхняя часть свёрточной сети и классификатор

Рисунок 3 – Схематичное представление двух режимов TL

В основном, fine-tuning подход применяется при наличии большего объёма данных или при существенном различии визуальных характеристик целевого и исходного доменов.

В данной диссертационной работе использовались оба режима трансферного обучения в зависимости от задачи. Первый режим извлекает признаки, при этом часть сверточной модели фиксируется и применяется в дискриминаторе генеративной модели для перцептивного анализа с помощью VGG19, ResNet50, InceptionV3 и DenseNet121. А второй режим, fine-tuning, используется в классификационной модели, где часть верхних сверточных слоев дообучается совместно с классификатором (VGG16, ResNet18, DenseNet121, MobileNetV2, EfficientNet-V2-S).

С точки зрения общей математической структуры, TL не меняет функциональную форму модели $f_{\theta}(X)$, описанную в разделе 2.1, а лишь уточняет процесс оптимизации параметров.

Таким образом, трансферное обучение позволяет адаптировать глубокие сверточные модели под задачи медицинской диагностики. Однако, несмотря на преимущества переноса обучения, TL не решает проблему дисбаланса данных и не обеспечивает достаточной вариативности изображений для устойчивого обучения. Поэтому для устранения ограничений, применяется вспомогательный

модуль в виде генеративной модели. При генеративной аугментации выполняется синтез новых, разнообразных и реалистичных примеров миноритарного класса.

2.3 Математическая модель генеративных сетей: от классического GAN к WGAN и WGAN-GP

В данном подразделе рассматривается более последовательно архитектура генеративно-состязательной сети и ее две модификации WGAN и WGAN-GP. Формализация GAN, WGAN и WGAN-GP является основой предлагаемого метода DGAN-WP-TL в данной работе

Развитие ИИ и его методов, а именно глубокого обучения, привело к появлению моделей, способных не только классифицировать, но и генерировать новые данные, статистически схожие с исходными. Впервые архитектура GAN была предложена I. Goodfellow et al. в 2014 году [12]. Идея GAN основана на состязательном обучении двух сетей: генератора и дискриминатора. Дискриминатор учится отличать реальные и синтетические выборки, а генератор – формировать данные, максимально приближённые к реальным, что формализуется минимакс-оптимизацией.

Генератор $G(z; \theta_G)$ – это нейронная сеть с параметрами (весами) θ_G , которая преобразует латентный вектор z в изображение [12]. Дискриминатор $D(x; \theta_D)$, в свою очередь, получает изображение x (реальное или сгенерированное) и возвращает скалярное значение $D(x) \in [0,1]$, которое показывает вероятность того, насколько изображение является настоящим. Его параметры θ_D также обновляются во время обучения.

Обе сети обучаются совместно, а именно одна максимизирует целевую функцию, другая минимизирует её. В этом и заключается минимакс-оптимизация: дискриминатор пытается увеличить вероятность правильного распознавания, а генератор, наоборот, уменьшить её, чтобы его изображения были труднее отличить от реальных.

Архитектура генератора состоит из нескольких транспонированных свёрточных слоёв (Conv2DTranspose), которые выполняют обратную свёртку. Дискриминатор, напротив, состоит из обычных свёрточных слоёв (Conv2D), которые классифицируют изображение как реальное или синтетическое.

Целевая функция и обучение. Состязание между $G(z)$ и $D(x)$ формулируется через целевую функцию (value function) (3) [12]:

$$\begin{aligned} & \min_G \max_D V(G, D) \\ & = \mathbb{E}_{x \sim P_{real}} [\log D(x)] + \mathbb{E}_{z \sim P_z} [\log (1 - D(G(z)))], \end{aligned} \quad (3)$$

где $V(G, D)$ – это функция, описывающая “игру” между сетями;

$\mathbb{E}_{x \sim P_{real}}$ – математическое ожидание, то есть среднее значение по всем реальным изображениям из распределения P_{real} (распределение реальных данных);

$\mathbb{E}_{z \sim P_z}$ – ожидание (среднее значение) по всем латентным векторам z , взятым из распределения латентного (скрытого) шума $P_z(z)$. Математическое ожидание в данном контексте означает усреднение – сеть оценивает не конкретное изображение, а поведение функции на множестве примеров.

$\mathbb{E}_{x \sim P_{real}}[\log D(x)]$ – показывает, насколько хорошо дискриминатор различает настоящие изображения;

$\mathbb{E}_{z \sim P_z}[\log(1 - D(G(z)))]$ – отражает, насколько успешно генератор “обманывает” дискриминатор;

В процессе обучения дискриминатор максимизирует функцию $V(G, D)$, стремясь отличить реальные изображения от синтетических, тогда как генератор минимизирует её, формируя данные, неотличимые от реальных. При $P_{gen}(x) \approx P_{real}(x)$ достигается равновесие, где ни одна сеть не может улучшить результат, не ухудшив работу другой.

Функция потерь. В классическом GAN минимакс-оптимизация (3) эквивалентна минимизации расхождения между распределениями P_{gen} и P_{real} , которое часто связывают с дивергенцией Дженсена-Шеннона. На практике при слабом перекрытии распределений градиенты для генератора могут становиться малыми, что приводит к нестабильности обучения и эффекту mode collapse.

Для преодоления этих проблем Arjovsky et al. (2017) [28] предложили использовать расстояние Вассерштейна (Wasserstein distance), также известное как “расстояние переноса массы” или Earth Mover’s Distance. Оно измеряет минимальные “затраты” на преобразование одного распределения в другое. Расстояние можно записать в практической форме (4) [28]:

$$L_W = \mathbb{E}_{x \sim P_{real}}[D(x)] - \mathbb{E}_{z \sim P_z}[D(G(z))], \quad (4)$$

где D – 1-Lipschitz-функция, ограничивающая скорость изменения выходов ($D(x)$ должна быть плавной, без резких скачков).

Таким образом, дискриминатор превращается в критика, который оценивает не вероятность, а “величину различия” между распределениями. Для того, чтобы соблюсти 1-Lipschitz-ограничение, в оригинальном WGAN использовался метод weight clipping (ограничение весов) $w_i \in [-c, c]$. Однако при слишком малом c критик терял способность различать данные, а при слишком большом – терялась устойчивость.

Критик должен максимизировать L_W , то есть усиливать различие между реальными и сгенерированными распределениями. Но поскольку оптимизатор минимизирует функцию потерь, то берётся отрицательный знак ($-L_W$), что приводит к функциям потерь (5)-(6) [28].

$$L_D = \mathbb{E}_{z \sim P_z}[D(G(z))] - \mathbb{E}_{x \sim P_{real}}[D(x)], \quad (5)$$

$$L_G = -\mathbb{E}_{z \sim P_z}[D(G(z))], \quad (6)$$

где L_D и L_G – функция потерь критика и генератора соответственно в WGAN.

Генератор минимизирует оценку расстояния Вассерштейна, увеличивая значения критика на синтетических примерах и тем самым сближая P_{gen} с P_{real} .

Позже Gulrajani et al. (2017) [13] предложили улучшение – WGAN с градиентным штрафом (WGAN-GP). В нём вместо обрезания весов вводится дополнительный регуляризационный член, называемый градиентным штрафом (Gradient Penalty) (7) [13]:

$$L_{GP} = \lambda_{GP} \mathbb{E}_{\hat{x}} (\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2, \quad (7)$$

где λ_{GP} – коэффициент, определяющий силу штрафа;

$\hat{x} = \varepsilon x_{real} + (1 - \varepsilon)x_{gen}$ – интерполяция между реальным и синтетическим изображением. Интерполяция означает, что берётся “точка между” реальным и синтетическим изображением, как будто их пиксели смешаны в пропорции ε . Таким образом формируются интерполированные сэмплы – промежуточные примеры, на которых проверяется, насколько плавно ведёт себя $D(x)$.

$\mathbb{E}_{\hat{x}}$ – усреднение по этим интерполированным сэмплам;

$\varepsilon \sim U[0,1]$ – случайный коэффициент смешивания;

$\nabla_{\hat{x}} D(\hat{x})$ – градиент функции $D(x)$ по её входу.

Регуляризация означает добавление в функцию потерь дополнительного “штрафующего” терма. Если условие $\|\nabla_{\hat{x}} D(\hat{x})\|_2 = 1$ нарушено, значение L_{GP} растёт, штрафует модель. Иными словами, ошибка (функция потерь) увеличивается, заставляя сеть корректировать свои веса так, чтобы вернуть градиенты к норме. Таким образом, итоговые функции потерь в WGAN-GP принимают вид (8) [13]:

$$L_D = \mathbb{E}_{z \sim P_z} [D(G(z))] - \mathbb{E}_{x \sim P_{real}} [D(x)] + \lambda_{GP} \mathbb{E}_{\hat{x}} (\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2, \quad (8)$$

где L_{GP} применяется только к критику, так как штраф контролирует норму его градиентов по входным данным. Функция потерь генератора в WGAN-GP совпадает с уравнением 6, поскольку градиентный штраф применяется исключительно к критику и не влияет на оптимизацию генератора.

В отличие от weight clipping, градиентный штраф обеспечивает непрерывность функции критика без ограничения диапазона весов, устраняя проблемы насыщения и повышая устойчивость обучения.

Модифицированная версия GAN, а именно WGAN-GP, позволила стабильно обучаться на малых датасетах и избегать эффекта mode collapse. Обучение стало плавным, способствуя генерации реалистичных синтетических изображений.

На рисунке 4 показана общая схема архитектуры GAN с двумя модификациями WGAN и WGAN-GP. Схема демонстрирует, что генератор $G(z)$ получает на вход латентный вектор и формирует синтетические изображения, тогда как критик $D(x)$ оценивает как реальные, так и сгенерированные данные, возвращая скаляр, используемый в соответствующих функциях потерь L_D и L_G .

Затем параметры обеих сетей обновляются по направлению градиента этих функций.

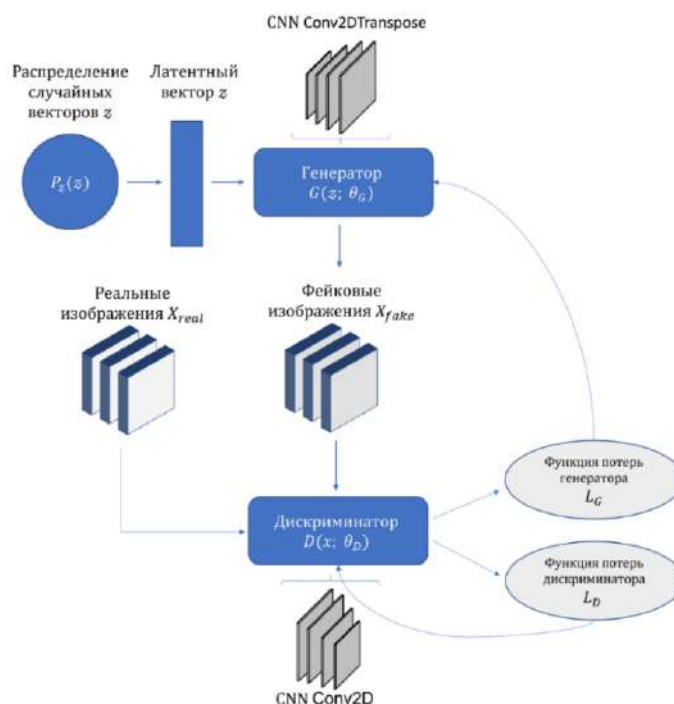


Рисунок 4 – Общая структура генеративно-состязательной сети в формулировке GAN / WGAN / WGAN-GP

Для наглядного сравнения основных отличий между архитектурами GAN, WGAN и WGAN-GP представлена таблица 4.

Таблица 4 – Сравнение архитектур GAN, WGAN и WGAN-GP

Модель	Авторы	Основная идея	Функция потерь	Особенности	Ограничения	Преимущества
1	2	3	4	5	6	7
GAN	Goodfellow et al.	Состязание генератора и дискриминатора	Логарифмические функции потерь, связанные с минимизацией дивергенции Дженсена–Шеннона	Простая реализация, вероятностный выход	Исчезающие градиенты, mode collapse	Гибкость, высокая выразительность
WGAN	Arjovsky et al.	Метрика Вассерштейна	Линейная разность математических ожиданий, аппроксимирующая расстояние Вассерштейна	Критик вместо дискриминатора, weight clipping	Потеря информативности, настройка c	Устойчивое обучение, понятная метрика

Продолжение таблицы 4

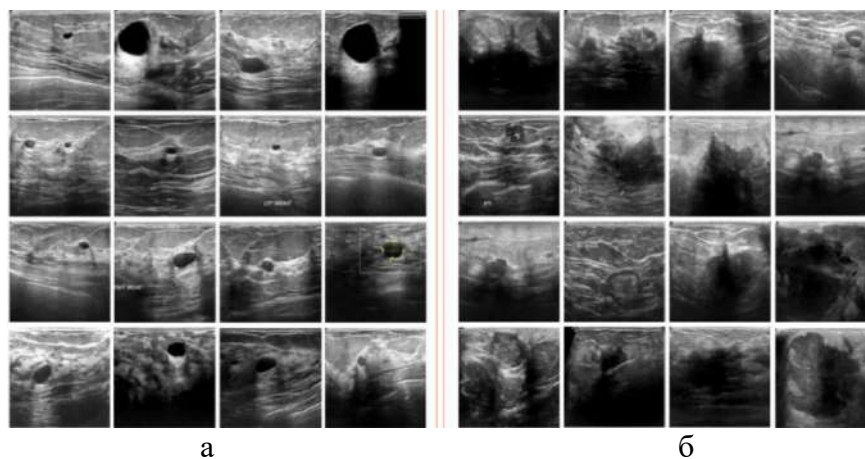
1	2	3	4	5	6	7
WGAN-GP	Gulrajani et al.	Добавление градиентного штрафа	Функция потерь WGAN с градиентным штрафом, обеспечивающим выполнение 1-Lipschitz-ограничения	Регуляризация градиентов, плавная сходимость	Повышенная вычислительная сложность	Максимальная стабильность и качество синтеза
Примечание – Составлено по источникам [12; 13; 28]						

Таким образом, в разделе 2.3 были рассмотрены основные теоретические принципы генеративно-сопоставительных сетей, начиная с классической модели GAN, WGAN и заканчивая устойчивой модификацией WGAN-GP. Эти архитектуры и соответствующие функции потерь образуют математическую основу для дальнейшего построения предлагаемой модели DGAN-WP-TL. Следующий раздел посвящён описанию архитектурной реализации предложенного метода DGAN-WP-TL, в котором рассмотренные теоретические принципы используются в модульной генеративной схеме с учётом специфики медицинских изображений.

2.4 Датасеты и экспериментальные данные

В качестве обучающих датасетов были выбраны три медицинских набора, которые отличаются по модальности, разрешению и визуальным признакам. Использование мультимодальных датасетов позволяет провести оценку устойчивости предлагаемого метода и синтетического расширения малых несбалансированных датасетов.

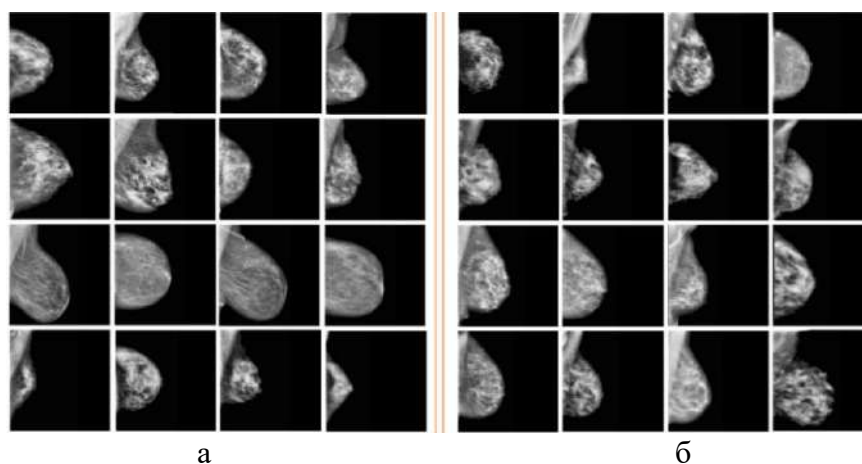
Breast Ultrasound Images Dataset (BUSI). Датасет BUSI содержит ультразвуковые изображения РМЖ с категориями *benign* и *malignant* опухолями (рисунки 5а, 5б) [47, р. 104863]. В наборе представлено 437 доброкачественных и 210 злокачественных изображений. Все изображения приведены к единому формату (PNG) и размеру 512×512.



а – доброкачественные опухоли РМЖ; б – злокачественные опухоли РМЖ

Рисунок 5 – Примеры ультразвуковых изображений из датасета BUSI

Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM). CBIS-DDSM представляет собой подвыборку базы DDSM, включающую цифровые маммограммы с отмеченными радиологами областями интереса [48]. В исследовании рассматривались случаи микрокальцификаций с использованием стандартных ракурсов Cranio-Caudal (CC) и Mediolateral Oblique (MLO). Как и УЗИ-изображения, маммографические снимки были приведены к единому формату (PNG) и размеру 224×224 пикселей. Конечная выборка включала 718 доброкачественных и 509 злокачественных изображений (рисунок 6а, 6б).

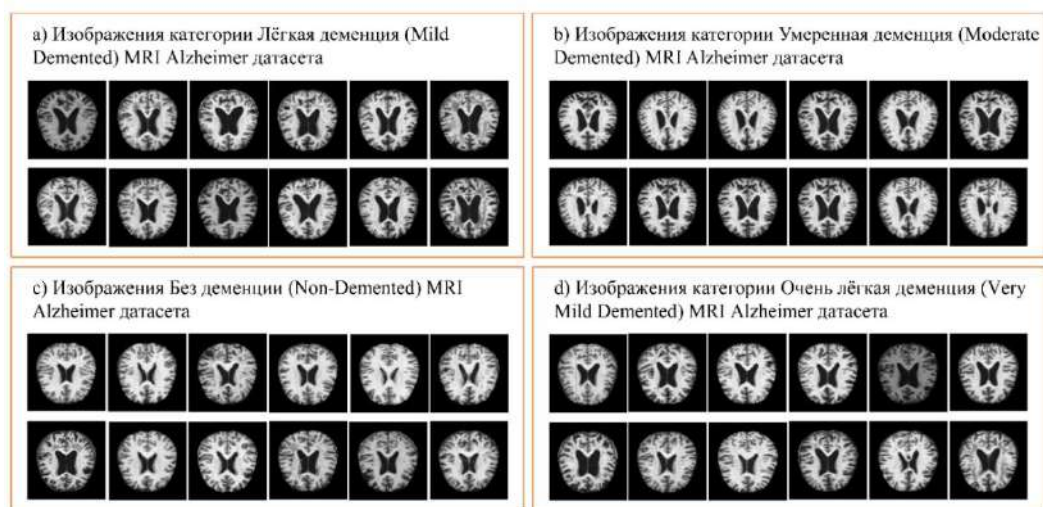


а – доброкачественные опухоли РМЖ; б – злокачественные опухоли РМЖ

Рисунок 6 – Примеры маммографических изображений из датасета CBIS-DDSM

Alzheimer MRI Preprocessed Dataset (Alzheimer MRI). Для того чтобы оценить, насколько метод можно применять к другим видам медицинских изображений, дополнительно использовался Alzheimer MRI Dataset [49]. Датасет включает 6400 предобработанных МРТ-изображений мозга размером 128×128 пикселя, распределённых по четырём категориям: Non-Demented (3200), Very Mild Demented (2240), Mild Demented (896) и Moderate Demented (64) (рисунок 7

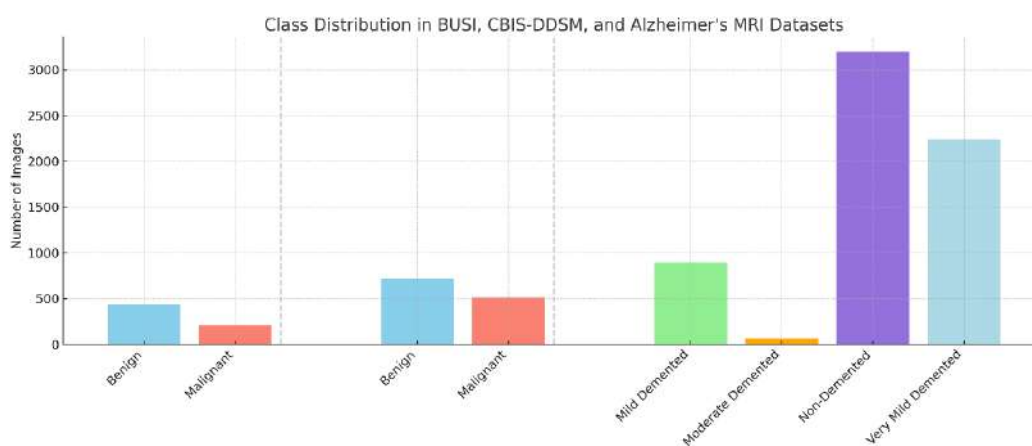
a-d). Наиболее редким является класс Moderate Demented, что создаёт выраженный дисбаланс классов.



a – Mild Demented; b – Moderate Demented; c – Non-Demented; d – Very Mild Demented

Рисунок 7 – Примеры изображений четырёх категорий датасета Alzheimer MRI

Несмотря на различия в модальности, разрешении и характере визуальных признаков, все рассмотренные датасеты объединяет выраженный дисбаланс классов. В ультразвуковом датасете BUSI число злокачественных опухолей существенно уступает числу доброкачественных случаев, в Alzheimer MRI крайне малочисленным является класс Moderate Demented, а в CBIS-DDSM ограничено количество изображений злокачественных микрокальцификаций.



BUSI – левый; CBIS-DDSM – центральный; Alzheimer's MRI – правый

Рисунок 8 – Распределение классов в медицинских датасетах

Рисунок 8 отображает распределение классов внутри трех датасетов. Из схемы видно, что датасеты несбалансированны и могут привести к смещению предсказаний при обучении генеративных и классификационных моделей. Таким образом, выбранные наборы данных показывают разнообразие по

медицинским модальностям и выраженный дисбаланс, что является методологически показательным для оценки эффективности предложенного подхода.

На основе рассмотренных характеристик экспериментальных данных в подразделе 2.5 формируется архитектура предлагаемой модели DGAN-WP-TL, которая в подразделе 2.6 используется для построения системы классификации РМЖ.

2.5 Интеграция генеративной модели и feature extractor в критике (двухветвевая архитектура DGAN-WP-TL)

В данном подразделе формализуется предложенная модель DGAN-WP-TL на базе модификации WGAN-GP с интеграцией трансферного обучения для перцептивного анализа с применением предобученных моделей. Такая интеграция именно в критике позволяет достичь высокой реалистичности медицинских изображений.

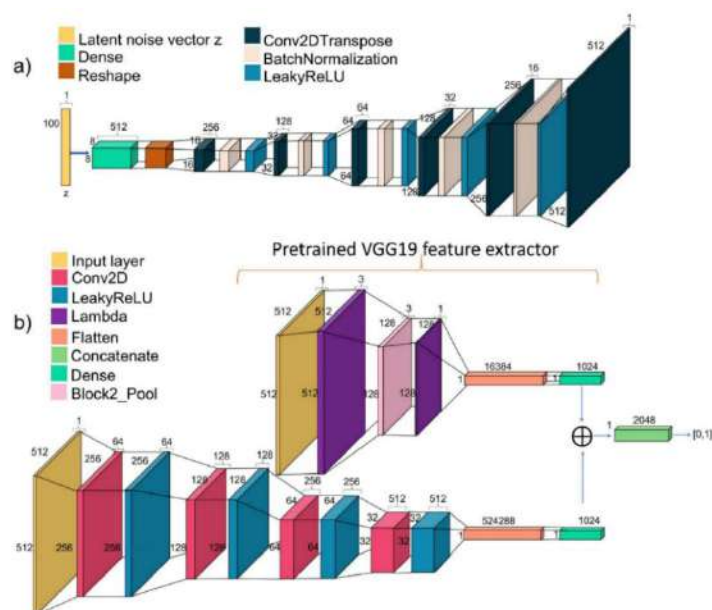
Математическая модель заключается в формализации процесса обучения DGAN-WP-TL и синтеза изображений РМЖ. DGAN-WP-TL использует оптимизацию на основе расстояния Вассерштейна (Wasserstein GAN) с добавлением градиентного штрафа (Gradient Penalty) и использованием TL через предварительно обученную сеть VGG19, выбранную в качестве основной перцептивной ветви критика. Как было определено ранее, такая интеграция позволяет извлекать низкоуровневые признаки, характерные для всех изображений. Модель функционирует в безусловном (unconditional) режиме: метки классов не подаются ни в генератор G , ни в критик D .

В критике DGAN-WP-TL реализованы две ветви. Первая ветвь выполняет роль перцептивного экстрактора признаков за счет усеченной модели VGG19 до блока *block2_pool*, при этом веса $\varphi(x)$ зафиксированы. Вторая ветвь состоит из собственных сверточных слоев, которые извлекают обучаемые признаки $\psi(x)$.

Далее признаки двух ветвей конкатенируются и передаются в полносвязные слои для формирования скалярного значения дискриминатора $D(x)$.

Используя введенные в подразделе 2.3 обозначения для генератора $G(z; \theta_G)$, критика $D(x; \theta_D)$ и распределений $P_{real}(x)$, $P_{gen}(x)$, далее формализуется интеграция перцептивных признаков из предобученной сети в критике модели DGAN-WP-TL.

Генератор реализует постепенное создание изображения из латентного пространства. После начального полносвязного слоя формируется тензор размером $8 \times 8 \times 512$, который проходит через шесть транспонированных свёрточных слоев (Conv2DTranspose, stride=2), последовательно увеличивающих разрешение изображения ($8 \rightarrow 16 \rightarrow 32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512$) (рисунок 9а). Подробная спецификация слоёв генератора приведена в таблице 5 [67, p. 2556687].



а – генератор, преобразующий латентный вектор в изображение с помощью последовательности транспонированных свёрточных слоёв; б – двухветвевой критик, объединяющий признаки, извлечённые предобученной VGG19 (верхняя ветвь), с обучаемыми признаками собственной CNN (нижняя ветвь)

Рисунок 9 – Архитектура модели DGAN-WP-TL

Примечание – Составлено по источнику [67, р. 2556687]

Таблица 5 – Архитектура генератора модели DGAN-WP-TL

Номер слоя	Тип слоя	Выход, пиксели
1	Входной полносвязный слой, Dense	8×8×512
2	Reshape – изменение формы	8×8×512
3	Транспонированный сверточный слой, Conv2DTranspose	16×16×256
4	Нормализация батча, Batch Normalization	16×16×256
5	Транспонированный сверточный слой, Conv2DTranspose	32×32×128
6	Нормализация батча, Batch Normalization	32×32×128
7	Транспонированный сверточный слой, Conv2DTranspose	64×64×64
8	Нормализация батча, Batch Normalization	64×64×64
9	Транспонированный сверточный слой, Conv2DTranspose	128×128×32
10	Нормализация батча, Batch Normalization	128×128×32
11	Транспонированный сверточный слой, Conv2DTranspose	256×256×16
12	Нормализация батча, Batch Normalization	256×256×16
13	Транспонированный сверточный слой	512×512×1

Критик имеет двухветвевую архитектуру, как описано выше. Первая ветвь состоит из усечённой сети VGG19, извлекающей предобученные признаки $\varphi_{VGG19}(x)$; её конфигурация показана в таблице 6А. Вторая ветвь из собственных

свёрточных блоков критика, формирующих обучаемые признаки $\psi(x)$; её структура приведена в таблице 6В. Эти признаки объединяются (конкатенируются) и подаются в полносвязные слои, формирующие итоговое значение функции критика (рисунок 9b). Конкатенация показана в таблице 6С.

Таблица 6 – Архитектура двухветвевго критика (дискриминатора) генеративной модели DGAN-WP-TL

Номер слоя	Тип слоя	Выход, пиксели
А. Первая ветвь с предобученными весами VGG19		
1	Вход, Input	512×512×1
2	Слой преобразования одноканального изображения в трехканальный, Lambda	512×512×3
3	Слой block2_pool модели VGG19	128×128×1
4	Слой Flatten	16384
5	Полносвязный слой, Dense	1024
В. Вторая ветвь с кастомным набором сверточных слоев		
6	Сверточный слой Conv2D, входной слой	512×512×1
7	Сверточный слой Conv2D	256×256×64
8	Сверточный слой Conv2D	128×128×128
9	Сверточный слой Conv2D	64×64×256
10	Сверточный слой Conv2D	32×32×512
11	Слой Flatten	512288
12	Полносвязный слой, Dense	1024
С. Объединение двух ветвей		
13	Конкатенация слоя 5 и 12	2048
14	Полносвязный слой Dense	Число, оценка критика

Уравнение (9) отображает математическую модель двухветвевго критики, в которой $\varphi_{VGG19}(x)$ представлены как признаки из предобученной модели VGG19, а $\psi(x)$ признаки из сверточной части [13, 25]:

$$D(x) = f_{disc}([\varphi_{VGG19}(x), \psi(x)]), \quad (9)$$

Функции потерь генератора и критика приведены в уравнениях (6) и (8) подраздела 2.3.

Согласно архитектуре критикиа перцептивный анализ проводится внутри критика, не затрагивая функцию потерь, что повышает оценку реалистичности изображений по человеческому восприятию. Алгоритм 1 показывает интеграцию перцептивных признаков в предложенную архитектуру DGAN-WP-TL (рисунок 10).

Алгоритм 1: Обучение генеративной модели DGAN-WP-TL для синтеза медицинских изображений	
Шаг 1: предварительная обработка изображений	
-	Необходимо выбрать медицинский датасет для генерации (BUSI или CBIS-DDSM)
-	Определить <i>миноритарный класс</i> (например, злокачественный класс)
-	Сделать предварительную обработку изображений:
1)	привести все изображения к единому размеру и формату (512×512 , 224×224 , I)
2)	провести нормализацию пикселей в диапазоне $[-1, 1]$
3)	конвертировать изображения в градациях серого в трехканальные RGB
Шаг 2: модель генератора G	
-	Входные данные: инициализация вектора латентного шума размерности 100 ($z \sim P_z(z)$)
-	Построить генератор с помощью следующих слоев:
1)	входной полносвязный слой, <i>Dense</i>
2)	<i>reshape</i> – изменение формы
3)	транспонированный сверточный слой, <i>Conv2DTranspose</i>
4)	нормализация батча, <i>Batch Normalization</i>
-	Выходные данные – <i>синтетические изображения</i> $x_{fake} = G(z)$
Шаг 3: модель дискриминатора (критика) D	
-	Входные данные: реальные и сгенерированные изображения
-	Необходимо загрузить предобученную модель VGG19 в качестве первой ветви
1)	Обрезать VGG19 до блока <i>block2_pool</i>
2)	Заморозить веса VGG19
-	Построить вторую ветвь с кастомным набором сверточных слоев <i>Conv2D</i>
-	Конкатенация признаков из двух ветвей
-	Выходные данные – <i>значение дискриминатора (критика)</i>
Шаг 4: обучение модели DGAN-WP-TL	
-	Инициализировать оптимизаторы <i>Adam</i> генератора и дискриминатора для обновления весов
-	Для каждой эпохи ($t = 1, 2, \dots, EPOCHS$):
•	Для каждого мини-батча в обучающей выборке:
○	Выбрать реальные изображения x_{real} из мини-батча;
○	Выбрать случайные латентные векторы $z \sim P_z(z)$;
○	Сгенерировать синтетические изображения $x_{fake} = G(z)$;
○	Посчитать функцию потерь для дискриминатора L_D и обновить веса D ;
○	Каждые $n_{critic} = 5$ шагов посчитать функцию потерь генератора L_G , и обновить веса G
-	Посчитать метрики генерации каждые 100 эпох (FID, KID, LPIPS, MS-SIM)
Шаг 5: сгенерировать и сохранить синтетические изображения	
-	С помощью обученной модели генератора сгенерировать синтетические изображения
-	Сохранить синтетические изображения

Рисунок 10 – Алгоритм обучения модели DGAN-WP-TL

Для обучения модели DGAN-WP-TL использовалась стратегия, основанная на схеме WGAN-GP. Оптимизация параметров генератора и критика выполняется с применением оптимизатора *Adam* при скорости обучения $lr = 1 \times 10^{-5}$ и коэффициентах $\beta_1 = 0.5$, $\beta_2 = 0.999$. Для обеспечения более плавной сходимости и уменьшения колебаний функции потерь применяется экспоненциальный коэффициент затухания скорости обучения (decay factor), равный 0.995 на эпоху. В ходе предварительного экспериментального анализа размер батча был установлен равным 2. Критик обновляется пять раз на один шаг обновления генератора ($n_{critic} = 5$). Общее количество эпох составило 500 и 1000 для ImageNet и RadImageNet весов, соответственно.

