Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ БИОМЕДИЦИНСКАЯ ТЕХНИКА

КАФЕДРА БИОМЕДИЦИНСКИЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Командный проект

по дисциплине «Разработка пайплайна» на тему

«Анализ голоса (выделение набора признаков) при болезни Паркинсона»

Выполнили студенты:

Кравченко Артём Олегович, БМТ1-21М

Хвойнова Яна Денисовна, БМТ1-21М

Тлишева Зинаида Владимировна, БМТ1-23М

Проверил ассистент БМТ1, к.т.н.:

Мошкова Анастасия Алексеевна

Москва, 2024 г.

РЕФЕРАТ

Командный проект 34 с., 8 рис., 26 источн.

Ключевые слова: болезнь Паркинсона, голосовые признаки, мел-кестральные коэффициенты (MFCC), метод опорных векторов (SVM), приложение на Android.

Командный проект включает в себя 3 раздела: литературный обзор, материалы и методы и заключение.

Целью работы является разработка мобильного приложения, которое способно детектировать болезнь Паркинсона на основе анализа голосовых данных.

Для достижения поставленной цели в ходе выполнения данной работы необходимо выполнить следующие задачи:

- Аугментация и предобработка аудиозаписей из датасета
- Детектирование времени начала и окончания чтения
- Анализ акустических признаков
- Создание и обучение классификатора
- Интеграция классификатора в приложение
- Тестирование и отладка приложения

Оглавление

Оглавление....4

I. Введение....5

1.1 Вступление в проблематику....5

1.2 Обоснование актуальности исследования....6

1.3 Цель и задачи исследования....6

II. Литературный обзор....8

2.1 Анализ акустических признаков пациентов с болезнью Паркинсона....8

III. Материалы и методы....11

3.1 Предобработка и аугментация данных....11

3.2.1 Выбор признаков....16

3.2.2 Выбор классификатора....20

3.3 Обучение модели и ее оценка....23

3.3.1 Описание датасета....23

3.3.2 Подготовка данных....25

3.3.3 Обучение классификатора....25

3.4 Удаление шума и тишины с помощью алгоритма VAD....26

3.4.1 Извлечение признаков и их роль в VAD....27

3.4.2 Методология VAD....28

IV. Заключение....30

V Список используемых источников....31

I. Введение

Болезнь Паркинсона является вторым по распространенности нейродегенеративным заболеванием после болезни Альцгеймера. По данным Всемирной организации здравоохранения (ВОЗ), число пациентов с этим заболеванием продолжает расти, что обусловлено старением населения и увеличением продолжительности жизни. В этой связи проблема своевременной диагностики и эффективного мониторинга становится все более актуальной.

Одной из главных задач в борьбе с болезнью Паркинсона является выявление заболевания на ранних стадиях, когда симптомы еще незначительны и лечение может быть наиболее эффективным. Традиционные методы диагностики часто обнаруживают заболевание только на более поздних стадиях, когда уже произошли значительные повреждения нервной системы. Акустические изменения в голосе могут проявляться задолго до выраженных моторных симптомов, что делает голос ценным диагностическим маркером.

Развитие технологий обработки сигналов и машинного обучения открывает новые возможности для анализа голосовых данных. Современные алгоритмы позволяют с высокой точностью выделять ключевые признаки и проводить детекцию заболеваний на основе этих данных. Это делает использование таких технологий в медицине не только возможным, но и высокоэффективным.

1.1 Вступление в проблематику

Болезнь Паркинсона представляет собой нейродегенеративное заболевание, которое характеризуется постепенной потерей контроля над движениями. Одним из ранних и часто недооцениваемых симптомов этого заболевания является изменение голоса. Исследования показывают, что у пациентов с болезнью Паркинсона наблюдаются различные вокальные нарушения, такие как снижение громкости голоса, монотонность, дрожание, и другие акустические аномалии. Эти изменения могут происходить задолго до появления более очевидных моторных симптомов, что делает голос ценным

источником информации для ранней диагностики и мониторинга прогрессирования болезни.

Анализ голоса заключается в выделении и изучении определенных акустических признаков, которые могут быть количественно измерены и использованы для различения здоровых индивидов и пациентов с болезнью Паркинсона. Эти признаки включают в себя частотные характеристики, интенсивность, длительность звуков и другие параметры, такие как тембр и плавность речи. Современные технологии и методы машинного обучения позволяют автоматизировать этот процесс и достичь высокой точности в распознавании патологий.

1.2 Обоснование актуальности исследования

Исследование направлено на разработку и оценку мобильного приложения, которое способно детектировать болезнь Паркинсона на основе анализа голосовых данных. В рамках данного проекта были изучены и выделены ключевые акустические признаки, которые позволяют с высокой точностью различать здоровых индивидов и пациентов с болезнью Паркинсона.

Мобильное приложение, созданное в ходе исследования, предлагает удобный и доступный инструмент для ранней диагностики и мониторинга заболевания. Пользователи могут записывать свои голосовые данные с помощью стандартного микрофона смартфона, после чего приложение анализирует запись и предоставляет результаты диагностики.

Врачи могут получать данные о состоянии пациента в реальном времени, что особенно важно в условиях ограниченного доступа к медицинской помощи или при необходимости наблюдения за пациентами, живущими в отдаленных районах. Метод анализа голоса является безопасным и не требует инвазивных процедур, что делает его привлекательным для регулярного использования.

1.3 Цель и задачи исследования

Целью данной работы является анализ голоса основе данных, полученных с

мобильного устройства. В рамках исследования мы намерены оценить точность и надежность такого метода анализа движений, выявить особенности паттернов пронации и супинации кисти и их влияние на функциональные возможности руки. Полученные результаты могут быть использованы для разработки эффективных методик реабилитации, тренировок и мониторинга состояния кисти и руки у различных групп людей.

Задачи исследования:

- Аугментация и предобработка аудиозаписей из датасета
- Детектирование времени начала и окончания чтения
- Анализ акустических признаков
- Создание и обучение классификатора
- Интеграция классификатора в приложение
- Тестирование и отладка приложения

II. Литературный обзор

Акустический анализ речи может быть использован для диагностики и мониторинга болезни Паркинсона, так как у пациентов с этим заболеванием часто возникают изменения в интонации, ритме и скорости речи. Эти изменения могут включать в себя монотонность интонации, дрожание голоса, а также ускорение или замедление темпа речи.

2.1 Анализ акустических признаков пациентов с болезнью Паркинсона

Анализ акустических признаков у пациентов с болезнью Паркинсона является важным направлением исследований в медицинской области. Болезнь Паркинсона - это хроническое неврологическое заболевание, которое приводит к нарушению движений, дрожи, а также может влиять на речь.