Дополнительной частью методологии экспериментов является исследование влияния градиентного штрафа на генерацию и классификацию. Выбраны пять вариаций градиентного штрафа в диапазоне $\{0.5; 3.0; 5.0; 7.0; 10.0\}$, что позволяет исследовать как регуляризация влияет на обнаружение РМЖ. Эксперименты по исследованию влияния градиентного штрафа позволили оценить качество синтеза и влияние на классификацию.

Ключевые отличия предлагаемой модели DGAN-WP-TL заключаются в следующем:

1. Предложена архитектурная модификация критика WGAN-GP в виде двухветвевой структуры, в которой объединяются признаки, извлекаемые предобученной сетью VGG19 (усечённой до блока *block2_pool*), и обучаемые признаки собственной свёрточной ветви критика. Такое архитектурное решение позволяет учитывать как предобученные признаки влияют на реалистичность изображений, обеспечивая более устойчивое обучение генеративной модели в условиях выраженного дисбаланса медицинских данных.

2. В рамках разработки модели DGAN-WP-TL проведено исследование влияния коэффициента градиентного штрафа λ в диапазоне $\{0.5; 3.0; 5.0; 7.0; 10.0\}$ на устойчивость обучения критика, качество синтетических изображений и эффективность последующей классификации. Полученные результаты представлены и проанализированы в разделе 3.

Таким образом, предложенная архитектура DGAN-WP-TL объединяет механизмы Wasserstein-регуляризации и TL для устойчивого синтеза медицинских изображений. Модель DGAN-WP-TL является вспомогательным инструментом генеративной аугментации, что позволяет дополнить реальные данные, сбалансировать миноритарный класс и повысить вариативность выборки.

В следующем подразделе 2.6 рассмотрен процесс интеграции сгенерированных синтетических изображений в задачу классификации опухолей РМЖ на основе CNN с применением TL.

2.6 Классификация рака молочной железы с использованием свёрточных нейронных сетей и синтетически расширенного набора данных

В данном подразделе формализован заключительный этап обнаружения РМЖ с использованием глубоких CNN. Классификационная модель применяет трансферное обучение в режиме *fine-tuning*. Обучение происходит на расширенной обучающей выборке с помощью генеративной модели DGAN-WP-TL. Для балансировки выбранных датасетов в подразделе 2.4, синтетические данные добавляются только в *train*-выборку, чтобы избежать искажений предсказательной способности классификационной модели на валидационных и тестовых реальных данных и обеспечить корректность результатов.

Расширенная обучающая выборка определяется как (10) [12; 13; 14, p. 101552]:

$$X^{train} = X_{real}^{train} \cup X_{gen}, Y^{train} = Y_{real}^{train} \cup Y_{gen}, \quad (10)$$

где X_{real}^{train} – реальные изображения обучающего множества, а

X_{gen} – синтетические изображения малочисленного (*malignant*) класса, сгенерированные моделью DGAN-WP-TL.

Y_{gen} – метки синтетических изображений, которые для всех X_{gen} задаются как принадлежащие малочисленному классу *malignant* (метка 1), поскольку генеративная модель DGAN-WP-TL обучается на изображениях данного класса.

Такое объединение обеспечивает увеличение доли изображений малочисленного класса в обучающей выборке.

Валидационное и тестовое подмножества остаются неизменными (11) [12; 13; 14, p. 101552]:

$$X^{val} = X_{real}^{val}, X^{test} = X_{real}^{test}, \quad (11)$$

На рисунке 11 показана схема добавления синтетических изображений только в обучающую выборку. Далее расширенная выборка X^{train} участвует в обучении CNN-классификаторов.

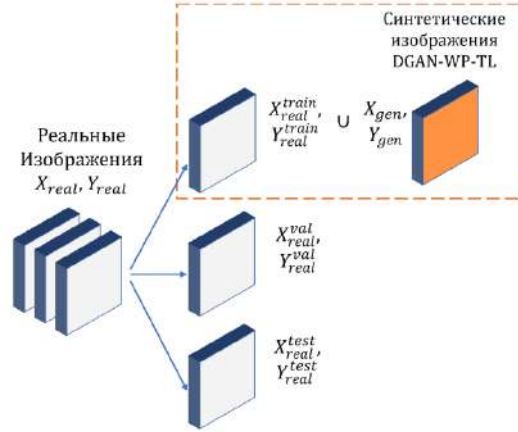


Рисунок 11 – Формирование обучающего, валидационного и тестового подмножеств при использовании синтетических изображений DGAN-WP-TL (синтетические данные добавляются только в train)

Классификатор предобучается на весах ImageNet и адаптируется под задачу бинарной классификации опухолей. Финальный полносвязный слой заменяется на слой с количеством нейронов, равным числу классов, и функцией активации Softmax на выходе (12) [102, p. 1345-1358; 103; 104]:

$$f_{\theta}(x) = g_{\theta_c}(\phi_{\theta_f}(x)), x \in X^{train}, \quad (12)$$

где ϕ_{θ_f} – сверточная часть сети, извлекающая признаки;

g_{θ_c} – классификатор, формирующий выходной вектор вероятностей.

Совокупность параметров всей сети обозначается как $\theta = \{\theta_f, \theta_c\}$, где θ_f – набор параметров (весов и смещений) свёрточных слоёв, инициализируемых из предобученной модели; θ_c – параметры полносвязных слоёв, обучаемых с нуля.

При режиме тонкой настройки (fine-tuning), предобученные веса θ_f частично размораживаются и дообучаются совместно с θ_c на расширенном множестве данных X^{train} .

Оптимизация параметров сети выполняется по функции потерь кросс-энтропии [102, p. 1345-1358; 103; 104, p. 533-535; 105]:

$$\theta_f, \theta_c = \arg \min_{\theta_f, \theta_c} L_{CE}(f_{\theta}(X^{train}), Y^{train}), \quad (13)$$

где Y – вектор истинных меток классов, а

L_{CE} – функция потерь кросс-энтропии, минимизация которой осуществляется методом Adam для ускорения и стабилизации сходимости.

Алгоритм 2 (рисунок 12) предоставляет процедуру обучения классификационной модели.

Алгоритм 2: Тонкая настройка классификатора на синтетически расширенной выборке
Шаг 1: Входные данные и подготовка данных
- Даны реальные изображения (X_{real}, Y_{real}) и синтетические изображения миноритарного класса (X_{gen}, Y_{gen}). - Разделить реальные изображения на три подвыборки train/validation/test subsets (70/20/10). - Расширить обучающую выборку с помощью синтетических изображений: • $X^{train} = X_{real}^{train} \cup X_{gen}, Y^{train} = Y_{real}^{train} \cup Y_{gen}$.
Шаг 2: Предварительная обработка изображений Image Pre-processing
- Привести все изображения к единому размеру и формату. - Провести нормализацию пикселей в диапазоне [0,1].
Шаг 3: Инициализация классификатора с использованием трансферного обучения
- Загрузить предобученную модель (например, VGG16). - Заменить последний слой новым полносвязным слоем с 2-мя категориями Dense (benign/malignant) + Softmax. - Обозначить θ_f в качестве весов экстрактора, θ_c – весов классификатора - Инициализировать θ_f из предобученной модели; а θ_c – случайно.
Шаг 4: Гиперпараметры обучения
- Функция потерь: Cross-Entropy Loss L_{CE} - Оптимизатор: Adam, $lr = 1 \times 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ - Размер батча = 8, обучение в течение 20 эпох.
Шаг 5: Обучение на расширенной обучающей выборке
- Для каждой эпохи ($t = 1, 2, \dots, EPOCHS$): • Для каждого мини-батча (x_b, y_b) из обучающей выборки (X^{train}, Y^{train}) ○ Выполнить прямой проход и сделать предсказания; ○ Посчитать функцию потерь; ○ Выполнить обратное распространение градиентов; ○ Обновить веса θ_f и θ_c. • Оценить модель на валидационной выборке и сохранить лучшую модель.
Шаг 6: Метрики классификации
- Загрузить обученную модель f_{θ}. - Классифицировать тестовую выборку (X^{test}, Y^{test}) - Посчитать метрики классификации

Рисунок 12 – Процедура обучения CNN-классификатора для задачи выявления опухолей РМЖ с использованием синтетически расширенного множества данных

Общий пайплайн обучения классификатора на расширенной выборке показан на рисунке 13.

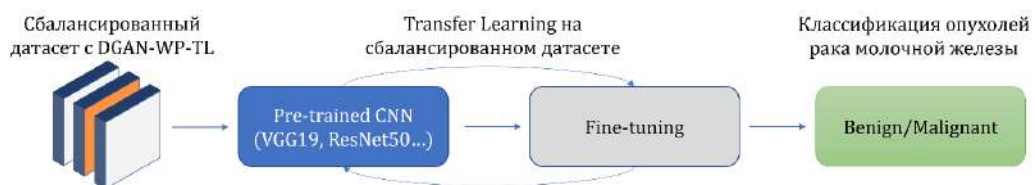


Рисунок 13 – Пайплайн классификации опухолей молочной железы на основе синтетически расширенной обучающей выборки и fine-tuning предобученной CNN

В рамках данного исследования предложена и реализована схема интеграции синтетических изображений, сгенерированных моделью DGAN-WP-TL, в процесс обучения свёрточного классификатора опухолей РМЖ. Ключевые особенности предлагаемого пайплайна:

1. Разработан метод DGAN-WP-TL, предназначенный для синтетического расширения обучающей выборки для балансировки миноритарного класса. При этом валидационная и тестовая выборки остаются без изменений с оригинальными изображениями для избежания смещения предсказательной способности модели к мажоритарному классу.

2. Разработан пайплайн классификатора с интеграцией трансферного обучения на базу тонкой настройки с использованием предобученных весов. Обучение классификатора происходит на расширенной обучающей выборке, которая включает только реальные и синтетические изображения генеративной модели DGAN-WP-TL.

Далее в третьей главе проведены эксперименты для валидации разработанного метода DGAN-WP-TL и его влияния на классификационную часть автоматизированной системы обнаружения РМЖ в качестве image-level classification. Для обеспечения сопоставимости экспериментальных результатов в подразделе 2.7 рассматриваются метрики, используемые для оценки качества генерации и классификации.

2.7 Метрики оценки качества синтетических изображений и классификации

Результаты генеративной и классификационной моделей необходимо оценить с помощью метрик генерации и классификации. Процесс синтеза оценивается с точки зрения реалистичности, разнообразия и восприятия человеком (т.е. перцептуальности), а классификация с точки зрения клинически значимых метрик.

Метрики генерации. Для оценки реалистичности и разнообразия используется метрика *Fréchet Inception Distance (FID)*, которая измеряет близость синтетических изображений к реальным. FID измеряет расстояние между распределениями признаков реальных и синтетических изображений [50]. Признаки извлекаются из промежуточных слоёв предобученной InceptionV3, после чего сравниваются их средние значения и ковариации. Формально FID определяется выражением (14):

$$FID = \|\mu_r - \mu_g\|_2^2 + Tr\left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}\right), \quad (14)$$

где μ_r, μ_g обозначают средние значения признаков;

Σ_r, Σ_g – их ковариационные матрицы;

$Tr(\cdot)$ – след матрицы. Низкие значения FID свидетельствуют о близком совпадении распределений признаков реальных и синтетических изображений и, следовательно, о высокой реалистичности синтеза. Следует отметить, что FID

чувствителен к размеру выборки, что требует осторожной интерпретации при анализе медицинских датасетов малого объёма.

Для определения сходства распределений между реальными x и синтетическими y изображениями именно в малых выборках применяется Kernel Inception Distance (KID). Для калькуляции KID применяется уравнение (15) [51]:

$$KID(x, y) = MMD^2(\varphi(x), \varphi(y)), \quad (15)$$

где MMD – ядровая функция;

φ – отображение признаков.

Высокий KID показывает, что синтетические изображения имеют сильное расхождение с реальными, тогда как низкие значения метрики о близком сходстве с реальными.

Для перцептивного анализа синтетических изображений используется метрика *Learned Perceptual Image Patch Similatory (LPIPS)* [52, p.586-594], которая позволяет определять, насколько хорошо синтетические изображения воспринимаются человеком. Чем ниже результаты метрики между реальными и сгенерированными изображениями, тем выше сходство синтетических изображений по восприятию человеком. При этом сравнение между синтетическими изображениями говорит о разнообразии и отсутствии коллапса мод.

Для оценки структурного качества, а именно совпадают ли текстуры, яркость, контраст с реальными изображениями, применяется *Multi-Scale Structural Similarity Index (MS-SSIM)* [53, p. 1398-1401]. Высокие значения MS-SSIM (r-f) указывают, что синтетические данные сходны по яркости, контрасту и текстуре с реальными. В свою очередь, низкие значения MS-SSIM (f-f) показывают рост разнообразия, что свидетельствует о генерации вариативных изображений.

Дополнительная структурная метрика Sliced Wasserstein Distance (SWD) оценивает совпадение локальных текстурных деталей между синтетическими и реальными изображениями [106]. Низкие значения SWD показывают совпадение с реальными, как на примере с FID и KID.

Метрики классификации. Эффективность классификаторов оценивается несколькими стандартными метриками. *Accuracy* отражает общую долю верно классифицированных изображений (16) [56, p. 325-326]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (16)$$

где TP (True Positive) – количество верно предсказанных положительных случаев;

TN (True Negative) – количество верно предсказанных отрицательных случаев;

FP (False Positive) – количество отрицательных случаев, ошибочно классифицированных как положительные;

FN (False Negative) – количество положительных случаев, ошибочно классифицированных как отрицательные.

Эффективность классификационных моделей в условиях выраженного дисбаланса классов оценивается не только по Accuracy, но прежде всего по метрикам, отражающим способность модели корректно выявлять злокачественные случаи и отличать их от доброкачественных. К таким метрикам относятся *Precision*, *Recall*, *F1-score*, а также графический и интегральный анализ ROC-кривой и площади под ней (AUC-ROC).

Precision отражает долю истинно злокачественных случаев среди всех изображений, которые модель отнесла к malignant. Формально она определяется как (17) [57, p. 35-38]:

$$Precision = \frac{TP}{TP + FP}, \quad (17)$$

Высокое значение *Precision* указывает на низкую частоту ложноположительных ошибок. Это важно в клинической практике, поскольку каждая ложноположительная классификация может приводить к дополнительным диагностическим процедурам и увеличению нагрузки на систему здравоохранения. Низкое значение *Precision* указывает на высокую долю ложноположительных решений.

Recall (Чувствительность или *Sensitivity*) измеряет способность модели обнаруживать злокачественные опухоли среди всех реально имеющихся malignant-случаев (18) [57, p. 35-38]:

$$Recall = \frac{TP}{TP + FN}, \quad (18)$$

Recall является ключевой метрикой, поскольку отражает долю выявленных malignant случаев и связан с риском пропуска ложноотрицательных случаев опухолей (FN).

Поскольку *Precision* и *Recall* описывают разные стороны работы классификатора, их необходимо рассматривать совместно. Для этой цели используется *F1-score*, являющийся гармоническим средним между ними (19) [57, p. 35-38]:

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}, \quad (19)$$

Высокий *F1-score* достигается только в том случае, если и *Precision*, и *Recall* имеют высокие значения одновременно.

Дополнительное понимание поведения модели обеспечивает анализ ROC-кривой (*Receiver Operating Characteristic*), графически описывающей зависимость между чувствительностью и уровнем ложноположительных

классификаций при изменении порога решения. ROC-кривая демонстрирует, как модель балансирует между обнаружением опухолей и предотвращением ложных тревог на всех возможных порогах. Модель, способная уверенно отделять злокачественные изображения от доброкачественных, будет иметь ROC-кривую, стремящуюся к левому верхнему углу графика [58, p. 1285-1292].

Количественно это поведение выражается через площадь под ROC-кривой (AUC-ROC). Значения AUC, близкие к 1, отражают высокую дискриминационную способность модели; значения около 0.5 соответствуют случайному угадыванию. В медицинской диагностике высокое значение AUC свидетельствует о том, что модель способна надёжно различать *benign* и *malignant* в широком диапазоне условий и остаётся устойчивой к выбору порога принятия решений. Низкие значения AUC указывают на то, что модель не обладает достаточной чувствительностью и спецификой для клинического использования [107].

Совместное применение Precision, Recall, F1-score и AUC-ROC позволяет получить всестороннее представление о качествах классификатора. В контексте данной работы эти метрики особенно важны, поскольку позволяют количественно выявить, усиливает ли синтетическая выборка, сгенерированная DGAN-WP-TL, способность модели к обнаружению редких злокачественных случаев, что соответствует основной цели устранения дисбаланса данных и повышения чувствительности автоматизированных методов обнаружения РМЖ.

Выводы по второму разделу

В результате проведённого исследования выполнено математическое моделирование и разработка метода обнаружения РМЖ, объединяющего генеративные и CNN модели. Основные результаты заключаются в следующем:

1. Формализована математическая основа предлагаемого метода, включающая постановку задачи классификации опухолей РМЖ и описание параметризованного отображения CNN в условиях ограниченных и несбалансированных медицинских данных.

2. Рассмотрены и систематизированы теоретические основы генеративных моделей GAN, WGAN и WGAN-GP, включая использование расстояния Вассерштейна и градиентного штрафа, обеспечивающих устойчивость обучения и снижение риска *mode collapse* при синтезе медицинских изображений в условиях дисбаланса классов, что послужило теоретическим обоснованием выбора архитектуры WGAN-GP в рамках предлагаемого метода.

3. Разработана двухветвевая архитектура генеративной модели DGAN-WP-TL, в которой объединяются признаки, извлекаемые усечённой предобученной сетью VGG19 (*block2_pool*), и обучаемые признаки собственной свёрточной ветви критика. Такое архитектурное решение расширяет пространство оценки реалистичности изображений за счёт одновременного учёта предобученных и адаптивных признаков представлений.

4. Предложен метод синтетического расширения обучающего множества, при котором изображения малочисленного класса, сгенерированные моделью DGAN-WP-TL, добавляются исключительно в обучающее подмножество.

Данная стратегия обеспечивает баланс классов без изменения валидационной и тестовой выборок и является составной частью общего метода обнаружения РМЖ.

5. Сформирована интегрированная схема обучения классификатора, основанная на fine-tuning предобученной CNN на синтетически расширенной обучающей выборке в условиях дисбаланса классов.

6. В рамках предложенной архитектуры DGAN-WP-TL выполнено исследование влияния коэффициента градиентного штрафа λ в диапазоне 0.5-10.0 на устойчивость обучения критика и качество синтезируемых изображений. Детальный анализ влияния параметра λ на процесс генерации и последующую классификацию представлен в разделе 3.

7. Сформирован и теоретически обоснован комплекс метрических критериев для оценки качества генерации и последующей классификации, включающий распределительные и перцептуальные метрики (FID, KID, LPIPS, MS-SSIM), а также диагностические показатели классификации (Precision, Recall, F1-score, AUC). Метрический аппарат обеспечивает объективную количественную характеристику реалистичности, разнообразия и диагностической значимости синтетических изображений.

3 ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ И РЕЗУЛЬТАТЫ

В данной диссертационной работе градиентный штраф обозначается только как λ и используется только в этом значении [13]. Другие коэффициенты генеративной модели не вводятся.

3.1 Экспериментальные исследования генеративной модели DGAN-WP-TL

В подразделе 3.1 отображены результаты генеративного синтеза для аугментации медицинских датасетов с целью расширения миноритарного класса и устранения дисбаланса. Генеративная аугментация реализована с помощью формализованного метода DGAN-WP-TL и является вспомогательным модулем в автоматизированной системе обнаружения РМЖ [67, р. 2556687]. В качестве выборок были использованы три медицинских датасета, представленных в подразделе 2.4: ультразвуковой BUSI [47, р. 104863], маммографический CBIS-DDSM [48] и набор MPT-изображений Alzheimer MRI [49]. В случае Alzheimer MRI датасет выбран для дополнительной апробации генеративного метода на другом домене.

Исходя из выводов второго раздела, метод DGAN-WP-TL применен только к миноритарному классу трех датасетов. Количество эпох составляло 500 с подсчетом метрик каждые 100 эпох при диапазоне $\lambda \in \{0.5; 3.0; 5.0; 7.0; 10.0\}$.

Из представленных метрик в подразделе 2.7 использован комплекс метрик генерации FID, KID, LPIPS и MS-SSIM. FID и KID являются распределительными метриками, показывающими сходство распределений реальных и синтетических изображений. Тогда как LPIPS используется для перцептивной оценки, которая показывает восприятие человеком. В свою очередь, MS-SSIM оценивает структурное сходство и разнообразие. Стоит отметить, что KID является приоритетной метрикой в малых выборках, чем FID, которая чувствительна и нестабильна в малых выборках. Наилучшие значения метрик FID и KID достигаются при минимальных значениях, тогда как высокие значения метрик говорят о низком сходстве распределений между реальными и сгенерированными изображениями.

Обе метрики LPIPS и MS-SSIM вычисляются попарно: «реальное-синтетическое» (r-f) и «синтетическое-синтетическое» (f-f). Если LPIPS (r-f) показывает низкие значения, то это означает, что синтетические изображения похожи на реальные в перцептуальном смысле, т.е. по восприятию человеком. А высокие значения LPIPS (f-f), наоборот, показывают, что синтетические изображения разнообразны перцептуально, тем самым исключается генерация одинаковых синтетических изображений. Метрика MS-SSIM является структурной метрикой и анализирует яркость, контрастность и текстуры. MS-SSIM (r-f) с низкими значениями показывает отсутствие копирования реальных изображений, а низкие значения MS-SSIM (f-f) соответствует разнообразию на структурном уровне, что также говорит об исключении коллапса мод.

Комбинация перцептуальной LPIPS и структурной MS-SSIM метрик показывает, что сгенерированные изображения вариативны, не копируют реальные и исключают одинаковые структуры.

В подразделе 3.1 эксперименты решают 4 задачи: 1) исследуется диапазон градиентного штрафа и проводится оценка качества генерации; 2) интерпретируются полученные результаты генерации на мультимодальных датасетах (BUSI, CBIS-DDSM, Alzheimer MRI) с целью выявления оптимального значения λ от типа медицинских изображений; 3) анализируется степень согласованности распределений реальных и синтетических данных в признаковом пространстве с использованием PCA и t-SNE, а также количественных мер Jensen-Shannon Divergence и Wasserstein Distance; и 4) выполняется сравнительная оценка предложенной модели DGAN-WP-TL с baseline-архитектурой StyleGAN2, позволяющая сопоставить устойчивость обучения и разнообразие генерации в условиях малых и экстремально малых выборок [67, p. 2556687].

3.1.1 Результаты DGAN-WP-TL на датасете BUSI

Эксперименты на датасете BUSI были организованы в соответствии с описанной выше схемой: для каждого значения λ модель обучалась 500 эпох, а метрики качества генерации вычислялись через каждые 100 эпох. Результаты сведены в таблицу 7.

Таблица 7 – Значения метрик качества генерации модели DGAN-WP-TL на датасете BUSI при различных λ и эпохах обучения

λ	Эпохи	KID ↓	FID ↓	LPIPS (r-f) ↓	LPIPS (f-f) ↑	MS-SSIM (r-f) ↓	MS-SSIM (f-f) ↓
1	2	3	4	5	6	7	8
0.5	100	0.4189	369.2397	0.5708	0.1849	0.1952	0.4743
	200	0.3892	351.1959	0.5853	0.2564	0.1908	0.4125
	300	0.3677	329.3237	0.5651	0.3873	0.1958	0.2029
	400	0.2451	257.9432	0.5364	0.4304	0.1921	0.1866
	500	0.2433	251.5650	0.5414	0.4369	0.1862	0.1735
3.0	100	0.4923	404.0282	0.7023	0.2862	0.1957	0.2110
	200	0.2825	282.1408	0.5371	0.4095	0.1833	0.1740
	300	0.2322	244.4819	0.5456	0.4544	0.1905	0.1995
	400	0.1680	198.4747	0.5261	0.4791	0.1787	0.1746
	500	0.1670	193.9128	0.5082	0.4689	0.1722	0.1795
5.0	100	0.4351	373.7834	0.6700	0.2913	0.2076	0.2757
	200	0.3279	304.9025	0.5371	0.4019	0.1687	0.1701
	300	0.2352	239.9935	0.5324	0.4610	0.1800	0.1787
	400	0.1714	198.0696	0.5146	0.4760	0.1912	0.2048
	500	0.1448	179.4158	0.5113	0.4862	0.1822	0.1760
7.0	100	0.4829	395.2223	0.7090	0.2890	0.2018	0.2271
	200	0.3025	295.5860	0.5294	0.3690	0.1927	0.1912
	300	0.2337	244.3256	0.5312	0.4296	0.1900	0.1891
	400	0.1913	212.9755	0.5228	0.4774	0.1866	0.1809

Продолжение таблицы 7

1	2	3	4	5	6	7	8
	500	0.1843	202.0724	0.5096	0.4770	0.1943	0.1882
10.0	100	0.3768	341.9732	0.7343	0.2915	0.2014	0.2721
	200	0.3037	293.7198	0.5595	0.4088	0.1547	0.1546
	300	0.2351	246.4043	0.5520	0.4411	0.1759	0.1635
	400	0.1667	198.1975	0.5386	0.4903	0.1638	0.1414
	500	0.1730	198.6400	0.5294	0.4993	0.1857	0.1941

Наилучшие показатели на BUSI достигнуты при $\lambda=5.0$ на 500-й эпохе: минимальное значение KID (0.1448), одно из самых низких значений FID (179.4158), высокая перцептивная вариативность LPIPS (f-f) = 0.4862 и низкие значения MS-SSIM как для r-f (0.1822), так и для f-f (0.1760) (таблица 7).

Качественная оценка синтетических изображений на датасете BUSI представлена на рисунке 14. На ранних этапах обучения наблюдаются сглаженные структуры и недостаточная выраженность границ опухолевых образований. По мере увеличения числа эпох визуальное качество синтеза улучшается, формируются более детализированные изображения с различимыми тканевыми структурами и очертаниями очагов.

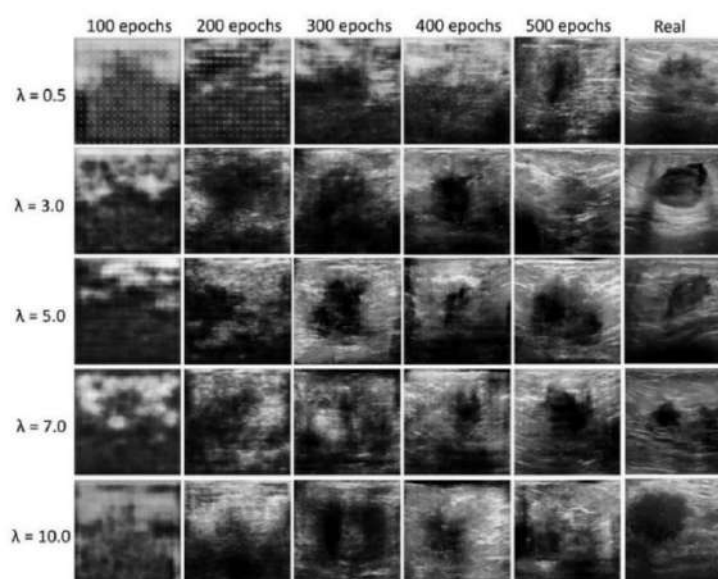
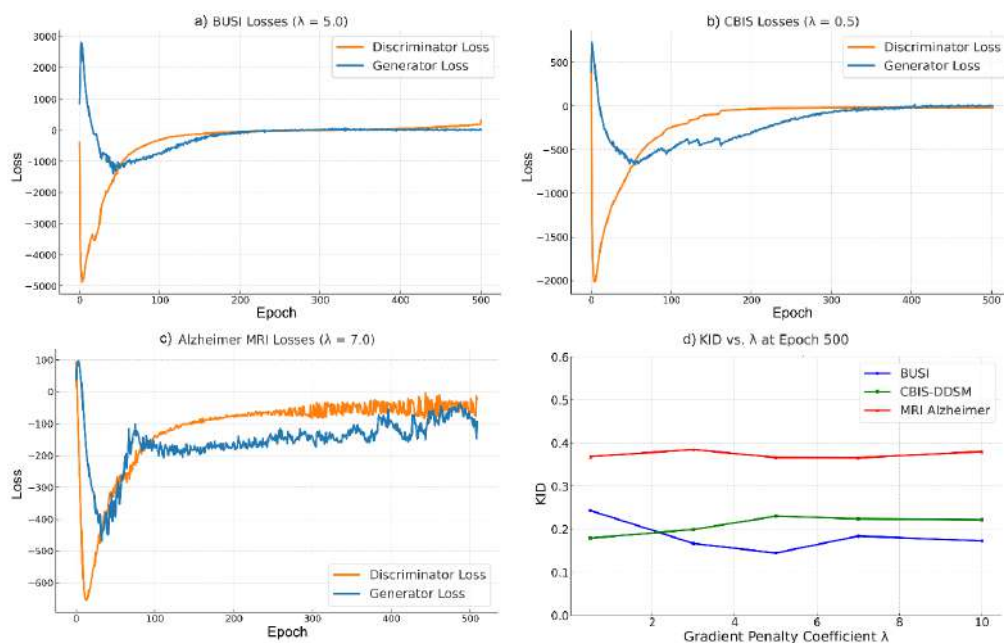


Рисунок 14 – Сгенерированные изображения malignant-класса BUSI датасета с 100 по 500 эпоху

Примечание – В правом столбце представлены реальные изображения

Функции потерь генератора и критика отображены на рисунке 15a при $\lambda = 5.0$. Графики демонстрируют резкий скачок функции потерь генератора, в то время как критик резко уменьшается. Данное поведение характеризуется тем, что в начале критик преобладает над генератором. Однако далее наблюдается стабилизация динамики потерь, что косвенно указывает на устойчивость обучения (рисунок 15b и 15c).



а – функции потерь генератора и критика (дискриминатора) для BUSI (оптимальное λ); б – функции потерь для CBIS-DDSM; с – функции потерь для Alzheimer MRI; д – значения KID на 500-й эпохе для различных λ и всех трёх датасетов

Рисунок 15 – Кривые обучения и влияние коэффициента λ

Следует сделать вывод, что в результате экспериментов над ультразвуковыми изображениями BUSI датасета, DGAN-WP-TL улучшает качество генерации синтетических изображений при соответствующем коэффициенте градиентного штрафа. Наилучший баланс между распределительным сходством и разнообразием сгенерированных данных достигается при значении $\lambda=5.0$, что подтверждается минимальными значениями KID и FID, а также показателями LPIPS и MS-SSIM.