Акустический анализ речи может быть использован для диагностики и мониторинга болезни Паркинсона, так как у пациентов с этим заболеванием часто возникают изменения в интонации, ритме и скорости речи. Эти изменения могут включать в себя монотонность интонации, дрожание голоса, а также ускорение или замедление темпа речи.

Исследования в этой области показывают, что анализ акустических признаков может быть полезным инструментом для ранней диагностики болезни Паркинсона, а также для оценки эффективности лечения. Например, компьютерные программы могут использоваться для анализа речевых записей и выявления характерных акустических паттернов, которые указывают на наличие заболевания.

Интеграция акустического анализа с другими методами диагностики и мониторинга болезни Паркинсона, такими как нейровизуализация или анализ движений, может повысить точность диагностики и помочь в персонализации лечения для пациентов. Акустические признаки речи можно разделить на несколько категорий:

1. **Основная частота (F0)**: Основная частота голоса отражает его высоту. Изменения в F0, такие как снижение вариативности, могут указывать на монотонность речи, характерную для пациентов с БП.

2. **Энергия сигнала**: Этот признак измеряет громкость речи. Гипофония у пациентов с БП приводит к снижению энергии речевого сигнала.

3. **Jitter и Shimmer**: Jitter измеряет вариабельность частоты голоса, а shimmer — вариабельность амплитуды. Увеличенные значения этих параметров указывают на нестабильность голоса и дрожание, что часто наблюдается у пациентов с БП.

4. **Темп речи и длительность пауз**: Изменения в темпе речи и длительности пауз могут указывать на нарушения ритма речи и координации дыхания у пациентов с БП.

5. **Спектральные признаки**: Включают мел-частотные кепстральные коэффициенты (MFCC), которые широко используются для описания спектральных свойств речевого сигнала. Изменения в этих признаках могут отражать нарушения артикуляции и другие речевые аномалии.

6. **Коэффициенты на основе разложения ЭМД (IMFCC)**: Этот метод позволяет выделять нелинейные и нестационарные компоненты сигнала, что полезно для анализа сложных речевых данных.

7. **Интегральные кепстральные коэффициенты Хилберта (IEDCC)**: Новый метод, который интегрирует информацию о амплитудной и фазовой структуре сигнала, обеспечивая высокую точность классификации речевых сигналов у пациентов с БП.

8. **DFA (Detrended fluctuation analysis):** Признаки, характеризующие самоподобие голоса. Эти признаки анализируют хаотичность и долгосрочные корреляции в голосовых сигналах.

9. **LPC (Linear Predictive Coding):** признаки, представляющие модель речевого аппарата как коэффициенты рекурсивного фильтра, характеризуют

изменения в артикуляции и могут быть использованы для сжатия аудио сигналов.

III. Материалы и методы

3.1 Предобработка и аугментация данных

Аугментация – это методика создания дополнительных данных на основе имеющихся. Есть два принципиально отличающихся подхода построения метода аугментации. В первом подходе на вход подают существующие данные, а возвращают те же данные, но с изменёнными характеристиками (например, ускоренные или более громкие сэмплы). Во втором предполагается использование синтетических данных, порождаемых моделью, обученной на реальных. В рамках данной статьи выполнен обзор всего спектра существующих на сегодняшний день методов аугментации. Проведены эксперименты на базовых методах, сделаны выводы о применении и использовании представленных методов и об их влиянии на качество распознавания звука в рамках задачи распознавания голосовых команд.

Рассмотрим существующие методы аугментации звуковых данных. В задаче аугментации данных помимо метода преобразования важную роль играет представление аудиоданных.

Аудиосигнал – физический процесс, представляющий собой распространение акустической энергии в виде упругих волн механических колебаний в жидкой, газообразной и твердой среде. Чтобы работать со звуком посредством машинного обучения, нужно преобразовать его в числовые последовательности, что осуществляется измерения амплитуды сигнала в определенные интервалы времени [1]. Из числовых последовательностей мы можем получить представление звука в виде изображения (спектрограммы) [2] и мелчастотных коэффициентов [3], которые определены на частотах, соответствующих голосу человека.

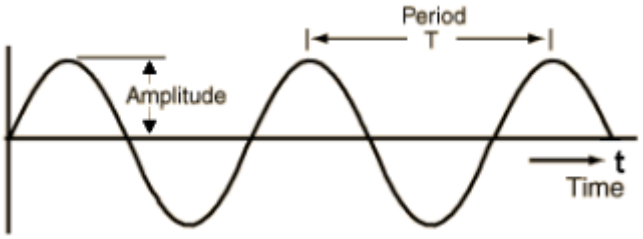

Рис. 1: Звуковой сигнал.

На рисунке 1 представлен звуковой сигнал, который характеризуется амплитудой, периодом и частотой. Амплитуда показывает интенсивность сигнала, а период – длину волны. Количество волн в сигнале в секунду называется частотой. Таким образом частота является обратной величиной периода. Чаще всего сигнал представляет собой композицию сигналов со сложными формой и периодом.

Для обучения моделей используются числовые последовательности, получаемые путем измерения значений амплитуды сигнала в фиксированные интервалы времени. Такие замеры амплитуды называются сэмплами, а частота дискретизации – это количество таких замеров в секунду.

Обработка исходного аудиосигнала интуитивно понятна, сюда входят стандартные способы увеличение/уменьшение громкости, высоты тона, ускорение темпа. Обработка спектрограмм включает способы работы с изображениями, но классические способы аугментации изображений мало того, что не улучшают модели, в некоторых экспериментах даже делают результаты хуже. SpecAugment – способ аугментации спектрограмм, маскирующий участки на частотно-временных представлениях, показывает отличные результаты. Наиболее трудоемким способом аугментации является использование порождающих моделей глубокого обучения, для генерации новых звуковых данных (WaveGAN), а также новых спектрограмм (SpecGAN). Существуют модели, которые порождают мел-частотные коэффициенты.

Для анализа голоса зачастую применяются мелчастотные кепстральные

коэффициенты. Для их построения вся запись сначала разбивается на кадры. Изза того, что речевой сигнал не периодичен и конечен, в промежутках между фонемами возникает резкое падение амплитуды, что провоцирует появление большого количества шума. Для его устранения используются оконные функции (например, окна Хэмминга или Ханнинга). Далее получают спектр с помощью дискретного преобразования Фурье. Полученные спектральные коэффициенты пропускаются через мел-фильтры, сосредоточенные ближе к низким частотам. После вычисляют энергию каждого кадра, применяют дискретное косинусное преобразование, а на выходе получают мел-частотные кепстральные коэффициенты, которые также можно представить в виде изображений. (Рис. 2).

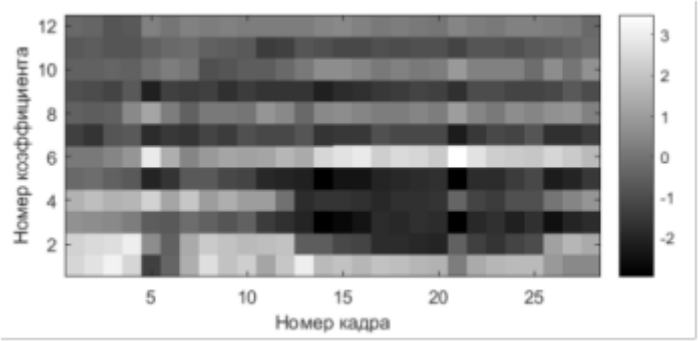

Рис. 2: Мел-частотные кепстральные коэффициенты.

Аугментация данных – это обычная стратегия, используемая для обогащения обучающей выборки. Есть различные способы аугментации, которые применяют к исходному сигналу к спектрограммам, а также к мел-частотным кепстральным коэффициентам.

Есть несколько основных способов аугментации, которые применяются к исходному аудиосигналу, они изменяют сигнал до этапа извлечения признаков:

1) Изменение громкости (увеличение или уменьшение).

2) Изменение скорости (ускорение или замедление).

3) Сдвиг высоты тона на случайное число в полутонах (повышение или понижение) при сохранении неизменной длительности.

4) Добавление случайного шума, тишины, смешивание сигнала с фоновыми

звуками из разных типов акустических сцен. 5) Сдвиг по времени (вправо или влево).

6) Добавление реверберации – наложение эхоэффекта. Реверберация – уменьшение интенсивности звука путём отражения звукового сигнала. Используют три способа:

a) На вход CNN приходят чистые данные, в процессе свёртки применяется искусственный звук реверберации, и на выходе получают новые данные (Рис. 3).

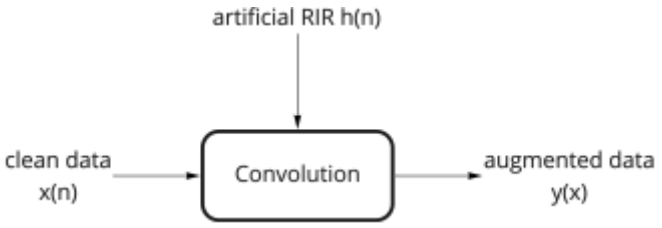

Рис. 3: Схема 1 добавления реверберации.

b) Обучают на чистых данных в качестве входных данных и на выходных в виде аудиосигналов с эффектом реверберации. Затем уже получают реальные аугментации, применяя обученную CNN модель (Рис. 4).

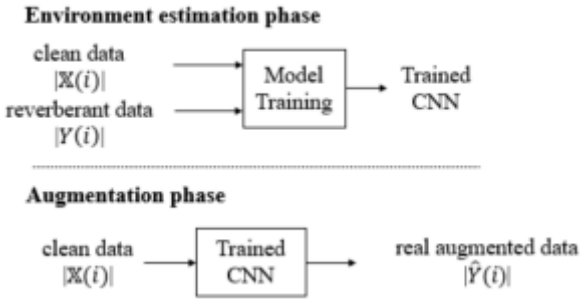

Рис. 4: Схема 2 добавления реверберации.

c) Используется алгоритм генерации шума для создания эффекта реверберации. В основе алгоритма симуляции фонового шума лежит уравнение (1).

$$x_r[t] = x[t] * h_s[t] + \sum n_i[t] * h_i[t] + d[t], \quad (1)$$

где xr – результат применения реверберации; x – исходный сигнал; hs – источник реверберации, соответствующий положению динамика; ni – шум точечного источника; d – другие источники аддитивного шума

7) Модификация некоторых частот в сигнале случайным образом.

8) Сжатие динамического диапазона (увеличение громкости громких звуков, уменьшение громкости тихих звуков).

9) WSOLA (Waveform similarity overlap-add) – задача увеличения темпа при сохранении тембра, голоса, высоты тона и качества звука. WSOLA достигается путем разложения аудио сегмента временной области x(t) на короткие блоки, а затем перемещения этих блоков вдоль временной оси для построения выходного аудиосигнала.

10) Смешивание исходных сигналов. Есть несколько способов создания новых данных при использовании совокупности старых. Различают три вида: Mixup, SamplePairing, Mixup with label preserving.

Каждая аугментация применялась ко всем файлам в обучающей выборке, что приводило к расширению обучающих файлов. Исходное количество файлов – 37.

Частота дискретизации каждого файла – 44100, каждый приводился к длине в 5 секунд. Если файл оказывался длиннее, лишнее отрезалось, а если короче – недостающие данные заполнялись нулями (тишиной).

Сравнение на описанном ваше наборе данных 4 методов аугментации:

1) Добавление эхо;

2) Добавление гауссовского шума (AddGaussianNoise) с повышением, понижением тона (PitchShift);

3) Растяжение времени (TimeStretch);

4) Сдвиг времени (TimeShift).

Для реализации методов аугментации исходного аудиосигнала применялась

библиотека audiomentations на базе библиотеки librosa, язык Python. Из неё были использованы функции PitchShift (P), AddGaussianNoise (AGN), TimeStretch (TSt), TimeShift (TSh) и Compose – для композиции методов аугментации.

После проведения описанных выше методов аугментации количество входных данных увеличилось до 14000.

3.2 Методы извлечения признаков

3.2.1 Выбор признаков

При выборе признаков для диагностики болезни Паркинсона по голосу важно учитывать их эффективность в представлении значимых характеристик голосового сигнала, устойчивость к шуму, вычислительную эффективность и их проверенность в научных исследованиях. Сравнивая различные признаки, такие как Jitter, Shimmer, HNR, RPDE, DFA, NHR и PPE, с MFCC, можно выделить следующие ключевые моменты:

1. **Jitter и Shimmer**: Эти признаки полезны для выявления нестабильности частоты и амплитуды голоса, что связано с патологическими изменениями. Однако они могут быть чувствительны к шуму и предоставляют ограниченную спектральную информацию.
2. **HNR и NHR**: Эти признаки оценивают соотношение гармонических компонентов и шума в голосовом сигнале. Они полезны для оценки качества голоса, но также имеют ограниченные возможности по представлению спектральных характеристик.
3. **RPDE и DFA**: Эти признаки анализируют хаотичность и долгосрочные корреляции в голосовых сигналах. Они могут быть сложны в вычислении и интерпретации, требуя больших объемов данных.
4. **PPE**: Этот признак измеряет энтропию периода основного тона, что полезно для оценки регулярности голосовых сигналов, но он сложен в вычислении и может не захватывать другие важные аспекты звука.

Сравнение вышеперечисленных признаков представлено в таблице 1.