3.1.2 Результаты DGAN-WP-TL на датасете CBIS-DDSM

Как и на BUSI датасете, на CBIS-DDSM метод был применён только к злокачественным новообразованиям. Диапазон коэффициента градиентного штрафа и полученные результаты генерации представлены в таблице 8.

Таблица 8 – Значения метрик качества генерации модели DGAN-WP-TL на датасете CBIS-DDSM

λ	Эпохи	KID ↓	FID ↓	LPIPS (r-f) ↓	LPIPS (f-f) ↑	MS-SSIM (r-f) ↓	MS-SSIM (f-f) ↓
1	2	3	4	5	6	7	8
0.5	100	0.4845	364.9807	0.6890	0.1739	0.4401	0.5684
	200	0.2770	245.9964	0.3106	0.2519	0.5777	0.5752
	300	0.2148	206.7753	0.2888	0.2592	0.5633	0.5630
	400	0.1961	195.0377	0.2699	0.2736	0.5910	0.5552
	500	0.1795	182.3454	0.2870	0.2875	0.5182	0.4877

Продолжение таблицы 8

1	2	3	4	5	6	7	8
3.0	100	0.3623	296.5409	0.7962	0.1626	0.3442	0.6062
	200	0.2638	241.1752	0.3187	0.2715	0.5666	0.5549
	300	0.2209	209.1812	0.2855	0.2669	0.5699	0.5583
	400	0.2058	196.0290	0.2787	0.2610	0.5800	0.5861
	500	0.1992	196.2727	0.2929	0.2986	0.5536	0.5179
5.0	100	0.4015	316.9537	0.7932	0.1619	0.3471	0.5762
	200	0.3424	287.5098	0.3890	0.2705	0.5545	0.5450
	300	0.2584	236.3212	0.3102	0.2705	0.5656	0.5460
	400	0.2324	219.2736	0.2917	0.2735	0.5712	0.5558
	500	0.2303	215.2661	0.2839	0.2652	0.5728	0.5562
7.0	100	0.5081	372.8909	0.6377	0.2440	0.4931	0.5205
	200	0.2974	259.9855	0.3473	0.2481	0.5639	0.5761
	300	0.2515	231.4126	0.2967	0.2609	0.5756	0.5526
	400	0.2372	222.4938	0.2927	0.2735	0.5739	0.5456
	500	0.2241	212.8649	0.2950	0.2752	0.5474	0.5313
10.0	100	0.4594	352.4138	0.6785	0.1717	0.4293	0.5439
	200	0.2936	256.5190	0.3738	0.2705	0.5471	0.5398
	300	0.2363	221.8855	0.3031	0.2732	0.5738	0.5356
	400	0.2353	220.6487	0.2915	0.2782	0.5727	0.5333
	500	0.2218	210.1113	0.2880	0.2647	0.5703	0.5578

Лучшие значения KID (0.1795) и FID (182.3454) достигнуты при $\lambda = 0.5$ на 500-й эпохе. В этой конфигурации также наблюдаются сравнительно низкие LPIPS (r-f) = 0.2870 и MS-SSIM (r-f) = 0.5182 при умеренных значениях LPIPS (f-f) и снижении MS-SSIM (f-f) до 0.4877 таблицы 8, что указывает на сохранение структурного разнообразия синтетических изображений.

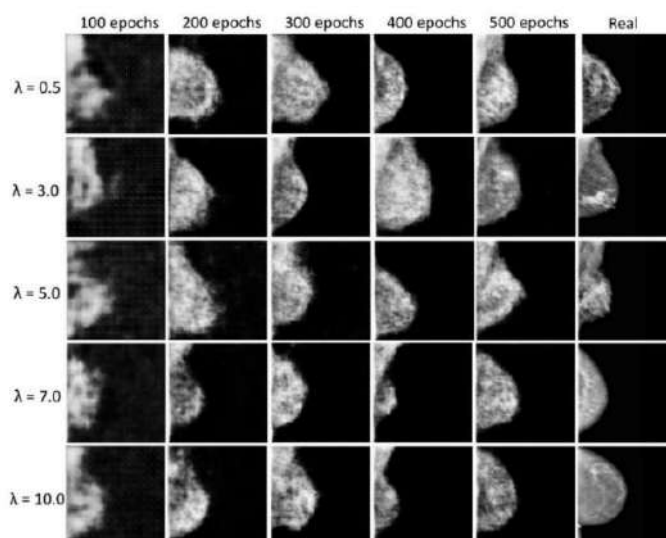


Рисунок 16 – Примеры сгенерированных маммографических изображений CBIS-DDSM для $\lambda = 0.5; 3.0; 5.0; 7.0; 10.0$ и эпох 100–500

Примечание – В крайнем правом столбце приведены реальные изображения для сравнения

Рисунок 16 показывает примеры сгенерированных изображений на 100-500 эпохах, где на 100-300 эпохах контуры размыты, на 400-500 визуальное изображение более похоже на реальные и количество артефактов уменьшается.

Функции потерь для маммографического датасета (рисунок 15b) медленно сходятся, колеблясь по сравнению с BUSI. Данное поведение связано со структурой молочной железы на маммографических снимках. Однако к 400-500 эпохам обучение стабилизируется, на что указывают метрики генерации и их улучшение.

Таким образом, согласно распределительным, перцептуальным и структурным метрикам, DGAN-WP-TL метод эффективно генерирует маммографические снимки. При этом, малые значения коэффициента градиентного штрафа, позволяют генерировать более реалистичные и разнообразные изображения, которые будут использованы для балансировки миноритарного класса в классификационной модели.

3.1.3 Дополнительная апробация метода DGAN-WP-TL на датасете Alzheimer MRI

Поскольку метод DGAN-WP-TL ориентирован на работу с медицинскими датасетами, для подтверждения его эффективности была проведена дополнительная апробация на независимом датасете Alzheimer MRI. Междоменный эксперимент позволяет проанализировать устойчивость переноса метода между модальностями, тем самым симитировать межцентровую гетерогенность.

Следует подчеркнуть, что данный эксперимент носит методический характер и не предполагает клинической интерпретации результатов для датасета Alzheimer MRI.

Согласно описанию датасетов в подразделе 2.4 для Alzheimer MRI миноритарным классом является Moderate Demented, который нуждается в балансировке с помощью генеративной аугментации. Диапазон коэффициента градиентного штрафа и процесс анализа остаются подобными, как и в двух предыдущих датасетах. Результаты экспериментов представлены в таблице 9.

Таблица 9 – Значения метрик качества генерации модели DGAN-WP-TL на датасете Alzheimer MRI

λ	Эпохи	KID ↓	FID ↓	LPIPS (r-f) ↓	LPIPS (f-f) ↑	MS-SSIM (r-f) ↓	MS-SSIM (f-f) ↓
1	2	3	4	5	6	7	8
0.5	100	0.4571	329.1758	0.6892	0.2247	0.2454	0.4182
	200	0.5307	368.3591	0.4754	0.2849	0.4442	0.5087
	300	0.5557	393.5698	0.3942	0.2720	0.4944	0.4656
	400	0.4336	333.2770	0.3199	0.2745	0.4908	0.4825
	500	0.3685	298.9347	0.2966	0.2754	0.5024	0.4877
3.0	100	0.5853	397.5363	0.6670	0.2835	0.2940	0.3085
	200	0.6256	415.3796	0.4880	0.2847	0.4409	0.4361
	300	0.5351	384.0563	0.3922	0.2978	0.4793	0.4507

Продолжение таблицы 9

1	2	3	4	5	6	7	8
	400	0.4254	329.8052	0.3298	0.2804	0.4942	0.4799
	500	0.3844	306.2459	0.2907	0.2654	0.5127	0.5022
5.0	100	0.5115	357.2624	0.6856	0.2679	0.2887	0.3203
	200	0.6011	405.0012	0.5014	0.2880	0.4305	0.4553
	300	0.5437	384.2440	0.3893	0.2880	0.4863	0.4610
	400	0.4596	342.3627	0.3178	0.2695	0.5037	0.4872
	500	0.3664	294.7207	0.2810	0.2487	0.5164	0.5269
7.0	100	0.5615	385.1168	0.6790	0.2587	0.2947	0.3361
	200	0.6088	409.2390	0.4904	0.2663	0.4235	0.4504
	300	0.5198	375.6805	0.3799	0.2855	0.4984	0.4860
	400	0.4098	323.3915	0.3131	0.2695	0.5172	0.4996
	500	0.3657	289.7888	0.2746	0.2428	0.5240	0.5240
10.0	100	0.4986	351.3085	0.7033	0.2375	0.2194	0.3191
	200	0.5925	399.2790	0.5353	0.2663	0.4299	0.4634
	300	0.5418	385.8337	0.3826	0.2702	0.4771	0.4787
	400	0.4206	327.5089	0.3087	0.2712	0.5259	0.5043
	500	0.3799	291.0406	0.2826	0.2530	0.5082	0.5083

В отличие от BUSI и CBIS-DDSM, где оптимальные значения λ были получены при значениях 0.5 и 5.0, в Alzheimer MRI оптимум достигается при $\lambda = 7.0$ на 500-й эпохе. В этой конфигурации модель демонстрирует минимальное значение KID (0.3657) и FID (289.7888), а также минимальное LPIPS(r-f) = 0.2746 отражает перцептивную близость синтетических изображений к реальным, тогда как относительно низкое LPIPS(f-f) = 0.2428 указывает на ограниченное перцептивное разнообразие по сравнению с альтернативными конфигурациями. При этом значения MS-SSIM и распределительные метрики свидетельствуют о приемлемом компромиссе между устойчивостью обучения и вариативностью генерации (таблица 9).

Рисунок 17 показывает качественные результаты сгенерированных изображений для выбранного диапазона λ . На эпохах 100-300 метод DGAN-WP-TL на Alzheimer MRI генерирует сглаженные структуры, на 400-500 изображения становятся более четче и похожими на реальные.

Генерация Alzheimer MRI является более сложной, так как человеческий мозг отличается по структуре по сравнению с более шумными УЗИ-изображениями. Это подтверждается значительными колебаниями на первых эпохах, однако далее обучение стабилизируется, что и демонстрируют метрики генерации. Полученные результаты подтверждают устойчивость метода DGAN-WP-TL на междоменном датасете, что косвенно свидетельствует о межцентровой гетерогенности и возможности применения для генеративной аугментации миноритарного класса.

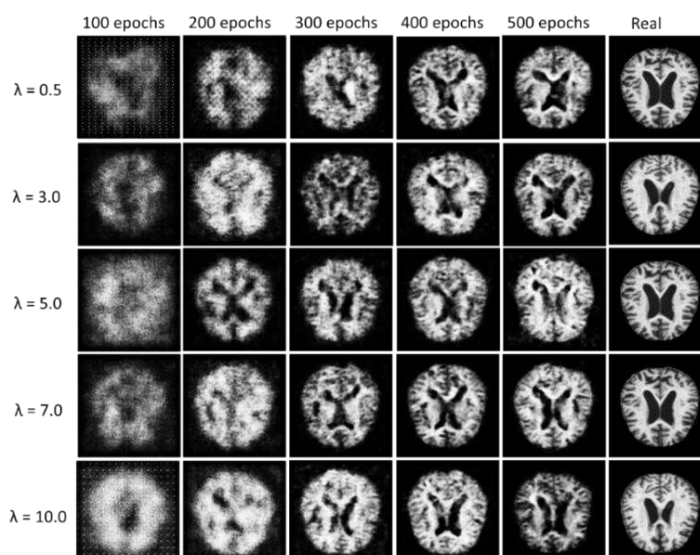


Рисунок 17 – Примеры сгенерированных изображений класса Moderate Demented на датасете Alzheimer MRI для различных значений λ и эпох обучения

Примечание – В правом столбце показаны реальные изображения

Дополнительно был проведён анализ влияния коэффициента градиентного штрафа λ . Обобщённые значения метрики KID на 500-й эпохе представлены на рисунке 15d. Анализ показывает, что оптимальное значение λ зависит от модальности данных: для BUSI наилучшие результаты достигаются при $\lambda = 5.0$, для CBIS-DDSM – при малых значениях $\lambda = 0.5$, а для Alzheimer MRI – при более высоких значениях $\lambda = 7.0$.

Таким образом, дополнительная апробация метода DGAN-WP-TL на независимом датасете Alzheimer MRI продемонстрировала его устойчивость и обобщающую способность в условиях экстремально малых и несбалансированных медицинских выборок. Показано, что модель способна формировать синтетические изображения, сохраняющие характеристики реальных данных, при корректном выборе коэффициента градиентного штрафа. Полученные результаты подтверждают применимость предложенного подхода не только в задачах генерации изображений РМЖ, но и при междоменном переносе, что повышает универсальность метода и обоснованность его использования для балансировки данных в задачах медицинской диагностики.

3.1.4 Анализ распределений синтетических и реальных изображений с использованием PCA и t-SNE

Наряду с количественными метриками FID, KID, LPIPS и MS-SSIM в данном подразделе рассматриваются методы анализа распределений и визуализации признакового пространства – Principal Component Analysis (PCA) [54, p. 20150202] и t-distributed Stochastic Neighbor Embedding (t-SNE) [55, p. 2579-2604]. Эти методы позволяют проанализировать взаимное расположение признаков реальных и синтетических изображений в двумерном пространстве и

оценить степень воспроизведения структуры исходного распределения (рисунок 18).

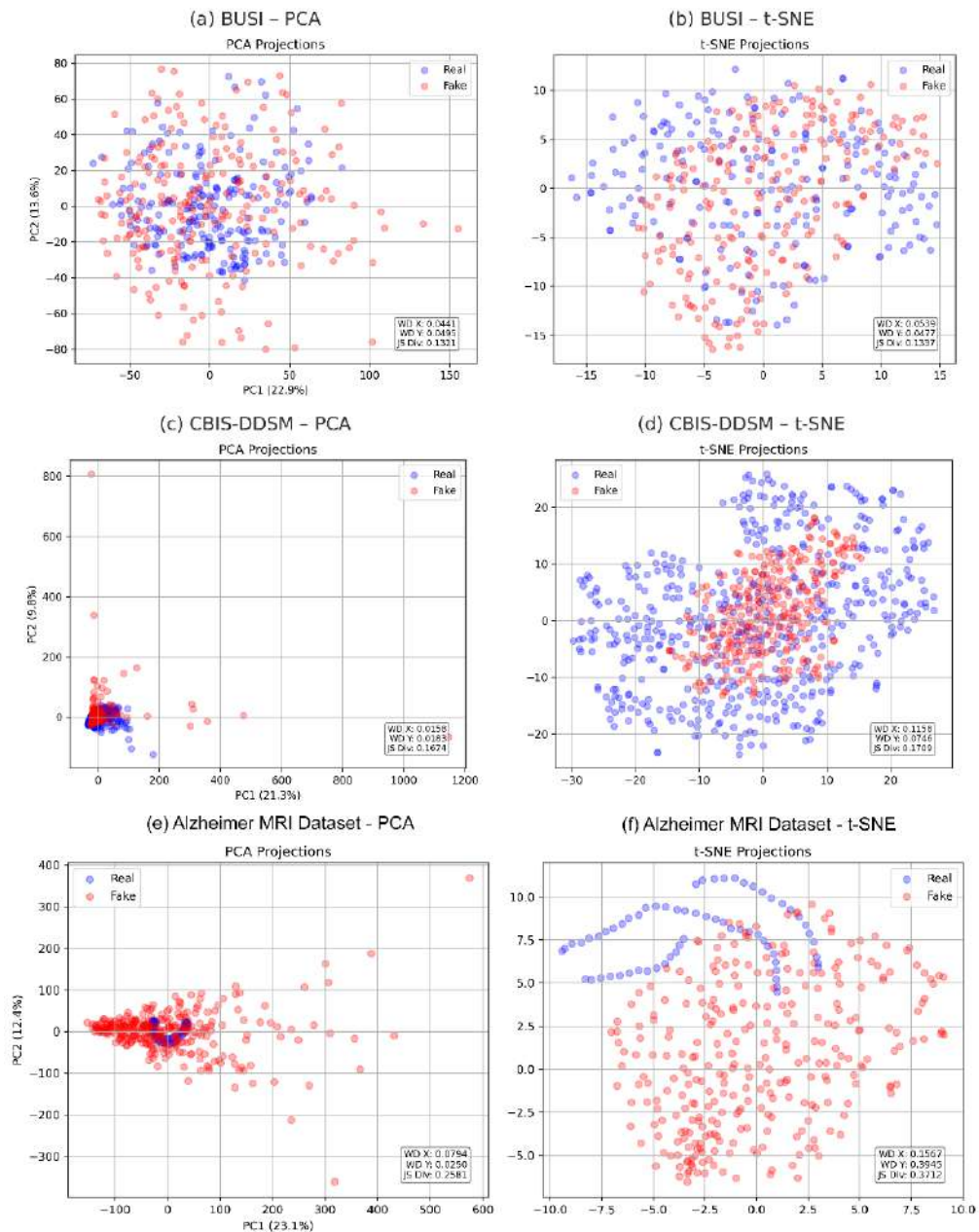


Рисунок 18 – PCA- и t-SNE-проекции реальных (синие точки) и синтетических (красные точки) изображений для датасетов BUSI, CBIS-DDSM и Alzheimer MRI, полученные на 500-й эпохе при оптимальных значениях коэффициента λ

Примечание - По каждому графику приведены значения Jensen-Shannon Divergence (JSD) и Wasserstein Distance (WD)

В качестве признаков представления изображений использовались пиксельные данные после приведения изображений к оттенкам серого (grayscale) и масштабирования до разрешения 224×224 с последующим разворачиванием в вектор. Проекция PCA и t-SNE строились на объединённой выборке, включающей реальные и синтетические изображения, что обеспечивало их

отображение в общем двумерном пространстве и позволяло напрямую сравнивать взаимное расположение распределений. Для метода t-SNE использовалось фиксированное значение random seed, что обеспечивало воспроизводимость результатов визуализации.

Анализ PCA и t-SNE выполнен для всех трёх датасетов с использованием синтетических изображений, сгенерированных моделью DGAN-WP-TL на 500-й эпохе при оптимальных значениях коэффициента градиентного штрафа λ . Для количественного сравнения распределений дополнительно вычислялись Jensen–Shannon Divergence (JSD) и Wasserstein Distance (WD), рассчитанные в пространстве пониженной размерности и используемые для сравнительного анализа согласованности распределений. Результаты представлены в таблице 10 и на рисунке 18.

Следует отметить, что PCA используется для анализа глобальной структуры распределений и направлений наибольшей дисперсии, тогда как t-SNE ориентирован на сохранение локальных взаимосвязей и выявление кластеров в данных. В данной работе данные методы применяются как инструменты качественного анализа и интерпретации, дополняющие количественные метрики качества генерации.

Таблица 10 – Сходство распределений реальных и синтетических данных по метрикам JSD и WD (500 эпох)

Датасет PCA / t-SNE	WD X ↓ PCA / t-SNE	WD Y ↓ PCA / t-SNE	JSD ↓ PCA / t-SNE
BUSI	0.0441 / 0.0539	0.0495 / 0.0477	0.1321 / 0.1337
CBIS-DDSM	0.0158 / 0.1158	0.0183 / 0.0746	0.1674 / 0.1709
Alzheimer MRI	0.0794 / 0.1567	0.0250 / 0.3945	0.2581 / 0.3712

Согласно таблице 10 распределение ультразвуковых изображений из реальной и синтетической выборки схоже. Рисунок 18а отображает визуальный анализ методов PCA, а рисунок 18b – t-SNE, из которых видно, что классы изображений перекрываются. Анализ PCA и t-SNE говорит о близком сходстве распределений классов реальных и синтетических изображений. Дополнительные метрики JSD и WD на рисунках 18а и 18b, также подтверждают низкое различие между кластерами и согласуются с численными метриками KID, FID для ультразвуковых изображений.

Для маммографического датасета CBIS-DDSM наблюдается различие между кластерами (рисунки 18с, 18d), на что указывают и распределительные метрики. Метрики JSD и WD выше в маммографических, чем в ультразвуковых изображениях, что вызвано вариативностью и сложностью маммографий.

Для MRI-датасета наблюдается наибольшее расхождение распределений. На рисунке 18е, 18f видны:

- значения WD существенно выше, особенно в t-SNE (0.3945 по оси Y).
- JSD достигает 0.3712, что значительно больше, чем для BUSI и CBIS.

В случае Alzheimer MRI более высокие результаты связаны с более сложной структурой мозга в отличие от УЗИ и маммографических снимков.

Однако данные результаты не рассматриваются как недостаток, так как основная цель DGAN-WP-TL на датасете Alzheimer MRI является валидация и проверка междоменного переноса.

Полученные результаты PCA и t-SNE показывают различную степень совпадения распределений для исследуемых датасетов. Наиболее тесное согласование наблюдается для BUSI, умеренное – для CBIS-DDSM, и наибольшее расхождение – для Alzheimer MRI, что отражает различия в сложности и модальности данных. Полученные результаты согласуются с количественными метриками качества генерации, представленными ранее в подразделе 3.1.

После установления оптимальных режимов обучения и подтверждения качества синтетических изображений, следующим этапом является анализ влияния источника предобученных весов критика на характеристики генерации, что рассматривается в подразделе 3.2.

3.1.5 Сравнение с baseline-архитектурой StyleGAN2

Для оценки эффективности разработанного метода DGAN-WP-TL проведено сравнение с StyleGAN2 как с одной из современных baseline-архитектур [108].

В отличие от DGAN-WP-TL, в StyleGAN2 используется количественная мера обучения в *king* (тысячи изображений). Для соответствия обучения DGAN-WP-TL и StyleGAN2, количество *king* эквивалентно 500 эпохам. Так, в датасете BUSI *king* = 105 при выборке 210 злокачественных изображений, в CBIS-DDSM *king* = 255 при выборке 509 изображений; для Alzheimer MRI *king* = 32 (64 изображения). Обучение StyleGAN2 проводилось с базовыми настройками и без применения аугментации данных (no-aug).

В рамках данного сравнения проверяется гипотеза о том, что модель DGAN-WP-TL обеспечивает более устойчивую и вариативную генерацию по сравнению с StyleGAN2 в условиях малых и экстремально малых медицинских выборок за счёт использования перцептивного критика и схемы WGAN-GP, стабилизирующей процесс обучения.

Датасет BUSI. Датасет BUSI содержит всего 210 изображений злокачественных опухолей и характеризуется высокой визуальной однородностью. Согласно полученным результатам генерации в таблице 11 StyleGAN2 демонстрирует более высокие распределительные метрики в BUSI датасете, а именно FID = 330.94 и KID = 0.3797 по сравнению с DGAN-WP-TL, где FID = 179.42 и KID = 0.1448. В перцептивном восприятии, StyleGAN2 проигрывает DGAN-WP-TL, о чем свидетельствует LPIPS (r-f) = 0.6016, в то время как LPIPS (r-f) в DGAN-WP-TL равен 0.5113.

Таблица 11 – Количественное сравнение DGAN-WP-TL и StyleGAN2 на датасетах BUSI, CBIS-DDSM и Alzheimer MRI

Датасет	Модель	KID ↓	FID ↓	LPIPS (r-f) ↓	LPIPS (f-f) ↑	MS-SSIM (r-f) ↓	MS-SSIM (f-f) ↓
<i>BUSI</i>	DGAN-WP-TL	0.1448	179.4158	0.5113	0.4862	0.1822	0.1760
	StyleGAN2	0.3797	330.9356	0.6016	0.0715	0.2577	0.9532
<i>CBIS-DDSM</i>	DGAN-WP-TL	0.1795	182.3454	0.2875	0.2986	0.5182	0.4877
	StyleGAN2	0.1201	142.7946	0.2753	0.2489	0.5890	0.6199
<i>Alzheimer MRI</i>	DGAN-WP-TL	0.3567	289.7888	0.2746	0.2428	0.5240	0.5240
	StyleGAN2	0.7097	458.7359	0.8275	0.2464	0.2932	0.8967

Согласно структурной метрике MS-SSIM, StyleGAN2 показал структурное сходство между синтетическими изображениями, о чем говорит MS-SSIM (f-f) = 0.9532 и указывает на одинаковые сгенерированные изображения (коллапс мод), а MS-SSIM (f-f) = 0.1760 в DGAN-WP-TL – отсутствие коллапса мод и вариативность синтетических изображений по яркости, контрасту и текстуре. Таким образом, StyleGAN2 демонстрирует худшее качество генерации и коллапс мод, чем DGAN-WP-TL в BUSI датасете при малом объеме данных.

Датасет CBIS-DDSM. Согласно описанию датасета CBIS-DDSM в подразделе в 2.4 в маммографическом наборе содержатся 509 изображений злокачественных опухолей. Таблица 11 демонстрирует, что StyleGAN2 достиг FID = 142.79 и KID = 0.1201, тогда как у DGAN-WP-TL FID составил 182.35, а KID = 0.1795. Однако на перцептивном и структурном уровне, DGAN-WP-TL генерирует более разнообразные изображения, о чем говорят метрики LPIPS (f-f) = 0.2986 и MS-SSIM (r-f) = 0.5182. Особенно, метрика MS-SSIM (f-f) = 0.4877 у DGAN-WP-TL значительна ниже, чем у StyleGAN2 со значением 0.6199 и означает частичный коллапс мод.

Датасет Alzheimer MRI. Датасет Alzheimer MRI относится к набору с наиболее экстремальным количеством изображений (64 изображений миноритарного класса). Практически все метрики генерации, кроме MS-SSIM (r-f), ухудшились при обучении StyleGAN2, что подтверждает неспособность StyleGAN2 обучаться на экстремально малых выборках. При этом DGAN-WP-TL показывает более устойчивое обучение, разнообразные и перцептивно схожие изображения согласно таблице 11.

На рисунке 19 представлена динамика изменения метрик KID и FID в процессе обучения DGAN-WP-TL и StyleGAN2 для всех трёх датасетов.

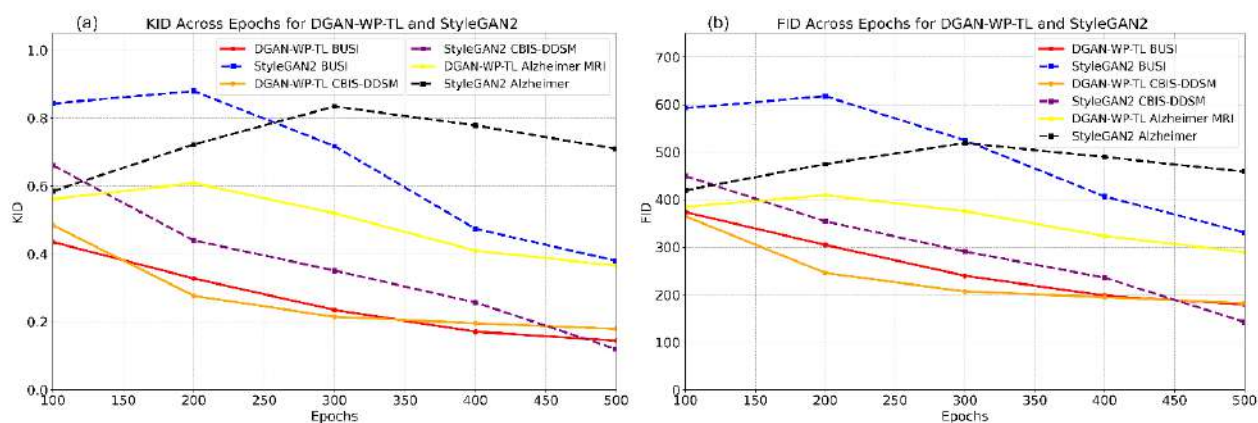


Рисунок 19 – Динамика изменения метрик KID и FID в процессе обучения моделей DGAN-WP-TL и StyleGAN2 на датасетах BUSI, CBIS-DDSM и Alzheimer MRI

Таким образом, сравнительный анализ с baseline-архитектурой StyleGAN2 показывает, что эффективность генеративных моделей существенно зависит от объёма и сложности данных. StyleGAN2 демонстрирует более низкие значения KID и FID на датасете CBIS-DDSM, что указывает на высокое качество аппроксимации распределения при наличии достаточного объёма данных. В то же время, DGAN-WP-TL имеет преимущества в генерации BUSI и Alzheimer MRI изображений в малых медицинских выборках, более того обеспечивая устойчивое обучение, высокое разнообразие. Кроме того, наблюдается коллапс мод у StyleGAN2 во всех трех датасетах. Полученные результаты подтверждают гипотезу, что необходимо оценивать генерацию в комплексе распределительных, перцептуальных и структурных метрик, особенно в задачах медицинской диагностики.

3.2 Влияние предобученных весов ImageNet и RadImageNet на генерацию

Эксперименты данного подраздела направлены на оценку влияния источника предобученных весов ImageNet/RadImageNet в экстракторе критика метода DGAN-WP-TL на качество генерации в маммографических изображениях. Дальнейшие результаты позволят ответить на вопрос: дают ли доменно-специфичные веса RadImageNet преимущество перед повседневными весами ImageNet, а также провести комплексную оценку и сравнение по распределительным, перцептивным и структурным метрикам при фиксированных параметрах обучения. Длительность обучения составляла 500 эпох для обоих весов. На этом этапе обучения с ImageNet показало устойчивое обучение и наилучшие метрики генерации, тогда как обучение с RadImageNet на 500-й эпохе не достигло сопоставимых результатов с генерацией на базе ImageNet весов. В связи с этим, обучение с RadImageNet было продолжено до 1000 эпох для оценки предельного потенциала и анализа поведения при длительной оптимизации [109].

Как было отмечено в подразделе 2.1, в экстракторе признаков применялись несколько CNN-архитектур: с весами ImageNet - VGG19 и DenseNet121; с весами RadImageNet - DenseNet121, InceptionV3 и ResNet50. Модель DenseNet121 позволяет сравнить напрямую генерацию, инициализированную с весами ImageNet и RadImageNet, в отличие от других архитектур.

Результаты генерации оценивались согласно ранее определенным метрикам (FID, KID, LPIPS и MS-SSIM). Дополнительно представлена метрика SWD для того, чтобы оценить структуры изображений и их визуальные особенности локально. Во всех случаях использовался единый протокол обучения: одинаковые параметры оптимизатора и общий диапазон коэффициента градиентного штрафа $\lambda \in \{0.5; 3.0; 5.0; 7.0; 10.0\}$.

Для оценки качества генерации использовались разные метрики, каждая из которых отвечала за свою сторону анализа. KID и FID применялись для сравнения распределений данных. Метрики LPIPS, MS-SSIM и SWD использовались для того, чтобы посмотреть на структуру изображений и их визуальные особенности.

3.2.1 Количественный анализ качества генерации: распределительные, перцептивные и структурные метрики

Как было заявлено ранее, сначала обучение генеративной модели составляло 500 эпох, результаты которого представлены в таблице 12. В данном режиме обучения модель VGG19 на базе ImageNet демонстрирует лучшее распределительное сходство между синтетическими и реальными изображениями со значениями KID=0.17 и FID=182.34 при $\lambda=0.5$ [67, p. 2556687]. Тогда как модели на базе RadImageNet не достигают сопоставимых результатов с ImageNet. Для трех моделей с RadImageNet весами значения KID остаются в пределах 0.21-0.23, а FID – 198-216, что уступает VGG19 и говорит о худшем сходстве между реальными и синтетическими изображениями.