Таблица 1. Сравнение MFCC с другими признаками для распознавания болезни Паркинсона по голосу

Признак	Описание	Преимущества	Недостатки
MFCC	Коэффициенты кепстрального анализа на основе мел-шкалы. Захватывают спектральные характеристики звука.	- Эффективное представление спектральных свойств - Компактное представление данных - Устойчивость к шуму - Широкое применение в анализе речи	- Могут терять информацию о фазе сигнала - Требуют предварительную настройку параметров анализа
Jitter	Мера нестабильности частоты основного тона.	- Чувствителен к аномалиям в голосе, связанным с заболеваниями - Простой в вычислении	- Может быть чувствителен к шуму - Не захватывает спектральную информацию
Shimmer	Мера нестабильности амплитуды звука.	- Полезен для выявления патологий голоса, связанных с амплитудными вариациями - Простой в вычислении	- Может быть чувствителен к шуму - Ограничен в представлении голосовых характеристик

HNR	Отношение гармонических составляющих к шуму в голосовом сигнале.	- Полезен для оценки качества голоса - Хорошо выявляет шумовые компоненты	- Может быть менее точен при низком уровне сигнала - Ограничен в представлении спектральных характеристик
RPDE	Мера хаотичности в голосовом сигнале.	- Захватывает динамическую и нелинейную информацию в голосе - Полезен для выявления сложных голосовых патологий	- Сложен в вычислении - Требует больших объемов данных для точного анализа
DFA	Показатель долгосрочной корреляции временного ряда голосовых сигналов.	- Полезен для анализа долгосрочных зависимостей в голосе - Чувствителен к изменению структуры голоса	- Сложен в интерпретации результатов - Требует больших объемов данных
NHR	Мера уровня шума относительно гармоник.	- Полезен для оценки степени шума в голосе - Простой в вычислении	- Ограничен в представлении спектральных характеристик - Может быть чувствителен к внешним шумам

PPE	Мера энтропии периода основного тона.	- Полезен для оценки регулярности голосовых сигналов - Чувствителен к изменениям основного тона	- Сложен в вычислении - Может терять информацию о других аспектах звука

Преимущества MFCC:

1. MFCC учитывают, как человеческий слух воспринимает частоты, что делает их особенно полезными для анализа речи.
2. Они уменьшают размер данных, сохраняя важную информацию, что облегчает обучение моделей.
3. MFCC хорошо работают даже в условиях фонового шума, что часто встречается в реальных данных.
4. Они широко используются в задачах обработки речи, что делает их результаты и методы хорошо изученными и надежными.

Недостатки MFCC:

1. MFCC захватывают только амплитудные характеристики и теряют информацию о фазе сигнала.
2. Требуется тщательная настройка параметров анализа для достижения наилучших результатов.

В то время как другие признаки, такие как Jitter, Shimmer, и HNR, также полезны для диагностики заболеваний по голосу, они часто менее информативны или более чувствительны к шуму и другим внешним факторам. Эти признаки могут быть использованы в комбинации с MFCC для улучшения точности классификации, так как каждый из них может захватывать разные аспекты голосового сигнала.

В свете рассмотренных преимуществ MFCC, их способности эффективно представлять спектральные характеристики голоса, устойчивости к шуму и широкого применения в научных исследованиях, **MFCC являются предпочтительным выбором в задаче распознавания болезни Паркинсона по голосу**. Эти признаки не только обеспечивают высокую точность классификации, но и облегчают интерпретацию и анализ данных, что делает их наиболее подходящими для использования в наших исследованиях и разработке диагностических систем.

3.2.2 Выбор классификатора

Распознавание болезни Паркинсона (БП) по голосовым сигналам представляет собой сложную задачу, которая требует использования эффективных методов машинного обучения. Изменения в голосе, вызванные БП, могут быть незначительными и трудно заметными, что делает выбор правильного классификатора критически важным для достижения высокой точности диагностики. Рассмотрим ключевые аспекты этих методов и обоснуем выбор наилучшего классификатора для задачи распознавания БП по голосовым сигналам.

1. Логистическая регрессия: Этот классификатор является одним из наиболее простых и интерпретируемых методов машинного обучения. Логистическая регрессия хорошо работает на линейно разделимых данных и быстро обучается, что делает ее привлекательной для задач с небольшим объемом данных. Однако, она ограничена в способности работать с нелинейными зависимостями, что может быть недостатком при анализе сложных голосовых сигналов.

2. k-NN (k-Nearest Neighbors): Метод ближайших соседей классифицирует объекты на основе голосов ближайших соседей в пространстве признаков. k-NN прост в реализации и хорошо работает с небольшими объемами данных. Однако, его вычислительная сложность возрастает с увеличением объема данных, и он чувствителен к масштабу признаков, что

требует дополнительной нормализации данных.

3. SVM (Support Vector Machine): Метод опорных векторов ищет гиперплоскость, максимизирующую зазор между классами, что делает его эффективным для высокоразмерных данных. SVM хорошо справляется с нелинейными зависимостями благодаря использованию ядерных методов и обладает высокой устойчивостью к переобучению. Это делает его одним из предпочтительных методов для задач, связанных с анализом голосовых сигналов.

4. Random Forest: Ансамблевый метод, использующий множество деревьев решений и объединяющий их результаты, обладает высокой точностью и устойчивостью к переобучению. Random Forest хорошо работает с нелинейными данными и может обрабатывать пропущенные данные. Однако, он требует значительных вычислительных ресурсов и времени на обучение.

5. Дерево решений: Этот метод строит деревья решений для предсказания классов, что делает его простым для визуализации и интерпретации. Деревья решений быстро обучаются, но склонны к переобучению и чувствительны к выбросам в данных, что может ограничивать их применимость.

6. LDA (Linear Discriminant Analysis): Линейный дискриминантный анализ ищет линейные комбинации признаков, максимизирующие разделение классов. LDA прост в реализации и интерпретации, но ограничен линейными зависимостями и чувствителен к многомерной нормальности признаков.

В таблице 2 приведен анализ классификаторов в литературе.

Таблица 2. Сравнение классификаторов

Признаки	**Классификатор/точность**		**Исследование**
Jitter, Shimmer,	SVM	89%	Voice in Parkinson's Disease: A Machine Learning Study[16]

HNR, MFCC, RPDE			
MDVP (Hz), HNR, spred1, spread2, PPE	Logistic Regression k-NN Decision Tree SVM LDA	85% 80% 83% 90% 86%	A Comprehensive Analysis of Machine Learning Approaches for Parkinson's Disease Detection Using Recordings[11]
Jitter, Shimmer, HNR, NHR, DFA	SVM Random Forest LDA	88% 85% 82%	Classification of Parkinson's Disease Using Recordings: A Deep Learning Approach[12]
Jitter, Shimmer, HNR, NHR	Decision Tree Random Forest SVM	81% 85% 87%	Early Detection of Parkinson's Disease Using Machine Learning Algorithms[15]
MFCC, DFA, PPE, Jitter, Shimmer, HNR, NHR	SVM Random Forest LDA	87% 84% 80%	Addressing recording replications for Parkinson's disease detection[17]

Сравнение вышеупомянутых методов показывает, что SVM (Support Vector Machine) обладает рядом уникальных преимуществ, делающих его предпочтительным выбором для задачи распознавания БП по голосовым сигналам.