Таблица 12 – Сравнение метрик KID и FID архитектур критика и значений коэффициента λ на датасете CBIS-DDSM (500 эпох)

Архитектура (веса)	$\lambda = 0.5$ KID↓ / FID↓	$\lambda = 3.0$ KID↓ / FID↓	$\lambda = 5.0$ KID↓ / FID↓	$\lambda = 7.0$ KID↓ / FID↓	$\lambda = 10.0$ KID↓ / FID↓
VGG19 (ImageNet)	0.17/182.34	0.19/196.27	0.23/215.26	0.22/213.87	0.22/212.86
DenseNet121 (ImageNet)	0.21/202.65	0.21/205.16	0.21/205.11	0.21/204.84	0.21/203.62
DenseNet121 (RadImageNet)	0.21/203.38	0.20/198.57	0.22/210.17	0.22/214.77	0.21/203.70
InceptionV3 (RadImageNet)	0.21/206.40	0.21/203.75	0.22/212.27	0.22/210.30	0.23/218.64
ResNet50 (RadImageNet)	0.21/216.27	0.21/203.94	0.21/205.26	0.22/211.30	0.21/203.94

Анализ полученных результатов таблицы 12 показал, что на 500-й эпохе наблюдается плато в генерации и модели с весами RadImageNet могут показать

улучшение при дальнейшем обучении. Для того, чтобы проверить данное предположение обучение продолжено до 1000 эпох. В результате дальнейшего обучения получены метрики, отображенные в таблице 13 [109, p. 254-268].

Согласно таблице 13 обучение моделей VGG19 и DenseNet121 с весами ImageNet дает незначительное улучшение FID при 1000 эпохах, тогда как FID у VGG19 не меняется, достигнув плато на 500-й эпохе. DenseNet121 (ImageNet) на 1000 эпохах улучшила результаты генерации чем на 500-й эпохе, однако уступая VGG19, несмотря на увеличение эпох в два раза.

Обучение с моделями на базе RadImageNet весами улучшается в течение 1000 эпох. Среди трех моделей с RadImageNet весами, наилучшие KID = 0.18 и FID = 179.35 демонстрирует модель DenseNet121 (RadImageNet) при $\lambda = 3.0$ по сравнению с KID = 0.20 и FID = 198.57 на 500 эпохах. Архитектуры InceptionV3 и ResNet50 также улучшили распределительные метрики на 1000 эпохах чем на 500, но выраженных преимуществ по сравнению с VGG19 (ImageNet) не достигают. Однако интеграция доменно-специфичных весов указывает на потенциал и важность выбора источника предобученных весов и служит основой для дальнейшего анализа перцептивных и структурных метрик.

Таблица 13 – Сравнение метрик KID и FID архитектур критика и значений коэффициента λ на датасете CBIS-DDSM (1000 эпох)

Архитектура (веса)	$\lambda = 0.5$ KID↓ / FID↓	$\lambda = 3.0$ KID↓ / FID↓	$\lambda = 5.0$ KID↓ / FID↓	$\lambda = 7.0$ KID↓ / FID↓	$\lambda = 10.0$ KID↓ / FID↓
VGG19 (ImageNet)	0.17/178.59	0.18/184.45	0.19/188.87	0.17/180.75	0.18/187.14
DenseNet121 (ImageNet)	0.18/181.60	0.19/188.71	0.18/195.31	0.18/187.44	0.19/189.44
DenseNet121 (RadImageNet)	0.18/179.99	0.18/179.35	0.19/191.15	0.19/191.55	0.18/184.67
InceptionV3 (RadImageNet)	0.17/180.01	0.18/186.16	0.19/189.38	0.19/190.23	0.19/192.36
ResNet50 (RadImageNet)	0.18/182.11	0.18/184.29	0.18/183.06	0.18/185.07	0.18/183.46

Дополнительно генерация была рассмотрена с точки зрения восприятия человеком и разнообразия.

Таблица 14 – Перцептивные и структурные метрики для лучших конфигураций DGAN-WP-TL на датасете CBIS-DDSM (1000 эпох)

Модель DGAN-WP-TL	SWD	LPIPS (r-f) ↓	LPIPS (f-f) ↑	MS-SSIM (r-f) ↓	MS-SSIM (f-f) ↓
VGG19 (ImageNet), ($\lambda = 0.5$)	2.10	0.27	0.28	0.57	0.53
DenseNet121 (RadImageNet), ($\lambda = 3.0$)	1.79	0.28	0.27	0.55	0.55
InceptionV3 (RadImageNet), ($\lambda = 0.5$)	2.14	0.27	0.26	0.58	0.55

В таблице 14 представлены метрики LPIPS, MS-SSIM и SWD [110] для 1000 эпох. Данные метрики изменяются незначительно для обоих типов весов,

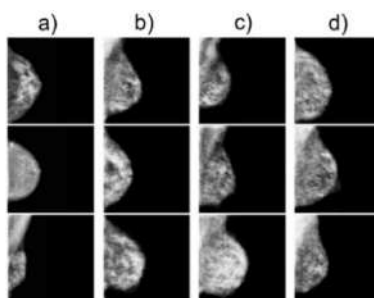
однако DenseNet121 (RadImageNet) обеспечивает минимальный $SWD = 1.79$, что указывает на близкое сходство синтетического распределения с реальным и реалистичность.

Таким образом, перцептивные и структурные метрики показывают, что доменно-специфическое предобучение RadImageNet улучшает текстурную согласованность синтетических изображений с реальными (минимальный SWD), при сопоставимом уровне разнообразия генерации ($LPIPS(f-f)$, $MS-SSIM(f-f)$). Полученные результаты дополняют выводы по распределительным метрикам и подчёркивают значимость выбора источника предобучения критика для повышения качества синтетических данных в задачах автоматизированного обнаружения РМЖ.

3.2.2 Качественные результаты и анализ признакового пространства

Кроме количественной оценки генерации, необходимо проводить количественный анализ синтетических изображений, а именно их сопоставление с реальными. Для анализа взаимного расположения реальных и синтетических данных в проекции использован метод PCA.

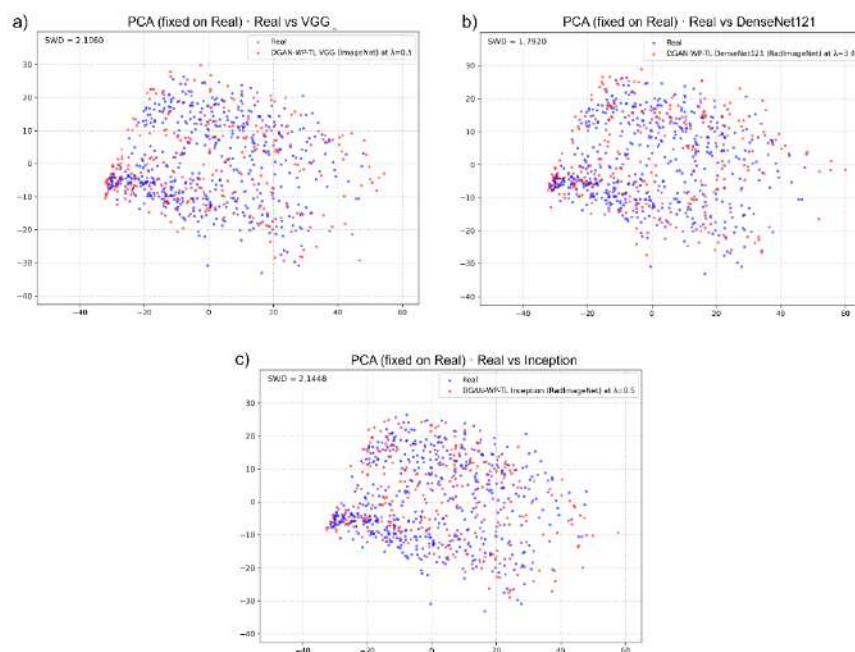
Качественное сравнение реальных и синтетических данных представлено на рисунке 20 для самых лучших конфигураций предобученных моделей в критике для 1000 эпох. При визуальном просмотре все три конфигурации из таблицы 14 воспроизводят общую структуру молочной железы. В то же время, изображения более сглажены, мелкие детали менее отчетливы, что характерно для генеративных моделей. Для генеративных моделей представляет сложность синтезирования четких контуров.



a – реальные изображения; b – DGAN-WP-TL (VGG19, ImageNet, $\lambda = 0.5$); c – DGAN-WP-TL (DenseNet121, RadImageNet, $\lambda = 3.0$); d – DGAN-WP-TL (InceptionV3, RadImageNet, $\lambda = 0.5$)

Рисунок 20 – Примеры реальных и синтетических маммограмм злокачественного класса на датасете CBIS-DDSM

Особо можно отметить DenseNet121 (RadImageNet) с более четкими контурами и деталями. Это связано с интеграцией доменно-специфичных весов в архитектуру критика. Модели InceptionV3 (RadImageNet) и VGG19 (ImageNet) генерируют более сглаженные с потерей детализации.



a – VGG19, ImageNet, $\lambda = 0.5$; b – DenseNet121, RadImageNet, $\lambda = 3.0$; c – InceptionV3, RadImageNet, $\lambda = 0.5$

Рисунок 21 – PCA-проекции злокачественных изображений из датасета CBIS-DDSM и соответствующие значения SWD для различных конфигураций DGAN-WP-TL

Примечание – Синие точки обозначают реальные изображения, красные – синтетические

В соответствии с рисунком 21, для визуального определения соответствия распределений реальных и синтетических данных выполнен анализ PCA. Анализ проекций лучших конфигураций из таблицы 14 показывает, что реальные и сгенерированные в целом перекрываются, т.е. модели попадают в нужную область признакового пространства. Наиболее плотное перекрытие кластеров наблюдается у DenseNet121 (RadImageNet), что соответствует минимальному значению SWD среди рассмотренных моделей. В этом случае синтетические изображения оказываются ближе к реальным и в признаковом пространстве и наблюдается более компактное распределение точек по сравнению с другими вариантами. Для VGG19 (ImageNet) и InceptionV3 (RadImageNet) ситуация немного иная.

Если объединить визуальный анализ PCA и значения SWD из таблицы 14, то результаты показывают стабильное и согласованное совпадение текстур у конфигурации DenseNet121 (RadImageNet) при $\lambda = 3.0$ в задачах распознавания опухолевых образований и микрокальцинатов на маммограммах.

В последующих подразделах главы 3 основной акцент сделан на проверке того, как синтетическое расширение обучающей выборки влияет на показатели downstream-классификации – Recall, F1-score и AUC – по сравнению с режимом обучения только на реальных данных. Также отдельно рассматривается, как

различается эффект от использования синтетики, полученной с применением перцептивных бэкбонов, предобученных на ImageNet и RadImageNet.

3.3 Абляционное исследование архитектуры DGAN-WP-TL

В подразделе представлено абляционное исследование, направленное на оценку вклада ключевых компонентов DGAN-WP-TL: (i) конфигурации критика (Full vs VGG19-Only vs Conv-Only), (ii) функции потерь (WGAN-GP vs GAN с BCE (Binary Cross Entropy)), (iii) режимов обучения на малых выборках (размер батча и объём данных), а также (iv) выбора backbone предобученной модели критика с учётом компромисса «распределительные метрики - вычислительные затраты». Все эксперименты выполнены при едином протоколе (500 эпох) и оптимальных λ из раздела 3.1 для каждого датасета [67]. Оптимальный λ выбран экспериментально из подраздела 3.1 при 500 эпохах.

3.3.1 Влияние архитектуры критика и функции потерь

Для оценки вклада отдельных компонентов гибридного критика DGAN-WP-TL было проведено абляционное исследование, в котором сравнивались три конфигурации:

Full (VGG19+Conv) – двухветвевая архитектура критика, включающая ветвь предобученного признакового экстрактора VGG19 и сверточную ветвь, обучаемую с нуля.

VGG19-Only – вариант, в котором использовалась только ветвь VGG19 без сверточной части.

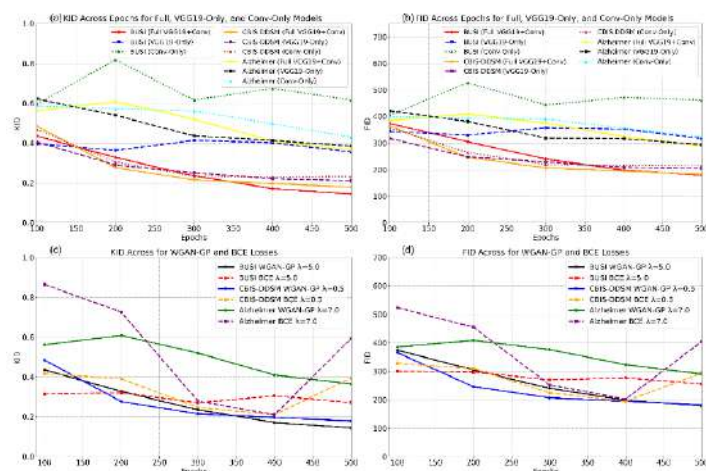
Conv-Only – критик, состоящий исключительно из сверточных слоёв, инициализируемых случайным образом.

Результаты генерации оценивались комплексно метриками генерации (KID, FID, LPIPS, MS-SSIM) на трёх выбранных ранее медицинских датасетах и отображены в таблице 15 согласно оптимальному значению коэффициента градиентного штрафа для каждого набора данных при 500 эпохах.

Таблица 15 – Результаты абляционного анализа архитектуры критика на датасетах BUSI, CBIS-DDSM и Alzheimer MRI (500 эпох)

Model	KID ↓	FID ↓	LPIPS (r-f) ↓	LPIPS (f-f) ↑	MS-SSIM (r-f) ↓	MS-SSIM (f-f) ↓
<i>BUSI ($\lambda = 5.0$)</i>						
Full	0.1448	179.4158	0.5113	0.4862	0.1822	0.1760
VGG19-Only	0.3559	316.8558	0.5508	0.4189	0.1510	0.1238
Conv-Only	0.6153	461.0453	0.7039	0.4885	0.0588	0.1698
<i>CBIS-DDSM ($\lambda = 0.5$)</i>						
Full	0.1795	182.3454	0.2875	0.2986	0.5182	0.4877
VGG19-Only	0.2105	205.3161	0.2602	0.2311	0.5791	0.5624
Conv-Only	0.2318	215.3497	0.2917	0.2764	0.5683	0.5410
<i>Alzheimer MRI ($\lambda = 7.0$)</i>						
Full	0.3567	289.7888	0.2746	0.2428	0.5240	0.5240
VGG19-Only	0.3857	294.3719	0.3816	0.3003	0.2677	0.3186
Conv-Only	0.4294	323.2034	0.3053	0.2478	0.5222	0.5469

Таблица 15 показывает, что Full конфигурация обеспечивает лучшие распределительные показатели (минимальные KID/FID) на всех трёх датасетах и стабильную сходимость для FID/KID при меньших колебаниях метрик (рисунок 22а, 22б). Урезанные конфигурации только с одной ветвью в критике (VGG19-Only, Conv-Only) сгенерировали сильно отличающиеся изображения, которые не схожи с реальными по распределительным метрикам (таблица 15) и качественному анализу на рисунке 23 для всех трех датасетов. Даже несмотря на некоторые улучшения в перцептивных и структурных метриках синтетические изображения в VGG19-Only и Conv-Only сильно искажены по восприятию.



a, b – KID и FID по эпохам для Full VGG19+Conv, VGG19-Only, Conv-Only для каждого датасета; c, d – KID и FID по эпохам для WGAN-GP и BCE-потерь при оптимальных значениях λ для каждого датасета

Рисунок 22 – Динамика метрик KID и FID в абляционном исследовании архитектуры критика

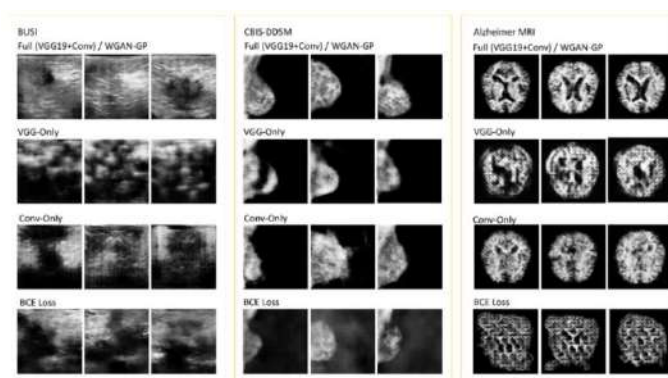


Рисунок 23 – Качественные результаты абляционного анализа на датасетах BUSI, CBIS-DDSM и Alzheimer MRI

Примечание – В каждом блоке показаны изображения при: Full, VGG19- Only, Conv-Only, а также модель с BCE Loss

Одним из важных абляционных экспериментов является анализ вклада функции потерь в генерацию, который и обосновывает выбор модификации

WGAN-GP. Были протестированы Full-конфигурация двухветвевое критика с WGAN-GP и с классической функцией потерь BCE в GAN. Результаты абляционного исследования функции потерь сведены в таблице 16, где видна сильная деградация в распределительных метриках всех трех датасетов в BCE-конфигурации.

Таблица 16 – Сравнение WGAN-GP и стандартного GAN (BCE-потеря) на датасетах BUSI, CBIS-DDSM и Alzheimer MRI (500 эпох)

Функция потерь	KID ↓	FID ↓	LPIPS (r-f) ↓	LPIPS (f-f) ↑	MS-SSIM (r-f) ↓	MS-SSIM (f-f) ↓
<i>BUSI ($\lambda = 5.0$)</i>						
WGAN-GP	0.1448	179.4158	0.5113	0.4862	0.1822	0.1760
BCE	0.2720	255.9382	0.5135	0.4042	0.1538	0.2207
<i>CBIS-DDSM ($\lambda = 0.5$)</i>						
WGAN-GP	0.1795	182.3454	0.2875	0.2986	0.5182	0.4877
BCE	0.3944	293.5161	0.5137	0.3115	0.5084	0.6189
<i>Alzheimer MRI ($\lambda = 7.0$)</i>						
WGAN-GP	0.3567	289.7888	0.2746	0.2428	0.5240	0.5240
BCE	0.5931	404.3077	0.4396	0.2931	0.2362	0.2773

3.3.2 Анализ устойчивости и вычислительной эффективности DGAN-WP-TL при малом объеме данных

Дополнительные абляционные исследования были проведены по выбору размера батча, чувствительность KID/FID к объёму выборки, а также компромисс “качество-скорость” при выборе бэкбона критика [67, p. 2556687].

Эксперименты по выбору размера батча показали, что наилучшим параметром батча является 2 из выбранного диапазона batch size $\in \{2,4,8\}$. Результаты представлены в таблице 17.

Таблица 17 – Влияние размера батча на качество генерации (CBIS-DDSM, 500 эпох, $\lambda = 0.5$)

Batch Size	KID ↓	FID ↓	LPIPS (r-f) ↓	LPIPS (f-f) ↑	MS-SSIM (r-f) ↓	MS-SSIM (f-f) ↓
2	0.1795	182.3454	0.2875	0.2986	0.5182	0.4877
4	0.2326	217.8881	0.3135	0.2594	0.5644	0.5440
8	0.4137	320.2039	0.4932	0.2342	0.5447	0.5799

Эксперимент был проведен на маммографическом датасете при $\lambda = 5.0$, 500-й эпохе. Данные значения батча 2, 4, 8 выбраны как типичные практические режимы обучения, ограниченные доступной GPU-памятью при работе с изображениями. В результате таблица 17 показывает распределение метрик в зависимости от размера батча, где минимальный размер демонстрирует наилучшие результаты в комплексе метрик. Это дополнительно указывает, что медицинские изображения нуждаются в подборе параметров для устойчивого обучения.

Следующим абляционным экспериментом является оценка устойчивости метрик KID и FID при различных размерах выборки на датасете BUSI. Использовались 100, 200 и 300 изображений ($\lambda = 5.0$, 500-я эпоха). Результаты показаны в таблице 18.

Таблица 18 – Поведение KID и FID при различных размерах выборки на датасете BUSI ($\lambda = 5.0$, 500 эпох)

Размер выборки	KID ↓	FID ↓
100	0.1392	212.7855
200	0.1412	193.6433
300	0.1448	179.4158

Анализ поведения метрик дополняет результаты экспериментов с размером батча. Согласно таблице 18, статистически KID меняется незначительно ($\Delta \approx 0.006$), что говорит об устойчивости к малым выборкам. В свою очередь, FID значительно колеблется в зависимости от размера выборки и изначально должен применяться в крупных датасетах. В данной диссертационной работе KID является ключевой метрикой в условиях малых медицинских выборок, тогда как FID носит более информативное значение, необходимое для сравнительной характеристики с существующими методами.

Завершающим абляционным экспериментом является выбор основной архитектуры предобученной модели в критике DGAN-WP-TL с ImageNet весами. Для сравнения были выбраны 5 архитектур, результаты которых показаны в таблице 19. Эксперимент проводился на CBIS-DDSM датасете при $\lambda = 0.5$ и 500 эпох. Согласно абляционному эксперименту, VGG19 показало наилучшие распределительные метрики, уступая в перцептуальных и структурных метриках ResNet50 незначительно. В свою очередь, ResNet50 не была выбрана основной моделью из-за более низкого сходства распределений между сгенерированными и реальными изображениями. Кроме того, длительность обучения превысила почти в 2 раза по сравнению с VGG19, что также повлияло на выбор архитектуры в критике, сохраняя баланс «качество-вычислительные затраты».

Таблица 19 – Сравнение бэкбонов критика на датасете CBIS-DDSM (500 эпох, $\lambda=0.5$)

Backbone	KID ↓	FID ↓	LPIPS (r-f) ↓	LPIPS (f-f) ↑	MS-SSIM (r-f) ↓	MS-SSIM (f-f) ↓	Training time, s/epoch
VGG19 (ours)	0.1795	182.3454	0.2875	0.2986	0.5182	0.4877	55.8
ResNet50	0.2119	221.8758	0.5375	0.4421	0.4899	0.5311	93.2
MobileNetV2	0.2295	216.8151	0.2894	0.2786	0.5666	0.5352	51.0
EfficientNet-B0	0.2247	210.8969	0.2902	0.2730	0.5588	0.5400	62.4
VGG16	0.2198	208.7327	0.2839	0.2763	0.5744	0.5518	49.5

Таким образом, VGG19 обеспечивает наилучшие значения KID/FID среди рассмотренных бэкбонов при приемлемом времени обучения. Более лёгкие модели (VGG16, MobileNetV2) обучаются быстрее, но уступают по

распределительным метрикам, тогда как ResNet50 существенно увеличивает время обучения без выигрыша в качестве генерации.

3.4 Классификация рака молочной железы на расширенной выборке

Раздел 3.4 оценивает практическую полезность синтетического расширения данных: как добавление malignant-изображений, сгенерированных DGAN-WP-TL, влияет на downstream-классификацию РМЖ. Качество классификации анализируется по ключевым метрикам Recall, F1-score и AUC, а также по Precision, Specificity и Accuracy.

3.4.1 Оценка влияния синтетических данных на классификацию

Оценка классификации РМЖ была проведена при фиксированном классификаторе ResNet18, являющаяся базовой архитектурой в медицинской диагностике. ResNet18 инициализирована весами ImageNet и обучалась в течение 20 эпох, результаты усреднялись по 5 независимым запускам. Синтетические malignant-изображения добавлялись в объёме 30, 50 и 100% от числа реальных malignant-примеров; генерация выполнялась при $\lambda \in \{0.5; 3.0; 5.0; 7.0; 10.0\}$ [67, p. 2556687].

Результаты на датасете BUSI. По сравнению с режимом Real Only в таблице 20 наблюдается улучшение метрики F1-score практически во всех режимах синтетики, а именно 30, 50 и 100%, при этом Recall растёт в основном с синтетикой 50% во всех значениях λ , кроме 10.0. Метрика Accuracy также растёт при генеративной аугментации в различных режимах синтетики, при этом Precision растёт только при 30 и 100% синтетики при малых $\lambda=0.5, 3.0$ и 5.0, при дальнейшем увеличении λ с 7.0 до 10.0 рост Precision наблюдается во всех режимах синтетики. В целом генеративная аугментация положительно влияет на обнаружение РМЖ, однако эффект зависит от λ и доли синтетического пополнения.

Таблица 20 – Результаты downstream-классификации BUSI (ResNet18)

Конфигурация	F1-score	AUC	Accuracy	Precision	Recall	Specificity
1	2	3	4	5	6	7
<i>Real Only</i>	0.743	0.920	0.833	0.825	0.722	0.885
$\lambda = 0.5$						
Syn 30%	0.757	0.910	0.866	0.893	0.663	0.961
Syn 50%	0.765	0.913	0.848	0.798	0.761	0.890
Syn 100%	0.720	0.882	0.827	0.755	0.702	0.885
$\lambda = 3.0$						
Syn 30%	0.742	0.890	0.850	0.834	0.683	0.929
Syn 50%	0.754	0.906	0.847	0.787	0.741	0.897
Syn 100%	0.778	0.908	0.877	0.917	0.683	0.968
$\lambda = 5.0$						
Syn 30%	0.778	0.915	0.875	0.902	0.688	0.963
Syn 50%	0.776	0.918	0.864	0.824	0.737	0.924
Syn 100%	0.776	0.907	0.867	0.900	0.668	0.961

Продолжение таблицы 20

1	2	3	4	5	6	7
$\lambda = 7.0$						
Syn 30%	0.769	0.922	0.867	0.872	0.693	0.949
Syn 50%	0.803	0.912	0.883	0.877	0.751	0.945
Syn 100%	0.771	0.913	0.875	0.929	0.668	0.972
$\lambda = 10.0$						
Syn 30%	0.776	0.903	0.869	0.862	0.707	0.945
Syn 50%	0.732	0.908	0.850	0.871	0.649	0.945
Syn 100%	0.803	0.934	0.891	0.946	0.702	0.979

CBIS-DDSM датасет. В датасете CBIS-DDSM генеративная аугментация также улучшает классификацию РМЖ. Однако существенные улучшения наблюдаются при добавлении 30% синтетики при $\lambda=5.0$ и 10.0 . Рост клинически значимых метрик F1-score с 0.574 в Real Only до 0.654 и Recall с 0.608 до 0.667 при $\lambda=0.5$ и 30% синтетики (таблица 21). При этом вторичная метрика Ассигасу достигает наилучших значений при $\lambda=10.0$ и синтетике 30%, хотя лучший KID из раздела 3.1 был достигнут при $\lambda=0.5$ в CBIS-DDSM. Это говорит о том, что метрики генерации и классификации не всегда напрямую коррелируют и требуют экспериментальной проверки для получения лучших результатов в зависимости от поставленной задачи.

Таблица 21 – Результаты downstream-классификации CBIS-DDSM (ResNet18)

Конфигурация	F1-score	AUC	Accuracy	Precision	Recall	Specificity
<i>Real Only</i>	0.574	0.647	0.605	0.516	0.647	0.575
$\lambda = 0.5$						
Syn 30%	0.602	0.692	0.669	0.596	0.608	0.712
Syn 50%	0.539	0.688	0.669	0.632	0.471	0.808
Syn 100%	0.420	0.682	0.621	0.567	0.333	0.822
$\lambda = 3.0$						
Syn 30%	0.463	0.676	0.645	0.613	0.373	0.836
Syn 50%	0.506	0.670	0.653	0.611	0.431	0.808
Syn 100%	0.527	0.691	0.653	0.600	0.471	0.781
$\lambda = 5.0$						
Syn 30%	0.654	0.694	0.710	0.642	0.667	0.740
Syn 50%	0.535	0.680	0.621	0.540	0.529	0.685
Syn 100%	0.522	0.683	0.645	0.585	0.471	0.767
$\lambda = 7.0$						
Syn 30%	0.607	0.712	0.645	0.557	0.667	0.630
Syn 50%	0.500	0.699	0.629	0.561	0.451	0.753
Syn 100%	0.512	0.676	0.661	0.629	0.431	0.822
$\lambda = 10.0$						
Syn 30%	0.639	0.707	0.718	0.674	0.608	0.795
Syn 50%	0.574	0.667	0.677	0.628	0.529	0.781
Syn 100%	0.405	0.645	0.597	0.515	0.333	0.781

Alzheimer MRI датасет. На данных Alzheimer MRI классификация с синтетическим расширением миноритарного класса также улучшилась

практически во всех режимах синтетики (30, 50 и 100%). Однако максимальные значения клинически значимых метрик (F1-score = 0.854 и Recall=0.857) достигаются при 50% синтетики и $\lambda=10.0$. При $\lambda=7.0$ также получены конкурентные результаты при 50% синтетики. При этом минимальный KID фиксируется при $\lambda=7.0$, однако соответствующие downstream-метрики оказываются несколько ниже, чем при $\lambda = 10.0$. Следовательно, согласованность между KID и качеством классификации зависит от режима синтетического пополнения (таблица 22).

Таблица 22 – Результаты downstream-классификации Alzheimer MRI (ResNet18)

Конфигурация	F1-score	AUC	Accuracy	Precision	Recall	Specificity
<i>Real Only</i>	0.763	0.944	0.783	0.856	0.783	0.910
$\lambda = 0.5$						
Syn 30%	0.773	0.967	0.782	0.833	0.782	0.910
Syn 50%	0.780	0.970	0.785	0.847	0.785	0.914
Syn 100%	0.821	0.972	0.827	0.852	0.827	0.927
$\lambda = 3.0$						
Syn 30%	0.842	0.973	0.847	0.865	0.847	0.933
Syn 50%	0.807	0.974	0.820	0.855	0.820	0.918
Syn 100%	0.738	0.968	0.748	0.825	0.748	0.896
$\lambda = 5.0$						
Syn 30%	0.831	0.968	0.833	0.862	0.833	0.930
Syn 50%	0.848	0.973	0.847	0.860	0.847	0.936
Syn 100%	0.821	0.964	0.822	0.850	0.822	0.927
$\lambda = 7.0$						
Syn 30%	0.797	0.966	0.802	0.850	0.802	0.921
Syn 50%	0.813	0.972	0.816	0.854	0.816	0.925
Syn 100%	0.756	0.961	0.767	0.827	0.767	0.906
$\lambda = 10.0$						
Syn 30%	0.726	0.956	0.738	0.818	0.738	0.893
Syn 50%	0.854	0.967	0.857	0.865	0.857	0.939
Syn 100%	0.819	0.968	0.823	0.851	0.823	0.929

Сравнение трёх датасетов показывает модально-зависимый характер эффекта: для BUSI улучшения F1/AUC чаще согласуются с метриками качества генерации, тогда как для CBIS-DDSM и Alzheimer MRI тенденция менее стабильна и зависит от доли синтетики. Это отражается в ROC-кривых (рисунок 24a) и в динамике F1-score при разных λ (рисунок 24b, 24d).

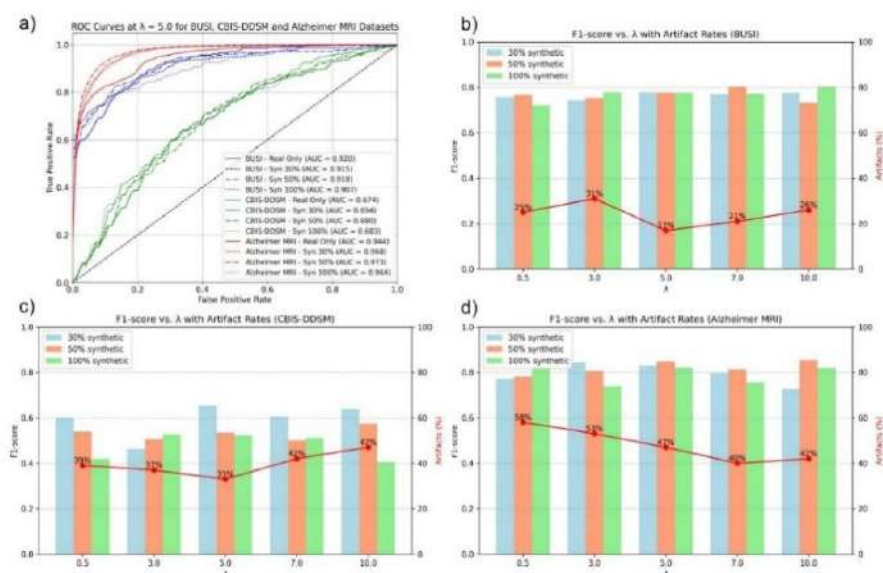


Рисунок 24 – Сравнение ROC-кривых для BUSI, CBIS-DDSM и Alzheimer MRI при $\lambda = 5.0$ (a), а также распределение F1 и уровня артефактов по значениям λ и уровням синтетического пополнения (30, 50, 100%) для BUSI (b), CBIS-DDSM (c) и Alzheimer MRI (d)

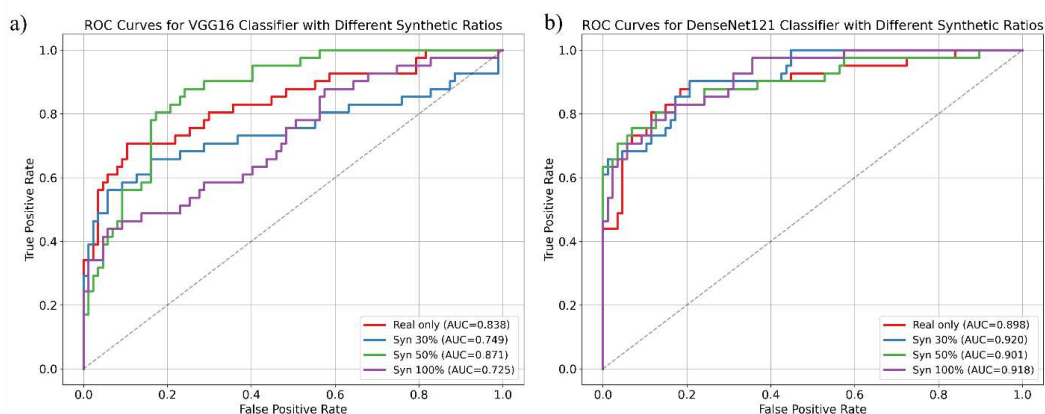
3.4.2 Классификация BUSI с использованием моделей на предобученных весах ImageNet

В данном подразделе анализируются 4 выбранные модели в классификации РМЖ (VGG16, DenseNet121, MobileNetV2 и EfficientNet-V2-S) с весами ImageNet, аугментированные синтетическими изображениями метода DGAN-WP-TL в УЗИ изображениях. Режим синтетики также составляет 30, 50 и 100%, синтетические изображения добавлены из модели с $\lambda=5.0$ с наилучшими метриками генерации с VGG19 предобученной моделью в двухветвевом критике.

VGG16 классификатор. При использовании VGG16 (таблица 23) генеративная аугментация улучшила результаты при 50% синтетики, достигнув F1-score = 0.721, Recall = 0.756 и AUC = 0.871 по сравнению с режимом Real Only (F1-score=0.675, AUC=0.838). При добавлении 30 и 100% результаты значительно ухудшаются. Об этом также свидетельствуют ROC-кривые на рисунке 25а, которая демонстрирует наибольшее смещение вверх и влево по сравнению с другими конфигурациями. При 100% синтетики качество резко ухудшается (F1-score 0.495, Ассигасу 0.617), что визуально подтверждается на ROC-кривой (рисунок 25а), смещающейся ближе к диагонали. В целом динамика VGG16 показывает, что умеренное количество синтетики повышает диагностическую надёжность, тогда как её избыток или недостаток приводят к нарушению клинического баланса между метриками.

Таблица 23 – Результаты классификации BUSI с использованием VGG16 (перцептивный бэкбон VGG19 в критике DGAN-WP-TL с ImageNet весами)

Доля синтетики	F1-score	AUC	Accuracy	Precision	Recall	Specificity
Real Only	0.675	0.838	0.805	0.722	0.634	0.885
Syn 30%	0.622	0.749	0.805	0.767	0.683	0.920
Syn 50%	0.721	0.871	0.812	0.689	0.756	0.839
Syn 100%	0.495	0.725	0.617	0.429	0.585	0.632



а – ROC-кривые для классификатора VGG16; б – ROC-кривые для классификатора DenseNet121

Рисунок 25 – ROC-кривые классификаторов BUSI при различных долях синтетических изображений (Real Only – 0, 30, 50, 100%), сгенерированных DGAN-WP-TL (перцептивный бэкбон VGG19, ImageNet)

DenseNet121 классификатор. При обучении DenseNet121 классификатора (таблица 24) генеративная аугментация улучшает результаты классификации, особенно миноритарного класса при режиме синтетики 50%, достигая максимальной Recall=0.878 и снижая вероятность нахождения ложноотрицательных случаев. Однако Precision падает до 0.610, увеличивая число ложноположительных случаев. В медицинской диагностике минимизация ложноотрицательных решений является более приоритетной задачей по сравнению с минимизацией ложноположительных. Добавление 30% синтетики практически не улучшает метрику Recall, а в 100% синтетики изменяется баланс Precision/Recall. При анализе F1-score наблюдается рост при всех режимах синтетики и необходимо выбирать подходящий режим согласно поставленной задаче. ROC-кривые отображены на рисунке 25b и показывают сохраняющуюся разделимость классов при 50 и 100% синтетики.

Таблица 24 – Результаты классификации BUSI с использованием DenseNet121 (перцептивный бэкбон VGG19 в критике DGAN-WP-TL с ImageNet весами)

Доля синтетики	F1-score	AUC	Accuracy	Precision	Recall	Specificity
Real Only	0.643	0.898	0.859	0.833	0.732	0.931
Syn 30%	0.723	0.920	0.820	0.714	0.732	0.862
Syn 50%	0.720	0.901	0.781	0.610	0.878	0.736
Syn 100%	0.743	0.918	0.867	0.897	0.634	0.966

MobileNetV2 классификатор. MobileNetV2 (таблица 25) сильнее остальных архитектур зависит от объёма синтетики, что может быть связано с ограниченной модельной ёмкостью и повышенной чувствительностью к изменению состава обучающей выборки. Синтетическое расширение в 30% позволило достичь Recall равный 0.780, а F1-score 0.660. Рисунок 26a демонстрирует ограниченный рост разделяемости классов, указывая на нестабильность MobileNetV2.

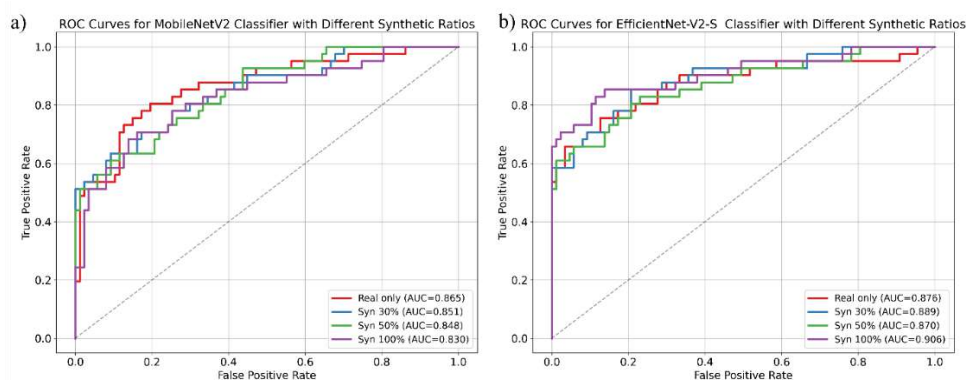
Таблица 25 – Результаты классификации BUSI датасета на классификаторе MobileNetV2 с перцептивной моделью VGG19 в критике модели DGAN-WP-TL с ImageNet весами

Доля синтетики	F-score	AUC	Accuracy	Precision	Recall	Specificity
Real only	0.513	0.865	0.805	0.674	0.756	0.828
Syn 30%	0.660	0.851	0.742	0.571	0.780	0.724
Syn 50%	0.644	0.848	0.750	0.592	0.707	0.770
Syn 100%	0.683	0.830	0.797	0.683	0.683	0.851

EfficientNet-V2-S классификатор: Согласно таблице 26 EfficientNet-V2-S показал рост Recall и других метрик при синтетическом расширении в 30%, однако F1-score стал равно 0.730. При синтетическом расширении в 50% наблюдается снижении Recall до 0.858. Наилучший результат по всем метрикам, в том числе и Recall = 0.707 достигнет при 100% доли синтетики. Однако, следует с осторожностью. Использовать 100%-ное соотношение расширение по отношению к реальным изображениям. Так как 100% синтетических изображений может вызвать рост доли артефактов и искажений относительно реальных изображений. У все трех режимов наблюдается одинаковое поведение в ROC-кривых (рисунок 26b).

Таблица 26 – Результаты классификации BUSI датасета на классификаторе EfficientNet-V2-S с перцептивной моделью VGG19 в критике модели DGAN-WP-TL с ImageNet весами

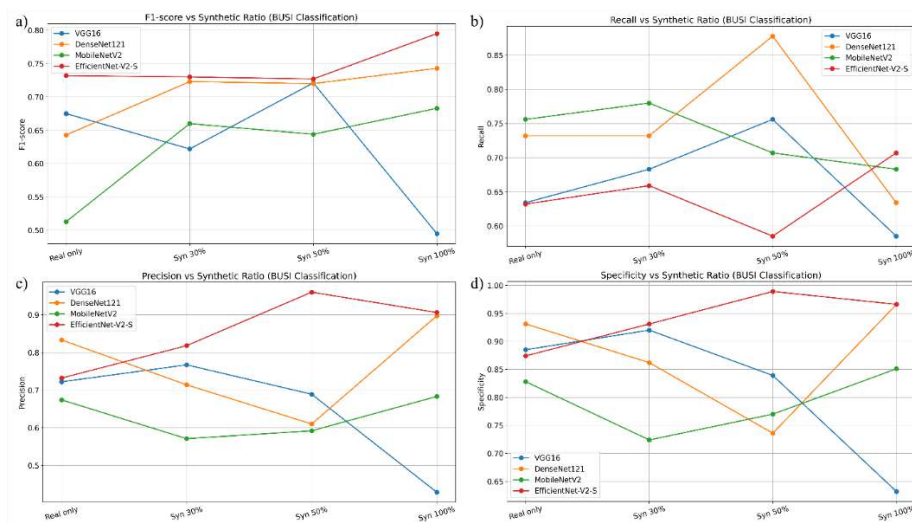
Доля синтетики	F-score	AUC	Accuracy	Precision	Recall	Specificity
Real only	0.732	0.876	0.828	0.732	0.632	0.874
Syn 30%	0.730	0.889	0.844	0.818	0.659	0.931
Syn 50%	0.727	0.870	0.859	0.960	0.858	0.989
Syn 100%	0.795	0.906	0.883	0.906	0.707	0.966



а – ROC-кривые для классификатора MobileNetV2; б – ROC-кривые для классификатора EfficientNet-V2-S

Рисунок 26 – ROC-кривые классификаторов BUSI при различных долях синтетических изображений (0, 30, 50, 100%), сгенерированных DGAN-WP-TL (перцептивный бэкбон VGG19, ImageNet)

Рисунок 27 отображает общую динамику при варьировании доли синтетического расширения. Результаты моделей зависят от самой архитектуры классификаторов, так, например, для VGG19 метрики F1-score и Recall увеличиваются при 50% и падают при полном синтетическом расширении в 100% (рисунки 27а, 27б). DenseNet121 демонстрирует более плавные и согласованные изменения F1-score и Recall при варьировании доли синтетических данных (рисунок 27а, 27б).



а – F1-score; б – Recall; в – Precision; г – Specificity

Рисунок 27 – Зависимость ключевых диагностических метрик классификации BUSI от доли синтетических данных, сгенерированных DGAN-WP-TL (VGG19, ImageNet) для классификаторов VGG16, DenseNet121, MobileNetV2 и EfficientNet-V2-S при синтетике 0, 30, 50 и 100%

MobileNetV2 характеризуется наибольшей чувствительностью к объёму синтетического пополнения, что проявляется в выраженных колебаниях

Precision и Specificity при переходе между режимами (рисунок 27с, 27d). В то же время EfficientNet-V2-S демонстрирует наиболее стабильную динамику всех рассмотренных метрик, включая Precision и Specificity, при увеличении доли синтетических изображений (рисунок 27а, 27b, 27с, 27d)

3.4.3 Классификация CBIS-DDSM на основе синтетических изображений, сгенерированных DGAN-WP-TL (VGG19 и DenseNet121 в качестве перцептивных бэкбонов с ImageNet весами)

В данном подразделе проведены эксперименты классификации с синтетическим расширением миноритарного класса с помощью метода DGAN-WP-TL в маммографическом датасете CBIS-DDSM. Дальнейшие результаты позволяют оценить, в какой степени синтетические malignant-изображения улучшают способность различных архитектур CNN обнаруживать злокачественные микрокальцинаты.

В данном эксперименте было проведено сравнение между двумя архитектурами в критике DGAN-WP-TL, а именно VGG19 и DenseNet121 с весами ImageNet, что позволяет проанализировать влияние архитектуры перцептивного представления на downstream-классификацию. Сводные результаты приведены в таблицах 27, 28, 29, 30.

Классификация микрокальцинатов в CBIS-DDSM демонстрирует типичное поведение для несбалансированных медицинских выборок. В режиме Real Only большинство моделей достигают высокой Specificity при резко сниженной чувствительности (Recall), что приводит к низким значениям F1-score.

При обучении классификатора VGG16 наилучший баланс достигнут при критике с обоими моделями VGG19 и DenseNet121 и 30%-м наполнении синтетикой в обоих случаях (таблица 27). Так, для VGG16 Recall увеличивается с 0.367 до 0.706 (VGG19-бэкбон) и до 0.725 (DenseNet121-бэкбон), при одновременном росте F1-score с 0.607 до 0.643 и 0.638 соответственно (таблица 27).

Таблица 27 – Результаты классификации CBIS-DDSM с использованием VGG16 (перцептивные бэкбоны VGG19 и DenseNet121 в критике DGAN-WP-TL с ImageNet весами)

Доля синтетики	F1-score VGG19/DenseNet121	AUC VGG19/DenseNet121	Accuracy VGG19/DenseNet121	Precision VGG19/DenseNet121	Recall VGG19/DenseNet121	Specificity VGG19/DenseNet121
Real Only	0.607/0.607	0.662/0.662	0.645/0.645	0.557/0.557	0.367/0.367	0.630/0.630
Syn 30%	0.643/0.638	0.662/0.671	0.677/0.661	0.590/0.569	0.706/0.725	0.658/0.616
Syn 50%	0.386/0.421	0.650/0.695	0.589/0.645	0.500/0.640	0.314/0.314	0.781/0.877
Syn 100%	0.467/0.460	0.672/0.674	0.613/0.621	0.538/0.556	0.412/0.392	0.753/0.781

Поведение DenseNet121 как классификатора характеризуется более сложной динамикой (таблица 28). При добавлении 30% синтетики чувствительность остаётся низкой или временно снижается, однако при

увеличении доли синтетических данных до 50% наблюдается резкий рост Recall (до 0.627 и 0.510) и F1-score (до 0.593 и 0.591). При 100% синтетики Recall продолжает расти, но большинство метрик снижается и указывает на ухудшение баланса. Оптимальным режимом для DenseNet121 с обоими моделями в критике является добавление 50% синтетических изображений.

Таблица 28 – Результаты классификации CBIS-DDSM с использованием DenseNet121 (перцептивные бэкбоны VGG19 и DenseNet121 в критике DGAN-WP-TL с ImageNet весами)

Доля синтетики	F1-score VGG19/DenseNet121	AUC VGG19/DenseNet121	Accuracy VGG19/DenseNet121	Precision VGG19/DenseNet121	Recall VGG19/DenseNet121	Specificity VGG19/DenseNet121
Real Only	0.487/0.487	0.733/0.733	0.677/0.677	0.704/0.704	0.373/0.373	0.890/0.890
Syn 30%	0.348/0.394	0.730/0.693	0.637/0.653	0.667/0.700	0.235/0.275	0.918/0.918
Syn 50%	0.593/0.591	0.679/0.728	0.645/0.710	0.561/0.703	0.627/0.510	0.658/0.849
Syn 100%	0.483/0.583	0.678/0.660	0.637/0.597	0.583/0.507	0.412/0.686	0.795/0.534

Более яркое влияние синтетики наблюдается в компактных архитектурах. MobileNetV2 – наиболее показательный пример: Recall увеличивается в 4-5 раз с 0.137 до 0.765 для VGG19 бэкбона и для DenseNet121 бэкбона (0.686, таблица 29), F1-score – с 0.237 до 0.583-0.645. Это отражает переход от модели с крайне низкой чувствительностью к режиму активного выявления злокачественных микрокальцинатов.

Таблица 29 – Результаты классификации CBIS-DDSM с использованием MobileNetV2 (перцептивные бэкбоны VGG19 и DenseNet121 в критике DGAN-WP-TL с ImageNet весами)

Доля синтетики	F1-score VGG19/DenseNet121	AUC VGG19/DenseNet121	Accuracy VGG19/DenseNet121	Precision VGG19/DenseNet121	Recall VGG19/DenseNet121	Specificity VGG19/DenseNet121
Real Only	0.237/0.237	0.701/0.701	0.637/0.637	0.875/0.875	0.137/0.137	0.986/0.986
Syn 30%	0.645/0.583	0.699/0.657	0.653/0.597	0.557/0.507	0.765/0.686	0.575/0.534
Syn 50%	0.522/0.308	0.701/0.655	0.645/0.637	0.585/0.714	0.471/0.196	0.767/0.945
Syn 100%	0.395/0.338	0.647/0.619	0.605/0.621	0.533/0.600	0.314/0.235	0.808/0.890

EfficientNet-V2-S демонстрирует похожую тенденцию (таблица 30). При добавлении синтетики 30 и 100% F1-score повышается до 0.624-0.626, а AUC достигает 0.769 и 0.739 для VGG19 бэкбона. Рост AUC показывает улучшение в разделимости классов, что повышает потенциальную надёжность модели. Recall также улучшается при генеративной аугментации во всех режимах синтетики.

Таблица 30 – Результаты классификации CBIS-DDSM с использованием EfficientNet-V2-S (перцептивные бэкбоны VGG19 и DenseNet121 в критике DGAN-WP-TL с ImageNet весами)

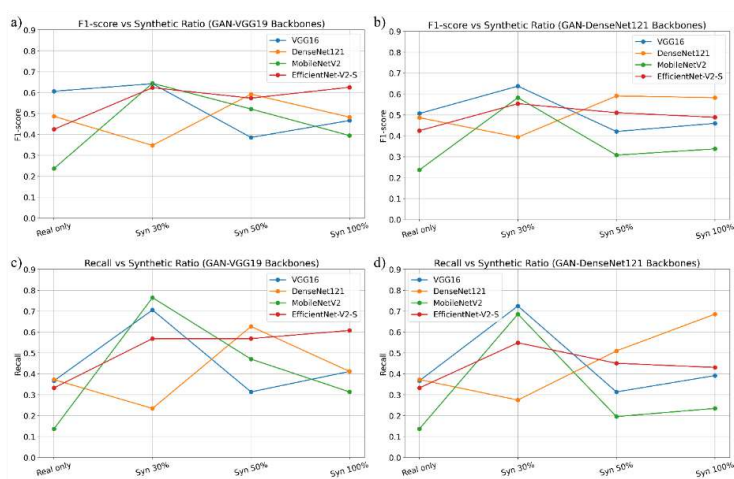
Доля синтетики	F1-score VGG19/DenseNet121	AUC VGG19/DenseNet121	Accuracy VGG19/DenseNet121	Precision VGG19/DenseNet121	Recall VGG19/DenseNet121	Specificity VGG19/DenseNet121
Real Only	0.425/0.425	0.677/0.677	0.629/0.629	0.586/0.586	0.333/0.333	0.836/0.836
Syn 30%	0.624/0.554	0.769/0.705	0.718/0.637	0.690/0.560	0.569/0.549	0.822/0.699
Syn 50%	0.574/0.511	0.670/0.722	0.653/0.645	0.580/0.590	0.569/0.451	0.712/0.781
Syn 100%	0.626/0.489	0.739/0.686	0.702/0.629	0.646/0.564	0.608/0.431	0.767/0.767

Выбор предобученной модели в двухветвевом критике метода DGAN-WP-TL напрямую влияет на классификационную модель обнаружения РМЖ. Синтетика, сгенерированная с помощью DGAN-WP-TL и VGG19 бэкбона в критике, улучшает клинически значимые метрики и обнаружение миноритарного класса.

Рисунок 28а, 28б, 28с, 28д показывает зависимость F1-score и Recall от доли синтетики. Графики показывают, что максимальные значения F1-score и Recall достигаются при 30% синтетики, кроме DenseNet121, для которой определено оптимальное добавление синтетики в размере 50%.

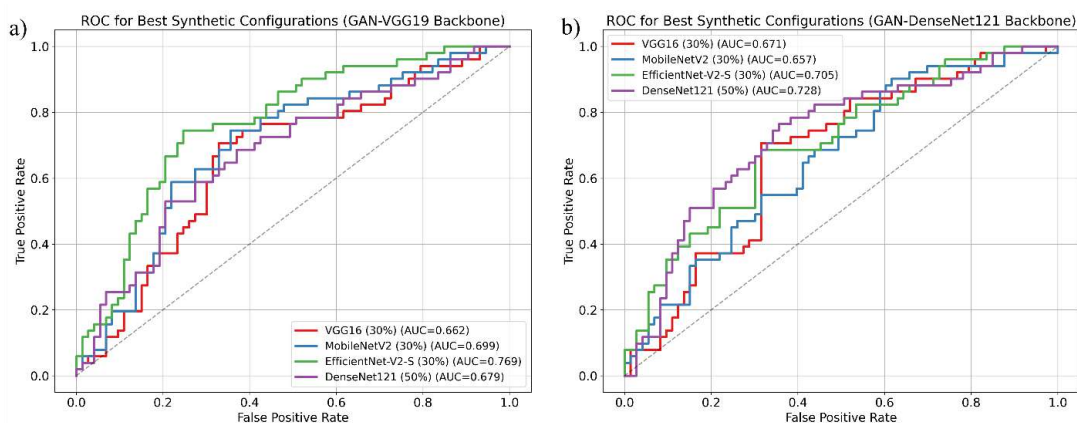
ROC-кривые отображены на рисунке 29 и показывают улучшение разделимости классов и совпадают с ростом метрик F1-score и Recall.

В условиях выраженного дисбаланса классов, характерного для датасета CBIS-DDSM, метрика Ассигасу оказывается слабоинформативной, поскольку отражает преимущественно корректную классификацию доминирующего benign-класса. В связи с этим при анализе результатов основное внимание уделяется метрикам Recall, F1-score и AUC, более чувствительным к изменениям способности моделей выявлять злокачественные случаи.



а – F1-score при бэкбоне VGG19; б – F1-score при бэкбоне DenseNet121; в – Recall при бэкбоне VGG19; д – Recall при бэкбоне DenseNet121

Рисунок 28 – Влияние доли синтетических изображений, сгенерированных DGAN-WP-TL, на качество классификации CBIS-DDSM



а – VGG19 бэкбон; б – DenseNet121 бэкбон

Рисунок 29 – ROC-кривые лучших по F1-score конфигураций классификаторов VGG16 (30%), MobileNetV2 (30%), EfficientNet-V2-S (30%), DenseNet121 (50%), обученных на синтетических изображениях, сгенерированных DGAN-WP-TL

3.4.4 Классификация CBIS-DDSM с использованием DGAN-WP-TL + RadImageNet-весов (DenseNet121 и InceptionV3 бэкбоны)

В подразделе 3.4.4 анализируется влияние синтетических изображений, сгенерированных DGAN-WP-TL с перцептивными бэкбонами критика, предобученными на RadImageNet. Результаты эксперимента позволят оценить влияние доменно-специфичных весов на классификацию РМЖ. Рассматриваются DenseNet121 и InceptionV3, которые были выбраны как наилучшие конфигурации при оптимальном λ среди RadImageNet моделей при 1000 эпох (DenseNet121 (RadImageNet), $\lambda=3.0$ и InceptionV3 (RadImageNet), $\lambda=0.5$).

Оценка выполнена для четырёх классификаторов (VGG16, DenseNet121, MobileNetV2, EfficientNet-V2-S) при уровнях синтетического пополнения 30, 50 и 100%; результаты приведены в таблицах 31, 32, 33, 34 и на рисунке 30.

Таблица 31 – Результаты классификации CBIS-DDSM с использованием VGG16 (перцептивные бэкбоны InceptionV3 и DenseNet121 в критике DGAN-WP-TL с RadImageNet весами)

Доля синтетики	F1-score InceptionV3/DenseNet121	AUC InceptionV3/DenseNet121	Accuracy InceptionV3/DenseNet121	Precision InceptionV3/DenseNet121	Recall InceptionV3/DenseNet121	Specificity InceptionV3/DenseNet121
Real Only	0.607/0.607	0.662/0.662	0.645/0.645	0.557/0.557	0.367/0.367	0.630/0.630
Syn 30%	0.624/0.661	0.678/0.686	0.669/0.694	0.586/0.607	0.667/0.725	0.671/0.671
Syn 50%	0.381/0.321	0.652/0.681	0.581/0.556	0.485/0.433	0.314/0.255	0.767/0.767
Syn 100%	0.536/0.511	0.675/0.676	0.637/0.629	0.565/0.558	0.510/0.471	0.726/0.740

Таблица 32 – Результаты классификации CBIS-DDSM с использованием DenseNet121 (перцептивные бэкбоны InceptionV3 и DenseNet121 в критике DGAN-WP-TL с RadImageNet весами)

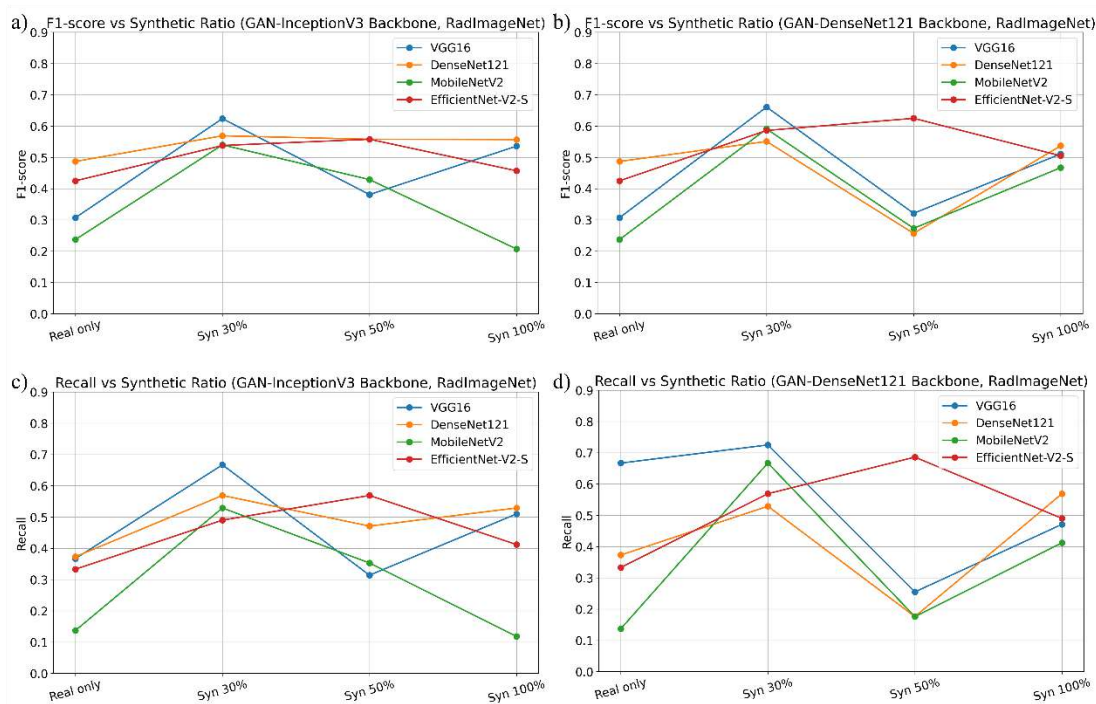
Доля синтетики	F1-score InceptionV3/DenseNet121	AUC InceptionV3/DenseNet121	Accuracy InceptionV3/DenseNet121	Precision InceptionV3/DenseNet121	Recall InceptionV3/DenseNet121	Specificity InceptionV3/DenseNet121
Real Only	0.487/0.487	0.733/0.733	0.677/0.677	0.704/0.704	0.373/0.373	0.890/0.890
Syn 30%	0.569/0.551	0.670/0.703	0.645/0.645	0.569/0.574	0.569/0.529	0.699/0.726
Syn 50%	0.558/0.257	0.696/0.610	0.694/0.581	0.686/0.474	0.471/0.176	0.849/0.863
Syn 100%	0.557/0.537	0.704/0.675	0.653/0.597	0.587/0.509	0.529/0.569	0.740/0.616

Таблица 33 – Результаты классификации CBIS-DDSM с использованием MobileNetV2 (перцептивные бэкбоны InceptionV3 и DenseNet121 в критике DGAN-WP-TL с RadImageNet весами)

Доля синтетики	F1-score InceptionV3/DenseNet121	AUC InceptionV3/DenseNet121	Accuracy InceptionV3/DenseNet121	Precision InceptionV3/DenseNet121	Recall InceptionV3/DenseNet121	Specificity InceptionV3/DenseNet121
Real Only	0.237/0.237	0.701/0.701	0.637/0.637	0.875/0.875	0.137/0.137	0.986/0.986
Syn 30%	0.540/0.591	0.647/0.685	0.629/0.621	0.551/0.531	0.529/0.667	0.699/0.589
Syn 50%	0.429/0.273	0.659/0.694	0.613/0.613	0.545/0.600	0.353/0.176	0.795/0.918
Syn 100%	0.207/0.467	0.655/0.650	0.629/0.613	0.857/0.538	0.118/0.412	0.986/0.753

Таблица 34 – Результаты классификации CBIS-DDSM с использованием EfficientNet-V2-S (перцептивные бэкбоны InceptionV3 и DenseNet121 в критике DGAN-WP-TL с RadImageNet весами)

Доля синтетики	F1-score InceptionV3/DenseNet121	AUC InceptionV3/DenseNet121	Accuracy InceptionV3/DenseNet121	Precision InceptionV3/DenseNet121	Recall InceptionV3/DenseNet121	Specificity InceptionV3/DenseNet121
Real Only	0.425/0.425	0.677/0.677	0.629/0.629	0.586/0.586	0.333/0.333	0.836/0.836
Syn 30%	0.538/0.586	0.710/0.702	0.653/0.669	0.595/0.604	0.490/0.569	0.767/0.740
Syn 50%	0.558/0.625	0.697/0.694	0.629/0.661	0.547/0.574	0.569/0.686	0.671/0.644
Syn 100%	0.457/0.505	0.676/0.668	0.597/0.605	0.512/0.521	0.412/0.490	0.726/0.685



a – F1-score при бэкбоне InceptionV3; b – F1-score при бэкбоне DenseNet121; c – Recall при бэкбоне InceptionV3; d – Recall при бэкбоне DenseNet121

Рисунок 30 – Влияние доли синтетических изображений, сгенерированных DGAN-WP-TL с RadImageNet весами, на качество классификации CBIS-DDSM

Далее приведены результаты downstream-классификации для синтетических выборок, сгенерированных с использованием RadImageNet-бэкбонов DenseNet121 и InceptionV3, при различных уровнях синтетического пополнения обучающей выборки.