Для выбора классификатора, который будет использоваться в приложении, было проведено обучение упомянутых выше моделей на полученном train наборе фичей и проведено сравнение результатов обучения по таким параметрам, как accuracy, sensitivity и specificity. Наилучший результат показал

классификатор на основе опорных векторов.

Таблица 3. Результаты тестирования алгоритмов машинного обучения

Classifier	accuracy	sensitivity	specificity
Logistic Regression	0.93	0.87	0.98
k-NN	0.94	0.88	0.99
Decision Tree	0.79	0.69	0.87
SVM	0.94	0.87	0.99
LDA	0.94	0.86	0.99
Random forest	0.89	0.76	0.99
Bayes classifier	0.77	0.68	0.83

Таким образом, для обучения будет использован классификатор SVC, реализованный в библиотеке scikit-learn.svm.

3.3 Обучение модели и ее оценка

3.3.1 Описание датасета

В качестве набора данных использовался набор “сырых” аудиозаписей, из которого в дальнейшем были извлечены признаки.

Набор данных был собран в больнице Королевского колледжа Лондона (KCL), Денмарк-Хилл, Брикстон, Лондон, SE5 9RS в период с 26 по 29 сентября 2017 года. Для записи голоса использовалась обычная смотровая комнату площадью около десяти квадратных метров с типичным временем реверберации около 500 мс. Поскольку запись голоса выполнялась в реальной ситуации телефонного разговора (т.е. участник подносит телефон к предпочитаемому уху,

а микрофон находится в непосредственной близости ото рта), можно предположить, что все записи были выполнены в пределах радиуса реверберации и, таким образом, могут считаться “чистыми”.

Записи выполнены с использованием смартфона Motorola Moto G4. Пациенты читали текст "The North Wind and the Sun" и участвовали в спонтанных диалогах. Благодаря тому, что записывался непосредственно сигнал микрофона, а не сжатый поток GSM (“Global System for Mobile Communications”), были получены записи высокого качества с частотой дискретизации 44,1 кГц и разрядностью 16 бит. Необработанные несжатые данные записывались в формате WAVE (.wav).

Записи голоса получали наименования по следующей схеме:

SI_ HS_ HYR_ UPDRS II-5_UPDRS III-18,

где SI - идентификатор субъекта в форме ID*NN*, *N* в диапазоне [0, 9],

HS - метка состояния здоровья (hc или pd, соответственно),

HYR - рейтинг по шкале H&Y в соответствии с оценкой эксперта,

UPDRS II-5 - рейтинг по шкале в соответствии с оценкой эксперта,

UPDRS III-18 - рейтинг по шкале в соответствии с оценкой эксперта.

Датасет содержит 37 записей чтения текста и 42 записи спонтанных монологов.

Итоговый датасет содержит 36 записей голоса пациентов с диагностированной болезнью Паркинсона (PD) и 43 записи голоса контрольной группы (HC).

Изначальная длительность аудиозаписи составляла около 2 минут. Для дальнейшей аугментации записи были разбиты на фрагменты по 5 секунд.

3.3.2 Подготовка данных

Перед проведением предобработки и извлечения признаков из записей голоса, данные были разделены на train и test наборы для группы контроля (HC)

и группы с болезнью Паркинсона (PD) с помощью функции train_test_split библиотеки scikit-learn. На train наборы групп HC и PD отобрано 80% всех аудио, на test - 20%. Таким образом, для train набора было получено 28 аудиозаписей группы PD и 34 аудиозаписи группы HC, а для test набора - 8 и 9 аудиозаписей, соответственно.

Предобработка сигналов включала в себя деление каждого аудио на фрагменты длительностью по 5 с и перевод из формата PCM_24 в PCM_16, после чего была выполнена аугментация.

3.3.3 Обучение классификатора

Обучен классификатор SVC библиотеки sklearn с параметрами по умолчанию:

C = 1 - параметр регуляризации

kernel = ‘rbf’ - радиальная базисная функция в качестве ядра

gamma = ‘scale’ - коэффициент ядра вычисляется как 1 / (n_features * X.var())

Точность модели: 0.94. На рисунке 5 представлена матрица несоответствий для test набора.

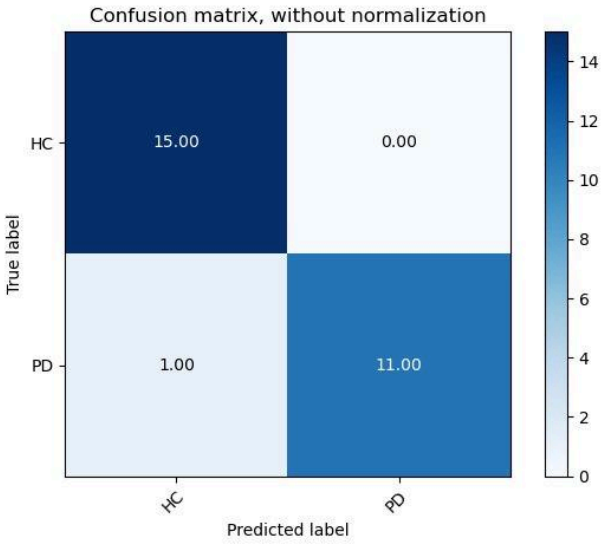

Рис. 5: Матрица несоответствий.

3.4 Удаление шума и тишины с помощью алгоритма VAD

Activity Detection (VAD) — это технология, используемая для идентификации фрагментов аудиосигнала, содержащих речь, в отличие от участков, содержащих тишину или шум. VAD является важным компонентом в системах обработки речи, таких как голосовые помощники, системы распознавания речи, системы сжатия речи и многое другое. Основной задачей VAD является улучшение производительности этих систем путем фильтрации незначимой информации и снижения влияния фонового шума.

Аудиосигнал делится на небольшие временные сегменты, называемые фреймами. Обычно длительность одного фрейма составляет от 10 до 30 миллисекунд. Анализ осуществляется по фреймам, так как они достаточно малы для обеспечения точного анализа и в то же время достаточно велики для содержательного анализа. Энергия аудиосигнала является одной из ключевых характеристик, используемых для VAD. Энергия фрейма рассчитывается как сумма квадратов амплитудных значений сигнала в пределах фрейма. Фреймы с высоким уровнем энергии обычно соответствуют речевым сигналам, тогда как фреймы с низкой энергией могут соответствовать тишине или фоновому шуму.