Для VGG16 в режиме 30% синтетики наблюдаются повышение Recall с 0.367 до 0.667-0.725 и F1-score с 0.607 до 0.624-0.661 (таблица 31). При 50 и 100% синтетики наблюдается деградация в метриках при генеративной аугментации, тогда как при 100% отмечается частичное восстановление Recall и F1-score. При этом лучшие результаты достигаются с использованием бэкбона DenseNet121 с RadImageNet весами в критике DGAN-WP-TL. При 100% синтетики в обеих моделях меняется баланс чувствительности и специфичности; отдельные режимы (в частности, 50% синтетики для DenseNet121-бэкбона) приводят к деградации метрик, что указывает на чувствительность модели к распределению синтетических признаков.

Для MobileNetV2 наблюдается значительный рост в Recall до 0.529 и 0.667 при InceptionV3 и DenseNet121 бэкбонов в критике, соответственно. Тогда как 50% и 100% только ухудшают обнаружение миноритарного класса.

EfficientNet-V2-S демонстрирует наиболее стабильную динамику метрик (таблица 34): при 30% синтетики F1-score возрастает до 0.538-0.586, а оптимальные значения достигаются при 50% (F1-score до 0.625, Recall до 0.686); при 100% синтетики наблюдается умеренное снижение метрик.

На рисунке 30 приведены зависимости F1-score и Recall от доли синтетики для всех классификаторов при бэкбонах InceptionV3 и DenseNet121. Сравнение результатов 3.4.4 с оценками качества генерации из раздела 3.2 показывает, что улучшение распределительных метрик (KID/FID) не всегда линейно отражается в максимуме F1-score.

Выводы по третьему разделу

В третьей главе представлены результаты автоматизированного обнаружения РМЖ с использованием глубоких сверточных нейронных сетей с применением генеративной аугментации метода DGAN-WP-TL. Показано, что разработанный метод DGAN-WP-TL и сгенерированные с его помощью синтетические данные положительно влияют на downstream-классификацию и улучшают обнаружение злокачественных опухолей, что является жизненно важным в медицинской диагностике. Классификация проведена на трех датасетах BUSI (УЗИ), CBIS-DDSM (маммография) и Alzheimer MRI (MPT), что позволяет сформулировать следующие выводы:

1. Экспериментально установлено, что разработанный метод DGAN-WP-TL устойчиво генерирует изображения миноритарного класса РМЖ, что подтверждается распределительными метриками, а также на перцептивном и структурном уровнях. Кроме того, результаты позволяют с уверенностью подтвердить переносимость модели и на другие медицинские наборы данных. Однако требует дополнительной экспериментальной апробации.

2. Определено оптимальное значение веса градиентного штрафа, который зависит от датасета. Так для BUSI оптимальное значение веса градиентного штрафа составило 5.0, для маммографического датасета 0.5, а для датасета с заболеванием Альцгеймера 7.0.

3. Установлено, что необходим комплексный подход к качеству генерации, основываясь на совокупности метрик, а именно FID, KID, LPIPS, MS-SIM, что позволяет достичь оптимального качества синтетических изображений, опираясь на статический и визуальный анализ.

4. Анализ признаков распределений (PCA и t-SNE) показывает перекрытие реальных и синтетических данных (наиболее выражено для BUSI; умеренно – для CBIS-DDSM), что согласуется с распределительными метриками и подтверждает применимость синтетики для расширения обучающей выборки.

5. Downstream-эксперименты по классификации показали, что синтетическое пополнение malignant-класса может приводить к росту Recall, F1-score и AUC, однако эффект зависит от доли синтетики, параметра λ и архитектуры классификатора.

6. Абляционные исследования подтвердили, что предложенная архитектура DGAN-WP-TL с двухветвевым критиком и использованием схемы WGAN-GP является ключевым фактором достижения устойчивого качества генерации синтетических медицинских изображений.

7. Сравнение со StyleGAN2 показывает, что при большем объеме данных (CBIS-DDSM) StyleGAN2 может достигать более низких KID/FID, однако наблюдается коллапс мод, что является критичным в генеративной аугментации

данных, особенно в медицинской диагностике. На малых и экстремально малых выборках (BUSI и Alzheimer MRI) DGAN-WP-TL демонстрирует более устойчивую динамику обучения и, по LPIPS/MS-SSIM, сопоставимое или более благоприятное разнообразие, что соответствует фокусу работы на практической полезности синтетики для downstream-классификации.

В целом результаты главы показывают, что синтетическая генерация медицинских изображений может рассматриваться как управляемый механизм балансировки данных, однако её практический эффект определяется модальностью изображений, режимом регуляризации и архитектурой модели, а также сценарием downstream-применения. Выявленные закономерности используются для дальнейшего обсуждения применимости предложенного подхода в главе 4.

4 АНАЛИЗ И ОБОБЩЕНИЕ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

Раздел 4 содержит анализ и интерпретацию результатов, полученных в разделе 3, и направлена на оценку эффективности автоматизированной системы обнаружения РМЖ при использовании глубоких CNN и генеративной аугментации методом DGAN-WP-TL для миноритарного класса. Дополнительно рассматриваются выявленные закономерности и причинно-следственные связи как в генеративной (вспомогательной), так и в классификационной (основной) части системы. По результатам первых трех глав сформированы следующие гипотезы:

Архитектурная гипотеза: предложенный метод DGAN-WP-TL обеспечивает более устойчивое обучение генератора и двухветвевых критика. Интеграция предобученной модели в первую ветвь критика и обучаемой сверточной ветви снижает риск появления коллапса мод по сравнению с одноветвевыми критиками.

Гипотеза о регуляризации: применение WGAN-GP функции потерь генерирует более качественные изображения по сравнению с классической функцией потерь при малом количестве медицинских изображений.

Гипотеза генеративной аугментации: аугментация миноритарного класса с помощью синтетических изображений улучшает обнаружение РМЖ с использованием CNN, о чем говорят рост метрик Recall, F1-score и AUC. Однако эффект зависит от домена (датасета), архитектуры классификатора и доли синтетики.

Гипотеза несоответствия метрик: наилучшие результаты метрик генерации (FID/KID) не приводят к лучшим результатам downstream-классификации, а также требуется соблюдать баланс между реалистичностью и внутриклассовым разнообразием.

Для проверки архитектурной гипотезы были проведены абляционные эксперименты двухветвевых критиков и его одноветвевых, составляющих с одновременным подтверждением гипотезы регуляризации.

Основные результаты предложенного подхода автоматизированного обнаружения РМЖ представлены в опубликованных автором ряде работ. Статьи представляют исследования по внедрению генеративной аугментации DGAN-WP-TL в классификацию РМЖ в малых медицинских датасетах, а также выбору функции потерь, оценке StyleGAN-семейства и переносу предобученных признаков на медицинский домен [59, р. 99-103; 60, р. 122-126; 67, р. 255-268; 75, р. 99-111; 109, р. 254-268].

4.1 Устойчивость и качество генерации: роль архитектуры, функции потерь и режима обучения

Одной из важных характеристик разработанного метода DGAN-WP-TL должна быть устойчивость генерации, которая зависит от архитектуры, функции потерь и режима оптимизации. Устойчивость метода заключается в способности генерировать реалистичные ультразвуковые и маммографические изображения с высокой вариативностью без коллапса мод, особенно в медицинских датасетах.

Согласно математической модели, предложенной в разделе 2 и экспериментам в разделе 3, архитектура критика DGAN-WP-TL реализует устойчивую генерацию реалистичных изображений за счет двух ветвей в критике, а именно перцептивной с предобученными весами трансферного обучения и кастомной ветви со сверточными слоями. Эффективность конкатенации признаков из двух ветвей доказана в абляционных экспериментах (таблица 15) в Full-конфигурации, обеспечивая качественную генерацию в трех датасетах согласно метрикам генерации. При этом более простые критики с одной ветвью, такие VGG19-Only и Conv-Only сильно ухудшают качество синтетических изображений, показывая неспособность критика к обучению и передаче деталей реальных изображений во всех трех датасетах (рисунок 23). В итоге только двухветвевой критик оказался способным генерировать синтетические изображения, которые отличаются реализмом. Гипотеза об устойчивости DGAN-WP-TL в малых медицинских датасетах подтверждено [67, p. 2556687].

Дополнительно абляционные эксперименты с функциями потерь WGAN-GP и BCE показали, что классическая функция потерь не справилась с задачей и столкнулась с коллапсом мод и с низким качеством сгенерированных изображений в отличие от WGAN-GP. Для генеративной аугментации, которая используется как вспомогательный инструмент, критически важно генерировать изображения, которые будут улучшать дальнейшую классификацию, а не ухудшать. Тем самым гипотеза о методологической предпочтительности WGAN-GP для генерации медицинских изображений в режиме малых выборок также подтверждается экспериментально [75, p. 99-111].

Дополнительным параметром, являющимся критически важным для устойчивости обучения DGAN-WP-TL, является параметр batch size, который чувствителен при дефиците данных. Согласно таблице 17 исследовано, что увеличение размера батча, существенно снижает распределительные метрики (FID и KID). Это связано с выраженным дефицитом данных, который характерен для медицинских датасетов. В экспериментальной постановке наилучшее сочетание реалистичности и контролируемого разнообразия достигается при batch size = 2, что согласуется с общей логикой обучения GAN в условиях малых выборок [67, p. 2556687]. При batch size=2, каждое изображение вносит существенный вклад в обучение модели, тем самым позволяя избегать усредненной оценки при большем размере батча и генерируя наиболее реалистичные изображения.

Наконец, важным моментом оценки полученных результатов, является корректный выбор метрик генерации в условиях малых медицинских датасетов, о чем свидетельствуют результаты экспериментов в таблице 18, согласно которой FID зависит от количества выборки и колеблется сильнее, требуя для устойчивых результатов тысячи изображений. Тогда как, KID направлен на калькуляцию метрик с малым количеством изображений, что позволяет установить KID как главную распределительную метрику в данной диссертационной работе.

В результате анализа, проведенного в данном подразделе, подтверждено что предложенная архитектура DGAN-WP-TL с двухветвевым критиком достигает более устойчивого результата в генерации, а именно предобученная ветвь, кастомная ветвь, функция потерь WGAN-GP и гиперпараметры обучения $\text{batch size} = 2$). Таким образом, гипотезы об архитектуре и регуляризации подтверждены экспериментами, что позволяет рекомендовать предложенный метод DGAN-WP-TL для генерации реалистичных ультразвуковых и маммографических изображений в малых медицинских датасетах для дальнейшей классификации РМЖ.

4.2 Сравнительный анализ: StyleGAN2 и влияние предобучения и перцептивного бэкбона

Данный подраздел посвящен двум составляющим в анализе устойчивости и эффективности разработанного метода DGAN-WP-TL, а именно сравнению с SOTA-архитектурой StyleGAN2 и анализу влияния перцептивного бэкбона в критике DGAN-WP-TL с весами ImageNet и RadImageNet.

Анализ с архитектурой StyleGAN2 и DGAN-WP-TL основывается на экспериментах в подразделе 3.1.5 и результатов в таблице 11. Так, результаты на датасете BUSI демонстрируют явные преимущества DGAN-WP-TL перед StyleGAN2, тогда как StyleGAN2 не справился с генерацией реалистичных и разнообразных изображений, демонстрируя выраженный коллапс мод при экстремально высоком значении метрики MS-SSIM (f-f) = 0.9532 и низким LPIPS (f-f) = 0.0715, указывающие на одинаковые сгенерированные изображения. Однако на датасете CBIS-DDSM, StyleGAN2 достиг лучших распределительных метрик (FID/KID), но с существенным ростом MS-SSIM (f-f) = 0.6199, что говорит о частичном коллапсе мод и повторяемости синтетических паттернов. В свою очередь, DGAN-WP-TL выигрывает за счет отсутствия коллапса мод и внутриклассового разнообразия. В датасете Alzheimer MRI StyleGAN2 демонстрирует подобное поведение, как и в BUSI датасете: выраженный коллапс мод и неспособность генерировать реалистичные изображения. DGAN-WP-TL, напротив, демонстрирует баланс между распределительными, перцептивными и структурными метриками, и способность генерировать изображения, улучшающие downstream-классификацию в дальнейшем. В совокупности, результаты универсальной архитектуры семейства StyleGAN демонстрируют неспособность генерировать реалистичные и разнообразные изображения в малых медицинских датасетах. В этом контексте преимущество DGAN-WP-TL проявляется прежде всего в устойчивости генерации и сохранении разнообразия синтетики, что делает модель более пригодной для практических задач downstream-классификации при дефиците данных.

Вторая составляющая анализа устойчивости предложенного метода является определение влияния выбора бэкбона и источника весов в предобученных моделях в методе DGAN-WP-TL. Эксперименты подраздела 3.2 были проведены только на датасете CBIS-DDSM, так как данная модальность медицинской диагностики представлена в доменно-специфичных весах RadImageNet, что позволило провести полноценное сравнение между ImageNet

и RadImageNet при одинаковых настройках обучения. Результаты экспериментов (таблицы 13, 14 раздела 3) показывают, что наилучшее глобальное распределительное соответствие достигается при использовании ImageNet-предобучения. В частности, конфигурация VGG19 (ImageNet) при $\lambda = 0.5$ демонстрирует минимальные значения $KID = 0.17$ и $FID = 178.59$, которые не были превзойдены ни одной из конфигураций с предобучением на RadImageNet. Можно с уверенностью утверждать, что доменно-специфичное обучение с RadImageNet не улучшает качество генерации, а является дополнительным источником весов и проверки устойчивости метода. Кроме того, обучение с RadImageNet требует большего количества эпох и вычислительных ресурсов, что также сказывается на выборе в сторону ImageNet. Перцептивно-структурные показатели и PCA-проекции не выявляют принципиальных различий в глобальной структуре распределений с ImageNet или RadImageNet весами, что согласуется с близкими значениями KID и FID . Автором диссертации также рассматривалась роль доменно-ориентированного предобучения (RadImageNet) и transfer learning при генерации редких патологий и было доказано экспериментально, что RadImageNet не доминирует в генеративной аугментации [67, p. 2556687; 109, p. 89-95].

Для рассматриваемой экспериментальной постановки на CBIS-DDSM наиболее устойчивым по совокупности характеристик оказался вариант с ImageNet-предобучением, тогда как конфигурации с RadImageNet следует рассматривать как дополнительную валидацию и перспективное направление дальнейшей оптимизации при расширении выборки или изменении постановки задачи.

Таким образом, сравнительный анализ DGAN-WP-TL с StyleGAN2 показывает эффективность и устойчивость разработанного метода в диссертационной работе при дефиците данных в медицинских датасетах. В свою очередь, StyleGAN2 продемонстрировал коллапс мод и неспособность генерировать реалистичные изображения и быть потенциально применимым в обнаружении РМЖ. Далее, анализ источников предобучения дополнительно показывает, что доменно-специфические веса не гарантируют улучшения качества безусловной генерации при малом объёме данных, тогда как ImageNet-предобучение обеспечивает более стабильную оптимизацию. В совокупности эти результаты подтверждают, что практическая ценность генеративной модели в медицинских задачах определяется не минимизацией распределительных метрик как таковой, а её способностью формировать синтетические данные, полезные для последующего обучения CNN-классификаторов.

4.3 Сравнение предложенного метода DGAN-WP-TL с существующими подходами

Важной составляющей анализа устойчивости и конкурентоспособности метода DGAN-WP-TL является сравнительный анализ с ранее опубликованными существующими методами (таблицы 35, 36) по ряду параметров (FID/KID , количество изображений, тип задач и др.), напрямую влиявшие на результаты генерации.

Таблица 35 – Сравнение GAN-подходов для синтеза изображений BUSI

Автор	Метод	FID	KID	Размер изображения	Датасет	Количество изображений	Тип задачи
Liu et al.	GSDA	68.78	–	256×256	BUSI	780	Полуконтролируемая генерация
Alruily et al.	IGAN	41.86	–	512×512	BUSI	780	Генерация
Luo et al.	2S-BUSGAN	101.00	0.6238	256×256	BUSI	780	Генерация + сегментация
Jiménez-Gaona et al.	BraNet	116.85	0.06	128×128	BUSI + DatasetB	1041	Аугментация
Jiménez-Gaona et al.	CycleGAN	116.03	0.20	128×128	BUSI + DatasetB	1041	Перенос стиля
<i>Предлагаемый метод</i>	<i>DGAN-WP-TL</i>	<i>179.4158</i>	<i>0.1448</i>	<i>512×512</i>	<i>BUSI</i>	<i>210</i>	<i>Безусловная генерация malignant-класса</i>
Примечание – Составлено по источникам [30, с19585; 32, р. 8599; 33, р. 8614; 34, р. 2737-2755; 35, р. 106255; 67, р. 2556687]							

Сравнение GAN-подходов для синтеза ультразвуковых изображений показывает существенное более низкие значения в метрике FID в методах GSDA [30, р. с19585], IGAN [32, р. 8599], и 2S-BUSGAN [33, р. 8614] из таблицы 35. Преимущество FID происходит за счет большего количества изображений в диапазоне 780-1041, что почти в 4 раза больше чем в DGAN-WP-TL с 210 изображениями. По метрике KID DGAN-WP-TL существенно выигрывает у 2S-BUSGAN [33, р. 8614] и CycleGAN [35, р. 106255], однако BraNet [34, р. 2737-2755] демонстрирует KID равный 0.06 при меньшем размере изображений, но большем количестве данных. На этом фоне DGAN-WP-TL обладает существенным преимуществом за счет малого количества изображений в размере 210 malignant-класса BUSI при разрешении 512×512, что закономерно приводит к более высоким значениям FID. При этом сохраняется стабильное значение KID, являющееся более устойчивой метрикой при оценке генерации на малых выборках, что дополнительно подтверждается анализом чувствительности (раздел 4.1).

Таблица 36 – Сравнение GAN-подходов для синтеза маммографических изображений (CBIS-DDSM)

Автор	Метод	FID	Размер изображения	Датасет	Количество изображений	Тип задачи
1	2	3	4	5	6	7
Garrucho et al.	CycleGAN	73.16	1332×800	OPTIMAM	≈894	Перенос плотности
Park et al.	StyleGAN2	4.38	512×512	Private	105948	Генерация норм

Продолжение таблицы 36

1	2	3	4	5	6	7
Oyelade et al.	ROIImammoGAN	1.2 млн	64-128	MIAS	≈300	Генерация масс
Fan et al.	BreastGAN	21.25	128×128	DDSM	2600 ROI	Benign/Malignant
Prodan et al.	StyleGAN-XL	13.66	256 / 512	ADMANI	11913	Генерация positive
<i>Предлагаемый метод</i>	DGAN-WP-TL (VGG19, ImageNet, 500 эпох)	182.35	224×224	CBIS-DDSM	509	Безусловная генерация malignant-класса
	DGAN-WP-TL (VGG19, ImageNet, 1000 эпох)	178.59	224×224	CBIS-DDSM	509	Безусловная генерация malignant-класса
Примечание – Составлено по источникам [36, р. 1044496; 37, р. 2925; 38, р. 6166; 40, р. 4272; 67, р. 2556687; 95, р. 431-436; 109, р. 254-268]						

Сравнительный анализ существующих разработок в области маммографии показывает, что результаты FID достигаются за счет крупных медицинских выборок от 2600 до 105 тысяч и малого размера изображений в таких методах, как StyleGAN [37, р. 2925], BreastGAN [95, р. 431-436], StyleGAN-XL [40, р. 4272]. Тогда как DGAN-WP-TL обучается на малом размере датасета и на 509 полноразмерных маммографиях.

В рамках настоящей работы конфигурация с 500 эпохами обучения рассматривается как более сбалансированный режим, обеспечивающий устойчивую генерацию при разумном компромиссе между качеством синтеза и вычислительной эффективностью. Данный баланс позволяет добиться стабильности и воспроизводимости результатов, а не просто минимальных значений метрик генерации.

Таким образом, необходимо сравнивать метод DGAN-WP-TL с другими методами в комплексе параметров, минимальными которыми являются, по крайней мере, количество изображений и их размер. DGAN-WP-TL обладает явным преимуществом устойчивой генерации в малых медицинских датасетах и не предназначен для прямой конкуренции с универсальными GAN-архитектурами, оптимизированными под масштабные выборки и условные постановки.

С точки зрения сформулированных в начале главы гипотез, результаты данного подраздела подтверждают гипотезы, подчёркивая необходимость архитектурной стабилизации и регуляризации генеративного процесса при работе с малыми медицинскими выборками. При этом, согласно гипотезе о минимизации метрик, подтверждено, что только минимальные распределительные метрики не достаточны для практической ценности синтетических данных, так как для автоматизированных систем в медицине критически важно разнообразие входных данных для CNN-классификаторов.

4.4 Анализ влияния синтетических данных на качество классификации опухолей

Главным результатом в данной диссертационной работе является обнаружение злокачественного класса с синтетическим расширением методом DGAN-WP-TL. Под обнаружением РМЖ выполняется классификация с генеративной аугментацией миноритарного класса с целью улучшения чувствительности CNN-классификаторов к редкому malignant-классу, что определяется доменом данных, архитектурой классификатора и долей синтетического пополнения обучающей выборки. Эффект синтетической балансировки как средства повышения чувствительности модели к редкому классу и необходимость контролируемой дозировки синтетики подтверждаются и в авторских исследованиях по мультидоменной генерации и балансировке медицинских датасетов [67, p. 2556687; 109, p. 89-95].

В результате downstream-классификация показала выраженную датасет-зависимость эффекта синтетического пополнения. Для ультразвукового датасета BUSI наблюдается относительно согласованная связь между улучшением качества генерации и ростом диагностических метрик. Это объясняется свойствами домена: УЗИ-изображения характеризуются более однородной визуальной структурой, а признаки злокачественности в меньшей степени зависят от сложных многомасштабных текстур. В результате повышение реалистичности и разнообразия синтетических изображений чаще напрямую транслируется в рост Recall и F1-score без резких провалов. Для BUSI конфигурации, демонстрирующие наилучшие значения распределительных метрик генерации, одновременно обеспечивают наиболее устойчивое улучшение downstream-показателей, что указывает на согласованность генеративных и классификационных критериев в данном домене.

Для маммографического датасета CBIS-DDSM наблюдается существенно более сложная и немонотонная картина. Маммограммы с микрокальцификациями характеризуются высокой вариативностью фоновой ткани и сложными локальными паттернами, что приводит к ослаблению прямой связи между качеством генерации и эффективностью классификации. Добавление синтетических malignant-изображений позволяет классификаторам «увидеть» редкий класс и существенно повысить Recall и F1-score при умеренных долях синтетического пополнения. Однако при дальнейшем увеличении доли синтетики эффект становится нестабильным, что указывает на риск смещения обучающего распределения в сторону узкого или артефактного синтетического подраспределения, особенно критичного для сложных маммографических данных.

Полученные результаты подтверждают, что практическая полезность синтетики достигается не в режиме максимального наращивания числа синтетических изображений, а в зоне умеренного пополнения обучающей выборки. В экспериментальной постановке для CBIS-DDSM наиболее стабильные улучшения downstream-метрик достигаются, как правило, при доле синтетических данных порядка 30-50%. Синтетическое расширение можно

рассматривать как регуляризацию обучения за счет варьирования процентного содержания синтетических данных в обучающей выборке.

На результаты классификации с синтетическим расширением также влияют и сам классификаторы. Согласно экспериментам установлено, что необходимо рассматривать генерацию и классификацию как единый пайплайн, анализируя генерацию и классификацию совместно, а не изолированно. В ряде случаев минимальные значения FID или KID достигаются при параметрах генерации, которые не приводят к наилучшим значениям Recall и F1-score. Это указывает на то, что оптимизация генеративной модели исключительно по распределительным метрикам является недостаточной, если синтетические изображения используются как обучающие данные. Решающее значение приобретает баланс между реалистичностью и внутриклассовым разнообразием: синтетика должна расширять распределение malignant-класса, а не воспроизводить ограниченный набор типовых паттернов. Выявленное несоответствие между оптимальностью распределительных метрик и максимизацией downstream-показателей также согласуется с выводами авторских работ, где подчёркивается значимость внутриклассового разнообразия и корректного протокола валидации при использовании синтетики для обучения классификаторов [67, p. 2556687].

Гипотеза генеративной аугментации подтверждается. Синтетическое пополнение обучающей выборки повышает чувствительность классификаторов к редкому malignant-классу, однако эффект является датасет-зависимым и требует контролируемой дозировки.

Гипотеза несоответствия метрик также подтверждается. Минимизация распределительных метрик генерации (FID/KID) сама по себе не гарантирует улучшения downstream-классификации; практическая полезность синтетики определяется сочетанием реалистичности, разнообразия (LPIPS/MS-SSIM) и фактического влияния на Recall, F1-score и AUC.

В целом результаты данного подраздела показывают, что синтетические данные, сгенерированные моделью DGAN-WP-TL, при корректном выборе конфигурации генерации и доли синтетического пополнения обеспечивают улучшение эффективности для задач обнаружения опухолей. Это подтверждает целесообразность использования генеративных моделей не как самоцели, а как вспомогательного инструмента повышения чувствительности CNN-классификаторов в условиях выраженного дефицита и дисбаланса медицинских данных.

4.5 Сравнительный анализ подходов к классификации РМЖ с использованием GAN-аугментации

В данном подразделе 4.5 выполнен сравнительный анализ подходов к классификации РМЖ с использованием GAN-аугментации, включая предложенный интегрированный метод DGAN-WP-TL (таблицы 37, 38). Сопоставление с опубликованными исследованиями показывает, что большинство существующих работ оценивают влияние синтетических данных на классификацию в условиях, отличающихся от рассматриваемой в настоящем

исследовании постановки. Как правило, используются более крупные и частично сбалансированные выборки, условные или полу-контролируемые схемы генерации, сегментационные ограничения, парные преобразования доменов и синтез ROI-фрагментов вместо полноразмерных изображений. Кроме того, нередко применяются частные или устаревшие датасеты (например, MIAS), что ограничивает воспроизводимость и практическую интерпретацию результатов [35, p. 106255; 41, p. 105456].

Принципиальным отличием предложенного подхода является решение более сложной задачи – безусловной генерации изображений редкого malignant-класса в условиях выраженного дефицита данных с последующим контролируемым использованием синтетики для обучения CNN-классификаторов. Отказ от условных меток, сегментационных масок и априорных ограничений повышает сложность генеративного обучения и делает прямое сопоставление абсолютных значений метрик генерации (например, FID) с результатами других работ методологически некорректным без учёта контекста экспериментальной постановки.

Таблица 37 – Сравнение подходов к классификации РМЖ на УЗИ (BUSI) с использованием GAN-аугментации

Автор	GAN-модель	Тип GAN	Датасет (images)	Роль GAN	Метрики генерации	Метрики классификации	CNN	Ограничения
Al-Dhabyani et al., 2019	DAGAN	Uncond.	BUSI, Dataset B (780, 163)	Аугментация	–	ACC ↑ до 99%	NASNetLarge	Нет метрик генерации
Liu et al., 2023	GSDA	Uncond.	BUSI (780)	Генерация	FID: 62.92	ACC: до 97.9%	VGG16	Ограниченное количество метрик
Jiménez-Gaona et al., 2024	SNGAN	Condit./Uncond.	BUSI (780)	Аугментация	FID: 116.03	ACC: 43.9% B / 55.4% B	ResNet18	Ограниченное количество метрик
<i>Предлагаемый метод</i>	<i>DGAN-WP-TL (VGG19, ImageNet)</i>	<i>Uncond.</i>	<i>BUSI (210)</i>	<i>Баланс malignant</i>	<i>FID: 179.4158 KID: 0.1448</i>	<i>ACC: 81.2%, F1: 72.1%, Recall: 75.6%, AUC: 87.1%</i>	<i>VGG16 + Syn 50%</i>	–
Примечание – Составлено по источникам [30, p. e19585; 35, p. 106255; 67, p. 2556687; 91]								

Таблица 38 – Сравнение подходов к классификации РМЖ на маммографических изображениях с интеграцией GAN-моделей

Автор	GAN-модель	Тип GAN	Датасет (images)	Роль GAN	Метрики генерации	Метрики классификации	CNN	Ограничения
1	2	3	4	5	6	7	8	9
Kim et al., 2018	ICADx	Condit.	DDSM (1088 ROI)	Генерация признаков	–	ACC: 70.6% AUC: 0.908	VGG	Нет метрик генерации
Fan & Jiao, 2021	BreastGAN	Condit.	DDSM (2600)	Аугментация	FID: 21.25	ACC: 86.6% F1: 0.863	TBMSL-Net	Нет валидации; низкое разрешение
Park et al., 2023	StyleGAN2	Uncond.	Asan Medical Center Dataset (105948)	Генерация нормальных случаев	–	AUC: 70.0%	–	Вычислительные ресурсы, нет метрик генерации

Продолжение таблицы 38

1	2	3	4	5	6	7	8	9
Joseph et al., 2024	Prior-guided cGAN	Condit	MIAS (322)	Аугментация	—	ACC: 98.53% F1: 98.53%	—	Нет метрик генерации
Jiménez-Gaona et al., 2024	SNGAN, WGAN-GP	Condit / Uncond.	Mini-MIAS, INbreast, CBIS-DDSM (10239)	Аугментация ROI	FID: 58.80 KID: 0.5959	ACC:33.8%, F1: 34%, Recall: 40%	ResNet18	Аугментация с GAN ухудшило классификацию
<i>Предлагаемый метод</i>	<i>DGAN-WP-TL (VGG19, ImageNet)</i>	<i>WGAN-GP</i>	<i>CBIS-DDSM (509)</i>	<i>Баланс malignant</i>	<i>FID: 182.345 KID: 0.1795</i>	<i>ACC: 71.8%, F1: 62.4%, Recall: 56.9%, AUC: 76.9%</i>	<i>Efficient Net-V2-Syn 30%</i>	—
Примечание – Составлено по источникам [35, p. 106255; 37, p. 2925; 41, p. 105456; 67, p. 2556687; 81, p 1057522; 95, p. 431-437]								

Сравнение подходов на ультразвуковом датасете BUSI (таблица 37) показывает, что в большинстве существующих работ эффект GAN-аугментации оценивается преимущественно по метрике Ассигасу, которая в условиях дисбаланса классов не отражает клинически значимую способность модели выявлять злокачественные случаи. В ряде работ отсутствует оценка распределительных и перцептивных характеристик генерации (FID, KID, LPIPS, MS-SSIM), а режим добавления синтетических данных в обучающую выборку либо не описывается, либо не поддается воспроизведению. Подход к оценке GAN-аугментации через клинически значимые метрики (Recall/F1/AUC) и анализ дозозависимого эффекта синтетики соответствует авторским публикациям, где GAN рассматривается как инструмент балансировки с обязательной downstream-валидацией, а не как самоцель снижения FID [67, p. 2556687; 109, p. 89-95]. Примечательно, что в отдельных исследованиях применение GAN-аугментации приводит к ухудшению классификационных метрик, что указывает на отсутствие универсального эффекта «GAN → улучшение классификации» [30, p. e19585; 35, p. 106255; 91].

На этом фоне предложенный метод DGAN-WP-TL демонстрирует принципиально иной уровень прозрачности и полноты оценки. Для BUSI используется всего 210 изображений malignant-класса, при этом анализ охватывает как распределительные и перцептивные метрики генерации (FID, KID), так и клинически значимые показатели классификации (Recall, F1-score, AUC). Дополнительно в явном виде рассматриваются различные режимы синтетического пополнения (30%, 50%, 100%), что позволяет проследить зависимый эффект синтетики от количества пополнения и определить оптимальные режимы применения [67, p. 2556687].

При анализе существующих разработок в области маммографии наблюдаются различия в методах, а именно количестве изображений, задаче генерации, размерах изображений, а также метрики для оценки полученных результатов (таблица 38). Существующие методы ограничиваются только распределительной метрикой генерации, а в классификации только метрикой Ассигасу без анализа влияния на миноритарный класс в отличие от DGAN-WP-TL. В работе [35, p. 106255] генерация повлияла отрицательно, ухудшив результаты классификации из-за балансировки датасета синтетическими изображениями.

В отличие от этого, предложенный подход DGAN-WP-TL обучается на 509 полноразмерных изображениях malignant-класса CBIS-DDSM и демонстрирует воспроизводимое улучшение Recall и F1-score при умеренных долях синтетического пополнения. Существенным преимуществом является то, что в рамках работы одновременно анализируются характеристики генерации (FID, KID) и downstream-метрики классификации, что позволяет выявить случаи их несоответствия и избежать переоценки качества синтетики на основе одной метрики и принципиально отличает данный подход от большинства существующих работ [67, p. 2556687].

Важным аспектом является также вычислительная и методологическая сопоставимость. В отличие от архитектур семейства StyleGAN, требующих

значительных вычислительных ресурсов и масштабных датасетов, предложенный метод ориентирован на работу в режиме малых выборок и ограниченных вычислительных условий, что повышает его практическую применимость в реальных медицинских сценариях.

В совокупности результаты данного подраздела показывают, что предложенный интегрированный подход DGAN-WP-TL обеспечивает более полную и клинически релевантную оценку эффективности GAN-аугментации по сравнению с существующими методами. С точки зрения сформулированных в начале главы гипотез, полученные результаты согласуются с гипотезами генеративной аугментации и несоответствия метрик, подтверждая, что практическая полезность синтетических данных определяется не минимизацией отдельных метрик генерации, а их влиянием на чувствительность классификаторов к редкому malignant-классу и балансом между реалистичностью и внутриклассовым разнообразием. Именно такой подход позволяет использовать синтетические данные как инструмент повышения диагностической эффективности, а не как самоцель оптимизации генеративных показателей.

4.6 Ограничения предложенного метода, практические рекомендации и направления дальнейших исследований

Несмотря на продемонстрированную эффективность предложенного метода автоматизированного обнаружения РМЖ, следует отметить ряд ограничений, которые необходимо учитывать при интерпретации полученных результатов и определении области применимости разработанного подхода [67, p. 2556687].

Качество синтетических изображений и их влияние на downstream-классификацию в значительной степени зависят от характеристик исходных данных и выбранных режимов обучения генеративной модели. В отдельных конфигурациях, особенно в условиях экстремально малого объёма обучающих выборок, наблюдалось формирование генеративных артефактов, включая размытость границ и искажения текстур. Наличие подобных артефактов может снижать диагностическую ценность синтетических данных и указывает на необходимость дальнейшего развития функций потерь и механизмов регуляризации, ориентированных на сохранение анатомически и диагностически значимых признаков.

Сравнительный анализ в настоящей работе выполнен с использованием StyleGAN2 как одного из наиболее распространённых и репрезентативных базовых методов генерации изображений, широко применяемых в современных исследованиях по медицинской визуализации. Несмотря на появление более новых генеративных архитектур, таких как StyleGAN-XL, BigGAN-deep и модели латентной диффузии, StyleGAN2 по-прежнему рассматривается в литературе как устойчивый baseline, особенно в задачах синтеза медицинских изображений в условиях ограниченного объёма данных. Более современные генеративные модели характеризуются существенно большей архитектурной сложностью и параметрической ёмкостью и, как правило, ориентированы на

обучение на больших объёмах данных. Их применение при малых выборках сопряжено с риском переобучения, снижением воспроизводимости результатов и повышенными вычислительными затратами, а также может приводить к генерации визуально правдоподобных, но диагностически некорректных структур. В связи с этим в рамках настоящего исследования приоритет был отдан архитектуре, обеспечивающей устойчивость обучения и сопоставимость результатов в условиях дефицита размеченных медицинских изображений.

Положительный эффект синтетических данных на качество downstream-классификации проявляется неравномерно и зависит как от архитектуры используемого CNN-классификатора, так и от доли синтетического пополнения обучающей выборки. В ряде экспериментальных сценариев увеличение объёма синтетических изображений приводило к насыщению или снижению значений Recall и F1-score, что указывает на существование оптимального диапазона использования синтетики. Данное обстоятельство ограничивает универсальность метода и подчёркивает необходимость адаптивного подбора стратегии синтетического расширения данных для конкретных архитектур и диагностических задач.

Оценка эффективности обнаружения РМЖ в настоящей работе основана на стандартных метриках машинного обучения (Recall, F1-score, AUC) и не включает клиническую интерпретацию ошибок классификации. Отсутствие клинической валидации с участием врачей-радиологов не позволяет напрямую экстраполировать полученные результаты на реальные диагностические сценарии и требует дополнительных исследований в формате reader-studies и мультицентровых экспериментов.

Экспериментальная часть работы выполнена преимущественно на двух открытых датасетах (BUSI и CBIS-DDSM), которые широко используются в современных исследованиях по автоматизированному обнаружению РМЖ и де-факто выступают в роли стандартизированных бенчмарков для оценки и сопоставления алгоритмов. В то же время данные наборы обладают рядом известных ограничений, включая гетерогенность оборудования и протоколов сканирования, а также возможный этнический и демографический bias, что может приводить к эффекту domain shift при переносе моделей на клинические данные из других источников. Проведённая междоменная апробация предложенного метода на наборе Alzheimer MRI продемонстрировала его устойчивость к смене домена, однако полноценная внешняя валидация на независимых мультицентровых клинических выборках остаётся задачей дальнейших исследований.

Практические рекомендации по использованию результатов работы

Наряду с выявленными ограничениями полученные экспериментальные результаты позволяют сформулировать практические рекомендации по использованию предложенного подхода. Метод DGAN-WP-TL целесообразно применять в задачах автоматизированного обнаружения РМЖ, характеризующихся выраженным дисбалансом классов и ограниченным объёмом данных, прежде всего как инструмент повышения чувствительности классификаторов к редкому malignant-классу.

Синтетические изображения рекомендуется использовать исключительно для расширения обучающего подмножества данных. Формирование валидационных и тестовых выборок должно осуществляться только на основе реальных клинических изображений, что позволяет избежать переоценки эффективности моделей и сохранить корректность оценки диагностических метрик. Наибольший практический эффект достигается при умеренной доле синтетического пополнения обучающей выборки (ориентировочно 30-50% от объёма реального malignant-класса). Использование более высоких долей синтетики требует обязательного контроля по валидационным метрикам, прежде всего Recall и F1-score, и не рекомендуется без дополнительной проверки устойчивости классификатора.

Эффективность синтетического расширения данных в значительной степени зависит от архитектуры используемого CNN-классификатора. Наиболее устойчивую динамику диагностических метрик при добавлении синтетических изображений демонстрируют архитектуры средней и высокой ёмкости (VGG16, DenseNet, EfficientNet), тогда как для более лёгких моделей применение синтетических данных требует дополнительного контроля специфичности и баланса ошибок. В целом синтетическое расширение следует рассматривать не как универсальную замену реальных данных, а как форму управляемой регуляризации обучающего распределения.

Направления дальнейших исследований

Дальнейшие исследования будут направлены на повышение устойчивости предложенного метода в условиях экстремального дефицита данных, а также на углублённый анализ факторов, влияющих на эффективность автоматизированного обнаружения злокачественных образований. Планируется проведение сравнительных экспериментов между безусловными и условными подходами к генерации изображений, включая cWGAN-GP, ControlNet-подобные архитектуры и text-to-image diffusion-модели, с оценкой их влияния не только на распределительные метрики генерации, но и на качество downstream-классификации на идентичных наборах данных [111].

В дальнейшем планируется апробация и других современных архитектур в критике DGAN-WP-TL и других GAN моделей, например Diffusion, StyleGAN3 и их адаптации к малым медицинским выборкам [112-116].

Особое внимание будет уделено оценке клинической значимости ошибок классификации, анализу взаимосвязи между разнообразием синтетических данных и устойчивостью downstream-моделей, а также интеграции предложенного подхода в более широкие диагностические пайплайны поддержки принятия врачебных решений.

Выводы по четвертому разделу

В данном подразделе сформулированы обобщающие выводы по результатам анализа и интерпретации экспериментальных исследований. Выводы отражают основные закономерности, выявленные в ходе сравнительного и абляционного анализа, а также оценивают влияние предложенного метода генерации синтетических медицинских изображений

DGAN-WP-TL на эффективность downstream-задач классификации обнаружения РМЖ. Особое внимание уделено тому, в какой степени использование синтетических данных способствует повышению качества обнаружения РМЖ с применением CNN в условиях дефицита и дисбаланса обучающих выборок.

1. Анализ и обобщение результатов экспериментальных исследований позволил подтвердить, что предложенный метод DGAN-WP-TL является эффективным инструментом расширения обучающей выборки и повышения качества автоматизированного обнаружения РМЖ с использованием CNN в условиях дефицита и дисбаланса медицинских данных.

2. Абляционное исследование показало, что устойчивое качество генерации достигается при использовании Full критика, обеспечивающей согласованное улучшение распределительных и перцептивных характеристик синтетических изображений, критически важных для последующего обучения CNN-классификаторов. Это подтверждает методологическую обоснованность выбранных архитектурных решений.

3. Экспериментально доказано, что замена функции потерь значительно влияет на качество синтетических изображений. Кроме того, подтверждена необходимость малого батча в малых медицинских наборах.

4. Установлено преимущество DGAN-WP-TL по сравнению со StyleGAN2 в качестве современной архитектуры. Так как StyleGAN2 предназначен для крупных наборов данных и происходит коллапс мод при его обучении. Возникновение коллапса мод не допустимо в медицинском ИИ.

5. В результате классификации датасетов с синтетическим расширением с помощью DGAN-WP-TL была повышена Recall-метрика по обнаружению миноритарного класса. В результате снижен риск пропуска злокачественных новообразований за счет синтетического расширения обучающей выборки.

6. Определено, что пайплайн генерации с классификацией необходимо оценивать комплексно с учетом метрик обеих задач. Недопустимо проводить оценку только по 1-2 метрикам. Кроме того, визуальный анализ требует дополнительной клинической валидации.

7. Проведенный анализ в главе 4 установил, что разработанный метод DGAN-WP-TL улучшает результаты классификации с целью балансировки миноритарного класса и его обнаружения в классификации РМЖ. Полученные экспериментальные и аналитические выводы демонстрируют практическую состоятельность предложенного подхода и его соответствие поставленным целям исследования.

ЗАКЛЮЧЕНИЕ

В данной диссертационной работе решена актуальная научно-практическая задача повышения эффективности автоматизированного обнаружения РМЖ на медицинских изображениях с использованием методов искусственного интеллекта и CNN. Актуальность исследования обусловлена высокой распространённостью РМЖ, критической значимостью ранней диагностики и сохраняющимися ограничениями алгоритмов машинного обучения, связанными с дефицитом и выраженным дисбалансом клинических данных, прежде всего для злокачественного класса.

В первом разделе выполнен аналитический обзор современного состояния проблемы автоматизированного обнаружения РМЖ. Показано, что, несмотря на высокую эффективность современных CNN, их диагностическая надёжность существенно ограничена дефицитом данных и смещением обучающих выборок в сторону benign-класса, что приводит к снижению чувствительности моделей к злокачественным образованиям. Обоснована целесообразность применения генеративных моделей не как самостоятельной цели, а как прикладного инструмента синтетического расширения обучающих выборок и повышения качества downstream-классификации.

Во втором разделе разработан метод автоматизированного обнаружения РМЖ, основанный на интеграции CNN и генеративной модели DGAN-WP-TL. В свою очередь, DGAN-WP-TL состоит из генератора и критика, который в последнюю очередь включает две ветви. Двухветвевой критик объединяет предобученную модель VGG19 для извлечения перцептивных признаков и сверточную часть для извлечения доменно-адаптированных признаков. Кроме того, функция потерь WGAN-GP позволяет адаптировать DGAN-WP-TL к малым выборкам. Разработан общий пайплайн для интеграции синтетического расширения с классификацией с целью управления долей синтетических изображений. Общий пайплайн включает в себя синтетические изображения, классификатор с тонкой настройкой трансферного обучения.

В третьем разделе проведены экспериментальные исследования, направленные на апробацию разработанного метода DGAN-WP-TL. Кроме того, экспериментально доказано эффективность варьирования доли синтетических изображений на этапе классификации. Устойчиво были сгенерированы синтетические изображения трех датасетов, в том числе и изображения заболевания Альцгеймера для валидации метода на междоменном датасете. Эксперименты продемонстрировали эффективность метода DGAN-WP-TL на классификацию. Были повышены метрики Recall, F1-score и другие, которые сигнализируют об улучшении обнаружения именно злокачественных опухолей РМЖ. Доказано, что DGAN-WP-TL более лучше адаптирован к малым выборкам в отличие от StyleGAN2.

В четвертом разделе проведен анализ полученных результатов из раздела 3. Согласно анализу, разработанный метод DGAN-WP-TL, метод обучения классификатора и методика оценки результатов улучшили классификацию миноритарного класса в малых медицинских наборах. При этом общий пайплайн

необходимо оценивать в комплексе ряда метрик, достигая баланса между реалистичностью, разнообразием и вычислительными ресурсами. Однако результаты классификации с синтетическим расширением зависят от датасета, архитектуры классификатора и доли формируемой обучающей выборки синтетическими изображениями. Сформулированы ограничения метода, включая зависимость от качества исходных данных, отсутствие клинической reader-study валидации и риск генеративных артефактов в экстремальных режимах обучения, а также обозначены практические рекомендации направления дальнейших исследований.

Основным научным результатом диссертации является разработка и экспериментальная валидация метода автоматизированного обнаружения РМЖ, в котором генеративная модель используется целенаправленно как инструмент синтетического расширения обучающих данных для повышения чувствительности CNN к злокачественному классу. В отличие от подходов, ориентированных преимущественно на визуальное качество генерации, в работе оценка синтетических изображений напрямую связана с их вкладом в downstream-классификацию и клинически значимые метрики.

Практическая значимость результатов заключается в возможности применения предложенного метода как вспомогательного инструмента в системах компьютерной поддержки принятия врачебных решений при анализе изображений молочной железы в условиях дефицита и дисбаланса клинических данных. Метод не предназначен для замены клинической диагностики, однако может повысить устойчивость и чувствительность автоматизированных систем на предпилотных и исследовательских этапах.

Таким образом, поставленные в диссертационной работе цель и задачи выполнены в полном объёме, а полученные результаты подтверждают целесообразность и эффективность управляемого синтетического расширения данных как механизма повышения чувствительности CNN при автоматизированном обнаружении РМЖ.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Kim J., Harper A., McCormack V. et al. Global patterns and trends in breast cancer incidence and mortality across 185 countries // *Nature Medicine*. – 2025. – Vol. 31, Issue 4. – P. 1154-1162.
- 2 Liao L. Inequality in breast cancer: Global statistics from 2022 to 2050 // *The Breast*. – 2025. – Vol. 79. – P. 103851.
- 3 Bray F., Laversanne M., Sung H. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries // *A Cancer Journal for Clinicians*. – 2024. – Vol. 74, Issue 3. – P. 229-263.
- 4 Xu Y., Gong M., Wang Y. et al. Global trends and forecasts of breast cancer incidence and deaths // *Scientific Data*. – 2023. – Vol. 10, Issue 1. – P. 334.
- 5 Toguzbayeva A., Telmanova Z., Khozhayev A. et al. Impact of screening on breast cancer incidence in Kazakhstan: Results of component analysis // *Asian Pacific Journal of Cancer Prevention*. – 2021. – Vol. 22, Issue 9. – P. 2807-2817.
- 6 Igissinov N., Toguzbayeva A., Khamidullina Z. et al. Epidemiology of breast cancer mortality in Kazakhstan, trends and geographic distribution // *Asian Pacific Journal of Cancer Prevention*. – 2023. – Vol. 24, Issue 10. – P. 3361-3371.
- 7 Dunenova G.A., Kalmataeva Z.A., Kaidarova D.R. et al. Breast cancer epidemiology in Kazakhstan for the period 2012–2021 // *Nauka i Zdravookhranenie*. – 2023. – Vol. 25, Issue 2. – P. 128-137.
- 8 Midlenko A., Mussina K., Zhakhina G. et al. (2023). Prevalence, incidence, and mortality rates of breast cancer in Kazakhstan: Data from the Unified National Electronic Health System, 2014–2019 // *Frontiers in Public Health*. – 2023. – Vol. 11. – P. 1-10.
- 9 Aidarov D.E., Есентаева С.Е., Khaidarov S.Zh. et al. (2025). Epidemiology of breast cancer in Kazakhstan: The analysis of morbidity, mortality, and disease staging in 2015–2024 // *Oncology and Radiology of Kazakhstan*. – 2025. – Vol. 2, Issue 76. – P. 9-16.
- 10 Kassymbekova F., Glushkova N., Dunenova G. et al. Burden of major cancer types in Almaty, Kazakhstan // *Scientific Reports*. – 2024. – Vol. 14, Issue 1. – P. 20536.
- 11 Akhmedullin R., Aimyshev T., Zhakhina G. et al. (2024). In-depth analysis and trends of cancer mortality in Kazakhstan: A joinpoint analysis of nationwide healthcare data 2014–2022 // *BMC Cancer*. – 2024. – Vol. 24, Issue 1. – P. 1340.
- 12 Goodfellow I.J., Pouget-Abadie J., Mirza M. et al. Generative Adversarial Nets // https://proceedings.neurips.cc/paper_files/paper/2014/. 10.09.2025.
- 13 Gulrajani I., Ahmed F., Arjovsky M. et al. Improved Training of Wasserstein GANs // https://proceedings.neurips.cc/paper_files/paper/2017. 10.09.2025.
- 14 Yi X., Walia E., Babyn P. Generative adversarial network in medical imaging: A review // *Medical Image Analysis*. – 2019. – Vol. 58. – P. 101552.
- 15 Wang J., Wang K., Yu Y. et al. Self-improving generative foundation model for synthetic medical image generation and clinical applications // *Nature Medicine*. – 2025. – Vol. 31, Issue 2. – P. 609-617.

- 16 Chen X., Duan Y., Houthoofd R. et al. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets // https://proceedings.neurips.cc/paper_files/paper. 10.10.2025.
- 17 Shin H.-C., Tenenholtz N.A., Rogers J.K. et al. Medical image synthesis for data augmentation and anonymization using generative adversarial networks // Simulation and synthesis in medical imaging: procced. 3rd internat. Workshop, (SASHIMI 2018), Held in Conjunction with (MICCAI 2018). – Granada: Springer, 2018. – P. 1-11.
- 18 McKinney S.M., Sieniek M., Godbole V. et al. International evaluation of an AI system for breast cancer screening // Nature. – 2020. – Vol. 577, Issue 7788. – P. 89-94.
- 19 Yala A., Lehman C., Schuster T. et al. A deep learning mammography-based model for improved breast cancer risk prediction // Radiology. – 2019. – Vol. 292, Issue 1. – P. 60-66.
- 20 Ehteshami Bejnordi B., Veta M., van Diest P.J. et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer // JAMA. – 2017. – Vol. 318, Issue 22. – P. 2199-2210.
- 21 Ehteshami Bejnordi B., Zuidhof G., Balkenhol M. et al. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images // Journal of Medical Imaging. – 2017. – Vol. 4, Issue 4. – P. 044504.
- 22 Janowczyk A., Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases // Journal of Pathology Informatics. – 2016. – Vol. 7, Issue 1. – P. 29.
- 23 Geras K.J., Wolfson S., Kim S. et al. High-resolution breast cancer screening with multi-view deep convolutional neural networks // <https://doi.org/10.48550/arXiv.1703.07047>. 10.09.2025.
- 24 Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks // https://proceedings.neurips.cc/paper_. 10.09.2025.
- 25 Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // <https://arxiv.org/abs/1409.1556>. 10.09.2025.
- 26 Szegedy C., Liu W., Jia Y. et al. Going deeper with convolutions // Procced. IEEE conf. on Computer Vision and Pattern Recognition (CVPR). – Boston, 2015. – P. 1-9.
- 27 He K., Zhang X., Ren S. et al. Deep residual learning for image recognition // Procced. IEEE conf. on Computer Vision and Pattern Recognition (CVPR). – Las Vegas, 2016. – P. 770-778.
- 28 Arjovsky M., Chintala S., Bottou L. Wasserstein GAN // <https://arxiv.org/abs/1701.07875>. 10.08.2025.
- 29 Chen R.J., Lu M.Y., Chen T.Y. et al. Synthetic data in machine learning for medicine and healthcare // Nature Biomedical Engineering. – 2021. – Vol. 5, Issue 6. – P. 493-497.
- 30 Liu Z., Lv Q., Lee C.H. et al. GSDA: Generative adversarial network-based semi-supervised data augmentation for ultrasound image classification // Heliyon. – 2023. – Vol. 9, Issue 9. – P. e19585.

- 31 Li C., Liu X., Li W. et al. U-KAN makes strong backbone for medical image segmentation and generation // <https://arxiv.org/abs/2406.02918>. 10.08.2025.
- 32 Alruily M., Said W., Mostafa A.M. et al. Breast Ultrasound Images Augmentation and Segmentation Using GAN with Identity Block and Modified U-Net 3+ // *Sensors*. – 2023. – Vol. 23, Issue 20. – P. 8599.
- 33 Luo J., Zhang H., Zhuang Y. et al. 2S-BUSGAN: A Novel Generative Adversarial Network for Realistic Breast Ultrasound Image with Corresponding Tumor Contour Based on Small Datasets // *Sensors*. – 2023. – Vol. 23, Issue 20. – P. 8614.
- 34 Jiménez-Gaona Y. BraNet: A mobil application for breast image classification based on deep learning algorithms // *Medical & Biological Engineering & Computing*. – 2024. – Vol. 62, Issue 9. – P. 2737-2756.
- 35 Jiménez-Gaona Y. et al. GAN-based data augmentation to improve breast ultrasound and mammography mass classification // *Biomedical Signal Processing and Control*. – 2024. – Vol. 94. – P. 106255.
- 36 Garrucho L., Kushibar K., Osuala R. et al. High-resolution synthesis of high-density breast mammograms: Application to improved fairness in deep learning based mass detection // *Frontiers in Oncology*. – 2023. – Vol. 12. – P. 1044496.
- 37 Park S., Lee K.H., Ko B. et al. Unsupervised anomaly detection with generative adversarial networks in mammography // *Scientific Reports*. – 2023. – Vol. 13, Issue 1. – P. 2925.
- 38 Oyelade O.N., Ezugwu A.E., Almutairi M.S. et al. A generative adversarial network for synthetization of regions of interest based on digital mammograms // *Scientific Reports*,. – 2022. – Vol. 12, Issue 1. – P. 6166.
- 39 Na I., Kim J., Ko E.S. et al. RadiomicsFill-Mammo: Synthetic mammogram mass manipulation with radiomics features // *Procceed. Medical Image Computing and Computer Assisted Intervention (MICCAI 2024)*. – Cham, 2024. – P. 723-733.
- 40 Prodan M., Paraschiv E., Stanciu A. Applying Deep Learning Methods for Mammography Analysis and Breast Cancer Detection // *Applied Sciences*. – 2023. – Vol. 13, Issue 7. – P. 4272.
- 41 Joseph A.J., Dwivedi P., Joseph J. et al. Prior-guided generative adversarial network for mammogram synthesis // *Biomedical Signal Processing and Control*. – 2024. – Vol. 87. – P. 105456.
- 42 Xing J., Li Z., Wang B. et al. Lesion segmentation in ultrasound using semi-pixel-wise Cycle generative adversarial nets // *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. – 2021. – Vol. 18, Issue 6. – P. 2555-2565.
- 43 Raghu M., Zhang C., Kleinberg J. et al. Transfusion: Understanding transfer learning for medical imaging // https://proceedings.neurips.cc/paper_files. 10.09.2025.
- 44 Mei X., Liu Z., Robson P.M. et al. RadImageNet: An open radiologic deep learning research dataset for effective transfer learning // *Radiology: Artificial Intelligence*. – 2022. – Vol. 4, Issue 5. – P. e210315.
- 45 Frid-Adar M., Diamant I., Klang E. et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification // *Neurocomputing*. – 2018. – Vol. 321. – P. 321-331.

- 46 Yang Q., Yan P., Zhang Y. et al. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss // *IEEE Transactions on Medical Imaging*. – 2018. – Vol. 37, Issue 6. – P. 1348-1357.
- 47 Al-Dhabyani W., Gomaa M., Khaled H. et al. Dataset of breast ultrasound images // *Data in Brief*. – 2020. – Vol. 28. – P. 104863.
- 48 Sawyer-Lee R., Gimenez F., Hoogi A. et al. Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) // <https://doi.org/10.7937/K9/TCIA.2016.7O02S9CY>. 10.09.2025.
- 49 Kumar S., Shastri S. Alzheimer MRI preprocessed dataset // <https://www.kaggle.com/datasets/sachinkumar413/alzheimer-mri-dataset>. 23.04.2023.
- 50 Heusel M., Ramsauer H., Unterthiner T. et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium // https://proceedings.neurips.cc/paper_files/paper/2017/. 23.04.2023.
- 51 Bińkowski M., Sutherland D.J., Arbel M. et al. Demystifying MMD GANs // <https://openreview.net/forum?id=r1lUOzWCW>. 23.04.2023.
- 52 Zhang R., Isola P., Efros A.A. et al. The unreasonable effectiveness of deep features as a perceptual metric // *Proc. IEEE/CVF conf. on Computer Vision and Pattern Recognition (CVPR)*. – Salt Lake City, 2018. – P. 586-595.
- 53 Wang Z., Simoncelli E.P., Bovik A.C. Multiscale structural similarity for image quality assessment // *Proc. 37th Asilomar conf. on Signals, Systems & Computers*. – Pacific Grove, 2003. – P. 1398-1402.
- 54 Jolliffe I.T., Cadima J. Principal component analysis: A review and recent developments // *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. – 2016. – Vol. 374, Issue 2065. – P. 20150202.
- 55 Van der Maaten L., Hinton G.E. Visualizing data using t-SNE // *Journal of Machine Learning Research*. – 2008. – Vol. 9. – P. 2579-2605.
- 56 Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: Data mining, inference, and prediction. – Ed. 2nd. – NY.: Springer, 2009. – 745 p.
- 57 Van Rijsbergen C.J. Information retrieval. – Ed. 2nd. – London: Butterworths, 1979. – 208 p.
- 58 Swets J.A. Measuring the accuracy of diagnostic systems // *Science*. – 1988. – Vol. 240, Issue 4857. – P. 1285-1293.
- 59 Ryspayeva M. Generative adversarial network as data balance and augmentation tool in histopathology of breast cancer // *Proc. IEEE internat. conf. on Smart Information Systems and Technologies (SIST)*. – Astana, 2023. – P. 99-104.
- 60 Ryspayeva M., Nishan A. Enhancing grayscale image synthesis with deep conditional GAN and transfer learning. // *Proc. IEEE AITU: Digital Generation*. – Astana, 2024. – P. 122-127.
- 61 Ryspayeva M., Salykova O. Effect of data balancing methods on MRI Alzheimer's classification // *Proc. IEEE 5th internat. conf. on Smart Information Systems and Technologies (SIST)*. – Astana, 2025. – P. 1-7.
- 62 Lauby-Secretan B., Scoccianti C., Loomis D. et al. Breast-cancer screening – Viewpoint of the IARC Working Group // *New England Journal of Medicine*. – 2015. – Vol. 372, Issue 24. – P. 2353-2358.

- 63 Lehman C.D., Wellman R.D., Buist D.S.M. et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection // JAMA Internal Medicine. – 2015. – Vol. 175, Issue 11. – P. 1828-1837.
- 64 Lee R.S., Gimenez F., Hoogi A. et al. A curated mammography data set for use in computer-aided detection and diagnosis research // Scientific Data. – 2017. – Vol. 4, Issue 1. – P. 170177.
- 65 LeCun Y., Boser B., Denker J.S. et al. Backpropagation applied to handwritten zip code recognition // Neural Computation. – 1989. – Vol. 1, Issue 4. – P. 541-551.
- 66 LeCun Y., Bottou L., Bengio Y. et al. Gradient-based learning applied to document recognition // Proceedings of the IEEE. – 1998. – Vol. 86, Issue 11. – P. 2278-2324.
- 67 Ryspayeva M., Salykova O. Multi-domain synthetic medical image generation and dataset balancing with DGAN-WP-TL // Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. – 2025. – Vol. 13, Issue 1. – P. 2556687.
- 68 Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position // Biological Cybernetics. – 1980. – Vol. 36, Issue 4. – P. 193-202.
- 69 Tan M., Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks // Proceed. of the 36th internat. conf. on Machine Learning. – Long Beach, 2019. – P. 6105-6114.
- 70 Deng J., Dong W., Socher R. et al. ImageNet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – Miami, 2009. – P. 248-255.
- 71 Russakovsky O., Deng J., Su H. et al. ImageNet Large Scale Visual Recognition Challenge // International Journal of Computer Vision. – 2015. – Vol. 115, Issue 3. – P. 211-252.
- 72 Hussain J., Båth M., Ivarsson J. Generative adversarial networks in medical image reconstruction: A systematic literature review // Computers in Biology and Medicine. – 2025. – Vol. 191. – P. 110094.
- 73 Mirza M., Osindero S. Conditional generative adversarial nets // <https://doi.org/10.48550/arXiv.1411.1784>. 10.08.2025.
- 74 Radford A. et al. Unsupervised representation learning with deep convolutional generative adversarial networks // [https://arxiv.org/abs. 10.08.2025](https://arxiv.org/abs.10.08.2025).
- 75 Ryspayeva M., Salykova O. Impact of loss function on synthetic breast ultrasound image generation // Scientific Journal of Astana IT University. – 2025. – Vol. 24. – P. 99-112.
- 76 Karras T., Laine S., Aila T. A style-based generator architecture for generative adversarial networks // Proceed. IEEE/CVF conf. on Computer Vision and Pattern Recognition (CVPR). – Long Beach, 2019. – P. 4396-4405.
- 77 Karras T., Laine S., Aila T. Analyzing and improving the image quality of StyleGAN // Proceed. of the IEEE/CVF conf. on Computer Vision and Pattern Recognition (CVPR). – Seattle, 2020. – P. 8107-8116.