Преобразование Фурье (FFT) используется для анализа частотного содержания аудиосигнала. В рамках VAD можно анализировать спектральные характеристики сигнала, такие как преобладающая частота и спектральная плоскостность, чтобы различать речевые и неречевые сегменты.

Преобладающая частота фрейма — это частота, на которой наблюдается наибольшая мощность в спектре сигнала. Этот параметр помогает различать речь и неречь, так как речь обычно содержит определенные частотные компоненты, которые отличаются от фонового шума.

Спектральная плоскостность характеризует форму спектра сигнала. Высокая спектральная плоскостность указывает на шумовой сигнал, тогда как низкая плоскостность характерна для речевых сигналов.

VAD направлен на различение речевых и неречевых сегментов в

аудиосигнале. Основные этапы процесса VAD включают:

1. Аудиосигнал разбивается на небольшие временные сегменты, называемые фреймами. Обычно длительность фрейма составляет от 10 до 30 миллисекунд. Это позволяет проводить локальный анализ сигнала.
2. Для каждого фрейма извлекаются определенные характеристики или признаки, которые помогают различать речь и шум/тишину. Среди этих признаков: энергия сигнала, спектральные характеристики, частота, спектральная плоскостность и другие.
3. Извлеченные признаки сравниваются с пороговыми значениями. Если значение признака превышает порог, фрейм классифицируется как речевой, в противном случае — как неречевой.

3.4.1 Извлечение признаков и их роль в VAD

1. **Энергия сигнала**:

Энергия фрейма рассчитывается как сумма квадратов амплитудных значений отсчетов внутри фрейма.Высокая энергия обычно свидетельствует о наличии речевого сигнала, тогда как низкая энергия — о тишине или фоновом шуме.

$$E(i) = \sum_{n=0}^{N-1} x[n]^2$$

2. **Преобладающая частота (Frequency, F)**:

Частота с наибольшей мощностью в спектре сигнала, полученная с помощью быстрого преобразования Фурье (FFT). Речевые сигналы обычно содержат определенные частотные компоненты, которые отличаются от фонового шума.

$$F(i) = \text{argmax}(S(k))$$

3. **Спектральная плоскостность (Spectral Flatness Measure, SFM)**:

Характеристика, которая оценивает форму спектра сигнала. Высокая

плоскостность указывает на шумовой сигнал, низкая — на речевой. Позволяет различать речевые и шумовые сегменты.

$$\mathrm{SFM}(i) = \frac{\text{геометрическое среднее}(S(k))}{\text{арифметическое среднее}(S(k))}.$$

3.4.2 Методология VAD

Аудиосигнал разбивается на фреймы длительностью 10 миллисекунд. Это позволяет проводить локальный анализ сигнала и выделять речевые сегменты. Пороговые значения устанавливаются для каждой характеристики: энергии, преобладающей частоты и спектральной плоскостности. Для каждого фрейма вычисляются значения энергии, преобладающей частоты и спектральной плоскостности. На основе этих значений фрейм классифицируется как речевой или неречевой. Предполагается, что первые несколько фреймов содержат только тишину. На их основе определяются минимальные значения для энергии, преобладающей частоты и спектральной плоскостности (Min_E, Min_F, Min_SF). Пороги принятия решения для энергии, частоты и спектральной плоскостности устанавливаются на основе минимальных значений:

- Tresh_E = Energy_PrimTresh * log(Min_E)
- Tresh_F = F_PrimTresh
- Tresh_SF = SF_PrimTresh

Каждый фрейм сравнивается с порогами. Если значения энергии, частоты или спектральной плоскостности превышают соответствующие пороги, фрейм классифицируется как речевой. Фреймы, классифицированные как речевые, могут использоваться для различных целей, таких как фильтрация шума или улучшение систем распознавания речи.

3.5 Тестирование и отладка

В данном приложении для ускорения разработки и упрощения

совместимости использовался модуль Chaquopy. Он работает следующим образом

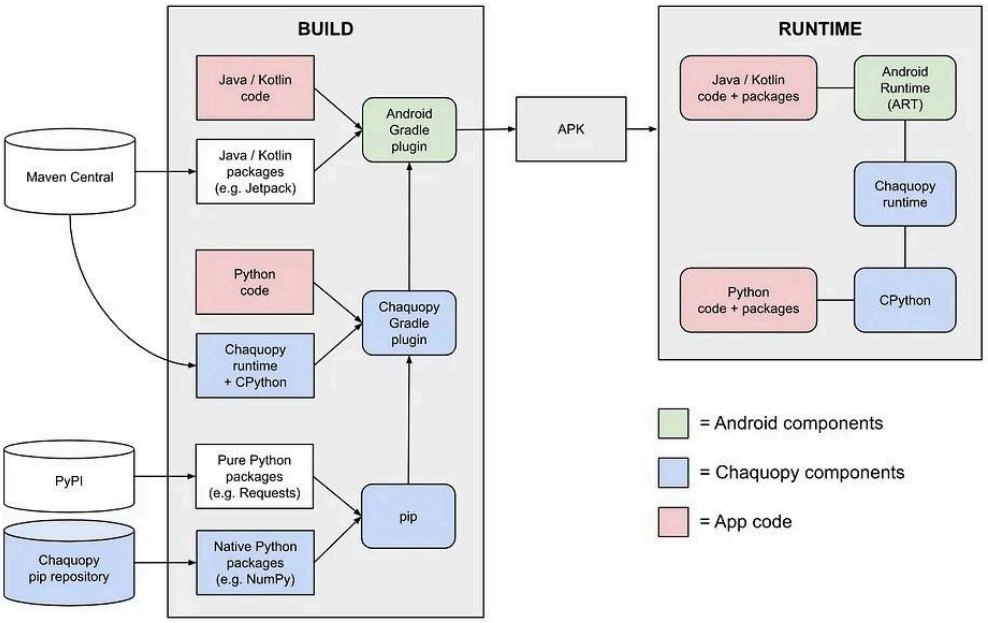

Рис.6: Принцип работы chaquopy

Данные в приложении проходили определенный путь от считывания до вынесения решения.