- 78 Karras T., Aittala M., Laine S et al. Alias-free generative adversarial networks // *Procceed. conf. Advances in Neural Information Processing Systems*. – New Orleans, 2021. – P. 852-863.
- 79 Odena A., Olah C., Shlens J. Conditional image synthesis with auxiliary classifier GANs // *Proceedings of Machine Learning Research*. – 2017. – Vol. 70. – P. 2642-2651.
- 80 Berthelot D., Schumm T., Metz L. BEGAN: Boundary equilibrium generative adversarial networks // <https://doi.org/10.48550/arXiv.1703.10102>. 10.10.2025.
- 81 Kim S.T., Lee H., Kim H.G. et al. ICADx: Interpretable computer aided diagnosis of breast masses // *Procceed. conf. "Medical Imaging 2018: Computer-Aided Diagnosis"*. – Houston, 2018. – P. 1057522.
- 82 Shen T., Gou C., Wang F.-Y. et al. Learning from adversarial medical images for X-ray breast mass segmentation // *Computer Methods and Programs in Biomedicine*. – 2019. – Vol. 180. – P. 105012.
- 83 Singh V.K., Rashwan H.A., Romani S. et al. Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network // *Expert Systems with Applications*. – 2020. – Vol. 139. – P. 112855.
- 84 Zhou Y., Wei J., Helvie M.A. et al. Generating high resolution digital mammogram from digitized film mammogram with conditional generative adversarial network // *Procceed. conf. "Medical Imaging 2020: Computer-Aided Diagnosis"*. – Houston, 2020. – P. 1131426.
- 85 Jiang G., Wei J., Xu Y. et al. Synthesis of mammogram from digital breast tomosynthesis using deep convolutional neural network with gradient guided cGANs // *IEEE Transactions on Medical Imaging*. – 2021. – Vol. 40, Issue 8. – P. 2080-2091.
- 86 Lee J., Nishikawa R.M. Simulating breast mammogram using conditional generative adversarial network: Application towards finding mammographically-occult cancer // *Procceed. conf. "Medical Imaging 2020: Computer-Aided Diagnosis"*. – Houston, 2020. – P. 1131418.
- 87 Swathi N., Bobby T.C. Breast cancer segmentation of mammographic images using generative adversarial networks // *Biomedical Sciences Instrumentation*. – 2021. – Vol. 57, Issue 2. – P. 247-255.
- 88 Montoya-del-Angel R. et al. MAM-E: Mammographic synthetic image generation with diffusion models // *Sensors*. – 2024. – Vol. 24, Issue 7. – P. 2076.
- 89 Zhou Y., Wei J., Wu D. et al. Generating full-field digital mammogram from digitized screen-film mammogram for breast cancer screening with high-resolution generative adversarial network // *Frontiers in Oncology*. – 2022. – Vol. 12. – P. 868257.
- 90 Zhou Q., Zuley M., Guo Y. et al. A machine and human reader study on AI diagnosis model safety under attacks of adversarial images // *Nature Communications*. – 2021. – Vol. 12, Issue 1. – P. 7281.
- 91 Al-Dhabyani W. et al. Deep learning approaches for data augmentation and classification of breast masses using ultrasound images // <https://doi.org/10.14569/IJACSA.2019.0100579>. 10.10.2025.
- 92 Han L., Huang Y., Dou H. et al. Semi-supervised segmentation of lesion from breast ultrasound images with attentional generative adversarial network // *Computer Methods and Programs in Biomedicine*. – 2020. – Vol. 189. – P. 105275.

- 93 Barkat L., Freiman M., Azhari H. Image Translation of Breast Ultrasound to Pseudo Anatomical Display by CycleGAN // Bioengineering. – 2023. – Vol. 10, Issue 3. – P. 388.
- 94 Yao Z., Luo T., Dong Y. et al. Virtual elastography ultrasound via generative adversarial network for breast cancer diagnosis // Nature Communications. – 2023. – Vol. 14, Issue 1. – P. 788.
- 95 Fan Y., Jiao J. Mammography data augmentation using ACGAN // Proceed. of the 2nd internat. sympos. on Artificial Intelligence for Medicine Sciences (ISAIMS '21). – Beijing, 2021. – P. 431-437.
- 96 Ohuchi N., Suzuki A., Sobue T. et al. Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): A randomised controlled trial // Lancet. – 2016. – Vol. 387, Issue 10016. – P. 341-348.
- 97 Huang G., Liu Z., van der Maaten L. et al. Densely connected convolutional networks // Proceed. 2017 IEEE conf. on Computer Vision and Pattern Recognition (CVPR). – Honolulu, 2017. – P. 2261-2269.
- 98 Szegedy C., Vanhoucke V., Ioffe S. et al. Rethinking the Inception architecture for computer vision // Proceed. IEEE conf. on Computer Vision and Pattern Recognition (CVPR). – Las Vegas, 2016. – P. 2818-2826.
- 99 Sandler M., Howard A., Zhu M. et al. MobileNetV2: Inverted residuals and linear bottlenecks // Proceed. IEEE/CVF conf. on Computer Vision and Pattern Recognition (CVPR). – Salt Lake City, 2018. – P. 4510-4520.
- 100 Tan M., Le Q.V. EfficientNetV2: Smaller models and faster training // Proceedings of Machine Learning Research. – 2021. – Vol. 139. – P. 10096-10106.
- 101 Goodfellow I. et al. A. Deep learning. – Cambridge, 2016. – 716 p.
- 102 Pan S.J., Yang Q. A survey on transfer learning // IEEE Transactions on Knowledge and Data Engineering. – 2010. – Vol. 22, Issue 10. – P. 1345-1359.
- 103 Yosinski J. et al. How transferable are features in deep neural networks? // https://proceedings.neurips.cc/paper_files/paper.10.10.2025.
- 104 Rumelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors // Nature. – 1986. – Vol. 323, Issue 6088. – P. 533-536.
- 105 Kingma D.P., Ba J. Adam: A method for stochastic optimization // <https://arxiv.org/abs/1412.6980>. 10.10.2025.
- 106 Deshpande I., Zhang Z., Schwing A.G. Generative modeling using the sliced Wasserstein distance // https://openaccess.thecvf.com/content_cvpr_2018. 10.10.2025.
- 107 Hanley J.A. et al. The meaning and use of the area under a receiver operating characteristic (ROC) curve // Radiology. – 1982. – Vol. 143, Issue 1. – P. 29-36.
- 108 Ryspayeva M.K., Salykova S.O. Evaluation of StyleGAN2 and StyleGAN3 for synthetic medical image generation on BUSI and CBIS-DDSM datasets // Bulletin of Shakarim University. – 2025. – Vol. 1, Issue 4(20). – P. 89-96.
- 109 Ryspayeva M., Salykova O. GAN-based medical image generation for rare pathologies using transfer learning and RadImageNet weights // International J of Information and Communication Technologies. – 2025. – Vol. 6, Issue 3. – P. 254-269.
- 110 Karras T. et al. Progressive growing of GANs for improved quality, stability, and variation // <https://openreview.net/forum?id=Hk99zCeAb>. 10.10.2025.

- 111 Wang S. et al. EPG-GAN: Edge-guided progressive growing generative adversarial network for mammogram synthesis // <https://doi.org/10.55092>. 10.10.2025.
- 112 Zhu J.-Y., Park T., Isola P. et al. Unpaired image-to-image translation using cycle-consistent adversarial networks // *Proc. IEEE internat. conf. on Computer Vision (ICCV)*. – Venice, 2017. – P. 2242-2251.
- 113 Rao V.M., Hla M., Moor M. et al. Multimodal generative AI for medical image interpretation // *Nature*. – 2025. – Vol. 639, Issue 8056. – P. 888-896.
- 114 Song Y. et al. Score-based generative modeling through stochastic differential equations // <https://openreview.net/forum?id=PXTIG12RRHS>. 10.10.2025.
- 115 Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models // *Proc. conf. Advances in Neural Information Processing Systems*. – Vancouver, 2020. – P. 6840-6851.
- 116 You G. et al. A cGAN-based tumor segmentation method for breast ultrasound images // *Physics in Medicine & Biology*. – 2023. – Vol. 68, Issue 13. – P. 135002.

ПРИЛОЖЕНИЕ А

Акт внедрения

Утверждаю

Проректор по академическим вопросам
НАО «Костанайский региональный университет
имени Ахмета Байтурсынулы»



Э.Наурызбаева
2026 г.

АКТ

**о внедрении результатов диссертационного исследования
Рыспаевой Марии Куанышевны**

на тему:

*«Искусственный интеллект для обнаружения рака груди с
использованием глубоких сверточных нейронных сетей»*

Настоящим актом подтверждается, что результаты диссертационного исследования Рыспаевой Марии Куанышевны, выполненного по образовательной программе 8D06102 – «Информационные технологии и робототехника», внедрены в учебный процесс факультета машиностроения, энергетики и информационных технологий НАО «Костанайский региональный университет имени Ахмета Байтурсынулы».

В рамках исследования разработана интеллектуальная система автоматизированного анализа медицинских изображений, основанная на модифицированных архитектурах сверточных нейронных сетей (CNN), предназначенная для выявления патологических изменений тканей молочной железы на маммографических и гистологических изображениях.

Научная новизна работы заключается в адаптации и оптимизации архитектур глубоких CNN для задач медицинской диагностики с учетом ограниченных и дисбалансированных датасетов; разработке процедуры комбинированной аугментации и регуляризации, направленной на

снижение переобучения; интеграции механизмов интерпретируемости (Grad-CAM) для визуального обоснования решений модели; формировании алгоритма оценки диагностической надежности на основе комплексного анализа метрик Accuracy, Recall, F1-score и AUC.

Методический вклад диссертационного исследования состоит в разработке учебно-методического комплекса, позволяющего интегрировать современные методы глубокого обучения в образовательный процесс, включая лабораторные модули по предобработке изображений, обучению моделей, настройке гиперпараметров, анализу ошибок классификации и оценке клинически значимых показателей (снижение доли ложных отрицательных результатов).

Полученные программные решения внедрены в учебный процесс в виде электронного учебного издания по дисциплине «Глубокое обучение» (свидетельство № 68063 от 26.02.2026) и используются как практический инструмент при изучении методов искусственного интеллекта.

В бакалавриате 6B06103 «Информационные технологии и робототехника» результаты применяются при освоении методов компьютерного зрения и нейросетевого анализа изображений.

В магистратуре 7M06103 «Информационные технологии и робототехника» разработанная система используется в исследовательских проектах по созданию интеллектуальных диагностических систем и анализу медицинских данных.

В докторантуре 8D06102 «Информационные технологии и робототехника» материалы исследования применяются при разработке и экспериментальной апробации моделей глубокого обучения, включая задачи оптимизации архитектур и повышения интерпретируемости ИИ-систем.

Внедрение результатов диссертационного исследования обеспечивает интеграцию современных технологий искусственного интеллекта в

образовательный процесс, способствует развитию научных компетенций обучающихся и формированию исследовательской базы в области интеллектуального анализа медицинских изображений.

Декан факультета  Б. Калаков

Заведующий кафедрой  О.Бижанова

Научный руководитель
Согласовано:  О.Салыкова

Руководитель управления  Т.Исмаилова
академической политики

Начальник отдела  М.Курмангалиева
образовательных программ

ПРИЛОЖЕНИЕ Б

Авторские свидетельства

КАЗАҚСТАН РЕСПУБЛИКАСЫ

РЕСПУБЛИКА КАЗАХСТАН

СВИДЕТЕЛЬСТВО
О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР
ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ

№ 29428 от «14» октября 2022 года

Фамилия, имя, отчество, (если оно указано в документе, удостоверяющем личность) автора (ов):
РЫСПАЕВА МАРЬЯ КУАНЫШЕВНА

Вид объекта авторского права: **программа для ЭВМ**

Название объекта: **Генерация гистопатологических изображений рака груди с помощью Generative-Adversarial Networks**

Дата создания объекта: **20.02.2022**

Курсы: <http://www.kazpatent.kz/ru> сайт
"Авторлық құқық" бөлімінде тексеруге болады: <https://copyright.kazpatent.kz>
Подлинность документа возможно проверить на сайте [kazpatent.kz](http://www.kazpatent.kz)
в разделе «Авторское право» <https://copyright.kazpatent.kz>

Подписано ЭЦП

Е. Оспанов

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ



РЕСПУБЛИКА КАЗАХСТАН

СВИДЕТЕЛЬСТВО

О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР
ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ

№ 57670 от «6» мая 2025 года

Фамилия, имя, отчество, (если оно указано в документе, удостоверяющем личность) автора (ов):

РЫСПАЕВА МАРЬЯ КУАНЫШЕВНА, САЛЫКОВА ОЛЬГА СЕРГЕЕВНА

Вид объекта авторского права: **программа для ЭВМ**

Название объекта: **Генеративно-сопоставительная сеть Deep Generative Adversarial Network with Wasserstein Gradient Penalty and Transfer Learning (DGAN-WP-TL)**

Дата создания объекта: **20.01.2024**



Документ опубликован на <http://www.kazpatent.kz/kz> сайтын
"Авторлық құқық" бөлімінде тексеруге болады. <https://copyright.kazpatent.kz>

Подлинность документа возможно проверить на сайте kazpatent.kz
в разделе «Авторское право» <https://copyright.kazpatent.kz>

Подписано ЭЦП

С. Ахметов

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ



РЕСПУБЛИКА КАЗАХСТАН

**СВИДЕТЕЛЬСТВО
О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР
ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ**

№ 67033 от «4» февраля 2026 года

Фамилия, имя, отчество, (если оно указано в документе, удостоверяющем личность) автора (ав):
РЫСПАЕВА МАРЬЯ КУАНЫШЕВНА, САЛЫКОВА ОЛЬГА СЕРГЕЕВНА

Вид объекта авторского права: **компьютерные программы (программное обеспечение)**

Название объекта: **Интеллектуальная система генерации и классификации медицинских изображений
рака молочной железы на основе DGAN-WP-TL**

Дата создания объекта: **01.05.2025**



Документ опубликован на <http://www.kazpatent.kz/ru> сайті
"Авторське право" Бюро текстових баз <https://copyright.kazpatent.kz>

Подлинность документа возможно проверить на сайте [kazpatent.kz](http://www.kazpatent.kz)
в разделе «Авторское право» <https://copyright.kazpatent.kz>

Подписано ЭЦП

Л. Утешева

ПРИЛОЖЕНИЕ В

Обзор генеративных моделей и их применения в диагностике рака молочной железы

Таблица В.1 – Эволюция архитектур генеративных моделей и их применение в медицинской визуализации

Архитектура	Авторы, год	Основной принцип	Ключевые особенности	Преимущества	Ограничения	Типичные области применения в медицине
GAN	Goodfellow et al., 2014	Противоборство генератора и дискриминатора	Базовая архитектура; обучение с нуля на реальных данных	Простота, универсальность	Неустойчивое обучение; mode collapse	Первые эксперименты по синтезу рентгеновских и МРТ-изображений
cGAN	Mirza & Osindero, 2014	Генерация изображений под условием меток	Управляемый синтез	Контроль типа патологии	Требует аннотированных данных	Генерация опухолей; сегментация
DCGAN	Radford et al., 2015	Использование свёрточных слоёв	Улучшает стабильность и качество текстур	Реалистичные изображения	Не решает проблему разнообразия	Аугментация данных; синтез КТ и УЗИ
WGAN	Arjovsky et al., 2017	Расстояние Васерштейна	Стабильная сходимость	Снижает mode collapse	Нарушение ограничения Липшица	Синтез редких патологий
WGAN-GP	Gulrajani et al., 2017	Градиентный штраф для стабилизации	Контроль нормы градиента	Устойчивость при малых выборках	Высокая нагрузка	Синтез УЗИ и маммограмм; балансировка
CycleGAN	Zhu et al., 2017	Обучение без парных данных	Перенос изображений между доменами	Перенос без парных данных	Артефакты при сложных морфологиях	Перенос между модальностями (МРТ↔КТ)
StyleGAN	Karras et al., 2019	Введение многомасштабного управления стилем (style-based generator)	Управление текстурой и структурой	Фотореалистичность	Ресурсоёмкость	Генерация фотореалистичных маммограмм
StyleGAN2	Karras et al., 2020	Улучшенное представление стиля и устранение артефактов	Высокая детализация и стабильность	Более реалистичные и стабильные изображения	Высокие требования к вычислительным ресурсам	Синтез изображений кожных заболеваний, маммограмм, патологии печени
StyleGAN3	Karras et al., 2021	Alias-free архитектура без пространственных искажений	Стабильность, отсутствие aliasing	Ещё более плавное и реалистичное движение в видео и 3D	Малая распространённость в медицинских исследованиях	Потенциально применим для 3D-визуализации и динамических сцен
Diffusion Models	Ho et al., 2020	Обратный диффузионный процесс	Итеративное восстановление из шума	Чёткие текстуры	Высокая сложность	Синтез редких случаев; восстановление низкодозовых КТ
Foundation /Multimodal	Wang et al., 2025	Объединение данных разных модальностей	Самообучение и мультидоменность	Интеграция данных	Требует больших ресурсов	Мультимодальная диагностика
Примечание – Составлено по источникам [12; 13; 15, p. 609-616; 28; 73; 74; 76, p. 4396-4404; 77, p. 8107-8115; 78, p. 852-862; 112, p. 2242-2250; 115, p. 6840-6850]						

Таблица В.2 – Генеративные модели и классификационные CNN в маммографии

Автор, год	Архитектура / тип	Тип условности	Датасет	Задача (Task)	Роль GAN и CNN	Основной результат	Ограничения
1	2	3	4	5	6	7	8
Kim et al., 2018	ICADx (interpreter GAN: dual-network)	Condit.	DDSM	Интерпретируемая классификация опухолей (CADx)	Генерация признаков опухолей и обучение CNN-классификатора BI-RADS.	Повышена точность классификации и интерпретируемость решений; GAN обучается соотносить опухоли с уровнями BI-RADS	Отсутствует количественная оценка; небольшой набор данных; без внешней валидации
Shen et al., 2019	cGAN + улучшенный U-Net	Condit.	INbreast, private (Nanfang Hospital, China)	Сегментация / Аугментация	GAN генерирует маски опухолей, CNN (U-Net) улучшает сегментацию при малых данных.	Улучшено качество сегментации U-Net после добавления синтетических данных	Нет сравнения с современными GAN; нет оценки на внешних данных; генерация ограничена ROI (256×256)
Singh et al., 2020	cGAN+CNN (сегментация и классификация формы)	Condit.	INbreast (106), DDSM (1168), private (300)	Сегментация и классификация формы	GAN генерирует контуры и формы опухолей; CNN классифицирует форму и выполняет сегментацию.	Dice = 94 %, IoU = 87 % (INbreast); Accuracy = 80 % (shape classification)	Требуется предварительного выделения ROI; снижает точность при неполных опухолях; малый объем размеченных данных
Zhou et al., 2020	Cycle-HDDM (conditional GAN, U-Net generator + dual-channel discriminator)	Condit.	DFM: 1651, DM: 813 (IRB-approved dataset)	Image-to-image translation / Domain adaptation	Синтез цифровых маммограмм высокой четкости (DM) из оцифрованных плёночных (DFM) для улучшения анализа плотности груди	Синтезированные DM улучшили точность сегментации плотности груди ($p < 0.001$); визуальное качество сохраняет структуру тканей; DNN на синтетических DM показывает более стабильные результаты	Ограничено до 100 μm разрешения; требует парных данных по категориям (breast/background/boundary); высокие вычислительные затраты
Lee & Nishikawa, 2020	cGAN (U-Net generator, left-right breast translation)	Condit.	1366 нормальных маммограмм (train), 333 (test)	Image-to-image translation / диагностика скрытых опухолей (MO)	GAN генерирует контралатеральные снимки; CNN выявляет скрытые опухоли по симметрии.	Обеспечивает CNN более информативные пары левый-правый снимок; AUC классификатора = 0.67 ± 0.09 ; синтез улучшает обнаружение МО-рака	Ограниченный размер тестовой выборки; нет сегментации или локализации опухолей
Swathi & Christy Bobby, 2021	Pix2Pix (cGAN, U-Net Generator + DNN Discriminator)	Condit.	Частный мини-датасет с GT-масками	Сегментация	GAN создаёт маски опухолей; CNN улучшает сегментацию.	Dice = 90 %, Jaccard = 88.38 %; визуально близкие к GT маски, стабильная сходимость GAN-обучения	Малый объем данных; нет кросс-валидации и внешних тестов; нет метрик FID/LPIPS; требуется расширение набора и fine-tuning

Продолжение таблицы В.2

1	2	3	4	5	6	7	8
Jiang et al., 2021	Gradient-Guided cGAN (GGGAN-VGG) + DCNN (U-Net-like, multi-scale, 3D SWGC)	Condit.	1077 пар (Hologic Selenia), тест: 122 с массами, 21 с кальцинатам и	Image-to-image translation / синтез SDM	GAN синтезирует синтетические цифровые маммограммы SDM из DBT; CNN анализирует маски и микрокальцинаты.	PSNR = 26.7, SSIM = 0.71; GGGAN-VGG сохраняет массы и микрокальцинаты с точностью ближе к FFDM; Dice ↑, чувствительность MC 75 % vs 83 % (FFDM); визуально сравнима с коммерческим C-view	Требуется парных DBT/FFDM; высокая вычислительная сложность (80 ч обучения); ограничена к данным Hologic; не оценена переносимость на другие системы
Zhou et al., 2021	Dual-GAN framework (two U-Net-based generators + VGG16 classifier)	Condit.	4346 изображений (1284 женщин, UPMC, США)	Состязательный синтез / оценка устойчивости AI-CAD	Изучение поведения AI-CAD и врачей при атаках GAN, вставляющих или удаляющих опухоли на маммограммах	GAN-изображения вызвали ложный диагноз у AI-CAD в 69.1% случаев; врачи определяли поддельные изображения в 29–71% случаев	Исследован только один AI-CAD; нет внешней валидации; не разработаны методы защиты; обучение GAN вычислительно дорого
Fan & Jiao, 2021	BreastGAN (ACGAN + Self-Attention + Spectral Normalization + Label Smoothing)	Condit.	Public mammogram dataset	Аугментация / Классификация	Генерация размеченных маммограмм по категориям (benign/malignant) с повышением качества и сбалансированности выборки	FID = 21.25; IS = 56.54; точность классификации ↑ на 4.9 % по сравнению с базовым ACGAN; синтетические данные улучшили обучение CNN	Отсутствует внешняя валидация; ограничено разрешение изображений; слабая интерпретируемость при мультиклассовых задачах
Zhou et al., 2022	HRGAN (conditional)	Condit.	CBIS-DDSM, INbreast	Синтез / Аугментация	Генерация полноформатных маммограмм, CNN использует их для аугментации.	Повышает разнообразие обучающих данных для классификаторов / детекторов	Потеря микрокальцинатов
Oyelade et al., 2022	ROImammoGAN)	Condit.	MIAS, DDSM, CBIS-DDSM	Генерация ROI-областей четырёх типов аномалий	Синтез ROI для баланса классов, аугментации; CNN классифицирует тип аномалий.	Модель устойчива к mode collapse, обеспечивает качественные ROI для аугментации.	Низкое разрешение 128×128; отсутствует клиническая оценка реалистичности.
Park et al., 2023	StyleGAN2-based	Uncondit.	DDSM, CBIS-DDSM	Аномалия детекция	GAN создаёт аномальные паттерны; CNN выполняет локализацию.	Повышена чувствительность	Высокие ресурсы

Продолжение таблицы В.2

1	2	3	4	5	6	7	8
Garrucho et al., 2023	CycleGAN (unpaired style-transfer, low→high breast density)	Condit.	OPTIMAM, CSAW, BCDR, INbreast	Low→high density translation / Data augmentation	Синтез изображений для устранения дисбаланса BI-RADS C/D и повышения точности DL-детекторов	Улучшена детекция масс в BI-RADS D; повышена обобщающая способность CNN. Радиологи не смогли надёжно отличить синтетические изображения (AUC = 0.61–0.74).	Возможны артефакты в MLO-виде при высокой плотности; нестабильность CycleGAN на малых данных; результаты зависят от сканера и уровня регистрации.
Prodan et al., 2023	StyleGAN-XL + CNN/ViT (ResNet, EfficientNet, MaxViT) + Grad-CAM	Uncondit.	RSNA Breast Cancer Detection / ADMANI dataset	Аугментация + Классификация + Explainable AI	Генерация 1000 синтетических положительных примеров для балансировки классов и повышения точности классификатора	FID _{50k} = 9.8 (256×256) / 13.66 (512×512); AUC ↑ до 0.85 (ResNet18); Precision ↑ до 0.35 / Recall ↑ до 0.59; Grad-CAM визуализировал решения CNN.	Синтетика только для позитивного класса; низкое разрешение; нет внешней клинической валидации
Na et al., 202	RadiomicsFILL-Mammo (cGAN)	Condit.	INbreast	Контролируемый синтез	GAN управляет структурой опухолей; CNN классифицирует.	Улучшение устойчивости CNN к морфовариациям.	Узкая специализация
Jiménez-Gaona et al., 2024	SNGAN (mammo), cGAN (УЗИ)	Condit. / Uncondit.	CBIS-DDSM, BUSI	Аугментация / Классификация	GAN расширяет малые выборки; CNN классифицирует.	Прирост AUC/чувствительности	Узкая специализация
Jiménez-Gaona et al., 2024	BraNet (с спектральной нормализацией) + SAM + ResNet18 в BraNet	Condit.	BUSI, Dataset A/B, UDIAT, Mini-MIAS, INbreast, CBIS-DDSM, VinDr-Mammo	Аугментация + Сегментация + Классификация (мобильное приложение)	Разработка мобильного приложения для сегментации и классификации 2D-изображений груди с использованием SNGAN для аугментации и классификации	SNGAN FID = 52.4 (маммо), 116.85 (УЗИ); KID = 0.051 / 0.06; ResNet Accuracy = 94.7 % (US benign), 93.6 % (US malignant), 80.9 % / 76.9 % (DM); F1≈0.95/0.92; реалистичность синтетических изображений подтверждена	Ограниченное кол-во классов (BI-RADS 1–3 vs 4–6); низкое разрешение (128×128 px); нет полноценной валидации на клинических данных; приложение в стадии pre-clinical тестирования
Joseph et al., 2024	Prior-guided cGAN (Deep Convolutional GAN c class-conditioning)	Condit.	MIAS (322 изображения)	Аугментация / Классификация	Синтез реалистичных маммограмм по классам (normal, benign, malignant) для решения проблемы дисбаланса данных	Точность классификации повышена на 3.9 %; 90 % сгенерированных изображений сохраняют классовую принадлежность; GAN-аугментация улучшила результаты CNN по MIAS	Малый размер набора (MIAS); отсутствие оценки по FID/KID; не проверена на других датасетах; низкое разрешение (1024×1024 → 256×256)

Продолжение таблицы В.2

1	2	3	4	5	6	7	8
Montoya -del- Angel et al., 2024	Stable Diffusion	Condit. (text-prompt + mask)	OPTIMAM (OMI-H), VinDr- Mammo	Синтез и inpainting маммографий / гене рация поражений (mass-like)	Генерация синтети ческих маммографий и реалистичных пораже ний с управлением по тексту и маске	Радиологическая оценка AUROC = 0.49 (неразличимы от реальных), MS-SSIM = 0.29–0.60 (высокое разнооб разие)	Ограничение разрешения, отсутствие микрокальци фикаций, длительное время генерации
Wang S. & Wang S., 2025	EPG-GAN (Edge- Guided Prog ressive Gro wing GAN)	Condit. (mask + edge sketch)	INbreast (410 снимков, 115 пациентов)	Синтез / Аугментация	Генерация реалистич ных маммограмм с управляемой структурой и текстурой	FID = 22.23; PSNR = 34.89 dB; SSIM = 0.983 (512×512). Реализована редактируемая генерация.	Проверено на INbreast; требуется GPU > 24 GB; нет метрик FID 50k/LPIPS; время генерации высокое.
Примечание – Составлено по источникам [34, p. 2737-2755; 35, p. 106255; 36, p. 1044496; 37, p. 2925; 38, p. 6166; 39, p. 723-732; 40, p. 4272; 41, p. 105456; 81, p. 1057522; 82, p. 105012; 83, p. 112855; 84, p. 1131426; 85, p. 2080-2090; 86, p. 1131418; 87, p. 247-254; 88, p. 2076; 89, p. 868257; 90, p. 7281; 95, p. 431-436; 111]							

Таблица В.3 – Генеративные модели и классификационные CNN в ультразвуковом исследовании

Авторы, год	Архитектура / тип	Тип услов ности	Датасет	Задача (Task)	Роль GAN и CNN	Основной результат	Ограничения
1	2	3	4	5	6	7	8
Al- Dhabyani et al., 2019	DAGAN (WGAN-GP- based) + CNN / Transfer Learning	Uncondit.	BUSI, Dataset B, BUSI + B	Аугмен тация и классифика ция	GAN генерирует снимки для увеличения выборки и уст ранения дисбаланса; CNN классифицирует опухоли.	Повышение accuracy до 99 % (NASNet TL); подтверждено влияние GAN-аугментации	Нет оценки реалистичности; низкое разрешение; высокая вычислительная сложность
Han et al., 2020	Attentional GAN (BUS- GAN, semi- supervised)	Condit.	In-house (2800), Public (163)	Сегмен тация	GAN генерирует маски и контуры опухолей, усиливая сегментирующую способность CNN при недостатке аннотаций.	Dice ↑ до 87.1 %, 79.8 %; улучшено качество границ	Зависимость от качества неаннотированных данных; снижение точности (~8 %)
Xing et al., 2021	Semi-pixel-wise CycleGAN (SPCGAN)	Condit.	In-house (Siemens, Toshiba) – 670 изобра жений	Сегмен тация	GAN генерирует пиксельно- точные BUS-структуры и воспроизводит границы опухолей; CNN выполняет сегментацию.	Dice = 0.92 ± 0.04; устойчивость при малом числе данных	Нет оценки межэкспертного согласия; требуется больше размеченных случаев
Luo et al., 2023 [33, p. 8614]	2S-BUSGAN (двухстадийная GAN)	Condit.	BUSI + 920 Collected	Синтез изоб ражений и контуров опухолей; аугмен тация	GAN генерирует BUS- изображения + маски опу холей; CNN выполняет сег ментацию и классификацию.	FID = 72.0 / 89.1, KID ×100 = 27.7 / 43.6, MS-SSIM = 0.15 / 0.13; Dice = 0.76, Accuracy = 0.85	Высокая вычислительная сложность; обучение занимает много времени

Продолжение таблицы В.3

1	2	3	4	5	6	7	8
Barkat et al., 2023	CycleGAN (unsupervised image translation)	Uncondit./ Unpaired Condit.	BUSI + оптические снимки тканей	Image-to-image translation	GAN переводит снимки в оптически-похожие изображения; CNN получает более структурированную картину.	Улучшена визуальная интерпретируемость; median Dice: 0.91 / 0.70	Использовались синтетические оптические данные; отсутствует оценка врачами
Yao et al., 2023	Conditional GAN + U-Net с двумя дискриминаторами + Color Rebalancing	Condit.	4580 пациентов из 15 мед центров	Image-to-image translation / синтез виртуальной эластографии	GAN синтезирует виртуальные эластографические изображения; CNN использует их для классификации.	SSIM = 0.903; AUC=0.752	Ограничение по выборке (только Азия); высокая вычислительная стоимость
You et al., 2023	cGAN + EfficientUNet (ResNet18 + Attention)	Condit.	Dataset B	Сегментация	GAN усиливает текстуру и улучшает качество снимков; CNN выполняет сегментацию.	Dice = 0.8856, IoU = 0.8111	Нет валидации; высокая сложность модели
Liu et al., 2023	GSDA (SGA + CNN, semi-supervised GAN with TL)	Uncondit. (semi-supervised)	BUSI	Аугментация / Классификация	GAN генерирует BUS-изображения и pseudo-labels; CNN классифицирует 3 класса.	Accuracy ↑ до 97.9 % (VGG backbone); FID = 62.92	Обучение SGA и CNN отдельно; отдельный GAN для класса
Alruily et al., 2023	GAN с нелинейным Identity Block + Modified U-Net 3+	Condit.	BUSI	Аугментация + Сегментация	GAN синтезирует BUS-структуры; CNN выполняет сегментацию.	IS = 14.32, FID = 41.86; Dice = 95.49%, Accuracy = 95.67%	Нет валидации на внешних данных; высокие вычислительные ресурсы
Примечание – Составлено по источникам [30, p. e19585; 32, p. 8599; 42, p. 2555-2564; 91; 92, p. 105275; 93, p. 388; 94, p. 788; 116, p. 135002]							