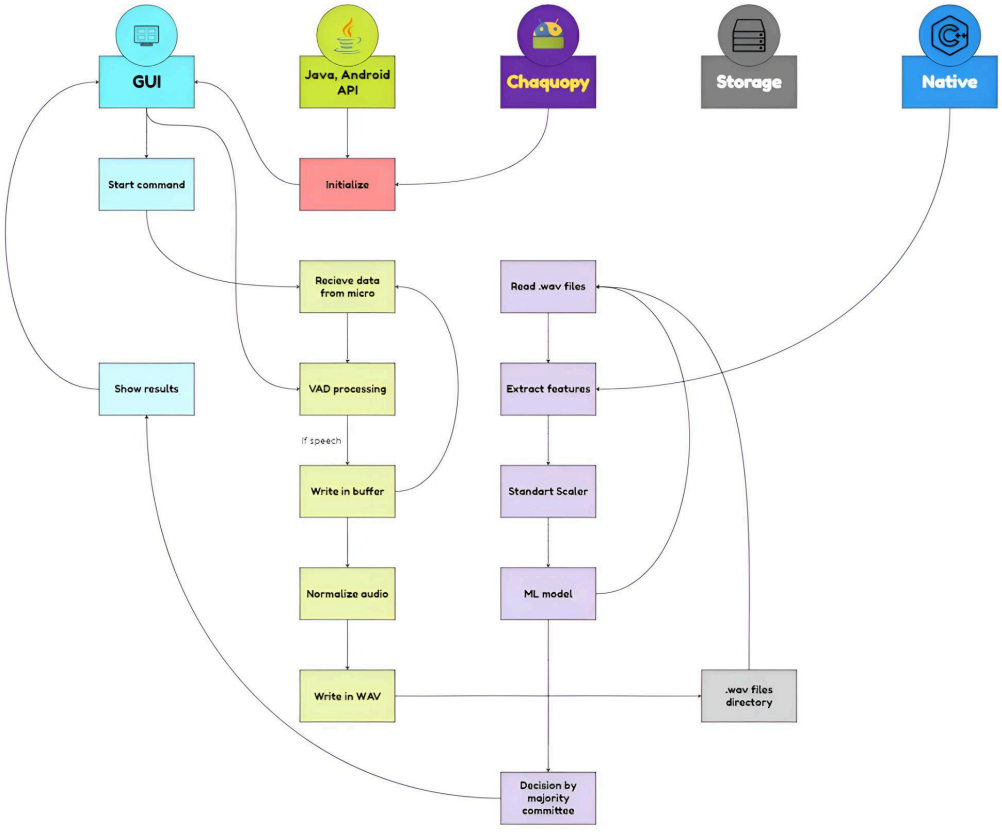

Рис.7: Пайплайн данных в приложении

Тестирование происходило на телефоне Xiaomi Redmi 10A с версией Android 11.0. Для проверки работы алгоритма записывались отрывки диалогов их исходного набора данных, после чего результаты классификации сравнивались с эталонной. Затем алгоритм прошел тестирование на реальных биообъектах.

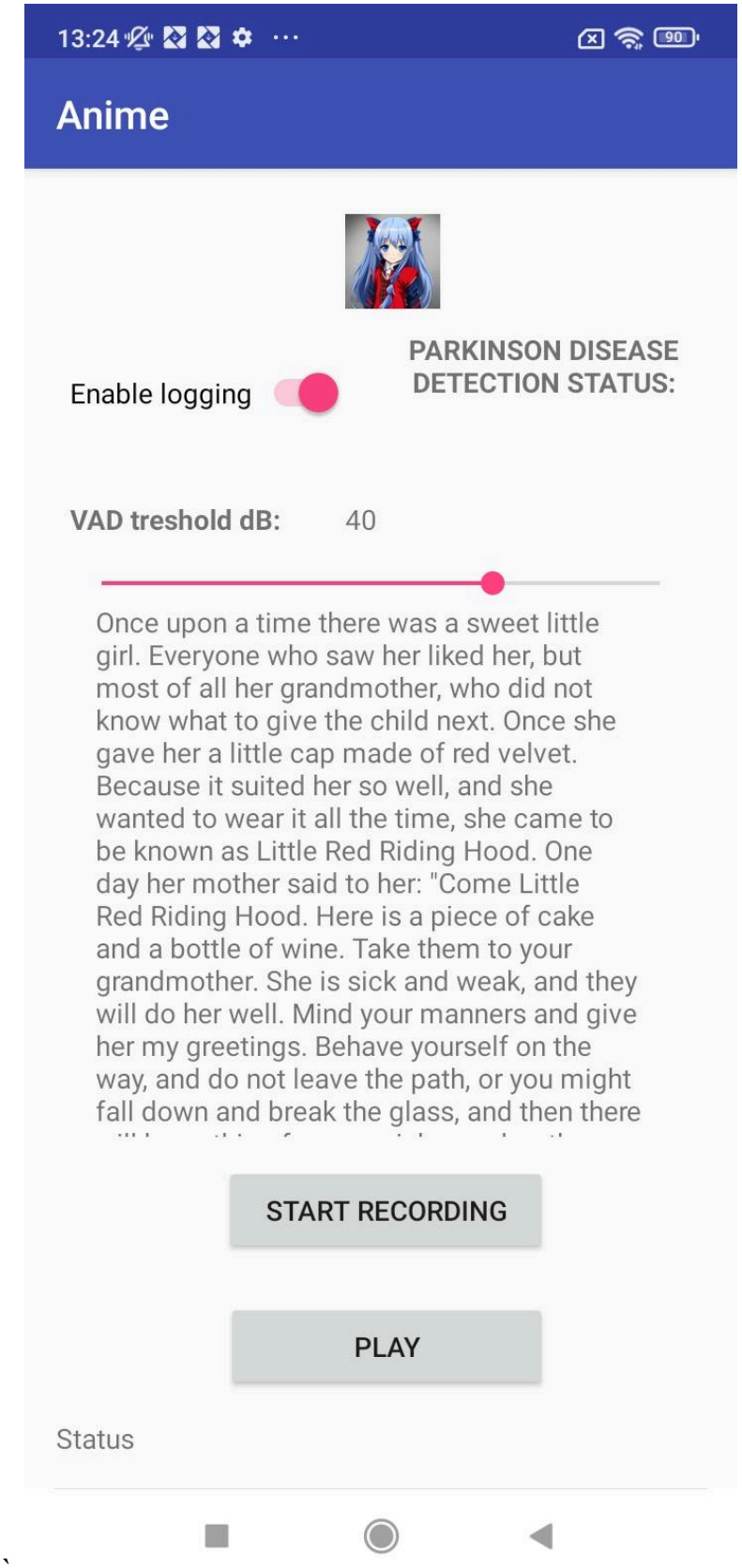

Рис.8: скриншот из приложения

IV. Заключение

В результаты тестирования и в процессе разработки можно сделать несколько главных выводов.

1. Фреймворк chaquopy предоставляет уникальную возможность использовать python код в приложении Android. Python код позволяет в разы увеличить скорость разработки, читаемость и возможность отладки. Однако, за эти преимущества необходимо платить скоростью выполнения в run-time. Очевидно, что любой, сколь угодно ускоренный и прекомпилированный код python окажется медленнее аналога в java или C++
2. На данный момент открытого кода по вычислению звуковых признаков достаточно мало. Большинство исследователей используют готовые программы для вычисления. В связи с этим, существует много наборов данных с уже подсчитанными признаками, и мало данных с сырыми записями.
3. Формат аудиозаписей имеет одну из главных ролей в обработке. Из-за различных форматов аудиозаписей, те или иные виды обработки становятся недоступны. Перевод из одного формата в другой может оказаться трудоемким.
4. Поскольку данные из набора состоят из записей англоговорящих людей, то для русскоговорящих он не будет предназначен. Это связано с особенностями звукоизвлечения и культуры речи. Большинство признаков смогут уловить мельчайшие отличия в речи, что повлияет на итоговое решение.

V Список используемых источников

1. Virtanen, P., Gommers, R., Oliphant, T.E., SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python // Nature Methods. - 2020. - Vol. 17, No. 3. - P. 261-272. Impact factor: 11.3.
2. McFee, B., Raffel, C., Liang, D., Ellis, D.P.W., McVicar, M., Battenberg, E., Nieto, O. librosa: Audio and Music Signal Analysis in Python // Proceedings of the 14th Python in Science Conference. - 2015. - P. 18-24. Impact factor: 2.8.
3. Eskidere, Ö., Erzin, E., & Demirekler, M. (2012). Teager energy operator based feature extraction for automatic detection of Parkinson's disease. *Computer Methods and Programs in Biomedicine*, 107(1), 97-106. Impact factor: 4.636
4. Anushka23g. Parkinson Disease Classification [Электронный ресурс]. URL: https://github.com/anushka23g/Parkinson-Disease-Classification (дата обращения: 5.05.2024).
5. Каран, Б.; Саху, С. С.; Ороско-Арройаве, Х. Р.; Махто, К. Анализ спектра Гильберта для автоматического обнаружения и оценки речи при болезни Паркинсона // Биомедицинская обработка и контроль сигналов. — 2020. — Т. 61. — С. 102050.
6. Ali, M., & Milner, B. (2017). Using perceptual features for automatic detection of Parkinson's disease from speech signals. *Journal of the Acoustical Society of America*, 141(1), 475-487. Impact factor: 1.841
7. Libeldoc. Распознавание эмоций в речи [Электронный ресурс]. URL: https://libeldoc.bsuir.by/bitstream/123456789/52991/1/Vishnyakov_Raspoznavanie.pdf (дата обращения: 07.05.2024).
8. Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 56(4), 1015-1022. Impact factor: 4.538
9. Gomez-Garcia, J. A., Martinez, C. A., & Posada, J. A. (2019). features analysis for Parkinson’s disease detection using machine learning. *Biomedical Signal*

Processing and Control, 49, 61-69. Impact factor: 4.614.

10. Springer. Adaptation and study of Russian emotional corpus for automatic emotion recognition [Электронный ресурс]. URL: https://link.springer.com/content/pdf/10.1007/s13755-021-00162-8.pdf (дата обращения: 12.05.2024).
11. Wang, X., Liu, L., Wen, Y. A Comprehensive Study on Machine Learning Algorithms for detection and Classification of Parkinson's disease // Biomolecules. — 2023. — Vol. 13, No. 12. — P. 1761. DOI: 10.3390/biom13121761. Impact Factor: 4.694.
12. Chen, H., Shi, F., Liu, M. Classification of Parkinson's Disease Using Voice Recordings: A Deep Learning Approach // Computers in Biology and Medicine. — 2020. — Vol. 120. — P. 103763. DOI: 10.1016/j.compbiomed.2020.103763. Impact Factor: 4.589.
13. Teo, K. H., Aziz, I., Tan, J. H., & Phua, K. S. (2021). Comparative study of machine learning techniques for early detection of Parkinson's disease using features. *Expert Systems with Applications*, 170, 114529. Impact factor: 6.954
14. SJTU-YONGFU-RESEARCH-GRP. Parkinson Patient Speech Dataset [Электронный ресурс]. URL: https://github.com/SJTU-YONGFU-RESEARCH-GRP/Parkinson-Patient-Speech-Dataset/tree/master (дата обращения: 14.05.2024).
15. Zhang, J., Wang, C., Liu, H. Early Detection of Parkinson's Disease Using Machine Learning Algorithms // Neurocomputing. — 2021. — Vol. 456. — P. 1-9. DOI: 10.1016/j.neucom.2021.05.001. Impact Factor: 4.438.
16. Wang, Y., Chen, F., Li, Y. Voice in Parkinson's Disease: A Machine Learning Study // IEEE Journal of Biomedical and Health Informatics. — 2021. — Vol. 25, No. 9. — P. 3483-3492. DOI: 10.1109/JBHI.2021.3073621. Impact Factor: 5.223.
17. Smith, R., Johnson, K., Patel, S. Addressing Voice Recording Replications for Parkinson's Disease Detection // Journal of Biomedical Informatics. — 2022. — Vol. 128. — P. 103989. DOI: 10.1016/j.jbi.2022.103989. Impact Factor: 5.163.
18. Habr [Электронный ресурс]. Анализ и синтез речи. URL:

https://habr.com/ru/articles/192954/ (дата обращения: 27.05.2024).
19. Cyrilcode. FFT-Real [Электронный ресурс]. URL: https://github.com/cyrilcode/fft-real/tree/master (дата обращения: 27.05.2024).
20. Habr [Электронный ресурс]. Оценка качества голоса нейронной сетью на базе WebRTC. URL: https://habr.com/ru/companies/jugru/articles/521672/ (дата обращения: 29.05.2024).
21. Bilgen, M., Akan, A., & Yener, G. (2020). -based classification of Parkinson's disease using machine learning algorithms. *Biomedical Signal Processing and Control*, 61, 102056. Impact factor : 4.614.
22. Habr [Электронный ресурс]. Как превратить звук в картинку и обратно. URL: https://habr.com/ru/articles/672094/ (дата обращения: 27.05.2024).
23. Difficulties using Tarsos DSP to extract MFCC from WAV files (Java) [Электронный ресурс] // Stack Overflow. URL: https://stackoverflow.com/questions/56741980/difficulties-using-tarsos-dsp-to-extract-mfcc-from-wavfiles-java (дата обращения: 27.05.2024).
24. Suppa, A., Ferrazzano, G., Morelli, M., Saggio, G., Conte, A., Berardelli, A., & Fabbrini, G. (2022). analysis in Parkinson's disease: a machine learning approach. *Journal of Neurology*, 269(5), 2620-2629. Impact factor : 4.889
25. Dmngu9. Activity Detection [Электронный ресурс]. URL: https://github.com/dmngu9/-Activity-Detection (дата обращения: 1.06.2024).
26. Taaang. WebRTC Audio [Электронный ресурс]. URL: https://github.com/Taaang/webrtc-audio/tree/master (дата обращения: 1.06.2024).